1

MicroD

National Library
of Canada

Bibliothèque nationale
du Canada

Canadian Theses Service

Service des thèses canadiennes

Ottawa, Canada
K1A 0N4

NOTICE

The quality of this microform is heavily dependent upon the quality of the original thesis submitted for microfilming. Every effort has been made to ensure the highest quality of reproduction possible.

If pages are missing, contact the university which granted the degree.

Some pages may have indistinct print especially if the original pages were typed with a poor typewriter ribbon or if the university sent us an inferior photocopy.

Previously copyrighted materials (journal articles, published tests, etc.) are not filmed.

Reproduction in full or in part of this microform is governed by the Canadian Copyright Act, R.S.C. 1970, c. C-30.

AVIS

La qualité de cette microforme dépend grandement de la qualité de la thèse soumise au microfilmage. Nous avons tout fait pour assurer une qualité supérieure de reproduction.

S'il manque des pages, veuillez communiquer avec l'université qui a conféré le grade.

La qualité d'impression de certaines pages peut laisser à désirer, surtout si les pages originales ont été dactylographiées à l'aide d'un ruban usé ou si l'université nous a fait parvenir une photocopie de qualité inférieure.

Les documents qui font déjà l'objet d'un droit d'auteur (articles de revue, tests publiés, etc.) ne sont pas microfilmés.

La reproduction, même partielle, de cette microforme est soumise à la Loi canadienne sur le droit d'auteur, SRC 1970, c. C-30.

ON THE COMPUTATION OF SIZE AND POWER FOR TOXICOLOGICAL BIOASSAYS

by

MIKELIS G. BICKIS, B.Sc., M.Sc.

A thesis submitted to
the Faculty of Graduate Studies and Research
in partial fulfilment of
the requirements for the degree of
Doctor of Philosophy

Department of Mathematics and Statistics

Carleton University
Ottawa, Ontario
June 7, 1988
© copyright
1988, Mikelis G. Bickis

Permission has been granted to the National Library of Canada to microfilm this thesis and to lend or sell copies of the film.

The author (copyright owner) has reserved other publication rights, and neither the thesis nor extensive extracts from it may be printed or otherwise reproduced without his/her written permission.

L'autorisation a été accordée à la Bibliothèque nationale du Canada de microfilmer cette thèse et de prêter ou de vendre des exemplaires du film.

L'auteur (titulaire du droit d'auteur) se réserve les autres droits de publication; ni la thèse ni de longs extraits de celle-ci ne doivent être imprimés ou autrement reproduits sans son autorisation écrite.

ISBN 0-315-46231-0

Thesis Acceptance Form
Doctoral Candidate

The undersigned hereby recommend to
The Faculty of Graduate Studies and Research
acceptance of the thesis,

"On the Computation of Size and Power for
Toxicological Bioassays"

submitted by

Mikelis G. Bickis

in partial fulfilment of the requirements
for the degree of Doctor of Philosophy

M. J. Moore
.....
Chairman, Department of
Mathematics & Statistics

J. N. Rao *P. Krenowski*
.....
Thesis Supervisor

J. Lawless
.....
External Examiner

Carleton University

June 27, 1988

Abstract

Operating characteristics are important considerations in the design and interpretation of toxicological bioassays. After a review of the role of statistics in toxicological research, the efficiencies of various bioassay designs involving intralitter correlation are compared by developing asymptotic methods that take into account randomization. The moments of a natural test statistic are calculated, and asymptotic normality is proved in the null case.

A class of discrete distributions that includes as special cases distributions commonly occurring in small-sample procedures is presented. The probabilities under these distributions can be computed using simple recursive relations, and bounds can be placed on probabilities in the tails. A linear-time algorithm for calculating distributions in this class is presented.

This class of distributions extends to multivariate situations. Computation in the multivariate case depends on defining a tree structure on the sample space. Several tree structures are defined, and their interrelationships are studied.

These distributions arise naturally in permutation tests with discrete data. A general treatment of such procedures is presented using the concept of a probability algebra invariant under a group of permutations. The family of conditional distributions induced by the permutation group is shown to have, in certain cases, a stochastic monotonicity property that allows for the development of efficient computational algorithms utilizing tree structures.

Finally, a data base of actual bioassays is studied empirically, considering four different decision rules. Methods are proposed for estimating false positive and false negative rates from the empirical distribution of responses over all the bioassays in the data base. These methods are used to compare the operating characteristics of the four decision rules.

Acknowledgments

I want to express gratitude to my former employer, the Health Protection Branch of Health and Welfare Canada, under whose aegis this investigation was begun, for appreciating the value of this kind of research; and to my present employer, the University of Saskatchewan, for encouraging me to complete this work. I want to thank Dr. J. N. K. Rao, my teacher and supervisor, for his advice, his patience, and his indulgence. I want to thank Dr. Dan Krewski, who has been, at various times, my employer, supervisor, colleague and friend, for his tremendous encouragement and help in finishing this task. And I express my appreciation to all the relatives, friends, and colleagues who have sustained me during this work. I especially owe a debt to the women in my life, Jane, Wendy, and Heidi, who have had to make do with a part-time husband and father while this thesis was in progress.

It is fitting that I should thank two people who had nothing to do with this dissertation: Dr. Michael Stephens, my first teacher of statistics, who showed me that one could spend an interesting life studying statistics; and Dr. Sudish Ghurye, who showed me the way when I was a confused graduate student. Without their encouragement, I would never have even thought to attempt such work.

Some of the expenses of this research were supported by grant number A8664 to Dr. Krewski from the Natural Sciences and Engineering Research Council, and by a grant from the President's NSERC fund at the University of Saskatchewan.

Contents

1	Introduction	1
1.1	Apology	1
1.2	Summary	2
2	An Overview of the Role of Statistics in Toxicity Bioassays	3
2.1	Study Initiation	3
2.2	Experimental Design	5
2.3	Statistical Analysis	9
2.4	Low Dose Risk Estimation	11
2.5	The Ultimate Issue	12
3	Asymptotic Theory for Comparing Two Treatments with Correlated Binary Responses	14
3.1	Introduction	14
3.2	Statistical Model	14
3.2.1	A correlated independent competing risks model	14
3.2.2	Three randomization schemes	17
3.2.3	Test statistic	18
3.3	Moments of the Test Statistic	18
3.3.1	Expected value	18
3.3.2	Variance—null case	19
3.3.3	Variance—alternative case	22
3.4	Asymptotic Distribution of the Test Statistic	27
3.4.1	Random standardization	27
3.4.2	Constant standardization	30
3.4.3	Sample-based standardization	33
3.4.4	Loose ends	35
4	A Class of Discrete Distributions	36
4.1	The Pólya Family	37
4.1.1	Definition	37
4.1.2	The natural parameter space	37
4.2	Properties of Pólya Distributions	42
4.2.1	Recursive relations	42
4.2.2	Shape of distribution	44

4.2.3	Truncation of tails	52
4.3	A Computational Algorithm	53
4.3.1	Factoring the probability function	53
4.3.2	Bounds on partial products	54
4.3.3	The algorithm	56
4.3.4	Complexity and error analysis	58
4.4	The Multivariate Pólya Distribution	59
4.4.1	Definition	59
4.4.2	Marginal and conditional distributions	60
4.4.3	Geometry of the support	63
4.5	Summary	63
5	Tree Structures on Discrete Multivariate Distributions	64
5.1	Configurations and Trees	65
5.1.1	Configurations	65
5.1.2	Trees	66
5.1.3	Trees of configurations	68
5.2	Recursive Calculations of Probabilities	72
5.2.1	Weighing offspring	72
5.2.2	Weighing complete branches	73
5.3	Probabilities of Critical Regions	77
5.3.1	Linear test statistics	77
5.3.2	Another tree structure	78
5.3.3	Case with some capacities null	84
5.3.4	Chains of branches	86
5.3.5	Calculating probabilities	88
5.4	Finding Critical Configurations	92
5.4.1	Semi-compact configurations	92
5.4.2	Lateral trees	93
5.4.3	Inverting the lateral tree	98
5.5	Grafted Trees	100
5.6	Summary	102
6	Permutation Tests with Binary Data	103
6.1	The Fisher-Irwin Test	103
6.1.1	Motivation	103
6.1.2	Power calculations	104
6.1.3	Stochastic monotonicity	105
6.1.4	Computing algorithm	107
6.2	Permutation Tests	109
6.2.1	Probability algebras	109
6.2.2	Permutation groups	110
6.2.3	Permutations of probability algebras	113
6.2.4	Lattices and products of algebras	115
6.2.5	Reduction of the conditional distribution	117

6.3	The Permutation Distribution under Three Designs	122
6.3.1	Completely randomized design	123
6.3.2	Nested design	124
6.3.3	Litter matched design	125
6.4	A Model For Correlated Binary Data	127
6.4.1	The beta binomial model	127
6.4.2	Independent competing risks	128
6.5	Calculation of Power	130
6.5.1	Completely randomized design	131
6.5.2	Nested design	138
6.5.3	Litter matched design	148
6.6	Summary	152
7	An Empirical Evaluation of Decision Rules	153
7.1	The Data	153
7.2	Analysis of the Data	155
7.2.1	Tests for trend	155
7.2.2	Incorporating historical information	157
7.2.3	Decision rules	158
7.2.4	Results	159
7.3	Estimation of False Positive Rates	161
7.3.1	Adjusting for multiplicity	161
7.3.2	A general formulation	163
7.3.3	Two phase sampling	168
7.3.4	The estimators	169
7.3.5	Variance of the estimator	170
7.3.6	Estimating the variance	176
7.4	Estimating the False Negative Rate	180
7.4.1	Abstraction	180
7.4.2	Distribution of P -values	183
7.4.3	Empirical distribution of P -values	185
7.5	Error Rates for the Decision Rules	188
A	Componentwise and Tensor Products of Matrices	190
B	Fourth Order Mixed Moments	192
C	Computational Considerations	195
C.1	Computational complexity	195
C.2	Round-off errors and underflows	196

List of Tables

7.1	Nominal significance levels for Rule III.	159
7.2	Number of studies in each classification using the four decision rules.	159
7.3	Croestabulation of classifications using the four decision rules.	160
7.4	Croestabulation of classifications using Rule II with and without historical controls.	161
7.5	Estimated error rates for four decision rules.	189

List of Figures

4.1	The natural parameter space for the Pólya family in the a - b plane when $n = 5$	40
4.2	The natural parameter space of the Pólya family in the μ - ρ plane when $n = 5$	43
4.3	Parameter regions in the a - b plane defining various shapes of the distribution	46
4.4	Parameter regions in the μ - ρ plane defining various shapes of the distribution	47
4.5	Region satisfying $(1 - \rho)\mu(1 - \mu) < \delta/n^2$	48
5.1	A tree of configurations	79
5.2	A steptree of configurations	81
5.3	An example of a lateral tree	94
6.1	Conditional tree with $k = 2$, $n_1 = 3$ and margin $(1,1,3)$	138
6.2	Test tree with $k = 2$ and $n_1 = 3$	139
6.3	Conditional tree with $k = 2$, $n_1 = 3$ and margin $(1,2,2)$	139
6.4	Conditional tree with $k = 2$, $n_1 = 3$ and margin $(1,3,1)$	140
6.5	Conditional tree with $k = 2$, $n_1 = 3$ and margin $(2,0,3)$	140

Chapter 1

Introduction

1.1 Apology

The title of this dissertation requires some explanation, since we present very few computations, and only the first and last chapters are explicitly concerned with toxicological bioassays. Indeed, the major part of this work could be considered parenthetical, dealing with some rather technical mathematical minutiae, yet bracketed on either side by the practical concerns that gave them birth.

The research presented here arose from the need to know the exact operating characteristics of common procedures in toxicological research, particularly those used to assess the health risks due to chemicals. Since toxicological data often involve low response rates, calculation of size and power using asymptotic arguments is suspect, and the need for exact small-sample computations becomes apparent. Moreover, the small sample distributions, though simple in theory, become computationally burdensome for even quite moderate samples. While attempting to answer some rather practical questions, we became entangled in trees, groups, Horvitz-Thompson estimators, and empirical processes. The consequences of these entanglements are found in the following pages.

Much of the work is concerned with permutation distributions. Many algorithms have been published for calculating various probabilities resulting from permutation tests. Some of these use rather brute force approaches, whereas others are quite sophisticated. It was our intention to add to this list, but first there appeared to be a need for a comprehensive theory. Many of the distributions appearing in toxicological data analysis can be considered to be members of a common family of probability distributions, all having properties that make them attractive candidates for recursive calculations. Recursive calculation requires a well-ordering of the sample space, whereas the definition of significance tests requires at least a partial order. The mathematical construct called a *rooted tree* can be used to define both these orderings. However, considerable computation can be saved by recognising symmetries. The study of such symmetries leads naturally to the theory of groups.

Although this research arose out of practical concerns, only in the final chapter do we present any numerical results. We do not include any computer programs, and discuss algorithms only in general terms. The main emphasis of this thesis is neither the practical implications of various statistical procedures, nor the implementation of algorithms, but

the *mathematical structure* which makes the development of efficient algorithms feasible, and which in turn will point to better understanding of the operating characteristics of bioassays.

1.2 Summary

Chapter 2 sets a background by summarizing the current state of the art of the use of statistics in toxicology. We will walk through the steps involved in the bioassay, at each step examining any questions for which statistical methodology might help. Included in this overview are considerations of experimental design, data analysis, and risk assessment. Only some of these issues are pursued in subsequent chapters.

In Chapter 3, we consider a problem of experimental design with correlated binary data, such as might occur in allocating litter mates between two treatments. The asymptotic efficiencies of the completely randomized, litter matched, and nested designs are compared. The calculation of the *exact* operating characteristics requires the development of efficient algorithms. The structure underlying such algorithms is developed in the next three chapters.

The development begins in Chapter 4, which examines a family of discrete probability distributions containing the binomial, hypergeometric, and beta-binomial distributions as special cases. The properties of the family, which allow for efficient computational algorithms, are presented. This family is then extended to the multivariate case using a conditioning argument.

To extend the computational algorithms to the multivariate case, we require an ordering on the sample space. This is achieved in Chapter 5 by considering the sample points to be the nodes of trees. Several constructions are presented, and their role in the computation of probabilities is examined. The structure of the trees is seen to be intimately related to the design of efficient algorithms.

In Chapter 6, we return to the bioassay design problem to consider the computation of exact operating characteristics of permutation tests. Using the concept of the algebra of events invariant under a group of permutations, we show how the computational burden can be reduced by restricting the computations to a subalgebra. Monotonicity properties of the distribution of certain conditional test statistics then allow for recursive calculations of rejection probabilities, employing the methodology developed in Chapter 5.

Chapter 7 discusses a collaborative empirical study involving the analysis of a large data base of actual toxicological studies. Although the data itself has some limitations, consideration of some of its deficiencies led to valuable lessons and some interesting problems. The issues highlighted here concern the estimation of false positive and false negative rates of screening procedures. In the process, we derive some results on two-phase sampling with estimated selection probabilities, and exploit the empirical distribution of *P*-values for estimating the proportion of carcinogens in a population of chemicals.

Chapter 2

An Overview of the Role of Statistics in Toxicity Bioassays

The growth of the chemical industry in the modern era has led to the introduction of many new chemicals into the environment. Although most of these chemicals are being manufactured for human benefit, many of them also have harmful effects. The study and regulation of these chemicals have given rise to the science of toxicology, one of whose major tools is the animal bioassay. Because such experiments are expensive, and have far-reaching consequences, there has been in recent years a great interest in the proper use of statistical methods in the design and analysis of such experiments, with the consequence that many professional statisticians have devoted their research efforts towards the development and evaluation of appropriate methodology. Much of this methodology is novel only in its manner of application, but much has been specifically designed for the toxicological bioassay. The resulting work has ranged from the very general, such as how to synthesize information from many studies [24], to the very specific, such as what continuity correction to use [81]; from the very practical, such as how to record, analyze, and interpret results [31], to the highly theoretical, such as the asymptotic theory of censored life distributions [1].

This dissertation will concentrate on some specific statistical problems involving the operating characteristics of bioassays, but in the interests of setting a background, we first will summarize the current state of the art of the use of statistics in toxicological testing.

2.1 Study Initiation

The general procedure of toxicological animal bioassay is as follows: A sample of animals is partitioned among a number of treatment groups. Each group is administered a different treatment, and a response is observed. Inferences about the toxicology of the treatments are based on comparing the distribution of responses among the treatment groups.

There is considerable diversity possible within this framework, and hence considerable input possible from the statistician. Some of the contributions of the statistician follow from standard theory of experimental design and statistical inference. Some contributions seem to be specific to the toxicological area, and some, indeed are intrinsically

4

multidisciplinary, requiring that the statistician possess considerable toxicological insight to understand the problem, and that the toxicologist have sufficient statistical training to appreciate the solution. And of course, there are problems that are strictly biological, for whose solution the statistical expert's advice provides little help.

The multidisciplinary nature of toxicological research sometimes leads to confusion about the proper jurisdiction of the various disciplines. One might thus encounter toxicologists attempting to redefine the meaning of statistical significance, pathologists arguing about experimental design, and statisticians discussing the importance of benign neoplasms. The ensuing discussions become philosophical musings at their best, or semantic quibbling at their worst. It would be an important contribution in this field just to define carefully what is a statistical issue and what is not.

A decision that must be made even before the experiment is planned is what chemicals to investigate. It is impractical to test all chemicals, so that prioritization becomes an issue, although largely a non-statistical one. The chemicals that are studied are those of greatest concern, where this concern may be based on social, scientific, political, or economic motives. Some considerations that are involved in the prioritization are: potential for exposure, potency, and toxicity of related chemicals. Statistical methods can assist in the quantification of some of these criteria. For example, survey data on dietary habits would give an estimate of exposure levels to food additives; concentrations of environmental contaminants in cadaverous fat samples would indicate the doses received.

In such cases, the methods of survey sampling would improve and evaluate the estimates obtained. The carcinogenic potency of a compound is a concept that would help quantify societal concerns. The definition and quantification of such a concept does involve statistical input [16,72]. There has been much recent work on the prediction of toxicology from chemical structure. Much of the work is highly statistical [38,33] involving, at its most sophisticated level, the quantification of structural parameters, and the application of non-standard discriminant analysis techniques. However, little of this work has appeared in the statistical literature, and its presentation in subject matter journals has skirted the technical details.

Once a chemical is selected, the first decisions an investigator must make involve the test animal to use, and the route of administration. Since the ultimate goal is usually to make inferences about toxicity to man, these choices are dictated by the need to make the animal model as similar to the human situation as possible. Practical considerations also have some bearing. Ethical issues and economic restraints usually limit the testing to small rodents. These decisions are mostly non-statistical, and are based on the toxicologist's understanding of the metabolism of the chemical in question. However, some attempts at using statistical ideas in this area have been made, particularly with reference to carcinogenicity. Historical data on other bioassays can be combined with epidemiological results to quantify the predictability of human susceptibility from that of various species. Such findings can be expressed in terms of sensitivity and specificity [44]. From such information it might appear that one species is a better mimic of human susceptibility than others. However, reliance on such data can be misleading. For one thing, historical data do not in any sense result from a random sample from some population of chemicals from which a new chemical is to be selected. Thus, there is little basis for believing that the predictive utility of a particular species in previous studies is indicative of

its ability at present. Moreover, such an approach would disregard important information about the metabolism of the chemical in question, which ought to be a better indicator of interspecies comparability than mere statistical association.

2.2 Experimental Design

It is at the time of setting up the experiment that the statistical contribution becomes important. Here the theory of experimental design can greatly assist the investigator in achieving scientific validity and statistical efficiency.

The first issue is the selection of the animals. Traditionally, animals have been selected from a single inbred strain, in the interests of maintaining as uniform a sample as possible. Although the desire for uniformity stems from a misunderstanding of the principles of experimental design, it is so firmly entrenched in the toxicological dogma that there is little chance of uprooting it in the near future. Some workers [34] have indicated advantages of experimental heterogeneity, suggesting that the use of outbred strains, or a mixture of inbred strains would be more informative. However, such ideas have not caught on. We will not address this issue further, but will assume that the test animals are a "quasi-random" sample from some fairly homogenous population.

The nature of rodent breeding is such that the test sample usually consists of a number of identifiable litters. Thus sibling relationships are known, and it is a germane question as to how this information should be used. Traditionally, such information has been ignored, or has been used in an *ad hoc* manner. However, if the response of interest has either a genetic or perinatal component, then using litter as a blocking factor could prove advantageous. On the other hand, there are experiments in which an entire litter of animals must be on the same treatment. Such would be the case, for example, in so-called two-generation studies [25], which provide for effects due to *in utero* exposure. We will return to this question in Chapter 3.

Another question of importance is the physical layout of the experiment. Toxicologists have been slow to accept the importance of randomization. The customary layout of animal bioassays is to arrange the treatment groups in systematic blocks, with all cages containing animals of one treatment group being physically contiguous, doses increasing as one moves across the room. Animals may be in separate cages, or might be grouped several to a cage. Thus all control animals would be clustered on one side of the room, and all high-dose animals on the other side. Although even undergraduate students of statistics should recognize such a design as flawed, it is defended by many toxicologists on practical grounds. One advantage of such a layout is that it makes dosing the animals particularly simple, and avoids the problem of inadvertent mixup due to bad record keeping or negligent staff. Additionally, it avoids the possibility of one animal's cage being contaminated by the spillage of feed or excreta from a differently treated neighbour. Moreover, it is believed that the toxicologists' obsessive efforts to control the experimental environment should eliminate any concern about environmental gradients.

Several workers [52,8,67] have presented experimental evidence that environmental gradients do exist, although vertical gradients appear to be of more importance than horizontal ones. These findings, however, have arisen as byproducts of other experiments. Uniformity trials, such as have been carried out in agricultural research [20], have not

been attempted with toxicological bioassays, partly, no doubt, because of their high cost. Nevertheless, even if such experiments did fail to find gradients, they would not justify the traditional design. The experimental unit in that design is the cage block rather than the individual animal, and in that light one sees that since there is no replication, there can be no valid estimate of experimental variation. The practical difficulties with a properly randomized design need to be addressed, and this is one area in which the statistician and toxicologist must work closely to achieve what is both practical and scientifically valid. If cross-contamination is a problem, then cage blocks of like treatments may have to be used. However, these blocks should be replicated and randomized, and the subsequent analysis must recognize the cage block as the experimental unit.

A similar situation already exists in layouts with multiple caging, in which the cage rather than the animal is the experimental unit. This clustering is often ignored in the statistical analysis: indeed, information that would allow for its inclusion is often withheld from the statistician. Thus, most toxicological bioassays are analyzed as if they were executed following a completely randomized design, which in many cases is far from the truth.

An issue akin to randomization, and even more controversial [7], is the question of treatment labelling: Should the animal records indicate the treatment received, or should they only give a code number? The final observations of these studies are generally a pathology report on each animal, which usually involves microscopic examination of various tissues. The somewhat subjective nature of diagnostic pathology would strongly suggest that the pathologist be blinded as to the treatment the animals were receiving. Otherwise, there is the risk that he may exercise more diligence in seeking abnormalities in the treated groups than in the control group. Moreover, there is a tendency among pathologists to examine first samples from the control group to get an idea of what normal tissues look like, and then to examine samples from the high dose group. Such a procedure could not only disturb the independence of observations that is assumed by the analysis, but could also introduce a bias due to a drift in the pathologist's perception. Such a drift could be caused by fatigue, or by novelty: Consider a pathologist who notices, for the first time, a certain interesting anomaly in a tissue part way through the examination. Might he not, from this point on, be likely to notice subtle instances of this kind of thing, whereas before he might have overlooked it?

Considerations of this sort lead the statistician to insist that animal records contain no treatment identification, and that pathological examinations be done in a random order [27]. However, such suggestions are violently opposed by most pathologists [60], partly because of the nuisance of dealing with coded samples, and also because of the apparent questioning of their scientific integrity.

There is nevertheless a valid point to the pathologists' hostility. The purpose of the experiment is to distinguish effects of the treatments, not to assess the discriminatory abilities of the pathologists. Hence, anything that impedes the identification of treatment effects, which, pathologists insist, coded samples do, should be avoided. One approach, which may satisfy both pathologists and statisticians, but has not been much used, would be to present the pathologist with animals in triples. One of the three would be a control animal, identified as such, while the other two would be unidentified. The pathologist could then make a judgment about the relative pathology of the two unidentified specimens,

using the control sample as a reference point. The unidentified pairs would comprise all pairs of treatment groups, including controls. For some pairs, both animals should come from the same treatment, thus providing a "pure" estimate of error. The statistical analysis of such an experiment should, of course, be based on this actual design.

Consideration of pathology leads one to ask about the relevant variables to be observed—i.e., what outputs of the experiment are to be analyzed? In the very simplest situation one is looking for a particular binary response. The simplest experiment is the so-called acute toxicity bioassay, in which one is interested in the effect of a single large dose of the chemical in question. The response of interest may in fact be death: the animal either survives the treatment, or does not. (In this case, pathology is not an issue.) In more subtle situations the response may be the appearance of some clinical sign, which does lead to the possibility of misdiagnosis.

In the chronic toxicity bioassay, on the other hand, the animals are exposed to the chemical throughout a period of time; often their entire lifetime, and the cumulative effects of this exposure need to be assessed. Usually the concern in this case is cancer, and the binary endpoint is the development of a malignant lesion. Two complications arise at this point.

The first is that the "binary" response develops over a course of time, and the time of development may itself be of interest. Thus a chemical that induces tumours early in life would be considered more toxic than one that only does so in senility. The time to tumour response itself presents several difficulties. Since most tumours are not directly observable in the living animal, the time of their occurrence can only be given an upper bound—i.e., the observations are left-censored. Furthermore, animals may die during the course of the experiment, and hence may not be at risk of tumour development throughout the experimental period. Observations on time-to-tumour for these animals are right-censored. Ignoring this censoring can lead to bias in either direction: Early detection of tumours may be an artefact due to early death (and hence necropsy) unrelated to tumour induction; on the other hand, early death could also explain a paucity of tumours. Tumour prevalence cannot be properly assessed from natural mortality alone, but requires that a subgroup of the animals be killed for necropsy at predefined intervals [82].

The second complication is that pathology results are essentially multivariate. Most animals will develop any number of lesions during the course of their lifetimes, most of them innocuous. Thus the endpoint observed will not be just the presence or absence of a tumour, but a profile of many lesions [31]. Recording the complete profile of individual animals can avoid some of the difficulties discussed in Chapter 7.

Most of the design considerations discussed above apply to all experiments. However, once one gets down to the choice of treatment groups, the purpose of the experiment becomes important. Toxicological experiments can be classified as to their purpose along statistical inference lines: either testing or estimation. An experiment for testing (or screening as toxicologists prefer to call it) has as its goal the classification of a compound as either toxic or innocuous. Thus a binary decision rule is invoked, and the terminology and methodology of hypothesis testing and decision theory can be employed. Estimation experiments, on the other hand, usually have as their goal the estimation of a dose-response curve for a chemical. Methods of optimal design theory and curve fitting then apply [45,42]. Other types of experiments can also be done, such as experiments to study the joint action of several chemicals, reversibility of lesions, initiation/promotion phenomena,

etc. A particularly important case of the estimation problem is the estimation of an extremely low quantile of the dose response curve—the so-called low-dose extrapolation problem which is discussed in Section 2.4.

Decision theoretic ideas have not been much employed in toxicity screening experiments. One reason for this lack is the difficulty of positing a realistic loss function. The consequences of incorrectly classifying a carcinogenic compound as innocuous are human suffering and premature death, whereas the losses of damning a useful chemical as carcinogenic involve profits of the manufacturer and possibly, social benefits. Such losses do not seem commensurable.

Screening studies are generally evaluated by means of significance tests, the significance level generally being the traditional 5%. The design problem then involves the maximization of the power of the procedure. At first sight, it would seem that power would be maximized with only two dose groups, one being a control, and the other dosed with as high a level as possible. There is a difficulty with this design, however, in chronic carcinogenicity studies. The response of interest in such experiments is the development of neoplasia in the *absence of any acute toxic effects*. Thus the dose must be low enough that the animals suffer no ill effects other than those due to induced tumours. The concept of a "maximum tolerated dose" (MTD) is an idealized one, although operational definitions have been proposed [64]. Under ideal conditions, the most efficient design would have two treatments, one on zero dose, and one at the MTD. In practice, however, the MTD can only be guessed at, so doses below the estimated MTD are also included. If the top dose proves to be too high then the treatment may cause premature death, thus precluding the development of late appearing tumours. In this case, data from the intermediate doses render the experiment salvageable. The use of several dose levels also allows for the estimation of a dose response curve. The current standard design adopted by the United States National Toxicology Program has four dose groups: MTD, half-MTD, quarter-MTD, and control.

If the purpose of the experiment is to estimate a dose response curve, or some aspect of it, then the dose groups should be chosen to optimize some criterion quantifying the goodness of the estimate. This criterion is usually based on variances, which are in general defined only asymptotically in non-linear models.

One difficulty with designs in non-linear models is that the optimal design generally involves the unknown parameters, thus some idea of the parameters (possibly from a preliminary experiment) is necessary before the design can be implemented. A natural development of this situation would be a sequential experiment, such as the designs considered by Wu[84], and McLeish and Tosh[59]. Unfortunately, such sequential procedures are not practical for many toxicological studies on account of the length of time required to observe a response.

The duration of the experiment is another design parameter that the experimenter can manipulate. Traditionally, chronic toxicology experiments have run for the major part of the animals' lifetime (about two years for rats). However, because of the costs involved, early termination may prove advantageous in screening experiments. If there is sufficient evidence of a substance's toxicity early in life, then there is little purpose in prolonging the investigation. Procedures for such early termination have been discussed by Berry[11] and Louis[56]. Early termination of a subset of animals may also be required to observe

the progressive development of lesions [54] or possibly, their regression [8,55].

2.3 Statistical Analysis

The statistical analysis of toxicological data normally proceeds with tests of significance and often, estimation of dose response curves. From the latter, estimates of low dose risk are obtained, and this information is combined with biological and environmental data to produce an estimate of the health hazard of the chemical in question. In the following we will examine some of the issues involved in this process.

The simplest statistical analysis involves the comparison of a binary response in a control and test group. The null hypothesis of equality of response probabilities, against a one-sided alternative, can be tested using the Fisher-Irwin test [13] (or *Fisher's exact test* as it is generally known in the toxicological community). This test is a permutation test whose validity depends solely on the random allocation of treatments to experimental units. This test has been criticized, however, [10] because it conditions on a statistic, (the sum of the responses in the two groups) that is not ancillary. One consequence is that this test can be extremely conservative when the response probabilities are low. Since the permutation test for balanced designs is closely approximated by the asymptotic normal theory test with continuity correction, whereas the uncorrected test has size closer to the nominal significance level; this controversy can be rephrased into whether one should use a continuity correction in Pearson's χ^2 [22]. Alternatives to the Fisher-Irwin test, however, do base their validity on the assumption of comparing two binomial populations, an assumption that may not be warranted in experiments with heterogeneous biological populations [40]. Moreover, in the toxicological area, the conservatism of the Fisher-Irwin test turns out to be an advantage because it compensates for the problem of multiplicity [31].

In most experiments, the active site of the chemical may not be known *a priori*. Thus the possibility of induced lesions is examined in a large number of tissues. There are typically two extreme strategies in this situation. One is to reduce the multivariate problem to a univariate one, by defining an animal as "responding" if it responds in any tissue, and performing a single test of significance. The other is to test each tissue marginally, and to reject the overall null hypothesis if it is rejected for any tissue. The first approach may lead to reduction of power: Typically experimental animals spontaneously develop lesions in various tissues, whereas chemical treatments tend to induce lesions at specific sites. Thus the aggregation of lesion types will tend to obscure a treatment-induced effect with the random occurrence of diverse spontaneous lesions.

On the other hand, the second approach has been criticized [68] as inflating the false positive rate since each lesion is typically tested at a nominal significance level of .05. However, Gart et al. [31] have pointed out that the conservatism of the Fisher-Irwin test generally compensates for the multiplicity in most situations. This observation depends on the fact that background incidences of most spontaneous lesions are very low. Haseman [26] has proposed modifying the nominal significance level for common lesions to maintain the required conservatism.

Much of the foregoing can be repeated for the situation where there are more than two treatment groups. In this case, the asymptotic theory test would correspond to the

Cochran-Armitage test [19,2], suitably modified for one-sided alternatives.

A small-sample test, analogous to the Fisher-Irwin procedure, can be constructed by permuting the trend statistic $\sum d_i y_i$, where the d_i 's represent dose, and the y_i 's the binary responses, conditioning on $\sum y_i$, the total number of responders in the experiment. This test is also the likelihood ratio test for zero slope if the responses are modelled as a logistic regression in dose [23]. The exact test can be computationally cumbersome unless efficient algorithms, such as those presented in Chapter 5 are used. Again it appears that the asymptotic test *corrected for continuity* approximates the permutation test very closely, and that both tend to be conservative, at least for balanced designs. These issues are explored in Chapter 7.

Another issue in such analyses is the incorporation of historical information on spontaneous incidence of the lesion in question. One recent proposal [80] naïvely suggests that bioassays in which the high dose group falls within the historical control range be discounted, statistical significance notwithstanding. The fact that such an approach, recommended by highly respected scientists, completely ignores the nature of random variability and misinterprets the meaning of significance tests illustrates how much confusion reigns in this area. A more reasonable procedure for incorporating historical controls was proposed by Tarone [78] in the context of a logistic regression. Although the author seems to have missed the empirical Bayes nature of this approach, the idea has been picked up by Krewski et al. [50]. The consensus appears to be that addition of historical control information can render insignificant results significant, but is unlikely to undermine the significance of an established effect. Indeed, a proper understanding of the concept of statistical significance would intuitively lead one to such a conclusion. There also appears to be evidence [76] that Tarone's procedure is questionable for rare tumours. Our empirical study discussed in Chapter 7 supports this view.

As mentioned before, most toxicological experiments, in addition to giving binary response data, also provide data on the time the response was first observed. In some cases this time corresponds to the time the response occurred, but in others, the occurrence time is only known to have preceded the observation time. Typically, the response of interest is a neoplastic lesion in an internal organ, which is only observed at necropsy. Such lesions can be classified as either contributing to the death of the animal (so-called *fatal* tumours) or being incidental to it. The importance of distinguishing the two types of response was pointed out by Peto [65], although similar considerations were discussed by Hoel and Walburg [36]. A good discussion of these issues is given by Lagakos and Louis [51].

There are essentially two aspects to the time-to-response problem, which could be briefly denoted as preservation of validity, and preservation of efficiency. The validity issue arises when the test chemical causes premature mortality of the animals. In this case, the treated animals are at decreased risk of developing tumours, so that one gets biased estimates of the differences in (age-specific) tumour incidence unless the competing risk is adjusted for. However, such an adjustment can have the opposite bias on incidental tumours, since the earlier death of animals from other causes would make these tumours observable sooner, without any difference in incidence. These biasing effects should not occur if non-tumour mortality was equal in the two groups—i.e., if the highest dose does not exceed the MTD.

However, even if the MTD is not exceeded, there remains the question of efficiency:

How much information is lost by ignoring the time-to-occurrence data, and just analyzing the lifetime incidence of tumours? Some recent work [12,32] suggests that at least in certain situations, the information loss is small. In the following, therefore, we shall concentrate on the binary response situation.

2.4 Low Dose Risk Estimation

From the point of view of public health, the most important use of dose response models is in estimating low dose risk. Because risks that are too low to be detected in experiments of practical size may still be large enough to cause public concern, this estimation necessarily involves interpolating an estimated dose response curve. The difficulty is that the low dose risk estimates are not robust to model specification errors. Thus, two models that fit the data equally well can give very different low dose estimates.

Suppose one wants to estimate the incremental risk at a low dose, say δ . That is, one wants to estimate $F(\delta) - F(0)$, where F is the dose response function giving the probability of a response as a function of dose. Assuming that F has sufficient derivatives, we can expand the response function in a Taylor series, giving

$$F(\delta) - F(0) \simeq \sum_j F^{(j)}(0) \delta^j.$$

Provided that $F'(0) > 0$, the dose response curve is essentially linear at sufficiently low doses, so that the added risk is proportional to dose. Such dose response curves are called *low-dose linear*. However, if the first $k - 1$ derivatives vanish at the origin, then the dose response curve resembles δ^k at low doses. This case, called *sublinear*, leads to rather small increments in risk. For this reason, estimation of low-dose risk assuming linearity is sometimes called *conservative*, although this use of the term is opposite in meaning to its use in the context of significance testing. Suppose that the dose-response curve is *convex* in the low dose region, (so that $F'' \geq 0$) and suppose that one interpolates the response at dose δ linearly,

$$\bar{F}(\delta) = \left(1 - \frac{\delta}{x}\right) F(0) + \frac{\delta}{x} F(x)$$

for some $x > \delta$. Then since the secant from $F(x)$ to $F(0)$ lies above the dose response curve, we have that $\bar{F}(\delta) \geq F(\delta)$. Thus, the low-dose risk is overestimated, and the error is on the side of safety.

Although many empirical dose response curves do appear to be convex in the low dose region, it is in fact impossible to ascertain their actual shape in the region of interest. Rather, the low dose shape is usually posited on theoretical grounds. One of the earliest models used in toxicology is the probit model, defined by assuming that each individual in the population has a tolerance to the chemical, and that this tolerance has a log-normal distribution. Thus the probability of a response at dose x is given by $F(x) = \Phi(\alpha + \beta \log x)$ where Φ represents the standard normal distribution function, and α and $\beta > 0$ are parameters. For this model, the dose response curve is indeed convex in the low dose region, and in fact all derivatives of F vanish at the origin, leading to extremely small incremental risks for low doses. Unfortunately, the probit model has no biological basis. One could just as well take another tolerance distribution $F(x) = \Psi(\alpha + \beta \log x)$ [47]. If

Ψ has an exponential left tail, i.e. $\log \Psi(x) = O(x)$ as $x \rightarrow -\infty$, then the low-dose shape depends on the parameter values: It is convex if $\beta > 1$ and concave otherwise. In fact, if $\beta < 1$ then $F'(0)$ is infinite, leading to very large incremental risks at low doses. If $\log \Psi(x) = o(x)$ as $x \rightarrow -\infty$ then $F'(0)$ is infinite for all parameter values.

The so-called *supralinear* case of $F'(0) = \infty$ is sometimes dismissed as biologically meaningless. The most popular theoretical model for carcinogenic dose response curves is the multi-stage model [3] first proposed by Armitage and Doll [4,5]. If no spontaneous incidence is allowed for, this model leads to a log-tolerance distribution with Ψ equal to the extreme value distribution, but β restricted to be a positive integer. Thus the annoying supralinear case of $\beta < 1$ is excluded. Moreover, although the case $\beta = 1$ is concave, the curve is quite straight near the origin: Extrapolating linearly from a 10% response point underestimates the risk by a *relative* error of less than 5%.

However, supralinear curves can arise from a mixed population of low-dose linear cases. For example, if $F(x) = 1 - e^{-\alpha - \beta x}$, where the parameter β itself is a random variable B having the log-logistic distribution $\Pr\{B \leq \beta\} = \beta(1 + \beta)^{-1}$, then the resulting curve will be supralinear.

In addition, strongly concave (albeit low-dose linear) curves can arise if non-linear pharmacokinetics are factored into the model [35,46]. The principle is that the amount of active chemical reaching the target site is not equal to the administered dose, but is possibly a non-linear function of it. If the administered chemical is not itself carcinogenic, but is metabolized to one that is via a pathway that is saturable, the resulting dose response curve can display pronounced downward curvature. In such a situation, linear extrapolation would not lead to a "conservative" estimate of low dose risk.

Linear extrapolation is becoming, however, the accepted procedure for estimating low-dose risk because it is conceptually simple, robust to model misspecification, and also partly because it is believed by some to give "conservative" estimates.

There are basically three classes of linear extrapolation methods [48]. The simplest is merely to draw a straight line between the control and lowest observed response, and to use that line for estimating low-dose risk. On the other hand, one could fit to the observed data a low-dose linear model, and use the estimated dose-response curve to predict the low dose risk. Although this estimate would be model dependent, the low-dose linearity would seem to minify such dependence. Or finally, one could combine the two procedures, by first fitting a dose-response model (not necessarily low-dose linear) and then drawing a straight line from the predicted control response to a point on the estimated curve corresponding to a not too low response, say 1-10%.

2.5 The Ultimate Issue

The final step in the toxicological risk assessment process is more like a leap: How does one take an estimate of risk based on a rodent study and extrapolate that to a risk for a human population? The current dogma is that the lifetime risk of a carcinogenic response is the same for man as for the test species at equivalent doses of the test compound. What constitute equivalent doses remains a bit of a controversy. The common solution is to equate doses on a weight per weight per day basis. However, physiological considerations lead one to consider equivalence on a weight per surface area basis, and there is some

empirical support for this model [49]. Since surface area is rarely measured, it is usually approximated by the $2/3$ power of weight.

A statistical methodology for systematically comparing the susceptibility of different species, as well as comparing related compounds, has been worked out by duMouchel and Harris[24]. This procedure, however, does not appear to have been applied on a wider basis than the examples done by the authors in their original paper.

Chapter 3

Asymptotic Theory for Comparing Two Treatments with Correlated Binary Responses

3.1 Introduction

Consider an experiment having two treatments, a control group and a test group, with a binary response variable. Traditionally, the statistical significance of a treatment effect is determined on the basis of a model assuming that the experimental units are mutually independent, and that the responses are identically distributed within a treatment group. This model leads to a test for the difference between two binomial parameters, which can be performed using either an asymptotic or exact distribution. In actual fact, however, the experimental units might not be independent. One solution to this difficulty is to use a permutation test, whose validity depends solely on experimental randomization. The justification of this approach will be discussed in Chapter 6. In the present chapter, we will examine, by asymptotic arguments, how the usual test statistic is affected by correlation between certain experimental units under three randomization schemes.

Motivating our terminology from the toxicological bioassay, we will refer to the experimental units as *animals* which come from a set of identifiable *litters*. Different litters are assumed to be independent, but responses of animals in the same litter are related by the *intra-litter correlation*, which may depend on the treatment. We propose a model that provides for this treatment-specific correlation, and derive the mean and variance of the usual test statistic under this model. We also prove the asymptotic normality under the null hypothesis.

3.2 Statistical Model

3.2.1 A correlated independent competing risks model

Suppose that there are m litters, each of size l , giving a total of $n = ml$ animals in the experiment. Although contrary to custom, we will find it convenient to code lack

of response (e.g. survival) as 1, and a response as 0. Each animal is assumed to have two possible modes of response: *Spontaneous*, unrelated to the treatment; or *induced*, due to the treatment. These two modes of response are considered to be independent competing risks with only their combined effect being observable. We thus define the following response variables to represent spontaneous, induced and observed responses:

$$\begin{aligned} X_{ij} &= \begin{cases} 0 & \text{if the } i^{\text{th}} \text{ animal in the } j^{\text{th}} \text{ litter responds spontaneously,} \\ 1 & \text{otherwise;} \end{cases} \\ Z_{ij} &= \begin{cases} 0 & \text{if the } i^{\text{th}} \text{ animal in the } j^{\text{th}} \text{ litter responds due to treatment,} \\ 1 & \text{otherwise;} \end{cases} \\ Y_{ij} &= \begin{cases} 0 & \text{if the } i^{\text{th}} \text{ animal in the } j^{\text{th}} \text{ litter is observed to respond,} \\ 1 & \text{otherwise;} \end{cases} \end{aligned}$$

for $i = 1, \dots, l$; $j = 1, \dots, m$.

The actual outcome of the experiment, of course, depends on the design. We thus define design variables

$$a_{ij} = \begin{cases} 0 & \text{if the } i^{\text{th}} \text{ animal in the } j^{\text{th}} \text{ litter is in the control group,} \\ 1 & \text{if the } i^{\text{th}} \text{ animal in the } j^{\text{th}} \text{ litter is in the test group.} \end{cases}$$

We will use $\mathbf{X}$, $\mathbf{Y}$, $\mathbf{Z}$, and $\mathbf{a}$ to represent the $n \times 1$ vectors with components X_{ij} , Y_{ij} , Z_{ij} , and a_{ij} respectively, in reverse lexicographic order.

Animals in the control group can only respond spontaneously. However, either mode of response is possible in the test group, so that animals under this treatment will fail to respond if they respond in neither fashion. Thus

$$\begin{aligned} Y_{ij} &= (1 - a_{ij})X_{ij} + a_{ij}X_{ij}Z_{ij} \\ &= X_{ij} + a_{ij}X_{ij}(Z_{ij} - 1). \end{aligned} \quad (3.1)$$

We assume that the probability of response is the same for all animals in a group, and that the responses in different litters are independent. As mentioned above, we also assume that induced and spontaneous responses are independent. The first and second moments of the response variables can be parametrized as follows:

$$\begin{aligned} E(X_{ij}) &= \mu_S, & E(Z_{ij}) &= \mu_I, \\ E(X_{ij}Z_{i'j'}) &= \mu_S\mu_I, \\ \text{Var}(X_{ij}) &= \mu_S(1 - \mu_S), & \text{Var}(Z_{ij}) &= \mu_I(1 - \mu_I); \\ \text{Cov}(X_{ij}, X_{i'j'}) &= \rho_S\mu_S(1 - \mu_S), & \text{Cov}(Z_{ij}, Z_{i'j'}) &= \rho_I\mu_I(1 - \mu_I) \\ & & & \text{(if } i \neq i'); \\ \text{Cov}(X_{ij}, X_{i'j'}) &= 0, & \text{Cov}(Z_{ij}, Z_{i'j'}) &= 0 \\ & & & \text{(if } j \neq j'). \end{aligned} \quad (3.2)$$

Here $1 - \mu_S$ and $1 - \mu_I$ are the individual animal probabilities for spontaneous and induced responses, respectively, and ρ_S and ρ_I are the corresponding intra-litter correlations.

The covariance structure of the X_{ij} 's and the Z_{ij} 's can be more compactly represented as a tensor product (see Appendix A) of covariance matrices:

$$\text{Cov}(\mathbf{X}) = \mu_S(1 - \mu_S)((1 - \rho_S)\mathbf{I}_l + \rho_S\mathbf{J}_l) \otimes \mathbf{I}_m, \quad (3.3)$$

$$\text{Cov}(\mathbf{Z}) = \mu_I(1 - \mu_I)((1 - \rho_I)\mathbf{I}_l + \rho_I\mathbf{J}_l) \otimes \mathbf{I}_m, \quad (3.4)$$

where, as customary, $\mathbf{I}_l$ represents an $l \times l$ identity matrix, and $\mathbf{J}_l$ represents an $l \times l$ matrix of 1's. (The subscripts on $\mathbf{I}$ and $\mathbf{J}$ may be suppressed when the dimension is clear from the context).

Now, if μ_0 and μ_1 represent the individual animal probabilities of non-response in control and test groups, respectively, we have

$$\mu_0 = E(Y_{ij} | a_{ij} = 0) = \mu_S$$

and

$$\mu_1 = E(Y_{ij} | a_{ij} = 1) = \mu_S\mu_I.$$

One can also readily calculate the covariances between the responses of litter mates when both are controls, when exactly one is treated, and when both are treated.

$$\text{Cov}(Y_{ij}, Y_{i'j} | a_{ij} = 0, a_{i'j} = 0) = \rho_S\mu_0(1 - \mu_0), \quad (3.5)$$

$$\begin{aligned} \text{Cov}(Y_{ij}, Y_{i'j} | a_{ij} = 0, a_{i'j} = 1) &= \rho_S\mu_S\mu_I(1 - \mu_0) \\ &= \rho_S\mu_1(1 - \mu_0), \end{aligned} \quad (3.6)$$

$$\begin{aligned} \text{Cov}(Y_{ij}, Y_{i'j} | a_{ij} = 1, a_{i'j} = 1) \\ = \mu_I\mu_S[(\mu_I + \rho_I(1 - \mu_I))(\mu_S + \rho_S(1 - \mu_S)) - \mu_I\mu_S] \\ (i \neq i'). \end{aligned} \quad (3.7)$$

From this one can derive the correlation between two animals in the same litter as

$$\rho_0 = \rho_S, \quad (3.8)$$

$$\rho_{01} = \rho_S \sqrt{\frac{\mu_1(1 - \mu_0)}{\mu_0(1 - \mu_1)}}, \quad (3.9)$$

$$\text{and } \rho_1 = \frac{(\rho_I(1 - \mu_I) + \mu_I)(\rho_S(1 - \mu_S) + \mu_S) - \mu_I}{1 - \mu_1} \quad (3.10)$$

for the cases of two control animals, one control and one test animal, and two test animals, respectively.

3.2.2 Three randomization schemes

We now consider randomized designs for such experiments. The design variables a_{ij} are assumed to be random variables with a joint distribution depending on the randomization scheme, and it is their realized values that define the actual design used. We consider, in general, randomization schemes in which the design variables are independent of the response variables, and the proportion π of test animals is a predetermined, non-random quantity. (Sequential designs, for example, are thus excluded.) We highlight, for special emphasis, three designs that are commonly used in toxicology.

The first design ignores the litter structure, and chooses a simple random sample of the n animals to constitute the test group. We call this the *completely randomized design*. It is the design most commonly used for bioassay because of its conceptual simplicity and ease of analysis. Assuming that a fraction π of the animals are to be put on the test treatment, the completely randomized design induces the following structure on the design variables.

$$E(a_{ij}) = \pi, \quad (3.11)$$

$$E(a_{ij}a_{i'j'}) = \frac{\pi(\pi n - 1)}{n - 1} \quad (i, j \neq i', j'), \quad (3.12)$$

$$\begin{aligned} \text{Cov}(a) &= \frac{\pi(1 - \pi)}{n - 1} (nI_{lm} - J_{lm}) \\ &= \frac{\pi(1 - \pi)}{n - 1} (nI_l \otimes I_m - J_l \otimes J_m). \end{aligned} \quad (3.13)$$

The *litter matched design* results from selecting a stratified sample for the test group, where the strata are the litters. This design is used in bioassays in an attempt to reduce genetic heterogeneity. Assuming that the number of litters divides the number of treated animals; we have

$$E(a_{ij}) = \pi, \quad (3.14)$$

$$E(a_{ij}a_{i'j'}) = \pi^2 \quad \text{if } j \neq j', \quad (3.15)$$

$$E(a_{ij}a_{i'j}) = \frac{\pi(\pi l - 1)}{l - 1} \quad (i \neq i') \quad (3.16)$$

from which

$$\text{Cov}(a) = \frac{\pi(1 - \pi)}{l - 1} ((I_l - J_l) \otimes I_m). \quad (3.17)$$

Finally the *nested design* is achieved by selecting the treated animals as a cluster sample, with litters as clusters, giving

$$E(a_{ij}) = \pi, \quad (3.18)$$

$$E(a_{ij}a_{i'j}) = \pi, \quad (3.19)$$

$$E(a_{ij}a_{i'j'}) = \pi \left(\frac{\pi n - l}{n - l} \right) \quad (j \neq j') \quad (3.20)$$

and hence

$$\text{Cov}(a) = \frac{\pi(1 - \pi)}{n - l} (J_l \otimes (nI_m - lJ_m)). \quad (3.21)$$

Such a design is mandated for any experiments involving perinatal effects [25], but is also used (sometimes inadvertently) in other situations.

3.2.3 Test statistic

Now, to test the null hypothesis $\mu_I = 1$ against the alternative $\mu_I < 1$ we will use the test statistic

$$T = \phi'Y \quad (3.22)$$

where the components of ϕ are given by

$$\phi_{ij} = \frac{\pi - a_{ij}}{\sqrt{\pi(1-\pi)}} \quad (3.23)$$

This statistic is proportional to the difference between the observed response rates in control and test groups, and thus is equivalent to the usual asymptotic test for comparing two binomial populations. The (design dependent) coefficients have been normalized so that

$$E(\phi) = 0, \quad \phi'1 = 0, \quad \text{and} \quad \phi'\phi = n. \quad (3.24)$$

Note that since there is a one-to-one correspondence between the design variables a and the coefficient vectors ϕ , the design is unambiguously defined by either one. The abuse of terminology is thus harmless when we talk about "the design ϕ ".

3.3 Moments of the Test Statistic

We now derive the unconditional mean and variance of the test statistic assuming a randomized design. The derivation proceeds by first conditioning on the design variables. The litter matched and nested designs could actually be handled in a simpler way, but we present them here as special cases of the general setup. We will, however, derive the variance first under the null hypothesis, because the derivation is much simpler in that case.

3.3.1 Expected value

The expected value of T can easily be derived by noting that

$$\begin{aligned} E(Y_{ij} | \phi) &= E(X_{ij} + a_{ij}X_{ij}(Z_{ij} - \bar{f}) | \phi) \\ &= \mu_S + a_{ij}\mu_S(\mu_I - 1) \end{aligned} \quad (3.25)$$

since the design variables are independent of the response variables X_{ij} and Z_{ij} . Combining (3.25) with (3.23) gives

$$\begin{aligned} E(T | \phi) &= \frac{1}{\sqrt{\pi(1-\pi)}} \sum_{ij} (\pi - a_{ij})(\mu_S + a_{ij}\mu_S(\mu_I - 1)) \\ &= \frac{1}{\sqrt{\pi(1-\pi)}} \sum_{ij} (\pi\mu_S + a_{ij}(\pi\mu_S\mu_I - \pi\mu_S - \mu_S\mu_I)) \\ &= n\sqrt{\pi(1-\pi)}\mu_S(1 - \mu_I), \end{aligned} \quad (3.26)$$

since $\sum_{ij} a_{ij} = n\pi$. The conditional expectation does not depend on the design, and hence

$$E(T) = E(E(T | \phi)) = n\sqrt{\pi(1-\pi)}\mu_S(1-\mu_I). \quad (3.27)$$

The hypothesis of no treatment effect is equivalent to there being no induced responses ($H_0: \mu_I = 1$). Under this null hypothesis, $E(T) = 0$.

3.3.2 Variance—null case

We next want to determine the variance of T under the null hypothesis $H_0: \mu_I = 1$. It turns out that the variance of T depends on the design only through the quantity

$$V = \frac{1}{m} \sum_j (\nu_j - l\pi)^2, \quad (3.28)$$

where $\nu_j = \sum_i a_{ij}$ is the number of treated animals in the j^{th} litter. V can be considered to be the variance of the finite population of ν_j , $j = 1, \dots, m$, conditional on ϕ .

Theorem 3.1 Under H_0 , the conditional variance of T given the design ϕ is

$$\text{Var}(T | \phi) = \frac{\mu_S(1-\mu_S)}{\pi(1-\pi)} (n(1-\rho_S)\pi(1-\pi) + m\rho_S V) \quad (3.29)$$

where V is defined by (3.28).

Proof:

$$\text{Var}(T) = E(\text{Var}(T | \phi)) + \text{Var}(E(T | \phi)). \quad (3.30)$$

However, from (3.26) we saw that $E(T | \phi)$ does not depend on ϕ , and thus

$$\text{Var}(E(T | \phi)) = 0.$$

Now $\text{Var}(T | \phi) = \phi' \Sigma \phi = \text{tr} \phi \phi' \Sigma$ where Σ is the covariance matrix of the Y_{ij} 's. But under H_0 , $Y_{ij} = X_{ij}$, so Σ is just the covariance matrix of the X_{ij} 's, given by (3.3).

The $n \times n$ matrix $\phi \phi'$ can be partitioned to conform to the tensor product form of Σ . Specifically we can write

$$\phi \phi' = \frac{1}{\pi(1-\pi)} \begin{bmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} & \dots & \mathbf{A}_{1m} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{A}_{m1} & \mathbf{A}_{m2} & \dots & \mathbf{A}_{mm} \end{bmatrix}, \quad (3.31)$$

where $\mathbf{A}_{jj'}$ is an $l \times l$ matrix whose ii^{th} entry is $(\pi - a_{ij})(\pi - a_{ij'})$. Thus

$$\begin{aligned} \phi \phi' \Sigma &= \frac{\mu_S(1-\mu_S)}{\pi(1-\pi)} \begin{bmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} & \dots & \mathbf{A}_{1m} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{A}_{m1} & \mathbf{A}_{m2} & \dots & \mathbf{A}_{mm} \end{bmatrix} \\ &\times \begin{bmatrix} (1-\rho_S)\mathbf{I}_l + \rho_S \mathbf{J}_l & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & (1-\rho_S)\mathbf{I}_l + \rho_S \mathbf{J}_l \end{bmatrix}, \quad (3.32) \end{aligned}$$

and

$$\text{tr } \phi \phi' \Sigma = \frac{\mu_S(1 - \mu_S)}{\pi(1 - \pi)} \left[(1 - \rho_S) \sum_j \text{tr } \mathbf{A}_{jj} + \rho_S \sum_j \text{tr } \mathbf{A}_{jj} \mathbf{J} \right]. \quad (3.33)$$

Now,

$$\text{tr } \mathbf{A}_{jj} = \sum_{i=1}^l (\pi - a_{ij})^2$$

so

$$\begin{aligned} \sum_j \text{tr } \mathbf{A}_{jj} &= \sum_{j=1}^m \sum_{i=1}^l (\pi - a_{ij})^2 \\ &= n\pi(1 - \pi). \end{aligned} \quad (3.34)$$

(The quantity $\pi(1 - \pi)$ can be interpreted as the variance of the finite population of 0's and 1's constituting the realized values of the design variables.)

The ii' -th element of $\mathbf{A}_{jj} \mathbf{J}$ is $\sum_k (a_{ij} - \pi)(a_{ik'} - \pi) = (a_{ij} - \pi)(\nu_j - l\pi)$. We then have that

$$\sum_j \text{tr } \mathbf{A}_{jj} \mathbf{J} = \sum_j \sum_i (a_{ij} - \pi)(\nu_j - l\pi) \quad (3.35)$$

$$= \sum_j (\nu_j - l\pi)^2 \quad (3.36)$$

$$= mV. \quad (3.37)$$

Thus

$$\begin{aligned} \text{Var}(T | \phi) &= \text{tr } \phi \phi' \Sigma \\ &= \frac{\mu_S(1 - \mu_S)}{\pi(1 - \pi)} \left(n(1 - \rho_S)\pi(1 - \pi) + m\rho_S V \right). \end{aligned}$$

QED

The formula (3.29) is the same for all three designs, the only difference being the value of V . For the litter matched design $\nu_j = l\pi$ for all j and so

$$V \equiv 0. \quad (3.38)$$

Thus

$$\text{Var}(T) = E(\text{Var}(T | \phi)) = n\mu_S(1 - \mu_S)(1 - \rho_S). \quad (3.39)$$

For the nested design $\nu_j = l$ for a fraction π of the litters and $\nu_j = 0$ for the remainder. Thus

$$V \equiv l^2\pi(1 - \pi) \quad (3.40)$$

and we have

$$\begin{aligned} \text{Var}(T) &= E(\text{Var}(T | \phi)) \\ &= \mu_S(1 - \mu_S) [n(1 - \rho_S) + m\rho_S l^2] \\ &= n\mu_S(1 - \mu_S) [1 + (l - 1)\rho_S], \end{aligned} \quad (3.41)$$

recalling that $ml = n$. For the completely randomized design, V is a random variable. Its expected value is given by the following theorem.

Theorem 3.2 *For the completely randomized design, the expected value of V is*

$$\frac{n(m-1)l\pi(1-\pi)}{m(n-1)} \quad (3.42)$$

Proof: V can be written as a quadratic form in the vector of ν_j 's.

$$V = \frac{1}{m} \nu' (\mathbf{I}_m - \frac{1}{m} \mathbf{J}_m) \nu \quad (3.43)$$

from which

$$\begin{aligned} E(V) &= \frac{1}{m} E(\nu') \left(\mathbf{I}_m - \frac{1}{m} \mathbf{J}_m \right) E(\nu) \\ &\quad + \frac{1}{m} \text{tr} \left(\mathbf{I}_m - \frac{1}{m} \mathbf{J}_m \right) \text{Cov}(\nu) \end{aligned} \quad (3.44)$$

Now $\nu = (\mathbf{1}'_l \otimes \mathbf{I}_m) \mathbf{a}$ so that

$$E(\nu) = \pi (\mathbf{1}'_l \otimes \mathbf{I}_m) (\mathbf{1}_l \otimes \mathbf{1}_m) = \pi \mathbf{1}_m \quad (3.45)$$

and

$$\begin{aligned} \text{Cov}(\nu) &= (\mathbf{1}' \otimes \mathbf{I})(\text{Cov}(\mathbf{a}))(\mathbf{1} \otimes \mathbf{I}) \\ &= \frac{\pi(1-\pi)}{n-1} (\mathbf{1}' \otimes \mathbf{I})(n\mathbf{I} \otimes \mathbf{I} - \mathbf{J} \otimes \mathbf{J})(\mathbf{1} \otimes \mathbf{I}) \\ &= \frac{\pi(1-\pi)}{n-1} (nl\mathbf{I}_m - l^2\mathbf{J}_m) \\ &= \frac{nl\pi(1-\pi)}{n-1} (\mathbf{I}_m - m^{-1}\mathbf{J}_m). \end{aligned} \quad (3.46)$$

Clearly, the first term in (3.44) vanishes, and the second term becomes

$$\frac{n(m-1)l\pi(1-\pi)}{m(n-1)}.$$

QED

Substituting (3.42) into (3.29) gives

$$\begin{aligned} \text{Var}(T) &= E(\text{Var}(T | \phi)) \\ &= \mu_S(1-\mu_S) \left[n(1-\rho_S) + \frac{nl(m-1)}{n-1} \rho_S \right] \\ &= n\mu_S(1-\mu_S) \left(1 - \frac{l-1}{n-1} \rho_S \right) \end{aligned} \quad (3.47)$$

for the completely randomized design.

The variance of T under H_0 for any design can be written as $n\mu_S(1 - \mu_S)/\epsilon$ where ϵ depends on the design. We then have

$$\epsilon_{CR} = \frac{n-1}{n-1-(l-1)\rho_S}, \quad (3.48)$$

$$\epsilon_{LM} = \frac{1}{1-\rho_S}, \quad (3.49)$$

$$\text{and } \epsilon_{NE} = \frac{1}{1+(l-1)\rho_S} \quad (3.50)$$

for the completely randomized, litter-matched, and nested designs, respectively. Clearly $\epsilon_{NE} \leq \epsilon_{CR} \leq \epsilon_{LM}$, and $\epsilon_{CR} \rightarrow 1$ as $n \rightarrow \infty$.

3.3.3 Variance—alternative case

To determine the variance under the alternative, we again condition on the design. The development is complicated here because the conditional variance itself depends on the design. Thus instead of $\text{Var}(T | \phi) = \text{tr } \phi \phi' \Sigma$ we have

$$\text{Var}(T | \phi) = \text{tr}(\phi \phi' \text{Cov}(Y | \phi)). \quad (3.51)$$

Using $*$ to represent the componentwise product (see Appendix A), we can write the vector of Y_{ij} 's as

$$\begin{aligned} Y &= X * (1 + a * (Z - 1)) \\ &= X * (1 + \Delta(Z - 1)) \end{aligned} \quad (3.52)$$

where Δ is a diagonal matrix whose diagonal elements are the a_{ij} 's. Because X and Z are independent, we can write

$$\begin{aligned} E(Y | \Delta) &= E(X) * (1 + \Delta(E(Z) - 1)) \\ &= \mu_S \mathbf{1} * (1 + (\mu_I - 1)\Delta \mathbf{1}) \\ &= \mu_S(\mathbf{I} + (\mu_I - 1)\Delta)\mathbf{1}. \end{aligned} \quad (3.53)$$

To find the conditional variance of Y we need the following lemma.

Lemma 3.1 *If X and Y are two (stochastically) independent random vectors of the same dimension, then*

$$\begin{aligned} \text{Cov}(X * Y) &= \text{Cov}(X) * \text{Cov}(Y) \\ &\quad + (E(X)E(X')) * \text{Cov}(Y) \\ &\quad + (E(Y)E(Y')) * \text{Cov}(X). \end{aligned} \quad (3.54)$$

Proof: The covariance matrix of $X * Y$ is

$$\begin{aligned} &E((X * Y)(X * Y)') - E(X * Y)E(X * Y)' \\ &= E((XX')(YY')) - (E(X)E(X')) * (E(Y)E(Y)') \end{aligned}$$

$$\begin{aligned}
&= (\text{Cov}(\mathbf{X}) + \mathbf{E}(\mathbf{X})\mathbf{E}(\mathbf{X})') * (\text{Cov}(\mathbf{Y}) + \mathbf{E}(\mathbf{Y})\mathbf{E}(\mathbf{Y})') \\
&\quad - (\mathbf{E}(\mathbf{X})\mathbf{E}(\mathbf{X})') * (\mathbf{E}(\mathbf{Y})\mathbf{E}(\mathbf{Y})') \\
&= \text{Cov}(\mathbf{X}) * \text{Cov}(\mathbf{Y}) + (\mathbf{E}(\mathbf{X})\mathbf{E}(\mathbf{X})') * \text{Cov}(\mathbf{Y}) \\
&\quad + (\mathbf{E}(\mathbf{Y})\mathbf{E}(\mathbf{Y})') * \text{Cov}(\mathbf{X}).
\end{aligned}$$

QED

Now, let us denote $\text{Cov}(\mathbf{X})$ and $\text{Cov}(\mathbf{Z})$ by Σ_S and Σ_I respectively. We then have the following theorem.

Theorem 3.3 *The conditional covariance matrix of $\mathbf{Y}$ given the design $\mathbf{a}$ is given by*

$$\begin{aligned}
\text{Cov}(\mathbf{Y} | \mathbf{a}) &= \text{Cov}(\mathbf{Y} | \Delta) \\
&= \Delta((\Sigma_S + \mu_S^2 \mathbf{J}) * \Sigma_I) \Delta \\
&\quad + (\mathbf{I} - (\mu_S - 1)\Delta) \Sigma_S (\mathbf{I} - (\mu_S - 1)\Delta).
\end{aligned}$$

Proof: $\mathbf{E}(\mathbf{X}) = \mu_S \mathbf{1}$, so $\mathbf{E}(\mathbf{X})\mathbf{E}(\mathbf{X})' = \mu_S^2 \mathbf{J}$. Also, since $\mathbf{E}(\mathbf{Z}) = \mu_I \mathbf{1}$ we have

$$\mathbf{E}(\mathbf{1} + \Delta(\mathbf{Z} - \mathbf{1}) | \mathbf{a}) = (\mathbf{I} + (\mu_I - 1)\Delta) \mathbf{1}. \quad (3.55)$$

Hence

$$\begin{aligned}
&\mathbf{E}(\mathbf{1} + \Delta(\mathbf{Z} - \mathbf{1}) | \Delta) \mathbf{E}(\mathbf{1} + \Delta(\mathbf{Z} - \mathbf{1}) | \Delta)' \\
&= (\mathbf{I} + (\mu_I - 1)\Delta) \mathbf{1} \mathbf{1}' (\mathbf{I} + (\mu_I - 1)\Delta) \\
&= (\mathbf{I} + (\mu_I - 1)\Delta) \mathbf{J} (\mathbf{I} + (\mu_I - 1)\Delta) \\
&= \mathbf{J} + (\mu_I - 1)\Delta \mathbf{J} + (\mu_I - 1)\mathbf{J} \Delta + (\mu_I - 1)^2 \Delta \mathbf{J} \Delta.
\end{aligned} \quad (3.56)$$

Moreover, $\text{Cov}(\mathbf{1} + \Delta(\mathbf{Z} - \mathbf{1}) | \Delta) = \Delta \text{Cov}(\mathbf{Z}) \Delta = \Delta \Sigma_I \Delta$. Hence

$$\begin{aligned}
&\text{Cov}(\mathbf{X} * (\mathbf{1} + \Delta(\mathbf{Z} - \mathbf{1})) | \Delta) \\
&= \Sigma_S * (\Delta \Sigma_I \Delta) + \mu_S^2 \mathbf{J} * (\Delta \Sigma_I \Delta) \\
&\quad + (\mathbf{J} + (\mu_I - 1)\Delta \mathbf{J} + (\mu_I - 1)\mathbf{J} \Delta + (\mu_I - 1)^2 \Delta \mathbf{J} \Delta) * \Sigma_S.
\end{aligned} \quad (3.57)$$

Now using the properties of the componentwise product summarized in Appendix A we have that $\Sigma_S * (\Delta \Sigma_I \Delta) = \Delta(\Sigma_S * (\Sigma_I \Delta)) = \Delta((\Sigma_I \Delta) * \Sigma_S) = \Delta(\Sigma_I * \Sigma_S) \Delta$. Similarly we get that $(\Delta \mathbf{J}) * \Sigma_S = \Sigma_S * (\Delta \mathbf{J}) = \Delta(\Sigma_S * \mathbf{J}) = \Delta \Sigma_S$, $\mathbf{J} \Delta \Sigma_S = \Sigma_S \Delta$, and $\Delta \mathbf{J} \Delta * \Sigma_S = \Delta \Sigma_S \Delta$. Thus

$$\begin{aligned}
&\text{Cov}(\mathbf{X} * (\mathbf{1} + \Delta(\mathbf{Z} - \mathbf{1})) | \mathbf{a}) \\
&= \Delta((\Sigma_S + \mu_S^2 \mathbf{J}) * \Sigma_I) \Delta + \Sigma_S + (\mu_I - 1)\Delta \Sigma_S \\
&\quad + (\mu_I - 1)\Sigma_S \Delta + (\mu_I - 1)^2 \Delta \Sigma_S \Delta \\
&= \Delta((\Sigma_S + \mu_S^2 \mathbf{J}) * \Sigma_I) \Delta \\
&\quad + (\mathbf{I} - (\mu_S - 1)\Delta) \Sigma_S (\mathbf{I} - (\mu_S - 1)\Delta).
\end{aligned}$$

QED

Theorem 3.4 The conditional variance of T given the design $\mathbf{a}$ is given by

$$\begin{aligned} \text{Var}(T | \mathbf{a}) = & m\mu_S \left[(1-\pi)\mu_I(1-\mu_I) \right. \\ & + (1-\pi)\mu_I(1-\mu_I)(1-\pi)(\mu_S + (1-\mu_S)\rho_S)\rho_I \\ & + (\pi + \mu_I^2(1-\pi))(1-\rho_S)(1-\mu_S) + \pi(1-\pi)(1-\mu_I)^2(1-\mu_S)\rho_S \left. \right] \\ & + \left[m\mu_S(1-\mu_S)(\pi + \mu_I - \pi\mu_I)^2\rho_S \right. \\ & \left. + (1-\pi)^2 m\mu_S\mu_I(1-\mu_I)(\mu_S + (1-\mu_S)\rho_S)\rho_I \right] \frac{V}{\pi(1-\pi)} \end{aligned} \quad (3.58)$$

Proof: Since $T = \phi'Y$, we have

$$\begin{aligned} \text{Var}(T | \mathbf{a}) &= \phi' \text{Cov}(Y | \mathbf{a}) \phi \\ &= \phi' \Delta ((\Sigma_S + \mu_S^2 J) * \Sigma_I) \Delta \phi \\ &+ \phi' (I - (\mu_S - 1)\Delta) \Sigma_S (I - (\mu_S - 1)\Delta) \phi \end{aligned} \quad (3.59)$$

by Theorem 3.3. To evaluate the terms in (3.59) first note that the ij^{th} element of $\Delta \phi$ is equal to

$$a_{ij} \frac{\pi - a_{ij}}{\sqrt{\pi(1-\pi)}} = -\sqrt{\frac{1-\pi}{\pi}} a_{ij},$$

so that we can write

$$\Delta \phi = -\sqrt{(1-\pi)/\pi} \mathbf{a}. \quad (3.60)$$

This gives

$$\begin{aligned} & \phi' \Delta ((\Sigma_S + \mu_S^2 J) * \Sigma_I) \Delta \phi \\ &= \text{tr} \Delta \phi \phi' \Delta ((\Sigma_S + \mu_S^2 J) * \Sigma_I) \\ &= \frac{1-\pi}{\pi} \text{tr} ((\mathbf{a}\mathbf{a}')((\Sigma_S + \mu_S^2 J) * \Sigma_I)) \end{aligned} \quad (3.61)$$

for the first term of (3.59). Similarly, the ij^{th} element of $(I + (\mu_I - 1)\Delta)\phi$ is

$$\frac{\pi - a_{ij}}{\sqrt{\pi(1-\pi)}} + \frac{(\mu_I - 1)(\pi - 1)}{\sqrt{\pi(1-\pi)}} a_{ij} = \frac{\pi - (1 - (1-\pi)(1-\mu_I))}{\sqrt{\pi(1-\pi)}} a_{ij} \quad (3.62)$$

so that

$$(I + (\mu_I - 1)\Delta)\phi = \sqrt{\frac{\pi}{1-\pi}} \mathbf{1} - \frac{\pi + \mu_I - \pi\mu_I}{\sqrt{\pi(1-\pi)}} \mathbf{a},$$

and the second term of (3.59) becomes

$$\begin{aligned} & \phi' (I + (\mu_I - 1)\Delta) \Sigma_S (I + (\mu_I - 1)\Delta) \phi \\ &= \text{tr} \left(\left(\sqrt{\frac{\pi}{1-\pi}} \mathbf{1} - \frac{\pi + \mu_I - \pi\mu_I}{\sqrt{\pi(1-\pi)}} \mathbf{a} \right) \right. \\ & \quad \times \left. \left(\sqrt{\frac{\pi}{1-\pi}} \mathbf{1} - \frac{\pi + \mu_I - \pi\mu_I}{\sqrt{\pi(1-\pi)}} \mathbf{a} \right)' \Sigma_S \right) \end{aligned}$$

$$= \text{tr} \left[\frac{\pi}{1-\pi} \mathbf{J} - 2 \frac{\pi + \mu_I - \pi \mu_I}{1-\pi} \mathbf{a} \mathbf{1}' + \frac{(\pi + \mu_I - \pi \mu_I)^2}{\pi(1-\pi)} \mathbf{a} \mathbf{a}' \right] \Sigma_S. \quad (3.63)$$

We complete the derivation by substituting the values of Σ_S from (3.3) and Σ_I from (3.4) into (3.61) and (3.63).

$$\Sigma_S = \mu_S(1 - \mu_S)((1 - \rho_S)\mathbf{I}_l + \rho_S \mathbf{J}_l) \otimes \mathbf{I}_m, \quad (3.64)$$

$$\Sigma_I = \mu_I(1 - \mu_I)((1 - \rho_I)\mathbf{I}_l + \rho_I \mathbf{J}_l) \otimes \mathbf{I}_m. \quad (3.65)$$

To evaluate the first term of (3.59), we note that since

$$\Sigma_S * \Sigma_I = \mu_S \mu_I (1 - \mu_S)(1 - \mu_I) \times ((1 - \rho_S \rho_I)\mathbf{I}_l + \rho_S \rho_I \mathbf{J}_l) \otimes \mathbf{I}_m, \quad (3.66)$$

and of course

$$\begin{aligned} \mu_S^2 \mathbf{J} * \Sigma_I &= \mu_S^2 \Sigma_I \\ &= \mu_S^2 \mu_I (1 - \mu_I)((1 - \rho_I)\mathbf{I}_l + \rho_I \mathbf{J}_l) \otimes \mathbf{I}_m, \end{aligned} \quad (3.67)$$

$$\begin{aligned} (\Sigma_S + \mu_S^2 \mathbf{J}) * \Sigma_I &= \mu_S \mu_I (1 - \mu_I) \left[(1 - \mu_S)(1 - \rho_S \rho_I)\mathbf{I}_l + (1 - \mu_S)\rho_S \rho_I \mathbf{J}_l \right. \\ &\quad \left. + \mu_S(1 - \rho_I)\mathbf{I}_l + \mu_S \rho_I \mathbf{J}_l \right] \otimes \mathbf{I}_m \end{aligned} \quad (3.68)$$

$$\begin{aligned} &= \mu_S \mu_I (1 - \mu_I) \left[(1 - (\mu_S + (1 - \mu_S)\rho_S)\rho_I)\mathbf{I}_l \right. \\ &\quad \left. + (\mu_S + (1 - \mu_S)\rho_S)\rho_I \mathbf{J}_l \right] \otimes \mathbf{I}_m. \end{aligned} \quad (3.69)$$

Clearly, $\text{tr}(\mathbf{a} \mathbf{a}'(\mathbf{I}_l \otimes \mathbf{I}_m)) = \sum_j \nu_j = m l \pi$. To evaluate $\text{tr}(\mathbf{a} \mathbf{a}'(\mathbf{J}_l \otimes \mathbf{I}_m))$, write

$$\mathbf{a} \mathbf{a}'(\mathbf{J}_l \otimes \mathbf{I}_m) = \begin{bmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} & \dots & \mathbf{A}_{1m} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{A}_{m1} & \mathbf{A}_{m2} & \dots & \mathbf{A}_{mm} \end{bmatrix} (\mathbf{J}_l \otimes \mathbf{I}_m) \quad (3.70)$$

using the notation of (3.31). Now $\mathbf{A}_{jj} \mathbf{J}_l = \mathbf{a}_j \mathbf{a}_j' \mathbf{1} \mathbf{1}' = \nu_j \mathbf{a}_j \mathbf{1}'$, so $\text{tr} \mathbf{A}_{jj} \mathbf{J}_l = \nu_j^2$, giving

$$\text{tr}(\mathbf{a} \mathbf{a}'(\mathbf{J}_l \otimes \mathbf{I}_m)) = \sum_j \nu_j^2 = mV + m l^2 \pi^2. \quad (3.71)$$

Combining this result with (3.68) in (3.61) gives

$$\begin{aligned} \text{tr} \left[\mathbf{a} \mathbf{a}' (\Sigma_S + \mu_S^2 \mathbf{J}) * \Sigma_I \right] &= \mu_S \mu_I (1 - \mu_I) \left[(1 - (\mu_S + (1 - \mu_S)\rho_S)\rho_I) m l \pi \right. \\ &\quad \left. + (\mu_S + (1 - \mu_S)\rho_S)\rho_I (m l^2 \pi^2 + mV) \right] \end{aligned} \quad (3.72)$$

$$+ (\mu_S + (1 - \mu_S)\rho_S)\rho_I (m l^2 \pi^2 + mV) \quad (3.73)$$

and we get

$$(1 - \pi)ml\mu_S\mu_I(1 - \mu_I)\left[1 + (l\pi - 1)(\mu_S + (1 - \mu_S)\rho_S)\rho_I\right] \\ (1 + \pi^{-1}(1 - \pi)m\mu_S\mu_I(1 - \mu_I)(\mu_S + (1 - \mu_S)\rho_S)\rho_IV \quad (3.74)$$

for the first term of (3.59).

To get the second term of (3.59), we have to simplify (3.63). Now

$$\begin{aligned} J\Sigma_S &= (J_l \otimes J_m)\Sigma_S \\ &= \mu_S(1 - \mu_S)((1 - \rho_S)J_l + l\rho_S J_l) \otimes J_m \\ &= \mu_S(1 - \mu_S)(1 + (l - 1)\rho_S)J_l \otimes J_m \end{aligned}$$

and so

$$\text{tr} \frac{\pi}{1 - \pi} J\Sigma_S = ml\mu_S(1 - \mu_S)(1 + (l - 1)\rho_S). \quad (3.75)$$

To simplify $a1'((1 - \rho_S)I_l + \rho_S J_l) \otimes I_m$, consider $a1'$ as a partitioned matrix

$$\begin{bmatrix} a_{11}' & a_{12}' & \dots & a_{1l}' \\ a_{21}' & a_{22}' & \dots & a_{2l}' \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1}' & a_{m2}' & \dots & a_{ml}' \end{bmatrix}$$

where each piece of the partition is $l \times l$. Now, for $j = 1, \dots, m$

$$a_{j1}'((1 - \rho_S)I_l + \rho_S J_l) = (1 + (l - 1)\rho_S)a_{j1}' \quad (3.76)$$

giving

$$a1'((1 - \rho_S)I_l + \rho_S J_l) \otimes I_m = (1 + (l - 1)\rho_S)a1', \quad (3.77)$$

from which

$$\begin{aligned} 2 \text{tr} \frac{\pi + \mu_I - \pi\mu_I}{1 - \pi} a1'\Sigma_S \\ &= 2 \frac{\pi + \mu_I - \pi\mu_I}{1 - \pi} \mu_S(1 - \mu_S)(1 + (l - 1)\rho_S) \text{tr} a1' \\ &= 2 \frac{(\pi + \mu_I - \pi\mu_I)ml\mu_S(1 - \mu_S)(1 + (l - 1)\rho_S)}{1 - \pi} \end{aligned} \quad (3.78)$$

Now finally, to simplify the third term in (3.63), note that

$$\begin{aligned} \text{tr} \left(\frac{(\pi + \mu_I - \pi\mu_I)^2}{\pi(1 - \pi)} aa'\Sigma_S \right) \\ &= \frac{(\pi + \mu_I - \pi\mu_I)^2}{\pi(1 - \pi)} \mu_S(1 - \mu_S) \\ &\quad \times \left[(1 - \rho_S) \text{tr}(aa'I_l \otimes I_m) + \rho_S \text{tr}(aa'J_l \otimes I_m) \right] \\ &= \frac{(\pi + \mu_I - \pi\mu_I)^2}{\pi(1 - \pi)} \mu_S(1 - \mu_S) \left[(1 - \rho_S)ml\pi + m\rho_S(V + l^2\pi^2) \right] \\ &= \frac{(\pi + \mu_I - \pi\mu_I)^2}{\pi(1 - \pi)} m\mu_S(1 - \mu_S) \left[l\pi(1 + (l\pi - 1)\rho_S) + \rho_S V \right] \end{aligned} \quad (3.79)$$

using (3.71). Inserting the results of (3.75), (3.78), and (3.79) into (3.63) gives

$$\begin{aligned}
 & \phi'(\mathbf{I} - (\mu_S - 1)\Delta)\Sigma_S(\mathbf{I} - (\mu_S - 1)\Delta)\phi \\
 &= \frac{m\mu_S(1 - \mu_S)}{1 - \pi} \left[\pi(1 + (l - 1)\rho_S) \right. \\
 &\quad - 2\pi(\pi + \mu_I - \pi\mu_I)(1 + (l - 1)\rho_S) \\
 &\quad \left. + \frac{(\pi + \mu_I - \pi\mu_I)^2}{\pi} (\pi(1 + (l\pi - 1)\rho_S) + l^{-1}\rho_S V) \right] \\
 &= m\mu_S(1 - \mu_S) \left[(\pi + \mu_I^2(1 - \pi))(1 - \rho_S) \right. \\
 &\quad \left. + \pi(1 - \pi)(1 - \mu_I)^2 l \rho_S \right] \\
 &\quad + \frac{m\mu_S(1 - \mu_S)(\pi + \mu_I - \pi\mu_I)^2 \rho_S}{\pi(1 - \pi)} V
 \end{aligned} \tag{3.80}$$

for the second term of (3.59).

Combining the components (3.74) and (3.80) gives us (3.58). QED

The unconditional variance of T under the litter matched, nested, or completely randomized designs can be evaluated by replacing V by its expected value (3.38), (3.40), or (3.42), respectively. However, the resulting expressions do not seem to simplify.

3.4 Asymptotic Distribution of the Test Statistic

We now prove the asymptotic normality of $T/\sqrt{\text{Var}(T)}$ under H_0 . We will derive this result by first considering the conditional distribution of T for a given value of ϕ , and then showing that under certain conditions the unconditional distribution is also asymptotically normal. Finally, we present consistent estimators of $\text{Var}(T)$ for the three designs considered.

3.4.1 Random standardization

The statistic T can be written as a sum: $T = \sum_{j=1}^m t_j$ where $t_j = \sum_{i=1}^l \phi_{ij} Y_{ij}$. Since the litters are assumed independent, the t_j 's will also be independent conditionally on a given design a . We will prove a conditional central limit theorem for the t_j 's, for which we will need the (conditional) mean and variance.

Theorem 3.5 *For a given design ϕ , and under H_0*

$$E(t_j) = \frac{\mu_S(l\pi - \nu_j)}{\sqrt{\pi(1 - \pi)}} \tag{3.81}$$

and

$$\text{Var}(t_j) = \frac{\mu_S(1 - \mu_S)}{\pi(1 - \pi)} \left((1 - \rho_S)(l\pi^2 + (1 - 2\pi)\nu_j) + \rho_S(l\pi - \nu_j)^2 \right) \tag{3.82}$$

where ν_j is the number of test animals in litter j .

Proof: Since

$$t_j = \frac{1}{\sqrt{\pi(1-\pi)}} \sum_i (\pi - a_{ij}) Y_{ij},$$

$$E(t_j | \mathbf{a}) = \frac{1}{\sqrt{\pi(1-\pi)}} \sum_i (\pi - a_{ij}) E(Y_{ij}) = \frac{\mu_S(l\pi - \nu_j)}{\sqrt{\pi(1-\pi)}},$$

and

$$\begin{aligned} \text{Var}(t_j | \mathbf{a}) &= \frac{1}{\pi(1-\pi)} (\pi \mathbf{1} - \mathbf{a})' \text{Cov}(\mathbf{Y}) (\pi \mathbf{1} - \mathbf{a}) \\ &= \frac{\mu_S(1-\mu_S)}{\pi(1-\pi)} \left[(1-\rho_S) \sum_i (\pi - a_{ij})^2 \right. \\ &\quad \left. + \rho_S \sum_i \sum_{i'} (\pi - a_{ij})(\pi - a_{i'j}) \right]. \end{aligned}$$

But

$$\begin{aligned} \sum_i (\pi - a_{ij})^2 &= \sum_{i=1}^l (\pi^2 - 2\pi a_{ij} + a_{ij}^2) \\ &= l\pi^2 - 2\pi\nu_j + \nu_j \\ &= l\pi^2 + (1-2\pi)\nu_j, \end{aligned}$$

and

$$\sum_i (\pi - a_{ij}) = (l\pi - \nu_j).$$

This completes the proof.

QED

Theorem 3.6 Consider a sequence of experiments with designs

$$\mathbf{a}_m, m = 1, 2, \dots,$$

and responses

$$Y_{ij}^{(m)}, i = 1, \dots, l; j = 1, \dots, m.$$

Let $F_m^{\mathbf{a}}$ be the conditional distribution of T_m/S_m under H_0 , where

$$T_m = \phi_m' \mathbf{Y}^{(m)},$$

$$S_m^2 = \text{Var}(T_m | \mathbf{a}_m) = \frac{\mu_S(1-\mu_S)}{\pi(1-\pi)} (n(1-\rho_S)\pi(1-\pi) + m\rho_S V_m),$$

and

$$V_m = m^{-1} \sum_{j=1}^m (\nu_j - l\pi)^2.$$

Then $F_m^{\mathbf{a}}$ converges weakly to the standard normal distribution.

Proof: Let $U_{jm} = t_j - \mu_S(l\pi - \nu_j)/\sqrt{\pi(1-\pi)}$. Then $E(U_{jm}) = 0$, and

$$\begin{aligned}\text{Var}(U_{jm}) &= \text{Var}(t_j) \\ &= \frac{\mu_S(1-\mu_S)}{\pi(1-\pi)} \left[(1-\rho_S)(l\pi^2 + (1-2\pi)\nu_j) + \rho_S(l\pi - \nu_j)^2 \right].\end{aligned}$$

Moreover, $\sum_j U_{jm} = T_m$ since $\sum \nu_j = n\pi = ml\pi$. Also $\text{Var}(\sum_j U_{jm}) = \sum_j \text{Var}(t_j) = \text{Var}(T_m | \mathbf{a}) = S_m^2$. Thus $E(U_{jm}/S_m) = 0$, $\sum_j \text{Var}(U_{jm}/S_m) = 1$, and the U_{jm}/S_m are independent. Moreover, the sequence U_{jm} is uniformly bounded since, by the Cauchy-Schwarz inequality

$$\begin{aligned}t_j^2 &\leq \frac{\sum_i (\pi - a_{ij})^2 \sum_i Y_{ij}^2}{\pi(1-\pi)} \\ &\leq \frac{(l\pi^2 + (1-2\pi)\nu_j)l}{\pi(1-\pi)} \\ &\leq \frac{l^2}{\pi(1-\pi)}\end{aligned}$$

so $|t_j| \leq l/\sqrt{\pi(1-\pi)}$, and $|\mu_S(l\pi - \nu_j)| \leq \mu_S l$. Thus

$$U_{jm} \leq \frac{l(1+\mu_S)}{\sqrt{\pi(1-\pi)}}.$$

However, $S_m \rightarrow \infty$ as $m \rightarrow \infty$, thus $\max_j U_{jm}/S_m \rightarrow 0$ as $m \rightarrow \infty$. This implies that the Liapounov condition of the central limit theorem is satisfied [18, Theorem 7.1.2] and it follows that $F_m^{\mathbf{a}}$ converges weakly to the standard normal distribution. QED

Theorem 3.7 *The unconditional distribution of T_m/S_m under H_0 also converges weakly to the standard normal distribution.*

Proof: We know that $F_m^{\mathbf{a}}(x) = \Pr\{T_m/S_m \leq x | \mathbf{a}\} \rightarrow \Phi(x)$ for every x . Let $G_m(\mathbf{a})$ be the probability of the (random) design $\mathbf{a}$ under some randomization scheme. Then the unconditional distribution can be written

$$F_m(x) = \Pr\left\{\frac{T_m}{S_m} \leq x\right\} = \sum_{\mathbf{a}} \Pr\left\{\frac{T_m}{S_m} \leq x \mid \mathbf{a}\right\} G_m(\mathbf{a}) = \sum_{\mathbf{a}} F_m^{\mathbf{a}}(x) G_m(\mathbf{a}),$$

and thus

$$\begin{aligned}|F_m(x) - \Phi(x)| &= \left| \sum_{\mathbf{a}} (F_m^{\mathbf{a}}(x) - \Phi(x)) G_m(\mathbf{a}) \right| \\ &\leq \sum_{\mathbf{a}} |F_m^{\mathbf{a}}(x) - \Phi(x)| G_m(\mathbf{a}) \\ &\leq \sup_{\mathbf{a}} |F_m^{\mathbf{a}}(x) - \Phi(x)|.\end{aligned}$$

Now, for each m , let $\mathbf{a}^m$ be that design for which $|F_m^{\mathbf{a}}(x) - \Phi(x)|$ is a maximum. (Such a design must exist since there are only a finite number of designs for a given m .) Then,

choosing this sequence, of designs in Theorem 3.6 we have that for any $\epsilon > 0$ there is an M such that

$$\max_x |F_m^a(x) - \Phi(x)| < \epsilon \quad \text{for } m > M.$$

Thus $|F(x) - \Phi(x)| < \epsilon$ for $m > M$, and the theorem is proved. QED

3.4.2 Constant standardization

Theorem 3.7 gives the asymptotic distribution of T standardized by a (possibly) random, design-dependent variance. In the next theorem we present a sufficient condition under which the asymptotic distribution is still standard normal even when T is standardized by a constant.

Theorem 3.8. If $\text{Var}(V_m) \rightarrow 0$ as $m \rightarrow \infty$ then

$$\frac{T_m}{\sqrt{\text{Var}(T_m)}} = \frac{T_m}{\sqrt{E(S_m^2)}} \rightarrow \Phi \quad \text{in distribution.}$$

Proof: Since S_m^2/m is a linear function of V_m ,

$$\frac{S_m^2}{E(S_m^2)} \xrightarrow{\text{pr}} 1 \quad \text{if } V_m \xrightarrow{\text{pr}} E(V_m).$$

This in turn will follow (by Chebyshev's inequality) if $\text{Var}(V_m) \rightarrow 0$. In that case we have

$$\frac{T_m}{\sqrt{E(S_m^2)}} = \frac{T_m}{S_m} \sqrt{\frac{S_m^2}{E(S_m^2)}} \xrightarrow{\text{dist}} \Phi \quad \text{by Slutsky's theorem.}$$

QED

For the litter-matched and nested designs, $\text{Var}(V_m) = 0$ for all m , so the theorem is automatically satisfied. We now show that it also holds for the completely randomized design. For this we will require the variance of V_m , a quadratic function of the design variables.

The variance of a quadratic form $W'QW$ in a random vector W is

$$E((W'QW)^2) - (E(W'QW))^2. \quad (3.83)$$

The first term in (3.83) can be written

$$E(W'QWW'QW) = E((W' \otimes W')Q \otimes Q(W \otimes W)). \quad (3.84)$$

Using the summation convention and the definitions described in Appendix B, we can write this expression as

$$E(W^i W^j W^k W^l q_{ij} q_{kl}) \quad (3.85)$$

which by (B.8) is equal to

$$\delta_{ij}^{kl} h^{uv} \gamma_{uv} q_{ij} q_{kl} \quad (3.86)$$

where δ_{ω}^{ijkl} , h_{ω}^{ij} , and γ_{ω}^{ij} are defined by (B.7), (B.4), and (B.1), and the q_{ij} 's are the components of Q . (The symbols i, j, k, l are used here merely as "dummy" indices, and of course, bear no relationship to the i, j, k, l used elsewhere in this chapter.) Thus, to find the variance of V , we require the matrix of the quadratic form, and the fourth order mixed moments. These are provided by the following two lemmas.

Lemma 3.2 V can be written as a quadratic form in a or in ν as

$$V = m^{-1} \nu' Q \nu = l m^{-1} a' R a \quad (3.87)$$

where $Q = I_m - m^{-1} J_m$ and $R = l^{-1} J_l \otimes Q$.

Proof: The form of Q follows from the fact that

$$m^{-1} J_m \nu = \left(m^{-1} \sum_{j=1}^m \nu_j \right) \mathbf{1}_m = m^{-1} n \pi \mathbf{1}_m = l \pi \mathbf{1}_m.$$

To find R , we note that $\nu = \mathbf{1}_l' \otimes J_m a$. Thus

$$\begin{aligned} \nu' Q \nu &= a' (\mathbf{1}_l \otimes I_m) (I_m - m^{-1} J_m) (\mathbf{1}_l' \otimes I_m) a \\ &= a' (\mathbf{1}_l \mathbf{1}_l' \otimes (I_m - m^{-1} J_m)) a \\ &= a' (J_l \otimes Q) a \\ &= l a' R a. \end{aligned}$$

QED

Remark: The important feature of the quadratic forms Q and R is that they are symmetric and idempotent, and that they annihilate the constant vector $\mathbf{1}$.

Lemma 3.3 The fourth-order mixed moments of a for the completely randomized design are given by

$$\gamma_{44} = \pi \quad (3.88)$$

$$\gamma_{31} = \frac{\pi(n\pi - 1)}{n - 1} \quad (3.89)$$

$$\gamma_{22} = \frac{\pi(n\pi - 1)}{n - 1} \quad (3.90)$$

$$\gamma_{211} = \frac{\pi(n\pi - 1)(n\pi - 2)}{(n - 1)(n - 2)} \quad (3.91)$$

$$\gamma_{1111} = \frac{\pi(n\pi - 1)(n\pi - 2)(n\pi - 3)}{(n - 1)(n - 2)(n - 3)} \quad (3.92)$$

Proof: Since the a_{ij} 's are binary variables, the mixed moments are nothing but the joint selection probabilities. Their form follows from the definition of simple random sampling.

Thus $\gamma_{44} = E(a_{ij}^4) = E(a_{ij}) = \pi$, the probability of selecting animal i in litter j for the test group. Similarly, $\gamma_{31} = E(a_{ij}^3 a_{i'j'}) = E(a_{ij} a_{i'j'}) = E(a_{ij}^2 a_{i'j'}) = \gamma_{22}$ is the

probability of selecting animal i in litter j and animal i' in litter j' . Since after selecting animal (i, j) we have to select $n\pi - 1$ more animals out of $n - 1$, the conditional probability of selecting $a_{i'j'}$ is $(n\pi - 1)/(n - 1)$. Thus

$$\gamma_{31} = \gamma_{22} = \frac{\pi(n\pi - 1)}{n - 1}. \quad (3.93)$$

Using similar arguments, γ_{211} and γ_{1111} can be derived. QED

Theorem 3.9 For the completely randomized design, $\text{Var}(V) \xrightarrow{\text{Pr}} 0$ as $m \rightarrow \infty$.

Proof: The expression (3.87) can be further simplified by exploiting the properties of $\mathbf{R}$. In particular, since $\mathbf{R}\mathbf{1} = \mathbf{R}'\mathbf{1} = \mathbf{0}$, the sum of the elements of any row or column of $\mathbf{R}$ will vanish. Hence (using the summation convention) we have

$$\delta_{ij}^{ijkl} r_{ij} r_{kl} = 0, \quad (3.94)$$

$$\delta_{12}^{ijkl} r_{ij} r_{kl} = r_{ii} \sum_k \sum_l r_{kl} = 0, \quad (3.95)$$

and similarly for δ_{13}^{ijkl} , δ_{14}^{ijkl} , δ_{23}^{ijkl} , δ_{24}^{ijkl} , δ_{34}^{ijkl} ,

$$\delta_{123}^{ijkl} r_{ij} r_{kl} = r_{ii} \sum_l r_{il} = 0, \quad (3.96)$$

and similarly for δ_{124}^{ijkl} , δ_{134}^{ijkl} , δ_{234}^{ijkl} . Thus we only require the terms $\delta_{1234}^{ijkl} r_{ij} r_{kl}$, $\delta_{\{12\}\{34\}}^{ijkl} r_{ij} r_{kl}$, $\delta_{\{13\}\{24\}}^{ijkl} r_{ij} r_{kl}$, and $\delta_{\{14\}\{23\}}^{ijkl} r_{ij} r_{kl}$. Now

$$\delta_{1234}^{ijkl} r_{ij} r_{kl} = \sum_i r_{ii}^2, \quad (3.97)$$

$$\delta_{\{12\}\{34\}}^{ijkl} r_{ij} r_{kl} = \sum_i r_{ii} \sum_k r_{kk} = (\text{tr } \mathbf{R})^2, \quad (3.98)$$

$$\delta_{\{13\}\{24\}}^{ijkl} r_{ij} r_{kl} = \sum_{ij} r_{ij} r_{ij} = \text{tr } \mathbf{R} \mathbf{R}' = \text{tr } \mathbf{R}, \quad (3.99)$$

$$\text{and } \delta_{\{14\}\{23\}}^{ijkl} r_{ij} r_{kl} = \sum_{ij} r_{ij} r_{ji} = \text{tr } \mathbf{R}^2 = \text{tr } \mathbf{R}, \quad (3.100)$$

because $\mathbf{R}$ is idempotent and symmetric. But $\text{tr } \mathbf{R} = m - 1$, and $\sum_i r_{ii}^2 = (m - 1)^2/n$. Thus

$$\mathbf{E}((\mathbf{a}' \mathbf{R} \mathbf{a})^2) = \eta_{\omega} h^{\omega\psi} \gamma_{\psi} \quad (3.101)$$

where $\omega \in \{1234, \{12\}\{34\}, \{13\}\{24\}, \{14\}\{23\}\}$,

$$\eta_{1234} = (m - 1)^2/n, \quad (3.102)$$

$$\eta_{\{12\}\{34\}} = (m - 1)^2, \quad (3.103)$$

$$\eta_{\{13\}\{24\}} = m - 1, \quad (3.104)$$

$$\eta_{\{14\}\{23\}} = m - 1, \quad (3.105)$$

h_{ij} are from the 1st, 6th, 7th, and 8th, rows of H in (B.4), and γ_{ij} is as defined in Lemma 3.3. A routine but tedious calculation then gives

$$\begin{aligned} E(V^2) &= \frac{l^2}{m^2} E((a'Ra)^2) \\ &= \frac{(m-1)l^3\pi(1-\pi)}{(ml-1)(ml-2)(ml-3)} \\ &\quad \times [l\pi(1-\pi)(ml(m+1) - 6(m-1)) - 2(l-1)] \end{aligned} \quad (3.106)$$

From Theorem 3.2 we have

$$\begin{aligned} E(V) &= \frac{n(m-1)l\pi(1-\pi)}{m(n-1)} \\ &= \frac{(m-1)l^2\pi(1-\pi)}{ml-1} \end{aligned} \quad (3.107)$$

Combining (3.106) and (3.107) then gives, after more tedious calculations,

$$\begin{aligned} \text{Var}(V) &= E(V^2) - (E(V))^2 \\ &= 2\pi(1-\pi)l^3(l-1) \\ &\quad \times \frac{(m-1)(ml\pi-1)(ml(1-\pi)-1)}{(ml-1)^2(ml-2)(ml-3)} \end{aligned} \quad (3.108)$$

Since the numerator is $O(m^3)$ but the denominator is $O(m^4)$, we have that $\text{Var}(V) \rightarrow 0$ as $m \rightarrow \infty$. QED

3.4.3 Sample-based standardization

Although these theorems demonstrate the asymptotic normality of T , they do not directly provide a test statistic, since the expression for $\text{Var}(T)$ involves the unknown parameters μ_S and ρ_S . To construct an asymptotic test, we therefore require consistent estimators of $\text{Var}(T)$. These are given by the following theorem.

Theorem 3.10 Let $A_j = \sum_i Y_{ij}$ and $\bar{A} = m^{-1} \sum_j A_j$. Then unbiased and consistent estimators of $m^{-1} \text{Var}(T)$ under H_0 are given by

$$\begin{aligned} s_{CR}^2 &= \frac{m\bar{A}(l-\bar{A})}{ml-1}, \\ s_{LM}^2 &= \frac{1}{m(l-1)} \sum_{j=1}^m A_j(l-A_j), \\ \text{and } s_{NE}^2 &= \frac{1}{m-1} \sum_{j=1}^m (A_j - \bar{A})^2 \end{aligned} \quad (3.109)$$

for the completely randomized, ~~litter~~ matched, and nested designs, respectively.

Proof: Under H_0 , the A_j 's are i.i.d. and independent of the design a . Moreover, $E(A_j) = l\mu_S$ and $\text{Var}(A_j) = l\mu_S(1 - \mu_S)(1 + (l - 1)\rho_S)$. But from (3.41) we see that this latter quantity is just $\text{Var}(T)/m$ for the nested design. Thus the sample variance s_{NE}^2 is an unbiased and consistent estimator of $\text{Var}(A_j) = m^{-1}\text{Var}(T)$ for the nested design.

To establish the estimator for the litter-matched case note that

$$\begin{aligned} E(A_j(l - A_j)) &= lE(A_j) - \text{Var}(A_j) - E(A_j)^2 \\ &= l^2\mu_S - l\mu_S(1 - \mu_S)(1 + (l - 1)\rho_S) - l^2\mu_S^2 \\ &= l\mu_S(1 - \mu_S)(l - 1)(1 - \rho_S), \end{aligned}$$

and hence

$$\begin{aligned} E((l - 1)^{-1}A_j(l - A_j)) &= l\mu_S(1 - \mu_S)(1 - \rho_S) \\ &= m^{-1}\text{Var}(T), \end{aligned} \quad (3.110)$$

using (3.39). By the (weak) law of large numbers,

$$\frac{1}{m(l - 1)} \sum_j A_j(l - A_j) \xrightarrow{\text{pr}} m^{-1}\text{Var}(T) \quad (3.111)$$

and thus s_{LM}^2 is an unbiased, consistent estimator in the litter matched case.

For the completely randomized design, note that $\bar{A} = m^{-1} \sum A_j$, and thus $\bar{A}$ is consistent for $E(A_j) = l\mu_S$. Hence

$$\bar{A}(l - \bar{A}) \xrightarrow{\text{pr}} l^2\mu_S(1 - \mu_S) \quad (3.112)$$

and

$$\frac{m\bar{A}(l - \bar{A})}{ml - 1} \xrightarrow{\text{pr}} l\mu_S(1 - \mu_S). \quad (3.113)$$

However,

$$m^{-1}\text{Var}(T) = l\mu_S(1 - \mu_S) \left(1 - \frac{l - 1}{ml - 1}\rho_S\right) \rightarrow l\mu_S(1 - \mu_S).$$

Hence $m(ml - 1)^{-1}\bar{A}(l - \bar{A}) \xrightarrow{\text{pr}} m^{-1}\text{Var}(T)$. Moreover

$$\begin{aligned} E(\bar{A}(l - \bar{A})) &= l^2\mu_S - \text{Var}(\bar{A}) - (E(\bar{A}))^2 \\ &= l^2\mu_S - l\mu_S(1 - \mu_S)(1 + (l - 1)\rho_S)/m - l^2\mu_S^2 \\ &= m^{-1}l\mu_S(1 - \mu_S)(ml - 1 - (l - 1)\rho_S) \end{aligned}$$

so

$$E\left(\frac{m\bar{A}(l - \bar{A})}{ml - 1}\right) = l\mu_S(1 - \mu_S) \left(1 - \frac{l - 1}{ml - 1}\rho_S\right) = \text{Var}(T), \quad (3.114)$$

by (3.47).

QED

Remark: Note that for the completely randomized and nested designs the divisors in (3.109) could be $m\bar{l}$ and m instead of $ml - 1$ and $m - 1$, and the estimators, although biased, would still be consistent. However, for the litter matched case the divisor must be $l - 1$ just to ensure consistency.

3.4.4 Loose ends

We have only proved the asymptotic normality for the null case. This is sufficient for determining rejection regions for large-sample tests, but in itself does not give a basis for calculating approximate power. It is a reasonable conjecture that asymptotic normality also holds in the alternative case. Presumably, the same argument could be carried out, but the form of (3.58) forebodes that the details could be quite gory.

We would like to call the c 's in (3.48) the *efficiencies* of the three designs. These quantities would in fact be the asymptotic relative efficiencies in the Pitman sense [73] if we could establish that the standardized test statistic $(T - E(T))/\sqrt{\text{Var}(T)}$ converges to the standard normal distribution *uniformly* in some neighbourhood of $H_0 : \mu_I = 1$. We have not attempted to prove this result.

Although the discussion in this chapter has centred on the completely randomized, litter-matched and nested designs, most of the development is not restricted to these three cases. The key design quantity seems to be V , the variance of the number of test animals in each litter. Since this is minimized for the litter matched design, and maximized for the nested design, it would appear that these are, respectively, the best and worst designs for this situation, at least for large samples. This argument, of course, assumes that the "usual" test statistic T is used. It would be interesting to know the efficiency of this *statistic* relative to something like, say, a likelihood ratio test under a beta-binomial model.

Chapter 4

A Class of Discrete Distributions

We will begin this chapter by examining a two-parameter family of discrete probability distributions which includes, as special cases, the hypergeometric, the binomial and the beta-binomial. Since a dense subset of this family can be generated by an urn model proposed by Pólya, we call this the Pólya family. Probability distributions from this family arise in various guises in the analysis of toxicological data. The beta-binomial branch provides a family for modelling intra-litter correlation (Chapter 6) as well as intra-study correlation (Chapter 7). The hypergeometric branch, on the other hand, arises as the distribution of conditional test statistics for permutation tests. The distributions in this family all have the property that simple recursive relations can be defined on their support, which allow for the development of efficient computational algorithms.

The recursive relations also provide a means for studying the shape of the distributions. Criteria can be derived to determine whether the distributions are bell-shaped, J-shaped or U-shaped, and bounds can be placed on the probabilities in the tails. Such information is useful in designing computer algorithms for calculating probabilities or expected values, because they give assurances about the magnitude of truncation errors and indicate where numerical problems might arise. We pay particular attention to distributions near the boundaries of the parameter space, since a naïve application of recursive relations could break down in such cases. Although these are not cases that would typically arise in practice, a general computer implementation which is to be used as part of another algorithm should be able to handle these situations sensibly.

For distributions in the Pólya family, the probability of any point in the support can be expressed as the product of factors chosen from either of two sequences. The properties of these sequences allow one to put bounds on probabilities without necessarily accumulating all the factors. These techniques provide the basis of an efficient algorithm for approximating expected values with a prespecified bound on the error.

The Pólya family can be extended from distributions on sets of non-negative integers to a class of distributions on the integer lattice points of a simplex. This extension can be achieved by considering a Markov chain (of finite length) whose transition probabilities come from the family. Members of this *multivariate Pólya family* arise naturally in significance tests of bioassays involving either clustering, or more than two treatments. The simple recursive relations carry over to the extended family, but the recursions are no longer linearly ordered. Since a *partially ordered* recursion is computationally more in-

volved, its development will be delayed to Chapter 5 in which the appropriate structures are developed.

4.1 The Pólya Family

4.1.1 Definition

For a , b , and n positive integers, the formula

$$P(x; a, b, n) = \frac{\binom{a}{x} \binom{b}{n-x}}{\binom{a+b}{n}} \quad (4.1)$$

defines a probability distribution on integers $x = l, l+1, \dots, u$, where $l = 0 \vee n - b$ and $u = n \wedge a$. Following Feller[28] we can define binomial coefficients for any real number a and any integer r by

$$\binom{a}{r} = \begin{cases} 0 & \text{if } r < 0 \\ 1 & \text{if } r = 0 \\ \prod_{i=0}^{r-1} \frac{a-i}{i+1} & \text{if } r > 0. \end{cases} \quad (4.2)$$

With this extension $P(x; a, b, n)$ is well defined for integers x and $n \geq 0$, and any real numbers a, b provided that: (i) not both a and b are zero, and (ii) $a + b$ is not a positive integer less than n . It is easy to derive the following relationships:

$$\sum_{x=0}^n P(x; a, b, n) = 1, \quad (4.3)$$

$$\sum_{x=0}^n x P(x; a, b, n) = \frac{na}{a+b}, \quad (4.4)$$

$$\sum_{x=0}^n \left(x - \frac{na}{a+b}\right)^2 P(x; a, b, n) = \frac{nab}{(a+b)^2} \left(1 - \frac{n-1}{a+b-1}\right). \quad (4.5)$$

If the function $P(\cdot; a, b, n)$ defines a probability measure on $\{0, 1, \dots, n\}$ then (4.4) and (4.5) define the mean and variance of this distribution. Because of (4.3), P will define such a distribution if and only if (4.1) gives a non-negative number for each integer $x \in \{0, 1, \dots, n\}$. Known properties about moments of probability distributions allow us to use (4.4) and (4.5) to help delineate the set of a, b such that (4.1) does define a probability distribution.

4.1.2 The natural parameter space

Since the expected value of a random variable must lie between its minimum and maximum values, (4.4) implies that

$$0 \leq \frac{na}{a+b} \leq n. \quad (4.6)$$

Also, the variance of a random variable is non-negative, and the variance of a random variable X bounded between 0 and n cannot exceed $nE(X)$. These observations give us

$$0 \leq \frac{nab}{(a+b)^2} \left(1 - \frac{n-1}{a+b-1}\right) \leq \frac{n^2 ab}{(a+b)^2}. \quad (4.7)$$

From this we see that $ab \geq 0$, and if $ab > 0$,

$$-(n-1) \leq \frac{n-1}{a+b-1} \leq 1. \quad (4.8)$$

This relation is obviously satisfied if $a+b < 0$, but if $a+b > 0$ then one of the inequalities will fail unless in fact $a+b \geq n$. These relations thus give necessary conditions for (4.1) to define a probability distribution. Sufficient conditions have been considered by Kemp and Kemp[39] and Sarkadi[71]. Necessary and sufficient conditions are given by the following theorem.

Theorem 4.1 *The expression (4.1) defines a probability distribution on $\{0, 1, \dots, n\}$ if and only if one of the following conditions is met.*

- (a) *Exactly one of a or b is zero.*
- (b) *a and b are both negative.*
- (c) *$a \geq n-1$ and $b \geq n-1$.*
- (d) *$a+b \geq n$ and a and b are both positive integers.*
- (e) *a is a positive integer and $b \geq n-1$.*
- (f) *b is a positive integer and $a \geq n-1$.*

Proof: From (4.1) and (4.2) we have $P(0; 0, b, n) = 1 = P(n; a, 0, n)$ and $P(x; 0, b, n) = 0$ for $x > 0$, and $P(x; a, 0, n) = 0$ for $x < n$. Thus condition (a) gives a degenerate probability distribution concentrated at 0 or n according as a or b , respectively, is zero.

In view of (4.3), it is sufficient to establish that $P(x; a, b, n)$ is non-negative for $0 \leq x \leq n$. If a and b are both negative then the expansion of (4.1) according to (4.2) gives an expression having n negative factors in the numerator and n negative factors in the denominator. Thus the quotient will be positive, and $P(x; a, b, n)$ defines a probability distribution under condition (b). Because of (4.8) we are left with the case where both a and b are positive and $a+b \geq n$, encompassing conditions (c) to (f) above.

If a and b are positive integers, then

$$\binom{a}{x} = \prod_{i=0}^{x-1} \frac{a-i}{i+1}$$

will never be negative, nor will

$$\binom{b}{n-x} = \prod_{i=0}^{n-x-1} \frac{b-i}{i+1}.$$

All the *factors* in those products will be non-negative if $a \geq n-1$ and $b \geq n-1$, whether or not a and b are integers. Since $a+b \geq n$, the denominator $\binom{a+b}{n}$ will be positive. Now,

$$P(x; a, b, n) = \binom{n}{x} \frac{\prod_{i=0}^{x-1} (a-i) \prod_{i=0}^{n-x-1} (b-i)}{\prod_{i=0}^{n-1} (a+b-i)} \quad (4.9)$$

so that the numerator will be non-negative if either a or b is a positive integer, or if either is at least as great as $n-1$, since the smallest factor in the numerator of (4.9) is no less than $a-(n-1)$ or $b-(n-1)$. Thus, any of the above conditions is sufficient for defining a probability distribution. It remains to prove that (at least) one of them is necessary.

Suppose that $a > 0$ and $b \neq 0$, i.e., we are not in cases (a) or (b). Because of (4.7) and (4.8), $b > 0$ and $a+b \geq n$. Thus if a and b are both integers we are in case (c). Suppose now that $a < n-1$, so that we are not in case (f) or (c), and that b is not an integer, so that we are not in case (d) either. We show that in this case $b \geq n-1$ and a must be an integer so that we are in case (e).

Let r and s be integers such that $r \leq a < r+1$ and $s < b < s+1$. By hypothesis, $0 < r \leq n-2$. Consider now the factors $(a-i)$, where i goes from 0 to $x-1$. These will be non-negative for $i \leq r$ and negative for $i > r$. Thus the product $\prod_{i=0}^{x-1} (a-i)$ will be non-negative if $x \leq r+1$. If a is an integer then $\prod_{i=0}^{x-1} (a-i)$ will be zero for $x \geq r+1$, and if a is not an integer this product will alternate in sign, being negative for $x = r+2$. Consider now the product $\prod_{i=0}^{n-x-1} (b-i)$. This product will be positive if $n-x-1 \leq s$, i.e. $x \geq n-s-1$, and will alternate in sign if $x < n-s-1$.

Consider now the expressions for the probabilities of $x=0$ and $x=1$. These involve, respectively, the products $\prod_{i=0}^{n-1} (b-i)$ and $a \prod_{i=0}^{n-2} (b-i)$, other factors being positive. In order for both of these products to be non-negative, it is necessary that both $\prod_{i=0}^{n-1} (b-i)$ and $\prod_{i=0}^{n-2} (b-i)$ be of the same sign, which requires that $b \geq n-1$.

Suppose now that a is not an integer, and consider the expressions for $x=n-1$ and $x=n-2$ involving the products $\prod_{i=0}^{n-2} (a-i)b$ and $\prod_{i=0}^{n-1} (a-i)$. Since $b > 0$ and $a-(n-1) < 0$, these products are of opposite sign, and we cannot have a probability distribution. Thus both parts of condition (e) must be satisfied. QED

The Pólya family provides a model for correlated binary variables. Consider a vector of n binary random variables whose joint distribution is given by

$$\Pr\{R_i = r_i, i = 1, \dots, n\} = \frac{\binom{a}{\sum r_i} \binom{b}{n - \sum r_i}}{\binom{n}{\sum r_i} \binom{a+b}{n}} \quad (4.10)$$

The distribution of their sum is then given by (4.1). The random variables are exchangeable by construction, and hence in particular are equicorrelated and have a common mean. Because of (4.4) this mean must be $a/(a+b)$, and from (4.5) it follows that the correlation

Figure 4.1: The natural parameter space for the Pólya family in the a - b plane when $n = 5$

between any two different R_i 's is $-1/(a+b-1)$. The case with a, b positive integers corresponds to the hypergeometric distribution, in which this correlation is negative. Negative a and b , on the other hand, give a positive correlation.

This latter case is called the *Pólya distribution* by Feller[28], and for rational a and b , can be generated by an urn model. In toxicological applications, however, this distribution is called the *beta-binomial*, and is derived from a different paradigm: Suppose that a random variable p is taken from a beta distribution with density $p^{\alpha-1}(1-p)^{\beta-1}/B(\alpha, \beta)$, and that conditionally on p we have n Bernoulli random variables. Then, as we show in Section 6.4, the marginal distribution of their sum is given by (4.1) with $a = -\alpha$ and $b = -\beta$.

The parametrization (4.1) in terms of a and b is convenient for some purposes, but is awkward for others. The singularity at the origin is annoying, and all points with $a = 0$ define the same distribution, as do all those with $b = 0$. Moreover, since the correlation goes to zero as either a or $b \rightarrow \pm\infty$, one might hope that the binomial distribution is achieved in the limit. This limit is easier to see after reparametrization.

Let $\mu = a/(a+b)$ and $\rho = 1/(1-a-b)$ be the mean and correlation of the underlying binary variables. Writing

$$\begin{aligned} P(x; a, b, n) &= \frac{\binom{a}{x} \binom{b}{n-x}}{\binom{a+b}{n}} \\ &= \binom{n}{x} \frac{\prod_{i=0}^{x-1} ((1-\rho)\mu + i\rho) \prod_{i=0}^{n-x-1} ((1-\rho)(1-\mu) + i\rho)}{\prod_{i=0}^{n-1} ((1-\rho) + i\rho)} \quad (4.11) \end{aligned}$$

we see that as $\rho \rightarrow 0$ (4.11) gives the binomial distribution. Moreover, by rewriting (4.11) as

$$P(x; a, b, n) = \begin{cases} \frac{(1-\mu) \prod_{i=1}^{n-1} ((1-\rho)(1-\mu) + i\rho)}{\prod_{i=2}^{n-1} ((1-\rho) + i\rho)} & \text{if } x = 0 \\ \binom{n}{x} \frac{\mu \prod_{i=1}^{x-1} ((1-\rho)\mu + i\rho) \prod_{i=0}^{n-x-1} ((1-\rho)(1-\mu) + i\rho)}{\prod_{i=2}^{n-1} ((1-\rho) + i\rho)} & \text{if } x > 0 \end{cases}$$

we can also take the limit as $\rho \rightarrow 1$, thus defining the distribution at the origin ($a = b = 0$) of the original parameter space. We find that this point actually corresponds to a family of two-point distributions, with a mass of $(1-\mu)$ at 0 and μ at n .

The parametrization in terms of μ and ρ gives a compact parameter space, with meaning given to the limiting cases. By transforming the restrictions of Theorem 4.1 to the

new parameters and including the boundary points we find that the permissible range of the parameters is

$$\begin{aligned} & \left\{ (\mu, \rho) : 0 \leq \mu \leq 1; -\frac{1-\mu}{n-2+\mu} \vee -\frac{\mu}{n-1-\mu} \leq \rho \leq 1 \right\} \\ \cup & \left\{ (\mu, \rho) : 0 \leq \mu \leq 1; \rho = -\frac{\mu}{k-\mu}, k \text{ integer}, \frac{(n-1)\mu}{1-\mu} \leq k < n-1 \right\} \\ \cup & \left\{ (\mu, \rho) : 0 \leq \mu \leq 1; \rho = -\frac{1-\mu}{k-1+\mu}, k \text{ integer}, \frac{(n-1)(1-\mu)}{\mu} \leq k < n-1 \right\} \\ \cup & \left\{ (\mu, \rho) : \mu = \frac{k}{k+l}, k=2, \dots, n-2; l=n-k, \dots, n-2; \rho = -\frac{1}{k+l-1} \right\}. \end{aligned}$$

We also see that the binomial family forms the boundary between the overdispersed (beta-binomial) and underdispersed (hypergeometric) distributions. (This parametrization, however, is still not continuous because all the points on the line segment $\mu = 0$ represent the same distribution regardless of the value of ρ , as do all those on $\mu = 1$. A continuous parametrization could be obtained by μ and $\varphi = \rho\mu(1-\mu)$. However, this parametrization is not as convenient for computational purposes, and so we shall not use it.)

4.2 Properties of Pólya Distributions

We now examine some properties of the distributions in the Pólya family. For simplicity, we will just write $P(x)$ for $P(x; a, b, n)$ when the parameters are clear from the context.

4.2.1 Recursive relations

From either (4.1) or (4.11) one can see a recursive relation that is useful for computing probabilities:

$$P(x+1) = c(x)P(x) \quad (4.12)$$

where

$$c(x) = \frac{(n-x)(a-x)}{(1+x)(b-n+1+x)} \quad (4.13)$$

or

$$c(x) = \frac{(n-x)((1-\rho)\mu + x\rho)}{(1+x)((1-\rho)(1-\mu) + (n-x-1)\rho)} \quad (4.14)$$

in the alternative parametrization. In the second form we see that the recursion also works for the binomial family, in which case the recursive factor reduces to

$$\frac{(n-x)\mu}{(x+1)(1-\mu)}$$

The recursive factor $c(x)$ is undefined in some cases. This happens in the first form for $x = n-1-b = (1-\mu)(1-\rho)/\rho$. Since $c(x)$ is required for $0 \leq x \leq n-1$ this can happen only if $0 \leq b \leq n-1$, which requires b to be an integer. In this case, the distribution has zero mass for $x \leq n-1-b$ and the recursive relationship is only needed for $x > n-1-b$.

Figure 4.2: The natural parameter space of the Pólya family in the μ - ρ plane when $n = 5$

In the second form, we also see the recursion breaking down if $\rho = 1$ and $x = n - 1$. This case gives a two-point distribution, so that a recursive formula is unnecessary. It should be borne in mind, however, that computational problems will occur using (4.12) for ρ close to 1.

4.2.2 Shape of distribution

Examination of the recursive factor $c(x)$ reveals the shape of the distribution. We can write

$$\begin{aligned} c(x) &= \frac{(n-x)(a-x)}{(1+x)(b-n+x+1)} \\ &= 1 + d(x), \end{aligned} \quad (4.15)$$

where

$$d(x) = \frac{n(a+1) - (b+1) - (a+b+2)x}{(x+1)(b-n+x+1)}. \quad (4.16)$$

Thus the probabilities will be increasing or decreasing according as $d(x)$ is positive or negative, respectively. If $a, b > 0$ then the denominator of $d(x)$ will be positive, as will the factor $a + b + 2$. Hence

$$d(x) \begin{cases} > 0 & \text{if } x < \frac{n(a+1) - (b+1)}{a+b+2} \\ < 0 & \text{if } x > \frac{n(a+1) - (b+1)}{a+b+2} \end{cases} \quad (4.17)$$

If $b+1 > n(a+1)$ then $d(x) < 0$ for all $x \geq 0$ and thus the probabilities will be everywhere decreasing. By contrast, if $(a+1) > n(b+1)$ then $d(x) > 0$ for all $x \leq n-1$ and the probabilities will be everywhere increasing, giving a "J-shaped" distribution. The distribution will be "bell-shaped" (in the sense of having a mode in $[1, n-1]$ and no other local maxima) if

$$0 < \frac{n(a+1) - (b+1)}{a+b+2} < n-1. \quad (4.18)$$

If $a, b < -1$ then the distribution will be decreasing if $b+1 < n(a+1)$, increasing if $a+1 < n(b+1)$ and again bell-shaped if (4.18) holds. If $a \leq -1$ but $-1 < b < 0$, then $d(x)$ is positive throughout the range $0 \leq x \leq n-1$, and the probabilities are everywhere increasing, whereas if $b \leq -1$ but $-1 < a < 0$ the probabilities are everywhere decreasing. Finally, if $-1 < a < 0$ and $-1 < b < 0$ then $d(x)$ is negative for $x < \frac{n(a+1) - (b+1)}{a+b+2}$ and positive for $x > \frac{n(a+1) - (b+1)}{a+b+2}$. Hence in this case the distribution will be increasing if $b+1 > n(a+1)$, decreasing if $a+1 > n(b+1)$, and U-shaped if $(a+1)/n < b+1 < n(a+1)$. The remaining case $a = b = -1$ corresponds to the discrete uniform distribution, which has a constant probability.

In the alternative parametrization the lower boundary of the U-shaped region is given by

$$\rho = \frac{\mu - \frac{1}{n+1}}{\mu + \frac{1}{n+1}} \quad \text{and}$$

$$\rho = \frac{\frac{n}{n+1} - \mu}{\frac{2n-1}{n+1} - \mu}$$

(see Figures 4.3 and 4.4).

At the boundaries of the parameter space, the U- and J-shaped distributions become two and one-point distributions, respectively. Distributions near these degenerate cases will assign negligible probabilities to points other than 0 and n . It is useful to have an upper bound on this negligible probability in terms of the parameters. Such a bound is given by the following theorem.

Theorem 4.2 Suppose that

$$(1 - \rho)\mu(1 - \mu) \leq \delta/n^2, \quad (4.19)$$

where $\delta < \frac{1}{8}$. Then $P(0) + P(n) \geq 1 - \delta$.

Proof: From (4.11) we can write

$$\begin{aligned} P(1) &= n(1 - \rho)\mu(1 - \mu) \prod_{j=2}^n \frac{(1 - \rho)(1 - \mu) + (n - j)\rho}{(1 - \rho) + (n - j)\rho} \\ &\leq n(1 - \rho)\mu(1 - \mu) \end{aligned} \quad (4.20)$$

and similarly $P(n - 1) \leq n(1 - \rho)\mu(1 - \mu)$. Now, if the distribution has mode at 0 or n and (4.19) holds then the probabilities of each x in $1 \leq x \leq n - 1$ are no greater than δ/n , and hence

$$\sum_{i=1}^{n-1} P(i) \leq \delta.$$

We now show that (4.19) cannot hold if the distribution is bell-shaped.

The inequality (4.19) can be written in the original parametrization as

$$k(a + b)^2 - k(a + b) - ab \leq 0 \quad (4.21)$$

where $k = \delta/n^2$. The equation $k(a + b)^2 - k(a + b) - ab = 0$ defines a hyperbola in the (a, b) plane, and the region satisfying (4.19) is the part between its two branches (see Figure 4.5). By a further reparametrization to $u = (a + b)/2$ and $v = (a - b)/2$, (4.21) can be written

$$v^2 \geq (1 - 4k)u^2 + 2ku. \quad (4.22)$$

With this parametrization the set of bell-shaped distributions (4.18) is represented by $B_1 \cup B_2$, where

$$B_1 = \left\{ u < -1, v^2 < \left(\frac{(n-1)(u+1)}{n+1} \right)^2 \right\}, \quad (4.23)$$

$$\text{and } B_2 = \left\{ u > 0, v^2 < \left(\frac{(n-1)(u+1)}{n+1} \right)^2 \right\}. \quad (4.24)$$

Figure 4.3: Parameter regions in the a - b plane defining various shapes of the distribution

Figure 4.4: Parameter regions in the μ - ρ plane defining various shapes of the distribution

Figure 4.5: Region satisfying $(1 - \rho)\mu(1 - \mu) < \delta/n^2$

Consider the function

$$\begin{aligned} f(u) &= (1 - 4k)u^2 + 2ku - \left(\frac{(n-1)(u+1)}{n+1} \right)^2 \\ &= \left(1 - 4k - \left(\frac{n-1}{n+1} \right)^2 \right) u^2 + 2 \left(k - \left(\frac{n-1}{n+1} \right)^2 \right) u - \left(\frac{n-1}{n+1} \right)^2 \end{aligned} \quad (4.25)$$

Since $k < 1/(6n^2) < n/(n+1)^2$ we have $\left(\frac{n-1}{n+1} \right)^2 < 1 - 4k$ and f is a convex function of u . Moreover, $f(-1) = 1 - 6k > 0$, since $k = \delta/n^2 < \frac{1}{6}$. Since the zeros of f will have opposite signs these two facts imply that $f(u) > 0$ for all $u < -1$. The definition (4.25) of f then implies that if $(u, v) \in B_1$ then

$$v^2 \leq \left(\frac{(n-1)(u+1)}{n+1} \right)^2 < (1 - 4k)u^2 + 2ku \quad (4.26)$$

so that (u, v) cannot satisfy (4.22). Since (4.22) is merely a reparametrization of (4.19), it follows that this inequality fails as well.

Consider now the set of bell shaped distributions in B_2 . By Theorem 4.1 the parameters (u, v) with $u > 0$ will not define a probability distribution unless $u \geq n/2$. Now,

$$\begin{aligned} f(n/2) &= \frac{(n^2 - 1)(1 - n(n+1)k) + n}{(n+1)^2} \\ &> \frac{(n-1)(1 - n(n+1)k)}{n+1} \\ &> 0 \end{aligned}$$

since $n(n+1) < 1/k$.

The arguments leading up to (4.26) can then be repeated to give that $(u, v) \in B_2$ implies that (4.22) and consequently (4.19) is false. Since every bell-shaped distribution is in B_1 or B_2 , and the only bimodal distributions in the Pólya family are U-shaped it follows that distributions for which (4.19) holds must have a mode at either 0 or n . QED

For practical purposes, distributions which satisfy the conditions of Theorem 4.2 can be considered to be two-point distributions with the probabilities of non-extreme outcomes set to zero. This way, numerical problems caused by recursively calculating probabilities over a wide range of magnitudes can be avoided. There is still a concern, however, about U-shaped distributions (which have $-1 < a, b < 0$) that have small but not negligible probabilities at 1 and $n-1$. It is conceivable that a recursive computation beginning at one end of the distribution could lead to a number so small as to cause underflow, (see Appendix C) which would prevent completion of the calculations at the other end. The following theorem indicates that this cannot occur, since the U-shaped distributions are relatively flat in the middle.

Theorem 4.3 Let P define the probability distribution in the family (4.1), with $-1 < a, b < 0$. Then $P(x)/P(1) \geq 4/n^2$.

Proof: Let

$$c_x(\rho) = \frac{P(x+1)}{P(x)} = \frac{\mu + (x-\mu)\rho}{(1-\mu) + (n-x-2-\mu)\rho} \cdot \frac{n-x}{x+1}.$$

This is a rational function of ρ , having a root at $\rho = -\mu/(x-\mu)$ and a pole at $\rho = -(1-\mu)/(n-x-2+\mu)$. When $\rho = 1$, $c_x = \frac{x(n-x)}{(x+1)(n-x-1)}$. As ρ increases to 1, $c_x(\rho)$ either decreases or increases to $c_x(1)$ according as $-\mu/(x-\mu) < -(1-\mu)/(n-x-2+\mu)$ or $-\mu/(x-\mu) > -(1-\mu)/(n-x-2+\mu)$ respectively.

For $-1 < a, b < 0$, or equivalently $\rho > \mu/(1+\mu) \vee (1-\mu)/(2-\mu)$, $c_x(\rho)$ is an increasing function of x , being less than unity for $x = x_1(\omega) = \frac{(n+1)\mu - 1 - (n-1)\omega}{1-2\omega}$, where $\omega = \rho/(1-\rho) = -1/(a+b)$. Suppose first that $\mu \geq \frac{1}{2}$. Then for $\rho > \mu/(1+\mu) \vee (1-\mu)/(2-\mu)$, $\omega > \frac{1}{2}$, and $x_1(\omega) < (n-1)/2 \leq (n-1)\mu$. Now since $c_x < 1$ for $x < x_1(\omega)$ and $c_x > 1$ thereafter, $P(x)$ must achieve its smallest value at $x = [x_1(\omega)] + 1$, and its value there is

$$P(1) \prod_{x=1}^{[x_1(\omega)]} c_x(\rho).$$

But $c_x(\rho)$ is decreasing with ρ (since $x < (n-1)\mu$) so that $c_x(\rho) \geq c_x(1)$. Furthermore, $x_1(\omega)$ is increasing with ρ , so that $\prod_{x=1}^{[x_1(\omega)]} c_x(\rho)$ has at least as many factors as $\prod_{x=1}^{[x_1(\omega)]} c_x(1)$, all of which are less than 1.

Hence,

$$\begin{aligned} P(x) &= P(1) \prod_{x=1}^{[x_1(\omega)]} c_x(\rho) \\ &\geq P(1) \prod_{x=1}^{[x_1(\omega)]} c_x(1) \\ &= P(1) \prod_{x=1}^{[(n-1)/2]} \frac{x(n-x)}{(x+1)(n-x+1)}. \end{aligned}$$

If n is odd, the above product equals

$$\frac{n-1}{\left(\frac{n-1}{2} + 1\right) \left(n - \frac{n-1}{2} - 1\right)} = \frac{4}{n+1} > \frac{4}{n^2}. \quad (4.27)$$

If n is even then the product equals

$$\frac{n-1}{\left(\frac{n}{2} - 1 + 1\right) \left(n - \frac{n}{2} + 1 - 1\right)} = \frac{4(n-1)}{n^2} > \frac{4}{n^2}. \quad (4.28)$$

Thus we have proved the theorem for $\mu \geq \frac{1}{2}$.

Now suppose that $\mu < \frac{1}{2}$. Letting $P_0(x; a, b, n) = P(n-x; b, a, n)$, we see that

$$\min_{1 \leq x \leq n-1} P(x)/P(n-1) \geq 4(n-1)/n^2$$

by (4.27) and (4.28). Hence to complete the proof we need to show that $P(n-1)/P(1) \geq 1/(n-1)$. Now

$$\begin{aligned} \frac{P(n-1)}{P(n)} &= \prod_{x=1}^{n-2} c_x(\rho) \\ &= \prod_{x=1}^{n-2} \frac{\mu + (x-\mu)\rho}{(1-\mu) + (n-x-2+\mu)\rho} \cdot \frac{n-x}{x+1}. \end{aligned} \quad (4.29)$$

This expression is a rational function of ρ , with roots at $\rho = -\mu/(x-\mu)$ and poles at $\rho = -(1-\mu)/(n-x-2+\mu)$, for $x = 1, \dots, n-2$. The largest root will exceed the largest pole if $-\mu/(n-2-\mu) > -(1-\mu)/(n-3+\mu)$, i.e. if $\mu < \frac{1}{2}$. In this case, (4.29) is increasing in ρ for $\rho > -(1-\mu)/(n-3+\mu)$. Thus for $\rho > \mu/(1-\mu) \vee (1-\mu)/(2-\mu)$, a lower bound for (4.29) can be obtained by setting $\rho = \mu/(1-\mu) \vee (1-\mu)/(2-\mu)$, which for $\mu < \frac{1}{2}$ means $\rho = (1-\mu)/(2-\mu)$, or in the original parametrization $b = -1$. Writing (4.29) in terms of the parameters a and b gives us

$$\prod_{x=1}^{n-2} \frac{(n-x)(x-a)}{(x+1)(n-1-x-b)} = \prod_{x=1}^{n-2} \frac{x-a}{x+1} \geq \prod_{x=1}^{n-2} \frac{x}{x+1} = \frac{1}{n-1}$$

since $b = -1$ and $a < 0$.

QED

We have demonstrated that the U-shaped distributions are relatively flat in the middle, with any large changes taking place between 0 and 1, and $n-1$ and n . These almost degenerate cases, however, are taken care of by Theorem 4.2. Specifically, if $n(1-\rho)\mu(1-\mu) > \delta/n$, then

$$\begin{aligned} P(1)/P(0) &= c(0) \\ &= \frac{n(1-\rho)\mu}{(1-\mu) + (n-2+\mu)\rho} \\ &> \frac{n(1-\rho)\mu(1-\mu)}{n-1} \\ &> \frac{\delta}{n(n-1)} \end{aligned}$$

and

$$\begin{aligned} P(n-1)/P(n) &= 1/c(n-1) \\ &= \frac{n(1-\rho)(1-\mu)}{\mu + (n-1-\mu)\rho} \\ &> \frac{n(1-\rho)\mu(1-\mu)}{n-1} \\ &> \frac{\delta}{n(n-1)}. \end{aligned}$$

Thus by the previous theorem, the range of probabilities does not exceed a factor of $n^3(n-1)/(4\delta)$.

4.2.3 Truncation of tails

In many situations extreme values have negligible probabilities, so that in the computation of tail probabilities and expected values, considerable computation can be avoided by truncation. If the distribution is decreasing, i.e. $c(x) < 1$, then a bound on the upper tail probability is given by

$$\sum_{i=x+1}^n P(i) \leq (n-x)P(x+1).$$

However, if $c(x)$ itself is decreasing then a tighter bound is possible.

By repeated recursion we have

$$P(x+k) = P(x) \prod_{i=0}^{k-1} c(x+i). \quad (4.30)$$

If $c(x)$ is decreasing in x then the product is less than $[c(x)]^k$ so that

$$\begin{aligned} \sum_{i=x+1}^n P(i) &\leq P(x) \sum_{i=1}^{n-x} [c(x)]^i \\ &\leq P(x) \sum_{i=1}^{\infty} [c(x)]^i \\ &= P(x) \frac{c(x)}{1-c(x)} \\ &= \frac{P(x+1)}{1-c(x)}. \end{aligned} \quad (4.31)$$

Suppose that we want to calculate a tail probability $\Pr\{X \geq x\}$ to an accuracy of ϵ . Then if $P(x+k)/(1-c(x+k-1)) < \epsilon$ it follows that

$$\sum_{i=x}^{x+k-1} P(i) \leq \sum_{i=x}^n P(i) \leq \sum_{i=x}^{x+k-1} P(i) + \epsilon. \quad (4.32)$$

Thus after having calculated $P(x+k)$ we can truncate the remaining values since they will have a negligible effect. Similar truncation can be done when calculating expected values, provided the random variable whose expectation is being calculated is bounded.

The recursive factor $c(x)$ will be decreasing if a and b are both positive, or if they are both less than -1 . If $-1 < a < 0$ and $-1 < b < 0$ then the recursive factor is in fact increasing, so that one could attempt to truncate by replacing $c(x)$ by $c(n-1) = (a-n+1)/(nb)$ in (4.31), although this tactic will fail if the distribution is in fact U-shaped, so that $c(n-1) > 1$. Although more exact bounds on the tail probabilities could be calculated in this case, their computation would tend to take more time than the calculation of the probabilities themselves.

If $a > -1$ but $b < -1$ the recursive factor initially increases to a maximum and then decreases. The point of maximum can be found by equating the derivative of (4.12) to zero, giving

$$x = \frac{n(a+1) - (b+1) + \sqrt{(n+1)(a+1)(b+1)(a+b+1)}}{a+b+2}. \quad (4.33)$$

However, in practice it is easier to test for a decreasing $c(x)$ during computation. If $a < -1$ but $b > -1$, the distribution is J-shaped, and it might be better to calculate the probabilities in the opposite sequence using the fact that $P(x; a, b, n) = P(n - x; b, a, n)$.

4.3 A Computational Algorithm

4.3.1 Factoring the probability function

A convenient computation scheme is based on the following factorization: Let

$$\begin{aligned} f_i &= \frac{(a+1-i)(n+1-i)}{i(a+b-n+i)} \quad \text{and} \\ g_i &= \frac{b-n+i}{a+b-n+i}. \end{aligned} \quad (4.34)$$

Then

$$P(x) = \prod_{i=1}^x f_i \prod_{i=x+1}^n g_i \quad (4.35)$$

is the product of n factors, and the recursive factor $c(x)$ is equal to f_{x+1}/g_{x+1} . The boundary cases can be handled by the alternative parametrization:

$$\begin{aligned} f_i &= \frac{(1-\rho)\mu + (i-1)\rho}{(1-\rho) + (n-i)\rho} \cdot \frac{n-j+i}{j} \quad \text{and} \\ g_i &= \frac{(1-\rho)(1-\mu) + (n-i)\rho}{(1-\rho) + (n-i)\rho}. \end{aligned} \quad (4.36)$$

The expression (4.35) for $P(x)$ is now well defined in all cases except for $\rho = 1$, in which cases f_n and g_n are undefined. Problems also may arise for $\mu = 1$, in which case $g(n) = 0$, and for $0 < b = k < n - 1$ (where k is an integer), for which $g(n - k) = 0$, so that the recursive factors are undefined.

For the cases $\rho = 1, \mu = 1$ as well as $\mu = 0$, the formula (4.35) is unnecessarily complicated, since we have a two-point or one-point distribution. To avoid numerical problems, one should appeal to Theorem 4.2 to approximate distributions close to these degenerate cases.

The properties of the f_i 's and g_i 's should be noted. $g_i < 1$ for $1 \leq i \leq n$ and if $\rho_i > 0$, g_i is decreasing with i . Hence its minimum value is $g_n = 1 - \mu$ if $\rho \neq 1$, and this will also be the limiting value when $\rho \rightarrow 1$. However, direct computation of g_n using (4.35) when ρ is close to 1 will lead to numerical problems. If $\rho < 0$ then g_i is increasing with i , hence its maximum value is $1 - \mu$. The restraints on the parameter space ensure that if $g_i < 0$, then $g_j = 0$ for some $j > i$. However, round-off errors may not give $g_j = 0$ exactly, so that situations where g_i approaches zero have to be handled specially.

If $a \leq -1$ or $a > 0$, then f_i is decreasing on $1 \leq i \leq n$. If $a > 0$ then f_i can become negative, but again, this can only happen if $f_i = 0$ for some i . If $-1 \leq a < 0$ but $a + b \leq -1$ then f_i can be non-monotone, but in this case $0 < f_i < 1$ for $1 \leq i \leq n$. If $-1 \leq a + b < 0$ then f_i is increasing, and in fact is unbounded as $a, b \rightarrow 0$. These cases correspond to

U-shaped distributions, however, so that in practice one would avoid the calculation of excessively large f_i 's by appealing to Theorem 4.2.

Provided that $a < -1$ or $a > 0$ (this is equivalent to $\rho < \mu/(1 + \mu)$ in the alternative parametrization) f_i is bounded above by

$$f_1 = \frac{na}{a+b-n+1} = \frac{n(1-\rho)\mu}{(1-\rho) + (n-1)\rho}.$$

This expression reaches its maximum at $a = n-1$, $b = 1$, where it equals $n(n-1)$.

If (4.19) holds, we can safely approximate the distribution by $P(0) = 1 - \mu$, $P(n) = \mu$, and $P(x) = 0$ for $0 < x < n$, and this will deviate from the true distribution by less than δ .

If (4.19) does not hold, then

$$f_n < \frac{\mu}{n} + \frac{n(n-1)\mu(1-\mu)}{\delta} - \frac{n(n-1)}{n} < \frac{n(n-1)}{4\delta}.$$

Similarly

$$f_1 = \frac{n(1-\rho)\mu}{1+(n-2)\rho} > \frac{n(1-\rho)\mu(1-\mu)}{n-1} > \frac{\delta}{n(n-1)}.$$

Thus the factors f_i will not become either excessively large or excessively small, provided we exclude the nearly degenerate distributions covered by Theorem 4.2.

4.3.2 Bounds on partial products

Excluding (4.19) as a special case, we can use (4.35) to produce an efficient, numerically stable procedure for calculating the probabilities. But first we require a few more results which will allow us to truncate negligible probabilities.

Theorem 4.4 Let $P(x) = \prod_{i=1}^x f_i \prod_{i=x+1}^n g_i$ be a probability distribution where $0 \leq f_i \leq 1$ and $0 \leq g_i \leq 1$. Suppose that $\prod_{i=k}^n g_i < \epsilon$. Then $\max_{0 \leq x < k} P(x) < \epsilon$.

Proof:

$$P(x) = \prod_{i=1}^x f_i \prod_{i=x+1}^n g_i = \prod_{i=1}^x f_i \prod_{i=x+1}^{k-1} g_i \prod_{i=k}^n g_i < \epsilon$$

since $\prod f_i, \prod g_i \leq 1$.

QED

Theorem 4.5 Let $P(x)$ be defined as above, with $0 \leq g_i \leq 1$ and $f_i \leq f_j$ if $i \leq j$. Suppose that $\prod_{i=k}^n g_i < \epsilon$. Then $\max_{0 \leq x < k} P(x) < \epsilon$.

Proof:

$$P(x) = \prod_{i=1}^x f_i \prod_{i=x+1}^n g_i = \prod_{i=1}^x f_i \prod_{i=x+1}^{k-1} g_i \prod_{i=k}^n g_i < \epsilon \prod_{i=1}^x f_i.$$

Now, suppose that $f_x \leq 1$. Then $f_i \leq 1$ for $i \leq x$ and hence $\prod_{i=1}^x f_i \leq 1$. On the other hand, if $f_x > 1$ then $f_i > 1$ for $i > x$ and $\prod_{i=1}^x f_i < \prod_{i=1}^n f_i = P(n) = 1$. QED

In the description of the algorithm we will use the following notation for partial products: For $0 \leq r < s \leq t$ define

$$Q(r, s, t) = \prod_{i=1}^r f_i \prod_{i=s}^{t-1} f_i \prod_{i=t}^n g_i. \quad (4.37)$$

Then $P(x) = Q(r, r+1, x+1)$ for any $r \leq x$.

Theorem 4.6 Suppose that $0 \leq g_i \leq 1$ and that $f_j \leq f_i$ for $j \geq i$. Let $u = \max\{i : f_i > 1\}$ and $0 < \epsilon < 1/(n+1)$.

(a) If

$$Q(r, s, s) < \epsilon/s \text{ for all } 0 \leq r < u < s \quad (4.38)$$

then $\sum_{i=0}^{t-1} P(i) < \epsilon$.

(b) If (4.38) holds and

$$Q(u, s, j) < \epsilon/t \text{ for all } s \leq j \leq t \quad (4.39)$$

then $\sum_{i=0}^{t-1} P(i) < \epsilon$.

(c) If (4.38) and (4.39) hold and $Q(u, s, t+1) < Q(u, s, t)$, then f_i/g_i exceeds 1 for some $i > t+1$.

(d) If (4.38) holds and

$$Q(u, s, j) < \epsilon/j \text{ for all } x \leq j \leq t \quad (4.40)$$

then $\sum_{i=0}^{t-1} P(i) < \epsilon$.

Proof: For any $j < s$,

$$P(j) = \prod_{i=1}^j f_i \prod_{i=j+1}^n g_i \leq \prod_{i=1}^j f_i \prod_{i=s}^n g_i = Q(j, s, s).$$

Now, if $j > u$, then

$$P(j) \leq \prod_{i=1}^j f_i \prod_{i=s}^n g_i \leq \prod_{i=1}^u f_i \prod_{i=s}^n g_i = Q(u, s, s).$$

Thus if $Q(j, s, s) < \epsilon/s$ for all $0 \leq j \leq u$ then $P(j) < \epsilon/s$ for all $0 \leq j \leq s-1$, and hence

$$\sum_{j=0}^{s-1} P(j) \leq s \max_{0 \leq j \leq s-1} P(j) < \epsilon,$$

thus proving (a).

Suppose now that (a) holds. Then

$$P(j) = \prod_{i=u+1}^{t-1} f_i Q(u, s, j) < Q(u, s, j) < \frac{\epsilon}{t}, \quad (4.41)$$

and hence $\sum_{i=0}^{t-1} P(i) < t\epsilon/t = \epsilon$.

Now, let (4.41) be true, and suppose that $Q(u, s, t+1) < Q(u, s, t)$. Then $Q(u, s, t+1) < \epsilon/t$. Now, for $j \leq t$, $P(j) = \prod_{i=u+1}^{t-1} f_i Q(u, s, j+1) \leq Q(u, s, j+1)$ and $Q(u, s, j+1) = (\prod_{i=1}^j (f_i/g_i)) Q(u, s, t)$. Suppose $f_i/g_i \leq 1$ for all $i > t+1$, then

$$Q(u, s, j+1) \leq Q(u, s, t) < \epsilon/t.$$

Therefore, $\sum_{j=t}^n P(j) \leq \sum_{j=t}^n Q(u, s, j+1) \leq \epsilon(n+1)/t$. But by hypothesis, $\sum_{j=1}^{t-1} P(j) \leq \epsilon$, giving us

$$1 = \sum_{j=0}^n P(j) \leq \epsilon + \frac{n-t+1}{t} \epsilon = \frac{n+1}{t} \epsilon \leq (n+1)\epsilon.$$

But ϵ was chosen less than $1/(n+1)$, and we have a contradiction. Thus f_i/g_i must exceed 1 for some $i > t$.

To prove (d), we first show that under the stated conditions $Q(u, s, j) \geq Q(u, s, t)$. For otherwise we would have a situation of condition (c) and thus f_i/g_i must exceed 1. Thus f_i/g_i must increase from less than 1 to greater than 1. But this can only happen if $-1 \leq a < 0$ and $-1 \leq b < 0$, but in such a case f_i is increasing, contrary to the hypothesis. Now, if $Q(u, s, j) \leq Q(u, s, t) < \epsilon/t$ then $P(j) < \epsilon/t$ for $0 \leq j \leq t-1$, and hence $\sum_{j=0}^{t-1} P(j) < \epsilon$. QED

The above theorems provide criteria for truncating the initial part of the distribution, criteria which can be checked dynamically while the probability is being calculated by (4.35). The truncation achieves two purposes. First, it reduces computational labour, and secondly, it avoids underflow errors by not attempting to calculate probabilities that are nearly zero. The tail bounds given in (4.31) provide a mechanism for truncating the upper tail of the distribution. Thus, only the middle part of the distribution needs to be calculated.

4.3.3 The algorithm

We now describe the computational algorithm. The algorithm is defined in terms of the partial products $Q(r, s, t)$ defined in (4.37). The basic operations involve the following transitions:

$$\begin{aligned} A &: Q(r, s, t) \longrightarrow Q(r, s, t+1) \\ B &: Q(r, s, s) \longrightarrow Q(r, s-1, s-1) \\ C &: Q(r, s, t) \longrightarrow Q(r+1, s, t) \end{aligned}$$

These operations are implemented by $AQ = c_{t-1}Q$, $BQ = g_{t-1}Q$, and $CQ = f_{r+1}Q$. We consider these operations as updating both the arguments r, s, t and the value of Q . Note

that Q is undefined unless $0 < r < s \leq t < n + 1$. Following the usual conventions with products, we define $Q(0, n + 1, n + 1) = 1$.

We present the algorithm for calculating the expected value of a bounded function $\phi(X)$ with $|\phi(x)| \leq M$. We assume nothing else about this function, although in most applications it will be non-negative. The algorithm will produce a number $\tilde{E}$ such that $\tilde{E} - \epsilon \leq E(\phi) \leq \tilde{E} + \epsilon$ for some predetermined absolute error ϵ . We assume that computation is done with relative error less than σ . Typically $\sigma = 2^{-22} \simeq 2 \times 10^{-7}$ for single precision.

The algorithm is presented in a pseudo-code with obvious semantics and no particular syntax. Value assignment is indicated by $\leftarrow$. Thus, for example, the statement $Q \leftarrow CQ$ is to be read "Perform operation C on Q and substitute the result for the current value of Q ".

step1. $\delta \leftarrow \epsilon / (2M) - 5n\sigma$. Initialize $r = 0$, $s = n + 1$, $t = n + 1$, $Q \leftarrow Q(0, n + 1, n + 1) = 1$.

step2. IF $Q \geq \delta/t$
 THEN IF $r + 1 < s \leq t$
 THEN $Q \leftarrow BQ$
 ELSE IF $r + 1 < s$
 THEN $Q \leftarrow CQ$
 ELSE GO TO step4
 ELSE IF $a - 1 \leq a < 0$
 THEN $Q \leftarrow C^{s-r-1}Q$
 ELSE IF $r + 1 < s$
 THEN IF $f_{r+1} > 1$
 THEN $Q \leftarrow CQ$
 ELSE $Q \leftarrow AQ$
 ELSE GO TO step4

step3. Repeat step2.

step4. IF $c_{t-1} < c_{t-2}$ and $Q < \delta(1 - 1/c_{t-1})$
 THEN $Q \leftarrow AQ$
 ELSE GO TO step6.

step5. Repeat step4.

step6. $\tilde{E} \leftarrow 0$.

step7. $Q \leftarrow AQ$

step8. IF $c_{t-1} < c_{t-2}$ and $Q < \delta(1 - c_{t-1})$
 THEN STOP.

step9. $\tilde{E} \leftarrow \tilde{E} + \phi(t - 1)Q$

step10. GO TO step7.

It should be noted that in some applications the function itself may be defined by a computer algorithm, possibly even by another instance of this same algorithm. *A priori* knowledge of the function's behaviour will allow refinements of the algorithm. For example, if one is calculating tail probabilities, then the function would be defined by $\phi(x) = 1$ if $x \geq q$ and $\phi(x) = 0$ otherwise. In this case one would replace the condition on the

second line of *step2* with $q < s = t$, since there is no point in calculating probabilities for $P(x) = Q(x, x+1, x+1)$ for $x < q$. If, in this case, the branch $Q \leftarrow AQ$ on the second last line of *step2* is reached, then we know that $\sum_{x=0}^{q-1} P(x) < \delta$, and hence the probability of interest is approximately 1.

Refinements are also possible if $0 < a \leq n-1$, or $0 < b \leq n-1$. As proved in Theorem 4.1, this requires that a (or b) be an integer. In such cases, cancellation takes place in the products of f_i 's and g_i 's so that some economization of computation is possible. Specifically, if a is a positive integer less than $n-1$, then f_i will be zero for $i = a+1$. Thus the product $\prod_{i=1}^x f_i \prod_{i=x+1}^n g_i$ will be nonzero only for $x \leq a$. Now if $a < n-a$, then

$$\prod_{i=a+1}^n g_i = \prod_{i=a+1}^n \frac{b-n+i}{a+b-n+i} \quad (4.42)$$

reduces, by cancellation, to

$$\prod_{i=1}^a \frac{a+b-n+i}{b+i}, \quad (4.43)$$

which has fewer factors. Similarly, if b is a positive integer less than $n-1$, then $g_i = 0$ for $i = n-b$. Thus, in calculating the product $\prod f_i \prod g_i$, we can begin with

$$\prod_{i=1}^{n-b} f_i \prod_{j=n-b+1}^n g_j.$$

Now, if $n-b \leq k \leq a$, then

$$\prod_{i=1}^k f_i = \prod_{i=1}^k \frac{(n-i+1)(a+1-i)}{k(a+b-n+i)} \quad (4.44)$$

$$= \prod_{i=1}^k \frac{n-i+1}{k} \prod_{i=1}^{k-n+b} \frac{a-k+i}{a+i} \quad (4.45)$$

which again has fewer factors.

4.3.4 Complexity and error analysis

The algorithm involves multiplication by g_i 's, f_i 's, and c_i 's. Each g_i appears at most once, and for each i , only one of f_i or c_i will appear. In the worst case of no truncation, the algorithm will involve n multiplications by g_i followed by n multiplications by c_i . Since the factors each involve a fixed number of operations, the complexity of the algorithm is thus linear.

Suppose that the relative round-off error of each factor is d . Then since $P(x)$ involves $n+x$ factors, its relative error is less than $(n+x)d$. Now

$$g_i = \frac{1 - \mu + (n-i-1+\mu)\rho}{1 + (n-i-1)\rho}.$$

For $i < n-2$, the relative error of g_i will be less than 4σ , where σ is the machine precision. The relative errors of $g_{n-1} = 1 - \mu + \mu\rho$ and $g_n = 1 - \mu$ can be substantial if μ is close

to 1. This problem can be solved by calculating $1 - \mu$ in double precision. If μ has single precision accuracy, then $1 - \mu$ calculated in double precision will also have single precision accuracy. If this is done, then g_{n-1} and g_n will have relative errors 2σ and σ respectively. Similarly, the relative errors of f_i and c_i are bounded by 6σ , provided that $1 - \rho$ is also calculated in double precision. The extra 2σ error in these factors is due to the ratio $(n-1)/(i+1)$. Thus $P(x)$ has round-off error less than $(4n+6x)\sigma$, and an upper bound on all round-off errors is $\tau = 10n\sigma$.

Two special cases deserve mention: If $\rho = 0$, (i.e. we have a binomial distribution) then $g_i = 1 - \mu$ for all i , and has error σ , whereas $c_i = (n-1)\mu/((i+1)(1-\mu))$ has error 4σ , provided that $1 - \mu$ is calculated in double precision. Thus the bound now becomes $\tau = 5n\sigma$. On the other hand, with the hypergeometric distribution, in which a and b are positive integers, the expressions for g_i , f_i , and c_i can be rewritten as quotients of integers. Using integer arithmetic for the numerator and denominator gives us an error of 2σ for g_i , f_i and c_i , so that the bound becomes $\tau = 4n\sigma$.

Provided that the accumulation of the sum (step9) is done in double precision, the absolute round-off error of the expected value will be bounded by $M\tau$. In addition, there is a truncation error bounded by $2\delta M$. Thus by defining δ as in step1, we bound the absolute error in the result by ϵ . If ϕ itself is subject to computational error (as it often will), this must be added to ϵ to give the total error.

4.4 The Multivariate Pólya Distribution

4.4.1 Definition

We can construct a k -dimensional generalization of the Pólya family in the following manner: Let n be a positive integer, and let $b_0, b_1, \dots, b_k$ be a list of $k+1$ real numbers. Define $a_i = \sum_{j=i}^k b_j$. Now define a sequence of random variables $S_1, \dots, S_k$ having the Markov property

$$\Pr \{S_i = s_i | S_0 = s_0, \dots, S_{i-1} = s_{i-1}\} = \Pr \{S_i = s_i | S_{i-1} = s_{i-1}\}$$

and for which

$$p_i(s_i | s_{i-1}) = \Pr \{S_i = s_i | S_{i-1} = s_{i-1}\} = \frac{\binom{a_i}{s_i} \binom{b_{i-1}}{s_{i-1} - s_i}}{\binom{a_i + b_{i-1}}{s_{i-1}}} \quad (4.46)$$

where by convention S_0 is a constant n . This definition will define a legitimate joint distribution for $(S_0, S_1, \dots, S_k)$ provided that each of the parameters (a_i, b_{i-1}) is in the permissible range as described in Section 4.1.2. If all the b_i 's are negative, or if they are all positive integers, then the conditions are clearly met. By analogy to the univariate case, we also want to include limiting cases when the $b_i \rightarrow \infty$.

Let us reparametrize by

$$\rho = \frac{1}{1 - a_0}, \quad \mu_i = \frac{a_i}{a_{i-1}}, \quad i = 1, \dots, k. \quad (4.47)$$

Then, the conditional distribution of S_i given S_{i-1} can be written in the form (4.11) with mean μ_i and correlation

$$\rho_i = \frac{\rho}{\rho + (1 - \rho) \prod_{j=1}^{i-1} \mu_j}.$$

In this form we can determine the probability distribution defined by the limiting case as all $a_i \rightarrow \infty$ but the ratios a_i/a_{i-1} approach finite limits. In this case the conditional distributions are binomial.

It is actually of more interest to study the increments, $X_i = S_i - S_{i+1}$. (Just as it was convenient to define S_0 to be equal to n to make (4.46) defined for all $1 \leq i \leq k$, it will always be understood that $S_{k+1} = 0$.) The joint distribution of the X_i 's can be written

$$\Pr \{X_i = x_i; i = 0, \dots, k\} = \frac{\prod_{i=0}^k \binom{b_i}{x_i}}{\binom{a_0}{n}}. \quad (4.48)$$

By definition, the X_i are non-negative random variables satisfying $\sum X_i = n$.

The limiting case $b_i \rightarrow \infty$ in the form (4.48) leads to the joint distribution

$$\Pr \{X_i = x_i; i = 0, \dots, k\} = \frac{n!}{\prod_{i=0}^k x_i!} \prod_{i=0}^k \left[(1 - \mu_{i+1}) \left(\prod_{j=1}^i \mu_j \right) \right]^{x_i}, \quad (4.49)$$

(where $\mu_{k+1} = 0$). This is the multinomial distribution with cell probabilities $p_i = (1 - \mu_{i+1}) \prod_{j=1}^i \mu_j$. In this case one can interpret the sequence S_i as the result of n independent sequences of binary experiments, continuing until failure occurs, where the probability of success on the i th trial is μ_i . The random variable X_i then represents the number of sequences in which the last success occurred on trial i .

4.4.2 Marginal and conditional distributions

We will have occasion to use marginal and conditional distributions from this family, where the concept of marginal distribution will be somewhat extended from the usual one. Because the support of the distribution is contained in the k -dimensional simplex $\{\sum_{i=0}^k x_i = n; x_i \geq 0, i = 0, \dots, k\}$ one of the variables x_i is redundant. We could thus consider the distribution to be k -dimensional with support contained in the k -dimensional simplex $\{\sum_{i=1}^k x_i \leq n; x_i \geq 0, i = 1, \dots, k\}$ with $x_0 = n - \sum_{i=1}^k x_i$, and then use the standard definition of marginal distributions of the k -dimensional vector $(x_1, \dots, x_k)$. However, for reasons of symmetry it will be convenient to maintain the redundancy of specifying the distribution on a k -simplex by a $(k+1)$ -dimensional vector.

Given a random vector $(X_0, \dots, X_k)$ and a partition $\{C_0, C_1, \dots, C_l\}$ of $\{0, \dots, k\}$, we define a marginal distribution of $(X_0, \dots, X_k)$ to be the distribution of $(Y_0, \dots, Y_l)$, where $Y_i = \sum_{j \in C_i} X_j$, $i = 0, \dots, l$. Thus, the marginal distribution is itself defined on an l -simplex. (Geometrically, one may think of the partition as defining a set of disjoint subsimplexes of the original simplex, and the operation of taking margins as collapsing the mass on the subsimplex into a single vertex.) The standard definition of marginal

distribution of the l -vector $(X_{j_1}, \dots, X_{j_l})$ is a special case of this more general definition with

$$\begin{aligned} C_i &= \{j_i\} \quad i=1, \dots, l \quad \text{and} \\ C_0 &= \{0, \dots, k\} \setminus \{j_1, \dots, j_l\}. \end{aligned}$$

The properties of marginal distributions then depend on the following lemma:

Lemma 4.1 For any real numbers a and b

$$\sum_{x=0}^n \binom{a}{x} \binom{b}{n-x} = \binom{a+b}{n} \quad (4.50)$$

Proof: The proof of this identity follows from multiplying the formal power series for the expressions $(1+t)^a$ and $(1+t)^b$ and equating coefficients. QED

Since this identity is true for all finite a and b it remains true for infinite a and b in the sense that the limit of the ratio of the left and right sides of (4.50) is unity.

The following theorems are well-known for the multivariate hypergeometric and multinomial cases, (see for example Bishop et al. [15, Chapter 13]) but in fact are true for all members of the multivariate Pólya family.

Theorem 4.7 Let P be a multivariate Pólya distribution with parameters $(b_0, \dots, b_k, n)$, and let P_C be the marginal distribution defined by the partition $C = (C_0, \dots, C_l)$. Then P_C is also a multivariate Pólya distribution with parameters $(b'_0, \dots, b'_l, n)$ where $b'_j = \sum_{i \in C_j} b_i$, $j = 0, \dots, l$.

Proof: Consider first the case $l = k - 1$. In this case, all the members but one of C must be singletons, and this exceptional member j will have exactly two elements i_1 and i_2 with $i_1 < i_2$. Then we have

$$\begin{aligned} P_C(y_0, \dots, y_l) &= \sum_{x=0}^{y_j} P(y_0, y_1, \dots, \underset{\substack{\uparrow \\ i_1^{\text{th}} \\ \text{position}}}{x}, \dots, \underset{\substack{\uparrow \\ i_2^{\text{th}} \\ \text{position}}}{y_j - x}, \dots, y_{k-1}) \\ &= \frac{\prod_{i \notin \{i_1, i_2\}} \binom{b_i}{y_i}}{\binom{\sum b_i}{n}} \sum_{x=0}^{y_j} \binom{b_{i_1}}{x} \binom{b_{i_2}}{y_j - x} \\ &= \frac{\prod_{i=0}^{k-1} \binom{b'_i}{y_i}}{\binom{\sum b'_i}{n}}, \end{aligned} \quad (4.51)$$

where $b'_i = b_i$ for $i < i_1$, $b'_{i_2} = b_{i_1} + b_{i_2}$, and $b'_i = b_{i+1}$ for $i_2 \neq i \geq i_1$.

Now, any marginal distribution on an l -simplex can also be written as a marginal of an $(l+1)$ -simplex, which in turn can be written as the marginal of an $(l+2)$ -simplex, and so on. For any marginal distribution, we can thus construct a sequence of distributions, each being of dimension one less than its predecessor, beginning with the original distribution on the k -simplex. Applying the above result to each step in this sequence of marginals then proves the theorem. QED

The preceding theorem shows that the marginal distributions of Pólya distributions with finite parameters remain in the same family. The same can be established for the limiting cases of multinomial distributions.

Theorem 4.8 *Let P be a multinomial distribution with parameters $(p_0, \dots, p_k, n)$, and let P_C be the marginal distribution defined by the partition $C = (C_0, \dots, C_l)$. Then P_C is also a multinomial distribution with parameters $(p'_0, \dots, p'_l, n)$ where $p'_j = \sum_{i \in C_j} p_i$, $j = 0, \dots, l$.*

Proof: The proof follows the same arguments as Theorem 4.7, replacing the identity (4.50) with

$$\sum_{x=0}^n \frac{a^x b^{n-x}}{x!(n-x)!} = \frac{(a+b)^n}{n!}. \quad (4.52)$$

QED

Conditioning on marginals also does not take us very far from the Pólya family.

Theorem 4.9 *The conditional distribution of any distribution in the multivariate Pólya family given a margin is the product of multivariate Pólya distributions.*

Proof: Given a partition $\{C_0, \dots, C_l\}$ of $\{0, \dots, k\}$, let $y_j = \sum_{i \in C_j} x_i$. Then for finite $(b_0, \dots, b_k)$

$$\begin{aligned} P(x_i, i=0, \dots, k \mid y_j, j=0, \dots, l) &= \frac{\prod_{j=0}^l \prod_{i \in C_j} \binom{b_i}{x_i}}{\binom{\sum_{i \in C_j} b_i}{y_j}} \\ &= \prod_{j=0}^l \frac{\prod_{i \in C_j} \binom{b_i}{x_i}}{\binom{\sum_{i \in C_j} b_i}{y_j}}. \end{aligned} \quad (4.53)$$

The limiting case can be handled similarly. QED

4.4.3 Geometry of the support

If the b_i are negative or infinite, then the support of the distribution coincides with the set of lists of non-negative integers $(x_0, x_1, \dots, x_k)$ such that $\sum x_i = n$. As mentioned above, these distributions can be considered to be defined on the lattice points of a k -dimensional simplex. The vertices of this simplex correspond to outcomes in which all but one of the x_i 's are zero. If the b_i 's are non-negative, then there is the further restriction that

$$0 \leq x_i < b_i + 1 \quad (i = 0, 1, \dots, k). \quad (4.54)$$

These restrictions of course have no effect if all of the $b_i \geq n$, so that in such cases the support still includes all of the lattice points of the k -dimensional simplex. If any of the b_i are less than n , however, then some of the vertices may be excluded from the support, giving a polyhedron instead of a simplex. The geometry of the support can become somewhat complicated, because the vertices of the polyhedron defined by the restriction need not be integers.

The algorithm of Section 4.3.3 can be adapted to calculating tail probabilities and expected values involving these multivariate distributions. However, to achieve this, we require more structure on the set of possible outcomes. This structure is discussed in the next chapter.

4.5 Summary

Although a number of probability distributions are used in the analysis of the sort of discrete data found in toxicological bioassays, the study of many properties of these distributions can be unified by considering them all to be members of the Pólya family. This family has its support on a set of non-negative integers, and extends in a natural way to families of multivariate discrete distributions supported on the integer lattice points of a k -dimensional simplex. The hypergeometric, binomial, and beta-binomial distributions are members of the Pólya family, and the multivariate hypergeometric and multinomial distributions are examples of the multivariate extensions.

Probabilities in the Pólya family are rational functions of the parameters of the distribution. Although in general these functions involve high-order polynomials, there exist simple relations relating the probabilities of neighboring integers. Thus, the complete probability distribution can be calculated recursively by an algorithm whose complexity is linear in the cardinality of the support.

With the exception of those distributions in the family that are U-shaped, bounds can be placed on the probabilities of the tails of the distribution. Often, a substantial part of the support will have a negligible probability, and thus approximations of expected values can be computed with a guaranteed accuracy by calculating only part of the probability distribution. This truncation could substantially reduce computational requirements in those cases where the determination of the function whose expected value is being computed itself requires extensive calculation. Such a case could arise in the determination of the unconditional power of certain permutation tests, which might be calculated as the expected value of conditional power given some random variable having a distribution in the Pólya family.

Chapter 5

Tree Structures on Discrete Multivariate Distributions

Recursive computations such as those discussed in the last chapter require an ordering on the set in which the computations are done. Although the recursive relations carry over to the multivariate Pólya family, the order inherited from the marginal distributions is only partial. A linear ordering on the support of the multivariate distribution can be constructed by defining a *tree* whose nodes represent points of the simplex. The recursive relations can then be implemented by following a tree traversal algorithm.

After an introduction to the concepts and terminology of trees, we examine in detail several tree structures that can be defined on the support of our distributions. We are mainly interested in how these structures can expedite the calculation of probabilities of certain events, in particular, of rejection regions defined by linear test statistics. This development relies on the possibilities of recursive calculations in the univariate Pólya family as well as on the Markovian nature of the multivariate family.

Even with efficient recursive algorithms, calculating the probabilities of all support points can be computationally extravagant for even moderately sized problems. Typically, however, the number of points with positive probabilities will exceed by several orders of magnitude the number of those whose aggregate probability exceeds some "threshold of negligibility". Sets that are negligible in this sense can be identified by the bounds on tail probabilities that can be evaluated during the recursive calculations. Thus, provided we are willing to accept some truncation error, only a fraction of the tree will need to be visited. The Markovian nature of the multivariate distributions, on the other hand, allows for direct computation of the probability of entire branches without the need to visit every node.

A somewhat different application of graph theoretic methods has been employed by Mehta et al. [61,62,63] for similar problems. They represent sample points as paths through a network, and use network flow arguments in the calculation of probabilities. Our development is independent of Mehta's work, and is not presented as either a generalization or an alternative. Although also motivated by computational considerations, our approach is more concerned with order relationships in the sample space which extend to the multivariate case the concept of the tail of a distribution. Much of this theory does not depend on a particular probabilistic model. Considerable simplification is possible, however, if this

model is from the multivariate Pólya family. In this case, some of the algorithms stemming from our development would be similar in spirit, if not in detail, to those presented by Mehta et al.

5.1 Configurations and Trees

5.1.1 Configurations

The basic objects of study in this chapter are lists of non-negative integers $x = (x_0, \dots, x_k)$ satisfying certain linear constraints. These lists represent the support points of the multivariate Pólya family expressed in terms of the increments (4.48), and thus the constraints are

$$\sum_{i=0}^k x_i = n, \quad (5.1)$$

and if $b_i \geq 0$,

$$x_i < b_i + 1. \quad (5.2)$$

For some purposes, the support points of the distribution are more conveniently represented in terms of the *partial sums* $s_i = \sum_{j=0}^i x_j$, which must satisfy

$$n = s_0 \geq s_1 \geq \dots \geq s_k \geq s_{k+1} = 0 \quad (5.3)$$

in addition to

$$s_{i+1} > s_i - b_i - 1 \quad (i = 0, 1, \dots, k)$$

if the b_i 's are non-negative. We will use both representations interchangeably, using x as a generic symbol for representations of the form (5.1) and s for those of the form (5.3). The term *configuration* will designate such an object in either representation.

Consider now the configuration $x = (x_0, x_1, \dots, x_k)$. One can metaphorically think of it as an array of $k+1$ boxes or *cells*, among which we have distributed n balls, x_i being the number of balls in the i^{th} cell. We will call the i^{th} coordinate of the configuration the i^{th} cell, and x_i the *content* of the i^{th} cell. If b_i is positive then we must have $x_i < b_i + 1$, which is a genuine restriction if $b_i \leq n - 1$. If $0 \leq b_i < n$ we will call v_i , the smallest integer not exceeded by b_i , the *capacity* of cell i . Otherwise, the capacity is irrelevant, but for definiteness we will define all the capacities to be equal to n in this case. If the content of a cell is zero it will be called *empty*, and if it is equal to the capacity, it is called *full*. Note that it is possible for some capacities to be zero, in which case the corresponding cells must be empty. Cells with zero capacity will be called *null*. Thus null cells are at the same time empty and full.

We can thus formally define a set of configurations of n balls in $k+1$ cells with capacities $(v_0, \dots, v_k)$ as

$$\left\{ x = (x_0, \dots, x_k) : \sum_{i=0}^k x_i = n \text{ and } 0 \leq x_i \leq v_i, \quad i = 0, \dots, k \right\}. \quad (5.4)$$

In calculating probabilities involving the multivariate distributions, we first require an algorithm for generating all the configurations in a manner that is both efficient and avoids

double-counting. Although naively one could generate all lists of integers not exceeding n and reject those that fail to meet the constraints, such an approach is both unaesthetic and computationally extravagant. The methods we will use take advantage of a tree structure defined on the set of configurations. This tree will provide a convenient algorithm for stepping through the configurations as well as simplifying the probability calculations.

5.1.2 Trees

In graph theory, a *tree* is defined as a special kind of graph. A *graph* is a finite set X of objects called *nodes* along with a symmetric relation $E \subset X \times X$ whose elements are called *edges*. A *path* of length ℓ from x to y is a finite sequence of distinct nodes $x_1, x_2, \dots, x_\ell$ such that $x_\ell = y$ and $(x_{i-1}, x_i) \in E$ for $i = 1, \dots, \ell$, where $x_0 = x$. A *cycle* is a path from x to x of length greater than two. A graph is *connected* if for any two nodes x and y , there is a path from x to y . A connected graph without cycles is called a *tree*. It is elementary that in a tree there is exactly one path between any two distinct nodes.

We can impose more structure on the tree by designating a distinguished node called the *root*. In that case we have a *rooted tree*. All the trees that will be discussed will have roots, so by the term *tree* we always mean a *rooted tree*. The root of a tree X will be designated by $\text{rt } X$. A rooted tree defines a partial order on its nodes, where we say $x < y$ if $x \neq y$ but x lies on the path from the root to y . ($x \leq y$ will mean $x < y$ or $x = y$.) Clearly, the root is the smallest node. If x is not the root, then there will be a node $\hat{x} < x$ such that $y < x$ implies that $y \leq \hat{x}$. This node will be called the *parent* of x . We thus have a *parental mapping* $\pi : x \mapsto \hat{x}$ with domain $X \setminus \text{rt } X$. Conversely, a parental mapping will define the structure of the tree, for we will have $(x, y) \in E$ if and only if $x = \pi(y)$ or $y = \pi(x)$. In fact, we will define all our trees in terms of parental mappings. To do this, we need some assurance that such a mapping does in fact define a tree. And to derive that we first require the following lemma.

Lemma 5.1 Consider a graph (X, E) where E is defined by a function $\pi : X_0 \subset X \rightarrow X$ such that $(x, y) \in E$ iff either $x = \pi(y)$ or $y = \pi(x)$. Let $x = x_0, x_1, \dots, x_n = y$ be a path of length $n > 2$ from x to y . Then if for some j , $(1 \leq j \leq n)$, $x_{j-1} = \pi(x_j)$, then $x_{i-1} = \pi(x_i)$ for all $i \geq j$.

Proof: Consider the set $\{i : j \leq i \leq n \text{ such that } x_{i-1} \neq \pi(x_i)\}$. If this set is non-empty, it has a least element k . By definition $j \neq k$, and $x_{i-1} = \pi(x_i)$ for all $i < k$. However, since $\pi(x_k) \neq x_{k-1}$ we must have $x_k = \pi(x_{k-1})$. On the other hand $k-1 < k$ and so $\pi(x_{k-1}) = x_{k-2}$. This implies that $x_k = x_{k-2}$ which contradicts the requirement that the nodes of the path be distinct. QED

Theorem 5.1 Let $\pi : X_0 \rightarrow X$ be a mapping defined on some subset X_0 of a finite set X . Then π is the parental mapping of a tree with X as nodes if and only if the following two conditions hold:

- (a) $X \setminus X_0$ consists of exactly one element.
- (b) There is some real-valued function $t : X \rightarrow \mathbb{R}$ such that $t \circ \pi(x) < t(x)$ for all $x \in X_0$.

Proof: The necessity of (a) follows immediately from the definition of a parental mapping. The necessity of (b) follows by letting $t(x)$ be the length of the path from the root to x .

To prove sufficiency, consider any path $x = x_0, x_1, \dots, x_n = y$ from x to y . Suppose that $x_0 = \pi(x_1)$. Then by the preceding lemma, we have that $x_{i-1} = \pi(x_i)$ and hence $t(x_{i-1}) < t(x_i)$ for all i . By induction, this implies that $t(x_0) < t(x_n)$ and thus x and y cannot be equal. If, on the other hand, $x_1 = \pi(x_0)$, then either $x_1 = \pi(x_{i-1})$ for all i , or there is a j such that $\pi(x_j) = x_{j-1}$. In the former case $t(x_i)$ is decreasing, so that $t(x) > t(y)$ and $x \neq y$. In the latter case, the above lemma implies that $\pi(y) = x_{n-1}$. But since $\pi(x) = x_1$ and $x_1 \neq x_{n-1}$ this precludes $x = y$. Thus no path can be a cycle.

To show connectedness we will construct a path between two elements x and y . Consider the sequence defined by $x_0 = x$ and $x_i = \pi(x_{i-1})$. Since $t(x_i)$ is decreasing, the elements of this sequence must all be distinct. If any of these nodes is y , then we have constructed a path. Otherwise since X is a finite set, this sequence must terminate with an element $\pi^k(x)$ such that $\pi(\pi^k(x))$ is not defined. Therefore $\pi^k(x) \in X \setminus X_0$. Consider now a similar sequence defined by $y_0 = y$ and $y_i = \pi(y_{i-1})$. This sequence must similarly terminate with a node $\pi^l(y)$ such that $\pi(\pi^l(y))$ is undefined. But since $X \setminus X_0$ has, by hypothesis, only a single element, we have that $\pi^l(y) = \pi^k(x)$. Thus the set $\{i : \pi^i(y) = \pi^j(x) \text{ for some } j\}$ is nonempty and has a least element m . Then the sequence $x, \pi(x), \dots, \pi^m(x) = \pi^m(y), \pi^{m-1}(y), \dots, \pi(y), y$ is a path from x to y . QED

The inverse image of the parental mapping $\pi^{-1}(x)$ is referred to as the set of *offspring* of x . Nodes that have no offspring are called *terminal nodes*, or *leaves*. The set $B_x = \{y \in X : \pi^k(y) = x \text{ for some } k\}$ can be considered to be a tree itself with root x and parental mapping the restriction of π to $B_x \setminus \{x\}$. This *subtree* is called a *branch* springing from x , and its nodes will be called the *descendants* of x . The node x itself will be called the *patriarch* of the branch. (We apologize for the promiscuous mingling of botanical and genealogical metaphors, which is not uncommon in this topic.)

We will use the tree structure to define a systematic way of performing (at least in principle) some operation on all its nodes. We will call such a generic operation *enumeration*. (This use of the word is motivated by the idea of enumerating people for a survey or election, but is at variance with its use by graph theorists who use it as a synonym for "counting".) Computer scientists describe techniques for enumerating the nodes as *tree traversal algorithms* [77]. These methods require the inverse of the parental mapping, so that for any node, the set of its offspring can be determined. The enumeration is then achieved using the concept of a *stack*.

Intuitively, a stack is the analogue of a stack of papers. New papers may be added to the top of the stack, or old ones may be removed. Thus the last sheet added is the first one to be removed. Mathematically, a stack could be defined as a sequence of ordered finite sets A_i such that either $A_i \subset A_{i+1}$ or $A_{i+1} \subset A_i$ and in either case $x \in A_i \cap A_{i+1}$ and $y \in A_i \Delta A_{i+1}$ implies that $x < y$.¹ (Note that the order relation here is different from that implied by the tree structure.) Computationally, a stack is represented by an array with a pointer pointing to its "largest" element. Items are added to the "top of the stack" by incrementing the pointer. Items are "popped" from the stack by decrementing the pointer.

¹The symbol Δ denotes the symmetric difference operation on sets.

The algorithm begins by placing the root onto the (empty) stack. It then proceeds by enumerating the top element of the stack and then replacing it by the set of its offspring (which might be empty). These steps are then repeated until the stack is once again empty.

Metaphorically, we can think that after enumerating the root, the algorithm chops it off, thereby breaking off its branches. Since each branch is itself a tree, the algorithm proceeds to enumerate each of these subtrees, and the process repeats itself. Eventually it will reach a subtree consisting of a single node, and after it is chopped, there is nothing left. It can then go back to enumerate any other subtrees that are still waiting. It is clear that every node has to be enumerated in this fashion, and no node will appear more than once.

Our main goal in playing with trees is the calculation of probabilities. We can continue the metaphor by attaching a weight (i.e. probability) to each node, and think of our task as being the weighing of the tree (or some part of it). This task can be facilitated if the tree structure has certain properties:

1. It should be a simple computation to derive the weight of the offspring from that of its parent.
2. It should be possible to calculate the weight of a branch without having to weigh each of its nodes.
3. Failing the above, it should be possible to put a simple upper (or lower) bound on the weight of a branch.
4. The parts of the tree we want to weigh would represent large branches.

If these properties hold, then we can think of our task as one of trimming the tree (i.e. cutting off its branches) and weighing the trimmings. Using property 3, we may also decide that certain trimmings are negligible, so not worth weighing. (We could call these "twigs".)

5.1.3 Trees of configurations

We now present some tree structures on the set of configurations which satisfy these properties. First we need some more terminology to make our descriptions precise, and some more metaphors to make them understandable.

Let X be the set of configurations $(x_0, x_1, \dots, x_k)$ satisfying $\sum x_i = n$ and $0 \leq x_i \leq v_i$. If the capacities v_i are all greater than n , we can define a parental mapping on X metaphorically as follows: Take a ball from the last non-empty cell and toss it one cell to the left. To simplify the mathematical description, we will use a circumflex to denote parenthood—i.e., we define $\hat{x} = \pi(x)$, and for any symbol denoting an attribute of a configuration we put a circumflex on that symbol to denote the corresponding attribute of the parent. We can now define the parental mapping on configurations mathematically by

$$\begin{aligned}\hat{x}_m &= x_m - 1, \\ \hat{x}_{m-1} &= x_{m-1} + 1, \text{ and} \\ \hat{x}_i &= x_i \text{ if } i \notin \{m, m-1\},\end{aligned}$$

where $m = \max\{i : x_i > 0\}$. Clearly a configuration will have a parent unless the last non-empty cell has no left neighbour—i.e. it is the first cell. In that case, we must have $x = (n, 0, \dots, 0)$, and this is the root configuration.

If some of the capacities do impose restrictions, then the definition of the parental mapping must be modified: We take the last non-empty cell, and toss one ball from it to the closest *unfull* cell to its left. Symbolically, we let

$$j = \max\{i < m : x_i < v_i\}$$

and then define

$$\begin{aligned}\hat{x}_m &= x_m - 1, \\ \hat{x}_j &= x_j + 1, \text{ and} \\ \hat{x}_i &= x_i \text{ if } i \notin \{j, m\}.\end{aligned}$$

Again a configuration will have a parent unless all the cells to the left of the last non-empty cell are full. This requires that $x_0 = v_0, x_1 = v_1, \dots, x_{m-1} = v_{m-1}$, and consequently $x_m = n - \sum_{i=0}^{m-1} v_i$. These requirements define a unique configuration which is the root.

It is instructive to consider the definition of the parental mapping in terms of the partial sums s . Clearly, $s_i = 0$ for $i > m$, so that $s_m = x_m$, and $\hat{s}_m = s_m - 1$. Moreover, $\hat{s}_j = \hat{x}_j + \dots + \hat{x}_m = x_j + 1 + \dots + x_m - 1 = s_j$. But for $j < i < m$, $\hat{s}_i = s_i - 1$.

The parental mapping in terms of the s 's is particularly simple in the case $j = m - 1$, (which always obtains if $v_i \geq n$). Just decrement by 1 the last non-zero component. To describe this mapping in the more complicated case in which $j < m - 1$, we define the concept of *minimal content*.

A cell i is *minimal* if $s_i = 0 \vee s_{i-1} = v_{i-1}$. Obviously, a non-empty cell is minimal if and only if its left neighbour is full. The case $j < m - 1$ occurs if the last non-empty cell is minimal. This means that s_m cannot be decremented because to do so would cause its left neighbour to exceed its capacity. Therefore, we step backwards from the m^{th} cell until we find a non-minimal cell ($j + 1$), and we decrement s_{j+1} and all s 's to the right up to s_m . (The monotonicity restraint on the s 's guarantees that $s_i > 0$ for $j < i < m$.)

To prove that the parental mapping we have defined does in fact define a tree, we need to display a real-valued function that is monotone in the partial order implied by the tree. One such function is

$$t(x) = \sum_{i=0}^k ix_i = \sum_{i=1}^k s_i$$

since

$$t(\hat{x}) = \sum_{i=1}^j s_i + \sum_{i=j+1}^m (s_i - 1) = t(x) - (m - j) < t(x).$$

To implement the algorithm, we require the inverse of the parental mapping, so that we can define offspring and branches. This is given by the following theorems.

Theorem 5.2 *Let x be a configuration in which the last non-empty cell m is not full. Let*

$$J_1 = \{j < m : x_j > 0 \text{ and } j < i < m \implies x_i = v_i\}.$$

Then for any $j \in J_1$ the configuration y defined by

$$y_j = x_j - 1, \quad y_m = x_m + 1, \text{ and } y_i = x_i \text{ for } i \notin \{j, m\}$$

is an offspring of x .

Proof: Since $x_j > 0$ and $x_m < v_m$, y will be a legitimate configuration. By definition m is the last non-empty cell of y , and by definition of j , there are no unfull cells between j and m . Thus the parental mapping will yield $\hat{y}_j = y_j + 1 = x_j$ and $\hat{y}_m = y_m - 1 = x_m$, and thus $\hat{y} = x$. QED

Theorem 5.3 Let m be the last non-empty cell of a configuration x . Suppose the set $M = \{i > m : v_i > 0\}$ is non-empty, and let $\bar{m}$ be the minimum of this set. Let

$$J_2 = \{j < \bar{m} : x_j > 0 \text{ and } j < i < \bar{m} \implies x_i = v_i\}.$$

Then for any $j \in J_2$ the configuration y with

$$y_j = x_j - 1, \quad y_{\bar{m}} = 1, \text{ and } y_i = x_i \text{ for } i \notin \{j, \bar{m}\}$$

is an offspring of x .

Proof: Again, the definitions of j and $\bar{m}$ ensure that y is a legitimate configuration. Moreover, $\bar{m}$ is the last non-empty cell of y , and there are no unfull cells between j and $\bar{m}$. Thus $\hat{y}_j = y_j + 1 = x_j$ and $\hat{y}_{\bar{m}} = y_{\bar{m}} - 1 = 0 = x_{\bar{m}}$. QED

These two theorems present two types of offspring, which could be termed daughters and sons respectively. There will be no daughters if either v_m is full or J_1 is empty. There will be no sons if M is empty (e.g. if $m = k$), but J_2 will never be empty since it will at least contain m . If there are any daughters, there can be only one son, since in this case $J_2 = \{m\}$. If all the capacities are greater than n , so that no cells are full, then each node has at most one daughter and one son.

Theorem 5.4 A node has no offspring besides the ones described in Theorems 5.2 and 5.3.

Proof: Let x be a configuration having m as the last non-empty cell, and let y be another configuration with $\hat{y} = x$. Let $\bar{m}$ be the last non-empty cell of y . By definition, $\hat{y}_m = y_m - 1$, and $\hat{y}_j = y_j + 1$ where $j = \max\{i < \bar{m} : y_i < v_i\}$. If $y_m > 1$ then $x_m = \hat{y}_m > 0$ and thus $m = \bar{m}$. In this case, moreover, $\hat{y}_i = y_i = v_i$ for $j < i < m$ so that $j \in J_1$ and y is a daughter of x . On the other hand, if $y_m = 1$ then $x_m = \hat{y}_m = 0$ and $m < \bar{m}$. Moreover $j \leq m$ by the definition of m , and $\hat{y}_i = y_i = v_i$ for $j < i < \bar{m}$. Thus $j \in J_2$ and y is a son of x . QED

At the risk of introducing too many metaphors, we will introduce some more terminology to facilitate the description of offspring. Any offspring is defined by decrementing one cell and incrementing a later cell—i.e., "tossing one ball to the right". The cell which is decremented will be called the *pitcher* of the offspring and the one that is incremented

will be called the *catcher*. Thus finding all the offspring of a configuration amounts to determining all the catchers and their corresponding pitchers. The above theorems indicate that there are at most two catchers, either the last non-empty cell or its non-null right neighbour, but a variable number of pitchers. We will encounter more complicated cases later.

A daughter can only exist if the last non-empty cell is not full. In that case, the only son will have this cell as the pitcher. Thus, for any configuration, distinct offspring will have distinct pitchers, and a particular offspring can be identified by its pitcher. To find all the offspring, we just need to step backwards from the last non-empty cell and ask if that cell can be a pitcher. The answer will be affirmative provided the cell is non-empty and there are no unfull cells to the right before the last non-empty cell. Once such an unfull cell appears, all the offspring have been found since no cell to the left can be a pitcher.

These considerations help to determine the maximum size of the stack required in implementing the tree traversal algorithm. But first we need the following lemma.

Lemma 5.2 *Let x be the offspring of a configuration $\hat{x}$ with pitcher j . Then none of the descendants of x can have a pitcher less than j .*

Proof: Let m be the last non-empty cell (catcher) of x , and consider an offspring $\hat{x}$ of x with catcher m . Clearly $j < m$ and $x_j = \hat{x}_j - 1 < \hat{x}_j \leq v_j$. Now let $\hat{j}$ be a pitcher of $\hat{x}$. Then necessarily $\{r : \hat{j} < r < m \text{ and } x_r < v_r\}$ is empty. However, $\hat{j}$ will belong to this set unless $\hat{j} \geq j$. Similarly, for an offspring $\hat{\hat{x}}$ of $\hat{x}$, the pitcher must satisfy $\hat{\hat{j}} \geq \hat{j} \geq j$, and the result follows by induction. QED

Theorem 5.5 *Suppose that in implementing the tree traversal algorithm, offspring are placed on the stack in order of increasing pitchers. Then the number of entries in the stack is always less than the number of cells in the configuration.*

Proof: To prove the theorem we show that all the pitchers in the stack at any one time form a strictly increasing sequence, and hence must designate distinct cells. Suppose the stack contains one entry, with pitcher j . Then after enumeration, this entry will be replaced by all its offspring, with pitchers in increasing order. Obviously, removing entries from the stack cannot affect this monotonicity. We now show that adding to it cannot either.

Suppose that the stack contains ν entries with pitchers $j_1 < j_2 < \dots < j_\nu$. After enumeration, the top entry will be replaced by its offspring, with pitchers $j_1 < j_2 < \dots < j_\nu$. But by Lemma 5.2 $j_\nu \leq j_1$. Hence the new stack will have pitchers $j_1 < j_2 < \dots < j_{\nu-1} < j_1 < j_2 < \dots < j_\nu$ in strictly ascending order. Since the pitchers thus represent distinct cells, and since the right-most cell can never be a pitcher, the theorem follows. QED

This theorem assures us that in implementing the enumeration algorithm, it is sufficient to reserve k array locations for defining the stack, regardless of the magnitude of n .

5.2 Recursive Calculations of Probabilities

The structures discussed in the preceding section do not depend on any probability distribution defined on the set of configurations. However, if such a distribution belongs to the multivariate Pólya family then the recursive calculations discussed in Section 4.2.1 generalize readily to the multivariate case. Although the computational labour of calculating probabilities is substantially reduced by this feature, it may still be excessive if the support has many points, as it will if either n or k is large. The computations with large n can be reduced by employing tail truncation as discussed in Section 4.2.3. The case with large k can be handled by exploiting the conditional form of the distribution given by (4.46). This form provides a method for calculating the probability for an entire branch without the need to visit every node.

5.2.1 Weighing offspring

One of the desiderata enunciated on page 68 was the ability to calculate the probabilities of offspring recursively from the probabilities of their parents. We now examine to what extent this property is satisfied for the defined tree structure.

From (4.48) the joint probability of a configuration is

$$P(x_0, x_1, \dots, x_k) = \frac{\prod_{i=0}^k \binom{b_i}{x_i}}{\binom{a_0}{n}} \quad (5.5)$$

The parental configuration will have the same values of the x_i 's except for the pitcher j and the catcher m . Thus all factors in the product will remain the same except for $\binom{b_j}{x_j}$ which will become $\binom{b_j}{x_j + 1}$ and $\binom{b_m}{x_m}$ which will become $\binom{b_m}{x_m - 1}$. Since

$$\binom{r}{s+1} = \binom{r}{s} \frac{(r-s)}{(s+1)},$$

$$\binom{b_j}{x_j + 1} \binom{b_m}{x_m - 1} = \binom{b_j}{x_j} \binom{b_m}{x_m} \frac{b_j - x_j}{x_j + 1} \cdot \frac{x_m}{b_m - x_m + 1}.$$

Thus

$$P(\hat{x}) = P(x) \frac{x_m}{x_j + 1} \cdot \frac{b_j - x_j}{b_m - x_m + 1}, \quad (5.6)$$

or alternatively

$$P(x) = P(\hat{x}) \frac{\hat{x}_j (b_m - \hat{x}_m)}{(\hat{x}_m + 1)(b_j - \hat{x}_j + 1)}. \quad (5.7)$$

Thus the probability of an offspring can be determined from the probability of its parent by two multiplications and two divisions.

This formula extends quite readily to the limiting case where $b_j, b_m \rightarrow \infty$. In this case, the only possible pitcher is $j = m - 1$, so that

$$\frac{b_m - \hat{x}_m}{b_{m-1} - \hat{x}_{m-1} + 1} = \frac{\frac{b_m}{b_{m-1}} - \frac{\hat{x}_{m-1}}{b_{m-1}}}{1 - \frac{\hat{x}_{m-1} - 1}{b_{m-1}}}$$

Now, using the reparametrization (4.47),

$$\begin{aligned} \frac{b_m}{b_{m-1}} &= \frac{a_m - a_{m+1}}{a_{m-1} - a_m} \\ &= \frac{(a_m/a_{m-1})(1 - a_{m+1}/a_m)}{1 - a_m/a_{m-1}} \\ &= \frac{\mu_m(1 - \mu_{m+1})}{1 - \mu_m}, \end{aligned}$$

so that as $b_m, b_{m-1} \rightarrow \infty$ we have that

$$P(x) = P(\hat{x}) \frac{\hat{x}_{m-1} \mu_m (1 - \mu_{m+1})}{(\hat{x}_m + 1)(1 - \mu_m)} \quad (5.8)$$

with the usual convention that $\mu_{k+1} = 0$.

Accumulated round-off error is always a concern with recursive calculations. The maximum round-off error is proportional to the number of steps in the recursion. The maximum number of steps required to calculate a probability is the longest path from the root to any node, called the *height* of the tree. A bound on the height can be obtained by considering the function $t(x) = \sum_{i=0}^k i x_i = \sum_{i=1}^k s_i$. This function is integer valued and increases with each offspring. Thus the longest path from the root cannot exceed the difference between the maximum and minimum values of this function. These values do not have simple expressions, but clearly $0 \leq t(x) \leq nk$, so that nk can be used as an upper bound on the number of recursions. (This bound should be fairly tight unless the v_i 's are small compared to n .)

The number of operations involved in calculating the probabilities of all configurations is proportional to the number of configurations. If the b_i 's are large or negative, then the number of configurations is $\binom{n+k}{k}$ which is of the order of n^k for large n , and of the order of k^n for large k . Thus the tree traversal algorithm is of polynomial complexity in either parameter if the other stays fixed, but is exponential if both parameters increase. (In our applications, k will usually be a small integer.)

There does not appear to be a closed formula for the number of configurations in the presence of bounds on the cell capacities. However, this number does not grow indefinitely with n , since to every configuration $(x_0, x_1, \dots, x_k)$ of n balls there corresponds another configuration $(v_0 - x_0, v_1 - x_1, \dots, v_k - x_k)$ of $\sum v_i - n$ balls constrained by the same capacities.

5.2.2 Weighing complete branches

The second desideratum for trees was that one could weigh a complete branch without enumerating all of its nodes. This is in fact possible, although to do that in the general

case requires some more structure, which we will now develop.

In Section 5.1.2 we defined B_x , the branch springing from a node (configuration) x . Alternatively, we can consider the branch as the set of all nodes $y \geq x$ under the partial order induced by the tree. We give now another characterization of branches which is useful. For any node x we shall say that a non-empty set A of nodes is a *covering set* of x if for any $y \in A$ such that $y \neq x$, $\hat{y} \in A$. We then have

Lemma 5.3 *If A is a covering set of x then $x \in A \subset B_x$.*

Proof: Since A is nonempty it has an element y . Consider the sequence $y, \hat{y}, \hat{\hat{y}}, \dots$ of distinct nodes, where $\hat{y}$ as usual denotes the parent of y . If all the terms of this sequence were different from x , then the definition of the covering set implies that the entire sequence must belong to A . However, the sequence will eventually reach the root r of the tree, but we cannot have $r \in A$ since r is undefined. We therefore conclude that one of the terms of the sequence is x , that is, $y \in B_x$. Since y was an arbitrary element of A , we have $A \subset B_x$.

Now if y is an element of A distinct from x , then $x = \pi^\ell(y)$ for some $\ell > 0$ and hence $x = \hat{z}$ where $z = \pi^{\ell-1}(y) \in A$. Thus $x \in A$. QED

Lemma 5.4 *The branch springing from x is the largest covering set of x .*

Proof: We have already shown that every covering set is a subset of the branch. It remains to show that a branch is a covering set. Obviously the branch is non-empty since it contains x . Moreover, if $y \in B_x$ then $x = \pi^\ell(y)$ for some $\ell \geq 0$, and $\ell > 0$ if $y \neq x$. Therefore $x = \pi^{\ell-1}(\pi(y)) = \pi^{\ell-1}(\hat{y})$ and thus $\hat{y} \in B_x$. QED

Lemma 5.5 *Two branches are either disjoint or one is wholly contained in the other.*

Proof: If $x \in B_y$ then clearly $B_x \subset B_y$. Suppose that $x \notin B_y$, and let $z \in B_x \cap B_y \neq \emptyset$. Then if $z \neq y$ then $\hat{z} \in B_y$. Also, since $z \neq x$, we have $\hat{z} \in B_x$. Thus $\hat{z} \in B_x \cap B_y$ and $B_x \cap B_y$ is a covering set for y . Therefore by Lemma 5.3 $y \in B_x \cap B_y$ and $B_y \subset B_x$. QED

We now characterize the branches of the concrete trees of the configurations which we have been studying. This characterization is simplest in terms of the partial sums $s = (s_0, s_1, \dots, s_k)$. Recall that a minimal cell is one for which $s_i = 0 \vee s_{i-1} = v_{i-1}$. Before we proceed we need the following lemma.

Lemma 5.6 *Let x and y be configurations with $x > y$. Let j and j^* be the last non-minimal cells of x and y respectively. Then $j \geq j^*$.*

Proof: Let $m = \max\{i : x_i > 0\}$. Then clearly $j \leq m$, and for any i such that $j < i \leq m$ the partial sums of the x_i 's satisfy

$$s_i = s_{i-1} - v_{i-1}.$$

Moreover, for such i , $\hat{s}_i = s_i - 1 = s_{i-1} - 1 - v_{i-1} = \hat{s}_{i-1} - v_{i-1}$, and thus i is still minimal in the parent. Therefore $\hat{j}$, the last non-minimal cell in the parent, cannot satisfy $j < \hat{j} \leq m$. We thus have $\hat{j} \leq j$. Similarly, for $\hat{\hat{j}}$, the last non-minimal cell in the grandparent, we have $\hat{\hat{j}} \leq \hat{j} \leq j$. By induction, it follows that $j^* \leq j$. QED

Theorem 5.6 Let x be a configuration with partial sums $s = (s_0, s_1, \dots, s_k)$, and let $J = \{i > 0 : s_i > 0 \vee s_{i-1} - v_{i-1}\}$ be the set of non-minimal cells. If J is empty then x is the root, and the branch springing from x is the entire tree. Otherwise, let $j = \max J$. Then the branch springing from x is the set of configurations with partial sums s^* , such that $s_i^* \geq s_i$ for $i \geq j$ and $s_i^* = s_i$ for $i < j$.

Proof: If the set J is empty, then all cells but the last one are full. Hence the parental mapping is undefined, and the configuration is the root. Now suppose that J is non-empty and let

$$A = \{s^* : s_i^* = s_i \text{ for } i < j \text{ and } s_i^* \geq s_i \text{ for } i \geq j\}.$$

First we will show that $B_x \subset A$.

Suppose that $s^* \in B_x$. Then there must be a chain $s^* > \hat{s}^* > \hat{\hat{s}}^* \dots > s$. Now, from the definition of the parental mapping it follows that

$$\hat{s}_i^* = s_i^* - 1 \quad \text{for } i \in D$$

and

$$\hat{s}_i^* = s_i^* \quad \text{for } i \notin D$$

where

$$\begin{aligned} D &= \{i : j^* \leq i \leq m^*\}, \\ j^* &= \max\{i : s_i^* > s_{i-1}^* - v_{i-1} \vee 0\}, \text{ and} \\ m^* &= \max\{i : s_i^* > 0\}. \end{aligned} \quad (5.9)$$

Thus for any cell $s_i^* \geq \hat{s}_i^* \geq \hat{\hat{s}}_i^* \dots \geq s_i$. Moreover by Lemma 5.6, $j \leq j^*$ and thus for $i \leq j$ we have $s_i^* = \hat{s}_i^* = \dots = s_i$. Thus $s^* \in A$.

We next show that A is a covering set for x . Then by Lemma 5.3 it will follow that $A \subset B_x$, and the theorem will be proved.

Obviously $s \in A$. Now let $s^* \in A$ where $s^* \neq s$. Define j^* , m^* , and D as in (5.9). Then the parent of s^* is defined by

$$\hat{s}_i^* = s_i^* - 1 \quad \text{for } i \in D$$

and

$$\hat{s}_i^* = s_i^* \quad \text{otherwise.}$$

Then $\hat{s}^*$ will be in A if $s_i^* > s_i$ for $i \in D$.

Let $G = \{i : s_i^* > s_i\}$. We will show that $D \subset G$. Since $s^* \neq s$, $G \neq \emptyset$, and thus we can define $r = \min G$. Then

$$s_r^* > s_r \geq s_{r-1} - v_{r-1} \vee 0 = s_{r-1}^* - v_{r-1} \vee 0$$

and thus $r \leq j^*$. Now suppose that $i \in G$ and $i < m$. Then $s_{i+1}^* \geq s_i^* - v_i > s_i - v_i$. However, from the definition of A it follows that $j \leq r \leq i$ and hence $s_{i+1} = s_i - v_i \vee 0 = s_i - v_i$ since $s_{i+1} > 0$. Thus $s_{i+1}^* > s_{i+1}$ and $i+1 \in G$. On the other hand, if $m \leq i < m^*$, then $s_{i+1}^* > 0 = s_{i+1}$ so again we have $i+1 \in G$. Thus $G = \{i : r \leq i \leq m^*\}$. Moreover, $s_r^* > s_r \geq s_{r-1} - v_{r-1} \vee 0 = s_{r-1}^* - v_{r-1} \vee 0$, and thus $r \leq j^*$.

Therefore $D \subset G$, and $s^* \in A$. QED

Theorem 5.6 shows that it is particularly simple to calculate the probability of the branch springing from a configuration. Using the notation of (4.46)

$$\begin{aligned}
 P(B_s) &= \sum_{s^* \geq s} P(s^*) \\
 &= \sum_{s^* \geq s} \prod_{i=1}^k p_i(s_i^* | s_{i-1}^*) \\
 &= \sum_{s_j^* = s_j}^{s_{j-1}^*} \sum_{s_{j+1}^* = s_{j+1}}^{s_j^*} \cdots \sum_{s_k^* = s_k}^{s_{k-1}^*} \prod_{i=1}^k p_i(s_i^* | s_{i-1}^*) \\
 &= \left(\prod_{i=1}^{j-1} p_i(s_i^* | s_{i-1}^*) \right) \left(\sum_{s_j^* = s_j}^{s_{j-1}^*} p_j(s_j^* | s_{j-1}^*) \right. \\
 &\quad \times \sum_{s_{j+1}^* = s_{j+1}}^{s_j^*} p_{j+1}(s_{j+1}^* | s_j^*) \cdots \sum_{s_k^* = s_k}^{s_{k-1}^*} p_k(s_k^* | s_{k-1}^*) \left. \right)
 \end{aligned} \tag{5.10}$$

where j is defined as in Theorem 5.6 and with the understanding that $p_i(u|v) = 0$ if $v > u$. However, for $i > j$, s_i is minimal, so that

$$\sum_{s_i^* = s_i}^{s_{i-1}^*} p_i(s_i^* | s_{i-1}^*) = 1. \tag{5.11}$$

Thus

$$P(B_s) = \prod_{i=1}^{j-1} p_i(s_i^* | s_{i-1}^*) \sum_{s_j^* = s_j}^{s_{j-1}^*} p_j(s_j^* | s_{j-1}^*). \tag{5.12}$$

The summation only involves the tail probability of a univariate distribution as discussed in Section 4.2, and thus the same recursive calculations and truncations can be performed in its computation. (The expression (5.10) is in essence the idea behind the network algorithms of Mehta et al. (op. cit.) in which each choice of the indices in the sequence of summations corresponds to a path through the network.)

The formula (5.12) also gives a means for putting an upper bound on the weight of the branch. Since none of the factors exceed unity, the probability will be bounded above by any partial product. Thus in calculating the product, we can abandon the calculation whenever the product drops below ϵ , for in that case we know that the entire branch will have probability less than ϵ .

5.3 Probabilities of Critical Regions

5.3.1 Linear test statistics

We will chiefly be interested in calculating the probabilities of linear halfspaces of the form

$$C = \{x : \sum w_i x_i \geq c\} \quad (5.13)$$

for some constant c and list of weights $w = \{w_0, w_1, \dots, w_k\}$. Although the set of configurations defined by this inequality depends on c and w , the correspondence is not one-to-one. If $u_i = \alpha + \beta w_i$ then $\sum u_i x_i = \alpha \sum x_i + \beta \sum w_i x_i = \alpha n + \beta \sum w_i x_i$, and the set $\{x : \sum w_i x_i \geq c\} = \{x : \sum u_i x_i \geq \alpha n + \beta c\}$. Thus an arbitrary origin shift and rescaling of weights can be made without loss of generality. It is convenient to choose $w_0 = 0$, so that we get

$$\sum_{i=0}^k w_i x_i = \sum_{i=1}^k w_i (s_i - s_{i+1}) = \sum_{i=1}^k (w_i - w_{i-1}) s_i.$$

If the sequence $(w_0, w_1, \dots, w_k)$ is nondecreasing, then $w_i - w_{i-1} \geq 0$ and it follows that the function

$$t : x \mapsto \sum w_i x_i$$

is also nondecreasing in terms of the partial order defined by the tree on the set of configurations. Thus the region $C = t^{-1}(c)$ will define a set of branches. The set of patriarchs of these branches will be the nodes x with $t(x) \geq c$ but $t(\hat{x}) < c$. Such nodes can be found by stepping through the tree using a modified version of the tree-visiting algorithm.

Once a node with $t(x) \geq c$ is found then all its descendants will similarly have $t(x) \geq c$ and there is no need to explicitly enumerate them. Thus the patriarch can be removed from the stack, and another branch traced to find another patriarch. This procedure will thus trace through the complement of C , and its probability can be accumulated using the recursive relations (5.7). If in fact C is a smaller set than its complement, then the calculation can be shortened by working with a "reversed" tree. That is, we enumerate the configurations x' with $t'(x') \leq c'$, where $x'_i = x_{k-i}$, $t'(x') = \sum (w_k - w_{k-i}) x'_i$ and $c' = w_k n - c$. This enumeration proceeds like the previous one, except that we include in our enumeration nodes having $t(x) = c$.

In general, there does not seem to be a simple procedure for identifying the patriarchs directly, without enumerating their ancestors, although heuristic search algorithms could be devised. However, in the special case that $w_i = i$, such a procedure can be devised, as will be discussed in the next section. In this case we will call the number $t(x) = \sum i x_i$ the *score* of the configuration x .

Actually, the case $w_i = i$ is not so special, since any set of *rational* increasing weights can be rescaled to this form by the artifice of adding dummy cells having capacities of zero. Let $w_i = r_i/q_i$ for integers r_i, q_i , where $r_0 = 0$ and q_i does not divide r_i . Let q be the least common multiple of the q_i , and let r be the greatest common divisor of the set of integers $\{q r_i/q_i, i = 1, \dots, k\}$. Let $k^* = (q r_k)/(r q_k)$ and define

$$\begin{aligned} v_j^* &= v_i & \text{if } j = \frac{q r_i}{r q_i}, \\ v_j^* &= 0 & \text{otherwise,} \end{aligned} \quad (j = 0, 1, \dots, k^*).$$

Then the set of configurations y with n balls in $k^* + 1$ cells with capacities $(v_0^*, \dots, v_k^*)$ corresponds exactly to configurations x of $k + 1$ cells with capacities $(v_0, \dots, v_k)$ except for the interposition of empty cells. Moreover,

$$\{x : \sum_{i=1}^k w_i x_i \geq c\} = \{y : \sum_{j=1}^{k^*} j y_j \geq c q / r\}.$$

The tree of configurations can be drawn such that the height of a node above a baseline ("the ground") is equal to the value of the function t . Because of the monotonicity of t , the branches of the tree will be always ascending (see Figure 5.1). The set $\{x : t(x) \geq c\}$ would then correspond to those branches that would fall off if we were to chop the tree with a horizontal slice at height c . For a statistical interpretation of this process, consider c to be the critical value of the statistic t , and the weight of the chopped branches to be the probability of rejection. The patriarchs of the chopped branches represent the *critical configurations*, those outcomes that are just extreme enough to lead to rejection.

Because of discreteness, there generally need not be any configurations with $t(x) = c$. However, in the case $w_i = i$, all the values of t will be integral, and we might hope that the critical configurations would have $t(x) = \kappa = \lceil c \rceil$ (where $\lceil c \rceil$ represents the smallest integer greater or equal to c). If the capacities all are (at least) n , then this is in fact the case, for each configuration will have at most two offspring, each of which will have a score one greater than its parent. In this case the "height" of a node will be equal to its graph-theoretic height—i.e., the length of the path to the root. Thus the critical configurations would be precisely those of height κ .

Unfortunately, this pleasant property is no longer enjoyed by configurations in which the capacities are genuinely limiting. In such a case there may exist "hidden criticals": configurations for which $t(x) > \kappa$ but $t(\hat{x}) < \kappa$. Thus it is not sufficient to collect all the branches springing from configurations with $t(x) = \kappa$.

5.3.2 Another tree structure

If all the capacities are *positive* then we can address this problem of hidden criticals by defining another relation which we will call *stepparenthood*. Let $x = (x_0, \dots, x_k)$ be a configuration. Then x^1 , the stepparent of x is defined as the configuration with

$$\begin{aligned} x^1_j &= x^1_{j+1} + 1, \\ x^1_{j+1} &= x^1_{j+1} - 1, \text{ and} \\ x^1_i &= x_i \quad \text{if } i \neq j, j+1 \end{aligned} \tag{5.14}$$

where $j = \max\{i : x_i < v_i \text{ and } x_{i+1} > 0\}$. Clearly, the stepparental mapping decreases the value of the function $\sum i x_i$, and is undefined only if the set $\{i : x_i < v_i \text{ and } x_{i+1} > 0\}$ is empty. However, if $x_i = v_i > 0$, then $i - 1$ belongs to this set unless $x_{i-1} = v_{i-1}$. If all $v_i > 0$, then by induction, there must be a $j < i$ satisfying the defining property unless $x_r = v_r$ for all $r \leq i$. Thus $\{i : x_i < v_i \text{ and } x_{i+1} > 0\} = \emptyset$ if and only if $x_i = v_i$ for $i < m$, and $x_m = n - \sum_{i=0}^{m-1} v_i$, where $m = \min\{i : \sum_{r=0}^i v_r \geq n\}$. Since these conditions define a unique configuration, by Theorem 5.1 the stepparental mapping also defines a tree structure. This structure will, in general, be different from the one defined in Section 5.1.3

Figure 5.1: An example of a tree of configurations. Critical configurations are indicated as solid circles. Capacities are $v = (2, 1, 1, 4, 1)$, $n = 6$.

although the roots coincide. The two structures will be identical if $\min v_i \geq n$. We will distinguish the two structures, and their attributes, by the adjectives "natural" and "step". We will, for example, refer to branches of the original tree as "natural" branches, whereas those defined by the stepparental mapping will be called "stepbranches". We will generally use the modifier "pure" to distinguish attributes that do differ in the two structures. In particular, if $\hat{x} \neq x^1$ then we will call x a *pure stepchild* of x^1 .

Obviously $t(x^1) = t(x) - 1$, so that $t(x) \geq \kappa$ and $t(x^1) < \kappa$ jointly imply that $t(x) = \kappa$. Therefore, to weigh the configurations with $t(x) \geq \kappa$ it is sufficient to weigh the stepbranches springing from configurations with $t(x) = \kappa$.

However, the definitions of offspring and branches are not quite as elegant for the steptree as for the natural tree. Therefore, we will use the natural definition whenever possible. We can do this without danger of double counting because of the following theorem.

Theorem 5.7 *For any configuration x with strictly positive capacities, either $\hat{x} = x^1$ or $\hat{x} = x^1$.*

Proof: Let $m = \max\{i : x_i > 0\}$. If $x_{m-1} < v_{m-1}$ then $x^1_{m-1} = x_{m-1} + 1 = \hat{x}_{m-1}$ and $x^1_m = x_m - 1 = \hat{x}_m$, from which $\hat{x} = x^1$. Otherwise, let $j = \max\{i < m : x_i < v_i\}$. Then $x^1_j = x_j + 1$, $x^1_{j+1} = x_{j+1} - 1$. On the other hand, $\hat{x}_j = x_j + 1$ and $\hat{x}_m = x_m - 1$. And since $x^1_{j+1} = x_{j+1} - 1 < v_{j+1}$, we have that $x^1_{j+1} = x^1_{j+1} + 1 = (x_{j+1} - 1) + 1 = x_{j+1} = \hat{x}_{j+1}$. But $x^1_m = x^1_m - 1 = x_m - 1 = \hat{x}_m$, and $x^1_j = x^1_j = x_j = \hat{x}_j$. Thus $\hat{x} = x^1$. QED

This theorem has several corollaries.

Corollary 5.7.1 *The natural branch springing from a node is contained in the stepbranch springing from the same node.*

Proof: Let y be an element of the natural branch B_x , $y \neq x$. Then either $y^1 = \hat{y} \in B_x$ or $y^1 = \hat{y} \in B_x$ which implies that $y^1 \in B_x$. Thus B_x is a covering set for x in the steptree, and the result follows from Lemma 5.3. QED

Corollary 5.7.2 *A natural branch springing from a configuration is disjoint from the stepbranch springing from a pure stepchild.*

Proof: The stepbranch B^1_x springing from a node x is the union of the node itself with the stepbranches springing from its stepchildren. These stepbranches are of course disjoint. The natural branch is the union of the node itself with the branches springing from the natural children. Natural branches springing from natural children that are also stepchildren are by Corollary 5.7.1 contained in stepbranches, which must of course be disjoint from those springing from pure stepchildren. Consider now a natural branch B_y springing from a natural child y that is not also a stepchild. Since by Corollary 5.7.1 $B_y \subset B^1_x$, y must have a stepparent $z = y^1$ that is contained in a stepbranch. By Theorem 5.7, we have that either $z = \hat{y}$ or $\hat{z} = \hat{y}$. But $\hat{y} = x$, and by hypothesis y is not a stepchild of x . Thus we must have $\hat{z} = x$, and thus y is part of the stepbranch springing

Figure 5.2: An example of the stepparental mapping. This is the same tree as Figure 5.1 with the pure stepchild relation shown as dotted lines

from another natural child. If $z^1 = x$ then the branch springing from it must be disjoint from that of any pure stepchildren. On the other hand, if $z^1 \neq x$ then we can repeat the argument. We thus get a chain of stepparents, which is contained in the stepbranch B_x^1 . Since the tree is finite, it must terminate in x . The second last node then is also a natural child, and thus the chain, and hence B_y is disjoint from any branch springing from a pure stepchild. QED

Corollary 5.7.3 *Every node is contained in the stepbranch springing from its natural parent.*

Proof: $x \in B_x \subset B_x^1$. QED

These corollaries assure us that we can augment a set of natural branches by the branches springing from pure stepchildren without thereby introducing double-counting. Our current goal is to find a simple procedure for defining stepbranches. Before we do that however, we require a few more results. First, we want a way of inverting the stepparental mapping.

Theorem 5.8 *Given a configuration x , let $m = \max\{i : x_i > 0\}$, $j = \max\{i < m : x_i < v_i\}$. Then if $x_{j-1} > 0$ and $x_j = v_j - 1$ then the configuration y with*

$$\begin{aligned} y_{j-1} &= x_{j-1} - 1, \\ y_j &= x_j + 1, \\ \text{and } y_i &= x_i \text{ if } i \neq j, j+1 \end{aligned} \quad (5.15)$$

is the unique pure stepchild of x . Otherwise x has no pure stepchildren.

Proof: It is clear that y so defined is a stepchild. Now, consider another configuration z such that $z^1 = x$, and let l be the cell such that $x_l = z_l - 1$. If $l = \max\{i : x_i > 0\}$ then x is the natural parent of z . Suppose that $l < \max\{i : x_i > 0\}$. Then by definition of stepparents we require that $z_i = v_i$ for $l \leq i < m$. Thus $x_l - z_l - 1 = v_l - 1$, and $l = \max\{i < m : x_i < v_i\}$. Thus $z = y$. QED

The preceding theorem begins to show where the natural tree structure differs from the steptree. Any node can have at most one pure stepchild, this stepchild itself can have at most one pure stepchild and so on. We will define the *stepchain* from x as a sequence of nodes $x^{(1)}, \dots, x^{(r)}$ such that $x = x^{(1)}$, $x^{(i+1)}$ is a pure stepchild of $x^{(i)}$, ($i = 1, \dots, r-1$); and $x^{(r)}$ has no pure stepchildren. We will ultimately show that a stepbranch can be considered to be the union of natural branches connected by a stepchain. However, to prove this, we require a few more results.

Theorem 5.9 *Let y be on the steppath from $\hat{z}$ to x , $y \neq \hat{z}$. Then $\hat{y} = \hat{z}$.*

Proof: Clearly the theorem is true for $y = x$. Now suppose that $\hat{z} = \hat{z}$. Then either $z^1 = \hat{z} = \hat{x}$ or $z^1 = \hat{z} = \hat{x}$. Thus if $z^1 \neq \hat{x}$ the theorem is true for z^1 as well. The result follows by induction. QED

Theorem 5.10 *If a configuration has no pure stepchildren, then the stepbranch springing from it is equal to the natural branch.*

Proof: Since $B_x \subset B^1_x$, the result follows if B^1_x is a covering set for x in the natural tree. This is obviously true if $B^1_x = \{x\}$, so suppose that $y \in B^1_x$, $y \neq x$.

Now $y \in B^1_y$ and $y \in B^1_x$, so $y \in B^1_y \cap B^1_x$. Now suppose that $\hat{y} \notin B^1_x$. Then we must have by Lemma 5.5 that $x \in B^1_y$ and $x \neq \hat{y}$. That is, x is on the steppath from $\hat{y}$ to y , and there must be a stepchild x_0 on this path. But by Theorem 5.10 $x_0 = \hat{y}$ but $x_0 = x$. Thus x_0 is a pure stepchild which contradicts the hypothesis. Therefore, we must have $\hat{y} \in B^1_x$, and thus B^1_x is a covering set for x in the natural tree. QED

Theorem 5.11 *The natural branch springing from x is equal to the union of $\{x\}$ with the stepbranches of the natural offspring of x .*

Proof: Let

$$S = \{x\} \cup \bigcup_{y: \hat{y}=x} B^1_y.$$

For y a natural offspring of x , let z be an element of B^1_y . If $z = y$, then $\hat{z} = \hat{y} = x$ because y is a natural offspring of x . Thus $\hat{z} \in S$. If $z \neq y$, then by Corollary 5.7.3 we have that $z \in B^1_z$ and hence $z \in B^1_z \cap B^1_y$. Thus by Lemma 5.5 either $B^1_z \subset B^1_y$ or $B^1_y \subset B^1_z$. If $B^1_z \subset B^1_y$, then $\hat{z} \in B^1_y$ and $\hat{z} \in S$. On the other hand, if $B^1_y \subset B^1_z$, then y is on the steppath from $\hat{z}$ to z and thus $\hat{z} = \hat{y} = x$. Thus once more $\hat{z} \in S$. Thus S is a covering set for x , and is contained in B_x . On the other hand, we know that $B_x \subset B^1_x$. But $B^1_x = B^1_w \cup S$, where w is a pure stepchild, and by Corollary 5.7.2 $B_x \cap B^1_w = \emptyset$, so $B_x \subset S$. QED

We now present the main theorem of this section.

Theorem 5.12 *The stepbranch springing from x is the union of a set of natural branches. The patriarchs of these branches form a chain of stepparents terminating in x .*

Proof: Let x be a node, and let B^1_x be the stepbranch springing from x . Let $x^{(1)}, \dots, x^{(r)}$ be the stepchildren of x . If there is no pure stepchild, then by Theorem 5.10 $B^1_x = B_x$ and we are finished. Otherwise, let $x^{(r)}$ be the pure stepchild. Then

$$B^1_x = \left(\bigcup_{i=1}^{r-1} B^1_{x^{(i)}} \right) \cup \{x\} \cup B^1_{x^{(r)}} = B_x \cup B^1_{x^{(r)}}$$

by Theorem 5.11. We apply the same argument to $B^1_{x^{(r)}}$, and thus the result follows by induction. QED

The last theorem provides a basis for enumerating the stepbranch by breaking it up into natural branches. The natural branches can then be enumerated using the methods of Section 5.1.3.

5.3.3 Case with some capacities null

This strategy of steptrees fails however, if we allow some of the capacities to be zero. The difficulty is that in this case the stepparent relationship no longer defines a tree. Indeed consider a configuration with $x_m = v_m > 0$, $x_{m-1} = v_{m-1} = 0$, $x_{m-2} = 0 < v_{m-2}$, and $x_i = 0$ for $i < m-2$ or $i > m$. In this case the set $\{i : x_i < v_i \text{ and } x_{i+1} > 0\}$ is empty even though this configuration is not the root of the natural tree. It is tempting to modify the definition of stepparent by defining $m = \max\{i : x_i > 0\}$, $j = \max\{i < m : x_i < v_i\}$, $l = \min\{i > j : x_i > 0\}$ and then making $x_l^1 = x_j + 1$, $x_l^1 = x_l - 1$, and $x_i^1 = x_i$ for $l \neq i \neq j$, but this definition lacks the attractive property of parents differing from their children by a score of 1. This property, after all, was the main motivation for introducing the steptree concept.

We get out of our difficulty by introducing the concept of "phantom configurations"—configurations having a probability of zero, but which are required to fill the gaps in the tree structure. One can picture the stepparental mapping as tossing a ball from a cell to the next cell to its left. In cases in which the stepparent differs from the natural parent, the donor cell will be full. We can maintain this idea even in those cases in which this full cell has a capacity of zero. After all, a null cell will be full. However, removing a ball from such a cell would create a deficit, i.e., a cell with negative content.

We will henceforth extend the definition of configuration to allow null cells to have a content of -1 . Since these *phantom* configurations have probability zero, it will not be necessary to enumerate them all. It will suffice to include enough of them to keep the tree together.

We thus redefine the stepparental mapping as follows: Let $m = \max\{i : x_i > 0\}$ and $j = \max\{i < m : x_i < v_i\}$. Then define

$$\begin{aligned} x_l^1 &= x_j^1 + 1, \\ x_{j+1}^1 &= x_{j+1}^1 - 1, \\ \text{and } x_i^1 &= x_i \text{ for } i \neq j, j+1. \end{aligned} \quad (5.16)$$

The definition of j implies that $x_{j+1} = 0$ can occur if and only if $v_{j+1} = 0$, so that only null cells can have deficits, and the deficit cannot exceed 1.

The definition also mandates that deficits can only be created in cells to the left of a non-empty cell. The stepparent of a configuration without deficits cannot have a deficit in any cells to the right of the last non-empty cell, and iteration of the stepparental mapping will not change the empty status of such cells. Moreover, the cell to the left of the deficient one will by definition be non-empty. Thus we only need to allow deficits in cells which have non-empty cells to both their left and right.

The set of phantom configurations can be further restricted. Note that the stepparent of a configuration without deficits cannot have non-full cells between a deficient cell and the last non-empty cell. Since a deficient cell is not full, such configurations can have only a single deficit. The stepparental mapping will eliminate the deficit, although if the next cell to the right is also null, it might create another one.

These considerations lead us to the following augmented definition. A configuration of n balls in $k+1$ cells with capacities $(v_0, \dots, v_k)$ is a list of integers $(x_0, \dots, x_k)$ with

$\sum_{i=1}^k x_i = n$ and the following constraints:

$$\begin{aligned}
 0 \leq x_i \leq v_i & \text{ if } v_i > 0; \\
 -1 \leq x_i \leq 0 & \text{ if } v_i = 0, \\
 & q < i < m, \text{ and} \\
 & x_r = v_r \text{ for } i < r < m \\
 & \text{where } q = \min\{i : x_i > 0\} \text{ and} \\
 & m = \max\{i : x_i > 0\}; \\
 & \text{and} \\
 x_i = 0 & \text{ otherwise.}
 \end{aligned} \tag{5.17}$$

Configurations with deficits will be called *phantom*, whereas those without deficits will be called *legitimate*.

Using the same arguments as previously, we can see that for this augmented set of configurations, the stepparental mapping x^1 defined in (5.16) defines a tree, whose root coincides with that of the natural tree. For this tree, the set of nodes with $t(x) \geq \kappa$ is equivalent to the union of the branches springing from nodes such that $t(x) = \kappa$, although some of the patriarchs of these branches may be phantoms.

The definition of the natural parental mapping can also of course be extended to phantom configurations. However, the phantom nodes have no structural importance in the natural tree because they are merely additional "leaves", as shown by the following theorem.

Theorem 5.13 *A phantom configuration cannot be the natural parent of another configuration.*

Proof: If x is a legitimate configuration, then $x_i \geq 0$ for all i . For the natural parent $\hat{x}$, $\hat{x}_i \geq x_i \geq 0$ except for $\hat{x}_m$ where $m = \max\{i : x_i > 0\}$. Thus $x_m > 0$ by definition, $x_m - 1 \geq 0$ and $\hat{x}$ is legitimate. On the other hand, if x is phantom then there exists a $j < m$ such that $x_j = -1$, and $x_i \geq 0$ for all other i . However, this implies that $x_i = v_i$ for all $j < i < m$. Thus, the natural parent has $\hat{x}_j = x_j + 1 = 0$. Thus the parent is also legitimate. QED

It should be noted that the characterizations of Theorem 5.4 for natural children apply only to legitimate children. However, since phantom children are weightless, failure to enumerate them will cause no problems.

There is, however, one situation in which such phantom offspring appear naturally (see Section 5.4.1). If x is a configuration with $m = \max\{i : x_i > 0\}$, and $v_j = 0$, with $j > m$, but $\{i > j : v_i > 0\} \neq \emptyset$, then the phantom configuration $\hat{x}$ with $\hat{x}_j = -1$, $\hat{x}_h = 1$, and $\hat{x}_i = x_i$ for $j \neq i \neq h$, where $h = \min\{i > j : v_i > 0\}$, will have x as a natural parent. We will call these phantom sons *bastards*. The bastard with $j = h - 1$ will also be a stepson.

The new definition of stepchildren differs from the previous one only in the sense that null cells are not considered empty if they have a non-empty cell to their right. Thus all the theorems in Section 5.3.2 about the steptree remain true, with the exception of Theorem 5.8 which requires a slight modification.

Theorem 5.14 Given a configuration x , let $m = \max\{i : x_i > 0\}$, $j = \max\{i < m : x_i < v_i\}$, and $h = \min\{i : x_i > 0\}$. Then if $h < j$, $x_j = v_j - 1$, and either $x_{j-1} > 0$ or $v_{j-1} = 0$ then the configuration y with

$$\begin{aligned} y_{j-1} &= x_{j-1} - 1, \\ y_j &= x_j + 1, \\ \text{and } y_i &= x_i \text{ if } i \neq j, j+1 \end{aligned} \quad (5.18)$$

is the unique pure stepchild of x . Otherwise x has no pure stepchildren.

Proof: The proof follows the same arguments as Theorem 5.8, and will not be repeated.

QED

5.3.4 Chains of branches

The enumeration of the stepbranch of a configuration again now is straightforward: One steps through the stepchain, enumerating in turn the natural branch springing from each legitimate stepchild in the chain. Since, however, there is no need to enumerate phantom stepchildren, we can terminate the chain once we are assured that there can be no more legitimate configurations in it.

We will thus use the term *virtual stepchain* for the sequence of pure stepchildren in the augmented tree. We will reserve the term *stepchain* for a sequence of nodes $x^{(1)}, \dots, x^{(f)}$ such that $x^{(1)}, \dots, x^{(r)}$ is a virtual stepchain ($r \geq f$), and $f = \max\{i : x^{(i)} \text{ is legitimate}\}$. This definition of stepchain will agree with the preceding one if there are no phantoms. It is possible that all the pure stepchildren of a configuration are phantoms, in which case the stepchain will be empty.

The next theorem gives a procedure for determining the stepchains at a node. It is stated in terms of the cumulative sums $(s_0, \dots, s_k)$.

Theorem 5.15 Let $L = \{i > 0 : s_i > s_{i-1} - v_{i-1} \vee 0\}$ be non-empty and define $l = \max L$. Let $H = \{i < l : s_i - s_l < \sum_{r=i}^{l-1} v_r - 1\}$ and define $h = 0$ if $H = \emptyset$ and $h = \max H$ otherwise. Then the following statements are true.

- (a) If $h < l - 1$ then $s_l = s_{l-1} - v_{l-1} + 1$.
- (b) $s_i = s_{i-1} - v_{i-1}$ for $h + 1 < i < l$.
- (c) If $h = l - 1$ then s is a legitimate configuration but has no pure stepchildren.
- (d) Let $s^{(1)} = s$ and for $r = 2, \dots, l - h$, $s_i^{(r)} = s_i + 1$ if $l - r + 1 \leq i < l$, but $s_i^{(r)} = s_i$ otherwise. Then if $s_h > s_{h+1}$, the stepchain at s is the sequence $(s^{(1)}, \dots, s^{(l-h)})$.
- (e) If $h < l - 1$, $s_h \leq s_{h+1}$ and $F = \{i : h < i < l, v_i > 0\} = \emptyset$ then the stepchain at s is empty, whereas if $F \neq \emptyset$, then the stepchain is $(s^{(1)}, \dots, s^{(l-f)})$ where $f = \min F$.

2

Proof: In order that $h < l - 1$ we require that $s_{l-1} - s_l \geq v_{l-1} - 1$ or equivalently, $s_l \leq s_{l-1} - v_{l-1} + 1$. However, since by definition of l , $s_l > s_{l-1} - v_{l-1}$, we must have $s_l = s_{l-1} - v_{l-1} + 1$, thus proving (a).

Now let i satisfy $h + 1 < i < l$. Such an i will exist only if $h < l - 2$. Thus *a fortiori* $h < l - 1$ and $s_l = s_{l-1} - v_{l-1} + 1$. But since $h < l - 2$ we also require that

$$s_{l-2} - s_l \geq v_{l-2} + v_{l-1} - 1.$$

Combining this with the previous result gives that $s_{l-1} \leq s_{l-2} - v_{l-2}$. However, since $s_{l-1} \geq s_{l-2} - v_{l-2}$ we must have $s_{l-1} = s_{l-2} - v_{l-2}$.

Now consider the inductive hypothesis:

$$\begin{aligned} s_l &= s_{l-1} - v_{l-1} + 1 \quad \text{and there exists } g > h + 1 \text{ such that} \\ s_i &= s_{i-1} - v_{i-1} \quad \text{for all } g < i < l. \end{aligned} \tag{5.19}$$

Then since $g - 1 > h$ we have that

$$s_{g-1} - s_l \geq \sum_{i=g-1}^{l-1} v_i - 1$$

which can be written

$$\begin{aligned} s_{g-1} - s_g + \sum_{i=g+1}^{l-1} (s_{i-1} - s_i) + s_{l-1} - s_l \\ \geq v_{g-1} + \sum_{i=g+1}^{l-1} v_{i-1} + v_{l-1} - 1. \end{aligned} \tag{5.20}$$

But by hypothesis, $v_{i-1} = s_{i-1} - s_i$ for $g < i < l$ and $s_{l-1} - s_l = v_{l-1} - 1$. Thus (5.20) implies that $s_{g-1} - s_g \geq v_{g-1}$. But since $s_g \geq s_{g-1} - v_{g-1}$ we have $s_g = s_{g-1} - v_{g-1}$. Hence by induction on g , (b) is established.

To prove (c), suppose that $h = l - 1$, so that the condition of Theorem 5.14 fails, and s has no pure stepchild. However, since $x_{l-1} \geq -1$, we must have $v_{l-1} > 0$ so that the configuration cannot be phantom.

To prove (d) and (e) consider the $l - h + 1$ cells from h^{th} to the l^{th} . Since the case $h = l - 1$ is addressed by (c) we will assume that $h < l - 1$, thus $h < l - 1 < l$. By (a), $x_{l-1} = s_{l-1} - s_l = v_{l-1} - 1$. If moreover there are any additional cells $h < i < l - 1$, then part (b) assures us that $x_i = s_i - s_{i+1} = v_i$. Thus all the cells in the range $h + 1, \dots, l - 1$ are full, except for the last one, which is "almost full".

The pure stepchild of this configuration will have $x_{l-1} = v_{l-1}$, $x_{l-2} = v_{l-2} - 1$, and all other cells the same as the stepparent. The pure *grand*-stepchild will have $x_{l-2} = v_{l-2}$, $x_{l-3} = v_{l-3} - 1$, and in general, the pure *grand* ^{$g-1$} -stepchild will have $x_{l-g-1} = v_{l-g-1} - 1$ and $x_i = v_i$ for $i \neq l - g - 1$, $h < i < l$.

(The pure stepchild is the configuration obtained by tossing a ball from the $l - 2^{\text{nd}}$ cell to the $l - 1^{\text{st}}$, or alternatively, we can think of a "hole" moving one cell to the left. Thus the stepchain is defined by the sequence obtained by progressively moving the hole to the left.)

This process will give a legitimate configuration if $v_{l-g-1} > 0$. If $v_{l-g-1} = 0$, then we have an admissible deficit as long as $x_i > 0$ for some $i < l - g - 1$. Thus if $x_h > 0$ i.e. $s_h > s_{h-1}$, the sequence obtained by producing pure stepchildren will always yield admissible configurations, although some of these may be phantoms. Moreover, the last configuration in this chain will be the configuration

$$(\dots, x_h, v_{h+1} - 1, v_{h+2}, \dots, v_{l-1}, x_l, \dots)$$

whose pure stepchild is

$$(\dots, x_h - 1, v_{h+1}, v_{h+2}, \dots, v_{l-1}, x_l, \dots),$$

which is a legitimate configuration. Since $x_h = s_h - s_{h-1} < v_h$, however, this configuration will have no pure stepchild itself, and hence terminates the stepchain. (d) is thus proved.

On the other hand, if $x_h = 0$, then this last configuration is not legitimate, and since $v_h > x_h = 0$, a deficit is not allowed at this position either. If $h < i < f$, then $v_i = 0$ and hence any configuration with $x_i = v_i - 1$ will be phantom. However, since $v_f > 0$, the configuration with $x_f = v_f - 1$ will be legitimate. This establishes (e). QED

5.3.5 Calculating probabilities

If our goal is merely to calculate the probability of the stepbranch, then the preceding theorem provides a ready technique. Let a configuration be represented in terms of the partial sums $s = (s_0, \dots, s_k)$, and define l to be the last non-minimal cell:

$$l = \max\{i : s_i > 0 \vee s_{i-1} - v_{i-1}\}.$$

Then from (5.12), we have that the probability of the natural branch can be calculated as

$$P(B_s) = \prod_{i=1}^{l-1} p_i(s_i | s_{i-1}) \sum_{u=s_l}^{s_{l-1}} p_l(u | s_{l-1}).$$

If s is a phantom configuration then the deficits admitted by (5.17) are such that if $x_g < 0$, then either $x_r = 0$ for all $r > g + 1$, or $x_r = v_r$ for all r such that $g + 1 \leq r < m = \max\{i : x_i > 0\}$. This implies that s_i is minimal for $i > g + 1$, so the last non-minimal cell $l = g + 1$. But if $x_{l-1} < 0$, then $s_{l-1} = s_l + x_{l-1} < s_l$, so that the summation is void. Thus the branch has probability zero, confirming Theorem 5.13.

Suppose now that s has a pure stepchild. Theorem 5.15 implies that $s_l = s_{l-1} - v_{l-1} + 1$, so that F is almost minimal.

Moreover, this stepchild $\bar{x}$ has $\bar{x}_{l-1} = x_{l-1} + 1$, $\bar{x}_{l-2} = x_{l-2} - 1$, and hence $\bar{s}_{l-1} = s_{l-1} + 1$, and $\bar{s}_l = s_l = s_{l-1} - v_{l-1} + 1 = \bar{s}_{l-1} - v_{l-1}$, so the l^{th} cell is minimal in the stepchild.

However, $l - 1$ is not minimal in the stepchild since

$$\bar{s}_{l-1} = s_{l-1} + 1 > s_{l-1} \geq s_{l-2} - v_{l-2} \vee 0.$$

Thus, the natural branch springing from the stepchild has probability

$$P(B_{[s]}) = \prod_{i=1}^{l-2} p_i(s_i | s_{i-1}) \sum_{u=s_{l-1}+1}^{s_{l-2}} p_{l-1}(u | s_{l-2}). \quad (5.21)$$

This configuration will itself have a further stepchild if and only if $s_{l-1} = s_{l-2} - v_{l-2}$ so that $s_{l-1} = s_{l-2} - v_{l-2} + 1$. In that case, the probability of the natural branch springing from the step-step-child is

$$P(B_{[[s]}) = \prod_{i=1}^{l-3} p_i(s_i | s_{i-1}) \sum_{u=s_{l-2}+1}^{s_{l-3}} p_{l-2}(u | s_{l-3}). \quad (5.22)$$

Using the results of Theorem 5.15, we can express the probability of a stepbranch as

$$P(B_s^{[l]}) = \sum_{i=1}^{l-f} P(B_{s^{(i)}}), \quad (5.23)$$

where

$$\begin{aligned} l &= \max\{i : s_i > 0 \vee s_{i-1} - v_{i-1}\}, \\ h &= \max\left(\{0\} \cup \left\{i < l : s_i - s_l < \sum_{r=1}^{l-1} v_{r-1}\right\}\right), \text{ and} \\ f &= \begin{cases} h & \text{if } s_h > s_{h+1} \text{ or } h = l-1 \\ \min\{i > h : v_i > 0\} & \text{otherwise.} \end{cases} \end{aligned}$$

The last non-minimal cell of $s^{(i)}$ is $l-i+1$ and so using (5.12)

$$P(B_{s^{(i)}}) = \prod_{r=1}^{l-i} p_r(s_r | s_{r-1}) \sum_{u=s_{l-i+1}^{(i)}}^{s_{l-i}} p_{l-i+1}(u | s_{l-i}). \quad (5.24)$$

However, if $h < l-1$, then $s_l^{(1)} = s_l = s_{l-1} - v_{l-1} + 1$ so that

$$\sum_{u=s_l}^{s_{l-1}} p_l(u | s_{l-1}) = 1 - p_l(s_{l-1} - v_{l-1} | s_{l-1}).$$

Also, for $1 < i < l-h$,

$$s_{l-i+1}^{(i)} = s_{l-i+1} + 1 = s_{l-i} - v_{l-i} + 1,$$

which implies

$$\begin{aligned} \sum_{u=s_{l-i+1}^{(i)}}^{s_{l-i}} p_{l-i+1}(u | s_{l-i}) &= 1 - p_{l-i+1}(s_{l-i} - v_{l-i} | s_{l-i}) \\ &= 1 - p_{l-i+1}(s_{l-i+1} | s_{l-i}), \end{aligned}$$

by part (b) of Theorem 5.15. Substituting into (5.24) we can thus write for $1 \leq i < l - h$

$$P(B_s(i)) = \left(\prod_{r=1}^{l-i} p_r(s_r | s_{r-1}) \right) (1 - p_{l-i+1}(s_{l-i+1} - \delta_{i1} | s_{l-i})). \quad (5.25)$$

To simplify the notation, write, for now, just

$$p_i \quad \text{for} \quad p_i(s_i | s_{i-1}) \text{ if } 1 \leq i < l \\ \text{and } p_l \quad \text{for} \quad p_l(s_l - 1 | s_{l-1}).$$

Then for $f < h$ we have

$$\begin{aligned} \Pr\{B_s\} &= \sum_{i=1}^{l-f} \left(\prod_{r=1}^{l-i} p_r \right) (1 - p_{l-i+1}) \\ &= \left(\prod_{r=1}^f p_r \right) \sum_{i=1}^{l-f} \left[\prod_{r=f+1}^{l-i} p_r - \prod_{r=f+1}^{l-i+1} p_r \right] \\ &= \left(\prod_{r=1}^f p_r \right) \left[\sum_{i=1}^{l-f} \prod_{r=f+1}^{l-i} p_r - \sum_{i=1}^{l-f} \prod_{r=f+1}^{l-i+1} p_r \right] \\ &= \left(\prod_{r=1}^f p_r \right) \left[\sum_{i=1}^{l-f} \prod_{r=f+1}^{l-i} p_r - \sum_{i=0}^{l-f-1} \prod_{r=f+1}^{l-i} p_r \right] \\ &= \left(\prod_{r=1}^f p_r \right) \left[1 - \prod_{r=f+1}^l p_r \right]. \end{aligned} \quad (5.26)$$

On the other hand, if $f = h$, we have

$$\begin{aligned} \Pr\{B_s\} &= \sum_{i=1}^{l-h-1} \left(\prod_{r=1}^{l-i} p_r \right) (1 - p_{l-i+1}) + \left(\prod_{r=1}^h p_r \right) \sum_{u=s_{h+1}+1}^{s_h} p_{h+1}(u | s_h) \\ &= \left(\prod_{r=1}^{h+1} p_r \right) \left[1 - \prod_{r=h+2}^l p_r \right] + \left(\prod_{r=1}^h p_r \right) \sum_{u=s_{h+1}+1}^{s_h} p_{h+1}(u | s_h) \\ &= \left(\prod_{r=1}^h p_r \right) \left[p_{h+1} - \prod_{r=h+1}^l p_r + \sum_{u=s_{h+1}+1}^{s_h} p_{h+1}(u | s_h) \right] \\ &= \left(\prod_{r=1}^h p_r \right) \left[\sum_{u=s_{h+1}+1}^{s_h} p_{h+1}(u | s_h) - \prod_{r=h+1}^l p_r \right]. \end{aligned} \quad (5.27)$$

However,

$$\left(\prod_{r=1}^h p_r \right) \sum_{u=s_{h+1}+1}^{s_h} p_{h+1}(u | s_h)$$

is in fact the probability of a natural branch. The patriarch of this branch is the configuration for which $h + 1$ is the last non-minimal cell. Now, l is the last non-minimal cell

for s , so to make $h+1$ the last non-minimal, we must decrement s_l (since l is almost minimal, we only need to decrement it by 1). But if $s_{l+1} > 0$ this action will make it non-minimal, so we must decrement s_{l+1} as well, and so on, up to s_m , the last non-empty cell. However, this action describes the natural parent of s . Thus

$$\prod_{r=1}^h p_r \sum_{u=s_{h+1}}^{s_h} p_{h+1}(u|s_h)$$

is nothing but $P(B_h)$. Thus in the case $f = h$

$$\Pr \{B_s\} = \Pr \{B_h\} - \prod_{i=1}^l p_i(s_i - \delta_{il}|s_{i-1}). \quad (5.28)$$

Calculations are further simplified by recognizing the form of p_i for $f < i \leq l$ in the case $b_i = v_i$. Since

$$\begin{aligned} p_i &= p_i(s_{i-1} - v_{i-1}|s_{i-1}) \\ &= \frac{\binom{a_i}{s_{i-1} - v_{i-1}} \binom{b_{i-1}}{v_{i-1}}}{\binom{a_{i-1}}{s_{i-1}}} \\ &= \frac{\binom{a_i}{s_i - \delta_{il}}}{\binom{a_{i-1}}{s_{i-1}}}, \end{aligned} \quad (5.29)$$

thus

$$\begin{aligned} \prod_{i=f+1}^l p_i &= \frac{\prod_{i=f+1}^l \binom{a_i}{s_i - \delta_{il}}}{\prod_{i=f}^{l-1} \binom{a_i}{s_i}} \\ &= \frac{\binom{a_l}{s_l - 1}}{\binom{a_f}{s_f}}. \end{aligned} \quad (5.30)$$

And of course, if $h < i < f$ then $v_i = 0$ so that $p_i = 1$. Thus (5.26) reduces to

$$\left(\prod_{r=1}^h p_r \right) p_f \left(1 - \frac{\binom{a_l}{s_l - 1}}{\binom{a_f}{s_f}} \right).$$

5.4 Finding Critical Configurations

The preceding discussion provides an efficient manner for enumerating and weighing the stepbranches of configurations. Thus the problem of enumerating or weighing $\{x : t(x) \geq \kappa\}$ is reduced to the problem of enumerating $\{x : t(x) = \kappa\}$. This can be accomplished by defining yet another tree structure which is based on a chain of configurations encompassing all possible values of $t(x)$. This chain may include phantoms if some of the cells are null. Each node of this chain then becomes the root of a *lateral subtree*, whose nodes all have the same value of $t(x)$. Traversing this lateral tree will then provide the means for identifying the branches which need to be weighed.

5.4.1 Semi-compact configurations

A configuration x will be called *compact* if $x_i = v_i$ for $q < i < m$, where $q = \min\{i : x_i > 0\}$ and $m = \max\{i : x_i > 0\}$. It will be called *para-compact* if there is a cell h , ($q < h < m$), with $x_h < v_h$ such that the configuration y with $y_h = x_h + 1$ and $y_i = x_i$ for $i \neq h$ is compact. The h^{th} cell will be called the *hole* in the para-compact configuration. (Note that y , having $n+1$ balls, will not be in the same set of configurations as x . We introduce it here only to simplify the definition.) Configurations that are either compact or para-compact will be called *semi-compact*.

If $\min_i v_i > n$, then the definition of compactness can obtain only if $\{i : q < i < m\} = \emptyset$. Thus in this case all semi-compact configurations are compact, and have at most two non-empty cells. Obviously, compact configurations cannot have deficits, and a deficit in a para-compact one must be at the hole. It is also clear that the root of the tree of configurations is compact. The set of semi-compact configurations forms a chain, as shown by the following two theorems.

Theorem 5.16 *The stepparent of a semi-compact configuration that is not the root is also semi-compact.*

Proof: If x is para-compact with a hole at cell h , then $x^1_h = x_h + 1 = v_h$, and $x^1_{h+1} = x_{h+1} - 1$. If $h+1 < m$, then x^1 will itself be para-compact with a hole at $h+1$, whereas if $h+1 = m$ then x^1 will be compact.

Suppose now that x is compact. By definition, $j = \max\{i < m : x_i < v_i\} \leq q$, provided that x is not the root, in which case j is undefined. Now $q^1 = j$, and $x^1_{j+1} = x_{j+1} - 1$. But $x_i = v_i$ for $j < i < m$ so that x^1 will be compact if $j+1 = m$, and will be para-compact with hole at $j+1$ if $j+1 < m$. QED

Remark: The stepparent of a compact configuration will be a phantom if $x_q = v_q$ and $v_{q-1} = 0$. The stepparent of a para-compact configuration with hole at h will be a phantom if $v_{h+1} = 0$.

Theorem 5.17 *Every semi-compact configuration but one has a unique semi-compact stepchild. The exceptional configuration has no stepchildren.*

Proof: If the configuration is para-compact with a hole at h , then it has a pure stepchild by Theorem 5.14, and it is easy to see that this stepchild is itself semi-compact.

However, a natural child $\bar{x}$ will have $\bar{x}_i = x_i - 1 \leq v_i$ for some $h \leq i \leq m$. If $i = h$ then $\bar{x}_i = v_i - 2$, whereas if $i > h$, then both $\bar{x}_h < v_h$ and $\bar{x}_i < v_i$ with $h < i < m$. In either case, $\bar{x}$ will not be semi-compact. Thus the pure stepchild is the only semi-compact stepchild.

A compact configuration, on the other hand, will not have a pure stepchild. If $x_m < v_m$, then all the natural daughters will be semi-compact, but the sons will not. However, only that daughter $\bar{x}$ with $\bar{x}_{m-1} = x_{m-1} - 1$ will also be a stepchild. If $x_m = v_m$ then there are no natural daughters, but all legitimate sons will be semi-compact. If $v_{m+1} = 0$, then any bastard son $\bar{x}$ with $\bar{x}_h = 1$, where $h = \min\{i > m: v_i > 0\}$ will also be semi-compact. However, the only natural son (legitimate or not) that is also a stepson is one with $\bar{x}_h = 1$, and $\bar{x}_{h-1} = x_{h-1} - 1$. Moreover, every compact configuration will have at least one son or daughter, unless $x_m = v_m$ and $\{i > m: v_i > 0\} = \emptyset$. However, there can be only one compact configuration with this property, and it is obviously terminal. QED

We thus see that the set of semi-compact configurations forms a chain in the steptree. The function $t(x) = \sum i x_i$ achieves its minimum value at the root, which is contained in this chain, and (as can be readily seen) reaches its maximum value at the compact configuration in Theorem 5.17 without offspring. It follows that this chain is in fact the longest path in the tree. Moreover, the value of t changes by 1 for each step along the path. Thus there is a semi-compact configuration for each value of this function.

5.4.2 Lateral trees

To enumerate all the configurations with $t(x) = \kappa$ we will embed them in a tree whose root will be the unique semi-compact configuration having $t(x) = \kappa$. We call this tree the *lateral tree* at κ .

This tree structure is more easily described in terms of the cumulative sums $s_i = \sum_{j=i}^k x_j$. Recall that $t(x) = \sum i x_i = \sum_{i=1}^k s_i$. Thus the configurations with $\sum i x_i = \kappa$ are also those with $\sum_{i=1}^k s_i = \kappa$. Thus we can model the lateral tree after the natural tree, which involves configurations with $\sum_{i=0}^k x_i = n$. The analogy is not as simple as it may appear, because the list of s_i 's has more complicated constraints. In particular, for legitimate configurations we require that

$$0 \vee s_{i-1} - v_{i-1} \leq s_i \leq s_{i-1} \text{ for } i = 1, \dots, k+1, \quad (5.31)$$

where for convenience we define $s_0 = n$ and $s_{k+1} = 0$. The tree structure is rendered simpler if we admit phantom configurations. In fact, we have to allow more general deficits than those specified in (5.17). For now, we will allow a deficit in any number of null cells, with the proviso that the deficit can appear in neither the first nor the last non-empty cell. Later, we will more precisely specify what deficits are required to keep the lateral tree from falling apart.

Deficits appear in the partial sums as violations of the right side inequality of (5.31). We thus relax the requirement that $s_i \leq s_{i-1}$ if $v_i = 0$. The set of configurations in the lateral tree can now be formally described as follows:

$$s = (s_0, s_1, \dots, s_k, s_{k+1}) \text{ where } s_0 = n, \quad s_{k+1} = 0, \quad \sum_{i=1}^k s_i = \kappa,$$

Figure 5.3: A tree of configurations showing the lateral trees at, each value of κ . (In this figure the node labels represent the cumulative sums s .) The vertical line represents the chain of semi-compact configurations. Note the phantom nodes. Nodes to the left represent lateral offspring. Capacities are $v = (2, 2, 0, 2, 2)$, $n = 5$.

$$\text{and } 0 \vee s_{i-1} - v_{i-1} \leq s_i \leq s_{i-1} + 1 \quad (i = 1, 2, \dots, k+1) \quad (5.32)$$

with equality $s_i = s_{i-1} + 1$ only if $v_{i-1} = 0$ and $s_{i-1} < n$.

The definition of lateral parent now resembles that of natural parent, in the sense of transferring a "ball" from one cell to another cell to its left. This operation necessarily preserves the constraint $\sum_{i=1}^k s_i = \kappa$. However, we also have to ensure that it does not violate the inequality constraints (5.32). The lateral parental mapping will be defined using the following terms (some of which have already been used).

Given a configuration $s = (s_0, s_1, \dots, s_k)$, the i^{th} cell will be called

- *minimal* if $s_i = s_{i-1} - v_{i-1} \vee 0$,
- *strictly minimal* if $s_i = s_{i-1} - v_{i-1}$,
- *almost minimal* if $s_i = s_{i-1} - v_{i-1} + 1$,
- *superminimal* if $s_i > s_{i-1} - v_{i-1} + 1$, and
- *submaximal* if $s_i \leq s_{i-1} < n$ with equality if and only if $v_{i-1} = 0$.

We are now prepared to define the lateral parent of a configuration s . Let

$$l = \max\{i: i \text{ is not minimal}\}, \quad (5.33)$$

$$J_0 = \{i < l: i \text{ is submaximal}\}, \quad (5.34)$$

$$J_1 = \{i < l-1: i+1 \text{ is not strictly minimal}\}, \quad (5.35)$$

$$J_2 = \{i = l-1: l \text{ is superminimal}\}, \text{ and} \quad (5.36)$$

$$j = \max J_0 \cap (J_1 \cup J_2). \quad (5.37)$$

(Note that $J_2 = \emptyset$ if and only if l is almost minimal).

We now define the lateral parent $\bar{s}$ of a configuration s by

$$\begin{aligned} \bar{s}_l &= s_l - 1, \\ \bar{s}_j &= s_j + 1, \text{ and} \\ \bar{s}_i &= s_i \text{ if } j \neq i \neq l. \end{aligned} \quad (5.38)$$

The configuration $\bar{s}$ will not be defined if and only if either of the sets $\{i: s_i > s_{i-1} - v_{i-1} \vee 0\}$ or $(J_1 \cup J_2) \cap J_0$ is empty. Obviously, the mapping $\bar{\cdot}$ leaves invariant $\sum_{i=1}^k s_i$ and decreases $\sum_{i=1}^k i s_i$. Thus to prove that it does in fact define a tree with a semi-compact root, we need to show that:

1. Its image is a configuration satisfying (5.32), and
2. It is well defined for all configurations except semicompact ones.

This we do in the following theorems.

Theorem 5.18 *If s satisfies (5.32) then so does $\bar{s}$.*

Proof: The mapping $\bar{\cdot}$ changes the contents of cells j and l , but leaves all other cells unchanged. It is thus necessary and sufficient to verify the inequalities involving cells j , $j+1$, l , and $l+1$. Consider first the $l+1^{\text{st}}$ cell. By definition

$$s_l - v_l \vee 0 = s_{l+1} \leq s_l + 1. \quad (5.39)$$

Now

$$\begin{aligned} \bar{s}_l - v_l \vee 0 &= s_l - 1 - v_l \vee 0 \\ &\leq s_l - v_l \vee 0 \\ &= s_{l+1} \\ &= \bar{s}_{l+1} \end{aligned} \quad (5.40)$$

so that $\bar{s}_l - v_l \vee 0 \leq \bar{s}_{l+1}$ as required.

If $v_l > s_l$ then

$$\bar{s}_{l+1} = s_{l+1} = 0 \leq s_l - 1 = \bar{s}_l$$

since $s_l > 0$, whereas if $0 < v_l \leq s_l$ then

$$\bar{s}_{l+1} = s_{l+1} = s_l - v_l \leq s_l - 1 = \bar{s}_l.$$

If $v_l = 0$ then

$$\bar{s}_{l+1} = s_{l+1} = s_l = \bar{s}_l + 1.$$

Thus we have

$$\bar{s}_l - v_l \vee 0 \leq \bar{s}_{l+1} \leq s_l + 1 \quad (5.41)$$

with $\bar{s}_{l+1} = \bar{s}_l + 1$ only if $v_l = 0$. Since, moreover, $\bar{s}$ will be undefined if $s_l = n$, (5.32) is satisfied for $i = l+1$.

Similarly, consider the j^{th} cell. We have

$$s_{j-1} - v_{j-1} \vee 0 \leq s_j \leq s_{j-1} + 1 \quad (5.42)$$

and by definition of j , $s_j < s_{j-1}$ if $v_{j-1} > 0$, and $s_j = s_{j-1}$ if $v_{j-1} = 0$. Thus $\bar{s}_j = s_j + 1 \leq s_{j-1} + 1$, with equality if and only if $v_{j-1} = 0$. Thus we have

$$\bar{s}_{j-1} - v_{j-1} \vee 0 \leq \bar{s}_j \leq \bar{s}_{j-1} + 1 \quad (5.43)$$

with the left hand inequality being obvious.

Now, consider the l^{th} cell. We have $s_{l-1} - v_{l-1} \vee 0 < s_l \leq s_{l-1} + 1$. Clearly,

$$\bar{s}_l = s_l - 1 \leq s_{l-1} + 1 \leq \bar{s}_{l-1} + 1$$

and if $j < l-1$ then

$$\bar{s}_l = s_l - 1 \geq s_{l-1} - v_{l-1} \vee 0 = \bar{s}_{l-1} - v_{l-1} \vee 0.$$

On the other hand, if $j = l-1$, then $j \in J_2$ and hence $s_{l-1} - v_{l-1} + 1 < s_l$ so that

$$\bar{s}_l = s_l - 1 \geq s_{l-1} - v_{l-1} + 1 = \bar{s}_{l-1} - v_{l-1}. \quad (5.44)$$

Now finally, consider the $j+1^{\text{st}}$ cell in the case $j < l-1$. We have $s_j - v_j < s_{j+1} \leq s_j + 1$. Clearly $\bar{s}_{j+1} = s_{j+1} \leq s_j + 1 = \bar{s}_j$, and also

$$\bar{s}_{j+1} = s_{j+1} > s_j - v_j \geq s_j + 1 - v_j = \bar{s}_j - v_j. \quad (5.45)$$

Thus all the conditions of (5.32) are satisfied. QED

We next show that $\bar{s}$ is undefined only for semi-compact configurations.

Theorem 5.19 *If s is not the root of the natural tree, then $\{i: s_i > s_{i-1} - v_{i-1} \vee 0\}$ is not empty.*

Proof: Suppose that $\{i: s_i > s_{i-1} - v_{i-1} \vee 0\} = \emptyset$. Then $s_i = s_{i-1} - v_{i-1}$ for all i such that $s_i > 0$. Now if for any i , $s_i = 0$, then $s_{i+1} > 0$ would only be possible if $v_i = 0$, in which case $s_{i+1} = s_i - v_i = s_i = 0$. Thus there must be an m such that $s_i = 0$ for $i > m$ and $s_i = s_{i-1} - v_{i-1}$ for $i \leq m$. That is for $i < m$, $x_i = s_i - s_{i+1} = v_i$. This configuration is the root of the natural tree. QED

The root of the natural tree is the only configuration at which t takes on its minimum value. Thus the lateral tree for this value of t is degenerate, consisting solely of this node.

Theorem 5.20 *If J_0 is empty, then s is compact.*

Proof: $J_0 = \emptyset$ implies that s_i is maximal for all $i < l$. Hence $s_i = s_{i-1}$ if $v_{i-1} > 0$, and $s_i = s_{i-1} + 1$ if $v_{i-1} = 0$ and $s_{i-1} < s_0$. However, since $s_i \geq s_{i-1}$ for all $i < l$, the condition $s_{i-1} < s_0$ cannot arise, and so we must have $s_i = s_{i-1}$ for all $i < l$, or $x_i = s_i - s_{i+1} = 0$ for all $i < l-1$. The definition of l , moreover, requires that $s_i = s_{i-1} - v_{i-1}$ for any $i > l$ such that $s_i > 0$, which is equivalent to $x_i = s_i - s_{i+1} = v_i$ for $l \leq i < \max\{i: s_i > 0\}$. These conditions make s compact. QED

Finally, we note that $(J_1 \cup J_2) \cap J_0 = \emptyset$ if and only if $J_1 \cap J_0 = \emptyset$ and $J_2 \cap J_0 = \emptyset$. Since $J_0 = \emptyset$ implies that s is compact, we can, for the next theorem assume that $J_0 \neq \emptyset$.

Theorem 5.21 *If J_0 is non-empty, then either $J_1 \cap J_0$ or $J_2 \cap J_0$ is nonempty unless s is semi-compact.*

Proof: Let $i \in J_0$. If $i < j-1$ then $i \notin J_2$ but $i \in J_1$ unless $s_{i+1} = s_i - v_i$. If $v_i > 0$ then $s_{i+1} < s_i$; whereas if $v_i = 0$ then $s_{i+1} = s_i$. In either case, $i+1$ is not maximal and $i+1 \in J_0$. Thus if $j = \min J_0$ then $j \leq i < l$ implies that $i \in J_0$, and $J_0 \cap J_1 = \emptyset$ only if $s_{i+1} = s_i - v_i$ for $j \leq i < l-1$. Now, $l-1 \in J_2$ unless $s_l = s_{l-1} - v_{l-1} + 1$.

Thus $J_1 \cap J_0$ and $J_2 \cap J_0$ are both empty only if

$$\begin{aligned} s_i &= s_{i+1} \text{ for } i < j-1, \\ s_{i+1} &= s_i - v_i \text{ for } j-1 \leq i < l-1, \\ s_l &= s_{l-1} - v_{l-1} + 1, \text{ and} \\ s_{i+1} &= s_i - v_i \text{ for } i \geq l \text{ if } s_{i+1} > 0. \end{aligned}$$

But these conditions are equivalent, in terms of the x_i 's to

$$x_i = \begin{cases} 0 & \text{for } i < j-1, \\ v_i & \text{for } j-1 < i < l-1, \\ v_i - 1 & \text{for } i = l-1, \\ v_i & \text{for } l-1 < i < m = \max\{i: s_i > 0\}. \end{cases} \quad (5.46)$$

which defines a semi-compact configuration. QED

Thus we see that the lateral parental mapping does define a tree, whose root is the unique semi-compact configuration having a given value of $t(x)$.

5.4.3 Inverting the lateral tree

In order to use the lateral tree for enumeration, we have to be able to invert the parental mapping. This is most easily done in terms of pitchers and catchers, as defined in Section 5.1.3. Every lateral offspring of a configuration is defined by a catcher and a pitcher. Identifying the offspring in the lateral tree is then equivalent to finding all possible catchers and their corresponding pitchers, such that the lateral parent of the result of tossing a ball from a pitcher to a catcher is the original configuration. Thus the set of offspring can be defined in terms of a list of lists:

$$\begin{aligned} & (r_1, \dots, r_f) \\ & (g_{11}, \dots, g_{1\nu_1}) \\ & \vdots \\ & (g_{f1}, \dots, g_{f\nu_f}), \end{aligned} \quad (5.47)$$

where r_ζ represents the ζ^{th} catcher and $g_{\zeta\eta}$, the η^{th} pitcher for this catcher. With this notation, the set of all lateral offspring $\bar{s}^{(\zeta\eta)}$; $\zeta = 1, \dots, f$; $\eta = 1, \dots, \nu_\zeta$ of a configuration s can be defined by

$$\begin{aligned} \bar{s}_{g_{\zeta\eta}}^{(\zeta\eta)} &= s_{g_{\zeta\eta}} - 1, \\ \bar{s}_{r_\zeta}^{(\zeta\eta)} &= s_{r_\zeta} + 1, \\ \text{and } \bar{s}_i^{(\zeta\eta)} &= s_i \text{ if } g_{\zeta\eta} \neq i \neq r_\zeta. \end{aligned} \quad (5.48)$$

We thus require necessary and sufficient conditions for a cell to appear in the list of catchers, and similarly, for a cell to appear in the list of pitchers for a particular catcher.

A cell l can be a catcher if the configuration resulting from incrementing that cell would give a permissible configuration with $l = \max\{i: i \text{ is not minimal}\}$. (This does not exclude the possibility that the list of pitchers for this catcher may be empty).

Theorem 5.22 *A cell can be a catcher if and only if all of the following conditions hold:*

- (a) l is submaximal,
- (b) $l+1$ is not strictly minimal,

(c) If $s_{l+1} > 0$ then $s_{l+1} = s_l - v_l + 1$, and

(d) If $s_i > 0$ for any $i > l + 1$ then $s_i > s_{i-1} - v_{i-1} + 1$.

Proof: We shall prove necessity first. If either (a) or (b) is false then the configuration $\bar{s}$ with $\bar{s}_l = s_l + 1$ would not satisfy (5.32). Now suppose that $s_{l+1} > 0$ but (c) fails. By (b) we must have that $s_{l+1} > s_l - v_l + 1 = \bar{s}_l - v_l$, so that $l + 1$ would not be minimal. Thus $l \neq \max\{i : i \text{ is not minimal}\}$. The necessity of (d) follows similarly.

Now suppose that (a)-(d) all hold. Then (a) and (b) guarantee that $\bar{s}$ is a configuration and (c) and (d) indicate that l is minimal for $i > l$. However, since $\bar{s}_l = s_l + 1 > 0 \vee s_{l-1} - v_{l-1}$, l is not minimal in $\bar{s}$. QED

Remark: Conditions (b)-(d) are trivially satisfied if $l = k$.

To identify a pitcher for a catcher l we similarly have to consider cells j such that a configuration $\bar{s}$ with $\bar{s}_j = s_j - 1$ and $\bar{s}_l = s_l + 1$ will satisfy (5.32) and will have $j = \max J_0 \cap (J_1 \cup J_2)$ as required by (5.37). Note that the case $j = l - 1$ requires special treatment. We define some more terms so as to simplify the statement of the next theorem.

We will call a cell $i < l - 1$

- *potential* if i is not minimal and $i + 1$ is submaximal,
- *acceptable* if i is submaximal and $i + 1$ is not strictly minimal, and
- *strongly acceptable* if it is acceptable and $s_i < s_{i-1} - 1$.

We use the same terms for $i = l - 1$ except that in this case we call i potential if it is not minimal and $s_{i+1} < s_i - 1$, and acceptable if it is submaximal and $s_{i+1} > s_i - v_i - 1$. We can now give necessary and sufficient conditions for pitchers.

Theorem 5.23 *A cell j can be a pitcher for a catcher l if and only if all the following conditions hold:*

- (a) j is potential,
- (b) if $j < l - 1$ then $j + 1$ is not strongly acceptable, and
- (c) if $j + 1 < i < l$ then i is not acceptable.

Proof: We prove necessity first. If (a) is violated, then the configuration $\bar{s}$ with $\bar{s}_j = s_j - 1$ will violate (5.32). Now suppose that $j < l - 1$ but $j + 1$ is strongly acceptable. Then $\bar{s}_{j+1} = s_{j+1} < s_j - 1 = \bar{s}_j$ so that $j + 1$ will be submaximal in $\bar{s}$ and by definition $j + 1$ will not be strictly minimal in $\bar{s}$. Thus $j + 1 \in J_0 \cap (J_1 \cup J_2)$ in (5.37) and $j \neq \max J_0 \cap (J_1 \cup J_2)$. For $j + 1 < i < l$, if i is merely acceptable then since $\bar{s}_i = s_i$ and $\bar{s}_{i-1} = s_{i-1}$, we again have $i \in J_0 \cap (J_1 \cup J_2)$, so again $j \neq \max J_0 \cap (J_1 \cup J_2)$.

Now, to consider sufficiency, we note that (a) implies that $\bar{s}$ will satisfy (5.32). Also, (b) implies that $j + 1 \notin J_0 \cap (J_1 \cup J_2)$, and for $j + 1 < i < l$, (c) implies that $i \notin J_0 \cap (J_1 \cup J_2)$. Thus $\max J_0 \cap (J_1 \cup J_2) \leq j$.

However, $\bar{s}_j = s_j - 1$ is clearly submaximal, and

$$\begin{aligned}\bar{s}_{j+1} &= \begin{cases} s_{j+1} & \text{if } j < l-1 \\ s_{j+1} + 1 & \text{if } j = l-1 \end{cases} \\ &\geq s_j - v_j \\ &> \bar{s}_j - v_j\end{aligned}$$

is not strictly minimal. Thus $j \in J_0 \cap (J_1 \cup J_2)$ and $j = \max J_0 \cap (J_1 \cup J_2)$. QED

Using these theorems, one can construct an algorithm for visiting the lateral tree, again using the stack ideas of Section 5.1.2. Each catcher satisfying the conditions of Theorem 5.22 would generate one stack entry for every pitcher satisfying the conditions of Theorem 5.23. These entries will then themselves be scanned for catchers and pitchers, and the process would iterate until the list of pitchers was empty, at which point the stack would be popped.

5.5 Grafted Trees

A tree can be *labelled* by defining a function from the set of nodes to some set of *labels*. We in fact have been implicitly using such labels by assigning to each node a configuration, and by assigning to each node a weight. Now it is possible for the labels themselves to be other trees, so that we can define a tree of trees. Given a tree T , and a set $\mathcal{T}$ of other trees, we can define a *grafting* as a function $\sigma: T \rightarrow \mathcal{T}$. The grafting in fact defines another tree, $\mathcal{T}$ *grafted onto* T , which we denote by $\sigma[T]$. The nodes of $\sigma[T]$ are ordered pairs (x, y) where $x \in T$ and $y \in \sigma(x)$. The parental mapping is defined by $(\widehat{x, y}) = (x, \hat{y})$ if y is not the root of $\sigma(x)$, and $(\hat{x}, \text{rt } \sigma(\hat{x}))$ if y is the root of $\sigma(x)$. It is obvious that this mapping defines a tree whose root is $(\text{rt } T, \text{rt } \sigma(\text{rt } T))$.

An important special case occurs if $\mathcal{T}$ is a singleton, containing a single tree S . In that case the set of nodes of the grafted tree is essentially the Cartesian product of the sets of nodes of T and S . We will call such a construction a *product tree* and denote it by $T \times S$. It should be noted that the partial order defined by this product of trees is not the same as would obtain from the Cartesian product of the trees considered merely as partially ordered sets. This latter ordering, in fact, would not in general even be definable by a tree structure.

The Cartesian product of trees is not commutative, so that $S \times T$ is not necessarily even isomorphic $T \times S$. However, the following lemma shows that the product is associative, so that if S , say, is itself the product of two trees S_0 and S_1 , we can unambiguously write $T \times S_0 \times S_1$ for $T \times (S_0 \times S_1)$.

Lemma 5.7 *If S_1, S_2 , and S_3 are three trees, then $(S_1 \times S_2) \times S_3$ and $S_1 \times (S_2 \times S_3)$ are isomorphic.*

Proof: Let $T_1 = (S_1 \times S_2) \times S_3$ and $T_2 = S_1 \times (S_2 \times S_3)$. Obviously T_1 and T_2 have the same nodes. We need to show that they also have the same parental mapping. Consider $y = (x_1, x_2, x_3)$ as a node of T_1 where $x_i \in S_i$, $i = 1, 2, 3$. Then if $x_3 \neq \text{rt } S_3$ then

$\hat{y} = (x_1, x_2, \hat{x}_3)$. If $x_3 = \text{rt } S_3$, then $\hat{y}$ will involve the parent of (x_1, x_2) , which will be $(x_1, \hat{x}_2)$ or $(\hat{x}_1, x_2)$ depending on whether $x_2 = \text{rt } S_2$.

Now, to find the parent of y considered as a node in T_2 we must determine whether $(x_2, x_3) = \text{rt } S_2 \times S_3$. This can happen only if $x_2 = \text{rt } S_2$ and $x_3 = \text{rt } S_3$. In this case, the parent of y will be $(\hat{x}_1, x_2, x_3)$ which equals the parent under T_1 . If however $x_3 \neq \text{rt } S_3$ then (x_2, x_3) cannot be $\text{rt } S_2 \times S_3$, but will have a parent $(x_2, \hat{x}_3)$ giving a parent $(x_1, x_2, \hat{x}_3)$ in T_2 , which again coincides with the parent in T_1 . A similar argument can be used for the case that $x_3 = \text{rt } S_3$ but $x_2 \neq \text{rt } S_2$. QED

Once tree structures have been defined on several sets of configurations, then the product of these trees defines a tree structure on the Cartesian product of the sets of configurations. Weights can be assigned to the nodes of the product trees as the product of the weights of the component trees. With this set-up, product trees can be used for calculation of probabilities involving the joint distribution of independent configurations.

If (x, y) is a node of a product tree $S \times T$, then the weight of the branch springing from (x, y) can be calculated simply from the weights of the component trees. If $y = \text{rt } T$ then $P(B_{(x,y)}) = P(B_x)$, since every node in B_x corresponds to a branch in $B_{(x,y)}$ whose weight is $P(x)P(T)$, and $P(T) = 1$. On the other hand, if $y \neq \text{rt } T$, then $P(B_{(x,y)}) = P(x)P(B_y)$ since every node of the branch will have x as its first component.

Product trees are particularly useful for calculating the probabilities of events defined in terms of linear functions of a pair of independent configurations. Given two trees of configurations denoted by $x = (x_0, \dots, x_k) \in S$ and $y = (y_0, \dots, y_l) \in T$, define linear functions $t_1 = \sum_{i=0}^k ix_i$ and $t_2 = \sum_{i=0}^l iy_i$. Then we could determine $\Pr\{t_1 + t_2 \geq c\}$ directly as a sum of branches of the product tree, but a more ingenious method would reduce the computational burden. This method involves conditioning on the value of t_1 . Clearly

$$\Pr\{t_1 + t_2 \geq c\} = \sum_i \Pr\{t_1 + t_2 \geq c | t_1 = i\} \Pr\{t_1 = i\}. \quad (5.49)$$

Now since $t_2 \geq 0$, $\Pr\{t_1 + t_2 \geq c | t_1 = i\} = 1$ if $i \geq c$. Moreover, if M_2 is the maximum possible value of t_2 then $\Pr\{t_1 + t_2 \geq c | t_1 = i\} = 0$ if $i < c - M_2$. Thus

$$\begin{aligned} \Pr\{t_1 + t_2 \geq c\} &= \sum_{i=c-M_2+1}^{c-1} \Pr\{t_1 + t_2 \geq c | t_1 = i\} \Pr\{t_1 = i\} + \Pr\{t_1 \geq c\} \\ &= \sum_{i=c-M_2+1}^{c-1} (1 - \Pr\{t_2 < c - i\}) \Pr\{t_1 = i\} + \Pr\{t_1 \geq c\}. \end{aligned} \quad (5.50)$$

The last term is just a sum of weights of branches of S . The summation involves the products of ascending slices of S and initial parts of T . These factors can be evaluated using lateral trees connected by the chain of semi-compact configurations. Such an evaluation can exploit truncation by stopping the summation once $1 - \Pr\{t_2 < c - i\}$ drops below a tolerance ϵ .

5.6 Summary

The recursive computation of probabilities in multivariate situations requires an ordering defined on the sample space. Such an ordering can be achieved by the concept of a tree, which defines a partial order on the sample points and provides a structure for sequentially calculating all their probabilities. Branches of the tree provide a natural generalization of the idea of the tail of a univariate discrete distribution. In statistical applications, one may be primarily interested in the probability of such branches, which may represent the rejection region of a hypothesis test.

One tree structure, which we have called the natural tree, provides a convenient formula for computing the probability of a branch in the case where the probability distribution is in the multivariate Pólya family. This multivariate distribution is defined in terms of a sequence of conditional distributions, each being itself a univariate Pólya distribution. The probability of a branch in this natural tree is then a marginal probability involving the tail of one of the conditional distributions. The computation of probabilities in the multivariate case is thus reduced to computations in the univariate case.

The rejection region of a hypothesis test defined in terms of a linear test statistic corresponds to a collection of entire branches of the natural tree. The calculation of its probability is thus particularly convenient. In the natural tree, however, the branches making up such a rejection region do not necessarily originate from points of the same (graph-theoretic) height. It may thus prove difficult to identify the boundary of this rejection region. We solved this problem by defining another structure, called a step-tree, in which the rejection region corresponds precisely to all the nodes of the tree above a given height. To make effective use of this new structure, we examined in detail the relationship between the natural tree and the step-tree. The definition of the step-tree required, in certain cases, the augmentation of the sample space by phantom points having a probability of zero.

Although the natural and step-trees prove convenient for computing probabilities, they do not directly give procedures for finding all the boundary points of the rejection region. A third tree structure, called a lateral tree, allows one to step-through precisely these points. The definition of this tree again may require the introduction of phantom points.

The tree concepts can be extended to more complex discrete probability spaces. By assigning to each node of one tree another entire tree, one can consider the situation of a discrete multivariate distribution being conditional on a random outcome on another simplex. We formalized this idea in the concept of a grafted tree. In the special case where these two multivariate outcomes are independent, this construction leads to the product of two trees. By grafting trees onto trees that are themselves grafted, these concepts extend naturally to any number of component trees. In the next chapter, such construction will in fact be useful for considering the distributions arising from multivariate conditional tests.

Chapter 6

Permutation Tests with Binary Data

In this chapter, we will consider the problem of computing exact operating characteristics for certain permutation tests, using the ideas developed in the previous two chapters. We begin by describing how a fast algorithm for computing the power of the Fisher-Irwin test can be constructed, using the recursive relations of the Pólya family and a stochastic monotonicity property of the conditional distributions on which the test is based. The randomization basis of the Fisher-Irwin test will be abstracted to a general case of hypothesis testing conditionally on an algebra of events invariant under a group of permutations. The structure of the group itself then provides the distribution of the test statistic, and indicates symmetries whereby the conditioning algebra can be reduced.

These ideas are applied to the computation of exact size and power of the bioassay designs discussed by asymptotic methods in Chapter 3, using the beta-binomial distribution to model intra-litter correlation. The monotonicity property of the family of conditional distributions noted for the Fisher-Irwin test generalizes to these cases. The conditional distributions in fact are members of the multivariate Pólya family, and hence the tree structures discussed in Chapter 5 can be used to implement computational algorithms.

6.1 The Fisher-Irwin Test

6.1.1 Motivation

Consider a bioassay of n animals in which n_0 are randomly assigned to a control treatment and n_1 are assigned to a test treatment. Suppose that x control animals respond and y test animals respond, giving a total of $z = x + y$ responders. The null hypothesis that the response rates are equal under the two treatments can be tested by a conditional test, which considers the distribution of y conditional on the total z . This conditional distribution will be hypergeometric under the null hypothesis if the responses under the two treatments are independent binomial random variables. However, the conditional distribution can be derived equally well from randomization arguments, as we will see in Section 6.3.1. This test was proposed by Irwin[37], and popularized by Fisher[29], from whom it derives its common name of "Fisher's exact test". If the test is modified to allow randomized critical regions, then it is in fact the uniformly most powerful unbiased test against the alternative hypothesis of two different independent binomial distributions

[53, p. 140]. This conditional test has also been justified on the basis of ancillarity [30], although these arguments are a little weak because the "ancillarity" of z holds only under the null hypothesis [10].

An alternative to the Fisher-Irwin test is the test statistic

$$t = \frac{\hat{p}_1 - \hat{p}_0}{\sqrt{\hat{p}(1-\hat{p})\left(\frac{n_1+n_0}{n_1n_0}\right)}} \quad (6.1)$$

where $\hat{p}_0 = x/n_0$, $\hat{p}_1 = y/n_1$ and $\hat{p} = z/n$, whose asymptotic distribution is standard normal under the null hypothesis. In fact its square is the usual Pearson χ^2 for testing for independence in a 2×2 contingency table. Yates' correction for discreteness is often included. This adjustment is based on approximating the cell frequencies by normal distributions, dividing in the usual way the continuous probabilities between half-integers, under the assumption that the margins are fixed. For a one-tailed test against the alternative $p_1 > p_0$ this adjustment replaces $\hat{p}_1$ by $\hat{p}_1 - \frac{1}{2n_1}$ and $\hat{p}_0$ by $\hat{p}_0 + \frac{1}{2n_0}$ (provided the inequality $\hat{p}_1 - \frac{1}{2n_1} > \hat{p}_0 + \frac{1}{2n_0}$ holds). It has been noted empirically by many workers, including this author, that the Yates'-corrected t -statistic very well approximates the Fisher-Irwin test, even for quite small sample sizes if $n_1 \approx n_0$.

There is, however, considerable controversy over the appropriateness of this adjustment [22,9,41]. The test without the correction has been presented as superior, on the empirical grounds that its size tends to be quite close to the nominal significance level. Since the uncorrected test always has a larger critical region, it will be more powerful. In fact it appears that the uncorrected test has operating characteristics quite similar to those of the uniformly most powerful unbiased test, while lacking the disgrace of having a randomized critical region [75].

The advantage of the conditional test, however, is that its justification does not depend on the assumption that x and y come from independent binomial populations, but only on the random allocation of animals to treatment groups as we shall see in Section 6.3.1.

6.1.2 Power calculations

Even if it does not justify the validity, the binomial model is still a reasonable reference point for evaluating and comparing operating characteristics of test procedures. At first sight, the calculation of power—even for a conditional test—seems to pose little problem. All one needs to do is to determine the critical region for each conditioning event, evaluate its probability under the alternative, and average these conditional probabilities over all possible conditions. However, a naïve approach to this calculation can take an inordinate amount of computer time even for moderate values of n_0 and n_1 . This is particularly troublesome if the evaluation of power is only an input to a procedure that is itself iterative, such as searching for an optimal design. However, the techniques of Chapter 4 provide an algorithm for the power of the Fisher-Irwin test whose complexity is at most linear in the sample size n .

Although the Fisher-Irwin test is based on a conditional distribution, it is actually simpler to evaluate the power in terms of the unconditional distributions of x and y , since these are independent. The rejection region is a subset of the $(n_0 + 1) \times (n_1 + 1)$ points representing outcomes of the experiment, which may be imbedded in the rectangle

$[0, n_0] \times [0, n_1]$ in the Cartesian plane. The calculation of power can be facilitated by conditioning on the control incidence:

$$\Pr\{\text{rejection}\} = \sum_{x=0}^{n_0} \Pr\{\text{rejection}|X=x\} \Pr\{X=x\}. \quad (6.2)$$

The probabilities of X can be calculated recursively using (4.12), and for large values of n_0 , truncation of negligible tails will avoid the calculation of many probabilities. The calculation would be further simplified if the rejection regions were themselves tails of the distribution. In particular, if the conditional critical region given $X=x$ is given by $\{Y : Y \geq \kappa(x)\}$ then the power is given by $\sum_{x=0}^{n_0} \Pr\{X=x\} \sum_{y=\kappa(x)}^{n_1} \Pr\{Y=y\}$. If, moreover, the critical value $\kappa(x)$ is increasing in x , then the power can be calculated by stepping once through the distribution of X and Y . The complexity of the algorithm is thus at most linear in the sample size n , and could conceivably be less because of the possibility of tail truncation. The following results demonstrate that the critical region of the Fisher-Irwin test does indeed have the desired property, as a consequence of a stochastic monotonicity property of the family of conditional distributions. The main theorem of this section (Theorem 6.3) is proved using intuitive probabilistic arguments. An algebraic demonstration of a more general result is found in Theorem 6.29.

6.1.3 Stochastic monotonicity

Let X and Y be two integer-valued random variables, and define

$$F(y; z) = \Pr\{Y \leq y | X + Y = z\}, \quad (6.3)$$

where the probability is calculated under the null hypothesis. Given a significance level α , each value of z will define a rejection region according to the conditional distribution (6.3). We want to define the union of all these rejection regions directly in terms of x and y .

A pair (x, y) will be called *subcritical* (at level α) if $F(y-1; x+y) < 1-\alpha$, and *supercritical* if $F(y-1; x+y) \geq 1-\alpha$. The pair will be called *critical* if it is supercritical but $(x+1, y-1)$ is subcritical. The family of distributions (6.3) will be said to have *critical tails* if for every x in the range of X there is a constant $\kappa(x)$ such that

$$\{y : (x, y) \text{ is supercritical}\} = \{y : y \geq \kappa(x)\}. \quad (6.4)$$

It will be said to have *monotone tails* if (6.4) holds and $\kappa(x)$ is non-decreasing in x . (It should be noted that the set (6.4) could be empty. While we will only deal with bounded random variables here, the definitions do not require boundedness provided we allow $\kappa(x) = \infty$.)

If a family of conditional distributions has monotone tails, then the rejection region of a conditional test can be found recursively. Given $\kappa(x)$ for a particular x , we know that $\kappa(x+1) \geq \kappa(x)$. We then determine whether $(x+1, \kappa(x))$ is supercritical. If so, it follows that $\kappa(x+1) = \kappa(x)$. If not, we then step through $(x+1, \kappa(x)+1), (x+1, \kappa(x)+2), \dots$, until we find the first supercritical pair $(x+1, \kappa(x)+j)$. Then $\kappa(x+1) = \kappa(x) + j$.

Theorem 6.1 *If the family of conditional distributions (6.3) satisfies*

$$F(y; z) \leq F(y+1; z+1) \quad (6.5)$$

then it has critical tails.

Proof: Suppose that (x, y) is subcritical. Since

$$F(y-2; x+y-1) \leq F(y-1; x+y) < 1-\alpha,$$

it follows that $(x, y-1)$ is also subcritical. Thus, by induction, (x, y') is subcritical for all $y' \leq y$. Defining

$$\kappa(x) = \inf\{y : (x, y) \text{ is supercritical}\}$$

we see that for no $y \geq \kappa(x)$ can (x, y) be subcritical. QED

Theorem 6.2 *If the family (6.3) has critical tails and is stochastically increasing:*

$$F(y; z) \geq F(y; z+1),$$

then it has monotone tails.

Proof: Since for any x , $(x, \kappa(x)-1)$ is subcritical, we have

$$F(\kappa(x)-2; x+\kappa(x)) \leq F(\kappa(x)-2; x+\kappa(x)-1) < 1-\alpha$$

so that $(x+1, \kappa(x)-1)$ is also subcritical. Thus $\kappa(x+1) \geq \kappa(x)$, and the result follows by induction. QED

Theorem 6.3 *If F is defined by the hypergeometric family:*

$$F(y; z) - F(y-1; z) = \frac{\binom{n_0}{z-y} \binom{n_1}{y}}{\binom{n_0+n_1}{z}}, \quad (6.6)$$

then

$$F(y; z+1) \leq F(y; z) \leq F(y+1; z+1). \quad (6.7)$$

Proof: We prove this theorem by an urn model argument. The hypergeometric formula (6.6) can be written in the alternative form

$$\frac{\binom{z}{y} \binom{n_0+n_1-z}{n_1-y}}{\binom{n_0+n_1}{n_1}}$$

which can be considered to be the probability of selecting y black balls in a sample of n_1 out of a population of x black balls and $n_0 + n_1 - x$ white balls. Suppose now that one of the white balls is distinguished in that it changes colour ("by magic") after the sample is drawn. In this way, any event involving a population having x black balls is mapped onto an event involving a population having $x + 1$ black balls.

Let $b_{y,x}$ be the probability of drawing y out of x black balls in addition to the magic white ball, and let $a_{y,x}$ be the probability of drawing y out of the x black balls and leaving the magic ball in the urn. Then

$$F(y; x) = \sum_{j=0}^y (a_{j,x} + b_{j,x}). \quad (6.8)$$

Now, we can obtain the event of drawing j black balls out of a population having $x + 1$ by either drawing j and leaving the magic one in the urn, or by drawing $j - 1$ and the magic one. Thus

$$\begin{aligned} F(y; x + 1) &= \sum_{j=0}^y (a_{j,x} + b_{j-1,x}) \\ &= \sum_{j=0}^y a_{j,x} + \sum_{j=0}^{y-1} b_{j,x} \end{aligned} \quad (6.9)$$

and hence

$$F(y; x + 1) = F(y; x) - b_{y,x} \leq F(y; x). \quad (6.10)$$

Similarly, we obtain the event of drawing $j + 1$ black balls out of a population having $x + 1$ by drawing j black balls and the magic one, or $j + 1$ black balls without the magic one. Thus

$$\begin{aligned} F(y + 1; x + 1) &= a_{0,x} + \sum_{j=0}^y (a_{j+1,x} + b_{j,x}) \\ &= \sum_{j=0}^{y+1} a_{j,x} + \sum_{j=0}^y b_{j,x} \\ &= F(y; x) + a_{y+1,x} \geq F(y; x). \end{aligned} \quad (6.11)$$

QED

Corollary 6.3.1 *The conditional distributions of the Fisher-Irwin test form a family with monotone tails.*

6.1.4 Computing algorithm

To calculate the power of the Fisher-Irwin test, we begin by finding $\kappa(0)$. In this case the pair (x, y) will be supercritical if

$$\prod_{i=0}^{y-1} \frac{n_1 - i}{n - i} < \alpha \quad (6.12)$$

since $F(y; y) = 1$. Thus $\kappa(0)$ is the smallest y satisfying (6.12) and can be determined by multiplying the factors until the product drops below α . Once $\kappa(x-1)$ is known, we can see if $\kappa(x) = \kappa(x-1)$ by determining whether the tail probability $1 - F(x; \kappa(x-1)) < \alpha$. This calculation can be expedited by using the recursion

$$\frac{\binom{n_0}{x} \binom{n_1}{y}}{\binom{n}{x}} = \frac{z(n_0 - x + 1)}{x(n - z + 1)} \frac{\binom{n_0}{x-1} \binom{n_1}{y}}{\binom{n}{x-1}} \quad (6.13)$$

to avoid recalculation of the binomial coefficients, and using the tail bounds discussed in Section 4.2.3 to reduce the number of probabilities that need to be calculated. If $(x, \kappa(x-1))$ turns out to be subcritical, the determination is repeated for $(x, \kappa(x-1) + 1)$, whose conditional probability can again be determined recursively from that of $(x, \kappa(x-1))$ by a formula similar to (6.13).

It should be noted that the function κ cannot in practice be calculated exactly using floating point arithmetic because of the problem of round-off error. One cannot determine with certainty that $1 - F(y; z) \leq \alpha$ for some y , but only that $1 - F(y; z) \leq \alpha + \delta$, where δ is some small but unknown round-off error. Thus it is possible to find, computationally, that (x, y) is subcritical at level α , when in fact (x, y) is critical at level α but is subcritical at level $\alpha - \delta$. Such a small error may cause an appreciable error in calculation of power because the probability distribution is discontinuous. Thus such an algorithm cannot really find the power corresponding to a precise level α but only to a "fuzzy" level that is approximately α . The only way out of this problem is to do the computations using a system like Maple[17] that performs rational arithmetic without round-off errors. Such computations are generally far more time-consuming, but with efficient algorithms based on the above recursive relations this solution is at least ponderable.

It should be noted that nothing in the proposed methodology requires that the distributions of X and Y be binomial. Thus the outlined procedure could be used to determine the power of the Fisher-Irwin test for any distribution. If these distributions are from the Pólya family, then the same ideas of recursion and tail truncation can be applied. The only requirement is that the responses under the two treatments be independent. A case in which independence is not required will be discussed in Section 6.5.1.

The approach will also work for calculation of the unconditional probability of the rejection region of any test statistic that can be defined in terms of a non-decreasing function $\kappa(x)$ as in (6.4), whether or not the test is explicitly conditional. There is thus a convenient approach for determining the exact operating characteristics of numerous approximate procedures that have been proposed. It can be shown that the rejection region defined by the statistic (6.1) as well as its continuity-corrected version has this monotonicity property, as does that defined by applying the arcsine transformation to the estimated response probabilities [83].

6.2 Permutation Tests

We now examine the properties of permutation tests in general, abstract terms. Although all our applications involve binary random variables, we do not restrict ourselves to this case here. To simplify the proofs, we will, however, usually require that the probability distributions have finite or (possibly countable) support, even though many of the theorems may be true under more general assumptions.

6.2.1 Probability algebras

Behind every probability statement there lies a probability space $(\Omega, \mathcal{A}, P)$ where $\mathcal{A}$ is a σ -algebra of events (i.e. subsets of Ω), and P is a probability measure on $\mathcal{A}$. It may happen that there exist events $A \in \mathcal{A}$ such that $P(A) > 0$ but $P(B) = 0$ for every $B \subsetneq A$. Events having that property will be called *atoms* of $\mathcal{A}$, and we will use the symbolism $A \propto \mathcal{A}$ to indicate that A is an atom of $\mathcal{A}$. If every non-empty event contains atoms, then $\mathcal{A}$ is said to be *atomic*, and the probability measure P is uniquely defined by its value on atoms. (This will always happen if Ω is at most countable.) We will only be considering probability spaces of this type. To avoid tedious trivialities, we do not distinguish two atoms A and B if $P(A \triangle B) = 0$.

It can be shown that the atoms must be (almost) disjoint, are at most countable in number, and that for atomic algebras their union has probability one. Thus (neglecting sets of measure zero) the set of atoms forms a partition of the sample space Ω , and every probabilistic statement can be expressed in terms of the atoms. Conversely, the set of all subsets of a finite (or countable) partition will define an atomic algebra which will be a subalgebra of $\mathcal{A}$ if the partition is measurable. We will call this set the (sub)algebra *generated* by the partition. The partition itself may be defined by the level sets of a (simple) random variable, in which case we say the algebra is generated by that random variable. Obviously, if we have a formula for computing the probability of any atom, we can compute the probability of any event.

Sometimes all events of interest lie in a subalgebra $\mathcal{B} \subset \mathcal{A}$. For us, such a situation is particularly helpful if there is a formula for computing the probability of the atoms of $\mathcal{B}$ directly, without having to compute probabilities of the atoms of $\mathcal{A}$, for then we can find the probabilities of interest with less work. We generalize this idea, borrowing terminology from the theory of stochastic processes, by defining a *filtration* as an increasing sequence $\mathcal{A}_0 \subset \mathcal{A}_1 \subset \dots \subset \mathcal{A}$ of subalgebras¹ of $\mathcal{A}$, where by convention, $\mathcal{A}_0 = \{\emptyset, \Omega\}$. If we have formulas for calculating probabilities for the atoms of each algebra in the filtration, then obviously it would be expeditious to find $P(A)$ for some event A by using the formula for the first algebra $\mathcal{A}_i$ such that $A \in \mathcal{A}_i$. The formulas for probabilities can sometimes be expressed in terms of *transition probabilities*, defined as a sequence of functions $P_1, P_2, \dots$ where $P_i: \Omega \times \mathcal{A}_i \rightarrow [0, 1]$ is defined by $P_i(\cdot, A) = \Pr\{A | \mathcal{A}_{i-1}\}$ for any $A \in \mathcal{A}_i$. In fact the definition of the multivariate Pólya distribution in (4.46) is of this form.

¹To be useful in probability theory, subalgebras are usually required to be σ -algebras, so that they are closed under limits of sequences of events. In this chapter all algebras will be finite, and thus every subalgebra is automatically a σ -algebra. For the sake of simplicity, we dispense with prefixing σ - everywhere that a probabilist would like to see it.

If we restrict ourselves to atomic algebras, then we do not even need the sample space Ω , for everything can be expressed in terms of the set of atoms. In particular, an A -measurable function will be (almost surely) constant on an atom, and thus could be considered to "be" a function on the set of atoms. The pair (A, P) , which satisfies the axioms of probability theory abstractly without any explicit reference to an underlying sample space on which the events in A live, is called a *probability algebra*. The symbol Ω will be retained to represent the sure event that contains all other events.

We will abuse notation by using the same symbol (such as A) to denote either an algebra of events or the set of its atoms, the meaning being clear from the context. However, we will always use the notation $B \propto \mathcal{B}$ to specify that B is an atom of $\mathcal{B}$, reserving $B \in \mathcal{B}$ to mean just that B is a $\mathcal{B}$ -measurable event. If $A \in \mathcal{A}$ and $\mathcal{B} \subset \mathcal{A}$, we define AB by

$$AB = \{B \in \mathcal{B} : B \subset A\}. \quad (6.14)$$

Note that we do not require that $A \in \mathcal{B}$ so that possibly $A \notin AB$. We will also say $B \propto AB$ to mean $B \propto \mathcal{B}$ and $B \subset A$.

6.2.2 Permutation groups

By a *permutation* we mean a one-to-one mapping of a set onto itself. A set of permutations G is called a *group* if it contains the identity map 1 , is closed under composition, and if every permutation $g \in G$ has an inverse $g^{-1} \in G$ such that $g^{-1}g = 1$. If G is finite, we denote by $|G|$ its cardinality, or *order*. Only finite groups will be considered here. The symbol S_n will designate the so-called *symmetric group* of all permutations of a finite set of n elements. Note that in general isomorphic groups will not be notationally distinguished, so the actual set being permuted is immaterial. We will freely use, without proof or citation, elementary properties of groups which may be found in standard texts [58].

A group of permutations of a set X induces permutations on related sets. We will discuss some general techniques for constructing such permutations which will be useful in the sequel. Given a set X , consider the set Y^X of all mappings $f: X \rightarrow Y$ where Y is some arbitrary non-empty set. Then for any permutation $g: X \rightarrow X$ we can define $g^*: Y^X \rightarrow Y^X$ by

$$(g^*f)(x) = f(g^{-1}x)$$

for $f \in Y^X$, $x \in X$. We then have

Theorem 6.4 *If G is a group of permutations of X , and*

$$G^* = \{g^*: Y^X \rightarrow Y^X : (g^*f)(x) = f(g^{-1}x)\}$$

then G^ is a group of permutations of Y^X , and G and G^* are isomorphic.*

Proof: Since every g^* is derived from some g , it is obvious that the mapping $g \mapsto g^*$ is onto. We now show that it is one-to-one. It obviously is if $|G| = 1$, so we assume that $|G| \geq 2$. Let $g_1, g_2 \in G$ such that $g_1 \neq g_2$. Then $g_1^{-1} \neq g_2^{-1}$. Thus, there must be an $x \in X$ such that $x_1 = g_1^{-1}x \neq g_2^{-1}x = x_2$. Choose an $f \in Y^X$ such that $f(x_1) \neq f(x_2)$. Then $g_1^*f(x) = f(g_1^{-1}x) = f(x_1) \neq f(x_2) = f(g_2^{-1}x) = g_2^*f(x)$. Thus $g_1^*f \neq g_2^*f$ and $g_1^* \neq g_2^*$.

We next need to show that G^* is a group, and that its elements are permutations. To see that every element in G^* has an inverse in G^* note that for $g \in G$, $f \in Y$, $(g^{-1})^* f(x) = f(gx)$ for all $x \in X$. Thus $g^*(g^{-1})^* f = f$ so that $g^*(g^{-1})^* = 1$, the identity map on Y^X . Therefore, $(g^{-1})^* = (g^*)^{-1}$.

If $f \in Y^X$, then so is $g^* f$ and $f = (g^{-1})^*(g^* f)$. Thus g^* is onto. If $f_1, f_2 \in Y^X$, and $g^* f_1 = g^* f_2 = f$ then $f_1 = (g^{-1})^* f = f_2$ so that g^* is one-to-one. Thus g^* is in fact a permutation.

Finally, to see that G^* is closed under composition, consider $g_1, g_2 \in G$. Then for any $f \in Y^X$, $(g_1 g_2)^* f(x) = f(g_2^{-1} g_1^{-1} x) = (g_2^* f)(g_1^{-1} x) = g_1^*(g_2^* f(x))$, and hence $(g_1 g_2)^* = g_1^* g_2^*$. QED

Since the groups G and G^* are isomorphic, we will henceforth not distinguish them, and will simply write gf for $g^* f$. There are several examples of how Theorem 6.4 will be useful.

1. If X is a finite set denoting experimental units, and Y is the set of possible observations on the units, then Y^X is the set of experimental outcomes. A permutation of the experimental units thus induces a permutation of experimental outcomes.
2. If X is as in 1 but Y is a set of random variables on some probability space, then an element of Y^X is a random vector indexed by experimental units. Permutation of experimental units permutes random vectors. (In this situation we will usually be only interested in a subset of Y^X .)
3. If $Y = \{0, 1\}$, then Y^X is equivalent to 2^X , the set of all subsets of X . Thus a permutation of X induces a permutation of the Boolean algebra of subsets of X . It is easy to show that these induced permutations are in fact automorphisms of the Boolean algebra.
4. If Y is itself a group, then so is the function set Y^X , where products are defined componentwise:

$$(f_1 f_2)(x) = f_1(x) f_2(x) \text{ for } f_1, f_2 \in Y^X.$$

The permutations of Y^X thus induced by a group of permutations of X form a group of automorphisms of Y^X .

In a similar fashion, permutations on Y will induce permutations on the function set X^Y . If H is a group of permutations on Y , and X is an arbitrary non-empty set, then consider the set H^X of mappings from X to H . For any $h \in H^X$, we can define a mapping $h_*: Y^X \rightarrow Y^X$ by

$$(h_* f)(x) = (h(x)) f(x) \text{ for } f \in Y^X, x \in X.$$

Theorem 6.5 *If H is a group of permutations of Y , and*

$$H_* = \{h_*: Y^X \rightarrow Y^X : (h_* f)(x) = (h(x)) f(x) \text{ where } h \in H^X\}$$

then H_ is a group of permutations of Y^X , isomorphic to H^X .*

Proof: The proof mimics the arguments of Theorem 6.4. To show that the mapping $h \mapsto h_*$ is one-to-one, suppose that $h_1, h_2 \in H^X$, $h_1 \neq h_2$, so that there is an $x \in X$ for which $h_1(x) \neq h_2(x)$. Now, since these are different permutations of Y , there must be a $y \in Y$ such that $h_1(x)y \neq h_2(x)y$. Now, if f is such that $f(x) \neq y$ then clearly $h_{1*}f \neq h_{2*}f$.

For $h \in H^X$, h^{-1} is defined by $h^{-1}(x) = (h(x))^{-1}$. Then clearly $(h^{-1})_*h_*f = f$, so that $(h_*)^{-1} = (h^{-1})_*$. The remainder of the proof unfolds analogously to that of Theorem 6.4.

QED

Again, since H_* and H^X are isomorphic, we will not notationally distinguish these groups or their elements.

The two techniques can be combined. If G is a group of permutations of X and H is a group of permutations of Y , then for any pair of permutations $g \in G$, $h \in H^X$, we can define the permutation hg on Y^X by

$$(hg)f(x) = (h(gf))(x) = h(x)f(g^{-1}x).$$

In general the effect of hg will differ from that of gh , since

$$(g(hf))(x) = (hf)(g^{-1}x) = h(g^{-1}x)f(g^{-1}x).$$

However, ghf can be written as $(g \star h)gf$, where by the $g \star h$ in parentheses we mean the result of the action of the permutation g on the set H^X . Thus for $g_1, g_2 \in G$, $h_1, h_2 \in H^X$, $h_1g_1h_2g_2 = h_1(g_1 \star h_2)g_1g_2$, and since $g_1 \star h_2 \in H^X$, $h_1g_1h_2g_2$ is also of the form hg . It follows that the set of such "products" is itself a group, containing isomorphic images of G and H^X as subgroups. Since the action of any g on H^X is an automorphism, the structure so defined is in fact the *semi-direct product* of H^X and G , which will be denoted by $H^X \odot G$. It can be shown that H^X is in fact a *normal* subgroup of $H^X \odot G$.

The preceding construction is precisely what is needed for discussing designs with litter structure. Letting X denote the set of litters, and Y the set of possible outcomes for a litter, the experimental outcome can be considered to be an element of Y^X . If G represents permissible permutations of litters, and H , permutations within a litter, then $H^X \odot G$ is the group of permutations that leave litters intact.

A permutation may leave some element x fixed $gx = x$. It can be easily shown that the set of permutations leaving an element fixed is a subgroup F_x called the *subgroup fixing x* . It follows that all members of the coset gF_x will carry x to the same image. Usually, however, F_x is not a normal subgroup, so that it cannot be used to construct a quotient group.

The following result will be useful.

Lemma 6.1 Let $F_x = \{g \in G : gx = x\}$. Then if $y = g_0x$ for some $g_0 \in G$ then $|F_x| = |F_y|$.

Proof: If $h \in F_y$ then $hg_0x = hy = y = g_0x$, so $g_0^{-1}hg_0x = x$ and $g_0^{-1}hg_0 \in F_x$. Conversely, if $g_0^{-1}hg_0 \in F_x$ for some $h \in G$ then $g_0^{-1}hg_0x = x$ so $hy = g_0x = y$ and $h \in F_y$. Thus $F_x = g_0^{-1}F_yg_0$. The conjugacy mapping $h \mapsto g_0^{-1}hg_0$, however, is an isomorphism, so that $|F_x| = |F_y|$.

QED

6.2.3 Permutations of probability algebras

Our main interest will be the effect of permutation groups on probability algebras. Given a probability space $(\Omega, \mathcal{A}, P)$, only permutations that are measurable are probabilistically meaningful. If $\mathcal{A}$ is atomic, then a measurable permutation of Ω is essentially a permutation of the atoms of $\mathcal{A}$. Using the construction of example 3 in Section 6.2.2 we see that such a permutation can also be considered to be a permutation of $\mathcal{A}$ itself. The underlying set Ω really plays no role, so we will consider permutation groups to be acting (ambiguously) either on an atomic algebra or on the set of its atoms.

We now define two useful concepts, and discuss their properties. Given a group G of permutations of an algebra $\mathcal{A}$ and $x \in \mathcal{A}$, the event

$$Gx = \bigcup \{gx : g \in G\}$$

is called the *orbit* of x under G , or less specifically, a G -orbit.² Clearly, every atom is contained in its own orbit, and it is easy to show that distinct orbits must be disjoint. Thus the set of orbits forms a partition of the algebra, and hence generates a subalgebra.

An event $A \in \mathcal{A}$ will be called G -invariant (or just *invariant* if the group is clear from the context) if $gA = A$ for every $g \in G$. Since $gA \subset A$ implies that $A \subset g^{-1}A$, invariance is equivalent to the apparently weaker condition that $gA \subset A$ for all $g \in G$. The next two theorems show that orbits and invariant events are closely related.

Theorem 6.6 *The set of invariant events is a subalgebra.*

Proof: Clearly Ω is invariant. If A and B are invariant, then $g(A \cup B) = (gA) \cup (gB) = A \cup B$ and so $A \cup B$ is also invariant. Finally, if A is invariant, then $gA^c = (gA)^c = A^c$, so its complement is also invariant. These conditions are sufficient to define an algebra.

QED

Theorem 6.7 *The atoms of the subalgebra of G -invariant events in an atomic algebra are precisely the G -orbits.*

Proof: Clearly every G -orbit is G -invariant. It thus suffices to prove that no non-empty invariant event can be properly contained in an orbit. Suppose that $\emptyset \neq B \neq Gx$ but $B \subset Gx$. Then either $x \in B$ or $x \in Gx \setminus B$. If $x \in B$ then there must be a $g \in G$ such that $gx \in Gx \setminus B$ and hence B is not invariant. Similarly, if $x \in Gx \setminus B$ there must be a $g \in G$ such that $gx \in B$, but $g^{-1}gx \notin B$, and again B cannot be invariant.

QED

Remark: Although we do not need this result, it can be shown that the set of invariant events under measurable permutations is a sub- σ -algebra even if the underlying σ -algebra

²Standard usage would be to call $\{gx : g \in G\}$ the orbit. Our definition is necessary because x is considered to be an event and not an element of some underlying sample space. Probabilistically, there is no advantage to the set-theoretic distinction between the element ω and the event $\{\omega\}$. Rigorous adherence to this distinction can cause notational nightmares when an object that is an event (i.e. a collection of ω 's) in one algebra becomes an object to be permuted in a subalgebra, thus taking on the role of a point in a sample space.

is not atomic. However, Theorem 6.7 fails in this case, for the orbits need not be atoms. The algebra whose atoms are G -orbits will be called the *invariant subalgebra*.

Given any subalgebra $\mathcal{B}$, we can define conditional probability in the usual way as $P(A|B) = P(A \cap B)/P(B)$ for $A \in \mathcal{A}$, $B \in \mathcal{B}$. For a given A , this expression defines a function on the atoms of $\mathcal{B}$. This function will be represented as $P(A|\mathcal{B})$. Note that if $A \in \mathcal{A}$ then $P(A|\mathcal{B}) = 0$ except on that $B \in \mathcal{B}$ for which $A \subset B$ where it equals $P(A)/P(B)$.

A group G of permutations of a probability algebra $\mathcal{A}$ is called *measure preserving* if $P(gA) = P(A)$ for every $g \in G$, $A \in \mathcal{A}$. The invariant subalgebra of a measure preserving group has a special structure.

Theorem 6.8 *Let $(\mathcal{A}, P)$ be a finite probability algebra, and let G be a measure-preserving group of permutations, with invariant subalgebra $\mathcal{B}$. Then for any $b \in \mathcal{B}$, $A \in \mathcal{A}$,*

$$P(A|B) = \frac{|F_A|}{|G|}.$$

Proof: By Theorem 6.7,

$$B = \bigcup_{g \in G} gA.$$

However, two permutations will map A into different atoms if and only if they belong to the same right coset of F_A . Thus the above union will contain each event $|F_A|$ times. Thus

$$P(B) = \frac{\sum_{g \in G} P(gA)}{|F_A|}.$$

But since the group is measure preserving $P(gA) = P(A)$ so that

$$P(B) = \frac{|G|}{|F_A|} P(A).$$

Thus

$$P(A|B) = \frac{P(A)}{P(B)} = \frac{|F_A|}{|G|}.$$

QED

Theorem 6.9 *The conditional probability given the invariant algebra under a measure preserving group of permutations is uniform on the orbits.*

By Lemma 6.1 $|F_{A_1}| = |F_{A_2}|$ if A_1 and A_2 are in the same orbit. Then by Theorem 6.8, every atom in the orbit will have the same conditional probability.

QED

Theorem 6.10 *Let G be a group of permutations of a finite probability algebra $(\mathcal{A}, P)$. Define for any $A \in \mathcal{A}$*

$$P_G(A) = \frac{1}{|G|} \sum_{g \in G} P(gA). \quad (6.15)$$

Then for $\mathcal{B}$ the invariant subalgebra, $P_G(A|\mathcal{B})$ does not depend P .

Proof: Since clearly G is measure preserving of (A, P_G) , the result follows immediately from Theorem 6.9. QED

Theorem 6.10 provides the mechanism for the construction of permutation tests. We suppose that the outcome y of an experiment is governed by some probability law P , but that instead of observing y , the experiment has been randomized by randomly choosing a permutation g from some group G , and then observing gy . The probability law of the observed y is then the P_G of Theorem 6.10. A hypothesis test not depending on P can be constructed by conditioning on the invariant subalgebra. To find the distribution of a test statistic for a particular randomized design, one then has only to determine the invariant subalgebra. The use of this technique for hypothesis testing depends on the assumption that under the null hypothesis the random permutation is stochastically independent of the algebra of observations, but that under an alternative hypothesis the choice of g will affect the outcome, so that (6.15) no longer defines the randomized distribution. The choice of test statistic of course would depend on the sort of dependence that is expected.

6.2.4 Lattices and products of algebras

The set of subgroups of a group G forms a lattice, where for subgroups H_1 and H_2 we write $H_1 \wedge H_2$ for the elements common to H_1 and H_2 , and $H_1 \vee H_2$ for the smallest subgroup containing both H_1 and H_2 . Similarly, the set of subalgebras of an algebra A forms a lattice, where for subalgebras B_1 and B_2 we write $B_1 \wedge B_2$ for the set of events common to both B_1 and B_2 , and $B_1 \vee B_2$ for the smallest subalgebra containing all the events of B_1 and B_2 . If A is an atomic algebra, then it is easy to show that the atoms of $B_1 \vee B_2$ are intersections of atoms of B_1 and atoms of B_2 . Although there is a correspondence between the lattice structure of subgroups and that of subalgebras, it is not perfect, as we see from the following two theorems.

Theorem 6.11 *Let G be a group of permutations of a finite algebra A . Let H_1 and H_2 be two subgroups of G , let $\mathcal{H}_i$ be the subalgebra of H_i -invariant events ($i = 1, 2$), and let $\mathcal{F}$ be the subalgebra of $(H_1 \vee H_2)$ -invariant events. Then $\mathcal{F} = \mathcal{H}_1 \wedge \mathcal{H}_2$.*

Proof: If $F \in \mathcal{F}$ then since $H_i \subset H_1 \vee H_2$ we have $H_i F = F$ for $i = 1, 2$ and therefore $F \in \mathcal{H}_1 \wedge \mathcal{H}_2$.

Conversely let $g \in H_1 \vee H_2$. Then g can be written as

$$g = \prod_{j=1}^n a_j b_j \quad (6.16)$$

for some $n \geq 0$ and $a_j \in H_1$, $b_j \in H_2$. If $F \in \mathcal{H}_1 \wedge \mathcal{H}_2$, then

$$\begin{aligned} gF &= \left(\prod_{j=1}^{n-1} a_j b_j \right) a_n b_n F \\ &= \left(\prod_{j=1}^{n-1} a_j b_j \right) a_n F \quad (\text{since } F \in \mathcal{H}_2) \end{aligned}$$

$$= \left(\prod_{j=1}^{n-1} a_j b_j \right) F \quad (\text{since } F \in \mathcal{H}_1).$$

It then follows, by induction, that $gF = F$ and hence $F \in \mathcal{F}$.

QED

Theorem 6.12 Let G , H_1 , H_2 , $\mathcal{H}_1$, and $\mathcal{H}_2$ be as in Theorem 6.11. Let $\mathcal{F}$ be the algebra of $(H_1 \wedge H_2)$ -invariant events. Then $\mathcal{H}_1 \vee \mathcal{H}_2 \subset \mathcal{F}$.

Proof: Let F be an atom of $\mathcal{H}_1 \vee \mathcal{H}_2$. Then $F = H_1 \cap H_2$ where $H_i \in \mathcal{H}_i$, ($i = 1, 2$). Now if $g \in H_1 \wedge H_2$ then

$$gF = g(H_1 \cap H_2) = gH_1 \cap gH_2 = H_1 \cap H_2 = F,$$

since $g \in H_1$ and $g \in H_2$. Thus $F \in \mathcal{F}$.

QED

The following counterexample shows that the conclusion of Theorem 6.12 cannot be strengthened. Let $X = \{1, 2, 3, 4\}$, and define

$$H_1 = \{(1), (123), (132)\},$$

$$H_2 = \{(1), (124), (142)\},$$

using the standard notation for cyclic permutations.³ Then

$$\mathcal{H}_1 = \{\emptyset, \{1, 2, 3\}, \{4\}, X\}$$

$$\text{and } \mathcal{H}_2 = \{\emptyset, \{1, 2, 4\}, \{3\}, X\}.$$

Now, since $H_1 \wedge H_2 = \{(1)\}$, every subset is $(H_1 \wedge H_2)$ -invariant, so that $\mathcal{F} = 2^X$. However, it can be seen that $\{1\} \notin \mathcal{H}_1 \vee \mathcal{H}_2$.

If G is a group of permutations of X and H is a group of permutations of Y , the direct product $G \times H$ of the groups defines in a natural way a group of permutations of the Cartesian product $X \times Y$ of the sets, by $(gh)(x, y) = (gx, hy)$. If $\mathcal{A}$ and $\mathcal{B}$ are subalgebras of 2^X and 2^Y respectively, then one can define the product algebra $\mathcal{A} \times \mathcal{B}$ as the algebra generated by the Cartesian products of the atoms of $\mathcal{A}$ and $\mathcal{B}$. It is evidently a subalgebra of $2^{X \times Y}$. It should be noted that the set of events in $\mathcal{A} \times \mathcal{B}$ is not the Cartesian product of the set of events in $\mathcal{A}$ and the set of events in $\mathcal{B}$. Thus the notation $\mathcal{A} \times \mathcal{B}$, though standard, is somewhat abusive. If P and Q are probability measures on the algebras $\mathcal{A}$ and $\mathcal{B}$, then the product measure $P \times Q$ on $\mathcal{A} \times \mathcal{B}$ is defined in the usual way by the requirement that $(P \times Q)(A \times B) = P(A)Q(B)$ for $A \in \mathcal{A}$, $B \in \mathcal{B}$.

Theorem 6.13 Let G and H be groups of permutations of X and Y respectively, with $\mathcal{G}$ and $\mathcal{H}$ the corresponding invariant subalgebras. Let $\mathcal{F}$ be the $(G \times H)$ -invariant subalgebra of $2^{X \times Y}$. Then $\mathcal{F} = \mathcal{G} \times \mathcal{H}$.

³The parenthesized list $(x_1 x_2 \dots x_p)$ means the permutation that maps $x_i \mapsto x_{i+1}$, ($i = 1, \dots, p-1$), $x_p \mapsto x_1$, and maps elements not in the list to themselves.

Proof: Let F be an atom of $\mathcal{G} \times \mathcal{H}$. Then $F = (Gx) \times (Hy)$ for some $(x, y) \in X \times Y$. Clearly $(G \times H)F = F$ so that $F \in \mathcal{F}$.

Now suppose that $\emptyset \neq F \in \mathcal{F}$ and let $(x, y) \in F$. Then for any $g \in G$, $h \in H$, $(gx, hy) \in F$, and so $(Gx) \times (Hy) \subset F$. However, $(Gx) \times (Hy) \propto \mathcal{G} \times \mathcal{H}$. Thus $\mathcal{F}$ contains all atoms of $\mathcal{G} \times \mathcal{H}$, and there are no non-empty events in $\mathcal{F}$ properly contained in these atoms. Thus the atoms of $\mathcal{G} \times \mathcal{H}$ are also atoms of $\mathcal{F}$ and hence $\mathcal{F} = \mathcal{G} \times \mathcal{H}$. QED

Theorem 6.14 *Let G be a measure preserving group of (A, P) , and H a measure preserving group of (B, Q) . Then $G \times H$ is a measure preserving group of $(A \times B, P \times Q)$.*

Proof: For any atoms $A \propto \mathcal{A}$, $B \propto \mathcal{B}$, and permutations $g \in G$, $h \in H$, we have $(P \times Q)(A \times B) = P(A)Q(B) = P(gA)Q(hB) = (P \times Q)((gA) \times (hB)) = (P \times Q)(gh(A \times B))$. QED

If X is a finite set with m elements, then clearly Y^X is equivalent to the m -fold Cartesian product $Y \times \cdots \times Y$, and the group H^X is isomorphic to the m -fold direct product $H \times \cdots \times H$. With this correspondence the following two theorems easily follow, by induction, from Theorems 6.13 and 6.14.

Theorem 6.15 *Let H be a group of permutations of the finite set Y with $\mathcal{H}$ being the invariant subalgebra of 2^Y . Let X be a finite set with m elements. Then the H^X -invariant subalgebra of 2^{Y^X} is isomorphic to the m -fold product $\mathcal{H} \times \cdots \times \mathcal{H}$.*

Theorem 6.16 *Let H be a measure preserving group of (A, P) , and let X be a finite set of m elements. Then H^X is measure preserving of $(\underbrace{A \times \cdots \times A}_{m \text{ factors}}, \underbrace{P \times \cdots \times P}_{m \text{ factors}})$.*

6.2.5 Reduction of the conditional distribution

The conditional distribution given the invariant subalgebra can often be expressed in a simpler form, by recognizing symmetries in the structure of the permutation group. In certain situations, a smaller conditioning algebra can be used. This can be an advantage if one is interested in unconditional operating characteristics, for then there will be fewer atoms in the conditioning algebra to consider. We will now present some theorems allowing such reduction, which are particularly relevant to the case of designs with litter structure.

In the following discussion we will be dealing with a group Z which is the semi-direct product of two subgroups W and B :

$$Z = W \odot B.$$

(Although the development is entirely abstract, it may help intuition to think of W as representing permutations of animals within a litter, and B as representing permutations of entire litters.) Any subgroup of permutations will define an invariant subalgebra. It will be useful to relate the structure of the group of permutations to the structure of the subalgebras of the probability algebra. We will use $\mathcal{W}$, $\mathcal{B}$, and $\mathcal{Z}$ to represent the invariant algebras under W , B , and Z , respectively.

Since $Z = W \vee B$, it follows from Theorem 6.11 that $Z \Rightarrow W \wedge B$. Our goal in the following theorems will be to derive conditions under which conditioning on either W or B can be replaced by conditioning on Z . In the process we will also exploit symmetries that will allow probability calculations to be restricted to either W or B . The following lemmas will be useful.

Lemma 6.2 Let $Z = W \odot B$, be a group of permutations of an algebra $\mathcal{A}$. Denote by Z , B , and W orbits under Z , B , and W . Then if $B \subset Z$ and $W \subset Z$ then $W \cap B \neq \emptyset$.

Proof: Let $A_1, A_2 \in \mathcal{A}$ such that $A_1 \subset B$ and $A_2 \subset W$. Then since $A_1 \in ZA$,

$$Z = \bigcup_{z \in Z} zA_1 = \bigcup_{w \in W} \bigcup_{b \in B} wbA_1. \quad (6.17)$$

But since $A_2 \in ZA$ (6.17) implies that $A_2 = wbA_2$ for some $w \in W, b \in B$. But then $w^{-1}A_2 = bA_1$ is an element of Z which is contained in both W and B . Hence $W \cap B \neq \emptyset$.

QED

Lemma 6.3 Given an atom $A \in \mathcal{A}$, let W, B , and Z be the orbits of A under W, B , and Z respectively. Then $Z = BW = WB$.

Proof: By definition,

$$\begin{aligned} Z &= \bigcup_{z \in Z} zA \\ &= \bigcup_{\substack{w \in W \\ b \in B}} wbA \\ &= \bigcup_{w \in W} w \bigcup_{b \in B} bA \\ &= \bigcup_{w \in W} wB \\ &= WB. \end{aligned}$$

Alternatively,

$$\begin{aligned} Z &= \bigcup_{\substack{w \in W \\ b \in B}} bb^{-1} * wA \\ &= \bigcup_{b \in B} \bigcup_{w \in W} b^{-1} * wA. \end{aligned}$$

But since $b^{-1} * w \in W$ this gives

$$\bigcup_{b \in B} bW = BW.$$

QED

Theorem 6.17 Let $Z = W \odot B$, be a group of permutations of a probability algebra (A, P) , where B is measure-preserving. Suppose that W and B are orbits under W and B , respectively, such that $\emptyset \neq A \subset W \cap B$. Denote by Z_0 and W_0 the subgroups of Z and B , respectively, fixing A . Then

$$P(W|B) = \frac{|Z_0|}{|W_0||B|}. \quad (6.18)$$

Proof: Let

$$H = \{w \in W : w^{-1}b \in Z_0 \text{ for some } b \in B\}.$$

Then $wA \subset W \cap B$ if and only if $w \in H$. Now clearly $W_0 \subset H$ and if $w_1 \in H$ then so is any element of the coset w_1W_0 . Thus

$$W \cap B = \bigcup_{w \in H} wA,$$

and each atom of $W \cap B$ occurs $|W_0|$ times in the union. Thus

$$P(W|B) = \frac{1}{|W_0|} \sum_{w \in H} P(wA|B) \quad (6.19)$$

$$= \frac{|H|}{|W_0|} P(A|B). \quad (6.20)$$

But since B is measure preserving, Theorem 6.8 implies that $P(A|B) = |B_0|/|B|$, where B_0 is the subgroup of B fixing A . Substituting into (6.19) gives

$$P(W|B) = \frac{|H||B_0|}{|W_0||B|}. \quad (6.21)$$

To complete the proof, we establish a one-to-one correspondence between $H \times B_0$ and Z_0 . For any $w \in H$, there exists a $b \in B$ such that $w^{-1}b \in Z_0$. Choose one such b . Then for any $b_0 \in B_0$, define the mapping $\pi: H \times B_0 \rightarrow Z_0$ by

$$\pi(w, b_0) = w^{-1}bb_0 \in Z_0. \quad (6.22)$$

This mapping is evidently one-to-one. To see that it is onto, let $z_0 = wb \in Z_0$. Then $w^{-1} = bz_0^{-1} \in H$. If b_1 is the chosen element of B for this w^{-1} , then $z_0 = wb_1b_1^{-1}b$. However, since $wbA = A = wb_1A$, we must have $b_1^{-1}bA = A$, implying that $b_1^{-1}b \in B_0$. Thus z_0 is in the range of the mapping (6.22).

Since π is a one-to-one correspondence between $H \times B_0$ and Z_0 , it follows that $|H||B_0| = |Z_0|$, and thus from (6.21)

$$P(W|B) = \frac{|Z_0|}{|W_0||B|}.$$

QED

Theorem 6.18 Let $Z = W \odot B$, be a group of permutations of a probability algebra (A, P) , where B is measure-preserving. Denote by Z , W , and B the invariant subalgebras under Z , W , and B , respectively. Then for any $W \in \mathcal{W}$,

$$P(W|B) = P(W|Z). \quad (6.23)$$

Proof: Since for $Z \propto Z, W \in \mathcal{W}$,

$$P(W|Z) = \sum_{B \propto Z \beta} P(W|B)P(B|Z),$$

the result will follow if we show that $P(W|B)$ is constant on the atoms of $Z\beta$. Moreover, by the summation rule of probabilities, (6.23) will hold for any $W \in \mathcal{W}$ if it holds for all atoms of $\mathcal{W}$.

Let $W \propto \mathcal{W}, B_0 \propto \beta$. If $W \cap B_0 = \emptyset$ then by Lemma 6.3 $B_0 \subset Z$ implies $W \cap Z = \emptyset$, so that $P(W|B_0) = 0 = P(W|Z)$ trivially. On the other hand, if $W \cap B_0 \neq \emptyset$ then by Theorem 6.17

$$P(W|B_0) = \frac{|Z_0|}{|W_0||B|}$$

where W_0 and Z_0 are the subgroups of W and Z , respectively, fixing some A -atom A_0 contained in $W \cap B_0$.

Now, let $B_1 \propto \beta$ be contained in the same Z -orbit as B_0 . Then since $W \cap B_1 \neq \emptyset$ it will contain an A -atom A_1 . Since W is a W -orbit, there is a permutation $w_1 \in W$ such that $A_1 = w_1 A_0$. Letting W_1 and Z_1 be the subgroups of W and Z , respectively, fixing A_1 , we again have from Theorem 6.17 that

$$P(W|B_1) = \frac{|Z_1|}{|W_1||B|}$$

But $z \in Z_1$ if and only if $zw_1 A_0 = z A_1 = A_1 = w_1 A_0$, that is, if $w_1^{-1} z w_1 \in Z_0$. Thus Z_1 is the image, under conjugation by w_1 , of Z_0 . Since conjugation is an isomorphism, it follows that $|Z_1| = |Z_0|$. Similarly, it can be shown that $|W_1| = |W_0|$. Thus

$$P(W|B_1) = \frac{|Z_0|}{|W_0||B|} = P(W|B_0).$$

QED

A group B will be said to be *conditionally measure preserving given $\mathcal{W}$* if $P(bA|bW) = P(A|W)$ for all $b \in B, A \in \mathcal{A}, W \propto \mathcal{W}$. The following theorem assures that the definition makes sense.

Theorem 6.19 *If $\mathcal{W}$ is the invariant algebra under a normal subgroup W of a group of permutations Z , then $zW \propto \mathcal{W}$ for every $z \in Z, W \propto \mathcal{W}$.*

Proof: By Theorem 6.7 any $W \propto \mathcal{W}$ can be written as $W = WA$ for some $A \propto \mathcal{A}$. Now $zW = z(WA) = (zW)A$. However, since W is a normal subgroup, $zW = Wz$. Thus $zW = (Wz)A = W(zA) \propto \mathcal{W}$.

QED

Theorem 6.20 *Let B be conditionally measure preserving of $(\mathcal{A}, P)$ given $\mathcal{W}$, and let two A -atoms A_1, A_2 be both contained in the same B -atom B . Then for any $W \propto \mathcal{W}$, $P(A_1|W) = P(A_2|W)$ unless $P(A_1|W)P(A_2|W) = 0$.*

Proof: Since A_1 and A_2 are contained in the same B -orbit, $P(A_2|W) = P(bA_1|W)$ for some $b \in B$. However, $P(A_1|W) = P(bA_1|bW)$ for all $b \in B$. By Theorem 6.19, either $bW = W$ or $bW \cap W = \emptyset$. In the former case we have $P(A_1|W) = P(bA_1|W) = P(A_2|W)$ as required. In the latter case, suppose that $P(A_1|W) > 0$, implying that $A_1 \cap W \neq \emptyset$. Then $P(A_2|W) = P(bA_1|W) = 0$ since $bA_1 \cap W = \emptyset$. **QED**

Theorem 6.21 Let B be conditionally measure preserving of (A, P) given $\mathcal{W}$. Then for any $B \propto \mathcal{B}$, $W \propto \mathcal{W}$

$$P(B|W) = \frac{|F_W|}{|F_A|} P(A|W), \quad (6.24)$$

where A is any A -atom contained in B .

Proof:

$$P(B|W) = \bigcup_{A \propto BA} P(A|W) = \bigcup_{b \in B} P(bA|W)$$

for some $A \propto BA$. But for $b_1, b_2 \in B$, $b_1A = b_2A$ if and only if $b_1F_A = b_2F_A$. Thus

$$\bigcup_{b \in B} P(bA|W) = \frac{1}{|F_A|} \sum_{b \in B} P(bA|W).$$

However, $P(bA|W) = P(bA|bW) = P(A|W)$ if $b \in F_W$ and $P(bA|W) = P(bA|bW) = 0$ if $b \notin F_W$. Thus

$$\sum_{b \in B} P(bA|W) = |F_W| P(A|W).$$

QED

Theorem 6.22 If B is conditionally measure preserving of (A, P) given $\mathcal{W}$ then for any B -invariant event B , $P(B|W) = P(B|bW)$ for all $b \in B$, $W \propto \mathcal{W}$.

Proof: If $B \propto \mathcal{B}$ then from (6.24)

$$P(B|W) = \frac{|F_W|}{|F_A|} P(A|W),$$

for some $A \propto BA$. Now, since $bB = B$,

$$P(B|bW) = P(bB|bW) = \frac{|F_{bW}|}{|F_{bA}|} P(bA|bW).$$

But $P(bA|bW) = P(A|W)$ by hypothesis, and $|F_{bW}| = |F_W|$, $|F_{bA}| = |F_A|$ by Lemma 6.1. Thus $P(B|W) = P(B|bW)$. If B is not an atom, then the result follows from the summation rule of probabilities. **QED**

Theorem 6.23 If B is conditionally measure preserving of (A, P) given $\mathcal{W}$, then for any B -invariant event B , $P(B|W) = P(B|Z)$.

Proof: For any $Z \propto Z$,

$$P(B|Z) = \sum_{W \propto ZW} P(B|W)P(W|Z).$$

Now, since Z is an orbit of Z , Lemma 6.3 and Theorem 6.19 imply that any $W \propto ZW$ is the image, under some $b \in B$, of a fixed $W_0 \propto W$. Hence $P(B|W) = P(B|bW_0) = P(B|W_0)$ by Theorem 6.22, and

$$P(B|Z) = P(B|W_0) \sum_{W \propto ZW} P(W|Z) = P(B|W_0).$$

QED

6.3 The Permutation Distribution under Three Designs

We will now apply the theory of the previous section to derive the permutation distributions of an experiment with m litters, each of size l , under the three randomization schemes discussed in Chapter 3. We again let $n = ml$ be the total number of animals of which n_1 are to be placed on the test treatment. For testing the null hypothesis we will use as the test statistic T_1 , defined as the total number of responders in the test treatment. Since the total number of responders in the experiment is invariant under the permutation groups of any of the designs considered, the statistic T_1 is equivalent, given the invariant subalgebra, to the statistic T presented in Chapter 3. Thus the present discussion is the permutation test analogue of the asymptotic procedures discussed in that chapter.

We discuss now procedures for evaluating the permutation distribution of T_1 under the three design schemes. A simple closed formula for this distribution does not seem to exist in general, so we will concentrate on efficient computational procedures for calculating tail probabilities and quantiles of this distribution, using the methodology of Chapter 5.

To construct the permutation distribution, we must determine the invariant subalgebra under the randomizations. To implement this strategy, however, we must first establish the link between the design variables a_{ij} and a permutation group.

It is helpful to consider the n experimental units as being arranged in an $m \times l$ array, with rows representing litters, and columns animals within a litter. A number n_1 of the positions are designated as receiving the test treatment, the remainder are to receive the control. We then randomly choose a permutation g from a group G , and rearrange the units according to that permutation before imposing the treatments. For the nested design, the designated units must constitute entire litters, the first n_1/l rows would do. Similarly, for the litter matched design, we designate the first n_1/m columns. Either designation would work for the completely randomized design.

The group of permutations depends on the design scheme. For the completely randomized design it would be S_n , the group of all permutations of the animals. For the nested design it would be $B = S_m \times 1_l$ which permutes litters, and for the litter matched design $W = S_l^m$ which independently permutes animals within a litter. In the latter two cases the groups can in fact be enlarged to $W \odot B$ using the theorems of Section 6.2.5.

A convenient way of specifying a subalgebra is by a random variable generating it. In this way, the atoms of the subalgebra can be indexed by the values of a random variable that is constant on the atoms and takes distinct values on distinct atoms. For such a random variable X we will use the corresponding lower case letter x as a generic symbol for an atom of the algebra generated by that variable. Thus we will happily write $P(y|x)$ instead of the more complete $P\{Y = y|X = x\}$, or the pedantic $P(Y^{-1}(y)|X^{-1}(x))$. The essential aspect is not the random variable actually used, but the subalgebra so defined.

To define the invariant subalgebra under a particular randomization scheme, it thus suffices to find a random variable that indexes the orbits under the permutation group. For a random variable to have this property the following conditions are obviously necessary and sufficient.

1. The random variable is invariant under all permutations in the group, and
2. If the random variable has the same value for two experimental outcomes, then there is a permutation in the group that transforms one outcome to the other.

6.3.1 Completely randomized design

For the completely randomized design, the invariant subalgebra is generated by Z , the total number of responders. It is obvious that Z satisfies the first condition, and it is easy to construct (though somewhat tedious to spell out the details) a permutation that transforms one experiment with a given number of responders to another. The conditional probability of any outcome $r = (r_0, \dots, r_n)$ given this subalgebra is given by Theorem 6.8:

$$P(r|Z) = \frac{Z!(n-Z)!}{n!} \quad (6.25)$$

since the permutations rearranging the Z responders and $n-Z$ nonresponders leave the outcome r fixed, and $|S_n| = n!$. To find the conditional probability of $\{T_1 = t\}$ given $\{Z = z\}$, we then add up the probabilities (6.25) of all the atoms contained in $\{T_1 = t\} \cap \{Z = z\}$. By standard combinatorial arguments there are evidently $\binom{n_1}{t} \binom{n_0}{z-t}$ such atoms, and by Theorem 6.9 they all have the same probability, giving us

$$\Pr\{T_1 = t|Z = z\} = \frac{\binom{n_0}{z-t} \binom{n_1}{t}}{\binom{n}{z}}, \quad (6.26)$$

the hypergeometric distribution. The Fisher-Irwin test is thereby vindicated.

It may appear that we have gone through a rather elaborate mathematical maze just to arrive at a simple and well-known result. However, the preceding discussion illustrates how the general theory of Section 6.2 applies in this case.

6.3.2 Nested design

We continue our discussion with the derivation of the permutation distribution for the nested design. The permutation group for this design permutes entire litters. The outcome of each litter can be represented as a binary vector of length l , and thus the atoms of the probability algebra can be represented as elements of $(2^l)^m$. By arguments similar to those for the completely randomized design one can show that the invariant subalgebra is generated by a (random) integer-valued function b with domain

$$2^l = \{(r_1, \dots, r_l) : r_i \in \{0, 1\}, i = 1, \dots, l\}$$

where $b(r_1, \dots, r_l)$ gives the number of litters in which the pattern of responses is $(r_1, \dots, r_l)$.

We can actually condition on a smaller algebra if we are willing to assume that the litters are exchangeable (such is the case under the beta binomial model of Section 6.24), for in that case the permutation group is measure-preserving. Note that since all litter-mates must be on the same treatment, the statistic T_1 is invariant under any within-litter permutations in W . Theorem 6.18 then tells us that for probability statements about T_1 we can just as well condition on the orbits of $W \odot B$. These orbits are indexed by integer valued vectors $(z_0, \dots, z_l)$ where z_i is the number of litters having i responders ($i = 0, \dots, l$).

Again using Theorem 6.8, the conditional probability for any outcome given the algebra invariant under $W \odot B$ can be derived, using combinatorial arguments, to be

$$P(r|z_0, \dots, z_l) = \frac{\prod_{i=0}^l (i!(l-i)!)^{z_i}}{m!(l!)^m} \quad (6.27)$$

In this case, however, there does not appear to be a simple procedure for counting all the atoms in the orbit that have the same value of T_1 . Some reduction is possible, however, by noting invariance properties of T_1 . Clearly, rearrangements of responses within a litter do not affect its value, nor do exchanges between litters in the same treatment.

Let y_i represent the number of litters under the test treatment having i responses ($i = 0, \dots, l$), and let $m_1 = \sum_i y_i$ be the number of test litters. Then $T_1 = \sum_i i y_i$, so that outcomes with the same list of y_i 's will have the same value of T_1 . (However, different y_i 's might not have different T_1 's.) For a particular outcome with a given value of $(z_0, \dots, z_l)$, there are $\prod_{i=0}^l \binom{l}{i}^{z_i}$ distinct within-litter rearrangements giving the same value of $(y_0, \dots, y_l)$, and there are similarly

$$\frac{m_1!(m-m_1)!}{\prod_{i=0}^l y_i!(z_i - y_i)!}$$

rearrangements of the litters themselves. There are thus a total of

$$m_1!(m-m_1)! \prod_{i=0}^l \frac{\binom{l}{i}^{z_i}}{y_i!(z_i - y_i)!} \quad (6.28)$$

outcomes in the orbit having the same value of $(y_0, \dots, y_l)$, and (by Theorem 6.9) the same probability (6.27). Multiplying (6.27) by (6.28) then gives

$$P(y_0, \dots, y_l | z_0, \dots, z_l) = \frac{\prod_{i=0}^l \binom{z_i}{y_i}}{\binom{m}{m_1}}, \quad (6.29)$$

a multivariate hypergeometric distribution.

Thus the atoms generated by the list of $(y_0, \dots, y_l)$, conditional on $(z_0, \dots, z_l)$ can be represented as configurations in the sense of Chapter 5, with the z_i 's being the capacities. The calculation of $\Pr\{T_1 = t | z_0, \dots, z_l\}$ then amounts to weighing the lateral tree as discussed in Section 5.4.2. If the only interest in the permutation distribution is the calculation of tail probabilities, then a simpler computation is possible, since the tail probability will correspond to the weights of branches. In this case, a procedure such as described in Section 5.3 could be used.

A hypothesis test based on the distribution (6.29) was proposed by Soms[74] who gave an algorithm for its implementation, essentially involving exhaustive enumeration. A network algorithm for such computations was proposed by Mehta et al.[62]. Although we have not attempted a direct comparison, we feel that algorithms based on tree structures could work faster because of the possibility of truncating negligible tails.

6.3.3 Litter matched design

In the litter matched design, the permutation group independently permutes animals within each litter. In a sense, each litter is like a mini-experiment with a completely randomized design. Thus we see that the invariant algebra is generated by a list $(w_1, \dots, w_m)$ specifying the number of responders in each litter. Here again we can reduce the conditioning algebra by the techniques of Section 6.2.5. Note that for this design, interchange of litters does not change the value of T_1 . Moreover, the permutation distribution, conditional on W -orbits, is the same in each litter, and thus B is conditionally measure-preserving. Appealing to Theorem 6.23 then allows to condition on $W \odot B$. Thus the conditional distribution of outcomes given the invariant subalgebra is again given by (6.27). The distribution of the test statistic T_1 differs, however. In the litter matched case T_1 is invariant under B , but not under W . The value of T_1 only depends on $(y_0, \dots, y_\nu)$, where y_i is the number of litters with i responders in the test portion, ($i = 0, \dots, \nu$), but the set of outcomes corresponding to such a list and consistent with a given condition $(z_0, \dots, z_l)$ depends also on the number of responders in the control portion. Denote by u_{ij} the number of litters having i responders in the test portion and j responders in total. Obviously $u_{ij} = 0$ if $i > j$ or if $i < \nu - l + j$, so we consider the domain of the u_{ij} 's to be

$$0 \vee \nu - l + j \leq i \leq j \wedge \nu.$$

We then have that $z_j = \sum_i u_{ij}$, and $y_i = \sum_j u_{ij}$.

To find the number of distinct outcomes having the same value of $(u_{00}, u_{01}, \dots, u_{\nu\nu})$ we notice that any litter having i test responders and j total responders can have $\binom{l-y}{i} \binom{\nu}{j}$

distinct arrangements, thus giving a total of

$$\prod_{j=0}^l \prod_{i=0}^{\nu \wedge j} \left[\binom{l-\nu}{j-i} \binom{\nu}{i} \right]^{u_{ij}} \quad (6.30)$$

within-litter rearrangements. For each of these, there are $m! / \prod_{ij} u_{ij}!$ distinct arrangements of entire litters giving the same u_{ij} values. Multiplying this by (6.30) and (6.27) then gives

$$\begin{aligned} & \Pr \{u_{00}, u_{01}, \dots, u_{\nu\nu} | z_0, \dots, z_l\} \\ &= \prod_{j=0}^l \frac{z_j!}{\binom{l}{j}^{x_j}} \prod_{i=0}^{\nu \wedge j} \frac{\left[\binom{l-\nu}{j-i} \binom{\nu}{i} \right]^{u_{ij}}}{u_{ij}!} \end{aligned} \quad (6.31)$$

This distribution is actually a product multinomial as can be seen by writing

$$\lambda_{ij} = \frac{\binom{l-\nu}{j-i} \binom{\nu}{i}}{\binom{l}{j}}$$

It is obvious that $\sum_i \lambda_{ij} = 1$ and substituting λ_{ij} into (6.31) gives

$$\Pr \{u_{00}, u_{01}, \dots, u_{\nu\nu} | z_0, \dots, z_l\} = \prod_{j=0}^l z_j! \prod_{i=0}^{\nu \wedge j} \frac{\lambda_{ij}^{u_{ij}}}{u_{ij}!} \quad (6.32)$$

The atoms generated by the u_{ij} 's do not form a single configuration, but in fact a list of configurations:

$$\begin{aligned} & (u_{00}), \\ & (u_{01}, \quad u_{11}), \\ & \vdots \\ & (u_{0j}, \quad \dots, \quad \dots, \quad u_{jj}), \\ & \vdots \\ & (u_{\nu-l+j', j'}, \quad \dots, \quad \dots, \quad u_{\nu j'}), \\ & \vdots \\ & (u_{\nu-1, l-1}, \quad u_{\nu, l-1}), \\ & (u_{\nu l}). \end{aligned} \quad (6.33)$$

Each of the configurations in the list can be considered to be the node of a tree, and the list of configurations thus becomes the node of a product tree. Since the first and last of these configurations have but a single cell (and hence form a degenerate tree), there are effectively only $l-1$ trees involved in the product. The distribution on each component

tree is multinomial, and the trees are stochastically independent. Thus the techniques of Section 5.5 can be used for the calculation of tail probabilities.

A hypothesis test involving the permutation distribution of T_1 under within-litter permutations was proposed by Gart[30]. Our development is a special case of this in which the litters are all of the same size. This specialization allows the reduction of the conditioning subalgebra. A network algorithm by Mehta et al.[63], which efficiently calculates combinatorial coefficients required for confidence intervals on a common odds ratio, exploits ideas similar to our product tree concept, although using a different graph-theoretic construction. Their method, intended for more general situations, does not exploit the symmetry of exchangeable litters.

6.4 A Model For Correlated Binary Data

Although the permutation test is *valid* in the sense that its size is bounded above by the nominal significance level irrespective of the probability distribution governing the responses, the operating characteristics of the procedure do nevertheless depend on this probability law. The permutation distribution only holds under the null hypothesis, and the distribution under alternatives will in general depend on the unrandomized probability. Moreover, when dealing with discrete data, it is in general not possible to have critical regions whose size exactly equals the desired significance level unless randomized decision rules are admitted. Since these are, however, disdained by many practitioners we will not consider them here, but will consider significance tests that are *conservative* in the sense that the actual size will be strictly less than the nominal significance level. Although the actual size conditional on the realized invariant event does not depend on the underlying probability law, the probability of the invariant event does, so that the *unconditional* size of the critical region will no longer be distribution free. In this section we propose a model for correlated binary variables appropriate to a bioassay with litter structure. This model can then be used as the basis for computing the exact operating characteristics of the permutation tests.

6.4.1 The beta binomial model

The beta binomial model has been used in many contexts to model correlated binary responses. Although it is actually a special case of the Pólya family, we discuss here its derivation under a different paradigm which is more intuitive in a bioassay context. This is given by the following theorem.

Theorem 6.24 *Let U be random variable with a beta distribution:*

$$\Pr\{U \leq u\} = \frac{1}{B(\alpha, \beta)} \int_0^u t^{\alpha-1} (1-t)^{\beta-1} dt$$

and let $S_1, \dots, S_n$ be binary random variables that are conditionally independent given U , with $E(S_i | U) = U$. Then the unconditional distribution of $X = \sum_{i=1}^n S_i$ is Pólya with parameters $(-\alpha, -\beta, n)$.

Proof: Since the S_i are conditionally independent, the conditional distribution of X is binomial. Thus

$$\begin{aligned}\Pr\{X = x\} &= E(\Pr\{X = x|U\}) \\ &= \binom{n}{x} E(U^x(1-U)^{n-x}) \\ &= \binom{n}{x} \frac{B(\alpha + x, \beta + n - x)}{B(\alpha, \beta)}.\end{aligned}\quad (6.34)$$

However, $B(\alpha, \beta) = \Gamma(\alpha)\Gamma(\beta)/\Gamma(\alpha + \beta)$, and for x an integer we have the recursive relation of the gamma function

$$\begin{aligned}\Gamma(\alpha + x) &= \Gamma(\alpha) \prod_{i=0}^{x-1} (\alpha + i), \\ &= (-1)^x x! \Gamma(\alpha) \binom{-\alpha}{x},\end{aligned}\quad (6.35)$$

and similarly for $\Gamma(\beta + n - x)$ and $\Gamma(\alpha + \beta + n)$. Substituting these relations into (6.34) then gives

$$\Pr\{X = x\} = \frac{\binom{-\alpha}{x} \binom{-\beta}{n-x}}{\binom{-\alpha - \beta}{n}} \quad (6.36)$$

which agrees with (4.1).

QED

This distribution can be given the following interpretation. The responses of animals in a litter represent independent identically distributed random variables. However, the probability of response may vary from litter to litter (because of maternal effects), and the litters in the experiment represent a random sample from a population of litters in which the response probability follows a beta distribution. The binomial case ($\alpha, \beta \rightarrow \infty$) arises as a limiting case in which the beta distribution degenerates to a point mass at $\mu = \alpha/(\alpha + \beta)$.

6.4.2 Independent competing risks

To model treatment effects, we consider the same setup as in Chapter 3, except that we specify the complete distribution instead of just the first two moments. Again, the derivation is simpler if we model non-response rather than response. Let $U_1, \dots, U_m$ and $V_1, \dots, V_m$ be independent samples from beta distributions. The U_i 's have parameters $\alpha_S = \mu_S(1 - \rho_S)/\rho_S$ and $\beta_S = (1 - \mu_S)(1 - \rho_S)/\rho_S$, and the V_i 's have parameters $\alpha_I = \mu_I(1 - \rho_I)/\rho_I$ and $\beta_I = (1 - \mu_I)(1 - \rho_I)/\rho_I$. Then with S_{ij} , Q_{ij} and R_{ij} defined, respectively, as X_{ij} , Z_{ij} , and Y_{ij} in Section 3.2.1 we specify

$$E(S_{ij} | U_1, \dots, U_m, V_1, \dots, V_m) = U_j, \quad (6.37)$$

$$E(Q_{ij} | U_1, \dots, U_m, V_1, \dots, V_m) = V_j, \quad (6.38)$$

with the S_{ij} 's and Q_{ij} 's being conditionally independent given the U_i 's and V_i 's. The unconditional joint distributions of the S_{ij} 's and Q_{ij} 's will have the same first two moments, (3.2), as those of the X_{ij} 's and Z_{ij} 's in Chapter 3. Moreover, the litter sums $\sum_{i=1}^l S_{ij}$ and $\sum_{i=1}^l Q_{ij}$ ($j = 1, \dots, m$) are independent random samples from beta binomial distributions. The actual response variable, however, is $R_{ij} = S_{ij} + a_{ij}S_{ij}(Q_{ij} - 1)$ where a_{ij} , as before, is a binary design variable specifying whether animal i in litter j is to receive the test treatment. We will derive the conditional joint distribution of the R_{ij} given a particular design a . Note that the S_{ij} 's and Q_{ij} 's are independent between litters, and exchangeable within them. Thus the conditional distribution of the R_{ij} 's given a particular design will also be (except for permutation of subscripts) the unconditional distribution under a randomized design, provided that the design variables also have an analogous exchangeability property. This is the case for litter matched and nested designs, but not for the completely randomized design. The unconditional distribution for the completely randomized design does not appear to be tractable by algebraic means. Algorithms for computing it are discussed in Section 6.5.1. Since given a particular design a , the R_{ij} 's in different litters are independent, it suffices to derive the joint distribution for one litter. This is provided by the following theorem. Because of exchangeability, the joint distribution only depends on the $\nu_j = \sum_i a_{ij}$, the number of animals in the litter on the test treatment.

Theorem 6.25 *Let U and V be independent random variables with beta distributions having parameters (α_S, β_S) and (α_I, β_I) respectively. Let S_i , $i = 1, \dots, l$ and Q_i , $i = 1, \dots, \nu$ be binary random variables with $E(S_i | U, V) = U$ and $E(Q_i | U, V) = V$, with the S_i 's and Q_i 's being conditionally mutually independent given (U, V) . Then the joint distribution of*

$$X = \sum_{i=\nu+1}^l S_i \quad (6.39)$$

$$\text{and } Y = \sum_{i=1}^{\nu} S_i Q_i \quad (6.40)$$

is given by

$$\begin{aligned} \Pr\{X = x, Y = y\} &= \binom{l-\nu}{x} \sum_{h=y}^{\nu} \binom{\nu}{h} \binom{h}{y} \\ &\times \frac{B(\alpha_S + h + x, \beta_S + l - h - x)}{B(\alpha_S, \beta_S)} \\ &\times \frac{B(\alpha_I + y, \beta_I + h - y)}{B(\alpha_I, \beta_I)}. \end{aligned} \quad (6.41)$$

Proof: Let $H = \sum_{i=1}^{\nu} S_i$. Conditioning on the value of H , we see that the event $\{Y = y\}$ will occur if and only if $Q_i = 1$ for exactly y of the H indices i for which $S_i = 1$. Because the Q_i 's are independent of H , and independent of each other given (U, V) , we have

$$\Pr\{Y = y | H, U, V\} = \binom{H}{y} V^y (1 - V)^{H-y}, \quad (6.42)$$

and taking the conditional expectation given (U, V) gives

$$\Pr\{Y = y|U, V\} = \sum_{h=y}^{\nu} \binom{h}{y} V^y (1-V)^{h-y} \binom{\nu}{h} U^h (1-U)^{\nu-h}, \quad (6.43)$$

because H is (conditionally) binomial. Because X is also conditionally binomial and independent of Y we get

$$\begin{aligned} \Pr\{X = x, Y = y|U, V\} &= \binom{l-\nu}{x} \sum_{h=y}^{\nu} \binom{h}{y} V^y (1-V)^{h-y} \\ &\quad \times \binom{\nu}{h} U^{h+x} (1-U)^{l-h-x}. \end{aligned} \quad (6.44)$$

Taking the expectation of (6.44) over (U, V) then gives (6.41). QED

For some applications, it will be useful to have the conditional distribution of Y given X . Since, X is beta-binomial with parameters $(\alpha_S, \beta_S, l-\nu)$, this conditional distribution can be obtained by dividing (6.41) by (6.34) giving

$$\begin{aligned} \Pr\{Y = y|X = x\} &= \sum_{h=y}^{\nu} \binom{\nu}{h} \binom{h}{y} \frac{B(\alpha_S + x + h, \beta_S + l - x - h) B(\alpha_I + y, \beta_I + h - y)}{B(\alpha_S + x, \beta_S + l - \nu - x) B(\alpha_I, \beta_I)}. \end{aligned} \quad (6.45)$$

It should be noted that Y does not have a beta binomial distribution, neither conditionally nor marginally. The conditional distribution of Y is a mixture of beta binomials, with the mixing distribution itself being beta binomial. As such, the distribution can be expressed in terms of μ_S, ρ_S, μ_I , and ρ_I using the relationship (4.11). The coefficient

$$\binom{\nu}{h} \binom{h}{y} = \frac{\nu!}{y! (h-y)! (\nu-h)!}$$

is the number of ways in which the ν test animals can be partitioned into three classes: $\nu - h$ that respond spontaneously, $h - y$ that have induced but no spontaneous responses, and y that respond in neither fashion.

Although the beta-binomial derivation of Theorem 6.24 implies that $0 < \rho < 1$, the formulas can equally well be extended to the complete Pólya family. We can thus also consider cases of negative intra-litter correlations.

6.5 Calculation of Power

To find the power (or for that matter the exact, unconditional size) of the permutation tests, we must find the unconditional probability of the rejection region. Although simple in principle, a naïve approach to this exercise could involve an astronomical amount of calculations unless both the litter size and number of litters are very small. With the completely randomized design the difficulty lies with the unconditional distribution, which

must be determined by averaging (6.41) over all possible randomizations. With the other two designs, the unconditional distribution is no problem. However, the rejection region is not defined by a simple formula, but must be determined dynamically by searching through the trees. An algorithm not based on tree structures, but using principles similar to those of Soms[74], has been used by Krewski et al.[43] to calculate power for some special cases of the nested design. These computations involved a considerable amount of computer time. In the following sections we present systematic procedures whereby such computations can be performed in an efficient way.

6.5.1 Completely randomized design

The distribution (6.41) depends on only three statistics from each litter: The number of test animals ν_j , the total number of responders w_j , and the number of test responders y_j . These three numbers define a random partition of the l litter mates of the j^{th} litter into the four classes representing $l - \nu_j - w_j + y_j$ control nonresponders, $w_j - y_j$ control responders, $\nu_j - y_j$ test nonresponders, and y_j test responders. Each litter can thus be in one of $\binom{l+3}{3}$ states, and the outcome of the experiment can be summarized by the distribution of the m litters among these states. All outcomes having the same distribution of litters have the same probability, and the rejection region is measurable with respect to the subalgebra generated by this distribution. Thus this subalgebra is sufficiently fine for calculating the probability of the rejection region. For future reference, denote this subalgebra as $\mathcal{A}$.

Let us partition the $\binom{l+3}{3}$ states into $l+1$ classes, where all the states in a class have the same value of ν , the number of test animals in the litter. Given that a litter has ν test animals, the probabilities of its falling into one of the $(\nu+1)(l-\nu+1)$ states in that class are given by (6.41), and by hypothesis, the litters are independent given the design. Thus given D_ν , the number of litters in the ν^{th} class, the conditional distribution within the class is multinomial, with cell probabilities given by (6.41). The conditional distribution of all the litters among the $\binom{l+3}{3}$ states, given the distribution $(D_0, \dots, D_l)$ of litters among the classes, is then product multinomial.

To complete the calculation of the unconditional probabilities we thus need the probability distribution of $(D_0, \dots, D_l)$. This is given by the following lemma.

Lemma 6.4 *Let n_1 units be selected by simple random sampling from a total of m strata of l units each. Let D_ν be the number of strata in which ν units have been selected ($\nu = 0, \dots, l$). Then*

$$\Pr\{D_\nu = d_\nu, \nu = 0, \dots, l\} = \frac{m!}{\binom{n}{n_1}} \prod_{\nu=0}^l \frac{\binom{l}{\nu}^{d_\nu}}{d_\nu!}. \quad (6.46)$$

Proof: The number of samples in which ν_j units are selected from the j^{th} stratum is

$$\prod_{j=1}^m \binom{l}{\nu_j} = \prod_{\nu=0}^l \binom{l}{\nu}^{d_\nu}. \quad (6.47)$$

Since there are $\binom{n}{n_1}$ equally likely samples, the probability of such a sample is

$$\prod_{\nu=0}^l \binom{l}{\nu}^{d_\nu} / \binom{n}{n_1}. \quad (6.48)$$

However, every partition of the m strata into $l + 1$ classes in which the ν^{th} class ($\nu = 0, \dots, l$) has d_ν strata will give a different sample with the probability (6.48). The number of such partitions is $m! / \prod_{\nu=0}^l d_\nu!$. Multiplying this quantity by (6.48) then gives (6.46).

QED

Note that because $\sum_\nu d_\nu = m$, the list $(d_0, \dots, d_l)$ can be considered to be a configuration in the sense of Chapter 5 with the additional constraint the $\sum \nu d_\nu = n_1$. The set of such configurations can be given a tree structure using the lateral parental mapping of Section 5.4.2. Let us call the tree whose nodes are the $(d_0, \dots, d_l)$'s the *allocation tree*. The tree structure in this case is particularly simple because the cell contents are not constrained. Although the distribution defined by (6.46) is not in the multivariate Pólya family, the same recursive computations can in fact be used. By defining

$$\theta_\nu = 2^{-l} \binom{l}{\nu} \quad (6.49)$$

$$\text{and } K = 2^{-n} \binom{n}{n_1} \quad (6.50)$$

we can rewrite (6.46) as

$$\Pr \{D_\nu = d_\nu, \nu = 0, \dots, l\} = \frac{m!}{K} \prod_{\nu=0}^l \frac{\theta_\nu^{d_\nu}}{d_\nu!} \quad (6.51)$$

which is, except for the proportionality constant K , a multinomial distribution.

(It is curious that this distribution also arises as a conditional distribution from an entirely different sampling scheme: Let each unit be selected independently with probability $\frac{1}{2}$. Then the conditional distribution of the D_ν 's given that a total of n_1 units are selected is precisely given by (6.51).)

The computation of the power can now be described in terms of a filtration. Let $\mathcal{A}_1$ be the algebra generated by $(D_0, \dots, D_l)$. Given $\mathcal{A}_1$, the conditional probabilities of the atoms of $\mathcal{A}$ are defined by product multinomial probabilities. The basic idea of the power calculation is to accumulate the conditional probabilities of all the atoms in the rejection region, multiply this sum by the probability of the containing $\mathcal{A}_1$ atom, and then accumulate these unconditional probabilities. Such a straightforward approach, however, would still require impractically many calculations for even moderately sized experiments.

Fortunately several shortcuts are possible. It may be possible to identify atoms of $\mathcal{A}_1$ that do not intersect the rejection region. Such atoms can be omitted from the calculations. Also, if the probability of an $\mathcal{A}_1$ -atom is less than some acceptable truncation error ϵ , then there is no need to identify the rejection region within that atom, nor to compute its conditional probability, since it will not contribute appreciably to the final result.

The conditioning argument could now be repeated. Let X be the algebra generated by the number of control responders, and define $\mathcal{F} = \mathcal{A}_1 \vee X$. Within each $x \in X$ the rejection region is defined by the critical value $\kappa(x)$, the minimum number of test responders required for rejection. Unfortunately, there do not seem to be simple formulas for computing either the conditional probabilities of the rejection region given $\mathcal{F}$, or of the $\mathcal{F}$ -atoms given $\mathcal{A}_1$. We must therefore go to a finer algebra.

Let v_ν represent the number of control responders among all the litters having ν animals on the test treatment, and denote by $\mathcal{V}$ the subalgebra generated by the pair of configurations

$$\begin{pmatrix} d_0, d_1, \dots, d_l \\ v_0, v_1, \dots, v_l \end{pmatrix}$$

Clearly $v_\nu \leq (l-\nu)d_\nu$, $X \subset \mathcal{V}$, and $\sum_\nu v_\nu$ is constant on X -atoms. The v_i are conditionally independent given $\mathcal{A}_1$, with the conditional distribution of v_ν being the convolution of d_ν beta binomials with parameters $(\beta_S, \alpha_S, l-\nu)$ (recall that we were modelling *non-response* in Section 6.4). This distribution again does not appear to have a simple formula, and thus $\mathcal{V}$ is still not fine enough to calculate the required probability.

Each v_ν represents the aggregate number of control responders among d_ν litters of $l-\nu$ control animals each, disregarding how they distribute among the litters. Denote by $\xi_{\nu j}$ the number of litters having ν test animals and j control responders ($j = 0, \dots, l-\nu$). Clearly

$$\sum_{j=0}^{l-\nu} \xi_{\nu j} = d_\nu, \quad (6.52)$$

$$\text{and } \sum_{j=0}^{l-\nu} j \xi_{\nu j} = v_\nu. \quad (6.53)$$

Let $\mathcal{C}_\nu$ represent the subalgebra generated by $(\xi_{\nu 0}, \dots, \xi_{\nu, (l-\nu)})$. The $\mathcal{C}_\nu$ are conditionally independent given $\mathcal{A}_1$, and the conditional probabilities of the atoms of $\mathcal{C}_\nu$ given $\mathcal{A}_1$ are multinomial, with cell probabilities defined by the beta binomial distribution with parameters $(\beta_S, \alpha_S, l-\nu)$. Moreover, letting

$$\mathcal{C} = \mathcal{C}_0 \vee \dots \vee \mathcal{C}_{l-1}, \quad (6.54)$$

we see that $\mathcal{A}_1 \subset \mathcal{C}$ and $X \subset \mathcal{C}$. ($\mathcal{C}_l$, which consists of events involving control responders among litters having no controls, is trivial and so can be omitted from (6.54).)

Let $\mathcal{A}_1 \propto \mathcal{A}_1$. Because of (6.52), the atoms of $\mathcal{A}_1 \mathcal{C}_\nu$ can be considered to be the nodes of a tree of configurations defined by the natural parental mapping. By grafting the tree representing $\mathcal{A}_1 \mathcal{C}_1$ onto each node of the tree representing $\mathcal{A}_1 \mathcal{C}_0$, we can represent $\mathcal{A}_1 (\mathcal{C}_0 \vee \mathcal{C}_1)$ as a product tree. Iterating this process then produces $\mathcal{A}_1 \mathcal{C}$ as the product of l trees, which we call a *control tree*. Each atom of X intersects a control tree in the set of all nodes at a given height. The height h of each node, representing the number of control responders, thus defines (by Corollary 6.3.1) a critical number $\kappa(h)$ of test responders, and thus the rejection region within that atom. Since the atoms of $\mathcal{A}_1$ themselves are nodes of the allocation tree, all the atoms of $\mathcal{C}$ form a single tree obtained by grafting the appropriate control tree onto each node of the allocation tree. These atoms can then be

enumerated by visiting, for every non-negligible node in the allocation tree, the nodes in the control tree attached thereto.

In stepping through the control tree, shortcuts again are possible. Any tree grafted onto a node with negligible probability can be ignored, and entire branches can be truncated if an upper bound on their probability shows them to be negligible. One can also omit any nodes representing atoms disjoint from the critical region. Note that the maximum number of test responders, given $(d_0, \dots, d_l)$ is $\sum \nu d_\nu$. Rejection is impossible in any atom of C for which $\kappa(h) > \sum \nu d_\nu$. Since κ is an increasing function of the height of the node, no higher nodes need to be visited.

For each atom of C that cannot be ignored, the conditional probability of the rejection region must be calculated. Let $\eta_{\nu ij}$ ($i = 0, \dots, \nu; j = 0, \dots, l - \nu$) represent the number of litters with ν test animals of which i respond and $l - \nu$ control animals of which j respond. Clearly $\sum_i \eta_{\nu ij} = \xi_{\nu j}$. Denoting by $\mathcal{T}_{\nu j}$ the subalgebra generated by $(\eta_{\nu 0j}, \dots, \eta_{\nu \nu j})$, we see that the algebras $\mathcal{T}_{\nu j}$ are conditionally independent given $A \vee C$, and the probabilities of their atoms are again given by multinomial distributions with cell probabilities defined by (6.45). Thus the atoms of $\mathcal{T} = \mathcal{T}_{10} \vee \dots \vee \mathcal{T}_{l0}$ can be attached to another product tree, called the *test tree*. (Again note that $\mathcal{T}_{00}, \dots, \mathcal{T}_{l0}$ are trivial.) To find the conditional probability of rejection given a node in the control tree of height h , we need to weigh all of the nodes of the test tree whose heights are at least $\kappa(h)$, i.e., for which

$$\sum_{\nu ij} i \eta_{\nu ij} \geq \kappa(h).$$

A solution to this problem was discussed in Section 5.5.

The entire calculation can be formalized as follows: Let

$$\begin{aligned} A_2 &= A_1 \vee C_0 \\ A_3 &= A_2 \vee C_1 \\ &\vdots \\ A_{l+1} &= A_l \vee C_{l-1} \\ A_{l+2} &= A_{l+1} \vee \mathcal{T}_{l0} \\ A_{l+3} &= A_{l+2} \vee \mathcal{T}_{l-1,1} \\ &\vdots \\ A_{(l+1)(l+2)/2} &= A_{(l+1)(l+2)/2-1} \vee \mathcal{T}_{10}. \end{aligned} \tag{6.55}$$

It can be seen that in fact $A_{(l+1)(l+2)/2} = A$ and thus the sequence $A_0, A_1, \dots, A_{(l+1)(l+2)/2}$ defines a filtration for which the various conditional probabilities defined above represent transition probabilities. (Thus, for example, P_{l+3} is defined by a multinomial distribution over l cells, with cell probabilities given by (6.45) with $\nu = l - 1$, $x = 0$, and the number of trials given by $\xi_{l-1,1}$ which is a A_{l+2} -measurable (indeed A_{l-1} -measurable) random variable.)

Now, our goal is to calculate the probability of the rejection region, which is an event $R \in A$, in as few steps as possible. The strategy is to attempt the calculation in an algebra early in the sequence, and to proceed to a finer algebra only if necessary. Now given an algebra A_i , we compare R with each $A \propto A_i$. There are three possibilities.

(a) $A \cap R = \emptyset$ and thus $P(R|A) = 0$.

(b) $A \subset R$ and thus $P(R|A) = 1$.

(c) $\emptyset \neq A \cap R \neq A$ in which case calculation of $P(R|A)$ requires a finer algebra.

The algorithm needs to look at only those atoms of $\mathcal{A}_{i+1}$ that are contained in $\mathcal{A}_i$ -atoms satisfying (c) above. The three types of atom can be formally defined by

$$\begin{aligned}\mathcal{E}_i &= \{A \propto R^c A_i\}, \\ \mathcal{G}_i &= \{A \propto R A_i\}, \text{ and} \\ \mathcal{F}_i &= \{A \propto A_i : A \notin \mathcal{E}_i \cup \mathcal{G}_i\}.\end{aligned}\tag{6.56}$$

(Note that $\mathcal{E}_i$, $\mathcal{F}_i$, and $\mathcal{G}_i$ have been explicitly defined as mere collections of atoms, and not algebras.) Then the algorithm can be summarized by the expression

$$\begin{aligned}P(R) &= \sum_{A_1 \in \mathcal{G}_1} P_1(A_1, \Omega) \\ &+ \sum_{A_1 \in \mathcal{F}_1} P_1(A_1, \Omega) \sum_{A_2 \in \mathcal{G}_2} P_2(A_2, A_1) \\ &+ \dots \\ &+ \sum_{A_1 \in \mathcal{F}_1} P_1(A_1, \Omega) \sum_{A_2 \in \mathcal{F}_2} P_2(A_2, A_1) \dots \sum_{A_k \in \mathcal{G}_k} P_k(A_k, A_{k-1}),\end{aligned}\tag{6.57}$$

where k is the number of algebras (excluding $\mathcal{A}_0$) in the filtration.

Although the definitions (6.56) are appropriate for exact calculations, economies are possible if some truncation error is permitted. Suppose there are events N_i and $M_i \in \mathcal{A}_i$ with $\Pr\{N_i | \mathcal{A}_{i-1}\} < \epsilon/2$ and $\Pr\{M_i \Delta R | \mathcal{A}_{i-1}\} < \epsilon/2$. Then atoms $A \propto N_i A_i$ with $\Pr\{A_i | \mathcal{A}_{i-1}\} > 0$, although strictly in $\mathcal{F}_i$, could harmlessly be assigned to $\mathcal{E}_i$, while those in $M_i A_i$ could equally well be placed in $\mathcal{G}_i$. We need some assurance that the errors induced by such approximations do not propagate to an unacceptable extent. This assurance is provided by the following theorem.

Theorem 6.26 Let $\{\emptyset, \Omega\} = \mathcal{A}_0 \subset \mathcal{A}_1 \subset \dots \subset \mathcal{A}_k = \mathcal{A}$ be a finite filtration of a probability algebra $\mathcal{A}$ and let $c_1, \dots, c_k$ be a sequence of random variables such that $E(|c_i| | \mathcal{A}_{i-1}) < \epsilon$ for some $\epsilon > 0$. Given a random variable X , define a perturbed sequence $\tilde{X}_0, \dots, \tilde{X}_k$ by

$$\begin{aligned}\tilde{X}_k &= X \\ \text{and } \tilde{X}_{i-1} &= E(\tilde{X}_i + c_i | \mathcal{A}_{i-1}) \text{ for } i = 1, \dots, k.\end{aligned}\tag{6.58}$$

Then $|E(X) - \tilde{X}_0| < k\epsilon$.

Proof: Define $X_i = E(X | \mathcal{A}_i)$. Since

$$X_{i-1} - \tilde{X}_{i-1} = E(X_i - \tilde{X}_i | \mathcal{A}_{i-1}) - E(c_i | \mathcal{A}_{i-1})$$

$$\begin{aligned}
|X_{i-1} - \tilde{X}_{i-1}| &\leq E(|X_i - \tilde{X}_i| | \mathcal{A}_{i-1}) + E(|e_i| | \mathcal{A}_{i-1}) \\
&< E(|X_i - \tilde{X}_i| | \mathcal{A}_{i-1}) + \epsilon.
\end{aligned}$$

Letting $\delta_i = E(|X_i - \tilde{X}_i| | \mathcal{A}_{i-1})$ we then have

$$|X_{i-1} - \tilde{X}_{i-1}| < \delta_i + \epsilon$$

which implies that $\delta_{i-1} < \delta_i + \epsilon$ because δ_i is $\mathcal{A}_{i-1}$ -measurable. Since $\delta_k = 0$, $\delta_1 < (k-1)\epsilon$. Thus

$$|E(X) - \tilde{X}_0| < \delta_1 + \epsilon < k\epsilon.$$

QED

Theorem 6.26 can be used to put a bound on the truncation error induced by ignoring atoms with negligible conditional probabilities by putting X equal to the indicator function of R , and e_i equal to the indicator function of events that are ignored at the i^{th} step in the filtration.

If the atoms of $\mathcal{A}_i$ contained in an $\mathcal{A}_{i-1}$ -atom are defined by a random variable whose conditional distribution given $\mathcal{A}_{i-1}$ is in the (univariate) Pólya family, then the negligible events N_i and M_i can be defined by the tail bounds discussed in Section 4.2.3. Such events can be identified in (multivariate) configurations by reducing weighing of branches to accumulating univariate tail areas as described by (5.12). In fact, the calculation (5.12) is actually based on a filtration where each of the variables S_i in (4.46) defines a subalgebra. Since Theorem 6.26 applies equally well to this case, negligible events in an algebra $\mathcal{A}_i$ generated by a configuration with $j+1$ cells can be identified by truncation, in the conditional univariate distributions, of tails with probabilities less than ϵ/j . Alternatively, one could consider as the filtration in Theorem 6.26, the sequence of algebras obtained from (6.55) by replacing each algebra in the sequence by its own filtration in terms of subalgebras generated by the S_i 's and adjusting the value of k accordingly. This tactic fails, however, for the algebra $\mathcal{A}$ whose structure is defined by a lateral tree for which there are no expressions analogous to (4.46) or (5.12). For this algebra, identification of negligible events must be done by less efficient, *ad hoc* methods.

6.5.2 Nested design

For the nested design, the algorithm for evaluating the power can be based on the same principle as that for the Fisher-Irwin test in the case of two binomial populations. In that case, we determined the rejection region in terms of a minimum number of test responders for a given number of control responders. Both the control and test probabilities can be calculated recursively, and the determination of the critical value for the next control outcome can be determined by updating that for the previous one.

In the nested design, we again have a similar situation with independent control and test litters. This time, however, the outcomes are not just the number of responders, but configurations of litters. Let x_i and y_i be the number of control and treated litters, respectively, having exactly i responders ($i = 0, \dots, l$). Then the outcome can be represented by the pair of configurations (x, y) with $x = (x_0, \dots, x_l)$ and $y = (y_0, \dots, y_l)$. The

distributions of these configurations are multinomial, with the cell probabilities given by the beta binomial distribution with parameters (β_s, α_s) for the control configuration, and by (6.41) with $\nu = l$ and $x = 0$ for the test configuration. Denote by X and Y the algebras generated by $\hat{x}$ and y , and define $A = X \vee Y$. Then the atoms of A can be considered to be nodes in a product tree which we call the *unconditional tree*.

Let Z be the subalgebra generated by $z = (x_0 + y_0, \dots, x_l + y_l)$. The total number of control and test responders is given by $T_0 = \sum_i ix_i$ and $T_1 = \sum_i iy_i$, respectively. Although the permutation test is defined by the conditional distribution of T_1 given Z , Corollary 6.3.1 asserts that the rejection region in the univariate case ($l = 1$) can in fact be described solely in terms of T_0 and T_1 . This is not true in the multivariate ($l > 1$) case as we shall see in an example presently. However, theorems analogous to those in Section 6.1.3, which involve not just T_0 and T_1 but the configurations x and y , can be derived. Although $\hat{x}$, y , and z are analogous to the quantities with the same designations in Section 6.1.3, here these outcomes have only a partial ordering defined by a tree structure. Thus the definitions of critical and monotone tails have to be generalized to handle this case.

We actually need to consider two tree structures simultaneously. The rejection region R is defined by the conditional probabilities given Z , and X and Y are not conditionally independent given Z . The atoms of Z , being defined by configurations, can themselves be embedded in a tree, which we call the *marginal tree*. The probabilities of the marginal tree are not very tractable in the case of the alternative hypothesis, but that need not concern us since it is only conditional probabilities that we will need. We now define, for each node $(z_0, \dots, z_l)$ of the marginal tree, a tree of configurations $(y_0, \dots, y_l)$ of n_1 balls with capacities given by $(z_0, \dots, z_l)$. The nodes of this tree, which we call the *conditional tree*, represent the atoms of zY , and hence their probabilities, conditional on Z (under the null hypothesis) are given by (6.29). We can define another tree structure on the atoms of A by grafting, onto each node of the marginal tree the corresponding conditional tree. We call this structure the *grafted tree*.

Although the nodes of the trees are in fact atoms of subalgebras, we will employ different symbols for the trees, so as not to confuse the different orderings. The unconditional tree, which is the product of control and test trees, X and Y , respectively, will be denoted by $X \times Y$, the conditional tree grafted onto $z \propto Z$ by Y_z , and the grafted tree by G . Although $X \times Y$ and G represent different trees, they can be considered to have the same nodes, since in either case they represent the atoms of A . The nodes of Y_z , representing atoms of zY , can be mapped in an obvious way to the nodes of Y , which represent atoms of Y . The order defined by the tree structure is not preserved by this mapping however, so that Y_z is not generally isomorphic to a subtree of Y .

For any $z \propto Z$, the conditional rejection region is defined as

$$R_z = \{y \propto Y_z : \Pr\{T_1 \geq t(y) | z\} \leq \alpha\}, \quad (6.59)$$

where $t(y) = \sum_i iy_i$ is a linear function on Y_z . The set of atoms in A leading to rejection can be defined as a set of nodes in G ,

$$R = \{(z, y) \in G : y \in R_z\}, \quad (6.60)$$

Figure 6.1: Conditional tree with $k = 2$, $n_1 = 3$ and margin $(1,1,3)$

but is obviously also a set of nodes in $X \times Y$. To determine the power, we need to calculate the probability of R . This computation is evidently easier to perform in $X \times Y$ than in G , but to do so we require a description of R in terms of the tree structure of $X \times Y$. Before presenting this description, we look at an example to elucidate some of the complications involved.

Suppose $m = 5$, $m_1 = 3$, $l = 2$ and $\alpha = .5$. (In practice, of course, one would not use such a large significance level, but it serves us well to illustrate the possibilities with a small example.) Consider the pair of outcomes $x \doteq (1, 0, 1)$ for the control configuration and $y = (0, 1, 2)$ for the test configuration. The marginal configuration is then $z = (1, 1, 3)$. To determine the rejection region we must consider the conditional distribution given z . The conditional tree is shown in Figure 6.1. As noted before, the height of a node corresponds to the value of $t(y) = \sum i y_i$, and beside the tree is written down the cumulative distribution of this statistic. Since $\Pr\{T_1 \geq 5\} = 0.4 < 0.5$, the rejection region, conditionally on $z = (1, 1, 3)$ contains the configurations $(0, 1, 2)$ and $(0, 0, 3)$, and in fact $(0, 1, 2)$ is the smallest (in the tree order) configuration in the rejection region given the marginal configuration $(1, 1, 3)$. The node $((1, 1, 3), (0, 1, 2))$ in the grafted tree maps to the node $((1, 0, 1), (0, 1, 2))$ in the unconditional tree. Figure 6.2 shows the test tree. Note that the order in the conditional tree is not a restriction of the order in the test tree. Thus in particular, $(1, 1, 1) < (0, 1, 2)$ in $Y_{(1,1,3)}$, whereas $(1, 1, 1)$ and $(0, 1, 2)$ are incomparable in Y .

Consider now the node $((1, 0, 1), (0, 2, 1))$ in $X \times Y$. The rejection region is here determined by the conditional distribution given $z = (1, 2, 2)$. The conditional tree is shown in Figure 6.3. With $\alpha = 0.5$ the critical region consists of $(1, 0, 2)$, $(0, 2, 1)$, and $(0, 1, 2)$. Thus the node $((1, 0, 1), (0, 2, 1))$ is also in the rejection region, and by a similar argument one can see that $((1, 0, 1), (0, 0, 3))$ is as well.

Consider, however, the node $((1, 0, 1), (0, 3, 0))$ of $X \times Y$. The conditional tree is given by Figure 6.4, from which we see that $(0, 3, 0)$ is not in the rejection region. Thus, for

Figure 6.2: Test tree with $k = 2$ and $n_1 = 3$ Figure 6.3: Conditional tree with $k = 2$, $n_1 = 3$ and margin $(1,2,2)$

Figure 6.4: Conditional tree with $k = 2$, $n_1 = 3$ and margin $(1,3,1)$ Figure 6.5: Conditional tree with $k = 2$, $n_1 = 3$ and margin $(2,0,3)$

a control configuration of $(1, 0, 1)$, all points in the branch springing from $(0, 2, 1)$ in the unconditional tree would be in the rejection region, whereas those points less than this node would not be. Thus the point $(0, 2, 1)$ can be considered a *critical configuration* for the control configuration $(1, 0, 1)$. Note that it is the *configuration* that is critical and not just the value of the test statistic. Thus, even though $y = (1, 0, 2)$ has the same value of T_1 as $y = (0, 2, 1)$, the relevant probability for the latter test configuration is conditional on the margin $(2, 0, 3)$. The conditional tree is then given in Figure 6.5 from which we see that $(1, 0, 2)$ is not a critical configuration for the control configuration $(1, 0, 1)$.

It should also be noted that even if two control configurations give the same value of T_0 , they do not necessarily have the same critical test configurations. For the control configuration $(0, 2, 0)$, the test configuration $(0, 2, 1)$ is not critical but $(1, 0, 2)$ is.

To proceed with the discussion we require a few more definitions regarding trees. Given any partially ordered set $(X, <)$, we define an *extremal subset* as a set $A \subset X$ such that if $x \in A$ and $x' > x$ then $x' \in A$. If the partial ordering is defined by a tree, then an extremal set will consist of a union of branches. It is obvious that the complement of an extremal set in a tree will be a subtree containing the root. Such a set will be called a *rooted subset* of the tree. In terms of the order relationship, a rooted subset will be *extremal* with respect to the converse order. Thus the concepts of rooted and extremal subsets are dual to each other.

Denote by τ^X and τ_X the sets of extremal and rooted subsets of X , respectively. If X is a tree and $A \in \tau^X$ then A is determined by its minimal elements which form a *boundary* for A

$$\partial A = \{x \in A : \hat{x} \notin A\} \quad (6.61)$$

in the sense that

$$\begin{aligned} A &= \{x \in X : x \succeq a \text{ for some } a \in \partial A\} \\ &= \bigcup_{a \in \partial A} B_a, \end{aligned}$$

where, as in Chapter 5, B_a represents the branch springing from a . It is clear from the definition of ∂A that if $a, a' \in \partial A$, $a \neq a'$, then $B_a \cap B_{a'} = \emptyset$.

Given two partially ordered sets X and Y , consider a mapping $\phi: X \rightarrow 2^Y$. For such a mapping we define an *adjoint mapping* $\phi^*: Y \rightarrow 2^X$ defined by

$$\phi^*(y) = \{x : y \in \phi(x)\}. \quad (6.62)$$

We will be particularly concerned with mappings such that $\phi(y)$ is an extremal (or dually, rooted) set for all y .

If $\phi: X \rightarrow \tau^Y$ then ϕ is said to be *antitonic* if $x' > x$ implies $\phi(x') \subset \phi(x)$. Dually, $\phi: Y \rightarrow \tau_X$ will be called *isotonic* if $y' < y$ implies $\phi(y') \subset \phi(y)$. The following results will be useful.

Lemma 6.5 *If $\phi: X \rightarrow \tau^Y$ is antitonic, then its adjoint ϕ^* is isotonic, and the range of the adjoint is contained in τ_X .*

Proof: Let $A = \phi^*(y) = \{x : y \in \phi(x)\}$. Suppose that $x \in A$ and $x' < x$. Then $y \in \phi(x) \subset \phi(x')$. Thus $x' \in A$ and A is a rooted subset. Now, suppose that $y' < y$ and

$A' = \phi^*(y')$. If $x \in A'$ then $y' \in \phi(x)$ which implies that $y \in \phi(x)$ since $\phi(x) \in \mathcal{T}^Y$. Hence $x \in A$ and we have $A' \subset A$. QED

Lemma 6.6 *If A is a subset of nodes in a tree such that for any $x \in A$, $\hat{x} \in A$, then A is a rooted set.*

Proof: If $x' < x$ then there is a sequence $x' = x_1, x_2, \dots, x_l = x$ such that $x_i = \hat{x}_{i+1}$. By hypothesis, $x_{l-1} \in A$ and the result then follows by induction. QED

Lemma 6.7 *If A is a set such that for any $x \in A$ all offspring of x are in A , then A is an extremal set.*

Proof: By duality. QED

Lemma 6.8 *If $\phi: X \rightarrow \mathcal{T}_Y$ is isotonic, then its adjoint ϕ^* is antitonic and the range of the adjoint is contained in $\mathcal{T}^X$.*

Proof: By duality. QED

Lemma 6.9 *Let X and Y be trees and let f be a real-valued function on $X \times Y$. Suppose that $\phi: X \rightarrow \mathcal{T}^Y$ is antitonic. Then*

$$\sum_{x \in X} \sum_{y \in \phi(x)} f(x, y) = \sum_{y \in Y} \sum_{x \in \phi^*(y)} f(x, y). \quad (6.63)$$

Proof: The summation on the left hand side is over $\{(x, y) : y \in \phi(x)\}$. But by definition of ϕ^* , this is precisely the set $\{(x, y) : x \in \phi^*(y)\}$ which is the range of summation on the right hand side. QED

The algorithm of Section 6.1.2 is based on the fact that the conditional distribution of the test statistic had monotone tails, and thus the critical region could be defined in terms of tail areas of the joint unconditional distributions of the control and test responders. Since the algebras generated by these two types of responders are independent, the probability computation requires only one pass through the atoms of each of them. We now examine whether this property can be extended to the case where the control and test responses are multivariate. To do this, we need to extend the definitions of Section 6.1.2.

Because t is an isotone function, it is clear that R_s is an extremal subset of Y_s . Also, (provided that $\alpha < 1$), $rt(Y_s) \notin R_s$, and thus R is an extremal subset of G . Extremal sets are thus the extension of the idea of the tail of a distribution. We can still call the pair (x, y) in the unconditional tree *supercritical* if it corresponds to a node in the rejection region R in the grafted tree. The definition of *critical* points does not extend so easily, for it requires one to determine whether the point obtained by decrementing y and incrementing x is still supercritical. The idea of decrementing y can easily be generalized to taking the parent of y in the test tree, but incrementing x would generalize to taking the offspring

of y , which is not in general unique. The choice of offspring could affect the determination of supercriticality. The essential aspect of the critical pairs is that for a family with critical tails they define the *boundary* of the rejection region in the unconditional tree. However, the boundary is only a meaningful concept for extremal and rooted subsets. Thus the essential property that generalizes the idea of critical tails is that the rejection region, which is an extremal subset in the grafted tree, is also extremal in the unconditional tree. The boundary of this extremal subset of the unconditional tree thus generalizes the role of the critical pairs.

Let $F: \mathfrak{R} \times Z \rightarrow \mathfrak{R}$ be the family of conditional distributions of the test statistic $T_1 = \sum_i y_i$:

$$F(s, z) = \Pr \{T_1 \leq s | z\}. \quad (6.64)$$

The rejection region in the grafted tree can then be defined as

$$\begin{aligned} R &= \{(z, y) \in G : F(t(y)^-, z) \geq 1 - \alpha\} \\ &= \{(z, y) \in G : F(t(y) - 1, z) \geq 1 - \alpha\}. \end{aligned} \quad (6.65)$$

since T_1 can only take integer values.

To continue our discussion, we also need to generalize the parental concept. For a configuration y , define the mapping L_j by $L_j y = y'$ where

$$\begin{aligned} y'_{j-1} &= y_{j-1} + 1, \\ y'_j &= y_j - 1, \\ \text{and } y'_i &= y_i \text{ for } i \notin \{j-1, j\}. \end{aligned} \quad (6.66)$$

For configurations with unconstrained capacities, $\hat{y} = L_m y$ where $m = \max\{i : y_i > 0\}$. In general, however, $L_j y$ need not even be a legitimate configuration. Necessary and sufficient conditions are that $y_j > 0$ and $y_{j-1} < v_{j-1}$ (where v_i , as before, represents the capacity of the i^{th} cell). If we just consider L_j as a mapping on the set of integer-valued vectors $(y_0, \dots, y_k)$ with $\sum y_i = n$, then it is in fact a one-to-one correspondence.

The expression $F(s, z)$ is only defined for legitimate configurations z . To avoid the tedium of exceptional cases, arbitrarily define $F(s, z) = 0$ for all s if z is not a configuration.

We now present the generalizations of Theorems 6.1 and 6.2.

Theorem 6.27 *If $F(s-1, L_j z) \leq F(s, z)$ for all $s \in \mathfrak{R}$, $z \in Z$, and all $j \in \{1, \dots, l\}$ such that $z_j > 0$, then the rejection region R is an extremal subset of $X \times Y$.*

Proof: We first show that for any $z \in X$, $D_z = \{y : (z, y) \notin R\}$ is a rooted subset of Y . If $(z, y) \notin R$ then $F(t(y) - 1, z + y) < 1 - \alpha$. Now $\hat{y} = L_m y$ where $m = \max\{i : y_i > 0\}$. Thus $z + \hat{y} = z + L_m y = L_m(z + y)$. Moreover, $t(\hat{y}) = t(y) - 1$. Thus

$$\begin{aligned} F(t(\hat{y}) - 1, z + \hat{y}) &= F(t(y) - 2, L_m(z + y)) \\ &\leq F(t(y) - 1, z + y) \\ &< 1 - \alpha. \end{aligned}$$

Thus $(z, \hat{y}) \notin R$, and $\hat{y} \in D_z$. By Lemma 6.6 D_z is a rooted subset, and it is nonempty since $F(t(\text{rt } Y) - 1, z + y) = 0 < 1 - \alpha$. Thus D_z is an extremal subset which does not

contain the root. Now $R = \{(x, y) : y \in D_x^c\}$. If $(x, y) \in R$ and $(x', y') > (x, y)$ then since $y \neq \text{rt } Y$ we must have $x' = x$ and $y' > y$. But then $y' \in D_x^c = D_{x'}^c$, so $(x', y') \in R$. QED

Theorem 6.28 *If $F(s, L_j z) \geq F(s, z)$ for all $s \in \mathfrak{R}$, $z \in Z$, and all $j \in \{1, \dots, l\}$ such that $z_j > 0$, then the mapping $\kappa : x \mapsto \{y \in Y : (x, y) \in R\}$ is antitonic.*

Proof: We actually prove that the complementary mapping is isotonic. The proof then follows by duality. Let $x' > x$. Then there is a chain $x = x^{(0)}, x^{(1)}, \dots, x^{(k)} = x'$ such that $x^{(i-1)} = \hat{x}^{(i)}$. Suppose that $(x, y) \notin R$. Then $F(t(y) - 1, x + y) < 1 - \alpha$. Now letting $m = \max\{i : x_i^{(1)} > 0\}$ we have that $\hat{x} = L_m x^{(1)}$. Thus

$$\begin{aligned} F(t(y) - 1, x^{(1)} + y) &\leq F(t(y) - 1, L_m(x^{(1)} + y)) \\ &= F(t(y) - 1, x + y) \\ &< 1 - \alpha. \end{aligned}$$

Hence $(x^{(1)}, y) \notin R$. By induction it thus follows that $(x', y) \notin R$. Thus $[\kappa(x)]^c \subset [\kappa(x')]^c$ or $\kappa(x') \subset \kappa(x)$. Therefore $\kappa(x)$ is antitonic. QED

We will presently show that the required conditions do hold for the multivariate hypergeometric family. We will do this by the next lemma, which breaks the probability of a node into two pieces. This idea is a formalization and generalization of the paradigm of the "magic ball" used in Theorem 6.3

Before we present these results, a caution on the application of the L_j operators is in order. We will consider nodes $(z, y) \in G$, where $y \in Y_z$. Since both z and y are configurations, we can apply the operators L_j to them, although the result will not always be a configuration. Since Z is a tree of unconstrained configurations, $L_j z \in Z$ if and only if $z_j > 0$. However, even if $L_j z \in Z$ it is not necessary that $y \in Y_{L_j z}$, and thus $(L_j z, y)$ may not be a node in G . The tidiest way of handling such difficulties is to ensure that such "pseudo-nodes" have probability zero. Although we could arbitrarily make such a definition, it would be better if the computational formulas used actually give zero probability in such cases.

The multivariate hypergeometric formula (4.48) employs binomial coefficients $\binom{n}{r}$, which are well defined for any real number n . The configuration y will fail to be a node in $Y_{L_j z}$ if and only if $y_j = z_j$. In this case, however, the formula will include a factor $\binom{z_j}{z_j}^{-1}$ which equals zero if $z_j \geq 1$. (However, if $z_j = 0$ then $\binom{z_j}{z_j}^{-1} = 1$. One must thus be careful when employing the formula for nonexistent nodes, because such pseudo-configurations may not always be assigned zero probability, and may in fact give negative "probabilities" — e.g. $\binom{-1}{1} = -1$.)

Let

$$P(y|z) = \frac{\prod_{i=0}^l \binom{z_i}{y_i}}{\binom{m}{m_1}} \quad (6.67)$$

be the conditional probability of the node y given z . The following lemma involves the following (conditional) subprobabilities defined on G : For any $j \in \{0, \dots, l\}$ such that $z_j > 0$ let

$$Q_j(y, z) = \frac{z_j - y_j}{z_j} P(y|z), \quad (6.68)$$

$$\text{and } R_j(y, z) = \frac{y_j}{z_j} P(y|z). \quad (6.69)$$

Obviously $P(y|z) = Q_j(y, z) + R_j(y, z)$.

Lemma 6.10 $P(y|z) = Q_{j-1}(y, L_j z) + R_{j-1}(L_j y, L_j z)$.

Proof: Using the relationship

$$\binom{n}{r} = \binom{n-1}{r} + \binom{n-1}{r-1} \quad (6.70)$$

we can write

$$P(y|z) = \left[\binom{z_{j-1}}{y_{j-1}} \left(\binom{z_j - 1}{y_j} + \binom{z_j - 1}{y_j - 1} \right) \right] \frac{\prod_{i \in \{j, j-1\}} \binom{z_i}{y_i}}{\binom{m}{m_1}}. \quad (6.71)$$

The expression in brackets can be rearranged to give

$$\frac{z_{j-1} + 1 - y_{j-1}}{z_{j-1} + 1} \binom{z_{j-1} + 1}{y_{j-1}} \binom{z_j - 1}{y_j} + \frac{y_{j-1} + 1}{z_{j-1} + 1} \binom{z_{j-1} + 1}{y_{j-1} + 1} \binom{z_j - 1}{y_j - 1}. \quad (6.72)$$

Substituting this back into (6.71) then gives

$$\begin{aligned} & \frac{z_{j-1} + 1 - y_{j-1}}{z_{j-1} + 1} P(y, L_j z) + \frac{y_{j-1} + 1}{z_{j-1} + 1} P(L_j y, L_j z) \\ &= Q_{j-1}(y, L_j z) + R_{j-1}(L_j y, L_j z). \end{aligned} \quad (6.73)$$

QED

Remark: If $y_j = z_j$ then as seen above $P(y, L_j z) = 0$ and hence $Q_{j-1}(y, L_j z) = 0$.

Theorem 6.29 If $F(s, z)$ is given by the multivariate hypergeometric family

$$F(s, z) = \sum_{y: \mathbf{t}(y) \leq s} P(y|z) \quad (6.74)$$

where $P(y|z)$ is defined by (6.67), then

$$F(s - 1, z) \leq F(s - 1, L_j z) \leq F(s, z). \quad (6.75)$$

Proof: Denoting by V_s the set $\{y \in Y : t(y) \leq s\}$ we have

$$\begin{aligned} F(s, z) &= \sum_{y \in Y_s \cap V_s} P(y|z) \\ &= \sum_{y \in Y_s \cap V_s} Q_{j-1}(y, L_j z) + \sum_{y \in Y_s \cap V_s} R_{j-1}(L_j y, L_j z) \end{aligned} \quad (6.76)$$

by Lemma 6.10. On the other hand,

$$\begin{aligned} F(s-1, L_j z) &= \sum_{y \in Y_{L_j s} \cap V_{s-1}} P(y|z) \\ &= \sum_{y \in Y_{L_j s} \cap V_{s-1}} Q_{j-1}(y, L_j z) + \sum_{y \in Y_{L_j s} \cap V_{s-1}} R_{j-1}(y, L_j z). \end{aligned} \quad (6.77)$$

Care must be taken in comparing these two expressions since the summations are not over the same sets. If

$$\begin{aligned} J_0 &= Y_s \cap Y_{L_j s}, \\ J_1 &= Y_s \setminus Y_{L_j s}, \\ \text{and } J_2 &= Y_{L_j s} \setminus Y_s \end{aligned}$$

then the summation in (6.76) is over $(J_0 \cup J_1) \cap V_s$ whereas that in (6.77) is over $(J_0 \cup J_2) \cap V_{s-1}$. However, $J_1 = \{y : y_j = z_j\}$, and for such y , $P(y, L_j z) = 0 = Q_{j-1}(y, L_j z)$. Also $J_2 = \{y : y_{j-1} = z_{j-1} + 1\}$ and for such y ,

$$Q_{j-1}(y_j, L_j z) = \frac{z_{j-1} + 1 - y_{j-1}}{z_{j-1} + 1} P(y_j, L_j z) = 0.$$

Thus for $y \in J_1 \cup J_2$, $Q_{j-1}(y, L_j z) = 0$, and thus

$$\sum_{y \in Y_s \cap V_s} Q_{j-1}(y, L_j z) = \sum_{y \in Y_{L_j s} \cap V_s} Q_{j-1}(y, L_j z). \quad (6.78)$$

Now since $V_{s-1} \subset V_s$ and all terms are non-negative (6.78) gives that

$$\sum_{y \in Y_s \cap V_s} Q_{j-1}(y, L_j z) \geq \sum_{y \in Y_{L_j s} \cap V_{s-1}} Q_{j-1}(y, L_j z). \quad (6.79)$$

Now consider $\sum_{y \in Y_s \cap V_s} R_{j-1}(L_j y, L_j z)$. Letting

$$\begin{aligned} H &= \{y : L_j y \in Y_{L_j s}\}, \\ K_0 &= H \cap Y_s, \\ K_1 &= Y_s \setminus H, \\ \text{and } K_2 &= H \setminus Y_s \end{aligned}$$

we see that the summation (6.76) is over $(K_0 \cup K_1) \cap V_s$. However, $K_1 = \{y : y_j = 0\}$. For such y , the expression for $P(L_j z, L_j y)$ will include a factor $\binom{z_j - 1}{-1} = 0$, and thus $R_{j-1}(L_j y, L_j z) = 0$. Moreover, $K_2 = \{y : y_{j-1} = -1\}$ and for such y ,

$$R_{j-1}(L_j y, L_j z) = \frac{y_{j-1} + 1}{z_{j-1} + 1} P(L_j y, L_j z) = 0.$$

Thus

$$\begin{aligned} \sum_{y \in Y_s \cap V_s} R_{j-1}(L_j y, L_j z) &= \sum_{y \in H \cap V_s} R_{j-1}(L_j y, L_j z) \\ &= \sum_{y \in Y_{L_j s} \cap V_{s-1}} R_{j-1}(y, L_j z), \end{aligned} \quad (6.80)$$

where the last equality follows from the definition of H and the fact that $t(L_j y) = t(y) - 1$. Combining (6.79) and (6.80) with (6.76) and (6.77) then gives

$$F(s-1, L_j z) \leq F(s, z). \quad (6.81)$$

On the other hand,

$$\begin{aligned} F(s-1, z) &= \sum_{y \in Y_s \cap V_{s-1}} Q_{j-1}(y, L_j z) + \sum_{y \in Y_s \cap V_{s-1}} R_{j-1}(L_j y, L_j z) \\ &= \sum_{y \in Y_s \cap V_{s-1}} Q_{j-1}(y, L_j z) + \sum_{y \in Y_{L_j s} \cap V_{s-2}} R_{j-1}(y, L_j z) \end{aligned} \quad (6.82)$$

using similar arguments. Comparing (6.82) with (6.77) we see that

$$F(s-1, z) \leq F(s-1, L_j z).$$

QED

We can now present a recursive algorithm for calculating the probability of the critical region R as a generalization of that presented in Section 6.1.2.

For any $x \in X$, let $\kappa(x) = \{y \in Y : (x, y) \in R\}$. Then by Theorem 6.27 we know that $\kappa(x)$ is an extremal set of Y , and by Theorem 6.28 that κ is an antitonic mapping. We thus have

$$\begin{aligned} P(R) &= \sum_{x \in X} P(x) P(\kappa(x)) \\ &= \sum_{x \in X} P(x) \sum_{y \in \partial \kappa(x)} \sum_{y' \geq y} P(y') \end{aligned} \quad (6.83)$$

Now the sum can be evaluated by stepping through the tree X , using the tree visiting algorithm described in Section 5.1.2. As usual, branches with negligible weights can be truncated. To each node $x \in X$, we attach the set $\partial \kappa(x)$ of critical configurations. These can be determined, dynamically, as the nodes in X are visited. Once $\partial \kappa(x)$ is defined, we

know that the critical configurations $\partial\kappa(\bar{x})$ corresponding to any offspring $\bar{x}$ of x can only be descendants of the nodes in $\partial\kappa(x)$.

To determine the elements of $\partial\kappa(\bar{x})$, we begin with a list of candidates M , equal to $\partial\kappa(x)$, and a list N of confirmed members of $\partial\kappa(\bar{x})$, which is initially empty. We then remove a candidate y from M and check whether

$$\Pr\{T_1 \geq t(y)|\bar{x} + y\} > \alpha. \quad (6.84)$$

If not, we append y to N . Otherwise we append all the offspring of y to M . This process continues until M is empty. The verification of (6.84) can be done by the procedures discussed in Section 5.3.

6.5.3 Litter matched design

In Section 6.3.3 we saw that although the invariant algebra for the litter matched design is $\mathcal{W}$, (generated by $w = (w_1, \dots, w_m)$, the number of responders in each litter), we can in fact condition on the smaller algebra $\mathcal{Z}$ (generated by $z = (z_0, \dots, z_l)$, the number of litters with a given number of responders) provided we only want the probabilities of $\mathcal{B}$ -invariant events. The rejection region is again defined in terms of the conditional distribution of T_1 , the total number of test responders.

Let $K_t = \{T_1 = t\}$. Then clearly $K_t \in \mathcal{B}$ and thus by Theorem 6.23 $P(K_t|\mathcal{W}) = P(K_t|\mathcal{Z})$. The rejection region is defined in terms of the family of conditional distributions

$$F(t, z) = \Pr\{T_1 \leq t|z\} = \sum_{r=0}^t P(K_r|z), \quad (6.85)$$

as the set of those outcomes for which $F(t-1, z) \geq 1 - \alpha$.

As in the nested design case, the rejection region is not determined solely by the total number of responders under each treatment, but requires the list of configurations $\{u_{ij}\}$ defined in (6.33). For example, consider the case $m = 5$, $l = 2$, $\nu = 1$, and the two lists of configurations:

j	z_j	u_{0j}	u_{1j}
0	2	2	
1	2	0	2
2	1		1

(6.86)

j	z_j	u_{0j}	u_{1j}
0	1	1	
1	4	1	3
2	0		0

(6.87)

In both cases there is one control responder and three test responders. However, in the first case the control response occurs in a litter that also has a test responder, whereas in the second case there is a litter having a control response but no test response.

In (6.86) there are two litters with exactly one responder, and in both of these this responder is in the treated portion. This list of configurations produces the maximum possible value of T_1 given the condition $z_0 = 2$, $z_1 = 2$, and $z_2 = 1$ and its conditional

probability can be calculated as $\frac{1}{4}$. In (6.87) there are four litters with exactly one responder, but in only three of them is the response in the treated portion. The conditional probability of this list of configurations is also $\frac{1}{4}$, but the list

$$\begin{array}{cc} j & z_j \\ 0 & 1 \\ 1 & 4 \\ 2 & 0 \end{array} \begin{array}{cc} u_{0j} & u_{1j} \\ \boxed{\begin{array}{cc} 1 & \\ & 0 \end{array}} & \boxed{\begin{array}{cc} & 4 \\ & 0 \end{array}} \end{array} \quad (6.88)$$

has a higher value of T_1 and has the same value of the condition (z_0, z_1, z_2) . The conditional probability of (6.88) is $\frac{1}{16}$, and thus the P -value corresponding to (6.87) is $\frac{5}{16}$. Thus if we were to choose $\alpha = \frac{9}{32}$, then (6.86) would be in the rejection region, whereas (6.87) would not be.

The unconditional distribution of the list of configurations is multinomial, with the cell probabilities of u_{ij} given by (6.41) with $\nu = n_1 l / m$, $x = l - \nu - j + i$, and $y = \nu - i$. From Theorem 4.8 it then follows that the marginal distribution of z is also multinomial. Unlike the nested case, the control and treated configurations are not independent, so there is no advantage to constructing an analogue of the unconditional product tree. Thus power calculations can be made directly by calculating the conditional power given Z and then averaging over the atoms of Z . In Section 6.3.3 we saw that the permutation distribution was product multinomial, and Theorem 4.9 shows that the conditional distribution is still product multinomial under the alternative hypothesis.

The marginal configurations z can be embedded in a tree, and given z , the list of configurations u , being independent, can be represented by a product tree. The entire list can then be represented by grafting the corresponding product tree onto each node of the marginal tree. To calculate the power, one can then step through the nodes of the marginal tree, determining the rejection region, using the null conditional distribution, and then computing its probability under the alternative conditional distribution. As before, truncation of negligible branches is possible. Again it would be convenient to be able to determine the rejection region recursively. To do this, we require the following result on the stochastic monotonicity of this family of distributions.

Theorem 6.30 *For the litter matched design, the family of conditional distributions (6.85) is stochastically increasing, in the sense that $F(t, z) \leq F(t, \hat{z})$.*

Proof: Given a vector $w \propto \mathcal{W}$ of total litter responders, the vector $y = (y_1, \dots, y_m)$ of responders in the test portion of each litter has the conditional probability

$$P(y, w) = \prod_{j=1}^m \frac{\binom{l - w_j}{\nu - y_j} \binom{w_j}{y_j}}{\binom{l}{\nu}} \quad (6.89)$$

For any j such that $w_j > 0$ define

$$Q_j(y, w) = \frac{w_j - y_j}{w_j} P(y|w) \text{ and} \quad (6.90)$$

$$R_j(y, w) = \frac{y_j}{w_j} P(y|w) \quad (6.91)$$

so that $P(y, w) = Q_j(y, w) + R_j(y, w)$. Now denoting by e^i the m -dimensional vector with 1 in the i^{th} position and zeroes elsewhere, we have that

$$P(y, w - e^i) = \frac{\binom{l - w_i + 1}{\nu - y_i} \binom{w_i - 1}{y_i}}{\binom{l}{\nu}} \times \prod_{j \neq i} \frac{\binom{l - w_j}{\nu - y_j} \binom{w_j}{y_j}}{\binom{l}{\nu}}. \quad (6.92)$$

The numerator of the factor before the product sign in (6.92) can be written

$$\begin{aligned} & \left[\binom{l - w_i}{\nu - y_i - 1} + \binom{l - w_i}{\nu - y_i} \right] \binom{w_i - 1}{y_i} \\ &= \binom{l - w_i}{\nu - y_i - 1} \binom{w_i}{y_i + 1} \frac{y_i + 1}{w_i} + \binom{l - w_i}{\nu - y_i} \binom{w_i}{y_i} \frac{w_i - y_i}{w_i}. \end{aligned} \quad (6.93)$$

Substituting (6.93) into (6.92) we have

$$P(y, w - e^i) = R_i(y + e^i, w) + Q_i(y, w) \quad (6.94)$$

provided that $w_i > 0$. Note that (6.94) is valid even in the case $y_i = w_i$ in which case both sides are formally, as well as logically, equal to zero.

Now consider a particular pair (y, w) . This pair represents an atom of $\mathcal{A}$, and will be contained in a particular K_i . Now, if $(y, w) \in K_i$, then $(y + e^i, w) \in K_{i+1}$. Also, given a vector of litter responses, w can be construed to be a subevent of z . If z is not the root of the marginal tree, then $z_i > 0$ for some $i > 0$. Letting $k = \max\{i : z_i > 0\}$ and $j^* = \min\{j : w_j = k\}$, we have that if $w \subset z$ then $w - e^{j^*} \subset z$.

Now

$$\begin{aligned} F(t, z) &= \sum_{r=0}^t P(K_r | z) \\ &= \sum_{r=0}^t P(K_r | w), \end{aligned} \quad (6.95)$$

for any $w \subset z$. But

$$\begin{aligned} P(K_r | w) &= \sum_{y \in K_r} P(y, w) \\ &= \sum_{y \in K_r} [Q_j(y, w) + R_j(y, w)]. \end{aligned} \quad (6.96)$$

On the other hand

$$\begin{aligned} F(t, \hat{z}) &= \sum_{r=0}^t P(K_r | \hat{z}) \\ &= \sum_{r=0}^t P(K_r | w - e^{j^*}) \end{aligned} \quad (6.97)$$

with

$$P(K_r | w - e^{j^*}) = \sum_{y \in K_r} Q_{j^*}(y, w) + \sum_{y \in K_r} R_{j^*}(y + e^{j^*}, w) \quad (6.98)$$

using (6.94). Now if $y \in K_t$, then $y + e^{j^*} \in K_{t+1}$, the illegitimate case $y_{j^*} = w_{j^*}$ causing no problem since in that case $R_{j^*}(y + e^{j^*}, w) = 0$. Moreover, every $y \in K_{t+1}$ with $y_{j^*} > 0$ is equal to $y' + e^{j^*}$ for some $y' \in K_t$, and if $y_{j^*} = 0$ then $R_{j^*} = 0$. Hence

$$\sum_{y \in K_r} R_{j^*}(y + e^{j^*}, w) = \sum_{y \in K_{r+1}} R_{j^*}(y, w) \quad (6.99)$$

and thus

$$F(t, \hat{z}) = \sum_{r=0}^t \sum_{y \in K_r} Q_{j^*}(y, w) + \sum_{r=1}^{t+1} \sum_{y \in K_r} R_{j^*}(y, w). \quad (6.100)$$

Substituting (6.96) into (6.95) and subtracting from (6.100) then gives

$$F(t, \hat{z}) - F(t, z) = \sum_{y \in K_{t+1}} R_{j^*}(y, w) - \sum_{y \in K_0} R_{j^*}(y, w). \quad (6.101)$$

However $y \in K_0$ only if $y_{j^*} = 0$, in which case $R_{j^*}(y, w) = 0$. Thus (6.101) gives a positive quantity, proving the theorem. QED

Given a marginal configuration z , the rejection region will again be represented by the trimmings of the conditional tree above a certain height. Theorem 6.30 then assures that this height can only increase for the offspring of this conditional configuration. Thus to determine the rejection region for the offspring, we only need to determine whether the weight of the trimmings in the parent is still less than α under the conditional distribution of this offspring. If so, then the critical region is unchanged, otherwise we increment the critical value, and test again.

This algorithm can actually be improved. The component trees of the conditional tree corresponding to the first and last rows of (6.33) are trivial, and thus the conditional trees grafted onto any marginal nodes with the same value of the sum $z_0 + z_l$ are isomorphic and have the same probabilities. Thus the marginal tree of configurations $(z_0, \dots, z_l)$ in $l+1$ cells can be reduced to one in l cells: $(z_0^*, \dots, z_{l-1}^*)$ where $z_0^* = z_0 + z_l$ and $z_i^* = z_i$ for $i = 1, \dots, l-1$.

Even greater economies are possible in the balanced case $\nu = 1/2$. Here the permutation distribution is invariant under the group that, in selected litters, interchanges simultaneously control with test animals and response and non-response. Using the arguments of Section 6.2.5 the invariant subalgebra can be further reduced to one involving configurations in only $l/2$ cells. The details of this reduction will be reported elsewhere.

6.6 Summary

In this chapter, we have examined the structure of permutation tests, and have indicated how efficient algorithms could be constructed to determine the unconditional operating characteristics of such procedures in certain cases commonly occurring in toxicological bioassays. The discussion draws heavily on the methods developed in the previous two chapters.

The simplest case involves the Fisher-Irwin test, which arises as the permutation test for a completely randomized design with binary response variables. Utilizing a stochastic monotonicity property of the hypergeometric distribution, we were able to describe an algorithm for calculating the exact power of this test under the assumption of independence of control and test responses. Since this algorithm is of no more than linear complexity in the sample size, it provides a practical tool for calculating power in the case of moderately sized experiments with low response rates, a situation often handled by various approximations of unknown accuracy.

The justification of permutation tests in general depends on conditioning the data on an algebra of events invariant under a group of permutations. The calculation of the resulting probabilities can be facilitated by recognizing symmetries. It was found that the structures of the permutation groups under litter matched and nested designs themselves displayed a symmetry that allowed a reduction of the conditioning algebra.

The computation of power in situations allowing for intralitter correlation was facilitated by decomposing the probability of rejection into a sequence of conditional probabilities defined on a filtration of algebras. Such a decomposition can then exploit the structure of grafted trees, and the conditional probabilities of rejection can be calculated efficiently using the strategies of recursive computation, truncation of negligible tails, and weighing of complete branches. The rejection region itself can be determined recursively, since a stochastic monotonicity property analogous to that of the hypergeometric distribution was seen to hold for the permutation distributions arising from the nested and litter matched designs as well. Thus algorithms for determining the power of permutation tests under completely randomized, nested, and litter matched designs can readily be implemented. Although we have proposed the beta binomial distribution as the basis for modelling intralitter correlation, the algorithms discussed do not really depend on this model. They would work equally well for any response distribution in which litters were independent and animals within a litter were exchangeable.

Chapter 7

An Empirical Evaluation of Decision Rules

In this final chapter, we will examine the analysis of a data base of carcinogenicity studies. The initial purpose of this analysis was to see whether statisticians given no information other than the data would reach the same conclusions as were reached by regulatory authorities. However, our present purpose in this data base is to see what empirical light it can shed on questions of assessing false positive and false negative rates.

Various technical details and complications about the analysis are described elsewhere [14,85], and we will not repeat them here. We merely summarize the essential aspects of the data base, and propose four possible decision rules for consideration. After a brief look at the possibility of incorporating historical information, we discuss methods for estimating the error rates of these decision rules as applied to data bases of this type.

7.1 The Data

The data were received in coded form from S. Stanley Young of the Greenfield Laboratories of Eli Lilly & Co., who had randomly selected bioassays on 25 compounds which had been tested as part of the Carcinogenesis Bioassay Program sponsored by the United States National Cancer Institute and National Toxicology Program. For each compound, there were two bioassays, one on rats, and one on mice, and each bioassay included both sexes. These records were sent to three statisticians: Thomas A. Louis[57], C. T. Lin[70], and Daniel Krewski[14], who represented members of academic institutions, the pharmaceutical industry, and government regulatory bodies, respectively. The bioassays were identified by species, but the identity of the compounds was concealed, as was the correspondence between rat and mouse bioassays on the same compound. The object was to see what concurrence might be found between these statisticians working with the same data and whether the rat and mouse studies yielded similar conclusions.

The data for each study consisted of a number of records. Each record contained information about a particular response, with separate records being presented for male and for female animals. For each treatment, the record stated the number of animals examined and the number displaying that particular response. The responses were of

three types: specific lesions, aggregated lesions, and survival. Each record of the first type contained information about a pathologically specific tumour occurring in a particular tissue. An average of twenty such records per study were reported for each sex, although the actual number varied from as low as 6 to as high as 43. There were two records for each sex of aggregated lesions, one giving the number of animals in each treatment suffering from any benign tumour, and the other giving the number suffering from any malignant tumour. The survival records gave the number of animals in each treatment that survived to the termination of the study.

There are two deficiencies in this data set that could cause problems in data analysis: the lack of individual animal records, and the inadequacy of the information on survival. The response in a carcinogenicity bioassay is an essentially multivariate one, with the incidence and lethality of various lesion types being intimately related both with each other and with survival. In this set of data, only the *marginal* responses were available, so there was no possibility of examining this interrelationship. Although there was information provided that could indicate the presence of dose dependent mortality, this information was not sufficient to properly accommodate such a dependency, should it exist. There was no way of determining how many of the tumour-bearing animals were among the survivors, and how many were among the early decedents. Moreover, the available mortality data made no distinction between those animals that died one week after study initiation, and those that died one week before study termination: they were both considered equivalently as "early deaths".

Although the best way of handling these deficiencies would be to avoid them by demanding more complete data, we will here consider the data to be given, and will discuss how these deficiencies affect the selection of a decision procedure, its operating characteristics and the interpretation of the results. The first concern is that the lack of time-to-tumour data may reduce the sensitivity of the experiment to detect oncogenicity. For the sake of this analysis, we assumed that loss of efficiency was not a problem. There is some theoretical support for this assumption [12,32], although it is by no means definitive. Even if procedures based on crude incidence data alone are reasonably efficient, however, there is still the possibility of bias if the test compound induces early mortality. In the present data there are two possible indications of such bias. The first, of course, is the presence of dose related mortality, the other is a non-monotone dose response curve in tumour occurrence, in which the incidence at the high dose is actually less than that at intermediate doses. One explanation of such a shape would be the possibility that there was an excess of early deaths at the high dose which reduced the effective number of animals at risk to oncogenesis. Of course, detailed mortality records could elucidate the cogency of such an explanation, but lacking such information, prudence demands that this possibility should not be discounted.

The second concern is that the multiplicity of responses may lead to an excessive false positive rate if decisions of carcinogenicity are made on the basis of dose related effects in any response marginally. This concern is complicated by the fact that the number of response types is itself a random variable. There are no records for responses which were not observed in any animals in a study, so that the data records represent a truncated sample of lesion types. Thus determination of false positive rates must take into account the potential number of responses analyzed, rather than just the number

actually analyzed in a given study. When only marginal data are available, the simplest adjustment for multiplicity is by the Bonferroni bound, in which the required significance level is divided by the number of significance tests made. However, such an adjustment in the face of a large number of response types can lead to extreme conservatism, especially since significance tests with binary data tend to be conservative to begin with. One would not want to discount a clearly significant effect in one lesion just because there were forty other lesion types noted as appearing in some animals, and therefore being available for significance testing. Indeed, the notion of "type of tumour" is not even clearly defined, since one pathologist may group together lesions that are related by either morphology or etiology, whereas another might record them as distinct types. Thus the possible number of response types is limitless. An attempt by a pathologist to be more specific in his diagnosis might be counterproductive in this context, for it could lead in the extreme case to a situation in which each lesion is considered to be distinctive, resulting in many response types with only a single animal responding according to any one type. Any marginal analysis of such data would fail to detect any significance, no matter how clear an oncogenic effect was present.

7.2 Analysis of the Data

Bearing in mind the foregoing discussion, we present four decision rules for deciding on the classification of each compound. Since the correspondence between rat and mouse studies was not available, we considered the studies on the different species as testing different compounds, thus making separate classifications for each. In practice, one would combine information across species to make decisions about carcinogenicity.

7.2.1 Tests for trend

Because the bioassays had treatments on several dose levels, all the decision rules were based on tests for trend in the responses. The rules differed in the manner in which they combined information across response types, as we will describe presently. But first, let us examine the trend test itself.

Given an experiment with $k + 1$ treatments on increasing doses $d_0, d_1, \dots, d_k$, let x_i be the number of animals on treatment i responding out of a total number of n_i observed in that treatment. The decisions were made according to the statistic $T = \sum d_i x_i$. This statistic (or statistics equivalent to it) was proposed by Cochran[19] and Armitage[2] on heuristic grounds. It can be derived under the model of a logistic tolerance distribution in which case T and $p = \sum x_i / \sum n_i$ are jointly sufficient [23]. More recently, Gart and Tarone[79] have studied this statistic under a general class of dose response models.

If the experiment was performed as a completely randomized design, then we can test the hypothesis of no treatment effect by the randomization distribution. The permutation group here would be all permutations of animal labels, and the invariant subalgebra would be that induced by $\bar{p}$. We thus consider the conditional distribution of T given $\bar{p}$. The

conditional distribution of $(x_0, \dots, x_k)$ is multivariate hypergeometric

$$\Pr\{X_i = x_i | p = x/n\} = \prod_{i=0}^k \binom{n_i}{x_i} / \binom{n}{x} \quad (7.1)$$

where $n = \sum n_i$ and $x = np = \sum x_i$, and is thus identical to that in the nested design with two treatments (Section 6.3.2), although of course the symbol x_i in (7.1) has a different meaning from y_i in (6.29). Thus although the permutation groups are in fact different, the conditional distribution for this design can be calculated by algorithms similar to those used for the nested design. In fact, if the dose levels are equally spaced at unit intervals, the permutation distribution of T under the completely randomized design is identical to that of T_1 , the total number of responders, under the nested design. If the doses are not equally spaced, one can usually make them equal by inserting phantom treatments having no observations. This artifice will only fail if the experimenter perversely decided to employ dose spacings at irrational ratios. It should be noted that the algorithm of Section 5.3.1 will work even in this case. Only the lateral tree algorithm will fail. (Interestingly enough, the permutation distribution of T also arises from purely likelihood arguments, assuming independent observations from a logistic tolerance distribution [23].)

Let P_e be the exact P -value determined from the distribution (7.1). In spite of efficient algorithms for its calculation, it was felt that it would be computationally excessive to determine it for each of the more than two thousand records in the data base. An approximation based on the asymptotic normality of T was calculated by defining

$$t_u = \frac{\sqrt{n}(T/n - \bar{d}p)}{\sqrt{p(1-p)\sigma_d^2}} \quad (7.2)$$

where $\bar{d} = \sum n_i d_i / n$ and $\sigma_d^2 = \sum n_i (d_i - \bar{d})^2 / n$ are respectively the mean and variance of the doses. An approximate P -value was then calculated as $P_u = \Phi(-t_u)$, (where Φ is the standard normal c.d.f.). Empirically, we have found that a continuity correction improves the approximation. If Δ is the largest spacing between doses we also calculated a corrected statistic

$$t_c = \begin{cases} \frac{\sqrt{n}\left(\left(\frac{T}{n} - \bar{d}p\right) - \frac{\Delta}{2n}\right)}{\sqrt{p(1-p)\sigma_d^2}} & \text{if } \left(\frac{T}{n} - \bar{d}p\right) > \frac{\Delta}{2n} \\ \frac{\sqrt{n}\left(\frac{T}{n} - \bar{d}p\right)}{\sqrt{p(1-p)\sigma_d^2}} & \text{if } -\frac{\Delta}{2n} \leq \left(\frac{T}{n} - \bar{d}p\right) \leq \frac{\Delta}{2n} \\ \frac{\sqrt{n}\left(\left(\frac{T}{n} - \bar{d}p\right) + \frac{\Delta}{2n}\right)}{\sqrt{p(1-p)\sigma_d^2}} & \text{if } \left(\frac{T}{n} - \bar{d}p\right) < -\frac{\Delta}{2n} \end{cases} \quad (7.3)$$

and a corresponding corrected P -value

$$P_c = \Phi(-t_c). \quad (7.4)$$

Since it has been found that in most cases $P_u \leq P_e \leq P_c$ if $P_c < \frac{1}{2}$, we adopted the following strategy: For a test at level α in which $np \geq 10$, we declared the result to be significant if $P_c \leq \alpha$ and not significant if $P_u > \alpha$. The exact P -value P_e was only computed if $P_u \leq \alpha < P_c$ or if $np < 10$. In those cases the result was declared significant if $P_e \leq \alpha$.

The above procedure was applied to all the records in the data base — i.e., all reported responses, on both sexes, on all studies. The choice of level α depended on the decision rule and the response as described later. First, we consider the possibility of combining information across studies.

7.2.2 Incorporating historical information

The control animals are ideally on the same treatment for all studies. Thus, information about the background incidence of any tumour is present in all the studies, not just the one being tested. However, since the studies were not identical, it is possible that the background rate might vary from study to study.

Suppose that the background incidence rate of a particular tumour is a random variable following a beta distribution, with parameters α, β , and that for any study, the true probability of a control animal getting that tumour is a realization of that random variable. In such a context, it makes sense to replace the maximum likelihood estimator of the control probability $\hat{p}_0 = x_0/n_0$ by the Bayes posterior mean

$$\tilde{p}_0 = \frac{x_0 + \alpha}{n_0 + \alpha + \beta} \quad (7.5)$$

Tarone[78] suggested replacing the control incidence x_0 by $x_0 + \alpha$ and n_0 by $n_0 + \alpha + \beta$ in calculating the test statistic t_u (7.2), although his justification was not based on a Bayesian argument. In practice, the parameters α and β are not known, but given data $(x_0^1, n_0^1; \dots; x_0^m, n_0^m)$ on control groups from m studies, it is possible to estimate them by maximum likelihood. It is possible for the estimates $\hat{\alpha}, \hat{\beta}$ to be such that $n_0 + \hat{\alpha} + \hat{\beta}$ exceeds the total number of control animals (and might even be infinite). In the limiting case $\alpha, \beta \rightarrow \infty$ the set of control groups becomes a set of binomial samples with the same parameter p_0 , whose best estimate would be X_0/N_0 where $X_0 = \sum x_0^i$ and $N_0 = \sum n_0^i$. So as not to overstate the information provided by the other control groups, the estimates should be truncated to $\tilde{\alpha} = \tau \hat{\alpha}, \tilde{\beta} = \tau \hat{\beta}$ where $\tau = \min(1, \frac{N_0 - n_0}{\alpha + \beta})$.

This analysis incorporating historical controls was performed on the data, except that we used the corrected t -statistic t_c (7.3) instead of t_u . This modified statistic we call t_h and the corresponding P -value P_h . The parameters α and β were estimated from all the records on the corresponding tumour in the same sex and species, where absent records were assumed to indicate a background incidence of zero. For sake of both estimation and interpretation, the beta distribution was reparametrized in terms of the mean incidence $\mu = \alpha(\alpha + \beta)^{-1}$ and the intra-study correlation $\rho = (1 + \alpha + \beta)^{-1}$. If $X_0 = 0$ then the maximum likelihood estimators are not unique, but we set $\hat{\mu} = \hat{\rho} = 0$.

7.2.3 Decision rules

To accommodate the possibility that early mortality precluded the occurrence of tumours in the high-dose group, we also tested the dose-response curve for non-linearity. Letting

$$\chi_h^2 = \sum_i \frac{(x_i - n_i p)^2}{n_i p (1 - p)} \quad (7.6)$$

be the usual statistic for heterogeneity, we define $\chi_{nl}^2 = \chi_h^2 - t_u^2$. If χ_{nl}^2 exceeded the upper .05 quantile of the χ^2 distribution with $k - 1$ degrees of freedom, we admitted the possibility that a reversal in the dose response curve might be due to early mortality (possibly undetectable from the survival summary), and repeated the original test after omitting the highest dose group.

The tumour types could be divided into two levels of severity: benign and malignant. We took this distinction into account in classifying the compounds, distinguishing between *carcinogens*, for which there was evidence of inducing malignant lesions, and *oncogens* which only appeared to cause benign lesions. In addition, there was a third "inconclusive" category, to which we assigned studies in which there was no clear indication of *carcinogenicity*, but that did display dose-dependent mortality which might have masked a carcinogenic effect.

In summary, each decision rule classified the compound in each study into one of four ordered categories:

1. Non-oncogen, if there was no evidence of any effect;
2. Oncogen, if there was evidence of induction of some benign tumour but no evidence of dose-dependent mortality nor of malignant tumour induction;
3. Inconclusive, if there was no evidence of carcinogenicity but there was evidence of dose-dependent mortality; and
4. Carcinogen, if there was evidence of induction of some malignant lesion.

The decision rules differed in the manner in which they combined information across the different types. Let P_j be the P -value used for the j^{th} response. (This might be P_s , P_u , P_c , or P_h .) Then the rules used the following definitions of "evidence of tumour induction".

Rule I: If $P_j < .05$ for any specific or aggregated lesion.

Rule II: If $P_j < .05$ for aggregated records. Ignore data on specific lesions.

Rule III: If $P_j < \alpha_j$ for any specific lesion where α_j depends on the rarity of the lesion as estimated from all the control groups on the same sex and species, according to Table 7.1. Data on aggregated lesions is ignored.

Rule IV: If $P_j < \alpha_j/2$ for specific lesions, or $P_j < .025$ for aggregated lesions.

Under any rule, the compound was classified for each sex, and if the classifications differed, was placed in the highest category. The distinctions of the four decision rules were not applied to the tests for dose-dependent mortality or non-linearity. These were always tested at $\alpha = .05$.

Table 7.1: Nominal significance levels for Rule III.

$\hat{\mu}_j$	α_j
0-.001	.20
.001-.002	.10
.002-.005	.05
.005-.010	.02
.010-.020	.01
>.02	.001

Table 7.2: Number of studies in each classification using the four decision rules.

	Classification			
	Non-oncogen	Oncogen	Inconclusive	Carcinogen
Rule I	7	5	9	28
Rule II	15	3	13	16
Rule III	12	5	11	21
Rule IV	16	2	11	20

7.2.4 Results

Table 7.2 gives the number of studies placed in each classification by each of the four rules when the historical control adjustment is not made. As would have been expected, Rule I is the most eager to find carcinogenicity, whereas Rule II is the most conservative. Table 7.3 gives a partial cross-classification. We see that there is a moderate agreement among the rules, and that Rule II is not necessarily always conservative. There were, for example, three compounds classified as carcinogens by Rule II which were not so classified by Rule III.

When the analysis was done including historical controls, almost all the compounds were classified as carcinogens by rules I, III, and IV. Closer inspection revealed that many of these classifications were due to a single animal on the high-dose treatment having a malignant tumour of a type that had never occurred in any control group. Substituting this outcome into Tarone's procedure inevitably led to a significant result. One should not give too much credibility to this finding, however, since the procedure is only valid asymptotically and might not give reasonable P -values for points so close to the boundary of the parameter space.

One might, though, consider this same situation from an exact testing perspective: Consider 25 studies such that there is no study to study variation ($\rho = 0$), so that control groups can be legitimately pooled. For simplicity, suppose also that there are two dosed groups each having the same number of animals as each of the control groups. If now a particular tumour were not to appear in any control group, but appeared in a single animal in the high-dose group of one study, then for this study this tumour would give $P_0 = \frac{1}{2^5} < .05$ if the control groups were combined. Thus we see that the idea of a single tumour leading to statistical significance is not as preposterous as it may sound to some practitioners.

Table 7.3: Crosstabulation of classifications using the four decision rules.

	Classification	Classification by Rule II			
		Non-oncogen	Oncogen	Inconclusive	Carcinogen
by Rule I	Non-oncogen	7	0	0	0
	Oncogen	4	1	0	0
	Inconclusive	0	0	9	0
	Carcinogen	4	2	4	16
by Rule III	Non-oncogen	10	0	0	2
	Oncogen	3	1	0	1
	Inconclusive	0	0	11	0
	Carcinogen	2	2	2	13
by Rule IV	Non-oncogen	14	1	0	1
	Oncogen	0	2	0	0
	Inconclusive	0	0	11	0
	Carcinogen	1	0	2	15

We must consider in this context, however, the role of multiplicity. Rare tumours will occasionally occur, and may, in a database of this size, occur only once. They might well appear in the high dose quite fortuitously, leading to a spurious finding of significance. Although such tumours are individually rare, because of the large number of possible types, it is highly likely that *every* study will contain some such tumours. On the other hand, one must not lightly dismiss the possibility that the test chemicals may in fact induce, with a small probability, non-specific rare tumours. The existence of such effects would be of particular concern to regulatory agencies, for they would be totally undetectable by conventional screening procedures; yet still might have a non-negligible impact on a large population.

We decided to see if the pattern of rare lesions in a large database such as this gave any evidence of such "sub-critical" effects. All the records reporting but a single animal with its particular type of tumour were extracted. It turned out that almost *half* the records were of this sort. If there were a tendency for weak, non-specific oncogenic effects in rare tumours, one would expect to find more of these rare tumours in the high dose group than in the controls. However, we found that the frequencies with which these tumours appeared in the various treatments were nearly proportional to the group sizes. It therefore appears that these rare lesions were occurring randomly, quite irrespective of the treatment. Thus multiplicity seems a better explanation of the carcinogenic classification than the presence of weak oncogenic effects.

For Rule II the issue of multiplicity is not of such great concern, since only aggregate records are involved. Moreover, since the background rates are relatively high, the parameters are farther from the boundary, and the asymptotic theory is probably more appropriate. Table 7.4 gives a summary of the effects of incorporating historical information for Rule II. Note that in only 5 out of 47 studies were classifications changed by incorporating historical information, and in four of these, the change was towards a higher category of carcinogenicity.

Table 7.4: Crosstabulation of classifications using Rule II with and without historical controls.

Ignoring Historical Controls	Including Historical Controls			
	Non-oncogen	Oncogen	Inconclusive	Carcinogen
Non-oncogen	13	0	0	2
Oncogen	0	2	0	1
Inconclusive	0	1	11	1
Carcinogen	0	0	0	16

7.3 Estimation of False Positive Rates

In interpreting the result of a bioassay, it is important to know the actual *size* or *false positive rate* of the procedure. In addition to the significance level chosen by a test, the false positive rate is influenced by two opposing tendencies—multiplicity and conservatism.

Suppose that a decision involves r hypothesis tests, where the i^{th} test is performed at level α_i . Then, by the Bonferroni inequality, the size of the overall procedure is bounded above by $\min(\sum_{i=1}^r \alpha_i, 1)$. This bound is sharp in the sense that it will be achieved if the tests are non-conservative and their rejection regions are mutually exclusive. In practice, such a situation is unlikely to arise, although it may be approached if the occurrence of one type of lesion precludes the occurrence of another type [69]. However, for nonconservative tests, the bound is still relatively tight in the case of stochastically independent tests, provided that $\sum \alpha_i$ is not too large. Applying the Bonferroni inequality to the rejection regions C_i gives

$$\sum_i \Pr(C_i) - \sum_{i < j} \Pr(C_i \cap C_j) \leq \Pr\left(\bigcup_i C_i\right) \leq \sum_i \Pr(C_i). \quad (7.7)$$

Under the null hypothesis, $\Pr(C_i) = \alpha_i$, and if the tests are independent,

$$\sum \alpha_i - \Pr\left(\bigcup_i C_i\right) \leq \sum_{i < j} \alpha_i \alpha_j \leq \frac{(\sum \alpha_i)^2}{2}.$$

If we decide to perform a significance test on each tumour type *observed* in a study, then the number of tests performed is a random variable. Letting a_j be the indicator of the event that tumour type j is observed, we will perform $R = \sum_j a_j$ tests. The domain of this summation is the set of all possible tumour types. Since this set is not clearly defined we will not even assume that it is finite.

7.3.1 Adjusting for multiplicity

Let T be the set of all possible tumour types. Then the outcome for a particular animal will be a map $y : T \rightarrow \{0, 1\}$ where $y_j = 0$ or 1 depending, respectively, on whether the animal does or does not develop tumour type j . We consider the y_j 's to be random variables,

with $E(y_j) = \mu_j$, and although we make no direct assumptions about the cardinality of T , we do assume that

$$\sum_{j \in T} \mu_j = M < \infty.$$

This restriction obviously implies that the number of tumours in any animal is (almost surely) finite.

Now, let a_j be the indicator of the event that tumour type j occurs in at least one animal in a particular study with n animals. Then the number of tumours observed is $R = \sum_{j \in T} a_j$. Moreover,

$$E(a_j) = (1 - (1 - \mu_j)^n) \leq n\mu_j \quad (7.8)$$

so that

$$E(R) \leq n \sum \mu_j = nM. \quad (7.9)$$

There are several ways of dealing with the multiplicity of tests. The most obvious is to test conditionally on the number of tumour types observed. Thus, if each tumour type is tested at level α/R , then the overall false positive rate will be bounded by α . This approach is what is implied by proponents of "Bonferroniized" significance levels. There is, however, something unsatisfying about making the significance of one tumour type depend on the appearance of other, possibly rare tumours. It comes down to saying that the more different types of tumour are found, the less significant that makes an apparent dose response in any particular type. With this strategy, a pathologist could easily render a study totally negative merely by meticulously considering each observed tumour to be a distinctive type.

Another approach is to put a universal bound on the significance level regardless of the number of tumour types actually observed. From (7.9) we see that testing every tumour at $\alpha/(nM)$ would again bound the overall false positive rate by α , regardless of the number of tumours tested. However, to implement this procedure, we require an estimate of M , the mean number of different tumour types per control animal. Such an estimate could be readily obtained from historical data.

This second approach, however, is still unsatisfactory, for it would demand rather high significance, and moreover, would require higher significance of larger studies. The high significance required would probably be hard to achieve for rare tumours, which might in fact be the very indicators of carcinogenic activity. It is a common belief among toxicologists that dose-related effects in rare tumours should be ascribed more significance than those in common tumours, regardless of what the statistics say. In what follows, we will justify this view, and will incorporate it into the decision procedure.

It is well known that conditional tests with binary data can be highly conservative. This conservatism has been noted by Gart et al.[31] and Haseman[26], for the Fisher-Irwin test. Our own computations have found the same to be true for the trend test. (Portier and Hoel[66] found the same to be true by Monte-Carlo methods.) We must thus distinguish between the nominal significance level α , and the true size of the test, which we will denote by φ .

The critical regions of the trend tests discussed in Section 7.2.1 depend on the significance level, the test employed (exact, corrected, or uncorrected), and also on the mean

observed incidence p . Given a significance level α , one can calculate, for a particular design and test statistic, the conditional size

$$\varphi_p = \Pr \{T \geq c_\alpha | p\} \quad (7.10)$$

using the null distribution (7.1). The unconditional size φ is then the expected value of φ_p over the distribution of p .

It should be noted that $\varphi_p \leq \alpha$ for the exact test, and this inequality should also hold, approximately even for asymptotic tests. Moreover, φ_p can in fact be much smaller than α for some values of p . (In the extreme case $p = 0$, φ_p is necessarily zero.) Generally speaking, φ_p will be an increasing function of μ , the expected value of p , although it will, to some extent, also depend on other aspects of the distribution. This relationship is the rationale for the "sliding scale" choice of significance levels used by Rules III and IV. In principle, one could define a rule based on a continuous relationship between background incidence μ , and significance level α . However, we propose a step function as a practical alternative.

7.3.2 A general formulation

For a given design and decision rule, let φ_j be the true unconditional size of the test for trend for tumour type j . Then the false positive rate for this experiment is bounded above by $\varphi^* = \sum_{j \in T} \varphi_j$. However, even if the true sizes φ_j were known, one could not calculate this bound since the set T is not known. If one calculates the sum of the φ_j over the lesions actually observed in an experiment, then we have the estimate $\sum a_j \varphi_j$ which is less than φ^* .

In the following, our attitude will be that the false positive rate is a parameter of the population of bioassays from which the database in question is a random sample. We could use a Horvitz-Thompson estimator [21] to estimate φ^* by

$$\hat{\varphi}^* = \sum_j a_j \frac{\varphi_j}{p_j} \quad (7.11)$$

where $p_j = E(a_j)$. The Horvitz-Thompson approach gives an unbiased estimator, but, unfortunately, neither p_j nor φ_j are known. In our case, they can, however, be estimated from the data. We estimated the true size φ_j by taking the expectation of (7.10) over p , postulating that for any particular response in a given sex and species np has a beta binomial distribution. The parameters of this distribution were estimated from the set of control group incidences in the data base, as discussed in Section 7.2.2. The probability of observation, p_j can be estimated by the proportion of studies having records for this response.

Let $\{X_i : i \in T\}$ be a family of random variables, and let $\Theta = \{\theta_i : i \in T\}$ be a (multidimensional) parameter of their joint distribution. Given estimates $\hat{\theta}_i$ of the θ_i , our modest goal is to estimate the sum

$$\theta^* = \sum_{i \in T} \theta_i.$$

However, the problem is complicated by the fact that some of the estimators $\hat{\theta}_i$ are randomly invisible, so that we only have estimators for $i \in S \subset T$, where S is a random finite subset of T . Let a_i be the indicator of the event $i \in S$, and define $p_i = E(a_i)$. Then the problem could be rephrased to say that we observe not $\hat{\theta}_i$ but $a_i \hat{\theta}_i$.

We will propose some estimators for this situation and will examine their properties under a variety of assumptions. We will then try to apply the insight so gained to our actual situation.

The first proposal is the *naïve estimator*

$$\hat{\theta}_1^* = \sum_{i \in S} \hat{\theta}_i = \sum_{i \in T} a_i \hat{\theta}_i. \quad (7.12)$$

Consider

Assumptions A: The $\hat{\theta}_i$ are unbiased and a_i and $\hat{\theta}_i$ are independent. (This includes the case that the $\hat{\theta}_i$ are *known*.)

Under Assumptions A

$$E(\hat{\theta}_1^*) = \sum p_i \theta_i \leq \theta^* \quad (7.13)$$

so that this estimator is biased. We could get an unbiased estimator by employing the Horvitz-Thompson technique to get

$$\hat{\theta}_2^* = \sum a_i \frac{\hat{\theta}_i}{p_i}, \quad (7.14)$$

but this estimator requires knowledge of the p_i . Suppose, however, that we have a *random sample* of the indicator variables a_i — i.e., let $\mathbf{a}_k = \{a_{ki} : i \in T\}$, $k = 1, \dots, m$ be a set of independent identically distributed sequences, where we identify the sequence $\{a_i, i \in T\}$ above with $\mathbf{a}_1$. (We mean here that the sequence $\mathbf{a}_k$ is independent of the sequence $\mathbf{a}_{k'}$ for $k \neq k'$. It is *not* assumed that the random variables a_{ki} , $i \in T$ are either independent or identically distributed.) In this case, we can estimate p_i by $\hat{p}_i = m^{-1} \sum_{k=1}^m a_{ki}$. Consider now the estimator

$$\hat{\theta}_3^* = \sum_{i \in S} \frac{\hat{\theta}_i}{\hat{p}_i} = \sum_{i \in T} a_i \frac{\hat{\theta}_i}{\hat{p}_i}. \quad (7.15)$$

The i^{th} term in the sum is undefined if $\hat{p}_i = 0$. However, this can only happen if $i \notin S$, so we adopt the convention that $a_i/\hat{p}_i = 0$ if $\hat{p}_i = 0$.

To examine the properties of $\hat{\theta}_3^*$, we require the following lemma and its corollary. Recall from Section 6.2 the idea of a group G of measurable permutations of a probability space $(\Omega, \mathcal{A}, P)$. As discussed in Section 6.2.2, this group also acts on the set of random variables.

Lemma 7.1 Let G be a group of measurable permutations of a probability space $(\Omega, \mathcal{A}, P)$ where P is G -invariant, and let a subalgebra $\mathcal{B}$ be contained in the subalgebra of G -invariant events. Then for any random variable Y , and $g \in G$,

$$E(gY | \mathcal{B}) = E(Y | \mathcal{B}). \quad (7.16)$$

Proof: Let $f = E(gY | \mathcal{B})$. By definition, for any $B \in \mathcal{B}$,

$$\int_B f dP = \int_B gY dP = \int_{g^{-1}B} Y dP g = \int_B Y dP \quad (7.17)$$

by the assumption of invariance. But this is just the defining property of $E(Y | \mathcal{B})$. QED

Corollary 7.1 Let $Y_1, \dots, Y_m$ be exchangeable random variables. Then

$$E\left(Y_i \mid \sum_j Y_j\right) = m^{-1} \sum_j Y_j.$$

Now, returning to the estimator $\hat{\theta}_3^*$, we note that not only $E(\hat{p}_i) = p_i$, but also $E(a_{1i} | \hat{p}_i) = \hat{p}_i$ by Corollary 7.1. However, these properties do not make $\hat{\theta}_3^*$ unbiased. For we have

$$E\left(\frac{a_{1i}}{\hat{p}_i} \mid \hat{p}_i\right) = \begin{cases} 1 & \text{if } \hat{p}_i > 0 \\ 0 & \text{if } \hat{p}_i = 0 \end{cases} \quad (7.18)$$

giving

$$E(\hat{\theta}_3^*) = \sum_{i \in T} \tau_i \theta_i \leq \theta^* \quad (7.19)$$

where $\tau_i = \Pr\{\hat{p}_i > 0\}$.

Instead of Assumptions A, let us now consider

Assumptions B: The $\hat{\theta}_i$ are unbiased and $a_{1i} = 0 \implies \hat{\theta}_i = 0$, but $\hat{\theta}_i$ is independent of a_{kj} for $k \neq 1$.

Under these assumptions, the invisibility of a parameter implies that its estimate is equal to zero. A consequence of Assumptions B is that in fact $a_i \hat{\theta}_i = \hat{\theta}_i$ and thus

$$E(\hat{\theta}_1^*) = \theta^*.$$

We seem to have come full circle, with the "naïve estimator" now being unbiased. We will not abandon the empirical Horvitz-Thompson estimator (7.15), however, for it will turn out to be useful under a third set of assumptions.

To continue the discussion we require the following results.

Lemma 7.2 If x is a random variable and a is a binary random variable such that $a = 0 \implies x = 0$ then $E(x | a) = \frac{aE(x)}{E(a)}$.

Proof:

$$\begin{aligned} E(x) &= E(E(x | a)) \\ &= E(a)E(x | a = 1) + (1 - E(a))E(x | a = 0). \end{aligned}$$

But $E(x | a = 0) = 0$, so $E(x | a = 1) = E(x) / E(a)$. Thus

$$E(x | a) = \frac{aE(x)}{E(a)}.$$

QED

Lemma 7.3 Let C be a subalgebra of the underlying probability space, a a binary random variable, and x a random variable such that $a = 0 \implies x = 0$. Suppose that x is conditionally independent of C given a . Then

$$E(x|C) = \frac{E(x)E(a|C)}{E(a)}.$$

Proof:

$$E(x|C) = E(x|C, a=1)E(a|C) = E(x|a=1)E(a|C).$$

But by the preceding lemma, $E(x|a=1) = E(x)/E(a)$. QED

Theorem 7.1 Under Assumptions B

$$E(a_{1i}\hat{\theta}_i | \hat{p}_i) = \frac{\hat{p}_i}{p_i}\theta_i.$$

Proof: The proof follows immediately from the application of Lemma 7.3. QED

We now consider the third set of assumptions.

Assumptions C: $\hat{\theta}_i = m^{-1} \sum_{k=1}^m \hat{\theta}_{ki}$, where $\hat{\theta}_{ki}$ is unbiased for θ_i , $\hat{\theta}_{ki}$ is independent of $a_{k'i}$ and $\hat{\theta}_{k'i}$ for $k' \neq k$, but $a_{ki} = 0 \implies \hat{\theta}_{ki} = 0$.

Theorem 7.2 Under Assumptions C,

$$E(a_{1i}\hat{\theta}_i | \hat{p}_i) = \frac{\hat{p}_i^2}{p_i}\theta_i.$$

Proof:

$$E(a_{1i}\hat{\theta}_i | \hat{p}_i) = \frac{1}{m} \sum_k E(a_{1i}\hat{\theta}_{ki} | \hat{p}_i). \quad (7.20)$$

Theorem 7.1 gives $E(a_{1i}\hat{\theta}_{1i} | \hat{p}_i)$. Let us now consider $E(a_{1i}\hat{\theta}_{ki} | \hat{p}_i)$ for $k \neq 1$. First, notice that

$$E(a_{1i}\hat{\theta}_{ki} | \hat{p}_i) = E(a_{1i} | \hat{p}_i) E(\hat{\theta}_{ki} | \hat{p}_i, a_{1i} = 1).$$

Now, the event

$$\{\hat{p}_i = r/m\} \cap \{a_{1i} = 1\}$$

is the same as

$$\left\{ \sum_{k \neq 1} a_{ki} = r - 1 \right\} \cap \{a_{1i} = 1\}.$$

Moreover, $\hat{\theta}_{ki}$ is independent of a_{1i} so we have

$$\begin{aligned} E(\hat{\theta}_{ki} | \hat{p}_i = \frac{r}{m}, a_{1i} = 1) &= \frac{r-1}{(m-1)p_i} \theta_i \\ &= \frac{m\hat{p}_i - 1}{(m-1)p_i} \theta_i. \end{aligned}$$

We thus have

$$E(a_{1i}\hat{\theta}_i | \hat{p}_i) = \frac{m\hat{p}_i - 1}{m - 1} \frac{\hat{p}_i}{p_i} \theta_i \quad (7.21)$$

for $k \neq 1$. Substituting (7.21) in (7.20) we have

$$\begin{aligned} E(a_{1i}\hat{\theta}_i | \hat{p}_i) &= m^{-1} \left[\frac{\hat{p}_i}{p_i} \theta_i + (m\hat{p}_i - 1) \frac{\hat{p}_i}{p_i} \theta_i \right] \\ &= \frac{\hat{p}_i^2}{p_i} \theta_i. \end{aligned}$$

QED

Corollary 7.2.1 Under Assumptions C, the estimator $\hat{\theta}_i^*$ is unbiased.

Proof:

$$E\left(\frac{a_i \hat{\theta}_i}{\hat{p}_i}\right) = E\left(E\left(\frac{a_i \hat{\theta}_i}{\hat{p}_i} \middle| \hat{p}_i\right)\right).$$

If $\hat{p}_i > 0$ then

$$E\left(\frac{a_i \hat{\theta}_i}{\hat{p}_i} \middle| \hat{p}_i\right) = \frac{\hat{p}_i^2 \theta_i}{p_i \hat{p}_i} = \frac{\hat{p}_i \theta_i}{p_i}.$$

If $\hat{p}_i = 0$, $a_i \hat{\theta}_i / \hat{p}_i$ is interpreted as zero, but in that case $\hat{p}_i \theta_i / p_i = 0$. Thus

$$E\left(\frac{a_i \hat{\theta}_i}{\hat{p}_i} \middle| \hat{p}_i\right) = \frac{\hat{p}_i \theta_i}{p_i}$$

for all $\hat{p}_i$ and

$$E\left(\frac{a_i \hat{\theta}_i}{\hat{p}_i}\right) = \frac{E(\hat{p}_i) \theta_i}{p_i} = \theta_i.$$

Since each term in the sum (7.15) is unbiased, so is the sum.

QED

In our practical problem, the invisible parameters represent the false positive rates for lesions that were not observed. If these were considered, however, their incidence would be unbiasedly estimated as zero, and hence so would their false positive rate. Thus Assumptions B are reasonable here. If, however, we estimate the parameters of the distribution of p from data pooled across many studies, then Assumptions C are the relevant ones to use, and $\hat{\theta}_i^*$ would appear to be a reasonable estimator. The difficulty is, however, that the false positive rates are non-linear functions of the estimated parameters, and so we require a more elaborate scheme.

Assumptions D: $\hat{\theta}_i = \varphi_i(\hat{\mu}_i)$ where φ_i is an increasing function, convex at μ_i , with $\varphi_i(0) = 0$; $\hat{\mu}_i = m^{-1} \sum_{k=1}^m \hat{\mu}_{ki}$, and $\hat{\mu}_{ki}$ is unbiased for $\mu_i = \varphi_i^{-1}(\theta_i)$, $\hat{\mu}_{ki}$ is independent of $a_{k'i}$ and $\hat{\mu}_{k'i}$ for $k' \neq k$, but $a_{ki} = 0 \Rightarrow \hat{\mu}_{ki} = 0$.

Theorem 7.3 Under Assumptions D, $E(\hat{\theta}_i^*) \geq \theta^*$.

Proof: To prove the theorem, we prove the inequality for each index $i \in T$. For simplicity, we suppress the subscript i in the following.

Since φ is convex at μ , we can write

$$\varphi(\hat{\mu}) \geq \varphi(\mu) + \beta(\hat{\mu} - \mu)$$

for some number β . Thus

$$\begin{aligned} E\left(\frac{a\varphi(\hat{\mu})}{\hat{p}} \mid \hat{p}\right) &\geq E\left(\frac{a}{\hat{p}} \mid \hat{p}\right) (\varphi(\mu) - \beta\mu) + \beta E(a\hat{\mu} \mid \hat{p}) \\ &= (1 - \tau)(\varphi(\mu) - \beta\mu) + \beta\mu \\ &= \varphi(\mu) + \tau(\beta\mu - \varphi(\mu)), \end{aligned}$$

where, as before, $\tau = \Pr\{\hat{p} = 0\}$. Also, by convexity $\varphi(0) \geq \varphi(\mu) - \beta\mu$. But since $\varphi(0) = 0$, $\beta\mu - \varphi(\mu) \geq 0$. QED

7.3.3 Two phase sampling

We now consider another reason to prefer $\hat{\theta}_3^*$. The determination of the false positive rates is very time consuming. Thus, instead of determining the false positive rate for every record in the database, we only did so for a random sample of records. Such a random sample, however, is really the second phase of a two-phase sampling scheme, since the records themselves constitute a sample from the population of all possible lesions. Thus, even if we take a simple random sample from the set of records, the selection probabilities of the tumour types will be unequal, being proportional to the "visibility probabilities" p_i . The estimator $\hat{\theta}_3^*$ is readily adaptable to this two phase scheme.

Let T be an index set, and let $\{a_i, i \in T\}$ be a sequence of binary random variables with $E(a_i) = p_i$, $\sum p_i < \infty$, and $N = \sum a_i$. As before, we can consider $\{a_i = 1\}$ as the event that unit i is visible. We now take a random sample of the visible units by defining another binary sequence $\{b_i, i \in T\}$ with $b_i \leq a_i$. If the second sample is taken with equal selection probabilities, and is of size at most n , then we have

$$E(b_i \mid a) = \min\left(a_i, \frac{n}{N} a_i\right). \quad (7.22)$$

Thus the unconditional selection probabilities are

$$E(b_i) = \min\left(p_i, nE\left(\frac{a_i}{N}\right)\right). \quad (7.23)$$

Theorem 7.4 If the subsampling indicators b_i satisfy (7.22), and $\{Y_i, i \in T\}$ is a sequence of random variables such the Y_i and b_i are conditionally independent given a , then

$$E\left(\frac{N}{n'} \sum b_i Y_i\right) = \eta E\left(\sum a_i Y_i\right) \quad (7.24)$$

where $n' = \min(n, N)$ and $\eta = \Pr\{N > 0\}$.

Proof:

$$E\left(\frac{N}{n'} \sum b_i Y_i\right) = E\left(E\left(\frac{N}{n'} \sum b_i Y_i \mid a\right)\right).$$

Now

$$E(b_i Y_i \mid a) = a_i \frac{n'}{N} E(Y_i \mid a)$$

if $N \neq 0$, and as before, we define $b_i/N = 0$ if $N = 0$. Thus

$$\begin{aligned} E\left(\frac{N}{n'} \sum b_i Y_i\right) &= \eta E(E(a_i Y_i \mid a) \mid N > 0) + (1 - \eta) \times 0 \\ &= \eta E\left(\sum a_i Y_i\right). \end{aligned}$$

QED

With this theorem, we can take any estimator using our original sample and adapt it to a random subsample merely by multiplication by N/n' . In our applications, the event $\{N = 0\}$ will have negligible probability, so we will ignore the factor η . In fact, we will assume that $\Pr\{N < n\}$ is negligible, thus replacing $n' = \min(n, N)$ by just n .

7.3.4 The estimators

We are now ready to apply the preceding theorems to the problem of estimating the false positive rate. As remarked previously, the false positive rate depends not only on the significance level and the distribution of background incidence, but also on the design. Since the studies in the database have various designs, they will also have various false positive rates. In principle, we could calculate the false positive rate for each design, but for the purpose of this investigation we will only estimate the average false positive rate over the set of designs appearing.

Because we were interested in the contribution of different kinds of responses to the overall false positive rate, we divided the database into twenty strata. For each sex-species combination, we defined the record categories of specific benign, aggregate benign, specific malignant, aggregate malignant, and mortality as five different strata. We then randomly selected (by simple random sampling) five records from each of the strata representing the aggregate benign, aggregate malignant, and mortality records, and fifteen records from each of the strata representing the specific lesions. The selection probability of any response type can then be estimated as

$$\hat{\pi}_i = n \hat{p}_i / N, \quad (7.25)$$

where N is the total number of records in the stratum of the i^{th} record, n is the number of records sampled from that stratum, and $\hat{p}_i$ is the observed proportion of studies in the stratum having records for that response type. The false positive rate was then estimated as

$$\hat{\alpha}^* = m^{-1} \sum_i \frac{\varphi(\theta_i)}{\hat{\pi}_i}, \quad (7.26)$$

where the summation is over all the sampled records, $\theta_i = (\hat{\mu}_i, \hat{\rho}_i)$ are the beta-binomial parameters for the response type in the i^{th} record (estimated from the collection of all control groups in that sex and species), and φ_i is the function expressing the exact size as a function of the beta-binomial parameters and the design used in the i^{th} record.

According to the theorems of this section, this estimator would be unbiased if certain oversimplifying assumptions are made: That φ_i is a linear function that depends only on $\hat{\mu}_i$, which itself is the arithmetic mean of m independent unbiased estimates, and that the probabilities of observing a lesion i are the same in all studies. If we weaken the assumptions requiring only that φ_i be convex rather than linear, then Theorem 7.3 shows that the estimator (7.26) would have a positive bias.

These assumptions are evidently not true. The function is definitely not linear, nor is it convex, and it depends on the intrastudy correlation ρ as well as the mean μ . The parameters were estimated by iterative procedures, and thus the estimates do not have a closed form expression, and certainly are not a simple average of individual study estimates. (It is not even known whether they are unbiased.) Furthermore, the "visibility probabilities" are not equal across studies, since larger studies will have more chance of finding rare lesions, as will studies with a treatment inducing that particular lesion.

Nevertheless, heuristic arguments suggest that the estimator (7.26) is an overestimate. The function φ was found empirically to depend only slightly on ρ , and appeared to be convex at least in the region corresponding to rare lesions. (The inflexion point for most designs examined appeared to be between $\mu = .01$ and $\mu = .05$). Moreover, since the beta-binomial parameters were estimated from fairly large samples, and the intrastudy correlations ρ were low, any bias in the parameters $\hat{\mu}_i$ would be expected to be small. Finally, the unequal visibility probabilities would tend to give more weight to φ_i values for the larger designs, and such designs would tend to have larger false positive rates.

7.3.5 Variance of the estimator

To gauge the precision of the estimator (7.26), we require a variance estimate. We will present an obvious estimator, and then examine its properties, again under simplifying assumptions. The application of the variance estimator in the present case again depends on heuristics.

For the variance estimator, we require the pairwise selection probabilities. First, consider the complete set of records. Denote by $p_{ij} = E(a_{ik}a_{kj})$ the probability that lesions i and j both appear in the same study. Since we are assuming the studies to be independent, the probability of these two lesions appearing in different studies k and k' is just $p_i p_j$.

These probabilities can be represented as a matrix

$$P_2 = I_m \otimes (P_1 - pp') + J_m \otimes pp' \quad (7.27)$$

where p is the vector of p_i 's and P_1 is the matrix of the p_{ij} 's. We will also denote by A the matrix in which the k^{th} row is the sequence a_k . We can now state the following theorem.

Theorem 7.5 *If the binary sequences a_k are independent and identically distributed, then*

an unbiased estimator of P_2 is given by

$$\hat{P}_2 = \frac{1}{m-1} [I \otimes A' (I - m^{-1}J) A + m^{-1}J \otimes A' (J - I) A]. \quad (7.28)$$

Proof: The ij^{th} element of $A'A$ is $\sum_k a_{ki}a_{kj}$ so that $E(A'A) = mP_1$. The ij^{th} element of $A'JA$ is

$$\sum_k \sum_{k'} a_{ki}a_{k'j} = \sum_k a_{ki}a_{kj} + \sum_{k' \neq k} \sum_k a_{ki}a_{k'j}$$

whose expected value is $mp_{ij} + m(m-1)p_i p_j$. Thus $E(A'JA) = mP_1 + m(m-1)pp'$. From this it follows that

$$E(A'(I - m^{-1}J)A) = (m-1)(P_1 - pp') \quad (7.29)$$

and

$$E(A'(J - I)A) = m(m-1)pp'. \quad (7.30)$$

Substituting (7.29) and (7.30) into (7.28) then gives (7.27). QED

The elements of the matrix $\hat{P}_2$ can be written as

$$\hat{p}_{ki,j} = \frac{m(1 - \delta_{kl})\hat{p}_i\hat{p}_j + (m\delta_{kl} - 1)\hat{p}_{ij}}{m-1} \quad (7.31)$$

where $\hat{p}_{ij} = m^{-1} \sum_k a_{ki}a_{kj}$. Note that $p_i p_j$ is not estimated by $\hat{p}_i \hat{p}_j$ but by

$$\frac{(m\hat{p}_i\hat{p}_j - \hat{p}_{ij})}{(m-1)}.$$

Theorem 7.6 If the $\{a_k, k = 1, \dots, m\}$ are exchangeable, then

$$E(a_{ki}a_{lj} | \hat{P}_2) = \frac{m(1 - \delta_{kl})\hat{p}_i\hat{p}_j + (m\delta_{kl} - 1)\hat{p}_{ij}}{m-1}. \quad (7.32)$$

Proof: Note that the map $(\hat{p}, \hat{P}_1) \mapsto \hat{P}_2$ is one-to-one, so that conditioning on $\hat{P}_2$ is equivalent to conditioning on $(\hat{p}, \hat{P}_1)$. Moreover, these statistics are invariant under permutations of the indices k of the set $\{a_k, k = 1, \dots, m\}$, as is their distribution, by the hypothesis of exchangeability. Thus by Lemma 7.1 it follows that the conditional expectation must also be invariant. Hence $E(a_{ki}a_{lj} | \hat{P}_2)$ must be of the form $\alpha_{ij}\delta_{kl} + \beta_{ij}$ for two $\hat{P}_2$ -measurable random variables α_{ij} and β_{ij} .

Now $m^{-2} \sum_{kl} a_{ki}a_{lj} = \hat{p}_i\hat{p}_j$. Taking conditional expectations given $\hat{P}_2$ of both sides gives

$$\begin{aligned} m^{-2} \sum_{kl} (\alpha_{ij}\delta_{kl} + \beta_{ij}) &= m^{-1}\alpha_{ij} + \beta_{ij} \\ &= \hat{p}_i\hat{p}_j. \end{aligned} \quad (7.33)$$

Similarly, $m^{-1} \sum_k \delta_{kl} a_{ki} a_{lj} = \hat{p}_{ij}$, so that

$$\begin{aligned} m^{-1} \sum_k (\alpha_{ij} \delta_{kl} + \beta_{ij} \delta_{kl}) &= \alpha_{ij} + \beta_{ij} \\ &= \hat{p}_{ij}. \end{aligned} \quad (7.34)$$

Solving the system of equations (7.33) and (7.34) we have

$$\begin{aligned} \alpha_{ij} &= \frac{m}{m-1} (\hat{p}_{ij} - \hat{p}_i \hat{p}_j), \\ \text{and } \beta_{ij} &= \frac{m \hat{p}_i \hat{p}_j - \hat{p}_{ij}}{m-1}. \end{aligned}$$

Thus

$$\begin{aligned} E(a_{ki} a_{lj} | P_2) &= \alpha_{ij} \delta_{kl} + \beta_{ij} \\ &= \frac{m(1 - \delta_{kl}) \hat{p}_i \hat{p}_j + (m \delta_{kl} - 1) \hat{p}_{ij}}{m-1}. \end{aligned}$$

QED

Corollary 7.6.1

$$\text{Cov}(a_{ki}, a_{lj} | \hat{P}_2) = \frac{(m \delta_{kl} - 1) (\hat{p}_{ij} - \hat{p}_i \hat{p}_j)}{m-1}.$$

The estimated pairwise selection probabilities thus have the property that

$$E\left(\frac{a_{ki} a_{lj}}{\hat{p}_{ki, lj}} \mid \hat{p}_{ki, lj}\right) = 1, \text{ if } \hat{p}_{ki, lj} \neq 0.$$

Unlike the case with the $\hat{p}_i$'s, however, visible lesions can have zero values of $\hat{p}_{ki, lj}$. Note that $\hat{p}_{ki, lj} = 0$ if and only if $a_{ki} a_{lj} = 0$ for all k , so that lesions i and j never occur together in the same study, whereas $\hat{p}_{ki, lj} = 0$ for $k \neq l$ if and only if $a_{ki} a_{lj} = 0$ for all $k \neq l$. If both the i^{th} and j^{th} lesions are visible, this latter event can happen only if each of the lesions i and j occurs in exactly one common study, so that $\hat{p}_{ij} = 1/m$.

Consider now sampling the visible lesions as in Section 7.3.3, again denoting by b_{ki} the indicator for selecting response type i in study k . Since sampling is conditional on visibility, we have that $E(b_{ki} | \mathbf{a}) = n a_{ki} / N$ and $E(b_{ki} b_{lj} | \mathbf{a}) = d_{ki, lj} a_{ki} a_{lj}$ where $d_{ki, lj}$ will depend on the sampling scheme. For simple random sampling,

$$d_{ki, lj} = \frac{n(\delta_{ij} \delta_{kl} (N - n) + n - 1)}{N(N - 1)}. \quad (7.35)$$

For stratified sampling, the same expression holds for i and j in the same stratum, (except that now n represents the stratum sample size, and N the total number of visible records in the stratum) whereas if i and j are in different strata (say r and s), then

$$d_{ki, lj} = \frac{n_r n_s}{N_r N_s}.$$

Note that with stratified sampling in this two phase form, b_{ki} and b_{kj} may be dependent even if i and j are in different strata, since a_{ki} and a_{kj} may be dependent.

With this notation, the estimator (7.26) can be written

$$\hat{\alpha}^* = \frac{1}{m} \sum_{ki} \frac{b_{ki}}{\hat{\pi}_i} \varphi_{ki}(\hat{\theta}_i) \quad (7.36)$$

from which we see that its variance would be

$$\text{Var}(\hat{\alpha}^*) = \frac{1}{m^2} \sum_{kij} \text{Cov} \left(\frac{b_{ki}}{\hat{\pi}_i} \varphi_{ki}(\hat{\theta}_i), \frac{b_{lj}}{\hat{\pi}_j} \varphi_{lj}(\hat{\theta}_j) \right). \quad (7.37)$$

Determination of the exact value of this variance is rather intractable, because of the complicated dependence between the indicator variable b_{ki} , the estimated selection probability $\hat{\pi}_i$, and the background parameters $\hat{\theta}_i$. We will re-express this variance into a form that will be useful for examining the bias of an estimator.

Define

$$\hat{\pi}_{ki,lj} = d_{ki,lj} \hat{p}_{ki,lj} \quad \text{and} \quad (7.38)$$

$$\hat{\theta}_{ki,lj} = \begin{cases} \frac{\hat{\pi}_{ki,lj}}{\hat{\pi}_i \hat{\pi}_j} - 1 & \text{if } \hat{\pi}_i \hat{\pi}_j > 0 \\ -1 & \text{if } \hat{\pi}_i \hat{\pi}_j = 0. \end{cases} \quad (7.39)$$

Our derivation will be done by a sequence of conditioning arguments, using the identity $\text{Cov}(X, Y) = E(\text{Cov}(X, Y | Z)) + \text{Cov}(E(X | Z), E(Y | Z))$. Ultimately, to achieve anything tractable, we will have to resort to the oversimplifying Assumptions C.

First, we condition on the visible lesions, and their response parameters:

$$\begin{aligned} & \text{Cov} \left(\frac{b_{ki}}{\hat{\pi}_i} \varphi_{ki}(\hat{\theta}_i), \frac{b_{lj}}{\hat{\pi}_j} \varphi_{lj}(\hat{\theta}_j) \mid a, \hat{\theta} \right) \\ &= \left[E(b_{ki} b_{lj} \mid a, \hat{\theta}) - E(b_{ki} \mid a, \hat{\theta}) E(b_{lj} \mid a, \hat{\theta}) \right] \frac{\varphi_{ki}(\hat{\theta}_i) \varphi_{lj}(\hat{\theta}_j)}{\hat{\pi}_i \hat{\pi}_j} \\ &= \left(d_{ki,lj} a_{ki} a_{lj} - \frac{n_r}{N_r} a_{ki} \frac{n_s}{N_s} a_{lj} \right) \frac{\varphi_{ki}(\hat{\theta}_i) \varphi_{lj}(\hat{\theta}_j)}{\hat{\pi}_i \hat{\pi}_j}, \end{aligned}$$

where r and s are the strata in which response types i and j appear. Now, if $\hat{p}_{ki,lj} \neq 0$, then $d_{ki,lj} = \frac{\hat{\pi}_{ki,lj}}{\hat{p}_{ki,lj}}$, $n_r/N_r = \hat{\pi}_i/\hat{p}_i$, and $n_s/N_s = \hat{\pi}_j/\hat{p}_j$. The case $\hat{p}_{ki,lj} = 0$ can happen only if $a_{ki} a_{lj} = 0$ in which case the whole expression would be interpreted as zero. Thus

$$\begin{aligned} & \text{Cov} \left(\frac{b_{ki}}{\hat{\pi}_i} \varphi_{ki}(\hat{\theta}_i), \frac{b_{lj}}{\hat{\pi}_j} \varphi_{lj}(\hat{\theta}_j) \mid a, \hat{\theta} \right) \\ &= \begin{cases} \frac{a_{ki} a_{lj}}{\hat{p}_{ki,lj}} \left(\frac{\hat{\pi}_{ki,lj}}{\hat{\pi}_i \hat{\pi}_j} - \frac{\hat{p}_{ki,lj}}{\hat{p}_i \hat{p}_j} \right) \varphi_{ki}(\hat{\theta}_i) \varphi_{lj}(\hat{\theta}_j) & \text{if } \hat{p}_{ki,lj} > 0 \\ 0 & \text{if } \hat{p}_{ki,lj} = 0. \end{cases} \quad (7.40) \end{aligned}$$

On the other hand

$$E \left(\frac{b_{ki}}{\hat{\pi}_i} \varphi_{ki}(\hat{\theta}_i) \mid \mathbf{a}, \hat{\Theta} \right) = \frac{a_{ki}}{\hat{p}_i} \varphi_{ki}(\hat{\theta}_i). \quad (7.41)$$

If we now take the conditional expectation of (7.40) given $\hat{\mathbf{P}}_2$ we get

$$E \left(a_{ki} a_{lj} \varphi_{ki}(\hat{\theta}_i) \varphi_{lj}(\hat{\theta}_j) \mid \hat{\mathbf{P}}_2 \right) \frac{1}{\hat{p}_{ki,lj}} \left(\frac{\hat{\pi}_{ki,lj}}{\hat{\pi}_i \hat{\pi}_j} - \frac{\hat{p}_{ki,lj}}{\hat{p}_i \hat{p}_j} \right), \quad (7.42)$$

whereas from (7.41) we have

$$\begin{aligned} & \text{Cov} \left(\frac{a_{ki}}{\hat{p}_i} \varphi_{ki}(\hat{\theta}_i), \frac{a_{lj}}{\hat{p}_j} \varphi_{lj}(\hat{\theta}_j) \right) \\ &= E \left(\frac{a_{ki} a_{lj}}{\hat{p}_i \hat{p}_j} \varphi_{ki}(\hat{\theta}_i) \varphi_{lj}(\hat{\theta}_j) \right) - E \left(\frac{a_{ki}}{\hat{p}_i} \varphi_{ki}(\hat{\theta}_i) \right) E \left(\frac{a_{lj}}{\hat{p}_j} \varphi_{lj}(\hat{\theta}_j) \right). \end{aligned} \quad (7.43)$$

Combining (7.42) with (7.43) we thus have

$$\begin{aligned} & \text{Cov} \left(\frac{b_{ki}}{\hat{\pi}_i} \varphi_{ki}(\hat{\theta}_i), \frac{b_{lj}}{\hat{\pi}_j} \varphi_{lj}(\hat{\theta}_j) \right) \\ &= E \left(E \left(a_{ki} a_{lj} \varphi_{ki}(\hat{\theta}_i) \varphi_{lj}(\hat{\theta}_j) \mid \hat{\mathbf{P}}_2 \right) \frac{\hat{\pi}_{ki,lj}}{\hat{p}_{ki,lj} \hat{\pi}_i \hat{\pi}_j} \right) \\ &\quad - E \left(\frac{a_{ki}}{\hat{p}_i} \varphi_{ki}(\hat{\theta}_i) \right) E \left(\frac{a_{lj}}{\hat{p}_j} \varphi_{lj}(\hat{\theta}_j) \right), \end{aligned} \quad (7.44)$$

where we are careful to interpret

$$\frac{E \left(a_{ki} a_{lj} \varphi_{ki}(\hat{\theta}_i) \varphi_{lj}(\hat{\theta}_j) \mid \hat{\mathbf{P}}_2 \right)}{\hat{p}_{ki,lj}}$$

as zero if $\hat{p}_{ki,lj} = 0$.

Now, to simplify (7.44) we must assume Assumptions C and replace $\varphi_{ki}(\hat{\theta}_i)$ and $\varphi_{lj}(\hat{\theta}_j)$ by simply $\hat{\theta}_i$ and $\hat{\theta}_j$. We also require the following results.

Lemma 7.4 Under Assumptions C,

$$E \left(a_{ki} a_{lj} \hat{\theta}_{ri} \hat{\theta}_{sj} \mid \hat{\mathbf{P}}_2 \right) = \frac{E(\hat{\theta}_{ri} \hat{\theta}_{sj}) E(a_{ki} a_{lj} a_{ri} a_{sj} \mid \hat{\mathbf{P}}_2)}{E(a_{ri} a_{sj})}. \quad (7.45)$$

Proof:

$$E \left(a_{ki} a_{lj} \hat{\theta}_{ri} \hat{\theta}_{sj} \mid \hat{\mathbf{P}}_2 \right) = E \left(\hat{\theta}_{ri} \hat{\theta}_{sj} \mid a_{ki} a_{lj} = 1, \hat{\mathbf{P}}_2 \right) E \left(a_{ki} a_{lj} \mid \hat{\mathbf{P}}_2 \right). \quad (7.46)$$

But by Lemma 7.3

$$E \left(\hat{\theta}_{ri} \hat{\theta}_{sj} \mid a_{ki} a_{lj} = 1, \hat{\mathbf{P}}_2 \right) = \frac{E(\hat{\theta}_{ri} \hat{\theta}_{sj}) E(a_{ri} a_{sj} \mid a_{ki} a_{lj} = 1, \hat{\mathbf{P}}_2)}{E(a_{ri} a_{sj})}, \quad (7.47)$$

and

$$E(a_{ki}a_{lj}a_{ri}a_{sj} | \hat{P}_2) = E(a_{ri}a_{sj} | a_{ki}a_{lj} = 1, \hat{P}_2) E(a_{ki}a_{lj} | \hat{P}_2). \quad (7.48)$$

Combining (7.46), (7.47), and (7.48) gives us (7.45). QED

Theorem 7.7 Under Assumptions C

$$\begin{aligned} E(a_{ki}a_{lj}\hat{\theta}_i\hat{\theta}_j | \hat{P}_2) \\ = \left[\frac{\hat{p}_{ij}}{mp_{ij}}\kappa_{ij} + \left(\frac{\hat{p}_{ij}}{mp_{ij}} + \frac{m\hat{p}_i\hat{p}_j - \hat{p}_{ij}}{mp_i p_j} \right) \theta_i\theta_j \right] \hat{p}_{ki,lj}, \end{aligned} \quad (7.49)$$

where $\kappa_{ij} = \text{Cov}(\hat{\theta}_{ki}, \hat{\theta}_{lj})$.

Proof:

$$\begin{aligned} m^2 E(a_{ki}a_{lj}\hat{\theta}_i\hat{\theta}_j | \hat{P}_2) &= \sum_{r,s} E(a_{ki}a_{lj}\hat{\theta}_{ri}\hat{\theta}_{sj} | \hat{P}_2) \\ &= \sum_{r,s} \frac{E(a_{ki}a_{lj}a_{ri}a_{sj} | \hat{P}_2) E(\hat{\theta}_{ri}\hat{\theta}_{sj})}{E(a_{ri}a_{sj})} \\ &= \sum_r \frac{E(a_{ki}a_{lj}a_{ri}a_{rj} | \hat{P}_2) E(\hat{\theta}_{ri}\hat{\theta}_{rj})}{p_{ij}} \\ &\quad + \sum_{r \neq s} \frac{E(a_{ki}a_{lj}a_{ri}a_{sj} | \hat{P}_2) E(\hat{\theta}_{ri}\hat{\theta}_{sj})}{p_i p_j}. \end{aligned} \quad (7.50)$$

Now $E(\hat{\theta}_{ri}\hat{\theta}_{rj}) = \theta_i\theta_j + \kappa_{ij}$ and

$$\sum_r E(a_{ki}a_{lj}a_{ri}a_{rj} | \hat{P}_2) = m\hat{p}_{ij} E(a_{ki}a_{lj} | \hat{P}_2). \quad (7.51)$$

However, for $r \neq s$, $E(\hat{\theta}_{ri}\hat{\theta}_{sj}) = \theta_i\theta_j$. And since

$$\sum_{\substack{r \\ r \neq s}} E(a_{ki}a_{lj}a_{ri}a_{sj} | \hat{P}_2) = E(a_{ki}a_{lj}a_{ri}(m\hat{p}_j - a_{rj}) | \hat{P}_2) \quad (7.52)$$

we have

$$\sum_r \sum_{s \neq r} E(a_{ki}a_{lj}a_{ri}a_{sj} | \hat{P}_2) = (m^2\hat{p}_i\hat{p}_j - m\hat{p}_{ij}) E(a_{ki}a_{lj} | \hat{P}_2). \quad (7.53)$$

Substituting (7.51) and (7.53) into (7.50) and invoking Corollary 7.6.1 then gives us (7.49). QED

We now apply Theorem 7.7 to the expression (7.44). Noting that $\hat{p}_{ki,lj} = 0$ implies that $\hat{\pi}_{ki,lj} = 0$, and that under Assumptions C

$$E\left(\frac{\hat{p}_{ij}}{mp_{ij}} + \frac{m\hat{p}_i\hat{p}_j - \hat{p}_{ij}}{mp_i p_j}\right) = 1, \quad (7.54)$$

we obtain

$$\begin{aligned} & \text{Cov} \left(\frac{b_{ki}}{\hat{\pi}_i} \varphi_{ki}(\hat{\theta}_i), \frac{b_{lj}}{\hat{\pi}_j} \varphi_{lj}(\hat{\theta}_j) \right) \\ &= E \left(\left(\frac{\hat{p}_{ij}}{mp_{ij}} + \frac{m\hat{p}_i\hat{p}_j - \hat{p}_{ij}}{mp_i p_j} \right) \hat{\sigma}_{ki,lj} \right) \theta_i \theta_j + E \left(\frac{\hat{p}_{ij} \hat{\pi}_{ki,lj}}{mp_{ij} \hat{\pi}_i \hat{\pi}_j} \right) \kappa_{ij}. \end{aligned} \quad (7.55)$$

7.3.6 Estimating the variance

Proceeding heuristically, we propose an estimator of $\text{Var}(\hat{\alpha}^*)$ which would be appropriate if the $\hat{\pi}_i$ and $\hat{\theta}_i$ were constants rather than random variables. Consider

$$v(\hat{\alpha}^*) = \frac{1}{m^2} \sum_{k,l,j} \frac{b_{ki} b_{lj}}{\hat{\pi}_{ki,lj}} \hat{\sigma}_{ki,lj} \varphi_{ki}(\hat{\theta}_i) \varphi_{lj}(\hat{\theta}_j). \quad (7.56)$$

This estimator of $\text{Var}(\hat{\alpha}^*)$ is actually somewhat biased. To determine the bias we must find the expectation of $v(\hat{\alpha}^*)$. We will do this by finding the expected value of a typical term in the summation, again by a sequence of conditional arguments, and then accumulating these terms in the end.

Conditioning first on $(\mathbf{a}, \hat{\Theta})$ we get for the conditional expectation of a typical term

$$E \left(\frac{b_{ki} b_{lj}}{\hat{\pi}_{ki,lj}} \hat{\sigma}_{ki,lj} \varphi_{ki}(\hat{\theta}_i) \varphi_{lj}(\hat{\theta}_j) \mid \mathbf{a}, \hat{\Theta} \right) = \frac{a_{ki} a_{lj}}{\hat{p}_{ki,lj}} \hat{\sigma}_{ki,lj} \varphi_{ki}(\hat{\theta}_i) \varphi_{lj}(\hat{\theta}_j). \quad (7.57)$$

To achieve further simplification, we again have to invoke Assumptions C, and replace $\varphi_{ki}(\hat{\theta}_i)$ by $\hat{\theta}_{ki}$. We then have by Theorem 7.7

$$\begin{aligned} & E \left(\frac{a_{ki} a_{lj}}{\hat{p}_{ki,lj}} \hat{\theta}_i \hat{\theta}_j \mid \hat{\mathbf{P}}_2 \right) \\ &= \begin{cases} \left(\frac{\hat{p}_{ij}}{mp_{ij}} + \frac{m\hat{p}_i\hat{p}_j - \hat{p}_{ij}}{mp_i p_j} \right) \theta_i \theta_j + \frac{\hat{p}_{ij}}{mp_{ij}} \kappa_{ij} & \text{if } \hat{p}_{ki,lj} > 0 \\ 0 & \text{if } \hat{p}_{ki,lj} = 0. \end{cases} \end{aligned} \quad (7.58)$$

(Note that $\hat{\sigma}_{ki,lj}$, being a function of $\hat{\mathbf{P}}_2$ can be factored out of the conditional expectation.) Thus using (7.58) we have

$$\begin{aligned} & E \left(\frac{b_{ki} b_{lj}}{\hat{\pi}_{ki,lj}} \hat{\sigma}_{ki,lj} \hat{\theta}_i \hat{\theta}_j \right) \\ &= E \left(E \left(\frac{b_{ki} b_{lj}}{\hat{\pi}_{ki,lj}} \hat{\sigma}_{ki,lj} \hat{\theta}_i \hat{\theta}_j \mid \hat{\mathbf{P}}_2 \right) \right) \\ &= E \left(\left(\frac{\hat{p}_{ij}}{mp_{ij}} + \frac{m\hat{p}_i\hat{p}_j - \hat{p}_{ij}}{mp_i p_j} \right) \hat{\sigma}_{ki,lj} \mid \hat{p}_{ki,lj} > 0 \right) \nu_{ki,lj} \theta_i \theta_j \\ &\quad + E \left(\frac{\hat{p}_{ij}}{mp_{ij}} \hat{\sigma}_{ki,lj} \mid \hat{p}_{ki,lj} > 0 \right) \nu_{ki,lj} \kappa_{ij}, \end{aligned} \quad (7.59)$$

where $\nu_{ki,lj} = \Pr \{ \hat{p}_{ki,lj} > 0 \}$.

For further simplifications we again need some more results.

Lemma 7.5 For any two random variables X and Y with finite expectations and $Y \geq 0$

$$E(X | Y > 0) \Pr\{Y > 0\} = E(X) - E(X | Y = 0) \Pr\{Y = 0\}.$$

Proof: Obvious.

Theorem 7.8 For $k \neq l$

$$E(\hat{p}_{ij} | \hat{p}_{ki,lj} = 0) \Pr\{\hat{p}_{ki,lj} = 0\} = p_{ij}(1 - p_i - p_j + p_{ij})^{m-1}$$

Proof: For $k \neq l$

$$\hat{p}_{ki,lj} = \frac{m\hat{p}_i\hat{p}_j - \hat{p}_{ij}}{m-1}. \quad (7.60)$$

Now $\hat{p}_{ij} = m^{-1} \sum_k a_{ki} a_{kj} \leq m^{-1} (\sum_k a_{ki} a_{kj})^2$ with equality if and only if $\sum_k a_{ki} a_{lj}$ is 0 or 1. But by the Cauchy-Schwarz inequality,

$$\left(\sum_k a_{ki} a_{kj} \right)^2 \leq \sum_k a_{ki}^2 \sum_l a_{lj}^2 = \sum_k a_{ki} \sum_l a_{lj} = m^2 \hat{p}_i \hat{p}_j.$$

with equality if and only if $a_{ki} = c a_{lj}$ for some constant c . However since the a_{ki} are binary, c must be 0 or 1. Thus $\hat{p}_{ij} \leq m \hat{p}_i \hat{p}_j$ with equality if and only if

$$\hat{p}_i \hat{p}_j = \hat{p}_{ij} = 0 \quad (7.61)$$

$$\text{or} \quad \hat{p}_i = \hat{p}_j = \hat{p}_{ij} = \frac{1}{m}. \quad (7.62)$$

Thus

$$E(\hat{p}_{ij} | \hat{p}_{ki,lj} = 0) \Pr\{\hat{p}_{ki,lj} = 0\} = \frac{1}{m} \Pr\left\{\hat{p}_{ij} = \hat{p}_i = \hat{p}_j = \frac{1}{m}\right\}.$$

Considering now, for any study k , the probabilities of the following four outcomes

$$\Pr\{a_{ki} = 0, a_{kj} = 0\} = 1 - p_i - p_j + p_{ij}$$

$$\Pr\{a_{ki} = 0, a_{kj} = 1\} = p_j - p_{ij}$$

$$\Pr\{a_{ki} = 1, a_{kj} = 0\} = p_i - p_{ij}$$

$$\Pr\{a_{ki} = 1, a_{kj} = 1\} = p_{ij}$$

and using (7.61) and (7.62) it can be determined that

$$\Pr\{\hat{p}_{ij} = \hat{p}_i = \hat{p}_j\} = m p_{ij} (1 - p_i - p_j + p_{ij})^{m-1}$$

and the result follows. QED

Theorem 7.9

$$E(\hat{p}_i \hat{p}_j | \hat{p}_{ij} = 0) = \frac{m-1}{m} \frac{(p_i - p_{ij})(p_j - p_{ij})}{(1 - p_{ij})^2}.$$

Proof: One can write the product $\hat{p}_i \hat{p}_j$ as $m^{-2} \sum_{kl} a_{ki} a_{lj}$. For $k \neq l$ we have

$$\begin{aligned} E(a_{ki} a_{lj} | \hat{p}_{ij} = 0) &= \frac{\Pr \left\{ a_{ki} = 1, a_{kj} = 0, a_{li} = 0, a_{lj} = 1, \sum_{\substack{r \neq k \\ r \neq l}} a_{ri} a_{rj} = 0 \right\}}{\Pr \left\{ \sum_{r=1}^m a_{ri} a_{rj} = 0 \right\}} \\ &= \frac{(p_i - p_{ij})(p_j - p_{ij})(1 - p_{ij})^{m-2}}{(1 - p_{ij})^m} \\ &= \frac{(p_i - p_{ij})(p_j - p_{ij})}{(1 - p_{ij})^2}, \end{aligned} \quad (7.63)$$

whereas $E(a_{ki} a_{lj} | \hat{p}_{ij} = 0) = 0$ by Theorem 7.6. Summing (7.63) over the $m(m-1)$ terms for which $k \neq l$ then completes the proof. QED

We can now proceed with the simplification of (7.59). From the second term, we have

$$\begin{aligned} E \left(\frac{\hat{p}_{ij}}{mp_{ij}} \delta_{ki,lj} \mid \hat{p}_{ki,lj} > 0 \right) \\ = E \left(\frac{\hat{p}_{ij} \hat{\pi}_{ki,lj}}{mp_{ij} \hat{\pi}_i \hat{\pi}_j} \mid \hat{p}_{ki,lj} > 0 \right) - E \left(\frac{\hat{p}_{ij}}{mp_{ij}} \mid \hat{p}_{ki,lj} > 0 \right). \end{aligned} \quad (7.64)$$

But if $\hat{p}_{ki,lj} = 0$ then $\hat{\pi}_{ki,lj} = 0$ so that

$$E \left(\frac{\hat{p}_{ij} \hat{\pi}_{ki,lj}}{mp_{ij} \hat{\pi}_i \hat{\pi}_j} \mid \hat{p}_{ki,lj} > 0 \right) \nu_{ki,lj} = E \left(\frac{\hat{p}_{ij} \hat{\pi}_{ki,lj}}{mp_{ij} \hat{\pi}_i \hat{\pi}_j} \right). \quad (7.65)$$

On the other hand, by Lemma 7.5 we have

$$E \left(\frac{\hat{p}_{ij}}{mp_{ij}} \mid \hat{p}_{ki,lj} > 0 \right) \nu_{ki,lj} = \frac{1}{m} - E \left(\frac{\hat{p}_{ij}}{mp_{ij}} \mid \hat{p}_{ki,lj} = 0 \right) (1 - \nu_{ki,lj}). \quad (7.66)$$

If $k = l$, then $\hat{p}_{ki,lj} = 0$ means that $\hat{p}_{ij} = 0$, whereas for $k \neq l$ we can apply Theorem 7.8 to get the conditional expectation. Denoting the left-hand side of (7.66) by $C_{ki,lj}$ we then have

$$C_{ki,lj} = \frac{1}{m} (1 - (1 - \delta_{kl})(1 - p_i - p_j + p_{ij})^{m-1}). \quad (7.67)$$

Similarly applying Lemma 7.5 to the first term of (7.59) we have

$$\begin{aligned} E \left(\left(\frac{\hat{p}_{ij}}{mp_{ij}} + \frac{m\hat{p}_i \hat{p}_j - \hat{p}_{ij}}{mp_i p_j} \right) \delta_{ki,lj} \mid \hat{p}_{ki,lj} > 0 \right) \nu_{ki,lj} \\ = E \left(\left(\frac{\hat{p}_{ij}}{mp_{ij}} + \frac{m\hat{p}_i \hat{p}_j - \hat{p}_{ij}}{mp_i p_j} \right) \delta_{ki,lj} \right) \\ - E \left(\left(\frac{\hat{p}_{ij}}{mp_{ij}} + \frac{m\hat{p}_i \hat{p}_j - \hat{p}_{ij}}{mp_i p_j} \right) \delta_{ki,lj} \mid \hat{p}_{ki,lj} = 0 \right) (1 - \nu_{ki,lj}). \end{aligned} \quad (7.68)$$

Now, if $\hat{p}_{ki,j} = 0$ then $\partial_{ki,j} = -1$. Noting that $\hat{p}_{ki,j} = 0$ means $\hat{p}_j = 0$ if $k = i$ and $\hat{p}_i = \hat{p}_i \hat{p}_j$ if $k \neq i$, we can write the second term in (7.68) as

$$-B_{ki,j} = \left[-\delta_{ki} E \left(\frac{\hat{p}_i \hat{p}_j}{\hat{p}_i \hat{p}_j} \middle| \hat{p}_{ij} = 0 \right) - (1 - \delta_{ki}) E \left(\frac{\hat{p}_{ij}}{m \hat{p}_{ij}} \middle| \hat{p}_{ki,j} = 0 \right) \right] (1 - \nu_{ki,j})$$

giving

$$\begin{aligned} & E \left(\left(\frac{\hat{p}_{ij}}{m \hat{p}_{ij}} + \frac{m \hat{p}_i \hat{p}_j - \hat{p}_{ij}}{m \hat{p}_i \hat{p}_j} \right) \partial_{ki,j} \middle| \hat{p}_{ki,j} > 0 \right) \nu_{ki,j} \\ &= E \left(\left(\frac{\hat{p}_{ij}}{m \hat{p}_{ij}} + \frac{m \hat{p}_i \hat{p}_j - \hat{p}_{ij}}{m \hat{p}_i \hat{p}_j} \right) \partial_{ki,j} \right) + B_{ki,j}. \end{aligned} \quad (7.69)$$

Substituting (7.64), (7.65), (7.67), and (7.69) into (7.59) then gives

$$\begin{aligned} & E \left(\frac{b_{ki,j}}{\hat{p}_{ki,j}} \partial_{ki,j} \hat{\theta}_i \hat{\theta}_j \right) \\ &= E \left(\left(\frac{\hat{p}_{ij}}{m \hat{p}_{ij}} + \frac{m \hat{p}_i \hat{p}_j - \hat{p}_{ij}}{m \hat{p}_i \hat{p}_j} \right) \partial_{ki,j} \right) + E \left(\frac{\hat{p}_{ij} \hat{p}_{ki,j}}{m \hat{p}_{ij} \hat{p}_i \hat{p}_j} \right) \kappa_{ij} \\ &\quad + B_{ki,j} - C_{ki,j} \kappa_{ij}. \end{aligned} \quad (7.70)$$

Comparing (7.70) with (7.55) we see that the bias of the estimator (7.56) is given by

$$E(\hat{\alpha} - \alpha^*) - \text{Var}(\hat{\alpha}^*) = \frac{1}{m^2} \sum_{k,i,j} (B_{ki,j} - C_{ki,j} \kappa_{ij}). \quad (7.71)$$

Noting that for $k = i$, $(1 - \nu_{ki,j}) = (1 - p_{ij})^m$, we can apply Theorem 7.9 to get

$$\begin{aligned} B_{ij} &= \sum_{k,i} B_{ki,j} \\ &= (m-1) \left[\left(1 - \frac{p_{ij}}{p_i}\right) \left(1 - \frac{p_{ij}}{p_j}\right) (1 - p_{ij})^{m-2} + (1 - p_i - p_j + p_{ij})^{m-1} \right]. \end{aligned} \quad (7.72)$$

Similarly, we can define

$$C_{ij} = \sum_{k,i} C_{ki,j}. \quad (7.73)$$

$$= m - (m-1)(1 - p_i - p_j + p_{ij})^{m-1}. \quad (7.74)$$

Thus the bias of the estimator (7.56) under Assumptions C is

$$\frac{1}{m^2} \sum_{i,j \in T \times T} (B_{ij} - C_{ij} \kappa_{ij}). \quad (7.75)$$

Since $B_{ij} = o(1)$ and $C_{ij} = O(m)$ the bias will go to zero as $m \rightarrow \infty$. Whether the bias is positive or negative for finite m is not clear. For moderate sample sizes C_{ij} would dominate B_{ij} , leading to a negative bias if $\kappa_{ij} > 0$. However, if p_{ij} is small, then B_{ij} will decline very slowly, and in such cases κ_{ij} indeed may well be negative.

7.4 Estimating the False Negative Rate

We consider now the estimation of the false negative rate of bioassays of this type. This quantity can be simply estimated as the proportion of carcinogenic compounds that were classified as noncarcinogenic by the decision rule in question. (A similar calculation could be made for mere oncogenicity as well, but we will restrict our attention to malignant responses for the purposes of this exercise.) The problem with this approach is that we do not know which compounds are "in fact" carcinogenic. Even if the identity of the compounds was revealed, it would still be inappropriate to estimate a false negative rate on the basis of compounds that have been *classified* (by whatever criteria) as carcinogenic, since such an approach would beg the question. Although one might investigate the degree of concurrence of various decision procedures, such as those relying solely on the statistics versus those involving biological judgment, it would be fallacious to pretend that one of these procedures gives "the truth" relative to which all others must be judged. Our approach will attempt to determine the false negative rates from the data base itself, without any external references.

To estimate a false negative rate, it is not necessary to know which compounds are carcinogenic. It is sufficient to know the *proportion* of carcinogenic compounds in the population being studied. If this population contains a fraction κ of carcinogens, then the probability of classifying a compound as carcinogenic by a decision rule with size α and power β is

$$c = \alpha(1 - \kappa) + \beta\kappa. \quad (7.76)$$

This equation can then be solved to give the false negative rate as

$$(1 - \beta) = (1 - \alpha) - (c - \alpha)/\kappa. \quad (7.77)$$

In general, the power β will not be a constant but will depend on the *potency* of the compound being studied. Thus β will be a random variable, and the quantity we call the false negative rate in fact represents the *expected* false negative rate for a *particular population of chemicals*. It must be stressed that, unlike the false positive rate which is a characteristic of the test procedure itself, this *average* false negative rate is specific to a particular population of chemicals being studied. For the purposes of this investigation, we will assume that the 49 bioassays in the data base represent a random sample from the relevant population of chemicals.

7.4.1 Abstraction

Let us abstract the nature of our problem. Consider a decision procedure based on a test statistic T whose distribution F_λ depends on a parameter λ from an arbitrary index set Λ . We distinguish an element $0 \in \Lambda$ as representing the *null case*, and call $\Lambda_1 = \Lambda \setminus \{0\}$ the set of *alternatives*. We assume that 0 is *stochastically dominated* by all $\lambda \in \Lambda$ in the sense that $F_\lambda(t) \leq F_0(t)$ for all $\lambda \in \Lambda$, $t \in \mathbb{R}$. We also assume that Λ is endowed with a probability measure H .

If T in fact has distribution F_0 , then $\Pr\{F_0(T) \leq u\} \leq u$. Moreover, if F_0 is continuous, then $\Pr\{F_0(T) \leq u\} = u$, so that $U = F_0(T)$ has a uniform distribution on $[0, 1]$. However,

if T in fact has a distribution $F_\lambda < F_0$, the $\Pr\{F_0(T) \leq u\} \leq \Pr\{F_\lambda(T) \leq u\} = u$, so that the c.d.f of U will be sub-uniform. If tests are performed on a large sample of alternatives chosen from Λ , then the empirical distribution function of U can be used to make inferences about H .

The decision procedures we are discussing are in fact based on the conditional distributions of the test statistic, and the distribution is not continuous. However, we can manoeuvre our situation into the present one by randomisation. The following two lemmas provide the basis for this tactic.

Lemma 7.6 Let T be a random variable with c.d.f. F . For any $u \in [0, 1]$ define $s(u) = \sup\{t : F(t) \leq u\}$. Then

$$\Pr\{F(T) \leq u\} = \begin{cases} u & \text{if } F(s(u)) \leq u \\ F(s(u)^-) & \text{if } F(s(u)) > u. \end{cases} \quad (7.78)$$

Proof: By definition, $T < s(u)$ implies $F(T) \leq u$ whereas $T > s(u)$ implies $F(T) > u$. Thus

$$\Pr\{F(T) \leq u\} = \Pr\{T < s(u)\} + \Pr\{T = s(u), F(T) \leq u\} \quad (7.79)$$

whence

$$\Pr\{F(T) \leq u\} = \begin{cases} \Pr\{T \leq s(u)\} = F(s(u)) & \text{if } F(s(u)) \leq u, \\ \Pr\{T < s(u)\} = F(s(u)^-) & \text{if } F(s(u)) > u. \end{cases} \quad (7.80)$$

However, since $F(s(u)) \geq u$ by definition, $F(s(u)) \leq u$ implies that $F(s(u)) = u$. QED

Lemma 7.7 Let T be a random variable with c.d.f. F . For any $u \in [0, 1]$ define

$$l(u) = \sup\{v \leq u : v = F(t) \text{ for some } t \in [-\infty, \infty]\}.$$

Then $\Pr\{F(T) \leq u\} = l(u)$.

Proof: By definition, for any $\epsilon > 0$, there must be a t_0 such that

$$l(u) - \epsilon < F(t_0) \leq l(u) \leq u.$$

Let $s(u)$ be defined as in Lemma 7.6. If $F(s(u)) > u$ then $t_0 < s(u)$. But in that case, for any t such that $t_0 < t < s(u)$, $F(t) \leq u$, and thus

$$l(u) - \epsilon < F(t_0) \leq F(t) \leq l(u).$$

Since ϵ is arbitrary, we have that $l(u) = F(s(u)^-)$. On the other hand, if $F(s(u)) \leq u$, then obviously $l(u) = u$. In either case, the previous Lemma 7.6 gives us $\Pr\{F(T) \leq u\} = l(u)$. QED

The function $s(u)$ in Lemma 7.6 indicates how the c.d.f. behaves at the value u . If $t > s(u)$ then $F(t^-) > u$ and conversely. On the other hand, if $t < s(u)$ then $F(t) \leq u$ although here the converse need not hold, and $F(t^-) \leq u < F(t)$ implies that $t = s(u)$. The only ambiguous case is $u = F(t)$ which may occur either with $t = s(u)$ or $t < s(u)$. We use these properties to prove the following theorem.

Theorem 7.10 Let T be a random variable with c.d.f. F , and let Z be a random variable, uniform on $[0, 1]$, independent of T . Then the random variable $U = ZF(T) + (1 - Z)F(T^-)$ has a uniform distribution on $[0, 1]$.

Proof:

$$\begin{aligned} \Pr\{U \leq u\} &= \Pr\{U \leq u | F(T) \leq u\} \Pr\{F(T) \leq u\} \\ &\quad + \Pr\{U \leq u | F(T^-) \leq u < F(T)\} \Pr\{F(T^-) \leq u < F(T)\} \\ &\quad + \Pr\{U \leq u | F(T^-) > u\} \Pr\{F(T^-) > u\}. \end{aligned}$$

Clearly the first conditional probability is unity, whereas the last is zero. Thus

$$\begin{aligned} \Pr\{U \leq u\} &= \Pr\{F(T) \leq u\} \\ &\quad + \Pr\{U \leq u | F(T^-) \leq u < F(T)\} \Pr\{F(T^-) \leq u < F(T)\}. \end{aligned} \quad (7.81)$$

Now,

$$\Pr\{U \leq u | F(T^-) \leq u < F(T)\} = \Pr\left\{Z \leq \frac{u - F(T^-)}{F(T) - F(T^-)} \mid F(T^-) \leq u < F(T)\right\}.$$

However, $F(T^-) \leq u < F(T)$ implies that $T = s(u)$, and thus

$$\begin{aligned} \Pr\{U \leq u | F(T^-) \leq u < F(T)\} &= \frac{u - F(s(u)^-)}{F(s(u)) - F(s(u)^-)} \\ &= \frac{u - l(u)}{F(s(u)) - F(s(u)^-)} \end{aligned}$$

by Lemmas 7.6 and 7.7.

Now $\Pr\{F(T) \leq u\} = l(u)$ by Lemma 7.7, and

$$\begin{aligned} \Pr\{F(T^-) \leq u < F(T)\} &= \Pr\{T = s(u)\} - \Pr\{T = s(u), F(T) = u\} \\ &= F(s(u)) - F(s(u)^-) - \Pr\{T = s(u), F(T) = u\}. \end{aligned}$$

Thus from (7.81) we have

$$\Pr\{U \leq u\} = u + \frac{u - l(u)}{F(s(u)) - F(s(u)^-)} \Pr\{T = s(u), F(T) = u\}.$$

However, if $F(s(u)) = u$, then $l(u) = u$, so we have $\Pr\{U \leq u\} = u$. QED

Now, let T be a test statistic whose null distribution conditional on S is F^S (not necessarily continuous), and let Z be a uniform random variable on $[0, 1]$ independent of (T, S) , and define

$$U = ZF^S(T) + (1 - Z)F^S(T^-).$$

Then, by Theorem 7.10

$$\Pr\{U \leq u\} = E(\Pr\{U \leq u | S\}) = E(u) = u$$

and thus U has a uniform distribution.

In the sequel, we will assume that this kind of "random fuzzing" has been done, and hence that the distribution of U is continuous.

7.4.2 Distribution of P -values

Suppose now that the distribution of T is F_λ , where λ is chosen randomly according to H . Define

$$S(u) = \Pr\{U \leq u\} = \int F_\lambda(F_0^{-1}(u)) dH(\lambda). \quad (7.82)$$

(If F_0 is not strictly increasing, we define $F_0^{-1}(u) = \inf\{t : F(t) \geq u\}$.) Let $\kappa = H(\Lambda_1)$ be the probability of alternatives, and write

$$S(u) = (1 - \kappa)u + \kappa S_1(u) \quad (7.83)$$

where

$$S_1(u) = \frac{1}{\kappa} \int_{\Lambda_1} F_\lambda(F_0^{-1}(u)) dH(\lambda). \quad (7.84)$$

Suppose that the decision rule states that $\lambda \neq 0$ whenever $U > u_0$. Then the false positive rate will be $1 - u_0$, and the false negative rate will be

$$\Pr\{U \leq u_0 | \lambda \neq 0\} = S_1(u_0). \quad (7.85)$$

Thus

$$S(u_0) = (1 - \kappa)u_0 + \kappa S_1(u_0) \quad (7.86)$$

or

$$\kappa = \frac{u_0 - S(u_0)}{u_0 - S_1(u_0)} \geq 1 - \frac{S(u_0)}{u_0} \quad (7.87)$$

since $S_1(u_0) \leq u_0$. We now consider conditions under which we can say more.

Theorem 7.11 *If F_λ is differentiable and*

$$\frac{F'_\lambda(x)}{F'_0(x)} \rightarrow 0$$

as $x \rightarrow L$ for all $\lambda \in \Lambda_1$, where $L = \inf\{x : F_0(x) > 0\}$, then $\lim_{u \downarrow 0} \frac{S(u)}{u} = S'(u) = 1 - \kappa$.

Proof:

$$\frac{d}{du} F_\lambda(F_0^{-1}(u)) = \frac{F'_\lambda(F_0^{-1}(u))}{F'_0(F_0^{-1}(u))}.$$

Letting $x = F_0^{-1}(u)$, as $u \rightarrow 0$, $x \rightarrow L$, so that

$$\lim_{u \downarrow 0} \frac{S(u)}{u} = \frac{d}{du} F_\lambda(F_0^{-1}(u)) = \lim_{x \rightarrow L} \frac{F'_\lambda(x)}{F'_0(x)} = 0.$$

Now

$$F_\lambda(F_0^{-1}(u)) \leq F_0(F_0^{-1}(u)) = u$$

3 of/de 3

so, by bounded convergence

$$\begin{aligned}
 \lim_{u \downarrow 0} \frac{S(u)}{u} &= \lim_{u \downarrow 0} \frac{1}{u} \int F_{\lambda}(F_0^{-1}(u)) dH(\lambda) \\
 &= \int \lim_{u \downarrow 0} \frac{F_{\lambda}(F_0^{-1}(u))}{u} dH(\lambda) \\
 &= \int_{\{\lambda=0\}} dH(\lambda) \\
 &= 1 - \kappa.
 \end{aligned}$$

QED

Theorem 7.12 *If F_{λ} is a (one-sided) location family, ($F_{\lambda}(x) = F_0(x - \lambda)$ for $\lambda \geq 0$) and $(\log F'(x))' \rightarrow \infty$ as $x \rightarrow -\infty$ then the conditions of Theorem 7.11 hold.*

Proof:

$$\log \frac{F'_{\lambda}(x)}{F'_0(x)} = \log F'_0(x - \lambda) - \log F'_0(x).$$

By the mean value theorem

$$\log F'_0(x - \lambda) = \log F'_0(x) - \lambda \frac{F''_0(\xi)}{F'_0(\xi)}$$

for some $x - \lambda \leq \xi \leq x$. Thus

$$\log \frac{F'_{\lambda}(x)}{F'_0(x)} = -\lambda \frac{F''_0(\xi)}{F'_0(\xi)} = -\lambda (\log F'_0(\xi))'.$$

Now, as $x \rightarrow -\infty$, $\xi \rightarrow -\infty$, so that $\log \frac{F'_{\lambda}(x)}{F'_0(x)} \rightarrow -\infty$ and $\frac{F'_{\lambda}(x)}{F'_0(x)} \rightarrow 0$.

QED

Theorem 7.13 *If $\{F_{\lambda}, \lambda \in \Lambda\}$ is a one-sided location family and $-\log F_0(x)$ is convex, then $S(u)/u$ is nondecreasing in u .*

Proof:

$$\frac{d}{du} \frac{F_{\lambda}(F_0^{-1}(u))}{u} = \frac{u F'_{\lambda}(F_0^{-1}(u)) / F'_0(F_0^{-1}(u)) - F_{\lambda}(F_0^{-1}(u))}{u^2}.$$

Now, letting $x = F_0^{-1}(u)$, we have

$$\frac{d}{du} \frac{F_{\lambda}(F_0^{-1}(u))}{u} = \frac{1}{F_0(x)} \left(\frac{F'_{\lambda}(x)}{F'_0(x)} - \frac{F_{\lambda}(x)}{F_0(x)} \right). \quad (7.88)$$

If $F_{\lambda}(x) = F_0(x - \lambda)$ then the factor in parentheses in (7.88) becomes

$$\frac{F'_0(x - \lambda)}{F'_0(x)} - \frac{F_0(x - \lambda)}{F_0(x)}.$$

If $-\log F_0(x)$ is convex, then $F'_0(x)/F_0(x)$ is decreasing, and so

$$\frac{F'_0(x-\lambda)}{F_0(x-\lambda)} > \frac{F'_0(x)}{F_0(x)}. \quad (7.89)$$

Comparing (7.89) with (7.88) shows that $\frac{d}{du} u^{-1} F_\lambda(F_0^{-1}(u)) > 0$. Integration over λ then establishes the result. QED

Many common distributions enjoy the properties required by these theorems. Theorem 7.11 requires that the power of the procedure goes to 1 as the size goes to 1, Theorem 7.12 that the distribution has "shorter than exponential" tails, and Theorem 7.13 that the c.d.f. is log concave. In particular, the normal distribution satisfies these properties. The following discussion therefore applies to it, and by approximation, to any decision procedure depending on an asymptotically normal test statistic.

7.4.3 Empirical distribution of P -values

Given a sample $\lambda_1, \dots, \lambda_m$ from Λ , with independent test statistics $T_1, \dots, T_m$, we can infer properties of H , such as the value of $\kappa = H(\Lambda_1)$, by considering $\hat{S}$, the empirical distribution function of $U_1, \dots, U_m$, where $U_i = F_0(T_i)$, ($i = 1, \dots, m$). We see from (7.87) that for any u (since $\hat{S}(u)$ is unbiased for $S(u)$), $\hat{S}(u)/u$ is an overestimate of $1 - \kappa$. However, if the assumptions of Theorems 7.11–7.13 hold, then this bias will diminish as $u \rightarrow 0$. By contrast, the variance of $\hat{S}(u)/u$ is

$$\frac{S(u)(1-S(u))}{mu^2} = \frac{S(u)}{u} \frac{(1-S(u))}{mu} \quad (7.90)$$

which, since $S(u)/u \geq 1 - \kappa$, will become infinite as $u \rightarrow 0$ if $\kappa < 1$. The attempt at estimating $1 - \kappa$ is thus caught between the jaws of reducing bias and improving precision. The point at which some compromise criterion such as the mean squared error is minimized cannot be determined in general since it would depend on S itself. If most of the alternatives are "distant", then $S(u)$ will be essentially $(1 - \kappa)u$ over a wide range of u 's, and thus a relatively large value of u will still have low bias. On the other hand, if the population contains a large proportion of "near alternatives" ($F_\lambda \approx F_0$), then S will have a pronounced curvature at the origin, and a lower value of u will be required to reduce bias.

Rather than choosing an arbitrary point u , and hoping it to be near the optimum, we propose an estimator which integrates the empirical distribution function over a range (a, b) :

$$\hat{\kappa} = 1 - (b-a)^{-1} \int_a^b \frac{\hat{S}(u)}{u} du. \quad (7.91)$$

Because $\hat{S}$ is a step function, the integral in (7.91) can actually be written as a finite sum. To derive this form, we require the following lemma.

Lemma 7.8 Let F be an increasing step function on (a, b) with unit jumps at $x_1, \dots, x_n$, ($a \leq x_1 < x_2 < \dots < x_n \leq b$) and let g be a Riemann integrable function with indefinite integral G . Then

$$\int_a^b F(x)g(x) dx = F(b)G(b) - F(a)G(a) - \sum_{i=1}^n G(x_i). \quad (7.92)$$

Proof:

$$\begin{aligned} \int_a^b F(x)g(x) dx &= \int_a^{x_1} F(a)g(x) dx \\ &\quad + \sum_{i=1}^{n-1} \int_{x_i}^{x_{i+1}} F(x_i)g(x) dx + \int_{x_n}^b F(x_n)g(x) dx \\ &= F(a)(G(x_1) - G(a)) \\ &\quad + \sum_{i=1}^{n-1} F(x_i)(G(x_{i+1}) - G(x_i)) + F(x_n)(G(b) - G(x_n)) \\ &= -F(a)G(a) + (F(a) - F(x_1))G(x_1) + \sum_{i=2}^{n-1} (F(x_{i-1}) - F(x_i))G(x_i) \\ &\quad + (F(x_{n-1}) - F(x_n))G(x_n) + F(b)G(b) \\ &= F(b)G(b) - F(a)G(a) - \sum_{i=1}^n G(x_i). \end{aligned}$$

QED

Theorem 7.14 Let $\hat{S}$ be the empirical distribution function of a sample $U_1, \dots, U_m$. Then

$$(b-a)(1-\hat{\kappa}) = \int_a^b \frac{\hat{S}(u)}{u} du = \sum_{j=r}^s w_j y_j \quad (7.93)$$

where

$$\begin{aligned} r &= m\hat{S}(a), & s &= m\hat{S}(b), \\ w_r &= \hat{S}(a), & y_r &= \log(b/a), \\ w_j &= \frac{1}{m}, & y_j &= \log\left(\frac{b}{U_{(j)}}\right), \quad j = r+1, \dots, s, \end{aligned}$$

$U_{(j)}$ denoting the j^{th} order statistic of the U_j 's.

Proof: Using Lemma 7.8 we have

$$\int_a^b \frac{\hat{S}(u)}{u} du = \hat{S}(b) \log b - \hat{S}(a) \log a - m^{-1} \sum_{j=r+1}^s \log U_{(j)}. \quad (7.94)$$

Substituting into (7.94) the fact that $\hat{S}(b) = \hat{S}(a) + \sum_{j=r+1}^s m^{-1}$ gives

$$\hat{S}(a) \log \frac{b}{a} + m^{-1} \sum_{j=r+1}^s \log \frac{b}{U_{(j)}}$$

from which (7.93) follows. QED

The variance of $\hat{K}$ can also be determined. One can write

$$\frac{\hat{S}(u)}{u} = \frac{S(u)}{u} + \frac{\hat{S}(u) - S(u)}{u} \quad (7.95)$$

$$= \frac{S(u)}{u} + \frac{m^{-1/2}B(u)}{u} \quad (7.96)$$

where $B(u)$ is an empirical process with mean 0 and covariance function

$$\text{Cov}(B(u), B(v)) = S(u) \wedge S(v) - S(u)S(v).$$

The variance of $\int_a^b B(u)/u$ is then given by

$$V = \int_a^b \int_a^b \frac{S(u) \wedge S(v) - S(u)S(v)}{uv} du dv \quad (7.97)$$

$$= 2 \int_a^b \frac{S(u)}{u} \int_u^b \frac{dv}{v} du - \left[\int_a^b \frac{S(u)}{u} du \right]^2 \quad (7.98)$$

Since $\hat{S}$ is consistent for S , this quantity can be estimated by substituting the empirical distribution function into (7.97). The resulting integral can be simplified by the following result.

Theorem 7.15

$$2 \int_a^b \frac{\hat{S}(u)}{u} \int_u^b \frac{dv}{v} du = \sum_{i=r}^s w_i y_i^2 \quad (7.99)$$

where r, s, w_i , and y_i are as in Theorem 7.14.

Proof:

$$\int_a^b \frac{\hat{S}(u)}{u} \int_u^b \frac{dv}{v} du = \log b \int_a^b \frac{\hat{S}(u)}{u} du - \int_a^b \frac{\hat{S}(u)}{u} \log u du. \quad (7.100)$$

Since the indefinite integral of $u^{-1} \log u$ is $\frac{1}{2}(\log u)^2$, we get, by Lemma 7.8, that

$$\begin{aligned} 2 \int_a^b \hat{S}(u) \frac{\log u}{u} du &= \hat{S}(b)(\log b)^2 - \hat{S}(a)(\log a)^2 - m^{-1} \sum_{i=r+1}^s (\log U_{(i)})^2 \\ &= \hat{S}(a) [(\log b)^2 - (\log a)^2] - m^{-1} \sum_{i=r+1}^s [(\log U_{(i)})^2 - (\log b)^2]. \end{aligned} \quad (7.101)$$

Now

$$\begin{aligned} (\log U_{(i)})^2 - (\log b)^2 &= \left(\log \frac{b}{U_{(i)}} \right)^2 - 2 \left[\log b \log \frac{b}{U_{(i)}} \right] \\ &= y_i^2 - 2y_i \log b. \end{aligned}$$

Thus

$$2 \int_a^b \hat{S}(u) \frac{\log u}{u} du = - \sum_{i=r}^s w_i y_i^2 + 2 \log b \sum_{i=r}^s w_i y_i$$

and

$$2 \int_a^b \frac{\hat{S}(u)}{u} \int_u^b \frac{dv}{v} du = 2 \log b \sum_{i=r}^n w_i y_i + \sum_{i=r}^n w_i y_i^2 - 2 \log b \sum_{i=r}^n w_i y_i$$

$$\pm \sum_{i=r}^n w_i y_i^2.$$

QED

Corollary 7.15.1 A consistent estimator of $\text{Var}(\hat{\kappa})$ is given by

$$\frac{\sum w_i y_i^2 - (\sum w_i y_i)^2}{m(b-a)}$$

where w_i and y_i are defined in Theorem 7.14.

7.5 Error Rates for the Decision Rules

We now return to the practical problem of calculating the error rates for the four decision rules of Section 7.2.3, using the methodology we have developed. The proportion of carcinogens in the data base is estimated as $\hat{\kappa}$ according to (7.91). For this purpose we used for U_i the quantity

$$U_i = (1 - \min(P_R^{im}, P_R^{if}))^2 \quad (7.102)$$

where P_R^{im} is the exact P -value for trend for aggregate malignant tumours for the males in study i , randomized as provided for in Theorem 7.10, and P_R^{if} is the corresponding quantity for females. Under the null hypothesis of no treatment effect in either sex, U_i will have a uniform distribution. The range (a, b) of integration of $\hat{S}(u)/u$ was somewhat arbitrarily chosen to be $(.5, .9025)$. The selection of the upper limit (equal to .95²) was prompted by the wish to unequivocally include in the estimation of κ all studies in which at least one sex had a significant trend ($P_e < .05$) in malignant tumours.

The false positive rate was estimated as described in Section 7.3.4 and then (7.77) was used to estimate the false negative rate, where c was estimated, for each decision rule, as the proportion of compounds that were classified as carcinogenic. Because the empirical process B is asymptotically a Gaussian process, it follows that $\hat{\kappa}$, being a linear functional, will be asymptotically normal. Thus $\text{Var}(\hat{\kappa})$, estimated as in Corollary 7.15.1 could be used for calculating the "confidence interval" on κ were it not for the unknown bias in $\hat{\kappa}$. The lower limit $\hat{\kappa} + \Phi^{-1}(\gamma)\sqrt{\text{Var}(\hat{\kappa})}$ would be a $1 - \gamma$ confidence limit for κ , but the upper limit would not be.

We are less comfortable about the distribution of the false positive rate estimates, but nevertheless use the variance estimator (7.56) to put a confidence interval on $\hat{\kappa}$. Because of these difficulties, we avoid the term "confidence limit" and speak instead of "plausible limits". The estimator of c is, of course, just a binomial proportion, so its variance presents no problem. We can then put 95% plausible limits on the false negative rate by putting

Table 7.5: Estimated error rates (%) for four decision rules. (Plausible limits are given in parentheses).

	False positive rate	False negative rate
Rule I	21 (6-26)	0 (0-54)
Rule II	10 (6-10)	23 (0-78)
Rule III	2 (0.3-2)	0 (0-62)
Rule IV	4 (3-6)	1 (0-68)

upper and lower 99.2% plausible limits on κ , α , and c , and taking the maximum and minimum values consistent with these bounds. Evidently, this is a conservative procedure.

The estimates of false positive and false negative rates are summarized for the four decision rules in Table 7.5. The results are quite vividly ambiguous. It does appear, as may have been anticipated, that Rule I has a rather high false positive rate, and Rule II a high false negative rate. The other two rules appear more reasonable, but the rather wide range of the plausible limits gives little cause for comfort. Evidently more research is required before making definitive statements about the operating characteristics of cancer bioassays.

Appendix A

Componentwise and Tensor Products of Matrices

It is convenient, for certain derivations, to have a notation for operations on matrices other than the usual matrix addition and multiplication. We describe two such operations, and list their basic properties. The proofs of these properties are elementary.

The *component-wise product* (sometimes called "Hadamard product") is defined for two $n \times m$ matrices A and B with elements a_{ij} and b_{ij} , as an $n \times m$ matrix $A * B$ whose ij^{th} component is $a_{ij}b_{ij}$. It is obvious that the product $*$ is commutative and associative, and distributive over matrix addition. The matrix of ones J is the identity element of $*$. The componentwise product does not have a simple relationship with matrix multiplication, except in special cases.

Matrix multiplication by *diagonal* matrices commutes with $*$: $A * (DB) = D(A * B)$ if D is a diagonal matrix. In general, however, $A * (BC) \neq (A * B)C$. Note that the $*$ -product and the matrix product are equivalent for diagonal matrices.

If A and B are vectors (i.e. $m = 1$) then $A * B$ can be considered as the ordinary matrix product $D_A B$ where D_A is a diagonal matrix with the diagonal elements equal to the components of A . Another useful case involves the *outer product* AB' of two vectors. It is easy to show that if A, B, C , and D , are (column) vectors then $(A * B)(C * D)' = (AC') * (BD')$.

The *tensor product*, (also called "direct product", or "Kronecker product") of two matrices is an extension of the outer product concept to matrices. If A is an $n_1 \times m_1$ matrix with components a_{ij} , and B is an $n_2 \times m_2$ matrix with components b_{kl} , then the tensor product $A \otimes B$ is defined as an $n_1 n_2 \times m_1 m_2$ matrix with components $a_{ij}b_{kl}$. It is convenient to refer to the components by double subscripts, where by the $(ik, jl)^{\text{th}}$ component we mean the number in the $i + (k - 1)n_1^{\text{th}}$ row and the $j + (l - 1)m_1^{\text{th}}$ column. With this notation, we say that $a_{ij}b_{kl}$ is the $(ik, jl)^{\text{th}}$ component of $A \otimes B$. (For most operations, the actual *order* of the rows and columns of $A \otimes B$ is irrelevant, provided a consistent convention is adopted. Some authors would write the above defined operation as $B \otimes A$. Although the tensor product is not commutative in the usual sense, $B \otimes A$ can be obtained from $A \otimes B$ merely by an appropriate permutation of rows and columns, and thus the two representations are equivalent for most purposes.)

The tensor product is associative, and is distributive over matrix addition. It also

"commutes" with both the componentwise and matrix product in the sense that

$$(A * B) \otimes (C * D) = (A \otimes C) * (B \otimes D)$$

and

$$(AB) \otimes (CD) = (A \otimes C)(B \otimes D)$$

whenever the products are defined. One more fact we find useful is that $\text{tr}(A \otimes B) = (\text{tr } A)(\text{tr } B)$.

Appendix B

Fourth Order Mixed Moments

Given a random vector $W = (W_1, W_2, \dots, W_m)$, consider the fourth order mixed moments,

$$\alpha_{ijkl} = E(W_i W_j W_k W_l).$$

If the random variables $W_1, W_2, \dots, W_m$ are exchangeable, then α_{ijkl} only depends on how many of the subscripts i, j, k, l are equal. We can thus define:

$$E(W_i W_j W_k W_l) = \begin{cases} E(W_i^4) = \gamma_{44} & \text{if } i = j = k = l; \\ E(W_i^3 W_j) = \gamma_{31} & \text{if } i = j = k \neq l, \text{ or } i = \\ & j = l \neq k, \text{ or } i = k = \\ & l \neq j, \text{ or } i = k = l \neq i; \\ E(W_i^2 W_j^2) = \gamma_{22} & \text{if } i = j \neq k = l, \text{ or } \\ & i = k \neq j = l, \text{ or } i = \\ & l \neq j = k; \\ E W_i^2 W_j W_k = \gamma_{211} & \text{if } l \neq i = j \neq k \neq l, \text{ or } \\ & i \neq j = k \neq l \neq i, \text{ or } \\ & j \neq i = k \neq l \neq k, \text{ or } \\ & i \neq j = l \neq k \neq i, \text{ or } \\ & j \neq i = l \neq k \neq j, \text{ or } \\ & i \neq l = k \neq j \neq i; \\ E(W_i W_j W_k W_l) = \gamma_{1111} & \text{if } i, j, k, l \text{ are all dis-} \\ & \text{tinct.} \end{cases} \quad (B.1)$$

It is convenient, for algebraic manipulations, to write $E(W_i W_j W_k W_l)$ as a linear combination of the γ 's with weights defined by generalized Kronecker deltas defined by

$$\delta^{i_1, i_2, \dots, i_r} = \begin{cases} 1 & \text{if } i_1 = i_2 = \dots = i_r \\ 0 & \text{otherwise.} \end{cases} \quad (B.2)$$

If we let $\gamma' = (\gamma_4, \gamma_{31}, \gamma_{22}, \gamma_{211}, \gamma_{1111})$,

$$\delta_{ijkl} = \begin{bmatrix} \delta^{ijkl} \\ \delta^{ijk} \\ \delta^{ijl} \\ \delta^{ikl} \\ \delta^{jkl} \\ \delta^{ij}\delta^{kl} \\ \delta^{ik}\delta^{jl} \\ \delta^{il}\delta^{jk} \\ \delta^{ij} \\ \delta^{ik} \\ \delta^{il} \\ \delta^{jk} \\ \delta^{jl} \\ \delta^{kl} \\ 1 \end{bmatrix}, \quad (B.3)$$

and

$$H = \begin{bmatrix} 1 & -4 & -3 & 12 & -6 \\ 0 & 1 & 0 & -3 & 2 \\ 0 & 1 & 0 & -3 & 2 \\ 0 & 1 & 0 & -3 & 2 \\ 0 & 1 & 0 & -3 & 2 \\ 0 & 0 & 1 & -2 & 1 \\ 0 & 0 & 1 & -2 & 1 \\ 0 & 0 & 1 & -2 & 1 \\ 0 & 0 & 0 & 1 & -1 \\ 0 & 0 & 0 & 1 & -1 \\ 0 & 0 & 0 & 1 & -1 \\ 0 & 0 & 0 & 1 & -1 \\ 0 & 0 & 0 & 1 & -1 \\ 0 & 0 & 0 & 1 & -1 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix} \quad (B.4)$$

then direct verification of the above possibilities shows that

$$E(W_i W_j W_k W_l) = \delta'_{ijkl} H \gamma. \quad (B.5)$$

As an example, consider the case $i \neq j = k = l$. Here

$$\delta'_{ijkl} = (0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 1, 1, 1, 1),$$

so $\delta'_{ijkl} H = (0, 1, 0, 0, 0)$, and $\delta'_{ijkl} H \gamma = \gamma_{31}$ as required.

When dealing with multiple subscripts, matrix notation becomes unworkable and use of summation symbols becomes unwieldy. We will thus employ the "summation convention" [58], where an expression such as

$$a_{ijk} b^{jkl} \quad (B.6)$$

containing both subscripts and superscripts is understood to be summed over all values of any indices that appear in both subscript and superscript. (The indices need not represent consecutive integers, but may be from any (finite) index set of symbols.) Thus the expression (B.6) means

$$\sum_j \sum_k a_{ijk} b^{jkl}$$

To exploit this notation, we will employ symbols

$$\delta_S^{i_1, \dots, i_q}$$

where S is a set of subsets of the integers $\{1, \dots, q\}$. We define

$$\delta_S^{i_1, \dots, i_q} = \begin{cases} 1 & \text{if for every } S \in S, j \in S, \text{ and } k \in S \text{ implies} \\ & \text{that } i_j = i_k. \\ 0 & \text{otherwise} \end{cases} \quad (\text{B.7})$$

Note that $\delta_{\{\emptyset\}}^{i_1, \dots, i_q} = 1$. When there is no confusion, we will abuse notation by omitting the braces and commas. Thus we write δ_{123}^{ijkl} for $\delta_{\{1,2,3\}}^{ijkl}$, and $\delta_{\{13\}\{24\}}^{ijkl}$ for $\delta_{\{\{1,3\}, \{2,4\}\}}^{ijkl}$, these symbols representing δ^{ijk} and $\delta^{ik}\delta^{jl}$ respectively. With these conventions, we can write

$$E(W, W_j W_k W_l) = \delta_{\omega}^{ijkl} h^{\omega\psi} \gamma_{\psi}, \quad (\text{B.8})$$

where ω is one of the symbols

1234	{12}{34}	14
123	{13}{24}	23
124	{14}{23}	24
134	12	34
234	13	0,

ψ is one of the symbols 4, 31, 22, 211, 1111 and $h^{\omega\psi}$ are the elements of the matrix defined in (B.4).

Appendix C

Computational Considerations

C.1 Computational complexity

Computational complexity refers to the time required to complete a given algorithm as a function of the "size" of the problem. The size is not a clearly defined concept but is some quantity that could be expected to be large for situations to which the algorithm applied. In our applications of computing univariate probabilities, size could be measured by $n + 1$, the number of support points for the distribution. In the computations of multivariate probabilities discussed in Chapters 5 and 6, size could be variously measured by the number of animals n , the number of litters m , or the litter size l .

Given a measure of size (which we denote by n for definiteness), an algorithm may have the property that T , the time required for completion can be estimated as

$$T(n) = An^k + o(n^k) \quad (C.1)$$

as $n \rightarrow \infty$. If such is the case, we say the algorithm is of *polynomial complexity* and call k the *order* of the algorithm. Orders of 1, 2, ... are called *linear*, *quadratic* etc.

For simple algorithms, the complexity can be determined merely by counting the number of operations. Although the actual running time will depend on the language in which the algorithm is implemented, programming techniques, and the hardware on which it is ultimately run, such differences generally affect only the constant A in (C.1). The time for ordinary arithmetic operations is usually fairly constant in most processors, so that in determining complexity one can take "an operation" as being a basic unit. Some machines may have appreciable differences between execution times for floating point as opposed to integer arithmetic, but unless the algorithm involves switching from one type of operation to another as a function of n , such differences will not affect the complexity.

There are situations, however, where complexity can be implementation dependent. For example, calculating x^n could be implemented as an iterative product (of complexity n), as an adaptive sequence of squarings and products (of complexity $\log n$), or by use of logarithms and exponentiation (of constant complexity). Typically, the constant terms of these algorithms would be increasing in the above stated order, so that the first strategy would be fastest for small n , whereas the last would be fastest for large n . The complexity of an algorithm that involved taking n^{th} powers would then depend on which strategy the compiler used in generating computer code from "Y=X**N".

C.2 Round-off errors and underflows

Two practical matters must be considered in using the recursive calculations in actual computations. One is the possibility of underflow, the other the accumulation of round-off errors. Both problems are the result of the way floating-point numbers are handled in computers. In general, a real number $x \neq 0$ is stored as

$$\tilde{x} = m \times r^c \text{ if } m \leq xr^{-c} < m+1, \quad (\text{C.2})$$

where m and c are integers, $2^t/r \leq |m| < 2^t$, and r is usually a power of 2. Such a representation has an error $|\tilde{x} - x| \leq r^c$ so that the *relative error*

$$\left| \frac{\tilde{x} - x}{x} \right| \leq \frac{r^c}{2^t r^{c-1}} = r2^{-t}. \quad (\text{C.3})$$

The parameters t and r depend on the hardware being used, but typically $t \approx 24$ and $r = 2$ or 16 . So-called *double-precision* numbers may have $t = 56$ or more. (For some computers, use of double-precision numbers may appreciably slow down computation.) The maximum relative error $\sigma = r2^{-t}$ is thus a function of the hardware, and is referred to as the *machine precision*.

Errors will propagate in calculations. If $\tilde{x}$ and $\tilde{y}$ are two floating-point representations with relative errors σ_1 and σ_2 , then the relative error of $\tilde{x} \times \tilde{y}$ or $\tilde{x}/\tilde{y}$ is approximately bounded by $\sigma_1 + \sigma_2$. Thus to estimate the error resulting from a long series of multiplications and divisions, one can multiply the machine precision by the number of operations. Such an estimate may be fairly crude, because it is based on the *maximum possible error*. (However, since the sum of the errors only *approximately* bounds the error of the product, this estimate is not, strictly speaking, the maximum error possible.) If one assumes that the errors are independent random variables, then a fuzzier but more optimistic error estimate can be obtained from the central limit theorem. Unfortunately, in the sort of recursive calculations we are using, the assumption of *independent* errors may be quite untenable, and it is not difficult to construct examples in which the "worst case" error estimates are nearly achieved.

Error propagation can be more serious with addition and subtraction. In this case it is the *absolute errors* that add, and the relative error can become quite startling if the magnitude of the result is much smaller than that of the larger operand. For this reason, addition of quantities with opposite signs requires special attention in the implementation of algorithms.

There are maximum and minimum values that the exponent c in the representation (C.2) can have. If an arithmetic operation gives a result for which c would have to be less than this minimum value, the operation is said to produce an *underflow*. The consequence of an underflow depends on the computer system. In some cases the result is set to zero, in other cases the program grinds to a halt. In any case, the result is useless for any further recursive calculations. Thus the implementation of any algorithm that might produce very small numbers has to take account of the possibility of underflows. (The same considerations apply, of course, if the *maximum* value of c is exceeded, producing an *overflow*. However, since we are dealing with probabilities, it is small numbers that are more likely to cause problems.)

The simplest way to handle underflows in a recursive calculation such as (4.12) is to rescale the probabilities when they become too small. Suppose that ζ is the smallest positive number that can be stored in the computer, and suppose that there exists a number r such that $\sqrt{\zeta} < r < c(x) = P(x+1)/P(x)$ for all x . Then we can represent a probability $P(x)$ by the pair $(P^*(x), k)$ where $r < P^* \leq 1$ and $P(x) = P^*(x)r^k$. Then $P(x+1) = c(x)P(x) = c(x)P^*(x)r^k$, and $c(x)P^*(x) > \zeta$. If $c(x)P^*(x) < r$, we represent $P(x+1)$ by $(P^*(x+1), k+1)$ where $P^*(x+1) = c(x)P^*(x)/r$. Should $c(x)P^*(x+1) > 1$, we represent $P(x+1)$ by $(P^*(x+1), k-1)$ where $P^*(x+1) = c(x)P^*(x)r$.

Difficulties can arise with rescaling when an appropriate r cannot be found. In the Pólya family (4.1), such a problem may occur if $a = -\epsilon$ and $b < -1$ for some small $\epsilon > 0$. Such a case, however, corresponds to a distribution having nearly all of its mass at the origin; and hence can be practically approximated by a one-point distribution.

Rescaling is not an ideal solution. Because it requires more computation, and contributes to round-off error, it should be avoided if possible. Since probabilities so small as to cause underflow are practically negligible, the only reason to calculate them at all is for recursive calculation of other probabilities. If these negligible probabilities can be truncated beforehand, then rescaling may be avoided. This is the rationale behind the algorithm presented in Section 4.3.3.

Bibliography.

- [1] Andersen, Per Kragh. "On the application of the theory of counting processes in the statistical analysis of censored survival data." In *Survival Analysis*. Edited by John Crowley and Richard A. Johnson. Hayward, California: Institute of Mathematical Statistics, 1982, p. 1-13.
- [2] Armitage, P. 1955. Tests for linear trend in proportions and frequencies. *Biometrics* 11: 375-86.
- [3] Armitage, P. 1985. Multistage models of carcinogenesis. *Env. Health Perspectives* 63: 195-201.
- [4] Armitage, P. and R. Doll. 1954. The age distribution of cancer and a multi-stage theory of carcinogenesis. *Br. J. Cancer* 8: 1-12.
- [5] Armitage, P. and R. Doll. "Stochastic models for carcinogenesis." In *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability*, vol 4. Edited by J. Neyman. Berkeley: U. of California Press, 1961, p. 19-38.
- [6] Arnold, D. L., D. Krewski, B. Junkins, P. F. McGuire, C. A. Moodie, I. C. Munro. 1983. Reversibility of ETU induced thyroid hyperplasia. *Toxicol. Appl. Pharmacol.* 67: 264-73.
- [7] Arnold, D. L., E. Farber, D. Krewski. 1988. Carcinogenicity Testing: Histopathology and the Blind Method. *Comments Toxicology* 2: 67-80.
- [8] Bellhorn, R. W. 1980. Lighting in the animal environment. *Lab. Anim. Sci.* 30: 440-50.
- [9] Berkson, Joseph. 1978. In dispraise of the exact test. *J. Stat. Planning and Inference* 2: 27-42.
- [10] Berkson, Joseph. 1978. Do the marginal totals of the 2×2 table contain relevant information respecting the table proportions?. *J. Stat. Planning and Inference* 2: 43-44.
- [11] Berry, G. 1975. Design of carcinogenesis experiments using the Weibull distribution. *Biometrika* 62: 321-7.
- [12] Bickis, M. "Efficiency of censoring and dichotomization." *Statistical Society of Canada, Annual Meeting*. Winnipeg, 1985.

- [13] Bickis, Mikelis and Daniel Krewski. "Statistical design and analysis of the long-term carcinogenicity bioassay." In *Toxicological Risk Assessment, vol. 1. Biological and Statistical Criteria*. Edited by D. B. Clayson, D. Krewski, I. Munro. Boca Raton, Florida: CRC Press, 1985, p. 125-47.
- [14] Bickis, M. and D. Krewski. 1988. Statistical issues in the analysis of the long term carcinogenicity bioassay in small rodents: An empirical evaluation of statistical decision rules. *Fund. Appl. Toxicol.* :in press.
- [15] Bishop, Yvonne M. M., Stephen E. Fienberg, and Paul W. Holland. *Discrete Multivariate Analysis*. Cambridge, Mass.: MIT Press, 1975.
- [16] Clayson, David B. "Problems in interspecies extrapolation." In *Toxicological Risk Assessment, vol. 1, Biological and Statistical Criteria*. Edited by D. B. Clayson, D. Krewski, and I. Munro. Boca Raton, Florida: CRC Press, 1985, p. 105-22.
- [17] Char, B. W., K. O. Geddes, G. H. Gonnet, and S. M. Watt. *Maple User's Guide, 4th ed.* Waterloo, Ontario: WATCOM Publications, 1985.
- [18] Chung K. L. *A Course in Probability Theory*. New York: Hartcourt Brace & World, 1986.
- [19] Cochran, W. G. 1954. Some methods of strengthening the common χ^2 tests. *Biometrics* 10: 417-51.
- [20] Cochran, W. G. "The vital role of randomization in experiments and surveys." In *The Heritage of Copernicus: Theories "Pleasing to the Mind"*. Edited by Jerzy Neyman. Cambridge, Mass.: MIT Press, 1974, p. 445-62.
- [21] Cochran W. G. *Sampling Techniques, 3rd ed.* New York: John Wiley & Sons, 1977.
- [22] Conover, W. J. 1974. Some reasons for not using the Yates continuity correction on 2×2 contingency tables. *J. Amer. Stat. Assoc.* 69: 374-6.
- [23] Cox, D. R. *The Analysis of Binary Data*. London: Methuen, 1970.
- [24] DuMouchel, William H. and Jeffrey E. Harris. 1983. Bayes methods for combining cancer experiments in humans and other species. *J. Amer. Stat. Assoc.* 78: 293-308.
- [25] Grice, H. C., I. C. Munro, D. R. Krewski, and H. Blumenthal. 1981. *In utero* exposure in chronic toxicity carcinogenicity studies. *Food Cosmet. Toxicol.* 19: 373-9.
- [26] Haseman, J. K. 1983. A re-examination of false positive rates for carcinogenicity bioassays. *Fund. Appl. Toxicol.* 3: 334-9.
- [27] Fears, T. R. and M. A. Schneiderman. 1974. Pathologic evaluation and the blind technique. *Science* 183: 1144-45.
- [28] Feller, William. *An Introduction to Probability Theory and its Applications, 3rd ed.* New York: John Wiley & Sons, 1968.

- [29] Fisher, R. A. *Statistical Methods for Research Workers*. Edinburgh: Oliver and Boyd, 1934.
- [30] Gart, J. J. 1971. The comparison of proportions: A review of significance tests, confidence intervals and adjustments for stratification. *Rev. Int. Stat. Inst.* 39:148-69.
- [31] Gart, J. J., K. C. Chu, and R. E. Tarone. 1979. Statistical issues in interpretation of chronic bioassay tests for carcinogenicity. *J. Nat. Cancer Inst.* 62:957-74.
- [32] Gart, J. J. and R. E. Tarone. 1987. On the efficiency of age-adjusted tests in animal carcinogenicity experiments. *Biometrics* 43:235-44.
- [33] Goldberg, Leon, editor. *Structure-Activity Correlation as a Predictive Tool in Toxicology*. Washington: Hemisphere Publishing, 1983.
- [34] Haseman, J. K. and D. G. Hoel. 1979. Statistical design of toxicity bioassays: Role of genetic structure of test animal populations. *J. Toxicol. Env. Health* 5:89-101.
- [35] Hoel, D. G. "Incorporation of pharmacokinetics in low-dose extrapolation." In *Toxicological Risk Assessment, vol. I. Biological and Statistical Criteria*. Edited by D. B. Clayson, D. Krewski, and I. Munro. Boca Raton, Florida: CRC Press, 1985, p. 125-47.
- [36] Hoel, D. G. and H. E. Walburg. 1972. Statistical analysis of survival experiments. *J. Nat. Cancer Inst.* 49:361-72.
- [37] Irwin, J. O. 1935. Tests of significance for differences between percentages based on small numbers. *Metron* 12:83-94.
- [38] Jurs, P. C., M. N. Hasan, D. R. Henry, T. R. Stouch, E. K. Whalen-Pedersen. 1983. Computer assisted studies of molecular structure and carcinogenic activity. *Fund. Appl. Toxicol.* 3:343-9.
- [39] Kemp, C. D. and A. W. Kemp. 1956. Generalized hypergeometric distributions. *J. Roy. Stat. Soc.—Series B* 18:202-11.
- [40] Kempthorne, O. 1977. Why randomize?. *J. Stat. Planning and Inference* 1:1-25.
- [41] Kempthorne, O. 1979. In dispraise of the exact test—reactions. *J. Stat. Planning and Inference* 3:199-213.
- [42] Krewski, D., M. Bickis, J. Kovar, and D. L. Arnold. 1986. Optimal experimental designs for low dose extrapolation I. The case of zero background. *Utilitas Mathematica* 29:245-62.
- [43] Krewski, D., J. Bremner, and M. Bickis. 1984. The power of the Fisher permutation test in $2 \times k$ tables. *Commun. Statist.—Simula. Computa.* 13:433-48.
- [44] Krewski, D., D. Clayson, B. Collins, and I. C. Munro. "Toxicological procedures for assessing the carcinogenic potential of agricultural chemicals." In *Genetic Toxicology*. Edited by Raymond A. Fleck and Alexander Hollaender. New York: Plenum Publishing, 1982, p. 461-97.

- [45] Krewski, D., J. G. Kovar, and M. Bickis. "Optimal experimental designs for low dose extrapolation II. The case of non-zero background." In *Topics in Applied Statistics*. Edited by Y. P. Chaubey and T. D. Dwivedi. Montreal: Concordia University, 1984, p. 167-91.
- [46] Krewski, D., Duncan J. Murdoch, and Jim R. Withey. "The application of pharmacokinetic data in carcinogenic risk assessment." *Pharmacokinetics in Risk Assessment: Drinking Water and Health* 8: 441-68.
- [47] Krewski, D. and J. Van Ryzin. "Dose response models for quantal response toxicity data." In *Statistics and Related Topics*. Edited by M. Csörgö, D. A. Dawson, J. N. K. Rao, A. K. Md. E. Saleh. New York: North-Holland, 1981, p. 210-31.
- [48] Krewski, D., D. Murdoch, and A. Dewanji. "Statistical modelling and extrapolation of carcinogenesis data." In *Modern Statistical Methods in Chronic Disease Epidemiology*. Edited by S. H. Moolgavkar and R. L. Prentice. New York: Wiley-Interscience, 1986, p. 259-82.
- [49] Krewski, D., D. Murdoch and J. Withey. 1988. Recent developments in carcinogenic risk assessment. *Health Physics* :in press.
- [50] Krewski, D., R. T. Smythe, A. Dewanji, and D. Colin. "Tests for trend in carcinogen bioassay with historical controls." *Proceedings of the Biopharmaceutical Section of the American Statistical Association*. Washington: American Statistical Association, 1986, p. 248-253.
- [51] Lagakos, S. W. and T. A. Louis. "Statistical analysis of rodent tumorigenicity experiments." In *Toxicological Risk Assessment, vol. 1, Biological and Statistical Criteria*. Edited by D. B. Clayson, D. Krewski, and I. Munro. Boca Raton, Florida: CRC Press, 1985, p. 105-22.
- [52] Lagakos, S. and F. Mosteller. 1981. A case study of statistics in the regulatory process. *J. Nat. Cancer Inst.* 66: 197-212.
- [53] Lehmann E. L. *Testing Statistical Hypotheses*. New York: John Wiley & Sons, 1959.
- [54] Littlefield, N. A., J. H. Farmer, D. W. Gaylor, and W. G. Sheldon. 1980. Effects of dose and time in a long-term, low-dose carcinogenic study. *J. Environ. Pathol. Toxicol.* 3: 17-34.
- [55] Littlefield, N. A., D. L. Greenman, and J. H. Farmer. 1980. Effects of continuous and discontinued exposure to 2-AAF on urinary bladder hyperplasia and neoplasia. *J. Environ. Pathol. Toxicol.* 3: 35-54.
- [56] Louis, T. A. "Adaptive sacrifice plans for rodent bioassay." Technical report. Department of Biostatistics, Harvard School of Public Health, Boston, Mass., 1982.
- [57] Louis, T. A. 1988. Analysis of a random sample of two-year carcinogen bioassays from the NCI data base. *Fund. Appl. Toxicol.* :in press.

- [58] MacLane, Saunders and Garrett Birkhoff. *Algebra*. New York: Macmillan, 1967.
- [59] McLeish, D. and D. Tosh. 1983. The estimation of extreme quantiles in logit bioassay. *Biometrika* 70: 625-32.
- [60] McConnell, E. E., R. Van Ryzin, J. M. Ward, and V. C. Glocklin. "Blinded' slide reading vs. individual slide evaluation, assignment of cause of death, relevance of animal dose to human dose. (Panel discussion)." *Symposium on Long-Term Animal Carcinogenicity Studies: A Statistical Perspective*. Washington, D.C.: American Statistical Association, 1985, p. 9-25.
- [61] Mehta, C. R. and N. R. Patel. 1980. A network algorithm for the exact treatment of the $2 \times k$ contingency table. *Comm. Stat. B* 9: 649-64.
- [62] Mehta, C. R. and N. R. Patel. 1983. A network algorithm for performing Fisher's exact test in $r \times c$ contingency tables. *J. Amer. Stat. Assoc.* 78: 427-34.
- [63] Mehta, C. R., N. R. Patel, and R. Gray. 1985. Computing an exact confidence interval for the common odds ratio in several 2×2 contingency tables. *J. Amer. Stat. Assoc.* 80: 969-73.
- [64] Munro, I. C. 1977. Considerations in chronic toxicity testing: The chemical, the dose, the design. *J. Env. Pathol. Toxicol.* 1: 183-97.
- [65] Peto, R. 1974. Guidelines on the analysis of tumour rates and death rates in experimental animals. *Br. J. Cancer* 29: 101-105.
- [66] Portier, C. and D. Hoel. 1984. Type 1 error of trend tests in proportions and the design of cancer screens. *Comm. Stat.-Th. Meth.* 13: 1-14.
- [67] Riley, V. 1984. Psychoneuroendocrine influences on immunocompetence and neoplasia. *Science* 212: 1100-109.
- [68] Salsburg, D. S. 1977. Use of statistics when examining lifetime studies in rodents to detect carcinogenicity. *J. Toxicol. Environ. Health* 3: 611-28.
- [69] Salsburg, D. S. 1979. The effects of life-time feeding studies on the pattern of senile lesions in mice and rats. *Drug and Chemical Toxicol.* 3: 1-33.
- [70] Sanathanan, L. P., C. T. Lin, and R. W. Carr. 1988. A blind reanalysis of a random subset of NCI bioassay studies: Do rats predict mice?. *Fund. Appl. Toxicol.* :in press.
- [71] Sarkadi, K. 1957. Generalized hypergeometric distributions. *Magyar Tudományos Akadémia Matematikai Kutató Közleményi* 2: 59-68.
- [72] Sawyer, C., R. Peto, L. Bernstein, and M. Pike. 1984. Calculation of carcinogenic potency from long-term animal carcinogenesis experiments. *Biometrics* 40: 27-40.
- [73] Serfling, Robert J. *Approximation Theorems of Mathematical Statistics*. New York: John Wiley & Sons, 1980.

- [74] Soma, A. P. 1977. An algorithm for the discrete Fisher's permutation test. *J. Amer. Stat. Assoc.* 72:662-4.
- [75] Starmer, C. F., J. E. Grizzle, and P. K. Sen. 1974. Comment [on [22]]. *J. Amer. Stat. Assoc.* 69:376-8.
- [76] Tamura, R. and S. Young. 1986. The incorporation of historical control information in tests of proportions: simulation study of Tarone's procedure. *Biometrics* 42:343-9.
- [77] Tarjan, Robert Endre. *Data Structures and Network Algorithms*. Philadelphia: Society for Industrial and Applied Mathematics, 1983.
- [78] Tarone, R. E. 1982. The use of historical control information in testing for a trend in proportions. *Biometrics* 38:215-20.
- [79] Tarone, Robert E., and John J. Gart. 1980. On the robustness of combined tests for trends in proportions. *J. Amer. Stat. Assoc.* 75:110-16.
- [80] Task Force of Past Presidents, Society of Toxicology. 1982. Animal data in hazard evaluation: paths and pitfalls. *Fund. Appl. Toxicol.* 2:101-.
- [81] Thomas, D. G., N. Breslow, and J. Gart. 1977. Trend and homogeneity analyses of proportions and life table data. *Computers and Biomed. Res.* 10:373-381.
- [82] Turnbull, B. W. and T. J. Mitchell. 1984. Nonparametric estimation of the distribution of time to onset for specific diseases in survival/sacrifice experiments. *Biometrics* 40:41-50.
- [83] Walters, D. E. 1979. In defence of the arc sine approximation. *The Statistician* 28:219-22.
- [84] Wu, C. F. J. 1985. Efficient sequential designs with binary data. *J. Amer. Stat. Assoc.* 80:974-84.
- [85] Young, S. S. 1988. A blind reanalysis of a random subset of NCI bioassay studies: Agreement between rats and mice. *Fund. Appl. Toxicol.* :in press.

END

29-06-89

FIN