1

PM-1 3 1/2" x 4" PHOTOGRAPHIC MICROCOPY TARGET
NBS 1010a ANSI/ISO #2 EQUIVALENT

PRECISIONSM RESOLUTION TARGETS

National Library
of Canada

Bibliothèque nationale
du Canada

Acquisitions and
Bibliographic Services Branch

Direction des acquisitions et
des services bibliographiques

395 Wellington Street
Ottawa, Ontario
K1A 0N4

395, rue Wellington
Ottawa (Ontario)
K1A 0N4

Your file - Votre référence

Our file - Notre référence

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.

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

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.

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

Weighted Approximations and Contiguous Weak Convergence of Parameters-Estimated Empirical Processes with Applications to Changepoint Analysis

by

José Andrés Correa Q., M.Math.

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
September 1995

© Copyright, 1995

José Andrés Correa Q.

National Library
of Canada

Acquisitions and
Bibliographic Services Branch

395 Wellington Street
Ottawa, Ontario
K1A 0N4

Bibliothèque nationale
du Canada

Direction des acquisitions et
des services bibliographiques

395, rue Wellington
Ottawa (Ontario)
K1A 0N4

Your file Votre référence

Our file Notre référence

The author has granted an irrevocable non-exclusive licence allowing the National Library of Canada to reproduce, loan, distribute or sell copies of his/her thesis by any means and in any form or format, making this thesis available to interested persons.

L'auteur a accordé une licence irrévocable et non exclusive permettant à la Bibliothèque nationale du Canada de reproduire, prêter, distribuer ou vendre des copies de sa thèse de quelque manière et sous quelque forme que ce soit pour mettre des exemplaires de cette thèse à la disposition des personnes intéressées.

The author retains ownership of the copyright in his/her thesis. Neither the thesis nor substantial extracts from it may be printed or otherwise reproduced without his/her permission.

L'auteur conserve la propriété du droit d'auteur qui protège sa thèse. Ni la thèse ni des extraits substantiels de celle-ci ne doivent être imprimés ou autrement reproduits sans son autorisation.

ISBN 0-612-08836-7

Canada

Name

José Andrés Correa Q.

Dissertation Abstracts International is arranged by broad, general subject categories. Please select the one subject which most nearly describes the content of your dissertation. Enter the corresponding four-digit code in the spaces provided.

Mathematics - Statistics

SUBJECT TERM

0430 U·M·I

SUBJECT CODE

Subject Categories

THE HUMANITIES AND SOCIAL SCIENCES

COMMUNICATIONS AND THE ARTS

Architecture	0729
Art History	0377
Cinema	0900
Dance	0378
Fine Arts	0357
Information Science	0723
Journalism	0391
Library Science	0399
Mass Communications	0708
Music	0413
Speech Communication	0459
Theater	0465

EDUCATION

General	0515
Administration	0514
Adult and Continuing	0516
Agricultural	0517
Art	0273
Bilingual and Multicultural	0282
Business	0688
Community College	0275
Curriculum and Instruction	0727
Early Childhood	0518
Elementary	0524
Finance	0277
Guidance and Counseling	0519
Health	0680
Higher	0745
History of	0520
Home Economics	0278
Industrial	0521
Language and Literature	0272
Mathematics	0280
Music	0522
Philosophy of	0998
Physical	0523

Psychology	0525
Reading	0535
Religious	0527
Sciences	0714
Secondary	0533
Social Sciences	0534
Sociology of	0340
Special	0529
Teacher Training	0530
Technology	0710
Tests and Measurements	0288
Vocational	0747

LANGUAGE, LITERATURE AND LINGUISTICS

Language	
General	0679
Ancient	0289
Linguistics	0290
Modern	0291
Literature	
General	0401
Classical	0294
Comparative	0295
Medieval	0297
Modern	0298
African	316
American	0591
Asian	0305
Canadian (English)	0352
Canadian (French)	0355
English	0593
Germanic	0311
Latin American	0312
Middle Eastern	0315
Romance	0313
Slavic and East European	0314

PHILOSOPHY, RELIGION AND THEOLOGY

Philosophy	0422
Religion	
General	0318
Biblical Studies	0321
Clergy	0319
History of	0320
Philosophy of	0322
Theology	0469

SOCIAL SCIENCES

American Studies	0323
Anthropology	
Archaeology	0324
Cultural	0326
Physical	0327
Business Administration	
General	0310
Accounting	0272
Banking	0770
Management	0454
Marketing	0338
Canadian Studies	0385
Economics	
General	0501
Agricultural	0503
Commerce-Business	0505
Finance	0508
History	0509
Labor	0510
Theory	0511
Folklore	0358
Geography	0366
Gerontology	0351
History	
General	0578

Ancient	0579
Medieval	0581
Modern	0582
Black	0328
African	0331
Asia, Australia and Oceania	0332
Canadian	0334
European	0335
Latin American	0336
Middle Eastern	0333
United States	0337
History of Science	0585
Law	0398
Political Science	
General	0615
International Law and	
Relations	0616
Public Administration	0617
Recreation	0814
Social Work	0452
Sociology	
General	0626
Criminology and Penology	0627
Demography	0938
Ethnic and Racial Studies	0631
Individual and Family	
Studies	0628
Industrial and Labor	
Relations	0629
Public and Social Welfare	0630
Social Structure and	
Development	0700
Theory and Methods	0344
Transportation	0709
Urban and Regional Planning	0999
Women's Studies	0453

THE SCIENCES AND ENGINEERING

BIOLOGICAL SCIENCES

Agriculture	
General	0473
Agronomy	0285
Animal Culture and	
Nutrition	0475
Animal Pathology	0476
Food Science and	
Technology	0359
Forestry and Wildlife	0478
Plant Culture	0479
Plant Pathology	0480
Plant Physiology	0817
Range Management	0777
Wood Technology	0746
Biology	
General	0306
Anatomy	0287
Biostatistics	0308
Botany	0309
Cell	0379
Ecology	0329
Entomology	0353
Genetics	0369
Limnology	0793
Microbiology	0410
Molecular	0307
Neuroscience	0317
Oceanography	0416
Physiology	0433
Radiation	0821
Veterinary Science	0778
Zoology	0472
Biophysics	
General	0786
Medical	0760

EARTH SCIENCES

Biogeochemistry	0425
Geochemistry	0996

Geodesy	0370
Geology	0372
Geophysics	0373
Hydrology	0388
Mineralogy	0411
Paleobotany	0345
Paleoecology	0426
Paleontology	0418
Paleozoology	0985
Polynology	0427
Physical Geography	0368
Physical Oceanography	0415

HEALTH AND ENVIRONMENTAL SCIENCES

Environmental Sciences	0768
Health Sciences	
General	0566
Audiology	0300
Chemotherapy	0992
Dentistry	0567
Education	0350
Hospital Management	0769
Human Development	0758
Immunology	0982
Medicine and Surgery	0564
Mental Health	0347
Nursing	0569
Nutrition	0570
Obstetrics and Gynecology	0380
Occupational Health and	
Therapy	0354
Ophthalmology	0381
Pathology	0571
Pharmacology	0419
Pharmacy	0572
Physical Therapy	0382
Public Health	0573
Radiology	0574
Recreation	0575

Speech Pathology	0460
Toxicology	0383
Home Economics	0386

PHYSICAL SCIENCES

Pure Sciences	
Chemistry	
General	0485
Agricultural	0749
Analytical	0486
Biochemistry	0487
Inorganic	0488
Nuclear	0738
Organic	0490
Pharmaceutical	0491
Physical	0494
Polymer	0495
Radiation	0754
Mathematics	0405
Physics	
General	0605
Acoustics	0986
Astronomy and	
Astrophysics	0606
Atmospheric Science	0608
Atomic	0748
Electronics and Electricity	0617
Elementary Particles and	
High Energy	0798
Fluid and Plasma	0759
Molecular	0609
Nuclear	0610
Optics	0752
Radiation	0756
Solid State	0611
Statistics	0463
Applied Sciences	
Applied Mechanics	0346
Computer Science	0984

Engineering	
General	0537
Aerospace	0538
Agricultural	0539
Automotive	0540
Biomedical	0541
Chemical	0542
Civil	0543
Electronics and Electrical	0544
Heat and Thermodynamics	0348
Hydraulic	0545
Industrial	0546
Marine	0547
Materials Science	0794
Mechanical	0548
Metallurgy	0743
Mining	0551
Nuclear	0552
Packaging	0549
Petroleum	0765
Sanitary and Municipal	0554
System Science	0790
Geotechnology	0428
Operations Research	0796
Plastics Technology	0795
Textile Technology	0994

PSYCHOLOGY

General	0621
Behavioral	0384
Clinical	0622
Developmental	0620
Experimental	0623
Industrial	0624
Personality	0625
Physiological	0989
Psychobiology	0349
Psychometrics	0632
Social	0451

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

**Weighted Approximations and Contiguous
Weak Convergence of Parameters-Estimated
Empirical Processes with Applications to
Changepoint Analysis**

submitted by

José Andrés Correa Q., M.Math.,

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

R.B. Riddle

for Chair, Department of Mathematics and Statistics

Miklós Orosz

Thesis Supervisor

Barbara Suprowitz

Thesis Supervisor

Howard

External Examiner

Carleton University

September 1995

To the memory of my mother,
to my grandmothers,
to Yolaine Dudoit.

Abstract

This dissertation is concerned with asymptotic methods in probability and mathematical statistics. In particular, we study weak convergence in weighted metrics of empirical processes based on observations from parametric families and defined after the parameters have been estimated by suitably chosen estimators. We consider independent and identically distributed observations, as well as independent observations whose distributions are governed by contiguous measures.

The main sources of inspiration and lines of thought that have led to the present exposition are to be found in two independent works that do not immediately relate to each other. These two main sources are the paper of Burke, Csörgő, Csörgő and Révész (1979) and the 1992 Ph.D. thesis of Barbara Szyszkowicz. It is hoped that we have succeeded in exploring and combining these two lines of research in a productive and insightful way.

The technique used is that of approximating the stochastic process of interest by its ideal version in the limit, usually a Gaussian process, rather than the traditional one of identifying finite dimensional distributions and proving tightness.

A more detailed outline of the thesis is as follows:

A more detailed outline of the thesis is as follows:

For an i.i.d. sequence $X_1, X_2, \dots$ whose common distribution function F belongs to the family

$$(1) \quad \{F(x; \theta); x \in \mathbb{R}, \theta \in \Theta \subseteq \mathbb{R}^p\},$$

let $\{\hat{\theta}_n\}$ be a sequence of estimators of θ , the vector of unknown parameters, based on $X_1, \dots, X_n$. The *estimated* empirical process is defined by:

$$(2) \quad \hat{\beta}_n(x) = n^{-1/2} \sum_{i=1}^n \left(1\{X_i \leq x\} - F(x; \hat{\theta}_n) \right), x \in \mathbb{R}.$$

In Chapter 1, which is expository, we present approximations of the estimated empirical process as in (2) when the estimation of the unknown parameters is done under certain conditions. The presentation is based on the results by Burke, Csörgő, Csörgő and Révész (1979).

In Chapter 2 we introduce a two-time parameter version of (2) defined by

$$(3) \quad \tilde{\beta}_n(x, t) = n^{-1/2} \sum_{i=1}^{[nt]} \left(1\{X_i \leq x\} - F(x; \hat{\theta}_n) \right), 0 \leq t \leq 1, x \in \mathbb{R},$$

and obtain approximations of such processes in weighted supremum norm, for an optimal class of weight functions, under the same type of estimation used in Chapter 1. This chapter also serves as an introduction to two-time parameter estimated empirical processes, as well as to the optimal class of weight functions to be considered in subsequent chapters.

In Chapter 3, we introduce a different version of (3), defined when the estimation is done in a “sequential” way, i.e., for any given $0 \leq t \leq 1$, we define the sequence of estimators $\{\hat{\theta}_{[nt]}\}$, where $\hat{\theta}_{[nt]}$ is based on $X_1, \dots, X_{[nt]}$, and then the two-time parameter *sequentially* estimated empirical process is defined by

$$(4) \quad \hat{\beta}_n(x, t) = n^{-1/2} \sum_{i=1}^{[nt]} \left(1\{X_i \leq x\} - F(x; \hat{\theta}_{[nt]}) \right), \quad 0 \leq t \leq 1, \quad x \in \mathbb{R}.$$

We obtain approximations of $\hat{\beta}_n(\cdot, \cdot)$, in $\|\cdot/q\|$ -metrics, for an optimal class of weight functions q , as well as that of a "bridge" type version of it, namely

$$\hat{\beta}_n(x, t) = \begin{cases} n^{1/2} \frac{[(n+1)t]}{n} \left(1 - \frac{[(n+1)t]}{n} \right) \hat{\zeta}_n(x, t), & 0 \leq t \leq 1, \quad x \in \mathbb{R}, \\ 0, & t = 1, \quad x \in \mathbb{R}, \end{cases}$$

with

$$\hat{\zeta}_n(x, t) = \frac{\sum_{i=1}^{[(n+1)t]} [1\{X_i \leq x\} - F(x; \hat{\theta}'_{[nt]})]}{[(n+1)t]} - \frac{\sum_{i=[(n+1)t]+1}^n [1\{X_i \leq x\} - F(x; \hat{\theta}''_{[nt]})]}{n - [(n+1)t]},$$

where, for each $0 < t < 1$, $\{\hat{\theta}'_{[nt]}\}$ and $\{\hat{\theta}''_{[nt]}\}$ are sequences of estimators of θ based on $X_1, \dots, X_{[(n+1)t]}$ and $X_{[(n+1)t]+1}, \dots, X_n$, respectively, in the case when (1) is a normal family with both parameters, the mean and the variance, unknown and being estimated by maximum likelihood estimators. From these results we note that we can have an application to testing for a change in the distribution of a random sequence at an unknown point, that will combine non-parametric results for empirical processes with the parametric information coming from the estimation.

Considering these normal empirical processes, which are of independent interest on their own, they have helped us establishing the kind of conditions needed in the case when we have a more general parametric family. Indeed, we also obtain results along these lines, under certain regularity conditions for the family, which

enable us to have conditions for the type of estimators to be used.

Based on Sections 4.2, 4.3, and 4.4 of Szyszkowicz (1992b), in Chapter 4 we discuss the notion of contiguity and show that for the normal family the results of Chapter 3 remain true under appropriate contiguous measures.

In Chapter 5 we generalize the results of Chapter 3 for the univariate normal family to multivariate normal families. We obtain approximations in $||\cdot/q||$ -metrics for the bivariate as well as for the multivariate normal families, using maximum likelihood estimation, for the same optimal class of functions. The applications to testing for a change in the distribution at an unknown point can be generalized to these multivariate normal observations.

The following is a list of the author's own results which are believed to be new:

Theorem 2.2.1	Lemma 4.5.3
Theorem 3.2.1	Lemma 4.5.4
Theorem 3.3.1	Lemma 4.5.5
Theorem 3.4.1	Lemma 5.2.2
Theorem 3.4.2	Lemma 5.5.2
Theorem 4.5.1	Corollary 2.2.1
Theorem 4.5.2	Corollary 3.2.1
Theorem 5.2.1	Corollary 3.3.1
Theorem 5.3.1	Corollary 3.4.1
Theorem 5.5.1	Corollary 3.4.2
Theorem 5.6.1	Corollary 5.2.1
Lemma 3.2.3	Corollary 5.3.1
Lemma 3.4.1	Corollary 5.5.1
Lemma 4.5.2	Corollary 5.6.1

In the next list the results amount to somewhat useful observations whose proofs may be new:

Lemma 3.2.2	Lemma 5.5.1
Lemma 4.5.1	Proposition A.2.1
Lemma 5.2.1	Proposition A.2.2

Acknowledgments

I would like to express my deep gratitude to my supervisor, Professor Miklós Csörgő, for his precious guidance, encouragement, and support. His dedication to mathematics has been a source of inspiration throughout my stay in Carleton University.

I also thank my co-supervisor, Professor Barbara Szyszkowicz, for her invaluable advices and support, as well as Dr. Rimas Norvaiša for his careful reading of part of the manuscript of this thesis.

I wish to thank my family in Caracas, Venezuela, as well as the Dudoit and Sutto families, my family here in Canada, for their affection and concern.

Last but not least, my very special thanks to Yolaine Dudoit for her love, patience, and all she has taught me.

My doctoral studies at Carleton University were supported by a scholarship from the Faculty of Graduate Studies and Research, by a research grant from Professor Miklós Csörgő, and by a scholarship from Consejo Nacional de Investigaciones Científicas y Tecnológicas (CONICIT) of the Government of Venezuela.

Contents

Notation	1
-----------------------	----------

Chapter 1. Estimated empirical processes

1.1. Introduction	2
1.2. Estimated empirical processes. Weak convergence	7

Chapter 2. Estimated empirical processes. Weighted approximations

2.1. Introduction	11
2.2. Estimated empirical processes. Weighted approximations	18

Chapter 3. Sequentially estimated empirical processes. Weighted approximations. Applications to changepoint analysis

3.1. Introduction	27
3.2. Sequentially estimated normal empirical processes	29

3.3. Sequentially estimated normal empirical processes tied down at $t=1$	48
3.4. Sequentially estimated empirical processes	75
 Chapter 4. Sequentially estimated normal empirical processes. Weighted weak convergence under contiguous measures	
4.1. Introduction	98
4.2. The notion of contiguity	99
4.3. A parametrization of contiguity	102
4.4. Some stochastic processes under contiguous measures	105
4.5. Sequentially estimated normal empirical processes under contiguous measures	109
 Chapter 5. Sequentially estimated multivariate normal empirical processes. Weighted approximations	
5.1. Introduction	132
5.2. Sequentially estimated bivariate normal empirical processes ..	146
5.3. Sequentially estimated bivariate normal empirical processes tied down at $t=1$	161
5.4. Notation	170
5.5. Sequentially estimated multivariate normal empirical processes	171

5.6. Sequentially estimated multivariate normal empirical processes tied down at $t=1$	182
--	------------

Appendices

A.1. Some maximal inequalities for partial sums.....	188
A.2. On Riemman-Stieltjes integrals in $\mathbb{R}^2$	189
A.3. Some results on stochastic integrals	191
References	200

Introduction

Let $X_1, X_2, \dots$ be a sequence of independent and identically distributed (i.i.d.) random variables (r.v.'s) with distribution function $F(\cdot)$. The empirical distribution function based on the sample $X_1, \dots, X_n$ is defined by

$$F_n(x) = \frac{1}{n} \sum_{i=1}^n \mathbf{1}\{X_i \leq x\}, \quad x \in \mathbb{R}.$$

The systematic study of the empirical distribution function $F_n(x)$, based on a random sample of size n , and its relationship of "closeness" to the underlying distribution function $F(x)$ began with the fundamental papers of Cantelli (1933), Glivenko (1933) and Kolmogorov (1933).

Let

$$\alpha_n(s) = n^{1/2}(E_n(s) - s), \quad 0 \leq s \leq 1,$$

where $E_n(s)$, is the empirical distribution function based on the first n ($n \geq 1$) elements of a sequence of independent uniform-(0,1) random variables. Since $E\alpha_n(s) = 0$ and $E(\alpha_n(s)\alpha_n(t)) = s(1-t)$, $0 \leq s \leq t \leq 1$, the covariance structure of $\{\alpha_n(s), 0 \leq s \leq 1\}$ is the same as that of a Brownian bridge process. A

Brownian bridge $\{B(s), 0 \leq s \leq 1\}$ is a separable Gaussian process with mean zero and covariance function given by

$$E B(s)B(t) = s(1-t), 0 \leq s \leq t \leq 1.$$

Furthermore, by the multivariate central limit theorem, the finite-dimensional distributions of $(\alpha_n(s_1), \dots, \alpha_n(s_k))$ tend, as $n \rightarrow \infty$, to the finite-dimensional distributions of $(B(s_1), \dots, B(s_k))$, $0 \leq s_1 \leq \dots \leq s_k \leq 1$. Reproving the results of Kolmogorov (1993) along these lines, Doob (1949) argued that in studying asymptotics of $\alpha_n(\cdot)$, as $n \rightarrow \infty$, one should replace it by $B(\cdot)$. Donsker (1952) was the first to justify Doob's "heuristic" approach. He showed, for instance, that $\alpha_n(s) \xrightarrow{\mathcal{D}} B(s)$ in $D[0, 1]$, where $\xrightarrow{\mathcal{D}}$ means convergence in distribution. Due to the a.s. continuity of $B(\cdot)$, this convergence implies weak convergence in $C[0, 1]$ in the supremum norm $\|\cdot\|$ topology. For an account of developments in the general methodology for weak convergence we refer to Billingsley (1968).

Since for any r.v. X having continuous distribution function F , $F(X)$ is a r.v. uniformly distributed in $[0, 1]$, for all F continuous we have

$$\{\beta_n(F^{-1}(s)), 0 \leq s \leq 1, n \geq 1\} = \{\alpha_n(s), 0 \leq s \leq 1, n \geq 1\},$$

where

$$(I.1) \quad \beta_n(x) = n^{-1/2} \sum_{i=1}^n (1\{X_i \leq x\} - F(x)).$$

Hence, all the results proved for $\alpha_n(s)$ hold automatically for $\beta_n(x)$. Indeed, for all $n \geq 1$ and $\omega \in \Omega$, $\beta_n(x) = \alpha_n(F(x))$, $x \in \mathbb{R}$.

Beginning with Breiman (1968) and Brillinger (1969), there has been much work done in approximating $\beta_n(x)$ almost surely by a sequence of Brownian bridges. Kiefer (1972) was the first to obtain a strong approximation of $\beta_n(x)$, with rates of convergence, in terms of a Gaussian process in both x and n , the so-called Kiefer process. A Kiefer process $\{K(s, t), 0 \leq s, t \leq 1\}$ is a separable Gaussian process with mean zero and covariance function

$$EK(s_1, t_1)K(s_2, t_2) = (s_1 \wedge s_2 - s_1 s_2)(t_1 \wedge t_2).$$

Clearly, we have, for any distribution function F ,

$$E(K(F(x_1), t_1)K(F(x_2), t_2)) = (F(x_1 \wedge x_2) - F(x_1)F(x_2))(t_1 \wedge t_2).$$

Komlós, Major and Tusnády (1975) obtained strong approximations of $\beta_n(\cdot)$ by a sequence of Brownian bridges, with the best possible rate of convergence, as well as by a Kiefer process, with the best available rate of approximation (cf. Theorem 1.1.B). These results can be used, for instance, to construct asymptotic confidence bands for $F(x)$, and to handle goodness-of-fit problems when $F(x)$ is completely specified. Concerning the latest problem, however, in most practical cases, one may only know the form of the distribution function $F(x; \theta)$ with $\theta \in \Theta \subseteq \mathbb{R}^p$, $p \geq 1$, unknown. One approach to this problem is to estimate θ from the given random sample of size n by a sequence of estimators $\hat{\theta}_n$ and then substitute it in the distribution function for θ . Consequently, the resulting *estimated* empirical process is defined by

$$(1.2) \quad \hat{\beta}_n(x) = n^{1/2}(F_n(x) - F(x; \hat{\theta}_n)), \quad x \in \mathbb{R}.$$

Durbin (1973) considered the question of weak convergence of $\hat{\beta}_n(\cdot)$ under a sequence of alternatives, for a wide class of estimators. The estimators had to satisfy certain maximum likelihood-like conditions, for example:

(I.3)

$$(i) \ n^{1/2}(\hat{\theta}_n - \theta_0) = n^{-1/2} \sum_{i=1}^n l(X_i, \theta_0) + \varepsilon_n, \text{ where } \theta_0 \text{ is the theoretical value}$$

of θ , $l(\cdot, \theta_0)$ is a measurable p -dimensional vector valued function, and

ε_n converges to zero in probability or a.s.,

$$(ii) \ E l(X_i, \theta_0) = 0,$$

(iii) $M(\theta) = E\{l(X_j, \theta)^\top l(X_j, \theta)\}$, evaluated at θ_0 , is a finite nonnegative definite matrix,

as well as several other regularity conditions on the underlying distribution function (cf. (1.2.2), page 7).

Burke, Csörgő, Csörgő and Révész (1979) obtained weak (in probability) and strong approximations of $\hat{\beta}_n(\cdot)$ in terms of a Gaussian processes in both x and n , for the same class of estimators as in Durbin (1973).

In Chapter 1 we expose in somewhat more detail the strong approximation results of Komlós, Major and Tusnády (1975), as well as the weak approximation result of Burke, Csörgő, Csörgő and Révész (1979) (cf. Theorem 1.2.A).

The two-time parameter version of the empirical process defined in (I.1) is defined by

$$(I.4) \quad \beta_n(x, t) = n^{-1/2} \sum_{i=1}^{[nt]} (1\{X_i \leq x\} - F(x)), \quad 0 \leq t \leq 1, \ x \in \mathbb{R}.$$

This process plays an important role in the so-called changepoint problem.

The problem of abrupt parameter changes arises in a variety of experimental and mathematical sciences. For instance, one may be interested in testing whether the incidence of a disease has remained constant over time, and if not, in estimating the time(s) of change(s). Detection of possible changepoints is also of interest in rhythm analysis in electrocardiograms, in ethological time series, in quality control, in seismic signal processing, and in the study of archaeological sites.

Such situations can usually be modeled by saying that we have a random process that generates independent observations indexed by time and we wish to detect, using statistical tests based on this process, whether a change could have occurred in the distribution that governs this random process as time goes by.

When studying problems of detecting abrupt changes one usually describes them as changepoint problems. These have originally arisen in the context of quality control, where one typically observes the output of a production line and would wish to signal deviation from an acceptable output level while observing the data. When one observes such a random process sequentially and stops observing at a random time of detecting change, then one speaks of a sequential procedure. Otherwise, one usually observes a large, chronologically ordered, finite sequence for the sake of determining possible change during the data collection. Thus, in this case, one studies a fixed, usually large sample, in various sequential ways, frequently keeping the chronological order of data as they came, or ordering them via some other schemes for further sequential studies. Based on the chronologically ordered set of data, the main idea is to construct stochastic processes which should be sensitive to changes in their probabilistic behaviour.

One approach, for instance, to studying these stochastic processes is to approximate them by their ideal asymptotic versions, usually Gaussian processes. These approximations should be done using metrics that would be able to accommodate changes early enough in time without ruining our ability to measure their distance (deviation) from each other. Consequently, the construction of such processes should be such that statistical tests based on various weighted functionals of them should be sensitive to deviations from some postulated, acceptable stochastic (distributional) behaviour.

Changepoint problems have been studied extensively from a parametric, as well as a non-parametric point of view. For more insight on this subject we refer the reader to Csörgő and Horváth (1988b) and Brodsky and Darkhovsky (1993), and to the references in these works.

When testing for a change in the distribution of an independent random sequence of observations, it is of interest to compare the empirical distribution function of the first k observations to that of the remaining $(n - k)$ observations. Thus, one may consider the statistics

$$\begin{aligned}
 (I.5) \quad & \sup_{1 \leq k < n} \sup_{x \in \mathbb{R}} n^{1/2} \left| \frac{1}{k} \sum_{i=1}^k \mathbf{1}\{X_i \leq x\} - \frac{1}{n-k} \sum_{i=k+1}^n \mathbf{1}\{X_i \leq x\} \right| \\
 &= \sup_{1 \leq k < n} \sup_{x \in \mathbb{R}} \left| \sum_{i=1}^k \mathbf{1}\{X_i \leq x\} - \frac{k}{n} \sum_{i=1}^n \mathbf{1}\{X_i \leq x\} \right| / n^{1/2} \left(\frac{k}{n} \left(1 - \frac{k}{n} \right) \right).
 \end{aligned}$$

The above sequence of statistics converges in probability to ∞ , even in the case when the r.v.'s are identically distributed and, moreover, even if the denominator $n^{1/2} \frac{k}{n} (1 - \frac{k}{n})$ were to be replaced by $n^{1/2} (\frac{k}{n} (1 - \frac{k}{n}))^{1/2}$. Hence, in order to have the above test statistics converge in distribution to a non-degenerate random variable

one has to renormalize them somehow. One may, for example, introduce weight functions which are less severe on the tails than $(\frac{k}{n}(1 - \frac{k}{n}))^{1/2}$ is (cf. Picard (1985), Deshayes and Picard (1986), and Szyszkowicz (1992b, 1994)).

Just like tests based on the classical Kolmogorov-Smirnov statistic, the ones based on

$$(I.6) \quad \sup_{1 \leq k < n} \sup_{x \in \mathbb{R}} \frac{k}{n} \left(1 - \frac{k}{n}\right) n^{1/2} \left| \frac{1}{k} \sum_{i=1}^k 1\{X_i \leq x\} - \frac{1}{n-k} \sum_{i=k+1}^n 1\{X_i \leq x\} \right|$$

$$= \sup_{1 \leq k < n} \sup_{x \in \mathbb{R}} n^{-1/2} \left| \sum_{i=1}^k 1\{X_i \leq x\} - \frac{k}{n} \sum_{i=1}^n 1\{X_i \leq x\} \right|$$

should be more powerful for detecting changes that occur in the middle, namely near $n/2$, where $\frac{k}{n}(1 - \frac{k}{n})$ has its maximum, than for noticing the ones occurring near the endpoints 0 and n . Thus, a weighted version of (I.6) should emphasize changes which may have occurred near the endpoints, while retaining sensitivity to possible changes in the middle as well. This has led to the idea of studying the weighted "bridge-type" process

$$(I.7) \quad \beta_n^*(x, t) = (\beta_n(x, t) - t\beta_n(x, 1))/q(t), \quad 0 \leq t \leq 1, \quad x \in \mathbb{R},$$

where $\beta_n(\cdot, \cdot)$ is as in (I.4), and F is assumed to be a continuous distribution function, in order to determine what class of weight functions are useful for having a non-degenerate limit for (I.5).

There are now complete characterizations available for describing the asymptotic behaviour of weighted sequential uniform empirical processes in supremum-

norm and L_p -metrics (cf. Szyszkowicz (1995)). In particular, considering the class of weight functions $q : (0, 1] \rightarrow (0, \infty)$ such that

$$(I.8) \quad \inf_{\delta \leq t \leq 1} q(t) > 0 \text{ for all } 0 < \delta < 1$$

(positive functions on $(0, 1]$), and non-decreasing near zero, we have that (cf. Theorem 2.1.A), as $n \rightarrow \infty$,

$$(I.9) \quad \sup_{0 < t \leq 1} \sup_{x \in \mathbb{R}} |\beta_n(x, t) - n^{-1/2} K(F(x), nt)| / q(t) = o_P(1)$$

if and only if

$$(I.10) \quad I_0(q, c) < \infty \text{ for all } c > 0,$$

where

$$(I.11) \quad I_0(q, c) = \int_0^1 t^{-1} e^{-ct^{-1} q^2(t)} dt, \quad c > 0.$$

The above result in (I.9) was proven by Szyszkowicz (1992b, 1994) under stronger conditions on the weight functions that result from the law of the iterated logarithm, or, equivalently, from the Adler and Brown (1986) results for lower and upper functions for suprema of Kiefer processes. The present optimal form of this result was obtained by using the integral test of Csörgő, Horváth and Szyszkowicz (1994) for suprema of Kiefer processes.

The integral in (I.11) was used by Csörgő, Csörgő, Horváth and Mason (1986), and Csörgő, Shao and Szyszkowicz (1991) for characterizing the class of functions for which

$$(I.12) \quad \limsup_{t \downarrow 0} |W(t)|/q(t) < \infty \text{ a.s.,}$$

where $\{W(t), 0 \leq t \leq \infty\}$ is a Wiener process. These are the so-called local functions of Wiener processes.

As to the class of functions for having a statement like (I.9) in terms of the process $\beta_n^*(\cdot, \cdot)$, as defined in (I.7), and its limiting process $\Gamma(\cdot, \cdot)$, where $\{\Gamma(F(x), t), 0 \leq t \leq 1, x \in \mathbb{R}\}$ is a Gaussian process with mean zero and covariance function

$$(I.13) \quad E\Gamma(F(x_1), t_1)\Gamma(F(x_2), t_2) = (F(x_1 \wedge x_2) - F(x_1)F(x_2))(t_1 \wedge t_2 - t_1 t_2),$$

consider the class of functions q which are positive on $(0, 1)$, i.e., $q : (0, 1) \rightarrow (0, \infty)$ such that

$$(I.14) \quad \inf_{\delta \leq t \leq 1-\delta} q(t) > 0 \text{ for all } 0 < \delta \leq \frac{1}{2},$$

and the integral

$$(I.15) \quad I_{0,1}(q, c) = \int_0^1 \frac{1}{t(1-t)} e^{-cq^2(t)/t(1-t)} dt, \quad c > 0.$$

Szyszkowicz (1995) has, for instance, the following result (cf. Theorem 3.3.A):

$$(I.16) \quad \sup_{0 < t < 1} \sup_{x \in \mathbb{R}} |\beta_n^*(x, t) - \Gamma_n(F(x), t)|/q(t) = o_P(1), \text{ as } n \rightarrow \infty,$$

if and only if

$$(I.17) \quad I_{0,1}(q, c) < \infty \text{ for all } c > 0,$$

where

$$\begin{aligned}
 & \{\Gamma_n(F(x), t), 0 \leq t \leq 1, x \in \mathbb{R}\} \\
 (I.18) \quad & := \{n^{-1/2}(K(F(x), nt) - tK(F(x), n), 0 \leq t \leq 1, x \in \mathbb{R}\} \\
 & \stackrel{D}{=} \{\Gamma(F(x), t) := K(F(x), t) - tK(F(x), 1), 0 \leq t \leq 1, x \in \mathbb{R}\}.
 \end{aligned}$$

This class of functions characterizes the local behaviour (near zero and near one) of $\sup_{x \in \mathbb{R}} |\Gamma_n(F(x), t)|/q(t)$.

The integral (I.15) was used by Csörgő, Csörgő, Horváth and Mason (1986) to characterize the so-called local functions of Brownian bridge processes, i.e., the class of functions for which

$$\limsup_{t \downarrow 0, t \uparrow 1} |B(t)|/q(t) < \infty \text{ a.s.},$$

where $\{B(t), 0 \leq t \leq 1\}$ is a Brownian bridge. The latter is due to the fact that with a given Wiener process $\{W(t), 0 \leq t < \infty\}$ a Brownian bridge $\{B(t), 0 \leq t \leq 1\}$ can be defined by

$$(I.19) \quad B(t) = W(t) - tW(1),$$

and each Brownian bridge can be written as in (I.19) for a suitably chosen Wiener process. For more insight, results, and further references on the subject of local functions and integral tests, we refer to Csörgő, Csörgő, Horváth and Mason (1986), Csörgő, Shao and Szyszkowicz (1991), Szyszkowicz (1992b), Csörgő and Horváth (1993), Csörgő, Horváth and Szyszkowicz (1994), and Szyszkowicz (1994, 1995).

Having in mind the idea of comparing the empirical distribution function of the first k observations with that one of the remaining $n - k$ observations when testing for a change in the distribution of a random sequence $X_1, X_2, \dots$ of independent observations (cf. (I.6)), we consider, for the case when $X_1, X_2, \dots$ are observations from a parametric family $\{F(x; \theta); \theta \in \Theta \subseteq \mathbb{R}^p, x \in \mathbb{R}\}$, the idea of comparing the respective *estimated* empiricals. To this end we define a two-time parameter *estimated* empirical process, as well as a “bridge-type” version of it, and obtain weighted approximations by Gaussian processes for a wide class of weight functions. These results will provide a test for a change in distribution that will combine the non-parametric results for empirical processes already available with the parametric information provided by the estimation of the unknown parameter(s).

We first study a two-time parameter version of the estimated empirical process in (I.2) defined by

$$(I.20) \quad \tilde{\beta}_n(x, t) = n^{-1/2} \sum_{i=1}^{[nt]} (1\{X_i \leq x\} - F(x; \hat{\theta}_n)), \quad 0 \leq t \leq 1, x \in \mathbb{R},$$

where for any t , the vector θ of unknown parameters is estimated by the sequence $\{\hat{\theta}_n\}_{n \geq 1}$ of estimators based on the sample $X_1, \dots, X_n$. The class of estimators considered is the same used by Durbin (1973) and subsequently used by Burke, Csörgő, Csörgő and Révész (1979). We obtain weighted approximations of $\tilde{\beta}(\cdot, \cdot)$, in $\|\cdot/q\|$ -metrics, for the same optimal class of weight functions q as in Szyszkowicz (1995), i.e., positive functions on $(0, 1]$ which are non-decreasing near zero and that satisfy the proper integral condition based on the integral $I_0(q, c)$ as in (I.11), by

a sequence of Gaussian processes. We present these results in Chapter 2.

Next, we study a different version of (I.20), defined by

$$(I.21) \quad \hat{\beta}_n(x, t) = n^{-1/2} \sum_{i=1}^{[nt]} (1\{X_i \leq x\} - F(x; \hat{\theta}_{[nt]})), \quad 0 \leq t \leq 1, \quad x \in \mathbb{R},$$

where, for each t , $\{\hat{\theta}_{[nt]}\}_{n \geq 1}$ is a sequence of estimators based on $X_1, \dots, X_{[nt]}$. We call this process *sequentially* estimated. The fact that the estimator $\hat{\theta}_{[nt]}$ is not only a function of n but also a function of t poses additional technical difficulties when studying approximations of the process $\hat{\beta}_n(\cdot, \cdot)$, in weighted metrics. For instance, considering the class of estimators used by Durbin (1973), we now have

$$[nt]^{1/2}(\hat{\theta}_{[nt]} - \theta_0) = [nt]^{-1/2} \sum_{i=1}^{[nt]} l(X_i, \theta_0) + \varepsilon_n(t),$$

with $\varepsilon_n(t) \rightarrow 0$ in probability or a.s, as $(nt) \rightarrow \infty$. However, to be able to approximate $\hat{\beta}_n(\cdot, \cdot)$ in $\|\cdot/q\|$ -metrics, for example, we should have additional conditions for the behaviour of $\varepsilon_n(t)/q(t)$, *uniformly* in t .

We first consider the case when the sequentially estimated empirical process as in (I.21) is defined from a sample $X_1, \dots, X_n$ of a sequence of independent normal random variables, and the unknown parameters, the mean and the variance, are estimated by maximum likelihood estimators. We obtain weighted approximations of the process $\hat{\beta}_n(\cdot, \cdot)$ by a sequence of Gaussian processes. These results are presented in Section 3.2. We then proceed, in Section 3.3, to obtain weighted approximations of the “bridge-type” version of $\hat{\beta}_n(\cdot, \cdot)$, denoted by $\hat{\hat{\beta}}_n(\cdot, \cdot)$, also by a sequence of Gaussian processes, and under the same assumption of normality.

The process $\{\hat{\beta}_n(x, t), 0 \leq t \leq 1, x \in \mathbb{R}\}$ is defined by

$$(1.22) \quad \hat{\beta}_n(x, t) = \begin{cases} n^{1/2} \frac{[(n+1)t]}{n} \left(1 - \frac{[(n+1)t]}{n}\right) \hat{\xi}_n(x, t), & 0 \leq t \leq 1, x \in \mathbb{R}, \\ 0 & t = 1, x \in \mathbb{R}, \end{cases}$$

where

$$(1.23) \quad \hat{\xi}_n(x, t) = \frac{\sum_{i=1}^{[(n+1)t]} (1\{X_i \leq x\} - F(x; \hat{\theta}'_{[nt]}))}{[(n+1)t]} - \frac{\sum_{i=[(n+1)t]+1}^n (1\{X_i \leq x\} - F(x; \hat{\theta}''_{[nt]}))}{n - [(n+1)t]},$$

and, for each t , $\{\hat{\theta}'_{[nt]}\}$ and $\{\hat{\theta}''_{[nt]}\}$ are sequences of maximum likelihood estimators based on the samples $X_1, \dots, X_{[(n+1)t]}$ and $X_{[(n+1)t]+1}, \dots, X_n$, respectively.

Building on the experience gained in sections 3.2 and 3.3, the just mentioned results are extended to more general parametric families, under certain regularity conditions for the family which determine the kind of estimation that is to be used. This is discussed in Section 3.4.

In addition to the results obtained for the sequentially estimated empirical process defined for the i.i.d. case, when the underlying distribution function is normal, as discussed in Sections 3.2 and 3.3, we obtain weak approximations for the same sequentially estimated empirical processes under a sequence of contiguous alternatives.

The notion of contiguity was introduced by Lucien Le Cam (1960) as a criterion of nearness of sequences of probability measures. From the point of view of mathematics, this notion is an asymptotic version of the idea of absolute continuity of measures.

Given a sequence of measurable spaces $\{(\mathcal{X}_n, \mathcal{F}_n)\}$, $n \geq 1$, with probability measures P_n , and Q_n on $(\mathcal{X}_n, \mathcal{F}_n)$, we say that the sequence of probability measures $\{Q_n\}$ is *contiguous* to the sequence $\{P_n\}$, if for every sequence of sets $\{A_n\}$, $A_n \in \mathcal{A}_n$,

$$\lim_{n \rightarrow \infty} P_n(A_n) = 0 \text{ implies } \lim_{n \rightarrow \infty} Q_n(A_n) = 0.$$

In Section 4.2, following Szyszkowicz (1992b), we describe briefly the mathematical notion of contiguity, and its use in statistics, focusing on Le Cam's *third lemma* (cf. Lemma 4.2.A), a main tool of contiguity theory in applications, which, roughly speaking, tells us that convergence properties of the statistics of interest can be determine under the contiguous alternative if there is enough information about its behaviour under the null hypothesis. In statistical terms, the sequence $\{P_n\}$ may, for example, correspond to the null hypothesis and the sequence $\{Q_n\}$ to the alternative.

Given a sequence $X_1, X_2, \dots$ of independent random variables, we consider the null hypothesis

$$H_0 : X_i, 1 \leq i \leq n, \text{ have the same distribution } F,$$

versus the alternative hypotheses

$$H_{1n} : X_i, 1 \leq i \leq n, \text{ have the respective distribution functions } F_{in},$$

where we assume that all F_{in} are absolutely continuous with respect to the distribution function F and

$$(1.24) \quad \left[\frac{dF_{in}}{dF}(F^{-1}(u)) \right]^{1/2} \equiv 1 + \frac{1}{2n^{1/2}} g_n(t, u), \quad \frac{i-1}{n} < t \leq \frac{i}{n},$$

where $F^{-1}(u) = \inf\{x : F(x) \geq u\}$, $0 < u \leq 1$, $F^{-1}(0) = F^{-1}(0+)$ (cf. 4.3.1).

We assume also that there exists a function $g \in L^2[0, 1]^2$ such that

$$(I.25) \quad \int_0^1 g(t, u) du = 0, \text{ for almost all } t \in [0, 1]$$

and

$$(I.26) \quad \int_0^1 \int_0^1 [g_n(t, u) - g(t, u)]^2 dudt \rightarrow 0, \text{ as } n \rightarrow \infty.$$

This parametrization, which was used by Khmaladze and Parjanadze (1986) in the context of studying changepoint problems using linear statistics of sequential ranks, renders the sequence of direct products $F_{1n} \times \dots \times F_{nn}$, $n = 1, 2, \dots$, contiguous to the sequence $F \times \dots \times F$, via Le Cam's *first lemma* (cf. Section 4.3).

Our presentations of Sections 4.2 and 4.3 are based on Sections 2.1, 2.2, and 2.3 of Szyszkowicz (1992b), which constitute a very detailed and complete summary of results concerning contiguity based on the books of Hájek and Šidak (1967), Roussas (1972), Greenwood and Shiryaev (1985), Shorack and Wellner (1986), Le Cam (1986) and Le Cam and Yang (1990).

In Section 4.4 we present some results of Szyszkowicz (1995) concerning the weighted asymptotic behaviour, under contiguous measures, of the processes $\beta_n(\cdot, \cdot)$, and $\beta_n^*(\cdot, \cdot)$, as in (I.4) and (I.7), respectively. In terms of the class of weight functions, these are improvements of the earlier results of Szyszkowicz (1994).

In Section 4.5 we then study the sequentially estimated empirical processes $\hat{\beta}_n(\cdot, \cdot)$ and $\hat{\beta}_n^*(\cdot, \cdot)$, as defined (I.21) and (I.22), respectively, under the assumption that $X_1, \dots, X_n$ are independent random variables with respective distribution functions F_{in} , contiguous to the distribution function F of a $N(\mu, \sigma^2)$ random va-

riable, with μ and σ unknown. These alternatives H_{1n} are assumed to be parametrized as in (I.24)-(I.26). Computing the thus arising covariances, we obtain that the weighted weak convergence results of Sections 3.2 and 3.3, i.e., the results under our null hypothesis H_0 , continue to hold for the same optimal class of weight functions.

Csörgő and Szyszkowicz (1994) obtained approximations in $\|\cdot/q\|$ -metrics of the $(d+1)$ -parameter, $d \geq 1$, empirical process defined by

$$(I.27) \quad \beta_n(\mathbf{x}, t) = n^{-1/2} \sum_{i=1}^{[nt]} (1\{X_i \leq \mathbf{x}\} - F(\mathbf{x})), \quad \mathbf{x} \in \mathbb{R}^d, \quad 0 \leq t \leq 1,$$

as well as that of the “bridge” type empirical process defined by

$$(I.28) \quad \beta_n^*(\mathbf{x}, t) = \beta_n(\mathbf{x}, t) - t\beta_n(\mathbf{x}, 1), \quad \mathbf{x} \in \mathbb{R}^d, \quad 0 \leq t \leq 1,$$

where $X_1, X_2, \dots$ is an i.i.d. sequence of random vectors with distribution function F on $\mathbb{R}^d$, $d \geq 1$. Here the inequalities for vectors in $\mathbb{R}^d$ are meant coordinatewise.

A Kiefer process $K_F(\mathbf{x}, t)$ on $\mathbb{R}^d \times [0, \infty)$ associated with a distribution F on $\mathbb{R}^d$, $d \geq 1$, is a separable $(d+1)$ -parameter real-valued Gaussian process with $K_F(\mathbf{x}, 0) = 0$, $EK_F(\mathbf{x}, t) = 0$, and

$$(I.29) \quad EK_F(\mathbf{x}, s)K_F(\mathbf{y}, t) = (s \wedge t)(F(\mathbf{x} \wedge \mathbf{y}) - F(\mathbf{x})F(\mathbf{y}))$$

for all $\mathbf{x}, \mathbf{y} \in \mathbb{R}^d$ and $s, t \geq 0$.

Considering the class of weight functions q , positive on $(0, 1]$ (cf. (I.8)), Csörgő and Szyszkowicz (1994) proved, for instance, that, as $n \rightarrow \infty$,

$$(I.30) \quad \sup_{0 < t \leq 1} \sup_{\mathbf{x} \in \mathbb{R}^d} |\beta_n(\mathbf{x}, t) - n^{-1/2} K_F(\mathbf{x}, nt)|/q(t) = o_P(1)$$

if

$$\lim_{t \downarrow 0} (t \log \log 1/t)^{1/2} / q(t) = 0.$$

Equivalently (cf. Lemma 2.1 and Theorem 2.2 of Csörgő, Horváth and Szyszkowicz (1994)), they also consider the class of weight functions $q : (0, 1] \rightarrow (0, \infty)$ such that $q \in C(0, 1]$ and $q(t)/t^{1/2}$ is nonincreasing near zero. They have, for instance, that for any such function (I.30) holds true if $I_d(q, c) < \infty$ for all $c > 0$, where

$$(I.31) \quad I_d(q, c) := \int_0^1 \frac{q^{2d}(t)}{t^{d+1}} \exp\left(-c \frac{q^2(t)}{t}\right) dt.$$

We note here that the Csörgő, Horváth and Szyszkowicz (1994) integral test for suprema of Kiefer processes is actually proven for Kiefer processes indexed by arbitrary distribution functions F on $\mathbb{R}^d$, $d \geq 1$. Consequently we have that, for instance, (I.30) holds if and only if $I_0(q, c) < \infty$ for all $c > 0$, where $I_0(q, c)$ is as in (I.11), and q is only assumed to be positive on $(0, 1]$ and non-decreasing near zero.

The results of Csörgő and Szyszkowicz (1994) are based on the following strong approximation result by Csörgő and Horváth (1988):

THEOREM. *Assume that $X_1, X_2, \dots$ are independent random vectors with an arbitrary distribution function F on $\mathbb{R}^d$. Then there exists a Kiefer process $K_F(\mathbf{x}, t)$ associated with F , such that*

$$(I.32) \quad \sup_{0 \leq t \leq 1} \sup_{\mathbf{x} \in \mathbb{R}^d} \left| n^{1/2} \beta_n(\mathbf{x}, t) - K_F(\mathbf{x}, nt) \right| \stackrel{a.s.}{=} O(n^{1/2-1/4d} (\log n)^{3/2}).$$

With these results in hand, we also extend our results for the univariate normal sequentially estimated empirical process $\hat{\beta}_n(\cdot, \cdot)$ as defined in (I.21), as well as for

the process $\hat{\beta}_n(\cdot, \cdot)$ as defined in (I.22) to multivariate normal distributions, approximating them by Gaussian processes, for the same class of weight functions q as in the univariate case. In particular, given a sequence $X_1, X_2, \dots$ of independent random vectors with distribution function F on $\mathbb{R}^d$, $d \geq 1$, we define the $(d + 1)$ -time parameter sequentially estimated empirical process by

$$(I.33) \quad \dot{\beta}_n(x, t) = n^{-1/2} \sum_{i=1}^{[nt]} (1\{X_i \leq x\} - F(x; \hat{\theta}_{[nt]})), \quad 0 \leq t \leq 1, \quad x \in \mathbb{R}^d,$$

where, for any $t \in (0, 1)$, $\{\hat{\theta}_{[nt]}\}$ is a sequence of estimators based on the sample $X_1, \dots, X_{[nt]}$.

Similarly, we define the "bridge-type" process $\{\hat{\beta}_n(x, t), 0 \leq t \leq 1, x \in \mathbb{R}^d\}$, $d \geq 1$, the proper multivariate version of the process as in (I.22).

In Sections 5.2 and 5.3 we obtain weighted approximations of $\hat{\beta}_n(\cdot, \cdot)$ and $\hat{\beta}_n(\cdot, \cdot)$, for the case when $X_1, X_2, \dots$ is a sequence of independent bivariate normal random vectors and the unknown parameters, the means, variances, and the covariance are all estimated by their maximum likelihood estimators. In Section 5.4 we introduce some further notation needed for the discussion of the general multivariate normal case. In Sections 5.5 and 5.6 we proceed to present the weighted approximation results for the multivariate normal case when the means, variances, and covariances are all estimated by using maximum likelihood estimators. We have presented our results first for the bivariate normal case, and only then for the multivariate normal case, for the sake of clarity. The later case of course does contain the bivariate case.

Each chapter has its own introduction. It is inevitable that some of the arguments presented here will be repeated time and again throughout the presentation of our work. With this Introduction we hope to have given a general overview of what has been accomplished in this thesis.

Notation

The following notation will be used throughout this work:

- (i) The transpose of a vector v will be denoted by v^T .
- (ii) The norm $\| \cdot \|$ on $\mathbb{R}^p$ is defined by $\|(y_1, \dots, y_p)\| = \max_{1 \leq i \leq p} |y_i|$.
- (iii) For a function $g(x; \theta)$, where $\theta = (\theta_1, \dots, \theta_p)$, $\nabla_{\theta} g(x; \theta_0)$ will denote the vector of partial derivatives $(\frac{\partial g}{\partial \theta_1}(x; \theta), \dots, \frac{\partial g}{\partial \theta_p}(x; \theta))$ evaluated at $\theta = \theta_0$.
- (iv) The matrix $[\frac{\partial^2 g}{\partial \theta_i \partial \theta_j}(x; \theta)]_{i,j}$ will be denoted by $g''_{\theta\theta}(x, \theta)$.
- (v) For a matrix or vector $V = (v_{ij})$, let $|V|$ denote the matrix $(|v_{ij}|)$, let $\int V$ denote $(\int v_{ij})$, and let V^{δ} denote (v_{ij}^{δ}) .
- (vi) $a \wedge b$ and $a \vee b$ will denote $\min\{a, b\}$ and $\max\{a, b\}$, respectively.
- (vii) For any two vectors $u = (u_1, \dots, u_p)$ and $v = (v_1, \dots, v_p) \in \mathbb{R}^p$, $u \cdot v$ denotes the dot product $u \cdot v = \sum_{i=1}^n u_i v_i$.

Chapter 1

Estimated empirical processes

1.1 Introduction.

Let $X_1, X_2, \dots$ be independent, identically distributed (i.i.d.) random variables with distribution function F . The empirical distribution function F_n based on $X_1, \dots, X_n$ is defined by

$$(1.1.1) \quad F_n(x) = \frac{1}{n} \sum_{i=1}^n \mathbf{1} \{X_i \leq x\}, \quad x \in \mathbb{R},$$

the proportion of the X_i 's ($1 \leq i \leq n$) which are less than or equal to x .

Clearly, for every fixed x , $F_n(x)$ is the relative frequency of successes in a Bernoulli sequence of trials with $E F_n(x) = F(x)$ and $\text{Var } F_n(x) = \frac{1}{n} F(x)(1 - F(x))$. Consequently, by the classical law of large numbers,

$$(1.1.2) \quad F_n(x) \xrightarrow{a.s.} F(x), \quad \text{for fixed } x.$$

Using the language of statistics, (1.1.2) means that $F_n(x)$ is an unbiased and strongly consistent estimator of $F(x)$ for each fixed x .

Glivenko (1933) and Cantelli (1933) proved that

$$(1.1.3) \quad \sup_{x \in \mathbb{R}} |F_n(x) - F(x)| \xrightarrow{a.s.} 0,$$

which tells us that sampling *ad infinitum*, $F(\cdot)$ can be uniquely determined with probability one. Hence, viewing $\{F_n(x), x \in \mathbb{R}, n \geq 1\}$ as a stochastic process, its sample functions are distributions which estimate $F(x)$ uniformly in x with probability one.

From a practical point of view it is also of interest to study the rate of convergence in (1.1.3). Towards this end the empirical process is defined by

$$(1.1.4) \quad \beta_n(x) = n^{1/2}(F_n(x) - F(x)), \quad x \in \mathbb{R},$$

i.e.,

$$(1.1.5) \quad \beta_n(x) = n^{-1/2} \sum_{i=1}^n (1\{X_i \leq x\} - F(x)), \quad x \in \mathbb{R}.$$

For each fixed x one has immediately the central limit theorem:

$$\beta_n(x) \xrightarrow{\mathcal{D}} N(0, F(x)(1 - F(x))).$$

As to convergence in distribution, uniformly in x , we have the following result of Kolmogorov (1933) and, respectively, Smirnov (1939).

Theorem 1.1.A. *If $F(x)$ is a continuous distribution function, then*

$$P \left\{ \sup_{x \in \mathbb{R}} |\beta_n(x)| \leq y \right\} \rightarrow K(y)$$

where

$$K(y) = \begin{cases} \sum_{k=-\infty}^{\infty} (-1)^k e^{-2k^2 y^2}, & y > 0 \\ 0, & \text{otherwise} \end{cases}$$

and

$$P \left\{ \sup_{x \in \mathbb{R}} \beta_n(x) \leq y \right\} \rightarrow S(y)$$

where

$$S(y) = \begin{cases} 1 - e^{-2y^2}, & y > 0 \\ 0, & \text{otherwise.} \end{cases}$$

Let $U_i = F(X_i)$, $i = 1, 2, \dots$. Then the U_i 's are $U(0, 1)$ random variables provided F is continuous. Let now $E_n(s)$ be the empirical distribution function of the sample $U_1, \dots, U_n$, and denote the resulting empirical process (uniform empirical process) in this case by

$$(1.1.6) \quad \alpha_n(s) = \sqrt{n}(E_n(s) - s), \quad 0 \leq s \leq 1,$$

i.e.,

$$(1.1.7) \quad \alpha_n(s) = n^{-1/2} \sum_{i=1}^n (1\{U_i \leq s\} - s), \quad 0 \leq s \leq 1.$$

Then $E\alpha_n(s) = 0$ and the covariance function of the process $\{\alpha_n(s), 0 \leq s \leq 1\}$ is

$$\rho(s_1, s_2) = E\alpha_n(s_1)\alpha_n(s_2) = s_1 \wedge s_2 - s_1 s_2,$$

which coincides with that of a Brownian bridge $\{B(s), 0 \leq s \leq 1\}$, a Gaussian process with mean zero and covariance function

$$(1.1.8) \quad EB(t_1)B(t_2) = t_1 \wedge t_2 - t_1 t_2.$$

Since for any random variable X having a continuous distribution function F , the random variable $F(X)$ is uniformly distributed on $[0, 1]$, for all continuous F we have

$$\{\beta_n(F^{-1}(s)), 0 \leq s \leq 1, n \geq 1\} = \{\alpha_n(s), 0 \leq s \leq 1, n \geq 1\},$$

i.e., these two empirical processes are identical. Hence all results proved for $\alpha_n(s)$ will hold automatically also for $\beta_n(x)$. Indeed, for every $\omega \in \Omega$, $\alpha_n(F(x)) = \beta_n(x)$, $x \in \mathbb{R}$, $n \geq 1$.

The weak convergence of $\alpha_n(\cdot)$ to Brownian bridge was proved by Donsker (1952). The first strong approximation of the empirical process is due to Brillinger (1969). Kiefer (1972) was the first to call attention to the fact that the empirical process should be viewed as a two-time parameter process in both x and n and that it should be approximated by an appropriate two-time parameter Gaussian process. He also gave a solution to this problem and proved the first two-time parameter strong approximation of $\alpha_n(\cdot)$.

A Kiefer process $\{K(s, t), 0 \leq s \leq 1, 0 \leq t < \infty\}$ is a separable, two-time parameter Gaussian process with mean zero and covariance function

$$(1.1.9) \quad EK(s_1, t_1)K(s_2, t_2) = (s_1 \wedge s_2 - s_1 s_2)(t_1 \wedge t_2).$$

We note here that $\{K(F(x), t), x \in \mathbb{R}, 0 \leq t \leq 1\}$, with F being a continuous distribution function, is a Kiefer process, i.e. a two-time parameter separable Gaussian process with mean zero and covariance function

$$EK(F(x_1), t_1)K(F(x_2), t_2) = (F(x_1 \wedge x_2) - F(x_1)F(x_2))(t_1 \wedge t_2).$$

The following theorem provides a strong approximation of the two-time parameter empirical process by a sequence of Brownian bridges, with the optimal rate, as well as by a Kiefer process, with the best available rate. It is due to Komlós, Major and Tusnády (1975).

Theorem 1.1.B. *Given i.i.d. random variables $X_1, X_2, \dots$, one can define a sequence of Brownian bridges $\{B_n(s), 0 \leq s \leq 1\}$ and a Kiefer process $K(\cdot, \cdot)$ such that*

$$\sup_{x \in \mathbb{R}} |\beta_n(x) - B_n(F(x))| \stackrel{a.s.}{=} O(n^{-1/2} \log n),$$

and

$$\sup_{x \in \mathbb{R}} |n^{1/2} \beta_n(x) - K(F(x), n)| \stackrel{a.s.}{=} O(\log^2 n).$$

From a statistical point of view, Theorem 1.1.B. is useful in constructing confidence intervals for an unknown distribution function F and also to construct goodness-of-fit tests for a completely specified F . However, in practice, instead of a completely specified F , we may be given a whole parametric family of distributions functions $\{F(x; \theta); \theta \in \Theta \subseteq \mathbb{R}^p\}$. From a goodness-of-fit point of view, the unknown parameters θ are a nuisance (nuisance parameters), which render most goodness-of-fit null hypotheses composite ones. As far as the empirical process is concerned, one natural way of "getting rid of θ " is to "estimate out θ " by using some kind of a "good estimator" sequence $\{\hat{\theta}_n\}$, based on random samples $X_1, \dots, X_n$ ($n = 1, 2, \dots$) from $F(x; \theta)$. The estimated empirical process is then defined by

$$\hat{\beta}_n(x) = n^{-1/2} [F_n(x) - F(x; \hat{\theta}_n)], \quad x \in \mathbb{R}.$$

Durbin (1973) considered the question of weak convergence of the empirical process under a given sequence of alternative hypotheses when parameters of a continuous *unspecified* distribution $F(x; \theta)$ are estimated from the data. The estimators themselves had to satisfy certain maximum likelihood-like conditions. He showed that, for such a general class of estimators, the estimated empirical process converges weakly to a Gaussian process, whose mean and covariance functions he also specified.

Burke, Csörgő, Csörgő and Révész (1979) obtained asymptotic in-probability and almost sure representations, in terms of a Gaussian process in both x and n , of the estimated empirical process under the same type of parameter estimation as used by Durbin (1973). We review their weak convergence result in the next section.

1.2 Estimated empirical processes. Weak convergence.

For an i.i.d. sequence $X_1, X_2, \dots$ whose common distribution function $F \in \{F(x; \theta); x \in \mathbb{R}, \theta \in \Theta \subseteq \mathbb{R}^p\}$, let $\{\hat{\theta}_n\}$ be a sequence of estimators of the row vector θ , based on $X_1, \dots, X_n$. The estimated empirical process is defined by:

$$(1.2.1) \quad \hat{\beta}_n(x) = n^{1/2}[F_n(x) - F(x; \hat{\theta}_n)],$$

where $x \in \mathbb{R}$ and F_n is the empirical distribution function (cf. (1.1.1)).

First we list the set of conditions which will be used in this and also in some sections of the following chapters. Only subsets of it will be used at different stages in the sequel.

(1.2.2)

- (i) $n^{1/2}(\hat{\theta}_n - \theta_0) = n^{-1/2} \sum_{i=1}^n l(X_i, \theta_0) + \varepsilon_{1n}$, where θ_0 is the unknown true value of θ , $l(\cdot, \theta_0)$ is a measurable p -dimensional vector valued function, and ε_{1n} converges to zero in a manner to be specified later on.

(ii) $E l(X_1, \theta_0) = 0$.

- (iii) $M(\theta) = E\{l(X_j, \theta)^T l(X_j, \theta)\}$, evaluated at θ_0 , is a finite nonnegative definite matrix.

- (iv) The vector $\nabla_{\theta} F(x; \theta)$ is uniformly continuous in x and $\theta \in \Lambda$,

where Λ is the closure of a given neighbourhood of θ_0 .

(v) Each component of the vector function $l(x, \theta_0)$ is of bounded variation on each finite interval.

(vi) The vector $\nabla_{\theta} F(x; \theta_0)$ is uniformly bounded in x , and each component of the matrix $F''_{\theta\theta}(x; \theta)$ is uniformly bounded in x , and $\theta \in \Lambda$, where Λ is as in (iv).

(vii) $\lim_{s \downarrow 0} (s \log \log 1/s)^{1/2} \|l(F^{-1}(s; \theta_0), \theta_0)\| = 0$, and
 $\lim_{s \uparrow 1} ((1-s) \log \log 1/(1-s))^{1/2} \|l(F^{-1}(s; \theta_0), \theta_0)\| = 0$,

where $F^{-1}(s; \theta_0) = \inf\{x : F(x; \theta_0) \geq s\}$.

(viii) $s \|(\frac{\partial}{\partial s}) l(F^{-1}(s; \theta_0), \theta_0)\| \leq C$, $0 < s < 1/2$, and
 $(1-s) \|(\frac{\partial}{\partial s}) l(F^{-1}(s; \theta_0), \theta_0)\| \leq C$, $1/2 < s < 1$,

for some positive constant C , where the vector of partial derivatives of the components of $l(F^{-1}(s; \theta_0), \theta_0)$, with respect to s , $\frac{\partial}{\partial s} l(F^{-1}(s; \theta_0), \theta_0)$, exists for all $s \in (0, 1)$.

In addition to their major interest on their own, the results of Chapters 3 and 5, in particular, illustrate also how maximum likelihood estimators satisfy the conditions (1.2.2) in case of normal families of distributions.

It is shown in Burke, Csörgő, Csörgő and Révész (1979) that the estimated empirical process $\hat{\beta}_n(x)$ of (1.2.1) can be approximated by the two-parameter Gaussian process $\{G(x, n), x \in \mathbb{R}, n \geq 1\}$ defined by

(1.2.3)

$$G(x, n) = n^{-1/2} K(F(x; \theta_0), n) - \left\{ \int l(x, \theta_0) d_x n^{-1/2} K(F(x; \theta_0), n) \right\} \cdot \nabla_{\theta} F(x; \theta_0)^T,$$

where $K(\cdot, \cdot)$ is the Kiefer process of Theorem 1.1.B.

$G(\cdot, \cdot)$ has mean function $EG(x, n) = 0$ and covariance function

$$\begin{aligned}
 EG(x, n)G(y, m) &= (n \wedge m)(nm)^{-\frac{1}{2}} \{ F(x \wedge y; \theta_0) - F(x; \theta_0)F(y; \theta_0) \\
 &\quad - J(x) \cdot \nabla_{\theta} F(y; \theta_0) - J(y) \cdot \nabla_{\theta} F(x; \theta_0) \\
 (1.2.4) \quad &\quad + \nabla_{\theta} F(x; \theta_0) M(\theta_0) \nabla_{\theta} F(y; \theta_0)^{\top} \},
 \end{aligned}$$

where $M(\theta_0)$ is defined by (1.2.2)(iii) and

$$J(x) = \int_{-\infty}^x l(z, \theta_0) d_z F(z; \theta_0).$$

Since $M(\theta_0)$ is nonnegative definite, there is a nonsingular matrix $D(\theta_0)$ such that

$$(1.2.5) \quad D(\theta_0)^{\top} M(\theta_0) D(\theta_0) = \begin{pmatrix} I & 0 \\ 0 & 0 \end{pmatrix},$$

where I is the identity matrix and $\text{rank } I = \text{rank } M(\theta_0)$. Hence $G(x, n)$ of (1.2.3) can be written as

$$(1.2.6) \quad G(x, n) = n^{-1/2} K(F(x; \theta_0), n) - n^{-1/2} W(n) \cdot D(\theta_0)^{-1} \cdot \nabla_{\theta} F(x; \theta_0)^{\top},$$

where $W(n) = \int l(x, \theta_0) d_x n^{-1/2} K(F(x; \theta_0), n) \cdot D(\theta_0)$ is a vector valued Wiener process with covariance structure: $(n \wedge m)$ multiplied by (1.2.5).

Clearly we have, for each n , that

$$(1.2.7) \quad G(x, n) \stackrel{D}{=} G(x) = B(F(x; \theta_0)) - \left\{ \int l(x, \theta_0) d_x B(F(x; \theta_0)) \right\} \cdot \nabla_{\theta} F(x; \theta_0)^{\top},$$

where $\stackrel{D}{=}$ stands for the equality of all the finite dimensional distributions, and $B(\cdot)$ is a Brownian bridge.

We have the following result, which is Theorem 3.1 (a) of Burke, Csörgő, Csörgő, and Révész (1979),

Theorem 1.2.A. *Suppose that the sequence $\{\hat{\theta}_n\}$ satisfies conditions (1.2.2)(i)-(v). Then, as $n \rightarrow \infty$,*

$$\sup_{x \in \mathbb{R}} |\hat{\beta}_n(x) - G(x, n)| = o_P(1),$$

provided $\varepsilon_{1n} \xrightarrow{P} 0$, where the process $G(\cdot, \cdot)$ is as in (1.2.3).

Remark 1.2.1. Remark 2 of Burke, Csörgő, Csörgő and Révész (1979) can be stated as

Corollary 1.2.A. *Under the conditions of Theorem 1.2.A. we have*

$$(1.2.8) \quad \hat{\beta}_n(F^{-1}(\cdot, \hat{\theta}_n)) \xrightarrow{D} G(F^{-1}(\cdot, \theta_0)) \text{ in } D[0, 1].$$

This result is Corollary 1 in Durbin (1973) (under his null hypothesis of $\theta = \theta_0$). It follows from Theorem 1.2.A. and (1.2.7).

Chapter 2

Estimated empirical processes Weighted approximations

2.1 Introduction.

As pointed out in Szyszkowicz (1992b), the motivation for studying the asymptotic behaviour of weighted empirical and quantile processes started with Anderson and Darling (1952), Rényi (1953), Chibisov (1964) and O'Reilly (1974). Csörgő, Csörgő, Horváth and Mason (1986) obtained the optimal weighted approximation of the uniform empirical and quantile processes by a sequence of Brownian bridges. Their result was based on the Csörgő and Révész (1978) inequality for the strong approximation of the uniform quantile process by a sequence of Brownian bridges, which, in turn, is based on the Komlós, Major and Tusnády (1976) inequality for the strong approximation of the partial sum process by a Wiener process. They emphasize strongly that the optimal class of weight functions for the weak convergence of these weighted processes is determined not by these empirical processes themselves but by the weighted tail behaviour of the approximating Gaussian processes.

Given a sequence $X_1, X_2, \dots$ of *i.i.d.* random variables with distribution function F , the two time parameter version of the empirical process defined by (1.1.5) is given by

$$(2.1.1) \quad \beta_n(x, t) = n^{-1/2} \sum_{i=1}^{[nt]} (1\{X_i \leq x\} - F(x)), \quad 0 \leq t \leq 1, \quad x \in \mathbb{R}.$$

If $U_1, U_2, \dots$ is a sequence of *i.i.d.* random variables with uniform $(0, 1)$ distribution, then the two time-parameter version of the uniform empirical process defined by (1.1.7) is given by

$$\alpha_n(s, t) = n^{-1/2} \sum_{i=1}^{[nt]} (1\{U_i \leq s\} - s), \quad 0 \leq t \leq 1, \quad 0 \leq s \leq 1.$$

Since for any random variable X having continuous distribution function F the random variable $F(X)$ is uniformly distributed on $(0, 1)$, for all continuous F we have $\{\beta_n(F^{-1}(s), t), 0 \leq s, t \leq 1, n \geq 1\} = \{\alpha_n(s, t), 0 \leq s, t \leq 1, n \geq 1\}$, i.e. the above two empirical processes are identical. Hence all the results proved for $\beta_n(x, t)$ will hold automatically for $\alpha_n(s, t)$. Indeed, for every $\omega \in \Omega$,

$$\alpha_n(F(x), t) = \beta_n(x, t), \quad x \in \mathbb{R}, \quad t \in [0, 1], \quad n \geq 1.$$

Szyszkowicz (1991a,b, 1992a,b, 1994, 1995) obtained approximations of the two time parameter empirical process with weights in the time parameter $t \in [0, 1]$. We will refer to these as weighted approximations or approximations in $\|\cdot/q\|$ -metrics.

We will call q a positive function on $(0, 1]$ if $q : (0, 1] \longrightarrow (0, \infty)$ is such that

$$(2.1.2) \quad \inf_{\delta \leq t \leq 1} q(t) > 0 \quad \text{for all } 0 < \delta < 1.$$

Let $\mathcal{Q}$ be the class of positive functions on $(0, 1]$ which are nondecreasing in a neighbourhood of zero. Consider the integral

$$(2.1.3) \quad I_0(q, c) = \int_0^1 \frac{1}{t} e^{-c \tau^2(t)/t} dt, \quad c > 0.$$

Assuming F to be continuous, we have the following result, which is Theorem 1.1 of Szyszkowicz (1995).

Theorem 2.1.A. *Let $X_1, X_2, \dots$ be a sequence of i.i.d. r.v.'s with a continuous distribution function F . There exists a Kiefer process $K(\cdot, \cdot)$ such that, with a weight function $q \in \mathcal{Q}$, we have, as $n \rightarrow \infty$,*

(a)

$$(2.1.4) \quad I_0(q, c) < \infty \text{ for all } c > 0$$

if and only if

$$\sup_{0 < t \leq 1} \sup_{x \in \mathbb{R}} |\beta_n(x, t) - n^{-1/2} K(F(x), nt)|/q(t) = o_P(1),$$

(b)

$$(2.1.5) \quad I_0(q, c) < \infty \text{ for some } c > 0$$

if and only if

$$\sup_{0 < t \leq 1} \sup_{x \in \mathbb{R}} |\beta_n(x, t) - n^{-1/2} K(F(x), nt)|/q(t) = O_P(1),$$

(c) (2.1.5) holds if and only if

$$\sup_{0 < t \leq 1} \sup_{x \in \mathbb{R}} |\beta_n(x, t)|/q(t) \xrightarrow{P} \sup_{0 < t \leq 1} \sup_{x \in \mathbb{R}} |K(F(x), t)|/q(t).$$

Remark 2.1.1. (b) and (c) do not follow from (a), nor does (c) from (b).

Remark 2.1.2. We note here that in Theorems 1.2.A and 2.1.A we are dealing with the same Kiefer process $K(\cdot, \cdot)$ as constructed by Komlós, Major and Tusnády (1975) (cf. Theorem 1.1.B).

Remark 2.1.3. We may conclude that on assuming (2.1.4), we will have a convergence in distribution result for any in $\|\cdot/q\|$ continuous functional of $\beta_n(x, t)/q(t)$. However, for the sup-functional, which is usually of interest in changepoint analysis, Theorem 2.1.A (c) gives the desired convergence under the weaker assumption (2.1.5). For instance, Theorem 2.1.A (c) is true with $q(t) = (t \log \log((1/t) \vee 3))^{1/2}$, $0 < t \leq 1$.

Corollary 2.1.A. *Let $K(\cdot, \cdot)$ be a Kiefer process. Then, as $n \rightarrow \infty$,*

(a) *If $q \in \mathcal{Q}$, (2.1.4) holds if and only if*

$$(2.1.6) \quad \beta_n(x, t)/q(t) \xrightarrow{\mathcal{D}} K(F(x), t)/q(t)$$

in $D(\mathbb{R} \times [0, 1])$,

(b) *if q is a positive function on $(0, 1]$, then (2.1.6) holds in $D(\mathbb{R} \times [\delta, 1])$ for any $0 < \delta < 1$.*

Remark 2.1.4. It is clear from the proof of Theorem 2.1.A that, as a consequence of the Komlós, Major and Tusnády (1975) theorem (cf. Theorem 1.1.B), for any function q which is positive on $(0, 1]$ and such that

$$(2.1.7) \quad \lim_{t \downarrow 0} t^{1/2}/q(t) = 0,$$

then, as $n \rightarrow \infty$,

$$(2.1.8) \quad \sup_{0 < t \leq 1} \sup_{x \in \mathbb{R}} |\beta_n(x, t) - n^{-1/2} \bar{K}(F(x), nt)| / q(t) \stackrel{a.s.}{=} o(1),$$

where

$$\bar{K}(F(x), nt) = \begin{cases} K(F(x), nt), & \text{for } t \in [1/n, 1], \\ 0, & \text{elsewhere,} \end{cases}$$

and $K(\cdot, \cdot)$ is the Kiefer process as in Theorem 1.1.B.

We note that either one of (2.1.4) and (2.1.5) implies (2.1.7) (cf. Proposition 3.1 of Csörgő, Csörgő, Horváth and Mason (1986)). Hence, since, by definition, $\beta_n(x, t) = 0$, for all $x \in \mathbb{R}$ and $t \in (0, 1/n)$, and

$$\sup_{x \in \mathbb{R}} |n^{-1/2} K(F(x), nt)| \stackrel{D}{=} \sup_{x \in \mathbb{R}} |K(F(x), t)| \text{ for each } n \geq 1,$$

any sufficient condition on the class of weight functions q for Theorem 2.1.A to hold must be given in terms of the local behaviour (near zero) of $\sup_{x \in \mathbb{R}} |K(F(x), t)| / q(t)$.

The following result, which follows immediately from Theorem 2.1 of Csörgő, Horváth and Szyszkowicz (1994), provides an integral test for the suprema of Kiefer processes (cf. Lemma 2.A of Szyszkowicz (1995)).

Lemma 2.1.A. *Let $q \in \mathcal{Q}$ and $K(\cdot, \cdot)$ be a Kiefer process. Then*

(a) *(2.1.4) holds if and only if*

$$(2.1.9) \quad \lim_{t \downarrow 0} \sup_{x \in \mathbb{R}} |K(F(x), t)| / q(t) = 0 \text{ a.s.,}$$

(b) (2.1.5) holds if and only if

$$(2.1.10) \quad \limsup_{t \downarrow 0} \sup_{s \in \mathbb{R}} |K(F(x), t)|/q(t) < \infty \text{ a.s.}$$

Remark 2.1.5. The integral in (2.1.3) was used for characterizing the class of functions for which

$$(2.1.11) \quad \limsup_{t \downarrow 0} |W(t)|/q(t) < \infty \text{ a.s.,}$$

where $\{W(t), 0 \leq t < \infty\}$ is a standard Wiener process (local functions of Wiener process) (cf. Csörgő, Csörgő, Horváth and Mason (1986), and Csörgő, Shao and Szyszkowicz (1991)).

We have the following result, which follows immediately from Theorems 3.3 and 3.4 of Csörgő, Csörgő, Horváth and Mason (1986).

Theorem 2.1.B. *Let $q \in \mathcal{Q}$ and $W(\cdot)$ be a Wiener process. Then*

(a) (2.1.4) holds if and only if

$$(2.1.12) \quad \lim_{t \downarrow 0} |W(t)|/q(t) = 0 \text{ a.s.,}$$

(b) (2.1.5) holds if and only if

$$(2.1.13) \quad \limsup_{t \downarrow 0} |W(t)|/q(t) < \infty \text{ a.s.}$$

Thus, in virtue of Lemma 2.1.A and Theorem 2.1.B, we have that the class of functions $q(\cdot)$ that characterize the local behaviour (near 0) of the suprema of Kiefer processes is the same class of functions that characterize the similar local behaviour of Wiener processes.

Remark 2.1.6. Theorem 2.1.A was first proved by Szyszkowicz (1992b, 1994) (cf. Theorem 2.1 of Szyszkowicz (1994)) under stronger conditions, resulting from the law of the iterated logarithm for suprema of Kiefer processes. Namely, assuming q to be a positive function on $(0, 1]$, and replacing condition (2.1.4) by

$$(2.1.14) \quad \lim_{t \downarrow 0} (t \log \log 1/t)^{1/2} / q(t) = 0,$$

and condition (2.1.5) by

$$(2.1.15) \quad \lim_{t \downarrow 0} (t \log \log 1/t)^{1/2} / q(t) < \infty,$$

she obtains the same results as in Theorem 2.1.A.

She also remarks that, by Theorem 5.2 of Adler and Brown (1986), Theorem 2.1.A can be stated under seemingly different conditions. In particular, one may assume that $q \in C(0, 1]$ and $q(t)/t^{1/2}$ is nonincreasing near 0 and replace (2.1.14) by the condition

$$(2.1.16) \quad \mathcal{I}(q, c) < \infty \text{ for all } c > 0,$$

and (2.1.15) by

$$(2.1.17) \quad \mathcal{I}(q, c) < \infty \text{ for some } c > 0,$$

where

$$(2.1.18) \quad \mathcal{I}(q, c) = \int_0^1 t^{-2} q^2(t) e^{-c q^2(t)/t} dt, \quad c > 0,$$

and then Theorem 2.1.A holds true.

Csörgő, Horváth and Szyszkowicz (1994) show, however, that assuming the above mentioned condition of monotonicity of $q(t)/t^{1/2}$ near zero, (2.1.14) is equivalent to (2.1.16) and (2.1.15) is equivalent to (2.1.17) (cf. the comments following Theorem 2.2 of the just mentioned paper).

Due to the integral test for suprema of Kiefer processes of Csörgő, Horváth and Szyszkowicz (1994), where the conditions of continuity of q and the monotonicity of $q(t)/t^{1/2}$ near zero from Adler and Brown (1986) are dropped, the results of Theorem 2.1.A are obtained for their largest possible respective classes of weight functions, thus making Theorem 2.1.A an improvement on the earlier Theorem 2.1 of Szyszkowicz (1994), with the additional very nice feature of the class being the same that characterizes local functions of Wiener processes.

For more insight on this subject of local functions and integral tests we refer the reader to Csörgő, Csörgő, Horváth and Mason (1986), Csörgő, Shao and Szyszkowicz (1991), Szyszkowicz (1992b), Csörgő and Horváth, 1993, Chapter 4, Csörgő, Horváth and Szyszkowicz (1994) and Szyszkowicz (1994, 1995).

In this chapter we define a two-time parameter version, in the manner of (2.1.1), of the estimated empirical process given by (1.2.1), when the sequence of estimators is constructed by using the whole sample $X_1, X_2, \dots, X_n$, for each $n \geq 1$, and study it in $\|\cdot/q\|$ -metrics for the optimal class of functions mentioned above, i.e., positive functions on $(0, 1]$ which are nondecreasing near zero and which satisfy appropriate integral conditions. The empirical process thus defined will be called, simply, estimator 1.

2.2 Estimated empirical processes. Weighted approximations.

Let $X_1, X_2, \dots$ be a sequence of i.i.d. random variables with common continuous distribution function $F \in \{F(x; \theta); x \in \mathbb{R}, \theta \in \Theta \subseteq \mathbb{R}^p\}$. Let $\{\tilde{\theta}_n\} = \{(\tilde{\theta}_{n1}, \dots, \tilde{\theta}_{np})\}$ be a sequence of estimators of the row vector θ based on the random sample $X_1, \dots, X_n$. Let θ_0 denote the true value of the unknown θ . Consider the two-time parameter estimated empirical process defined by

$$(2.2.1) \quad \tilde{\beta}_n(x, t) = n^{-1/2} \sum_{i=1}^{[nt]} (1\{X_i \leq x\} - F(x; \tilde{\theta}_n)), \quad 0 \leq t \leq 1, x \in \mathbb{R}.$$

Remark 2.2.1. If $t = 1$ in (2.2.1) we have the estimated empirical process as in (1.2.1).

Recall that $q : (0, 1] \rightarrow (0, \infty)$ is called a positive function on $(0, 1]$ if

$$\inf_{\delta \leq t \leq 1} q(t) > 0 \text{ for all } 0 < \delta < 1,$$

$\mathcal{Q}$ is the class of positive functions on $(0, 1]$ which are nondecreasing in a neighbourhood of zero, and

$$I_0(q, c) = \int_0^1 \frac{1}{t} e^{-c q^2(t)/t} dt, \quad c > 0.$$

Provided that for the sequence $\{\tilde{\theta}_n\}$ conditions (1.2.2)(i)-(v) hold, $\tilde{\beta}_n(x, t)$ will be approximated, in $\|\cdot/q\|$ -metrics by the sequence of two-time parameter Gaussian processes

$$(2.2.2) \quad \tilde{G}_n(x, t) = n^{-1/2} K(F(x; \theta_0), nt) - \frac{[nt]}{n} \left\{ \int l(x, \theta_0) d_x n^{-1/2} K(F(x; \theta_0), n) \right\} \cdot \nabla_{\theta} F(x; \theta_0)^T,$$

where $K(\cdot, \cdot)$ is the Kiefer process of Theorem 2.1.A. This approximation is a consequence of Theorem 1.2.A. and Theorem 2.1.A.

We have

Theorem 2.2.1. *Let $X_1, X_2, \dots$ be i.i.d. r.v.'s with a continuous distribution function F . Suppose that the sequence $\{\tilde{\theta}_n\}$ satisfies conditions (1.2.2)(i)-(v), with $\epsilon_{1n} \xrightarrow{P} 0$, as $n \rightarrow \infty$, and let $\tilde{\beta}_n(x, t)$ be the two time-parameter estimated empirical process as in (2.2.1). Then there exists a Kiefer process $K(\cdot, \cdot)$ such that with the sequence of stochastic processes $\{\tilde{G}_n(x, t), 0 \leq t \leq 1, x \in \mathbb{R}\}$, and a weight function $q \in \mathcal{Q}$, we have, as $n \rightarrow \infty$*

(a) if

$$(2.2.3) \quad I_0(q, c) < \infty \text{ for all } c > 0,$$

then

$$\sup_{0 < t \leq 1} \sup_{x \in \mathbb{R}} |\tilde{\beta}_n(x, t) - \tilde{G}_n(x, t)|/q(t) = o_P(1),$$

(b) if

$$(2.2.4) \quad I_0(q, c) < \infty \text{ for some } c > 0,$$

then

$$\sup_{0 < t \leq 1} \sup_{x \in \mathbb{R}} |\tilde{\beta}_n(x, t) - \tilde{G}_n(x, t)|/q(t) = O_P(1),$$

(c) if (2.2.4) holds, then

$$\sup_{0 < t \leq 1} \sup_{x \in \mathbb{R}} |\tilde{\beta}_n(x, t)|/q(t) \xrightarrow{P} \sup_{0 < t \leq 1} \sup_{x \in \mathbb{R}} |\tilde{G}(x, t)|/q(t),$$

where

$$(2.2.5) \quad \tilde{G}(x, t) = K(F(x; \theta_0), t) - t \left\{ \int l(x, \theta_0) d_x K(F(x; \theta_0), 1) \right\} \cdot \nabla_{\theta} F(x; \theta_0)^T.$$

Remark 2.2.2. As in Remark (2.1.1), (b) and (c) do not follow from (a), nor does (c) from (b).

Remark 2.2.3. If $t = 1$ and $q(t) \equiv 1$ (i.e. no weight function), Theorem 2.2.1 is the same as Theorem 1.2.A. with $\tilde{G}_n(x, 1) \equiv G(x, n)$ of (1.2.3).

Proof of Theorem 2.2.1. By (2.1.1) and (2.2.1), we have

$$\begin{aligned}\tilde{\beta}_n(x, t) &= n^{-1/2} \sum_{i=1}^{[nt]} (1\{X_i \leq x\} - F(x; \theta_0)) - \frac{[nt]}{n} n^{1/2} (F(x; \tilde{\theta}_n) - F(x; \theta_0)) \\ &= \beta_n(x, t) + \frac{[nt]}{n} n^{1/2} (F_n(x) - F(x; \tilde{\theta}_n)) - \frac{[nt]}{n} n^{1/2} (F_n(x) - F(x; \theta_0)),\end{aligned}$$

where $F_n(x)$ is the empirical distribution function based on $X_1, X_2, \dots, X_n$.

By (1.1.4), (1.2.1) and (1.2.3) we have

$$\begin{aligned}\tilde{\beta}_n(x, t) &= \beta_n(x, t) - n^{-1/2} K(F(x; \theta_0), nt) + n^{-1/2} K(F(x; \theta_0), nt) \\ &\quad + \frac{[nt]}{n} \left(\hat{\beta}_n(x) - G(x, n) \right) - \frac{[nt]}{n} \left(\beta_n(x) - n^{-1/2} K(F(x; \theta_0), n) \right) \\ (2.2.6) \quad &\quad - \frac{[nt]}{n} \int l(x, \theta_0) d_x n^{-1/2} K(F(x; \theta_0), n) \cdot \nabla_{\theta} F(x; \theta_0)^{\top},\end{aligned}$$

where $K(\cdot, \cdot)$ is the Kiefer process from Theorem 1.1.B.

Let q be a positive function on $(0, 1]$. From (2.2.2) and (2.2.6) we have

$$(2.2.7) \quad \sup_{0 < t \leq 1} \sup_{x \in \mathbb{R}} |\tilde{\beta}_n(x, t) - \tilde{G}_n(x, t)| / q(t) \leq$$

$$\begin{aligned}
& \sup_{0 < t \leq 1} \sup_{x \in \mathbb{R}} |\beta_n(x, t) - n^{-1/2} K(F(x; \theta_0), nt)| / q(t) \\
& + \sup_{0 < t \leq 1} \frac{[nt]/n}{q(t)} \sup_{x \in \mathbb{R}} |\beta_n(x) - n^{-1/2} K(F(x; \theta_0), n)| \\
& + \sup_{0 < t \leq 1} \frac{[nt]}{n} / q(t) \sup_{x \in \mathbb{R}} |\hat{\beta}_n(x) - G(x, n)| := I_1^{(n)} + I_2^{(n)} + I_3^{(n)}.
\end{aligned}$$

Let $0 < \delta < 1$ be fixed. By Proposition 3.1 of Csörgő, Csörgő, Horváth and Mason (1986), either (2.2.3) or (2.2.4) implies that

$$(2.2.8) \quad \lim_{t \downarrow 0} t^{1/2} / q(t) = 0.$$

Also, since $[nt] \leq nt$, we have $\frac{[nt]}{n} / q(t) \leq t / q(t)$, for all n , and, hence

$$\begin{aligned}
\sup_{0 < t \leq 1} \frac{[nt]}{n} / q(t) & \leq \sup_{0 < t \leq 1} t / q(t) \leq \sup_{0 < t < \delta} t / q(t) \vee \sup_{\delta \leq t \leq 1} t / q(t) \\
& = O(1) \vee \sup_{\delta \leq t \leq 1} t / q(t) \\
& = O(1), \text{ for any } \delta > 0.
\end{aligned}$$

By Theorem 1.1.B and Theorem 1.2.A, we have that, as $n \rightarrow \infty$,

$$(2.2.9) \quad I_i^{(n)} = o_P(1), \quad i = 2, 3.$$

Now, since $I_1^{(n)} = \sup_{0 < t \leq 1} \sup_{x \in \mathbb{R}} |\beta_n(x, t) - n^{-1/2} K(F(x; \theta_0), nt)| / q(t)$,

(a) and (b) follow by Theorem 2.1.A.

We note, once more, that the Kiefer process $K(\cdot, \cdot)$ we are dealing with is the same as that constructed by Komlós, Major and Tusnády (1975) (cf. Theorem 1.1.B).

As for the proof of (c) we note that we can write

$$(2.2.10) \quad \tilde{G}_n(x, t) = G_n^*(x, t) + R_n(x, t),$$

where

$$(2.2.11) \quad G_n^*(x, t) = n^{-1/2} K(F(x; \theta_0), nt) - t \left\{ \int l(x, \theta_0) d_x n^{-1/2} K(F(x; \theta_0), n) \right\} \cdot \nabla_\theta F(x; \theta_0)^\top,$$

and

$$(2.2.12) \quad R_n(x, t) = \left(t - \frac{[nt]}{n} \right) \left\{ \int l(x, \theta_0) d_x n^{-1/2} K(F(x; \theta_0), n) \right\} \cdot \nabla_\theta F(x; \theta_0)^\top.$$

It can be easily checked that

$$\begin{aligned} \sup_{0 < t \leq 1} \sup_{x \in \mathbb{X}} |R_n(x, t)|/q(t) &\leq p \left\| \int l(x, \theta_0) d_x n^{-1/2} K(F(x; \theta_0), n) \right\| \\ &\quad \times \sup_{x \in \mathbb{X}} \|\nabla_\theta F(x; \theta_0)\| \times \sup_{0 < t \leq 1} \left| t - \frac{[nt]}{n} \right|/q(t). \end{aligned}$$

By (1.2.2) (ii),(iii) and the Central Limit Theorem,

$$\left| \int l_j(x, \theta_0) d_x n^{-1/2} K(F(x; \theta_0), n) \right| = O_P(1), \text{ as } n \rightarrow \infty, \text{ for each } j = 1, 2, \dots, p.$$

Also, by (1.2.2)(iv), $\sup_{x \in \mathbb{X}} \|\nabla_\theta F(x; \theta_0)\| < \infty$.

On the other hand,

$$\begin{aligned} \sup_{0 < t \leq 1} \left| t - \frac{[nt]}{n} \right|/q(t) &\leq \sup_{0 < t \leq 1} \left| t - \frac{[nt]}{n} \right|/t^{1/2} \cdot \sup_{0 < t \leq 1} t^{1/2}/q(t) \\ &\leq \left(\sup_{0 < t < \frac{1}{n}} \left| t - \frac{[nt]}{n} \right|/t^{1/2} \vee \sup_{\frac{1}{n} \leq t \leq 1} \left| t - \frac{[nt]}{n} \right|/t^{1/2} \right) \sup_{0 < t \leq 1} t^{1/2}/q(t). \end{aligned}$$

Also,

$$\sup_{0 < t < \frac{1}{n}} |t - \frac{[nt]}{n}| / t^{1/2} = \sup_{0 < t < \frac{1}{n}} t^{1/2} = n^{-1/2},$$

and

$$\sup_{\frac{1}{n} \leq t \leq 1} |t - \frac{[nt]}{n}| / t^{1/2} \leq \sup_{\frac{1}{n} \leq t \leq 1} |t - \frac{[nt]}{n}| \left(\inf_{\frac{1}{n} \leq t \leq 1} t^{1/2} \right)^{-1} \leq n^{-1/2}.$$

Hence, since condition (2.2.4) implies that $\lim_{t \downarrow 0} t^{1/2}/q(t) = 0$ and, consequently, $\sup_{0 < t \leq 1} t^{1/2}/q(t) < \infty$, we have, as $n \rightarrow \infty$,

$$\sup_{0 < t \leq 1} |t - \frac{[nt]}{n}| / q(t) = o(1).$$

Thus, if $q \in \mathcal{Q}$ is such that (2.2.4) holds,

$$\sup_{0 < t \leq 1} \sup_{x \in \mathbb{R}} |R_n(x, t)| / q(t) = o_P(1), \text{ as } n \rightarrow \infty.$$

Consequently, by (2.2.10), as $n \rightarrow \infty$,

$$(2.2.13) \quad \sup_{0 < t \leq 1} \sup_{x \in \mathbb{R}} |\tilde{G}_n(x, t) - G_n^*(x, t)| / q(t) = o_P(1).$$

It is clear, by (2.1.8) of Remark 2.1.3, (2.2.7), (2.2.9), (2.2.13) and the triangular inequality, that, if (2.2.4) holds, then, as $n \rightarrow \infty$,

$$(2.2.14) \quad \sup_{0 < t \leq 1} \sup_{x \in \mathbb{R}} |\tilde{\beta}_n(x, t) - \bar{G}_n(x, t)| / q(t) = o_P(1),$$

where

$$\bar{G}_n(x, t) = \begin{cases} G_n^*(x, t), & t \in [1/n, 1], x \in \mathbb{R}, \\ 0, & \text{elsewhere,} \end{cases}$$

for each $n \geq 1$. In fact (2.2.14) holds assuming only that q be positive on $(0, 1]$

and such that (2.2.8) holds.

Since

$$\{n^{-1/2}K(F(x), nt), 0 \leq t \leq 1, x \in \mathbb{R}\} \stackrel{\mathcal{D}}{=} \{K(F(x), t), 0 \leq t \leq 1, x \in \mathbb{R}\}$$

for each $n \geq 1$, we have,

(2.2.15)

$$\{G_n^*(x, t), 0 \leq t \leq 1, x \in \mathbb{R}\} \stackrel{\mathcal{D}}{=} \{\tilde{G}(x, t), 0 \leq t \leq 1, x \in \mathbb{R}\} \text{ for each } n \geq 1,$$

where $\tilde{G}(x, t)$ is as in (2.2.5). Now, by (2.2.14), we have

(2.2.16)

$$\lim_{n \rightarrow \infty} \left| P \left\{ \sup_{0 < t \leq 1} \sup_{x \in \mathbb{R}} |\tilde{\beta}_n(x, t)|/q(t) \leq x \right\} - P \left\{ \sup_{\frac{1}{n} \leq t \leq 1} \sup_{x \in \mathbb{R}} |\tilde{G}_n(x, t)|/q(t) \leq x \right\} \right| = 0,$$

for any $x \in \mathbb{R}$.

We also have, for any $q \in \mathcal{Q}$ and such that (2.2.4) holds,

$$\lim_{n \rightarrow \infty} P \left\{ \sup_{\frac{1}{n} \leq t \leq 1} \sup_{x \in \mathbb{R}} |\tilde{G}(x, t)|/q(t) \leq x \right\} = P \left\{ \sup_{0 < t \leq 1} \sup_{x \in \mathbb{R}} |\tilde{G}(x, t)|/q(t) \leq x \right\},$$

for any $x \in \mathbb{R}$. This, together with (2.2.16), implies (c). ■

Corollary 2.2.1. *Assuming conditions of Theorem 2.2.1, we have, as $n \rightarrow \infty$,*

(a) *if $q \in \mathcal{Q}$ is such that (2.2.3) holds, then*

$$(2.2.17) \quad \tilde{\beta}_n(x, t)/q(t) \xrightarrow{\mathcal{D}} \tilde{G}(x, t)/q(t)$$

in $D(\mathbb{R} \times [0, 1])$,

(b) if q is positive on $(0, 1]$, then (2.2.17) holds in $D(\mathbb{R} \times [\delta, 1])$ for any $0 < \delta < 1$.

Proof. (a) follows from Theorem 2.2.1 (a).

(b) By Theorem 2.2.1 (a) with $q(t) \equiv 1$, and (2.1.2). ■

Chapter 3

Sequentially estimated empirical processes Weighted approximations Applications to change-point analysis

3.1 Introduction.

Suppose that $X_1, X_2, \dots$ is a sequence of i.i.d. random variables with common continuous distribution function $F \in \{F(x; \theta); x \in \mathbb{R}, \theta \in \Theta \subseteq \mathbb{R}^p\}$. In Section 2.2 we have defined the two time-parameter estimated empirical process by

$$\tilde{\beta}_n(x, t) = n^{-1/2} \sum_{i=1}^{\lfloor nt \rfloor} (1\{X_i \leq x\} - F(x; \hat{\theta}_n)), \quad 0 \leq t \leq 1, x \in \mathbb{R},$$

where the sequence $\{\tilde{\theta}_n\} = \{(\hat{\theta}_{n1}, \dots, \hat{\theta}_{np})\}$ of estimators of the row vector θ is based on the random sample $X_1, \dots, X_n$, regardless of the value of t .

In this chapter we define a two-time parameter estimated empirical process when the estimation of the unknown θ is done sequentially, by using the sample $X_1, \dots, X_{\lfloor nt \rfloor}$, for each t . The process thus defined will be called *sequentially estimated*.

The sequentially estimated empirical process is defined by

$$(3.1.1) \quad \hat{\beta}_n(x, t) = n^{-1/2} \sum_{i=1}^{[nt]} \left(1\{X_i \leq x\} - F(x; \hat{\theta}_{[nt]}) \right), \quad 0 \leq t \leq 1, x \in \mathbb{R}.$$

We will study weighted approximations of this process, assuming that, for each $\frac{1}{n} \leq t \leq 1$, the vector of unknown parameters, θ , is estimated by a sequence $\{\hat{\theta}_{[nt]}\}$ of estimators based on the random sample $X_1, \dots, X_{[nt]}$, under conditions given by (1.2.2)(i)-(v) in Chapter I, where (1.2.2)(i) is satisfied for each fixed t , i.e.,

$$(3.1.2) \quad [nt]^{\frac{1}{2}}(\hat{\theta}_{[nt]} - \theta_0) = [nt]^{-\frac{1}{2}} \sum_{i=1}^{[nt]} l(X_i, \theta_0) + \varepsilon_{1n}(t),$$

and $\varepsilon_{1n}(t)$ converges to zero, either in probability or almost surely, as $(nt) \rightarrow \infty$. In order to get a Gaussian approximation of $\hat{\beta}_n(x, t)$ using the strong approximation methodology that we have been using, we need additional assumptions on the estimation. For instance, for the sequence $\{\varepsilon_{1n}(t)\}$, we will need to assume further, in (3.1.2), that, as $n \rightarrow \infty$,

$$(3.1.3) \quad \sup_{\delta \leq t \leq 1} \|\varepsilon_{1n}(t)\| = o_P(1), \text{ and } \sup_{\frac{1}{n} \leq t \leq \delta} \|\varepsilon_{1n}(t)\| = O_P(1),$$

for any $0 < \delta < 1$.

In order to get more insight, and because the processes $\hat{\beta}_n$ themselves are of interest, we will first investigate the sequentially estimated empirical process and its “bridge” or “tied down” version, in the case of a normal family and maximum likelihood estimation. This will be done in Sections 3.2 and 3.3. In Section 3.4 we discuss the above processes in a more general context.

3.2 Sequentially estimated normal empirical processes.

Let $X_1, X_2, \dots$ be an independent sequence of random variables with common distribution function given by

$$F(x; \theta) = \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^x e^{-(t-\mu)^2/2\sigma^2} dt,$$

where $\theta = (\mu, \sigma^2)$. Let $\theta_0 = (\mu_0, \sigma_0^2)$ denote the true value of the vector of unknown parameters θ . We assume that, for each $\frac{1}{n} \leq t \leq 1$, θ is estimated by a sequence $\{\hat{\theta}_{[nt]}\}$, of maximum likelihood estimators based on the sample $X_1, \dots, X_{[nt]}$.

It is known that

$$(3.2.1) \quad \hat{\theta}_{[nt]} = (\bar{X}_{[nt]}, S_{[nt]}^2),$$

where

$$(a) \quad \bar{X}_{[nt]} = \frac{1}{[nt]} \sum_{i=1}^{[nt]} X_i, \quad \text{and} \quad (b) \quad S_{[nt]}^2 = \frac{1}{[nt]} \sum_{i=1}^{[nt]} (X_i - \bar{X}_{[nt]})^2.$$

Adding and subtracting μ_0 we get, after a short calculation,

$$\begin{aligned} S_{[nt]}^2 &= \frac{1}{[nt]} \sum_{i=1}^{[nt]} [(X_i - \mu_0) + (\mu_0 - \bar{X}_{[nt]})]^2 \\ &= \frac{1}{[nt]} \sum_{i=1}^{[nt]} (X_i - \mu_0)^2 - (\bar{X}_{[nt]} - \mu_0)^2 \end{aligned}$$

Hence,

$$\begin{aligned} [nt]^{\frac{1}{2}}(\hat{\theta}_{[nt]} - \theta_0) &= [nt]^{\frac{1}{2}} \left(\bar{X}_{[nt]} - \mu_0, \frac{1}{[nt]} \sum_{i=1}^{[nt]} (X_i - \mu_0)^2 - \sigma_0^2 \right) \\ &\quad - [nt]^{\frac{1}{2}} (0, (\bar{X}_{[nt]} - \mu_0)^2) \\ &= \left([nt]^{-\frac{1}{2}} \sum_{i=1}^{[nt]} (X_i - \mu_0), [nt]^{-\frac{1}{2}} \sum_{i=1}^{[nt]} [(X_i - \mu_0)^2 - \sigma_0^2] \right) \\ &\quad - (0, [nt]^{\frac{1}{2}} (\bar{X}_{[nt]} - \mu_0)^2), \end{aligned}$$

i.e., for each $\frac{1}{n} \leq t \leq 1$, (3.1.2) holds with

$$(3.2.2) \quad l(x, \theta_0) := (l_1(x, \theta_0), l_2(x, \theta_0)) = (x - \mu_0, (x - \mu_0)^2 - \sigma_0^2)$$

and

$$(3.2.3) \quad \varepsilon_{1n}(t) = \left(0, -[nt]^{1/2} (\bar{X}_{[nt]} - \mu_0)^2\right).$$

The process $\hat{\beta}_n(x, t)$ defined in (3.1.1) will be approximated, in $\|\cdot/q\|$ -metrics, for the class of weight functions $\mathcal{Q}$ as used in Chapter 2, by the sequence of Gaussian processes $\{G_n(x, t), 0 \leq t \leq 1, x \in \mathbb{R}\}$, where

$$(3.2.4) \quad G_n(x, t) = n^{-1/2} K(F(x; \theta_0), nt) - \left\{ \int l(x, \theta_0) d_x n^{-1/2} K(F(x; \theta_0), nt) \right\} \cdot \nabla_{\theta} F(x; \theta_0)^T,$$

and $K(\cdot, \cdot)$ is a Kiefer process.

We note that, for each n , G_n is a Gaussian process with mean zero and covariance function $E G_n(x, s) G_n(y, t) = (s \wedge t) \times \{ \}$, where $\{ \}$ equals the right-hand side of (1.2.4), with $n = m$.

We recall that $q : (0, 1] \rightarrow (0, \infty)$ is said to be positive on $(0, 1]$, if

$$(3.2.5) \quad \inf_{\delta \leq t \leq 1} q(t) > 0 \text{ for all } 0 < \delta < 1,$$

$\mathcal{Q}$ is the class of positive functions on $(0, 1]$ that are non-decreasing near 0, and

$$I_0(q, c) := \int_0^1 \frac{1}{t} e^{-c q^2(t)/t} dt, \quad c > 0.$$

We have the following result

Theorem 3.2.1. *Let $X_1, X_2, \dots$ be a sequence of i.i.d. $N(\mu, \sigma^2)$ random variables. Let $\hat{\beta}_n(x, t)$ be the sequentially estimated normal empirical process defined as in (3.1.1), with $\{\hat{\theta}_{[nt]}\}$ as in (3.2.1). Then, there exists a Kiefer process $K(\cdot, \cdot)$ such that with the sequence of stochastic processes $\{G_n(x, t), 0 \leq t \leq 1, x \in \mathbb{R}\}$, as in (3.2.4), and a weight function $q \in \mathcal{Q}$ we have, as $n \rightarrow \infty$,*

(a) if

$$(3.2.6) \quad I_0(q, c) < \infty \text{ for all } c > 0,$$

then

$$\sup_{0 < t \leq 1} \sup_{x \in \mathbb{R}} |\hat{\beta}_n(x, t) - G_n(x, t)|/q(t) = o_P(1),$$

(b) if

$$(3.2.7) \quad I_0(q, c) < \infty \text{ for some } c > 0,$$

then

$$\sup_{0 < t \leq 1} \sup_{x \in \mathbb{R}} |\hat{\beta}_n(x, t) - G_n(x, t)|/q(t) = O_P(1),$$

(c) if (3.2.7) holds, then

$$\sup_{0 < t \leq 1} \sup_{x \in \mathbb{R}} |\hat{\beta}_n(x, t)|/q(t) \xrightarrow{D} \sup_{0 < t \leq 1} \sup_{x \in \mathbb{R}} |G(x, t)|/q(t),$$

where

$$(3.2.8) \quad G(x, t) = K(F(x; \theta_0), t) - \left\{ \int l(x, \theta_0) d_x K(F(x; \theta_0), t) \right\} \cdot \nabla_{\theta} F(x; \theta_0)^T.$$

For the proof of Theorem 3.2.1 we need the following lemmas:

Lemma 3.2.1. *The series $\sum_{k=1}^{\infty} k^r e^{-\log^2 k/2}$ is convergent for any $r \in \mathbb{R}$.*

Proof. Consider the function $\varphi(x) = x^r e^{-\log^2 x/2}$, $x \geq 1$. We have that its derivative $\varphi'(x) = x^{r-1} e^{-\log^2 x/2} (r - \log x)$. Hence, $\varphi(x)$ is a continuous, non-negative and decreasing function for all $x \geq e^r$. Furthermore,

$$\begin{aligned} \int_{e^r}^{\infty} x^r e^{-\log^2 x/2} dx &= \int_r^{\infty} e^{(r+1)u} e^{-u^2/2} du = \int_r^{\infty} e^{(r+1)^2/2} e^{-(u-(r+1))^2/2} du \\ &\leq e^{(r+1)^2/2} \int_{-\infty}^{\infty} e^{-(u-(r+1))^2/2} du = \sqrt{2\pi} e^{\frac{(r+1)^2}{2}}, \end{aligned}$$

since $\frac{1}{\sqrt{2\pi}} e^{-(u-(r+1))^2/2}$ is the density of a $N((r+1), 1)$ random variable.

Thus, by the integral test for numerical series, $\sum_{k \geq e^r} k^r e^{-\log^2 k/2}$ is convergent, which implies the result. ■

For the vector-valued function $l(x, \theta_0)$ in (3.2.2) and $F(x; \theta_0)$, the distribution function of a $N(\mu_0, \sigma_0^2)$ random variable, we have

Lemma 3.2.2. *The series $\sum_{k=1}^{\infty} k^{-1} \int_{|x| > \log k} l_j^2(x, \theta_0) d_x F(x; \theta_0)$ is convergent for any $j = 1, 2$.*

Proof. We may assume, without loss of generality, that $\theta_0 = (0, 1)$. Hence, in (3.2.2) we have $l(x, \theta_0) = (x, x^2 - 1)$ and the distribution function $F(x; \theta_0) \equiv \Phi(x)$, the distribution function of a standard normal random variable.

Thus,

$$\begin{aligned} (3.2.9) \quad \int_{|x| > \log k} l_1^2(x, \theta_0) d_x F(x; \theta_0) &= \frac{1}{\sqrt{2\pi}} \int_{|x| > \log k} x^2 e^{-x^2/2} dx = \frac{2}{\sqrt{2\pi}} \int_{\log k}^{+\infty} x^2 e^{-x^2/2} dx \\ &= 2 \left(1 - \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\log k} x^2 e^{-x^2/2} dx \right). \end{aligned}$$

Using integration by parts, we get

$$\begin{aligned} \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\log k} x^2 e^{-x^2/2} dx &= \frac{1}{\sqrt{2\pi}} \left[-x e^{-x^2/2} \Big|_{-\infty}^{\log k} + \int_{-\infty}^{\log k} e^{-x^2/2} dx \right] \\ &= \Phi(\log k) - \frac{1}{\sqrt{2\pi}} \log k e^{-\log^2 k/2}. \end{aligned}$$

Putting this in (3.2.9) we get

$$\begin{aligned} \frac{1}{\sqrt{2\pi}} \int_{|x| > \log k} x^2 e^{-x^2/2} dx &= 2 \left[1 - \Phi(\log k) + \frac{1}{\sqrt{2\pi}} \log k e^{-\log^2 k/2} \right] \\ &\leq \frac{2}{\sqrt{2\pi}} \left[e^{-\log^2 k/2} + \log k e^{-\log^2 k/2} \right] \end{aligned}$$

for any $k \geq 2$, by the well known estimation (cf. Feller, 1968, p. 175),

$$(3.2.10) \quad 1 - \Phi(x) \leq \frac{1}{\sqrt{2\pi}x} e^{-x^2/2}, \quad x > 0.$$

Hence,

$$\begin{aligned} \sum_{k=1}^{\infty} k^{-1} \int_{|x| > \log k} l_1^2(x, \theta_0) d_x F(x; \theta_0) &= \sum_{k=1}^{\infty} k \frac{1}{\sqrt{2\pi}} \int_{|x| > \log k} x^2 e^{-x^2/2} dx \\ (3.2.11) \quad &\leq 2(1 - \Phi(0)) + 2 \sum_{k=2}^{\infty} k^{-1} \left[e^{-\log^2 k/2} + \log k e^{-\log^2 k/2} \right] \\ &\leq 2(1 - \Phi(0)) + 4 \sum_{k=2}^{\infty} e^{-\log^2 k/2} < \infty, \end{aligned}$$

by Lemma 3.2.1, with $r = 0$.

On the other hand,

$$(3.2.12) \quad \int_{|x| > \log k} l_2^2(x, \theta_0) d_x F(x; \theta_0) = \frac{1}{\sqrt{2\pi}} \int_{|x| > \log k} (x^2 - 1)^2 e^{-x^2/2} dx$$

$$\leq \frac{1}{\sqrt{2\pi}} \int_{|x| > \log k} (x^4 + 1) e^{-x^2/2} dx = \frac{2}{\sqrt{2\pi}} \int_{\log k}^{\infty} (x^4 + 1) e^{-x^2/2} dx.$$

Using integration by parts we get,

$$\begin{aligned} \frac{2}{\sqrt{2\pi}} \int_{\log k}^{+\infty} x^4 e^{-x^2/2} dx &= \frac{2}{\sqrt{2\pi}} \left[-x^3 e^{-x^2/2} \Big|_{\log k}^{+\infty} + 3 \int_{\log k}^{+\infty} x^2 e^{-x^2/2} dx \right] \\ (3.2.13) \qquad &= \frac{2}{\sqrt{2\pi}} \log^3 k e^{-\log^2 k/2} + \frac{3}{\sqrt{2\pi}} \int_{|x| > \log k} x^2 e^{-x^2/2} dx. \end{aligned}$$

Also, for $k \geq 2$, by (3.2.10),

$$\begin{aligned} (3.2.14) \qquad \frac{2}{\sqrt{2\pi}} \int_{\log k}^{\infty} e^{-x^2/2} dx &= 2[1 - \Phi(\log k)] \\ &\leq \frac{2}{\sqrt{2\pi} \log k} e^{-\log^2 k/2} \leq 2e^{-\log^2 k/2}. \end{aligned}$$

Hence, from (3.2.12), (3.2.13) and (3.2.14), we have,

$$\begin{aligned} \sum_{k=2}^{\infty} k^{-1} \int_{|x| > \log k} l_2^2(x, \theta_0) d_x F(x; \theta_0) &\leq \sum_{k=2}^{\infty} k^{-1} \frac{1}{\sqrt{2\pi}} \int_{|x| > \log k} (x^4 + 1) e^{-x^2/2} dx \\ &\leq \sum_{k=2}^{\infty} k^{-1} \left(\frac{2}{\sqrt{2\pi}} \log^3 k e^{-\log^2 k/2} + \frac{3}{\sqrt{2\pi}} \int_{|x| > \log k} x^2 e^{-x^2/2} dx + 2e^{-\log^2 k/2} \right) \\ &\leq \sum_{k=2}^{\infty} k^{-1} \left(k^3 e^{-\log^2 k/2} + \frac{3}{\sqrt{2\pi}} \int_{|x| > \log k} x^2 e^{-x^2/2} dx \right). \end{aligned}$$

By (3.2.11) and Lemma 3.2.1, with $r = 2$, we have that the latter series is convergent; and, since $\frac{1}{\sqrt{2\pi}} \int (x^4 + 1) e^{-x^2/2} dx < \infty$, the results follows. ■

Introduce the following

(3.2.15)

$$\epsilon_{2n}(F(x; \theta_0), t) = n^{-1/2} \sum_{i=1}^{[nt]} (1\{X_i \leq x\} - F(x; \theta_0)) - n^{-1/2} K(F(x; \theta_0), nt)$$

and

$$(3.2.16) \quad L_n(t) = \int l(x, \theta_0) d_x \epsilon_{2n}(F(x; \theta_0), t),$$

where $K(\cdot, \cdot)$ is a Kiefer process.

Lemma 3.2.3. Let $X_1, X_2, \dots$ be a sequence of i.i.d. r.v.'s with distribution function given by

$$F(x; \theta_0) = \frac{1}{\sqrt{2\pi}\sigma_0} \int_{-\infty}^x e^{-(t-\mu_0)^2/2\sigma_0^2} dt,$$

where $\theta_0 = (\mu_0, \sigma_0^2)$. Then, for any function q , positive on $(0, 1]$ and such that $\lim_{t \downarrow 0} t^{1/2}/q(t) = 0$, we have, as $n \rightarrow \infty$,

$$\sup_{\frac{1}{n} \leq t \leq 1} \|L_n(t)\|/q(t) \rightarrow 0 \text{ in probability.}$$

Proof. Let $u_n(t) = \log(nt)$, $\frac{1}{n} \leq t \leq 1$, $n \geq 1$. Hence,

$$L_n(t) = \int_{|x| \leq u_n(t)} l(x, \theta_0) d_x \left[n^{-1/2} \sum_{i=1}^{[nt]} (1\{X_i \leq x\} - F(x; \theta_0)) - n^{-1/2} K(F(x; \theta_0), nt) \right]$$

(3.2.17)

$$+ \int_{|x| > u_n(t)} l(x, \theta_0) d_x n^{-1/2} \sum_{i=1}^{[nt]} (1\{X_i \leq x\} - F(x; \theta_0))$$

$$- \int_{|x| > u_n(t)} l(x, \theta_0) d_x n^{-1/2} K(F(x; \theta_0), nt)$$

$$:= L_{1n}(t) + L_{2n}(t) - L_{3n}(t).$$

Let $0 < \delta < 1$ be fixed. Using integration by parts, for each $i = 1, 2$, we have

$$\begin{aligned} \sup_{\delta \leq t \leq 1} \|L_{1n}(t)\|/q(t) &\leq \left\{ \sup_{\delta \leq t \leq 1} \|l(x, \theta_0) \varepsilon_{2n}(F(x; \theta_0), t) \Big|_{x=-u_n(t)}^{x=u_n(t)}\| \right. \\ &\quad \left. + \sup_{\delta \leq t \leq 1} \left\| \int_{-u_n(t)}^{u_n(t)} \varepsilon_{2n}(F(x; \theta_0), t) d_x l(x, \theta_0) \right\| \right\} \times \left(\inf_{\delta \leq t \leq 1} q(t) \right)^{-1}. \end{aligned}$$

Since q is positive on $(0, 1]$, $\inf_{\delta \leq t \leq 1} q(t) > 0$; and, by (3.2.2),

$$\begin{aligned} l(x, \theta_0) \Big|_{-u_n(t)}^{u_n(t)} &= (x - \mu_0, (x - \mu_0)^2 - \sigma_0^2) \Big|_{-u_n(t)}^{u_n(t)} = (2u_n(t), -4\mu_0 u_n(t)) \\ (3.2.18) \quad &= (2 \log(nt), -4\mu_0 \log(nt)). \end{aligned}$$

By Theorem 1.1.B, as $(nt) \rightarrow \infty$,

$$\sup_{x \in \mathbb{R}} |\varepsilon_{2n}(F(x; \theta_0), t)| \stackrel{a.s.}{=} O(n^{-1/2} \log^2(nt)).$$

Hence,

$$\begin{aligned} \sup_{\delta \leq t \leq 1} \|l(x, \theta_0) \varepsilon_{2n}(F(x; \theta_0), t) \Big|_{x=-u_n(t)}^{x=u_n(t)}\| &+ \sup_{\delta \leq t \leq 1} \left\| \int_{-u_n(t)}^{u_n(t)} \varepsilon_{2n}(F(x; \theta_0), t) d_x l(x, \theta_0) \right\| \\ &\stackrel{a.s.}{=} \sup_{\delta \leq t \leq 1} n^{-1/2} \log^3(nt) \leq n^{-1/2} \log^3 n \rightarrow 0, \text{ as } n \rightarrow \infty. \end{aligned}$$

Thus,

$$\sup_{\delta \leq t \leq 1} \|L_{1n}(t)\|/q(t) \stackrel{a.s.}{=} o(1), \text{ as } n \rightarrow \infty.$$

On the other hand, arguing as above,

$$\begin{aligned} \sup_{\frac{1}{n} \leq t \leq \delta} \|L_{1n}(t)\|/q(t) &\leq \sup_{\frac{1}{n} \leq t \leq \delta} t^{1/2}/q(t) \sup_{\frac{1}{n} \leq t \leq \delta} \|L_{1n}(t)\|/t^{1/2} \\ &\stackrel{a.s.}{=} \sup_{0 < t \leq \delta} t^{1/2}/q(t) \sup_{\frac{1}{n} \leq t \leq \delta} \frac{\log^3(nt)}{(nt)^{1/2}} \\ &\leq M \sup_{0 < t \leq \delta} t^{1/2}/q(t) \text{ for some } M > 0, \text{ and all } n \geq 1. \end{aligned}$$

Now, since q is such that $\lim_{t \downarrow 0} t^{1/2}/q(t) = 0$, by taking $\delta > 0$ sufficiently small, we get that

$$\sup_{\frac{1}{n} \leq t \leq \delta} \|L_{1n}(t)\|/q(t) \stackrel{a.s.}{=} o(1), \text{ as } n \rightarrow \infty.$$

Let again $0 < \delta < 1$ be fixed. Given $\epsilon > 0$, we have, by (3.2.17), that

$$\begin{aligned} P\left\{\sup_{\frac{1}{n} \leq t \leq \delta} \|L_{2n}(t)\|/q(t) > \epsilon\right\} &= P\left\{\sup_{\frac{1}{n} \leq t \leq \delta} \frac{\|L_{2n}(t)\|}{t^{1/2}} \sup_{\frac{1}{n} \leq t \leq \delta} t^{1/2}/q(t) > \epsilon\right\} \\ &\leq P\left\{\max_{1 \leq [nt] \leq [n\delta]} \|[nt]^{-1/2} \sum_{i=1}^{[nt]} \int_{|x| > \log(nt)} l(x, \theta_0) d_x (1\{X_i \leq x\} - F(x; \theta_0))\| > \right. \\ &\quad \left. \epsilon \left(\sup_{\frac{1}{n} \leq t \leq \delta} \frac{t^{1/2}}{q(t)}\right)^{-1}\right\}. \end{aligned}$$

Since

$$E\left(\int_{|x| > \log k} l_j(x, \theta_0) d_x (1\{X_i \leq x\} - F(x; \theta_0))\right) = 0,$$

and

$$E\left(\int_{|x| > \log k} l_j(x, \theta_0) d_x (1\{X_i \leq x\} - F(x; \theta_0))\right)^2 \leq \int_{|x| > \log k} l_j^2(x, \theta_0) d_x F(x; \theta_0)$$

for each $j = 1, 2$, we have, by the Hájek-Rényi inequality (cf. Appendix A.1, Theorem A.1.2),

$$\begin{aligned} &P\left\{\sup_{\frac{1}{n} \leq t \leq \delta} \|L_{2n}(t)\|/q(t) > \epsilon\right\} \\ (3.2.19) \quad &\leq \left(\sup_{\frac{1}{n} \leq t \leq \delta} t^{1/2}/q(t)\right)^2 \frac{1}{\epsilon^2} \sum_{j=1}^2 \sum_{k=1}^n k^{-1} \int_{|x| > \log k} l_j^2(x, \theta_0) d_x F(x; \theta_0). \end{aligned}$$

By Lemma 3.2.2 and the fact that $\lim_{t \rightarrow 0} t^{1/2}/q(t) = 0$, we have, by taking $\delta > 0$ sufficiently small, that, as $n \rightarrow \infty$,

$$\sup_{\frac{1}{n} \leq t \leq 1} \|L_{2n}(t)\|/q(t) = o_P(1).$$

On the other hand, by Kolmogorov's inequality (cf. Appendix A.1, Theorem A.1.1),

(3.2.20)

$$\begin{aligned} & P\left\{\sup_{\delta \leq t \leq 1} \|L_{2n}(t)\|/q(t) > \epsilon\right\} \\ & \leq P\left\{\max_{[n\delta] \leq [nt] \leq n} \left\|\sum_{i=1}^{[nt]} \int_{|x| > \log(nt)} l(x, \theta_0) d_x (1\{X_i \leq x\} - F(x; \theta_0))\right\| \right. \\ & \quad \left. > 2\epsilon n^{1/2} \inf_{\delta \leq t \leq 1} q(t)\right\} \\ & \leq \left(\inf_{\delta \leq t \leq 1} q(t)\right)^{-2} \frac{1}{\epsilon^2} \sum_{j=1}^2 \sum_{k=[n\delta]}^n k^{-1} \int_{|x| > \log k} l_j^2(x, \theta_0) d_x F(x; \theta_0). \end{aligned}$$

Now, by (3.2.5) and Lemma 3.2.2, we have that

$$\sup_{\delta \leq t \leq 1} \|L_{2n}(t)\|/q(t) = o_P(1), \text{ as } n \rightarrow \infty.$$

Since for each $n \geq 1$ and $\frac{1}{n} \leq t \leq 1$,

$$n^{-1/2} K(u, nt) \stackrel{D}{=} \left(\frac{t}{[nt]}\right)^{1/2} \sum_{i=1}^{[nt]} B_i(u), \quad 0 \leq u \leq 1,$$

where $\{B_i\}_{i=1}^{[nt]}$ is a family of independent Brownian bridges, it can be easily shown that, for each $j = 1, 2$,

$$\begin{aligned} & E\left(\int_{|x| > \log k} l_j(x, \theta_0) d_x n^{-1/2} K(F(x; \theta_0), nt)\right) = 0, \\ & E\left(\int_{|x| > \log k} l_j(x, \theta_0) d_x n^{-1/2} K(F(x; \theta_0), nt)\right)^2 \leq \int_{|x| > \log k} l_j^2(x, \theta_0) d_x F(x; \theta_0), \end{aligned}$$

and that we have the same inequalities (3.2.19) and (3.2.20) for $L_{3n}(t)$. Thus,

$$\sup_{\frac{1}{n} \leq t \leq 1} \|L_{3n}(t)\|/q(t) = o_P(1), \text{ as } n \rightarrow \infty. \quad \blacksquare$$

Proof of Theorem 3.2.1. By (3.1.1) and a one-term Taylor expansion around θ_0 , we get

$$\begin{aligned} \hat{\beta}_n(x, t) &= n^{-1/2} \sum_{i=1}^{[nt]} (1\{X_i \leq x\} - F(x; \theta_0)) - [nt]n^{-1/2} (F(x; \hat{\theta}_{[nt]}) - F(x; \theta_0)) \\ (3.2.21) \quad &= n^{-1/2} \sum_{i=1}^{[nt]} (1\{X_i \leq x\} - F(x; \theta_0)) + n^{-1/2} [nt]^{1/2} \epsilon_{3n}(x, t) \\ &\quad - n^{-1/2} [nt] (\hat{\theta}_{[nt]} - \theta_0) \cdot \nabla_{\theta} F(x; \theta_0)^{\top}, \end{aligned}$$

where

$$(3.2.22) \quad \epsilon_{3n}(x, t) = [nt]^{1/2} (\hat{\theta}_{[nt]} - \theta_0) \cdot (\nabla_{\theta} F(x; \theta_0)^{\top} - \nabla_{\theta} F(x; \hat{\theta}_{[nt]})^{\top}),$$

and

$$(3.2.23) \quad \sup_{0 \leq t \leq 1} \|\hat{\theta}_{[nt]} - \theta_0\| \leq \sup_{0 \leq t \leq 1} \|\hat{\theta}_{[nt]} - \theta_0\|.$$

Since $El(X_i, \theta_0) = 0$, by (3.1.2), we have,

$$\begin{aligned} n^{-1/2} [nt] (\hat{\theta}_{[nt]} - \theta_0) &= n^{-1/2} \sum_{i=1}^{[nt]} l(X_i, \theta_0) + n^{-1/2} [nt]^{1/2} \epsilon_{1n}(t) \\ (3.2.24) \quad &= n^{-1/2} \sum_{i=1}^{[nt]} \int l(x, \theta_0) d_x 1\{X_i \leq x\} + n^{-1/2} [nt]^{1/2} \epsilon_{1n}(t) \end{aligned}$$

$$\begin{aligned}
&= \int l(x, \theta_0) d_x n^{-1/2} \sum_{i=1}^{[nt]} (1\{X_i \leq x\} - F(x; \theta_0)) + n^{-1/2} [nt]^{1/2} \epsilon_{1n}(t) \\
&= \int l(x, \theta_0) d_x n^{-1/2} K(F(x; \theta_0), nt) + L_n(t) + n^{-1/2} [nt]^{1/2} \epsilon_{1n}(t),
\end{aligned}$$

where $L_n(t)$ is as in Lemma 3.2.3, and $K(\cdot, \cdot)$ is the same Kiefer process as that of Theorem 2.1.A (cf. also Remark 2.1.2). Hence, putting (3.2.24) in (3.2.21) we have,

$$\begin{aligned}
(3.2.25) \quad \hat{\beta}_n(x, t) &= n^{-1/2} \sum_{i=1}^{[nt]} (1\{X_i \leq x\} - F(x; \theta_0)) - n^{-1/2} K(F(x; \theta_0), nt) \\
&\quad + G_n(x, t) + n^{-1/2} [nt]^{1/2} \epsilon_{3n}(x, t) - L_n(t) \cdot \nabla_{\theta} F(x; \theta_0)^{\top} \\
&\quad + n^{-1/2} [nt]^{1/2} \epsilon_{1n}(t) \cdot \nabla_{\theta} F(x; \theta_0)^{\top},
\end{aligned}$$

where $G_n(x, t)$ is as in (3.2.4).

We first assume only that q is positive on $(0, 1]$ and such that $\lim_{t \downarrow 0} t^{1/2}/q(t) = 0$. Let $0 < \delta < 1$ be fixed. By (3.2.1)(a), (3.2.3) and the fact that $\sup_{\delta \leq t \leq 1} n^{-1/2} [nt]^{1/2} \leq 1$, we have

$$\sup_{\delta \leq t \leq 1} \frac{n^{-1/2} [nt]^{1/2}}{q(t)} \|\epsilon_{1n}(t)\| \leq \left(\inf_{\delta \leq t \leq 1} q(t) \right)^{-1} \sup_{[n\delta] \leq [nt] \leq n} \left([nt]^{-3/4} \sum_{i=1}^{[nt]} (X_i - \mu_0) \right)^2.$$

Let $S_k = \sum_{i=1}^k (X_i - \mu_0)$. By the Hájek-Rényi inequality, given $\epsilon > 0$,

$$P \left\{ \sup_{[n\delta] \leq k \leq n} k^{-3/4} |S_k| \geq \epsilon \right\} \leq \frac{\sigma_0^2}{\epsilon^2} \sum_{k=[n\delta]}^n \frac{1}{k^{3/2}} \rightarrow 0, \text{ as } n \rightarrow \infty,$$

i.e.

$$(3.2.26) \quad \sup_{\delta \leq t \leq 1} \|\epsilon_{1n}(t)\| = o_P(1), \text{ as } n \rightarrow \infty.$$

Thus, since q is positive on $(0, 1]$, by (3.2.5) and (3.2.26), we have that, as $n \rightarrow \infty$,

$$(3.2.27) \quad \sup_{\delta \leq t \leq 1} n^{-1/2} [nt]^{1/2} \|\varepsilon_{1n}(t)\| / q(t) = o_P(1).$$

On the other hand, since $[nt] \leq nt$, by (3.2.1)(a) and (3.2.3), we have

$$\sup_{\frac{1}{n} \leq t \leq \delta} \frac{n^{-1/2} [nt]^{1/2}}{q(t)} \|\varepsilon_{1n}(t)\| \leq \sup_{\frac{1}{n} \leq t \leq \delta} t^{1/2} / q(t) \sup_{1 \leq [nt] \leq [n\delta]} \left([nt]^{-3/4} \sum_{i=1}^{[nt]} (X_i - \mu_0) \right)^2.$$

We note that, for $S_k = \sum_{i=1}^k (X_i - \mu_0)$, and any $C > 0$, by Kolmogorov's inequality,

$$P \left\{ \sup_{1 \leq k \leq [n\delta]} k^{-3/4} |S_k| \geq C \right\} \leq \frac{\sigma_0^2}{C^2} \sum_{k=1}^n \frac{1}{k^{3/2}}.$$

Thus, as $n \rightarrow \infty$,

$$(3.2.28) \quad \sup_{\frac{1}{n} \leq t \leq \delta} \|\varepsilon_{1n}(t)\| = O_P(1).$$

Now, by the assumption that $\lim_{t \downarrow 0} t^{1/2} / q(t) = 0$, we have, by taking $\delta > 0$ sufficiently small,

$$(3.2.29) \quad \sup_{\frac{1}{n} \leq t \leq \delta} \sup_{x \in \mathbb{R}} n^{-1/2} [nt]^{1/2} \|\varepsilon_{1n}(t)\| / q(t) = o_P(1), \text{ as } n \rightarrow \infty.$$

Furthermore, since the vector $\nabla_{\theta} F(x; \theta_0)$ is uniformly bounded in x , we have, from

(3.2.27) and (3.2.29), that, as $n \rightarrow \infty$,

$$(3.2.30) \quad \sup_{\frac{1}{n} \leq t \leq 1} \sup_{x \in \mathbb{R}} n^{-1/2} [nt]^{1/2} |\varepsilon_{1n}(t) \cdot \nabla_{\theta} F(x; \theta_0)^{\top}| / q(t) = o_P(1).$$

We note that, for each $\frac{1}{n} \leq t \leq 1$, by (3.1.2),

$$(3.2.31) \quad \hat{\theta}_{[nt]} - \theta_0 = [nt]^{-1} \sum_{i=1}^{[nt]} l(X_i, \theta_0) + [nt]^{-1/2} \varepsilon_{1n}(t).$$

Let $0 < \delta < 1$ be fixed, and put $T_k = \sum_{i=1}^k l(X_i, \theta_0)$. By (3.2.2), $E l(X_i, \theta_0) = 0$, coordinate-wise, and, hence, T_k is a partial sum of i.i.d. random variables with mean zero. By the Hájek-Rényi inequality, given $\epsilon > 0$,

$$P \left\{ \sup_{[n\delta] \leq k \leq n} k^{-1} |T_k| \geq \epsilon \right\} \leq \frac{M}{\epsilon^2} \sum_{k=[n\delta]}^n \frac{1}{k^2},$$

for some $M > 0$. Thus, by (3.2.26) and (3.2.31),

$$(3.2.32) \quad \sup_{\delta \leq t \leq 1} \|\hat{\theta}_{[nt]} - \theta_0\| = o_P(1), \text{ as } n \rightarrow \infty.$$

Also, for any $C > 0$, by Kolmogorov's inequality,

$$P \left\{ \sup_{1 \leq k \leq [n\delta]} k^{-1} |T_k| \geq C \right\} \leq \frac{M}{C^2} \sum_{k=1}^n \frac{1}{k^2},$$

and, thus, by (3.2.28) and (3.2.31),

$$(3.2.33) \quad \sup_{\frac{1}{n} \leq t \leq \delta} \|\hat{\theta}_{[nt]} - \theta_0\| = O_P(1), \text{ as } n \rightarrow \infty.$$

Since $\nabla_{\theta} F(x; \theta_0)$ is uniformly continuous in x and $\theta \in \Lambda$, where Λ is the closure of a neighbourhood of θ_0 , we get, by (3.2.23) and (3.2.32), that

$$(3.2.34) \quad \sup_{\delta \leq t \leq 1} \|\nabla_{\theta} F(x; \theta_0)^T - \nabla_{\theta} F(x; \tilde{\theta}_{[nt]})^T\| = o_P(1), \text{ as } n \rightarrow \infty.$$

From (3.1.2) and (3.2.3), by the Central Limit Theorem and the Hájek-Rényi inequality, we have

$$\begin{aligned}
\sup_{\delta \leq t \leq 1} \|[nt]^{1/2}(\hat{\theta}_{[nt]} - \theta_0)\| &\leq \sup_{\delta \leq t \leq 1} \|[nt]^{-1/2} \sum_{i=1}^{[nt]} \dot{l}(X_i, \theta_0)\| + \sup_{\delta \leq t \leq 1} \|\varepsilon_{1n}(t)\| \\
(3.2.35) \quad &\leq [n\delta]^{-1/2} \sum_{i=1}^n \|\dot{l}(X_i, \theta_0)\| + \sup_{\delta \leq t \leq 1} \|\varepsilon_{1n}(t)\| \\
&= O_P(1) + o_P(1) = O_P(1), \text{ as } n \rightarrow \infty.
\end{aligned}$$

Thus, from (3.2.22), we have, by (3.2.34) and (3.2.35), as $n \rightarrow \infty$,

$$(3.2.36) \quad \sup_{\delta \leq t \leq 1} \sup_{x \in \mathbb{X}} |\varepsilon_{3n}(x, t)| = o_P(1).$$

Now, since $\sup_{\delta \leq t \leq 1} n^{-1/2}[nt]^{1/2} \leq 1$ and, by (3.2.5), $(\inf_{\delta \leq t \leq 1} q(t))^{-1} < \infty$, we have that, as $n \rightarrow \infty$,

$$(3.2.37) \quad \sup_{\delta \leq t \leq 1} \sup_{x \in \mathbb{X}} n^{-1/2}[nt]^{1/2} |\varepsilon_{3n}(x, t)|/q(t) = o_P(1).$$

On the other hand, after a one-term Taylor expansion around θ_0 , we have

$$\begin{aligned}
&[nt]^{1/2} (\hat{\theta}_{[nt]} - \theta_0) \cdot (\nabla_{\theta} F(x; \theta_0)^{\top} - \nabla_{\theta} F(x; \bar{\theta}_{[nt]})^{\top}) \\
(3.2.38) \quad &= [nt]^{1/2} (\hat{\theta}_{[nt]} - \theta_0) F''_{\theta\theta}(x; \bar{\theta}_{[nt]}) (\bar{\theta}_{[nt]} - \theta_0)^{\top},
\end{aligned}$$

where $F''_{\theta\theta}(x; \alpha) = \begin{pmatrix} \frac{\partial^2 F}{\partial \mu^2}(x; \alpha) & \frac{\partial^2 F}{\partial \sigma^2 \partial \mu}(x; \alpha) \\ \frac{\partial^2 F}{\partial \sigma^2 \partial \mu}(x; \alpha) & \frac{\partial^2 F}{\partial (\sigma^2)^2}(x; \alpha) \end{pmatrix},$
and

$$(3.2.39) \quad \sup_{0 \leq t \leq 1} \|\bar{\theta}_{[nt]} - \theta_0\| \leq \sup_{0 \leq t \leq 1} \|\hat{\theta}_{[nt]} - \theta_0\|.$$

Furthermore, by (3.2.23),

$$(3.2.40) \quad [nt]^{1/2} \left| (\hat{\theta}_{[nt]} - \theta_0) F''_{\theta\theta}(x; \bar{\theta}_{[nt]}) (\bar{\theta}_{[nt]} - \theta_0)^T \right| \\ \leq 4 \| [nt]^{1/4} (\hat{\theta}_{[nt]} - \theta_0) \|^2 \| F''_{\theta\theta}(x; \bar{\theta}_{[nt]}) \|,$$

where $\| \begin{pmatrix} \alpha_1 & \alpha_2 \\ \alpha_3 & \alpha_4 \end{pmatrix} \| = \max_{1 \leq i \leq 4} |\alpha_i|$.

Now, by (3.2.31),

$$[nt]^{1/4} (\hat{\theta}_{[nt]} - \theta_0) = [nt]^{-3/4} \sum_{i=1}^{[nt]} l(X_i, \theta_0) + [nt]^{-1/4} \epsilon_{1n}(t),$$

and, hence, by the Hájek-Rényi inequality (cf. (3.2.33)) and (3.2.28),

$$\sup_{\frac{1}{n} \leq t \leq \delta} [nt]^{1/4} \|\hat{\theta}_{[nt]} - \theta_0\| = O_P(1), \text{ as } n \rightarrow \infty.$$

Since each component of the matrix $F''_{\theta\theta}(x; \theta)$ is uniformly bounded in x and $\theta \in \Lambda$, where Λ is a neighbourhood of θ , by (3.2.23) and (3.2.39), we have

$$\sup_{\frac{1}{n} \leq t \leq \delta} \sup_{x \in \mathbb{R}} \|F''_{\theta\theta}(x; \bar{\theta}_{[nt]})\| = O_P(1), \text{ as } n \rightarrow \infty.$$

Finally, since $[nt] \leq nt$, $\sup_{\frac{1}{n} \leq t \leq \delta} t^{1/2}/q(t) \leq \sup_{0 < t \leq \delta} t^{1/2}/q(t)$ and $\lim_{t \downarrow 0} t^{1/2}/q(t) = 0$, we have,

$$(3.2.41) \quad \sup_{\frac{1}{n} \leq t \leq \delta} \sup_{x \in \mathbb{R}} n^{-1/2} [nt]^{1/2} |\epsilon_{3n}(x, t)|/q(t) = o_P(1), \text{ as } n \rightarrow \infty.$$

Now, by (3.2.37) and (3.2.41), we have

$$(3.2.42) \quad \sup_{\frac{1}{n} \leq t \leq 1} \sup_{x \in \mathbb{R}} n^{-1/2} [nt]^{1/2} |\epsilon_{3n}(x, t)|/q(t) = o_P(1), \text{ as } n \rightarrow \infty.$$

It is clear from (3.2.25) that, by (3.2.30), (3.2.42), Lemma 3.2.3 and the proof of Theorem 2.1.A (cf. (2.1.8)), what we have shown so far is that for any q positive on $(0, 1]$ and such that $\lim_{t \downarrow 0} t^{1/2}/q(t) = 0$, as $n \rightarrow \infty$,

$$(3.2.43) \quad \sup_{0 < t \leq 1} \sup_{x \in \mathbb{R}} |\hat{\beta}_n(x, t) - \bar{G}_n(x, t)|/q(t) \stackrel{a.s.}{=} o(1),$$

where

$$\bar{G}_n(x, t) = \begin{cases} G_n(x, t), & x \in \mathbb{R}, \frac{1}{n} \leq t \leq 1, \\ 0, & \text{elsewhere.} \end{cases}$$

By definition,

$$(3.2.44) \quad \hat{\beta}_n(x, t) = 0 \text{ for all } x \in \mathbb{R} \text{ and } 0 < t < \frac{1}{n}.$$

Hence, to prove (a) and (b), it suffices to show that

$$(3.2.45) \quad \sup_{0 < t < \frac{1}{n}} \sup_{x \in \mathbb{R}} |G_n(x, t)|/q(t) = \begin{cases} o_P(1), & \text{if (3.2.6) holds,} \\ O_P(1), & \text{if (3.2.7) holds.} \end{cases}$$

To this end, let

$$(3.2.46) \quad M(\theta_0) = E\{l(X_j, \theta_0)^\top l(X_j, \theta_0)\},$$

where $l(\cdot, \theta_0)$ is as in (3.2.2). Since $M(\theta_0)$ is nonnegative definite, there is a nonsingular matrix $D(\theta_0)$ such that

$$(3.2.47) \quad D(\theta_0)^\top M(\theta_0) D(\theta_0) = I,$$

where I is the 2×2 identity matrix.

Since, $E l(X_i, \theta_0) = 0$, we have that, for each $n \geq 1$ and $0 \leq s, t \leq 1$ (cf. (A.2.6)),

$$(3.2.48) \quad E \left\{ \int l(x, \theta_0) dx n^{-1/2} K(F(x; \theta_0), ns) \right\}^\top \cdot \left\{ \int l(y, \theta_0) dy n^{-1/2} K(F(y; \theta_0), nt) \right\} \\ = (t \wedge s) M(\theta_0).$$

Hence, from (3.2.4), we have that

$$(3.2.49) \quad G_n(x, t) = n^{-1/2} K(F(x; \theta_0), nt) - n^{-1/2} W(nt) \cdot D(\theta_0)^{-1} \cdot \nabla_\theta F(x; \theta_0)^\top,$$

where

$$(3.2.50) \quad n^{-1/2} W(nt) = \int l(x, \theta_0) dx n^{-1/2} K(F(x; \theta_0), nt) \cdot D(\theta_0)$$

is a vector valued Wiener process with covariance function given by $t \wedge s$ multiplied by (3.2.47), for each $n \geq 1$.

Next, we have that, since

$$\sup_{0 < t \leq 1} |n^{-1/2} W(nt)|/q(t) \stackrel{\mathcal{D}}{=} \sup_{0 < t \leq 1} |W(t)|/q(t),$$

for each $n \geq 1$, by Theorem 2.1.B (cf. our Remark 2.1.5),

$$\sup_{0 < t < \frac{1}{n}} |n^{-1/2} W(nt)|/q(t) = \begin{cases} o_P(1), & \text{if (3.2.6) holds,} \\ O_P(1), & \text{if (3.2.7) holds.} \end{cases}$$

Similarly, since, for each $n \geq 1$,

$$(3.2.51)$$

$$\{n^{-1/2} K(F(x; \theta_0), nt), 0 \leq t \leq 1, x \in \mathbb{R}\} \stackrel{\mathcal{D}}{=} \{K(F(x; \theta_0), t), 0 \leq t \leq 1, x \in \mathbb{R}\},$$

we have that

$$\sup_{\frac{1}{n} < t \leq 1} \sup_{x \in \mathbb{R}} |n^{-1/2} K(F(x; \theta_0), nt)|/q(t) \stackrel{D}{=} \sup_{\frac{1}{n} < t \leq 1} \sup_{x \in \mathbb{R}} |K(F(x; \theta_0), t)|/q(t).$$

By Lemma 2.1.A (cf. our Remark 2.1.4), we get, as $n \rightarrow \infty$,

$$\sup_{\frac{1}{n} < t \leq 1} \sup_{x \in \mathbb{R}} |n^{-1/2} K(F(x; \theta_0), nt)|/q(t) = \begin{cases} o_P(1), & \text{if (3.2.6) holds,} \\ O_P(1), & \text{if (3.2.7) holds.} \end{cases}$$

Hence, (3.2.45) holds and (a) and (b) follow.

As to the proof of (c), we note that, by (3.2.4), (3.2.8) and (3.2.51), for any $n \geq 1$,

$$(3.2.52) \quad \{G_n(x, t), 0 \leq t \leq 1, x \in \mathbb{R}\} \stackrel{D}{=} \{G(x, t), 0 \leq t \leq 1, x \in \mathbb{R}\}.$$

Next, we note that, by (3.2.43),

$$(3.2.53)$$

$$\lim_{n \rightarrow \infty} \left| P \left\{ \sup_{0 < t \leq 1} \sup_{x \in \mathbb{R}} |\hat{\beta}_n(x, t)|/q(t) \leq x \right\} - P \left\{ \sup_{\frac{1}{n} < t \leq 1} \sup_{x \in \mathbb{R}} |G_n(x, t)|/q(t) \leq x \right\} \right| = 0,$$

for any $x \in \mathbb{R}$. Also, with any function $q \in \mathcal{Q}$ and such that (3.2.7) holds, we have

$$\lim_{n \rightarrow \infty} P \left\{ \sup_{\frac{1}{n} < t \leq 1} \sup_{x \in \mathbb{R}} |G(x, t)|/q(t) \leq x \right\} = P \left\{ \sup_{0 < t \leq 1} \sup_{x \in \mathbb{R}} |G(x, t)|/q(t) \leq x \right\},$$

for any $x \in \mathbb{R}$. This, together with (3.2.52) and (3.2.53), implies (c) ■

Corollary 3.2.1. *Assuming the conditions of Theorem 3.2.1, we have, as $n \rightarrow \infty$,*

(a) *if $q \in \mathcal{Q}$ is such that (3.2.6) holds, then*

$$(3.2.54) \quad \hat{\beta}_n(x, t)/q(t) \xrightarrow{D} G(x, t)/q(t)$$

in $D(\mathbb{R} \times [0, 1])$,

(b) *if q is positive on $(0, 1]$, then (3.2.54) holds in $D(\mathbb{R} \times [\delta, 1])$ for any $0 < \delta < 1$.*

Proof. (a) follows from Theorem 3.2.1 (a).

(b) By Theorem 3.2.1 (a) with $q(t) \equiv 1$, and (3.2.5). ■

3.3 Sequentially estimated normal empirical processes tied down at $t=1$.

When testing for the possibility of a change in distribution in a random sequence of chronologically ordered observations $X_1, X_2, \dots, X_n$, it is of interest to make use of test statistics based on comparing the empirical distribution function of the first k observations to that of the last $n - k$ observations (cf. our Introduction for some discussion on the changepoint problem). For example, Picard (1985), and Deshayes and Picard (1986) consider

(3.3.1)

$$\begin{aligned} & \sup_{1 \leq k < n} \sup_{x \in \mathbb{R}} n^{1/2} \left| \frac{1}{k} \sum_{i=1}^k 1\{X_i \leq x\} - \frac{1}{n-k} \sum_{i=k+1}^n 1\{X_i \leq x\} \right| \\ &= \sup_{1 \leq k < n} \sup_{x \in \mathbb{R}} n^{1/2} \left| \sum_{i=1}^k 1\{X_i \leq x\} - \frac{k}{n} \sum_{i=1}^n 1\{X_i \leq x\} \right| / (k(n-k)/n). \end{aligned}$$

Since, given $H_0 : X_1, \dots, X_n$ are i.i.d., as $n \rightarrow \infty$ we have

$$\sup_{1 \leq k < n} \sup_{x \in \mathbb{R}} \left| \sum_{i=1}^k 1\{X_i \leq x\} - \frac{k}{n} \sum_{i=1}^n 1\{X_i \leq x\} \right| / \left[n^{1/2} \frac{k}{n} \left(1 - \frac{k}{n} \right) \right] \xrightarrow{P} \infty,$$

and even

$$\sup_{1 \leq k < n} \sup_{x \in \mathbb{R}} \left| \sum_{i=1}^k 1\{X_i \leq x\} - \frac{k}{n} \sum_{i=1}^n 1\{X_i \leq x\} \right| / \left[n^{1/2} \left(\frac{k}{n} \left(1 - \frac{k}{n} \right) \right)^{\frac{1}{2}} \right] \xrightarrow{P} \infty,$$

in order to have the above test statistics converge in distribution to a nondegenerate random variable, we have to renormalize them somehow. We may, for example, wish to introduce weight functions which are less severe on the tails than $((k/n)(1 - k/n))^{\frac{1}{2}}$ is.

Just like tests based on the classical Kolmogorov-Smirnov statistic, the ones based on

(3.3.1*)

$$\begin{aligned} & \sup_{1 \leq k < n} \sup_{x \in \mathbb{R}} \frac{k}{n} \left(1 - \frac{k}{n} \right) n^{1/2} \left| \frac{1}{k} \sum_{i=1}^k 1\{X_i \leq x\} - \frac{1}{n-k} \sum_{i=k+1}^n 1\{X_i \leq x\} \right| \\ &= \sup_{1 \leq k < n} \sup_{x \in \mathbb{R}} n^{-1/2} \left| \sum_{i=1}^k 1\{X_i \leq x\} - \frac{k}{n} \sum_{i=1}^n 1\{X_i \leq x\} \right| \end{aligned}$$

should be more powerful for detecting changes that occur in the middle, namely near $n/2$ where $\frac{k}{n} (1 - \frac{k}{n})$ has its maximum, than for noticing the ones occurring near the endpoints 0 and n . Thus, a weighted version of (3.3.1*) should emphasize changes which may have occurred near the endpoints, while retaining sensitivity to possible changes in the middle as well. To see what kind of weight functions are possible for the statistics in (3.3.1), or in (3.3.1*), Szyszkowicz (1991b, 1992b, 1994), and Csörgő and Szyszkowicz (1994) studied the asymptotic behaviour of the processes

$$(\alpha_n(s, t) - t\alpha_n(s, 1))/q(t), \quad 0 \leq s, t \leq 1$$

for a wide class of weight functions q , where

$$\begin{aligned}\alpha_n(F(x), t) &= n^{-1/2} \sum_{i=1}^{[nt]} (1\{F(X_i) \leq F(x)\} - F(x)), \\ &= \beta_n(x, t), \quad 0 \leq t \leq 1, \quad x \in \mathbb{R}.\end{aligned}$$

We note that, in our terminology, the process $\{\beta_n(x, t), 0 \leq t \leq 1, x \in \mathbb{R}\}$ is defined as

$$\beta_n(x, t) = n^{-1/2} \sum_{i=1}^{[nt]} (1\{X_i \leq x\} - F(x; \theta_0)).$$

We will call q a positive function on $(0, 1)$ if $q : (0, 1) \rightarrow (0, \infty)$ is such that

$$(3.3.2) \quad \inf_{\delta \leq t \leq 1-\delta} q(t) > 0 \text{ for all } 0 < \delta \leq 1/2.$$

Let $\mathcal{Q}^*$ be the class of positive functions on $(0, 1)$ that are non-decreasing in a neighbourhood of 0 and non-increasing in a neighbourhood of 1. Consider the integral

$$(3.3.3) \quad I_{0,1}(q, c) = \int_0^1 \frac{1}{t(1-t)} e^{-cq^2(t)/t(1-t)} dt, \quad c > 0.$$

With $\{K(s, t), 0 \leq s, t \leq 1\}$, a Kiefer process, we define $\Gamma(\cdot, \cdot)$ by

$$\Gamma(s, t) = K(s, t) - tK(s, 1), \quad 0 \leq s, t \leq 1.$$

Consequently, $\{\Gamma(s, t), 0 \leq s, t \leq 1\}$ is a separable Gaussian process with mean zero and covariance function

$$E \Gamma(s_1, t_1) \Gamma(s_2, t_2) = (s_1 \wedge s_2 - s_1 s_2)(t_1 \wedge t_2 - t_1 t_2).$$

We note that $\{\Gamma(F(x), t), 0 \leq t \leq 1, x \in \mathbb{R}\}$ is a separable Gaussian process with mean zero and covariance function

$$(3.3.4) \quad E \Gamma(F(x_1), t_1) \Gamma(F(x_2), t_2) = (F(x_1 \wedge x_2) - F(x_1)F(x_2))(t_1 \wedge t_2 - t_1 t_2).$$

Consider the process

$$(3.3.5) \quad \beta_n^*(x, t) = \begin{cases} n^{-1/2} \left(\sum_{i=1}^{[(n+1)t]} 1\{X_i \leq x\} - \frac{[(n+1)t]}{n} \sum_{i=1}^n 1\{X_i \leq x\} \right), & 0 \leq t < 1, \\ & x \in \mathbb{R}, \\ 0, & t = 1. \end{cases}$$

(cf. Szyszkowicz, 1994, pg. 283, where, in our terminology, we have $1\{X_i \leq x\}$, instead of $1\{F(X_i; \theta_0) \leq F(x; \theta_0)\}$ there).

We have,

Theorem 3.3.A. *Let $X_1, X_2, \dots$ be a sequence of i.i.d random variables with a continuous distribution function F . Then, there exists a Kiefer process $K(\cdot, \cdot)$ such that with the sequence of stochastic processes $\Gamma_n(\cdot, \cdot)$*

$$(3.3.6) \quad \begin{aligned} & \{\Gamma_n(F(x; \theta_0), t), 0 \leq t \leq 1, x \in \mathbb{R}\} \\ &= \{n^{-1/2}(K(F(x; \theta_0), nt) - tK(F(x; \theta_0), n)), 0 \leq t \leq 1, x \in \mathbb{R}\} \\ &\stackrel{D}{=} \{\Gamma(F(x; \theta_0), t), 0 \leq t \leq 1, x \in \mathbb{R}\} \text{ for each } n \geq 1, \end{aligned}$$

and with a weight function $q \in \mathcal{Q}^*$, we have, as $n \rightarrow \infty$,

(a)

$$(3.3.7) \quad I_{0,1}(q, c) < \infty \text{ for all } c > 0$$

if and only if

$$\sup_{0 < t < 1} \sup_{x \in \mathbb{R}} |\beta_n^*(x, t) - \Gamma_n(F(x), t)|/q(t) = o_P(1),$$

(b)

$$(3.3.8) \quad I_{0,1}(q, c) < \infty \text{ for some } c > 0$$

if and only if

$$\sup_{0 < t < 1} \sup_{x \in \mathbb{R}} |\beta_n^*(x, t) - \Gamma_n(F(x), t)|/q(t) = O_P(1),$$

(c) (3.3.8) holds if and only if

$$\sup_{0 < t < 1} \sup_{x \in \mathbb{R}} |\beta_n^*(x, t)|/q(t) \xrightarrow{D} \sup_{0 < t < 1} \sup_{x \in \mathbb{R}} |\Gamma(F(x), t)|/q(t),$$

where $\Gamma(\cdot, \cdot)$ is a Gaussian process with mean zero and covariance function as in (3.3.4).

Remark 3.3.1. (b) and (c) do not follow from (a), nor does (c) from (b).

Remark 3.3.2. Similarly as in the case of Theorem 2.1.A (cf. Remark 2.1.4), from the proof of Theorem 3.3.A, it can be seen that, provided that q is a positive function on $(0, 1)$ and such that $\lim_{t \downarrow 0} t^{1/2}/q(t) = \lim_{t \uparrow 1} (1 - t)^{1/2}/q(t) = 0$, we have

$$(3.3.9) \quad \sup_{\frac{1}{n} < t < 1} \sup_{x \in \mathbb{R}} |\beta_n^*(x, t) - \Gamma_n(F(x), t)|/q(t) \stackrel{a.s.}{=} o(1).$$

(cf. also Lemma 3.1 of Csörgő, Horváth and Szyszkowicz, 1994).

Since, by definition, $\beta_n^*(x, t) = 0$, for any $t \in A_n := (0, \frac{1}{n+1}] \cup [\frac{n}{n+1}, 1)$, and

$$\sup_{x \in \mathbb{R}} |\Gamma_n(F(x), t)| \stackrel{P}{=} \sup_{x \in \mathbb{R}} |\Gamma(F(x), t)| \text{ for each } n \geq 1,$$

any sufficient condition on the class of weight functions q for which Theorem 3.3.A holds must be given in terms of the local behaviour (near zero and near one) of $\sup_{x \in \mathbb{R}} |\Gamma(F(x), t)|/q(t)$.

Remark 3.3.3. Theorem 3.3.A is Theorem 3.1 of Szyszkowicz (1995). It was first proved by Szyszkowicz (1992b, 1994) (cf. Theorem 2.2 of Szyszkowicz (1994)), under stronger conditions, resulting from the law of the iterated logarithm for the suprema of Kiefer processes, or, equivalently, from the Adler and Brown (1980) test for upper and lower functions for the suprema of Kiefer processes (cf. Remark 2.1.6 here).

Theorem 3.3.A is an improvement of Theorem 2.2 of Szyszkowicz (1994), which is due to the following integral test of Csörgő, Horváth and Szyszkowicz (1994) for suprema of the process $\Gamma(F(x), t)$ (cf. their Theorem 2.3, with $d = 1$).

Theorem 3.3.B. *Let $q \in \mathcal{Q}^*$ and $K(\cdot, \cdot)$ be a Kiefer process. Then,*

(a) (3.3.7) holds if and only if

$$\lim_{t \downarrow 0} \sup_{x \in \mathbb{R}} |\Gamma(F(x), t)|/q(t) = 0 \text{ a.s.}$$

and

$$\lim_{t \uparrow 1} \sup_{x \in \mathbb{R}} |\Gamma(F(x), t)|/q(t) = 0 \text{ a.s.,}$$

(b) (3.3.8) holds if and only if

$$\lim_{t \downarrow 0} \sup_{x \in \mathbb{R}} |\Gamma(F(x), t)|/q(t) < \infty \text{ a.s.}$$

and

$$\lim_{t \uparrow 1} \sup_{x \in \mathbb{R}} |\Gamma(F(x), t)|/q(t) < \infty \text{ a.s.}$$

Remark 3.3.4. The integral in (3.3.3) was used in Csörgő, Csörgő, Horváth and Mason (1986) for characterizing the class of functions for which

$$\limsup_{t \downarrow 0, t \uparrow 1} |B(t)|/q(t) < \infty \text{ a.s.},$$

where $\{B(t), 0 \leq t \leq 1\}$ is a Brownian bridge process (cf. 1.1.8) (local functions of Brownian bridge).

We have the following result (cf. Theorems 3.3 and 3.4 of Csörgő, Csörgő, Horváth and Mason (1986))

Theorem 3.3.C. *Let $q \in \mathcal{Q}^*$ and $B(\cdot)$ be a Brownian bridge process. Then,*

(a) (3.3.7) holds if and only if

$$\lim_{t \downarrow 0} |B(t)|/q(t) = 0 \text{ a.s.}$$

and

$$\lim_{t \uparrow 1} |B(t)|/q(t) = 0 \text{ a.s.},$$

(b) (3.3.8) holds if and only if

$$\limsup_{t \downarrow 0} |B(t)|/q(t) < \infty \text{ a.s.}$$

and

$$\limsup_{t \uparrow 1} |B(t)|/q(t) < \infty \text{ a.s.}$$

On account of Theorems 3.3.B and 3.3.C, we have that the class of functions q that characterize the local behaviour (near 0 and near 1) of the process $\sup_{x \in \mathbb{R}} |\Gamma(F(x), t)|$ is the same class of functions that characterize the similar local behaviour of Brownian bridge processes.

Remark 3.3.5. We assume that $\{W(t), 0 \leq t < \infty\}$ is a Wiener process and $\{B(t), 0 \leq t \leq 1\}$ is a Brownian bridge define by

$$B(t) = W(t) - tW(1).$$

For each Wiener process, $B(t)$, as above, is always a Brownian bridge, and each Brownian bridge can be written in the above given form for a suitably chosen Wiener process. Hence, when studying weighted Brownian bridges, it will be enough to study the similar behaviour of a Wiener process only. In the light of the above, the class of functions that characterize the local behaviour (near 0) of Wiener processes is the same class of functions that characterize the local behaviour (near 0 and near 1) of Brownian bridge processes. Thus, the results of Lemma 2.1.A and Theorem 2.1.B (cf. Remarks 2.1.4 and 2.1.5), as well as the results of Theorems 3.3.B and 3.3.C here, are all equivalent. (cf. Szyszkowicz (1992b) and Csörgő, and Horváth, 1993, Chapter 4, for a general survey and further references).

In the case where we have a sequence of i.i.d. random variables from a family of continuous distribution functions $\{F(x; \theta); x \in \mathbb{R}, \theta \in \Theta \subseteq \mathbb{R}^p\}$, where θ is unknown, we propose to study an *estimated* version of (3.3.1*) in $\|\cdot/q\|$ -metrics, for the same class of functions q as considered in Theorem 3.3.A. We should note, however, that if the estimation of the unknown parameter(s) is done as in Chapter 2, i.e. by using the complete random sample $X_1, \dots, X_n$, and if we define the two time parameter estimated empirical process as in (2.2.1), then our *estimated* process for the statistics as in the first line of (3.3.1*) becomes

$$\begin{aligned} \frac{k}{n} \left(1 - \frac{k}{n}\right) n^{1/2} & \left(\frac{1}{k} \sum_{i=1}^k (1\{X_i \leq x\} - F(x; \hat{\theta}_n)) \right. \\ & \left. - \frac{1}{n-k} \sum_{i=k+1}^n (1\{X_i \leq x\} - F(x; \hat{\theta}_n)) \right). \end{aligned}$$

But since the term $F(x; \hat{\theta}_n)$ above vanishes, we get back the statistics in (3.3.1*), as if there was no estimation at all. Hence, on account of this, it is inviting to entertain the notion of estimating the parameters indexing the underlying distribution the way the distribution function itself was estimated in (3.3.1*), namely to estimate parameters "before" and "after" the change. Thus, we propose to make the following comparisons

(3.3.10)

$$\frac{1}{k} \sum_{i=1}^k (1\{X_i \leq x\} - F(x; \hat{\theta}_k)) - \frac{1}{n-k} \sum_{i=k+1}^n (1\{X_i \leq x\} - F(x; \hat{\theta}_{n-k})), \quad 1 \leq k \leq n,$$

with a proper normalization, where $\hat{\theta}_k$ and $\hat{\theta}_{n-k}$ are estimators based on the samples $X_1, X_2, \dots, X_k$ and $X_{k+1}, X_{k+2}, \dots, X_n$, respectively.

In terms of two-time parameter estimated empirical processes, what we propose is to study the asymptotic behaviour of the processes $\{\hat{\beta}_n(x, t), 0 \leq t \leq 1, x \in \mathbb{R}\}$ defined by

$$(3.3.11) \quad \hat{\beta}_n(x, t) = \begin{cases} n^{1/2} \frac{[(n+1)t]}{n} \left(1 - \frac{[(n+1)t]}{n}\right) \hat{\zeta}_n(x, t), & 0 \leq t \leq 1, x \in \mathbb{R}, \\ 0, & t = 1, x \in \mathbb{R}, \end{cases}$$

with

(3.3.11*)

$$\hat{\zeta}_n(x, t) = \frac{\sum_{i=1}^{[(n+1)t]} [1\{X_i \leq x\} - F(x; \hat{\theta}'_{[nt]})]}{[(n+1)t]} - \frac{\sum_{i=[(n+1)t]+1}^n [1\{X_i \leq x\} - F(x; \hat{\theta}''_{[nt]})]}{n - [(n+1)t]},$$

in weighted metrics, along the lines of Theorem 3.3.A, where, for each $0 < t < 1$, $\{\hat{\theta}'_{[nt]}\}$ and $\{\hat{\theta}''_{[nt]}\}$ are sequences of estimators of θ based on $X_1, \dots, X_{[(n+1)t]}$ and $X_{[(n+1)t]+1}, \dots, X_n$, respectively, and $F(\cdot; \cdot)$ denotes the common underlying distribution function. Thus, we are to study properly normalized differences between

the sequentially estimated empirical process "before" a certain point, and the sequentially estimated empirical process "after" that point, for a suitably chosen type of estimator. The results can provide applications to testing for a change in the distribution of a random sequence at an unknown point that will combine the non-parametric results for empirical processes with the additional parametric information coming from the estimation.

We note, furthermore that, as in the non-parametric case, a properly weighted version of (3.3.11) should emphasize changes near the endpoints while retaining sensitivity in the middle. This is why we are aiming at results like those of Theorem 3.3.A in this context as well. Indeed, we obtain results for the same optimal class of weight functions q as in Theorem 3.3.A, namely, positive functions on $(0, 1)$, non-decreasing near zero, non-increasing near one, that satisfy the same integral tests.

As in the previous section, we will initiate this study under the assumption of $X_1, X_2, \dots$ being a sequence of i.i.d. r.v.'s from a normal family.

Let $X_1, X_2, \dots$ be an independent sequence of random variables with common distribution function given by

$$F(x; \theta) = \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^x e^{-(t-\mu)^2/2\sigma^2} dt, \quad \theta := (\mu, \sigma^2).$$

With $\frac{1}{n+1} \leq t \leq \frac{n}{n+1}$, let $\{\hat{\theta}'_{[nt]}\}$ and $\{\hat{\theta}''_{[nt]}\}$ be sequences of maximum likelihood estimators of θ based on $X_1, \dots, X_{[(n+1)t]}$ and $X_{[(n+1)t]+1}, \dots, X_n$, respectively. It is known that under our assumption of normality

$$(3.3.12) \quad \hat{\theta}'_{[nt]} := \left(\bar{X}_{[nt]}, S_{[nt]}^2 \right) = \left(\frac{1}{[(n+1)t]} \sum_{i=1}^{[(n+1)t]} X_i, \frac{1}{[(n+1)t]} \sum_{i=1}^{[(n+1)t]} (X_i - \bar{X}_{[nt]})^2 \right)$$

and

(3.3.13)

$$\begin{aligned}\hat{\theta}_{[nt]}'' &:= (\bar{X}_{[nt]}, \bar{S}_{[nt]}^2) \\ &= \left(\frac{1}{n - [(n+1)t]} \sum_{i=[(n+1)t]+1}^n X_i, \frac{1}{n - [(n+1)t]} \sum_{i=[(n+1)t]+1}^n (X_i - \bar{X}_{[nt]})^2 \right).\end{aligned}$$

Consider now the two-time parameter process $\{\hat{\beta}_n(x, t), 0 \leq t \leq 1, x \in \mathbb{R}\}$ given by (3.3.11). This process will be approximated, in $\|\cdot/q\|$ -metrics, by the sequence of stochastic processes $\{\Psi_n(x, t), 0 \leq t \leq 1, x \in \mathbb{R}\}$ defined by

$$(3.3.14) \quad \Psi_n(x, t) = \Gamma_n(F(x; \theta_0), t) - \left\{ \int l(x, \theta_0) d_x \Gamma_n(F(x; \theta_0), t) \right\} \cdot \nabla_{\theta} F(x; \theta_0)^{\top},$$

where $\theta_0 = (\mu_0, \sigma_0^2)$ denotes the true value of the unknown parameter θ , and $\Gamma_n(\cdot, \cdot)$ is as in Theorem 3.3.A, i.e.,

$$\begin{aligned}\{\Gamma_n(F(x; \theta_0), t), 0 \leq t \leq 1, x \in \mathbb{R}\} = \\ \{n^{-1/2}(K(F(x; \theta_0), nt) - tK(F(x; \theta_0), n)), 0 \leq t \leq 1, x \in \mathbb{R}\},\end{aligned}$$

with $K(\cdot, \cdot)$ being the Kiefer process constructed in Komlós, Major and Tusnády (1975) (cf. Theorem 1.1.B).

We have

Theorem 3.3.1. *Let $X_1, X_2, \dots$ be a sequence of i.i.d. $N(\mu, \sigma^2)$ r.v.'s. Let $\hat{\beta}_n(x, t)$ be the sequentially estimated normal empirical process, tied down at $t=1$, defined as in (3.3.11), with $\{\hat{\theta}_{[nt]}'\}$ and $\{\hat{\theta}_{[nt]}''\}$ as in (3.3.12) and (3.3.13), respectively. Then, there exists a Kiefer process $K(\cdot, \cdot)$ such that with the sequence of Gaussian processes $\{\Psi_n(x, t), 0 \leq t \leq 1, x \in \mathbb{R}\}$ and a weight function $q \in \mathcal{Q}^*$, we have, as $n \rightarrow \infty$,*

(a) if

$$(3.3.15) \quad I_{0,1}(q, c) < \infty \text{ for all } c > 0,$$

then

$$\sup_{0 < t < 1} \sup_{x \in \mathbb{R}} |\hat{\beta}_n(x, t) - \Psi_n(x, t)|/q(t) = o_P(1),$$

(b) if

$$(3.3.16) \quad I_{0,1}(q, c) < \infty \text{ for some } c > 0,$$

then

$$\sup_{0 < t < 1} \sup_{x \in \mathbb{R}} |\hat{\beta}_n(x, t) - \Psi_n(x, t)|/q(t) = O_P(1),$$

(c) if (3.3.16) holds then

$$\sup_{0 < t < 1} \sup_{x \in \mathbb{R}} |\hat{\beta}_n(x, t)|/q(t) \xrightarrow{P} \sup_{0 < t < 1} \sup_{x \in \mathbb{R}} |\Psi(x, t)|/q(t),$$

where the process $\{\Psi(x, t), 0 \leq t \leq 1, x \in \mathbb{R}\}$ is defined by

$$(3.3.17) \quad \Psi(x, t) = \Gamma(F(x; \theta_0), t) - \left\{ \int l(x, \theta_0) d_x \Gamma(F(x; \theta_0), t) \right\} \cdot \nabla_{\theta} F(x; \theta_0)^{\top}.$$

Proof of Theorem 3.3.1. From (3.3.11*) we have, after adding and subtracting $F(x; \theta_0)$,

$$(3.3.18) \quad \begin{aligned} \hat{\zeta}_n(x, t) = & \frac{\sum_{i=1}^{[(n+1)t]} (1\{X_i \leq x\} - F(x; \theta_0))}{[(n+1)t]} - \frac{\sum_{i=[(n+1)t]+1}^n (1\{X_i \leq x\} - F(x; \theta_0))}{n - [(n+1)t]} \\ & + (F(x; \theta_0) - F(x; \hat{\theta}'_{[nt]})) + (F(x; \hat{\theta}''_{[nt]}) - F(x; \theta_0)). \end{aligned}$$

Now, applying a one-term Taylor expansion to $F(x; \theta)$, around θ_0 , we get

$$\begin{aligned}
 F(x; \theta_0) - F(x; \hat{\theta}'_{[nt]}) &= (\hat{\theta}'_{[nt]} - \theta_0) \cdot \left(\nabla_{\theta} F(x; \theta_0)^{\top} - \nabla_{\theta} F(x; \tilde{\theta}'_{[nt]})^{\top} \right) \\
 &\quad - (\hat{\theta}'_{[nt]} - \theta_0) \cdot \nabla_{\theta} F(x; \theta_0)^{\top} \\
 (3.3.19) \qquad &= \epsilon'_{3n}(x, t) - (\hat{\theta}'_{[nt]} - \theta_0) \cdot \nabla_{\theta} F(x; \theta_0)^{\top},
 \end{aligned}$$

where

$$(3.3.20) \qquad \epsilon'_{3n}(x, t) = (\hat{\theta}'_{[nt]} - \theta_0) \cdot \left(\nabla_{\theta} F(x; \tilde{\theta}'_{[nt]})^{\top} - \nabla_{\theta} F(x; \theta_0)^{\top} \right)$$

and

$$\sup_{0 \leq t \leq 1} \|\hat{\theta}'_{[nt]} - \theta_0\| \leq \sup_{0 \leq t \leq 1} \|\tilde{\theta}'_{[nt]} - \theta_0\|.$$

Similarly, we have

$$\begin{aligned}
 F(x; \hat{\theta}''_{[nt]}) - F(x; \theta_0) &= (\hat{\theta}''_{[nt]} - \theta_0) \cdot \left(\nabla_{\theta} F(x; \theta_0)^{\top} - \nabla_{\theta} F(x; \tilde{\theta}''_{[nt]})^{\top} \right) \\
 &\quad + (\hat{\theta}''_{[nt]} - \theta_0) \cdot \nabla_{\theta} F(x; \theta_0)^{\top} \\
 (3.3.21) \qquad &= \epsilon''_{3n}(x, t) + (\hat{\theta}''_{[nt]} - \theta_0) \cdot \nabla_{\theta} F(x; \theta_0)^{\top},
 \end{aligned}$$

where

$$(3.3.22) \qquad \epsilon''_{3n}(x, t) = (\hat{\theta}''_{[nt]} - \theta_0) \cdot \left(\nabla_{\theta} F(x; \theta_0)^{\top} - \nabla_{\theta} F(x; \tilde{\theta}''_{[nt]})^{\top} \right)$$

and

$$\sup_{0 \leq t \leq 1} \|\hat{\theta}''_{[nt]} - \theta_0\| \leq \sup_{0 \leq t \leq 1} \|\tilde{\theta}''_{[nt]} - \theta_0\|.$$

It can be easily shown from (3.3.12) that (cf. (3.2.1), (3.2.2) and (3.2.3))

$$(3.3.23) \qquad [(n+1)t]^{\frac{1}{2}}(\hat{\theta}'_{[nt]} - \theta_0) = [(n+1)t]^{-\frac{1}{2}} \sum_{i=1}^{[(n+1)t]} l(X_i; \theta_0) + \epsilon'_{1n}(t),$$

where

$$\epsilon'_{1n}(t) = \left(0, -[(n+1)t]^{\frac{1}{2}}(\bar{X}_{[nt]} - \mu_0)^2\right)$$

and

$$l(x, \theta_0) = (x - \mu_0, (x - \mu_0)^2 - \sigma_0^2).$$

Also, by (3.3.13) we have

(3.3.24)

$$(n - [(n+1)t])^{\frac{1}{2}}(\hat{\theta}_{[nt]}'' - \theta_0) = (n - [(n+1)t])^{-\frac{1}{2}} \sum_{i=[(n+1)t]+1}^n l(X_i; \theta_0) + \epsilon''_{1n}(t),$$

where

$$(3.3.25) \quad \epsilon''_{1n}(t) = \left(0, -(n - [(n+1)t])^{\frac{1}{2}}(\bar{X}_{[nt]} - \mu_0)^2\right).$$

Thus, by (3.3.11), (3.3.18), (3.3.19), (3.3.20), (3.3.21), (3.3.22), (3.3.23) and (3.3.24) we get,

$$(3.3.26) \quad \hat{\beta}_n(x, t) = \beta_n^*(x, t) - n^{-1/2} \left\{ \left(1 - \frac{[(n+1)t]}{n}\right) \sum_{i=1}^{[(n+1)t]} l(X_i; \theta_0) \cdot \nabla_{\theta} F(x; \theta_0)^{\top} \right.$$

$$\begin{aligned} & - \frac{[(n+1)t]}{n} \sum_{i=[(n+1)t]+1}^n l(X_i; \theta_0) \cdot \nabla_{\theta} F(x; \theta_0)^{\top} \} \\ & + \frac{[(n+1)t]}{n} \left(1 - \frac{[(n+1)t]}{n}\right) n^{1/2} \{\epsilon'_{3n}(x, t) + \epsilon''_{3n}(x, t)\} \\ & + \frac{[(n+1)t]}{n} n^{-1/2} (n - [(n+1)t])^{\frac{1}{2}} \epsilon''_{1n}(t) \cdot \nabla_{\theta} F(x; \theta_0)^{\top} \\ & - \left(1 - \frac{[(n+1)t]}{n}\right) n^{-1/2} [(n+1)t]^{\frac{1}{2}} \epsilon'_{1n}(t) \cdot \nabla_{\theta} F(x; \theta_0)^{\top}, \end{aligned}$$

where, $\beta_n^*(x, t)$ is as in (3.3.5).

We assume first that q is a positive function on $(0, 1)$ and such that

$$(3.3.27) \quad \lim_{t \downarrow 0} t^{1/2}/q(t) = \lim_{t \uparrow 1} (1-t)^{1/2}/q(t) = 0.$$

Let

$$(3.3.28) \quad \varphi_n(t) = n^{-1/2} \left\{ \left(1 - \frac{[(n+1)t]}{n}\right) \sum_{i=1}^{[(n+1)t]} l(X_i; \theta_0) - \frac{[(n+1)t]}{n} \sum_{i=[(n+1)t]+1}^n l(X_i; \theta_0) \right\}.$$

Adding and subtracting $\Psi_n(x, t)$ (cf. (3.3.14)), we get, in (3.3.26),

$$(3.3.29) \quad \begin{aligned} & \sup_{\frac{1}{n+1} \leq t \leq \frac{n}{n+1}} \sup_{x \in \mathbb{R}} |\hat{\beta}_n(x, t) - \Psi_n(x, t)|/q(t) \\ & \leq \sup_{\frac{1}{n+1} \leq t \leq \frac{n}{n+1}} \sup_{x \in \mathbb{R}} |\beta_n^*(x, t) - \Gamma_n(F(x; \theta_0), t)|/q(t) \\ & \quad + \sup_{\frac{1}{n+1} \leq t \leq \frac{n}{n+1}} \sup_{x \in \mathbb{R}} \left| \left\{ \varphi_n(t) - \int l(x, \theta_0) d_x \Gamma_n(F(x; \theta_0), t) \right\} \cdot \nabla_{\theta} F(x; \theta_0)^T \right|/q(t) \\ & \quad + \sup_{\frac{1}{n+1} \leq t \leq \frac{n}{n+1}} \sup_{x \in \mathbb{R}} \frac{[(n+1)t]}{n} \left(1 - \frac{[(n+1)t]}{n}\right) n^{1/2} |\varepsilon'_{3n}(x, t) + \varepsilon''_{3n}(x, t)|/q(t) \\ & \quad + \sup_{\frac{1}{n+1} \leq t \leq \frac{n}{n+1}} \sup_{x \in \mathbb{R}} \frac{[(n+1)t]}{n} n^{-1/2} (n - [(n+1)t])^{\frac{1}{2}} |\varepsilon''_{1n}(t) \cdot \nabla_{\theta} F(x; \theta_0)^T|/q(t) \\ & \quad + \sup_{\frac{1}{n+1} \leq t \leq \frac{n}{n+1}} \sup_{x \in \mathbb{R}} \left(1 - \frac{[(n+1)t]}{n}\right) n^{-1/2} [(n+1)t]^{\frac{1}{2}} |\varepsilon'_{1n}(t) \cdot \nabla_{\theta} F(x; \theta_0)^T|/q(t) \\ & := I_n^{(1)} + I_n^{(2)} + I_n^{(3)} + I_n^{(4)} + I_n^{(5)}. \end{aligned}$$

Let $0 < \delta < 1/2$ be fixed. We have

$$\begin{aligned}
& \sup_{\frac{1}{n+1} \leq t \leq \frac{n}{n+1}} \left(1 - \frac{[(n+1)t]}{n}\right) n^{-1/2} [(n+1)t]^{1/2} \|\varepsilon'_{1n}(t)\| / q(t) \\
& \leq \sup_{\frac{1}{n+1} \leq t \leq 1-\delta} \left(1 - \frac{[(n+1)t]}{n}\right) \sup_{\frac{1}{n+1} \leq t \leq 1-\delta} n^{-1/2} [(n+1)t]^{1/2} \|\varepsilon'_{1n}(t)\| / q(t) \\
& \quad \vee \sup_{1-\delta \leq t \leq \frac{n}{n+1}} \left(1 - \frac{[(n+1)t]}{n}\right) / q(t) \sup_{1-\delta \leq t \leq 1} n^{-1/2} [(n+1)t]^{1/2} \|\varepsilon'_{1n}(t)\| \\
& \leq \left(1 - \frac{1}{n}\right) \sup_{\frac{1}{n+1} \leq t \leq 1-\delta} n^{-1/2} [(n+1)t]^{1/2} \|\varepsilon'_{1n}(t)\| / q(t) \\
& \quad \vee \left(1 + \frac{1}{n}\right)^{1/2} \sup_{1-\delta \leq t \leq \frac{n}{n+1}} \left(1 - \frac{[(n+1)t]}{n}\right) / q(t) \sup_{1-\delta \leq t \leq 1} \|\varepsilon'_{1n}(t)\|.
\end{aligned}$$

It is easy to check that, as $n \rightarrow \infty$,

$$\sup_{1-\delta \leq t \leq \frac{n}{n+1}} \left(1 - \frac{[(n+1)t]}{n}\right) / q(t) \leq O(1) \sup_{1-\delta \leq t < 1} (1-t)^{1/2} / q(t).$$

Also, we note that $\varepsilon'_{1n}(t) = \varepsilon_{1(n+1)}(t)$, for all $\frac{1}{n+1} \leq t \leq \frac{n}{n+1}$, where $\varepsilon_{1n}(t)$, is as in (3.2.3). Hence, by (3.2.30) and (3.3.27), we have, as $n \rightarrow \infty$,

(3.3.30)

$$\begin{aligned}
I_n^{(5)} &:= \sup_{\frac{1}{n+1} \leq t \leq \frac{n}{n+1}} \sup_{x \in \mathbb{R}} \left(1 - \frac{[(n+1)t]}{n}\right) n^{-1/2} [(n+1)t]^{1/2} \|\varepsilon'_{1n}(t) \cdot \nabla_{\theta} F(x; \theta_0)^T\| / q(t) \\
&= o_P(1).
\end{aligned}$$

On the other hand, since, by (3.3.13) and (3.3.25), we can write,

$$(3.3.31) \quad \varepsilon''_{1n}(t) = (0, \left(\frac{1}{[(n+1)(1-t)]^{3/4}} \sum_{i=1}^{[(n+1)(1-t)]} (X_{n-i+1} - \mu_0) \right)^2),$$

by symmetry and (3.3.27), arguing similarly as above, we get, as $n \rightarrow \infty$,

**PM-1 "1/2"x4" PHOTOGRAPHIC MICROCOPY TARGET
NBS 1010a ANSI/ISO #2 EQUIVALENT**

(3.3.32)

$$I_n^{(4)} := \sup_{\frac{1}{n+1} \leq t \leq \frac{n}{n+1}} \sup_{x \in \mathbb{R}} \frac{[(n+1)t]}{n} n^{-1/2} (n - [(n+1)t])^{1/2} |\varepsilon_{1n}''(t) \cdot \nabla_{\theta} F(x; \theta_0)^T| / q(t) \\ = o_P(1).$$

Next we have that, for any fixed $0 < \delta < 1/2$,

$$\begin{aligned} & \sup_{\frac{1}{n+1} \leq t \leq \frac{n}{n+1}} \sup_{x \in \mathbb{R}} \frac{[(n+1)t]}{n} \left(1 - \frac{[(n+1)t]}{n}\right) n^{1/2} |\varepsilon'_{3n}(x, t)| / q(t) \\ & \leq \sup_{\frac{1}{n+1} \leq t \leq 1-\delta} \left(1 - \frac{[(n+1)t]}{n}\right) n^{-1/2} [(n+1)t]^{1/2} \\ & \quad \times \sup_{\frac{1}{n+1} \leq t \leq 1-\delta} \sup_{x \in \mathbb{R}} [(n+1)t]^{1/2} |\varepsilon'_{3n}(x, t)| / q(t) \\ & \vee \sup_{1-\delta \leq t \leq \frac{n}{n+1}} \left(1 - \frac{[(n+1)t]}{n}\right) / q(t) \sup_{1-\delta \leq t \leq 1} n^{-1/2} [(n+1)t]^{1/2} \\ & \quad \times \sup_{1-\delta \leq t \leq 1} \sup_{x \in \mathbb{R}} [(n+1)t]^{1/2} |\varepsilon'_{3n}(x, t)| \\ & \leq O(1) \sup_{\frac{1}{n+1} \leq t \leq 1-\delta} \sup_{x \in \mathbb{R}} [(n+1)t]^{1/2} |\varepsilon'_{3n}(x, t)| / q(t) \\ & \vee O(1) \sup_{1-\delta \leq t \leq \frac{n}{n+1}} (1-t)^{1/2} / q(t) \sup_{1-\delta \leq t \leq 1} \sup_{x \in \mathbb{R}} [(n+1)t]^{1/2} |\varepsilon'_{3n}(x, t)|. \end{aligned}$$

Since

$$[(n+1)t]^{1/2} \varepsilon'_{3n}(x, t) = \varepsilon_{3(n+1)}(x, t), \text{ for all } x \in \mathbb{R} \text{ and } \frac{1}{n+1} \leq t \leq \frac{n}{n+1},$$

where $\varepsilon_{3n}(x, t)$ is as in (3.2.22), we have that, by (3.2.42) and (3.3.27), as $n \rightarrow \infty$,

$$(3.3.33) \quad \sup_{\frac{1}{n+1} \leq t \leq \frac{n}{n+1}} \sup_{x \in \mathbb{R}} \frac{[(n+1)t]}{n} \left(1 - \frac{[(n+1)t]}{n}\right) n^{1/2} |\varepsilon'_{3n}(x, t)| / q(t) = o_P(1).$$

Moreover, we note that, by (3.3.24), we can write

(3.3.34)

$$(n - [(n+1)t])^{\frac{1}{2}}(\hat{\theta}_{[nt]}'' - \theta_0) = [(n+1)(1-t)]^{-\frac{1}{2}} \sum_{i=1}^{[(n+1)(1-t)]} l(X_{n+i-1}, \theta_0) + \varepsilon_{1n}''(t),$$

where $\varepsilon_{1n}''(t)$ is as in (3.3.28). Hence, since the X_i 's are i.i.d. random variables, by reasons of symmetry, following the same lines, from (3.2.31) to (3.2.41) in the proof of (3.2.42) (and consequently of (3.3.33)), and by (3.2.27), we also have that, as $n \rightarrow \infty$,

$$(3.3.35) \quad \sup_{\frac{1}{n+1} \leq t \leq \frac{n}{n+1}} \sup_{x \in \mathbb{R}} \frac{[(n+1)t]}{n} \left(1 - \frac{[(n+1)t]}{n}\right) n^{1/2} |\varepsilon_{3n}''(x, t)| / q(t) = o_P(1).$$

Thus, by (3.3.33), (3.3.35) and the triangular inequality (cf. (3.3.29)), we have, as $n \rightarrow \infty$,

$$(3.3.36) \quad I_n^{(3)} = o_P(1).$$

We proceed now to show (cf. 3.3.29) that, as $n \rightarrow \infty$,

$$(3.3.37) \quad I_n^{(2)} = o_P(1).$$

To this end we note that, after adding and subtracting $F(x; \theta_0)$ in (3.3.28), we get, since $E l(X_i, \theta_0) = 0$ and $l(X_i, \theta_0) = \int l(x, \theta_0) d_x \mathbf{1}\{X_i \leq x\}$,

$$(3.3.38) \quad \begin{aligned} \varphi_n(t) = & \left(1 - \frac{[(n+1)t]}{n}\right) \int l(x, \theta_0) d_x n^{-1/2} \sum_{i=1}^{[(n+1)t]} (\mathbf{1}\{X_i \leq x\} - F(x; \theta_0)) \\ & - \frac{[(n+1)t]}{n} \int l(x, \theta_0) d_x n^{-1/2} \sum_{i=[(n+1)t]+1}^n (\mathbf{1}\{X_i \leq x\} - F(x; \theta_0)). \end{aligned}$$

For $\frac{1}{n+1} \leq t \leq 1/2$, by Lemma 3.2.3, there exists a Kiefer process $K^{(1)}(\cdot, \cdot)$ such

that, for any $0 < \delta \leq 1/2$, as $n \rightarrow \infty$,

$$(3.3.39) \quad \sup_{\frac{1}{n+1} \leq t \leq \delta} \left\| \int l(x, \theta_0) d_x \left[n^{-1/2} \sum_{i=1}^{[(n+1)t]} (1\{X_i \leq x\} - F(x; \theta_0)) \right. \right. \\ \left. \left. - n^{-1/2} K^{(1)}(F(x; \theta_0), (n+1)t) \right] \right\| / q(t) = o_P(1).$$

Also, for $1/2 \leq t \leq n/(n+1)$, there exists a Kiefer process $K^{(2)}(\cdot, \cdot)$, independent of $K^{(1)}(\cdot, \cdot)$, such that, as $n \rightarrow \infty$,

$$(3.3.40) \quad \sup_{1-\delta \leq t \leq \frac{n}{n+1}} \left\| \int l(x, \theta_0) d_x \left[n^{-1/2} \sum_{i=1}^{[(n+1)(1-t)]} (1\{X_{n+i-1} \leq x\} - F(x; \theta_0)) \right. \right. \\ \left. \left. - n^{-1/2} K^{(2)}(F(x; \theta_0), (n+1)(1-t)) \right] \right\| / q(t) = o_P(1).$$

To see this we note that, by (3.3.27), we have $\lim_{t \downarrow 0} t^{1/2}/q(1-t) = 0$. Hence, it follows from Lemma 3.2.3 that, as $n \rightarrow \infty$,

$$\sup_{\frac{1}{n+1} \leq t \leq \delta} \left\| \int l(x, \theta_0) d_x \left[n^{-1/2} \left(\sum_{i=1}^{[(n+1)t]} (1\{X_i \leq x\} - F(x; \theta_0)) \right. \right. \right. \\ \left. \left. \left. - n^{-1/2} K^{(2)}(F(x; \theta_0), (n+1)t) \right) \right] \right\| / q(1-t) = o_P(1).$$

For each $n \geq 1$, we define a Kiefer process $\{K(s, nt), 0 \leq t \leq 1, 0 \leq s \leq 1\}$, by

$$(3.3.41) \quad K(s, nt) = \begin{cases} K^{(1)}(s, nt), & 0 \leq t \leq 1/2, 0 \leq s \leq 1, \\ K^{(1)}\left(s, \frac{n}{2}\right) + K^{(2)}\left(s, \frac{n}{2}\right) - K^{(2)}(s, n - nt), & 1/2 < t \leq 1, 0 \leq s \leq 1. \end{cases}$$

After some algebra in (3.3.38) we can show that

$$\begin{aligned}
 \varphi_n(t) &= \int l(x, \theta_0) d_x n^{-1/2} \sum_{i=1}^{[(n+1)t]} (1\{X_i \leq x\} - F(x; \theta_0)) \\
 &\quad - \frac{[(n+1)t]}{n} \int l(x, \theta_0) d_x n^{-1/2} \sum_{i=1}^n (1\{X_i \leq x\} - F(x; \theta_0)) \\
 (3.3.42) \quad &:= I_{1n}(t) - I_{2n}(t).
 \end{aligned}$$

For any $\frac{1}{n+1} \leq t \leq 1/2$, by (3.3.41), we have

$$\begin{aligned}
 I_{1n}(t) &= \int l(x, \theta_0) d_x [n^{-1/2} \sum_{i=1}^{[(n+1)t]} (1\{X_i \leq x\} - F(x; \theta_0)) \\
 &\quad - n^{-1/2} K^{(1)}(F(x; \theta_0), (n+1)t)] \\
 &\quad + \int l(x, \theta_0) d_x n^{-1/2} K^{(1)}(F(x; \theta_0), (n+1)t) \\
 (3.3.43) \quad &:= I_{1n}^{(1)}(t) + \int l(x, \theta_0) d_x n^{-1/2} K(F(x; \theta_0), (n+1)t).
 \end{aligned}$$

On the other hand,

$$\begin{aligned}
 I_{2n}(t) &= \frac{[(n+1)t]}{n} \int l(x, \theta_0) d_x [n^{-1/2} \sum_{i=1}^{[\frac{n+1}{2}]} (1\{X_i \leq x\} - F(x; \theta_0))] \\
 &\quad + \frac{[(n+1)t]}{n} \int l(x, \theta_0) d_x [n^{-1/2} \sum_{i=[\frac{n+1}{2}]+1}^n (1\{X_i \leq x\} - F(x; \theta_0))] \\
 (3.3.44) \quad &:= I_{2n}^{(1)}(t) + I_{2n}^{(2)}(t).
 \end{aligned}$$

Now,

$$\begin{aligned}
 I_{2n}^{(1)}(t) &= \frac{[(n+1)t]}{n} \int l(x, \theta_0) d_x [n^{-1/2} \sum_{i=1}^{[\frac{n+1}{2}]} (1\{X_i \leq x\} - F(x; \theta_0)) \\
 &\quad - n^{-1/2} K^{(1)}(F(x; \theta_0), \frac{n+1}{2})]
 \end{aligned}$$

$$\begin{aligned}
& + \frac{[(n+1)t]}{n} \int l(x, \theta_0) d_x n^{-1/2} K^{(1)}(F(x; \theta_0), \frac{n+1}{2}) \\
(3.3.45) \quad & = \frac{[(n+1)t]}{n} I_{1n}^{(1)}(1/2) + \frac{[(n+1)t]}{n} \int l(x, \theta_0) d_x n^{-1/2} K^{(1)}(F(x; \theta_0), \frac{n+1}{2}).
\end{aligned}$$

Also,

$$\begin{aligned}
I_{2n}^{(2)}(t) & = \frac{[(n+1)t]}{n} \int l(x, \theta_0) d_x [n^{-1/2} \sum_{i=1}^{[\frac{n+1}{2}]} (1\{X_{n-i+1} \leq x\} - F(x; \theta_0)) \\
& \quad - n^{-1/2} K^{(2)}(F(x; \theta_0), \frac{n+1}{2})] \\
& \quad + \frac{[(n+1)t]}{n} \int l(x, \theta_0) d_x n^{-1/2} K^{(2)}(F(x; \theta_0), \frac{n+1}{2}) \\
(3.3.46) \quad & := \frac{[(n+1)t]}{n} I_{2n}^{(2,1)} + \frac{[(n+1)t]}{n} \int l(x, \theta_0) d_x n^{-1/2} K^{(2)}(F(x; \theta_0), \frac{n+1}{2}).
\end{aligned}$$

By (3.3.41), with $t = 1$, we have

$$(3.3.47) \quad K^{(1)}(s, \frac{n+1}{2}) + K^{(2)}(s, \frac{n+1}{2}) = K(s, n+1), \quad 0 \leq s \leq 1.$$

Hence, from (3.3.42), by (3.3.43), (3.3.44), (3.3.45), (3.3.46) and (3.3.47), we have

$$\begin{aligned}
\varphi_n(t) & = \int l(x, \theta_0) d_x n^{-1/2} [K(F(x; \theta_0), (n+1)t) - tK(F(x; \theta_0), n+1)] \\
& \quad + I_{1n}^{(1)}(t) - \frac{[(n+1)t]}{n} I_{1n}^{(1)}(1/2) - \frac{[(n+1)t]}{n} I_{2n}^{(2,1)} \\
(3.3.48) \quad & \quad + (t - \frac{[(n+1)t]}{n}) \int l(x, \theta_0) d_x n^{-1/2} K(F(x; \theta_0), n+1).
\end{aligned}$$

By (3.3.39) with $\delta = 1/2$, we have, as $n \rightarrow \infty$,

$$(3.3.49) \quad \sup_{\frac{1}{n+1} \leq t \leq 1/2} \|I_{1n}^{(1)}(t)\|/q(t) = o_P(1).$$

On the other hand, since $\sup_{\frac{1}{n+1} \leq t \leq 1/2} \frac{[(n+1)t]}{n}/q(t) \leq \sup_{0 < t \leq 1/2} t/q(t) < \infty$, by (3.3.39) and (3.3.40), with $q(t) \equiv 1$ (i.e. no weight function), or Lemma 3.1 of Burke, Csörgő, Csörgő and Révész (1979), we have as $n \rightarrow \infty$,

$$(3.3.50) \quad \sup_{\frac{1}{n+1} \leq t \leq 1/2} \frac{[(n+1)t]}{n}/q(t) \|I_{1n}^{(1)}(1/2) + I_{2n}^{(2,1)}\| = o_P(1).$$

Lastly, since

$$(3.3.51) \quad \sup_{\frac{1}{n+1} \leq t \leq 1/2} \left| t - \frac{[(n+1)t]}{n} \right|/t^{1/2} \leq \left| 1 - \frac{n+1}{n} \right| \frac{1}{\sqrt{2}} + \frac{(n+1)^{1/2}}{n},$$

we have, as $n \rightarrow \infty$,

$$(3.3.52) \quad \begin{aligned} & \sup_{\frac{1}{n+1} \leq t \leq 1/2} \left| t - \frac{[(n+1)t]}{n} \right|/q(t) \left\| \int l(x, \theta_0) d_x n^{-1/2} K(F(x; \theta_0), n+1) \right\| \\ & \leq O_P(1) \sup_{\frac{1}{n+1} \leq t \leq 1/2} \left| t - \frac{[(n+1)t]}{n} \right|/t^{1/2} \sup_{\frac{1}{n+1} \leq t \leq 1/2} t^{1/2}/q(t) \\ & = o_P(1). \end{aligned}$$

Hence, from (3.3.48) we get, by (3.3.6), (3.3.49), (3.3.50) and (3.3.52), that, for any function q positive on $(0, 1)$ and such that (3.3.27) holds, as $n \rightarrow \infty$,

$$(3.3.53) \quad \sup_{\frac{1}{n+1} \leq t \leq 1/2} \sup_{x \in \mathbb{R}} |\varphi_n(t) - \int l(x, \theta_0) d_x n^{-1/2} \Gamma(F(x; \theta_0), nt)|/q(t) = o_P(1).$$

Consider again $\varphi_n(t)$ in (3.3.38). It can also be written as

$$\begin{aligned}
\varphi_n(t) &= \left(1 - \frac{[(n+1)t]}{n}\right) \int l(x, \theta_0) d_x n^{-1/2} \sum_{i=1}^n (1\{X_i \leq x\} - F(x; \theta_0)) \\
&\quad - \int l(x, \theta_0) d_x n^{-1/2} \sum_{i=1}^{[(n+1)(1-t)]} (1\{X_{n+i-1} \leq x\} - F(x; \theta_0)) \\
(3.3.54) \quad &:= J_{1n}(t) - J_{2n}(t).
\end{aligned}$$

For $1/2 \leq t \leq n/n+1$, by (3.3.41) and (3.3.47), we have

$$\begin{aligned}
J_{1n}(t) &= (1-t) \int l(x, \theta_0) d_x \left[n^{-1/2} \sum_{i=1}^n (1\{X_i \leq x\} - F(x; \theta_0)) \right. \\
&\quad \left. - n^{-1/2} K(F(x; \theta_0), n+1) \right] \\
&\quad + \left(t - \frac{[(n+1)t]}{n}\right) \int l(x, \theta_0) d_x n^{-1/2} \sum_{i=1}^n (1\{X_i \leq x\} - F(x; \theta_0)) \\
&\quad + (1-t) \int l(x, \theta_0) n^{-1/2} K(F(x; \theta_0), n+1) \\
(3.3.55) \quad &:= J_{1n}^{(1)}(t) + J_{1n}^{(2)}(t) + (1-t) \int l(x, \theta_0) n^{-1/2} K(F(x; \theta_0), n+1).
\end{aligned}$$

On the other hand,

$$\begin{aligned}
J_{2n}(t) &= \int l(x, \theta_0) d_x \left[n^{-1/2} \sum_{i=1}^{[(n+1)(1-t)]} (1\{X_{n-i+1} \leq x\} - F(x; \theta_0)) \right. \\
&\quad \left. - n^{-1/2} K^{(2)}(F(x; \theta_0), (n+1)(1-t)) \right] \\
&\quad + \int l(x, \theta_0) d_x n^{-1/2} K^{(2)}(F(x; \theta_0), (n+1)(1-t)) \\
(3.3.56) \quad &:= J_{2n}^{(1)}(t) + \int l(x, \theta_0) d_x n^{-1/2} K^{(2)}(F(x; \theta_0), (n+1)(1-t)).
\end{aligned}$$

Hence, from (3.3.54), by (3.3.41), (3.3.47), (3.3.55) and (3.3.56), we have

$$\begin{aligned}
 \varphi_n(t) &= \int l(x, \theta_0) d_x n^{-1/2} [K(F(x; \theta_0), (n+1)t) - tK(F(x; \theta_0), n+1)] \\
 (3.3.57) \quad &+ J_{1n}^{(1)}(t) + J_{1n}^{(2)}(t) + J_{2n}^{(1)}(t).
 \end{aligned}$$

Since, by (3.3.27), $\sup_{1/2 \leq t \leq \frac{n}{n+1}} (1-t)/q(t) \leq \sup_{0 < t \leq 1/2} t/q(1-t) < \infty$, we have, by Lemma 3.2.3, with $q(t) \equiv 1$ (i.e. no weight function), as $n \rightarrow \infty$,

$$(3.3.58) \quad \sup_{1/2 \leq t \leq \frac{n}{n+1}} \|J_{1n}^{(1)}(t)\|/q(t) = o_P(1).$$

By (3.3.40) we have, as $n \rightarrow \infty$,

$$(3.3.59) \quad \sup_{1/2 \leq t \leq \frac{n}{n+1}} \|J_{2n}^{(1)}(t)\|/q(t) = o_P(1).$$

Also, since

$$(3.3.60) \quad \sup_{1/2 \leq t \leq \frac{n}{n+1}} \left| t - \frac{[(n+1)t]}{n} \right| / (1-t)^{1/2} \leq \frac{1}{n} \frac{n}{n+1} (n+1)^{1/2} + \frac{1}{n} (n+1)^{1/2},$$

we have, as $n \rightarrow \infty$,

$$\begin{aligned}
 \sup_{1/2 \leq t \leq \frac{n}{n+1}} \|J_{1n}^{(2)}(t)\|/q(t) &\leq O_P(1) \sup_{1/2 \leq t \leq \frac{n}{n+1}} \left| t - \frac{[(n+1)t]}{n} \right| / (1-t)^{1/2} \\
 (3.3.61) \quad &\times \sup_{1/2 \leq t \leq \frac{n}{n+1}} (1-t)^{1/2}/q(t) = o_P(1).
 \end{aligned}$$

Hence, from (3.3.57) we get, by (3.3.6), (3.3.58), (3.3.59) and (3.3.61), that, for any function q positive on $(0, 1)$ and such that (3.3.27) holds, as $n \rightarrow \infty$,

$$(3.3.62) \quad \sup_{\frac{1}{2} \leq t \leq \frac{n}{n+1}} \sup_{x \in \mathbb{R}} |\varphi_n(t) - \int l(x, \theta_0) d_x \Gamma_n(F(x; \theta_0), t)|/q(t) = o_P(1).$$

This, together with (3.3.53), implies (3.3.37), since $\sup_{x \in \mathbb{R}} |\nabla_{\theta} F(x; \theta_0)| < \infty$.

Lastly, in the light of (3.3.9) of Remark 3.3.2, we have from (3.3.29) that, by (3.3.30), (3.3.32), (3.3.35) and (3.3.37), as $n \rightarrow \infty$,

$$(3.3.63) \quad \sup_{\frac{1}{n+1} \leq t \leq \frac{n}{n+1}} \sup_{x \in \mathbb{R}} |\hat{\beta}_n(x, t) - \Psi_n(x, t)|/q(t) = o_P(1),$$

provided that the function q is positive on $(0, 1]$ and such that (3.3.27) holds.

Since, by definition, $\hat{\beta}_n(x, t) = 0$ for any $t \in A_n := (0, \frac{1}{n+1}] \cup [\frac{n}{n+1}, 1)$, in virtue of (3.3.63), in order to prove (a) and (b), it now suffices to show, that, as $n \rightarrow \infty$,

$$(3.3.64) \quad \sup_{t \in A_n} \sup_{x \in \mathbb{R}} |\Psi_n(x, t)|/q(t) = \begin{cases} o_P(1), & \text{if (3.3.15) holds,} \\ O_P(1), & \text{if (3.3.16) holds.} \end{cases}$$

To this end we note that, by (3.2.46), (3.2.47), (3.2.48) and (3.3.14),

$$(3.3.65)$$

$$\Psi_n(x, t) = \Gamma_n(F(x; \theta_0), t) - n^{-1/2}(W(nt) - tW(n)) \cdot D(\theta_0)^{-1} \cdot \nabla_{\theta} \Gamma(x; \theta_0)^T,$$

where, for each $n \geq 1$, $n^{-1/2}W(nt) = \int l(x, \theta_0) d_x K(F(x; \theta_0), nt) \cdot D(\theta_0)$ is the vector valued Wiener process as in (3.2.50).

Next, we have that, by (3.3.6),

$$\sup_{0 < t < \frac{1}{n}} \sup_{x \in \mathbb{R}} |\Gamma_n(F(x; \theta_0), t)| \stackrel{P}{=} \sup_{0 < t < \frac{1}{n}} \sup_{x \in \mathbb{R}} |\Gamma(F(x; \theta_0), t)| \text{ for each } n \geq 1.$$

Hence, by Theorem 3.3.B (cf. Remark 3.3.3), we have, as $n \rightarrow \infty$,

$$(3.3.66) \quad \sup_{0 < t < \frac{1}{n}} \sup_{x \in \mathbb{R}} |\Gamma(F(x; \theta_0), t)|/q(t) = \begin{cases} o_P(1), & \text{if (3.3.15) holds,} \\ O_P(1), & \text{if (3.3.16) holds.} \end{cases}$$

Similarly, since $n^{-1/2}(W(nt) - tW(n)) \stackrel{D}{=} W(t) - tW(1) =: B(t)$, for each n , where $B(\cdot)$ is a Brownian bridge (cf. Remark 3.3.5), we have,

$$\sup_{0 < t < \frac{1}{n}} n^{-1/2} |W(nt) - tW(n)| \stackrel{D}{=} \sup_{0 < t < \frac{1}{n}} |B(t)|.$$

Thus, by Theorem 3.3.C (cf. Remark 3.3.4), we have, as $n \rightarrow \infty$,

$$(3.3.67) \quad \sup_{0 < t < \frac{1}{n}} |n^{-1/2}W(nt) - tW(1)|/q(t) = \begin{cases} o_P(1), & \text{if (3.3.15) holds,} \\ O_P(1), & \text{if (3.3.16) holds.} \end{cases}$$

Hence, by (3.3.66) and (3.3.67), we get that (3.3.64) holds, and (a) and (b) follow.

As for the proof of (c) we note that, by (3.3.6), (3.3.14) and (3.3.17), we have, for any $n \geq 1$,

$$(3.3.68) \quad \{\Psi_n(x, t), 0 \leq t \leq 1, x \in \mathbb{R}\} \stackrel{D}{=} \{\Psi(x, t), 0 \leq t \leq 1, x \in \mathbb{R}\}.$$

Next, we note that, (3.3.63),

$$(3.3.69) \quad \lim_{n \rightarrow \infty} \left| P \left\{ \sup_{0 < t < 1} \sup_{x \in \mathbb{R}} |\hat{\beta}_n(x, t)|/q(t) \leq x \right\} - P \left\{ \sup_{\frac{1}{n} \leq t < 1} \sup_{x \in \mathbb{R}} |\Psi_n(x, t)|/q(t) \leq x \right\} \right| = 0,$$

Also, with any function $q \in \mathcal{Q}$ and such that (3.3.16) holds, we have

$$\lim_{n \rightarrow \infty} P \left\{ \sup_{\frac{1}{n} \leq t < 1} \sup_{x \in \mathbb{R}} |\Psi(x, t)|/q(t) \leq x \right\} = P \left\{ \sup_{0 < t < 1} \sup_{x \in \mathbb{R}} |\Psi(x, t)|/q(t) \leq x \right\},$$

for any $x \in \mathbb{R}$. This, together with (3.3.68) and (3.3.69), implies (c). ■

Corollary 3.3.1. *Under the conditions of Theorem 3.3.1 we have, as $n \rightarrow \infty$,*

(a) *if $q \in \mathcal{Q}^*$ and is such that (3.3.15) holds then,*

$$(3.3.70) \quad \hat{\beta}_n(x, t)/q(t) \xrightarrow{\mathcal{D}} \Psi(x, t)/q(t)$$

in $D(\mathbb{R} \times [0, 1])$,

(b) *if q is a positive function on $(0, 1)$, then, (3.3.70) holds in $D(\mathbb{R} \times [\delta, 1 - \delta])$ for any $0 < \delta < 1/2$.*

We should note that on assuming condition (3.3.15), part (a) of Corollary 3.3.1 gives the limiting distribution of any continuous functional of $\hat{\beta}_n(x, t)/q(t)$. In addition to that, Theorem 3.3.1 (c) gives convergence in distribution of the sup – sup functional in $D(\mathbb{R} \times (0, 1])$ of $\hat{\beta}_n(x, t)/q(t)$, with the weaker condition (3.3.16). In particular, Theorem 3.3.1 (c) is true for the function

$$(3.3.71) \quad q(t) = \left(t(1-t) \log \log 1/t(1-t) \right)^{1/2}, \quad 0 < t < 1.$$

On account of this, we propose the following:

Suppose that $X_1, X_2, \dots, X_n$ are independent normal random variables. For testing

$$(3.3.72) \quad H_0 : X_1, X_2, \dots, X_n \text{ have the same normal distribution}$$

against

$$(3.3.73) \quad H_1 : \text{There is a } \lambda \in (0, 1) \text{ such that}$$

$$P(X_1 \leq t) = \dots = P(X_{[n\lambda]} \leq t),$$

$$P(X_{[n\lambda]+1} \leq t) = \dots = P(X_n \leq t), \quad -\infty < t < \infty, \text{ and}$$

$$P(X_{[n\lambda]} \leq t_0) \neq P(X_{[n\lambda]+1} \leq t_0), \text{ for some } t_0,$$

reject H_0 , in favour of H_1 , for large values of

$$(3.3.74) \quad \sup_{0 < t < 1} \sup_{x \in \mathbb{R}} |\hat{\beta}_n(x, t)| / \left(t(1-t) \log \log 1/(t(1-t)) \right)^{1/2},$$

where $\hat{\beta}_n(\cdot, \cdot)$ is as in (3.3.11).

3.4 Sequentially estimated empirical processes.

Building on the experience gained in Sections 3.2 and 3.3, in this section we study the sequentially estimated empirical process $\hat{\beta}_n(x, t)$ as in (3.1.1) as well as the process $\hat{\hat{\beta}}_n(x, t)$ as in (3.3.8), in the case of a general parametric family.

Suppose that $X_1, X_2, \dots$ is a sequence of i.i.d. random variables with continuous distribution function $F \in \{F(x; \theta); x \in \mathbb{R}, \theta \in \Theta \subseteq \mathbb{R}^p\}$. We assume that, for each $\frac{1}{n} \leq t \leq 1$, the unknown θ is estimated by a sequence $\{\hat{\theta}_{[nt]}\}$ of estimators based on the random sample $X_1, \dots, X_{[nt]}$, under conditions given by (1.2.2) (i)-(vi) in Chapter 1, where (1.2.2)(i) is satisfied for each fixed t , i.e.,

$$(3.4.1) \quad [nt]^{\frac{1}{2}}(\hat{\theta}_{[nt]} - \theta_0) = [nt]^{-\frac{1}{2}} \sum_{i=1}^{[nt]} l(X_i, \theta_0) + \varepsilon_{1n}(t),$$

where θ_0 denotes the true value of the parameter θ , and $\varepsilon_{1n}(t)$ converges to zero, in probability, as $(nt) \rightarrow \infty$.

We will assume further in (3.4.1), that, as $n \rightarrow \infty$,

$$(3.4.1^*) \quad \sup_{\delta \leq t \leq 1} \|\varepsilon_{1n}(t)\| = o_P(1) \text{ and } \sup_{\frac{1}{n} \leq t \leq \delta} \|\varepsilon_{1n}(t)\| = O_P(1),$$

for any $0 < \delta < 1$.

Let $T_j(x)$ denote the total variation of the j th component $l_j(\cdot, \theta_0)$ of $l(\cdot, \theta_0)$ (cf. (1.2.2)(v)) on the interval $[-x, x]$, $j = 1, \dots, p$, and let $T(x) = (T_1(x), \dots, T_p(x))$. We can choose a non-negative function $u(n, t) \equiv u(nt)$, $n \geq 1$, $0 \leq t \leq 1$, such that, for each n fixed, $u(n, \cdot)$ is increasing, for each t fixed, $u(\cdot, t)$ is increasing, and $u(n, t) \rightarrow \infty$, as $(nt) \rightarrow \infty$ so slowly that $\|T(u(nt))\|(nt)^{-1/2} \log^2(nt) \rightarrow 0$. (If $\|T(n)\|$ is bounded, then any constant function in t , $u(n, t) \equiv u_n \rightarrow \infty$ will suffice, while if $\|T(n)\| \nearrow \infty$ we can take $u(nt) = T^{-1}(v(nt))$, where $v(nt) \nearrow \infty$, as $(nt) \rightarrow \infty$ so that $v(nt) = o((nt)^{1/2} / \log^2(nt))$ and $T^{-1}(y) = \inf\{x : \|T(x)\| \geq y\}$).

Consider the series

$$(3.4.2) \quad \sum_{k=1}^{\infty} \frac{1}{k} \int_{|x| > u(k)} l_j^2(x, \theta_0) d_x F(x; \theta_0), \quad j = 1, \dots, p,$$

where $F(x; \theta_0)$ denotes the underlying distribution function with $\theta = \theta_0$.

We recall that the two time parameter sequentially estimated empirical process, based on the observations $X_1, \dots, X_{[nt]}$, is defined by

$$(3.4.3) \quad \hat{\beta}_n(x, t) = n^{-1/2} \sum_{i=1}^{[nt]} \left(\mathbf{1}\{X_i \leq x\} - F(x; \hat{\theta}_{[nt]}) \right), \quad 0 \leq t \leq 1, \quad x \in \mathbb{R}.$$

We have

Theorem 3.4.1. *Let $X_1, X_2, \dots$, be a sequence of i.i.d. r.v.'s with continuous distribution function $F \in \{F(x; \theta); x \in \mathbb{R}, \theta \in \Theta \subseteq \mathbb{R}^p\}$. Suppose that, $\{\hat{\theta}_{[nt]}\}$ satisfies conditions (1.2.2)(i)-(vi), with (3.4.1) and (3.4.1*). Suppose further that the series in (3.4.2) converges for each $j = 1, \dots, p$. Then, there exists a Kiefer process $K(\cdot, \cdot)$ such that with the sequence of stochastic processes $\{G_n(x, t), 0 \leq t \leq 1, x \in \mathbb{R}\}$, and a weight function $q \in \mathcal{Q}$ we have, as $n \rightarrow \infty$,*

(a) if

$$(3.4.4) \quad I_0(q, c) < \infty \quad \text{for all } c > 0,$$

then

$$\sup_{0 < t \leq 1} \sup_{x \in \mathbb{R}} |\hat{\beta}_n(x, t) - G_n(x, t)|/q(t) = o_P(1),$$

(b) if

$$(3.4.5) \quad I_0(q, c) < \infty \quad \text{for some } c > 0,$$

then

$$\sup_{0 < t \leq 1} \sup_{x \in \mathbb{R}} |\hat{\beta}_n(x, t) - G_n(x, t)|/q(t) = O_P(1),$$

(c) if (3.4.5) holds, then

$$\sup_{0 < t \leq 1} \sup_{x \in \mathbb{R}} |\hat{\beta}_n(x, t)|/q(t) \xrightarrow{P} \sup_{0 < t \leq 1} \sup_{x \in \mathbb{R}} |G(x, t)|/q(t),$$

where

(3.4.6)

$$G_n(x, t) = n^{-1/2} K(F(x; \theta_0), nt) - \left\{ \int l(x, \theta_0) d_x n^{-1/2} K(F(x; \theta_0), nt) \right\} \cdot \nabla_{\theta} F(x; \theta_0)^T$$

and

$$(3.4.7) \quad G(x, t) = K(F(x; \theta_0), t) - \left\{ \int l(x, \theta_0) d_x K(F(x; \theta_0), t) \right\} \cdot \nabla_{\theta} F(x; \theta_0)^T.$$

Remark 3.4.1. We note that, for each n , G_n is a Gaussian process with mean zero and covariance function $EC_n(x, s)G_n(y, t) = (s \wedge t) \times \{ \}$, where $\{ \}$ equals the right-hand side of (1.2.4), with $n = m$.

Recall the following

$$(3.4.8) \quad \varepsilon_{2n}(F(x; \theta_0), t) = n^{-1/2} \sum_{i=1}^{[nt]} (1\{X_i \leq x\} - F(x; \theta_0)) - n^{-1/2} K(F(x; \theta_0), nt)$$

and

$$(3.4.9) \quad L_n(t) = \int l(x, \theta_0) d_x \varepsilon_{2n}(F(x; \theta_0), t).$$

For the proof of Theorem 3.4.1 we need the following

Lemma 3.4.1. *If the series $\sum_{k=1}^{\infty} \frac{1}{k} \int_{|z| > u(k)} l_j^2(x, \theta_0) d_x F(x; \theta_0)$ converges for each $j = 1, \dots, p$, then, for any q which is a positive function on $(0, 1]$, and such that $\lim_{t \downarrow 0} t^{1/2}/q(t) = 0$, as $n \rightarrow \infty$,*

$$(3.4.10) \quad \sup_{\frac{1}{n} \leq t \leq 1} \|L_n(t)\|/q(t) \rightarrow 0 \text{ in probability.}$$

Proof. Consider

$$(3.4.11) \quad \begin{aligned} L_n(t) = & \int_{|z| \leq u(nt)} l(x, \theta_0) d_x \left[n^{-1/2} \sum_{i=1}^{[nt]} (1\{X_i \leq x\} - F(x; \theta_0)) - \right. \\ & \left. n^{-1/2} K(F(x; \theta_0), nt) \right] \\ & + \int_{|z| > u(nt)} l(x, \theta_0) d_x n^{-1/2} \sum_{i=1}^{[nt]} (1\{X_i \leq x\} - F(x; \theta_0)) \\ & - \int_{|z| > u(nt)} l(x, \theta_0) d_x n^{-1/2} K(F(x; \theta_0), nt) \end{aligned}$$

$$:= L_{1n}(t) + L_{2n}(t) - L_{3n}(t).$$

Let $0 < \delta < 1$ be fixed. Using integration by parts, we have

$$\begin{aligned} \sup_{\delta \leq t \leq 1} \|L_{1n}(t)\|/q(t) &\leq \left\{ \sup_{\delta \leq t \leq 1} \|l(x, \theta_0) \varepsilon_{2n}(F(x; \theta_0), t) \Big|_{x=-u_n(t)}^{x=u_n(t)}\| \right. \\ &\quad \left. + \sup_{\delta \leq t \leq 1} \left\| \int_{-u_n(t)}^{u_n(t)} \varepsilon_{2n}(F(x; \theta_0), t) d_x l(x, \theta_0) \right\| \right\} \times \left(1 / \inf_{\delta \leq t \leq 1} q(t) \right). \end{aligned}$$

By Theorem 1.1.B., as $(nt) \rightarrow \infty$,

$$\sup_{x \in \mathbb{R}} |\varepsilon_{2n}(F(x; \theta_0), t)| \stackrel{a.s.}{=} O(n^{-1/2} \log^2(nt))$$

and hence,

$$\sup_{\delta \leq t \leq 1} \|L_{1n}(t)\| \stackrel{a.s.}{=} \sup_{\delta \leq t \leq 1} \|T(u(nt))\| n^{-1/2} \log^2(nt).$$

Since q is positive on $(0, 1]$, $\inf_{\delta \leq t \leq 1} q(t) > 0$, and thus $\sup_{\delta \leq t \leq 1} \|L_{1n}(t)\|/q(t) = o_P(1)$, as $n \rightarrow \infty$.

Also,

$$\begin{aligned} \sup_{\frac{1}{n} \leq t \leq \delta} \|L_{1n}(t)\|/q(t) &\leq \sup_{0 < t \leq \delta} t^{1/2}/q(t) \sup_{\frac{1}{n} \leq t \leq \delta} \|L_{1n}(t)\|/t^{1/2} \\ &\stackrel{a.s.}{=} \sup_{0 < t \leq \delta} t^{1/2}/q(t) \sup_{\frac{1}{n} \leq t \leq 1} \|T(u(nt))\| (nt)^{-1/2} \log^2(nt) \\ &\stackrel{a.s.}{=} \sup_{0 < t \leq \delta} t^{1/2}/q(t) O(1). \end{aligned}$$

Thus, since q is such $\lim_{t \downarrow 0} t^{1/2}/q(t) = 0$, we have that taking $\delta > 0$ sufficiently small, we get, as $n \rightarrow \infty$,

$$\sup_{\frac{1}{n} \leq t \leq \delta} \|L_{1n}(t)\|/q(t) = o_P(1).$$

On the other hand, since $E\left(\int_{|x|>u(k)} l_j(x, \theta_0) d_x (1\{X_i \leq x\} - F(x; \theta_0))\right) = 0$ and

$$E\left(\int_{|x|>u(k)} l_j(x, \theta_0) d_x (1\{X_i \leq x\} - F(x; \theta_0))\right)^2 \leq \int_{|x|>u(k)} l_j^2(x, \theta_0) d_x F(x; \theta_0),$$

for each $j = 1, \dots, p$, following the same steps used to show (3.2.19) and (3.2.20),

for any $0 < \delta < 1$ fixed,

$$\begin{aligned} & P\left\{\sup_{\frac{1}{n} \leq t \leq \delta} \|L_{2n}(t)\|/q(t) > \epsilon\right\} \\ (3.4.12) \quad & \leq \left(\sup_{\frac{1}{n} \leq t \leq \delta} t^{1/2}/q(t)\right)^2 \frac{1}{\epsilon^2} \sum_{j=1}^p \sum_{k=1}^n \frac{1}{k} \int_{|x|>u(k)} l_j^2(x, \theta_0) d_x F(x; \theta_0). \end{aligned}$$

Also,

$$\begin{aligned} & P\left\{\sup_{\delta \leq t \leq 1} \|L_{2n}(t)\|/q(t) > \epsilon\right\} \\ (3.4.13) \quad & \leq \left(\inf_{\delta \leq t \leq 1} q(t)\right)^{-2} \frac{1}{\epsilon^2} \frac{1}{n} \sum_{j=1}^p \sum_{k=[n\delta]}^n \frac{1}{k} \int_{|x|>u(k)} l_j^2(x, \theta_0) d_x F(x; \theta_0). \end{aligned}$$

Now, since the series $\sum_{k=1}^{\infty} \frac{1}{k} \int_{|x|>u(k)} l_j^2(x, \theta_0) d_x F(x; \theta_0)$ is convergent, for any $j = 1, \dots, p$, it follows by (3.2.5) and the fact that $\lim_{t \downarrow 0} t^{1/2}/q(t) = 0$, that

$$\sup_{\frac{1}{n} \leq t \leq 1} \|L_{2n}(t)\|/q(t) = o_P(1), \text{ as } n \rightarrow \infty.$$

Since for each $n \geq 1$ and $\frac{1}{n} \leq t \leq 1$,

$$n^{-1/2}K(u, nt) \stackrel{D}{=} \left(\frac{t}{[nt]}\right)^{1/2} \sum_{i=1}^{[nt]} B_i(u),$$

where $\{B_i\}_{i=1}^{[nt]}$ is a family of independent Brownian bridges, it can be easily shown that, for each $j = 1, 2$,

$$E\left(\int_{|x|>\log k} l_j(x, \theta_0) d_x n^{-1/2}K(F(x; \theta_0), nt)\right) = 0,$$

$$E\left(\int_{|x|>\log k} l_j(x, \theta_0) d_x n^{-1/2}K(F(x; \theta_0), nt)\right)^2 \leq \int_{|x|>\log k} l_j^2(x, \theta_0) d_x F(x; \theta_0),$$

and that we have the same inequalities (3.4.12) and (3.4.13) for $L_{3n}(t)$. Thus,

$$\sup_{\frac{1}{n} \leq t \leq 1} \|L_{3n}(t)\|/q(t) = o_P(1), \text{ as } n \rightarrow \infty. \quad \blacksquare$$

Proof of Theorem 3.4.1. By (3.4.3) and a one-term Taylor expansion around θ_0 , following the same steps as in (3.2.21) and (3.2.24), we get

$$\begin{aligned} \hat{\beta}_n(x, t) &= n^{-1/2} \sum_{i=1}^{[nt]} (1\{X_i \leq x\} - F(x; \theta_0)) - n^{-1/2}K(F(x; \theta_0), nt) \\ &\quad + G_n(x, t) + n^{-1/2}[nt]^{1/2}\epsilon_{3n}(x, t) - L_n(t) \cdot \nabla_{\theta} F(x; \theta_0)^T \\ (3.4.14) \quad &\quad + n^{-1/2}[nt]^{1/2}\epsilon_{1n}(t) \cdot \nabla_{\theta} F(x; \theta_0)^T, \end{aligned}$$

where $G_n(x, t)$ is as in (3.4.9), $L_n(t)$ is as in Lemma 3.4.1,

$$(3.4.15) \quad \epsilon_{3n}(x, t) = [nt]^{1/2}(\hat{\theta}_{[nt]} - \theta_0) \cdot (\nabla_{\theta} F(x; \theta_0)^T - \nabla_{\theta} F(x; \hat{\theta}_{[nt]})^T),$$

and

$$(3.4.16) \quad \sup_{0 \leq t \leq 1} \|\hat{\theta}_{[nt]} - \theta_0\| \leq \sup_{0 \leq t \leq 1} \|\hat{\theta}_{[nt]} - \theta_0\|.$$

We first assume that q is a positive function on $(0, 1]$, and such that $\lim_{t \downarrow 0} \frac{t^{1/2}}{q(t)} = 0$. Let $0 < \delta < 1$ be fixed. Then, for each $n \geq 1$,

$$\sup_{\delta \leq t \leq 1} n^{-1/2} [nt]^{1/2} \|\varepsilon_{1n}(t)\| / q(t) \leq \sup_{\delta \leq t \leq 1} t^{1/2} / q(t) \sup_{\delta \leq t \leq 1} \|\varepsilon_{1n}(t)\|.$$

Since $\sup_{\delta \leq t \leq 1} t^{1/2} / q(t) \leq \left(\inf_{\delta \leq t \leq 1} q(t) \right)^{-1}$, we have, by (3.2.5) and (3.4.1*), that

$$(3.4.17) \quad \sup_{\delta \leq t \leq 1} n^{-1/2} [nt]^{1/2} \|\varepsilon_{1n}(t)\| / q(t) = o_P(1), \text{ as } n \rightarrow \infty.$$

Also, since $[nt] \leq nt$ and $\lim_{t \downarrow 0} t^{1/2} / q(t) = 0$, by (3.4.1*), taking $\delta > 0$ sufficiently small, we have

$$(3.4.18) \quad \sup_{\frac{1}{n} \leq t \leq \delta} \sup_{x \in \mathbb{R}} n^{-1/2} [nt]^{1/2} \|\varepsilon_{1n}(t)\| / q(t) = o_P(1), \text{ as } n \rightarrow \infty.$$

Furthermore, by (1.2.1)(vi), $\nabla_{\theta} F(x; \theta_0)$ is uniformly bounded in x , thus, we have, from (3.4.17) and (3.4.18), that

$$(3.4.19) \quad \sup_{\frac{1}{n} \leq t \leq 1} \sup_{x \in \mathbb{R}} n^{-1/2} [nt]^{1/2} |\varepsilon_{1n}(t) \cdot \nabla_{\theta} F(x; \theta_0)^T| / q(t) = o_P(1), \text{ as } n \rightarrow \infty.$$

We note that, for $\frac{1}{n} \leq t \leq 1$, by (3.4.1),

$$(3.4.20) \quad \hat{\theta}_{[nt]} - \theta_0 = [nt]^{-1} \sum_{i=1}^{[nt]} l(X_i, \theta_0) + [nt]^{-1/2} \varepsilon_{1n}(t).$$

Let $0 < \delta < 1$ be fixed. Since, by (1.2.2)(ii), $El(X_i, \theta_0) = 0$, an application of the Hájek-Rényi inequality and (3.4.1*) show that

$$(3.4.21) \quad \sup_{\delta \leq t \leq 1} \|\hat{\theta}_{[nt]} - \theta_0\| = o_P(1), \text{ as } n \rightarrow \infty.$$

Since by (1.2.2)(iv), $\nabla_{\theta} F(x; \theta_0)$ is uniformly continuous in x and $\theta \in \Lambda$, where Λ is the closure of a neighbourhood of θ_0 , we get, by (3.4.16), that

$$(3.4.22) \quad \sup_{\delta \leq t \leq 1} \|\nabla_{\theta} F(x; \theta_0)^{\top} - \nabla_{\theta} F(x; \tilde{\theta}_{[nt]})^{\top}\| = o_P(1), \text{ as } n \rightarrow \infty.$$

Again from (3.4.1), by (3.4.1*) and the Central Limit Theorem, we have

$$\begin{aligned} \sup_{\delta \leq t \leq 1} \|[nt]^{1/2}(\hat{\theta}_{[nt]} - \theta_0)\| &\leq \sup_{\delta \leq t \leq 1} \|[nt]^{-\frac{1}{2}} \sum_{i=1}^{[nt]} l(X_i, \theta_0)\| + \sup_{\delta \leq t \leq 1} \|\varepsilon_{1n}(t)\| \\ (3.4.23) \quad &\leq [n\delta]^{-1/2} \sum_{i=1}^n \|l(X_i, \theta_0)\| + \sup_{\delta \leq t \leq 1} \|\varepsilon_{1n}(t)\| \\ &= O_P(1) + o_P(1) = O_P(1), \text{ as } n \rightarrow \infty. \end{aligned}$$

Thus, from (3.4.15), we have, by (3.2.5), (3.4.22) and (3.4.23),

$$(3.4.24) \quad \sup_{\delta \leq t \leq 1} \sup_{x \in \mathbb{R}} n^{-1/2} [nt] |\varepsilon_{3n}(x, t)| / q(t) = o_P(1), \text{ as } n \rightarrow \infty.$$

On the other hand, after a one-term Taylor expansion around θ_0 , we have

$$\begin{aligned} &[nt]^{1/2} (\hat{\theta}_{[nt]} - \theta_0) \cdot (\nabla_{\theta} F(x; \theta_0)^{\top} - \nabla_{\theta} F(x; \tilde{\theta}_{[nt]})^{\top}) \\ (3.4.25) \quad &= [nt]^{1/2} (\hat{\theta}_{[nt]} - \theta_0) F''_{\theta\theta}(x; \tilde{\theta}_{[nt]}) (\tilde{\theta}_{[nt]} - \theta_0)^{\top}, \end{aligned}$$

where $F''_{\theta\theta}(x; \alpha)$ is the $p \times p$ matrix of second partial derivatives of $F(x; \theta)$ with respect to $\theta = (\theta_1, \dots, \theta_p)$, evaluated at $\alpha = (\alpha_1, \dots, \alpha_p)$ (cf. (iv), page 1), and

$$(3.4.26) \quad \sup_{0 \leq t \leq 1} \|\bar{\theta}_{[nt]} - \theta_0\| \leq \sup_{0 \leq t \leq 1} \|\hat{\theta}_{[nt]} - \theta_0\|.$$

Furthermore, by (3.4.16),

$$(3.4.27) \quad [nt]^{1/2} \left| (\hat{\theta}_{[nt]} - \theta_0) F''_{\theta\theta}(x; \bar{\theta}_{[nt]}) (\hat{\theta}_{[nt]} - \theta_0)^T \right| \\ \leq 4 \| [nt]^{1/4} (\hat{\theta}_{[nt]} - \theta_0) \|^2 \| F''_{\theta\theta}(x; \bar{\theta}_{[nt]}) \|,$$

where for any $p \times p$ matrix $A = [a_{ij}]$, $\|A\| = \max_{1 \leq i, j \leq p} |a_{ij}|$.

Now, by (3.4.20),

$$[nt]^{1/4} (\hat{\theta}_{[nt]} - \theta_0) = [nt]^{-3/4} \sum_{i=1}^{[nt]} l(X_i, \theta_0) + [nt]^{-1/4} \varepsilon_{1n}(t),$$

and, hence, by the Hájek-Rényi inequality and (3.4.1*),

$$\sup_{\frac{1}{n} \leq t \leq \delta} [nt]^{1/4} \|\hat{\theta}_{[nt]} - \theta_0\| = O_P(1), \text{ as } n \rightarrow \infty.$$

Since, by (1.2.2)(vi), each component of the matrix $F''_{\theta\theta}(x; \theta)$ is uniformly bounded in x and $\theta \in \Lambda$, where Λ is a neighbourhood of θ , by (3.4.16) and (3.4.26), we have, as $n \rightarrow \infty$,

$$(3.4.28) \quad \sup_{\frac{1}{n} \leq t \leq \delta} \sup_{x \in \mathbb{R}} \|F''_{\theta\theta}(x; \bar{\theta}_{[nt]})\| = O_P(1).$$

Finally, since $[nt] \leq nt$, $\sup_{\frac{1}{n} \leq t \leq \delta} t^{1/2}/q(t) \leq \sup_{0 < t \leq \delta} t^{1/2}/q(t)$ and $\lim_{t \downarrow 0} t^{1/2}/q(t) = 0$, we have

$$(3.4.29) \quad \sup_{\frac{1}{n} \leq t \leq 1} \sup_{x \in \mathbb{R}} n^{-1/2} [nt]^{1/2} |\varepsilon_{3n}(x, t)| / q(t) = o_P(1), \text{ as } n \rightarrow \infty.$$

Now, by (3.4.24) and (3.4.29), we have, as $n \rightarrow \infty$,

$$(3.4.30) \quad \sup_{\frac{1}{n} \leq t \leq 1} \sup_{x \in \mathbb{R}} n^{-1/2} [nt]^{1/2} |\varepsilon_{3n}(x, t)| / q(t) = o_P(1).$$

We note that, from (3.4.14), by the proof of Theorem 2.1.A (cf. 2.1.8), Lemma 3.2.3, (3.4.19) and (3.4.30), we have, for any q positive on $(0, 1]$ and such that $\lim_{t \downarrow 0} t^{1/2}/q(t) = 0$, as $n \rightarrow \infty$,

$$(3.4.31) \quad \sup_{0 < t \leq 1} \sup_{x \in \mathbb{R}} |\hat{\beta}_n(x, t) - \bar{G}_n(x, t)| / q(t) \stackrel{a.s.}{=} o(1),$$

where

$$\bar{G}_n(x, t) = \begin{cases} G_n(x, t), & x \in \mathbb{R}, \frac{1}{n} \leq t \leq 1, \\ 0, & \text{elsewhere,} \end{cases}$$

and $G_n(x, t)$ is as in (3.4.6).

By definition,

$$(3.4.32) \quad \hat{\beta}_n(x, t) = 0 \text{ for all } x \in \mathbb{R} \text{ and } 0 < t < \frac{1}{n}.$$

Hence, to prove (a) and (b), it suffices to show that

$$(3.4.33) \quad \sup_{0 < t < \frac{1}{n}} \sup_{x \in \mathbb{R}} |G_n(x, t)| / q(t) = \begin{cases} o_P(1), & \text{if (3.4.4) holds,} \\ O_P(1), & \text{if (3.4.5) holds.} \end{cases}$$

Now, the end of the proof follows exactly the lines of the end of the proof of Theorem 3.2.1 (cf. (3.2.45) and on).

The proof of (c) is also similar to that of (c) of Theorem 3.2.1 (cf. (3.2.52) and (3.2.53)). ■

Corollary 3.4.1. *Assuming conditions of Theorem 3.4.1, we have, as $n \rightarrow \infty$,*

(a) *if $q \in \mathcal{Q}$ is such that (3.4.4) holds, then*

$$(3.4.34) \quad \hat{\beta}_n(x, t)/q(t) \xrightarrow{\mathcal{D}} G(x, t)/q(t)$$

in $D(\mathbb{R} \times [0, 1])$,

(b) *if q is positive on $(0, 1]$, then (3.4.34) holds in $D(\mathbb{R} \times [\delta, 1])$ for any $0 < \delta < 1$.*

Proof. (a) follows from Theorem 3.4.1 (a).

(b) By Theorem 3.4.1 (a) with $q(t) \equiv 1$, and (3.2.5). ■

Consider now the process $\{\hat{\beta}_n(x, t), 0 \leq t \leq 1, x \in \mathbb{R}\}$ defined by

$$(3.4.35) \quad \hat{\beta}_n(x, t) = \begin{cases} n^{1/2} \frac{[(n+1)t]}{n} \left(1 - \frac{[(n+1)t]}{n}\right) \hat{\zeta}_n(x, t), & 0 \leq t \leq 1, x \in \mathbb{R}, \\ 0, & t = 1, x \in \mathbb{R}, \end{cases}$$

with

$$(3.4.36) \quad \hat{\zeta}_n(x, t) = \frac{\sum_{i=1}^{[(n+1)t]} [1\{X_i \leq x\} - F(x; \hat{\theta}'_{[nt]})]}{[(n+1)t]} - \frac{\sum_{i=[(n+1)t]+1}^n [1\{X_i \leq x\} - F(x; \hat{\theta}''_{[nt]})]}{n - [(n+1)t]},$$

where, for each $\frac{1}{n+1} < t < \frac{n}{n+1}$, $\{\hat{\theta}'_{[nt]}\}$ and $\{\hat{\theta}''_{[nt]}\}$ are sequences of estimators of θ based on $X_1, \dots, X_{[(n+1)t]}$ and $X_{[(n+1)t]+1}, \dots, X_n$, respectively.

We assume that the estimation is done under conditions given by (1.2.2) (i)-(vi) in Chapter I, where, for each fixed t ,

$$(3.4.37) \quad [nt]^{\frac{1}{2}}(\hat{\theta}'_{[nt]} - \theta_0) = [nt]^{-\frac{1}{2}} \sum_{i=1}^{[nt]} l(X_i; \theta_0) + \epsilon'_{1n}(t),$$

and

$$(3.4.38) \quad (n - [nt])^{\frac{1}{2}}(\hat{\theta}''_{[nt]} - \theta_0) = (n - [nt])^{-\frac{1}{2}} \sum_{i=[nt]+1}^n l(X_i; \theta_0) + \epsilon''_{1n}(t).$$

We will assume further that, as $n \rightarrow \infty$,

$$(3.4.39) \quad \sup_{\delta \leq t \leq 1} \|\epsilon'_{1n}(t)\| = o_P(1) \text{ and } \sup_{\frac{1}{n} \leq t \leq \delta} \|\epsilon'_{1n}(t)\| = O_P(1),$$

and

$$(3.4.40) \quad \sup_{\delta \leq t \leq 1} \|\epsilon''_{1n}(t)\| = o_P(1) \text{ and } \sup_{\frac{1}{n} \leq t \leq \delta} \|\epsilon''_{1n}(t)\| = O_P(1),$$

for any $0 < \delta < 1$.

In the next result we obtain weighted approximations of the process $\hat{\beta}_n(x, t)$, as defined in (3.4.35), by the sequence of Gaussian processes $\{\Psi_n(x, t), 0 \leq t \leq 1, x \in \mathbb{R}\}$, where

$$(3.4.41) \quad \Psi_n(x, t) = \Gamma_n(F(x; \theta_0), t) - \left\{ \int l(x, \theta_0) d_x \Gamma_n(F(x; \theta_0), t) \right\} \cdot \nabla_{\theta} F(x; \theta_0)^T,$$

with

$$(3.4.42) \quad \begin{aligned} & \{\Gamma_n(F(x; \theta_0), t), 0 \leq t \leq 1, x \in \mathbb{R}\} \\ &= \{n^{-1/2}(K(F(x; \theta_0), nt) - tK(F(x; \theta_0), n)), 0 \leq t \leq 1, x \in \mathbb{R}\} \\ &\stackrel{D}{=} \{\Gamma(F(x; \theta_0), t), 0 \leq t \leq 1, x \in \mathbb{R}\} \text{ for each } n \geq 1. \end{aligned}$$

Theorem 3.4.2. Let $X_1, X_2, \dots$ be a sequence of i.i.d. r.v.'s with continuous distribution function $F \in \{F(x; \theta); x \in \mathbb{R}, \theta \in \Theta \subseteq \mathbb{R}^p\}$. Suppose that $\{\hat{\theta}'_{[n]}\}$ and $\{\hat{\theta}''_{[n]}\}$ satisfy conditions (1.2.2)(i)-(vi), with (3.4.37) – (3.4.40). Suppose further that the series in (3.4.2) converges for each $j = 1, \dots, p$. Then, there exists a Kiefer process $K(\cdot, \cdot)$ such that with the sequence of Gaussian processes $\{\Psi_n(x, t), 0 \leq t \leq 1, x \in \mathbb{R}\}$, and a weight function $q \in Q^*$, we have, as $n \rightarrow \infty$,

(a) if

$$(3.4.43) \quad I_{0,1}(q, c) < \infty \quad \text{for all } c > 0,$$

then

$$\sup_{0 < t < 1} \sup_{x \in \mathbb{R}} |\hat{\beta}_n(x, t) - \Psi_n(x, t)|/q(t) = o_P(1),$$

(b) if

$$(3.4.44) \quad I_{0,1}(q, c) < \infty \quad \text{for some } c > 0,$$

then

$$\sup_{0 < t < 1} \sup_{x \in \mathbb{R}} |\hat{\beta}_n(x, t) - \Psi_n(x, t)|/q(t) = O_P(1),$$

(c) if (3.4.44) holds then

$$\sup_{0 < t < 1} \sup_{x \in \mathbb{R}} |\hat{\beta}_n(x, t)|/q(t) \xrightarrow{D} \sup_{0 < t < 1} \sup_{x \in \mathbb{R}} |\Psi(x, t)|/q(t),$$

where the process $\{\Psi(x, t), 0 \leq t \leq 1, x \in \mathbb{R}\}$ is defined by

$$(3.4.45) \quad \Psi(x, t) = \Gamma(F(x; \theta_0), t) - \left\{ \int l(x, \theta_0) d_x \Gamma(F(x; \theta_0), t) \right\} \cdot \nabla_{\theta} F(x; \theta_0)^T.$$

Proof. From (3.4.36) we have, after adding and subtracting $F(x; \theta_0)$,

$$\begin{aligned} \hat{\zeta}_n(x, t) &= \frac{\sum_{i=1}^{[(n+1)t]} (1\{X_i \leq x\} - F(x; \theta_0))}{[(n+1)t]} - \frac{\sum_{i=[(n+1)t]+1}^n (1\{X_i \leq x\} - F(x; \theta_0))}{n - [(n+1)t]} \\ (3.4.46) \quad &+ (F(x; \theta_0) - F(x; \hat{\theta}'_{[nt]})) + (F(x; \hat{\theta}''_{[nt]}) - F(x; \theta_0)). \end{aligned}$$

Now, applying a one-term Taylor expansion to $F(x; \theta)$, around θ_0 , we get

$$\begin{aligned} F(x; \theta_0) - F(x; \hat{\theta}'_{[nt]}) &= (\hat{\theta}'_{[nt]} - \theta_0) \cdot \left(\nabla_{\theta} F(x; \theta_0)^{\top} - \nabla_{\theta} F(x; \hat{\theta}'_{[nt]})^{\top} \right) \\ &\quad - (\hat{\theta}'_{[nt]} - \theta_0) \cdot \nabla_{\theta} F(x; \theta_0)^{\top} \\ (3.4.47) \quad &= \epsilon'_{3n}(x, t) - (\hat{\theta}'_{[nt]} - \theta_0) \cdot \nabla_{\theta} F(x; \theta_0)^{\top}, \end{aligned}$$

where

$$(3.4.48) \quad \epsilon'_{3n}(x, t) = (\hat{\theta}'_{[nt]} - \theta_0) \cdot \left(\nabla_{\theta} F(x; \hat{\theta}'_{[nt]})^{\top} - \nabla_{\theta} F(x; \theta_0)^{\top} \right)$$

and

$$\sup_{0 \leq t \leq 1} \|\hat{\theta}'_{[nt]} - \theta_0\| \leq \sup_{0 \leq t \leq 1} \|\hat{\theta}'_{[nt]} - \theta_0\|.$$

Similarly, we have

$$\begin{aligned} F(x; \hat{\theta}''_{[nt]}) - F(x; \theta_0) &= (\hat{\theta}''_{[nt]} - \theta_0) \cdot \left(\nabla_{\theta} F(x; \theta_0)^{\top} - \nabla_{\theta} F(x; \hat{\theta}''_{[nt]})^{\top} \right) \\ &\quad + (\hat{\theta}''_{[nt]} - \theta_0) \cdot \nabla_{\theta} F(x; \theta_0)^{\top} \\ (3.4.49) \quad &= \epsilon''_{3n}(x, t) + (\hat{\theta}''_{[nt]} - \theta_0) \cdot \nabla_{\theta} F(x; \theta_0)^{\top}, \end{aligned}$$

where

$$(3.4.50) \quad \epsilon''_{3n}(x, t) = (\hat{\theta}''_{[nt]} - \theta_0) \cdot \left(\nabla_{\theta} F(x; \theta_0)^{\top} - \nabla_{\theta} F(x; \hat{\theta}''_{[nt]})^{\top} \right)$$

and

$$\sup_{0 \leq t \leq 1} \|\tilde{\theta}_{[nt]}'' - \theta_0\| \leq \sup_{0 \leq t \leq 1} \|\tilde{\theta}_{[nt]}'' - \theta_0\|.$$

Thus, in (4.4.34), we get,

$$\begin{aligned} \hat{\beta}_n(x, t) &= \beta_n^*(x, t) - n^{-1/2} \left\{ \left(1 - \frac{[(n+1)t]}{n}\right) \sum_{i=1}^{[(n+1)t]} l(X_i; \theta_0) \cdot \nabla_{\theta} F(x; \theta_0)^{\top} \right. \\ (3.4.51) \quad &\quad \left. - \frac{[(n+1)t]}{n} \sum_{i=[(n+1)t]+1}^n l(X_i; \theta_0) \cdot \nabla_{\theta} F(x; \theta_0)^{\top} \right\} \\ &+ \frac{[(n+1)t]}{n} \left(1 - \frac{[(n+1)t]}{n}\right) n^{1/2} \{\epsilon'_{3n}(x, t) + \epsilon''_{3n}(x, t)\} \\ &+ \frac{[(n+1)t]}{n} n^{-1/2} (n - [(n+1)t])^{\frac{1}{2}} \epsilon''_{1n}(t) \cdot \nabla_{\theta} F(x; \theta_0)^{\top} \\ &- \left(1 - \frac{[(n+1)t]}{n}\right) n^{-1/2} [(n+1)t]^{\frac{1}{2}} \epsilon'_{1n}(t) \cdot \nabla_{\theta} F(x; \theta_0)^{\top}, \end{aligned}$$

where, $\beta_n^*(x, t)$ is as in (3.3.5).

We assume first that q is positive on $(0, 1)$ and such that (3.3.27) holds. Adding and subtracting $\Psi_n(x, t)$ (cf. (3.4.41)) we get in (3.4.51),

$$\begin{aligned} (3.4.52) \quad &\sup_{\frac{1}{n+1} \leq t \leq \frac{n}{n+1}} \sup_{x \in \mathbb{R}} |\hat{\beta}_n(x, t) - \Psi_n(x, t)|/q(t) \\ &\leq \sup_{\frac{1}{n+1} \leq t \leq \frac{n}{n+1}} \sup_{x \in \mathbb{R}} |\beta_n^*(x, t) - \Gamma_n(F(x; \theta_0), t)|/q(t) \\ &+ \sup_{\frac{1}{n+1} \leq t \leq \frac{n}{n+1}} \sup_{x \in \mathbb{R}} |\{\varphi_n(t) - \int l(x, \theta_0) d_x \Gamma_n(F(x; \theta_0), t)\} \cdot \nabla_{\theta} F(x; \theta_0)^{\top}|/q(t) \\ &+ \sup_{\frac{1}{n+1} \leq t \leq \frac{n}{n+1}} \sup_{x \in \mathbb{R}} \frac{[(n+1)t]}{n} \left(1 - \frac{[(n+1)t]}{n}\right) n^{1/2} |\epsilon'_{3n}(x, t) + \epsilon''_{3n}(x, t)|/q(t) \end{aligned}$$

$$\begin{aligned}
& + \sup_{\frac{1}{n+1} \leq t \leq \frac{2}{n+1}} \sup_{z \in \mathbb{R}} \frac{[(n+1)t]}{n} n^{-1/2} (n - [(n+1)t])^{\frac{1}{2}} |\varepsilon''_{1n}(t) \cdot \nabla_{\theta} F(x; \theta_0)^T| / q(t) \\
& + \sup_{\frac{1}{n+1} \leq t \leq \frac{2}{n+1}} \sup_{z \in \mathbb{R}} (1 - \frac{[(n+1)t]}{n}) n^{-1/2} [(n+1)t]^{\frac{1}{2}} |\varepsilon'_{1n}(t) \cdot \nabla_{\theta} F(x; \theta_0)^T| / q(t) \\
& := I_n^{(1)} + I_n^{(2)} + I_n^{(3)} + I_n^{(4)} + I_n^{(5)}.
\end{aligned}$$

Let $0 < \delta < 1/2$ be fixed. We have

(3.4.53)

$$\begin{aligned}
& \sup_{\frac{1}{n+1} \leq t \leq \frac{2}{n+1}} (1 - \frac{[(n+1)t]}{n}) n^{-1/2} [(n+1)t]^{1/2} \|\varepsilon'_{1n}(t)\| / q(t) \\
& \leq \sup_{\frac{1}{n+1} \leq t \leq 1-\delta} (1 - \frac{[(n+1)t]}{n}) \sup_{\frac{1}{n+1} \leq t \leq 1-\delta} n^{-1/2} ([(n+1)t])^{1/2} \|\varepsilon'_{1n}(t)\| / q(t) \\
& \quad \vee \sup_{1-\delta \leq t \leq \frac{2}{n+1}} (1 - \frac{[(n+1)t]}{n}) / q(t) \sup_{1-\delta \leq t \leq 1} n^{-1/2} ([(n+1)t])^{1/2} \|\varepsilon'_{1n}(t)\| \\
& \leq (1 - \frac{1}{n}) \sup_{\frac{1}{n+1} \leq t \leq 1-\delta} n^{-1/2} ([(n+1)t])^{1/2} \|\varepsilon'_{1n}(t)\| / q(t) \\
& \quad \vee (1 + \frac{1}{n})^{1/2} \sup_{1-\delta \leq t \leq \frac{2}{n+1}} (1 - \frac{[(n+1)t]}{n}) / q(t) \sup_{1-\delta \leq t \leq 1} \|\varepsilon'_{1n}(t)\|.
\end{aligned}$$

It is easy to check that, as $n \rightarrow \infty$,

$$\sup_{1-\delta \leq t \leq \frac{2}{n+1}} (1 - \frac{[(n+1)t]}{n}) / q(t) \leq O(1) \sup_{1-\delta \leq t < 1} (1-t)^{1/2} / q(t).$$

Hence, by (3.3.27), (3.4.1) and the fact that the vector $\nabla_{\theta} F(x; \theta_0)$ is uniformly bounded in x , in sup-norm (cf. (1.2.2)(vi)), we have, as $n \rightarrow \infty$,

(3.4.54)

$$I_n^{(5)} := \sup_{\frac{1}{n+1} \leq t \leq \frac{n}{n+1}} \sup_{x \in \mathbb{R}} \left(1 - \frac{[(n+1)t]}{n}\right) n^{-1/2} [(n+1)t]^{\frac{1}{2}} |\epsilon'_{1n}(t) \cdot \nabla_{\theta} F(x; \theta_0)^{\top}| / q(t)$$

$$= o_P(1).$$

On the other hand, arguing as above, substituting $\epsilon'_{1n}(t)$ by $\epsilon''_{1n}(t)$ in (3.4.53), we have, by (1.2.2)(vi), (3.3.27) and (3.4.40), as $n \rightarrow \infty$,

(3.4.55)

$$I_n^{(4)} := \sup_{\frac{1}{n+1} \leq t \leq \frac{n}{n+1}} \sup_{x \in \mathbb{R}} \frac{[(n+1)t]}{n} n^{-1/2} (n - [(n+1)t])^{1/2} |\epsilon''_{1n}(t) \cdot \nabla_{\theta} F(x; \theta_0)^{\top}| / q(t)$$

$$= o_P(1).$$

Next we have that, for any fixed $0 < \delta < 1/2$,

$$\begin{aligned} & \sup_{\frac{1}{n+1} \leq t \leq \frac{n}{n+1}} \sup_{x \in \mathbb{R}} \frac{[(n+1)t]}{n} \left(1 - \frac{[(n+1)t]}{n}\right) n^{1/2} |\epsilon'_{3n}(x, t)| / q(t) \\ & \leq \sup_{\frac{1}{n+1} \leq t \leq 1-\delta} \left(1 - \frac{[(n+1)t]}{n}\right) n^{-1/2} [(n+1)t]^{1/2} \\ & \quad \times \sup_{\frac{1}{n+1} \leq t \leq 1-\delta} \sup_{x \in \mathbb{R}} [(n+1)t]^{1/2} |\epsilon'_{3n}(x, t)| / q(t) \\ & \vee \sup_{1-\delta \leq t \leq \frac{n}{n+1}} \left(1 - \frac{[(n+1)t]}{n}\right) / q(t) \sup_{1-\delta \leq t \leq 1} n^{-1/2} [(n+1)t]^{1/2} \\ & \quad \times \sup_{1-\delta \leq t \leq 1} \sup_{x \in \mathbb{R}} [(n+1)t]^{1/2} |\epsilon'_{3n}(\cdot, t)| \\ & \leq O(1) \sup_{\frac{1}{n+1} \leq t \leq 1-\delta} \sup_{x \in \mathbb{R}} [(n+1)t]^{1/2} |\epsilon'_{3n}(x, t)| / q(t) \\ & \vee O(1) \sup_{1-\delta \leq t \leq \frac{n}{n+1}} (1-t)^{1/2} / q(t) \sup_{1-\delta \leq t \leq 1} \sup_{x \in \mathbb{R}} [(n+1)t]^{1/2} |\epsilon'_{3n}(x, t)|. \end{aligned}$$

We also note that, $[(n+1)t]^{1/2} \epsilon'_{3n}(x, t) = \epsilon_{3(n+1)}(x, t)$, for all $x \in \mathbb{R}$ and $\frac{1}{n+1} \leq t \leq \frac{n}{n+1}$, where $\epsilon_{3n}(x, t)$ is as in (3.4.15), we have, by (3.3.27) and (3.4.30), as $n \rightarrow \infty$,

$$(3.4.56) \quad \sup_{\frac{1}{n+1} \leq t \leq \frac{n}{n+1}} \sup_{x \in \mathbb{R}} \frac{[(n+1)t]}{n} \left(1 - \frac{[(n+1)t]}{n}\right) n^{1/2} |\epsilon'_{3n}(x, t)|/q(t) = o_P(1).$$

Furthermore, we note that, by (3.4.38), we can write

$$(3.4.57) \quad (n - [(n+1)t])^{\frac{1}{2}} (\hat{\theta}_{[nt]}'' - \theta_0) = [(n+1)(1-t)]^{-\frac{1}{2}} \sum_{i=1}^{[(n+1)(1-t)]} l(X_{n+i-1}, \theta_0) + \epsilon''_{1n}(t).$$

Thus, by symmetry, (3.3.27) and (3.4.40) and (3.4.50), following the same steps as in the proof of (3.4.56), we get, as $n \rightarrow \infty$,

$$(3.4.58) \quad \sup_{\frac{1}{n+1} \leq t \leq \frac{n}{n+1}} \sup_{x \in \mathbb{R}} \frac{[(n+1)t]}{n} \left(1 - \frac{[(n+1)t]}{n}\right) n^{1/2} |\epsilon''_{3n}(x, t)|/q(t) = o_P(1).$$

Thus, by (3.4.56), (3.4.58) and the triangle inequality (cf. (3.4.52)), we have, as $n \rightarrow \infty$,

$$(3.4.59) \quad I_n^{(3)} = o_P(1).$$

The proof of

$$(3.4.60) \quad I_n^{(2)} = o_P(1), \text{ as } n \rightarrow \infty,$$

is similar, line by line, to that of (3.3.37), substituting there Lemma 3.2.3 by Lemma 3.4.1.

In the light of (3.3.9) of Remark 3.3.2, we have, from (3.4.52), that, by (3.4.54), (3.4.56), (3.4.58) and (3.4.60), as $n \rightarrow \infty$,

$$(3.4.61) \quad \sup_{\frac{1}{n+1} \leq t \leq \frac{n}{n+1}} \sup_{x \in \mathbb{R}} |\hat{\beta}_n(x, t) - \Psi_n(x, t)|/q(t) = o_P(1),$$

provided the function q is positive on $(0, 1]$ and such that (3.3.27) holds.

Since, by definition, $\hat{\beta}_n(x, t) = 0$ for any $t \in A_n := (0, \frac{1}{n+1}] \cup [\frac{n}{n+1}, 1)$, in virtue of (3.4.61), in order to prove (a) and (b), it now suffices to show, that, as $n \rightarrow \infty$,

$$(3.4.62) \quad \sup_{t \in A_n} \sup_{x \in \mathbb{R}} |\Psi_n(x, t)|/q(t) = \begin{cases} o_P(1), & \text{if (3.4.43) holds,} \\ O_P(1), & \text{if (3.4.44) holds.} \end{cases}$$

The proof of (a) and (b) is similar, line by line from (3.3.64) to (3.3.67), to that of (a) and (b) of Theorem 3.3.1.

The proof of (c) is also similar to that of (c) of Theorem 3.3.1 (cf. (3.3.68) and (3.3.69)). ■

Corollary 3.4.2. *Under the conditions of Theorem 3.4.2 we have, as $n \rightarrow \infty$,*

(a) *if $q \in Q^*$ and is such that (3.4.43) holds then,*

$$(3.4.63) \quad \hat{\beta}_n(x, t)/q(t) \xrightarrow{D} \Psi(x, t)/q(t)$$

in $D(\mathbb{R} \times [0, 1])$,

(b) if q is a positive function on $(0, 1)$, then, (3.4.63) holds in $D(\mathbb{R} \times [\delta, 1 - \delta])$ for any $0 < \delta < 1/2$.

Remark 3.4.2. We presented in Section 3.2 the approximation of the sequentially estimated empirical process based on observations from a normal family. Although this process is of interest on its own, we also wanted to illustrate the kind of conditions needed in case of a general parametric family. We have, for instance that (3.4.1) and (3.4.1*) hold (cf. (3.2.1), (3.2.2) and (3.2.3), and (3.2.26) and (3.2.28), respectively), as well as the condition for Lemma 3.4.1 (cf. Lemma 3.2.3).

We note that the latter condition, the convergence of the series in (3.4.2), looks like a "Lindeberg-type" of a condition based on the second moments of the vector-valued function l produced by the estimation.

As far as conditions (3.4.1) and (3.4.1*) are concerned, we can also refer to Section 4 of Burke, Csörgő, Csörgő and Révész (1979). They point out that maximum likelihood estimators often satisfy (3.4.1) with $l(x; \theta_0) = \nabla_{\theta} \log f(x; \theta_0) \cdot I^{-1}(\theta_0)$, where f is the density function of F and $I^{-1}(\theta_0)$ is the inverse of the Fisher information matrix:

$$(3.4.64) \quad I(\theta_0) = E (\nabla_{\theta} \log f(X_1; \theta_0))^{\top} \cdot (\nabla_{\theta} \log f(X_1; \theta_0)).$$

They illustrate that, under certain regularity conditions, maximum likelihood estimators have a sum representation, as in (3.4.1), with $\epsilon_{1n}(t) = O((nt)^{-\epsilon})$ a.s., for some $\epsilon > 0$.

Suppose that the vector θ is estimated by $\{\hat{\theta}_n\}$, a sequence of maximum likelihood estimators. Assume the following conditions:

- (i) $F(x; \theta)$ has a density function $f(x; \theta)$.
- (ii) The parameter set Θ is an open cube (bounded or unbounded) in $\mathbb{R}^p$.
- (iii) The density $f(x; \theta)$ is measurable with respect to $X \times \xi$, where X is the collection of Borel subsets of $\mathbb{R}$ and ξ is a σ -algebra of measurable subsets of Θ .
- (iv) If $\theta \neq \theta'$, then $\int |f(x; \theta) - f(x; \theta')| dx > 0$.
- (v) There is a $\delta > 0$, such that

$$\sup_{\theta} \|\theta - \theta_0\|^\delta \int [f(x; \theta)f(x; \theta_0)]^{1/2} dx < \infty,$$

where θ_0 is the true value of θ .

- (vi) All second partial derivatives of the function $g(x; \theta) = \log f(x; \theta)$ with respect to the components of θ exist.
- (vii) First and second partial differentiation of $\int f(x; \theta) dx$, with respect to the components of θ , can be taken under the integral sign.
- (viii) The matrix $E |g''_{\theta\theta}(X_1; \theta)|^{1+\delta}$ and the vector $E |\nabla_{\theta} g(X_1, \theta)|^{2+\delta}$ have bounded components, for some $\delta > 0$, on compact subsets of Θ (cf. page 1 for notation).
- (ix) There exists a function $H(x)$ and $\beta > 0$ such that $\|g''_{\theta\theta}(x; \theta_2) - g''_{\theta\theta}(x; \theta_1)\| < H(x)\|\theta_2 - \theta_1\|^\beta$, and $E H(X_1)$ exists for each $\theta \in \Theta$.

Conditions (i)-(v) correspond to conditions I and III of Ibragimov and Has'minskiĭ (1972). The following Theorem is a generalization of Theorem 1 of Ibragimov and Has'minskiĭ (1973b). It is Theorem 4.1 in Burke, Csörgő, Csörgő and Révész (1979).

Theorem 3.4.A. *If the above conditions (i)-(ix) are satisfied, then for some $\epsilon > 0$*

$$n^{1/2}(\hat{\theta}_n - \theta_0) \cdot I(\theta_0) - n^{-1/2} \sum_{j=1}^n \nabla_{\theta} g(X_j, \theta_0) \stackrel{a.s.}{=} O(n^{-\epsilon}),$$

where $\nabla_{\theta} g(x, \theta_0) = \nabla_{\theta} \log f(x; \theta_0)$.

We conclude from the above theorem that for a sequence $X_1, X_2, \dots$ of i.i.d. random variables from a family of continuous distribution functions $\{F(x; \theta); x \in \mathbb{R}, \theta \in \Theta \subseteq \mathbb{R}^p\}$, and such that the regularity conditions (i)-(ix) are satisfied, by estimating θ by the sequence of maximum likelihood estimators $\{\hat{\theta}_{[nt]}\}$, based on the sample $X_1, \dots, X_{[nt]}$, for each $\frac{1}{n} \leq t \leq 1$, conditions (3.4.1) and (3.4.1*) are satisfied.

Furthermore, (3.4.2) becomes

$$(3.4.65) \quad \sum_{k=1}^{\infty} \frac{1}{k} \int_{|x| > u(k)} \frac{\partial}{\partial \theta_j} \log f(x; \theta_0) f(x; \theta_0) dx, \quad j = 1, \dots, p.$$

Remark 3.4.3. Under the conditions (i)-(ix), for each n , the covariance function of the process $G_n(\cdot, \cdot)$ in (3.4.6) (cf. Remark 3.4.1) is given by

$$(3.4.66) \quad \begin{aligned} E G_n(x, s) G_n(y, t) = (s \wedge t) \{ & F(x \wedge y; \theta_0) - F(x; \theta_0) F(y; \theta_0) \\ & - \nabla_{\theta} F(x; \theta_0) \cdot I^{-1}(\theta_0) \cdot \nabla_{\theta} F(y; \theta_0)^T \}. \end{aligned}$$

Chapter 4

Sequentially estimated normal empirical processes

Weighted weak convergence under contiguous measures

4.1 Introduction.

In the previous chapter we studied the asymptotic behavior of parameter-estimated empirical processes based on independent and identically distributed observations from a normal family. In this chapter we are to study the same processes under the assumption of *contiguity*, a condition of mathematical and statistical interest when studying asymptotics involving independent but not necessarily identically distributed observations.

In Section 4.2 we discuss the notion of contiguity and Le Cam's *third lemma*, a very useful result in statistical applications. It provides a tool for determining convergence properties of statistics of interest under contiguous alternatives via their joint convergence together with the log-likelihood ratio process under the null assumption of i.i.d. sampling.

In Section 4.3 we present a parametrization of contiguity which was used by Khmaladze and Parjanadze (1986). Under these contiguity conditions, Szyszkowicz (1991a,b,c, 1992a,b, 1994, 1995) studied the weighted asymptotic behavior of processes like partial sums and empirical processes based on observations, their ranks and sequential ranks. We present some of her results in Section 4.4.

In Section 4.5 we study the weighted asymptotic behavior of the sequentially estimated normal empirical processes studied in Sections 3.2 and 3.3, under the parametrization of contiguity of Section 4.3.

In Sections 4.2, 4.3 and 4.4 we follow the presentation of Szyszkowicz (1992b, Sections 2.1, 2.2 and 2.3) (cf. also Szyszkowicz (1991a,b,c, 1992a and 1994) for the presentation of Section 4.3).

4.2 The notion of contiguity.

The notion of contiguity, which was introduced by Lucien Le Cam (1960) as a criterion of nearness of sequences of probability measures, has an enduring place among the fundamental concepts of mathematical statistics. From the point of view of mathematics, contiguity can be thought of as an asymptotic version of the idea of absolute continuity of measures.

Definition 4.2.A. Let $\{(\mathcal{X}_n, \mathcal{F}_n)\}$, $n \geq 1$, be a sequence of measurable spaces and let P_n, Q_n be probability measures on $(\mathcal{X}_n, \mathcal{F}_n)$.

The sequence of probability measures $\{Q_n\}$ is said to be *contiguous* to the sequence $\{P_n\}$, denoted $\{Q_n\} \triangleleft \{P_n\}$, if for every sequence of sets $\{A_n\}$, $A_n \in \mathcal{A}_n$,

$$\lim_{n \rightarrow \infty} P_n(A_n) = 0 \text{ implies } \lim_{n \rightarrow \infty} Q_n(A_n) = 0.$$

This definition is not symmetric in P and Q . Initially, the concept of contiguity, as introduced by Le Cam (1960) and then adopted by Roussas (1972), was used to describe a symmetric relation - that may be termed as *mutual contiguity*. Namely we will say that $\{Q_n\}$ and $\{P_n\}$ are mutually contiguous, denoted $\{Q_n\} \triangleleft \triangleright \{P_n\}$, if both $\{Q_n\} \triangleleft \{P_n\}$ and $\{P_n\} \triangleleft \{Q_n\}$, i.e., if for every sequence of sets $\{A_n\}$, $A_n \in \mathcal{A}_n$,

$$\lim_{n \rightarrow \infty} P_n(A_n) = 0 \text{ if and only if } \lim_{n \rightarrow \infty} Q_n(A_n) = 0.$$

The one-sided definition is used by Hájek and Šidák (1967), Greenwood and Shirayev (1985) and also in the new book by Le Cam and Yang (1990). The notation is due to Witting and Nölle (1970) (cf. Oosterhoff and Van Zwet, 1975).

Remark 4.2.A. We note that contiguity implies that any sequence of random variables $\{T_n\}$ which converges to zero in P_n -probability, converges to zero also in Q_n -probability. This provides an alternative definition of contiguity and in fact it is used as such by Le Cam (1960), Le Cam and Yang (1990) and Roussas (1972).

Characterizations of contiguity are usually given in terms of some properties of the likelihood ratio sequence.

Having $\{P_n\}$ and $\{Q_n\}$ on $(\mathcal{X}_n, \mathcal{A}_n)$, let μ_n be a σ -finite measure dominating P_n and Q_n , i.e., P_n and Q_n are absolutely continuous with respect to μ_n for each n ($P_n \ll \mu_n$ and $Q_n \ll \mu_n$). Let X_n be the identity map from $\mathcal{X}_n$ onto $\mathcal{X}_n$ for each n . Let the random variables

$$p_n = \frac{dP_n}{d\mu_n} = \frac{dP_n}{d\mu_n}(X_n), \quad q_n = \frac{dQ_n}{d\mu_n} = \frac{dQ_n}{d\mu_n}(X_n)$$

be finite versions of the corresponding Randon-Nikodym derivatives, and let

$$\lambda_n = \begin{cases} q_n/p_n & \text{if } p_n > 0 \\ \infty & \text{if } p_n = 0 < q_n \\ 0 & \text{if } p_n = 0 = q_n. \end{cases}$$

Remark 4.2.B. If $Q_n \ll P_n$ then λ_n is simply a version of the Radon-Nikodym derivative dQ_n/dP_n .

For more details on contiguity and its characterizations we refer to Szyszkowicz (1992b), Chapter 2, where there is quite a complete and detailed account on the matter.

The following result, that is known as Le Cam's *third lemma* (cf. Hájek and Šidak, 1967), is very useful in statistical applications.

We say that a pair (T_n, Λ_n) is asymptotically jointly normal $(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \sigma_{12})$ if it converges in distribution to a normal vector (Z_1, Z_2) such that $EZ_i = \mu_i$, $\text{var}Z_i = \sigma_i^2$, $i = 1, 2$, and $\text{cov}(Z_1, Z_2) = \sigma_{12}$, where $\Lambda_n = \log \lambda_n$.

Lemma 4.2.A. Assume that the pair (T_n, Λ_n) under P_n is asymptotically jointly normal $(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \sigma_{12})$ with $\mu_2 = -\frac{1}{2}\sigma_2^2$. Then T_n under Q_n is asymptotically normal $(\mu_1 + \sigma_{12}, \sigma_1^2)$, i.e. if

$$\begin{pmatrix} T_n \\ \Lambda_n \end{pmatrix} \xrightarrow{D} N\left(\mu_1, -\frac{1}{2}\sigma_2^2, \sigma_1^2, \sigma_2^2, \sigma_{12}\right) \text{ under } P_n,$$

then

$$T_n \xrightarrow{D} N(\mu_1 + \sigma_{12}, \sigma_1^2) \text{ under } Q_n.$$

We note that even if contiguity is not assumed here explicitly, it is implied by the requirement that $\Lambda_n \xrightarrow{D} N(-\frac{1}{2}\sigma^2, \sigma^2)$, due to Le Cam's *first lemma* (cf. Oosterhoff and Van Zwet (1975), or Szyszkowicz (1992b)).

The multivariate version of Le Cam's third lemma reads as follows (cf. Shorack and Wellner, 1986, p.156).

Lemma 4.2.B. *If*

$$\begin{pmatrix} \Lambda_n \\ \tilde{T}_n \end{pmatrix} \xrightarrow{\mathcal{D}} N\left(\begin{bmatrix} -\frac{1}{2}\sigma^2 \\ \bar{\mu} \end{bmatrix}, \begin{bmatrix} \sigma^2 & \bar{\sigma} \\ \bar{\sigma} & \Sigma \end{bmatrix}\right) \text{ under } P_n,$$

then

$$\tilde{T}_n \xrightarrow{\mathcal{D}} N([\bar{\mu} + \bar{\sigma}], [\Sigma]), \text{ under } Q_n.$$

Note. It is clear that when T_n is a stochastic process then the weak convergence of T_n under Q_n can be derived if we have the joint weak convergence of Λ_n and T_n under P_n with a specified limit. Namely, the multivariate version of Le Cam's third lemma will give us limits of all finite dimensional distributions under Q_n , and tightness will be implied by the weak convergence of T_n under P_n and contiguity of $\{Q_n\}$ to $\{P_n\}$.

4.3 A parametrization of contiguity.

Let $X_1, X_2, \dots$ be independent random variables. Suppose we wish to test the null hypothesis

$$H_0 : X_i, 1 \leq i \leq n, \text{ have the same distribution } F,$$

versus the alternative hypotheses

$$H_{1n} : X_i, 1 \leq i \leq n, \text{ have the respective distribution functions } F_{in},$$

where we assume that all F_{in} are absolutely continuous with respect to the distribution function F and

$$(4.3.1) \quad \left[\frac{dF_{in}}{dF}(F^{-1}(u)) \right]^{1/2} \equiv 1 + \frac{1}{2n^{1/2}} g_n(t, u), \quad \frac{i-1}{n} < t \leq \frac{i}{n},$$

where $F^{-1}(u) = \inf\{x : F(x) \geq u\}$, $0 < u \leq 1$, $F^{-1}(0) = F^{-1}(0+)$. We assume also that there exists a function $h \in L^2[0, 1]^2$ such that

$$(4.3.2) \quad \int_0^1 h(t, u) du = 0, \text{ for almost all } t \in [0, 1]$$

and

$$(4.3.3) \quad \int_0^1 \int_0^1 [g_n(t, u) - h(t, u)]^2 dudt \rightarrow 0, \text{ as } n \rightarrow \infty.$$

The sequence of direct products $F_{1n} \times \dots \times F_{nn}$, $n = 1, 2, \dots$, thus parametrized, is contiguous to the sequence $F \times \dots \times F$, via Le Cam's *first lemma* (cf. Oosterhoff and Van Zwet (1975), or Szyszkowicz (1991a, 1992b)).

We assume without loss of generality that under H_0 the X_i are of the form $F^{-1}(U_i)$, where the U_i are independent identically distributed uniform- $(0, 1)$ random variables, and so $X_i \stackrel{\mathcal{D}}{=} F^{-1}(U_i)$, $i = 1, 2, \dots$.

From (4.3.1) we have, after a two-term Taylor expansion,

$$(4.3.4) \quad \begin{aligned} \log \left(\frac{dF_{1n}}{dF}(F^{-1}(u)) \right) &= 2 \log \left(1 + \frac{1}{2n^{1/2}} g_n(t, u) \right) \\ &= \frac{1}{n^{1/2}} g_n(t, u) + R_n^*(t, u), \quad \frac{i-1}{n} < t \leq \frac{i}{n}, \end{aligned}$$

where

$$(4.3.5) \quad |R_n^*\left(\frac{i}{n}, u\right)| \leq C(g_n\left(\frac{i}{n}, u\right))^2 / 4n,$$

for some constant $C > 0$.

Define the centered log-likelihood ratio process $\mathcal{L}_n$ by

$$(4.3.6) \quad \mathcal{L}_n(t) = \sum_{i=1}^{[nt]} \left\{ \log \frac{dF_{in}}{dF} (F^{-1}(U_i)) - E \log \frac{dF_{in}}{dF} (F^{-1}(U_i)) \right\}.$$

By (4.3.4) we have,

$$(4.3.7) \quad \mathcal{L}_n(t) := \sum_{i=1}^{[nt]} n^{-1/2} h_n\left(\frac{i}{n}, U_i\right) + R_n\left(\frac{i}{n}, U_i\right),$$

where,

$$(4.3.8) \quad \begin{aligned} h_n(t, u) &:= g_n(t, u) - E(g_n(t, U)) \text{ and} \\ R_n(t, u) &:= R_n^*(t, u) - E(R_n^*(t, U)), \quad \frac{i-1}{n} < t \leq \frac{i}{n}. \end{aligned}$$

Since

$$\int_0^1 \int_0^1 [h_n(t, u) - h(t, u)]^2 du dt \leq \int_0^1 \int_0^1 [g_n(t, u) - h(t, u)]^2 du dt,$$

by (4.3.3) we have, as $n \rightarrow \infty$,

$$(4.3.9) \quad \int_0^1 \int_0^1 [h_n(t, u) - h(t, u)]^2 du dt \rightarrow 0.$$

Also, from (4.3.5) and (4.3.7),

$$(4.3.10) \quad \lim_{n \rightarrow \infty} E(\mathcal{L}_n(t))^2 < \infty.$$

Let $\mathcal{L}$ be a Gaussian process with mean zero and covariance function $Q_{\mathcal{L}}(t_1 \wedge t_2)$, where

$$(4.3.11) \quad Q_{\mathcal{L}}(t) = \int_0^t \int_0^1 h^2(s, u) du ds.$$

Then, by (4.3.1), (4.3.2), (4.3.3), Theorem 2 of Oosterhoff and Van Zwet, 1975, (cf. Szyszkowicz, 1991a, proof of Theorem 3.1, or Szyszkowicz, 1992b, Lemma 2.3.1) under H_0 , we have

$$(4.3.12) \quad \mathcal{L}_n(t) \xrightarrow{\mathcal{D}} \mathcal{L}(t) \text{ in } D[0, 1].$$

4.4 Some stochastic processes under contiguous measures.

Szyszkowicz (1991a,b,c, 1992b, 1994) studied weighted weak convergence of, among other processes, partial sums, the process $\beta_n(x, t)$ in (2.1.1) and the "bridge" type process $\beta_n^*(x, t)$ in (3.3.5), under the sequence of contiguous alternatives of H_1 as parametrized in (4.3.1)-(4.3.3), via Le Cam's *third lemma*, for the same class of weight functions as used in previous chapters. In this section, for the sake of illustrating applications of Le Cam's *third lemma*, and for facilitating the presentation and proofs of our propositions in Section 4.5, we present some of her results.

Let $(W, \mathcal{L})$ be the two-dimensional Gaussian process, where W is a standard Wiener process, $\mathcal{L}$ is a Gaussian process with mean zero and covariance function $Q_{\mathcal{L}}(t_1 \wedge t_2)$, where

$$(4.4.1) \quad Q_{\mathcal{L}}(t) = \int_0^t \int_0^1 h^2(s, u) du ds$$

(cf. (4.3.11), and the covariance function of W and $\mathcal{L}$ is

$$(4.4.2) \quad Q_{W, \mathcal{L}}(t_1, t_2) = EW(t_1)\mathcal{L}(t_2) = C(t_1 \wedge t_2),$$

with

$$(4.4.3) \quad C(t) = \int_0^t \int_0^1 h(s, u) F^{-1}(u) du ds.$$

We have the following (cf. Szyszkowicz (1991a))

Lemma 4.4.A. *Let $X_1, X_2, \dots$ be i.i.d.r.v.'s with distribution F such that $\int x dF = 0$, $\int x^2 dF = 1$ and $S(nt) = X_1 + \dots + X_{[nt]}$. Then, as $n \rightarrow \infty$,*

$$(n^{-1/2}S(nt), \mathcal{L}_n(t)) \xrightarrow{\mathcal{D}} (W(t), \mathcal{L}(t)) \text{ in } D^2[0, 1].$$

By Lemma 4.4.A and Le Cam's third lemma, under contiguous alternatives, all finite dimensional distributions of S_n converge to those of $W + d$, where $d(t) = \text{Cov}(W(t), \mathcal{L}(1)) = C(t)$. Since contiguity preserves tightness, the following result holds (cf. Szyszkowicz (1991c, 1992a,b)).

Theorem 4.4.A. *Let $X_1, X_2, \dots$ be independent random variables and $S(nt) = X_1 + \dots + X_{[nt]}$. We assume that, under H_0 , $\int x dF = 0$ and $\int x^2 dF = 1$. Then, as $n \rightarrow \infty$, under H_1 we have*

$$n^{-1/2}S(nt) \xrightarrow{\mathcal{D}} W(t) + C(t) \text{ in } D[0, 1],$$

where

$$C(t) = \int_0^t \int_0^1 h(s, u) F^{-1}(u) du ds.$$

Let

$$(4.4.4) \quad c(s, t) = \int_0^t \int_0^s h(\tau, u) du d\tau, \quad 0 \leq s, t \leq 1,$$

where h is the function as given in the parametrization of H_1 .

In terms of the empirical process $\beta_n(x, t)$ as in (2.1.1), we have the following result (cf. Szyszkowicz, 1994, Theorem 5.1).

Theorem 4.4.B. *We assume that, under H_0 , $X_1, X_2, \dots$, are i.i.d. r.v.'s with continuous distribution F . Let $q \in \mathcal{Q}$ be such that*

$$(4.4.5) \quad I_0(q, c) < \infty \text{ for all } c > 0.$$

Then, under H_1 , we have, as $n \rightarrow \infty$,

$$(4.4.6) \quad \beta_n(x, t)/q(t) \xrightarrow{\mathcal{D}} (K(F(x), t) + c(F(x), t))/q(t) \text{ in } D(\mathbb{R} \times [0, 1]),$$

where $K(\cdot, \cdot)$ is a Kiefer process.

Introduce the Gaussian process $(K(s, t), \mathcal{L}(t))$, where $K(s, t)$ is a Kiefer process, and the covariance function of the process $(K, \mathcal{L})$ is

$$Q_{K, \mathcal{L}}(s, t_1, t_2) = E K(s, t_1) \mathcal{L}(t_2) = c(s, t_1 \wedge t_2),$$

where $c(s, t)$ is defined by (4.4.4).

As in the proof of Theorem 4.4.A, the proof of Theorem 4.4.B is a consequence of Le Cam's third lemma and the following joint weak convergence result, under the null assumption, of $\beta_n(x, t)$ and the centered log-likelihood process $\mathcal{L}_n(t)$ given in (4.3.6) (cf. Szyszkowicz, 1994, Lemma 5.1).

Lemma 4.4.B. *We assume that H_0 holds. Then, as $n \rightarrow \infty$,*

$$(\beta_n, \mathcal{L}_n) \xrightarrow{\mathcal{D}} (K, \mathcal{L})$$

in $D(\mathbb{R} \times [0, 1]) \times D[0, 1]$.

As for the "bridge" type process $\beta_n^*(x, t)$, in (3.3.5), we have (cf. Szyszkowicz, 1994, Theorem 5.2)

Theorem 4.4.C. Assume that, under H_0 , $X_1, X_2, \dots$ are i.i.d. r.v.'s with continuous distribution F . If $q \in \mathcal{Q}^*$ is such that

$$I_{0,1}(q, c) < \infty \text{ for all } c > 0,$$

then, under H_1 , we have, as $n \rightarrow \infty$,

$$(4.4.7) \quad \beta_n^*(x, t)/q(t) \xrightarrow{\mathcal{D}} (\Gamma(F(x), t) + c^*(F(x), t))/q(t) \text{ in } \mathbf{D}(\mathbb{R} \times [0, 1]),$$

where $\{\Gamma(s, t), 0 \leq s \leq 1, 0 \leq t \leq 1\}$ is a Gaussian process with mean zero and covariance as in (3.3.4), and

$$c^*(s, t) = c(s, t) - tc(s, 1),$$

with $c(s, t)$ as in (4.4.4).

Remark 4.4.1. Similarly as in our Remarks 2.1.5 and 3.3.4, we note here that Theorems 4.4.B and 4.4.C, were first proved under stronger conditions. Assuming that q is a positive function on $(0, 1]$, (4.4.6) was obtained under condition

$$\lim_{t \downarrow 0} (t \log \log 1/t)^{1/2}/q(t) = 0,$$

while (4.4.7) was proved for positive functions on $(0, 1)$ and such that

$$\lim_{t \downarrow 0} (t(1-t) \log \log 1/t(1-t))^{1/2}/q(t) = 0$$

and

$$\lim_{t \uparrow 1} (t(1-t) \log \log 1/t(1-t))^{1/2}/q(t) = 0$$

(cf. also Remarks 2.1.5 and 3.3.3 here).

Theorems 4.4.B and 4.4.C are improvements of the earlier Theorems 5.1 and 5.2 of Szyszkowicz (1994), respectively, on account of the integral test for suprema of Kiefer processes of Csörgő, Horváth and Szyszkowicz (1994).

4.5 Sequentially estimated normal empirical processes under contiguous measures.

In this section we study the sequentially estimated normal empirical process $\hat{\beta}_n(x, t)$ as in (3.1.1), as well as the process $\hat{\hat{\beta}}_n(x, t)$ as in (3.3.11), under the sequence of contiguous alternatives of H_1 , as given by (4.3.1)-(4.3.3) in Section 4.3.

Let $X_1, X_2, \dots$ be an independent sequence of random variables with common distribution function given by

$$F(x; \theta) = \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^x e^{-(t-\mu)^2/2\sigma^2} dt,$$

where $\theta = (\mu, \sigma^2)$ is unknown. Consider the null hypothesis

$$H_0 : X_i, 1 \leq i \leq n, \text{ have the same distribution } F(\cdot; \theta), \text{ and } \theta = \theta_0,$$

with $\theta_0 = (\mu_0, \sigma_0^2)$ denoting the true value of the unknown parameter θ , and the alternative hypotheses

$$H_{1n} : X_i, 1 \leq i \leq n, \text{ have the respective distribution functions } F_{in},$$

where we assume that the F_{in} 's are absolutely continuous with respect to the distribution function $F(\cdot; \theta_0)$ and are parametrized as in (4.3.1)-(4.3.3).

Define the centered log-likelihood ratio process $\mathcal{L}_n(\cdot)$ as in (4.3.6), namely,

$$\mathcal{L}_n(t) = \sum_{i=1}^{[nt]} \left\{ \log \frac{dF_{in}}{dF} (F^{-1}(U_i; \theta_0)) - E \log \frac{dF_{in}}{dF} (F^{-1}(U_i; \theta_0)) \right\},$$

where the U_i 's, $i = 1, 2, \dots$ are i.i.d. uniform-(0,1) random variables.

Let

$$(4.5.1) \quad d(s, t) = \int_0^t \int_0^s h(\tau, u) du d\tau \\ - \left\{ \int_0^t \int_0^1 l(F^{-1}(u; \theta_0), \theta_0) h(\tau, u) du d\tau \right\} \cdot \nabla_{\theta} F(x; \theta_0)^T, \quad 0 \leq s, t \leq 1.$$

We have,

Theorem 4.5.1. *We assume that under H_0 , $X_1, \dots, X_n$, are independent $N(\mu, \sigma^2)$ r.v's with $\theta = (\mu, \sigma^2)$ unknown, and $\theta_0 = (\mu_0, \sigma_0^2)$ denoting its true value. Let $\hat{\beta}_n(\cdot, \cdot)$ be the sequentially estimated normal empirical process defined as in (3.1.1), and suppose that $\{\hat{\theta}_{[nt]}\}$ is as in (3.2.1), and Let $q \in \mathcal{Q}$ be such that*

$$I_0(q, c) < \infty \text{ for all } c > 0.$$

Then, under H_1 , we have, as $n \rightarrow \infty$,

$$(4.5.2) \quad \hat{\beta}_n(x, t)/q(t) \xrightarrow{\mathcal{D}} G(x, t)/q(t) + d(F(x; \theta_0), t)/q(t)$$

in $D(\mathbb{R} \times [0, 1])$, where $G(\cdot, \cdot)$ is as in (3.2.8).

Let

$$(4.5.3) \quad e(s, t) = d(s, t) - td(s, 1), \quad 0 \leq s, t \leq 1,$$

where $d(s, t)$ is defined by (4.5.1).

For the process $\{\hat{\beta}_n(x, t) | 0 \leq t \leq 1, x \in \mathbb{R}\}$, defined by (3.3.11) we have

Theorem 4.5.2. *We assume that under H_0 , $X_1, \dots, X_n$, are independent $N(\mu, \sigma^2)$ r.v.'s, with $\theta = (\mu, \sigma^2)$ unknown and $\theta_0 = (\mu_0, \sigma_0^2)$ denoting its true value. Suppose that, $\{\hat{\theta}'_{[nt]}\}$ and $\{\hat{\theta}''_{[nt]}\}$ are given by (3.3.12) and (3.3.13), and let $\hat{\beta}_n(\cdot, \cdot)$ be the sequentially estimated normal empirical process, tied down at $t=1$, defined as in (3.3.11). Let $q \in \mathcal{Q}^*$ be such that*

$$I_{0,1}(q, c) < \infty \text{ for all } c > 0.$$

Then, under H_1 , we have, as $n \rightarrow \infty$,

$$(4.5.4) \quad \hat{\beta}_n(x, t)/q(t) \xrightarrow{D} \Psi(x, t)/q(t) + e(F(x; \theta_0), t)/q(t)$$

in $D(\mathbb{R} \times [0, 1])$, where $\Psi(\cdot, \cdot)$ is defined by (3.3.17).

For the proof of our results we need to prove several lemmas.

Lemma 4.5.1. *Let $m \geq 1$ be an integer. Let χ_n^2 be a chi-squared random variable with n degrees of freedom, where $n \geq 2m + 1$. Then,*

$$E(1/\chi_n^2)^m = 1/(n-2)(n-4)\dots(n-2m).$$

Proof. The density function of χ_n^2 is given by

$$f(x|n) = \frac{1}{\Gamma(\frac{n}{2})2^{\frac{n}{2}}} x^{\frac{n}{2}-1} e^{-x/2}, \quad x > 0.$$

Hence,

$$\begin{aligned} E(1/\chi_n^2)^m &= \frac{1}{\Gamma(\frac{n}{2})2^{\frac{n}{2}}} \int_0^\infty x^{-m} x^{\frac{n}{2}-1} e^{-x/2} dx \\ (4.5.5) \quad &= \frac{1}{\Gamma(\frac{n}{2})2^{\frac{n}{2}}} \int_0^\infty x^{(\frac{n-2m}{2})-1} e^{-x/2} dx = \frac{\Gamma(\frac{n-2m}{2})2^{\frac{n-2m}{2}}}{\Gamma(\frac{n}{2})2^{\frac{n}{2}}} = \frac{\Gamma(\frac{n-2m}{2})}{\Gamma(\frac{n}{2})} 2^{-m}. \end{aligned}$$

Now, since $\Gamma(\alpha + 1) = \alpha\Gamma(\alpha)$, for all $\alpha > 0$, we have

$$\begin{aligned}\Gamma\left(\frac{n}{2}\right) &= \Gamma\left(\frac{n-2}{2} + 1\right) = \frac{n-2}{2}\Gamma\left(\frac{n-2}{2}\right) \\ &= \frac{n-2}{2}\Gamma\left(\frac{n-4}{2} + 1\right) = \frac{n-2}{2}\frac{n-4}{2}\Gamma\left(\frac{n-4}{2}\right) \\ &= \frac{n-2}{2}\frac{n-4}{2}\cdots\frac{n-2m}{2} = \frac{(n-2)(n-4)\cdots(n-2m)}{2^m}\Gamma\left(\frac{n-2m}{2}\right),\end{aligned}$$

and hence the result follows by (4.5.5). ■

Lemma 4.5.2. *Assume H_0 to hold. Under the conditions of Theorem 4.5.1, for any $x \in \mathbb{R}$ and $0 \leq t_1, t_2 \leq 1$, we have, as $n \rightarrow \infty$,*

$$E\left(\hat{\beta}_n(x, t_1)\mathcal{L}_n(t_2)\right) \longrightarrow d(F(x; \theta_0), t_1 \wedge t_2),$$

Proof. By (3.2.21) we have

$$\begin{aligned}(4.5.6) \quad E\left(\hat{\beta}_n(x, t_1)\mathcal{L}_n(t_2)\right) &= E(\beta_n(x, t_1)\mathcal{L}_n(t_2)) \\ &\quad + n^{-1/2}[nt_1]^{1/2}E\left([nt_1]^{1/2}(\hat{\theta}_{[nt_1]} - \theta_0) \cdot (\nabla_{\theta}F(x; \theta_0)^{\top} - \nabla_{\theta}F(x; \hat{\theta}_{[nt_1]}))^{\top}\mathcal{L}_n(t_2)\right) \\ &\quad - n^{-1/2}[nt_1]^{1/2}E\left([nt_1]^{1/2}(\hat{\theta}_{[nt_1]} - \theta_0)\mathcal{L}_n(t_2)\right) \cdot \nabla_{\theta}F(x; \theta_0)^{\top} \\ &:= \Delta_n^{(1)}(x, t_1, t_2) + \Delta_n^{(2)}(x, t_1, t_2) - \Delta_n^{(3)}(x, t_1, t_2).\end{aligned}$$

By Lemma 4.4.B, as $n \rightarrow \infty$,

$$(4.5.7) \quad \Delta_n^{(1)}(x, t_1, t_2) \longrightarrow \int_0^{t_1 \wedge t_2} \int_0^{F(x; \theta_0)} h(\tau, u) du d\tau.$$

By (4.3.7), $\mathcal{L}_n(t) = \sum_{i=1}^{[nt]} n^{-1/2} h_n(\frac{i}{n}, U_i) + R_n(\frac{i}{n}, U_i)$, with

$$(4.5.8) \quad |R_n(\frac{i}{n}, U_i)| \leq \frac{C}{4n} \left(g_n^2(\frac{i}{n}, U_i) + E g_n^2(\frac{i}{n}, U_i) \right).$$

On the other hand, by (3.1.2),

$$(4.5.9) \quad \begin{aligned} \Delta_n^{(3)}(x, t_1, t_2) &= E \left(n^{-1/2} \sum_{i=1}^{[nt_1]} l(X_i, \theta_0) \mathcal{L}_n(t_2) \right) \cdot \nabla_\theta F(x; \theta_0)^\top \\ &\quad - n^{-1/2} [nt_1]^{1/2} E(\varepsilon_{1n}(t_1) \mathcal{L}_n(t_2)) \cdot \nabla_\theta F(x; \theta_0)^\top. \end{aligned}$$

By (3.2.2), (4.3.7) and (4.5.8),

$$(4.5.10) \quad \begin{aligned} &E \left(n^{-1/2} \sum_{i=1}^{[nt_1]} l(X_i, \theta_0) \mathcal{L}_n(t_2) \right) \\ &= \frac{1}{n} \sum_{i=1}^{[nt_1]} \sum_{j=1}^{[nt_2]} E \left(l(F^{-1}(U_i; \theta_0), \theta_0) h_n(\frac{j}{n}, U_j) \right) \cdot \nabla_\theta F(x; \theta_0)^\top \\ &\quad + n^{-1/2} \sum_{i=1}^{[nt_1]} \sum_{j=1}^{[nt_2]} E \left(l(F^{-1}(U_i; \theta_0), \theta_0) R_n(\frac{j}{n}, U_j) \right) \cdot \nabla_\theta F(x; \theta_0)^\top \\ &:= \left\{ \Delta_{n,1}^{(3)}(t_1, t_2) + \Delta_{n,2}^{(3)}(t_1, t_2) \right\} \cdot \nabla_\theta F(x; \theta_0)^\top. \end{aligned}$$

By independence and (1.2.2)(ii),

$$(4.5.11) \quad \begin{aligned} \Delta_{n,1}^{(3)}(t_1, t_2) &= \frac{1}{n} \sum_{k=1}^{[n(t_1 \wedge t_2)]} E \left(l(F^{-1}(U_k; \theta_0), \theta_0) h_n(\frac{k}{n}, U_k) \right) \\ &= \frac{1}{n} \sum_{k=1}^{[n(t_1 \wedge t_2)]} \int_0^1 l(F^{-1}(u; \theta_0), \theta_0) h_n(\frac{k}{n}, u) du \\ &\longrightarrow \int_0^{t_1 \wedge t_2} \int_0^1 l(F^{-1}(u; \theta_0), \theta_0) h(\tau, u) du d\tau, \text{ as } n \rightarrow \infty. \end{aligned}$$

On the other hand, by independence, (1.2.2)(ii), (4.3.3) and (4.5.8), we have

$$\begin{aligned}
 \Delta_{n,2}^{(3)}(t_1, t_2) &\leq Cn^{-1/2} \frac{1}{4n} \sum_{k=1}^{[n(t_1 \wedge t_2)]} E \left(|l(F^{-1}(U_k; \theta_0), \theta_0)| g_n^2\left(\frac{k}{n}, U_k\right) \right) \\
 (4.5.12) \quad &= Cn^{-1/2} \frac{1}{4n} \sum_{k=1}^{[n(t_1 \wedge t_2)]} \int_0^1 |l(F^{-1}(u; \theta_0), \theta_0)| g_n^2\left(\frac{k}{n}, u\right) du \\
 &\longrightarrow 0, \text{ as } n \rightarrow \infty.
 \end{aligned}$$

Hence, as $n \rightarrow \infty$,

$$(4.5.13) \quad E \left(n^{-1/2} \sum_{i=1}^{[nt_1]} l(X_i, \theta_0) \mathcal{L}_n(t_2) \right) \rightarrow \int_0^{t_1 \wedge t_2} \int_0^1 l(F^{-1}(u; \theta_0), \theta_0) h(\tau, u) du d\tau.$$

By (3.2.3), (4.3.7) and (4.5.8),

$$\begin{aligned}
 (4.5.14) \quad &n^{-1/2} [nt_1]^{1/2} E(\varepsilon_{1n}(t_1) \mathcal{L}_n(t_2)) \\
 &= n^{-1} [nt_1]^{-1/2} \left[E \left(\sum_{j=1}^{[nt_2]} h_n\left(\frac{j}{n}, U_j\right) \left(\sum_{i=1}^{[nt_1]} (X_i - \mu_0) \right)^2 \right) \right] \frac{\partial F}{\partial(\sigma^2)}(x; \theta_0) \\
 &\quad + n^{-1/2} [nt_1]^{-1/2} \left[E \left(\sum_{j=1}^{[nt_2]} R_n\left(\frac{j}{n}, U_j\right) \left(\sum_{i=1}^{[nt_1]} (X_i - \mu_0) \right)^2 \right) \right] \frac{\partial F}{\partial(\sigma^2)}(x; \theta_0) \\
 &:= \left\{ \Delta_{n,3}^{(3)}(t_1, t_2) + \Delta_{n,4}^{(3)}(t_1, t_2) \right\} \frac{\partial F}{\partial(\sigma^2)}(x; \theta_0).
 \end{aligned}$$

By independence,

$$\begin{aligned}
 \Delta_{n,3}^{(3)}(t_1, t_2) &= n^{-1} [nt_1]^{-1/2} E \sum_{j=1}^{[nt_2]} h_n\left(\frac{j}{n}, U_j\right) \left(\sum_{i=1}^{[nt_1]} (X_i - \mu_0)^2 \right. \\
 (4.5.15) \quad &\quad \left. + \sum_{i=1}^{[nt_1]} \sum_{k \neq i} (X_i - \mu_0)(X_k - \mu_0) \right)
 \end{aligned}$$

$$\begin{aligned}
&= n^{-1}[nt_1]^{-1/2} \sum_{j=1}^{[nt_2]} \sum_{i=1}^{[nt_1]} E(h_n(\frac{j}{n}, U_j)(X_i - \mu_0)^2) \\
&= n^{-1}[nt_1]^{-1/2} \sum_{j=1}^{[n(t_1 \wedge t_2)]} \int_0^1 (F^{-1}(u; \theta_0) - \mu_0)^2 h_n(\frac{j}{n}, u) du \\
&\longrightarrow 0, \text{ as } n \rightarrow \infty.
\end{aligned}$$

Also, arguing as in (4.5.12), we have

$$\begin{aligned}
\Delta_{n,4}^{(3)}(t_1, t_2) &\leq n^{-1/2}[nt_1]^{-1/2} \frac{\sigma_0^2}{4n} \sum_{i=1}^{[n(t_1 \wedge t_2)]} \int_0^1 (F^{-1}(u; \theta_0) - \mu_0)^2 g_n^2(\frac{i}{n}, u) du \\
(4.5.16) \quad &+ n^{-1/2}[nt_1]^{-1/2} \frac{\sigma_0^2}{4n} \sum_{i=1}^{[n(t_1 \wedge t_2)]} \int_0^1 g_n^2(\frac{i}{n}, u) du \longrightarrow 0, \text{ as } n \rightarrow \infty.
\end{aligned}$$

Thus, as $n \rightarrow \infty$,

$$(4.5.17) \quad \Delta_n^{(3)}(x, t_1, t_2) \longrightarrow \left\{ \int_0^{t_1 \wedge t_2} \int_0^1 l(F^{-1}(u; \theta_0), \theta_0) h(\tau, u) du d\tau \right\} \cdot \nabla_\theta F(x; \theta_0)^\top.$$

It remains to show that, as $n \rightarrow \infty$,

$$(4.5.18) \quad \Delta_n^{(2)}(x, t_1, t_2) \longrightarrow 0, \text{ as } n \rightarrow \infty.$$

To this end we note that, by the Cauchy-Schwartz inequality,

$$\begin{aligned}
(4.5.19) \quad \Delta_n^{(2)}(x, t_1, t_2) &= \\
&n^{-1/2}[nt_1]^{1/2} E \left([nt_1]^{1/2} (\hat{\theta}_{[nt_1]} - \theta_0) \cdot (\nabla_\theta F(x; \theta_0)^\top - \nabla_\theta F(x; \tilde{\theta}_{[nt_1]})^\top \mathcal{L}_n(t_2)) \right) \\
&\leq n^{-1/2}[nt_1]^{1/2} \left(E \left| [nt_1]^{1/2} (\hat{\theta}_{[nt_1]} - \theta_0) \cdot (\nabla_\theta F(x; \theta_0)^\top - \nabla_\theta F(x; \tilde{\theta}_{[nt_1]})^\top \right|^2 \right)^{1/2} \\
&\quad \times \left(E(\mathcal{L}_n(t_2))^2 \right)^{1/2} \\
&:= n^{-1/2}[nt_1]^{1/2} \Delta_{n,1}^{(2)}(x, t_1) \left(E(\mathcal{L}_n(t_2))^2 \right)^{1/2}.
\end{aligned}$$

We also have that, for each $k \geq 1$,

(4.5.20)

$$\nabla_{\theta} F(x; \theta_0) - \nabla_{\theta} F(x; \tilde{\theta}_k) = \left(\frac{\partial}{\partial \mu} (F(x; \theta_0) - F(x; \tilde{\theta}_k)), \frac{\partial}{\partial \sigma^2} (F(x; \theta_0) - F(x; \tilde{\theta}_k)) \right).$$

Applying a one term Taylor expansion to the first coordinate above we get

$$\begin{aligned} \frac{\partial}{\partial \mu} (F(x; \theta_0) - F(x; \tilde{\theta}_k)) &= \frac{\partial^2 F}{\partial \mu^2} (x; \xi_k) (\tilde{\theta}_k^{(1)} - \mu_0) + \frac{\partial^2 F}{\partial \sigma^2 \partial \mu} (x; \xi_k) (\tilde{\theta}_k^{(2)} - \sigma_0^2) \\ (4.5.21) \quad &= (\tilde{\theta}_k - \theta_0) \cdot \nabla_{\theta} \frac{\partial F}{\partial \mu} (x; \xi_k)^{\top}, \end{aligned}$$

where $\tilde{\theta}_k = (\tilde{\theta}_k^{(1)}, \tilde{\theta}_k^{(2)})$, and $\xi_k = (\xi_k^{(1)}, \xi_k^{(2)})$ is in the straight line joining $\tilde{\theta}_k$ and θ_0 , i.e.,

$$(4.5.22) \quad \xi_k^{(1)} \in (\mu_0 \wedge \tilde{\theta}_k^{(1)}, \mu_0 \vee \tilde{\theta}_k^{(1)}) \text{ and } \xi_k^{(2)} \in (\mu_0 \wedge \tilde{\theta}_k^{(2)}, \mu_0 \vee \tilde{\theta}_k^{(2)}).$$

Similarly,

$$\begin{aligned} \frac{\partial}{\partial \sigma^2} (F(x; \theta_0) - F(x; \tilde{\theta}_k)) &= \frac{\partial^2 F}{\partial \mu \partial \sigma^2} (x; \eta_k) (\tilde{\theta}_k^{(1)} - \mu_0) + \frac{\partial^2 F}{\partial (\sigma^2)^2} (x; \eta_k) (\tilde{\theta}_k^{(2)} - \sigma_0^2) \\ (4.5.23) \quad &= (\tilde{\theta}_k - \theta_0) \cdot \nabla_{\theta} \frac{\partial F}{\partial \sigma^2} (x; \eta_k)^{\top}, \end{aligned}$$

with $\eta_k = (\eta_k^{(1)}, \eta_k^{(2)})$ also satisfying (4.5.22).

We may assume, without loss of generality, that $\xi_k = \eta_k$ for all k .

From (4.5.20), (4.5.21) and (4.5.23), we have

(4.5.24)

$$\begin{aligned} &k^{1/2} (\tilde{\theta}_k - \theta_0) \cdot \left(\nabla_{\theta} F(x; \theta_0)^{\top} - \nabla_{\theta} F(x; \tilde{\theta}_k)^{\top} \right) \\ &= k^{1/2} (\tilde{\theta}_k - \theta_0) \cdot \left((\tilde{\theta}_k - \theta_0) \cdot \nabla_{\theta} \frac{\partial F}{\partial \mu} (x; \xi_k)^{\top}, (\tilde{\theta}_k - \theta_0) \cdot \nabla_{\theta} \frac{\partial F}{\partial \sigma^2} (x; \xi_k)^{\top} \right)^{\top} \end{aligned}$$

$$= k^{1/2}(\hat{\theta}_k - \theta_0)F''_{\theta\theta}(x; \xi_k)(\hat{\theta}_k - \theta_0)^\top,$$

$$\text{where } F''_{\theta\theta}(x; \alpha) = \begin{pmatrix} \frac{\partial^2 F}{\partial \mu^2}(x; \alpha) & \frac{\partial^2 F}{\partial \sigma^2 \partial \mu}(x; \alpha) \\ \frac{\partial^2 F}{\partial \sigma^2 \partial \mu}(x; \alpha) & \frac{\partial^2 F}{\partial (\sigma^2)^2}(x; \alpha) \end{pmatrix}.$$

It can be easily shown, that

$$\begin{aligned} \left| k^{1/2}(\hat{\theta}_k - \theta_0)F''_{\theta\theta}(x; \xi_k)(\hat{\theta}_k - \theta_0)^\top \right| &\leq 4k^{1/2} \|\hat{\theta}_k - \theta_0\| \|\hat{\theta}_k - \theta_0\| \|F''_{\theta\theta}(x; \xi_k)\| \\ (4.5.25) \qquad \qquad \qquad &\leq 4k^{1/2} \|\hat{\theta}_k - \theta_0\|^2 \|F''_{\theta\theta}(x; \xi_k)\|, \end{aligned}$$

where $\left\| \begin{pmatrix} \alpha_1 & \alpha_2 \\ \alpha_3 & \alpha_4 \end{pmatrix} \right\| = \max_{1 \leq i \leq 4} |\alpha_i|$, and the last inequality is due to (3.2.23).

Hence, by (4.5.19), (4.5.24) and (4.5.25), and the Cauchy-Schwartz inequality,

$$\begin{aligned} \Delta_{n,1}^{(2)}(x, t_1) &\leq 4 \left[E \left([nt_1] \|\hat{\theta}_{[nt_1]} - \theta_0\|^4 \|F''_{\theta\theta}(x; \xi_{[nt_1]})\|^2 \right) \right]^{1/2} \\ (4.5.26) \qquad \qquad &\leq 4[nt_1]^{1/2} \left[E \|\hat{\theta}_{[nt_1]} - \theta_0\|^8 \right]^{1/4} \left[E \|F''_{\theta\theta}(x; \xi_{[nt_1]})\|^4 \right]^{1/4}. \end{aligned}$$

By (3.1.2), (3.2.2) and (3.2.3), we have, for each $k \geq 1$,

$$\hat{\theta}_k - \theta_0 = \left(\frac{1}{k} \sum_{i=1}^k (X_i - \mu_0), \frac{1}{k} \sum_{i=1}^k ((X_i - \mu_0)^2 - \sigma_0^2) - (\bar{X}_k - \mu_0)^2 \right).$$

Hence,

$$\begin{aligned} k^{1/2} \|\hat{\theta}_k - \theta_0\|^8 &\leq k^{1/2} \max \left\{ \left| \frac{1}{k} \sum_{i=1}^k (X_i - \mu_0) \right|^8, \right. \\ (4.5.27) \qquad \qquad &\left. \left| \frac{1}{k} \sum_{i=1}^k ((X_i - \mu_0)^2 - \sigma_0^2) - (\bar{X}_k - \mu_0)^2 \right|^8 \right\}. \end{aligned}$$

We have,

$$(4.5.28) \quad k^{1/2} \left[E \left(\frac{1}{k} \sum_{i=1}^k (X_i - \mu_0) \right)^8 \right]^{1/4} = k^{-3/2} \left[E \left(\sum_{i=1}^k (X_i - \mu_0) \right)^8 \right]^{1/4}.$$

Let $S_k = \sum_{i=1}^k Y_i$, with $Y_i = X_i - \mu_0$. S_k is a partial sum of i.i.d random variables with mean zero, and thus,

$$E(S_k)^8 = \sum E(Y_i Y_l Y_m Y_n Y_p Y_q Y_s),$$

with i, j, n, l, m, p, q, s ranging independently from 1 to k . Since $EY = 0$, it follows, by independence, that any summand above vanishes if any index occurs an odd number of times in the product. This leaves terms of the form

- (1) $E(Y_i)^8$,
- (2) $E(Y_i)^6 E(Y_j)^2$,
- (3) $E(Y_i)^4 E(Y_j)^4$,
- (4) $E(Y_i)^4 E(Y_j)^2 E(Y_m)^2$ and
- (5) $E(Y_i)^2 E(Y_j)^2 E(Y_m)^2 E(Y_l)^2$.

The number of terms of each kind is bounded above by a positive constant, independent of k , times the number of permutations of different indices in each kind from the group of k indices, e.g., the number of terms of the form

$$\begin{aligned} E(Y_i)^6 E(Y_j)^2 &\leq C(\text{# of permutations of two objects from a group of } k) \\ &= Ck(k-1), \end{aligned}$$

with C independent of k . Thus,

$$\begin{aligned} E(S_k)^8 &\leq k + M_1 k(k-1) + M_2 k(k-1)(k-2) + M_3 k(k-1)(k-2)(k-3) \\ &\leq Mk^4, \end{aligned}$$

where M is a positive constant independent of k .

Hence, by (4.5.28), we have

$$(4.5.29) \quad k^{-3/2} \left[E \left(\sum_{i=1}^k (X_i - \mu_0) \right)^8 \right]^{1/4} = O(k^{-1/2}) \rightarrow 0, \text{ as } k \rightarrow \infty.$$

Also, by Minkowski's inequality, putting $Z_i = (X_i - \mu_0)^2 - \sigma_0^2$,

$$(4.5.30) \quad \begin{aligned} & k^{1/2} \left[E \left| \frac{1}{k} \sum_{i=1}^k ((X_i - \mu_0)^2 - \sigma_0^2) - (\bar{X}_k - \mu_0)^2 \right|^8 \right]^{1/4} \\ & \leq k^{1/2} \left\{ \left[E \left(\frac{1}{k} \sum_{i=1}^k Z_i \right)^8 \right]^{1/8} + \left[E (\bar{X}_k - \mu_0)^{16} \right]^{1/8} \right\}^2 \\ & = k^{1/2} \left[E \left(\frac{1}{k} \sum_{i=1}^k Z_i \right)^8 \right]^{1/4} + 2k^{1/2} \left[E \left(\frac{1}{k} \sum_{i=1}^k Z_i \right)^8 \right]^{1/8} \left[E (\bar{X}_k - \mu_0)^{16} \right]^{1/8} \\ & \quad + k^{1/2} \left[E (\bar{X}_k - \mu_0)^{16} \right]^{1/4}. \end{aligned}$$

Now, by the same argument used to show (4.5.29), we have

$$(4.5.31) \quad k^{1/2} \left[E \left(\frac{1}{k} \sum_{i=1}^k Z_i \right)^8 \right]^{1/4} \rightarrow 0, \text{ as } k \rightarrow \infty.$$

Similarly, we have

$$(4.5.32) \quad \begin{aligned} & k^{1/2} \left[E (\bar{X}_k - \mu_0)^{16} \right]^{1/4} = k^{1/2} \left[E \left(\frac{1}{k} \sum_{i=1}^k (X_i - \mu_0) \right)^{16} \right]^{1/4} \\ & = k^{1/2} \frac{1}{k^4} \left[E \left(\sum_{i=1}^k (X_i - \mu_0) \right)^{16} \right]^{1/4} \\ & = k^{-7/2} O(k^2) = O(k^{-3/2}) \rightarrow 0, \text{ as } k \rightarrow \infty. \end{aligned}$$

By using the same argument above,

$$\begin{aligned}
 (4.5.33) \quad & k^{1/2} \left[E \left(\frac{1}{k} \sum_{i=1}^k Z_i \right)^8 \right]^{1/8} \left[E (\bar{X}_k - \mu_0)^{16} \right]^{1/8} \\
 & = k^{-1/2} \left[E \left(\sum_{i=1}^k Z_i \right)^8 \right]^{1/8} \left[E (\bar{X}_k - \mu_0)^{16} \right]^{1/8} \\
 & \leq k^{-1/2} O(k^{1/2}) O(k^{-1}) = O(k^{-1}) \rightarrow 0, \text{ as } k \rightarrow \infty.
 \end{aligned}$$

Hence, from (4.5.28) we can see that (4.5.29), (4.5.30), (4.5.31), (4.5.32) and (4.5.33) imply that

$$(4.5.34) \quad [nt_1]^{1/2} E \|\hat{\theta}_{[nt_1]} - \theta_0\|^8 \rightarrow 0, \text{ as } (nt_1) \rightarrow \infty.$$

Now, we proceed to show that

$$(4.5.35) \quad \lim_{k \rightarrow \infty} E \|F''_{\theta\theta}(x; \xi_k)\|^4 < \infty.$$

To this end, we note that, since, for any $\theta = (\mu, \sigma^2)$,

$$F(x; \theta) = \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^x e^{-(t-\mu)^2/2\sigma^2} dt, \quad x \in \mathbb{R},$$

we have

$$\begin{aligned}
 (4.5.36) \quad (a) \quad & \frac{\partial^2 F}{\partial \mu^2}(x; \theta) = \frac{(x - \mu)}{\sqrt{2\pi\sigma^3}} e^{-(x-\mu)^2/2\sigma^2}, \\
 (b) \quad & \frac{\partial^2 F}{\partial \sigma^2 \partial \mu}(x; \theta) = \frac{-1}{2\sqrt{2\pi\sigma^3}} e^{-(x-\mu)^2/2\sigma^2} \left(1 - \frac{(x - \mu)^2}{\sigma^2} \right), \text{ and} \\
 (c) \quad & \frac{\partial^2 F}{\partial (\sigma^2)^2}(x; \theta) = \frac{(x - \mu)}{2\sqrt{2\pi\sigma^5}} e^{-(x-\mu)^2/2\sigma^2} \left(3 - \frac{(x - \mu)^2}{\sigma^2} \right).
 \end{aligned}$$

The function $\varphi(x) = (x - \mu)e^{-(x-\mu)^2/2\sigma^2}$, $x \in \mathbb{R}$, has a maximum at $x = \mu + \sigma$ equal to $\sigma e^{-1/2}$. Hence, if $\xi_k = (\xi_k^{(1)}, \xi_k^{(2)})$,

$$(4.5.37) \quad \left| \frac{\partial^2 F}{\partial \mu^2}(x; \xi_k) \right|^4 \leq \left[\frac{(\sqrt{2\pi})^{-1}}{(\xi_k^{(2)})^{3/2}} (\xi_k^{(2)})^{1/2} e^{-1/2} \right]^4 \\ \leq \frac{1}{(\xi_k^{(2)})^4}.$$

By (4.5.22) we have that $\xi_k^{(2)} \in (\sigma_0^2 \wedge \tilde{\theta}_k^{(2)}, \sigma_0^2 \vee \tilde{\theta}_k^{(2)})$, and, by (3.2.1) and (3.2.23), we have that $\tilde{\theta}_k^{(2)} \in (\sigma_0^2 \wedge S_k^2, \sigma_0^2 \vee S_k^2)$. Consequently, $\xi_k^{(2)} \in (\sigma_0^2 \wedge S_k^2, \sigma_0^2 \vee S_k^2)$.

Thus,

$$0 < \left(\frac{1}{\sigma_0^2 \vee S_k^2} \right)^m < \frac{1}{(\xi_k^{(2)})^m} < \left(\frac{1}{\sigma_0^2 \wedge S_k^2} \right)^m,$$

for any integer $m \geq 1$. However, $1/(\sigma_0^2 \wedge S_k^2) = 1/\sigma_0^2 \vee 1/S_k^2$; and thus, $1/(\xi_k^{(2)})^m < 1/(\sigma_0^2)^m \vee 1/(S_k^2)^m$. This implies,

$$(4.5.38) \quad E\left(\frac{1}{\xi_k^{(2)}}\right)^m \leq E\left(\frac{1}{(\sigma_0^2)^m} \vee \frac{1}{(S_k^2)^m}\right) \\ = \int_{\{1/\sigma_0^2 \geq 1/S_k^2\}} \left(\frac{1}{(\sigma_0^2)^m} \vee \frac{1}{(S_k^2)^m}\right) dP + \int_{\{1/\sigma_0^2 < 1/S_k^2\}} \left(\frac{1}{(\sigma_0^2)^m} \vee \frac{1}{(S_k^2)^m}\right) dP \\ \leq \frac{1}{(\sigma_0^2)^m} + E\left[\frac{1}{(S_k^2)^m}\right].$$

Since $\frac{k}{\sigma_0^2} S_k^2 = \sum_{i=1}^k (X_i - \bar{X}_k)^2 / \sigma_0^2$, we have

$$E\left(\frac{1}{S_k^2}\right)^m = \frac{k^m}{\sigma_0^{2m}} E\left(\frac{1}{\sum_{i=1}^k (X_i - \bar{X}_k)^2 / \sigma_0^2}\right)^m;$$

and, since $\sum_{i=1}^k (X_i - \bar{X}_k)^2 / \sigma_0^2$ has a chi-squared distribution with $(k-1)$ degrees of freedom, we can apply Lemma 4.5.1 to get that, for $k \geq 2(m+1)$,

(4.5.39)

$$E\left(\frac{1}{S_k^2}\right)^m = \frac{k^m}{\sigma_0^{2m}} \frac{1}{(k-3)(k-5)\dots(k-(2m+1))} \rightarrow \frac{1}{\sigma_0^{2m}}, \text{ as } k \rightarrow \infty.$$

Hence, from (4.5.37), (4.5.38), and (4.5.39) with $m = 4$, we get

$$(4.5.40) \quad \lim_{k \rightarrow \infty} \left[E \left| \frac{\partial^2}{\partial \mu^2} F(x; \xi_k) \right|^4 \right]^{1/4} < \infty.$$

Consider now the function $\psi(x) = (x - \mu)^2 e^{-(x-\mu)^2/2\sigma^2}$, $x \in \mathbb{R}$. This function has a maximum at $x = \mu + \sqrt{2}\sigma$, equal to $2\sigma^2 e^{-1}$. Hence, from (4.5.36)(b), we have

$$\begin{aligned} \left| \frac{\partial^2 F}{\partial \sigma^2 \partial \mu}(x; \xi_k) \right|^4 &\leq \left[\frac{1}{2\sqrt{2\pi}} \left(\frac{1}{(\xi_k^{(2)})^{3/2}} + \frac{2\xi_k^{(2)} e^{-1}}{(\xi_k^{(2)})^{5/2}} \right) \right]^4 \\ &\leq \left(\frac{1}{(\xi_k^{(2)})^{3/2}} \right)^4 = \frac{1}{(\xi_k^{(2)})^6}. \end{aligned}$$

Thus, by (4.5.38) and (4.5.39) with $m = 6$, we have

$$(4.5.41) \quad \lim_{k \rightarrow \infty} \left[E \left| \frac{\partial^2 F}{\partial \sigma^2 \partial \mu}(x; \xi_k) \right|^4 \right]^{1/4} < \infty.$$

Finally, if we take the function $\vartheta(x) = (x - \mu)^3 e^{-(x-\mu)^2/2\sigma^2}$, $x \in \mathbb{R}$, we have that this function has a maximum at $x = \mu + \sqrt{3}\sigma$, equal to $3\sqrt{3}\sigma^3 e^{-3/2}$. Hence, from (4.5.36)(c) (recalling that $\varphi(x) = (x - \mu)e^{-(x-\mu)^2/2\sigma^2}$, has a maximum equal

to $\sigma e^{-1/2}$),

$$\begin{aligned} \left| \frac{\partial^2 F}{\partial(\sigma^2)^2}(x; \xi_k) \right|^4 &\leq \left[\frac{1}{2\sqrt{2\pi}} \left(\frac{3}{2} \frac{(\xi_k^{(2)})^{1/2} e^{-1/2}}{(\xi_k^{(2)})^{5/2}} + 3\sqrt{3} \frac{(\xi_k^{(2)})^{3/2} e^{-3/2}}{(\xi_k^{(2)})^{7/2}} \right) \right]^4 \\ &\leq \frac{1}{(\xi_k^{(2)})^8}. \end{aligned}$$

Thus, by (4.5.38) and (4.5.39), with $m = 8$, we have that

$$(4.5.42) \quad \lim_{k \rightarrow \infty} \left[E \left| \frac{\partial^2}{\partial(\sigma^2)^2} F(x; \xi_k) \right|^4 \right]^{1/4} < \infty.$$

Now, from (4.5.40), (4.5.41) and (4.5.42) we get (4.5.35), while (4.5.34) and (4.5.35) in turn imply

$$(4.5.43) \quad \Delta_{n,1}^{(2)}(x, t_1) \rightarrow 0, \text{ as } (nt_1) \rightarrow \infty.$$

Since, by (4.3.10), $\lim_{n \rightarrow \infty} E(\mathcal{L}_n(t))^2 < \infty$, as $(nt) \rightarrow \infty$, we have, by (4.5.19), that (4.5.18) holds, and thus, the result now follows from (4.5.7), (4.5.17) and (4.5.18). ■

Let $(G, \mathcal{L})$ be the two-dimensional Gaussian process, where G is the Gaussian process in (3.2.8), $\mathcal{L}$ is a Gaussian process with mean zero and covariance function $E(\mathcal{L}(t_1)\mathcal{L}(t_2)) = Q_{\mathcal{L}}(t_1 \wedge t_2)$, where

$$Q_{\mathcal{L}}(t) = \int_0^t \int_0^1 h^2(s, u) du ds,$$

and the covariance function of G and $\mathcal{L}$ is $EG(x, t_1)\mathcal{L}(t_2) = d(F(x; \theta_0), t_1 \wedge t_2)$, where $d(s, t)$ is as in (4.5.1).

Lemma 4.5.3. *Assume the conditions of Theorem 4.5.1. Then, under H_0 , we have, as $n \rightarrow \infty$,*

$$(\hat{\beta}_n(x, t), \mathcal{L}_n(t)) \xrightarrow{\mathcal{D}} (G(x, t), \mathcal{L}(t)) \text{ in } D(\mathbb{R} \times [0, 1]).$$

Proof. From Lemma 4.5.2 we have that all finite-dimensional distributions of the process $(\hat{\beta}_n(x, t), \mathcal{L}_n(t)) \xrightarrow{\mathcal{D}} (G(x, t), \mathcal{L}(t))$ converge to those of the process $(G(x, t), \mathcal{L}(t))$ with covariance function

$$\begin{aligned} EG(x, t_1)\mathcal{L}(t_2) &= \int_0^{t_1 \wedge t_2} \int_0^{F(x; \theta_0)} h(\tau, u) du d\tau \\ &\quad - \left\{ \int_0^{t_1 \wedge t_2} \int_0^1 l(F^{-1}(u; \theta_0), \theta_0) h(\tau, u) du d\tau \right\} \cdot \nabla_{\theta} F(x; \theta_0)^{\top}. \end{aligned}$$

Tightness of $(\hat{\beta}_n(x, t), \mathcal{L}_n(t))$ follows from the weak convergence of $\hat{\beta}_n(x, t)$ (cf. (3.2.45) of Corollary 3.2.1) and that of $\mathcal{L}_n(t)$ (cf. 4.3.12). ■

Proof of Theorem 4.5.1. By Lemma 4.5.3 and the fact that

$$E\left(\frac{\hat{\beta}_n(x, t_1)}{q(t_1)} \mathcal{L}_n(t_2)\right) = \frac{1}{q(t_1)} E(\hat{\beta}_n(x, t_1) \mathcal{L}_n(t_2)) \rightarrow \frac{1}{q(t_1)} d(F(x; \theta_0), t_1 \wedge t_2),$$

as $n \rightarrow \infty$, we have that

$$\left(\frac{\hat{\beta}_n(x, t)}{q(t)}, \mathcal{L}_n(t)\right) \xrightarrow{\mathcal{D}} \left(\frac{G(x, t)}{q(t)}, \mathcal{L}(t)\right) \text{ in } D(\mathbb{R} \times [0, 1]).$$

We proceed to show that the limiting covariance function of the process $\left(\frac{\hat{\beta}_n(x, t)}{q(t)}, \mathcal{L}_n(t)\right)$ is finite, uniformly in t . To this end we note first that, by Fubini's Theorem and the Cauchy-Schwartz inequality,

$$\begin{aligned} \left| \int_0^t \int_0^s h(\tau, u) du d\tau \right| &\leq \int_0^s \left[\left(\int_0^t h^2(\tau, u) d\tau \right)^{1/2} \left(\int_0^t d\tau \right)^{1/2} \right] \\ &\leq t^{1/2} s^{1/2} \left(\int_0^1 \int_0^1 h^2(\tau, u) du d\tau \right)^{1/2}. \end{aligned}$$

It is also clear that

$$\begin{aligned} &\left| \left\{ \int_0^t \int_0^1 l(F^{-1}(u; \theta_0), \theta_0) h(\tau, u) du d\tau \right\} \cdot \nabla_{\theta} F(x; \theta_0)^T \right| \\ &\leq 2 \left\| \int_0^t \int_0^1 l(F^{-1}(u; \theta_0), \theta_0) h(\tau, u) du d\tau \right\| \|\nabla_{\theta} F(x; \theta_0)^T\|. \end{aligned}$$

By the Cauchy-Schwartz inequality,

$$\begin{aligned} &\left\| \int_0^t \int_0^1 l(F^{-1}(u; \theta_0), \theta_0) h(\tau, u) du d\tau \right\| \\ &\leq \int_0^t \left[\int_0^1 \|l(F^{-1}(u; \theta_0), \theta_0)\|^2 du \right]^{1/2} \left[\int_0^1 h^2(\tau, u) du \right]^{1/2} d\tau \\ &\leq \left(\int_0^1 \|l(F^{-1}(u; \theta_0), \theta_0)\|^2 du \right)^{1/2} \int_0^t \left(\int_0^1 h^2(\tau, u) du \right)^{1/2} d\tau \\ &\leq t^{1/2} \left(\int_0^1 \|l(F^{-1}(u; \theta_0), \theta_0)\|^2 du \right)^{1/2} \left(\int_0^1 \int_0^1 h^2(\tau, u) du d\tau \right)^{1/2}. \end{aligned}$$

Hence,

$$\begin{aligned} &\sup_{0 < t \leq 1} \sup_{x \in \mathbb{R}} |d(F(x; \theta_0), t)| / q(t) \\ &\leq \sup_{0 < t \leq 1} \frac{t^{1/2}}{q(t)} \sup_{0 \leq t \leq 1} \left(\int_0^1 \int_0^1 h^2(\tau, u) du d\tau \right)^{1/2} \{1 + 2 \int_0^1 \|l(F^{-1}(u; \theta_0), \theta_0)\|^2 du \\ &\quad \times \sup_{x \in \mathbb{R}} \|\nabla_{\theta} F(x; \theta_0)\|\} < \infty, \end{aligned}$$

due to the fact that if $q \in \mathcal{Q}$ and (4.4.5) holds, it implies $\sup_{0 < t \leq 1} t^{1/2}/q(t) < \infty$.

Now Le Cam's third lemma implies the result. ■

Lemma 4.5.4. Assume H_0 to hold. Under the conditions of Theorem 4.5.2, for any $x \in \mathbb{R}$ and $0 \leq t_1, t_2 \leq 1$, as $n \rightarrow \infty$,

$$E\left(\hat{\beta}_n(x, t_1) \mathcal{L}_n(t_2)\right) \longrightarrow e(F(x; \theta_0), t_1 \wedge t_2),$$

where $e(\cdot, \cdot)$ is defined by (4.5.3).

Proof. By (3.3.25) and (3.3.32), we have

(4.5.44)

$$\begin{aligned} E\left(\hat{\beta}_n(x, t_1) \mathcal{L}_n(t_2)\right) &= E(\beta_n^*(x, t_1) \mathcal{L}_n(t_2) - E(\varphi(t_1) \mathcal{L}_n(t_2)) \cdot \nabla_\theta F(x; \theta_0)^\top) \\ &\quad + \frac{[(n+1)t_1]}{n} \left(1 - \frac{[(n+1)t_1]}{n}\right) n^{1/2} E(\varepsilon'_{3n}(x, t_1) + \varepsilon''_{3n}(x, t_1)) \mathcal{L}_n(t_2) \\ &\quad + \frac{[(n+1)t_1]}{n} n^{-1/2} (n - [(n+1)t_1])^{\frac{1}{2}} E(\varepsilon''_{1n}(t_1) \mathcal{L}_n(t_2)) \cdot \nabla_\theta F(x; \theta_0)^\top \\ &\quad - \left(1 - \frac{[(n+1)t_1]}{n}\right) n^{-1/2} [(n+1)t_1]^{\frac{1}{2}} E(\varepsilon'_{1n}(t_1) \mathcal{L}_n(t_2)) \cdot \nabla_\theta F(x; \theta_0)^\top \\ &:= \Lambda_n^{(1)}(x, t_1, t_2) - \Lambda_n^{(2)}(x, t_1, t_2) + \Lambda_n^{(3)}(x, t_1, t_2) + \Lambda_n^{(4)}(x, t_1, t_2) - \Lambda_n^{(5)}(x, t_1, t_2). \end{aligned}$$

By Lemma 4.4.C, we have, as $n \rightarrow \infty$,

(4.5.45)

$$\Lambda_n^{(1)}(x, t_1, t_2) \rightarrow \int_0^{t_1 \wedge t_2} \int_0^{F(x; \theta_0)} h(\tau, u) du d\tau - t_1 \int_0^{t_2} \int_0^{F(x; \theta_0)} h(\tau, u) du d\tau.$$

Also, by (3.3.32) and (4.5.13),

$$\begin{aligned} \Lambda_n^{(2)}(x, t_1, t_2) &= E\left(n^{-1/2} \sum_{i=1}^{[(n+1)t_1]} l(X_i, \theta_0) \mathcal{L}_n(t_2)\right) \\ &\quad - \frac{[(n+1)t_1]}{n} E\left(n^{-1/2} \sum_{i=1}^n l(X_i, \theta_0) \mathcal{L}_n(t_2)\right), \end{aligned} \quad (4.5.46)$$

which tends to

$$\int_0^{t_1 \wedge t_2} \int_0^1 l(F^{-1}(u; \theta_0), \theta_0) h(\tau, u) du d\tau \\ - t_1 \int_0^{t_2} \int_0^1 l(F^{-1}(u; \theta_0), \theta_0) h(\tau, u) du d\tau, \text{ as } n \rightarrow \infty.$$

Since $\epsilon'_{1n}(t) = \epsilon_{1(n+1)}(t)$, where $\epsilon_1(t)$ is as in (4.5.14), we have, by (4.5.15) and (4.5.16), that

$$(4.5.47) \quad \Lambda_n^{(5)}(x, t_1, t_2) \rightarrow 0, \text{ as } n \rightarrow \infty.$$

By (3.3.28) and symmetry, we have that (4.5.47) implies

$$(4.5.48) \quad \Lambda_n^{(4)}(x, t_1, t_2) \rightarrow 0, \text{ as } n \rightarrow \infty.$$

Similarly, since, by (3.3.19), $n^{-1/2}[nt_1]\epsilon'_{3n}(t)$ is the same as $\Delta_n^{(2)}(x, t_1, t_2)$ in (4.5.6), we have, by (4.5.18) and symmetry, that

$$(4.5.49) \quad \Lambda_n^{(3)}(x, t_1, t_2) \rightarrow 0, \text{ as } n \rightarrow \infty.$$

Thus, the result follows. ■

Let $(\Psi, \mathcal{L})$ be a two-dimensional Gaussian process, where Ψ is the Gaussian process in (3.3.17), $\mathcal{L}$ is a Gaussian process with mean zero and covariance function $E(\mathcal{L}(t_1)\mathcal{L}(t_2)) = Q_{\mathcal{L}}(t_1 \wedge t_2)$, where

$$Q_{\mathcal{L}}(t) = \int_0^t \int_0^1 h^2(s, u) du ds,$$

and the covariance function of Ψ and $\mathcal{L}$ is $E\Psi(x, t_1)\mathcal{L}(t_2) = e(F(x; \theta_0), t_1 \wedge t_2)$, where $e(s, t)$ is as in (4.5.3).

Lemma 4.5.5. *Assume the conditions of Theorem 4.5.2. Then, under H_0 , we have, as $n \rightarrow \infty$,*

$$\left(\hat{\beta}_n(x, t), \mathcal{L}_n(t)\right) \xrightarrow{\mathcal{D}} \left(\Psi(x, t), \mathcal{L}(t)\right) \text{ in } D(\mathbb{R} \times [0, 1]).$$

Proof. The proof is similar to that of Lemma 4.5.3. ■

Proof of Theorem 4.5.2. The proof is similar to that of Theorem 4.5.1, using Lemma 4.5.5 instead of lemma 4.5.3. ■

Example 4.5.1. Let $X_1, \dots, X_n$ be independent $N(\mu, \sigma^2)$ random variables. Consider the null hypothesis

$$H_0 : X_1, \dots, X_n \text{ have the same distribution } F(\cdot; \mu, \sigma_0^2),$$

where $F(\cdot; \mu, \sigma_0^2)$ denotes the distribution function of a $N(\mu, \sigma_0^2)$ random variable, with μ unknown, and σ_0^2 denotes the assumed hypothetical value of the unknown parameter σ^2 . As alternative hypotheses, we consider

$$H_{1n} : X_1, \dots, X_n \text{ have the respective distribution functions } F_{in}(\cdot; \mu_n, \sigma_0^2),$$

where, for each n , $F_{in}(\cdot; \mu_n, \sigma_0^2)$ denotes the distribution function of a $N(\mu + c(i/n)n^{-1/2}, \sigma_0^2)$ random variable, and $c(\cdot)$ is a Riemman square integrable function.

We have

(4.5.50)

$$\frac{dF_{in}}{dF}(x) = \exp \{ (2c(i/n)n^{-1/2}(x - \mu) + c^2(i/n)n^{-1}) / 2\sigma_0^2 \},$$

$$x = F^{-1}(s; \mu, \sigma_0^2), \quad 0 \leq s \leq 1,$$

and therefore

(4.5.51)

$$\mathcal{L}_n(t) := \sum_{i=1}^{[nt]} \left\{ \log \frac{dF_{in}}{dF}(X_i) - E \log \frac{dF_{in}}{dF}(X_i) \right\} = n^{-1/2} \sum_{i=1}^{[nt]} c(i/n)(X_i - \mu) / \sigma_0^2.$$

We note that $\mathcal{L}_n(t) \xrightarrow{\mathcal{D}} \mathcal{L}(t)$, where $\mathcal{L}(t)$ is a Gaussian process with mean zero and covariance function

$$(4.5.52) \quad E\mathcal{L}(t_1)\mathcal{L}(t_2) = \int_0^{t_1 \wedge t_2} c^2(t)dt.$$

Suppose now that under H_0 , given $\frac{1}{n} \leq t \leq 1$, the unknown parameter μ is estimated, from the random sample $X_1, \dots, X_{[nt]}$, by its maximum likelihood estimator $\hat{\mu}_{[nt]} := \frac{1}{[nt]} \sum_{i=1}^{[nt]} X_i$, and consider the sequentially estimated empirical process $\{\hat{\beta}_n(x, t), 0 \leq t \leq 1, x \in \mathbb{R}\}$ defined by

$$(4.5.53) \quad \hat{\beta}_n(x, t) = n^{-1/2} \sum_{i=1}^{[nt]} \left(1\{X_i \leq x\} - F(x; \hat{\mu}_{[nt]}, \sigma_0^2) \right), \quad 0 \leq t \leq 1, x \in \mathbb{R}.$$

Let μ_0 be the unknown true value of μ . Then, after adding and subtracting $F(x; \mu_0, \sigma_0^2)$, and applying a one-term Taylor expansion to $F(x; \mu, \sigma_0^2)$ around μ_0 , we get that (4.5.53) can be written as

$$(4.5.54) \quad \begin{aligned} \hat{\beta}_n(x, t) &= n^{-1/2} \sum_{i=1}^{[nt]} (1\{X_i \leq x\} - F(x; \mu_0, \sigma_0^2)) \\ &\quad + n^{-1/2} [nt] (\hat{\mu}_{[nt]} - \mu_0) \left(\frac{d}{d\mu} F(x; \mu_0, \sigma_0^2) - \frac{d}{d\mu} F(x; \hat{\mu}_{[nt]}, \sigma_0^2) \right) \\ &\quad - n^{-1/2} [nt] (\hat{\mu}_{[nt]} - \mu_0) \frac{d}{d\mu} F(x; \mu_0, \sigma_0^2), \end{aligned}$$

where $\sup_{0 \leq t \leq 1} |\hat{\mu}_{[nt]} - \mu_0| \leq \sup_{0 \leq t \leq 1} |\hat{\mu}_{[nt]} - \mu_0|$.

Furthermore, it is easy to check that,

$$(4.5.55) \quad [nt]^{1/2}(\hat{\mu}_{[nt]} - \mu_0) = [nt]^{-1/2} \sum_{i=1}^{[nt]} (X_i - \mu_0).$$

By independence, for any $0 \leq t_1, t_2 \leq 1$,

$$\begin{aligned} &= E n^{-1/2} \sum_{i=1}^{[nt_1]} (1\{X_i \leq x\} - F(x; \mu_0, \sigma_0^2)) \mathcal{L}_n(t_2) \\ &= \frac{n}{\sigma_0^2} \sum_{i=1}^{[n(t_1 \wedge t_2)]} E(1\{X_i \leq x\} - F(x; \mu_0, \sigma_0^2))(X_i - \mu_0) c(i/n) \\ (4.5.56) \quad &= \frac{n}{\sigma_0^2} \sum_{i=1}^{[n(t_1 \wedge t_2)]} c(i/n) \int_{-\infty}^x (z - \mu_0) dF(z; \mu_0, \sigma_0^2) \\ &\rightarrow \frac{1}{\sigma_0^2} \int_0^{t_1 \wedge t_2} \int_{-\infty}^x (z - \mu_0) c(\tau) dF(z; \mu_0, \sigma_0^2) d\tau, \text{ as } n \rightarrow \infty. \end{aligned}$$

On the other hand, by (4.5.55),

$$\begin{aligned} &E n^{-1/2} [nt_1] (\hat{\mu}_{[nt_1]} - \mu_0) \mathcal{L}_n(t_2) \\ (4.5.57) \quad &= E n^{-1} \sum_{i=1}^{[nt_1]} (X_i - \mu_0) \sum_{j=1}^{[nt_2]} c(j/n) (X_j - \mu_0) / \sigma_0^2 \\ &= \frac{n}{\sigma_0^2} \sum_{i=1}^{[n(t_1 \wedge t_2)]} c(i/n) E(X_i - \mu_0)^2 \rightarrow \int_0^{t_1 \wedge t_2} c(\tau) d\tau, \text{ as } n \rightarrow \infty. \end{aligned}$$

Finally, since

$$(4.5.58) \quad \lim_{n \rightarrow \infty} E \mathcal{L}_n^2(t) = \int_0^t c^2(\tau) d\tau < \infty,$$

we have that, as $n \rightarrow \infty$, (cf. (4.5.19) and (4.5.43))

$$En^{-1/2}[nt_1](\hat{\mu}_{[nt_1]} - \mu_0) \left(\frac{d}{d\mu} F(x; \mu_0, \sigma_0^2) - \frac{d}{d\mu} F(x; \tilde{\mu}_{[nt_1]}, \sigma_0^2) \right) \mathcal{L}_n(t_2) \longrightarrow 0.$$

Thus, by (4.5.56), (4.5.57), and (4.5.58), we have that, under H_0 ,

$$(4.5.59) \quad E\hat{\beta}_n(x, t_1)\mathcal{L}_n(t_2) \longrightarrow \frac{1}{\sigma_0^2} \int_0^{t_1 \wedge t_2} \int_{-\infty}^x (z - \mu_0) c(\tau) dF(z; \mu_0, \sigma_0^2) d\tau \\ - \left\{ \int_0^{t_1 \wedge t_2} c(\tau) d\tau \right\} \frac{d}{d\mu} F(x; \mu_0, \sigma_0^2), \text{ as } n \rightarrow \infty.$$

Let q be a positive function on $(0, 1]$ (cf. 3.2.5), non-decreasing near zero, and such that (3.2.6) holds. By Corollary (3.2.1), we have that, under H_0 ,

$$\hat{\beta}_n(x, t)/q(t) \xrightarrow{\mathcal{D}} G(x, t)/q(t) \text{ in } D(\mathbb{R} \times [0, 1]),$$

where

$$G(x, t) = K(F(x; \mu_0, \sigma_0^2), t) - \left\{ \int (x - \mu_0) d_x K(F(x; \mu_0, \sigma_0^2), t) \right\} \frac{d}{d\mu} F(x; \mu_0, \sigma_0^2).$$

Hence, by (4.5.59) and Theorem 4.5.1, we have that, under H_{1n} ,

$$\hat{\beta}_n(x, t)/q(t) \xrightarrow{\mathcal{D}} (G(x, t) + d(x, t))/q(t) \text{ in } D(\mathbb{R} \times [0, 1]),$$

where

$$d(x, t) = \frac{1}{\sigma_0^2} \int_0^t \int_{-\infty}^x (z - \mu_0) c(\tau) dF(z; \mu_0, \sigma_0^2) d\tau - \left\{ \int_0^t c(\tau) d\tau \right\} \frac{d}{d\mu} F(x; \mu_0, \sigma_0^2).$$

Chapter 5

Sequentially estimated multivariate normal empirical processes Weighted approximations

5.1 Introduction.

In this chapter we discuss the sequentially estimated empirical processes, as in Chapter 3, for the multivariate normal family. Before proceeding, we introduce some definitions and preliminary results, based on Sections 2 and 3 of Csörgő and Szyszkowicz (1994).

For an arbitrary distribution function H on $\mathbb{R}^1$ we define the inverse (quantile function) of H by

$$H^{-1}(y) = \inf\{x \in \mathbb{R}^1 : H(x) \geq y\}, \quad 0 < y \leq 1, \quad H^{-1}(0) = H^{-1}(0+).$$

Hence, for $x \in \mathbb{R}^1$ and $0 < y < 1$, we have $H(x) \geq y$ if and only if $x \geq H^{-1}(y)$, and $H(H^{-1}(y)-) \leq y \leq H(H^{-1}(y))$. Consequently, if H is continuous, then $H(H^{-1}(y)) = y$.

Let F be an arbitrary, right continuously defined distribution function on $\mathbb{R}^d$, $F_{(j)}(x_j)$, $1 \leq j \leq d$, be the j th marginal of $F(\mathbf{x}) = F(x_1, \dots, x_d)$, $\mathbf{x} =$

$(x_1, \dots, x_d) \in \mathbb{R}^d$, $d \geq 1$, and let $F_{(j)}^{-1}(u_j)$, $1 \leq j \leq d$, $\mathbf{u} = (u_1, \dots, u_d) \in I^d := [0, 1]^d$, $d \geq 1$, be the inverses (quantile functions) of these marginals of F .

Define the map $L^{-1} : I^d \rightarrow \mathbb{R}^d$ by

(5.1.1)

$$L^{-1}(\mathbf{u}) = L^{-1}(u_1, \dots, u_d) := (F_{(1)}^{-1}(u_1), \dots, F_{(d)}^{-1}(u_d)), \quad \mathbf{u} = (u_1, \dots, u_d) \in I^d.$$

Then the map $L : \mathbb{R}^d \rightarrow I^d$, defined by

$$(5.1.2) \quad \begin{aligned} L(\mathbf{x}) &= L(x_1, \dots, x_d) := (F_{(1)}(x_1), \dots, F_{(d)}(x_d)), \\ \mathbf{x} &= (x_1, \dots, x_d) \in \mathbb{R}^d, \end{aligned}$$

is an inverse to L^{-1} in that, for $\mathbf{x} \in \mathbb{R}^d$ and $\mathbf{u} \in (0, 1)^d$, we have the component-wise inequalities $L(\mathbf{x}) \geq \mathbf{u}$ if and only if $\mathbf{x} \geq L^{-1}(\mathbf{u})$, and $L(L^{-1}(\mathbf{u})) \leq \mathbf{u} \leq L(L^{-1}(\mathbf{u}))$. Consequently, if the components of L are continuous, then $L(L^{-1}(\mathbf{u})) = \mathbf{u}$.

The following is a well-known lemma.

Lemma 5.1.A. *Let $\mathbf{X}$ be a random vector in $\mathbb{R}^d$, with an arbitrary distribution function F . Then there is a random vector $\mathbf{U} = (U^{(1)}, \dots, U^{(d)})$ with values in I^d , whose components are uniformly distributed over $[0, 1]$ such that $\mathbf{X} \stackrel{\mathcal{D}}{=} L^{-1}(\mathbf{U})$, i.e.,*

$$\mathbf{X} = (X^{(1)}, \dots, X^{(d)}) \stackrel{\mathcal{D}}{=} (F_{(1)}^{-1}(U^{(1)}), \dots, F_{(d)}^{-1}(U^{(d)})) = L^{-1}(\mathbf{U}).$$

Let G be the distribution function of the random vector $\mathbf{U} \in I^d$ of Lemma 5.1.A. Then, with F on $\mathbb{R}^d$ arbitrary, as in Lemma A, we have

$$(5.1.3) \quad G(L(\mathbf{x})) = P\{\mathbf{U} \leq L(\mathbf{x})\} = P\{L^{-1}(\mathbf{U}) \leq \mathbf{x}\} = F(\mathbf{x}), \quad \mathbf{x} \in \mathbb{R}^d,$$

and if F is continuous, then we have also

$$(5.1.4) \quad G(u) = F(L^{-1}(u)), \quad u \in I^d, \quad d \geq 1.$$

We note also that if F is continuous and $F(x) = \prod_{j=1}^d F_{(j)}(x_j)$ for all $x \in \mathbb{R}^d$, then

$$G(u) = F(L^{-1}(u)) = \prod_{j=1}^d u_j = \lambda(u),$$

the Lebesgue measure (uniform distribution) on $I^d, d \geq 1$.

Let $X = (X^{(1)}, \dots, X^{(d)})$, $X_i = (X_i^{(1)}, \dots, X_i^{(d)})$, $i = 1, 2, \dots$ be independent random vectors in $\mathbb{R}^d, d \geq 1$, with distribution function F , and let $F_n(x)$ be the empirical distribution function of $X_1, \dots, X_n$, i.e.,

$$F_n(x) = n^{-1} \sum_{i=1}^n 1(X_i \leq x), \quad x \in \mathbb{R}^d, \quad d \geq 1,$$

and define the $(d+1)$ -time parameter empirical process $\beta_n(x, t)$ by

$$(5.1.5) \quad \beta_n(x, t) = n^{-1/2} \sum_{i=1}^{[nt]} (1(X_i \leq x) - F(x)), \quad x \in \mathbb{R}^d, \quad 0 \leq t \leq 1.$$

Let $U = (U^{(1)}, \dots, U^{(d)})$, $U_i = (U_i^{(1)}, \dots, U_i^{(d)})$, $i = 1, 2, \dots$, be independent identically distributed random vectors with distribution as in Lemma 5.1.A, and define the $(d+1)$ -time parameter empirical process $\alpha_n(u, t)$ of the first $n \geq 1$ of these random vectors by

$$(5.1.6) \quad \alpha_n(\mathbf{u}, t) = n^{-1/2} \sum_{i=1}^{[nt]} (1(\mathbf{U}_i \leq \mathbf{u}) - G(\mathbf{u})), \quad \mathbf{u} \in I^d, \quad 0 \leq t \leq 1,$$

where G is the distribution function of $\mathbf{U} \in I^d$, $d \geq 1$. By Lemma 5.1.A, the above sequence of independent random vectors $\mathbf{U}_i = (U_i^{(1)}, \dots, U_i^{(d)}) \in I^d$, $i = 1, 2, \dots$, whose marginals are uniformly distributed over $[0, 1]$, are such that we have

$$(5.1.7) \quad \{\mathbf{X}_i, i \geq 1\} \stackrel{\mathcal{D}}{=} \{L^{-1}(\mathbf{U}_i), i \geq 1\}.$$

Consequently, by (5.1.7) and (5.1.3), with $F = G \circ L$ on $\mathbb{R}^d$ arbitrary, we have

$$(5.1.8) \quad \{\beta_n(\mathbf{x}, t), \mathbf{x} \in \mathbb{R}^d, 0 \leq t \leq 1, n \geq 1\} \stackrel{\mathcal{D}}{=} \{\alpha_n(L(\mathbf{x}), t), \mathbf{x} \in \mathbb{R}^d, 0 \leq t \leq 1, n \geq 1\}.$$

Moreover, if F on $\mathbb{R}^d$ is *continuous*, then $\mathbf{U}$ of Lemma 5.1.A with distribution function $G = F \circ L^{-1}$ on I^d can be taken to be $L(\mathbf{X})$, and (5.1.7) can be replaced by

$$(5.1.9) \quad \{L(\mathbf{X}_i), i \geq 1\} = \{\mathbf{U}_i, i \geq 1\}.$$

Consequently, by (5.1.9) and (5.1.4), with F on $\mathbb{R}^d$ *continuous*, we have $G = F \circ L^{-1}$ and

$$(5.1.10) \quad \{\beta_n(L^{-1}(\mathbf{u}), t), \mathbf{u} \in I^d, 0 \leq t \leq 1, n \geq 1\} = \{\alpha_n(\mathbf{u}, t), \mathbf{u} \in I^d, 0 \leq t \leq 1, n \geq 1\}.$$

A Kiefer process $K_F(\mathbf{x}, t)$ on $\mathbb{R}^d \times [0, \infty)$ associated with a distribution function F on $\mathbb{R}^d$, $d \geq 1$, is a separable $(d + 1)$ -parameter real valued Gaussian process with $K_F(\mathbf{x}, 0) = 0$, $E K_F(\mathbf{x}, t) = 0$, and

$$(5.1.11) \quad E K_F(\mathbf{x}, s) K_F(\mathbf{y}, t) = (s \wedge t)(F(\mathbf{x} \wedge \mathbf{y}) - F(\mathbf{x})F(\mathbf{y}))$$

for all $\mathbf{x}, \mathbf{y} \in \mathbb{R}^d$ and $s, t \geq 0$, where and throughout the symbol $\wedge$ means minimum, component-wise in higher dimensions.

Let F be an arbitrary distribution function on $\mathbb{R}^d$, $d \geq 1$ and let $\{K_F(\mathbf{x}, t), \mathbf{x} \in \mathbb{R}^d, t \geq 0\}$ be a Kiefer process associated with this distribution function.

Let $\mathbf{U} \in I^d$ be a random vector as in Lemma 5.1.A, with distribution function G on I^d , $d \geq 1$, and let $\{K_G(\mathbf{u}, t), \mathbf{u} \in I^d, t \geq 0\}$ be the Kiefer process associated with this distribution function G .

Consequently, if an arbitrary F on $\mathbb{R}^d$ is written as $F = G \circ L$ with G on I^d , $d \geq 1$, as in (5.1.3), then we have

$$\begin{aligned} \{K_F(\mathbf{x}, t), \mathbf{x} \in \mathbb{R}^d, t \geq 0\} &= \{K_{G \circ L}(\mathbf{x}, t), \mathbf{x} \in \mathbb{R}^d, t \geq 0\} \\ (5.1.12) \quad &\stackrel{D}{=} \{K_G(L(\mathbf{x}), t), \mathbf{x} \in \mathbb{R}^d, t \geq 0\}. \end{aligned}$$

On the other hand, if F on $\mathbb{R}^d$ is *continuous*, we can write $G = F \circ L^{-1}$ on I^d as in (5.1.4), and we have

$$\begin{aligned} \{K_G(\mathbf{u}, t), \mathbf{u} \in I^d, t \geq 0\} &= \{K_{F \circ L^{-1}}(\mathbf{u}, t), \mathbf{u} \in I^d, t \geq 0\} \\ (5.1.13) \quad &\stackrel{D}{=} \{K_F(L^{-1}(\mathbf{u}), t), \mathbf{u} \in I^d, t \geq 0\}. \end{aligned}$$

Csörgő and Horváth (1988) proved the following strong approximation result.

Theorem 5.1.A. . Assume that $\mathbf{X}_1, \mathbf{X}_2, \dots$ are independent random vectors with an arbitrary distribution function F on $\mathbb{R}^d$, $d \geq 1$. Then, associated with F on $\mathbb{R}^d$, there exists a Kiefer process $\{K_F(\mathbf{x}, t), \mathbf{x} \in \mathbb{R}^d, t \geq 0\}$ such that

$$(5.1.14) \quad \sup_{0 \leq t \leq 1} \sup_{\mathbf{x} \in \mathbb{R}^d} |n^{1/2} \beta_n(\mathbf{x}, t) - K_F(\mathbf{x}, nt)| \stackrel{a.s.}{=} O(n^{1/2-\varepsilon/(4d)} (\log n)^{3/2}).$$

Csörgő and Szyszkowicz (1994) studied weighted $\|\cdot/q\|$ -metric approximations of (5.1.5) for the class of weight functions used in previous sections.

Recall that q is said to be positive on $(0, 1]$ if $q : (0, 1] \rightarrow (0, \infty)$ is such that

$$(5.1.15) \quad \inf_{\delta \leq t \leq 1} q(t) > 0 \text{ for all } 0 < \delta < 1,$$

and $\mathcal{Q}$ is the class of positive functions q on $(0, 1]$ that are non-decreasing in a neighbourhood of 0. Consider the integral

$$(5.1.16) \quad I_0(q, c) = \int_0^1 \frac{1}{t} e^{-cq^2(t)/t} dt, \quad c > 0.$$

They prove the following

Theorem 5.1.B. *Assume that $X_1, X_2, \dots$ are independent random vectors with arbitrary distribution function F on $\mathbb{R}^d$, $d \geq 1$. Then, associated with F on $\mathbb{R}^d$, there exists a Kiefer process $\{K_F(\mathbf{x}, t), \mathbf{x} \in \mathbb{R}^d, t \geq 0\}$ such that, with a weight function $q \in \mathcal{Q}$, we have, as $n \rightarrow \infty$,*

(a)

$$(5.1.17) \quad I_0(q, c) < \infty \text{ for all } c > 0$$

if and only if

$$\sup_{0 < t \leq 1} \sup_{\mathbf{x} \in \mathbb{R}^d} |\beta_n(\mathbf{x}, t) - n^{-1/2} K_F(\mathbf{x}, nt)|/q(t) = o_P(1),$$

(b)

$$(5.1.18) \quad I_0(q, c) < \infty \text{ for some } c > 0$$

if and only if

$$\sup_{0 < t \leq 1} \sup_{\mathbf{x} \in \mathbb{R}^d} |\beta_n(\mathbf{x}, t) - n^{-1/2} K_F(\mathbf{x}, nt)|/q(t) = O_P(1),$$

(c) (5.1.18) holds if and only if

$$\sup_{0 < t \leq 1} \sup_{x \in \mathbb{R}^d} |\beta_n(x, t)|/q(t) \xrightarrow{P} \sup_{0 < t \leq 1} \sup_{x \in \mathbb{R}^d} |K_F(x, t)|/q(t),$$

where $K_F(\cdot, \cdot)$ is a Kiefer process associated with F on $\mathbb{R}^d$.

Remark 5.1.1. (b) and (c) do not follow from (a), nor does (c) from (b).

Remark 5.1.2. It is clear from the proof of Theorem 5.1.B that, as a consequence of Theorem 5.1.A, for any function q which is positive on $(0, 1]$ and such that

$$(5.1.19) \quad \lim_{t \downarrow 0} t^{1/2}/q(t) = 0,$$

then, as $n \rightarrow \infty$,

$$(5.1.20) \quad \sup_{0 < t \leq 1} \sup_{x \in \mathbb{R}^d} |\beta_n(x, t) - n^{-1/2} \bar{K}_F(x, nt)|/q(t) \stackrel{a.s.}{=} o(1),$$

where

$$\bar{K}_F(x, nt) = \begin{cases} K_F(x, nt), & \text{for } t \in [1/n, 1], \\ 0, & \text{elsewhere,} \end{cases}$$

and $K_F(\cdot, \cdot)$ is a Kiefer process as in Theorem 5.1.A.

We note that either one of (2.1.4) and (2.1.5) implies (2.1.7) (cf. Proposition 3.1 of Csörgő, Csörgő, Horváth and Mason (1986)). Hence, since, by definition, $\beta_n(x, t) = 0$, for all $x \in \mathbb{R}^d$ and $t \in (0, 1/n)$, and for each $n \geq 1$,

$$\sup_{x \in \mathbb{R}^d} |n^{-1/2} K_F(x, nt)| \stackrel{P}{=} \sup_{x \in \mathbb{R}^d} |K_F(x, t)|,$$

the optimal integral conditions of Theorem 5.1.B are due to the following result, which, itself, follows from Theorem 2.1 of Csörgő, Horváth and Szyszkowicz (1994), and provides an integral test for the suprema of Kiefer processes indexed by arbitrary distribution functions F on $\mathbb{R}^d$, $d \geq 1$.

Theorem 5.1.C. *Let $q \in \mathcal{Q}$ and $K_F(\cdot, \cdot)$ be a Kiefer process indexed by the distribution function F on $\mathbb{R}^d$, $d \geq 1$. Then*

(a) (5.1.17) holds if and only if

$$\lim_{t \downarrow 0} \sup_{x \in \mathbb{R}^d} |K_F(x, t)|/q(t) = 0 \text{ a.s.},$$

(b) (5.1.18) holds if and only if

$$\limsup_{t \downarrow 0} \sup_{x \in \mathbb{R}^d} |K_F(x, t)|/q(t) < \infty \text{ a.s.}$$

Remark 5.1.3. On account of Theorem 5.1.C and our Remarks 2.1.5 and 3.3.5, we note that, the class of functions that characterize the local behaviour (near 0) of suprema of Kiefer processes indexed by arbitrary distribution functions F on $\mathbb{R}^d$, $d \geq 1$, is the same class of functions that characterize the similar local behaviour of Wiener processes.

Remark 5.1.4. Theorem 5.1.A is Theorem 3.1 of Csörgő and Szyszkowicz (1994). It was proved there under stronger conditions, similar to those discussed in our Remark 2.1.5. Namely, assuming q to be positive on $(0, 1]$, and replacing condition (5.1.17) by (2.1.9) and condition (5.1.18) by (2.1.10), as well as by assuming that $q \in C(0, 1]$ and $q(t)/t^{1/2}$ is nonincreasing near 0, and replacing (2.1.9) and (2.1.10) by $\mathcal{I}_d(q, c) < \infty$ for all $c > 0$, and, respectively, $\mathcal{I}_d(q, c) < \infty$ for some $c > 0$, where

$$\mathcal{I}_d(q, c) = \int_0^1 \frac{q^{2d}(t)}{t^{d+1}} e^{-cq^2(t)/t} dt, \quad c > 0.$$

Due to the integral test for suprema of Kiefer processes of Csörgő, Horváth and Szyszkowicz (1994), which, as we noted, was proved for Kiefer processes indexed by arbitrary distribution functions F on $\mathbb{R}^d$, $d \geq 1$, Theorem 5.1.B is an improvement to Theorem 3.1 of Csörgő and Szyszkowicz (1994) (cf. Theorem 8.1 of Szyszkowicz (1995)).

As discussed in Section 3.3, for the univariate case, when testing for the possibility of having a change in distribution of a sequence of chronologically ordered d -dimensional observations $\mathbf{X}_i = (X_i^{(1)}, \dots, X_i^{(d)})$, $i = 1, \dots, n$, $d \geq 1$, at an unknown time $1 \leq k < n$, it is natural to compare the empirical distributions “before” to those “after”, via studying the asymptotic distribution of the sequence of statistics

$$\begin{aligned}
 & \sup_{1 \leq k < n} \sup_{\mathbf{x} \in \mathbb{R}^d} n^{1/2} \left| \frac{1}{k} \sum_{i=1}^k \mathbf{1}(\mathbf{X}_i \leq \mathbf{x}) - \frac{1}{n-k} \sum_{i=k+1}^n \mathbf{1}(\mathbf{X}_i \leq \mathbf{x}) \right| \\
 (5.1.21) \quad & = \sup_{1 \leq k < n} \sup_{\mathbf{x} \in \mathbb{R}^d} \left| \sum_{i=1}^k \mathbf{1}(\mathbf{X}_i \leq \mathbf{x}) - \frac{k}{n} \sum_{i=1}^n \mathbf{1}(\mathbf{X}_i \leq \mathbf{x}) \right| / \left(n^{1/2} \left(\frac{k}{n} \left(1 - \frac{k}{n} \right) \right) \right),
 \end{aligned}$$

which, however converge in distribution to ∞ , as $n \rightarrow \infty$, even if the null assumption of no change in distribution were true. This remains true even if we were to replace the denominator in the right hand side of (5.1.21) by $n^{1/2}((k/n)(1-k/n))^{1/2}$, $1 \leq k < n$.

Csörgő and Szyszkowicz (1994) studied the asymptotic behavior of the process

$$(\beta_n(\mathbf{x}, t) - t\beta_n(\mathbf{x}, 1))/q(t), \quad 0 \leq t \leq 1, \quad \mathbf{x} \in \mathbb{R}^d.$$

for a wide class of weight functions q , where $\beta_n(\mathbf{x}, t)$ is as in (5.1.5).

Define the "tied down in t " multi-time parameter empirical bridge process $\{\beta_n^*(x, t), x \in \mathbb{R}^d, 0 \leq t \leq 1, n \geq 1\}$ by

$$(5.1.22) \quad \beta_n^*(x, t) = \begin{cases} n^{-1/2} \left(\sum_{i=1}^{[(n+1)t]} 1(X_i \leq x) - \frac{[(n+1)t]}{n} \sum_{i=1}^n 1(X_i \leq x) \right), & 0 \leq t < 1, x \in \mathbb{R}^d, \\ 0, & t = 1, x \in \mathbb{R}^d. \end{cases}$$

Let $U_i = (U_i^{(1)}, \dots, U_i^{(d)}) \in I^d$, $i = 1, 2, \dots$, be a sequence of independent identically distributed random vectors with distribution function G as that of the random vector $U \in I^d$ of Lemma 5.1.A, and define the "tied down in t " multi-time parameter empirical bridge process $\{\alpha_n^*(s, t), s \in I^d, 0 \leq t \leq 1, n \geq 1\}$ by

$$(5.1.23) \quad \alpha_n^*(s, t) = \begin{cases} n^{-1/2} \left(\sum_{i=1}^{[(n+1)t]} 1(U_i \leq s) - \frac{[(n+1)t]}{n} \sum_{i=1}^n 1(U_i \leq s) \right), & 0 \leq t < 1, s \in I^d, \\ 0, & t = 1, s \in I^d. \end{cases}$$

With F on $\mathbb{R}^d$ arbitrary, via (5.1.3) and (5.1.7), we have

(5.1.24)

$$\{\beta_n^*(x, t), x \in \mathbb{R}^d, 0 \leq t \leq 1, n \geq 1\} \stackrel{D}{=} \{\alpha_n^*(L(x), t), x \in \mathbb{R}^d, 0 \leq t \leq 1, n \geq 1\}.$$

Moreover, if F on $\mathbb{R}^d$ is continuous, then U of Lemma 5.1.A with distribution function G on I^d can be taken to be $L(X)$. Consequently, by (5.1.4) and (5.1.9), with F on $\mathbb{R}^d$ continuous, we have

(5.1.25)

$$\{\beta_n^*(L^{-1}(u), t), u \in I^d, 0 \leq t \leq 1, n \geq 1\} = \{\alpha_n^*(u, t), u \in I^d, 0 \leq t \leq 1, n \geq 1\}.$$

Let $\{K_F(\mathbf{x}, t), \mathbf{x} \in \mathbb{R}^d, t \geq 0\}$ be a Kiefer process associated with an arbitrary distribution function F on $\mathbb{R}^d, d \geq 1$, and define the Gaussian process $\{\Gamma_F(\mathbf{x}, t), \mathbf{x} \in \mathbb{R}^d, 0 \leq t \leq 1\}$ by

$$(5.1.26) \quad \Gamma_F(\mathbf{x}, t) = K_F(\mathbf{x}, t) - tK_F(\mathbf{x}, 1), \mathbf{x} \in \mathbb{R}^d, 0 \leq t \leq 1.$$

Consequently, $\Gamma_F(\cdot, \cdot)$ is a separable mean zero Gaussian process with covariance function

$$(5.1.27) \quad E\Gamma_F(\mathbf{x}_1, t_1)\Gamma_F(\mathbf{x}_2, t_2) = (F(\mathbf{x}_1 \wedge \mathbf{x}_2) - F(\mathbf{x}_1)F(\mathbf{x}_2))(t_1 \wedge t_2 - t_1 t_2).$$

Clearly, if an arbitrary F on $\mathbb{R}^d$ is written as $F = G \circ L$ with G on $I^d, d \geq 1$, as in (5.1.3), then we have (cf. (5.1.12))

$$(5.1.28) \quad \begin{aligned} \{\Gamma_F(\mathbf{x}, t), \mathbf{x} \in \mathbb{R}^d, 0 \leq t \leq 1\} &= \{\Gamma_{G \circ L}(\mathbf{x}, t), \mathbf{x} \in \mathbb{R}^d, 0 \leq t \leq 1\} \\ &\stackrel{\mathcal{D}}{=} \{\Gamma_G(L(\mathbf{x}), t), \mathbf{x} \in \mathbb{R}^d, 0 \leq t \leq 1\}. \end{aligned}$$

On the other hand, if F on $\mathbb{R}^d$ is continuous, we can write $G = F \circ L^{-1}$ on I^d as in (5.1.4), and we have (cf. (5.1.13))

$$(5.1.29) \quad \begin{aligned} \{\Gamma_G(\mathbf{u}, t), \mathbf{u} \in I^d, 0 \leq t \leq 1\} &= \{\Gamma_{F \circ L^{-1}}(\mathbf{u}, t), \mathbf{u} \in I^d, 0 \leq t \leq 1\} \\ &\stackrel{\mathcal{D}}{=} \{\Gamma_F(L^{-1}(\mathbf{u}), t), \mathbf{u} \in I^d, 0 \leq t \leq 1\}. \end{aligned}$$

Recall that a function $q: (0, 1) \rightarrow (0, \infty)$ is said to be positive on $(0, 1)$ if

$$(5.1.30) \quad \inf_{\delta \leq t \leq 1-\delta} q(t) > 0 \text{ for all } 0 < \delta \leq 1/2,$$

and that Q^* is the class of positive functions on $(0, 1)$ that are non-decreasing in a neighbourhood of zero and non-increasing in a neighbourhood of one.

Consider the integral

$$I_{0,1}(q, c) = \int_0^1 \frac{1}{t(1-t)} e^{-cq^2(t)/(1-t)} dt, \quad c > 0.$$

Theorem 5.1.D. Assume that $X_1, X_2, \dots$ are independent random vectors with an arbitrary distribution function F on $\mathbb{R}^d$, $d \geq 1$. Then, associated with F on $\mathbb{R}^d$, there exists a Kiefer process $\{K_F(x, t), x \in \mathbb{R}^d, t \geq 0\}$ such that with

(5.1.31)

$$\begin{aligned} & \{\Gamma_{F,n}(x, t), x \in \mathbb{R}^d, 0 \leq t \leq 1, n \geq 1\} \\ & := \{n^{-1/2}(K_F(x, nt) - tK_F(x, n)), x \in \mathbb{R}^d, 0 \leq t \leq 1, n \geq 1\} \\ & \stackrel{\mathcal{D}}{=} \{\Gamma_F(x, t), x \in \mathbb{R}^d, 0 \leq t \leq 1\} \text{ for each } n \geq 1, \end{aligned}$$

and $q \in Q^*$, we have, as $n \rightarrow \infty$,

(a)

$$(5.1.32) \quad I_{0,1}(q, c) < \infty \text{ for all } c > 0$$

if and only if

$$\sup_{0 < t < 1} \sup_{x \in \mathbb{R}^d} |\beta_n^*(x, t) - \Gamma_{F,n}(x, t)|/q(t) = o_P(1),$$

(b)

$$(5.1.33) \quad I_{0,1}(q, c) < \infty \text{ for some } c > 0$$

if and only if

$$\sup_{0 < t < 1} \sup_{x \in \mathbb{R}^d} |\beta_n^*(x, t) - \Gamma_{F,n}(x, t)|/q(t) = O_P(1),$$

(c) (5.1.33) holds if and only if

$$\sup_{0 < t < 1} \sup_{x \in \mathbb{R}^d} |\beta_n^*(x, t)|/q(t) \xrightarrow{\mathcal{D}} \sup_{0 < t < 1} \sup_{x \in \mathbb{R}^d} |\Gamma_F(x, t)|/q(t),$$

where $\Gamma_F(\cdot, \cdot)$ is a Gaussian process as in (5.1.27).

Remark 5.1.5. (b) and (c) do not follow from (a), nor does (c) from (b).

Remark 5.1.6. Similarly as in the case of Theorem 3.3.A, from the proof of Theorem 5.1.D, we have that, provided that q is a positive function on $(0, 1)$ and such that $\lim_{t \downarrow 0} t^{1/2}/q(t) = 0$, we have

$$(5.1.34) \quad \sup_{0 < t < 1} \sup_{\mathbf{x} \in \mathbb{R}^d} |\beta_n^*(\mathbf{x}, t) - \bar{\Gamma}_{F,n}(\mathbf{x}, t)|/q(t) \stackrel{a.s.}{=} o(1),$$

where

$$\bar{\Gamma}_{F,n}(\mathbf{x}, t) = \begin{cases} \Gamma_{F,n}(\mathbf{x}, t), & \text{for } t \in (0, \frac{1}{n+1}] \cup [\frac{n}{n+1}, 1), \\ 0, & \text{elsewhere,} \end{cases}$$

and $\Gamma_{F,n}(\cdot, \cdot)$ is as in (5.1.31).

Since, by definition, $\beta_n^*(\mathbf{x}, t) = 0$, for any $t \in A_n := (0, \frac{1}{n+1}] \cup [\frac{n}{n+1}, 1)$, and for each $n \geq 1$,

$$\sup_{\mathbf{x} \in \mathbb{R}^d} |\Gamma_{F,n}(\mathbf{x}, t)| \stackrel{D}{=} \sup_{\mathbf{x} \in \mathbb{R}^d} |\Gamma_F(\mathbf{x}, t)|,$$

the present optimal form of Theorem 5.1.D follows from the following theorem, due to Csörgő, Horváth and Szyszkowicz (1994).

Theorem 5.1.E. Let $q \in \mathcal{Q}$ and $K_F(\cdot, \cdot)$ be a Kiefer process indexed by the distribution functions F on $\mathbb{R}^d$, $d \geq 1$. Then

(a) (5.1.32) holds if and only if

$$\lim_{t \downarrow 0} \sup_{\mathbf{x} \in \mathbb{R}^d} |\Gamma_F(\mathbf{x}, t)|/q(t) = 0 \text{ a.s.}$$

and

$$\lim_{t \downarrow 1} \sup_{x \in \mathbb{R}^d} |\Gamma_F(x, t)|/q(t) = 0 \text{ a.s.},$$

(b) (5.1.33) holds if and only if

$$\limsup_{t \downarrow 0} \sup_{x \in \mathbb{R}^d} |\Gamma_F(x, t)|/q(t) < \infty \text{ a.s.}$$

and

$$\limsup_{t \downarrow 1} \sup_{x \in \mathbb{R}^d} |\Gamma_F(x, t)|/q(t) < \infty \text{ a.s.},$$

where $\Gamma_F(\cdot, \cdot)$ is as in (5.1.31).

Remark 5.1.7. On account of Theorem 5.1.D and our Remarks 3.3.4 and 3.3.5, we note that the class of functions that characterize the local behaviour (near 0 and near 1) of suprema of the processes $\Gamma_F(x, t)$, indexed by arbitrary distribution functions F on $\mathbb{R}^d$, $d \geq 1$, is the same class of functions that characterize the similar local behaviour of Brownian bridge processes.

Remark 5.1.8. Theorem 5.1.D is Theorem 3.2 of Csörgő and Szyszkowicz (1994). It was proved under the stronger conditions, resulting from the law of iterated logarithm for $\Gamma_F(x, t)$, or, equivalently, from the Adler and Brown (1986) test for local functions of $\Gamma_F(x, t)$ (cf. our Remark 5.1.4).

In virtue of Theorem 5.1.E, Theorem 5.1.D is an improvement of Theorem 3.2 of Csörgő and Szyszkowicz (1994) (cf. Theorem 8.2 of Szyszkowicz (1995)).

In the case where $X_1, X_2, \dots$ is a sequence of i.i.d. random vectors from a family of continuous distribution functions $\{F(x; \theta); x \in \mathbb{R}^d, \theta \in \Theta \subseteq \mathbb{R}^p\}$ and θ , the vector of unknown parameters, is estimated, for each $\frac{1}{n} \leq t \leq 1$, by a sequence

$\{\hat{\theta}_{[nt]}\} = \{(\hat{\theta}_{[nt]1}, \dots, \hat{\theta}_{[nt]p})\}$, the $(d+1)$ -time parameter sequentially estimated empirical process is defined by

$$(5.1.35) \quad \hat{\beta}_n(\mathbf{x}, t) = n^{-1/2} \sum_{i=1}^{[nt]} \left(\mathbf{1}\{\mathbf{X}_i \leq \mathbf{x}\} - F(\mathbf{x}, \hat{\theta}_{[nt]}) \right), \quad \mathbf{x} \in \mathbb{R}^d, \quad 0 \leq t \leq 1.$$

In Section 5.2 we will study weighted $\|\cdot/q\|$ -metric approximations of $\hat{\beta}_n$, when $\mathbf{X}_1, \mathbf{X}_2, \dots$ is a sequence of i.i.d. bivariate normal random vectors. In Section 5.3 we introduce a "tied down at $t = 1$ " version, $\hat{\hat{\beta}}_n$, of $\hat{\beta}_n$, also for the bivariate normal family. In Section 5.4 we introduce some further notation needed for the development of Sections 5.5 and 5.6, where we generalize the results of Sections 5.2 and 5.3 to the multivariate normal family. In all cases, the unknown parameters are estimated by using maximum likelihood estimation.

5.2 Sequentially estimated bivariate normal empirical processes.

Let $(X_1, Y_1), (X_2, Y_2), \dots$ be independent bivariate normal random vectors with vector of means (ζ, η) and covariance matrix $\begin{pmatrix} \sigma^2 & \alpha \\ \alpha & \tau^2 \end{pmatrix}$. Let θ denote the vector of unknown parameters, i.e. $\theta := (\zeta, \eta, \sigma^2, \tau^2, \alpha)$. For each $\frac{1}{n} \leq t \leq 1$ and $n \geq 1$, let $\{\hat{\theta}_{[nt]}\}$ be the maximum likelihood estimator of θ based on the sample $(X_1, Y_1), \dots, (X_{[nt]}, Y_{[nt]})$. It is known that,

$$(5.2.1) \quad \hat{\theta}_{[nt]} := (\bar{X}_{[nt]}, \bar{Y}_{[nt]}, \hat{\sigma}_{[nt]}^2, \hat{\tau}_{[nt]}^2, \hat{\alpha}_{[nt]}),$$

where

$$\bar{X}_{[nt]} = \frac{1}{[nt]} \sum_{i=1}^{[nt]} X_i, \quad \bar{Y}_{[nt]} = \frac{1}{[nt]} \sum_{i=1}^{[nt]} Y_i, \quad \hat{\sigma}_{[nt]}^2 = \frac{1}{[nt]} \sum_{i=1}^{[nt]} (X_i - \bar{X}_{[nt]})^2,$$

$$\hat{\tau}_{[nt]}^2 = \frac{1}{[nt]} \sum_{i=1}^{[nt]} (Y_i - \bar{Y}_{[nt]})^2 \text{ and } \hat{\alpha}_{[nt]} = \frac{1}{[nt]} \sum_{i=1}^{[nt]} (X_i - \bar{X}_{[nt]})(Y_i - \bar{Y}_{[nt]}).$$

Let $\theta_0 := (\zeta_0, \eta_0, \sigma_0^2, \tau_0^2, \alpha_0)$ be the true value of the parameter θ , and let $F(x, y; \theta_0)$ denote the distribution function of a bivariate normal random variable with parameter θ_0 . It can be easily checked, after a short calculation, that

$$\frac{1}{[nt]} \sum_{i=1}^{[nt]} (X_i - \bar{X}_{[nt]})(Y_i - \bar{Y}_{[nt]}) = \frac{1}{[nt]} \sum_{i=1}^{[nt]} (X_i - \zeta_0)(Y_i - \eta_0) - (\bar{X}_{[nt]} - \zeta_0)(\bar{Y}_{[nt]} - \eta_0).$$

Hence, we have (cf. (3.2.1), (3.2.2) and (3.2.3)),

$$(5.2.2) \quad [nt]^{1/2}(\hat{\theta}_{[nt]} - \theta_0) = [nt]^{-1/2} \sum_{i=1}^{[nt]} l(X_i, Y_i, \theta_0) + \varepsilon_{1n}(t),$$

where

(5.2.3)

$$l(x, y, \theta_0) = (x - \zeta_0, y - \eta_0, (x - \zeta_0)^2 - \sigma_0^2, (y - \eta_0)^2 - \tau_0^2, (x - \zeta_0)(y - \eta_0) - \alpha_0)$$

and

(5.2.4)

$$\varepsilon_{1n}(t) = -[nt]^{1/2} (0, 0, (\bar{X}_{[nt]} - \zeta_0)^2, (\bar{Y}_{[nt]} - \eta_0)^2, (\bar{X}_{[nt]} - \zeta_0)(\bar{Y}_{[nt]} - \eta_0)).$$

The process $\hat{\beta}_n(x, y, t)$, defined in (5.1.35), for $d = 2$, will be approximated, in $\|\cdot\|_q$ -metrics by the sequence of Gaussian processes $\{G_{F_0, n}(x, y, t), 0 \leq t \leq 1, (x, y) \in \mathbb{R}^2\}$, $n \geq 1$, where

$$(5.2.5) \quad G_{F_0, n}(x, y, t) = n^{-1/2} K_{F_0}(x, y, nt) - \left\{ \int_{\mathbb{R}^2} l(x, y, \theta_0) d_x d_y n^{-1/2} K_{F_0}(x, y, nt) \right\} \cdot \nabla_{\theta} F(x, y; \theta_0)^T,$$

and K_{F_0} denotes a Kiefer process associated with the distribution function $F(x, y, \theta_0)$, as in Theorem 5.1.A, with $d = 2$, for the class of weight functions q considered in the previous chapters (cf. (5.1.15) and (5.1.16)).

Let us denote each component of the vector function l in (5.2.3) by l_j , $j = 1, \dots, 5$. We note in (5.2.3), that, without loss of generality, coordinate-wise, we can consider $l_j(x, y, \theta_0) = l_j^{(1)}(x, \theta_0)l_j^{(2)}(y, \theta_0)$. Hence, each component of the integral in (5.2.5) is meant as the iterated integral

$$(5.2.6) \quad \int_{\mathbb{R}} l_j^{(1)}(x, \theta_0) dx \left[\int_{\mathbb{R}} l_j^{(2)}(y, \theta_0) dy n^{-1/2} K_{F_0}(x, y, nt) \right],$$

defined by an integration by parts formula (cf. Appendix A.2).

We have

Theorem 5.2.1. *Let $(X_1, Y_1), (X_2, Y_2), \dots$ be independent bivariate normal random vectors with vector of means (ζ, η) and covariance matrix $\begin{pmatrix} \sigma^2 & \alpha \\ \alpha & \tau^2 \end{pmatrix}$. Suppose that, $\{\hat{\theta}_{[nt]}\}$ is as in (5.2.1), and let $\hat{\beta}_n(x, y, t)$ be the sequentially estimated bivariate normal empirical process defined as in (5.1.35), with $d = 2$. Then, there exists a Kiefer process K_{F_0} , associated with the distribution function, $F(x, y, \theta_0)$, of a bivariate normal random vector with parameters given by θ_0 , such that with the sequence of stochastic processes $\{G_{F_0, n}(x, y, t), 0 \leq t \leq 1, (x, y) \in \mathbb{R}^2\}$, as in (5.2.5), and r weight function $q \in \mathcal{Q}$ we have, as $n \rightarrow \infty$,*

(a) if

$$(5.2.7) \quad I_0(q, c) < \infty \text{ for all } c > 0,$$

then

$$\sup_{0 < t \leq 1} \sup_{(x, y) \in \mathbb{R}^2} |\hat{\beta}_n(x, y, t) - G_{F_0, n}(x, y, t)|/q(t) = o_P(1),$$

(b) if

$$(5.2.8) \quad I_0(q, c) < \infty \text{ for some } c > 0,$$

then

$$\sup_{0 < t \leq 1} \sup_{(x, y) \in \mathbb{R}^2} |\hat{\beta}_n(x, y, t) - G_{F_0, n}(x, y, t)|/q(t) = O_P(1),$$

(c) if (5.2.8) holds, then

$$\sup_{0 < t \leq 1} \sup_{(x, y) \in \mathbb{R}^2} |\hat{\beta}_n(x, y, t)|/q(t) \xrightarrow{P} \sup_{0 < t \leq 1} \sup_{(x, y) \in \mathbb{R}^2} |G_{F_0}(x, y, t)|/q(t),$$

where

(5.2.9)

$$G_{F_0}(x, y, t) = K_{F_0}(x, y, t) - \left\{ \int_{\mathbb{R}^2} l(x, y, \theta_0) d_x d_y K_{F_0}(x, y, t) \right\} \cdot \nabla_{\theta} F(x, y; \theta_0)^T.$$

The proof of Theorem 5.2.1 hinges on two lemmas which we state and prove in the sequel, before proving Theorem 5.2.1.

Let $A_2^*(k)$ denote the complement of the rectangle $[-\log k, \log k] \times [-\log k, \log k]$, $k \geq 1$. For the vector-valued function $l(x, \theta_0)$ in (5.2.3) and the distribution function $F(x, y; \theta_0)$, we have

Lemma 5.2.1. *The series $\sum_{k=1}^{\infty} \frac{1}{k} \int_{A_2^*(k)} [l_j(x, y, \theta_0)]^2 d_x d_y F(x, y; \theta_0)$ is convergent, for any $j = 1, \dots, 5$.*

Proof. We have, from (5.2.3), that, for $j = 1, 3$, $l_j(x, y, \theta_0) \equiv l_j(x, \theta_0)$. We may also assume, without loss of generality, that $\zeta_0 = \eta_0 = 0$ and $\sigma_0 = \tau_0 = 1$. For $j = 1, 3$,

$$\int_{|x| > \log k} \int [l_j(x, y, \theta_0)]^2 f(x, y; \theta_0) dy dx = \int_{|x| > \log k} [l_j(x, \theta_0)]^2 f_X(x; \theta_0) dx,$$

where $f_{X_i}(x; \theta_0)$ denotes the marginal density of X_i . Hence, by Lemma 3.2.2, we have that $\sum_{k=1}^{\infty} \frac{1}{k} \int_{|x| \geq \log k} [l_j(x, \theta_0)]^2 f_X(x; \theta_0) dx$ is convergent, for $j = 1, 3$.

It can be easily shown, using integration by parts, that, for some constant $C > 0$, and $j = 1, 3$,

$$\int_{|y| \geq \log k} \int_{|x| \leq \log k} [l_j(x, y, \theta_0)]^2 f(x, y; \theta_0) dx dy \leq C k^3 e^{-\log^2 k / 2}.$$

Now, an application of Lemma 3.2.1 shows that the result holds for $j = 1, 3$.

Similarly, by (5.2.3) and symmetry, since $l_j(x, y, \theta_0) \equiv l_j(y, \theta_0)$, for $j = 2, 4$, the result also holds for the cases $j = 2, 4$.

Consider now the function $g(x, y) = xy + a$, where a is a constant. We have, for instance, that

$$\int_{|x| \geq \log k} \int_{|y| \leq \log k} x^2 y^2 f(x, y; \theta_0) dx dy \leq \log^2 k \int_{|x| \geq \log k} x^2 f_{X_i}(x; \theta_0) dx.$$

On the other hand,

$$\begin{aligned} & \int_{|x| \geq \log k} \int_{|y| \geq \log k} x^2 y^2 f(x, y; \theta_0) dx dy \\ &= 4 \int_{\log k}^{\infty} \int_{\log k}^{\infty} x^2 y^2 e^{-[(x - \rho_0 y)^2 + y^2(1 - \rho_0^2)]/2(1 - \rho_0^2)} dx dy \\ &= 4 \int_{\log k}^{\infty} y^2 e^{-y^2} dy \int_{\log k}^{\infty} x^2 e^{-(x - \rho_0 y)^2/2(1 - \rho_0^2)} dx \\ &= 4\sqrt{2\pi}(1 - \rho_0^2) \int_{\log k}^{\infty} y^2 e^{-y^2} dy. \end{aligned}$$

Similar bounds can be obtained for the integrals $\int_{A_2^*(k)} xy dx dy F(x, y; \theta_0)$ and $\int_{A_2^*(k)} dx dy F(x, y; \theta_0)$. Then, applying again Lemma 3.2.2 we get that the result holds for $j = 5$ as well. ■

Introduce the following

$$(5.2.10) \quad \epsilon_{2n}(x, y, t) = n^{-1/2} \left[\sum_{i=1}^{[nt]} (1\{(X_i, Y_i) \leq (x, y)\}) - F(x, y; \theta_0) \right] - K_{F_0}(x, y, nt)$$

and

$$(5.2.11) \quad L_n(t) = \int_{\mathbb{R}^2} l(x, y, \theta_0) d_x d_y \epsilon_{2n}(x, y, t),$$

where K_{F_0} is a Kiefer process associated with $F(\cdot, \cdot; \theta_0)$.

Lemma 5.2.2. *Let $(X_1, Y_1), (X_2, Y_2), \dots$ be independent bivariate normal random vectors with vector of means (ζ_0, η_0) and covariance matrix $\begin{pmatrix} \sigma_0^2 & \alpha_0 \\ \alpha_0 & \tau_0^2 \end{pmatrix}$. Then, for any function q , positive on $(0, 1]$ and such that $\lim_{t \downarrow 0} t^{1/2}/q(t) = 0$, we have, as $n \rightarrow \infty$,*

$$\sup_{\frac{1}{n} \leq t \leq 1} \|L_n(t)\|/q(t) \rightarrow 0 \text{ in probability.}$$

Proof. For each $n \geq 1$, let $A_n^{(2)}(t) = [-\log(nt), \log(nt)]^2$, $\frac{1}{n} \leq t \leq 1$. Hence,

$$\begin{aligned} L_n(t) &= \int_{A_n^{(2)}(t)} l(x, y, \theta_0) d_x d_y \epsilon_{2n}(x, y, t) \\ &\quad + \int_{[A_n^{(2)}(t)]^c} l(x, y, \theta_0) d_x d_y n^{-1/2} \sum_{i=1}^{[nt]} (1\{(X_i, Y_i) \leq (x, y)\}) - F(x, y; \theta_0) \\ (5.2.12) \quad &\quad - \int_{[A_n^{(2)}(t)]^c} l(x, y, \theta_0) d_x d_y n^{-1/2} K_{F_0}(x, y, nt) \\ &:= L_{1n}(t) + L_{2n}(t) - L_{3n}(t), \end{aligned}$$

where A^c denotes the complement of A .

Consider the rectangle $R = [a, a'] \times [b, b'] \in \mathbb{R}^2$. We have, by using integration by parts, for any $j = 1, \dots, 5$,

(5.2.13)

$$\begin{aligned} & \int_a^{a'} l_j^{(1)}(x, \theta_0) dx \left[\int_b^{b'} l_j^{(2)}(y, \theta_0) dy \varepsilon_{2n}(x, y, t) \right] \\ &= l_j^{(1)}(a') l_j^{(2)}(b') \varepsilon_{2n}(a', b', t) - l_j^{(1)}(a) l_j^{(2)}(b') \varepsilon_{2n}(a, b', t) - l_j^{(1)}(a') l_j^{(2)}(b) \varepsilon_{2n}(a', b, t) \\ & \quad + l_j^{(1)}(a) l_j^{(2)}(b) \varepsilon_{2n}(a, b, t) - l_j^{(1)}(a') \int_b^{b'} \varepsilon_{2n}(a', y, t) dy l_j^{(2)}(y) \\ & \quad + l_j^{(1)}(a) \int_b^{b'} \varepsilon_{2n}(a, y, t) dy l_j^{(2)}(y) + \int_a^{a'} \int_b^{b'} \varepsilon_{2n}(x, y, t) dx l_j^{(1)}(x) dy l_j^{(2)}(y). \end{aligned}$$

By (5.2.3), it is clear that, for some constant $C > 0$,

$$(5.2.14) \quad \sup_{(x, y) \in A_n^{(2)}(t)} |l_j(x, \theta_0)| \leq C \log^2(nt).$$

By Theorem 5.1.A, with $d = 2$, as $(nt) \rightarrow \infty$,

$$(5.2.15) \quad \sup_{(x, y) \in \mathbb{K}^2} |\varepsilon_{2n}(x, y, t)| \stackrel{a.s.}{=} O(n^{-1/2}(nt)^{3/8}(\log(nt))^{3/2}).$$

Hence, for any fixed $0 < \delta < 1$,

$$\begin{aligned} \sup_{\delta \leq t \leq 1} \|L_{1n}(t)\| & \stackrel{a.s.}{=} \sup_{\delta \leq t \leq 1} n^{-1/2}(nt)^{3/8}(\log(nt))^{7/2} \\ & = n^{-1/8}(\log n)^{7/2} \longrightarrow 0, \text{ as } n \rightarrow \infty. \end{aligned}$$

Thus, since $\inf_{\delta \leq t \leq 1} q(t) > 0$, we have

$$\sup_{\delta \leq t \leq 1} \|L_{1n}(t)\|/q(t) = o_P(1), \text{ as } n \rightarrow \infty.$$

Also,

$$\begin{aligned}
 \sup_{\frac{1}{n} \leq t \leq \delta} \|L_{1n}(t)\|/q(t) &\leq \sup_{0 < t \leq \delta} t^{1/2}/q(t) \sup_{\frac{1}{n} \leq t \leq \delta} \|L_{1n}(t)\|/t^{1/2} \\
 &\stackrel{a.s.}{=} \sup_{0 < t \leq \delta} t^{1/2}/q(t) \sup_{\frac{1}{n} \leq t \leq 1} (nt)^{-1/8} (\log(nt))^{7/2} \\
 (5.2.16) \quad &\stackrel{a.s.}{=} \sup_{0 < t \leq \delta} t^{1/2}/q(t) \cdot O(1).
 \end{aligned}$$

Since q is such that $\lim_{t \downarrow 0} t^{1/2}/q(t) = 0$, taking $\delta > 0$ sufficiently small, we get that, as $n \rightarrow \infty$,

$$\sup_{\frac{1}{n} \leq t \leq \delta} \|L_{1n}(t)\|/q(t) = o_P(1).$$

Next, we note that, since for any $h : \mathbb{R}^2 \rightarrow \mathbb{R}$,

$$\int_{\mathbb{R}^2} h(x, y) d_x d_y 1\{(X_i, Y_i) \leq (x, y)\} = h(X_i, Y_i),$$

we have

$$E\left(\int_{A_2^+(k)} l_j(x, y, \theta_0) d_x d_y (1\{(X_i, Y_i) \leq (x, y)\} - F(x, y; \theta_0))\right) = 0$$

and

$$\begin{aligned}
 &E\left(\int_{A_2^+(k)} l_j(x, y, \theta_0) d_x d_y (1\{(X_i, Y_i) \leq (x, y)\} - F(x, y; \theta_0))\right)^2 \\
 (5.2.17) \quad &\leq \int_{A_2^+(k)} [l_j(x, y, \theta_0)]^2 d_x d_y F(x, y; \theta_0), \text{ for each } j = 1, \dots, 5.
 \end{aligned}$$

For any fixed $0 < \delta < 1$, we have

$$P\left\{\sup_{\frac{1}{n} \leq t \leq \delta} \|L_{2n}(t)\|/q(t) > \epsilon\right\} = P\left\{\sup_{\frac{1}{n} \leq t \leq \delta} \frac{\|L_{2n}(t)\|}{t^{1/2}} \sup_{\frac{1}{n} \leq t \leq \delta} t^{1/2}/q(t) > \epsilon\right\}$$

$$\leq P\left\{\max_{1 \leq [nt] \leq [n\delta]} \left| \left[[nt]^{-1/2} \sum_{i=1}^{[nt]} \int_{[A_n^{(2)}(t)]^c} l(x, y, \theta_0) d_x d_y (1\{(X_i, Y_i) \leq (x, y)\} - F(x, y; \theta_0)) \right] \right| > \epsilon \left(\sup_{\frac{1}{n} \leq t \leq \delta} t^{1/2}/q(t) \right)^{-1}\right\}.$$

Thus, by (5.2.17) and the Hájek-Rényi inequality, we have

$$\begin{aligned} & P\left\{\sup_{\frac{1}{n} \leq t \leq \delta} \|L_{2n}(t)\|/q(t) > \epsilon\right\} \\ (5.2.18) \quad & \leq \left(\sup_{\frac{1}{n} \leq t \leq \delta} t^{1/2}/q(t)\right)^2 \sum_{k=1}^{[n\delta]} \frac{1}{\epsilon^2} \sum_{j=1}^5 \sum_{i=1}^k \int_{A_2^*(k)} [l_j(x, y, \theta_0)]^2 d_x d_y F(x, y; \theta_0) \\ & \leq \left(\sup_{\frac{1}{n} \leq t \leq \delta} t^{1/2}/q(t)\right)^2 \frac{1}{\epsilon^2} \sum_{j=1}^5 \sum_{k=1}^n \frac{1}{k} \int_{A_2^*(k)} [l_j(x, y, \theta_0)]^2 d_x d_y F(x, y; \theta_0). \end{aligned}$$

Also, by (5.2.17) and Kolmogorov's inequality, we have

$$\begin{aligned} & P\left\{\sup_{\delta \leq t \leq 1} \|L_{2n}(t)\|/q(t) > \epsilon\right\} \\ (5.2.19) \quad & \leq P\left\{\max_{[n\delta] \leq [nt] \leq n} \left| \sum_{i=1}^{[nt]} \int_{[A_n^{(2)}(t)]^c} l(x, y, \theta_0) d_x d_y (1\{(X_i, Y_i) \leq (x, y)\} - F(x, y; \theta_0)) \right| > \epsilon n^{1/2} \inf_{\delta \leq t \leq 1} q(t) \right\} \\ & \leq \left(\inf_{\delta \leq t \leq 1} q(t)\right)^{-2} \frac{1}{\epsilon^2} \sum_{j=1}^5 \sum_{k=[n\delta]}^n \frac{1}{k} \int_{A_2^*(k)} [l_j(x, y, \theta_0)]^2 d_x d_y F(x, y; \theta_0). \end{aligned}$$

Now, since $\lim_{t \downarrow 0} t^{1/2}/q(t) = 0$, by (5.1.15), and Lemma 5.2.1, it follows that, as $n \rightarrow \infty$,

$$\sup_{\frac{1}{n} \leq t \leq 1} \|L_{2n}(t)\|/q(t) = o_P(1).$$

Let $\{B_{F_0}(x, y), x, y \geq 0\}$ be a Brownian bridge associated with the distribution function $F(\cdot; \theta_0)$, i.e. B_{F_0} is a Gaussian process such that $E B_{F_0}(x, y) = 0$, and

$$E B_{F_0}(x, y) B_{F_0}(z, w) = F(x \wedge z, y \wedge w; \theta_0) - F(x, y; \theta_0) F(z, w; \theta_0).$$

Since, for each t and n fixed,

$$n^{-1/2} K_{F_0}(x, y, nt) \stackrel{D}{=} \left(\frac{t}{[nt]} \right)^{1/2} \sum_{i=1}^{[nt]} B_{F_0}^{(i)}(x, y),$$

where $\{B_{F_0}^{(i)}\}$ is a family of independent Brownian bridges associated with $F(\cdot, \cdot; \theta_0)$, it follows, by (A.3.6) (cf. Appendix A.3), that inequalities (5.2.17) and (5.2.18) hold also for $L_{3n}(t)$. Hence, as $n \rightarrow \infty$,

$$\sup_{\frac{1}{n} \leq t \leq 1} \|L_{3n}(t)\|/q(t) = o_P(1).$$

Proof of Theorem 5.2.1. Adding and subtracting $F(x, y; \theta_0)$ and applying a one term Taylor expansion around θ_0 , we have,

(5.2.20)

$$\begin{aligned} \hat{\beta}_n(x, y, t) &= n^{-1/2} \sum_{i=1}^{[nt]} (1\{(X_i, Y_i) \leq (x, y)\} - F(x, y; \theta_0)) \\ &\quad - [nt]n^{-1/2} (F(x, y; \hat{\theta}_{[nt]}) - F(x, y; \theta_0)) \\ &:= \beta_n(x, y, t) + [nt]n^{-1/2} (\hat{\theta}_{[nt]} - \theta_0) \cdot \left(\nabla_{\theta} F(x, y; \theta_0)^\top \right. \\ &\quad \left. - \nabla_{\theta} F(x, y; \hat{\theta}_{[nt]})^\top \right) - [nt]n^{-1/2} (\hat{\theta}_{[nt]} - \theta_0) \cdot \nabla_{\theta} F(x, y; \theta_0)^\top, \end{aligned}$$

where

$$(5.2.21) \quad \sup_{0 \leq t \leq 1} \|\hat{\theta}_{[nt]} - \theta_0\| \leq \sup_{0 \leq t \leq 1} \|\hat{\theta}_{[nt]} - \theta_0\|,$$

and $\beta_n(x, y, t)$ is as in (5.1.5), with $d = 2$.

Since $El(X, Y, \theta_0) = 0$, by (5.2.2), we have

$$\begin{aligned}
 (5.2.22) \quad n^{-1/2}[nt](\hat{\theta}_{[nt]} - \theta_0) &= \sum_{i=1}^{[nt]} l(X_i, Y_i, \theta_0) + n^{-1/2}[nt]^{1/2}\varepsilon_{1n}(t) \\
 &= \int_{\mathbb{R}^2} l(x, y, \theta_0) d_x d_y n^{-1/2} \left[\sum_{i=1}^{[nt]} (1\{(X_i, Y_i) \leq (x, y)\}) - F(x, y; \theta_0) \right] \\
 &\quad - n^{-1/2}K_{F_0}(x, y, nt)] \\
 &\quad + \int_{\mathbb{R}^2} l(x, y, \theta_0) d_x d_y n^{-1/2} K_{F_0}(x, y, nt) + n^{-1/2}[nt]^{1/2}\varepsilon_{1n}(t).
 \end{aligned}$$

From (5.2.5), (5.2.20), and (5.2.22) we have

$$\begin{aligned}
 (5.2.23) \quad \hat{\beta}_n(x, y, t) &= \beta_n(x, y, t) - n^{-1/2}K_{F_0}(x, y, nt) + G_{F_0}(x, y, nt) \\
 &\quad + [nt]^{1/2}n^{-1/2}\varepsilon_{3n}(x, y, t) - L_n(t) \cdot \nabla_{\theta}F(x, y; \theta_0)^{\top} \\
 &\quad - [nt]^{1/2}n^{-1/2}\varepsilon_{1n}(t) \cdot \nabla_{\theta}F(x, y; \theta_0)^{\top},
 \end{aligned}$$

where

$$(5.2.24) \quad \varepsilon_{3n}(x, y, t) = [nt]^{1/2}(\hat{\theta}_{[nt]} - \theta_0) \cdot (\nabla_{\theta}F(x, y; \theta_0)^{\top} - \nabla_{\theta}F(x, y; \hat{\theta}_{[nt]})^{\top}).$$

We first assume that q is a positive function on $(0, 1]$, and such that $\lim_{t \downarrow 0} t^{1/2}/q(t) = 0$. Let $\varepsilon_{1n}^{(j)}(t)$ denote the j th component ($j = 1, \dots, 5$) of $\varepsilon_{1n}(t)$. Comparing (5.2.1) and (5.2.4) with (3.2.1) and (3.2.3), we conclude that, by (3.2.26), (3.2.29), (5.1.15) and the fact that the vector $\nabla_{\theta}F(x, y; \theta_0)$ is uniformly bounded in (x, y) , in sup-norm, in order to have that, as $n \rightarrow \infty$,

$$(5.2.25) \quad \sup_{\frac{1}{n} \leq t \leq 1} \sup_{(x, y) \in \mathbb{R}^2} n^{-1/2}[nt]^{1/2} |\varepsilon_{1n}(t) \cdot \nabla_{\theta}F(x, y; \theta_0)^{\top}| / q(t) = o_P(1),$$

it suffices to show (cf. (5.2.4)) that, as $n \rightarrow \infty$,

(5.2.26)

$$\begin{aligned} & \sup_{\frac{1}{n} \leq t \leq 1} n^{-1/2} [nt]^{1/2} \|\varepsilon_{1n}^{(5)}(t)\| / q(t) \\ & := \sup_{\frac{1}{n} \leq t \leq 1} n^{-1/2} [nt]^{1/2} \|[nt]^{1/2} (\bar{X}_{[nt]} - \zeta_0)(\bar{Y}_n(t) - \eta_0)\| / q(t) = o_P(1). \end{aligned}$$

To this end, we note that, by (5.2.1) and the Hájek-Rényi inequality (cf. proof of (3.2.28), with $\delta = 1$) we have that, as $n \rightarrow \infty$,

$$\begin{aligned} & \sup_{\frac{1}{n} \leq t \leq 1} [nt]^{1/4} |\bar{X}_{[nt]} - \zeta_0| := \sup_{\frac{1}{n} \leq t \leq 1} |[nt]^{-3/4} \sum_{i=1}^{[nt]} (X_i - \mu_0)| \\ (5.2.27) \quad & = O_P(1). \end{aligned}$$

Also, by (5.2.1) and proofs of (3.2.26), (3.2.27), (3.2.28), (3.2.29) and (3.2.30), we have that, as $n \rightarrow \infty$,

$$(5.2.28) \quad \sup_{\frac{1}{n} \leq t \leq 1} n^{-1/2} [nt]^{1/2} / q(t) \sup_{\frac{1}{n} \leq t \leq 1} |[nt]^{-3/4} \sum_{i=1}^{[nt]} (Y_i - \eta_0)| = o_P(1).$$

Thus, by (5.2.27) and (5.2.28), we get that (5.2.26) holds.

Similarly, in order to show that, as $n \rightarrow \infty$,

$$(5.2.29) \quad \sup_{\frac{1}{n} \leq t \leq 1} \sup_{(x,y) \in \mathbb{R}^2} n^{-1/2} [nt]^{1/2} |\varepsilon_{3n}(x,y,t)| / q(t) = o_P(1),$$

it suffices to show that, as $n \rightarrow \infty$,

(5.2.30)

$$\begin{aligned} & \sup_{\frac{1}{n} \leq t \leq 1} \sup_{(x,y) \in \mathbb{R}^2} n^{-1/2} [nt]^{1/2} |\varepsilon_{3n}^{(5)}(x,y,t)| / q(t) \\ & := \sup_{\frac{1}{n} \leq t \leq 1} \sup_{(x,y) \in \mathbb{R}^2} [nt]^{1/2} |(\hat{\alpha}_{[nt]} - \alpha_0) \cdot \frac{\partial}{\partial \alpha} (F(x,y; \theta_0) - F(x,y; \tilde{\theta}_{[nt]}))| / q(t) \\ & = o_P(1), \end{aligned}$$

where $\epsilon_{3n}^{(5)}(t)$ denotes the 5th component of $\epsilon_{3n}(x, y, t)$ (cf. 5.2.24).

We note that, by (5.2.2),

$$(5.2.31) \quad [nt]^{1/2} \hat{\alpha}_{[nt]} - \alpha_0 = [nt]^{-1/2} \sum_{i=1}^{[nt]} l_5(X_i, Y_i, \theta_0) + \epsilon_{1n}^{(5)}(t).$$

Furthermore, by (5.2.4), (5.2.27) and (5.2.28), we have that, for any fixed $0 < \delta < 1$, as $n \rightarrow \infty$,

$$(5.2.32) \quad \sup_{\frac{1}{n} \leq t \leq \delta} \|\epsilon_{1n}^{(5)}(t)\| = O_P(1)$$

and

$$(5.2.33) \quad \sup_{\delta \leq t \leq 1} |\epsilon_{1n}^{(5)}(t)| = o_P(1).$$

Hence, in virtue of (5.2.21), (5.2.31), (5.3.32) and (5.2.33), the proof of (5.2.30) is similar, line by line (cf. (3.4.20)-(3.4.29)), to that of (3.4.29), taking there $F''_{\theta\theta}(x, y; \theta)$ as the 5×5 matrix of second partial derivatives of $F(x, y; \theta)$ with respect to $\theta = (\theta_1, \dots, \theta_5)$. We note now that, from (5.2.23), by (5.1.20) (cf. remark 5.1.2), Lemma 5.2.2, (5.2.25) and (5.2.29), we have, as $n \rightarrow \infty$,

$$\sup_{\frac{1}{n} \leq t \leq 1} \sup_{(x, y) \in \mathbb{R}^2} |\hat{\beta}_n(x, y, t) - G_{F_0}(x, y, nt)|/q(t) \stackrel{a.s.}{=} o(1),$$

provided only that q be positive on $(0, 1]$ and such that $\lim_{t \downarrow 0} t^{1/2}/q(t) = 0$.

By definition,

$$(5.2.34) \quad \hat{\beta}_n(x, y, t) = 0 \text{ for all } (x, y) \in \mathbb{R}^2 \text{ and } 0 < t < \frac{1}{n}.$$

Hence, to have (a) and (b) of Theorem 5.2.1, it suffices to show that

$$(5.2.35) \quad \sup_{0 < t < \frac{1}{n}} \sup_{(x,y) \in \mathbb{R}^2} |G_{F_0}(x,y,nt)|/q(t) = \begin{cases} o_P(1), & \text{if (5.2.7) holds,} \\ O_P(1), & \text{if (5.2.8) holds.} \end{cases}$$

To this end, we note that, similarly to the one dimensional case, if $M(\theta_0)$ is the nonnegative definite matrix defined by

$$(5.2.36) \quad M(\theta_0) = E\{l(X_j, Y_j, \theta_0)^T l(X_j, Y_j, \theta_0)\},$$

where $l(\cdot, \cdot, \theta_0)$ is as in (5.2.3), there is a nonsingular matrix $D(\theta_0)$ such that

$$(5.2.37) \quad D(\theta_0)^T M(\theta_0) D(\theta_0) = I,$$

where I is the 5×5 identity matrix.

Consider now the Gaussian process $\{B_{F_0}(x, y), x, y \geq 0\}$ with $E B_{F_0}(x, y) = 0$, and

$$E B_{F_0}(x, y) B_{F_0}(z, w) = F(x \wedge z, y \wedge w; \theta_0) - F(x, y; \theta_0) F(z, w; \theta_0).$$

For each $n \geq 1$ and $0 \leq t \leq 1$,

$$n^{-1/2} K_{F_0}(x, y, nt) \stackrel{D}{=} t^{1/2} B_{F_0}(x, y), \quad (x, y) \in \mathbb{R}^2.$$

Using integration by parts (cf. A.2.6), since $E l(X_i, Y_i, \theta_0) = 0$, we have that, for each $n \geq 1$ and $0 \leq s, t \leq 1$,

$$(5.2.38) \quad E \left\{ \int l(x, y, \theta_0) d_x d_y n^{-1/2} K_{F_0}(x, y, ns) \right\}^T \cdot \left\{ \int l(x, y, \theta_0) d_x d_y n^{-1/2} K_{F_0}(x, y, nt) \right\} \\ = (s \wedge t) M(\theta_0).$$

Hence, from (5.2.5), we have that

(5.2.39)

$$G_{F_0,n}(x, y, t) = n^{-1/2} K_{F_0}(x, y, nt) - n^{-1/2} W(nt) \cdot D(\theta_0)^{-1} \cdot \nabla_{\theta} F(x, y; \theta_0)^T,$$

where

$$(5.2.40) \quad n^{-1/2} W(nt) = \int l(x, y, \theta_0) d_x d_y n^{-1/2} K_{F_0}(x, y, nt) \cdot D(\theta_0)$$

is a vector valued Wiener process with covariance function $(s \wedge t)$ multiplied by (5.2.37), for each $n \geq 1$.

Next, since

$$\sup_{0 < t \leq 1} |n^{-1/2} W(nt)|/q(t) \stackrel{D}{=} \sup_{0 < t \leq 1} |W(t)|/q(t)$$

for each $n \geq 1$, by Theorem 2.1.B (cf. our Remark 2.1.5), we have that

$$(5.2.41) \quad \sup_{0 < t < \frac{1}{n}} |n^{-1/2} W(nt)|/q(t) = \begin{cases} o_P(1), & \text{if (5.2.7) holds,} \\ O_P(1), & \text{if (5.2.8) holds.} \end{cases}$$

Similarly, since, for each $n \geq 1$,

$$\sup_{0 < t < \frac{1}{n}} \sup_{(x, y) \in \mathbb{R}^2} |n^{-1/2} K_{F_0}(x, y, nt)| \stackrel{D}{=} \sup_{0 < t < \frac{1}{n}} \sup_{(x, y) \in \mathbb{R}^2} |K_{F_0}(x, y, t)|,$$

by theorem 5.1.C, with $d = 2$ (cf. Remark 5.1.2), we have,

$$(5.2.42) \quad \sup_{0 < t < \frac{1}{n}} \sup_{(x, y) \in \mathbb{R}^2} |n^{-1/2} K_{F_0}(x, y, nt)|/q(t) = \begin{cases} o_P(1), & \text{if (5.2.7) holds,} \\ O_P(1), & \text{if (5.2.8) holds.} \end{cases}$$

Now, (5.2.35) follows from (5.2.39), (5.2.41) and (5.2.42).

The proof of (c) is similar to the proof of (c) of Theorem 3.2.1. ■

3

OF/DE

3

PM-1 3½"x4" PHOTOGRAPHIC MICROCOPY TARGET
NBS 1010a ANSI/ISO #2 EQUIVALENT

PRECISION™ RESOLUTION TARGETS

Corollary 5.2.1. *Assuming the conditions of Theorem 5.2.1, we have, as $n \rightarrow \infty$,*

(a) *if $q \in Q$ is such that (5.2.7) holds, then*

$$(5.2.43) \quad \hat{\beta}_n(x, y, t)/q(t) \xrightarrow{D} G_{F_0}(x, y, t)/q(t)$$

in $D(\mathbb{R}^2 \times [0, 1])$,

(b) *if q is positive on $(0, 1]$, then (5.2.43) holds in $D(\mathbb{R}^2 \times [\delta, 1])$ for any $0 < \delta < 1$.*

5.3 Sequentially estimated bivariate normal empirical processes tied down at $t=1$.

In an analogous way to the univariate case of Section 3 of Chapter 3, we propose, for the change-point problem, to study an estimated "bridge" empirical process.

We introduce the following:

$$(5.3.1) \quad \hat{\beta}_n(x, y, t) = \begin{cases} n^{1/2} \frac{[(n+1)t]}{n} \left(1 - \frac{[(n+1)t]}{n}\right) \hat{\zeta}_n(x, y, t), & 0 \leq t < 1, (x, y) \in \mathbb{R}^2, \\ 0, & t = 1, (x, y) \in \mathbb{R}^2, \end{cases}$$

where

$$(5.3.2) \quad \begin{aligned} \hat{\zeta}_n(x, y, t) = & \frac{1}{[(n+1)t]} \sum_{i=1}^{[(n+1)t]} \left(\mathbf{1}\{(X_i, Y_i) \leq (x, y)\} - F(x, y; \hat{\theta}'_{[nt]}) \right) \\ & - \frac{1}{n - [(n+1)t]} \sum_{i=[(n+1)t]+1}^n \left(\mathbf{1}\{(X_i, Y_i) \leq (x, y)\} - F(x, y; \hat{\theta}''_{[nt]}) \right), \end{aligned}$$

and, for each $0 < t < 1$, $\{\hat{\theta}'_{[nt]}\}$ and $\{\hat{\theta}''_{[nt]}\}$ are sequences of maximum likelihood estimators of θ based on the samples $(X_1, Y_1), \dots, (X_{[(n+1)t]}, Y_{[(n+1)t]})$ and $(X_{[(n+1)t]+1}, Y_{[(n+1)t]+1}), \dots, (X_n, Y_n)$, respectively. Here, $F(x, y; \theta)$ denotes the bivariate normal distribution function with parameter $\theta := (\zeta, \eta, \sigma^2, \tau^2, \alpha)$ as in Section 5.2.

It can be checked that

$$(5.3.3) \quad \hat{\theta}'_{[nt]} = \hat{\theta}_{[(n+1)t]},$$

where $\hat{\theta}_{[t]}$ is as in (5.2.1), and

$$(5.3.4) \quad \hat{\theta}''_{[nt]} := \left(\bar{X}_{[nt]}, \bar{Y}_{[nt]}, \hat{\sigma}_{[nt]}^2, \hat{\tau}_{[nt]}^2, \hat{\alpha}_{[nt]} \right),$$

where

$$\bar{X}_{[nt]} = \frac{1}{n - [(n+1)t]} \sum_{i=[(n+1)t]+1}^n X_i, \quad \bar{Y}_{[nt]} = \frac{1}{n - [(n+1)t]} \sum_{i=[(n+1)t]+1}^n Y_i,$$

$$\hat{\sigma}_{[nt]}^2 = \frac{1}{n - [(n+1)t]} \sum_{i=[(n+1)t]+1}^n (X_i - \bar{X}_{[nt]})^2,$$

$$\hat{\tau}_{[nt]}^2 = \frac{1}{n - [(n+1)t]} \sum_{i=[(n+1)t]+1}^n (Y_i - \bar{Y}_{[nt]})^2,$$

and

$$\hat{\alpha}_{[nt]} = \frac{1}{n - [(n+1)t]} \sum_{i=[(n+1)t]+1}^n (X_i - \bar{X}_{[nt]})(Y_i - \bar{Y}_{[nt]}).$$

The process $\{\hat{\beta}_n(x, y, t), 0 \leq t \leq 1, (x, y) \in \mathbb{R}^2\}$ defined by (5.3.1) will be approximated, in $\|\cdot\|_q$ -metrics, by the sequence of stochastic processes $\{\Psi_{F_0, n}(x, y, t), 0 \leq t \leq 1, (x, y) \in \mathbb{R}^2\}$ defined by

(5.3.5)

$$\Psi_{F_0,n}(x, y, t) = \Gamma_{F_0,n}(x, y, t) - \left\{ \int l(x, y, \theta_0) d_x d_y \Gamma_{F_0,n}(x, y, t) \right\} \cdot \nabla_{\theta} F(x; \theta_0)^{\top},$$

where $\Gamma_{F_0,n}(\cdot, \cdot)$ is as in (5.1.31), with $d = 2$, and $F = F_0 \equiv F(\cdot, \cdot; \theta_0)$, the distribution function of a bivariate normal random variable with parameter θ_0 .

We have

Theorem 5.3.1. *Let $(X_1, Y_1), (X_2, Y_2), \dots$ be independent bivariate normal random vectors with vector of means (ζ, η) and covariance matrix $\begin{pmatrix} \sigma^2 & \alpha \\ \alpha & \tau^2 \end{pmatrix}$. Suppose that $\{\hat{\theta}'_{[nt]}\}$ and $\{\hat{\theta}''_{[nt]}\}$ are as in (5.3.3) and (5.3.4), and let $\{\hat{\beta}_n(x, y, t)\}$ be the sequentially estimated bivariate normal empirical process, tied down at $t = 1$, defined as in (5.3.1). Then, there exists a Kiefer process $K(\cdot, \cdot)$ such that with the sequence of Gaussian processes $\Psi_{F_0,n}$, and a weight function $q \in \mathcal{Q}^*$, we have, as $n \rightarrow \infty$,*

(a) if

$$(5.3.6) \quad I_{0,1}(q, c) < \infty \text{ for all } c > 0,$$

then

$$\sup_{0 < t < 1} \sup_{(x, y) \in \mathbb{R}^2} |\hat{\beta}_n(x, y, t) - \Psi_{F_0,n}(x, y, t)|/q(t) = o_P(1),$$

(b) if

$$(5.3.7) \quad I_0(q, c) < \infty \text{ for some } c > 0,$$

then

$$\sup_{0 < t < 1} \sup_{(x, y) \in \mathbb{R}^2} |\hat{\beta}_n(x, y, t) - \Psi_{F_0,n}(x, y, t)|/q(t) = O_P(1),$$

(c) if (5.3.6) holds then

$$\sup_{0 < t < 1} \sup_{(x, y) \in \mathbb{R}^2} |\hat{\beta}_n(x, y, t)|/q(t) \xrightarrow{P} \sup_{0 < t < 1} \sup_{(x, y) \in \mathbb{R}^2} |\Psi_{F_0}(x, y, t)|/q(t),$$

where the process $\{\Psi_{F_0}(x, y, t), 0 \leq t \leq 1, (x, y) \in \mathbb{R}^2\}$ is defined by

$$(5.3.8) \quad \Psi_{F_0}(x, y, t) = \Gamma_{F_0}(x, y, t) - \left\{ \int l(x, y, \theta_0) d_x d_y \Gamma_{F_0}(x, y, t) \right\} \cdot \nabla_{\theta} F(x; \theta_0)^{\top}.$$

Proof of Theorem 5.3.1. Using the usual decomposition, we have

(5.3.9)

$$\begin{aligned} \hat{\beta}_n(x, y, t) &= \beta_n^*(x, y, t) - n^{-1/2} \left\{ \left(1 - \frac{[(n+1)t]}{n} \right) \sum_{i=1}^{[(n+1)t]} l(X_i, Y_i, \theta_0) \right. \\ &\quad \left. - \frac{[(n+1)t]}{n} \sum_{i=[(n+1)t]+1}^n l(X_i, Y_i, \theta_0) \right\} \cdot \nabla_{\theta} F(x, y; \theta_0)^{\top} \\ &\quad + \frac{[(n+1)t]}{n} \left(1 - \frac{[(n+1)t]}{n} \right) n^{1/2} \{ \varepsilon'_{3n}(x, y, t) + \varepsilon''_{3n}(x, y, t) \} \\ &\quad + \frac{[(n+1)t]}{n} n^{-1/2} (n - [(n+1)t])^{\frac{1}{2}} \varepsilon''_{1n}(t) \cdot \nabla_{\theta} F(x, y; \theta_0)^{\top} \\ &\quad - \left(1 - \frac{[(n+1)t]}{n} \right) n^{-1/2} [(n+1)t]^{\frac{1}{2}} \varepsilon'_{1n}(t) \cdot \nabla_{\theta} F(x, y; \theta_0)^{\top}, \end{aligned}$$

where

$$\beta_n^*(x, y, t) = \begin{cases} n^{-1/2} \left(\sum_{i=1}^{[(n+1)t]} \mathbf{1}\{(X_i, Y_i) \leq (x, y)\} \right. \\ \quad \left. - \frac{[(n+1)t]}{n} \sum_{i=1}^n \mathbf{1}\{(X_i, Y_i) \leq (x, y)\} \right), \\ \text{for } 0 \leq t < 1, (x, y) \in \mathbb{R}^2, \\ 0, \text{ for } t = 1 \end{cases}$$

(cf. (5.1.22), with $d = 2$); $l(x, y, \theta_0)$ is as in (5.2.3);

$$\varepsilon'_{1n}(t) = \left(0, 0, -[(n+1)t]^{1/2}(\bar{X}_{[nt]} - \zeta_0)^2, -[(n+1)t]^{1/2}(\bar{Y}_{[nt]} - \eta_0)^2, \varepsilon'_n(t)\right),$$

with

$$(5.3.10) \quad \varepsilon'_n(t) = [(n+1)t]^{1/2}(\bar{X}_{[nt]} - \zeta_0)(\bar{Y}_{[nt]} - \eta_0);$$

$$\varepsilon'_{2n}(t) = \left(0, 0, -(n - [(n+1)t])^{1/2}(\bar{\bar{X}}_{[nt]} - \zeta_0)^2, -(n - [(n+1)t])^{1/2}(\bar{\bar{Y}}_{[nt]} - \eta_0)^2, \varepsilon''_n(t)\right),$$

with

$$(5.3.11) \quad \varepsilon''_n(t) = -(n - [(n+1)t])^{1/2}(\bar{\bar{X}}_{[nt]} - \zeta_0)(\bar{\bar{Y}}_{[nt]} - \eta_0);$$

$$(5.3.12) \quad \varepsilon'_{3n}(x, y, t) = (\hat{\theta}'_{[nt]} - \theta_0) \cdot \left(\nabla_{\theta} F(x, y; \hat{\theta}'_{[nt]})^{\top} - \nabla_{\theta} F(x, y; \theta_0)^{\top} \right),$$

with

$$(5.3.13) \quad \sup_{0 \leq t \leq 1} \|\hat{\theta}'_{[nt]} - \theta_0\| \leq \sup_{0 \leq t \leq 1} \|\hat{\theta}'_{[nt]} - \theta_0\|;$$

and

$$(5.3.14) \quad \varepsilon''_{3n}(x, y, t) = (\hat{\theta}''_{[nt]} - \theta_0) \cdot \left(\nabla_{\theta} F(x, y; \hat{\theta}''_{[nt]})^{\top} - \nabla_{\theta} F(x, y; \theta_0)^{\top} \right),$$

with

$$(5.3.15) \quad \sup_{0 \leq t \leq 1} \|\hat{\theta}''_{[nt]} - \theta_0\| \leq \sup_{0 \leq t \leq 1} \|\hat{\theta}''_{[nt]} - \theta_0\|.$$

Let

$$(5.3.16) \quad \varphi_n(t) = n^{-1/2} \left(\left(1 - \frac{[(n+1)t]}{n}\right) \sum_{i=1}^{[(n+1)t]} l(X_i, Y_i, \theta_0) - \frac{[(n+1)t]}{n} \sum_{i=[(n+1)t]+1}^n l(X_i, Y_i, \theta_0) \right).$$

Adding and subtracting $\Psi_{F_0, n}(x, y, t)$ (cf. (5.3.5)) in (5.3.9), we get

$$(5.3.17) \quad \begin{aligned} & \sup_{\frac{1}{n+1} \leq t \leq \frac{n}{n+1}} \sup_{(x, y) \in \mathbb{R}^2} |\hat{\beta}_n(x, y, t) - \Psi_{F_0, n}(x, y, t)|/q(t) \\ & \leq \sup_{\frac{1}{n+1} \leq t \leq \frac{n}{n+1}} \sup_{(x, y) \in \mathbb{R}^2} |\beta_n^*(x, y, t) - \Psi_{F_0, n}(x, y, t)|/q(t) \\ & + \sup_{\frac{1}{n+1} \leq t \leq \frac{n}{n+1}} \sup_{(x, y) \in \mathbb{R}^2} |\{\varphi_n(t) - \int l(x, y, \theta_0) d_x d_y \Gamma_{F_0, n}(x, y, t)\} \cdot \nabla_{\theta} F(x, y; \theta_0)^T|/q(t) \\ & + \sup_{\frac{1}{n+1} \leq t \leq \frac{n}{n+1}} \sup_{(x, y) \in \mathbb{R}^2} \frac{[(n+1)t]}{n} \left(1 - \frac{[(n+1)t]}{n}\right) n^{1/2} |\varepsilon'_{3n}(x, y, t) + \varepsilon''_{3n}(x, y, t)|/q(t) \\ & + \sup_{\frac{1}{n+1} \leq t \leq \frac{n}{n+1}} \sup_{(x, y) \in \mathbb{R}^2} \frac{[(n+1)t]}{n} n^{-1/2} (n - [(n+1)t])^{\frac{1}{2}} |\varepsilon''_{1n}(t) \cdot \nabla_{\theta} F(x, y; \theta_0)^T|/q(t) \\ & + \sup_{\frac{1}{n+1} \leq t \leq \frac{n}{n+1}} \sup_{(x, y) \in \mathbb{R}^2} \left(1 - \frac{[(n+1)t]}{n}\right) n^{-1/2} [(n+1)t]^{\frac{1}{2}} |\varepsilon'_{1n}(t) \cdot \nabla_{\theta} F(x, y; \theta_0)^T|/q(t). \end{aligned}$$

We first assume that q is a positive function on $(0, 1)$ and such that (3.3.27) holds, i.e., $\lim_{t \downarrow 0} t^{1/2}/q(t) = \lim_{t \uparrow 1} (1-t)^{1/2}/q(t) = 0$. We note that, since $\varepsilon'_n(t) = \varepsilon'_{1(n+1)}(t)$, where $\varepsilon_{1n}(t)$ is as in (5.2.4), by (3.3.27) and (5.2.25), we have, as $n \rightarrow \infty$,

$$(5.3.18) \quad \begin{aligned} & \sup_{\frac{1}{n+1} \leq t \leq \frac{n}{n+1}} \left(1 - \frac{[(n+1)t]}{n}\right) n^{-1/2} [(n+1)t]^{1/2} \sup_{(x, y) \in \mathbb{R}^2} |\varepsilon'_{1n}(t) \cdot \nabla_{\theta} F(x, y; \theta_0)^T|/q(t) \\ & = o_P(1). \end{aligned}$$

Similarly, by symmetry and (3.3.27) (cf. (3.3.31)), we also have, as $n \rightarrow \infty$,

(5.3.19)

$$\sup_{\frac{1}{n+1} \leq t \leq \frac{2}{n+1}} \sup_{(x,y) \in \mathbb{R}^2} \left(1 - \frac{[(n+1)t]}{n}\right) n^{-1/2} [(n+1)t]^{1/2} |\varepsilon_{1n}''(t) \cdot \nabla_{\theta} F(x, y; \theta_0)^T| / q(t) = o_P(1).$$

On the other hand, since $\varepsilon_{3n}'(t) = \varepsilon_{3(n+1)}(t)$, where $\varepsilon_{3n}(t)$ is as in (5.2.24), by (3.3.27) and (5.2.29), we have, as $n \rightarrow \infty$,

(5.3.20)

$$\sup_{\frac{1}{n+1} \leq t \leq \frac{2}{n+1}} \sup_{(x,y) \in \mathbb{R}^2} \frac{[(n+1)t]}{n} \left(1 - \frac{[(n+1)t]}{n}\right) n^{1/2} |\varepsilon_{3n}'(x, y, t)| / q(t) = o_P(1),$$

and, similarly, by symmetry and (3.3.27), we also have, as $n \rightarrow \infty$,

(5.3.21)

$$\sup_{\frac{1}{n+1} \leq t \leq \frac{2}{n+1}} \sup_{(x,y) \in \mathbb{R}^2} \frac{[(n+1)t]}{n} \left(1 - \frac{[(n+1)t]}{n}\right) n^{1/2} |\varepsilon_{3n}''(x, y, t)| / q(t) = o_P(1).$$

Carrying through all the steps from (3.3.38) to (3.3.62) with the corresponding changes of $K_{F_0}^{(1)}(x, y, nt)$ instead of $K^{(1)}(F(x; \theta_0), nt)$; $K_{F_0}^{(2)}(x, y, nt)$ instead of $K^{(2)}(F(x; \theta_0), nt)$, and using Lemma 5.2.2 instead of Lemma 3.2.3, we have, as $n \rightarrow \infty$,

(5.3.22)

$$\sup_{\frac{1}{n+1} \leq t \leq \frac{2}{n+1}} \sup_{(x,y) \in \mathbb{R}^2} |\varphi_n(t) - \int_{\mathbb{R}^2} l(x, y, \theta_0) d_x d_y n^{-1/2} \Gamma_{F_0}(x, y, nt)| / q(t) = o_P(1).$$

Since, by definition (cf. (5.3.1) and (5.3.2)),

$$\hat{\beta}_n(x, y, t) = 0 \text{ for any } t \in A_n := (0, 1/n + 1] \cup [n/n + 1, 1),$$

to prove (a) and (b), it now suffices to show, in virtue of (5.1.20) (cf. Remark 5.1.2) and Theorem 5.1.C, with $d = 2$, that

$$(5.3.23) \quad \sup_{t \in A_n} \sup_{(x,y) \in \mathbb{R}^2} |\Psi_{F_0,n}(x,y,t)|/q(t) = \begin{cases} o_P(1), & \text{if (5.3.6) holds} \\ O_P(1), & \text{if (5.3.7) holds.} \end{cases}$$

To this end, we note that, by (5.2.36), (5.2.37), (5.2.38) and (5.3.5), we have

(5.3.24)

$$\Psi_{F_0,n}(x,y,t) = \Gamma_{F_0,n}(x,y,t) - n^{-1/2}(W(nt) - tW(n)) \cdot D(\theta_0)^{-1} \cdot \nabla_{\theta} F(x,y)^T,$$

where $W(\cdot)$ is as in (5.2.40).

Next, we have that, by (5.1.31),

$$\sup_{0 < t < \frac{1}{n}} \sup_{(x,y) \in \mathbb{R}^2} |\Gamma_{F_0,n}(x,y,t)| \stackrel{P}{=} \sup_{0 < t < \frac{1}{n}} \sup_{(x,y) \in \mathbb{R}^2} |\Gamma_{F_0}(x,y,t)|.$$

Hence, by Theorem 5.1.E, with $d = 2$ (cf. Remark 5.1.6), we get, as $n \rightarrow \infty$,

$$(5.3.25) \quad \sup_{0 < t < \frac{1}{n}} \sup_{(x,y) \in \mathbb{R}^2} |\Gamma_{F_0}(x,y,t)|/q(t) = \begin{cases} o_P(1), & \text{if (5.3.6) holds,} \\ O_P(1), & \text{if (5.3.7) holds.} \end{cases}$$

Similarly, since $n^{-1/2}(W(nt) - tW(n)) \stackrel{D}{=} B(t)$, for each n , where $B(\cdot)$ is a Brownian bridge (cf. Remark 3.3.5), we have, by Theorem 3.3.C (cf. Remark 3.3.4), as $n \rightarrow \infty$,

$$(5.3.26) \quad \sup_{0 < t < \frac{1}{n}} |n^{-1/2}W(nt) - tW(1)|/q(t) = \begin{cases} o_P(1), & \text{if (5.3.6) holds,} \\ O_P(1), & \text{if (5.3.7) holds.} \end{cases}$$

Thus, (a) and (b) follow.

The proof of (c) is similar to the proof of (c) of Theorem 3.3.1. ■

Corollary 5.3.1. (a) if $q \in \mathcal{Q}^*$ and is such that (5.3.6) holds then, as $n \rightarrow \infty$,

$$(5.3.27) \quad \hat{\beta}_n(x, y, t)/q(t) \xrightarrow{\mathcal{D}} \Psi_{F_0}(x, y, t)/q(t)$$

in $D(\mathbb{R}^2 \times (0, 1])$,

(b) if q is a positive function on $(0, 1)$, then, as $n \rightarrow \infty$, (5.3.27) holds in $D(\mathbb{R}^2 \times [\delta, 1 - \delta])$ for any $0 < \delta < 1/2$.

We should note that on assuming condition (5.3.6), part (a) of Corollary 5.3.1 gives limiting distribution of any continuous functional of $\hat{\beta}_n(x, y, t)/q(t)$. In addition to that, Theorem 5.3.1 (c) gives convergence in distribution of the sup – sup-functional in $D(\mathbb{R}^2 \times (0, 1])$ of $\hat{\beta}_n(x, y, t)/q(t)$, with the weaker condition (5.3.7); in particular, Theorem 5.3.1 (c) is true for the function

$$(5.3.28) \quad q(t) = \left(t(1-t) \log \log 1/t(1-t) \right)^{1/2}, \quad 0 < t < 1.$$

Thus, similarly as in the univariate case, for testing (3.3.72) against (3.3.73), where in this case $X_1, X_2, \dots, X_n$ are bivariate normal random vectors, we propose:

Reject H_0 , in favour of H_1 , for large values of

$$(5.3.29) \quad \sup_{0 < t < 1} \sup_{(x, y) \in \mathbb{R}^2} |\hat{\beta}_n(x, y, t)| / \left(t(1-t) \log \log 1/(t(1-t)) \right)^{1/2},$$

where $\hat{\beta}_n(\cdot, \cdot, \cdot)$ is as in (5.3.1).

Before considering the multivariate case, we need to introduce some further notation.

5.4 Notation.

- (5.4.1) The transpose of a vector $\mathbf{x}$ or a matrix A will be denoted by $\mathbf{x}^\top$ and $A^\top$, respectively.
- (5.4.2) For any vector $\mathbf{x} = (x_1, \dots, x_d)$ in $\mathbb{R}^d$, $\|\mathbf{x}\| = \max_{1 \leq i \leq d} |x_i|$, and for any matrix $A = [\sigma_{ij}]$, $\|A\| = \max_{i,j} |\sigma_{ij}|$.
- (5.4.3) The trace of a matrix A , i.e., the sum of its diagonal elements will be denoted by $\text{tr}(A)$.
- (5.4.4) For any *symmetric* matrix D , D^* will denote the matrix D with the upper diagonal block equal to zero, i.e. if $D = [a_{ij}]$, $1 \leq i, j \leq d$, then $D^* = [b_{ij}]$, where $b_{ij} = 0$ for all $i < j$, and $b_{ij} = a_{ij}$ for $i \geq j$.
- (5.4.5) We will call $\mathbf{v}$ a vector-matrix, and will denote by $\mathbf{v} = (\mathbf{x}, A)$, where $\mathbf{x}$ is a vector in $\mathbb{R}^d$ and A is a $d \times d$ symmetric matrix.
- (5.4.6) If $\mathbf{z} = (\mathbf{x}, A)$ and $\mathbf{w} = (\mathbf{y}, B)$ are two vector-matrices, their sum will be defined by

$$\mathbf{z} + \mathbf{w} = (\mathbf{x} + \mathbf{y}, A + B),$$

where the sums inside parenthesis are the vector addition in $\mathbb{R}^d$ and the matrix addition, and their dot product will be defined by

$$\mathbf{z} \cdot \mathbf{w} = \mathbf{x} \cdot \mathbf{y} + \text{tr}[A(B^*)],$$

where the products on the right hand side are the dot product in $\mathbb{R}^d$ and the matrix product.

- (5.4.7) For a function $g(\mathbf{x}; \theta)$, where $\theta = (\mu, A)$, we define the vector-matrix

$$V_\theta g(\mathbf{x}; \theta) = (\nabla_\mu g(\mathbf{x}; \theta), Ag(\mathbf{x}; \theta)),$$

where, for any $\mu = (\mu_1, \dots, \mu_d)$,

$$\nabla_\mu g(\mathbf{x}; \theta) = \left(\frac{\partial}{\partial \mu_1} g(\mathbf{x}; \theta), \dots, \frac{\partial}{\partial \mu_d} g(\mathbf{x}; \theta) \right) \text{ and,}$$

for any $d \times d$ matrix $A = [a_{ij}]$, $Ag(x; \theta)$ is the $d \times d$ symmetric matrix $[\alpha_{ij}]$, with

$$\alpha_{ij} = \frac{\partial g}{\partial a_{ij}}(x; \theta).$$

(5.4.8) For any matrix $A = [a_{ij}]$, let $\int A$ denote the matrix $[\int a_{ij}]$, and for any vector-matrix $\mathbf{v} = (\mathbf{x}, A)$, let $\int \mathbf{v}$ denote the vector-matrix $(\int \mathbf{x}, \int A)$.

(5.4.9) For any vector-matrix $\mathbf{v} = (\mathbf{x}, A)$, $\|\mathbf{v}\| = \max \{\|\mathbf{x}\|, \|A\|\}$, with $\|\mathbf{x}\|$ and $\|A\|$ as in (5.4.2)

The motivation for this notation will become clear in the next section.

5.5 Sequentially estimated multivariate normal empirical processes.

Let $\mathbf{X} = (X^{(1)}, \dots, X^{(d)})$, $\mathbf{X}_i = (X_i^{(1)}, \dots, X_i^{(d)})$, $i = 1, 2, \dots$ be an independent sequence of normal random vectors in $\mathbb{R}^d$, $d \geq 1$, with unknown $\theta = (\mu, \Sigma)$, where $\mu = (\mu_1, \dots, \mu_d)$ is the vector of means and Σ is the $d \times d$ covariance matrix.

It is clear that, for any fixed $d \geq 1$, the number of parameters equals $d(d+1)/2$. Instead of writing θ as a $d(d+1)/2$ -dimensional vector, it is simpler, in terms of notation, to write θ as a pair or vector-matrix, according to our definition (cf. (5.4.5)): on one hand, a d -dimensional vector μ (the vector of "means"), and on the other hand, a $d \times d$ symmetric matrix (the covariance-matrix Σ , with main diagonal formed by the "variances").

For any $\frac{1}{n} \leq t \leq 1$ and $n \geq 1$, consider the maximum likelihood estimator of θ , denoted by $\hat{\theta}_{[nt]}$, based on the random sample $X_1, \dots, X_{[nt]}$. We have

$$(5.5.1) \quad \hat{\theta}_{[nt]} = (\hat{\mu}_{[nt]}, \hat{\Sigma}_{[nt]}),$$

where

$$\hat{\mu}_{[nt]} := (\bar{X}_{[nt]}^{(1)}, \dots, \bar{X}_{[nt]}^{(d)}), \text{ with } \bar{X}_{[nt]}^{(k)} = \frac{1}{[nt]} \sum_{i=1}^{[nt]} X_i^{(k)}, \quad 1 \leq k \leq d,$$

and the matrix $\hat{\Sigma}_{[nt]} := [\hat{\alpha}_{kj}(t)]_{kj}$, with

$$(5.5.2) \quad \hat{\alpha}_{kj}(t) = \frac{1}{[nt]} \sum_{i=1}^{[nt]} (X_i^{(k)} - \bar{X}_{[nt]}^{(k)})(X_i^{(j)} - \bar{X}_{[nt]}^{(j)}), \quad 1 \leq k, j \leq d.$$

Let us denote by $\theta_0 := (\mu^{(0)}, \Sigma_0)$ the true value of the vector-matrix of unknown parameters θ , where $\mu^{(0)} = (\mu_1^{(0)}, \dots, \mu_d^{(0)})$ and $\alpha_{kj}^{(0)}$ is the kj -th component of the matrix Σ_0 . We have, coordinatewise (cf. (5.2.1), (5.2.2) and (5.2.3)),

$$(5.5.3) \quad [nt]^{1/2}(\hat{\theta}_{[nt]} - \theta_0) = [nt]^{-1/2} \sum_{i=1}^{[nt]} l(X_i, \theta_0) + \epsilon_{1n}(t),$$

where, for any $\mathbf{x} = (x_1, \dots, x_d) \in \mathbb{R}^d$,

$$(5.5.4) \quad l(\mathbf{x}, \theta_0) = (\mathbf{x} - \mu^{(0)}, \Sigma_l),$$

with the kj -th component of Σ_l , λ_{kj} , given by

$$(5.5.5) \quad \lambda_{kj} = (x_k - \mu_k^{(0)})(x_j - \mu_j^{(0)}) - \alpha_{kj}^{(0)},$$

and

$$(5.5.6) \quad \epsilon_{1n}(t) = (\mathbf{0}, \Sigma_\epsilon^{[nt]}),$$

where $\mathbf{0} = (\overbrace{0, \dots, 0}^d)$ and the kj -th component of the matrix $\Sigma_\epsilon^{[nt]}$ given by

$$(5.5.7) \quad [nt]^{1/2} \left(\bar{X}_{[nt]}^{(k)} - \mu_k^{(0)} \right) \left(\bar{X}_{[nt]}^{(j)} - \mu_j^{(0)} \right).$$

From (5.5.4) we have that, for any $\mathbf{x} = (x_1, \dots, x_d)$, the vector-matrix $l(\mathbf{x}, \theta_0)$ can be written as

$$(5.5.8) \quad l(\mathbf{x}, \theta_0) = \left((l_1(\mathbf{x}, \theta_0), \dots, l_d(\mathbf{x}, \theta_0)), \Sigma_l \right),$$

where

$$(5.5.9) \quad l_i(\mathbf{x}, \theta_0) = l_i(x_i, \theta_0) = x_i - \mu_i^{(0)},$$

and the kj -th component of the matrix Σ_l is given by

$$(5.5.10) \quad \lambda_{kj} := l_{kj}(x_k, x_j, \theta_0) = l_k(x_k, \theta_0)l_j(x_j, \theta_0) - \alpha_{kj}^{(0)}, \quad 1 \leq j, k \leq d.$$

The next result provides approximations, in $\|\cdot/q\|$ -metrics, of the process $\hat{\beta}_n(\mathbf{x}, t)$ as defined in (5.1.35), by the sequence of Gaussian processes

$$(5.5.11) \quad G_{F_0, n}(x, t) = n^{-1/2} K_{F_0}(\mathbf{x}, nt) - \left\{ \int l(\mathbf{x}, \theta_0) d\mathbf{x} \right\} n^{-1/2} K_{F_0}(\mathbf{x}, nt) \cdot V_\theta F(\mathbf{x}; \theta_0),$$

where $K_{F_0}(\cdot, \cdot)$ is a Kiefer process associated with the distribution function $F_0 \equiv F(\cdot; \theta_0)$ of a multivariate normal random vector with parameter θ_0 , as in Theorem 5.1.A. It is also a generalization of Theorem 5.2.1.

Theorem 5.5.1. *Let $\mathbf{X}_1, \mathbf{X}_2, \dots$ be a sequence of independent normal random vectors in $\mathbb{R}^d$, $d \geq 1$, with $\theta := (\mu, \Sigma)$, where $\mu = (\mu_1, \dots, \mu_d)$ is the vector of means and Σ is the $d \times d$ covariance matrix. Suppose that, $\{\hat{\theta}_{[nt]}\}$ is as in (5.5.1), and let $\hat{\beta}_n(\mathbf{x}, t)$ be the sequentially estimated multivariate normal empirical process defined as in (5.1.35). Then, there exists a Kiefer process K_{F_0} , associated with the distribution function, $F(\mathbf{x}; \theta_0)$, of a multivariate normal random vector with parameters given by θ_0 , such that with the sequence of stochastic processes*

$\{G_{F_0,n}(\mathbf{x}, t), 0 \leq t \leq 1, \mathbf{x} \in \mathbb{R}^d\}$, as in (5.5.11), and a weight function $q \in \mathcal{Q}$ we have, as $n \rightarrow \infty$,

(a) if

$$(5.5.12) \quad I_0(q, c) < \infty \text{ for all } c > 0,$$

then

$$\sup_{0 < t \leq 1} \sup_{\mathbf{x} \in \mathbb{R}^d} |\hat{\beta}_n(\mathbf{x}, t) - G_{F_0,n}(\mathbf{x}, t)|/q(t) = o_P(1),$$

(b) if

$$(5.5.13) \quad I_0(q, c) < \infty \text{ for some } c > 0,$$

then

$$\sup_{0 < t \leq 1} \sup_{\mathbf{x} \in \mathbb{R}^d} |\hat{\beta}_n(\mathbf{x}, t) - G_{F_0,n}(\mathbf{x}, t)|/q(t) = O_P(1),$$

(c) if (5.5.13) holds, then

$$\sup_{0 < t \leq 1} \sup_{\mathbf{x} \in \mathbb{R}^d} |\hat{\beta}_n(\mathbf{x}, t)|/q(t) \xrightarrow{P} \sup_{0 < t \leq 1} \sup_{\mathbf{x} \in \mathbb{R}^d} |G_{F_0}(\mathbf{x}, t)|/q(t),$$

where

$$(5.5.14) \quad G_{F_0}(\mathbf{x}, t) = K_{F_0}(\mathbf{x}, t) - \left\{ \int l(\mathbf{x}, \theta_0) d\mathbf{x} K_{F_0}(\mathbf{x}, t) \right\} \cdot V_{\theta} F(\mathbf{x}, \theta_0).$$

For the proof of Theorem 5.5.1 we need the following lemmas.

Let $A_d^*(k)$ denote the complement of the d -dimensional rectangle $[-\log k, \log k]$ $\times \dots \times [-\log k, \log k]$, $k \geq 1$. For the vector-valued function $l(\mathbf{x}, \theta_0)$ in (5.5.4) (cf. also (5.5.8)) and the distribution function $F(\mathbf{x}; \theta_0)$, we have

Lemma 5.5.1. The series $\sum_{k=1}^{\infty} \frac{1}{k} \int_{A_2^*(k)} l_{ij}(\mathbf{x}, \theta_0) d\mathbf{x} F(\mathbf{x}; \theta_0)$ is convergent for any $1 \leq j \leq d$ and $1 \leq i \leq d$, where $l_0 \equiv 1$ and $l_{0j} \equiv l_j$.

Proof. Without loss of generality, we may assume $\mu_i^{(0)} = 0$ for all $i = 1, \dots, d$, and $\alpha_{kj}^{(0)} = 1$ if $k = j$. Now, by (5.5.8) and (5.5.9), denoting the joint density of $(X_1, \dots, X_d)$ by $f(x_1, \dots, x_d; \theta_0)$, we have

$$\begin{aligned} \int_{A_2^*(k)} l_{ij}^2(\mathbf{x}, \theta_0) d\mathbf{x} F(\mathbf{x}; \theta_0) &= \int_{A_2^*(k)} l_{ij}^2(x_i, x_j, \theta_0) f(x_1, \dots, x_d; \theta_0) dx_1 dx_2 \dots dx_d \\ &\leq \int_{A_2^*(k)} l_{ij}^2(x_i, x_j, \theta_0) f_{(X_i, X_j)}(x_i, x_j; \theta_0) dx_i dx_j, \end{aligned}$$

where $f_{(X_i, X_j)}(\cdot, \cdot, \theta_0)$ is the density of a bivariate normal random vector with mean vector $(0, 0)$ and covariance matrix $\begin{pmatrix} 1 & \alpha_0 \\ \alpha_0 & 1 \end{pmatrix}$, with $\alpha_0 = E(X_i - \mu_i^{(0)})(X_j - \mu_j^{(0)})$. (cf. Theorem 2.4.3 in Anderson, 1958). The result now follows from Lemma 5.2.1.

■

By (5.5.8)–(5.5.10) we can consider, without loss of generality, that, for any $\mathbf{x} = (x_1, \dots, x_d)$,

$$(5.5.15) \quad l_{kj}(\mathbf{x}, \theta_0) = l_k(x_k, \theta_0) l_j(x_j, \theta_0), \quad 1 \leq j \leq d, \quad 0 \leq k \leq d,$$

where $l_{0j} \equiv l_j$.

Introduce the following

$$(5.5.16) \quad \varepsilon_{2n}(\mathbf{x}, t) = n^{-1/2} \left[\sum_{i=1}^{[nt]} (1\{\mathbf{X}_i \leq \mathbf{x}\} - F(\mathbf{x}; \theta_0)) - K_{F_0}(\mathbf{x}, nt) \right]$$

and

$$(5.5.17) \quad L_n(t) = \int_{\mathbb{R}^d} l(\mathbf{x}, \theta_0) d_{\mathbf{x}} \varepsilon_{2n}(\mathbf{x}, t),$$

where K_{F_0} is a Kiefer process associated with $F(\cdot; \theta_0)$ in $\mathbb{R}^d$.

The following lemma is a generalization of Lemma 5.2.2.

Lemma 5.5.2. *Let $\mathbf{X}_1, \mathbf{X}_2, \dots$ be a sequence of independent normal random vectors in $\mathbb{R}^d$, $d \geq 1$, with $\theta_0 := (\mu_0, \Sigma_0)$. Then, for any function q which is positive on $(0, 1]$ and such that $\lim_{t \downarrow 0} t^{1/2}/q(t) = 0$, we have, as $n \rightarrow \infty$,*

$$\sup_{\frac{1}{n} \leq t \leq 1} \|L_n(t)\|/q(t) \longrightarrow 0 \text{ in probability.}$$

Proof. For each $n \geq 1$, let $A_n^{(d)}(t)$ denote the d -dimensional rectangle $[-\log(nt), \log(nt)]^d$, $d \geq 1$. Then,

$$\begin{aligned} L_n(t) &= \int_{A_n^{(d)}(t)} l(\mathbf{x}, \theta_0) d_{\mathbf{x}} \varepsilon_{2n}(\mathbf{x}, t) \\ &\quad + \int_{[A_n^{(d)}]^c(t)} l(\mathbf{x}, \theta_0) d_{\mathbf{x}} n^{-1/2} \sum_{i=1}^{[nt]} (1\{\mathbf{X}_i \leq \mathbf{x}\} - F(\mathbf{x}; \theta_0)) \\ &\quad - \int_{[A_n^{(d)}]^c(t)} l(\mathbf{x}, \theta_0) d_{\mathbf{x}} n^{-1/2} K_{F_0}(\mathbf{x}, nt) \\ &:= L_{1n}(t) + L_{2n}(t) - L_{3n}(t). \end{aligned}$$

We have, for any $1 \leq j \leq d$, $0 \leq k \leq d$,

$$(5.5.18) \quad \int_{A_n^{(d)}(t)} l_{kj}(\mathbf{x}, \theta_0) d_{\mathbf{x}} \varepsilon_{2n}(\mathbf{x}, t) = \int_{A_n^{(2)}(t)} l_k(x_k) l_j(x_j) dx_k dx_j \left[\int_{A_n^{(d-2)}(t)} d_{\mathbf{x}_k} \varepsilon_{2n}(\mathbf{x}, t) \right].$$

Using integration by parts we get that

$$\int_{A_n^{(d-2)}(t)} d\mathbf{x}_{kj} \varepsilon_{2n}(\mathbf{x}, t) = \Delta_{A_n^{(d-2)}(t)} \varepsilon_{2n}(\mathbf{x}, t),$$

where $\Delta_A F$ is the difference of F around the vertices of A . (In the case $k = 1$, $A = [a, b]$ and $\Delta_A F = F(b) - F(a)$. In the case $k = 2$, $\Delta_A F = F(b_1, b_2) - F(b_1, a_2) - F(a_1, b_2) + F(a_1, a_2)$). Also,

$$(5.5.19) \quad \Delta_{A_n^{(d-2)}(t)} \varepsilon_{2n}(\mathbf{x}, t) = \sum_{i=1}^{2^{d-2}} \delta_i \varepsilon_{2n}(\mathbf{x}, t),$$

where $\delta_i = \pm 1$ and $\mathbf{x} = (x_1, \dots, x_d)$ has the form $x_i = \pm \log(nt)$ for all i except for $i = k, j$. Hence, by (5.5.18) and (5.5.19), we can write

$$(5.5.20) \quad \int_{A_n^{(d)}(t)} l_{kj}(\mathbf{x}, \theta_0) d\mathbf{x} \varepsilon_{2n}(\mathbf{x}, t) \\ = \sum_m \delta_m \int_{A_n^{(2)}(t)} l_k(x) l_j(y) d_x d_y \varepsilon_{2n}(\pm \log(nt), \dots, x, \pm \log(nt), \dots, y, \dots, \pm \log(nt)),$$

where the sum is taken over all possible combinations of $\pm \log(nt)$ in the argument of ε_n , and $\delta_m = \pm 1$. This now reduces to a two-dimensional integral, as in (5.2.11).

We can use (5.2.13) with $R = [-\log(nt), \log(nt)]^2$, $l_j^{(1)} \equiv l_j$, $l_j^{(2)} \equiv l_k$, and ε_{2n} as in (5.2.10).

By (5.5.9) and (5.5.10), we have that

$$(5.5.21) \quad \max_{i,j} \sup_{\mathbf{x} \in A_n^{(d)}(t)} |l_{kj}(\mathbf{x}, \theta_0)| \leq C \log^2(nt),$$

where C is a positive constant. Also, by Theorem 5.1.A,

$$\sup_{x \in \mathbb{R}^d} |\varepsilon_{2n}(x, t)| \stackrel{a.s.}{=} O(n^{-1/2}(nt)^{1/2-1/4d}(\log(nt))^{3/2}).$$

Let $0 < \delta < 1$ be fixed. Then,

$$\begin{aligned} \sup_{\delta \leq t \leq 1} \|L_{1n}(t)\| &\stackrel{a.s.}{=} \sup_{\delta \leq t \leq 1} n^{-1/2}(nt)^{1/2-1/4d}(\log(nt))^{7/2} \\ &= n^{-1/4d}(\log n)^{7/2} \rightarrow 0, \end{aligned}$$

as $n \rightarrow \infty$. Thus, since $\inf_{\delta \leq t \leq 1} q(t) > 0$, we have, as $n \rightarrow \infty$,

$$\sup_{\delta \leq t \leq 1} \|L_{1n}(t)\|/q(t) = o_P(1).$$

Also,

$$\begin{aligned} \sup_{\frac{1}{n} \leq t \leq \delta} \|L_{1n}(t)\|/q(t) &\leq \sup_{0 < t \leq \delta} t^{1/2}/q(t) \sup_{\frac{1}{n} \leq t \leq \delta} \|L_{1n}(t)\|/t^{1/2} \\ &\stackrel{a.s.}{=} \sup_{0 < t \leq \delta} t^{1/2}/q(t) \sup_{\frac{1}{n} \leq t \leq 1} (nt)^{-1/4d}(\log(nt))^{7/2} \\ &\stackrel{a.s.}{=} \sup_{0 < t \leq \delta} t^{1/2}/q(t) \times O(1). \end{aligned}$$

Since q is such that $\lim_{t \downarrow 0} t^{1/2}/q(t) = 0$, taking $\delta > 0$ sufficiently small, we get that, as $n \rightarrow \infty$,

$$\sup_{\frac{1}{n} \leq t \leq \delta} \|L_{1n}(t)\|/q(t) = o_P(1).$$

For any $1 \leq i, j \leq d$ and $0 \leq k \leq d$ we have, by (5.5.15),

$$\int_{A_n^{(d)}(t)} l_{kj}(x, \theta_0) d_x \mathbf{1}\{X_i \leq x\} = \int_{[A_n^{(2)}(t)]^c} l_k(x) l_j(y) d_x d_y \mathbf{1}\{X_i^{(k)} \leq x, X_i^{(j)} \leq y\},$$

since

$$\begin{aligned} 1\{(X_i^{(1)}, \dots, X_i^{(d)}) \leq (x_1, \dots, x_d)\} &= 1\{X_i^{(1)} \leq x_1, \dots, X_i^{(d)} \leq x_d\} \\ &= \prod_{m=1}^d 1\{X_i^{(m)} \leq x_m\}. \end{aligned}$$

Also,

$$\int l_{kj}(\mathbf{x}, \theta_0) d\mathbf{x} F(\mathbf{x}, \theta_0) = \int l_k(x) l_j(y) d_x d_y \tilde{F}(x, y; \theta_0) = 0,$$

where $\tilde{F}(x, y; \theta_0)$ is the distribution function of a bivariate normal random vector.

We have, thus, that this case reduces to the two-dimensional case. We will get, for instance, the same type of inequalities (5.2.17), (5.2.18) and (5.2.19). Hence, by Lemma 5.5.1, we have, similarly as in the proof of Lemma 5.2.1, that, as $n \rightarrow \infty$,

$$\sup_{\frac{1}{n} \leq t \leq 1} \|L_{2n}(t)\|/q(t) = o_P(1).$$

We can, similarly, generalize the result for the two-dimensional case (cf. (5.1.1)–(5.1.4)), to show that

$$\begin{aligned} \int l_{kj}(\mathbf{x}, \theta_0) d\mathbf{x} n^{-1/2} K_{F_0}(\mathbf{x}, nt) = \\ \sum_m \delta_m \int_0^1 \int_0^1 l_k(F_{(k)}^{-1}(u_k; \theta_0)) l_j(F_{(j)}^{-1}(u_j; \theta_0)) |J_{kj}(u_k, u_j)| d_{u_k} d_{u_j} n^{-1/2} K_{G_0}(\mathbf{u}, nt), \end{aligned}$$

where $\delta_m = \pm 1$, the sum is taken over all possible $\mathbf{u} = (u_1, \dots, u_k, \dots, u_j, \dots, u_d)$, with $u_i = 0$ or 1 for all $i \neq k, j$, $|J_{kj}(u_k, u_j)|$ is the Jacobian, and $G_0 = F_0 \circ L^{-1}$, where $L^{-1} : I^d \rightarrow \mathbb{R}^d$ is defined by (5.1.1). Thus, we also have that, as $n \rightarrow \infty$,

$$\sup_{\frac{1}{n} \leq t \leq 1} \|L_{3n}(t)\|/q(t) = o_P(1),$$

and the result follows. ■

Proof of Theorem 5.5.1. After a one-term Taylor expansion around θ_0 in (5.1.35), we have

(5.5.22)

$$\begin{aligned}\hat{\beta}_n(\mathbf{x}, t) &= n^{-1/2} \sum_{i=1}^{[nt]} (1\{X_i \leq \mathbf{x}\} - F(\mathbf{x}; \theta_0)) \\ &\quad - [nt]n^{-1/2} (F(\mathbf{x}; \hat{\theta}_{[nt]}) - F(\mathbf{x}; \theta_0)) \\ &:= \beta_n(\mathbf{x}, t) + [nt]n^{-1/2} (\hat{\theta}_{[nt]} - \theta_0) \cdot (V_\theta F(\mathbf{x}; \theta_0) - V_\theta F(\mathbf{x}; \hat{\theta}_{[nt]})) \\ &\quad - [nt]n^{-1/2} (\hat{\theta}_{[nt]} - \theta_0) \cdot V_\theta F(\mathbf{x}; \theta_0).\end{aligned}$$

Also, since $El(\mathbf{X}, \theta_0) = 0$,

(5.5.23)

$$\begin{aligned}n^{-1/2}[nt](\hat{\theta}_{[nt]} - \theta_0) &= \sum_{i=1}^{[nt]} l(\mathbf{X}_i, \theta_0) + n^{-1/2}[nt]^{1/2}\varepsilon_{1n}(t) \\ &= \int_{\mathbb{R}^d} l(\mathbf{x}, \theta_0) d\mathbf{x} n^{-1/2} \left[\sum_{i=1}^{[nt]} (1\{X_i \leq \mathbf{x}\} - F(\mathbf{x}; \theta_0)) \right. \\ &\quad \left. - n^{-1/2} K_{F_0}(\mathbf{x}, nt) \right] \\ &\quad + \int_{\mathbb{R}^d} l(\mathbf{x}, \theta_0) d\mathbf{x} n^{-1/2} K_{F_0}(\mathbf{x}, nt) + n^{-1/2}[nt]^{1/2}\varepsilon_{1n}(t).\end{aligned}$$

Hence, after adding and subtracting $G_{F_0, n}(\cdot, \cdot)$ (cf. (5.5.11)), we get, by (5.5.22),

(5.5.24)

$$\begin{aligned}\hat{\beta}_n(\mathbf{x}, t) &= \beta_n(\mathbf{x}, t) - n^{-1/2} K_{F_0}(\mathbf{x}, nt) + G_{F_0, n}(\mathbf{x}, nt) + [nt]^{1/2} n^{-1/2} \varepsilon_{3n}(\mathbf{x}, t) \\ &\quad - L_n(t) \cdot V_\theta F(\mathbf{x}; \theta_0)^\top - [nt]^{1/2} n^{-1/2} \varepsilon_{1n}(t) \cdot V_\theta F(\mathbf{x}; \theta_0),\end{aligned}$$

where

$$\varepsilon_{3n}(\mathbf{x}, t) = [nt]^{1/2} (\hat{\theta}_{[nt]} - \theta_0) \cdot (V_\theta F(\mathbf{x}; \theta_0) - V_\theta F(\mathbf{x}; \hat{\theta}_{[nt]})),$$

and $L_n(t)$ is as in Lemma 5.5.2.

By (5.5.6) and (5.5.7) we note that coordinatewise we can use (5.2.4) and the fact that $\sup_{x \in \mathbb{R}^d} \|V_\theta F(x; \theta_0)\| < \infty$, to show that (cf. (5.2.25)), as $n \rightarrow \infty$,

$$(5.5.25) \quad \sup_{\frac{1}{n} \leq t \leq 1} \sup_{x \in \mathbb{R}^d} n^{-1/2} [nt]^{1/2} |\varepsilon_{1n}(t) \cdot V_\theta F(x; \theta_0)| / q(t) = o_P(1).$$

This also applies to $\varepsilon_{3n}(x, t)$. Thus, as in the proof of (5.2.29) we can show that, as $n \rightarrow \infty$,

$$(5.5.26) \quad \sup_{\frac{1}{n} \leq t \leq 1} \sup_{x \in \mathbb{R}^d} n^{-1/2} [nt]^{1/2} |\varepsilon_{3n}(x, t)| / q(t) = o_P(1).$$

Hence, from (5.5.24), we have, by (5.1.20) (cf. Remark 5.1.2), Lemma 5.5.2, (5.5.25) and (5.5.26), that, as $n \rightarrow \infty$,

$$\sup_{\frac{1}{n} \leq t \leq 1} \sup_{x \in \mathbb{R}^d} |\hat{\beta}_n(x, t) - G_{F_{0,n}}(x, t)| / q(t) = o_P(1),$$

provided only that q be positive on $(0, 1]$ and such that $\lim_{t \downarrow 0} t^{1/2} / q(t) = 0$.

Since, by definition,

$$\hat{\beta}_n(x, t) = 0 \text{ for all } x \in \mathbb{R}^d \text{ and } 0 < t < 1/n,$$

to prove (a) and (b), it now suffices to show that, as $n \rightarrow \infty$,

$$(5.5.27) \quad \sup_{0 < t < \frac{1}{n}} \sup_{x \in \mathbb{R}^d} |G_{F_{0,n}}(x, t)| / q(t) = \begin{cases} o_P(1) & \text{if (5.1.12) holds,} \\ O_P(1) & \text{if (5.1.13) holds.} \end{cases}$$

The proof of this is similar to that of (5.2.35), in the two-dimensional case.

The proof of (c) is analogous to the proof of (c) of Theorem 5.2.1. ■

Corollary 5.5.1. *Assuming the conditions of Theorem 5.5.1, we have, as $n \rightarrow \infty$,*

(a) *if $q \in Q$ is such that (5.5.12) holds, then*

$$(5.5.28) \quad \hat{\beta}_n(\mathbf{x}, t)/q(t) \xrightarrow{P} G_{F_0}(\mathbf{x}, t)/q(t)$$

in $D(\mathbb{R}^d \times [0, 1])$,

(b) *if q is positive on $(0, 1]$, then (5.5.28) holds in $D(\mathbb{R}^d \times [\delta, 1])$ for any $0 < \delta < 1$.*

5.6 Sequentially estimated multivariate normal empirical processes tied down at $t=1$.

In this section we study the multivariate form of the process in (5.3.1).

Introduce the following:

$$(5.6.1) \quad \hat{\beta}_n(\mathbf{x}, t) = \begin{cases} n^{1/2} \frac{[(n+1)t]}{n} \left(1 - \frac{[(n+1)t]}{n}\right) \hat{\zeta}_n(\mathbf{x}, t), & 0 \leq t < 1, \mathbf{x} \in \mathbb{R}^d, \\ 0, & t = 1, \mathbf{x} \in \mathbb{R}^d, \end{cases}$$

where

$$(5.6.2) \quad \begin{aligned} \hat{\zeta}_n(\mathbf{x}, t) = & \frac{1}{[(n+1)t]} \sum_{i=1}^{[(n+1)t]} \left(1\{X_i \leq \mathbf{x}\} - F(\mathbf{x}; \hat{\theta}'_{[nt]})\right) \\ & - \frac{1}{n - [(n+1)t]} \sum_{i=[(n+1)t]+1}^n \left(1\{X_i \leq \mathbf{x}\} - F(\mathbf{x}; \hat{\theta}'_{[nt]})\right), \end{aligned}$$

and, for each $\frac{1}{n+1} \leq t \leq \frac{n}{n+1}$, $\{\hat{\theta}'_{[nt]}\}$ and $\{\hat{\theta}''_{[nt]}\}$ are sequences of maximum likelihood estimators of θ based on the samples $X_1, \dots, X_{[(n+1)t]}$ and $X_{[(n+1)t]+1}, \dots, X_n$, respectively. Here, $F(\mathbf{x}; \theta)$ denotes the multivariate normal distribution function with parameter θ .

In this case we have

$$(5.6.3) \quad \hat{\theta}'_{[nt]} = \hat{\theta}_{[(n+1)t]},$$

where $\hat{\theta}_{[t]}$ is as in (5.5.1). Also,

$$(5.6.4) \quad \hat{\theta}''_{[nt]} = (\hat{\mu}_{[nt]}, \hat{\Sigma}_{[nt]}),$$

where

$$\hat{\mu}_{[nt]} := (\overline{X}_{[nt]}^{(1)}, \dots, \overline{X}_{[nt]}^{(d)}),$$

with

$$\overline{X}_{[nt]}^{(k)} = \frac{1}{n - [(n+1)t]} \sum_{i=[(n+1)t]+1}^n X_i^{(k)}, \quad 1 \leq k \leq d,$$

and each component of the matrix $\hat{\Sigma}_{[nt]} := [\hat{\alpha}_{kj}(t)]_{kj}$ is given by

$$\hat{\alpha}_{kj}(t) = \frac{1}{n - [(n+1)t] + 1} \sum_{i=[(n+1)t]}^n (X_i^{(k)} - \overline{X}_{[nt]}^{(k)})(X_i^{(j)} - \overline{X}_{[nt]}^{(j)}), \quad 1 \leq k, j \leq d.$$

The next result is a generalization of Theorem 5.3.1. It provides approximations, in $\|\cdot/q\|$ -metrics, of the process $\hat{\beta}_n(\mathbf{x}, t)$ as defined in (5.6.1), by the sequence of Gaussian processes $\{\Psi_{F_{0,n}}(\mathbf{x}, t), 0 \leq t \leq 1, \mathbf{x} \in \mathbb{R}^d\}$ defined by

$$(5.6.5) \quad \Psi_{F_{0,n}}(\mathbf{x}, t) = \Gamma_{F_{0,n}}(\mathbf{x}, t) - \left\{ \int l(\mathbf{x}, \theta_0) d_{\mathbf{x}} \Gamma_{F_{0,n}}(\mathbf{x}, t) \right\} \cdot V_{\theta} F(\mathbf{x}; \theta_0),$$

where $\Gamma_{F_0,n}(\cdot, \cdot)$ is as in (5.1.31), and $F = F_0 \equiv F(\cdot; \theta_0)$, is the distribution function of a multivariate normal random variable with $\theta_0 = (\mu^{(0)}, \Sigma_0)$.

Theorem 5.6.1. *Let $X_1, X_2, \dots$ be a sequence of independent normal random vectors in $\mathbb{R}^d$, $d \geq 1$, with $\theta = (\mu, \Sigma)$, where $\mu = (\mu_1, \dots, \mu_d)$ is the vector of means and Σ is the $d \times d$ covariance matrix. Suppose that, $\{\hat{\theta}'_{[nt]}\}$ and $\{\hat{\theta}''_{[nt]}\}$ are as in (5.6.3) and (5.6.4). There exists a Kiefer process $K_{F_0}(\cdot, \cdot)$, associated with the distribution function $F(x; \theta_0)$, such that with the sequence of Gaussian processes $\{\Psi_{F_0,n}(x, t), 0 \leq t \leq 1, x \in \mathbb{R}^d\}$ defined by (5.6.5), and a weight function of $q \in Q^*$, we have, as $n \rightarrow \infty$,*

(a) if

$$(5.6.6) \quad I_{0,1}(q, c) < \infty \text{ for all } c > 0,$$

then

$$\sup_{0 < t < 1} \sup_{x \in \mathbb{R}^d} |\hat{\beta}_n(x, t) - \Psi_{F_0,n}(x, t)|/q(t) = o_P(1),$$

(b) if

$$(5.6.7) \quad I_0(q, c) < \infty \text{ for some } c > 0,$$

then

$$\sup_{0 < t < 1} \sup_{x \in \mathbb{R}^d} |\hat{\beta}_n(x, t) - \Psi_{F_0,n}(x, t)|/q(t) = O_P(1),$$

(c) if (5.6.7) holds then

$$\sup_{0 < t < 1} \sup_{x \in \mathbb{R}^d} |\hat{\beta}_n(x, t)|/q(t) \xrightarrow{D} \sup_{0 < t < 1} \sup_{x \in \mathbb{R}^d} |\Psi_{F_0}(x, t)|/q(t),$$

where

$$(5.6.8) \quad \Psi_{F_0}(\mathbf{x}, t) = \Gamma_{F_0}(\mathbf{x}, t) - \left\{ \int l(\mathbf{x}, \theta_0) d_{\mathbf{x}} \Gamma_{F_0}(\mathbf{x}, t) \right\} \cdot V_{\theta} F(\mathbf{x}; \theta_0),$$

and $\Gamma_{F_0}(\cdot, \cdot)$ is as in (5.1.91), with $F = F_0 \equiv F(\cdot; \theta_0)$, the multivariate normal distribution with $\theta_0 = (\mu^{(0)}, \Sigma_0)$.

Proof of Theorem 5.6.1. After adding and subtracting $\Gamma_{F_0, n}(\cdot, \cdot)$ (cf (5.6.5)), and a one-term Taylor expansion around θ_0 we have, from (5.6.1), we get

$$(5.6.9) \quad \begin{aligned} \hat{\beta}_n(\mathbf{x}, t) = & \beta_n^*(\mathbf{x}, t) - \Gamma_{F_0, n}(\mathbf{x}, t) + \Gamma_{F_0, n}(\mathbf{x}, t) - n^{-1/2} \left(1 - \frac{[(n+1)t]}{n} \right) \\ & \sum_{i=1}^{[(n+1)t]} l(\mathbf{X}_i, \theta_0) - \frac{[(n+1)t]}{n} \sum_{i=[(n+1)t]+1}^n l(\mathbf{X}_i, \theta_0) \cdot V_{\theta} F(\mathbf{x}; \theta_0) \\ & + \frac{[(n+1)t]}{n} \left(1 - \frac{[(n+1)t]}{n} \right) n^{1/2} \{ \epsilon'_{3n}(\mathbf{x}, t) + \epsilon''_{3n}(\mathbf{x}, t) \} \\ & + \frac{[(n+1)t]}{n} n^{-1/2} (n - [(n+1)t])^{\frac{1}{2}} \epsilon''_{1n}(t) \cdot V_{\theta} F(\mathbf{x}; \theta_0) \\ & - \left(1 - \frac{[(n+1)t]}{n} \right) n^{-1/2} [(n+1)t]^{\frac{1}{2}} \epsilon'_{1n}(t) \cdot V_{\theta} F(\mathbf{x}; \theta_0), \end{aligned}$$

where $\beta_n^*(\mathbf{x}, t)$ is as in (5.1.22);

$\epsilon'_{1n}(t) = (0, \Sigma_{\epsilon'}^{[(n+1)t]})$, with the kj -th component of $\Sigma_{\epsilon'}^{[(n+1)t]}$ given by

$$(5.6.10) \quad [(n+1)t]^{1/2} (\bar{X}_{[nt]}^{(k)} - \mu_k^{(0)}) (\bar{X}_{[nt]}^{(j)} - \mu_j^{(0)});$$

$\epsilon''_{1n}(t) = (0, \Sigma_{\epsilon''}^{[(n+1)t]})$, with the kj -th component of $\Sigma_{\epsilon''}^{[(n+1)t]}$ given by

$$(5.6.11) \quad (n - [(n+1)t])^{1/2} (\bar{X}_{[nt]}^{(k)} - \mu_k^{(0)}) (\bar{X}_{[nt]}^{(j)} - \mu_j^{(0)});$$

$$(5.6.12) \quad \epsilon'_{3n}(\mathbf{x}, t) = (\hat{\theta}'_{[nt]} - \theta_0) \cdot \left(V_{\theta} F(\mathbf{x}; \hat{\theta}'_{[nt]}) - V_{\theta} F(\mathbf{x}; \theta_0) \right),$$

with

$$(5.6.13) \quad \sup_{0 \leq t \leq 1} \|\hat{\theta}'_{[nt]} - \theta_0\| \leq \sup_{0 \leq t \leq 1} \|\hat{\theta}'_{[nt]} - \theta_0\|;$$

and

$$(5.6.14) \quad \epsilon''_{3n}(\mathbf{x}, t) = (\hat{\theta}''_{[nt]} - \theta_0) \cdot \left(V_{\theta} F(\mathbf{x}; \hat{\theta}''_{[nt]}) - V_{\theta} F(\mathbf{x}; \theta_0) \right),$$

with

$$(5.6.15) \quad \sup_{0 \leq t \leq 1} \|\hat{\theta}''_{[nt]} - \theta_0\| \leq \sup_{0 \leq t \leq 1} \|\hat{\theta}''_{[nt]} - \theta_0\|.$$

We note that coordinatewise this case reduces to the two-dimensional case. The proof of Theorem 5.6.1 then follows the lines of the proof of Theorem 5.3.1 (cf. (5.3.9)–(5.3.23), using here Theorem 5.1.D, Theorem 5.1.E and Lemma 5.5.2. ■

Corollary 5.6.1. (a) if $q \in Q^*$ and is such that (5.6.6) holds then, as $n \rightarrow \infty$,

$$(5.6.16) \quad \hat{\beta}_n(\mathbf{x}, t)/q(t) \xrightarrow{\mathcal{D}} \Psi_{F_0}(\mathbf{x}, t)/q(t)$$

in $D(\mathbb{R}^d \times [0, 1])$,

(b) if q is a positive function on $(0, 1)$, then, as $n \rightarrow \infty$, (5.6.16) holds in $D(\mathbb{R}^d \times [\delta, 1 - \delta])$ for any $0 < \delta < 1/2$.

We should note that on assuming condition (5.6.6), part (a) of Corollary 5.6.1 gives limiting distribution of any continuous functional of $\hat{\beta}_n(\mathbf{x}, t)/q(t)$. Furthermore, Theorem 5.6.1 (c) gives convergence in distribution of the supsup-functional in $D(\mathbb{R}^d \times [0, 1])$ of $\hat{\beta}_n(\mathbf{x}, t)/q(t)$, with the weaker condition (5.6.7); in particular,

Theorem 5.6.1 (c) is true for the function

$$q(t) = \left(t(1-t) \log \log 1/t(1-t) \right)^{1/2}, \quad 0 < t < 1.$$

Thus, similarly as in the bivariate case, for testing (3.3.72) against (3.3.73), where in this case $X_1, X_2, \dots, X_n$ are multivariate normal random vectors, we propose:

Reject H_0 , in favour of H_1 , for large values of

$$\sup_{0 < t < 1} \sup_{x \in \mathbb{R}^d} |\hat{\beta}_n(x, t)| / \left(t(1-t) \log \log 1/(t(1-t)) \right)^{1/2},$$

where $\hat{\beta}_n(\cdot, \cdot)$ is as in (5.6.1).

Appendices

A.1 Some maximal inequalities for partial sums.

Here we quote two results that are used frequently throughout our work. They concern partial sum of independent random variables.

A.1.1. Kolmogorov's inequality.

Theorem A.1.1. *Let $X_1, X_2, \dots$ be a sequence of independent random variables, with $EX_i = 0$ and $EX_i^2 = \sigma_i^2 < \infty$. For any $k \geq 1$, let $S_k = \sum_{i=1}^k X_i$. Then, for any $\alpha > 0$,*

$$P\left\{\max_{1 \leq k \leq n} |S_k| \geq \alpha\right\} \leq \frac{1}{\alpha^2} \sum_{i=1}^n \sigma_i^2.$$

Proof. See Theorem 22.4 in Billingsley, 1986, page 296.

The next result is a generalization of Theorem A.1.1.

A.1.2. Hájek-Rényi inequality.

Theorem A.1.2. *Assume the same conditions of Theorem A.1.1. Let $\{c_k\}_{k \geq 1}$ be a non-increasing sequence of positive real numbers. Then, for any $\alpha > 0$, and any positive integers n and m ($n > m$),*

$$P\left\{\max_{m \leq k \leq n} c_k |S_k| \geq \alpha\right\} \leq \frac{1}{\alpha^2} \left\{c_m^2 \sum_{k=1}^m \sigma_k^2 + \sum_{k=m+1}^n c_k^2 \sigma_k^2\right\}.$$

Proof. See Hájek and Rényi (1955).

A.2 On Riemann-Stieltjes integrals in $\mathbb{R}^2$.

Here we formulate an integration by parts (cf. Theorem A.2.1) of functions of intervals in $\mathbb{R}^2$. For more details we refer to Hildebrandt, 1963, Chapter III.

Let J denote the basic rectangle $[a_1, b_1] \times [a_2, b_2] \subseteq \mathbb{R}^2$. Let I denote any subrectangle of J . Consider the set function $F(I)$ defined by

$$\begin{aligned} F(I) &= f(d_1, d_2) - f(c_1, d_2) - f(c_2, d_1) + f(c_1, c_2) \\ &:= \Delta_{c_1}^{d_1} \Delta_{c_2}^{d_2} f(x_1, x_2), \text{ where } x_1, x_2 \in I, \end{aligned}$$

and $I = [c_1, d_1] \times [c_2, d_2]$.

Definition A.2.1. *A subdivision σ of an interval I consists of a finite number of intervals having at most edges or vertices in common, whose sum or union is I . A subdivision of the fundamental interval J obviously induces a subdivision on any subinterval I . A subdivision will be called a net if it is determined by lines completely across I parallel to the coordinate axes, so that a net is determined by subdivision σ_{x_1} and σ_{x_2} of the projections of I on the coordinate axes. We shall denote a net by $\sigma_1 \times \sigma_2$.*

Definition A.2.2. A real-valued function f on the rectangle $J := [a_1, b_1] \times [a_2, b_2]$ is of bounded variation on J if $\sum_{\sigma} F(I) = \sum_{\sigma} |\Delta_1 \Delta_2 f(I)|$ is bounded for all subdivisions σ of J , where, if $I = [c_1, d_1] \times [c_2, d_2]$, then

$$F(I) = |\Delta_1 \Delta_2 f(I)| = |f(d_1, d_2) - f(c_1, d_2) - f(d_1, c_2) + f(c_1, c_2)|.$$

What follows is a definition of bounded variation defined by Fréchet (1915). It is based on net subdivisions of intervals.

Definition A.2.3. A real-valued function f on J is said to be of Fréchet bounded variation if the expression

$$\begin{aligned} & \sum_{ij} \epsilon_i \epsilon'_j \Delta_1 \Delta_2 f(I_{ij}) \\ &= \sum_{ij} \epsilon_i \epsilon'_j [f(x_{i-1}, y_{i-1}) - f(x_i, y_{i-1}) - f(x_{i-1}, y_j) + f(x_i, y_j)] \end{aligned}$$

is bounded in σ_x and σ_y for all choices of $\epsilon_i = \pm 1$, $\epsilon'_i = \pm 1$. If $f(x, y)$ is of Fréchet bounded variation on J , then the Fréchet total variation $FV(U)$ is the l.u.b. $\sum_{\sigma_x \times \sigma_y} \epsilon_i \epsilon'_j \Delta_1 \Delta_2 f(I_{ij})$ for all $\epsilon_i = \pm 1$, $\epsilon'_j = \pm 1$ and $\sigma_x \times \sigma_y$ of J . We shall denote the class of functions of Fréchet bounded variation by FBV .

We note that $FV(J)$ is also - g.l.b. $[\sum_{\sigma_x \times \sigma_y} \epsilon_i \epsilon'_j \Delta_1 \Delta_2 f(I_{ij})]$ since a change in the signs of all ϵ_i gives a negative value corresponding to any positive value of the sum.

If $f(x, y) = g(x) + h(y)$, for all $x, y \in I$, then $\Delta_1 \Delta_2 f(I) = 0$ for all I so that $FV(J) = 0$. If $f(x, y) = g(x)h(y)$, for all $x, y \in I$, then

$$\begin{aligned} & \sum_{ij} \epsilon_i \epsilon'_j \Delta_1 \Delta_2 f(I_{ij}) \\ &= \left[\sum_i \epsilon_i (g(x_i) - g(x_{i-1})) \right] \left[\sum_j \epsilon'_j (h(y_j) - h(y_{j-1})) \right]. \end{aligned}$$

It follows that $f(x, y) = g(x)h(y)$ is of FBV if and only if g and h are of bounded variation on $[a, a']$ and $[b, b']$, respectively, provided f and g are not constant functions. Since

$$\sum_{ij} \varepsilon_i \varepsilon'_j \Delta_1 \Delta_1 f(I_{ij}) \leq \sum_{ij} |\Delta_1 \Delta_2 f(I_{ij})|,$$

f will be of FBV if f is of bounded variation on J , and $FV(I) \leq V(J)$.

Next, we present a formula of integration by parts. For the proof of it we refer the reader to Hildebrandt, 1963, page 122.

Theorem A.2.1. *If $f(x)$ is continuous on $[a, a']$; if $g(y)$ is bounded on $[b, b']$; if $h(x, y)$ is of FBV on $[a, a'] \times [b, b']$ and of bounded variation in x for $y = b$ (and so for all y); and if $\int_b^{b'} g(y) d_x h(x, y)$ exists for all x , then $\int_a^{a'} f(x) d_x \int_b^{b'} g(y) d_y h(x, y)$ and $\int_b^{b'} g(y) d_y \int_a^{a'} f(x) d_x h(x, y)$ both exist and are equal.*

Proof. See Theorem 7.14, page 122, in Hildebrandt 1963, Chapter III.

A.3 Some results on stochastic integrals.

Definition A.3.1. *A stochastic process $\{X_t\}_{t \in T}$ is said to be of second order if $EX^2(t) < \infty$ for all $t \in T$.*

Definition A.3.2. *Let $\{X_n\}$ be a sequence of random variables such that $EX_n^2 < \infty$ for each n . $\{X_n\}$ is said to converge in quadratic mean (or in mean square) to a random variable X , denoted $X_n \xrightarrow{L^2} X$, if and only if $E|X_n - X|^2 \rightarrow 0$, as $n \rightarrow \infty$.*

Theorem A.3.1. *Let $\{X_n\}$ be a sequence of random variables which converges in quadratic mean to a random variable X . Then, $\lim_{n \rightarrow \infty} EX_n = EX$ and $\lim_{n \rightarrow \infty} EX_n^2 = EX^2$.*

Proof. See Theorem 2.3.1 of Lukacs, 1975, Chapter II.

Theorem A.3.2. A sequence $\{X_n\}_{n \geq 1}$ of random variables such that $EX_n^2 < \infty$ for all n , converges in quadratic mean if and only if $\lim_{n,m \rightarrow \infty} EX_n X_m$ exists and is independent of the manner in which n and m tend to infinity.

Proof. See Corollary to Theorem 2.4.5 of Lukacs, 1975, Chapter II.

Proposition A.3.1. Let $\{\psi(x), x \in \mathbb{R}\}$ be a second order stochastic process with mean zero and covariance function $C(x, z) := E\psi(x)\psi(z)$. Let $l \in C^1[0, 1]$, the set of continuous function on $[0, 1]$, having continuous first derivative, be a function of bounded variation over closed subintervals $[a, b] \subseteq [0, 1]$. Then

$$(a) \quad E \left(\int_a^b \psi(x) dx l(x) \right)^2 = \int_a^b \int_a^b C(x, z) d_x l(x) d_z l(z) \text{ and}$$

$$(b) \quad E \left(\psi(\xi) \int_a^b \psi(x) d_x l(x) \right) = \int_a^b C(\xi, x) d_x l(x),$$

provided the Riemann-Stieltjes integrals exist.

Proof. (a) Let $S_n := \sum_{i=1}^n \psi(t_{i,n}^*) (l(t_{i,n}) - l(t_{i-1,n}))$, where $a = t_{0,n} < t_{1,n} < \dots < t_{n,n} = b$ is a partition of $[a, b]$, such that $\lim_{n \rightarrow \infty} \max_{1 \leq k \leq n} \Delta_k := \lim_{n \rightarrow \infty} \max_{1 \leq k \leq n} (t_{k,n} - t_{k-1,n}) = 0$, and $t_{k,n}^*$ is any point in each subinterval $[t_{k-1,n}, t_{k,n}]$. That the integral exists means that there exists an S such that, for any $\varepsilon > 0$, there exists $\delta > 0$ such that $|S_n - S| < \varepsilon$ when $\max_{1 \leq k \leq n} \Delta_k < \delta$, or, in other words, there exists $\lim_{\max_{1 \leq k \leq n} \Delta_k \rightarrow 0} S_n$, for any choice of points $t_{k,n}^*$ in each subinterval $[t_{k-1,n}, t_{k,n}]$.

Since $E\psi(\xi) = 0 \forall \xi \in \mathbb{R}$, we have that $ES_n = 0$ for all n . Also,

$$\begin{aligned} E(S_n S_m) &= E\left(\sum_{i=1}^n \sum_{j=1}^m \psi(t_{i,n}^*) \psi(t_{j,m}^*) (l(t_{i,n}) - l(t_{i-1,n})) (l(t_{j,m}) - l(t_{j-1,m}))\right) \\ &= \sum_{i=1}^n \sum_{j=1}^m E(\psi(t_{i,n}^*) \psi(t_{j,m}^*)) (l(t_{i,n}) - l(t_{i-1,n})) (l(t_{j,m}) - l(t_{j-1,m})). \end{aligned}$$

As $n, m \rightarrow \infty$,

$$(A.3.1) \quad E(S_n S_m) \rightarrow \int_a^b \int_a^b E(\psi(t) \psi(s)) d_t l(t) d_s l(s)$$

(independently of the manner in which n and m tend to infinity).

By Theorem A.3.2, $S_n \xrightarrow{L^2} X$ for some random variable X . Since this implies $S_n \rightarrow X$ in probability, we have $S = X$ a.s., i.e.,

$$S_n \xrightarrow{L^2} \int_a^b \psi(x) d_x l(x), \text{ as } n \rightarrow \infty.$$

By (A.3.1), $\lim_{n \rightarrow \infty} E(S_n^2) = \int_a^b \int_a^b C(x, z) d_x l(z) d_z l(x)$, and, by Theorem A.3.1, this limit also equals $E(S^2) = E\left(\int_a^b \psi(x) d_x l(x)\right)^2$.

Part (b) can be proved in a similar way. ■

Proposition A.3.2. Assume the same conditions of Proposition A.3.1. Then,

$$\begin{aligned} E\left(\int_a^b l(x) d_x \psi(x)\right)^2 &= \int_a^b l(x) d_x \left[\int_a^b l(z) d_z C(x, z)\right] \\ &=: \int_a^b \int_a^b l(x) l(z) d_x d_z C(x, z). \end{aligned}$$

Proof. Using integration by parts we get

$$\begin{aligned}
 E\left(\int_a^b l(x) d_x \psi(x)\right)^2 &= E\left(l(b)\psi(b) - l(a)\psi(a) - \int_a^b \psi(x) d_x l(x)\right)^2 \\
 &= l^2(b)C(b, b) + l^2(a)C(a, a) + E\left(\int_a^b \psi(x) d_x l(x)\right)^2 \\
 &\quad - 2l(a)l(b)C(a, b) - 2l(b)E\left(\psi(b) \int_a^b \psi(x) d_x l(x)\right) \\
 &\quad + 2l(a)E\left(\psi(a) \int_a^b \psi(x) d_x l(x)\right).
 \end{aligned}$$

On the other hand,

$$\begin{aligned}
 &\int_a^b l(x) d_x \left[\int_a^b l(z) d_z C(x, z) \right] \\
 &= \int_a^b l(x) d_x \left\{ l(b)C(x, b) - l(a)C(x, a) - \int_a^b C(x, z) d_z l(z) \right\} \\
 &= l(b) \int_a^b l(x) d_x C(x, b) - l(a) \int_a^b l(x) d_x C(x, a) - \int_a^b l(x) d_x \left[\int_a^b C(x, z) d_z l(z) \right] \\
 &= l(b) \left[l(b)C(b, b) - l(a)C(a, b) - \int_a^b C(x, b) d_x l(x) \right] - l(a) \left[l(b)C(b, a) - l(a) \right. \\
 &\quad \left. C(a, a) - \int_a^b C(x, a) d_x l(x) \right] - l(x) \int_a^b C(x, z) d_z l(z) \Big|_a^b + \int_a^b \int_a^b C(x, z) d_x l(x) d_z l(z) \\
 &= l^2(b)C(b, b) - 2l(a)l(b)C(a, b) - l(b) \int_a^b C(x, b) d_x l(x) + l^2(a)C(a, a) \\
 &\quad + l(a) \int_a^b C(x, a) d_x l(x) - l(b) \int_a^b C(b, z) d_z l(z) + l(a) \int_a^b C(a, z) d_z l(z)
 \end{aligned}$$

$$+ \int_a^b \int_a^b C(x, z) d_z l(z) d_x l(x).$$

Since $C(x, y) = C(y, x)$ for all x, y , we have that the result follows by Proposition A.3.1. ■

Let $\{B_G(x, y), x, y \in [0, 1]\}$ be a Brownian bridge associated with the distribution function G , i.e. B_G is a separable Gaussian process with $EB_G(x, y) = 0$ and

$$EB_G(x, y)B_G(s, t) = G(x \wedge s, y \wedge t) - G(x, y)G(s, t).$$

Let $[a, b] \times [c, d]$ be a subinterval of $[0, 1] \times [0, 1]$ and

$$(A.3.2) \quad \psi(x) := \int_c^d l_2(y) d_y B_G(x, y), \quad x \in [a, b],$$

where $l_2 \in C^1[0, 1]$ and is of bounded variation on any $[c, d] \subseteq [0, 1]$. We compute $E\psi(x)\psi(z)$.

Using integration by parts we have

$$\begin{aligned} E\psi(x)\psi(z) &= E\left(l_2(d)B_G(x, d) - l_2(c)B_G(x, c) - \int_c^d B_G(x, y) d_y l_2(y)\right) \\ &\quad \times \left(l_2(d)B_G(z, d) - l_2(c)B_G(z, c) - \int_c^d B_G(z, w) d_w l_2(w)\right) \\ &= l_2^2(d)(G(x \wedge z, d) - G(x, d)G(z, d)) - l_2(c)l_2(d)[G(x \wedge z, c) - G(x, d)G(z, c)] \\ &\quad - l_2(d) \int_c^d [G(x \wedge z, w) - G(x, d)G(z, w)] d_w l_2(w) \\ &\quad - l_2(c)l_2(d)(G(x \wedge z, c) - G(x, c)G(z, d)) \\ &\quad + l_2^2(c)(G(x \wedge z, c) - G(x, c)G(z, c)) \\ &\quad + l_2(c) \int_c^d [G(x \wedge z, c) - G(x, c)G(z, w)] d_w l_2(w) \end{aligned}$$

$$\begin{aligned}
& -l_2(d) \int_c^d [G(x \wedge z, y) - G(z, d)G(x, y)] d_y l_2(y) \\
& + l_2(c) \int_c^d [G(x \wedge z, c) - G(z, c)G(x, y)] d_y l_2(y) \\
& + \int_c^d \int_c^d [G(x \wedge z, w \wedge y) - G(x, y)G(z, w)] d_w l_2(w) d_y l_2(y) \\
& = \left\{ l_2^2(d)G(x \wedge z, d) - l_2(c)l_2(d)G(x \wedge z, c) - l_2(d) \int_c^d G(x \wedge z, y) d_y l_2(y) \right. \\
& \quad - l_2(c)l_2(d)G(x \wedge z, c) + l_2^2(c)G(x \wedge z, c) \\
& \quad + l_2(c) \int_c^d G(x \wedge z, c) d_y l_2(y) - l_2(d) \int_c^d G(x \wedge z, w) d_w l_2(w) \\
& \quad + l_2(c) \int_c^d G(x \wedge z,) d_w l_2(w) + \int_c^d \int_c^d G(x \wedge z, y \wedge w) d_w l_2(w) d_y l_2(y) \Big\} \\
& \quad - \left\{ l_2^2(d)G(z, d)G(x, d) - l_2(d)l_2(c)G(z, d)G(x, c) \right. \\
& \quad - l_2(d) \int_c^d G(z, d)G(x, y) d_y l_2(y) - l_2(c)l_2(d)G(x, d)G(z, c) \\
& \quad + l_2^2(c)G(x, c)G(z, c) + l_2(c) \int_c^d G(z, c)G(x, y) d_y l_2(y) \\
& \quad - l_2(d) \int_c^d G(x, d)G(z, w) d_w l_2(w) + l_2(c) \int_c^d G(x, c)G(z, w) d_w l_2(w) \\
& \quad \left. + \int_c^d \int_c^d G(x, y)G(z, w) d_w l_2(w) d_y l_2(y) \right\} \\
& = \int_c^d l_2(y) d_y \left[l_2(d)G(x \wedge z, y) - l_2(c)G(x \wedge z, c) - \int_c^d G(x \wedge z, y \wedge w) d_w l_2(w) \right]
\end{aligned}$$

$$\begin{aligned}
& - \int_c^d l_2(y) d_y \left[l_2(c) G(x, y) G(z, d) - l_2(c) G(x, y) G(z, c) \right. \\
& \quad \left. - \int_c^d G(x, y) G(z, w) d_w l_2(w) \right] \\
& = \int_c^d l_2(y) d_y \left[\int_c^d l_2(w) d_w G(x \wedge z, y \wedge w) \right] \\
& \quad - \int_c^d l_2(y) d_y \left[\int_c^d l_2(w) d_w G(x, y) G(z, w) \right] \\
& = \int_c^d l_2(y) d_y \left[\int_c^d l_2(w) d_w [G(x \wedge z, y \wedge w) - G(x, y) G(z, w)] \right],
\end{aligned}$$

i.e.

$$(A.3.3) \quad E(\psi(x)\psi(y)) = \int_c^d l_2(y) d_y \left[\int_c^d l_2(w) d_w E B_G(x, y) B_G(z, w) \right].$$

Let $l_1 \in C^1[0, 1]$ be of bounded variation on any interval $[a, b] \subseteq [0, 1]$. By (A.3.3) and Proposition A.3.2, we have

$$\begin{aligned}
& E \left(\int_a^b l_1(x) d_x \left[\int_c^d l_2(y) d_y B_G(x, y) \right] \right)^2 = \\
& = \int_a^b l_1(x) d_x \left[\int_a^b l_1(z) d_z \left[\int_c^d l_2(y) d_y \left[\int_c^d l_2(w) d_w E B_G(x, y) B_G(z, w) \right] \right] \right] \\
& = \int_a^b l_1(x) d_x \left[\int_a^b l_1(z) d_z \left[\int_c^d l_2(y) d_y \left[\int_c^d l_2(w) d_w [G(x \wedge z, y \wedge w) - \right. \right. \right. \\
& \quad \left. \left. \left. G(x, y) G(z, w) \right] \right] \right] \right] \\
& = \int_a^b l_1(x) d_x \left[\int_a^b l_1(z) d_z \left[\int_c^d l_2(y) d_y \left[\int_c^d l_2(w) d_w [G(x \wedge z, y \wedge w) - \right. \right. \right.
\end{aligned}$$

$$\begin{aligned}
& G(x \wedge z, y)G(x \wedge z, w)] \Big] \Big] \\
& + \int_a^b l_1(x) d_x \left[\int_a^b l_1(z) d_z \left[\int_c^d l_2(y) d_y \left[\int_c^d l_2(w) d_w G(x \wedge z, y)G(x \wedge z, w) \right] \right] \right] \\
& - \int_a^b l_1(x) d_x \left[\int_a^b l_1(z) d_z \left[\int_c^d l_2(y) d_y \left[\int_c^d l_2(w) d_w G(x, y)G(z, w) \right] \right] \right] \\
& = \int_a^b l_1(x) d_x \left[\int_a^b l_1(z) d_z E \left(\int_c^d l_2(y) B_G(x \wedge z, y) \right)^2 \right] \\
& + \int_a^b l_1(x) d_x \left[\int_a^b l_1(z) d_z \left(\int_c^d l_2(y) d_y G(x \wedge z, y) \right)^2 \right] \\
& - \int_a^b l_1(x) d_x \left[\int_a^b l_1(z) d_z \left[\int_c^d l_2(y) d_y \left[\int_c^d l_2(w) d_w G(x, y)G(z, w) \right] \right] \right],
\end{aligned}$$

i.e.,

$$\begin{aligned}
E \left(\int_a^b l_1(x) d_x \left[\int_c^d l_2(y) d_y B_G(x, y) \right] \right)^2 &= \int_a^b l_1(x) d_x \int_a^b l_1(z) d_z \left[\int_c^d l_2^2(y) d_y G(x \wedge z, y) \right] \\
&\quad - \left(\int_a^b l_1(x) d_x \left[\int_c^d l_2(y) d_y G(x, y) \right] \right)^2 \\
(A.3.4) \qquad \qquad \qquad &:= I_1 - I_2.
\end{aligned}$$

By Theorem A.2.1 and Proposition A.3.1, we have

(A.3.5)

$$\begin{aligned}
I_1 &= \int_c^d l_2^2(y) d_y \left[\int_a^b l_1(x) d_x \left[\int_a^b l_1(z) d_z [G(x \wedge z, y) - G(x, y)G(z, y)] \right] \right] \\
&\quad + \int_c^d l_2^2(y) d_y \left[\int_a^b l_1(x) d_x \left[\int_a^b l_1(z) d_z G(x, y)G(z, y) \right] \right] \\
&= \int_c^d l_2^2(y) d_y E \left(\int_a^b l_1(x) d_x B_G(x, y) \right)^2
\end{aligned}$$

$$\begin{aligned}
& + \int_c^d l_2^2(y) dy \left[\int_a^b l_1(x) dx \left[\int_a^b l_1(z) dz G(x, y) G(z, y) \right] \right] \\
& = \int_c^d l_2^2(y) dy \left[\int_a^b l_1^2(x) dx G(x, y) - \left(\int_a^b l_1(x) dx G(x, y) \right)^2 \right] \\
& + \int_c^d l_2^2(y) dy \left[\int_a^b l_1(x) dx \left[\int_a^b l_1(z) dz G(x, y) G(z, y) \right] \right] \\
& = \int_a^b \int_c^d l_1^2(x) l_2^2(y) dx dy G(x, y).
\end{aligned}$$

Hence, by (A.3.4) and (A.3.5), we have that

$$\begin{aligned}
& E \left(\int_a^b l_1(x) dx \left[\int_c^d l_2(y) dy B_G(x, y) \right] \right)^2 \\
\text{(A.3.6)} \quad & = \int_a^b \int_c^d l_1^2(x) l_2^2(y) dx dy G(x, y) - \left(\int_a^b l_1(x) dx \left[\int_c^d l_2(y) dy G(x, y) \right] \right)^2.
\end{aligned}$$

References

- ADLER, R.J. and BROWN, L.D. (1986). Tail behaviour for suprema of empirical processes. *Ann. Probab.* 14 1-30.
- ANDERSON, T.W. (1958). *Introduction to Multivariate Statistical Analysis*, J. Wiley, New York.
- BICKEL, P.J. and DOKSUM, K.A. (1977). *Mathematical Statistics: Basic ideas and selected topics*. Holden-Day, California.
- BILLINGSLEY, P. (1968). *Convergence of Probability Measures*, J. Wiley, New York.
- BILLINGSLEY, P. (1986). *Probability and Measure*, second edition, J. Wiley, New York.
- BRILLINGER, D.L. (1969). An asymptotic representation of the sample distribution function. *Bull. Amer. Math. Soc.* 75 545-547.
- BRODSKY, B.E. and DARKHOVSKY, B.S (1993). *Nonparametric methods in Change-point Problems*, Kluwer, Dordrecht.
- BURKE, M.D., CSÖRGÖ, M., CSÖRGÖ, S., and RÉVÉSZ, P. (1979) Approximations of the empirical process when parameters are estimated *Ann. Prob.* 7 790-810.
- CANTELLI, F.P. (1933). Sulla determinazione empirica delle leggi di probabilità. *Giorn. Ist. Ital. Attuari* 4 421-424.

- CHIBISOV, D. (1964). Some theorems on the limiting behaviour of empirical distribution functions, *Selected Transl. Math. Statist. Prob.* **6** 147-156.
- CHUNG, K.L. (1949). An estimate concerning the Kolmogoroff limit distribution. *Trans. Amer. Math. Soc.* **67** 36-50.
- CHUNG, K.L. and WILLIAMS, R.J. (1983). *Introduction to Stochastic Integration*. Birkhäuser, Boston.
- CSÖRGÖ, M. (1983). *Quantile processes with statistical applications*. SIAM 42, Philadelphia, Pennsylvania.
- CSÖRGÖ, M., CSÖRGÖ, S., HORVÁTH, L. and MASON, D. (1986). Weighted empirical and quantile processes. *Ann. Probab.* **14** 31-85.
- CSÖRGÖ, M. and HORVÁTH, L. (1986b). Approximations of weighted empirical and quantile processes. *Stat. Probab. Letters* **4** 275-280.
- CSÖRGÖ, M. and HORVÁTH, L. (1987). Detecting change in a random sequence. *J. Multivariate Anal.* **23** 119-130.
- CSÖRGÖ, M. and HORVÁTH, L. (1988a). Invariance principles for changepoint problems. *J. Multivariate Anal.* **27** 151-168.
- CSÖRGÖ, M. and HORVÁTH, L. (1988b). Nonparametric methods for changepoint problems. *Handbook of Statistics*, Vol. **7** 403-425, Elsevier Science Publishers B.V. (North-Holland).
- CSÖRGÖ, M. and HORVÁTH, L. (1988c). A note on strong approximations of multivariate empirical processes. *Stoch. Proc. Appl.* **27** 101-109.
- CSÖRGÖ, M. and HORVÁTH, L. (1993). *Weighted Approximations in Probability and Statistics*. J. Wiley, Chichester.
- CSÖRGÖ, M. and RÉVÉSZ, P. (1975). A new method to prove Strassen type laws of invariance principle, II. *Z. Wahrsch. Verw. Gebiete* **31** 261-269.
- CSÖRGÖ, M. and RÉVÉSZ, P. (1978). Strong approximations of the quantile process. *Ann. Statist.* **6** 882-894.

- CSÖRGŐ, M. and RÉVÉSZ, P. (1981). *Strong Approximations in Probability and Statistics*. Academic Press, New York.
- CSÖRGŐ, M., SHAO, Q.M. and SZYSZKOWICZ, B. (1991). A note on local and global functions of a Wiener process and some Rényi-type statistics. *Studia Sci. Math. Hungar.* **26** 239-259.
- CSÖRGŐ, M., HORVÁTH, L. and SZYSZKOWICZ, B. (1994). Integral tests for suprema of Kiefer processes with applications *Preprint*.
- CSÖRGŐ, M. and SZYSZKOWICZ, B. (1994). Applications of multi-time parameter processes to change-point analysis *Prob. Theory and Math. Stat.* 159-222.
- DARLING, D.A. and ERDŐS, P. (1956). A limit theorem for the maximum of normalized sums of independent random variables. *Duke Math. J.* **23** 143-145.
- DESHAYES, J. and PICARD, D. (1983). Ruptures de modèles en statistique. Thèses d'état, Université Paris Sud.
- DESHAYES, J. and PICARD, D. (1986). Off-line statistical analysis of change-point models using non parametric and likelihood methods. In *Lecture Notes in Control and Information Sciences* (eds. M. Thoma and A. Wyner) **77: Detection of Abrupt Changes in Signals and Dynamical Systems** (eds. M. Basseville and A. Benveniste 103-168. Springer-Verlag, Berlin.
- DONSKER, M. (1952). Justification and extension of Doob's heuristic approach to the Kolmogorov-Smirnov theorems. *Ann. Math. Statist.* **23** 277-283.
- DOOB, J.L. (1949). Heuristic approach to the Kolmogorov-Smirnov theorems. *Ann. Math. Statist.* **20**, 393-403.
- DURBIN, J. (1973). Weak convergence of the sample distribution function when parameters are estimated. *Ann. Statist.* **1** 279-290.
- FELLER, W. (1968). *An Introduction to Probability Theory and its Applications*. Vol. 1, Third edition, Wiley, New York.

- GLIVENKO, V. (1933). Sulla determinazione empirica delle leggi di probabilità. *Giorn. Ist. Ital. Attuari* 4 92-99.
- HÁJEK, J. and RÉNYI, A. (1955). Generalization of an inequality of Kolmogorov, *Acta Math. Acad. Sci. Hungar.* 6 281-283.
- HÁJEK, J. and ŠIDÁK, Z. (1967). *Theory of Rank Test*. Academic Press, New York.
- HILDEBRANT, T.H. (1963). *Introduction to the theory of integration*. Academic Press, New York.
- IBRAGIMOV, I.A. and HAS'MINSKIĬ, R.Z. (1972). Asymptotic behavior of some statistical estimators in the smooth case I. Study of the likelihood ratio. *Theory Probab. Appl.* 17 445-462.
- IBRAGIMOV, I.A. and HAS'MINSKIĬ, R.Z. (1973a). Asymptotic behavior of some statistical estimators II. Limit theorems for the a posteriori density and Baye's estimators. *Theory Probab. Appl.* 18 76-91.
- IBRAGIMOV, I.A. and HAS'MINSKIĬ (1973b). On the approximation of some statistical estimators by sums of independent random variables. *Soviet Math. Dokl.* 14 883-887.
- KHMALADZE, E.V. and PARJANADZE, A.M. (1986). Functional limit theorems for linear statistics from sequential ranks. *Probab. Theory Related Field* 73 585-595.
- KIEFER, J. (1972). Skorohod embedding of multivariate RV's and the sample D.F. *Z. Wahrsch. Verw. Gebiete* 24 1-35.
- KOMLÓS, J., MAJOR, P. and TUSNÁDY, G. (1975). An approximation of partial sums of independent R.V.'s and the sample D.F.I. *Z. Wahrsch. Verw. Gebiete* 32 111-131.
- KOMLÓS, J., MAJOR, P. and TUSNÁDY, G. (1976). An approximation of partial sums of independent R.V.'s and the sample D.F.II. *Z. Wahrsch. Verw. Gebiete* 34 33-58.

- KOLMOGOROV, A.N. (1933). Sulla determinazione empirica di una legge di distribuzione. *Giorn. Ist. Ital. Attuari* 4 83-91.
- LE CAM, L. (1960). Locally asymptotically normal families of distributions. *University of California Publications in Statistics*, 3 37-98.
- LE CAM, L. (1986). *Asymptotic Methods in Statistical Decision Theory*. Springer-Verlag, New York.
- LE CAM, L. and YANG, G.L. (1990). *Asymptotics in Statistics; Some Basic Concepts*. Springer-Verlag, New York.
- LEHMANN, E.L. (1991). *Theory of point estimation*. Wadsworth and Brooks/Cole. California.
- LUKACS, E. (1975). *Stochastic convergence*. Academic Press, New York.
- OOSTERHOFF, J. and VAN ZWET, W.R. (1975). A note on contiguity and Hellinger distance. In: J. Jurechkova, ed., *Contributions to Statistics*. Reidel, Dordrecht 157-166.
- O'REILLY, N. (1974). On the weak convergence of empirical processes in sup-norm metrics. *Ann. Probab.* 2 642-651.
- PARDZHANADZE, A.M. and KHMALADZE, E.V. (1986). On the asymptotic theory of statistics of sequential ranks. *Theory Probab. Appl.* 31 669-682.
- PICARD, D. (1985). Testing and estimating change-points in time series. *Adv. Appl. Prob.*, 17 126-135.
- POLLARD, D. (1984). *Convergence of stochastic processes*. Springer-Verlag, New York.
- PROHOROV, YU. V. (1956). Convergence of random processes and limit theorems in probability theory. *Theory Probab. Appl.* 1 157-214.
- PROTTER, M.H. and MORREY, C.B. (1991). *A first course in real analysis*. Second edition. Springer-Verlag, New York.

- PYKE, R. and SHORACK, G.R. (1968). Weak convergence of a two-sample empirical process and a new approach to Chernoff-Savage theorem. *Ann. Math. Statist.* **39** 755-771.
- RÉNYI, A. (1953). On the theory of order statistics. *Acta. Math. Acad. Sci. Hung.* **4** 191-232.
- ROUSSAS, G.G. (1972). *Contiguity of Probability Measures: Some Applications in Statistics*. Cambridge University Press.
- SHORACK, G.R. and WELLNER, J.A. (1986). *Empirical Processes with Applications to Statistics*. Wiley, New York.
- SMIRNOV, N.V. (1939). On the estimation of the discrepancy between empirical curves of distribution for two independent samples. *Bull. Math. de l'Université de Moscow* **2** (Fasc. 2).
- SZYSZKOWICZ, B. (1991a). Changepoint problem and contiguous alternatives. *Statist. Probab. Lett.* **11** 299-308.
- SZYSZKOWICZ, B. (1991b). Empirical type processes and contiguity. *C.R. Math. Rep. Acad. Sci. Canada. XIII* No.4.
- SZYSZKOWICZ, B. (1991c). Weighted Stochastic Processes under Contiguous alternatives. *C.R. Math. Rep. Acad. Sci. Canada. XIII* No. 5.
- SZYSZKOWICZ, B. (1992a). Weighted L_p , $0 < p \leq \infty$, convergence of Partial Sum type Processes. In *Tech. Rep. Ser. Lab. Res. Stat. Prob.* No. 191 Carleton Univ. and Univ. of Ottawa.
- SZYSZKOWICZ, B. (1992b). *Weak Convergence of Stochastic Processes in Weighted Metrics and their Applications to Contiguous Changepoint Analysis*. Ph.D Thesis, Carleton University.
- SZYSZKOWICZ, B. (1994). Weak convergence of weighted empirical type processes under contiguous and changepoint alternatives. *Stoch. Proc. Appl.* **50** 281-313.

SZYSZKOWICZ, B. (1995). Weighted sequential empirical type processes with applications to change-point problems. To appear in *Handbook of Statistics 16, Order Statistics and their Applications*.

END

3 1 - 0 5 - 9 / 6

FIN