CWI Tract 44

Statistical estimation in large parameter spaces

A.W. van der Vaart



Centrum voor Wiskunde en Informatica
Centre for Mathematics and Computer Science

1980 Mathematics Subject Classification: 62G05, 62G20, 62F35.
ISBN 90 6196 329 X
NUGI-code: 811

Copyright © 1988, Stichting Mathematisch Centrum, Amsterdam
Printed in the Netherlands

CWI Tracts

Managing Editors

J.W. de Bakker (CWI, Amsterdam)
M. Hazewinkel (CWI, Amsterdam)
J.K. Lenstra (CWI, Amsterdam)

Editorial Board

W. Albers (Enschede)
P.C. Baayen (Amsterdam)
R.T. Boute (Nijmegen)
E.M. de Jager (Amsterdam)
M.A. Kaashoek (Amsterdam)
M.S. Keane (Delft)
J.P.C. Kleijnen (Tilburg)
H. Kwakernaak (Enschede)
J. van Leeuwen (Utrecht)
P.W.H. Lemmens (Utrecht)
M. van der Put (Groningen)
M. Rem (Eindhoven)
A.H.G. Rinnooy Kan (Rotterdam)
M.N. Spijker (Leiden)

Centrum voor Wiskunde en Informatica
Centre for Mathematics and Computer Science
P.O. Box 4079, 1009 AB Amsterdam, The Netherlands

The CWI is a research institute of the Stichting Mathematisch Centrum, which was founded on February 11, 1946, as a nonprofit institution aiming at the promotion of mathematics, computer science, and their applications. It is sponsored by the Dutch Government through the Netherlands Organization for Scientific Research (NWO).

PREFACE

This manuscript is basically a reprint of my Ph.D. thesis, which appeared in May 1987. However, having fallen in the habit of sitting before a computer screen, I have not resisted the temptation to rewrite numerous parts of the original manuscript. Perhaps I should have.

Most of the changes are additions and many concern presentation. There are also some corrections. The most of the latter concerns the measurability problem with the statement of the former Theorem 2.17. I am grateful to one of the referees of my paper (1986b) for pointing this out.

In a University of Leiden thesis sentences containing the word *thank* are strictly forbidden. I am glad to have the present opportunity to thank Chris Klaassen and W.R. van Zwet for suggesting the model treated in Chapter 5, as well as for their numerous remarks, which have improved the presentation at many places. My thanks also go to Richard Gill, in particular for the stimulating discussions which motivated me to write Chapter 4; and to Jon Wellner for helpful discussions when revising the manuscript, especially concerning the subject of Section A.3.

Amsterdam, December 1987
Aad van der Vaart

CONTENTS

STATISTICAL ESTIMATION IN LARGE PARAMETER SPACES

1. INTRODUCTION 1

1.1. large parameter spaces 1

1.2. a Cramér-Rao bound 7

1.3. estimating location under symmetry 10

1.4. notation 13

2. ASYMPTOTIC BOUNDS ON THE PERFORMANCE OF ESTIMATORS WITH VALUES IN $\mathbb{R}^k$ 15

2.1. introduction 15

2.1.1. i.i.d. models 17

2.2. generalized convolution theorems 21

2.3. local asymptotic minimax theorems 26

2.4. convolution theorem 32

2.5. a theorem on asymptotic variance 33

2.6. asymptotic optimality 36

2.7. theorems for locally asymptotically normal models 43

3. ASYMPTOTIC BOUNDS ON THE PERFORMANCE OF ESTIMATORS WITH VALUES IN A VECTOR SPACE 47
3.1. introduction 47
3.1.1. differentiable functionals, influence functions 49
3.1.2. cylinder measures 50
3.1.3. (weakly regular) estimator sequences 53
3.2. convolution theorem 54
3.2.1. convolution of measures on $(B,\mathcal{U}(B'))$ 54
3.2.2. convolution theorem 55
3.3. local asymptotic minimax theorems 58
3.3.1. LAM theorems 58
3.3.2. subconvex loss functions 60
3.4. two results on the optimal limiting measure 62
3.5. special cases 64
3.5.1. separable normed spaces 64
3.5.2. D[a,b] with the projection σ-field 66
3.5.3. D[a,b] with Skorohod topology 70
3.5.4. the space B(T) 70
3.5.5. separable Hilbert space 74
3.6. examples of differentiable functionals 75
3.6.1. distribution function on the real line 75
3.6.2. set-indexed distributions 76
3.6.3. semi-parametric models I 77
3.6.4. semi-parametric models II 80
3.7. a characterization of D^* 82
3.8. some results on spaces of real functions 85

4. HADAMARD DIFFERENTIABLE FUNCTIONALS OF EFFICIENT ESTIMATORS 89
4.1. introduction 89
4.1.1. convergence in distribution 89
4.1.2. Hadamard differentiability 91
4.2. efficiency 94
4.2.1. definition of efficiency 94
4.2.2. efficiency, marginal efficiency and tightness 95
4.3. main results 101
4.4. examples 104
4.4.1. models with right censored observations 104
4.4.2. estimating a point symmetric distribution on $\mathbb{R}^k$ 111

5. SEMI-PARAMETRIC MODELS WITH A SUFFICIENT STATISTIC FOR THE NUISANCE PARAMETER 115
5.1. introduction 115
5.1.1. the i.i.d. model 118
5.1.2. the non i.i.d. model 121
5.1.3. some regularity conditions 123
5.2. an estimator for θ 124
5.3. estimation of location scores 129
5.4. efficiency 132
5.4.1. i.i.d. models 132
5.4.1.1. i.i.d. models convex in η 133
5.4.2. non i.i.d. models 135
5.5. mixture models, models with incidental parameters 138
5.5.1. efficiency for mixture models 140
5.6. estimation of some other functionals 145
5.7. examples 149
5.7.1. symmetric location in $\mathbb{R}$ 149
5.7.2. symmetric location in higher dimensions 152
5.7.3. elliptic distributions 153
5.7.4. two sample location scale 153
5.7.5. a paired hazard model 155
5.7.6. generalizations of example 5.7.5 160
5.7.7. errors in variables 162
5.7.8. Neyman-Scott model 163
5.8. technical results 164
5.8.1. some useful lemmas concerning mixture models 164
5.8.2. technical results 166
5.8.3. proofs 168

APPENDIX 175
A.1. contiguity 175
A.2. LAN, differentiability in quadratic mean and related results 181
A.3. differentiability, L_2-continuity, measurable transformations 188

REFERENCES 195
SUBJECT INDEX 201
SYMBOL INDEX 205

CHAPTER 1

INTRODUCTION

1.1. LARGE PARAMETER SPACES

A major part of statistics is concerned with parametric models. Here by *parametric* it is understood that the observable random variables possess a probability distribution which is known up to a vector θ of finitely many real numbers. On the basis of observed values of the random variables the statistician wishes to make inference about the value of the parameter θ or some functional thereof.

One of the simplest models of this kind refers to an experimenter who measures a natural constant θ. After n replications of his experiment he will have n different values, all approximately equal to θ. How is he to combine his data into a single estimate of θ? A classical formulation of the problem would be to assume that the measurement errors, which have apparently been made, are independent and normally distributed. Formally, letting X_j be the observations and e_j the errors,

$$X_j = \theta + e_j \qquad\qquad j = 1,2,\ldots,n \ ,$$

where it is assumed that the probability of having an error in the interval (a,b) equals

$$P(e_j \in (a,b)) = \int_a^b \sigma^{-1}(2\pi)^{-\frac{1}{2}} e^{-\frac{1}{2}x^2/\sigma^2}\, dx \ .$$

In this case the statistical model is completely fixed except for the parameter $(\theta,\sigma^2) \in \mathbb{R}^2$. It can be shown that a variety of optimality criteria lead to the *average* $\bar{X} = n^{-1}\Sigma_{j=1}^{n} X_j$ as the most accurate estimate of θ.

It is unreasonable to expect that a model as simple as this can reflect reality in an accurate way. Though the assumption of normality is sometimes motivated by reference to the central limit theorem, it seems clear that it is often made for mathematical convenience rather than accurate description. Indeed, historically the normal distribution was introduced to motivate the method of least squares and obtain the average as an optimal estimate.

Of course the criticism of oversimplification applies to a certain extent to any parametric model. Two attempts to remove the limitations of a finite dimensional parametrization are distribution free methods and robust methods.

Robust methods continue to consider parametric models, adopting as basic assumption that the true distribution of the observable variables is perhaps not a member of an *ideal* parametric model, but is at least in the neighbourhood of such a model.

For instance, in a popular model in robust statistics for our simple example, the distribution of the error terms e_j is a member of the set of distributions

$$\{(1-\varepsilon)N(0,\sigma^2) + \varepsilon H\colon\ \sigma > 0,\ H \text{ arbitrary distribution on } \mathbb{R}\},$$

where $\varepsilon > 0$ is a fixed constant.

Now that the true distribution is only known up to some neighbourhood system, it is not clear at first which quantity is to be estimated. However, estimators are compared as to the stability of their behaviour when the underlying distribution varies over the neighbourhood. This usually leads to estimators which are little affected by extreme observations, such as *trimmed means* $(n-2\alpha n)^{-1}\Sigma\ X_j$, where the α-th fractions of largest and smallest observations are deleted from the sum.

Given a class P of underlying distributions of the observations, a statistic (i.e. any measurable function of the observations) is called *distribution free* over P, if its distribution is the same under any member of the class. Any method based on such a statistic may be called a

distribution free method. Typically, however, the name distribution free is used for the case that P is a large class. In the latter case it is necessary to disregard much of the quantitative information available in the data and to base inference on e.g. ranks and signs of (translations of) the observations. This explains why distribution free methods are often considered to belong to *nonparametric statistics*.

For our example one assumes for instance that the underlying distribution of the observations belongs to the class P of all distributions which are symmetric about θ. In that case statistics such as $I(\theta)$, the number of X-values larger than θ, and $R(\theta)$, the sum of the ranks of the X-values which are larger than θ, are distribution free over the class P. Though these statistics depend on the unknown value of θ, they can be used in inference, as relatively large values of $I(\theta_0)$ or $R(\theta_0)$ for some fixed θ_0 are indicative of the true value of θ being larger than θ_0.

Distribution free methods are best known in statistical testing theory. Then one wishes to test the *null-hypothesis* that the underlying distribution of the observations belongs to the hypothesized class P against some alternative hypothesis. Thus in our example the null hypothesis might be that this underlying distribution is symmetric about some given value θ. Then the main attraction is that under the null-hypothesis, the distribution of the test statistic is completely known, so that the *level* of the test (the probability of rejecting the null hypothesis though it is true) can be completely controlled.

The name *nonparametric statistics* is used in several different meanings. Logically a nonparametric model is any model which is not parametric in the restricted sense introduced in the beginning of this section. In this sense nonparametric is close to *infinite dimensional*. Historically, the name seems also to be used in two more specific meanings. First to denote statistical methods based on ranks and signs, in which meaning it is close to distribution free. Second to denote statistical models where the underlying distribution is completely unknown. It is in the latter sense, for instance, that the empirical distribution function is called the nonparametric maximum likelihood estimator of a distribution function.

Robustness and nonparametric methods in its narrow sense have been studied extensively in the past decades. Procedures have been suggested for many statistical problems and are widely applied.

Chapter 1

Recently a new interest has developed for nonparametric models of an intermediate type, wherein the underlying distribution of the observations is not completely unknown, but the unknown part is larger than a Euclidean vector. Actually some of these models have been known for long and are implicitly or explicitly present in robust and distribution free statistics, but many new ones have also been introduced. Many of these have their roots in applied statistics. An explanation for their popularity may be that they are a natural extension of parametric models.

In a new approach to models of this type one starts with giving a full description of a model. Next, though there seems to be no intrinsic reason for this, one is usually concerned with estimation theory. Then one aims at finding optimal estimators for functionals which are defined explicitly on the model. The emphasis on explicitly defined functionals is an aspect that seems to set the new approach apart from distribution free and robust statistics. Indeed, reinterpreting these paradigms from the point of view of the new approach, one can say that in robust and distribution free statistics, the functionals one makes inference about, are chosen according to the criteria of stability or invariance, rather than fixed from the beginning. In the examples considered above, these functionals are for instance $\int_{\alpha}^{1-\alpha} F^{-1}(s)\,ds$ and $P(X_1-\theta_0 > 0)$. The precise relation of the new approach to robust and distribution free statistics remains to be investigated.

For our example a model might for instance be specified by the requirement that the distribution of the error terms is symmetric about zero, but otherwise unknown. The observable quantities X_j would therefore have distributions symmetric about θ and the problem of estimating the symmetry point θ is well-defined. We return to this example in Section 1.3.

Corresponding to historically grown terminology, a model of intermediate type is referred to as a *parametric-nonparametric model* or *semi-parametric model*. Though we do not have a short alternative name, we note that the latter terminology is not particularly satisfactory. *Parameter* is often used in a general sense to denote any quantity which (partly) determines a model. In combination with the above terminology one even speaks of nonparametric (i.e. infinite dimensional) nuisance parameters. If parameter is used in its general sense, any model is parametric and calling models *nonparametric* must be considered an abuse of language.

Introduction

In this manuscript a parameter is not necessarily a quantity in $\mathbb{R}^k$ and we can therefore speak of *statistical estimation in large parameter spaces.* While a major part of the manuscript is of a general nature, our main concern will be with the intermediate case where the underlying distribution is not completely unknown, nor can be specified by a finite dimensional parametric model in a smooth way.

Of course, our results concern only a limited area of the rather formidable part of statistics which could be ranked under our general title. We close this section with a short overview of the manuscript.

A main attraction of the statistical theory for large parameter spaces treated here, is the possibility of defining a notion of *asymptotic efficiency* of estimators, which designates essentially one estimator as an optimal estimator. Here *asymptotic* refers to the number of observations n, which approaches infinity. *Efficiency* is defined in terms of bounds on the performance of estimators. First one establishes such bounds and next one constructs estimators which attain these bounds. This general scheme is analogous to that of the finite dimensional parametric set-up.

Asymptotics is a necessary ingredient of the approach. Even for parametric models, considerations for a finite number of observations often do not lead to decidable problems, usually leaving the statistician with a large number of incomparable and/or intractable possibilities. For the more general problems under consideration asymptotics is indispensable to induce simplifications which lead to a mathematically attractive theory. Of course the question is of interest, as to how well the asymptotic results approximate the finite sample situation of the practical statistician. At present this question is not mathematically tractable and it is doubtful that it ever will be. One can but resort to computer simulations for the necessary selection and fine-tuning of procedures. This area of work is beyond our immediate interest, so that there are no tables in this manuscript.

Bounds on the performance of estimators of parameters and functions thereof in finite dimensional parametric models, have a long history. In general, these consist of a statement that a measure of discrepancy between the estimator and the estimated quantity is larger than some number. A well-known example is the Cramér-Rao bound. In two papers Hájek (1970,1972) gave a description of asymptotic lower bounds. His theorems were particularly suitable for generalization to general statistical models.

This has been accomplished in the past ten years, the results being in analogy with the parametric case. In Chapters 2 and 3 we discuss results of this type for estimators of smooth functionals in regular statistical models. Here regular models are those which can be approximated in a local sense by normal models. The latter terminology stems from the general theory of asymptotic statistics due to Le Cam (cf. Le Cam (1986)). While it should be possible to deduce a number of the results in Chapters 2 and 3 within the framework of Le Cam, our presentation is self-contained.

The results in Chapter 2 concern the estimation of functionals of the underlying distribution with values in $\mathbb{R}^k$. Next in Chapter 3 we extend some of the results of Chapter 2 to functionals with values in a general vector space.

With lower bounds given, an estimator is called *efficient* if it has the discrepancy measure equal to the lower bound. Now, while the theory of lower bounds is nearly complete, also for large parameter spaces, the construction of efficient estimators is at present an important open problem, except for an increasing number of special cases. This is markedly different from the parametric situation, where it has been proved in general that the known lower bounds are sharp, i.e. that asymptotically efficient estimators exist. In particular, maximum likelihood estimators or modifications thereof are classical examples of efficient estimators. No such results exist for the more general case, though efficient estimators have been constructed for special models, usually either by *adaptive procedures* or by *nonparametric maximum (or partial) likelihood.* In chapter 5 we construct efficient estimators in a class of models by a method which falls in the first category.

Given an estimator one expects that the property of being efficient is retained under the application of a smooth functional to the estimator and the estimated quantity. A theorem in this direction is established in Chapter 4. Its usefulness is illustrated by two examples.

An appendix containing results on contiguity and differentiability in quadratic mean, which are used many times, completes the manuscript.

As a further introduction to the manuscript, we discuss a generalization of the Cramér-Rao bound in Section 1.2, which may serve to introduce Chapters 2 and 3. Next we consider estimation of θ in the example introduced at page 1, in Section 1.3, which prepares for Chapter 5.

1.2. A CRAMÉR-RAO BOUND

Let $\mathcal{P}$ be a set of probability measures on a measurable space $(X,\mathcal{B})$ and let κ: $\mathcal{P} \to \mathbb{R}$ be a functional which is to be estimated by an estimator T_n based on an i.i.d. sample from some $P \in \mathcal{P}$. We assume that $\mathcal{P}$ is dominated by a σ-finite measure μ on $(X,\mathcal{B})$ and let $p = dP/d\mu$. The Cramér-Rao theorem gives a lower bound for the variance of *unbiased* estimators of $\kappa(P)$, i.e. $T_n = t_n(X_1,\ldots,X_n)$, satisfying

$$(1.1) \qquad E_P T_n = \kappa(P) \qquad \text{all } P \in \mathcal{P}.$$

To obtain the bound we define *differentiable submodels* of $\mathcal{P}$ and *differentiable functionals* κ.

A map $t \to P_t$ from $[0,1] \subset \mathbb{R}$ to $\mathcal{P}$ is called a (one-dimensional) *differentiable submodel* if there exists a measurable function g: $(X,\mathcal{B}) \to \mathbb{R}$ with as $t \downarrow 0$

$$(1.2) \qquad \int [t^{-1}(p_t^{\frac{1}{2}} - p^{\frac{1}{2}}) - \tfrac{1}{2} g\, p^{\frac{1}{2}}]^2 \, d\mu \to 0.$$

Disregarding the integral in (1.2) and considering g as a pointwise limit, we would have

$$(1.3) \qquad g(x) = 2p^{-\frac{1}{2}}(x)\, \partial/\partial t\, p_t^{\frac{1}{2}}(x) \,|_{t=0} = \partial/\partial t \log p_t(x) \,|_{t=0}.$$

The function g defined by (1.2) can therefore be considered a quadratic mean version of a *score function* in the one dimensional parametric model $\{P_t\colon t \in [0,1]\}$. Though this is inspired by asymptotics rather than the finite sample situation, definition of scores in terms of (1.2) instead of (1.3) precludes many of the awkward regularity conditions usually attached to the Cramér-Rao bound.

When $t \to P_t$ is a differentiable submodel leading to a score g, then for any $a > 0$ $t \to P_{at\wedge 1}$ is a differentiable submodel with score ag. These submodels are said to be in the same direction. If $\mathcal{P}$ is a large set of probability measures there may be submodels in many directions. The set of all scores is called a *tangent cone* and is denoted $T(P)$. Note that (1.2) implies that $T(P)$ is a subset of the Hilbert space $L_2(P)$, i.e. $\int g^2\, dP < \infty$ for every $g \in T(P)$.

Next we introduce *differentiable functionals*. Let $\{P_t\}$ satisfy (1.2). The idea of differentiation is to approximate differences $\kappa(P_t) - \kappa(P)$ by a linear functional $d\kappa$ in the sense that

$$\kappa(P_t) - \kappa(P) = d\kappa(P_t - P) + o(t). \tag{1.4}$$

However the fact that $\mathcal{P}$ is not a vector space makes a definition of type (1.4) inconvenient. One could embed $\mathcal{P}$ into a vector space, but this would not take the special characteristics of a set of probability measures into account. Following Pfanzagl (1982), we choose to define a derivative of κ at $P \in \mathcal{P}$ as a linear map $\kappa'_P: T(P) \to \mathbb{R}$, satisfying

$$\kappa(P_t) - \kappa(P) = t\, \kappa'_P(g) + o(t)\ , \tag{1.5}$$

when $\{P_t\}$ satisfies (1.2). Usually κ'_P admits a representation as an inner product

$$\kappa'_P(g) = \int g(x)\, \dot{\kappa}(x,P)\, dP(x). \tag{1.6}$$

for some element $\dot{\kappa}(x,P)$ of $L_2(P)$. Indeed, (1.5)-(1.6) relate to similar definitions in robust statistics, where $\dot{\kappa}(\cdot,P)$ is known as an *influence function* of κ.

We are ready to obtain the Cramér-Rao bound. By (1.1), (1.3) and (1.5), informally

$$\begin{aligned}
\kappa'_P(g) &= \lim_{t \downarrow 0}\ t^{-1}(\kappa(P_t) - \kappa(P)) \\
&= \lim_{t \downarrow 0} \int T_n(x_1,\ldots,x_n)\, t^{-1}\Big(\prod_{j=1}^n p_t(x_j) - \prod_{j=1}^n p(x_j)\Big)\, d\otimes\mu(x_j) \qquad (1.7) \\
&= \int T_n(x_1,\ldots,x_n)\, \Big(\sum_{j=1}^n g(x_j)\Big)\, \prod_{j=1}^n p(x_j)\, d\otimes\mu(x_j)\ .
\end{aligned}$$

Hence

$$E_P\, T_n\, \Big(\sum_{j=1}^n g(X_j)\Big) = \int g(x)\, \dot{\kappa}(x,P)\, dP(x). \tag{1.8}$$

Repeating the argument with $T_n = 1$ we also see

$$(1.9) \qquad E_p\Big(\sum_{j=1}^{n} g(X_j)\Big) = 0.$$

By (1.8)-(1.9)

$$(1.10) \qquad \mathrm{Cov}_p\Big(T_n, \sum_{j=1}^{n} g(X_j)\Big) = \int g(x)\, \dot{\kappa}(x,P)\, dP(x).$$

Applying the Cauchy-Schwarz inequality we obtain

$$(1.11) \qquad \sigma_p^2(T_n) \geq [n \int g^2(x)\, dP(x)]^{-1} \Big[\int g(x)\, \dot{\kappa}(x,P)\, dP(x)\Big]^2.$$

Relation (1.11) has been derived for an arbitrary $g \in T(P)$. In fact, using linearity of the covariance operator, we can derive it for any g in the linear span of $T(P)$.

To find the supremum of the right hand side of (1.11) when g varies over lin $T(P)$, we decompose $\dot{\kappa}(\cdot,P)$ into the sum of its orthogonal projection (in $L_2(P)$) onto the closure of lin $T(P)$ and a remainder, which is orthogonal to this space,

$$\dot{\kappa}(\cdot,P) = \tilde{\kappa}(\cdot,P) + (\dot{\kappa}(\cdot,P) - \tilde{\kappa}(\cdot,P)),$$

where

$$\int (\dot{\kappa}(x,P) - \tilde{\kappa}(x,P))\, g(x)\, dP(x) = 0 \qquad \text{all } g \in T(P).$$

Then by the Cauchy-Schwarz inequality

$$(1.12) \qquad \frac{\left(\int g(x)\, \dot{\kappa}(x,P)\, dP(x)\right)^2}{\int g^2(x)\, dP(x)} = \frac{\left(\int g(x)\, \tilde{\kappa}(x,P)\, dP(x)\right)^2}{\int g^2(x)\, dP(x)} \leq \int \tilde{\kappa}^2(x,P)\, dP(x).$$

It follows that the best bound in (1.11) is obtained by inserting $g = \tilde{\kappa}(\cdot,P)$, yielding

$$(1.13) \qquad \sigma_p^2(T_n) \geq n^{-1} \int \tilde{\kappa}^2(x,P)\, dP(x).$$

We can prove the validity of (1.13) under a weak regularity condition on T_n.

PROPOSITION 1.1. *For every* $g \in T(P)$ *let* κ *satisfy* (1.5)-(1.6) *for a sequence* $\{P_t\} \subset \mathcal{P}$ *satisfying* (1.2). *Let* T_n *be an unbiased estimator for* κ *satisfying*

$$\sup_{Q \in B(P,\varepsilon)} E_Q T_n^2 < \infty, \tag{1.14}$$

for some neighbourhood $B(P,\varepsilon) = \{Q \in \mathcal{P}: \int |q-p|\, d\mu < \varepsilon\}$ *of* P. *Then* (1.13) *holds true.* □

PROOF. It suffices to make the derivation of (1.7) rigorous. By a standard argument one sees

$$\int [t^{-1}(\prod_{j=1}^n p_t^{\frac{1}{2}}(x_j) - \prod_{j=1}^n p^{\frac{1}{2}}(x_j)) - \tfrac{1}{2}(\sum_{j=1}^n g(x_j)) \prod_{j=1}^n p^{\frac{1}{2}}(x_j)]^2 \, d\otimes\mu(x_j) \to 0 .$$

Next under (1.14), (1.7) follows from Lemma 5.21. ■

The reader should not be overly impressed by the above result. First it gives a bound for variance only. Secondly even for finite dimensional parametric models the bound is seldom sharp. Finally for models with large parameter spaces the set of unbiased estimators of κ will typically be empty.

However, the bias of an estimator and the discrepancy between its variance and the bound may disappear if $n \to \infty$. This is what makes the asymptotic theory of bounds, as presented in Chapter 2, possible. Proposition 1.1 does give us -at least partially- the correct idea of what to expect when deducing these results.

1.3. ESTIMATING LOCATION UNDER SYMMETRY

We illustrate the large parameter space approach in a model for the simple statistical experiment introduced in Section 1.1. Let H be a set of probability densities on $\mathbb{R}$ which are symmetric about zero. Based on an i.i.d. sample $X_1,\ldots,X_n$ from $\eta(\cdot-\theta)$, where θ is unknown and $\eta \in H$, it is required to estimate θ.

In the formulation of Section 1.2 we may choose

$\mathcal{P} = \{P_{\theta\eta}: \eta \in H, \theta \in \mathbb{R}\}$, where $P_{\theta\eta}$ is the distribution with density $\eta(\cdot-\theta)$. We set $\kappa(P_{\theta\eta}) = \theta$. To derive the lower bound of Proposition 1.1, first suppose that the shape parameter η is completely known, i.e. $H = \{\eta\}$ for a given η. Differentiable submodels can be obtained by varying θ and scores are of the form

$$\partial/\partial t \log \eta(x-\theta-at) \,|_{t=0} = -a\eta'(x-\theta)/\eta(x-\theta) \qquad (a \in \mathbb{R}).$$

Let

$$I(\eta) = \int [\eta'(x)/\eta(x)]^2 \, \eta(x) \, dx \,,$$

which is asummed to be finite. Then an influence function of κ equals (cf. (1.5)-(1.6))

$$(1.15) \qquad \tilde{\kappa}(x,P_{\theta\eta}) = -I^{-1}(\eta) \, \eta'(x-\theta)/\eta(x-\theta).$$

Hence by Proposition 1.1, for any unbiased estimator T_n

$$(1.16) \qquad \sigma^2_{\theta\eta}(T_n) \geq n^{-1}I^{-1}(\eta).$$

If η is known, we may estimate θ by the maximum likelihood estimator T_n, maximizing

$$(1.17) \qquad \prod_{j=1}^{n} \eta(X_j-\theta)$$

with respect to θ. Typically T_n satisfies

$$(1.18) \qquad L_{\theta\eta}(\sqrt{n}(T_n-\theta)) \to N(0,I^{-1}(\eta)).$$

Moreover, using the method of LeCam (1969), one can construct modifications of the maximum likelihood estimator, that satisfy (1.18) under the minimal regularity condition that $I(\eta) < \infty$. This is usually interpreted in the sense that the bound (1.16) is asymptotically sharp. A more precise statement can be based on the theorems in Chapter 2.

In Section 1.1 we have argued that finite dimensional parametric models are often unrealistic. Let us now assume that η is symmetric about zero, but otherwise unknown, i.e. $H = \{\eta: \eta(x) = \eta(|x|)\}$. Then to obtain a

bound for the variance we may also consider submodels of the form $t \to \eta_t(x-\theta)$, where η_t is symmetric about zero for every t. These lead to scores of the form

$$\partial/\partial t \log \eta_t(x-\theta) \mid_{t=0} = \partial/\partial t \log \eta_t(|x-\theta|) \mid_{t=0} = b(|x-\theta|).$$

This rather informal derivation may be made precise in the sense of (1.2). Indeed, it can be shown that any quadratically integrable function with zero expectation of the form $b(|x-\theta|)$ can be obtained as a limit in the sense of (1.2).

For the lower bound it is of great importance that scores for θ and η are orthogonal, i.e.

$$(1.19) \qquad \int -\eta'/\eta(x-\theta)\, b(|x-\theta|)\, \eta(x-\theta)\, dx = 0.$$

Indeed (1.19) implies that $\tilde{\kappa}(\cdot,\theta,\eta)$ given by (1.15) is again the influence function of κ. It suffices to check (1.5)-(1.6) for all submodels. But

$$t^{-1}[\kappa(P_{\theta+at,\eta_t}) - \kappa(P_{\theta\eta})] = t^{-1}(\theta-at-\theta) = a$$

and

$$\int [-a\eta'/\eta(x-\theta) + b(|x-\theta|)]\, \tilde{\kappa}(x,P_{\theta\eta})\, \eta(x-\theta)\, dx = a.$$

Thus, rather surprisingly, Proposition 1.1 yields the same bound for $\sigma^2_{\theta\eta}(T_n)$ for the models with $H = \{\eta\}$ and $H = \{\eta\colon \eta(x) = \eta(|x|)\}$ respectively. This fact was noted in a paper by Stein (1956), which has lead to a series of papers which are of great importance for the theory of estimation for large parameter spaces. At first thought, it is remarkable that the statistical problem should not become more difficult when going from a problem with η known to a problem with η essentially unknown. Yet this is the case, at least in an asymptotic sense. In increasing generality van Eeden (1970), Takeuchi (1971), Beran (1974) and Stone (1975) have shown the existence of estimators T_n satisfying (1.18), which are defined independently of η. Note that the latter restriction precludes maximum likelihood estimators defined by maximizing (1.17) with respect to θ.

The keyword in these constructions is *adaptation*. Based on the observations one first obtains a suitable estimate for the unknown,

symmetric shape η (or rather its score function η'/η). Next one takes (a modification of) the maximum likelihood estimator corresponding to the estimated shape. In this sense the estimator adapts itself to the underlying shape.

In its original meaning *adaptation* refers to statistical procedures which are based on an initial estimate of (part of) the underlying distribution of the observations. Bickel (1982) uses this term in a much more restrictive manner. Given a set $\{P_{\theta\eta}: \theta \in \Theta, \eta \in H\}$ of probability measures he calls an estimator for θ adaptive (to $\eta \in H$) if its asymptotic performance is not worse than that of the best estimator for the case that η is completely known. Because of the special orthogonality property (1.19), the estimators constructed by the above-mentioned authors are also adaptive in this sense.

In general the transition from a parametric to a semi-parametric model involves a 'loss of information'. Then the efficient influence functions for the two problems are not equal and the bound which can be obtained from Proposition 1.1 is larger for the semi-parametric model. In such cases estimators which are adaptive in the sense of Bickel (1982) cannot exist. However, at least in a number of cases, the bound given by Proposition 1.1 is asymptotically attainable. In Chapter 5 we construct estimators which are efficient in terms of this bound, in a class of models which generalizes the symmetric location model considerably. Using the word once again in its general sense, these estimators are adaptive.

1.4. NOTATION

Much of the notation we use is standard and/or given below when needed. For easy reference we gather the most important notation in this section.

In some parts of the manuscript measurability plays an important role and we often make this explicit by speaking about measurable functions $f: (X,\mathcal{B}) \to (Z,\mathcal{A})$, where $(X,\mathcal{B})$ and $(Z,\mathcal{A})$ are measurable spaces. We make an exception to this rule if one or both of the measurable spaces is a Euclidean space $\mathbb{R}^k$, in which case it is tacitly understood that measurability is with respect to the Borel σ-algebra on $\mathbb{R}^k$. A $(Z,\mathcal{A})$- valued random element G is a measurable map $G: (\Omega,\mathcal{B},P) \to (Z,\mathcal{A})$, where $(\Omega,\mathcal{B},P)$ is

some probability space. With the same notation $L(G)$ is the distribution of G on $(\mathcal{Z},\mathcal{A})$. This is sometimes also denoted by G(P).

For a measure P on a measurable space $(X,\mathcal{B})$, $L_2(P)$ denotes the set of all measurable functions g: $(X,\mathcal{B}) \to \mathbb{R}$ with $\int g^2(x)\, dP(x) < \infty$. $L_{2*}(P)$ is the subset of $L_2(P)$ of functions which also satisfy $\int g(x)\, dP(x) = 0$. $\|\cdot\|_P$ denotes the norm and $\langle\cdot,\cdot\rangle_P$ the inner product in the Hilbert space $L_2(P)$. If $C \subset L_2(P)$ then lin C is the smallest linear subspace containing C.

For $-\infty \le a < b \le \infty$ C[a,b] and D[a,b] denote the set of all continuous functions, and right continuous functions with left limits f: $[a,b] \to \mathbb{R}$, respectively. $\|f\|_\infty = \sup\{|f(u)|: u \in [a,b]\}$.

A matrix A may be denoted (α_{ij}), in which case $(A)_{ij} = \alpha_{ij}$. A prime ' is used to denote the transpose of a matrix.

$N_k(\mu,\Sigma)$ denotes the normal distribution on $\mathbb{R}^k$. δ_z the distribution which is degenerate at z.

When B' is a set of real functions on a set B, $\tau(B')$ denotes the weakest topology on B making all elements of B' continuous with respect to the usual topology on $\mathbb{R}$. $\mathcal{U}(B')$ is the smallest σ-algebra making all elements of B' measurable. If τ is a topology on a set B, then $\mathcal{U}(\tau)$ denotes the Borel σ-algebra on B.

Outer measure is denoted by P^*, inner integral by $\int_*$ or E_*.

When (B,τ) is a topological space, $\mathcal{U}$ a σ-field on B and L a probability measure on $(B,\mathcal{U})$, L is called (τ-)tight if for all $\varepsilon > 0$ there exists a compact set K_ε such that $L^*(K_\varepsilon) > 1-\varepsilon$. L is called separable if $L(S) = 1$ for a separable set $S \in \mathcal{U}$.

The dual space of a topological vector space (B,τ) is denoted by B^* or B^*_τ.

In the case that a topology τ is generated by a metric d, we may use the notation d, where logically one would expect τ. For instance $\mathcal{U}(d)$ denotes the Borel σ-field of the topology generated by d; $B^*_{\|\cdot\|}$ is the dual space with respect to a norm topology.

The indicator of a set C is denoted $1\{C\}$ or 1_C. C^c is the complement of C.

CHAPTER 2

ASYMPTOTIC BOUNDS ON THE PERFORMANCE OF ESTIMATORS WITH VALUES IN $\mathbb{R}^k$

2.1. INTRODUCTION

The subject matter of this chapter has its roots in two papers by Hájek (1970,1972). In these papers Hájek considers estimation of the parameter in parametric models $\{P_\theta: \theta \in \Theta \subset \mathbb{R}^k\}$ and he presents two types of theorems: the *convolution theorem* and the *local asymptotic minimax theorem*. Both theorems give bounds on the performance of sequences of estimators $\{T_n\}$ as $n \to \infty$ and, in a way, can be seen as asymptotic versions of the well-known Cramér-Rao bound for the variance of unbiased estimators. Since the publication of Hájek's papers many generalizations have been obtained, the extensions relating to the estimation of both infinite dimensional functionals, and finite dimensional functionals in models parametrized by a possibly infinite dimensional parameter. In this chapter we consider a general set-up for the estimation of functionals with values in $\mathbb{R}^k$.

The convolution theorem shows that a certain class of *regular* estimator sequences are asymptotically distributed as the sum of a normal random variable and an independent 'noise' variable, thus leading to the interpretation that an estimator sequence is best within this class if it has a certain normal limiting distribution. The local asymptotic minimax (LAM) theorem is not restricted to some class of estimators and gives a lower bound for the limit of the maximum risk over a shrinking neighbourhood of the true underlying distribution. We think most people like the convolution theorem better, as it is concerned with limiting

distributions and is easier to state (and prove). This theorem does however apply only to a subclass of the class of all estimator sequences. Our first aim in this chapter will be to weaken the regularity requirement, but nevertheless obtain a *generalized convolution theorem* (Section 2.2).

Next a convolution and a local asymptotic minimax (LAM) theorem are presented in Sections 2.3 and 2.4, both extending known results to the situation where the *tangent cone* is convex and not necessarily a linear space. This weakening will be useful when discussing estimation in mixture models in Chapter 5. The basic result obtained for convex tangent cones is that the lower bound for the LAM risk equals the bound that we get when we calculate as if the tangent cone were equal to the linear space spanned by itself. This implies that in general the minimax risk is larger than the lower bound that one gets by considering the most difficult one-dimensional submodel.

The three types of theorems have in common that they give bounds on many aspects of the asymptotic performance of a sequence of estimators. In Section 2.5 we have a more modest aim and give a theorem on the variance of the limiting distribution only, of course under weaker conditions than before. Of all the theorems this one is closest to being the asymptotic Cramér-Rao bound.

Basing ourselves on the results obtained in Sections 2.2-2.5, we discuss asymptotic optimality of estimators in Section 2.6, starting from a theorem which is closely connected to Hájek's (1972) characterization of LAM estimators as asymptotically linear estimators. A few simple examples are included here for illustration.

All results in Sections 2.2-2.6 are stated for models with i.i.d. observations. This is probably the most interesting case. Moreover, asymptotic lower bounds can be cast in an attractive form through use of terminology due to Koshevnik and Levit (1976) and Pfanzagl (1982), in particular the notion of a tangent cone and functionals which are differentiable with respect to it. However, as is well-known, the methods developed for i.i.d models have a much wider applicability. In Section 2.7 more general theorems are presented, which are based on *local asymptotic normality* and may be applied to models with non-identically distributed and even dependent observations.

In the remainder of this introduction we restrict ourselves to the set-up with i.i.d. observations.

2.1.1. i.i.d. models

Let $\mathcal{P}$ be a set of probability measures on a measurable space $(X,\mathcal{B})$ and let $\kappa\colon \mathcal{P} \to \mathbb{R}^k$ be a functional. Given an i.i.d. sample $X_1,X_2,\ldots,X_n$ from an unknown $P \in \mathcal{P}$ it is required to estimate $\kappa(P)$, which is done by an estimator $T_n = t_n(X_1,X_2,\ldots X_n)$. Here $t_n\colon (X^n,\mathcal{B}^n) \to \mathbb{R}^k$ is a measurable map.

Intuitively, as n grows large one can determine the unknown underlying distribution P (almost) exactly. The asymptotic difficulty of estimating $\kappa(P)$ is therefore determined by the local structure of $\mathcal{P}$ near P and asymptotic bounds on the performance of $\{T_n\}$ at $P \in \mathcal{P}$ are based on a linear approximation of $\mathcal{P}$ at P. This notion is made precise in the following definition. Recall that a cone C in a vector space over the reals is a subset which is closed under multiplication by nonnegative scalars: if $g \in C$, then $ag \in C$ for all $a \geq 0$.

DEFINITION 2.1. *A cone* T(P) *in* $L_2(P)$ *is called a tangent cone at* $P \in \mathcal{P}$ *if for all* $g \in T(P)$ *there exists* $\{P_t\} \subset \mathcal{P}$ *with*

$$\int [t^{-1}((dP_t)^{\frac{1}{2}}-(dP)^{\frac{1}{2}}) - \tfrac{1}{2}\, g\, (dP)^{\frac{1}{2}}]^2 \to 0 \qquad \text{as } t\downarrow 0. \quad \square \tag{2.1}$$

Formula (2.1) should be interpreted as follows. Let P_t and P have densities p_{tt} and p_t respectively with respect to an arbitrary σ-finite measure μ_t dominating P_t+P. Then we have

$$\int [t^{-1}(p_{tt}^{\frac{1}{2}} - p_t^{\frac{1}{2}}) - \tfrac{1}{2}\, g\, p_t^{\frac{1}{2}}]^2 \, d\mu_t \to 0.$$

It can be checked that this definition is independent of the choice of the dominating measure μ_t (cf. the argument in the proof of Proposition A.12). Thus, when the class $\mathcal{P}$ is dominated by a single σ-finite measure μ, then any $g \in T(P)$ is a limit in the sense that

$$\int [t^{-1}(p_t^{\frac{1}{2}} - p^{\frac{1}{2}}) - \tfrac{1}{2}\, g\, p^{\frac{1}{2}}]^2 \, d\mu \to 0 \,. \tag{2.2}$$

It is useful to note that (2.1) may in fact always be replaced by (2.2), in the sense that for any $g \in T(P)$ there exists a sequence $\{P_t\} \subset \mathcal{P}$ and a σ-finite measure μ dominating $\{P_t\}$ and P such that (2.2) holds. Indeed, we can without loss of generality replace the continuous paths $\{P_t\}$ in Definition 2.1 by sequences $\{P_{tn}\}$ $(n=1,2,\ldots)$ and set μ equal to a convex

linear combination of P and the P_{tn}.

Relation (2.2) can be split in the following two assertions

(2.2.a) $\quad P_t(\{x: p(x) = 0\}) = o(t^{-2})$

(2.2.b) $\quad \int [t^{-1}(p_t^{\frac{1}{2}} - p^{\frac{1}{2}}) - \alpha]^2 \, d\mu \to 0$,

for some $\alpha \in L_2(\mu)$. In the literature (2.2.a) is sometimes omitted, but cannot be dispensed with in general. (The situation differs if one considers *two-sided* paths p_t, $t \in (-1,1)$).

It is easy to see that there exists a *maximal* tangent cone $T_m(P)$, consisting of *all* $g \in L_2(P)$ satisfying (2.2) for some $\{P_t\} \subset \mathcal{P}$. When establishing lower bounds the maximal cone gives the best results. However in applications one frequently encounters the situation that one knows that a certain cone is *a* tangent cone, but is unable (or unwilling) to prove that it equals the maximal cone $T_m(P)$. Moreover functionals κ may be differentiable as defined below with respect to T(P), but not relative to $T_m(P)$. Definition 2.1 is meant to accomodate this situation.

A tangent cone is therefore not uniquely determined according to our definition. However, we assume throughout that a given tangent cone T(P) is fixed from the beginning. All definitions and theorems relate to this T(P), even when this is not explicit in the notation or terminology.

In the sequel we use repeatedly that $T(P) \subset L_{2*}(P)$, i.e. any $g \in T(P)$ satisfies $\int g \, dP = 0$. This is a consequence of

$$0 = t^{-1}(\int dP_t - \int dP) = \int (p_t^{\frac{1}{2}}+p^{\frac{1}{2}}) \, t^{-1}(p_t^{\frac{1}{2}} - p^{\frac{1}{2}}) \, d\mu \to \int g \, dP.$$

Definition 2.1 corresponds to a definition of differentiable functionals.

DEFINITION 2.2. *A functional* $\kappa: \mathcal{P} \to \mathbb{R}^k$ *is called* differentiable *at* $P \in \mathcal{P}$ *relative to* T(P), *if there exist an element* $\dot{\kappa}(\cdot,P) \in L_2(P)^k$ *such that*

(2.3) $\quad t^{-1}(\kappa(P_t)-\kappa(P)) \to \int \dot{\kappa}(x,P) \, g(x) \, dP(x)$ *in* $\mathbb{R}^k$,

for every $g \in T(P)$ *and some sequence* $\{P_t\} \subset \mathcal{P}$ *satisfying* (2.1). □

The element $\dot{\kappa}(\cdot,P)$ is called a *gradient* or *influence function* of κ. Note that (2.3) only specifies inner products of the gradient with elements of T(P). Consequently a gradient is not uniquely defined. However, the vector of orthogonal projections of its components onto the closure of the subspace lin T(P) of $L_2(P)$ *is* well-defined. This will be denoted $\tilde{\kappa}(\cdot,P)$ and is called *canonical gradient* or *efficient influence function*.

As an example consider semi-parametric models (cf. Begun, Hall, Huang, Wellner (1983)).

EXAMPLE 2.3. Let $\mathcal{P}$ be a set of probability distributions $P_{\theta\eta}$ given by densities $p(\cdot,\theta,\eta)$ with respect to a σ-finite measure μ on $(X,\mathcal{B})$, where $\theta \in \Theta \subset \mathbb{R}^m$, open, and $\eta \in H$, arbitrary. Suppose there exist $\dot{\ell}(\cdot,\theta,\eta) \in L_2(P_{\theta\eta})^m$ $(i=1,\ldots,m)$ such that for every $h \in \mathbb{R}^m$

$$(2.4) \qquad \int [t^{-1}(p^{\frac{1}{2}}(x,\theta+th,\eta)-p^{\frac{1}{2}}(x,\theta,\eta)) - \tfrac{1}{2}h'\dot{\ell}(x,\theta,\eta)p^{\frac{1}{2}}(x,\theta,\eta)]^2 d\mu(x) \to 0,$$

as $t\to 0$. Next let $T_\eta(P_{\theta\eta})$ be a cone of $b \in L_2(P_{\theta\eta})$ for which there exists $\{\eta_t\} \subset H$ with as $t\downarrow 0$

$$(2.5) \qquad \int [t^{-1}(p^{\frac{1}{2}}(x,\theta,\eta_t)-p^{\frac{1}{2}}(x,\theta,\eta)) - \tfrac{1}{2}b(x)p^{\frac{1}{2}}(x,\theta,\eta)]^2 \, d\mu(x) \to 0.$$

Let $\kappa(P_{\theta\eta}) = \psi(\theta)$, where $\psi\colon \Theta \to \mathbb{R}^k$ is differentiable in the ordinary sense with derivative $\psi'(\theta)\colon \mathbb{R}^m \to \mathbb{R}^k$ and set

$$(2.6) \qquad \tilde{\ell}(\cdot,\theta,\eta) = \dot{\ell}(\cdot,\theta,\eta) - b_0(\cdot,\theta,\eta),$$

where $b_0(\cdot,\theta,\eta)$ is the vector of $L_2(P_{\theta\eta})$ projections of the components of $\dot{\ell}(\cdot,\theta,\eta)$ onto the closure of lin $T_\eta(P_{\theta\eta})$. The functions $\tilde{\ell}_i(\cdot,\theta,\eta)$ are called the *efficient scores* for θ. Finally define an $(m\times m)$ matrix by

$$\tilde{I}(\theta,\eta) = E_{\theta\eta}\tilde{\ell}(X_1,\theta,\eta)\tilde{\ell}(X_1,\theta,\eta)'$$

and assume that $\tilde{I}(\theta,\eta)$ is nonsingular. If the tangent cone is given by $\{h'\dot{\ell}(x,\theta,\eta)\colon h \in \mathbb{R}^m\} \cup T_\eta(P_{\theta\eta})$, then $\kappa\colon \mathcal{P} \to \mathbb{R}^k$ is differentiable with

$$(2.7) \qquad \tilde{\kappa}(\cdot,P_{\theta\eta}) = \psi'(\theta)\,\tilde{I}^{-1}(\theta,\eta)\,\tilde{\ell}(\cdot,\theta,\eta).$$

As we shall see, to obtain lower bounds we usually need tangent cones that are *convex.* A better candidate for the tangent cone is

$$(2.8) \qquad T(P_{\theta\eta}) = \{\, h'\dot{\ell}(\cdot,\theta,\eta) + b(\cdot)\colon\ h \in \mathbb{R}^m,\ b \in T_\eta(P_{\theta\eta}) \,\}\ .$$

To ensure that this is a tangent cone and that κ is differentiable with respect to it, we impose the additional condition that for every $b \in T_\eta(P_{\theta\eta})$ and $h \in \mathbb{R}^m$ there exists $\{\eta_t\} \subset H$ satisfying

$$\int [t^{-1}(p^{\frac{1}{2}}(x,\theta+th,\eta_t)-p^{\frac{1}{2}}(x,\theta,\eta)) - \tfrac{1}{2}(h'\dot{\ell}(x,\theta,\eta)+b(x))\ p^{\frac{1}{2}}(x,\theta,\eta)]^2\ d\mu(x) \to 0\ .$$

A version of joint differentiability of $p(\cdot,\theta,\eta)$ in (θ,η) such as in Begun et al.(1983) will ensure this. $\kappa\colon P \to \mathbb{R}^k$ is differentiable with respect to this choice of a tangent cone and canonical gradients are still given by (2.7). □

We close this introduction by stating two preparatory lemmas, for easy reference.

LEMMA 2.4. *Let* $(H,\langle\cdot,\cdot\rangle)$ *be a Hilbert space and let* $b_1,\ldots,b_m$ *be linearly independent elements of* H. *Define the* $(m\times m)$ *matrix* B *by* $(B)_{ij} = \langle b_i,b_j\rangle$. *Then for* $a \in H$ *the orthogonal projection of* a *on* $\operatorname{lin}\{b_1,\ldots,b_m\}$ *is given by* $\Pi a = \Sigma_{i=1}^m \alpha_i b_i$, *where*

$$\alpha = B^{-1}(\langle a,b_1\rangle,\ldots,\langle a,b_m\rangle)'.$$

Hence

$$\langle \Pi a_1, \Pi a_2\rangle = (\langle a_1,b_1\rangle,\ldots,\langle a_1,b_m\rangle)B^{-1}(\langle a_2,b_1\rangle,\ldots,\langle a_2,b_m\rangle)'.\ \square\ \blacksquare$$

We shall apply this lemma with $H = L_2(P)$. The next lemma is a combination of the *local asymptotic normality lemma* and a version of *Le Cam 's third lemma.*

Asymptotic bounds for $\mathbb{R}^k$

LEMMA 2.5. *Let* $X_1,\ldots,X_n$ *be i.i.d. with* $L(X_1)$ *equal to* P, *let* $\{P_n\}$ *satisfy* (2.11) *(see below) and define* P*-a.s. (letting* log 0 = *arbitrary)*

$$\Lambda_n(P_n,P) = \log \prod_{i=1}^{n} \frac{p_n}{p}(X_i),$$

where p_n *and* p *are densities of* P_n *and* P *with respect to a* σ*-finite dominating measure* μ. *Then*

$$\Lambda_n(P_n,P) - n^{-\frac{1}{2}} \sum_{i=1}^{n} g(X_i) + \tfrac{1}{2}E_P g^2(X_1) \to 0.$$

Moreover assume that $\{S_n\}$, $S_n = S_n(X_1,\ldots,X_n)$, *is a sequence of random elements with respect to the Borel* σ*-algebra* $\mathcal{A}$ *in a separable metric space* $(\mathcal{Y},d)$, *such that*

$$(2.9) \qquad L_P((S_n,\Lambda_n(P_n,P))) \to L$$

for some probability measure L *on* $(\mathcal{Y}\times\mathbb{R},\mathcal{A}\times\mathcal{B})$. *Then*

$$L_{P_n}(S_n) \to L'$$

where L′ *is the probability measure on* $(\mathcal{Y},\mathcal{A})$ *given by*

$$L'(A) = \int_{A\times\mathbb{R}} e^{\lambda}\, dL(y,\lambda) \qquad , \qquad (A \in \mathcal{A}) \ . \ \square$$

PROOF. This is a combination of the Propositions in Appendix A.1. ∎

We shall apply the second part of the lemma with $\mathcal{Y}$ equal to $\bar{\mathbb{R}}^k$, where $\bar{\mathbb{R}} = [-\infty,\infty]$ is the usual compactification of $\mathbb{R}$.

2.2. A GENERALIZED CONVOLUTION THEOREM

Suppose that in the set-up introduced in 2.1.1 one considers sequences of estimators $\{T_n\}$ which converge in distribution to a limit distribution L = L(P) in the following sense

(2.10) $\quad L_P(\sqrt{n}(T_n-\kappa(P))) \to L, \qquad$ as $n \to \infty$.

What is the best limit distribution L (depending on P) that is attainable by an estimator sequence?

Unfortunately, even if we require (2.10) to hold for all $P \in \mathcal{P}$, the question of asymptotic efficiency phrased in this manner does not have a useful answer. Examples have been exhibited with estimators that have L(P) equal to a distribution degenerated at 0 for some $P \in \mathcal{P}$ and to well-behaved limit distributions for other P. On the other hand these *superefficient estimators* turn out not to be particularly good estimators, as they have rather irregular behaviour. A less naive approach to asymptotic efficiency is therefore needed.

The line we take is to look at *averages* of limiting distributions. Because taking averages obviously destroys information, they will be taken over small, and even shrinking neighbourhoods of a distribution $P \in \mathcal{P}$. We describe what this means.

In this section we restrict ourselves to sequences of estimators $T_n = t_n(X_1,X_2,\ldots X_n)$ for which for every $g \in T(P)$ there exists a probability distribution L_g on $\mathbb{R}^k$ and a sequence $\{P_n\} \subset \mathcal{P}$ such that

(2.11) $\quad \int [\sqrt{n}((dP_n)^{\frac{1}{2}}-(dP)^{\frac{1}{2}}) - \tfrac{1}{2}g(dP)^{\frac{1}{2}}]^2 \to 0$

(2.12) $\quad \sqrt{n}(\kappa(P_n)-\kappa(P)) \to \int \dot{\kappa}(x,P)\, g(x)\, dP(x)$

and

(2.13) $\quad L_{P_n}(\sqrt{n}(T_n-\kappa(P_n))) \to L_g$.

We call such estimator sequences $\{T_n\}$ *weakly regular at* $P \in \mathcal{P}$, reserving the name *regular* for the subclass of weakly regular sequences that have L_g equal to the same probability distribution for every $g \in T(P)$, as usual. One does not loose much by restricting oneself to weakly regular estimator sequences. Indeed, the requirement of weak regularity is only slightly stronger than tightness of $\{ L_P(\sqrt{n}(T_n-\kappa(P))) \}$.

The essential hypotheses are (2.11) and (2.13). The technical assumption (2.12) is necessary, because in Definition 2.2 we have required κ to be differentiable only along *some* sequence $\{P_t\}$ satisfying (2.1).

Often (2.12) is satisfied along every sequence $\{P_n\}$ satisfying (2.11). Note on the other hand, that we should not expect (2.13) to hold if (2.12) fails.

The asymptotic behaviour of a weakly regular estimator sequence $\{T_n\}$ near $P \in \mathcal{P}$ is characterized by the set of limiting distributions $\{L_g: g \in T(P)\}$. It will now be proved that an average of the limiting distributions is the convolution of a certain normal distribution and another distribution. For simplicity we restrict ourselves to averages over finite dimensional subsets of T(P).

THEOREM 2.6. *Let* T(P) *be a linear space, let* $\{g_1, g_2, \ldots, g_m\} \subset T(P)$ *be linearly independent and set* $(\Sigma)_{ij} = E_P g_i(X_1) g_j(X_1)$ *for the* $(m \times m)$ *matrix* Σ. *Let* $\kappa: \mathcal{P} \to \mathbb{R}^k$ *be differentiable at* $P \in \mathcal{P}$ *relative to* T(P) *and let the* $(k \times m)$ *matrix* $D\kappa$ *be defined by*

$$(2.14) \qquad (D\kappa)_{ij} = \langle \dot{\kappa}_i(\cdot, P), g_j \rangle_P \ .$$

Then for any at $P \in \mathcal{P}$ *weakly regular estimator sequence, any* $\tau \in \mathbb{R}^m$ *and any positive definite* $(m \times m)$ *matrix* Λ, *there exists a probability distribution* M *on* $\mathbb{R}^k$ *with*

$$(2.15) \qquad \int_{\mathbb{R}^m} L_{\Sigma \alpha_i g_i} \, dN(\tau, \Lambda)(\alpha) = N(0, D\kappa(\Sigma + \Lambda^{-1})^{-1} D\kappa') * M. \ \square$$

The interpretation of Theorem 2.6 is that the (average) limiting distribution of a weakly regular estimator sequence is more spread out than a $N(0, D\kappa(\Sigma+\Lambda^{-1})^{-1}D\kappa')$ distribution. This intuitive interpretation can be made quantitative by means of Anderson's Lemma (Anderson (1955); see Pfanzagl (1985) for a version where the covariance matrix of the normal distribution is allowed to be singular). Indeed (2.15) implies that for any convex set $C \subset \mathbb{R}^k$, which is symmetric about the origin

$$\int_{\mathbb{R}^m} L_{\Sigma \alpha_i g_i}(C^c) \, dN(\tau, \Lambda)(\alpha) \geq N(0, D\kappa(\Sigma + \Lambda^{-1})^{-1} D\kappa')(C^c).$$

To study the role of Λ in this expression, note that

$$D\kappa \Sigma^{-1} D\kappa' - D\kappa(\Sigma + \Lambda^{-1})^{-1} D\kappa'$$

$(= D\kappa(\Sigma+\Sigma\Lambda\Sigma)^{-1}D\kappa')$ is nonnegative definite for all Λ and converges to zero if e.g. $\Lambda = \lambda I$ and $\lambda \to \infty$. Hence for any C as above

$$(2.16) \qquad \sup_{\alpha\in\mathbb{R}^m} L_{\Sigma\alpha_i g_i}(C^c) \geq N(0, D\kappa\Sigma^{-1}D\kappa')(C^c).$$

Next by Lemma 2.4 the covariance matrix in the right hand side is the covariance matrix of the projection of the gradient of κ on $\text{lin}\,\{g_1, g_2, \ldots, g_m\}$. Thus we can choose $\{g_1, g_2, \ldots, g_m\}$ such that the difference

$$(2.17) \qquad E_P\tilde{\kappa}(X_1,P)\tilde{\kappa}(X_1,P)' - D\kappa\Sigma^{-1}D\kappa'$$

is arbitrarily small. If the linear space spanned by the components of $\tilde{\kappa}(\cdot,P)$ is contained in T(P) we may even choose $\{g_1, g_2, \ldots, g_m\}$ equal to these components and have equality of the two matrices in (2.17).

Theorem 2.6 can be strengthened to the following theorem, where we remove the requirements of weak regularity of $\{T_n\}$ and of linearity of T(P). Let $\overset{v}{\to}$ denote vague convergence.

THEOREM 2.7. *Let* $\{g_1, g_2, \ldots, g_m\} \subset T(P)$ *be linearly independent and let* $\kappa: \mathcal{P} \to \mathbb{R}^k$ *be differentiable at* $P \in \mathcal{P}$ *relative to* T(P). *Let* $\{T_n\}$ *be an arbitrary sequence of estimators. Then for every subsequence of* $\{n\}$ *there exists a further subsequence* $\{n'\}$ *such that*

$$(2.18) \qquad L_{P_{n'}}\left(\sqrt{n'}(T_{n'}-\kappa(P)) - \langle\dot{\kappa}(\cdot,P), g\rangle_P\right) \overset{v}{\to} L_g .$$

for every $\{P_n\}$ *satisfying* (2.11) *and every* $g \in \text{lin}\,\{g_1, \ldots, g_m\}$. *Here* L_g *is a possibly defective probability distribution on* $\mathbb{R}^k$. *Moreover for any set of limiting distributions thus obtained, any* $\tau \in \mathbb{R}^m$ *and any positive definite* $(m\times m)$ *matrix* Λ, *there exists a (possibly defective) probability distribution* M *on* $\mathbb{R}^k$ *such that* (2.15) *holds.* □

To understand the second assertion of Theorem 2.7, note that for every $g \in \text{lin}\,T(P)$, one can always find a sequence $\{P_n\}$ of probability measures, (though not necessarily in $\mathcal{P}$) which satisfies (2.11). Thus Theorem 2.7 generates a distribution L_g for every $g \in \text{lin}\,T(P)$. Here the L_g for which

$g \in T(P)$ are the ones of interest. Indeed, if κ is differentiable at $P \in \mathcal{P}$ and $g \in T(P)$, then we can choose $\{P_n\}$ such that it satisfies (2.11) and (2.12), and is contained in $\mathcal{P}$. Then (2.18) implies

$$L_{P_{n'}}\left(\sqrt{n'}(T_{n'}-\kappa(P_{n'})) \right) \xrightarrow{v} L_g .$$

The last assertion of Theorem 2.7 is that an average over all L_g thus obtained, is a convolution as in (2.15). If $T(P)$ does not contain $\operatorname{lin}\{g_1,g_2,\ldots,g_m\}$, this average will always contain L_g's in which we are not interested. However under a convexity assumption on $T(P)$ it is possible to find τ and Λ such that the normal distribution $N(\tau,\Lambda)$ in (2.15) gives arbitrarily small mass to the set of L_g's for which $g \notin T(P)$. We use this to obtain a strengthened LAM theorem in Section 2.3.

PROOF OF THEOREM 2.7. Consider T_n as measurable maps in $\bar{\mathbb{R}}^k$. By Prohorov's Theorem, for any subsequence of $\{n\}$ there exists a further subsequence, abusing notation denoted $\{n\}$, such that

$$(2.19) \qquad \bar{L}_P(\sqrt{n}(T_n-\kappa(P)),\ n^{-\frac{1}{2}}\textstyle\sum_{i=1}^{n} g_1(X_i),\ldots,\ n^{-\frac{1}{2}}\sum_{i=1}^{n} g_m(X_i)) \to \bar{L}(S,V) ,$$

weakly as laws on $\bar{\mathbb{R}}^k\times\mathbb{R}^m$. For $\alpha \in \mathbb{R}^m$ let $\{P_{n\alpha}\}$ satisfy

$$(2.20) \qquad \int [\sqrt{n}((dP_{n\alpha})^{\frac{1}{2}}-(dP)^{\frac{1}{2}}) - \tfrac{1}{2}\textstyle\sum_{i=1}^{m}\alpha_i g_i (dP)^{\frac{1}{2}}]^2 \to 0$$

By (2.19) and the first part of Lemma 2.5, we see that for any $\alpha \in \mathbb{R}^m$

$$(2.21) \qquad \bar{L}_P(\sqrt{n}(T_n-\kappa(P)),\ \Lambda_n(P_{n\alpha},P)) \to \bar{L}(S,\ \alpha'V-\tfrac{1}{2}\alpha'\Sigma\alpha) ,$$

weakly as laws on $\bar{\mathbb{R}}^k\times\mathbb{R}$. Define the matrix $D\kappa$ as in the statement of Theorem 2.6. By (2.21) and the second part of Lemma 2.9 there exist distributions $\bar{L}_{\Sigma\alpha_i g_i}$ on $\bar{\mathbb{R}}^k$ such that

$$(2.22) \qquad \bar{L}_{P_{n\alpha}}\left(\sqrt{n}\ (T_n - \kappa(P)) - D\kappa\alpha\ \right) \to \bar{L}_{\Sigma\alpha_i g_i} ,$$

weakly as laws on $\bar{\mathbb{R}}^k$. But this implies vague convergence of the

corresponding laws on $\mathbb{R}^k$. The first assertion of the theorem follows.

Here for every Borel set $B \subset \mathbb{R}^k$,

$$L_{\Sigma\alpha_i g_i}(B) = \bar{L}_{\Sigma\alpha_i g_i}(B) = \int_{B\times\mathbb{R}} e^{\lambda}\, d\bar{L}(S-D\kappa\alpha,\ \alpha'V-\tfrac{1}{2}\alpha'\Sigma\alpha)(y,\lambda)$$

(2.23)

$$= E1_B(S-D\kappa\alpha)\ e^{\alpha'V\ -\ \frac{1}{2}\alpha'\Sigma\alpha}\ .$$

Next we use

$$e^{\alpha'v-\frac{1}{2}\alpha'\Sigma\alpha}\ dN(\tau,\Lambda)(\alpha) = c(v)\ dN(e(v),(\Sigma+\Lambda^{-1})^{-1})(\alpha), \tag{2.24}$$

where

$$e(v) = (\Sigma+\Lambda^{-1})^{-1}(\Lambda^{-1}\tau+v)$$

$$c(v) = \det(\Sigma+\Lambda^{-1})^{-\frac{1}{2}}\det\Lambda^{-\frac{1}{2}}\ e^{-\frac{1}{2}[\tau'\Lambda^{-1}\tau-(\Lambda^{-1}\tau+v)'e(v)]}\ .$$

By (2.23)-(2.24)

$$\int \bar{L}_{\Sigma\alpha_i g_i}(B)\ dN(\tau,\Lambda)(\alpha)$$

$$= \int E1_B(S\ -\ D\kappa e(V)\ -\ D\kappa\alpha)\ c(V)\ dN(0,(\Sigma+\Lambda^{-1})^{-1})(\alpha)\ .$$

But the last expression is the convolution of the required normal distribution and the possibly defective law given by

$$M(B) = E1_B(S\ -\ D\kappa e(V))\ c(V)\ . \ \blacksquare$$

2.3. LAM THEOREMS

As a corollary of Theorem 2.7 we obtain LAM theorems in this section, under the assumption that the tangent cone is convex.

We obtain these for *bowl-shaped* loss functions. These are functions ℓ: $\mathbb{R}^k \to \mathbb{R}$ satisfying

$$
\begin{aligned}
&\ell(0) = 0 \\
&\ell(x) = \ell(-x) \qquad (2.25) \\
&\{x: \ell(x) \le c\} \text{ is convex for all } c \in \mathbb{R}.
\end{aligned}
$$

The reader who is acquainted with the theory of limiting experiments due to Le Cam (cf. Le Cam (1972,1986)), may gain some insight in the role played by convexity by considering the following simple experiment. One observes a single observation Y from a $N_m(\alpha,I)$ distribution, where α is known to belong to a set $C \subset \mathbb{R}^m$. Then for a known $(k\times m)$ matrix $D\kappa$ the minimax risk for the estimation of $D\kappa\alpha$ is defined as

$$R_C = \inf_t \sup_{\alpha\in C} E_\alpha \ell(t(Y)-D\kappa\alpha),$$

where t ranges over the set of measurable maps from $\mathbb{R}^m$ to $\mathbb{R}^k$.

It is well-known how to evaluate R_C in the case that $C = \mathbb{R}^m$. Indeed, R_C must be larger than the Bayes risk corresponding to any prior on C, in particular Bayes risks of the form

$$\inf_t \int E_\alpha \ell(t(Y)-D\kappa\alpha)\, dN(\tau,\Lambda)(\alpha) .$$

But the infimum is attained by the Bayes estimator

$$t(y) = D\kappa\, (I+\Lambda^{-1})^{-1}\, (\Lambda^{-1}\tau+y).$$

(Use (2.24) and Anderson's Lemma (cf. Ibragimov and Has'minskii (1981))). Hence

$$
\begin{aligned}
R_C &\ge \int E_\alpha \ell(D\kappa[(I+\Lambda^{-1})^{-1}(\Lambda^{-1}\tau+Y)-\alpha])\, dN(\tau,\Lambda)(\alpha) \\
&= \int \ell(D\kappa y)\, dN(0,(I+\Lambda^{-1})^{-1})(y) .
\end{aligned}
$$

Setting $\Lambda = \lambda I$ and letting $\lambda \to \infty$ we get

$$R_C \ge \int \ell(D\kappa y)\, dN(0,I)(y) .$$

Since $E_\alpha \ell(D\kappa Y - D\kappa\alpha)$ equals the right hand side of this expression for any

α, we must have equality.

If C is reduced to a proper subset of $\mathbb{R}^m$, then estimating $D\kappa\alpha$ should be easier. However if C contains a convex, m-dimensional cone, there is no corresponding decrease in minimax risk. One can see this by reworking the above argument with priors $N(\tau,\lambda I)$ that give probability one to C in the limit, when $\tau \to \infty$ within C, and next $\lambda \to \infty$.

When there is no m-dimensional convex subcone in C this argument breaks down. It may be difficult to calculate the minimax risk for these cases, the more so because we expect the form of the risk to depend, not only on the structure of C, but also on the loss function ℓ.

We now return to the asymptotic problem. Our first result is restricted to weakly regular estimator sequences.

THEOREM 2.8. *Let* T(P) *be convex and let* $\kappa: \mathcal{P} \to \mathbb{R}^k$ *be differentiable at* $P \in \mathcal{P}$. *Then for any sequence of estimators which is weakly regular at* P *and bowl-shaped loss function* ℓ

$$\sup_{g\in T(P)} \int \ell(x)\, dL_g(x) \geq \int \ell(x)\, dN(0, E_p\tilde{\kappa}(X_1,P)\tilde{\kappa}(X_1,P)')(x). \quad \square \tag{2.26}$$

An advantage of Theorem 2.8 over the usual LAM theorem is that it is much easier to work with, as it concerns limiting distributions rather than limits of risk. Instead of the limit of the local maximum risk it considers the maximum risk over the set of local limit distributions L_g.

We omit the proof (cf. the discussion following Theorem 2.6 and the proof of Theorem 2.9 below).

Consider the special case of estimating a real-valued functional. Then, in general, Theorem 2.8 gives a larger lower bound than that obtained by applying Hájek's (1972) theorem to one-dimensional submodels. Indeed, define a one-dimensional submodel as a map $t \to P_t$ where $\{P_t\} \subset \mathcal{P}$ satisfies (2.1), (2.3) for some $g \in T(P)$. For each fixed $g \in T(P)$ the tangent cone $T_g(P) = \{hg: h \geq 0\}$ is certainly convex. For the risk over $T_g(P)$ Theorem 2.8 yields a lower bound $[E_p g(X_1)\tilde{\kappa}(X_1,P)]^2/E_p g^2(X_1)$. Thus the difficulty of the estimation problem as measured by the difficulty of the one-dimensional submodels equals

$$(2.27) \qquad \sup_{g \in T(P)} [E_P g(X_1)\tilde{\kappa}(X_1,P)]^2/E_P g^2(X_1).$$

A g for which the supremum is obtained would correspond to a *most difficult one dimensional submodel* (Stein (1956)) or a *least favorable direction* (Begun et al.(1983)). When T(P) is a linear space, the right hand side of (2.27) gives the best known lower bound for the minimax risk of the problem. (The least favorable direction is the gradient and the bound (2.27) reduces to the right hand side of (2.26)). However in the case that T(P) is non-linear, a better lower bound holds true in general. For instance it follows from Theorem 2.8 that a larger lower bound can be obtained when $\tilde{\kappa}(\cdot,P) \notin \overline{T(P)}$ and T(P) is convex, indeed the right hand side of (2.26). Thus an estimation problem can be more difficult than its most difficult one-dimensional subproblem. (cf. Example 2.17 below).

The classical LAM theorem can also be obtained as a corollary to Theorem 2.7. To obtain a stronger result we initially restrict ourselves to finite dimensional submodels of $\mathcal{P}$.

THEOREM 2.9. *Let* T(P) *be convex and let* $\kappa: \mathcal{P} \to \mathbb{R}^k$ *be differentiable at* $P \in \mathcal{P}$. *Let* $\{g_1,\ldots,g_m\} \subset T(P)$ *be linearly independent and for* $\alpha \in \mathbb{R}^m$ *such that* $\Sigma_{i=1}^m \alpha_i g_i \in T(P)$ *let* $\{P_{n\alpha}\} \subset \mathcal{P}$ *satisfy*

$$(2.28) \qquad \int [\sqrt{n}((dP_{n\alpha})^{\frac{1}{2}}-(dP)^{\frac{1}{2}}) - \tfrac{1}{2}\Sigma_{i=1}^m \alpha_i g_i (dP)^{\frac{1}{2}}]^2 \to 0$$

$$(2.29) \qquad \sqrt{n}(\kappa(P_{n\alpha})-\kappa(P)) \to D\kappa\alpha = \langle \Sigma_{i=1}^m \alpha_i g_i, \dot{\kappa}(\cdot,P)\rangle_P \ .$$

Then for any bowl-shaped loss function ℓ *and any estimator sequence* $\{T_n\}$

$$(2.30) \qquad \lim_{c\to\infty} \liminf_{n\to\infty} \sup_{\|\alpha\|\le c} E_{P_{n\alpha}} \ell(\sqrt{n}(T_n-\kappa(P_{n\alpha}))) \ge \int \ell(x)\, dN(0,D\kappa\Sigma^{-1}D\kappa')(x).$$

Here $D\kappa$ *and* Σ *are defined as in Theorem* 2.6. □

PROOF. Let $R(\ell)$ denote the left hand side of (2.30). For r = 1,2,... set

$$\ell_r(x) = 2^{-r} \sum_{j=1}^{r2^r} 1\{x: \ell(x) > j2^{-r}\} = 2^{-r} \sum_{j=1}^{r2^r} 1\{C_{rj}^c\}(x) \ ,$$

(say). We have $0 \le \ell_r(x) \uparrow \ell(x)$ as $r \to \infty$. It therefore suffices to prove the theorem for ℓ_r, for if the theorem is true for ℓ_r $(r = 1,2,\ldots)$ then

$$R(\ell) \ge \limsup_{r\to\infty} R(\ell_r)$$

would be larger than

$$\limsup_{r\to\infty} \int \ell_r(x)\, dN(0,D\kappa\Sigma^{-1}D\kappa')(x),$$

which equals the right hand side of (2.30). Furthermore we may assume that the C_{rj} are closed, because $R(\ell_r)$ decreases if we replace each C_{rj} by its closure, while $\int \ell_r(x)\, dN(0,D\kappa\Sigma^{-1}D\kappa')(x)$ is not affected.

In the sequel the non-compact C's cause us some trouble. We replace them by compact ones as follows. By the separating hyperplane theorem any closed, convex C can be written as $\cap_{i=1}^{\infty}\{x: |\beta_i'x| \le \gamma_i\}$. Next by a similar approximation argument as above we see that it is no loss of generality to replace the countable intersection by a finite intersection.

Thus we prove the theorem for a loss function ℓ of the form

$$\ell(x) = \Sigma_{j=1}^{r} 1\{x: B_j x \in K_j^c\} ,$$

where B_j: $\mathbb{R}^k \to \mathbb{R}^p$ is a linear map and K_j a cube $\otimes_{i=1}^{p} [-\gamma_i,\gamma_i]$ in $\mathbb{R}^p$.

For $\alpha \in \mathbb{R}^m$ such that $g = \Sigma_{i=1}^{m}\alpha_i g_i \notin T(P)$ it is easy to construct a sequence $\{P_{n\alpha}\}$ (not necessarily in $\mathcal{P}$), satisfying (2.28). For such α write $\kappa(P) + n^{-\frac{1}{2}}D\kappa\alpha$ for $\kappa(P_{n\alpha})$ (formal notation, we don't *define* $\kappa(P_{n\alpha})$). Then (2.28)-(2.29) will hold for every $\alpha \in \mathbb{R}^m$.

Any $B_j\circ T_n$ is an estimator of the differentiable functional $B_j\circ\kappa$: $\mathcal{P} \to \mathbb{R}^p$. By Theorem 2.7 any subsequence of $\{n\}$ has a further subsequence (abusing notation denoted $\{n\}$) such that

$$L_{P_{n\alpha}}\left(\sqrt{n}\,(B_j\circ T_n - B_j\circ\kappa(P_{n\alpha}))\right) \overset{v}{\to} L_{\Sigma\alpha_i g_i,j} , \qquad (j = 1,2,\ldots,r)$$

for limiting distributions $L_{g,j}$ satisfying

$$\int_{\mathbb{R}^m} L_{\Sigma\alpha_i g_i,j}\, dN(\tau,A)(\alpha) = N(0,B_jD\kappa(\Sigma+A^{-1})^{-1}(B_jD\kappa)') * M_j .$$

Thus for every $\alpha \in \mathbb{R}^m$

$$\limsup_{n\to\infty} P_{n\alpha}(B_j \circ \sqrt{n}(T_n - \kappa(P_{n\alpha})) \in K_j) \le L_{\Sigma\alpha_i g_i, j}(K_j) .$$

Let $A_c = \{\alpha \in \mathbb{R}^m: \alpha \ge 0, \|\alpha\| \le c\}$. Note that $g = \Sigma_{i=1}^m \alpha_i g_i \in T(P)$ for every $\alpha \ge 0$, by convexity of $T(P)$ and choice of $\{g_1,\dots,g_m\}$. Clearly

$$\liminf_{n\to\infty} \sup_{\alpha\in A_c} E_{P_{n\alpha}} \ell(\sqrt{n}(T_n - \kappa(P_{n\alpha})))$$

$$\ge \Sigma_{j=1}^r \{1 - \int L_{\Sigma\alpha_i g_i, j}(K_j)\, dN(\tau,\Lambda)(\alpha)\} - r\, N(\tau,\Lambda)(\mathbb{R}^m \setminus A_c) .$$

Thus

$$R(\ell) \ge \Sigma_{j=1}^r \{1 - N(0, B_j D\kappa(\Sigma+\Lambda^{-1})^{-1} D\kappa' B_j') * M_j(K_j)\} - r\, N(\tau,\Lambda)(\mathbb{R}^m \setminus A_\infty)$$

$$\ge \Sigma_{j=1}^r N(0, B_j D\kappa(\Sigma+\Lambda^{-1})^{-1} D\kappa' B_j')(K_j^c) - r\, N(\tau,\Lambda)(\mathbb{R}^m \setminus A_\infty)$$

$$= \int \ell(x)\, dN(0, D\kappa(\Sigma+\Lambda^{-1})^{-1} D\kappa')(x) - r\, N(\tau,\Lambda)(\mathbb{R}^m \setminus A_\infty) ,$$

by Anderson's Lemma. Now choose $\tau = d(1,1,\dots,1)$ and let $d \to \infty$. Then $N(\tau,\Lambda)(\mathbb{R}^m \setminus A_\infty) \to 0$. Finally set $\Lambda = \lambda I$ and let $\lambda \to \infty$. ■

We next obtain a LAM theorem which is not limited to submodels. For $n = 1,2,\dots$ and each $c > 0$ let $H_n(P,c)$ be the intersection of a Hellinger ball of radius $cn^{-\frac{1}{2}}$ with $\mathcal{P}$, i.e.

$$H_n(P,c) = \{Q \in \mathcal{P}: n \int (\,(dQ)^{\frac{1}{2}} - (dP)^{\frac{1}{2}}\,)^2 \le c^2\},$$

(The integral is interpreted as $\int (q^{\frac{1}{2}} - p^{\frac{1}{2}})^2\, d\mu$, where q and p are densities with respect to some measure μ).

THEOREM 2.10. *Let* T(P) *be convex and let* $\kappa: \mathcal{P} \to \mathbb{R}^k$ *be differentiable at* $P \in \mathcal{P}$. *Then for any bowl-shaped loss function* ℓ *and any estimator sequence* $\{T_n\}$

$$\lim_{c\to\infty} \liminf_{n\to\infty} \sup_{Q\in H_n(P,c)} E_Q \ell(\sqrt{n}(T_n - \kappa(Q))) \tag{2.32}$$
$$\geq \int \ell(x)\, dN(0, E_P \tilde{\kappa}(X_1,P)\tilde{\kappa}(X_1,P)')(x). \quad \square$$

PROOF. We can choose a finite, linearly independent subset from T(P) such that the expression in (2.17) is arbitrarily small. We next copy the proof of Theorem 2.9, with minor changes. ■

Though (2.32) is still reasonably simple to interprete, it may in some examples be difficult to construct estimator sequences for which the left hand side actually equals the right hand side. The main reason for this is that the shrinking Hellinger balls over which the risk is maximized are still rather large. Choosing for each n = 1,2,... a Q_n from $H_n(P,c)$ one may for instance end up with a sequence $\{Q_n\}$ for which $\otimes_{j=1}^n dQ_n(x_j)$ is not contiguous with respect to $\otimes_{j=1}^n dP(x_j)$ (cf. Oosterhoff and van Zwet (1979)).

A possible solution is to define the local maximum risk of $\{T_n\}$ as the supremum of the local maximum risks over (suitable) finite dimensional submodels. A lower bound for this supremum is then given by the corresponding supremum over the right hand side of (2.30). This will equal the right hand side of (2.32) if we take the supremum over all finite dimensional submodels, or if the set of finite dimensional submodels is sufficiently rich. We do not formalize the latter here. However see Theorem 3.10.

2.4. CONVOLUTION THEOREM

A sequence of estimators is called *regular at* $P \in \mathcal{P}$ if it satisfies (2.11)-(2.13) for all $g \in T(P)$ with L_g equal to the same probability distribution L for all $g \in T(P)$.

The following theorem states the convolution theorem for the case that the tangent cone is convex. In fact, using a proof by characteristic

functions and analytic continuation of functions of several complex variables, the convexity assumption can be weakened. For estimating the parameter in finite dimensional parametric models this is done in Droste and Wefelmeyer (1984) (Also see Theorem 2.21). For general statistical models the strengthening is not easy to formulate. With convexity the convolution theorem is also a corollary of Theorem 2.7.

THEOREM 2.11. *Let* $T(P)$ *be convex and let* $\kappa: \mathcal{P} \to \mathbb{R}^k$ *be differentiable at* $P \in \mathcal{P}$. *Then any limiting distribution* L *of a sequence of estimators which is regular at* $P \in \mathcal{P}$ *satisfies*

$$L = N(0,\ E_P\tilde{\kappa}(X_1,P)\tilde{\kappa}(X_1,P)') * M,$$

where M *is a probability measure on* $\mathbb{R}^k$. □

PROOF. By Theorem 2.21 below (cf. Lemma 2.5) we infer the existence of a probability distribution M_m on $\mathbb{R}^k$ with

$$L = N(0, D\kappa\Sigma^{-1}D\kappa') * M_m.$$

Here $D\kappa\Sigma^{-1}D\kappa'$ is the matrix of inner products of the orthogonal projections of $\tilde{\kappa}(\cdot,P)$ onto $\dot{P}_i$ $(i = 1,2,\ldots,k)$ (cf. Lemma 2.4). Now by the definition of $\tilde{\kappa}(\cdot.P)$ and the convexity of $T(P)$ we can choose C such that

$$E_P\tilde{\kappa}(X_1,P)\tilde{\kappa}(X_1,P)' - D\kappa\Sigma^{-1}D\kappa'$$

is arbitrarily close to zero. ■

2.5. A THEOREM ON ASYMPTOTIC VARIANCE

The well-known Crámer-Rao theorem gives a lower bound for the variance of an unbiased estimator for fixed sample size n. In this section we give the 'true' asymptotic version of this theorem. One interesting feature of this theorem is that one does not need conditions on the tangent cone. This is a consequence of linearity of the covariance operator.

Call an at $P \in \mathcal{P}$ weakly regular sequence of estimators *asymptotically*

unbiased at P if $\int x_i \, dL_g(x)$ exists and equals zero ($i = 1,2,\ldots,k$) for every $g \in T(P)$. Call it *asymptotically of constant bias* if $\int x_i \, dL_g(x)$ exists and is constant in $g \in T(P)$. A regular sequence of estimators is clearly asymptotically of constant bias, provided the expectation of the limiting distribution exists.

Remember that L_0 is L_g for $g = 0$.

THEOREM 2.12. *Let* κ: $\mathcal{P} \to \mathbb{R}^k$ *be differentiable at* $P \in \mathcal{P}$ *and suppose that* $\{T_n\}$ *is weakly regular and asymptotically of constant bias at* $P \in \mathcal{P}$. *Then, if the covariance matrix* $\Sigma(L_0)$ *of* L_0 *exists, we have that*

$$\Sigma(L_0) - E_P\tilde{\kappa}(X_1,P)\tilde{\kappa}(X_1,P)' \tag{2.33}$$

is nonnegative definite. Moreover an asymptotically unbiased estimator sequence can have equality to zero in (2.33) *only if*

$$\sqrt{n}(T_n - \kappa(P)) = n^{-\frac{1}{2}} \sum_{j=1}^{n} \tilde{\kappa}(X_j,P) + o_P(1). \quad \square$$

PROOF. The first assertion of the theorem is equivalent to

$$\beta'\Sigma(L_0)\beta - \beta'E_P\tilde{\kappa}(X_1,P)\tilde{\kappa}(X_1,P)'\beta \geq 0 \qquad \text{all } \beta \in \mathbb{R}^k.$$

This concerns the asymptotic variance of the estimator $\beta'T_n$ of the differentiable, real-valued functional $\beta'\kappa$ (with gradient $\beta'\tilde{\kappa}(\cdot,P)$). Without loss of generality we assume that $k = 1$.

We give the proof for the case that $T(P)$ is infinite dimensional; for the finite dimensional case the proof is easier. Let $\{g_1,g_2,\ldots\} \subset T(P)$ be linearly independent and such that $\tilde{\kappa}(\cdot,P)$ is contained in the closure of its linear span. Given a subsequence of $\{n\}$ there exists a further subsequence (denoted $\{n\}$) such that

$$\begin{aligned} &L_P\left(\sqrt{n}(T_n - \kappa(P)),\ n^{-\frac{1}{2}} \sum_{i=1}^{n} \tilde{\kappa}(X_i,P),\ n^{-\frac{1}{2}} \sum_{i=1}^{n} g_1(X_i),\ n^{-\frac{1}{2}} \sum_{i=1}^{n} g_2(X_i),\ldots\right) \\ &\qquad \to L(S,\tilde{V},V_1,V_2,\ldots)\ , \end{aligned} \tag{2.34}$$

in $\mathbb{R}^\infty$. Here $L(V_1,V_2,\ldots) = N_\infty(0,\Sigma)$.

Asymptotic bounds for $\mathbb{R}^k$

In view of (2.23) with $\alpha = he_j$ $(h \geq 0)$

$$\int x \, dL_{hg_j}(x) = \int_{\mathbb{R}^2} x \, e^{\lambda} \, dL(S - D\kappa_j h, \; hV_j - \tfrac{1}{2}\Sigma_{jj}h^2)(x,\lambda)$$

$$= E(S - D\kappa_j h) \; e^{hV_j - \frac{1}{2}\Sigma_{jj}h^2} \; .$$

Differentiation of this expression from the right with respect to h at h = 0 yields

$$0 = ESV_j - D\kappa_j . \tag{2.35}$$

Let Σ^m be the upper $(m \times m)$ matrix of Σ and $D\kappa^m = (D\kappa_1, \ldots, D\kappa_m)$. Set $\tilde{V}^m = D\kappa^m \, (\Sigma^m)^{-1} \, (V_1, \ldots, V_m)$. By the Cauchy-Schwarz inequality

$$\sigma^2(S) \geq \sigma^{-2}(\tilde{V}^m) \; \mathrm{Cov}^2(S, \tilde{V}^m) = D\kappa^m \, (\Sigma^m)^{-1} \, (D\kappa^m)' \; ,$$

by (2.35). By (2.34), choice of $\{g_1, g_2, \ldots\}$ and Lemma 2.4

$$E[\tilde{V} - \tilde{V}^m]^2 = E_P[\tilde{\kappa}(X_1, P) - D\kappa^m \, (\Sigma^m)^{-1} \, (g_1(X_1), \ldots, g_m(X_1))']^2 \to 0 \; ,$$

as $m \to \infty$. Furthermore by (2.35)

$$E[S - \tilde{V}^m]^2 = ES^2 - D\kappa^m \, (\Sigma^m)^{-1} (D\kappa^m)' \; .$$

If the expression in (2.33) is zero and ES = 0, then this is smaller than

$$E_P \tilde{\kappa}^2(X_1, P) - D\kappa^m \, (\Sigma^m)^{-1} (D\kappa^m)' \to 0 \; .$$

We conclude that $S = \tilde{V}$ a.s.. Thus by (2.34)

$$L_P(\sqrt{n}(T_n - \kappa(P)) - n^{-\frac{1}{2}} \sum_{i=1}^{n} \tilde{\kappa}(X_i, P)) \xrightarrow{W} \delta_0 \; . \blacksquare$$

2.6. ASYMPTOTIC OPTIMALITY

Since all our results are about the local behaviour of sequences of estimators $\{T_n\}$ near some $P \in \mathcal{P}$, our discussion of optimality must necessarily be local too. It should be silently understood, though, that we would want to apply the discussion to all $P \in \mathcal{P}$ simultaneously. Also, while in the examples in this section we are mainly interested in the behaviour of a sequence of estimators near some fixed P, these estimators have been chosen so as to behave well globally over $\mathcal{P}$.

To judge the relative asymptotic performance of sequences of estimators we separate three cases:

- (i) The tangent cone contains the linear space spanned by the components of the efficient influence function.
- (ii) The tangent cone does not satisfy the condition under (i) but is convex.
- (iii) The tangent cone satisfies none of the possibilities (i) or (ii).

The discussion below may be summarized as follows. Suppose that we want our estimator sequence $\{T_n\}$ to be LAM for all reasonable loss functions. Then in case (i) we necessarily have to use a regular sequence. In case (ii) we could use the best regular estimator sequence (which is LAM), but if this is possible without too much effort we should use a nonregular improvement. In case (iii) the best regular estimator sequence is probably not LAM; it may be hard to find a LAM sequence. Fortunately this case is of little importance.

Here we say that an at $P \in \mathcal{P}$ regular sequence of estimators is *best regular* (*at* P) if its limiting distribution is $N(0, E_P\tilde{\kappa}(X_1,P)\tilde{\kappa}(X_1,P)')$. When we speak of LAM we usually mean LAM in the sense of Theorem 2.8.

As we have seen, in case (i) and (ii) the lower bound for the minimax risk in Theorems 2.8 and 2.10 equals the expected loss under the limiting distribution of a best regular estimator sequence. In this case a best regular estimator sequence is therefore minimax in the sense of Theorem 2.8 and when it attains its normal limit distribution sufficiently uniform in a neighbourhood of P, the same conlusion holds with respect to Theorem 2.10. In case (i) this statement on the optimality of the best regular estimator can be strengthened. Indeed for $k = 1$ any sequence of estimators which is LAM, is necessarily regular. For the LAM Theorem 2.10 this assertion is

part of Hájek's (1972) LAM theorem. An analogous result in terms of the minimax risk of Theorem 2.8 is given by

THEOREM 2.13. *Let* $\kappa: \mathcal{P} \to \mathbb{R}$ *be differentiable at* $P \in \mathcal{P}$ *relative to* $T(P)$ *and suppose that* $\operatorname{lin}\{\tilde{\kappa}(\cdot,P)\}$ *is contained in* $T(P)$. *Let* $\ell: \mathbb{R} \to \mathbb{R}$ *be bowl-shaped and satisfy* $0 < \int \ell(x)e^{\eta x}\, dN(0,E_P\tilde{\kappa}^2(X_1,P))(x) < \infty$, *for some* $\eta > 0$. *Then a weakly regular estimator sequence* $\{T_n\}$ *can have*

$$\sup_{g\in T(P)} \int \ell(x)\, dL_g(x) \le \int \ell(x)\, dN(0,E_P\tilde{\kappa}^2(X_1,P))(x) ,$$

only if

$$(2.36) \qquad \sqrt{n}(T_n-\kappa(P)) = n^{-\frac{1}{2}} \sum_{j=1}^{n} \tilde{\kappa}(X_j,P) + o_P(1). \ \square$$

Relation (2.36) indeed implies that $\{T_n\}$ is regular at $P \in \mathcal{P}$, i.e. $L_g = N(0,E_P\tilde{\kappa}^2(X_1,P))$ for every $g \in T(P)$ (cf. the second assertion of Lemma 2.5).

PROOF OF THEOREM 2.13. Set $J = E_P\tilde{\kappa}^2(X_1,P)$. According to Theorem 2.6, applied with $m = 1$ and $g_1 = \tilde{\kappa}(\cdot,P)/\{E_P\tilde{\kappa}^2(X_1,P)\}^{\frac{1}{2}}$, there exists for every $\lambda > 0$ a probability measure M_λ on $\mathbb{R}$ with

$$(2.37) \qquad \int L_{\alpha g_1}\, dN(0,\lambda)(\alpha) = N(0,(1+\lambda^{-1})^{-1}J) * M_\lambda .$$

It follows from the proof of Theorem 2.6 that

$$(2.38) \qquad M_\lambda = \int L(S-(1+\lambda^{-1})^{-1}J^{\frac{1}{2}}V \mid V=v)\, c(v)\, dN(0,1)(v) ,$$

where $L(V) = N(0,1)$,

$$(2.39) \qquad L_P(\sqrt{n}(T_n-\kappa(P)),\ n^{-\frac{1}{2}} \sum_{j=1}^{n} \tilde{\kappa}(X_j,P)) \to L(S,\ J^{\frac{1}{2}}V)$$

(at least along subsequences of $\{n\}$) and

$$c(v) = (\lambda+1)^{-\frac{1}{2}} \exp(\tfrac{1}{2}(1+\lambda^{-1})^{-1}v^2) .$$

We have for any $\varepsilon \ge 0$ (cf. Ibragimov and Has'minskii (1981),

Sect.2.10)

$$\inf_{|y|\geq\varepsilon} \int \ell(x+y)\, dN(0,(1+\lambda^{-1})^{-1}J)(x) = \int \ell(x+\varepsilon)\, dN(0,(1+\lambda^{-1})^{-1}J)(x).$$

Furthermore for any $0 < \varepsilon \leq J\eta$, as $\lambda \to \infty$

$$\int \ell(x+\varepsilon)\, dN(0,(1+\lambda^{-1})^{-1}J)(x) \to \int \ell(x+\varepsilon)\, dN(0,J)(x).$$

Set

$$\gamma_\varepsilon = \tfrac{1}{2}\, [\, \int \ell(x+\varepsilon)\, dN(0,J)(x) - \int \ell(x)\, dN(0,J)(x)\,].$$

Then $\gamma_\varepsilon > 0$. We now have that for sufficiently small $\varepsilon > 0$ and large λ

$$\inf_{|y|>\varepsilon} \int \ell(x+y)\, dN(0,(1+\lambda^{-1})^{-1}J)(x) \geq \int \ell(x)\, dN(0,(1+\lambda^{-1})^{-1}J)(x) + \gamma_\varepsilon. \tag{2.40}$$

By the assumption of the theorem, (2.37) and (2.40)

$$\int \ell(x)\, dN(0,J)(x) \geq \iint \ell(x)\, dL_{\alpha g_1}(x)\, dN(0,\lambda)(\alpha)$$

$$= \iint \ell(x+y)\, dN(0,(1+\lambda^{-1})^{-1}J)(x)\, dM_\lambda(y)$$

$$\geq \int [\, \int \ell(x)\, dN(0,(1+\lambda^{-1})^{-1}J)(x) + \gamma_\varepsilon 1\{|y|>\varepsilon\}\,]\, dM_\lambda(y).$$

Next inserting (2.38) and using that $c(v) \geq (\lambda+1)^{-\frac{1}{2}}$, we see that this is larger than

$$\int \ell(x)\, dN(0,(1+\lambda^{-1})^{-1}J)(x)$$

$$+ \gamma_\varepsilon(\lambda+1)^{-\frac{1}{2}} \int P(\,|S-(1+\lambda^{-1})^{-1}J^{\frac{1}{2}}V|>\varepsilon \mid V=v)\, dN(0,1)(v)$$

$$= \int \ell(x)\, dN(0,(1+\lambda^{-1})^{-1}J)(x) + \gamma_\varepsilon(\lambda+1)^{-\frac{1}{2}}\, P(\,|S-(1+\lambda^{-1})^{-1}J^{\frac{1}{2}}V| > \varepsilon).$$

We infer

$$P(\,|S-(1+\lambda^{-1})^{-1}J^{\frac{1}{2}}V|>\varepsilon) \leq \gamma_\varepsilon^{-1}(\lambda+1)^{\frac{1}{2}} \int \ell(x)\, d[N(0,J)-N(0,(1+\lambda^{-1})^{-1}J)](x).$$

But this converges to zero as $\lambda \to \infty$. Hence $S = J^{\frac{1}{2}}V$ a.s. Combination with

(2.39) yields (2.36). ■

Theorem 2.13 is restricted to the case k = 1. In fact for k ≥ 2 the theorem may fail, as is shown in

EXAMPLE 2.14. Let $X_1, X_2, \ldots, X_n$ be i.i.d. multivariate normal with mean vector μ and covariance matrix the identity. Consider the Stein shrinkage estimator $T_n = \{ 1 - (k-2)/n\|\bar{X}\| \} \bar{X}$. It is well-known that for dimension larger than two and for joint quadratic loss ($\ell(x) = \|x\|^2$), $\{T_n\}$ globally improves upon the mean $\{\bar{X}_n\}$. Hence it is certainly LAM for this loss function at any $\mu \in \mathbb{R}^k$. At $\mu = 0$ it is not regular, though, so that the generalization of Theorem 2.12 to higher dimensions fails.

However Theorem 2.13 does imply that the k components of the Stein shrinkage estimator can, as estimators of the corresponding components of μ, be LAM at $\mu = 0$ for no reasonable loss function. Indeed LAM would force all components to be asymptotically linear and thus jointly asymptotically normal and regular. □

It seems reasonable to require that the estimator sequences we would use, are componentwise LAM, at least for one, but preferably for all reasonable loss functions. Then for case (i), Theorem 2.13 implies that, as for first order asymptotic behaviour of our sequence, we have a choice out of one: a LAM estimator sequence is necessarily best regular.

In case (ii) this is not true and in general we have a choice between LAM estimator sequences with different asymptotic behaviour. To study their relative performance, it is convenient to restrict ourselves to weakly regular estimator sequences, in which case is it possible to compare the sets of limiting distributions $\{L_g: g \in T(P)\}$ of different sequences of estimators in terms of concentration near the origin. Probably there will not be a single best estimator sequence within the class of LAM estimator sequences (not even locally at one $P \in \mathcal{P}$!). Our choice would necessarily be a matter of taste. However, choosing the best regular estimator sequence could be bad taste, as there may exist weakly regular LAM estimator sequences with every L_g more concentrated near zero than a $N(0, E_P\tilde{\kappa}(X_1,P)\tilde{\kappa}(X_1,P)')$ distribution. (cf. Example 2.15). Note that such a non-regular estimator sequence is better and thus LAM in terms of *any* loss function. The improvement is therefore quite different from that achieved

by e.g. the Stein shrinkage estimator, which is better than the mean for joint quadratic loss, but worse (in fact not even LAM) for other loss functions.

In the above strong sense a best regular estimator may be asymptotically inadmissible within the class of all LAM estimators. Simple situations of this type arise with finite dimensional parametric models, for values of the parameter on the boundary of the parameter set. Typically truncating or projecting the best regular estimator into the parameter set, will mean an improvement, but destroy regularity. A more interesting example is estimation of a parameter in a mixture distribution when the mixing distribution has finite support. (cf. Chapter 5). However, for this case it is yet unknown whether sensible improvements of the best regular estimator sequence exist.

EXAMPLE 2.15. Let $X_1,\ldots,X_n$ be i.i.d. $N(\theta,1)$, where it is known that $\theta \geq 0$. For $\theta > 0$ the tangent cone equals the linear space $\{h\dot{\ell}(\cdot,\theta);\ h \in \mathbb{R}\}$, where $\dot{\ell}(\cdot,\theta) = -(x-\theta)$. For $\theta = 0$, however, the tangent cone is the half space $\{h\dot{\ell}(\cdot,0)\colon h \geq 0\}$. The functional κ given by $\kappa(N(\theta,1)) = \theta$ is differentiable at every $N(\theta,1)$ with canonical gradient $I^{-1}(\theta)\dot{\ell}(\cdot,\theta)$, where $I(\theta) = E_\theta \dot{\ell}^2(X_1,\theta) = 1$. Hence the minimax risk is $\int \ell(x)\, dN(0,1)(x)$ at every $N(\theta,1)$. $T_n = n^{-1}\Sigma X_j$ is an obvious estimator. The sequence $\{T_n\}$ is best regular and LAM at every $N(\theta,1)$ ($\theta \geq 0$). However, for $\theta = 0$ its performance can be improved. One improvement is $\{T_n^*\} = \{T_n 1\{T_n \geq 0\}\}$, which is regular at every $N(\theta,1)$ with $\theta > 0$ but has different behaviour from $\{T_n\}$ at $N(0,1)$. Letting Φ denote the cumulative distribution function of the standard normal distribution we have

$$P_{hn^{-\frac{1}{2}}}\left(\sqrt{n}(T_n^* - hn^{-\frac{1}{2}}) \leq x\right) \to \Phi(x)1_{[-h,\infty)}(x)\ .$$

Hence for any $h \geq 0$ the limiting distributions $L^*_{h\dot{\ell}(\cdot,0)}$ of $\{T_n^*\}$ are more concentrated around zero then the corresponding $N(0,1)$ limiting distributions of $\{T_n\}$. (We remark that $\{T_n^*\}$ itself is asymptotically inadmissible with respect to quadratic loss in the sense that there exist estimator sequences $\{T_n^{**}\}$ with for all $h > 0$

$$L_{hn^{-\frac{1}{2}}}\left(\sqrt{n}(T_n^{**} - hn^{-\frac{1}{2}})\right) \to L^{**}_{h\dot{\ell}(\cdot,0)}$$

and

$$\int x^2 \, dL^{**}_{h\ell(\cdot,0)}(x) \le \int x^2 \, dL^{*}_{h\ell(\cdot,0)}(x) ,$$

with strict inequality for some $h > 0$ (cf. Sacks (1963))). □

Hence in case (ii) it may be sensible not to require that $\{T_n\}$ is regular. However since the best regular estimator is still LAM in this case, the gain in efficiency by using a non-regular estimator sequence can never be dramatically large.

On the contrary, in case (iii) where T(P) is not convex, the gain in efficiency can be substantial, as is shown in Example 2.15 below. Note in this connection that a regular estimator sequence is asymptotically of constant bias, so that by Theorem 2.12 for $k = 1$, the variance of its limiting distribution (i.e. its local maximum quadratic risk in terms of Theorem 2.8) is always bounded from below by $E_P\tilde{\kappa}^2(X_1,P)$. Without convexity of the tangent cone the latter quantity will generally not be a lower bound for the local maximum quadratic risk of an arbitrary estimator sequence, i.e. the assertion of Theorem 2.8 fails to be true. Indeed the example below shows that there may be weakly regular sequences of estimators such that

$$\sup_{g\in T(P)} \int x^2 \, dL_g(x) < E_P\tilde{\kappa}^2(X_1,P) .$$

When the tangent cone is not convex it can certainly pay to allow the estimator sequence to be asymptotically biased.

However, one usually encounters T(P) that are convex.

EXAMPLE 2.16. Let $X_1,\dots,X_n$ be i.i.d. $N(\theta,1+\sqrt{2}\tau)$, where (θ,τ) is known to be in $B = \{(u,v): u = 0, v \ge 0, \text{ or } u \ge 0, v = 0\}$, the boundary of the positive quadrant in $\mathbb{R}^2$. Letting $\ell_\theta(x) = x$ and $\ell_\tau(x) = 2^{-\frac{1}{2}}(x^2-1)$, we may set $T(N(0,1)) = \{u\ell_\theta + v\ell_\tau: (u,v) \in B\}$. Consider estimation of the functional κ given by $\kappa(N(\theta,1+\sqrt{2}\tau)) = \theta+\tau$. This functional is differentiable at $N(0,1)$ with efficient influence function $\tilde{\kappa}(x,N(0,1)) = \ell_\theta(x) + \ell_\tau(x)$ and $E_{N(0,1)}\tilde{\kappa}^2(X_1,N(0,1)) = 2$.

Thus by Theorem 2.12 a lower bound for the asymptotic variance at

N(0,1) of any regular sequence of estimators of κ is given by 2. We shall now exhibit an at N(0,1) weakly regular sequence of estimators $\{T_n\}$ such that

$$(2.41) \qquad \sup_{(u,v)\in B} \int \ell(x)\, dL_{u\ell_\theta + v\ell_\tau}(x) \le \int \ell(x)\, dN(0,1)(x)$$

for every bowl-shaped loss function ℓ. Moreover $\{T_n\}$ will be best regular at every member of P other than N(0,1). Hence, specializing to $\ell(x) = x^2$, the local maximum quadratic risk of $\{T_n\}$ at N(0,1) is half the quadratic risk of the best regular estimator. (By applying Theorem 2.8 to the one-dimensional submodels corresponding to the half lines in T(N(0,1)) one obtains a lower bound $\int \ell(x)\, dN(0,1)$ for the left hand side of (2.41), so that $\{T_n\}$ is actually LAM at N(0,1) for this problem).

Define $T_n = \max\{\bar{X}_n,\ 2^{-\frac{1}{2}}(n^{-1}\Sigma(X_j-\bar{X}_n)^2-1),\ 0\}$, where $\bar{X}_n$ is the average of $X_1,\dots,X_n$. Then for all $h \ge 0$ both $L_{N(hn^{-\frac{1}{2}},1)}(\sqrt{n}(T_n-hn^{-\frac{1}{2}}))$ and $L_{N(0,1+\sqrt{2}hn^{-\frac{1}{2}})}(\sqrt{n}(T_n-hn^{-\frac{1}{2}}))$ converge weakly to the distribution with cumulative distribution function

$$G_h(x) = \Phi(x)\,\Phi(x+h)\,1_{[-h,\infty)}(x) .$$

When G and H are symmetric distributions around zero, call G more concentrated (near zero) if $G(x) \ge H(x)$ for all $x \ge 0$. Suppose that it has been proved that the symmetrized $G_h(x)$, i.e. $\frac{1}{2}[\,G_h(x) + 1-G_h((-x)-)\,]$, is more concentrated than Φ for any $h \ge 0$. Then, using symmetry of ℓ

$$\int \ell(x)\, dG_h(x) = \int \ell(x)\, \tfrac{1}{2}\, d[G_h(x) + 1-G_h((-x)-)] \le \int \ell(x)\, d\Phi(x) .$$

But this implies (2.41).

Finally, the symmetrized $G_h(x)$ is more concentrated than the symmetrized $\Phi(x)\Phi(x+h)$ $(h \ge 0)$, which is more concentrated than Φ iff

$$\tfrac{1}{2}[\Phi(x)\Phi(x+h) + 1-\Phi(-x)\Phi(-x+h)] \ge \Phi(x) , \qquad x \ge 0.$$

This inequality is equivalent to

$$\Phi(x)\Phi(-x-h) \le \Phi(x-h)\Phi(-x) , \qquad x \ge 0 ,$$

which follows from concavity of Φ. □

EXAMPLE 2.17. Let $X_1,\ldots,X_n$ and κ be as it Example 2.15, but let (θ,τ) range over $B' = \{(u,v)\colon u = 0,\ v \geq 0,\ \text{or } u \geq 0,\ v \leq cu\}$, where c is a positive constant. If c is small then B' looks only slightly larger than B. However, for the present situation there exists no estimator sequence with LAM risk smaller than $\int \ell(x)\, dN(0,2)(x)$ at $N(0,1)$. Also note that $\tilde{\kappa}(\cdot,P) \notin T(N(0,1)) = \{u\dot{\ell}_\theta + v\dot{\ell}_\tau\colon (u,v) \in B'\}$ if $c < 1$. □

2.7. THEOREMS FOR LOCALLY ASYMPTOTICALLY NORMAL MODELS

In this section we generalize some of the theorems presented so far to the non-i.i.d case.

For each $n = 1,2\ldots$ let $(\mathcal{X}_n,\mathcal{B}_n)$ be a measurable space and $\mathcal{P}_n$ be a set of probability measures on $(\mathcal{X}_n,\mathcal{B}_n)$. Furthermore let $\kappa_n\colon \mathcal{P}_n \to \mathbb{R}^k$ be a sequence of functionals, to be estimated by a sequence of estimators (measurable maps) $T_n\colon (\mathcal{X}_n,\mathcal{B}_n) \to \mathbb{R}^k$. The i.i.d. set-up studied in the earlier sections corresponds to $(\mathcal{X}_n,\mathcal{B}_n,P_n)$ being the product space of n copies of $(\mathcal{X},\mathcal{B},P)$. In the general set-up we do not have a tangent cone. The first result of Lemma 2.5 is now postulated.

DEFINITION 2.18. *Let* C *be a cone in* $\mathbb{R}^m$. *A sequence of maps* $P_n\colon C \to \mathcal{P}_n$ *is called a differentiable submodel (of* $\mathcal{P}_n$*) if for all* $h \in C$ *and a positive definite (m×m) matrix* Σ

$$(2.42)\qquad \Lambda_n(P_n(h),P_n(0)) = h'\Delta_n - \tfrac{1}{2}h'\Sigma h + o_{P_n(0)}(1),$$

where $\Delta_n\colon (\mathcal{X}_n,\mathcal{B}_n) \to \mathbb{R}^m$ *(n = 1,2,...) are measurable maps with*

$$(2.43)\qquad \mathcal{L}_{P_n(0)}(\Delta_n) \to N_m(0,\Sigma).\ \square$$

Here $\Lambda_n(Q_n,P_n)$ is the log likelihood ratio of Q_n with respect to P_n (cf. Definition A.2). Definition 2.18 is a version of a definition of a locally asymptotically normal model.

DEFINITION 2.19. *A sequence of functionals* $\kappa_n: P_n \to \mathbb{R}^k$ *is called differentiable on the differentiable submodel* (P_n,C), *if there exists a sequence* $\{R_n\}$ *of positive definite* $(k\times k)$ *matrices and a* $(k\times m)$ *matrix* $D\kappa$ *such that for all* $h \in C$

$$R_n\,(\kappa_n(P_n(h))-\kappa_n(P_n(0))) \to D\kappa h. \quad \square$$

The first result is a generalized convolution theorem.

THEOREM 2.20. *Let* $(P_n,\mathbb{R}^m)$ *be a differentiable submodel and* $\kappa_n: P_n \to \mathbb{R}^k$ *a differentiable sequence of functionals on* $(P_n,\mathbb{R}^m)$. *Let* $\{T_n\}$ *be an arbitrary sequence of estimators. Then for each subsequence of* $\{n\}$ *there exists a further subsequence* $\{n'\}$ *such that*

$$L_{P_{n'}(h)}(\,R_{n'}\,(T_{n'}-\kappa_{n'}(P_{n'}(h)))\,) \overset{v}{\to} L_h\,,$$

where for each $h \in \mathbb{R}^m$, L_h *is a (possibly defective) probability distribution on* $\mathbb{R}^k$. *Moreover for any* $\tau \in \mathbb{R}^m$ *and any positive definite* $(m\times m)$ *matrix* Λ,

$$\int_{\mathbb{R}^m} L_h \, dN(\tau,\Lambda)(h) = N(0,D\kappa(\Sigma+\Lambda^{-1})^{-1}D\kappa') * M\,.$$

Here M *is a (possibly defective) distribution on* $\mathbb{R}^k$. $\square$

PROOF. Analogous to the proof of Theorem 2.8. ∎

Call a cone $C \subset \mathbb{R}^m$ a *uniqueness set* (for analytic continuation) if every complex valued function which is analytic on $\mathbb{C}^m$ and is constant on C, is necessarily constant on $\mathbb{C}^m$. There is no simple alternative description of a uniqueness set. However it is well-known that it is sufficient that C has non-empty interior as a subset of $\mathbb{R}^m$. A convex cone C with $\operatorname{lin} C = \mathbb{R}^m$ is therefore a uniqueness set.

Asymptotic bounds for $\mathbb{R}^k$

THEOREM 2.21. *Let* C *be a uniqueness set,* (P_n,C) *be a differentiable submodel and* $\kappa_n: P_n \to \mathbb{R}^k$ *a differentiable sequence of functionals on* (P_n,C). *Let* $\{T_n\}$ *be a sequence of estimators satisfying for all* $h \in C$

$$L_{P_n(h)}(R_n (T_n-\kappa_n(P_n(h)))) \to L ,$$

where L *is a probability measure on* $\mathbb{R}^k$. *Then*

$$L = N(0, D\kappa\Sigma^{-1}D\kappa') * M. \quad \square$$

PROOF. By Prohorov's theorem there exists a subsequence of {n}, abusing notation denoted {n}, such that

$$L_{P_n(0)}(R_n(T_n-\kappa_n(P_n(0))), \Delta_n), \to L(S,V)$$

where $L(V) = N_m(0,\Sigma)$. Hence for all $h \in C$

$$L_{P_n(0)}(R_n(T_n-\kappa_n(P_n(h))), \Lambda_n(P_n(h),P_n(0))) \to L(S-D\kappa h, h'V-\tfrac{1}{2}h'\Sigma h) .$$

Thus by Proposition A.6,

$$L_{P_n(h)}(R_n(T_n-\kappa_n(P_n(h)))) \to L_h ,$$

where

$$\int e^{it'x} dL_h(x) = \int_{\mathbb{R}\times\mathbb{R}} e^{it'x + \lambda} dL(S-D\kappa h, h'V-\tfrac{1}{2}h'\Sigma h)(x,\lambda)$$

$$= E\, e^{it'(S-D\kappa h) + h'V - \frac{1}{2}h'\Sigma h} .$$

But by assumption the left hand side is constant in $h \in C$. Since C is a uniqueness set and the right hand side analytic on $\mathbb{C}^m$, we may substitute $h = -i\Sigma^{-1}D\kappa't$, to obtain

$$\int e^{it'x} dL(x) = E\, e^{it'(S-D\kappa\Sigma^{-1}V)} \; e^{-\frac{1}{2}t'D\kappa\Sigma^{-1}D\kappa't} . \blacksquare$$

Other theorems for the i.i.d. case can be generalized as well. In particular we have the following version of the LAM theorem.

THEOREM 2.22. *Let* C *be a convex cone with* $\lin C = \mathbb{R}^m$, *let* (P_n, C) *be a differentiable submodel and* $\kappa_n: P_n \to \mathbb{R}^k$ *a differentiable sequence of functionals on* (P_n, C). *Then for any estimator sequence* $\{T_n\}$ *and any bowl-shaped loss function* ℓ

$$\lim_{c\to\infty} \liminf_{n\to\infty} \sup_{\|h\|<c, h\in C} E_{P_n(h)} \ell\big(R_n(T_n - \kappa(P_n(h)))\big)$$

$$\geq \int \ell(x)\, dN(0, D\kappa\Sigma^{-1}D\kappa')(x). \quad \square$$

PROOF. Similar to Theorem 2.9. ■

CHAPTER 3

ASYMPTOTIC BOUNDS ON THE PERFORMANCE OF ESTIMATORS WITH VALUES IN A VECTOR SPACE

3.1. INTRODUCTION

Let $\mathcal{P}$ be a set of probability measures on the measurable space $(X,\mathcal{B})$ and let B be a vector space. In this chapter we consider asymptotic lower bounds for estimators of functionals $\kappa\colon \mathcal{P} \to B$, where we shall restrict ourselves to estimators based on i.i.d. random variables $X_1,X_2,\ldots,X_n$ with distribution $P \in \mathcal{P}$.

The convolution and local asymptotic minimax theorems obtained in Sections 3.2 and 3.3 extend and unify results of Beran (1977), Wellner (1982), Begun et al. (1983) and Millar (1983,1985). Millar (1983,1985) obtains the most general results, applicable to separable Banach spaces. His method relies on the 'abstract' approach of Le Cam (cf. Le Cam (1972, 1986)).

Our set-up generalizes the one used for the finite dimensional situation in Chapter 2. It is based on an extension of Definition 2.2 of an $\mathbb{R}^k$-valued differentiable functional (basically due to Pfanzagl (1982)) to the case of infinite dimensional spaces, and leads to easily applicable theorems. The results obtained depend to a certain extent on the choice of a σ-algebra with respect to which estimators are defined. Whereas in the Euclidean situation the Borel σ-field is considered natural, for the infinite dimensional case the choice to be made is less clear. Some of the statistically most interesting applications need Banach spaces B which are not separable. For instance, estimation of a distribution function on the real line is best studied in $D[-\infty,\infty]$ under its supremum norm; estimation of

a measure by a set-indexed empirical process is often considered within the context of (a subspace of) the space of all bounded functions $B(\mathcal{C})$ under the supremum metric. Unfortunately, for nonseparable normed spaces the Borel σ-field is too large for many applications. Which σ-algebra *is* suitable is not always clear. For $D[a,b]$ the projection σ-field (which equals the Skorohod Borel σ-field) is generally accepted. For $B(T)$ one usually chooses a σ-field generated by coordinate projections and/or closed balls.

The LAM and convolution theorem in this chapter apply to estimators with respect to σ-algebras induced by linear, real-valued functions on B. For $D[a,b]$ and for separable normed spaces, this leads to a satisfactory theory. For $B(T)$ this is true under additional considerations.

The choice of this type of σ-field has been motivated by examples, where σ-fields are often generated by linear maps. A more compelling motivation, though, is first, that convolution may not be defined on other σ-fields (the addition map need not be measurable with respect to the product σ-field). Secondly, the methods to obtain convolution and LAM theorems that are presently known simply do not apply to other σ-fields. These methods describe a probability measure on an infinite dimensional space by specifying its finite dimensional marginal distributions (which are the measures induced on $\mathbb{R}^k$ under linear maps). Next the marginals are analysed with the help of the theorems of Chapter 2. To a certain extent the theorems in the present chapter are therefore theorems about marginals only. As a consequence we need not discuss *weak convergence* of measures on B in this chapter (we will in Chapter 4).

The organization of the chapter is as follows. In the remainder of this introduction we generalize the definition of a differentiable functional from $\mathbb{R}^k$-valued κ to κ with values in a topological vector space. Next we recall some facts concerning (cylinder) measures on σ-fields generated by linear functionals. In Section 3.2 and 3.3 we obtain general convolution and LAM theorems, respectively. In Section 3.5 we discuss these theorems with additional detail for some special vector spaces B. Section 3.6 contains examples of differentiable functionals.

The remaining sections are of a more technical nature, and include a result on the support of the optimal limiting measure (Section 3.4), and a characterization of the dual space of $(D[a,b], \|\cdot\|_\infty)$ (Section 3.7).

3.1.1. Differentiable functionals, influence functions.

Recall Definition 2.1 of a tangent cone T(P).

DEFINITION 3.1. *Let* (B,τ) *be a topological vector space. A functional* $\kappa: \mathcal{P} \to (B,\tau)$ *is called differentiable at* $P \in \mathcal{P}$ *relative to* $T(P)$, *if there exists a continuous, linear map* $\kappa_P' : \text{lin } T(P) \to (B,\tau)$, *such that for every* $g \in T(P)$

$$(3.1) \qquad t^{-1}(\kappa(P_t)-\kappa(P)) \to \kappa_P'(g),$$

for some sequence $\{P_t\} \subset \mathcal{P}$ *satisfying* (2.1). □

For the case that (B,τ) is $\mathbb{R}^k$ with the usual topology, Definition 3.1 is equivalent to Definition 2.2. This follows from the well-known characterization of continuous, linear, real-valued functionals on a Hilbert space as inner products (cf. e.g. Jameson (1974), Prop.32.7). In our case we first note that a continuous, linear, *real*-valued functional $g \to \kappa_P'(g)$ defined on lin T(P) can be extended to a continuous, linear functional defined on the closure of lin T(P) in $L_2(P)$ (cf. Jameson (1974), Prop. 8.10). Next the extension, which is defined on a Hilbert space, admits a *gradient* $\dot{\kappa}(\cdot,P) \in L_2(P)$ in the sense that

$$\kappa_P'(g) = \int g(x)\, \dot{\kappa}(x,P)\, dP(x).$$

Gradients have been useful to describe lower bounds in $\mathbb{R}^k$. For arbitrary topological vector spaces we define gradients in the following way. Let B_τ^* be the dual space of the topological vector space (B,τ), i.e. the set of all τ-continuous, linear functions $b^*: B \to \mathbb{R}$.

DEFINITION 3.2. *Let* $\kappa : \mathcal{P} \to (B,\tau)$ *be differentiable at* $P \in \mathcal{P}$ *and* $b^* \in B_\tau^*$. *A gradient of* κ *in the direction* b^* *is a function* $\dot{\kappa}_{b^*}(.,P) \in L_2(P)$ *such that*

$$(3.2) \qquad b^* \circ \kappa_P'(g) = \int g(x)\, \dot{\kappa}_{b^*}(x,P)\, dP(x),$$

for all $g \in T(P)$. □

Since every $b^* \circ \kappa_P'$ is a continuous, linear, real-valued function on

lin T(P) gradients indeed exist in all directions (cf. discussion above). When compared to Definition 2.2 there is a slight abuse of notation, as the components of the gradient $\dot{\kappa}(\cdot,P)$ of Chapter 2 correspond to the $\dot{\kappa}_{e_i}(\cdot,P)$ of Definition 3.2 (where e_i is the i-th unit vector in $(\mathbb{R}^k)^* = \mathbb{R}^k$). Just as in Chapter 2 $\tilde{\kappa}_{b^*}(\cdot,P)$ denotes a *canonical gradient* and is defined as the projection of an arbitrary gradient $\dot{\kappa}_{b^*}(x,P)$ onto the closure of lin T(P) in $L_2(P)$.

Many commonly used functionals are differentiable in the sense of Definition 3.1. Some examples are given in Section 3.6.[1)]

3.1.2. Cylinder measures.

Given an arbitrary vector space B, estimators will be defined as measurable functions of the observations in $(B,\mathcal{U})$, with respect to some σ-algebra $\mathcal{U}$. In this chapter we mainly consider $\mathcal{U}$ of the following form. A vector space B' of linear functions $b'\colon B \to \mathbb{R}$ is given and $\mathcal{U} = \mathcal{U}(B')$, the smallest σ-algebra making all elements of B' measurable.

So-called cylinder σ-algebras and -measures play an important role. Given a finite set $\{b_1',b_2',\ldots,b_k'\} \subset B'$, $\mathcal{U}(b_1',b_2',\ldots,b_k')$ is called a *cylinder σ-algebra*. The union of all cylinder σ-algebras

$$(3.3) \qquad \mathcal{A}(B') = \cup\, \mathcal{U}(b_1',b_2',\ldots,b_k')$$

is an *algebra* which generates $\mathcal{U}(B')$. A *cylinder (probability) measure* M on $(B,\mathcal{A}(B'))$ is a function $M\colon \mathcal{A}(B') \to \bar{\mathbb{R}}$, of which the restriction to every cylinder σ-algebra is a σ-additive (probability) measure.

Since $\mathcal{A}(B')$ generates $\mathcal{U}(B')$, by Carathéodory's theorem any probability measure M on $(B,\mathcal{U}(B'))$ is uniquely determined by its finite dimensional *marginal distributions*

$$(b_1',b_2',\ldots,b_k')(M) \ ,$$

which are defined as the image measures of M on $\mathbb{R}^k$ under the maps given by

---

1) *An alternative definition would be to require only that* $b^* \circ \kappa_{\dot{}}\colon P \to \mathbb{R}$ *is differentiable in the sense of Definition 2.2, for every* $b^* \in B^*_T$. *In fact this would be sufficient to obtain a convolution and LAM theorem. Definition 3.1 with* τ *as in Theorem 3.7 requires slightly more. However this will turn out to be useful in a number of other results.*

$b \to (b_1'(b), b_2'(b), \ldots, b_k'(b))$ from B to $\mathbb{R}^k$. Moreover, as we assume that B' is a vector space, it is in fact true that M is already determined by the set of its *one*-dimensional marginals $\{b'(M): b' \in B'\}$.

It follows that any cylinder probability measure M on $(B, \mathcal{A}(B'))$ can in *at most* one manner be extended to a probability measure on $(B, \mathcal{U}(B'))$. Such an extension does not always exist. The following theorem, which is a version of a theorem due to Prohorov (1956), gives a sufficient condition. It will be used in the proof of Theorem 3.7. Recall that B' separates points of B if for any pair b_1 and b_2 in B there exists $b' \in B'$ with $b'(b_1) \neq b'(b_2)$. Furthermore, let $\tau(B')$ be the weakest topology on B making all elements of B' continuous, and $\mathcal{U}(\tau(B'))$ its Borel σ-field. Then $\mathcal{U}(B') \subset \mathcal{U}(\tau(B'))$.

PROPOSITION 3.3. *Let* B *be a vector space and* B' *a vector space of linear, real-valued functions on* B*, separating points of* B*. Let* M *be a cylinder probability measure on* $(B, \mathcal{A}(B'))$*, such that for any* $\varepsilon > 0$*, there exists a* $\tau(B')$*-compact set* $K_\varepsilon \subset B$ *such that for all finite sets* $\{b_1', b_2', \ldots b_k'\} \subset B'$

$$(3.4) \qquad (b_1', b_2', \ldots, b_k')(M)\ ((b_1', b_2', \ldots, b_k')K_\varepsilon) \geq 1-\varepsilon.$$

Then M *can be extended to a probability measure* $\tilde{M}$ *on* $(B, \mathcal{U}(\tau(B')))$*. Moreover* $\tilde{M}(\ \bigcup_{m=1}^{\infty} K_{1/m}\ ) = 1$. □

PROOF. Because B' separates points of B we have that B' is the dual space of the locally convex topological vector space $(B, \tau(B'))$ (cf. Rudin (1973), Th.3.10). Next see Araujo and Giné (1980), Th.(1.4.23). ■

Any cylinder measure M on $(B, \mathcal{A}(B'))$ defines *a system of finite dimensional distributions*

$$\{M_{b_1', b_2', \ldots, b_k'}\} \quad (\{b_1', b_2', \ldots, b_k'\} \subset B',\ k = 1,2,\ldots)\ ,$$

through

$$(3.5) \qquad M_{b_1', b_2', \ldots, b_k'} = (b_1', b_2', \ldots, b_k')(M).$$

This system is *consistent* in the sense that for any linear map R: $\mathbb{R}^k \to \mathbb{R}^r$

(3.6) $\quad R(M_{b_1',b_2',\ldots,b_k'}) = M_{(R(b_1',b_2',\ldots,b_k')')'}$.

Another result that we need in the proof of Theorem 3.7 is that (3.6) actually *characterizes* those systems that are induced by cylinder measures.

PROPOSITION 3.4. *Let* B *be a vector space and* B' *a vector space of linear, real-valued functions on* B. *For any finite set* $\{b_1',b_2',\ldots,b_k'\} \subset B'$ *let* $M_{b_1',b_2',\ldots,b_k'}$ *be a measure on* $\mathbb{R}^k$ *and suppose that* (3.6) *holds. Then there exists a cylinder measure* M *on* $(B,\mathcal{A}(B'))$ *satisfying* (3.5). □

PROOF. For $A_k \in \mathcal{B}^k$ and $A = (b_1',b_2',\ldots,b_k')^{-1}(A_k) \in \mathcal{A}(B')$ define

$$M(A) = M_{b_1',b_2',\ldots,b_k'}(A_k).$$

If M is well-defined then it clearly fulfils the requirements. We must show that the relation

(3.7) $\quad A = (b_1',b_2',\ldots,b_k')^{-1}(A_k) = (\underline{b}_1',\underline{b}_2',\ldots,\underline{b}_m')^{-1}(\underline{A}_m)$

implies

$$M_{b_1',b_2',\ldots,b_k'}(A_k) = M_{\underline{b}_1',\underline{b}_2',\ldots,\underline{b}_m'}(\underline{A}_m).$$

Let $\{b_1',b_2',\ldots,b_r'\}$ be a basis for lin $\{b_1',b_2',\ldots,b_k', \underline{b}_1',\underline{b}_2',\ldots,\underline{b}_m'\}$ and let $R: \mathbb{R}^r \to \mathbb{R}^k$ be a map with $(b_1',b_2',\ldots,b_k')' = R(b_1',b_2',\ldots,b_r')'$. Then by (3.6)

(3.8) $\quad M_{b_1',b_2',\ldots,b_k'}(A_k) = M_{b_1',b_2',\ldots,b_r'}(R^{-1}(A_k))$.

Repeating this argument we also find a map $\underline{R}: \mathbb{R}^r \to \mathbb{R}^m$ with

(3.9) $\quad M_{\underline{b}_1',\underline{b}_2',\ldots,\underline{b}_m'}(\underline{A}_m) = M_{b_1',b_2',\ldots,b_r'}(\underline{R}^{-1}(\underline{A}_m))$.

By (3.7) we have that A equals

(3.10) $\quad \{b: (b_1',b_2',\ldots,b_r')(b) \in R^{-1}(A_k)\} = \{b: (b_1',b_2',\ldots,b_r')(b) \in \underline{R}^{-1}(\underline{A}_m)\}$.

Since $b \to (b_1', b_2', \ldots, b_r')(b)$ is a map of B onto $\mathbb{R}^r$ we have $R^{-1}(A_k) = \underline{R}^{-1}(\underline{A}_m)$. Combination with (3.8)-(3.9) concludes the proof. ∎

3.1.3. (Weakly regular) Estimator sequences.

Let $X_1, \ldots, X_n$ be i.i.d. random elements in a measurable space $(X,\mathcal{B})$ with distribution $P \in \mathcal{P}$. Let $\mathcal{U}$ be a σ-algebra on B. An *estimator sequence* in $(B,\mathcal{U})$ is a sequence $\{T_n\}$, $T_n = t_n(X_1, \ldots, X_n)$, where t_n: $(X^n, \mathcal{B}^n) \to (B,\mathcal{U})$ are measurable maps. As in Chapter 2 we single out a set of estimator sequences satisfying a weak regularity condition. Let B' be a vector space of linear, real-valued functions on B. An estimator sequence $\{T_n\}$ in $(B,\mathcal{U}(B'))$ is called (B'-) *weakly regular at* $P \in \mathcal{P}$ if for any $g \in T(P)$ there exists a probability measure L_g on $(B,\mathcal{U}(B'))$ and a sequence $\{P_n\} \subset \mathcal{P}$ such that

$$\int [\sqrt{n}((dP_n)^{\frac{1}{2}} - (dP)^{\frac{1}{2}}) - \tfrac{1}{2} g\,(dP)^{\frac{1}{2}}]^2 \to 0 \tag{3.11}$$

$$\sqrt{n}\,(\kappa(P_n) - \kappa(P)) \to \kappa_P'(g) \qquad \text{in } \tau(B') \tag{3.12}$$

and

$$L_{P_n}(\, b'[\sqrt{n}(T_n - \kappa(P_n))]\,) \to b'(L_g) \qquad \text{for all } b' \in B'. \tag{3.13}$$

An estimator sequence is called (B'-) *regular*, if it satisfies (3.11)-(3.13) with $L_g = L$ for a probability measure L and every $g \in T(P)$.

Note that (3.13) only concerns *marginal convergence* of $\{L_{P_n}(\, \sqrt{n}(T_n - \kappa(P_n))\,)\}$. In the convolution and one version of the LAM theorem we aim at comparing the quality of different estimator sequences by comparing limiting distributions. It may be argued that to make this type of comparison valuable, convergence of $L_{P_n}(\, \sqrt{n}(T_n - \kappa(P_n))\,)$ to its limit should be in a stronger sense than convergence of marginals. In Chapter 4 we therefore study *weak convergence* of the estimator sequences in a suitable sense. However, this stronger sense of convergence implies marginal convergence, and as it happens, the convolution theorem depends on marginal convergence only. Thus in this chapter we speak about weak convergence only occasionally.

3.2. CONVOLUTION THEOREM

3.2.1. Convolution of measures on $(B,\mathcal{U}(B'))$.

We first recall some facts concerning convolution of measures. For two probability measures μ and ν on $(B,\mathcal{U}(B'))$ define the product measure $\mu\otimes\nu$ on the measurable space $(B\times B,\mathcal{U}(B')\times\mathcal{U}(B'))$ as usual. The map $S\colon (B\times B,\mathcal{U}(B')\times\mathcal{U}(B')) \to (B,\mathcal{U}(B'))$ given by $S(x,y) = x+y$ is measurable, because for each $b' \in B'$ the map in $\mathbb{R}$

$$(x,y) \to (b'(x),b'(y)) \to b'\circ S(x,y) = b'(x) + b'(y)$$

is $\mathcal{U}(B')\times\mathcal{U}(B')$-measurable. Hence the image measure $\mu*\nu = S(\mu\otimes\nu)$ is well defined as a measure on $(B,\mathcal{U}(B'))$. As always $\mu*\nu$ is called the *convolution* of μ and ν. By Fubini's theorem, for any $A \in \mathcal{U}(B')$

$$(\mu*\nu)(A) = (\mu\otimes\nu)(S^{-1}(A)) = \int \mu(A-y)\, d\nu(y).$$

The following lemma is basic for Theorem 3.7.

LEMMA 3.5. *Let* L,μ,ν *be measures on* $(B,\mathcal{U}(B'))$, *where* B' *is a vector space of linear, real-valued functions on* B. *Then*

$$L = \mu*\nu \quad \textit{iff} \quad b'(L) = b'(\mu)*b'(\nu) \quad \textit{for all } b' \in B'. \ \square$$

PROOF. On a suitable probability space $(\Omega,\mathcal{A},P)$ define $(B,\mathcal{U}(B'))$-valued random elements (measurable maps) X and Y such that X and Y are independent with $\mathcal{L}(X) = \mu$ and $\mathcal{L}(Y) = \nu$. Then $\mathcal{L}(X+Y) = \mu*\nu$ and since for every $b' \in B'$, $b'(X)$ and $b'(Y)$ are independent random variables in $\mathbb{R}$ we have

$$(3.14)\quad b'(\mu*\nu)=\mathcal{L}(b'(X+Y)) = \mathcal{L}(b'(X)+b'(Y)) = \mathcal{L}(b'(X))*\mathcal{L}(b'(Y))=b'(\mu)*b'(\nu).$$

Now suppose $L = \mu*\nu$. Then $b'(L) = b'(\mu*\nu) = b'(\mu)*b'(\nu)$ by (3.14). Conversely assume that $b'(L) = b'(\mu)*b'(\nu)$ for all $b' \in B'$. Set $L_0 = \mu*\nu$. By (3.14) $b'(L_0) = b'(\mu)*b'(\nu)$. Hence $b'(L_0) = b'(L)$ for all $b' \in B'$, which implies $L_0 = L$ by Carathéodory's theorem. ∎

In view of Lemma 3.5 the following is a natural extension of the

definition of convolution to cylinder probability measures.

DEFINITION 3.6. *Let* L, μ *and* ν *be cylinder probability measures on* $(B,\mathcal{A}(B'))$, *where* B' *is a vector space of linear, real-valued functions on* B. *Then we call* L *the convolution of* μ *and* ν *if* $b'(L) = b'(\mu)*b'(\nu)$ *for all* $b' \in B'$. *We denote this by* $L = \mu*\nu$. □

3.2.2. Convolution theorem.

Call a probability measure L on $(B,\mathcal{U}(B'))$ $\tau(B')$-tight if to any $\varepsilon > 0$ there exists a $\tau(B')$-compact set K_ε such that

$$L^*(K_\varepsilon) = \inf \{L(G): G \in \mathcal{U}(B'),\ G \supset K_\varepsilon\} \geq 1-\varepsilon.$$

We are ready for the convolution theorem.

THEOREM 3.7. *Let* B *be a vector space and* B' *a vector space of linear, real-valued functions on* B *separating points of* B. *Assume that* T(P) *is convex and that* $\kappa: \mathcal{P} \to (B,\tau(B'))$ *is differentiable at* $P \in \mathcal{P}$ *relative to* T(P). *Then any limiting distribution* L *of a regular estimator sequence in* $(B,\mathcal{U}(B'))$ *(at* P*) satisfies*

$$L = N*M \qquad \text{on } (B,\mathcal{A}(B')),$$

where N *and* M *are cylinder probability measures on* $(B,\mathcal{A}(B'))$, *with*

$$b'(N) = N(0,\|\tilde{\kappa}_{b'}(\cdot,P)\|_P^2) \qquad \text{for all } b' \in B'. \tag{3.15}$$

Furthermore, if L *is* $\tau(B')$*-tight, then* N *and* M *can be extended to probability measures on* $(B,\mathcal{U}(B'))$. □

We have introduced $\tau(B')$-tightness of the limit measure L in Theorem 3.7 as a sufficient condition for existence of N and M as measures on $(B,\mathcal{U}(B'))$, rather than as cylinder measures on $(B,\mathcal{A}(B'))$. If indeed one is interested in existence of N and M as measures, this assumption restricts the applicability of the theorem to a still smaller class of estimators than the already restricted class of regular estimators. Fortunately in many cases B' can be chosen in such a way that *any* probability measure is

$\tau(B')$-tight. For instance any probability measure on a Polish space is tight with respect to the topology (cf. Parthasarathy (1967, Th.3.2)), hence certainly $\tau(B')$-tight if B' is a set of continuous maps. Other examples where the tightness condition can be omitted are discussed in Section 3.5, and include $D[a,b]$ and $B(T)$ with their projection σ-field.

We do not know whether the second conclusion of Theorem 3.7 remains true if the $\tau(B')$-tightness condition is deleted from the assumptions. However, one can also reverse the above argument and interprete the possible nonexistence of M as a measure on $(B,\mathcal{U}(B'))$ as a mark of the bad quality of the estimator with limiting distribution L. In that case Theorem 3.7 may be used to say that the existence of but one regular estimator with a $\tau(B')$-tight limiting distribution, implies existence of N as a measure on $(B,\mathcal{U}(B'))$. Next it follows that any limiting distribution of a regular estimator is the convolution of N and a cylinder measure M. Now note that B can in a natural way be embedded in $\mathbb{R}^{B'}$ and that by Kolmogorov's Theorem M can always be extended to a probability measure on $(\mathbb{R}^{B'},\mathcal{U}(B'))$. Then an interpretation of the convolution theorem is that any regular estimator behaves asymptotically as the sum of the optimal estimator and some independent 'noise', which possibly takes values outside B.[2)]

Equality (3.15) completely determines a probability measure N on $\mathcal{U}(B')$. Using the fact that B' is a vector space we can see that (3.15) implies

$$(b_1',b_2',\ldots,b_k')(N) = N_k(0,(\langle\tilde{\kappa}_{b_i'}(\cdot,P),\tilde{\kappa}_{b_j'}(\cdot,P)\rangle_P)) ,$$

for every $\{b_1',\ldots,b_k'\} \subset B'$. Alternatively (3.15) can be expressed by $N = L(G)$ where G is a random element in $(B,\mathcal{U}(B'))$ with zero mean normal *marginals* $L((b_1'(G),\ldots,b_k'(G)))$ and *covariance function*

$$Eb_1'(G)b_2'(G) = E_P\tilde{\kappa}_{b_1'}(X_1,P)\tilde{\kappa}_{b_2'}(X_1,P) \qquad \{b_1',b_2'\} \subset B' .$$

PROOF OF THEOREM 3.7. For every finite set $\{b_1',b_2',\ldots,b_k'\} \subset B'$,

---

2) *Actually the second assertion of Theorem 3.7 can be proved under the weaker condition that for every sequence* $\{bi'\} \subset B'$, *there exists* $A' \subset B'$, *separating points of* B, *such that* $\mathcal{U}(b_1',b_2',\ldots) \subset \mathcal{U}(A')$ *and* L *is* $\tau(A')$*-tight.*

$(b_1',b_2',\ldots,b_k')\circ\kappa\colon P \to \mathbb{R}^k$ is differentiable with gradients $\tilde{\kappa}_{b_i'}(\cdot,P)$ $(i = 1,2,\ldots,k)$. Hence applying Theorem 2.11 to the regular estimator sequence $\{(b_1',b_2',\ldots,b_k')\circ T_n\}$, we have

$$(3.16)\qquad (b_1',b_2',\ldots,b_k')(L) = N_{b_1',b_2',\ldots,b_k'} * M_{b_1',b_2',\ldots,b_k'},$$

where

$$N_{b_1',b_2',\ldots,b_k'} = N_k(0,(<\tilde{\kappa}_{b_i'}(\cdot,P),\tilde{\kappa}_{b_j'}(\cdot,P)>_P)).$$

If $R = (\rho_{ij})$ is an arbitrary $(r\times k)$ matrix then

$$(3.17)\qquad (R(<\tilde{\kappa}_{b_i'}(\cdot,P),\tilde{\kappa}_{b_j'}(\cdot,P)>_P)R') = (<\tilde{\kappa}_{\Sigma\rho_{il}b_l'}(\cdot,P),\tilde{\kappa}_{\Sigma\rho_{jl}b_l'}(\cdot,P)>_P).$$

This implies that the system $\{N_{b_1',b_2',\ldots,b_k'}\}$ is consistent in the sense of (3.6). Next by considering characteristic functions and using (3.16) we deduce consistency of $\{M_{b_1',b_2',\ldots,b_k'}\}$. By Proposition 3.4 these systems correspond to cylinder measures N and M on $(B,A(B'))$, concluding the proof of the first assertion of the theorem.

Finally we prove that N and M are extendible to $(B,U(B'))$ under the assumption that L is $\tau(B')$-tight (cf. Araujo and Giné (1980), p.33, Ex.5). For any $\varepsilon > 0$ there exists a $\tau(B')$-compact set K_ε such that for all $\{b_1',b_2',\ldots,b_k'\} \subset B'$

$$(b_1',b_2',\ldots,b_k')(L)(\,(b_1',b_2',\ldots,b_k')K_\varepsilon\,) \geq L^*(K_\varepsilon) \geq 1-\varepsilon.$$

Let $\underline{R}_k\colon B \to \mathbb{R}^k$ denote the map $b \to (b_1'(b),b_2'(b),\ldots,b_k'(b))$. Then by (3.16)

$$(3.18)\qquad 1-\varepsilon \leq \underline{R}_k(L)(\underline{R}_kK_\varepsilon) = \int \underline{R}_k(N)(\underline{R}_kK_\varepsilon - x)\, d\underline{R}_k(M)(x).$$

Hence there exists $x_0 \in \mathbb{R}^k$ (possibly depending on $\{b_1',b_2',\ldots,b_k'\}$) with $\underline{R}_k(N)(\underline{R}_kK_\varepsilon - x_0) \geq 1-\varepsilon$. By symmetry of N $\underline{R}_k(N)(x_0-\underline{R}_kK_\varepsilon) \geq 1-\varepsilon$, so that $\underline{R}_k(N)(\underline{R}_k(\tfrac{1}{2}(K_\varepsilon-K_\varepsilon))) \geq 1-2\varepsilon$. Being the image of $K_\varepsilon\times K_\varepsilon$ under the $\tau(B')$-continuous map $(x,y) \to \tfrac{1}{2}(x-y)$, $\underline{K}_\varepsilon = \tfrac{1}{2}(K_\varepsilon-K_\varepsilon)$ is $\tau(B')$-compact. By Proposition 3.3 N is extendible to $(B,U(\tau(B')))$, hence certainly to $(B,U(B'))$.

For M we proceed in an analogous manner. First we have

$$1-\varepsilon \le \underline{R}_k(L)(\underline{R}_k K_\varepsilon) = \int \underline{R}_k(M)(\underline{R}_k K_\varepsilon - x)\, d\underline{R}_k(N)(x)$$

(3.19)

$$\le \underline{R}_k \underline{K}_\varepsilon \int \underline{R}_k(M)(\underline{R}_k K_\varepsilon - x)\, d\underline{R}_k(N)(x) + 2\varepsilon.$$

We infer existence of $x_1 \in \underline{R}_k \underline{K}_\varepsilon$ with $\underline{R}_k(M)(\underline{R}_k K_\varepsilon - x_1) \ge 1-3\varepsilon$. Hence $\underline{R}_k(M)(\underline{R}_k(K_\varepsilon - \underline{K}_\varepsilon)) \ge 1-3\varepsilon$ and a second application of Proposition 3.3 concludes the proof. ∎

3.3. LOCAL ASYMPTOTIC MINIMAX THEOREMS

The convolution theorem given above is essentially a theorem on marginal distributions. Actually the same is true for the LAM theorem below, a fact which now results from our choice of loss functions. These are introduced in two steps. First we define a class $\mathfrak{L}(B',N)$ for which the LAM theorems are proved. Next Lemma 3.10 asserts that some well-known loss functions are contained in this class.

3.3.1. LAM theorems.

Call a function $\ell_k: B \to \mathbb{R}$ a $(B'\text{-})$ *cylinder loss function* if there exist $\{b_1', \ldots, b_k'\} \subset B'$ and a bowl-shaped function $\underline{\ell}_k: \mathbb{R}^k \to \mathbb{R}$ (cf. (2.25)) such that

$$\ell_k(b) = \underline{\ell}_k(b_1'(b), \ldots, b_k'(b)). \tag{3.20}$$

Next, given a probability measure N on $(B,\mathcal{U})$ where $\mathcal{U}$ is a σ-algebra containing $\mathcal{U}(B')$, let $\mathfrak{L}(B',N)$ be the class of all functions $\ell: B \to \mathbb{R}$ for which there exists a sequence $\{\ell_k\}$ of cylinder loss functions such that

$$\begin{aligned} &\ell_k \uparrow \ell && \text{N-a.e.} \\ &\ell_k \le \ell && \text{on } B\ . \end{aligned} \tag{3.21}$$

For loss functions in $\mathfrak{L}(B',N)$ minimax theorems can be obtained as simple corollaries of the corresponding theorems for $\mathbb{R}^k$. Below E_* and $\int_*$

denote inner integral.

THEOREM 3.8. *Let* B *be a vector space and* B' *a vector space of linear, real-valued functions on* B. *Assume that* T(P) *is convex and let* κ: $\mathcal{P} \to (B,\tau(B'))$ *be differentiable at* $P \in \mathcal{P}$. *Furthermore assume existence of a measure* N *on* (B,U) *satisfying* (3.15) *for all* $b' \in B'$. *Then for any* U*-measurable* $\ell \in \mathcal{L}(B',N)$ *and every estimator sequence* $\{T_n\}$ *in* (B,U(B'))

$$\lim_{c\to\infty} \liminf_{n\to\infty} \sup_{Q\in H_n(P,c)} E_{*Q}\ell(\sqrt{n}(T_n-\kappa(Q))) \geq \int \ell(b)dN(b). \tag{3.22}$$

Here $H_n(P,c)$ *is defined as in Theorem 2.10.* □

PROOF. Let ℓ_k be cylinder loss functions satisfying (3.20)-(3.21). Then the left hand side of (3.22) is larger than

$$\lim_{c\to\infty} \liminf_{n\to\infty} \sup_{Q\in H_n(P,c)} E_Q\, \underline{\ell}_k(\sqrt{n}(b_1',\ldots,b_k')(T_n-\kappa(Q))).$$

By applying Theorem 2.10 to the estimator sequence $\{(b_1',\ldots,b_k')\circ T_n\}$ of $(b_1',\ldots,b_k')\circ\kappa$ we see that this is larger than

$$\int \underline{\ell}_k(x)\, dN_k(0,(<\tilde{\kappa}_{b_i'}(\cdot,P),\tilde{\kappa}_{b_j'}(\cdot,P)>_P))(x) = \int \underline{\ell}_k(b_1'(b),\ldots,b_k'(b))dN(b)$$

by (3.15). Finally let $k \to \infty$. ■

For some applications Theorem 3.8 has the same drawbacks as Theorem 2.10. In the following theorem the risk is considered as a supremum over finite dimensional submodels. For an arbitrary linearly independent subset $G = \{g_1,g_2,\ldots,g_m\}$ of T(P) and $\alpha \in \mathbb{R}^m$ such that $\Sigma\alpha_i g_i \in T(P)$ let $\{P_{n\alpha}^G\} \subset \mathcal{P}$ satisfy

$$\int [\sqrt{n}\,((dP_{n\alpha}^G)^{\frac{1}{2}}-(dP)^{\frac{1}{2}}) - \tfrac{1}{2}\Sigma\alpha_i g_i\,(dP)^{\frac{1}{2}}\,]^2 \to 0 \tag{3.23}$$

$$\sqrt{n}\,(\kappa(P_{n\alpha}^G)-\kappa(P)) \to \kappa_P'(\Sigma\alpha_i g_i) \qquad \text{in } \tau(B'). \tag{3.24}$$

THEOREM 3.8'. *Under the assumptions of Theorem 3.8*

$$(3.25)\qquad \sup_{G}\ \lim_{c\to\infty}\ \liminf_{n\to\infty}\ \sup_{\|\alpha\|\le c}\ E_{*\,P_{n\alpha}^{G}}\,\ell(\sqrt{n}(T_n-\kappa(P_{n\alpha}^{G}))) \ge \int \ell(b)\,dN(b)\ .$$

Here the supremum on the left contains a subset G as in (3.23)-(3.24) *for every finite subset of* $T(P)$. □

PROOF. cf. Theorem 2.9. ■

Just as in Chapter 2 we can also define LAM risk of a weakly regular estimator sequence as the maximum risk over the local limit distributions.

THEOREM 3.9. *Let* B *be a vector space and* B' *a vector space of linear, real-valued functions on* B. *Assume that* $T(P)$ *is convex and let* $\kappa\colon \mathcal{P} \to (B,\tau(B'))$ *be differentiable at* $P \in \mathcal{P}$. *Furthermore assume existence of a measure* N *on* $(B,\mathcal{U})$ *satisfying* (3.15) *for all* $b' \in B'$. *Then for any* $\mathcal{U}$*-measurable* $\ell \in \mathfrak{L}(B',N)$ *and every (at* P*) weakly regular estimator sequence* $\{T_n\}$ *in* $(B,\mathcal{U}(B'))$

$$\sup_{g\in T(P)} \int_{*} \ell(b)\ dL_g(b) \ge \int \ell(b)\ dN(b).\ \square$$

PROOF. Theorem 3.9 is a corollary of Theorem 2.8 in the same manner as Theorem 3.8 of Theorem 2.10. ■

3.3.2. Subconvex loss functions

With the aid of the Hahn-Banach Theorem we can give sufficient topological conditions for a function ℓ to be a member of $\mathfrak{L}(B^{*},N)$. Let (B,τ) be a topological vector space. A function $\ell\colon (B,\tau) \to \mathbb{R}$ is called *subconvex* if it satisfies

$$(3.26)\qquad \begin{aligned} &\ell(0) = 0 \le \ell(b)\\ &\ell(b) = \ell(-b)\\ &\{b \in B\colon \ell(b) \le c\} \text{ is convex and } \tau\text{-closed for every } c \in \mathbb{R}. \end{aligned}$$

LEMMA 3.10. *Let* $(B,\|\cdot\|)$ *be a normed space*[3] *and* B^* *its dual space. Let* $\mathcal{U}$ *be a* σ*-field containing* $\mathcal{U}(B^*)$ *and let* N *be a probability measure on* $(B,\mathcal{U})$ *such that* $N(S) = 1$ *for a* $\|\cdot\|$*-separable set* $S \in \mathcal{U}$. *Then any subconvex function* $\ell\colon (B,\|\cdot\|) \to \mathbb{R}$ *is a member of* $\mathfrak{L}(B^*,N)$. □

PROOF. We can approximate ℓ from below by elementary subconvex functions of the form

$$\ell_r(b) = 2^{-r} \sum_{j=1}^{r2^r} 1\{b\colon \ell(b) > j2^{-r}\} ,$$

$(r = 1,2,\ldots)$. Therefore it suffices to prove the theorem for ℓ of the form 1_{C^c}, where C is a $\|\cdot\|$-closed, convex set with $C = -C$. By the Hahn-Banach Theorem (cf. Rudin (1973), Th.3.4) any such C can be written as

$$C = \bigcap_{b'\in B^*} \{b\colon |b'(b)| \le \alpha_{b'}\},$$

where $\alpha_{b'} \in \bar{\mathbb{R}}$. Hence

$$C^c \cap S = \bigcup_{b'\in B^*} \{b \in S\colon |b'(b)| > \alpha_{b'}\}.$$

Since S is separable we have that the relative $\|\cdot\|$-topology on $C^c \cap S$ has a countable base, so that $C^c \cap S$ is Lindelöf (cf. Jameson (1974), Sect.10). We infer that there exists a sequence $\{b_i'\} \subset B^*$ such that

$$C^c \cap S = \bigcup_{i=1}^{\infty} \{b \in S\colon |b_i'(b)| > \alpha_{b_i'}\}.$$

Setting

$$\underline{\ell}_k(x) = \sup_{i=1..k} 1\{x \in \mathbb{R}^k\colon |x_i| > \alpha_{b_i'}\} ,$$

we have that $\underline{\ell}_k(b_1'(b),\ldots,b_k'(b)) \uparrow 1_{C^c}(b)$ on S as $k \to \infty$. Moreover $1_{C^c}(b) \ge \underline{\ell}_k(b_1'(b),\ldots,b_k'(b))$. ■

---

[3] *More generally, it suffices that* (B,τ) *is a locally convex topological vector space and that* N *concentrated on a set* S, *on which the relative* τ*-topology is hereditary Lindelöf (cf. Jameson (1974, Sect.10).*

3.4. TWO RESULTS ON THE OPTIMAL LIMITING MEASURE

The first theorem of this section characterizes, under some conditions, the support of the optimal limiting measure determined by (3.15) as the closure of the range of the derivative κ_P'.

THEOREM 3.11. *Let* $(B,\|\cdot\|)$ *be a normed space*[3] *with dual space* B^* *and let* $\kappa: \mathcal{P} \to (B,\|\cdot\|)$ *be differentiable at* $P \in \mathcal{P}$. *Let* N *be a probability measure on* $(B,\mathcal{U}(\|\cdot\|))$, *such that* $N(S) = 1$ *for a closed* $\|\cdot\|$*-separable set* S, *and such that* (3.15) *holds for every* $b' \in B^*$. *Then*

$$N\left(\overline{\kappa_P'(\text{lin } T(P))}\right) = 1 \ . \ \square$$

PROOF. As a consequence of the Hahn-Banach theorem (cf. Rudin (1973, Th.3.5))

$$\overline{\kappa_P'(\text{lin } T(P))} = \bigcap_{\substack{b' \in B^* \\ b' \circ \kappa_P' \equiv 0}} \{b: b'(b) = 0\} \ .$$

Then

$$\overline{\kappa_P'(\text{lin } T(P))} \cap S = \bigcap_{\substack{b' \in B^* \\ b' \circ \kappa_P' \equiv 0}} \{b \in S: b'(b) = 0\} \ .$$

As in the proof of Lemma 3.10, the fact that S is separable allows us to replace this intersection by a countable intersection. Thus

$$\overline{\kappa_P'(\text{lin } T(P))} \cap S = \bigcap_{i=1}^{\infty} \{b \in S: b_i'(b) = 0\} \ ,$$

for a sequence $\{b_i'\} \subset B^*$ with $\tilde{\kappa}_{b_i'}(\cdot,P) \equiv 0$.

By (3.15) $N(b: b_i'(b) = 0) = b_i'(N)\{0\} = 1$ for every i. The result follows. ■

The second result concerns relation (3.15). This completely determines a probability measure N as a measure on $(B,\mathcal{U}(B'))$. However, we ask whether (3.15) also holds for linear maps which are not members of B'. This is for instance of interest when we want to apply Theorem 3.11, but know the

validity of (3.15) only for a subset of the dual space (such as the coordinate projections in a subspace of B(T) (cf. Section. 3.5.4)).

One method to establish (3.15) for a larger class than B', would be to exhibit an appropriate estimator sequence which is regular with respect to the larger class, and next apply Theorem 3.7. The following result addresses the problem more directly and will be used in some of the proofs later on.

Assume that N is a probability measure on $(B,\mathcal{U})$, where $\mathcal{U}$ is a σ-field which contains $\mathcal{U}(B')$.

LEMMA 3.12. *Let* (B,τ) *be a topological vector space and let* $\kappa: \mathcal{P} \to (B,\tau)$ *be differentiable at* $P \in \mathcal{P}$ *with canonical gradients* $\tilde{\kappa}_{b^*}(\cdot,P)$. *Suppose* $B' \subset B_\tau^*$ *is a vector space and* N *a probability measure on* $(B,\mathcal{U})$ *such that* (3.15) *holds for all* $b' \in B'$. *Then* (3.15) *is true for all* $\mathcal{U}$*-measurable* $b^* \in B_\tau^*$ *for which there exists a sequence* $\{b_j'\} \subset B'$ *such that* $b_j'(b) \to b^*(b)$ *for all* b *in a set* $S \in \mathcal{U}$ *which contains the range of* κ_P' *and has* $N(S) = 1$. □

PROOF. Let H be the closure of $\dot{l}\text{in}\, T(P)$ in $L_2(P)$. Then H is a Hilbert space and for any $b^* \in B_\tau^*$ the extension of $b^* \circ \kappa_P'$ to H is an element of the dual space H^*. Furthermore

$$\|\tilde{\kappa}_{b^*}(\cdot,P)\|_P^2 = \|b^* \circ \kappa_P'\|_{H^*}^2. \tag{3.27}$$

Let G be a $(B,\mathcal{U})$-valued random element with $L(G) = N$. Suppose $b_j'(b) \to b^*(b)$ for all b in the support of N. Then

$$b_i'(G) - b_j'(G) \to 0 \qquad \text{a.s. if } i,j \to \infty.$$

Hence using (3.15) and (3.27)

$$N(0, \|b_i' \circ \kappa_P' - b_j' \circ \kappa_P'\|_{H^*}^2) = L((b_i' - b_j')(G)) \to \delta_0, \tag{3.28}$$

weakly as measures on $\mathbb{R}$. We conclude that $\{b_i' \circ \kappa_P'\}$ is a $\|\cdot\|_{H^*}$-Cauchy sequence in H^*. By completeness there exists $h^* \in H^*$ with $\|b_i' \circ \kappa_P' - h^*\|_{H^*} \to 0$. On the other hand we assume that $b_i' \circ \kappa_P'(h) \to b^* \circ \kappa_P'(h)$ for all $h \in T(P)$, so that $h^* = b^* \circ \kappa_P'$. Thus

(3.29) $\quad L(b_i'(G)) = N(0, \|b_i' \circ \kappa_P'\|_{H^*}^2) \to N(0, \|b^* \circ \kappa_P'\|_{H^*}^2).$

Since $b_i'(G) \to b^*(G)$ a.s., we also have

(3.30) $\quad L(b_i'(G)) \to L(b^*(G)).$

Finally conclude equality of the right hand sides of (3.29) and (3.30), and use (3.27). ■

3.5. SPECIAL CASES

In this section we specify the convolution and LAM theorems obtained in Sections 3.2 and 3.3 to some special spaces B, which are often encountered in applications. Moreover we obtain some additional results.

3.5.1. Separable normed spaces.

In this section $(B, \|\cdot\|)$ is a separable normed space and B^* its dual space.

For separable normed spaces $(B, \|\cdot\|)$ one usually considers estimators in $(B, \mathcal{U}(\|\cdot\|))$, where $\mathcal{U}(\|\cdot\|)$ is the Borel σ-algebra. *$\|\cdot\|$-Regularity* is defined by the requirement that for every $g \in T(P)$ there exists some sequence $\{P_n\} \subset \mathcal{P}$ satisfying (3.11),

(3.31) $\quad \sqrt{n}\,(\kappa(P_n)-\kappa(P)) \to \kappa_P'(g) \qquad$ in $\|\cdot\|$,

and

(3.32) $\quad L_{P_n}(\sqrt{n}(T_n-\kappa(P_n))) \to L$,

weakly in the usual sense on the metric space $(B, \|\cdot\|)$ (cf. Billingsley (1968)). Of course (3.32) implies

(3.33) $\quad L_{P_n}(\, b'[\sqrt{n}(T_n-\kappa(P_n))]\,) \to b'(L) \qquad$ for all $b' \in B^*$.

Separable normed spaces can be accomodated in the set-up of Sections 3.2-3.3 by setting B' equal to B^* in which case $\mathcal{U}(B')$ equals the Borel σ-field. The following lemma is well-known.

LEMMA 3.13. *Let* $(B,\|\cdot\|)$ *be a separable normed space. Then* $U(B^*)$ *equals the Borel* σ*-algebra on* B. □

PROOF. It suffices to show that any $\|\cdot\|$-open set is contained in $U(B^*)$. By separability every open set is the union of a countable number of open balls. Any open ball is a countable union of closed balls. By the Hahn-Banach theorem any closed ball is the intersection of closed half spaces of the form $\{b: b^*(b) \leq \alpha\}$. By separability we can replace the intersection by a countable intersection (Lindelöf property). ■

The following convolution theorem can now be obtained.

THEOREM 3.14. *Let* $(B,\|\cdot\|)$ *be a Banach space, let* $T(P)$ *be convex and let* $\kappa: \mathcal{P} \to (B,\|\cdot\|)$ *be differentiable at* $P \in \mathcal{P}$. *Suppose that* $\{T_n\}$ *is* $\|\cdot\|$*-regular as described above with limiting distribution* L. *Then there exists a Gaussian measure* N *on* $(B,U(\|\cdot\|))$ *satisfying* (3.15) *for every* $b' \in B^*$. *Moreover*

$$L = N*M \qquad \text{on } (B,U(\|\cdot\|)) \ . \ \square$$

PROOF. It was already noted that (3.32) implies (3.33). Furthermore, every probability measure on the Borel σ-algebra of a Banach space is tight with respect to the norm topology (cf. Parthasarathy (1967, Th.3.2)), hence certainly tight with respect to the weak topology. Apply Theorem 3.7 with B' equal to B^*. ■

For separable Banach spaces (3.32) is equivalent to (3.33) together with $\|\cdot\|$-tightness of the sequence of measures $\{L_{P_n}(\sqrt{n}(T_n-\kappa(P_n)))\}$. Actually the proof of Theorem 3.14 relies on (3.33) only. It may be considered surprising, or disappointing, that the convolution theorem is not related to tightness of $\{L_{P_n}(\sqrt{n}(T_n-\kappa(P_n)))\}$. Still, tightness of its sequence of laws is clearly a good property of an estimator sequence. In Chapter 4 we therefore include tightness of the sequence of measures $\{L_{P_n}(\sqrt{n}(T_n-\kappa(P_n)))\}$ on $(B,U(\|\cdot\|))$ in the definition of efficiency.

We conclude this sub-section with a LAM theorem. Here we assume existence of a measure N on the Borel σ-field, satisfying (3.15) for every

b' from the dual space. By Theorems 3.14 and 3.7, a convenient method to establish this, is to exhibit a tight $\|\cdot\|$-regular estimator sequence for κ.

THEOREM 3.15. *Let* $(B,\|\cdot\|)$ *be a separable normed space, let* $T(P)$ *be convex and let* $\kappa\colon \mathcal{P} \to (B,\|\cdot\|)$ *be differentiable at* $P \in \mathcal{P}$. *Furthermore assume the existence of a measure* N *on* $(B,\mathcal{U}(\|\cdot\|))$ *satisfying* (3.15) *for every* $b' \in B^*$. *Then for any subconvex loss function* $\ell\colon (B,\|\cdot\|) \to \mathbb{R}$ *and any estimator sequence* $\{T_n\}$ *in* $(B,\mathcal{U}(\|\cdot\|))$

$$\lim_{c\to\infty} \liminf_{n\to\infty} \sup_{Q\in H_n(P,c)} E_Q \ell(\sqrt{n}(T_n - \kappa(Q))) \geq \int \ell(b)\,dN(b).$$

Moreover for the limiting distributions of a weakly regular estimator sequence in $(B,\mathcal{U}(\|\cdot\|))$

$$\sup_{g\in T(P)} \int \ell(b)\, dL_g(b) \geq \int \ell(b)\, dN(b)\ . \ \square$$

PROOF. The theorem follows from Theorems 3.8-3.9, Lemma 3.10 and Lemma 3.13. ■

3.5.2. The space D[a,b] with the projection σ-field.

Given $-\infty \leq a \leq b \leq \infty$ let $D = D[a,b]$ be the space of all right continuous functions $d\colon [a,b] \to \mathbb{R}$ with left limits everywhere in $(a,b]$. Let $\pi_t\colon D \to \mathbb{R}$ denote the coordinate projections

$$\pi_t d = d(t)\ , \qquad t \in [a,b].$$

Let Π be the linear space spanned by the coordinate projections.

We consider estimator sequences in $(D,\mathcal{U}(\Pi))$. Here $\mathcal{U}(\Pi) = \mathcal{U}(\pi_t\colon t \in [a,b])$ is the *projection σ-field*. Let $\kappa\colon \mathcal{P} \to (D,\tau(\Pi))$ be differentiable at $P \in \mathcal{P}$ with gradients $\dot\kappa_{d*}(\cdot,P)$. An estimator sequence $\{T_n\}$ in $(D,\mathcal{U}(\Pi))$ is Π-*regular at* $P \in \mathcal{P}$ if for every $g \in T(P)$ there exists $\{P_n\} \subset \mathcal{P}$ satisfying (3.11),

$$\sqrt{n}\,(\pi_u\circ\kappa(P_n) - \pi_u\circ\kappa(P)) \to \int \dot\kappa_{\pi_u}(x,P)\, g(x)\, dP(x) \qquad \text{all } u \in [a,b] \tag{3.34}$$

and convergence of marginals

(3.35) $$L_{P_n}(\pi_{u_1\dots u_m}\sqrt{n}(T_n-\kappa(P_n))) \to \pi_{u_1\dots u_m}L\ ,$$

where $\pi_{u_1\dots u_m}: D \to \mathbb{R}^k$ is the evaluation map $\pi_{u_1\dots u_m}d = (d(u_1),\dots,d(u_m))$ and L is a probability measure on $(D,\mathcal{U}(\Pi))$.

The convolution theorem now takes the form of

THEOREM 3.16. *Assume that* T(P) *is convex and that* $\kappa: P \to (D,\tau(\Pi))$ *is differentiable at* $P \in P$. Let $\{T_n\}$ *be* Π*-regular at* $P \in P$ *with limiting distribution* L. *Then there exists a probability distribution* $N = L(G)$ *on* $(D,\mathcal{U}(\Pi))$ *with zero-mean normal marginals* $L(G(t_1),..,G(t_m))$ *and covariance function*

(3.36) $$EG(u)G(v) = <\tilde{\kappa}_{\pi_u}(\cdot,P),\ \tilde{\kappa}_{\pi_v}(\cdot,P)>_P \qquad u,v \in [a,b].$$

Moreover

$$L = N * M \qquad \text{on } (D,\mathcal{U}(\Pi)).\ \square$$

PROOF. Theorem 3.16 differs from Theorem 3.7 in that it asserts existence of N and M as probability measures on $(D,\mathcal{U}(\Pi))$, an assertion which is warranted in Theorem 3.7 only under a tightness condition on L. However we can apply Theorem 3.7 to a suitable subset of Π, for which the tightness condition is automatically satisfied. The details are as follows.

Let J denote the Skorohod topology on D and $\mathcal{U}(J)$ its Borel σ-field. Then $\mathcal{U}(J) = \mathcal{U}(\pi_u: u \in U)$ for every $U \subset [a,b]$ which is dense in [a,b] and contains b. Also, given an arbitrary probability measure L on $(D,\mathcal{U}(J))$, for all except possibly countably many $u \in (a,b)$

(3.37) $$L(\{d \in D[a,b]: d \text{ is continuous at } u\}) = 1,$$

(cf. Billingsley (1968), Th. 14.5 and p.124).

From the set of $u \in [a,b]$ satisfying (3.37) extract a countable dense subset U containing b. Next let D_U be the set of $d \in D[a,b]$ that are continuous at all $u \in U$ and let $B' = \text{lin}\{\pi_u: u \in U\}$. Then $D_U \in \mathcal{U}(J)$ and $L(D_U) = 1$. Combining this with the fact that the Skorohod topology is Polish, we infer existence of a J-compact set $K_\varepsilon \subset D_U$ with $L(K_\varepsilon) \geq 1-\varepsilon$ (cf. Parthasarathy (1967), Th.3.1). Now $d_n \to d \in D_U$ with respect to J implies

$d_n(u) \to d(u)$ for every $u \in U$. The latter is equivalent to $d_n \to d$ with respect to $\tau(B')$. It follows for the relative topologies that $\tau(B') \cap D_U \subset \tau(J) \cap D_U$, so that K_ε is $\tau(B')$-compact too. We conclude that L is $\tau(B')$-tight.

Suppose that L is the limit distribution of a Π-regular estimator. Apply Theorem 3.7 with B' as above to find that $L = N * M$ on $(B, \mathcal{U}(J))$. Here we know from Theorem 3.7 that N satisfies (3.15) for all $b' \in B'$. By Lemma 3.12, choosing $\tau = \tau(\Pi)$, we can conclude that (3.15) must hold for all $b' \in \Pi$. ■

Next we discuss the LAM theorem. Let $\|\cdot\|_\infty$ be the uniform metric on D,

$$\|d\|_\infty = \sup_{t \in [a,b]} |d(t)| .$$

Any measure N on $(D, \mathcal{U}(\Pi))$ for which there exists a $\|\cdot\|_\infty$-separable, closed set $S \in \mathcal{U}(\Pi)$ such that $N(S) = 1$, has a unique extension to the Borel σ-field $\mathcal{U}(\|\cdot\|_\infty)$ (cf. Gaenssler (1983, p.44)). Call such a measure *separable.*

THEOREM 3.17. *Assume that* T(P) *is convex and let* $\kappa: \mathcal{P} \to (D, \|\cdot\|_\infty)$ *be differentiable at* $P \in \mathcal{P}$. *Furthermore assume the existence of a separable measure* $N = L(G)$ *on* $(D, \mathcal{U}(\|\cdot\|_\infty))$ *with zero mean normal marginals and covariance function given by* (3.36). *Then for any subconvex loss function* $\ell: (D, \|\cdot\|_\infty) \to \mathbb{R}$ *and every estimator sequence* $\{T_n\}$ *in* $(D, \mathcal{U}(\Pi))$

$$\lim_{c \to \infty} \liminf_{n \to \infty} \sup_{Q \in H_n(P,c)} E_{*Q} \ell(\sqrt{n}(T_n - \kappa(Q))) \geq \int \ell(b) dN(b). \quad \square$$

PROOF. Let D^* be the dual space of $(D, \|\cdot\|_\infty)$. It is shown in Lemma 3.26 below that for every $d^* \in D^*$ there exists a sequence $\{d_n'\}$ of the form

$$d_n' = \Sigma_{i=1}^{n} \alpha_{ni} \pi_{t_{ni}} ,$$

such that

$$d^*(d) = \lim_{n \to \infty} d_n'(d) \qquad \text{every } d \in D.$$

As a first consequence of this we have $\mathcal{U}(D^*) = \mathcal{U}(\Pi)$. Next combination with

Lemma 3.12 and (3.36) yields

$$d^*(N) = N(0, \|\tilde{\kappa}_{d^*}(\cdot,P)\|_P^2) \qquad \text{every } d^* \in D^* .$$

By Lemma 3.10 any subconvex ℓ is a member of $\mathfrak{L}(D^*,N)$. The result follows from Theorem 3.8, applied with $B' = D^*$ and $\mathcal{U}$ equal to the $\|\cdot\|_\infty$-Borel σ-field. ■

Examples of subconvex functions $\ell:(D,\|\cdot\|_\infty) \to \mathbb{R}$ are

$$\ell(d) = \ell_1(\|d\|_\infty)$$

$$\ell(d) = \ell_1(\sup_t |q^{-1}(t)d(t)|)$$

$$\ell(d) = \int d^2(t)\, d\mu(t) ,$$

where $\ell_1:[0,\infty) \to [0,\infty)$ is lower semi-continuous and nondecreasing with $\ell_1(0) = 0$, q is arbitrary, and μ a finite, positive Borel measure on $[a,b]$.

The optimal limiting measure N seems usually to be separable. Therefore Theorem 3.17 is satisfactory for most situations. However, also if N is not separable $\mathfrak{L}(D^*,N)$[4] contains interesting loss functions. For instance the first of the above examples is contained in $\mathfrak{L}(D^*,N)$ for any nondecreasing function ℓ_1: $[0,\infty) \to [0,\infty)$ with $\ell_1(0) = 0$, for which $N(b: \|b\|_\infty = c) = 0$ for every discontinuity point c of ℓ_1.

3.5.3. D[a,b] with Skorohod topology

Under its Skorohod topology J (cf. Billingsley (1968)) D[a,b] is not a topological vector space: addition $(d_1,d_2) \to d_1+d_2$ is not $J\times J$-J continuous. This fact causes some difficulty when obtaining a convolution theorem, but these can be overcome. Since the Skorohod Borel σ-field $\mathcal{U}(J)$ equals the projection σ-field $\mathcal{U}(\Pi)$ (cf. Gaenssler (1983, p.91), addition is $\mathcal{U}(J)\times\mathcal{U}(J)$-$\mathcal{U}(J)$ measurable, and convolution on $\mathcal{U}(J)$ well-defined.

The following convolution theorem is closely related to Theorem 3.16. The difference is that convergence of marginals of the estimator sequence

4) *We consider* $\mathfrak{L}(D^*,N)$ *istead of* $\mathfrak{L}(\Pi,N)$ *because* $\mathcal{U}(D^*) = \mathcal{U}(\Pi)$ *(cf. Lemma 3.26).*

is replaced by weak convergence of $L_{P_n}(\sqrt{n}(T_n - \kappa(P_n)))$ with respect to the Skorohod topology (cf. Billingsley (1968)). Call an estimator sequence $\{T_n\}$ in $(D,U(J))$ for a differentiable functional $\kappa: \mathcal{P} \to (D,\tau(\Pi))$ *Skorohod-regular at* $P \in \mathcal{P}$ if for every $g \in T(P)$ there exists a sequence $\{P_n\} \subset \mathcal{P}$ satisfying (3.11),

$$\sqrt{n}\,(\pi_u \circ \kappa(P_n) - \pi_u \circ \kappa(P)) \to \int \tilde{\kappa}_{\pi_u}(x,P)\, g(x)\, dP(x) \qquad \text{all } u \in [a,b]$$

and

$$(3.38) \qquad L_{P_n}(\sqrt{n}(T_n - \kappa(P_n))) \to L\ ,$$

weakly with respect to the Skorohod topology.

THEOREM 3.18. *Assume that* $T(P)$ *is convex and let* $\kappa: \mathcal{P} \to (D,\tau(\Pi))$ *be differentiable at* $P \in \mathcal{P}$. *Furthermore let* $\{T_n\}$ *be an estimator sequence in* $(D,U(J))$ *which is Skorohod-regular at* $P \in \mathcal{P}$. *Then there exists a probability distribution* $N = L(G)$ *on* $(D,U(J))$ *with zero-mean normal marginals* $L(G(t_1),..,G(t_m))$ *and covariance function given by* (3.36). *Moreover*

$$L = N * M \qquad \text{on } (D,U(J)).\ \square$$

PROOF. Let B' be the linear space spanned by all L-a.e. J-continuous coordinate projections π_u. Then (3.38) implies for all $b' \in B'$

$$L_{P_n}(\ b'[\sqrt{n}(T_n - \kappa(P_n))]\ ) \to b'(L).$$

It is well-known that B' excludes at most countably many π_u, contains π_b and, moreover, $U(B') = U(J)$ (cf. Billingsley (1968), Th.14.5 and p.124). Finish the proof as the proof of Theorem 3.16. ■

3.5.4. The space B(T).

Given an arbitrary index set T, let B(T) be the set of all functions $h: T \to \mathbb{R}$ with

$$\|h\|_\infty := \sup\{|h(t)|: t \in T\} < \infty\ .$$

This space has gained importance as the support of empirical processes indexed by sets or functions, which can be viewed as estimators of an underlying probability measure (See Section 3.6.2).

The choice of a σ-field in this example is not obvious. On the one hand the Borel σ-field $\mathcal{U}(\|\cdot\|_\infty)$ is too large for most applications. On the other hand a random element Z in B(T) is usually viewed as a T-indexed process, meaning that the coordinates Z(t) are measurable maps in $\mathbb{R}$. Hence it is natural to choose a σ-field which is intermediate between $\mathcal{U}(\|\cdot\|_\infty)$ and the projection σ-field $\mathcal{U}(\Pi)$. Here $\Pi = \text{lin}\,\{\pi_t\colon t \in T\}$ and $\pi_t h = h(t)$ for every $t \in T$ and $h \in B(T)$.

Of course the larger the σ-field, the stronger the convolution statement. Here we restrict ourselves to the projection σ-field. This seems to be small, but turns out to contain every interesting set under an additional tightness assumption on the limiting measures.

The definition of a Π-regular estimator sequence (cf. Section 3.1.3) is similar to the one in Section 3.5.2. Call a probability measure L on $(B(T),\mathcal{U}(\Pi))$ $\|\cdot\|_\infty$-*tight*, if to every $\varepsilon > 0$ there exists a $\|\cdot\|_\infty$-compact set K_ε with $L^*(K_\varepsilon) \geq 1-\varepsilon$.

THEOREM 3.19. *Assume that* T(P) *is convex and that* $\kappa\colon \mathcal{P} \to (B(T),\tau(\Pi))$ *is differentiable at* $P \in \mathcal{P}$. *Let* $\{T_n\}$ *be* Π*-regular at* $P \in \mathcal{P}$ *with limiting distribution* L. *Then there exists a probability distribution* N = L(G) *on* $(B(T),\mathcal{U}(\Pi))$ *with zero-mean normal marginals* $L(G(t_1),..,G(t_m))$ *and covariance function*

$$EG(u)G(v) = \langle \tilde{\kappa}_{\pi_u}(\cdot,P),\ \tilde{\kappa}_{\pi_v}(\cdot,P) \rangle_P \qquad u,v \in T\ .$$

Moreover

$$L = N * M \qquad \text{(3.38)}$$

on $(B(T),\mathcal{U}(\Pi))$.

Finally if L *is* $\|\cdot\|_\infty$-*tight, then* N *and* M *are* $\|\cdot\|_\infty$-*tight too.* □

PROOF. Given $k \in \mathbb{R}^+$ set $K_k = [-k,k]^T$. Then $K_k \subset B(T)$ and $K_k \uparrow B(T)$ as $k \to \infty$. As $\tau(\Pi)$ induces the product topology on K_k, K_k is $\tau(\Pi)$-compact by Tychonov's theorem. It follows that any probability measure on $(B(T),\mathcal{U}(\Pi))$ is $\tau(\Pi)$-tight. Now apply Theorem 3.7.

The last assertion follows from the proof of Theorem 3.7 and the last assertion of Proposition 3.3. ■

Consider the situation of $\|\cdot\|_\infty$-tight limiting measures. Such measures concentrate on a $\|\cdot\|_\infty$-σ-compact subset of B(T). Lemma 3.21 asserts that the $\|\cdot\|_\infty$-Borel σ-field and the projection σ-field have equal trace on every $\|\cdot\|_\infty$-σ-compact set. In view of this, the convolution statement (3.38) is satisfactory for $\|\cdot\|_\infty$-tight measures L and N.

To underpin this we note the possibility to compare their concentration near the origin in the spirit of an Anderson's Lemma for B(T).

THEOREM 3.20. *Let* L, N *and* M *be* $\|\cdot\|_\infty$*-tight measures on* $(B(T),U(\Pi))$, *where* N *has zero-mean normal marginals and* L = N*M *on* $(B(T),U(\Pi))$. *Then there exist unique* $\|\cdot\|_\infty$*-tight extensions* $\tilde{L}$, $\tilde{N}$ *of* L *and* N *to* $U(\|\cdot\|_\infty)$. *Moreover*

$$\int \ell(x)\, d\tilde{L}(x) \geq \int \ell(x)\, d\tilde{N}(x) ,$$

for every subconvex ℓ: $(B(T),\|\cdot\|_\infty) \to \mathbb{R}$. □

The proof is based on the following lemma. Let $B(T)^*$ be the $\|\cdot\|_\infty$-dual space of B(T).

LEMMA 3.21. *Let* $S \subset B(T)$ *be* $\|\cdot\|_\infty$*-σ-compact. Then for every* $b^* \in B(T)^*$, *there exists a sequence* $\{b_n'\}$ *of the form* $b_n' = \Sigma_{i=1}^{k_n} \alpha_{ni}\, \pi_{t_{ni}}$ *such that*

$$(3.39) \qquad b^*(b) = \lim_{n\to\infty} b_n'(b) \qquad \text{for every } b \in S .$$

Moreover

$$(3.40) \qquad U(\|\cdot\|_\infty) \cap S = U(\Pi) \cap S. \ \square$$

PROOF. See Section 3.8. ■

PROOF OF THEOREM 3.20. Let S be $\|\cdot\|_\infty$-σ-compact and such that $L^*(S) = N^*(S) = M^*(S) = 1$. Let $A \in U(\|\cdot\|_\infty)$. By (3.40) there exists $A' \in U(\Pi)$ with $A \cap S = A' \cap S$. Set

$$\tilde{L}(A) = L(A').$$

This is well-defined because $L^*(S) = 1$. Extend N and M in a similar manner.

Any $\|\cdot\|_\infty$-tight extensions are equal, because if S is σ-compact and supports both $\tilde{L}_1$ and $\tilde{L}_2$, then $\tilde{L}_i(A) = \tilde{L}_i(A \cap S) = \tilde{L}_i(A') = L(A')$ for every $A \in \mathcal{U}(\|\cdot\|_\infty)$.

Next let $b^* \in B(T)^*$ arbitrary and let $\{b_n'\}$ satisfy (3.39). Then $b_n'(\tilde{L}) \xrightarrow{W} b^*(\tilde{L})$ and similarly for $\tilde{N}$ and $\tilde{M}$. Because of the special form of b_n' and (3.38) $b_n'(\tilde{L}) = b_n'(\tilde{N}) * b_n'(\tilde{M})$ for all n. We conclude

$$b^*(\tilde{L}) = b^*(\tilde{N}) * b^*(\tilde{M}) \quad . \tag{3.41}$$

Moreover $b^*(\tilde{N})$ is zero-mean normal. By Lemma 3.10 any subconvex loss function ℓ is a member of $\mathfrak{L}(B(T)^*, \tilde{N})$. Let $\{\ell_k\}$ be as in (3.21). Then

$$\int \ell(b)\, d\tilde{L}(b) \geq \int \ell_k(b)\, d\tilde{L}(b) = \int_{\mathbb{R}^k} \underline{\ell}_k(x)\, d(b_1', \ldots, b_k')(\tilde{L})(x)$$

which, by Anderson's Lemma for $\mathbb{R}^k$ and (3.41), is larger than

$$\int_{\mathbb{R}^k} \underline{\ell}_k(x)\, d(b_1', \ldots, b_k')(\tilde{N})(x) = \int \ell_k(b)\, d\tilde{N}(b) \quad .$$

Finally let $k \to \infty$. ∎

Finally consider the LAM Theorem. In general it will be too much to ask that an estimator sequence is measurable with respect to $\mathcal{U}(B(T)^*)$. Therefore, Lemma 3.10 applied with $\|\cdot\|_\infty$ is of little help for exhibiting loss functions to which the LAM theorems in Section 3.4 apply. However, loss functions may be shown to belong to $\mathfrak{L}(\Pi, N)$ either by direct arguments, or by applying Lemma 3.10 to B(T) with τ equal to the weak topology generated by the coordinate projections (cf. the footnote to Lemma 3.10). As an example we have

LEMMA 3.22. *Let* $\ell_1: [0,\infty) \to [0,\infty)$ *be non-decreasing and lower semi-continuous with* $\ell_1(0) = 0$. *Then the function* $\ell: b \to \ell_1(\|b\|_\infty)$ *is contained in* $\mathfrak{L}(\Pi, N)$ *for every* $\|\cdot\|_\infty$*-tight probability measure* N *on* $(B(T), \mathcal{U}(\|\cdot\|_\infty))$. □

PROOF. By lower semi-continuity of ℓ_1

$$\{b: \ell(b) \le c\} = \{b: \|b\|_\infty \le d\} = \bigcap_{t\in T} \{b: |\pi_t b| \le d\} .$$

We conclude that ℓ: $(B(T),\tau(\Pi)) \to \mathbb{R}$ is subconvex. Furthermore $(B(T),\tau(\Pi))$ is locally convex with dual space Π (cf. Rudin (1973, Th.3.10)), and N concentrates on a set S which is hereditary Lindelöf in the norm topology, hence certainly in the weaker $\tau(\Pi)$-topology (cf. Jameson (1974, Sect.10)). Now see the footnote to Lemma 3.10. ■

3.5.5. Separable Hilbert space.

Separable Hilbert space is among the simplest infinite dimensional generalizations of Euclidean space. In analogy to the finite dimensional situation Gaussian measures on such a space are often given through their mean vector and *covariance operator*. There is a simple relation between the covariance operator of the optimal limiting measure N and the derivative of the functional κ.

Assume existence of N satisfying (3.15) on the Borel σ-field of a separable Hilbert space H, i.e. $N = L(G)$ where $\langle G,h\rangle_H$ is a one-dimensional zero-mean normal variable with variance $\|\tilde{\kappa}_{\langle\cdot,h\rangle}(\cdot,P)\|_P^2$. Since

$$\langle \kappa_P'(g),h\rangle_H = \langle g,\ (\kappa_P')^* h\rangle_P ,$$

we have

$$\tilde{\kappa}_{\langle\cdot,h\rangle}(\cdot,P) = (\kappa_P')^* \circ h \qquad (h \in H) .$$

Here we identify a Hilbert space and its dual as usual, and $(\kappa_P')^*$ is the adjoint of κ_P': lin $T(P) \to H$. Then

$$\int_H \langle h,h_1\rangle_H \langle h,h_2\rangle_H \, dN(h)$$

$$= \langle (\kappa_P')^* h_1, (\kappa_P')^* h_2\rangle_P = \langle \kappa_P' \circ (\kappa_P')^* h_1, h_2\rangle_H \qquad (h_1,h_2 \in H) .$$

This shows that $\kappa_P' \circ (\kappa_P')^*$ is the covariance operator of N. It is well-known that this must be a *kernel operator* (i.e. have finite trace) to ensure existence of N (cf. Gihman and Skorohod (1974)).

3.6. EXAMPLES OF DIFFERENTIABLE FUNCTIONALS

3.6.1. Distribution Function on the Real Line.

Let $(X,\mathcal{B}) = (\mathbb{R},\mathcal{B})$ and for a probability measure P on $\mathbb{R}$ let $\kappa(P)$ be its cumulative distribution function. It is easily checked that $\kappa\colon \mathcal{P} \to (D[-\infty,\infty], \|\cdot\|_\infty)$ is differentiable at all $P \in \mathcal{P}$ with derivative

$$\kappa_P'(g) = \int 1_{[-\infty,\cdot]}(x)\, g(x)\, dP(x).$$

It follows that gradients of κ in the directions π_u are given by

$$\dot{\kappa}_{\pi_u}(x,P) = 1_{[-\infty,u]}(x). \tag{3.42}$$

The best limiting distribution N, given by (3.15), is the distribution of a zero-mean Gaussian process G with covariance function

$$EG(u)G(v) = \int \dot{\Pi}_P 1_{[-\infty,u]}(x)\, \dot{\Pi}_P 1_{[-\infty,v]}(x)\, dP(x). \tag{3.43}$$

Here $\dot{\Pi}_P$ is the orthogonal projection of $L_2(P)$ onto the closure of lin T(P). If $\mathcal{P}$ is the set of all distributions on $\mathbb{R}$ then we may take

$$T(P) = \{g \in L_2(P)\colon \int g(x)\, dP(x) = 0\} \tag{3.44}$$

In this case (3.43) reduces to (writing F(u) for $P(-\infty,u]$)

$$EG(u)G(v) = F(u\wedge v) - F(u)F(v),$$

the covariance function of $B\circ F$, where B is Brownian Bridge on [0,1]. This is the limit distribution of the empirical process $\sqrt{n}(\hat{F}_n - F)$ in an appropriate sense. After defining efficiency in Chapter 4, we may conclude that the empirical distribution function is an efficient estimator for κ if P is completely unknown.

It is well-known that the distribution of $B\circ F$ exists as a measure on $(D[-\infty,\infty], \mathcal{U}(\pi_u\colon u \in [-\infty,\infty]))$. We note that this is true for the measure N in this example no matter the tangent cone T(P). This follows since for any $u \le v \le w$

$$E(G(w)-G(v))^2(G(v)-G(u))^2 \le \{E(G(w)-G(v))^4\ E(G(v)-G(u))^4\}^{\frac{1}{2}}$$

$$= 3E(G(w)-G(v))^2E(G(v)-G(u))^2 = 3E_P(\mathbb{H}_P 1_{(v,w]}(X_1))^2E_P(\mathbb{H}_P 1_{(u,v]}(X_1))^2$$

$$\le 3(F(w)-F(v))(F(v)-F(u)).$$

Next see Billingsley (1968), p.133, (15.39). □

3.6.2. Set-Indexed Distributions

Cumulative distribution functions are convenient tools to summarize distributions on $\mathbb{R}$. The following set-up is convenient for studying distributions on $\mathbb{R}^k$ or abstract spaces.

Let $(X,\mathcal{B})$ be a measure space and let $\mathcal{C} \subset \mathcal{B}$. For $\mathcal{P}$ a set of probability measures on $(X,\mathcal{B})$, define κ: $\mathcal{P} \to (B(\mathcal{C}),\|\cdot\|)$ in a natural way by

$$\kappa(P)(C) = P(C) \qquad\qquad C \in \mathcal{C}\ .$$

(See Section 3.5.4 for a discussion of $B(\mathcal{C})$). If $X = \mathbb{R}^k$ then e.g. choose $\mathcal{C}$ equal to {all closed balls} or {all intervals of the form [a,b]}, etc.

As in Example 3.6.1 it is easily checked that κ is differentiable with derivative

$$\kappa_P'(g)(C) = \int 1_C(x)\ g(x)\ dP(x).$$

A gradient in the direction π_C is 1_C and the best limiting distribution N is the distribution of a zero-mean, $\mathcal{C}$-indexed Gaussian process G with covariance function

$$(3.45) \qquad EG(C_1)G(C_2) = \int \mathbb{H}_P 1_{C_1}(x)\ \mathbb{H}_P 1_{C_2}(x)\ dP(x).$$

Under (3.44) this reduces to

$$EG(C_1)G(C_2) = P(C_1 \cap C_2) - P(C_1)P(C_2),$$

the covariance of $\mathcal{C}$-*Indexed* P-*Bridge* (cf. Pollard (1984)).

It does not follow from our theorems that $L(G)$ exists as a probability

measure on $(B(T),U(\Pi))$. Indeed, it can be shown that C has to satisfy certain conditions for existence of $L(G)$. However Theorem 3.19 guarantees the existence of $L(G)$ on $(B(C),U(\Pi))$ if there exists a Π-regular estimator sequence for κ. Under additional conditions, which can be conveniently cast in the form of covering integrals (cf. Dudley (1984), Pollard (1984), Gaenssler (1983)), it is possible to prove that *the empirical C-process*

$$\sqrt{n}(\hat{P}_n(C)-P(C)) = n^{-\frac{1}{2}}\Sigma_{i=1}^{n} (\delta_{X_i}(C)-P(C)),$$

converges weakly in the sense of Dudley (1966,1967) (also cf. Chapter 4) to C-indexed Brownian P-Bridge. In Chapter 4 this will be taken as saying that the empirical C-process is efficient as an estimator of κ.

For special choices of C it may be fruitful to consider the functional κ as a functional in a subset of $B(C)$. Usually though, to ensure existence of regular estimator sequences, this subset can not be taken separable under the supremum metric. Hence this example can usually not be dealt with in the context of a separable normed space. □

3.6.3. Semi-Parametric Models I

This example considers the situation of Begun et al.(1983), though in a slightly different notation.

Let $\theta \subset \mathbb{R}^k$ be open and H a collection of densities with respect to a σ-finite measure ν on $\mathbb{R}$. For every pair $(\theta,\eta) \in \theta\times H$ let $P_{\theta\eta}$ be a measure on $(X,\mathcal{B})$ with a density $p(\cdot,\theta,\eta)$ with respect to a σ-finite measure μ. We let $\mathcal{P} = \{P_{\theta\eta}: (\theta,\eta) \in \theta\times H\}$ and consider $\kappa: \mathcal{P} \to D[-\infty,\infty]$ given by

$$\kappa(P_{\theta\eta}) = \int_{[-\infty,\cdot]} \eta(z)\, d\nu(z).$$

Suppose there exists $\ell(\cdot,\theta,\eta) \in L_{2*}(P_{\theta\eta})^k$ and a continuous, linear operator $A = A(\theta,\eta): L_{2*}(\eta) \to L_{2*}(P_{\theta\eta})$ such that for every $h \in \mathbb{R}^k$

$$\int [t^{-1}(p^{\frac{1}{2}}(x,\theta+th,\eta_t)-p^{\frac{1}{2}}(x,\theta,\eta)) \tag{3.46}$$
$$- \tfrac{1}{2}(h'\ell(x,\theta,\eta)+A\underline{b}(x))\, p^{\frac{1}{2}}(x,\theta,\eta)]^2\, d\mu(x) \to 0,$$

whenever $t \downarrow 0$ and

$$\int [t^{-1}(\eta_t^{\frac{1}{2}}(z)-\eta^{\frac{1}{2}}(z)) - \tfrac{1}{2}\underline{b}(z)\eta^{\frac{1}{2}}(z)]^2 \, d\nu(z) \to 0. \tag{3.47}$$

Then if $T(\eta,H)$ is a cone of $\underline{b}$ in $L_{2*}(\eta) = \{\underline{b} \in L_2(\eta): \int \underline{b}\eta \, d\nu = 0\}$ satisfying (3.46) for some sequence $\{\eta_t\} \subset H$, we may choose

$$T(P_{\theta\eta}) = \{h'\dot{\ell}(x,\theta,\eta)+A\underline{b}(x): h \in \mathbb{R}^k, \ \underline{b} \in T(\eta,H)\}. \tag{3.48}$$

(It should be noted that (3.47) severely restricts the set of sequences $\{\eta_t\}$. In fact the maximal tangent cone $T_m(P_{\theta\eta})$ (cf. p.18) may be considerably larger than (3.48)).

Analogous to Example 2.3 we define b_0 as the vector of orthogonal projections of $\dot{\ell}(\cdot,\theta,\eta)$ onto the closure of lin $AT(\eta,H)$ in $L_{2*}(P_{\theta\eta})$ and set

$$\tilde{\ell}(\cdot,\theta,\eta) = \dot{\ell}(\cdot,\theta,\eta) - b_0(\cdot). \tag{3.49}$$

We assume that $\tilde{I}(\theta,\eta) = \int \tilde{\ell}(x,\theta,\eta)\tilde{\ell}^2(x,\theta,\eta)' \, dP_{\theta\eta}(x)$ is nonsigular. Setting $H(u) = {}_{(-\infty,u]}\int \eta(z) \, d\nu(z)$ we have by (3.47)

$$t^{-1}(\kappa(P_{\theta+t,\eta_t})-\kappa(P_{\theta\eta})) \to \int (1_{(-\infty,\cdot]}(z)-H(\cdot)) \, \underline{b}(z) \, \eta(z) \, d\nu(z) \tag{3.50}$$

in $\|\cdot\|_\infty$. Now let $A^*: L_{2*}(P_{\theta\eta}) \to L_{2*}(\eta)$ be the adjoint of A and suppose that $A^*A: L_{2*}(\eta) \to L_{2*}(\eta)$ is one-to-one and onto. Then

$$\begin{aligned} &< 1_{(-\infty,\cdot]}-H(\cdot) \, , \, \underline{b} >_\eta = < A^*A(A^*A)^{-1}(1_{(-\infty,\cdot]}-H(\cdot)) \, , \, \underline{b} >_\eta \\ &= < A(A^*A)^{-1}(1_{(-\infty,\cdot]}-H(\cdot)) \, , \, A\underline{b} >_{P_{\theta\eta}} \qquad (3.51) \\ &= \int G_{\theta\eta}(x,\cdot) \, (h'\dot{\ell}(x,\theta,\eta) + A\underline{b}(x)) \, dP_{\theta\eta}(x). \end{aligned}$$

Here

$$\begin{aligned} G_{\theta\eta}(x,u) &= A(A^*A)^{-1}(1_{(-\infty,u]}(x)-H(u)) \\ &\quad - <A(A^*A)^{-1}(1_{(-\infty,u]}(\cdot)-H(u)),\dot{\ell}(\cdot,\theta,\eta)>'_{P_{\theta\eta}} \cdot \tilde{I}^{-1}(\theta,\eta)\tilde{\ell}(x,\theta,\eta). \end{aligned} \tag{3.52}$$

Relations (3.50) -(3.52) prove differentiability of $\kappa: P \to (D[-\infty,\infty], \tau(\pi_u: u \in [-\infty,\infty]))$ with derivative

(3.53) $\quad \kappa'_P(g) = \int G_{\theta\eta}(x,\cdot)\, g(x)\, dP_{\theta\eta}(x).$

Hence gradients are given by

(3.54) $\quad \dot{\kappa}_{\pi_u}(\cdot, P_{\theta\eta}) = G_{\theta\eta}(\cdot, u).$

If $T(\eta, H) = \{\underline{b} \in L_2(\eta): \int \underline{b}\eta \; d\nu = 0\}$, then the gradients given by (3.54) are canonical. Also, in this case

$$<A^*\dot{\ell}(\cdot,\theta,\eta), \underline{b}>_\eta = <\dot{\ell}(\cdot,\theta,\eta), A\underline{b}>_{P_{\theta\eta}} = 0, \qquad \text{all } \underline{b} \in L_{2^*}(\eta).$$

Thus $A^*\dot{\ell}(\cdot,\theta,\eta) = 0$. This can be seen to imply

(3.55) $\quad b_0(\cdot) = A(A^*A)^{-1}A^*\dot{\ell}(\cdot,\theta,\eta)\,, \qquad$ (componentwise) .

The optimal limiting measure N is $L(G)$ where G has zero mean normal marginals and covariance function

$$\begin{aligned} EG(u)G(v) &= E_{P_{\theta\eta}} \dot{\kappa}_{\pi_u}(X_1, P_{\theta\eta})\, \dot{\kappa}_{\pi_v}(X_1, P_{\theta\eta}) \\ &= < (1_{(-\infty,u]}(\cdot) - H(u)),\ (A^*A)^{-1}(1_{(-\infty,v]}(\cdot) - H(v)) >_\eta \\ &\quad + < (1_{(-\infty,u]}(\cdot) - H(u)), (A^*A)^{-1}A^*\dot{\ell}(\cdot,\theta,\eta) >'_\eta\ \tilde{I}^{-1}(\theta,\eta) \\ &\qquad \cdot < (1_{(-\infty,v]}(\cdot) - H(v)),\ (A^*A)^{-1}A^*\dot{\ell}(\cdot,\theta,\eta) >'_\eta\,. \end{aligned} \tag{3.56}$$

Finally we have

$$\begin{aligned} E(G(v)-G(v))^2 &= E_{P_{\theta\eta}} \left(\dot{\kappa}_{\pi_u}(X_1, P_{\theta\eta}) - \dot{\kappa}_{\pi_v}(X_1, P_{\theta\eta}) \right)^2 \\ &\le \|A\|^2\, \|(A^*A)^{-1}\|^2\, (1 + \|\dot{\ell}(\cdot,\theta,\eta)\|^2\, \|\tilde{I}^{-1}(\theta,\eta)\|)\ \ |H(v)-H(u)|\,. \end{aligned}$$

(Note that $(A^*A)^{-1}$ is a bounded operator by the Bounded Inverse Theorem

(Rudin (1973), 2.12(b))). By the same argument as in Section 3.6.1 we conclude that $L(G)$ exists as a measure on $(D[-\infty,\infty], \mathcal{U}(\Pi))$. □

3.6.4. Semi-Parametric Models II

Example 3.6.3 replicates results of Begun et al. (1983). It is possible to extend this in the following way. Let $\mathcal{P} = \{P_{\theta\eta} : (\theta,\eta) \in \Theta\times H\}$ as in Example 3.6.3, except that H is now a collection of densities with respect to a σ-finite measure ν on an arbitrary measure space. Again assume existence of $\ell(\cdot,\theta,\eta) \in L_{2*}(P_{\theta\eta})^k$ and of a continuous, linear operator $A = A(\theta,\eta)\colon L_{2*}(\eta) \to L_{2*}(P_{\theta\eta})$ such that (3.46)-(3.47) hold. Furthermore choose $T(P_{\theta\eta})$ as in (3.48). Now consider a functional κ of the form

$$\kappa(P_{\theta\eta}) = \psi(\eta), \tag{3.58}$$

where $\psi\colon H \to (B,\tau)$ and (B,τ) is a topological vector space. Let $N(A)$ be the null space of the operator A and $R(A)$ its range.

LEMMA 3.23. *Let* $\psi\colon H \to (B,\tau)$ *be differentiable in* $\eta \in H$ *with respect to the tangent cone* $T(\eta,H)$ *and derivative* ψ'_η, *assume* $\tilde{I}(\theta,\eta)$ *is nonsingular and suppose that for* $A = A(\theta,\eta)$

$$N(A) \subset N(\psi'_\eta) \tag{3.59}$$

$$R(A) \text{ is closed in } L_{2*}(P_{\theta\eta}). \tag{3.60}$$

Then $\kappa\colon \mathcal{P} \to (B,\tau)$ *given by* (3.58) *is differentiable at* $P_{\theta\eta}$ *with gradients given by* (3.68). *Furthermore* $A^*A\colon L_{2*}(\eta) \to L_{2*}(\eta)$ *is one-to-one and onto if and only if* $N(A) = 0$ *and* (3.60) *holds.* □

PROOF. From (3.58) we see that κ is differentiable at $P_{\theta\eta} \in \mathcal{P}$ if and only if there exists a continuous, linear map $\kappa'_{P_{\theta\eta}} : \operatorname{lin} T(P_{\theta\eta}) \to (B,\tau)$ with for all $h \in \mathbb{R}^k$ and $\underline{b} \in \operatorname{lin} T(\eta,H)$

$$\kappa'_{P_{\theta\eta}}\left(h'\ell(\cdot,\theta,\eta) + A\underline{b} \right) = \psi'_\eta(\underline{b}). \tag{3.61}$$

Since A is not necessarily one-to-one we introduce the quotient space

$L_{2*}(\eta)/N(A)$. Writing elements as $[\underline{b}] = \underline{b}+N(A)$, we have that $L_{2*}(\eta)/N(A)$ is a Hilbert space with respect to the inner product

$$(3.62) \qquad <[\underline{b}_1],[\underline{b}_2]> = <\pi^{\perp}\underline{b}_1, \pi^{\perp}\underline{b}_2>_{\eta},$$

where $\pi^{\perp}$: $L_{2*}(\eta) \to N(A)^{\perp} \subset L_{2*}(\eta)$ is the orthogonal projection onto $N(A)^{\perp}$. Define continuous, linear maps $\bar{A}$: $L_{2*}(\eta)/N(A) \to R(A) \subset L_2(P_{\theta\eta})$ and $\dot{\bar{\psi}}'_{\eta}$: lin $T(\eta,H)/N(A) \to (B,\tau)$ by

$$(3.63) \qquad \bar{A}[\underline{b}] = A\underline{b}$$

and

$$(3.64) \qquad \dot{\bar{\psi}}'_{\eta}([\underline{b}]) = \dot{\psi}'_{\eta}(\underline{b}).$$

Note that $\dot{\bar{\psi}}'_{\eta}$ is well-defined by (3.59). $\bar{A}$ is a continuous, one-to-one, linear map onto R(A), which is a Banach space by (3.60). Hence it has a continuous inverse $\bar{A}^{-1}$: $R(A) \to L_{2*}(\eta)/N(A)$ (cf. Rudin (1973), 2.12(b)). By (3.60) the space H given by

$$(3.65) \qquad H = \text{lin } \{\dot{\ell}(\cdot,\theta,\eta)\} + R(A)$$

is a Banach space. Therefore the projection $\underline{P}$ of H onto R(A) along $\dot{\ell}(\cdot,\theta,\eta)$ is continuous (cf. Jameson (1974), 29.2). Now for $g \in \text{lin } T(P_{\theta\eta}) \subset H$ set

$$(3.66) \qquad \kappa'_{P_{\theta\eta}}(g) = \dot{\bar{\psi}}'_{\eta} \circ \bar{A}^{-1} \circ \underline{P}(g).$$

This concludes the proof that κ is differentiable in $P_{\theta\eta} \in \mathcal{P}$.

Next let $\bar{A}^*$ be the adjoint of $\bar{A}$: $L_{2*}(\eta)/N(A) \to L_{2*}(P_{\theta\eta})$ and $(\bar{A}^{-1})^*$ the adjoint of $\bar{A}^{-1}$: $R(A) \to L_{2*}(\eta)/N(A)$. Then

$$(\bar{A}^*\bar{A})(\bar{A}^{-1}(\bar{A}^{-1})^*) = I = (\bar{A}^{-1}(\bar{A}^{-1})^*)(\bar{A}^*\bar{A}).$$

Clearly $\bar{A}^*\bar{A}$: $L_{2*}(\eta)/N(A) \to L_{2*}(\eta)/N(A)$ is invertible. Let $\dot{\psi}_{b^*}(\cdot,\eta) \in L_{2*}(\eta)$ denote gradients of ψ. By (3.59) we have that $\dot{\psi}_{b^*}(\cdot,\eta) \perp N(A)$. Hence by (3.61) for any $\underline{b} \in \text{lin } T(\eta,H)$ and $h \in \mathbb{R}^k$

$$b^* \circ \kappa'_{P_{\theta\eta}}(h'\dot{\ell}(\cdot,\theta,\eta)+A\underline{b}) = \langle \dot{\psi}_{b^*}(\cdot,\eta),\underline{b}\rangle_\eta = \langle [\dot{\psi}_{b^*}(\cdot,\eta)],[\underline{b}]\rangle$$

(3.67)

$$= \langle \bar{A}(\bar{A}^*\bar{A})^{-1}[\dot{\psi}_{b^*}(\cdot,\eta)],\ \bar{A}[\underline{b}]\rangle_{P_{\theta\eta}} = \langle \dot{\kappa}_{b^*}(\cdot,P_{\theta\eta}),\ h'\dot{\ell}(\cdot,\theta,\eta)+A\underline{b}\rangle_{P_{\theta\eta}}.$$

Here

$$\dot{\kappa}_{b^*}(x,P_{\theta\eta}) = \bar{A}(\bar{A}^*\bar{A})^{-1}[\dot{\psi}_{b^*}(x,\eta)]$$

(3.68)

$$- \langle \bar{A}(\bar{A}^*\bar{A})^{-1}[\dot{\psi}_{b^*}(\cdot,\eta)],\dot{\ell}(\cdot,\theta,\eta)\rangle'_{P_{\theta\eta}} \cdot \tilde{I}^{-1}(\theta,\eta)\dot{\ell}(x,\theta,\eta).$$

Finally if A^*A is one-to-one, then A must be so too, hence $N(A) = 0$, implying (3.59). If A^*A is onto then $R(A^*) = L_2(\eta)$ is closed, implying (3.60) (cf. Rudin (1973), 4.14). Conversely if $N(A) = 0$ then $\bar{A} = A$. If, furthermore $R(A)$ is closed then $A^*A = \bar{A}^*\bar{A}$: $L_2(\eta) \to L_2(\eta)$ is invertible by the above argument, so must be one-to-one and onto. ∎

The results obtained by Wellner (1982) on estimating a distribution function under right censoring can also be obtained within the set-up of this chapter. We defer a discussion of this to Section 4.4, where an indirect approach is taken towards showing differentiability of a functional. This is decomposed there in a functional which is differentiable in the sense of this chapter, and a *Hadamard differentiable* functional.

3.7. A CHARACTERIZATION OF D^*

In this section D denotes $D[a,b]$ $(-\infty \le a < b \le \infty)$; it is equipped with its supremum metric $\|d\|_\infty$; D^* is the set of all $\|\cdot\|_\infty$-continuous, linear maps $d^*: D \to \mathbb{R}$. C denotes the continuous functions on $[a,b]$. The first result paves the way to the characterization of D^* in Proposition 3.25 below. A corollary of that proposition is Lemma 3.26, which has been used in the proof of Theorem 3.17, and furthermore implies equality of the projection σ-field and the σ-field generated by D^*.

LEMMA 3.24. *If* $d^* \in D^*$ *and* $d^*c = 0$ *for all* $c \in C$, *then*

$$d^*(1_{[u,b]}(\cdot)) \neq 0,$$

for at most countably many $u \in (a,b]$. *Moreover if* $\{u_1,u_2,\ldots\}$ *is this set, then*

$$\Sigma_{i=1}^{\infty} |d^*(1_{[u_i,b]}(\cdot))| < \infty. \quad \square$$

PROOF. For an arbitrary countable set $\{v_1,v_2,\ldots\} \subset [a,b]$ define

$$d_n = \Sigma_{i=1}^{n} 1_{[v_i,b]} \operatorname{sgn}(d^*(1_{[v_i,b]})).$$

Then $d_n \in D$. There exists $c_n \in C$ with $\|d_n+c_n\| = 1$. Hence

$$\|d^*\| \geq |d^*(d_n+c_n)| = |d^*(d_n)| = \Sigma_{i=1}^{n} |d^*(1_{[v_i,b]})|.$$

We conclude that the sum of any countable subset of the set of numbers $\{|d^*(1_{[v,b]})|: v \in [a,b]\}$ is finite. ■

PROPOSITION 3.25. D^* *equals the set of* d^* *of the form*

$$d^*(d) = \int d(u)\, d\mu(u) + \Sigma_{i=1}^{\infty} \alpha_i \,(d(u_i)-d(u_i-)) \,. \tag{3.69}$$

Here μ *is a finite signed mesasure on* $[a,b]$, $\{u_i\}$ *is a sequence in* $(a,b]$ *and* $\{\alpha_i\}$ *a sequence in* $\mathbb{R}$ *with* $\Sigma_{i=1}^{\infty} |\alpha_i| < \infty$. $\square$

PROOF. It is not difficult to see that any d^* given by (3.69) is an element of D^*. We prove the converse. Let $d^*|_C$ be the restriction of $d^* \in D^*$ to C. Since $d^*|_C \in C^*$ we have by the Riesz Representation Theorem (Rudin (1966), Th.3.19) the existence of a finite signed measure μ on $[a,b]$ with

$$d^*|_C(c) = \int c(u)\, d\mu(u) \qquad c \in C.$$

Let

$$d_C^*(d) = \int d(u)\, d\mu(u) \qquad d \in D \tag{3.70}$$

(3.71) $\quad d^*_{C_\perp}(d) = d^*(d) - d^*_C(d).$

Clearly $d^*_{C_\perp}(c) = 0$ for all $c \in C$. Let $\{u_1, u_2, \ldots\}$ be the countable set guaranteed to exist by Lemma 3.24. For $d \in D$ let $\Delta d(u) = d(u) - d(u-)$ and set

$$d_n^d(u) = \sum_{i:\, |\Delta d(u_i)| \geq n^{-1}} \Delta d(u_i)\, 1_{[u_i, b]}(u)$$

$$d_n(u) = d(u) - d_n^d(u).$$

As d_n has jumps of absolute magnitude smaller than n^{-1}, there exists $d_n^c \in C$ with

$$\|d_n^c - d_n\|_\infty \leq 2n^{-1},$$

(cf. Billingsley (1968), L.14.1). Since $d^*_{C_\perp}$ is $\|\cdot\|_\infty$-continuous we have

$$d^*_{C_\perp}(d) = d^*_{C_\perp}(d_n^d) + d^*_{C_\perp}(d_n^c) + o(1) = \sum_{i:\, |\Delta d(u_i)| \geq n^{-1}} \Delta d(u_i)\alpha_i + o(1),$$

where $\alpha_i = d^*_{C_\perp}(1_{[u_i, b]})$. Since $\sup\{|\Delta d(u)|: u \in [a,b]\} \leq 2\|d\|_\infty$ and $\Sigma|\alpha_i| < \infty$ by Lemma 3.24, we conclude that

(3.72) $\quad d^*_{C_\perp}(d) = \Sigma_{i=1}^{\infty} \Delta d(u_i)\, \alpha_i$.

Combination of (3.70)-(3.72) yields (3.69). ■

LEMMA 3.26. *Given* $d^* \in D^*$ *there exists a sequence* $\{d_n'\}$ *of the form*

$$d_n' = \Sigma_{i=1}^{n} \alpha_{ni}\, \pi_{t_{ni}},$$

such that

$$d^*(d) = \lim_{n \to \infty} d_n'(d) \qquad \textit{every } d \in D. \ \square$$

PROOF. Let d^* be given by (3.69). For simplicity of notation assume that $[a,b] = [0,1]$. Define $\beta_{ik} = \mu[(i-1)2^{-k}, i2^{-k})$ for $i = 1,2,\ldots,2^k-1$ and set

$\beta_{ik} = \mu[1-2^{-k},1]$ for $i = 2^k$. For an arbitrary $d \in D[0,1]$ let

$$d_k(u) = \Sigma_{i=1}^{2^k-1} 1_{[(i-1)2^{-k},i2^{-k})}(u)\, d(i2^{-k}) + 1_{[1-2^{-k},1]}(u)\, d(1).$$

Since $d_k(u) \to d(u)$ for all $u \in [0,1]$, $(k \to \infty)$, we have for all $d \in D[0,1]$

$$\Sigma_{i=1}^{2^k} \beta_{ik}\, d(i2^{-k}) = \int d_k(u)\, d\mu(u) \to \int d(u) d\mu(u).$$

Next by dominated convergence for all d as $k \to \infty$

$$\Sigma_{i=1}^{k} \alpha_i\, (d(u_i)-d(u_i-\tfrac{1}{k})) \to \Sigma_{i=1}^{\infty} \alpha_i\, (d(u_i)-d(u_i-)).$$

Combination of these assertions implies existence of $\{d_n'\}$ as required. ■

3.8. SOME RESULTS ON SPACES OF REAL FUNCTIONS

This section contains a proof of Lemma 3.21 and a characterization of the support of a (tight) optimal limiting measure on a subspace of B(T).

PROOF OF LEMMA 3.21. Let $S = \cup_{m=1}^{\infty} K_m$, where K_m is compact for every m.

(i) We first show that there exists a semi-metric ρ on T which makes T into a totally bounded metric space with $S \subset UC(T,\rho)$, the elements of B(T) which are uniformly continuous with respect to ρ (cf. Andersen and Dobric (1987)).

For $m = 1,2,\ldots$ set

$$\rho_m(s,t) = \sup_{b \in K_m} |b(s) - b(t)|, \qquad s,t \in T.$$

Then (T,ρ_m) is totally bounded.

To see this, first cover K_m with finitely many balls of arbitrary small radius δ centered at $b_1,\ldots,b_k$ (say). Next divide $\mathbb{R}^k$ into k-dimensional cubes of edge δ. For every cube pick at most one $s \in T$ such that $(b_1(s),\ldots,b_k(s))$ is in the cube. Since $b_1,\ldots,b_k$ are uniformly bounded we obtain finitely many points $s_1,\ldots,s_p$ (say). It can be seen that

the balls $\{s \in T: \rho_m(s,s_j) < 3\delta\}$ cover T.

Next set

$$\rho(s,t) = \Sigma_{m=1}^{\infty} 2^{-m} (\rho_m(s,t) \wedge 1) \quad .$$

Take $b \in K_m$. Then for $2^m\rho(s,t) < \delta < 1$ we have $|b(s)-b(t)| \le \rho_m(s,t) \le 2^m\rho(s,t) < \delta$. Thus $S \subset UC(T,\rho)$.

Finally we show that (T,ρ) is totally bounded. Fix $\delta \in (0,1)$. Let $2^{-M} < \delta$. Choose $\{s_1,\ldots,s_p\}$ such that any $s \in T$ is within ρ_M-distance δ from a s_j. Then $\rho(s,s_j) \le \Sigma_{m=1}^{M}(\rho_m(s,s_j) \wedge 1)\, 2^{-m} + 2^{-M} < 2\delta$.

(ii). Any $b \in UC(T,\rho)$ has a unique extension to a continuous function $\bar{b}$ on the completion $\bar{T}$ of T under ρ. The identification $b \leftrightarrow \bar{b}$ is a norm isomorphism between $UC(T,\rho)$ and $C(\bar{T},\rho)$ under the supremum norm. By the Riesz Representation Theorem (Rudin (1966), Th. 6.19) any $f^* \in C(\bar{T},\rho)$ has the form

$$f^*(\bar{b}) = \int_{\bar{T}} \bar{b}(s)\, d\mu(s) \quad ,$$

for some finite, signed Borel measure μ on $(\bar{T},\rho)$. The restriction of an arbitrary $b^* \in B(T)^*$ to $UC(T,\rho)$ is in $UC(T,\rho)^*$. Thus

$$b^*(b) = \int_{\bar{T}} \bar{b}(s)\, d\mu(s) \qquad\qquad b \in UC(T,\rho) \quad ,$$

for some μ as above.

(iii). We prove (3.39) by discretizing μ. Let $\{t_{n1},\ldots t_{nk_n}\} \subset T$ be such that the balls around these points of radius $(2n)^{-1}$ cover T. Then the balls of radius n^{-1} certainly cover $\bar{T}$. Let $A_{n1} = \{s \in T: \rho(s,t_{n1}) \le n^{-1}\}$, $A_{ni} = \{s \in T: \rho(s,t_{ni}) \le n^{-1}\} \setminus \cup_{j=1}^{i-1} A_{nj}$. The A_{ni} partition $\bar{T}$. Set $b_n' = \Sigma_{i=1}^{k_n} \mu(A_{ni})\, \pi_{t_{ni}}$. Then

$$\left| \int_{\bar{T}} \bar{b}(s)\, d\mu(s) - b_n'(b) \right| \le \Sigma_{i=1}^{k_n} \int_{A_{ni}} |\bar{b}(s)-b(t_{ni})|\, d\|\mu\|(s) \quad .$$

For fixed b this tends to zero if $n \to \infty$.

(iv). Finally we prove (3.40). $UC(T,\rho)$ is separable under the supremum norm. Hence the Borel σ-field in $UC(T,\rho)$ is generated by the closed balls.

Every closed ball can be written as $\cap_{s\in S}\ \{b \in UC(T,\rho)\colon |b(s)-b_0(s)| \le r\}$, where S is a countable ρ-dense subset of T. Hence the Borel σ-field in $UC(T,\rho)$ equals $\mathcal{U}(\Pi)\cap UC(T,\rho)$. ■

The proof of the following lemma is based on the same idea as the proof of Lemma 3.2.

LEMMA 3.27. *Let* B *be a subspace of* B(T) *and let* $\kappa\colon P \to (B,\|\cdot\|_\infty)$ *be differentiable at* $P \in \mathcal{P}$. *Suppose that* $N = L(G)$ *is a* $\|\cdot\|_\infty$*-tight probability measure on* $(B,\mathcal{U}(\|\cdot\|_\infty))$ *with zero-mean normal marginals* $L(G(t_1),\ldots,G(t_m))$ *and covariance function*

$$EG(u)G(v) = \langle \tilde{\kappa}_{\pi_u}(\cdot,P),\ \tilde{\kappa}_{\pi_v}(\cdot,P)\rangle_P \qquad u,v \in T\ .$$

Then

$$b^*(N) = N(0,\ \|\tilde{\kappa}_{b^*}(\cdot,P)\|_P^2)\ , \qquad \textit{for every } b^* \in B^*.\ \square$$

PROOF. We can view N as a tight probability measure on the Borel σ-field on B(T). Moreover, by the Hahn-Banach theorem any $b^* \in B^*$ can be extended to an element of $B(T)^*$. It therefore suffices to prove the theorem for B equal to B(T).

Let S be a $\|\cdot\|_\infty$-σ-compact set such that $N(S) = 1$. Construct ρ as in the proof of Lemma 3.21. Then $S \subset UC(T,\rho)$ and for every $b^* \in B(T)^*$ there exists $\{b_n'\}$ of the form $b_n' = \Sigma_{i=1}^{k_n} \alpha_{ni}\,\pi_{t_{ni}}$ with $b_n'(b) \to b^*(b)$ for all b in $UC(T,\rho)$. In view of Lemma 3.12 with $B' = \Pi$ and $\tau = \tau(\|\cdot\|_\infty)$, it suffices to show that $\kappa_P'(T(P)) \subset UC(T,\rho)$.

Let $g \in T(P)$. Then

$$\kappa_P'(g)(t) = \pi_t\circ\kappa_P'(g) = \langle \tilde{\kappa}_{\pi_t}(\cdot,P),\ g\rangle_P = \langle \tilde{\kappa}_{\pi_t}(\cdot,P),\ Pg\rangle_P,$$

where P is the orthogonal projection of lin T(P) onto the closure of the linear space spanned by $\{\tilde{\kappa}_{\pi_t}(\cdot,P)\colon t \in T\}$. Suppose that $\Sigma_{i=1}^{p_n} \beta_{ni}\,\tilde{\kappa}_{\pi_{s_{in}}}(\cdot,P) \to Pg$ in $L_2(P)$ as $n \to \infty$. Then

$$\|\kappa_P'(g) - \Sigma_{i=1}^{p_n} \beta_{ni}\, EG(\cdot)G(s_{ni})\|_\infty$$

$$\leq \sup_{t\in T} \|\tilde{\kappa}_{\pi_t}(\cdot,P)\|_P \, \|Pg - \Sigma_{i=1}^{p_n} \beta_{ni}\, \tilde{\kappa}_{\pi_{s_{in}}}(\cdot,P)\|_P$$

$$\leq \sup_{t\in T} \|\pi_t \circ \kappa_P'\| \; o(1) \leq \|\kappa_P'\| \; o(1) \; .$$

Hence it suffices to show that the map $t \to EG(t)G(t_{ni})$ is ρ-uniformly continuous.

Now G can be extended to a $\bar{T}$-indexed process $\bar{G}$ with ρ-continuous paths. Clearly $L(\bar{G}(t)) = N(0,E\bar{G}^2(t))$. If $t_n \to t$, then $\bar{G}(t_n) \to \bar{G}(t)$ a.s., so that $L(\bar{G}(t_n)) \xrightarrow{W} L(\bar{G}(t))$. We conclude that $E\bar{G}^2(t_n) \to E\bar{G}^2(t)$. But then $E(\bar{G}(t_n)-\bar{G}(t))^2 \to 0$, so that $E\bar{G}(t_n)\bar{G}(u) \to E\bar{G}(t)\bar{G}(u)$ for every u. Thus $t \to E\bar{G}(t)G(u)$ is (uniformly) continuous on $\bar{T}$. ■

Combination of Theorem 3.11 and Lemma 3.27 yields the following characterization of the support of the optimal limiting distribution on a subspace of B(T).

THEOREM 3.28. *Let* B *be a subspace of* B(T) *and let* κ: $\mathcal{P} \to (B,\|\cdot\|_\infty)$ *be differentiable at* $P \in \mathcal{P}$. *Suppose that* $N = L(G)$ *is a* $\|\cdot\|_\infty$*-tight probability measure on* $(B,\mathcal{U}(\|\cdot\|_\infty))$ *with zero-mean normal marginals* $L(G(t_1),\ldots,G(t_m))$ *and covariance function*

$$EG(u)G(v) = < \tilde{\kappa}_{\pi_u}(\cdot,P), \tilde{\kappa}_{\pi_v}(\cdot,P) >_P \qquad u,v \in T \; .$$

Then

$$N(\overline{\kappa_P'(\text{lin } T(P))}) = 1 \; . \; \square$$

CHAPTER 4

HADAMARD DIFFERENTIABLE FUNCTIONALS OF EFFICIENT ESTIMATORS

4.1 INTRODUCTION. CONVERGENCE IN DISTRIBUTION. HADAMARD DIFFERENTIABILITY.

Let $\mathcal{P}$ be a set of probability measures on the measurable space $(X,\mathcal{B})$ and let B_1 and B_2 be vector spaces. The following is a commonly occurring situation. An 'efficient' estimator sequence $\{T_n\}$ for a functional $\kappa: \mathcal{P} \to B_1$ is given, but the quantity of interest is the value $\phi(\kappa(P))$, where $\phi: B_1 \to B_2$. An obvious estimator for $\phi \circ \kappa: \mathcal{P} \to B_2$ is $\phi(T_n)$. In this chapter we prove that $\{\phi(T_n)\}$ is 'efficient' for $\phi \circ \kappa$, if $\{T_n\}$ is 'efficient' for κ. More generally we investigate which properties of $\{T_n\}$ and ϕ would render $\{\phi(T_n)\}$ 'efficient'. We again restrict ourselves to estimators T_n based on an i.i.d. sample from $P \in \mathcal{P}$.

We put the word 'efficient' in quotes, because so far we have not defined this notion. It is clear that any reasonable definition must take the theorems of Chapter 3 into account, but this does not unequivocally fix one definition. Below we choose a definition based on convergence in distribution (weak convergence), a choice which seems to be appropriate for applications.

4.1.1. Convergence in distribution.

Thus to define efficiency of estimators $\{T_n\}$ of $\kappa: \mathcal{P} \to B$, we need a theory of *convergence in distribution* on the space B. To set this up we suppose that (B,d) is a metric topological vector space. Then the usual way to continue is to consider measures on the Borel σ-algebra $\mathcal{U}(d)$ and to

define weak convergence in the sense of Billingsley (1968). It turns out that this is a convenient set-up when (B,d) is separable, but may run into trouble with nonseparable (B,d). Without the separability the topological formation of (open) sets no longer possesses a countable character, which causes the Borel σ-algebra to be very large, allowing only few functionals to be measurable. Hence (too) few estimators exist.

For instance if (B,d) is D[a,b] with the supremum norm, it turns out that the empirical distribution function is not measurable with respect to the Borel σ-algebra. For this special space one could put the supremum norm aside and to work with the weaker Skorohod topology J instead. However, the Skorohod topology has the disadvantage, besides its complexity, that it does not make D[a,b] into a topological vector space: addition: $(d_1,d_2) \to d_1+d_2$ is not continuous in J.

The problem sketched above has been known for long and possible solutions have been offered by several authors, notably Dudley (1966,1967). Using key-concepts of Dudley, Pollard (1984, Chapters IV and V) and Gaenssler (1983, Section 3) give accounts of a theory of weak convergence for measures which are not necessarily defined on Borel σ-algebras. This body of theory may be seen as an alternative to Billingsley (1968). We include a short account of the concepts that will be needed later.

Given a metric space (B,d), let $\mathcal{U}(d)$ denote its Borel σ-algebra and let $\mathcal{U}$(d-balls) be the smallest σ-algebra containing all closed (and hence all open) d-balls $\{b: d(b,b_0) \le \varepsilon\}$. Following Dudley (1966) and Gaenssler (1983), we consider probability measures defined on $(B,\mathcal{A})$, where throughout the chapter $\mathcal{A}$ is a σ-algebra satisfying

$$\mathcal{U}(\text{d-balls}) \subset \mathcal{A} \subset \mathcal{U}(d). \tag{4.1}$$

DEFINITION 4.1. *Let* (B,d) *be a metric space and let* $\mathcal{A}$ *be a* σ*-algebra of subsets of* B *satisfying* (4.1). *A sequence of probability measures* $\{P_n\}$ *on* $(B,\mathcal{A})$ *is said to* converge weakly *on* $(B,\mathcal{A},d)$ *to a probability measure* P *on* $(B,\mathcal{A})$ *if*

$$\int f\, dP_n \to \int f\, dP,$$

for all bounded, d*-continuous and* $\mathcal{A}$*-measurable functions* $f: B \to \mathbb{R}$. □

We denote weak convergence by

$$P_n \to P \qquad \text{on } (B,\mathcal{A},d).$$

If the metric space (B,d) is separable, then Definition 4.1 reduces to the definition of weak convergence in Billingsley (1968). This follows from the fact that in this case the topology generated by d has a countable base of open balls, implying that the inclusions in (4.1) are in fact equalities.

A probability measure P on $(B,\mathcal{A})$ is *separable* if there exists a separable, closed set $S \subset B$ with $P(S) = 1$. (Under (4.1) any such S is automatically a member of $\mathcal{A}$). A probability measure P on $(B,\mathcal{A})$ is *tight* if to any $\varepsilon > 0$ there exists a compact set K_ε such that $P(K_\varepsilon) \geq 1-\varepsilon$. (Again $K_\varepsilon \in \mathcal{A}$, automatically). For complete metric spaces the notions separability and tightness coincide (cf. Parthasarathy (1967), Th. 3.2). Next a *sequence* of probability measures $\{P_n\}$ on $(B,\mathcal{A})$ is *tight* if for every $\varepsilon > 0$ there exists a compact set K_ε such that $\liminf P_n(G) > 1-\varepsilon$ for every open $\mathcal{A}$-measurable set G containing K_ε. We note that this requirement is essentially weaker than the usual tightness condition. In particular, every weakly converging sequence with a tight limit is tight. Finally $\{P_n\}$ is *relatively compact* if every subsequence has a weakly convergent further subsequence.

Now in Pollard (1984) and Gaenssler (1983) it is shown that many of the properties of weak convergence on Borel σ-algebras are retained in the new set-up, provided that the limit measure P is separable. (In Gaenssler this requirement is part of the definition of weak convergence). This holds true in particular for the continuous mapping theorem, the existence of almost sure representations and a version of Prohorov's theorem (cf. Pollard (1984),IV-12,13,29).

4.1.2. Hadamard differentiability

It is clear that the result mentioned in the first paragraph cannot be true unless ϕ is 'smooth' in an appropriate sense. We make this precise by Hadamard differentiability.

DEFINITION 4.2. *Let* (B_i,τ_i) $(i=1,2)$ *be topological vector spaces. A map* ϕ: $(B_1,\tau_1) \to (B_2,\tau_2)$ *is called Hadamard differentiable at* $b \in B_1$, *if there exists a continuous linear map* ϕ_b': $(B_1,\tau_1) \to (B_2,\tau_2)$, *such that*

$$t^{-1}[\phi(b+th) - \phi(b)] \to \phi_b'(h),$$

as $t \downarrow 0$, *uniformly in* $h \in K_1$, *for each* τ_1*-compact set* $K_1 \subset B_1$. □

An equivalent definition is that for all converging sequences $h_n \to h$ in B_1 and $t_n \downarrow 0$ in $\mathbb{R}$, it holds that

$$t_n^{-1}[\phi(b+t_nh_n) - \phi(b)] \to \phi_b'(h). \tag{4.2}$$

This leads to the following adaptation of Definition 4.2. A map ϕ: $(B_1,\tau_1) \to (B_2,\tau_2)$ is called *Hadamard differentiable* at $b \in B_1$ *tangentially to a subset* $S \subset B_1$, if there exists a continuous linear map ϕ_b': $(B_1,\tau_1) \to (B_2,\tau_2)$, such that (4.2) holds for all $h_n \to h$ for which the limit h is in S. [1)]

Hadamard differentiability is tailored to combination with the tightness of sequences of measures and therefore suitable for use in statistics. Its introduction in statistics is due to Reeds (1976) and meant to replace the more restrictive notion of Fréchet differentiability which has been used in robust statistics. Hadamard differentiability tangentially to a subset is due to Gill (1986), who makes fruitful use of this concept. Reeds (1976), Gill and Johansen (1987), and Fernholz (1983) give many examples of Hadamard differentiable functionals.

Given σ-fields A_i on the spaces B_i $(i = 1,2)$ and a measurable and Hadamard differentiable functional ϕ, a derivative ϕ_b' is not automatically measurable too. For the most commonly occurring types of σ-fields, it is, though.

---

1) *The derivative is assumed to exist on* B_1, *though in principle its values are now fixed on* S *only.*

LEMMA 4.3. *Let* (B_i,τ_i) *be topological vector spaces with* σ*-algebras* A_i *(i=1,2) which make translation and scalar multiplication measurable. Assume that* A_2 *is generated by a set of* τ_2*-continuous real maps on* B_2*. If* ϕ: $(B_1,\tau_1) \to (B_2,\tau_2)$ *is Hadamard differentiable at* $b \in B_1$ *and* A_1-A_2*-measurable, then its derivative* ϕ_b' *is* A_1-A_2*- measurable.* □

PROOF. Using the well-known fact that a pointwise limit of a sequence of measurable real functions, is measurable, we easily obtain that ϕ_b' is A_1-$\mathcal{U}$ measurable, where $\mathcal{U}$ is the σ-field generated by the A_2-measurable, τ_2-continuous, real functions on B_2. ■

This lemma applies e.g. to Borel σ-fields in metrizable spaces, to σ-fields generated by continuous, linear maps, and to σ-fields generated by closed balls in a normed space.

Thus derivatives are more often measurable than the original maps. An extreme and very useful case of this holds for (products of) the space D. Then *every* derivative is measurable with respect to (products of) the projection σ-field.

LEMMA 4.4. *Let* ϕ: $(D[a,b],\|\cdot\|_\infty) \to (D[c,d],\|\cdot\|_\infty)$ *be continuous and linear. Then it is* $\mathcal{U}(\Pi)$-$\mathcal{U}(\Pi)$ *-measurable. This remains true if one or both of the* D*-spaces is replaced by a finite product of* D*-spaces with product topology and product projection* σ*-field, or by Euclidean space.* □

PROOF. By Lemma 3.26 $\mathcal{U}(D[a,b]^*) = \mathcal{U}(\Pi)$. To show $\mathcal{U}(\Pi)$-$\mathcal{U}(\Pi)$-measurability of ϕ it suffices to show $\mathcal{U}(\Pi)$-measurability of $\pi_t \circ \phi$ for every $t \in [c,d]$. Since $\pi_t \circ \phi \in D^*$, the first result follows.

The second result follows by a similar argument, combined with the identity $\mathcal{U}((D_1\times\ldots\times D_k)^*) = \mathcal{U}(D_1^*)\times\ldots\times\mathcal{U}(D_k^*) = \mathcal{U}(\Pi)\times\ldots\times\mathcal{U}(\Pi)$. To see the first equality note that any $d^* \in (D_1\times\ldots\times D_k)^*$ is of the form $d^*(d_1,\ldots,d_k) = d_1^*(d_1) +\ldots+ d_k^*(d_k)$, for $d_i^* \in D_i^*$. ■

4.2. EFFICIENCY

4.2.1. Definition of efficiency.

We now define *efficiency* of an estimator sequence in $(B,\mathcal{A},d)$, where throughout the chapter $(B,\mathcal{A},d)$ satisfies the following conditions, which will be referred to as the *standard conditions*.

Let (B,d) be a metric topological vector space with dual space B^*. Assume that $\mathcal{A}$ is a σ-algebra on B which satisfies (4.1) and makes the translation and scalar multiplication map from B to B given by

$$b \to tb+b_0 \qquad \text{(fixed } t \in \mathbb{R} \text{ and } b_0 \in B\text{)}$$

$\mathcal{A}$-$\mathcal{A}$-measurable. Moreover assume that any separable probability measure L on $(B,\mathcal{A})$ is completely determined by the set of marginals $\{b^*(L)\colon b^* \in B^*,\ b^* \text{ is } \mathcal{A}\text{-measurable}\}$.[2)]

Let $X_1,\dots,X_n$ be i.i.d. random elements in a measurable space $(\mathcal{X},\mathcal{B})$ with distribution $P \in \mathcal{P}$. As before call $\{T_n\}$ an estimator sequence in $(B,\mathcal{A})$ if $T_n = t_n(X_1,\dots,X_n)$ for measurable maps $t_n\colon (\mathcal{X}^n,\mathcal{B}^n) \to (B,\mathcal{A})$.

DEFINITION 4.5. *Let* $\kappa\colon \mathcal{P} \to (B,d)$ *be differentiable at* $P \in \mathcal{P}$ *relative to* $T(P)$. *Assume the existence of a separable probability measure* N *on* $(B,\mathcal{A})$ *with*

$$b^*(N) = N(0,\|\tilde{\kappa}_{b^*}(\cdot,P)\|_P^2) \qquad \textit{for all } \mathcal{A}\textit{-measurable } b^* \in B^*. \tag{4.3}$$

Then an estimator sequence $\{T_n\}$ *in* $(B,\mathcal{A})$ *is called efficient for* κ *at* $P \in \mathcal{P}$ *if for every* $g \in T(P)$

$$L_{P_n}\left(\sqrt{n}(T_n-\kappa(P_n))\right) \to N \qquad \textit{on } (B,\mathcal{A},d), \tag{4.4}$$

for all $\{P_n\} \subset \mathcal{P}$ *satisfying* (2.11) *and*

$$\sqrt{n}(\kappa(P_n)-\kappa(P)) \to \kappa_P'(g) \qquad \textit{in } d. \ \square \tag{4.5}$$

---

2) *The standard conditions are satisfied e.g. for Borel* σ*-fields in normed spaces (cf. Lemma 3.13), and complete subspaces of* $B(T)$ *with the* σ*-field generated by the coordinate projections and norm balls (cf. Theorem 4.9).*

Hadamard differentiable functionals of efficient estimators

Efficiency is defined relatively to a particular choice of a tangent cone T(P). When T(P) is convex then an efficient estimator is a *best regular* estimator according to Theorem 3.7 and a *locally asymptotically minimax* estimator in the sense of Theorem 3.9. However, the requirement that $L(\sqrt{n}(T_n-\kappa(P_n)))$ *converges weakly* on $(B,\mathcal{A},d)$, is unrelated to any of these theorems. Some motivation to include weak convergence of $L(\sqrt{n}(T_n-\kappa(P_n)))$ in the definition of efficiency can be based on the minimax Theorem 3.8 and 3.10. (In general, though, the convergence (4.4) is not sufficient to ensure that the estimator attains equality in (3.22)).

Note that efficiency is also defined relative to a σ-field $\mathcal{A}$ and a metric d. This may be made explicit by saying that $\{T_n\}$ *in* $(B,\mathcal{A},d)$ is efficient.

The following warning is appropriate: an *efficient* estimator in our definition is not necessarily a *best* estimator, even if it is accepted that best should be defined in terms of limiting distributions of $\{L(\sqrt{n}(T_n-\kappa(P_n)))\}$. This is connected with the structure of the tangent cone and was discussed for the case that $B = \mathbb{R}^k$ in Section 2.6. This discussion immediately carries over to general B and is not repeated here. However, roughly, for convex tangent cones, efficiency is to be considered as an abbreviation of *LAM and best regular*.

4.2.2. Efficiency, marginal efficiency and tightness.

In the remainder of this section we obtain some useful alternative characterizations of efficiency, giving special attention to product spaces and subspaces of B(T). Unlike in Chapter 3 we do not assume that the σ-algebra $\mathcal{A}$ is generated by linear maps. This has complicated some of the proofs. The main results are Theorems 4.8 and 4.9.

Relations (4.3)-(4.4) imply

$$(4.6) \qquad L_{P_n}\big(\sqrt{n}(b^*\circ T_n - b^*\circ\kappa(P_n))\big) \to N(0,\|\tilde{\kappa}_{b^*}(\cdot,P)\|_P^2),$$

for all $\mathcal{A}$*-measurable* $b^* \in B^*$.

This can be seen to say that $\{b^*\circ T_n\}$ is efficient for $b^*\circ\kappa$ at $P \in \mathcal{P}$, for every $\mathcal{A}$-measurable $b^* \in B^*$. Thus efficiency implies *marginal efficiency* (which we define by (4.6)). This conclusion can, of course, not be reversed, since marginal efficiency does not imply relative compactness.

However, if $\{T_n\}$ is marginally efficient, and $\{L_{P_n}(\sqrt{n}(T_n-\kappa(P_n)))\}$ is relatively compact with separable limit points, for every $\{P_n\}$ satisfying (2.11) and (4.5), then it is efficient. To see this, note that marginal efficiency gives that every limit point N satisfies (4.3). Furthermore, by the last of the standard conditions (4.3) *uniquely* determines the measure N on A.

In applications one usually establishes relative compactness by showing tightness. By contiguity arguments it can be shown that tightness under the fixed P suffices.

LEMMA 4.6. *Let* $\kappa: \mathcal{P} \to (B,d)$ *be differentiable at* $P \in \mathcal{P}$. *Let* $\{T_n\}$ *be an estimator sequence in* $(B,\mathcal{A})$ *such that* $\{L_P(\sqrt{n}(T_n-\kappa(P)))\}$ *is tight. Then* $\{L_{P_n}(\sqrt{n}(T_n-\kappa(P_n)))\}$ *is relatively compact with tight limiting points for every* $\{P_n\}$ *satisfying* (2.11) *and* (4.5). □

PROOF. Consider the laws $\{L_P(\sqrt{n}(T_n-\kappa(P)), \Delta_n(P_n,P))\}$ on the product space $(B\times\mathbb{R},\mathcal{A}\times\mathcal{B})$, where $\mathcal{B}$ are the Borel sets in $\mathbb{R}$. Equip $B\times\mathbb{R}$ with the metric

$$\underline{d}((b,r),(b',r')) = d(b,b') \vee |r-r'| .$$

Then (4.1) holds for $(B\times\mathbb{R},\mathcal{A}\times\mathcal{B},\underline{d})$. Moreover the well-known implication, marginal tightness implies joint tightness, holds, also under the weaker tightness concept of 4.1.1.

Thus we conclude that $\{L_P(\sqrt{n}(T_n-\kappa(P)), \Delta_n(P_n,P))\}$ is tight on $(B\times\mathbb{R},\mathcal{A}\times\mathcal{B},\underline{d})$. By the compactness theorem (Pollard (1984), IV-29) it is relatively compact with tight limit points. Let L be a limit point along a subsequence of $\{n\}$. By Proposition A.6 (appendix) $\{L_{P_n}(\sqrt{n}(T_n-\kappa(P)))\}$ has a limit along the same subsequence, given by

$$L'(A) = \int_{A\times\mathbb{R}} e^{\lambda}\, dL(y,\lambda) \qquad (A \in \mathcal{A}).$$

Given two numbers $\varepsilon,\varepsilon' > 0$ there exist compact sets $K_1 \subset B$ and $K_2 \subset \mathbb{R}$ such that $L(K_1^c\times\mathbb{R}) < \varepsilon$ and $\int_{B\times K_2^c} e^{\lambda}\, dL(y,\lambda) < \varepsilon'$. Then

$$L'(K_1^c) \le \int_{B\times K_2^c} e^{\lambda}\, dL(y,\lambda) + \int_{K_1^c\times K_2} e^{\lambda}\, dL(y,\lambda) < \varepsilon' + \sup_{\lambda\in K_2} e^{\lambda}\, \varepsilon .$$

This can be made arbitrarily small by first choosing ε' and next ε sufficiently close to zero. We conclude that L' is tight.

Combination with (4.5) gives that $\{L_{P_n}(\sqrt{n}(T_n-\kappa(P_n)))\}$ is relatively compact with tight limiting points. ■

Thus, to show that an estimator sequence is efficient one usually shows tightness under P, and efficiency of marginals. Here $\{b^*\circ T_n\}$ is efficient for $b^*\circ\kappa$ at $P \in \mathcal{P}$ if and only if it is asymptotically linear in the sense of

$$\sqrt{n}(b^*\circ T_n - b^*\circ\kappa(P)) = n^{-\frac{1}{2}} \Sigma_{j=1}^n \tilde{\kappa}_{b^*}(X_j,P) + o_p(1) . \tag{4.7}$$

The if part of this assertion follows easily with the help of the *third Lemma of Le Cam* (Corollary A.7). The only if part is a consequence of Theorem 2.12.

It is usually sufficient to show efficiency of only a selected set of marginals. Rather than giving a general result, we consider special cases. First consider the case that $B = \mathbb{R}^k$. Then efficiency of coordinates implies joint efficiency. The following lemma generalizes Corollary 9.3.6 of Pfanzagl (1983).

LEMMA 4.7. *Let* $\kappa: \mathcal{P} \to \mathbb{R}^k$ *be differentiable at* $P \in \mathcal{P}$. *Suppose that* $\{T_n\}$ *is an estimator sequence in* $\mathbb{R}^k$ *such that* $\{e_i'\circ T_n\}$ *is efficient for* $e_i'\circ\kappa: \mathcal{P} \to \mathbb{R}$ *at* $P \in \mathcal{P}$ *(i = 1,2,...,k). Then* $\{T_n\}$ *is efficient for* κ *at* $P \in \mathcal{P}$. □

PROOF. Efficiency of $\{e_i'\circ T_n\}$ implies asymptotic unbiasedness. Apply Theorem 2.12 to see that

$$\sqrt{n}(e_i'\circ T_n - e_i'\circ\kappa(P)) = n^{-\frac{1}{2}}\Sigma_{j=1}^n \tilde{\kappa}_{e_i'}(X_j,P) + o_p(1) .$$

But this implies joint asymptotic normality of $\{\sqrt{n}(T_n-\kappa(P_n))\}$ under every $\{P_n\}$ satisfying (2.11) and (4.5), with zero mean and covariance matrix $(\langle\tilde{\kappa}_{e_i^*}(\cdot,P),\tilde{\kappa}_{e_j^*}(\cdot,P)\rangle_P)$ (cf. Corollary A.7). ■

Lemma 4.7 is a curiosity rather than a result which can facilitate a

proof of efficiency. It will be used to obtain two really useful theorems.

First we consider general product spaces. For $i = 1,2,\ldots,k$ let κ_i be a functional defined on $\mathcal{P}$ with values in B_i. In view of Lemma 4.7 we expect efficiency of $\{T_{ni}\}$ for κ_i, to imply efficiency of $\{T_n\} = \{(T_{n1},..,T_{nk})\}$ for $\kappa = (\kappa_1,\ldots,\kappa_k)$. This is essentially true. Let d_i and $\mathcal{A}_i$ be a metric and a σ-field on B_i, respectively, satisfying the standard conditions of Section 4.2.1. Define a metric $d_1 v \ldots v d_k$ on the product space by

$$d_1 v \ldots v d_k((b_1,\ldots,b_k),(b_1',\ldots,b_k')) = \sup_i d_i(b_i,b_i') \ .$$

THEOREM 4.8. *For* $i = 1,2,\ldots,k$ *let* $\kappa_i: \mathcal{P} \to (B_i,d_i)$ *be differentiable at* $P \in \mathcal{P}$ *and let* $\{T_{ni}\}$ *be efficient in* $(B_i,\mathcal{A}_i,d_i)$ *for* κ_i *at* $P \in \mathcal{P}$, *where the limiting measure is tight. Then* $\{T_n\} = \{(T_{n1},\ldots,T_{nk})\}$ *is efficient in* $(B_1\times\ldots\times B_k,\mathcal{A}_1\times\ldots\times\mathcal{A}_k,d_1 v \ldots v d_k)$ *for* $\kappa = (\kappa_1,\ldots,\kappa_k)$ *at* $P \in \mathcal{P}$. □

PROOF. We first check that the standard conditions imposed on $(B_i,\mathcal{A}_i,d_i)$ carry over to the product. Set $B = B_1\times\ldots\times B_k$, $\mathcal{A} = \mathcal{A}_1\times\ldots\times\mathcal{A}_k$ and $d = d_1 v \ldots v d_k$.

The metric d generates the product topology. It is easily seen that (4.1) is satisfied for d and $\mathcal{A}$; and translation and scalar multiplication are measurable with respect to $\mathcal{A}$.

Let $b^* \in B^*$ be $\mathcal{A}$- measurable. Set $b_i^*(b_i) = b^*(0,..,0,b_i,0,..,0)$, (where b_i is on the i-th position). It is easily checked that $b_i^* \in B_i^*$ and is $\mathcal{A}_i$-measurable. Clearly $b^*(b) = b_1^*(b_1) +\ldots+ b_k^*(b_k)$. Set $\mathcal{A}_i^* = U(b_i^* \in B_i^*$: b_i^* is $\mathcal{A}_i$-measurable). It can be checked that $\mathcal{A}_1^*\times\ldots\times\mathcal{A}_k^* = \mathcal{A}^* = U(b^* \in B^*$: b^* is $\mathcal{A}$-measurable).

Suppose that N is a separable measure on $\mathcal{A}$. The values of N on $\mathcal{A}^*$ determine the measure $B \to N(B\times A_2\times\ldots\times A_k)$ on $\mathcal{A}_1^*$ for every $A_2,\ldots,A_k$ from $\mathcal{A}_2^*,\ldots,\mathcal{A}_k^*$. By the standard conditions on $(B_1,\mathcal{A}_1,d_1)$ this measure is then determined on the whole of $\mathcal{A}_1$. Repeating the argument we see that N is completely determined by its values on $\mathcal{A}^*$. This concludes the proof that the standard conditions hold for the product space.

Next we show marginal efficiency of $\{T_n\}$. Let $b^* \in B^*$ be $\mathcal{A}$-measurable. Define b_i^* as above. By assumption $\{b_i^* \circ T_{ni}\}$ is efficient for $b_i^* \circ \kappa_i$. By Lemma 4.7 $\{(b_1^* \circ T_{n1},\ldots,b_k^* \circ T_{nk})\}$ is efficient for $(b_1^* \circ \kappa_1,\ldots,b_k^* \circ \kappa_k)$. It can be seen that this implies efficiency of

$\{b_1^* \circ T_{n1} + \ldots + b_k^* \circ T_{nk}\}$ for $b_1^* \circ \kappa_1 + \ldots + b_k^* \circ \kappa_k = b^* \circ \kappa$. (In fact this is a special case of the main result of this chapter, Theorem 4.10).

Finally $\{L_P(\sqrt{n}(T_n - \kappa(P)))\}$ is marginally, hence jointly tight. By Lemma 4.6 $\{L_{P_n}(\sqrt{n}(T_n - \kappa(P_n)))\}$ is relatively compact with tight (and hence separable) limiting points. Let N be a limit point. $b^*(N)$ is the weak limit of the corresponding subsequence of $\{L_{P_n}(\sqrt{n}(b^* \circ T_n - b^* \circ \kappa(P_n)))\}$. By efficiency of $\{b^* \circ T_n\}$, N must satisfy (4.3). Since (4.3) completely determines N on $(B, \mathcal{A})$ by the standard conditions, we can conclude that every subsequence of $\{L_{P_n}(\sqrt{n}(T_n - \kappa(P_n)))\}$ has a further subsequence, which converges weakly to N on $(B, \mathcal{A}, d)$. This implies (4.4). ■

Next we consider spaces of real functions. Let B be a subspace of B(T) (the space of all bounded real functions on T, discussed in Section 3.5.2). Then $\|\cdot\|_\infty$-tightness together with efficiency of all coordinates $\{\pi_t \circ T_n\}$ is sufficient for efficiency of $\{T_n\}$.

THEOREM 4.9. *Let* B *be a complete subspace of* $(B(T), \|\cdot\|_\infty)$ *and let* $\mathcal{A}$ *render all coordinate projections measurable, satisfy* (4.1) *and make translation and scalar multiplication measurable. Then the standard conditions are satisfied. Moreover, assume that* $\kappa: \mathcal{P} \to (B, \|\cdot\|_\infty)$ *is differentiable at* $P \in \mathcal{P}$. *Let* $\{T_n\}$ *be an estimator sequence in* $(B, \mathcal{A})$. *Then* $\{T_n\}$ *is efficient for* κ *at* $P \in \mathcal{P}$ *if and only if* $\{L_P(\sqrt{n}(T_n - \kappa(P)))\}$ *is* $\|\cdot\|_\infty$*-tight and* $\{\pi_t \circ T_n\}$ *is efficient for* $\pi_t \circ \kappa$ *at* $P \in \mathcal{P}$ *for every* $t \in T$. □

PROOF. Because B is complete, any separable probability measure on $(B, \mathcal{A})$ is tight. Now any tight measure on $(B, \mathcal{A})$ is uniquely determined by its values on the projection σ-field $\mathcal{U}(\Pi)$. To see this, let S be σ-compact with $L(S) = 1$. By Lemma 3.21, for every $A \in \mathcal{A}$ there exists $A' \in \mathcal{U}(\Pi)$ with $A \cap S = A' \cap S$. Since $S \in \mathcal{A}$, we have $L(A) = L(A \cap S) = L(A' \cap S) = L(A')$. Thus the standard conditions are satisfied.

Necessity of the conditions of the theorem follows, as separability of the limiting measure in (4.4) implies its tightness, and hence the tightness of the sequence $\{L_P(\sqrt{n}(T_n - \kappa(P)))\}$ (in the weakened sense of 4.1.1). Furthermore, that efficiency implies marginal efficiency was argued above (and, more rigorously, follows from Theorem 4.10). We now show sufficiency.

Let $\{P_n\}$ satisfy (2.11) and (4.5). By tightness and Lemma 4.6 $\{L_{P_n}(\sqrt{n}(T_n - \kappa(P_n)))\}$ is relatively compact with tight limiting points. Let L be a limiting point. For every $\{t_1, t_2, \ldots, t_k\} \subset T$

$$L_{P_n}(\sqrt{n}(\pi_{t_1..t_k} \circ T_n - \pi_{t_1..t_k} \circ \kappa(P_n)) \to \pi_{t_1..t_k}(L) ,$$

at least along a subsequence of $\{n\}$. By Lemma 4.7 and efficiency of coordinates $\{\pi_{t_i} \circ T_n\}$

$$\pi_{t_1..t_k}(L) = N(0, (<\tilde{\kappa}_{\pi_{t_i}}(\cdot,P), \tilde{\kappa}_{\pi_{t_j}}(\cdot,P)>_P)) . \tag{4.8}$$

But (4.8) uniquely determines L on $\dot{A}$. We conclude that all the limiting points of $\{L_{P_n}(\sqrt{n}(T_n - \kappa(P_n)))\}$ equal the tight measure L determined on $\dot{A}$ by (4.8). Finally we prove that this measure satisfies (4.3). Indeed, we give *two* proofs.

First we can invoke Lemma 3.27. Since L is tight it has a tight extension $\tilde{L}$ to the Borel σ-field. (Letting S be σ-compact with $L(S) = 1$, this can be defined by $\tilde{L}(A) = L(A')$, if $A \in \mathcal{U}(\|\cdot\|_\infty)$, $A' \in \dot{A}$ and $A \cap S = A' \cap S$ (cf. Lemma 3.21)). Lemma 3.27 shows that $b^*(\tilde{L}) = N(0, \|\tilde{\kappa}_{b^*}(\cdot,P)\|_P^2)$ for every $b^* \in B^*$.

Secondly we can apply Theorem 3.7 with B' equal to the set of all $\dot{A}$-measurable $b^* \in B^*$. By the above argument $\{T_n\}$ is B'-regular with limiting distribution L, which is $\|\cdot\|_\infty$-tight, hence certainly $\tau(B')$-tight. By Theorem 3.7 we conclude that there exists a measure N on $(B, \mathcal{U}(B'))$ that satisfies (4.3). It follows from the proof of this theorem that N is $\|\cdot\|_\infty$-tight. But then L must be an extension of N to $\dot{A}$. To see this, let S be σ-compact with $N^*(S) = 1$. Since $\mathcal{U}(\Pi) \cap S = \mathcal{U}(B') \cap S = \dot{A} \cap S$ (cf. Lemma 3.21), N has a tight extension $\tilde{N}$ to $\dot{A}$ (given by: $\tilde{N}(A) = N(A')$ if $A \cap S = A' \cap S$ and $A' \in \mathcal{U}(B')$). Now $\tilde{N}$ equals L on $\dot{A}$ by the above argument, as they are equal on $\mathcal{U}(\Pi)$. ∎

Perhaps tightness together with coordinate efficiency is what the reader had in mind for efficiency of an estimator sequence in (a subspace) of B(T). In that case Theorem 4.9 is to be considered as part of the main result of this chapter, together with the theorems of Section 4.3. Indeed,

Theorem 4.9 can be given the interpretation that efficiency of coordinates of an estimator sequence in a subspace of B(T), implies efficiency of all measurable, continuous, linear maps in $\mathbb{R}$ applied to the estimator sequence. These are the simplest examples of Hadamard differentiable maps. We now turn to general Hadamard differentiable maps.

4.3. MAIN RESULTS

In this section (B_i,d_i) are metric topological vector spaces with σ-fields A_i satisfying the standard conditions of Section 4.2.1 ($i = 1,2$).

The following theorem is the promised result on efficiency of functionals of efficient estimators.

THEOREM 4.10. *Let* $\kappa: \mathcal{P} \to (B_1,d_1)$ *be differentiable at* $P \in \mathcal{P}$ *and* $\phi: (B_1,d_1) \to (B_2,d_2)$ *be Hadamard differentiable at* $\kappa(P)$, *with an* A_1-A_2-*measurable derivative. Assume that* $\{T_n\}$ *is efficient for* κ *at* $P \in \mathcal{P}$. *Then* $\phi\circ\kappa: \mathcal{P} \to (B_2,\tau_2)$ *is differentiable at* $P \in \mathcal{P}$ *and* $\{\phi(T_n)\}$ *is efficient for* $\phi\circ\kappa$ *at* $P \in \mathcal{P}$, *provided that* $\phi\circ t_n: (X^n,\mathcal{B}^n) \to (B_2,A_2)$ *is measurable.* □

PROOF. Let $\{P_n\}$ satisfy (2.11) and (4.5). Then by Hadamard differentiability

$$
\begin{aligned}
&\sqrt{n}(\phi\circ\kappa(P_n) - \phi\circ\kappa(P)) \\
&\sqrt{n}[\phi(\kappa(P) + n^{-\frac{1}{2}}\sqrt{n}(\kappa(P_n)-\kappa(P))) - \phi\circ\kappa(P)] \to \phi'_{\kappa(P)}\circ\kappa'_P(g),
\end{aligned}
\tag{4.9}
$$

in d_2. Hence $\phi\circ\kappa: \mathcal{P} \to (B_2,d_2)$ is differentiable at $P \in \mathcal{P}$ with derivative $\phi'_{\kappa(P)}\circ\kappa'_P$. If $b_2^* \in B_2^*$ then $b_2^*\circ\phi'_{\kappa(P)} \in B_1^*$. Hence a gradient of $\phi\circ\kappa$ in the direction b_2^* is given by

$$
\dot{\phi\circ\kappa}_{b^*}(\cdot,P) = \dot{\kappa}_{b_2^*\circ\phi'_{\kappa(P)}}(\cdot,P).
$$

By assumption $L_{P_n}(\sqrt{n}(\phi(T_n)-\phi\circ\kappa(P_n)))$ is well-defined. We show that

$$
L_{P_n}(\sqrt{n}(\phi(T_n)-\phi\circ\kappa(P_n))) \to N_2 \qquad \text{on } (B_2,A_2,d_2),
\tag{4.10}
$$

where N_2 is a d_2-separable probability measure on $(B_2, \mathcal{A}_2)$ satisfying

$$(4.11) \qquad b_2^*(N_2) = N(0, \|\tilde{\kappa}_{b_2^* \circ \phi'_{\kappa(P)}}(\cdot, P)\|_P^2) \qquad \text{for all } \mathcal{A}_2\text{-measurable } b_2^* \in B_2^*.$$

For this we invoke a Skorohod-Dudley representation theorem (cf. Pollard (1984, IV-13). Because of (4.3)-(4.4) there exists a probability space $(\Omega, \mathcal{U}, \underline{P})$ and measurable maps G, Y_n: $(\Omega, \mathcal{U}) \to (B_1, \mathcal{A}_1)$ $(n = 1,2,\ldots)$ with

$$(4.12) \qquad \mathcal{L}(Y_n) = \mathcal{L}_{P_n}\left(\sqrt{n}(T_n - \kappa(P_n)) \right)$$

$$(4.13) \qquad \begin{aligned} &\mathcal{L}(b_1^*(G)) = N(0, \|\tilde{\kappa}_{b_1^*}(\cdot, P)\|_P^2) \qquad \text{for all } \mathcal{A}_1\text{-measurable } b_1^* \in B_1^*. \\ &\mathcal{L}(G) \text{ is separable} \end{aligned}$$

$$(4.14) \qquad Y_n \to G \qquad \text{a.s. in } d_1.$$

Under P_n, the random element $\sqrt{n}\,(\phi(T_n) - \phi\circ\kappa(P_n))$ has the same distribution as $Z_n = \sqrt{n}\,[\phi(n^{-\frac{1}{2}}Y_n + \kappa(P_n)) - \phi\circ\kappa(P_n)]$. By Hadamard differentiability of ϕ at $\kappa(P)$, (4.14), (4.5) and (4.9) we have

$$Z_n = \sqrt{n}[\phi\{\kappa(P) + n^{-\frac{1}{2}}(Y_n + \sqrt{n}(\kappa(P_n) - \kappa(P)))\} - \phi\circ\kappa(P_n)] \to \phi'_{\kappa(P)}(G),$$

a.s. in d_2. Hence

$$\mathcal{L}_{P_n}\left(\sqrt{n}(\phi(T_n) - \phi\circ\kappa(P_n)) \right) \to \mathcal{L}(\phi'_{\kappa(P)}(G)) \qquad \text{on } (B_2, \mathcal{A}_2, d_2).$$

If $b_2^* \in B_2^*$ is $\mathcal{A}_2$-measurable then $b_2^* \circ \phi'_{\kappa(P)}$ is $\mathcal{A}_1$-measurable, according to the assumptions. Hence by efficiency of $\{T_n\}$

$$(4.15) \qquad \mathcal{L}\left(b_2^* \circ \phi'_{\kappa(P)}(G) \right) = N(0, \|\tilde{\kappa}_{b_2^* \circ \phi'_{\kappa(P)}}(\cdot, P)\|_P^2).$$

This concludes the proof that for every $g \in T(P)$ *there exists* $\{P_{ng}\} \subset \mathcal{P}$ satisfying (2.11) and (4.9), such that (4.10)-(4.11) hold. Now suppose that $\{P_n^*\}$ satisfies (2.11) for some $g \in T(P)$. Then $\Lambda_n(P_n^*, P_{ng}) \to 0$ in P_{ng}-probability by Lemma 2.5. Hence (cf. Pollard (1984), p.69)

Hadamard differentiable functionals of efficient estimators

$$L_{P_{ng}}(\sqrt{n}(\phi(T_n)-\phi\circ\kappa(P_{ng})),A_n(P_n^*,P_{ng})) \to L(\phi'_{\kappa(P)}(G),0)$$

on $(B_2\times\mathbb{R},\mathcal{A}_2\times\mathcal{B},d_2\vee|\cdot|)$.

By Proposition A.6 (appendix)

$$L_{P_n^*}(\sqrt{n}(\phi(T_n)-\phi\circ\kappa(P_{ng}))) \to L(\phi'_{\kappa(P)}(G)) \quad \text{on } (B_2,\mathcal{A}_2,d_2). \tag{4.16}$$

Thus if $\{P_n^*\}$ satisfies (4.9), we have

$$L_{P_n^*}(\sqrt{n}(\phi(T_n)-\phi\circ\kappa(P_n^*))) \to L(\phi'_{\kappa(P)}(G)) \quad \text{on } (B_2,\mathcal{A}_2,d_2). \ \blacksquare$$

Some of the conditions of Theorem 4.10 can be relaxed. One very useful weakening is that ϕ needs only be Hadamard differentiable tangentially to a subspace.

Secondly, of course $\{T_n\}$ need not be efficient to render $\{\phi(T_n)\}$ efficient. Given tightness of $\{L_P(\sqrt{n}(T_n-\kappa(P)))\}$, efficiency of $\{T_n\}$ requires efficiency of all marginals $\{b_1^*\circ T_n\}$, whereas for $\{\phi(T_n)\}$ to be efficient, efficiency of a subset of marginals suffices. This subset is generated by the linear maps in $R(\phi'^{*}_{\kappa(P)})$, the range of the adjoint of the derivative of ϕ. ($b_1^* \in R(\phi'^{*}_{\kappa(P)})$ iff it has the form $b_2^*\circ\phi'_{\kappa(P)}$ for a $b_2^* \in B_2^*$).

A close look at the proof of Theorem 4.10 yields

THEOREM 4.11. *Let* $\kappa: \mathcal{P} \to (B_1,d_1)$ *be differentiable at* $P \in \mathcal{P}$ *and* $\phi: (B_1,d_1) \to (B_2,d_2)$ *be Hadamard differentiable at* $\kappa(P)$ *tangentially to* $\text{lin}\{S, \kappa'_P(T(P))\}$, *with an* $\mathcal{A}_1$-$\mathcal{A}_2$- *measurable derivative. Assume that* $\{L_P(\sqrt{n}(T_n-\kappa(P)))\}$ *is tight with limiting distribution concentrating on* S *and that* $\{b_1^*\circ T_n\}$ *is efficient for* $b_1^*\circ\kappa$ *for every* $\mathcal{A}_1$*-measurable* $b_1^* \in R(\phi'^{*}_{\kappa(P)})$. *Then* $\phi\circ\kappa: \mathcal{P} \to (B_2,d_2)$ *is differentiable at* $P \in \mathcal{P}$ *and* $\{\phi(T_n)\}$ *is efficient for* $\phi\circ\kappa$ *at* $P \in \mathcal{P}$, *provided that* $\phi\circ t_n: (X^n,\mathcal{B}^n) \to (B_2,\mathcal{A}_2)$ *is measurable.* $\square$

PROOF. Tightness of $\{L_P(\sqrt{n}(T_n-\kappa(P)))\}$ and (4.5) imply relative compactness of $\{L_{P_n}(\sqrt{n}(T_n-\kappa(P_n)))\}$. Next the proof of Theorem 4.10 applies with minor changes, except that we need to show that the limit

points of $\{L_{P_n}(\sqrt{n}(T_n-\kappa(P_n)))\}$ concentrate on lin $\{S, \kappa_P'(T(P))\}$ for every $\{P_n\}$ satisfying (2.11) and (4.5), if this is true under the fixed P. This follows by the argument in the proof of Lemma 4.6. ■

The condition on efficiency of marginals in Theorem 4.11 may be replaced by: $\{\phi'_{\kappa(P)}\circ T_n\}$ is efficient for $\phi'_{\kappa(P)}\circ\kappa$ at $P \in \mathcal{P}$. Checking efficiency of $\{\phi'_{\kappa(P)}\circ T_n\}$ may next be simplified by Theorems 4.8-4.9. This will be utilized in Example 4.4.2.

4.4. EXAMPLES

From a range of possible applications we include two examples.

4.4.1. Models with right censored observations.

Let $\mathcal{F}$ and $\mathcal{G}$ be classes of distributions on $(0,\infty) \subset \mathbb{R}$ and let Y and C be independent random variables with distributions $F \in \mathcal{F}$ and $G \in \mathcal{G}$ respectively. Let $\mathcal{P} = \{ P_{FG}: F \in \mathcal{F}, G \in \mathcal{G} \}$ be the set of distributions P_{FG} of the pair $(Y\wedge C, 1\{Y\le C\})$ on $((0,\infty)\times\{0,1\}, \mathcal{B}\times\{\emptyset,\{0\},\{1\},\{0,1\}\})$.

This is the most popular censoring model (cf. e.g. Gill (1980)). Y is the variable of interest, the distribution of which we want to estimate. However, we observe the precise value of Y only on the stochastic interval $[0,C]$ and we must estimate (functionals of) F based on a sample from P_{FG}.

If μ is a σ-finite measure dominating both F and G then P_{FG} has a density

$$(4.17)\qquad f(x,\delta,F,G) = \delta f(x)(1-G(x-)) + (1-\delta)g(x)(1-F(x)) \quad (x\in(0,\infty),\delta\in\{0,1\})$$

with respect to $\mu\otimes$(counting measure on $\{0,1\}$). Here F and G denote both the measures and the cumulative distribution functions, and f and g are their densities with respect to μ.

The distribution of the pair $(Y\wedge C, 1\{Y\le C\})$ is completely determined by the *sub-distribution functions*

$$(4.18)\qquad \underline{\kappa}_0(P_{FG}) = P_{FG}(Y\wedge C \le \cdot,\ 1\{Y\le C\} = 0) = \int_{[0,\cdot]} (1-F(x))\ dG(x).$$

$$(4.19)\qquad \underline{\kappa}_1(P_{FG}) = P_{FG}(Y\wedge C \le \cdot,\ 1\{Y\le C\} = 1) = \int_{[0,\cdot]} (1-G(x-))\ dF(x).$$

Any functional of P_{FG} can therefore be expressed as a function $\phi\circ\underline{\kappa}(P_{FG})$, where $\underline{\kappa}(P_{FG}) = (\underline{\kappa}_0(P_{FG}),\underline{\kappa}_1(P_{FG}))$.

Now view the sub-distribution functions as elements of $D[0,\infty]$ and suppose that ϕ: $(D[0,\infty]\times D[0,\infty], \|\cdot\|_\infty v \|\cdot\|_\infty) \to (B,d)$ is Hadamard differentiable. Below we prove that $\underline{\kappa}$ is differentiable. Then Theorem 4.10 implies that efficient estimator sequences of $\phi\circ\underline{\kappa}$ can be found as $\{\phi(T_n)\}$, where $\{T_n\} = \{(T_{n0},T_{n1})\}$ is efficient for $\underline{\kappa}$. This is convenient, as estimation of $\underline{\kappa}(P_{FG})$ may be easier than direct estimation of $\phi\circ\underline{\kappa}$. Moreover once Hadamard differentiability has been established for a specific function, Theorem 4.10 can be applied to obtain efficient estimators for $\phi\circ\underline{\kappa}$ in a range of problems $\mathcal{P}$, each time applying ϕ to a $\{T_n\}$ which is efficient for $\underline{\kappa}$ at $P_{FG} \in \mathcal{P}$.

Many interesting functionals in the censoring problem *can* be written as Hadamard differentiable functions ϕ of $\underline{\kappa}$. An important example is the cumulative distribution function F. It is well-known how to obtain F from $\underline{\kappa}(P_{FG})$. First

$$1 - \underline{\kappa}_0(P_{FG}) - \underline{\kappa}_1(P_{FG}) = (1-F)(1-G)\ .$$

Next the *cumulative hazard function* of F is defined as

$$(4.20)\qquad \Lambda_F(t) = \int_{[0,t]} (1 - F(s-))^{-1}\ dF(s)\ ,$$

and equals

$$= \int_{[0,t]} \frac{d\underline{\kappa}_1(P_{FG})(s)}{(1-\underline{\kappa}_0(P_{FG})(s-)-\underline{\kappa}_1(P_{FG})(s-))}\ .$$

The final step is

$$F(t) = \prod_{s\le t} (1 - d\Lambda_F(s))\ ,$$

which is a *product integral*, as defined in Gill and Johansen (1987).

Now let $\tau \in \mathbb{R}$ be such that $\max(F(\tau),G(\tau)) < 1$. It is shown in Gill and Johansen (1987) (also cf. Gill (1987)) that the maps

$$(K_0(\cdot),K_1(\cdot)) \to \int_{[0,\cdot]} \frac{dK_1(s)}{(1-K_0(s-)-K_1(s-))} = \Lambda \to \prod_{s\le\cdot} (1 - d\Lambda(s)) ,$$

extend from their natural domains of definition, to maps

$$D[0,\tau]\times D[0,\tau] \to D[0,\tau] \to D[0,\tau] ,$$

which are Hadamard differentiable at $\underline{\kappa}(P_{FG})$ and Λ_F, respectively[3]. Thus by the Chain rule for Hadamard differentiation the composite map taking $\underline{\kappa}(P_{FG})$ into F can be extended to an (at $\underline{\kappa}(P_{FG})$) Hadamard differentiable map ϕ: $(D[0,\infty]\times D[0,\infty],\|\cdot\|_\infty v\|\cdot\|_\infty) \to (D[0,\tau],\|\cdot\|_\infty)$.

Combination with results in Reeds (1976), Gill (1987) and Fernholz (1983) and the chain rule for Hadamard differentiation, gives a large class of interesting functionals, such as the quantile function of F, moments and L, M and R-functions. We do not discuss these or any other examples any further. Instead we consider differentiability of $\underline{\kappa}$, and efficient estimation of $\underline{\kappa}$ in a special model.

LEMMA 4.12. *Let* $\underline{\kappa}$: $\mathcal{P} \to (D[0,\infty]\times D[0,\infty],\|\cdot\|_\infty v\|\cdot\|_\infty)$ *be given by* $\underline{\kappa} = (\underline{\kappa}_0,\underline{\kappa}_1)$, *where* $\underline{\kappa}_0$ *and* $\underline{\kappa}_1$ *are given by* (4.18)-(4.19). *Then* $\underline{\kappa}$ *is differentiable at* $P_{FG} \in \mathcal{P}$ *(with respect to any tangent cone) with derivative*

$$\underline{\kappa}'_{P_{FG}}(g) = (\langle(1-\delta)1_{[0,\cdot]}(x),g(x,\delta)\rangle_{P_{FG}}, \langle\delta 1_{[0,\cdot]}(x),g(x,\delta)\rangle_{P_{FG}}). \quad (4.22)$$ □

PROOF. Let $\{P_{F_tG_t}\}$ satisfy the specialization of (2.11) to the present situation, i.e., assuming without loss of generality that F_t, G_t, F and G have densities f_t, g_t, f and g with respect to a σ-finite dominating measure μ, we have

---

3) *The extensions depend on* $\underline{\kappa}(P_{FG})$ *and* Λ_F, *and Hadamard differentiability is tangentially to a subset. This is of no consequence for the following.*

$$\int [t^{-1}(g_t^{\frac12}(1-F_t)^{\frac12} - g^{\frac12}(1-F)^{\frac12}) - \tfrac12 g(\cdot,0)g^{\frac12}(1-F)^{\frac12}]^2\, d\mu$$

(4.23)

$$+ \int [t^{-1}(f_t^{\frac12}(1-G_t(\cdot\cdot))^{\frac12} - f^{\frac12}(1-G(\cdot\cdot))^{\frac12}) - \tfrac12 g(\cdot,1)f^{\frac12}(1-G(\cdot\cdot))^{\frac12}]^2 d\mu \to 0.$$

Then uniformly in $u \in [0,\infty]$

$$t^{-1}(\underline{\kappa}_0(P_{F_tG_t}) - \underline{\kappa}_0(P_{FG}))(u)$$

(4.24)

$$= t^{-1}\int 1_{\{[0,u]\}}\, [g_t^{\frac12}(1-F_t)^{\frac12} - g^{\frac12}(1-F)^{\frac12}]\, [g_t^{\frac12}(1-F_t)^{\frac12} + g^{\frac12}(1-F)^{\frac12}]\, d\mu$$

$$\to \int 1_{\{[0,u]\}}\, g(\cdot,0)\, g(1-F)\, d\mu = \langle (1-\delta)\, 1_{\{[0,u]\}}(x),\, g(x,\delta)\rangle_{P_{FG}}.$$

This proves half of the assertion. The proof of the other half is almost identical. ■

In the special case that F and G are both equal to the class of all distributions on $(0,\infty)$, it can be proved that P is the class of all distributions on $(0,\infty)\times\{0,1\}$. Then we may set the tangent cone equal to $L_{2*}(P_{FG})$. For a more general treatment we compute a tangent cone explicitly. Suppose that $\{F_t\} \subset F$ and $\{G_t\} \subset G$ satisfy

$$\int [t^{-1}((dF_t)^{\frac12}-(dF)^{\frac12}) - \tfrac12 a_1(dF)^{\frac12}]^2 \to 0, \tag{4.25}$$

respectively

$$\int [t^{-1}((dG_t)^{\frac12}-(dG)^{\frac12}) - \tfrac12 a_2(dG)^{\frac12}]^2 \to 0. \tag{4.26}$$

It follows from Proposition A.11 that (4.25)-(4.26) imply

$$\int [t^{-1}((dP_{F_tG_t})^{\frac12}-(dP_{FG})^{\frac12}) - \tfrac12(A_1a_1+A_2a_2)(dP_{FG})^{\frac12}]^2 \to 0, \tag{4.27}$$

where $A_i = A_i(F,G)$ are continuous, linear operators from $L_{2*}(F)$, respectively $L_{2*}(G)$ to $L_{2*}(P_{FG})$ given by

$$A_1a_1(x,\delta) = \delta a_1(x) + (1-\delta)\int_{(x,\infty)} a_1\, dF\, /\, (1-F(x)) \tag{4.28}$$

$$(4.29)\qquad A_2a_2(x,\delta) = \delta \int_{[x,\infty)} a_2 \, dG /(1-G(x-)) + (1-\delta)a_2(x).$$

As explained in Section A.3 this has the intuitive meaning that the scores $a_1(Y) + a_2(C)$ measure information in the situation that one would observe (Y,C) instead of $(Y\wedge C, 1\{Y\le C\})$, while

$$A_1a_1(x,\delta) + A_2a_2(x,\delta) = E_{P_{FG}}(a_1(Y) + a_2(C) \mid Y\wedge C = x, 1\{Y\le C\} = \delta) .$$

If $T(F,\mathcal{F})$ respectively $T(G,\mathcal{G})$ are tangent cones at $F \in \mathcal{F}$ and $G \in \mathcal{G}$ then we may set

$$(4.30)\qquad T(P_{FG}) = \{A_1a_1+A_2a_2:\ a_1 \in T(F,\mathcal{F}),\ a_2 \in T(G,\mathcal{G})\}.$$

In this model Y is the variable of interest which is censored by C and it is therefore natural to consider functionals $\chi: \mathcal{P} \to (B_2,d_2)$ which are in fact functionals of F only, i.e.

$$(4.31)\qquad \chi(P_{FG}) = \psi(F).$$

Then, if χ is differentiable, for any $a_2 \in T(G,\mathcal{G})$ and $\{G_t\}$ as in (4.26),

$$(4.32)\qquad \chi'_{P_{FG}}(A_2a_2) = \lim_{t\to 0} t^{-1}(\chi(P_{FG_t}) - \chi(P_{FG})) = 0.$$

This implies that for any $b^* \in B_2^*$ and $P \in \mathcal{P}$ the gradients of χ satisfy

$$(4.33)\qquad \dot{\chi}_{b^*}(\cdot,P) \perp A_2T(G,\mathcal{G}).$$

Relation (4.33) has important consequences when applying Theorems 4.10 and 4.11 to functionals ϕ of $\underline{\kappa}$ with

$$(4.34)\qquad \phi\circ\underline{\kappa}(P_{FG}) = \psi(F).$$

LEMMA 4.13. *Let $\mathcal{G}'$ be the set of all probability distributions on* $(0,\infty)$, $\mathcal{P}' = \{P_{FG}: F \in \mathcal{F}, G \in \mathcal{G}'\}$ *and suppose that* $\{T_n\} = \{(T_{n0},T_{n1})\}$ *is efficient for a differentiable functional* $\kappa: \mathcal{P} \to (B_1,d_1)$ *at* $P_{FG} \in \mathcal{P}'$. *Furthermore let* $\phi: (B_1,d_1) \to (B_2,d_2)$ *of the form* (4.34) *be Hadamard differentiable at* $\kappa(P)$, *measurable and with a measurable derivative. Then* $\{\phi(T_n)\}$ *is efficient for* $\phi\circ\kappa$ *at* $P_{FG} \in \mathcal{P} = \{P_{FG}: F \in \mathcal{F}, G \in \mathcal{G}\}$, *for any* $\mathcal{G}$. □

PROOF. Since $\mathcal{P} \subset \mathcal{P}'$ the canonical gradients of $\phi\circ\kappa$ for the model $\mathcal{P}$ can be obtained as the projection of the canonical gradients for the model $\mathcal{P}'$ onto the closure of the linear space spanned by the tangent cone for $\mathcal{P}$, (which is given by (4.30)). Now the tangent cones for $\mathcal{P}$ and $\mathcal{P}'$ differ only by a subset of $A_2L_{2*}(P_{FG})$ and by (4.33) the canonical gradients of $\phi\circ\kappa$ for $\mathcal{P}'$ are orthogonal to this set. We conclude that the canonical gradients of $\phi\circ\kappa$ for the models $\mathcal{P}'$ and $\mathcal{P}$ are identical. Next if $\{T_n\}$ is efficient for κ in $P_{FG} \in \mathcal{P}'$, then by Theorem 4.10 $\{\phi(T_n)\}$ is efficient for $\phi\circ\kappa$ at $P_{FG} \in \mathcal{P}'$. As the canonical gradients of $\phi\circ\kappa$ for the two models are identical, $\{\phi(T_n)\}$ is efficient for $\phi\circ\kappa$ at $P_{FG} \in \mathcal{P}$. ■

By the same argument we see that to estimate a functional χ of the form (4.31), knowing G is no advantage in terms of the asymptotic lower bound. This result depends on the special form of χ as well as on the fact that the contributions of $\mathcal{F}$ and $\mathcal{G}$ to the tangent cone are additive (cf. (4.30)).

Finally we consider an efficient estimator sequence $\{T_n\}$ for $\underline{\kappa}$ for the situation that $T(F,\mathcal{F}) = L_{2*}(F)$ and $T(G,\mathcal{G}) = L_{2*}(G)$. The latter corresponds roughly to the situation wherein F and G are completely unknown. As remarked before, in this case P_{FG} is completely unknown too, so that we expect $T(P_{FG}) = L_{2*}(P_{FG})$. The validity of this equation can be shown under the above condition on $T(F,\mathcal{F})$ and $T(G,\mathcal{G})$ by using (4.28)-(4.29). If P_{FG} is completely unknown, then the empirical distribution is efficient as an estimator for P_{FG}. This is usually recast in terms of the pair of *empirical subdistribution functions*, $T_n = (T_{n0},T_{n1})$, where,

$$(4.35) \qquad T_{ni}(u) = n^{-1}\Sigma_{j=1}^{n}\, 1_{\{\Delta_j=i\}}\, 1_{[0,u]}(\underline{X}_j) \qquad (i = 0,1),$$

and $(\underline{X}_1,\Delta_1),\ldots,(\underline{X}_n,\Delta_n)$ is an i.i.d. sample from P_{FG}.

Tightness of $\{L_P(\sqrt{n}(T_{ni}-\underline{\kappa}_i(P)))\}$ can be shown by standard methods for proving tightness of empirical processes (cf. Pollard (1984)). By Theorems 4.8-4.9 we can next conclude that $\{T_n\}$ is efficient for $\underline{\kappa}$ in $(D[0,\tau]\times D[0,\tau], U(\Pi)\times U(\Pi), \|\cdot\|_\infty v \|\cdot\|_\infty)$, if $\pi_u \circ T_{ni}$ is efficient for $\pi_u \circ \underline{\kappa}_i$ ($u \in [0,\infty]$, $i = 0,1$). By Lemma 4.12

$$\dot{\underline{\kappa}}_{i,\pi_u}(x,\delta,P) = 1\{\delta = i\}\, 1_{[0,u]}(x) \tag{4.36}$$

Under the assumption that $T(P_{FG}) = L_{2^*}(P_{FG})$, the canonical gradients equal

$$\tilde{\underline{\kappa}}_{i,\pi_u}(x,\delta,P) = 1\{\delta = i\}\, 1_{[0,u]}(x) - \pi_u \circ \underline{\kappa}_i(P) \ .$$

Clearly

$$\sqrt{n}\ (\pi_u \circ T_{ni} - \pi_u \circ \underline{\kappa}_i(P)) = n^{-\frac{1}{2}} \Sigma_{j=1}^n \tilde{\underline{\kappa}}_{i,\pi_u}(\underline{X}_j, \Delta_j, P) \qquad (i = 0,1),$$

so that $\{\pi_u \circ T_{ni}\}$ is efficient for $\pi_u \circ \underline{\kappa}_i(P)$ at any $P \in \mathcal{P}$ ($i = 0,1$, cf.(4.7)).

Application of the functional ϕ given through (4.21) to T_n given by (4.35) yields

$$\hat{F}_n(t) = \prod_{s \le t} \left(1 - \frac{dT_{n1}(s)}{(1-T_{n0}(s-)-T_{n1}(s-))}\right),$$

which is well-defined on $[0,\tau]$ as soon as there is an observation greater than τ. In that case this formula precisely gives the *product limit estimator* (cf. Gill and Johansen (1987)). Thus the product limit estimator is efficient in the case that $T(F,G) = L_{2^*}(F)$ (G arbitrary).

Note how the above argument clarifies this fact, which is easily obscured by lengthy computations. The censoring problem considered here is a fully nonparametric problem. Hence the empirical distribution is an efficient estimator of the underlying distribution P_{FG}. Next the functional of interest F is a smooth (though somewhat complicated) function of P_{FG}. As to be expected this functional applied to the empirical gives an efficient estimator of F.

Weak convergence of the product limit estimator to a Gaussian process

was first proved in Breslow and Crowley (1974); efficiency in Wellner (1982).

4.4.2. Estimating a Point Symmetric Distribution on $\mathbb{R}^k$.

Let H be a class of densities with respect to Lebesgue measure on $\mathbb{R}^k$, which are pointsymmetric about zero

$$\eta(x) = \eta(-x) \ . \tag{4.37}$$

Given $\theta \in \mathbb{R}^k$ and $\eta \in H$ let $P_{\theta\eta}$ be the probability distribution on $\mathbb{R}^k$ with density $p(\cdot,\theta,\eta) = \eta(\cdot-\theta)$ and set $\mathcal{P} = \{P_{\theta\eta}: \theta \in \mathbb{R}^k, \eta \in H\}$.

Suppose that $\mathcal{C}$ is a collection of Borel subsets of $\mathbb{R}^k$, which is symmetric: $\mathcal{C} = -\mathcal{C}$ and translation invariant: $\mathcal{C} = \mathcal{C} + a$. The space $B(\mathcal{C})$ is discussed in Sections 3.5.4 and 3.6.2.

We consider estimation of the functional $\chi: \mathcal{P} \to B(\mathcal{C})$ given by

$$\chi(P_{\theta\eta})(C) = P_{\theta\eta}(C) = \int_{C-\theta} \eta(x)\, dx \ .$$

(The functional $C \to \int_C \eta(x)\, dx$ can be handled in a similar manner). Given an i.i.d. sample from $X_1,\ldots,X_n$ an efficient estimator for χ is based on the empirical measure

$$\hat{P}_n(C) = n^{-1}\Sigma_{j=1}^n 1_C(X_j) \ ,$$

and an efficient estimator for θ.

To avoid difficulties with measurability we consider the subspace B of $B(\mathcal{C})$ consisting of all $h \in B(\mathcal{C})$ satisfying

$$h(C_n) \to h(C) \qquad \text{if } 1_{C_n}(x) \to 1_C(x) \text{ for all } x \in \mathbb{R}^k \ .$$

Next we assume that the class $\mathcal{C}$ is such that the projection σ-field on B satisfies (4.1) and

$$L_{P_{\theta\eta}}(\sqrt{n}(\hat{P}_n - \chi(P_{\theta\eta}))) \to L(B_{\theta\eta}) \qquad \text{on } (B,\mathcal{U}(\Pi),\|\cdot\|_\infty) \ , \tag{4.38}$$

where $B_{\theta\eta}$ is $P_{\theta\eta}$-Bridge as defined in e.g. Pollard (1984, p.149). Thus

$L(B_{\theta\eta})$ is supported by the subspace $UC(C,\rho_{\theta\eta})$ of B consisting of all functions $F: C \to \mathbb{R}$ which are uniformly continuous with respect to the semi-metric

$$\rho_{\theta\eta}(C_1,C_2) = P_{\theta\eta}(C_1 \Delta C_2)^{\frac{1}{2}} .$$

Because $(C,\rho_{\theta\eta})$ is necessarily totally bounded, the space $UC(C,\rho_{\theta\eta})$ is separable (cf. Pollard (1984, Exercises VII-3,-7).

The above assumptions are satisfied if C e.g. equals the set of all closed balls or rectangles. (For condition (4.1) use that any $h \in B$ is determined by its values on a countable subset of C).

To obtain an efficient estimator $\{\hat{\theta}_n\}$ for θ requires generalization of the construction of e.g. Stone (1975) (cf. Section 1.3) to higher dimensions. This is done in Section 5.7.2. Suppose that η is absolutely continuous on $\mathbb{R}^k$ in the sense that

$$\eta(x+h) - \eta(x) = \int_0^1 h'\nabla\eta(x+uh)\, du \qquad (x,h \in \mathbb{R}^k) , \tag{4.39}$$

for some function $\nabla\eta: \mathbb{R}^k \to \mathbb{R}^k$ satisfying

$$\int \|\nabla\eta/\eta^{\frac{1}{2}}(x)\|^2 \, dx < \infty. \tag{4.40}$$

Then $\{\hat{\theta}_n\}$ is efficient for θ if

$$\sqrt{n}(\hat{\theta}_n - \theta) = n^{-\frac{1}{2}} \Sigma_{j=1}^n I^{-1}(\eta)\, \nabla\eta/\eta(X_j-\theta) + o_{P_{\theta\eta}}(1) , \tag{4.41}$$

(cf. Chapter 5). Here $I(\eta) = \int \nabla\eta/\eta(x)\nabla\eta'/\eta(x)\, \eta(x)\, dx$ is assumed to be nonsingular.

Now set

$$T_n(C) = \tfrac{1}{2} [\hat{P}_n(C) + 1 - \hat{P}_n(2\hat{\theta}_n - C)] .$$

We show that Theorem 4.11 implies efficiency of $\{T_n\}$ for χ in $(B,U(\Pi),\|\cdot\|_\infty)$.

First note that we may choose as a tangent cone

$$T(P_{\theta\eta}) = \{h'\nabla\eta/\eta(x-\theta) + b(x-\theta):\ h \in \mathbb{R}^k,\ b \in L_{2^*}(P_{0\eta}),\ b(x) = b(-x)\}.$$

Next set $\kappa(P_{\theta\eta}) = (\chi(P_{\theta\eta}),\theta)$. We have

LEMMA 4.14. *Let* (4.37) *and* (4.39)-(4.40) *hold. Then the map* ϕ: $(B(C)\times\mathbb{R}^k, \|\cdot\|_\infty v |\cdot|) \to (B(C), \|\cdot\|_\infty)$ *given by*

$$\phi(F,\mu) = \tfrac{1}{2}\,[F(\cdot) + 1 - F(2\mu-\cdot)]$$

is Hadamard differentiable at $\kappa(P_{\theta\eta})$, *tangentially to* $UC(C,\rho_{\theta\eta})\times\mathbb{R}^k$. *Moreover*

$$(4.41)\qquad \phi'_{\kappa(P_{\theta\eta})}(F,\mu) = \tfrac{1}{2}[F(\cdot)-F(2\theta-\cdot)] + \int \mu'\nabla\eta(x-\theta)\,dx\ .\ \square$$

PROOF. Let $\{G_n\} \subset B(C)$, $\|G_n-G\|_\infty \to 0$, where $G \in UC(C,\rho_{\theta\eta})$; $\nu_n \to \nu$ in $\mathbb{R}^k$ and $t_n \downarrow 0$ in $\mathbb{R}$. It is easily seen that

$$\sup_{C\in\mathcal{C}} P_{\theta\eta}((C + 2t_n\nu_n)\ \Delta\ C) \to 0\ .$$

Hence, since $G \in UC(C,\rho_{\theta\eta})$

$$\|G_n(2\theta+2t_n\nu_n-\cdot) - G(2\theta-\cdot)\|_\infty \le$$

$$\|G(2\theta+2t_n\nu_n-\cdot) - G(2\theta-\cdot)\|_\infty + \|G_n(\cdot) - G(\cdot)\|_\infty \to 0\ .$$

It can now be checked that

$$t_n^{-1}[\phi(\chi(P_{\theta\eta})+t_nG_n,\ \theta+t_n\nu_n) - \phi(\chi(P_{\theta\eta}),\theta)]$$

$$\to \tfrac{1}{2}[G(\cdot)-G(2\theta-\cdot)] + \int \nu'\nabla\eta(x-\theta)\,dx\ .\ \blacksquare$$

Finally we show efficiency of $\{T_n\} = \{\phi(\hat{P}_n,\hat{\theta}_n)\}$ for $\phi\circ\kappa(P_{\theta\eta}) = \chi(P_{\theta\eta})$, under the assumption that H is the class of all η satisfying (4.37) and (4.39)-(4.40).

$\{L_{P_{\theta\eta}}(\sqrt{n}(\hat{P}_n - \chi(P_{\theta\eta})), \sqrt{n}(\hat{\theta}_n - \theta))\}$ is tight by completeness of B,

(4.38) and (4.40)-(4.41). Its limit distributions concentrate on $UC(C,\rho_{\theta\eta})\times\mathbb{R}^k$. It can be checked that the range of $\kappa'_{P_{\theta\eta}}$ is also contained in this set. Furthermore $\phi'_{\kappa(P_{\theta\eta})}$ is $\mathcal{U}(\Pi)\times\mathcal{B}^k$-$\mathcal{U}(\Pi)$ measurable, since for every $C \in \mathcal{C}$ the composite map

$$(F,\mu) \to (F(C),\ F(2\theta-C),\ \mu'_C\textstyle\int \nabla\eta(x-\theta)\,dx) \to \phi'_{\kappa(P_{\theta\eta})}(C)\ ,$$

is $\mathcal{U}(\Pi)\times\mathcal{B}^k$-$\mathcal{B}$ measurable.

By Theorem 4.11 it suffices to show efficiency of $\{\phi'_{\kappa(P_{\theta\eta})}(\hat{P}_n,\ \hat{\theta}_n)\}$ for $\phi'_{\kappa(P_{\theta\eta})}\circ\kappa$ at $P_{\theta\eta}$. By Theorem 4.9 it next suffices to show efficiency of $\{\pi_C\circ\phi'_{\kappa(P_{\theta\eta})}(\hat{P}_n,\ \hat{\theta}_n)\}$ for $\pi_C\circ\phi'_{\kappa(P_{\theta\eta})}\circ\kappa$ at $P_{\theta\eta}$ for every $C \in \mathcal{C}$.

Now

$$\sqrt{n}\ \{\pi_C\circ\phi'_{\kappa(P_{\theta\eta})}(\hat{P}_n,\ \hat{\theta}_n) - \pi_C\circ\phi'_{\kappa(P_{\theta\eta})}\circ\kappa\ (P_{\theta\eta},\ \theta)\} =$$

$$n^{-\frac{1}{2}}\ \Sigma_{j=1}^{n}\tfrac{1}{2}\{1_C(X_j)-1_{2\theta-C}(X_j)\} + \sqrt{n}(\hat{\theta}_n-\theta)'_C\textstyle\int \nabla\eta(x-\theta)\,dx\ .$$

Hence $\{\pi_C\circ\phi'_{\kappa(P_{\theta\eta})}(\hat{P}_n,\ \hat{\theta}_n)\}$ is asymptotically linear with asymptotic influence function

$$\tfrac{1}{2}\{1_C(\cdot)-1_{2\theta-C}(\cdot)\}\ +\ _C\textstyle\int \nabla\eta(x-\theta)\,dx'I^{-1}(\eta)\ \nabla\eta/\eta(\cdot-\theta)\ .$$

This is an influence function of the functional $\pi_C\circ\phi\circ\kappa$ at $P_{\theta\eta}$, hence of $\pi_C\circ\phi'_{\kappa(P_{\theta\eta})}\circ\kappa$, and is contained in the tangent cone.

CHAPTER 5

SEMI-PARAMETRIC MODELS WITH A SUFFICIENT STATISTIC FOR THE NUISANCE PARAMETER

5.1. INTRODUCTION

Having obtained bounds on the asymptotic behaviour of estimators in the foregoing chapters, we now turn to the construction of estimators. Since no general construction method is available at present, we focus on a special class of models of the semi-parametric type. Here our main aim is the estimation of a Euclidean parameter θ, in models characterized by the existence of a suitable sufficient statistic for the (typically) infinite dimensional parameter η.

The constructions in this chapter pertain both to models with i.i.d. and with independent but not necessarily identically distributed observations, where a new nuisance parameter is introduced with every new observation. The study of lower bounds in the i.i.d. version of the model goes back to Klaassen and van Zwet (1985) (also cf. Klaassen, van der Vaart and van Zwet (1987)). The construction of estimators given below is based on van der Vaart (1986a), though the non-i.i.d. version of the model in the latter paper is in a sense orthogonal to the non-i.i.d. case considered below.

Probability distributions in this chapter are usually given through their densities. For this reason we write, with a slight abuse of notation, $L_2(p(\cdot,\theta,\eta))$ for $L_2(P_{\theta\eta})$, a tangent cone as $T(p(\cdot,\theta,\eta))$ instead of $T(P_{\theta\eta})$, and so on.

Let Θ be an open subset of $\mathbb{R}^k$ and H an arbitrary set. For every

$(\theta,\eta) \in \Theta\times H$ let $p(\cdot,\theta,\eta)$ be a density with respect to a σ-finite measure μ on a measurable space $(X,\mathcal{B})$. For easy notation X denotes a random element in $(X,\mathcal{B})$ with density $p(\cdot,\theta,\eta)$. This general situation is discussed in Example 2.3 and we adopt the notation introduced there.

The model is further specified by the existence, for every fixed θ, of an $\mathbb{R}^m$-valued statistic $\psi(X,\theta)$ which is sufficient for $\eta \in H$. By the factorization theorem this property of the model can also be expressed in the following way. There exist measurable functions $h(\cdot,\theta)$: $(X,\mathcal{B}) \to \mathbb{R}$ and $g(\cdot,\theta,\eta)$: $\mathbb{R}^m \to \mathbb{R}$ and a measure ν_θ on $\mathbb{R}^m$ with

$$p(\cdot,\theta,\eta) = h(\cdot,\theta)\, g(\psi(\cdot,\theta),\theta,\eta) \qquad \text{a.e. } [\mu] \tag{5.1}$$

$$\psi(X,\theta) \text{ has density } g(\cdot,\theta,\eta) \text{ w.r.t. } \nu_\theta. \tag{5.2}$$

The two ways of characterizing the model will be used interchangeably.

As in Example 2.3 we assume the existence of $\dot{\ell}(\cdot,\theta,\eta) \in L_2(p(\cdot,\theta,\eta))^k$ such that for every $h \in \mathbb{R}^k$ as $t \to 0$

$$\int [t^{-1}(p^{\frac{1}{2}}(x,\theta+th,\eta)-p^{\frac{1}{2}}(x,\theta,\eta)) - \tfrac{1}{2}h'\dot{\ell}(x,\theta,\eta)p^{\frac{1}{2}}(x,\theta,\eta)]^2 \, d\mu(x) \to 0. \tag{5.3}$$

Also, we define scores for η as elements b of $L_2(p(\cdot,\theta,\eta))$ for which there exists $\{\eta_t\} \subset H$ with as $t\downarrow 0$

$$\int [t^{-1}(p^{\frac{1}{2}}(x,\theta,\eta_t)-p^{\frac{1}{2}}(x,\theta,\eta)) - \tfrac{1}{2}b(x)p^{\frac{1}{2}}(x,\theta,\eta)]^2 \, d\mu(x) \to 0. \tag{5.4}$$

Since $p(\cdot,\theta,\eta)$ depends on η only through $\psi(\cdot,\theta)$, scores for η are necessarily functions of the sufficient statistic $\psi(\cdot,\theta)$. Formally we have

LEMMA 5.1. *Let* (5.1)-(5.2) *hold. Then* (5.4) *holds if and only if there exists* $\underline{b} \in L_2(g(\cdot,\theta,\eta))$ *with*

$$\int [t^{-1}(g^{\frac{1}{2}}(s,\theta,\eta_t)-g^{\frac{1}{2}}(s,\theta,\eta)) - \tfrac{1}{2}\underline{b}(s)g^{\frac{1}{2}}(s,\theta,\eta)]^2 \, d\nu_\theta(s) \to 0 \tag{5.5}$$

$$b(x) = \underline{b}(\psi(x,\theta)). \quad \square \tag{5.6}$$

PROOF. (5.4) is equivalent to the pair of assumptions

$$P_{\theta\eta_t}(p(X,\theta,\eta) = 0) = o(t^2) \tag{5.7}$$

$$E_{\theta\eta}[t^{-1}p^{-\frac{1}{2}}(X,\theta,\eta)(p^{\frac{1}{2}}(X,\theta,\eta_t)-p^{\frac{1}{2}}(X,\theta,\eta)) - \tfrac{1}{2}b(X)]^2 \to 0, \tag{5.8}$$

as $t\downarrow 0$. On the other hand, writing V for $\psi(X,\theta)$,we have that (5.5) is equivalent to

$$P_{\theta\eta_t}(g(V,\theta,\eta) = 0) = o(t^2) \tag{5.9}$$

$$E_{\theta\eta}[t^{-1}g^{-\frac{1}{2}}(V,\theta,\eta)(g^{\frac{1}{2}}(V,\theta,\eta_t)-g^{\frac{1}{2}}(V,\theta,\eta)) - \tfrac{1}{2}\underline{b}(V)]^2 \to 0. \tag{5.10}$$

By (5.1)-(5.2) we see that (5.7) and (5.9) are equivalent and that (5.8) follows from (5.10) when (5.6) holds. Finally assume that (5.8) holds. Using (5.1) and the L_2-projection property of a conditional expectation, we see

$$\begin{aligned} &E_{\theta\eta}\,[E_\theta(b(X)\,|V) - b(X)]^2 \\ &\le 4E_{\theta\eta}[t^{-1}g^{-\frac{1}{2}}(V,\theta,\eta)(g^{\frac{1}{2}}(V,\theta,\eta_t)-g^{\frac{1}{2}}(V,\theta,\eta)) - \tfrac{1}{2}b(X)]^2 \to 0. \end{aligned} \tag{5.11}$$

Hence $b(x) = E_\theta(b(X)\,|V = \psi(x,\theta))$ a.e. $[p(\cdot,\theta,\eta)]$. ■

Let $\underline{T}_\eta(g(\cdot,\theta,\eta)) \subset L_2(g(\cdot,\theta,\eta))$ be a cone. We consider a tangent cone for the model given by (5.1)-(5.2), of the form

$$T(p(\cdot,\theta,\eta)) = \{h'\dot{\ell}(\cdot,\theta,\eta) + \underline{b}(\psi(\cdot,\theta)):\ h \in \mathbb{R}^k,\ \underline{b} \in \underline{T}_\eta(g(\cdot,\theta,\eta))\}. \tag{5.12}$$

Here we assume that for every $h \in \mathbb{R}^k$ and $\underline{b} \in \underline{T}_\eta(g(\cdot,\theta,\eta))$ there exists $\{\eta_t\} \subset H$ with

$$\begin{aligned} \int [t^{-1}&(p^{\frac{1}{2}}(x,\theta+th,\eta_t)-p^{\frac{1}{2}}(x,\theta,\eta)) \\ &- \tfrac{1}{2}(h'\dot{\ell}(x,\theta,\eta)+\underline{b}(\psi(x,\theta)))\ p^{\frac{1}{2}}(x,\theta,\eta)]^2\ d\mu(x) \to 0. \end{aligned} \tag{5.13}$$

as $t\downarrow 0$. The further analysis of the model is inspired by a property, shared by quite a number of examples, which we call *local completeness.* Informally, $\psi(X,\theta)$ is strongly locally complete at (θ,η) if any function $\underline{b}$

of $\psi(\cdot,\theta)$ can be obtained as an η-score. Because of Lemma 5.1 and (5.5) this essentially entails that the density $g(\cdot,\theta,\eta)$ can be locally approximated by a sequence $\{g(\cdot,\theta,\eta_t)\}$, from any direction in $L_{2*}(g(\cdot,\theta,\eta))$, just as in the situation where $g(\cdot,\theta,\eta)$ would be completely unknown.

The precise definition is a bit more complicated, due to the fact that a tangent cone ought to be convex so as to have a convolution and LAM theorem (cf. Chapter 2).

DEFINITION 5.2. *Let* (5.1)-(5.2) *hold. Then* $\psi(X,\theta)$ *is called locally complete at* (θ,η) *if there exists a convex cone* $\underline{T}_\eta(g(\cdot,\theta,\eta)) \subset L_2(g(\cdot,\theta,\eta))$, *for which* $\operatorname{lin} \underline{T}_\eta(g(\cdot,\theta,\eta))$ *is dense in* $L_{2*}(g(\cdot,\theta,\eta))$ *and for which there exists for every* $h \in \mathbb{R}^k$ *and* $\underline{b} \in \underline{T}_\eta(g(\cdot,\theta,\eta))$ *a sequence* $\{\eta_t\} \subset H$ *satisfying* (5.13). *Furthermore we call* $\psi(X,\theta)$ *strongly locally complete if* $\underline{T}_\eta(g(\cdot,\theta,\eta))$ *can be chosen equal to* $L_{2*}(g(\cdot,\theta,\eta))$. □[1)]

5.1.1. The i.i.d. model

Consider estimation of θ based on i.i.d. random elements $X_1,\ldots,X_n$ with density $p(\cdot,\theta,\eta)$ with respect to μ. In this case asymptotic bounds on the performance of estimators are determined by the tangent cone. Let

$$(5.14)\quad T_\eta(p(\cdot,\theta,\eta)) = \{b \in L_2(p(\cdot,\theta,\eta)):\ b(x) = \underline{b}(\psi(x,\theta)),\ \underline{b} \in \underline{T}_\eta(g(\cdot,\theta,\eta))\}.$$

Under local completeness the closure of $\operatorname{lin} T_\eta(p(\cdot,\theta,\eta))$ is the set of all functions of $\psi(\cdot,\theta)$ in $L_{2*}(p(\cdot,\theta,\eta))$. By the properties of a conditional expectation the projection of $\dot{\ell}(\cdot,\theta,\eta)$ onto this set equals its conditional expectation given $\psi(\cdot,\theta)$. Hence the efficient influence function for the estimation of $\kappa(p(\cdot,\theta,\eta)) = \theta$ is given by (cf. (2.6)-(2.7))

$$(5.15)\quad \tilde{\kappa}(\cdot,P_{\theta\eta}) = \tilde{I}^{-1}(\theta,\eta)\, \dot{\tilde{\ell}}(\cdot,\theta,\eta),$$

where

1) Strictly speaking local completeness is not a property of $\psi(X,\theta)$, nor of its set of distributions. The motivation for the terminology is that it *would* be a property of $\psi(X,\theta)$ if (5.3)-(5.6) were equivalent to (5.13).

$$\dot{\ell}(\cdot,\theta,\eta) = \ell(\cdot,\theta,\eta) - E_\theta(\ell(X,\theta,\eta)\,|\,\psi(X,\theta) = \psi(\cdot,\theta)). \tag{5.16}$$

Next we give the idea behind the construction of an estimator sequence $\{T_n\}$, $T_n = t_n(X_1,\ldots,X_n)$, for θ. Under (5.15)-(5.16) $\{T_n\}$ is efficient for θ (in the sense of Definition 4.3; also cf. Section 5.4) if

$$\sqrt{n}(T_n-\theta) = n^{-\frac{1}{2}} \Sigma_{j=1}^n \tilde{I}^{-1}(\theta,\eta)\, \dot{\ell}(X_j,\theta,\eta) + o_{P_{\theta\eta}}(1). \tag{5.17}$$

One idea to obtain T_n would be to define it as the solution to the system of estimating equations

$$\Sigma_{j=1}^n \dot{\ell}(X_j,T_n,\eta) = 0. \tag{5.18}$$

Indeed by the usual Taylor expansion we should have

$$0 \approx n^{-\frac{1}{2}} \Sigma_{j=1}^n \dot{\ell}(X_j,\theta,\eta) - \hat{I}_n(\theta,\eta)\, \sqrt{n}(T_n-\theta), \tag{5.19}$$

where

$$(\hat{I}_n(\theta,\eta))_{rs} = -n^{-1}\Sigma_{j=1}^n \frac{\partial}{\partial\theta_s} \dot{\ell}_r(X_j,\theta,\eta) \to (\tilde{I}(\theta,\eta))_{rs}. \tag{5.20}$$

The latter would presumably follow from the law of large numbers and

$$\begin{aligned} 0 &= \frac{\partial}{\partial\theta_s} \int \dot{\ell}_r(x,\theta,\eta)\, p(x,\theta,\eta)\, d\mu(x) \\ &= \int \dot{\ell}_r(x,\theta,\eta)\ell_s(x,\theta,\eta)\, p(x,\theta,\eta)d\mu(x) + \int \frac{\partial}{\partial\theta_s} \dot{\ell}_r(x,\theta,\eta)\, p(x,\theta,\eta)d\mu(x). \end{aligned} \tag{5.21}$$

Relations (5.19)-(5.20) would imply (5.17). However as η is unknown, (5.18) cannot serve as estimating equations defining T_n. A way around this problem is to replace $\dot{\ell}(\cdot,\theta,\eta)$ in (5.18) by an estimated version $\hat{\ell}(\cdot,\theta)$ and to solve T_n from

$$\Sigma_{j=1}^n \hat{\ell}(X_j,T_n) = 0. \tag{5.22}$$

Here $\hat{\ell}(\cdot,\theta)$ may, but need not, be based on an estimate for η (for fixed θ).

This route will be followed, though with some modifications. First,

handling (5.22) by way of a Taylor expansion (as in (5.19)) requires quite a number of regularity conditions. Now it is usually possible to obtain an accurate initial estimate $\hat{\theta}_n$ for θ. Using $\hat{\theta}_n$ as the starting point for solving (5.22) by the Newton-Raphson scheme, we obtain as a second estimate

$$(5.23) \qquad T_n = \hat{\theta}_n + n^{-1} \Sigma_{j=1}^{n} \hat{I}_n^{-1}(\hat{\theta}_n) \dot{\ell}(X_j, \hat{\theta}_n).$$

Here, in view of (5.20), $\hat{I}_n(\hat{\theta}_n)$ should estimate $\tilde{I}(\theta,\eta)$. The next step is to forget about the foregoing motivation and *define* T_n by (5.23), choosing a convenient estimator $\hat{I}_n(\hat{\theta}_n)$ for $\tilde{I}(\theta,\eta)$. It turns out that this *one step-method* works well if $\{L_{\theta\eta}(\sqrt{n}(\hat{\theta}_n-\theta))\}$ is tight, a property which is usually called $\sqrt{n}$-*consistency*. Furthermore it works particularly well when combined with another trick, *discretization*. This consists of using an initial estimator $\hat{\theta}_n$ for which $\sqrt{n}(\hat{\theta}_n-\theta)$ has a discrete support, the number of support points within each ball $\{s \in \mathbb{R}^k: \|s\| \leq M\}$ being bounded uniformly in n. Any $\sqrt{n}$-consistent estimator can be discretized without destroying $\sqrt{n}$-consistency, by projecting it on a grid with mesh-width $n^{-\frac{1}{2}}$. There is little motivation for discretization, except that it leads to simple proofs under weak regularity conditions.

The one-step method and discretization are clever devices introduced by Le Cam to handle maximum likelihood estimators in parametric models (cf. Le Cam (1969)) . For semi-parametric models they have to be complemented with a method for estimating $\tilde{\ell}(\cdot,\theta,\eta)$, for given θ. In the special model determined by (5.1)-(5.2) this is usually possible. Suppose that ν_θ in (5.2) is Lebesgue measure and that $g(\cdot,\theta,\eta)$ is smooth. Informally we have that

$$(5.24) \qquad \dot{\ell}(x,\theta,\eta) = \dot{h}/h(x,\theta) + \dot{\psi}(x,\theta)\circ\nabla g/g(\psi(x,\theta),\theta,\eta) + \dot{g}/g(\psi(x,\theta),\theta,\eta).$$

Here $\nabla g = (g^{(1)}(\cdot,\theta,\eta),\ldots,g^{(m)}(\cdot,\theta,\eta))'$ is the vector of partial derivatives with respect to s of $g(s,\theta,\eta)$: $\mathbb{R}^m \to \mathbb{R}$ and $\dot{h}$, $\dot{g}$ and $\dot{\psi}$ are the vectors, respectively a (k×m) matrix, of which the i-th rows contain partial derivatives with respect to θ_i. Hence using (5.16)

$$(5.25) \qquad \tilde{\ell}(x,\theta,\eta) = \tilde{H}(x,\theta) + \tilde{\psi}(x,\theta)\circ\nabla g/g(\psi(x,\theta),\theta,\eta),$$

where

$$\tilde{H}(x,\theta) = \dot{h}/h(x,\theta) - E_\theta(\dot{h}/h(X,\theta)\,|\,\psi(X,\theta) = \psi(x,\theta)) \tag{5.26}$$

$$\tilde{\psi}(x,\theta) = \dot{\psi}(x,\theta) - E_\theta(\dot{\psi}(X,\theta)\,|\,\psi(X,\theta) = \psi(x,\theta)). \tag{5.27}$$

The key to the construction of an estimate $\hat{\ell}(\cdot,\theta)$ for $\ell(\cdot,\theta,\eta)$ is that (5.25) depends on η only through $\nabla g/g$. It therefore suffices to estimate $g(\cdot,\theta,\eta)$. Now, for given θ, $\psi(X_1,\theta),\ldots,\psi(X_n,\theta)$ is an i.i.d. sample from the distribution with density $g(\cdot,\theta,\eta)$. One way to estimate this density is to use the *kernel method*, i.e. we let

$$\hat{g}_n(s,\theta) = n^{-1}\Sigma_{j=1}^n\ \sigma_n^{-m}\ w(\sigma_n^{-1}(s-\psi(X_j,\theta))) \ , \tag{5.28}$$

where the kernel w is a probability density on $\mathbb{R}^m$. If w is a well-behaved kernel and g sufficiently regular, then $\nabla\hat{g}_n/\hat{g}_n(s,\theta)$ should give an accurate estimate of $\nabla g/g(s,\theta,\eta)$. Substituting this in (5.25) we get a candidate for $\hat{\ell}(\cdot,\theta)$.

In Section 5.2 we discuss the first part of the above construction, assuming a suitable estimator for $\nabla g/g$ given. Next a construction of a suitable estimator $\nabla\hat{g}_n/\hat{g}_n$ follows in Section 5.3. Here we have restricted ourselves to kernel estimators, but of course other estimators, perhaps better tuned to the special structure of $g(\cdot,\theta,\eta)$ could have performed the same role.

The resulting estimator sequence $\{T_n\}$ is efficient if (5.15)-(5.16) *do* give the efficient influence function for the model. This matter is discussed in Section 5.4. As we shall not discuss any other estimators and influence functions, the notation (5.16) will be used throughout this chapter, whether it gives the efficient score function or not.

We consider the estimation of other functionals than $\kappa(p(\cdot,\theta,\eta)) = \theta$ in Section 5.6. Concrete examples of the model can be found in Section 5.7. In Section 5.5 we treat the case of mixture models with a sufficient statistic, a class which generates a large number of examples. Mixture models are also called *structural models* as opposed to *functional models.* The latter are non-i.i.d. models and partly motivate the following non-i.i.d. model.

5.1.2. The non-i.i.d. model

Let $(X,\mathcal{B})$, μ, Θ and H be as before. We consider the following model.

for each $n = 1,2,\ldots$

(5.29) $X_{n1},X_{n2},\ldots,X_{nn}$ are independent random elements
$(\theta_n,\eta_{n1},\ldots,\eta_{nn}) \in \Theta\times H^n$
X_{nj} has density $p(\cdot,\theta_n,\eta_{nj})$ w.r.t. μ on $(X,\mathcal{B})$
$p(\cdot,\theta,\eta)$ satisfies (5.1)-(5.2).

We again focus on the estimation of θ and use the same method of constructing an estimator as in Section 5.1.1. Let (5.3) hold and define (letting 0/0 be 0, or arbitrary)

$$\bar{p}_n(x,\theta) = \bar{p}_n(x,\theta,\eta_{n1},..,\eta_{nn}) = n^{-1}\Sigma_{j=1}^n\, p(x,\theta,\eta_{nj}) \tag{5.30}$$

$$\bar{\ell}_n(x,\theta) = \bar{\ell}_n(x,\theta,\eta_{n1},..,\eta_{nn}) = n^{-1}\Sigma_{j=1}^n \ell(x,\theta,\eta_{nj})p(x,\theta,\eta_{nj})/\bar{p}_n(x,\theta) \tag{5.31}$$

$$\tilde{\ell}_n(x,\theta) = \tilde{\ell}_n(x,\theta,\eta_{n1},..,\eta_{nn}) = \bar{\ell}_n(x,\theta) - E_\theta(\bar{\ell}_n(X,\theta)\,|\psi(X,\theta) = \psi(x,\theta)). \tag{5.32}$$

Under the conditions we shall impose later on, as $\|\theta_n-\theta\| = O(n^{-\frac{1}{2}})$

$$\int [\sqrt{n}(\bar{p}_n^{\frac{1}{2}}(x,\theta_n)-\bar{p}_n^{\frac{1}{2}}(x,\theta)) - \tfrac{1}{2}\sqrt{n}(\theta_n-\theta)'\bar{\ell}_n(x,\theta)\bar{p}_n^{-\frac{1}{2}}(x,\theta)]^2\, d\mu(x) \to 0. \tag{5.33}$$

Hence $\bar{\ell}_n(\cdot,\theta)$ are the scores for θ for the *average* density. Consequently under a local completeness condition $\tilde{\ell}_n(\cdot,\theta)$ may be interpreted as the *efficient* score for the *average* density. In consequence of (5.25) we have, informally,

$$\tilde{\ell}_n(x,\theta) = \tilde{H}(x,\theta) + \tilde{\psi}(x,\theta)\circ\bar{Q}_n(\psi(x,\theta),\theta,\eta_{n1},\ldots,\eta_{nn}). \tag{5.34}$$

Here $\tilde{H}(x,\theta)$ and $\tilde{\psi}(x,\theta)$ are as in (5.26)-(5.27) and

$$\bar{Q}_n(s,\theta,\eta_{n1},\ldots,\eta_{nn}) = \Sigma_{j=1}^n\, \nabla g(s,\theta,\eta_{nj}) \;/\; \Sigma_{j=1}^n g(s,\theta,\eta_{nj}). \tag{5.37}$$

When defining an estimator T_n in the same manner as in Section 5.1.1 we see that the most important difference is, that the sufficient statistics $\psi(X_{n1},\theta),\ldots,\psi(X_{nn},\theta)$ are no longer i.i.d. A kernel estimate as in (5.28), which is now based on $\psi(X_{n1},\theta),\psi(X_{n2},\theta),\ldots,\psi(X_{nn},\theta)$, does not estimate a common density $g(\cdot,\theta,\eta)$, but rather the *average* density

$$\bar{g}_n(s,\theta) = \bar{g}_n(s,\theta,\eta_{n1},\ldots,\eta_{nn}) = n^{-1}\Sigma_{j=1}^{n}\, g(s,\theta,\eta_{nj}). \tag{5.38}$$

Thus from (5.34)-(5.37) we conclude that exactly the same construction should now yield an estimator satisfying

$$\sqrt{n}(T_n-\theta) = n^{-\frac{1}{2}}\Sigma_{j=1}^{n}\, \tilde{I}_n^{-1}(\theta)\dot{\ell}_n(X_{nj},\theta,\eta_{n1},\ldots,\eta_{nn}) + o_{P_{\theta\eta_{n1}\ldots\eta_{nn}}}(1), \tag{5.39}$$

where

$$\tilde{I}_n(\theta) = \tilde{I}_n(\theta,\eta_{n1},\ldots,\eta_{nn}) = \int \dot{\ell}_n(x,\theta)\dot{\ell}_n(x,\theta)'\, \bar{p}_n(x,\theta)\, d\mu(x). \tag{5.40}$$

In the next section (5.39) will be proved rigorously under some conditions.

Thus the construction of an estimator for θ in the non-i.i.d. model is carried through under the working hypothesis that the model is i.i.d.. This may seem peculiar. Notably, though, due to the 'adaptation' as above, this construction improves upon other constructions in the literature in several examples. $\{T_n\}$ satisfying (5.39) can also be shown to possess certain optimality properties.. We give a heuristic discussion of this in Section 5.4.2.

5.1.3. Some regularity conditions.

It is clear that in the non-i.i.d. mdoel we need conditions to ensure that the averages become reasonably stable if $n \to \infty$. As a first set of assumptions we require that, for $\|\theta_n-\theta\| = O(n^{-\frac{1}{2}})$ and every $\varepsilon > 0$,

$$\Sigma_{j=1}^{n}\int[p^{\frac{1}{2}}(x,\theta_n,\eta_{nj})-p^{\frac{1}{2}}(x,\theta,\eta_{nj})-\tfrac{1}{2}(\theta_n-\theta)'\dot{\ell}(x,\theta,\eta_{nj})p^{\frac{1}{2}}(x,\theta,\eta_{nj})]^2 d\mu(x)\to 0 \tag{5.41}$$

$$n^{-1}\Sigma_{j=1}^{n}\int \|\dot{\ell}(x,\theta,\eta_{nj})\|^2\, p(x,\theta,\eta_{nj})\, d\mu(x) = O(1) \tag{5.42}$$

$$n^{-1}\Sigma_{j=1}^{n}\int \|\dot{\ell}(x,\theta,\eta_{nj})\|^2\, 1_{\{\|\dot{\ell}(x,\theta,\eta_{nj})\|\geq\varepsilon\sqrt{n}\}}\, p(x,\theta,\eta_{nj})\, d\mu(x) \to 0 \tag{5.43}$$

$$\int \|\dot{\ell}_n(x,\theta_n)\|^2\, 1_{\{\|\dot{\ell}_n(x,\theta_n)\|\geq\varepsilon\sqrt{n}\}}\, \bar{p}_n(x,\theta_n)\, d\mu(x) \to 0 \tag{5.44}$$

$$\int \|\dot{\ell}_n(x,\theta_n)\bar{p}_n^{-\frac{1}{2}}(x,\theta_n) - \dot{\ell}_n(x,\theta)\bar{p}_n^{-\frac{1}{2}}(x,\theta)\|^2\, d\mu(x) \to 0 \tag{5.45}$$

(5.46) the limit points of $\{\tilde{I}_n(\theta)\}$ are non-singular.

(It follows from (5.42) that limsup $\|\tilde{I}_n(\theta)\| < \infty$).

For the i.i.d. model described in Section 5.1.1 we impose these conditions with $\eta_{nj} = \eta$, fixed. In this case (5.41) reduces to a slightly stronger statement than (5.3), (5.42)-(5.44) may be omitted, and (5.45) and (5.46) reduce to

$$\int \|\tilde{\ell}(x,\theta_n,\eta)p^{\frac{1}{2}}(x,\theta_n,\eta) - \tilde{\ell}(x,\theta,\eta)p^{\frac{1}{2}}(x,\theta,\eta)\|^2 \, d\mu(x) \to 0$$

and

$$\tilde{I}(\theta,\eta) \text{ is non-singular,}$$

respectively.

5.2. AN ESTIMATOR FOR θ

For the model given by (5.29) it is natural in view of (5.25)-(5.27), to assume the existence of measurable functions $\tilde{H}(\cdot,\theta)$: $(X,\mathcal{B}) \to \mathbb{R}^k$, $\tilde{\psi}(\cdot,\theta)$: $(X,\mathcal{B}) \to \mathbb{R}^{k\times m}$ (in the form of a (k×m) matrix) and $Q(\cdot,\theta,\eta)$: $\mathbb{R}^m \to \mathbb{R}^m$ such that

$$\tilde{\ell}(x,\theta,\eta) = \tilde{H}(x,\theta) + \tilde{\psi}(x,\theta)\circ Q(\psi(x,\theta),\theta,\eta) \tag{5.47}$$

$$E_\theta(\tilde{\psi}(X,\theta)\,|\psi(X,\theta)) = 0. \tag{5.48}$$

We then have that (5.34) (cf. (5.32)) holds with

$$\bar{Q}_n(s,\theta,\eta_{n1},..,\eta_{nn}) = n^{-1}\Sigma_{j=1}^n Q(s,\theta,\eta_{nj})g(s,\theta,\eta_{nj}) \;/\; \Sigma_{j=1}^n g(s,\theta,\eta_{nj}).$$

The one step method requires an initial estimator $\hat{\theta}_n = \hat{\theta}_n(X_{n1},X_{n2},\ldots,X_{nn})$ which is *$\sqrt{n}$-consistent*, i.e.

$$\{L_{\theta\eta_{n1}..\eta_{nn}}(\sqrt{n}(\hat{\theta}_n-\theta))\} \text{ is tight on } \mathbb{R}^k. \tag{5.49}$$

Finally we need a suitable estimator for $\bar{Q}_n(\cdot,\theta,\eta_{n1},\ldots,\eta_{nn})$. For every fixed θ we assume existence of measurable functions $Q_n(s,\theta,v_1,\ldots,v_{n-1})$: $\mathbb{R}^m\times(\mathbb{R}^m)^{(n-1)} \to \mathbb{R}^m$ such that for independent random

vectors $V_{n1},\dots,V_{nn}$ in $\mathbb{R}^m$, where V_{nj} has density $g(\cdot,\theta_n,\eta_{nj})$ with respect to ν_{θ_n}, and $\|\theta_n-\theta\| = O(n^{-\frac{1}{2}})$

$$(5.50)\quad \int \max_{j=1,n} E\,\|\hat{D}_n^j(s,\theta_n)\|^2\, E_{\theta_n}(\|\tilde{\psi}(X,\theta_n)\|^2 \mid \psi(X,\theta_n)=s)\,\bar{g}_n(s,\theta_n)\,d\nu_{\theta_n}(s) \to 0,$$

where

$$(5.51)\quad \hat{D}_n^j(s,\theta) = Q_n(s,\theta,V_{n1},..,V_{nj-1},V_{nj+1},..,V_{nn}) - \bar{Q}_n(s,\theta,\eta_{n1},\dots,\eta_{nn}).$$

Here for a (k×m) matrix $A = (\alpha_{ij})$, $\|A\|$ may be any norm, e.g. $(\Sigma\Sigma\,\alpha_{ij}^2)^{\frac{1}{2}}$.

Condition (5.50) will be treated in detail in the next section. As for $\sqrt{n}$-consistent estimators, it is usually not too difficult to find candidates in specific models. A general method that may work is the following. By sufficiency of $\psi(X,\theta)$, for any $\theta \in \Theta$ and $(\eta_{n1},\dots,\eta_{nn},\eta'_{n1},\dots,\eta'_{nn}) \in H^n \times H^n$

$$E_{\theta\eta_{n1}..\eta_{nn}} \Sigma_{j=1}^n \ell(X_{nj},\theta,\eta'_{nj}) = 0.$$

Therefore for fixed, conveniently chosen $(\eta'_{n1},\dots,\eta'_{nn}) \in H^n$, one may try defining an estimator $\hat{\theta}_n$ as the solution to

$$\Sigma_{j=1}^n \ell(X_{nj},\theta,\eta'_{nj}) = 0.$$

In the same spirit it may work to solve $\hat{\theta}_n$ from

$$\Sigma_{j=1}^n \tilde{H}(X_{nj},\hat{\theta}_n) = 0.$$

The main result of this section is

THEOREM 5.3. *Let* (5.1)-(5.3), (5.29) *and* (5.41)-(5.50) *hold. Then there exists an estimator sequence* $\{T_n\}$, $T_n = T_n(X_{n1},X_{n2},\dots,X_{nn})$, *satisfying* (5.39). □

A candidate for T_n can be constructed as follows. Let $V_{nj}(\theta) = \psi(X_{nj},\theta)$ and

(5.52) $$\hat{Q}_n^j(s,\theta) = Q_n(s,\theta,V_{n1}(\theta),..,V_{nj-1}(\theta),V_{nj+1}(\theta),..,V_{nn}(\theta))$$

(5.53) $$\hat{\ell}_n^j(x,\theta) = \tilde{H}(x,\theta) + \tilde{\psi}(x,\theta)\circ\hat{Q}_n^j(\psi(x,\theta),\theta) \ .$$

Define a (k×k) matrix by

(5.54) $$\hat{I}_n(\theta) = n^{-\frac{1}{2}} \Sigma_{j=1}^n \ (\hat{\ell}_n^j(X_{nj},\theta-n^{-\frac{1}{2}}e_1)-\hat{\ell}_n^j(X_{nj},\theta),\ldots,$$
$$\hat{\ell}_n^j(X_{nj},\theta-n^{-\frac{1}{2}}e_k)-\hat{\ell}_n^j(X_{nj},\theta)) \ ,$$

where e_i is the i-th unit vector in $\mathbb{R}^k$. Let $\hat{\theta}_n$ be a discretized $\sqrt{n}$-consistent estimator for θ and set

(5.55) $$T_n = \hat{\theta}_n + n^{-1} \Sigma_{j=1}^n \hat{I}_n^{-1}(\hat{\theta}_n) \hat{\ell}_n^j(X_{nj},\hat{\theta}_n),$$

whenever $\hat{I}_n(\hat{\theta}_n)$ is nonsingular and 0 otherwise.

The proof that $\{T_n\}$ satisfies (5.39) is accomplished through a series of lemmas. By (5.3), (5.41)-(5.43), Proposition A.8 and Corollary A.5 we have contiguity of the laws of $(X_{n1},\ldots,X_{nn})$ under $(\theta,\eta_{n1},\ldots,\eta_{nn})$ and $(\theta_n,\eta_{n1},\ldots,\eta_{nn})$ if $\|\theta_n-\theta\| = O(n^{-\frac{1}{2}})$. This means that convergence to zero in $P_{\theta\eta_{n1}..\eta_{nn}}$-probability is equivalent to convergence to zero in $P_{\theta_n\eta_{n1}..\eta_{nn}}$-probability, a fact that we use throughout the proofs.

LEMMA 5.4. *Under the conditions of Theorem 5.3, for* $\|\theta_n-\theta\| = O(n^{-\frac{1}{2}})$ *in* $P_{\theta_n\eta_{n1}..\eta_{nn}}$*-probability*

$$n^{-\frac{1}{2}} \Sigma_{j=1}^n \ (\mathcal{L}_n(X_{nj},\theta_n,\eta_{n1},\ldots,\eta_{nn}) - \mathcal{L}_n(X_{nj},\theta,\eta_{n1},\ldots,\eta_{nn}))$$
$$+ \tilde{I}_n(\theta,\eta_{n1},\ldots,\eta_{nn})\sqrt{n}(\theta_n-\theta) \to 0. \ \square$$

PROOF. See Section 5.8.3. ■

LEMMA 5.5. *Under the conditions of Theorem 5.3, for* $\|\theta_n-\theta\| = O(n^{-\frac{1}{2}})$

$$E_{\theta_n\eta_{n1}..\eta_{nn}} \|n^{-\frac{1}{2}} \Sigma_{j=1}^n \ (\hat{\ell}_n^j(X_{nj},\theta_n)-\mathcal{L}_n(X_{nj},\theta_n,\eta_{n1},\ldots,\eta_{nn}))\|^2 \to 0. \ \square$$

PROOF. Write V_{nj} and $\tilde{V}_{nj}$ for $\psi(X_{nj},\theta_n)$ and $\tilde{\psi}(X_{nj},\theta_n)$ respectively and E_n

for $E_{\theta_n \eta_{n1}..\eta_{nn}}$. By (5.34) and (5.53) we must show convergence to zero of (cf. (5.51)-(5.52))

$$E_n \| n^{-\frac{1}{2}} \Sigma_{j=1}^n \tilde{V}_{nj} \circ \hat{D}_n^j(V_{nj},\theta_n) \|^2$$

$$= n^{-1} \Sigma_{i=1}^n \Sigma_{j=1}^n E_n \hat{D}_n^i(V_{ni},\theta_n)' \tilde{V}_{ni}' \circ \tilde{V}_{nj} \hat{D}_n^j(V_{nj},\theta_n).$$

Taking first the conditional expectation with respect to $V_{n1},\ldots,V_{nn}$ and remembering that $\hat{D}_n^j(V_{nj},\theta_n)$ depends on $V_{n1},\ldots,V_{nn}$ only, we see that this equals

$$n^{-1}\Sigma_{i=1}^n \Sigma_{j=1}^n E_n \hat{D}_n^i(V_{ni},\theta_n)' E_{\theta_n}(\tilde{V}_{ni}' \circ \tilde{V}_{nj} | V_{ni},V_{nj}) \hat{D}_n^j(V_{nj},\theta_n). \tag{5.56}$$

By (5.48) for $i \neq j$

$$E_{\theta_n}(\tilde{V}_{ni}' \circ \tilde{V}_{nj} | V_{ni},V_{nj}) = E_{\theta_n}(\tilde{V}_{ni} | V_{ni})' \circ E_{\theta_n}(\tilde{V}_{nj} | V_{nj}) = 0.$$

Next the sum over the diagonal terms in (5.56) is smaller than

$$n^{-1}\Sigma_{j=1}^n E_n \|\hat{D}_n^j(V_{nj},\theta_n)\|^2 \|E_{\theta_n}(\tilde{V}_{nj}' \circ \tilde{V}_{nj} | V_{nj})\|$$

$$\leq n^{-1}\Sigma_{j=1}^n E_n \|\hat{D}_n^j(V_{nj},\theta_n)\|^2 E_{\theta_n}(\|\tilde{V}_{nj}' \circ \tilde{V}_{nj}\| | V_{nj})$$

$$\leq n^{-1}\Sigma_{j=1}^n E_n \int \|\hat{D}_n^j(s,\theta_n)\|^2 E_{\theta_n}(\|\tilde{V}_{nj}\|^2 | V_{nj} = s)\, g(s,\theta_n,\eta_{nj})\, d\nu_{\theta_n}(s)$$

which converges to zero by (5.50). ■

LEMMA 5.6. *Under the conditions of Theorem 5.3, for* $\|\theta_n - \theta\| = O(n^{-\frac{1}{2}})$ *we have* $\tilde{I}_n(\theta_n) - \tilde{I}_n(\theta) \to 0$ *and for all* $\varepsilon > 0$

$$P_{\theta_n \eta_{n1}..\eta_{nn}}(\hat{I}_n^{-1}(\theta_n) \text{ exists}, \|\hat{I}_n^{-1}(\theta_n) - \tilde{I}_n^{-1}(\theta,\eta_{n1},\ldots,\eta_{nn})\| < \varepsilon) \to 1. \quad \square$$

PROOF. By (5.42) limsup $\|\tilde{I}_n(\theta)\| < \infty$. Combination with (5.45) yields the first assertion. Next by (5.54) and Lemmas 5.5 and 5.4, the i-th column of

$\hat{I}_n(\theta_n)$ (i=1,2,...,k) equals

$$n^{-\frac{1}{2}} \Sigma_{j=1}^{n} (\dot{\ell}_n(X_{nj},\theta_n - n^{-\frac{1}{2}}e_i) - \dot{\ell}_n(X_{nj},\theta_n)) + o_{P_{\theta_n\eta_{n1}..\eta_{nn}}}(1) \tag{1}$$

$$= \tilde{I}_n(\theta)e_i + o_{P_{\theta_n\eta_{n1}..\eta_{nn}}}(1).$$

Hence $\hat{I}_n(\theta_n)-\tilde{I}_n(\theta) \to 0$ in $P_{\theta_n\eta_{n1}..\eta_{nn}}$-probability. Combination with (5.46) shows that for any subsequence $\{n_j\}$ there exists a further subsequence $\{n_{jk}\}$ and a nonsingular limit point $\tilde{I}$ of $\{\tilde{I}_n(\theta_n)\}$ such that $\hat{I}_n(\theta_n) \to \tilde{I}$ in $P_{\theta_n\eta_{n1}..\eta_{nn}}$-probability, where we write n for n_{jk}. Hence for all $\delta,\varepsilon > 0$ there exists $K_{\delta\varepsilon} \in \mathbb{N}$ such that for all $k > K_{\delta\varepsilon}$

$$P_{\theta_n\eta_{n1}..\eta_{nn}}(\|\hat{I}_n(\theta_n)-\tilde{I}\| < \delta) \geq 1-\varepsilon.$$

For $\delta > 0$ sufficiently small this implies for $k > K_{\delta\varepsilon}$

$$P_{\theta_n\eta_{n1}..\eta_{nn}}(\hat{I}_n^{-1}(\theta_n) \text{ exists}, \ \|\hat{I}_n^{-1}(\theta_n) - \tilde{I}^{-1}\| < \varepsilon) \geq 1-\varepsilon. \ \blacksquare$$

PROOF OF THEOREM 5.3. For T_n given by (5.55) we have

$$P_{\theta\eta_{n1}..\eta_{nn}}(\| \sqrt{n}(T_n-\theta) - n^{-\frac{1}{2}}\Sigma_{j=1}^{n}\tilde{I}_n^{-1}(\theta)\dot{\ell}_n(X_{nj},\theta) \| \geq \varepsilon)$$

(5.57)

$$\leq P_{\theta\eta_{n1}..\eta_{nn}}(\|\sqrt{n}(\hat{\theta}_n-\theta)\|\geq M) + \Sigma\, P_{\theta_n\eta_{n1}..\eta_{nn}}(\|\hat{L}_n(\theta_n)\|\geq\varepsilon,\ \hat{\theta}_n=\theta_n) + o(1),$$

where the sum is over the set of $\theta_n \in \mathbb{R}^k$ in the support of $\hat{\theta}_n$ with $\sqrt{n}\|\theta_n-\theta\| \leq M$,

$$\hat{L}_n(\theta_n) = [\sqrt{n}(\theta_n-\theta) + n^{-\frac{1}{2}}\Sigma_{j=1}^{n}(\hat{I}_n^{-1}(\theta_n)\hat{\ell}_n^{j}(X_{nj},\theta_n) - \tilde{I}_n^{-1}(\theta)\dot{\ell}_n(X_{nj},\theta))]1_{A_n(\theta_n)}$$

and $A_n(\theta) = \{\hat{I}_n^{-1}(\theta) \text{ exists}\}$. By $\sqrt{n}$-consistency of $\hat{\theta}_n$, M can be chosen such that the first term in (5.57) is arbitrarily small. Then, as $\hat{\theta}_n$ is discretized, the number of terms in the sum is finite and bounded uniformly in n, and it suffices to prove that the maximum over the terms converges to

zero. This would follow if for any sequence of vectors $\{\theta_n\}$ in $\mathbb{R}^k$ with $\|\theta_n-\theta\| = O(n^{-\frac{1}{2}})$

$$\hat{L}_n(\theta_n) \to 0,$$

in $P_{\theta_n \eta_{n1}..\eta_{nn}}$-probability. By Lemma 5.4, (5.46) and contiguity this can be reduced to proving

$$(5.58) \qquad [\, n^{-\frac{1}{2}}\Sigma_{j=1}^n (\hat{I}_n^{-1}(\theta_n)\hat{\ell}_n^j(X_{nj},\theta_n) - \tilde{I}_n^{-1}(\theta)\tilde{\ell}_n(X_{nj},\theta_n)) \,] \, 1_{A_n(\theta_n)} \to 0$$

in $P_{\theta_n \eta_{n1}..\eta_{nn}}$-probability. Relation (5.58) is a consequence of Lemmas 5.5 and 5.6, (5.46) and tightness of $\{L_{\theta_n \eta_{n1}..\eta_{nn}}(n^{-\frac{1}{2}}\Sigma_{j=1}^n \tilde{\ell}_n(X_{nj},\theta_n))\}$. ■

5.3. ESTIMATION OF LOCATION SCORES

In this section it is shown that estimators for $\nabla\bar{g}_n/\bar{g}_n$ needed for the construction in Section 5.2, typically exist. More precisely, we present a set of sufficient conditions for (5.50) where, motivated by (5.37), $\bar{Q}_n$ is replaced by $\nabla\bar{g}_n/\bar{g}_n$ and where it is assumed that ν_θ is Lebesgue measure on $\mathbb{R}^m$. The estimator for $\nabla\bar{g}_n/\bar{g}_n$ is obtained from a kernel estimate for $\bar{g}_n$. Though we do not discuss this here, we note that in special examples it may be fruitful to use an ad hoc method, which may yield an estimator which is better adapted to the structure of the model. In that sense, the result of this section should be taken as an existence result, rather than a recipe for practical implementation.

We state the main theorem in an abstract notation. This will be specified to the model in Corollary 5.9.

For each $n = 1,2,\ldots$, $V_{n1},\ldots,V_{nn}$ are independent random vectors, V_{nj} having a density g_{nj} with respect to Lebesgue measure λ on $\mathbb{R}^m$. We assume that the densities vanish outside a convex open set $S \subset \mathbb{R}^m$, while for any fixed n and j there exists a vector $\nabla g_{nj} = (g_{nj}^{(1)},\ldots,g_{nj}^{(m)})'$ of functions from $\mathbb{R}^m$ to $\mathbb{R}$, vanishing outside S and such that for all $c,d \in S$

$$(5.59) \qquad g_{nj}(d)-g_{nj}(c) = \int_0^1 \nabla g_{nj}(c+(d-c)u)'\circ(d-c)\, du.$$

Define

$$\bar{g}_n(s) = n^{-1} \Sigma_{j=1}^{n} g_{nj}(s)$$

$$\nabla\bar{g}_n(s) = n^{-1} \Sigma_{j=1}^{n} \nabla g_{nj}(s).$$

Let β_n: $\mathbb{R}^m \to [0,\infty) \subset \mathbb{R}$ be measurable functions satisfying

$$\text{(5.60)} \qquad \limsup_{h\to 0} \sup_{n} k(\|h\|)^{-1} |\beta_n(s+h)-\beta_n(s)| < \infty \qquad \text{a.e. } [\lambda],$$

for a non-decreasing function k: $[0,\infty) \to [0,\infty)$ with $k(h)\downarrow 0$ if $h\downarrow 0$. A sufficient condition for (5.60) is for instance

$$|\beta_n(s+h)-\beta_n(s)| \le M_s \|h\|^{\kappa} \qquad \text{a.e. } [\lambda],\ n = 1,2,\ldots, h \to 0,$$

for some constant $\kappa > 0$ and constants M_s. Finally assume for any $h_n \to 0$ and $b_n \to \infty$

$$\text{(5.61)} \qquad \int \| \nabla\bar{g}_n/\bar{g}_n^{\frac{1}{2}}\beta_n(s+h_n) - \nabla\bar{g}_n/\bar{g}_n^{\frac{1}{2}}\beta_n(s) \|^2 \, d\lambda(s) \to 0$$

$$\text{(5.62)} \qquad \{ \|\nabla\bar{g}_n/\bar{g}_n^{\frac{1}{2}}\beta_n(s)\|^2 : n=1,2..\} \text{ is equi-}\lambda\text{-integrable (cf. Bauer(1981))}$$

$$\text{(5.63)} \qquad \int \| \nabla\bar{g}_n/\bar{g}_n^{\frac{1}{2}}\beta_n(s) \|^2 \, 1\{\|\nabla\bar{g}_n/\bar{g}_n\beta_n(s)\| \ge b_n\} \, d\lambda(s) \to 0$$

$$\text{(5.64)} \qquad \int \| \nabla\bar{g}_n/\bar{g}_n^{\frac{1}{2}}\beta_n(s) \|^2 \, 1\{\beta_n(s) \notin [h_n,b_n]\} \, d\lambda(s) \to 0.$$

Let ω: $\mathbb{R}^m \to \mathbb{R}$ be any twice continuously differentiable probability density with respect to Lebesgue measure, with support contained in the unit ball $\{s \in \mathbb{R}^m: \|s\| \le 1\}$. Given $\sigma \in (0,\infty) \subset \mathbb{R}$ define

$$\hat{g}^j_{n\sigma}(s) = n^{-1}\Sigma^{n}_{\substack{i=1\\ i\neq j}} \sigma^{-m} \omega(\sigma^{-1}(s-V_{ni})).$$

Let ∂S be the boundary of the set S and set

$$\|s-\partial S\| = \inf \{\|s-y\|: y \in \partial S\}.$$

THEOREM 5.7. *Let* (5.59)-(5.64) *hold. Define*

$$\hat{Q}_n^j(s) = \nabla \hat{g}^j_{n\sigma_n}(s) \,/\, (\hat{g}^j_{n\sigma_n}(s)+\delta_n)\; 1_{C_n}(s),$$

where for sequences $\alpha_n, \gamma_n, \delta_n, \varepsilon_n, \sigma_n \downarrow 0$ *and* $b_n, c_n \to \infty$ *in* $\mathbb{R}$

$$C_n = \{s \in S:\ \alpha_n < \beta_n(s) < b_n,\ \sup\{|\beta_n(s+\sigma_n y)-\beta_n(s)| : \|y\|<1\} \le \gamma_n,$$

$$\|s-\partial S\| > \varepsilon_n,\ \|\nabla \hat{g}^j_{n\sigma_n}\beta_n(s)\| < c_n(\hat{g}^j_{n\sigma_n}(s)+\delta_n)\ \}.$$

If $\delta_n^{-2}\sigma_n^{-2m-2}b_n^2 n^{-1} \to 0$, $\gamma_n^{-1}k(\sigma_n) \to 0$, $\gamma_n\alpha_n^{-1} \to 0$, $\sigma_n\varepsilon_n^{-1} \to 0$ *and* $c_n^2\sigma_n\alpha_n^{-1} \to 0$, *then*

$$\int \max_{j=1..n} E\,\|\hat{Q}_n^j(s) - \nabla\bar{g}_n/\bar{g}_n(s)\|^2\, \beta_n^2(s)\,\bar{g}_n(s)\,d\lambda(s) \to 0. \quad \square$$

The following lemma is helpful, especially to simplify the conditions of Theorem 5.7 for i.i.d. models.

LEMMA 5.8. *Suppose that for a measurable function* $\beta: \mathbb{R}^m \to [0,\infty)$ *and a probability density* g *with respect to* λ *satisfying* (5.59)

(5.65) $\int \|\nabla g/g^{\frac{1}{2}}\beta(s)\|^2\, d\lambda(s) < \infty$

(5.66) $\int \|\nabla\bar{g}_n/\bar{g}_n^{\frac{1}{2}}\beta_n(s) - \nabla g/g^{\frac{1}{2}}\beta(s)\|^2\, d\lambda(s) \to 0$

(5.67) *the measures on* $\mathbb{R}^m$ *with densities* $\bar{g}_n$ *and* g *respectively, are contiguous*

(5.68) $\beta_n(\cdot) - \beta(\cdot) \to 0$ *in* λ*-measure.*

Then (5.61)-(5.64) *hold.* $\square$

From these general theorems we infer for the model given by (5.29):

COROLLARY 5.9. *For the model* (5.29) *suppose that* (5.59)-(5.64) *hold with* $g_{nj}(\cdot) = g(\cdot,\theta_n,\eta_{nj})$ *and* $\beta_n^2(\cdot) = E_{\theta_n}(\|\tilde{\psi}(X,\theta_n)\|^2 \mid \psi(X,\theta_n) =\cdot)$. *Then* (5.50) *holds with the replacement*

$$\bar{Q}_n(s,\theta,\eta_{n1},\ldots,\eta_{nn}) = \Sigma_{j=1}^n \nabla g(s,\theta,\eta_{nj}) \,/\, \Sigma_{j=1}^n g(s,\theta,\eta_{nj})$$

(cf. (5.37)*). In particular, for the i.i.d. model described in Section 5.1.1 (i.e.* $\eta_{nj} = \eta$ *for all* n *and* j*),* (5.50) *holds if* $\beta_n^2(\cdot)$ *satisfies* (5.60), $g_{nj}(\cdot) = g(\cdot,\theta,\eta_{nj})$ *satisfies* (5.59) *and*

(5.69) $\int \| \nabla g/g^{\frac{1}{2}}(s,\theta,\eta)\beta(s,\theta) \|^2 \, d\lambda(s) < \infty$

(5.70) $\int \| \nabla g/g^{\frac{1}{2}}(s,\theta_n,\eta)\beta(s,\theta_n) - \nabla g/g^{\frac{1}{2}}(s,\theta,\eta)\beta(s,\theta) \|^2 \, d\lambda(s) \to 0$

(5.71) *the measures on* $\mathbb{R}^m$ *with densities* $g(\cdot,\theta_n,\eta)$ *and* $g(s,\theta,\eta)$ *respectively, are contiguous*

(5.72) $\beta(\cdot,\theta_n) - \beta(\cdot,\theta) \to 0$ *in* λ*-measure.* □

The proofs of Theorem 5.7 and Lemma 5.8 are rather involved and can be found in Section 5.8.3.

5.4. EFFICIENCY

5.4.1. I.i.d. models

Let $X_1,\ldots,X_n$ be independent, identically distributed random elements with a density $p(\cdot,\theta,\eta)$ satisfying (5.1)-(5.2) and let a *convex* tangent cone $T(p(\cdot,\theta,\eta))$ be given by (5.12)-(5.13). Suppose $\{T_n\}$ satisfies the specialization of (5.39) to the i.i.d. case, i.e.

(5.73) $\sqrt{n}(T_n-\theta) = n^{-\frac{1}{2}}\Sigma_{j=1}^n \tilde{I}^{-1}(\theta,\eta)\, \tilde{\ell}(X_j,\theta,\eta) + o_{P_{\theta\eta}}(1).$

By Corollary A.7 and Lemma 2.5 we conclude

$$L_{\theta_n\eta_n}(\sqrt{n}(T_n-\theta_n)) \to N(0,\tilde{I}^{-1}(\theta,\eta)),$$

for any $\{(\theta_n,\eta_n)\} \subset \Theta\times H$ with $\sqrt{n}(\theta_n-\theta) \to h$ and (cf. (5.13))

$$\int [\sqrt{n}(p^{\frac{1}{2}}(x,\theta_n,\eta_n)-p^{\frac{1}{2}}(x,\theta,\eta)) - \tfrac{1}{2}(h'\dot{\ell}(x,\theta,\eta)+\underline{b}(\psi(x,\theta)))\, p^{\frac{1}{2}}(x,\theta,\eta)]^2\, d\mu(x) \to 0.$$

It follows that $\{T_n\}$ is efficient for θ if the components $\dot{\ell}_i(\cdot,\theta,\eta)$ of $\dot{\ell}(\cdot,\theta,\eta)$ are contained in the closure of $\operatorname{lin} T(p(\cdot,\theta,\eta))$ (cf. (5.15)-(5.16)). An equivalent condition is

$$(5.74)\quad E_\theta(\dot{\ell}_i(X,\theta,\eta)\,|\,\psi(X,\theta)=\cdot) \in \overline{\operatorname{lin} \underline{T}_\eta(g(\cdot,\theta,\eta))} \qquad (i=1,\ldots,k).$$

Local completeness (cf. Definition 5.1) is sufficient for (5.74) but not necessary. Since convexity of $T(p(\cdot,\theta,\eta))$ makes part of the assumption of local completeness, an estimator satisfying (5.73) will be LAM in the sense of Theorem 2.8 in the case of local completeness (case (i) or (ii) of Section 2.6). Furthermore, if the sufficient statistic is strongly locally complete, then we have that the linear span of $\dot{\ell}(\cdot,\theta,\eta)$ is contained in the tangent cone itself. The latter property implies a stronger form of efficiency of an estimator satisfying (5.73), because any LAM estimator necessarily satisfies (5.73) in this case (case (i) of Section 2.6).

We now turn to a special class of models.

5.4.1.1. Models that are convex in η

Assume that the model given by (5.1)-(5.2) is *convex in* η in the following sense. For every pair $(\eta,\eta') \in H\times H$ and $t \in [0,1] \subset \mathbb{R}$, there exists $\eta_t \in H$ with

$$(5.75)\quad p(\cdot,\theta,\eta_t) = t\, p(\cdot,\theta,\eta') + (1-t)\, p(\cdot,\theta,\eta).$$

Under this convexity condition, local completeness (but not strong local completeness), can be deduced from *ordinary* completeness. For given $(\theta,\eta) \in \Theta\times H$ define a set $H_{\theta\eta} \subset H$ by

$$H_{\theta\eta} = \{\eta' \in H:\ p(\cdot,\theta,\eta') << p(\cdot,\theta,\eta) \text{ and } \int p^2(x,\theta,\eta')/p(x,\theta,\eta)\, d\mu(x) < \infty\ \}.$$

THEOREM 5.10. *Let* (5.1)-(5.3) *hold for all* $(\theta,\eta) \in \Theta\times H$, *let the model be convex in* η *and suppose that* $\{g(\cdot,\theta,\eta'):\ \eta' \in H_{\theta\eta}\}$ *is complete. Then* $\psi(X,\theta)$ *is locally complete at* (θ,η). □

PROOF. Let

$$\underline{b}_{\eta'}(\cdot) = g^{-1}(\cdot,\theta,\eta)\,(g(\cdot,\theta,\eta')-g(\cdot,\theta,\eta)).$$

The idea is that

$$\underline{b}_{\eta'}(\psi(\cdot,\theta)) = \partial/\partial t \log p(\cdot,\theta,\eta_t)\ |_{t=0}$$

and hence is a score for η. However to show joint differentiability in (θ,η) as in (5.13) we have to do some work. Trivially (cf. (5.75))

$$\int [t^{-1}p^{-1}(x,\theta,\eta)(p(x,\theta,\eta_t)-p(x,\theta,\eta))-\underline{b}_{\eta'}(\psi(x,\theta))]^2 p(x,\theta,\eta)d\mu(x)=0.$$

By Proposition A.9

$$(5.76)\qquad \int [\ t^{-1}(p^{\frac12}(x,\theta,\eta_t)-p^{\frac12}(x,\theta,\eta)) - \tfrac12\underline{b}_{\eta'}(\psi(x,\theta))p^{\frac12}(x,\theta,\eta)\ ]^2 d\mu(x) \to 0.$$

Next we can infer from Lemma 5.18 (taking $Z_{nj} = \{\eta,\eta'\}$), that for any $h \in \mathbb{R}^k$

$$\int [t^{-1}(p^{\frac12}(x,\theta+th,\eta_t)-p^{\frac12}(x,\theta,\eta_t)) - \tfrac12 p^{-\frac12}(x,\theta,\eta_t)$$

$$\cdot \{(1-t)h'\dot{\ell}(x,\theta,\eta)p(x,\theta,\eta) + th'\dot{\ell}(x,\theta,\eta')p(x,\theta,\eta')\}\ ]^2\, d\mu(x) \to 0.$$

But then

$$(5.77)\qquad \int [t^{-1}(p^{\frac12}(x,\theta+th,\eta_t)-p^{\frac12}(x,\theta,\eta_t)) - \tfrac12 h'\dot{\ell}(x,\theta,\eta)p^{\frac12}(x,\theta,\eta)]^2 d\mu(x) \to 0.$$

Combination of (5.76)-(5.77) yields

$$\int [t^{-1}(p^{\frac12}(x,\theta+th,\eta_t)-p^{\frac12}(x,\theta,\eta))$$

$$- \tfrac{1}{2}(h'\dot{\ell}(x,\theta,\eta)+\underline{b}_{\eta'}(\psi(x,\theta)))\, p^{\frac{1}{2}}(x,\theta,\eta)]^2 d\mu(x) \to 0.$$

Let $\underline{T}_\eta(g(\cdot,\theta,\eta)) = \{\alpha \underline{b}_{\eta'}: \alpha \geq 0,\ \eta' \in H_{\theta\eta}\}$. Then $\underline{T}_\eta(g(\cdot,\theta,\eta))$ is a convex cone which satisfies (5.13). We conclude the proof by showing that $\operatorname{lin} \underline{T}_\eta(g(\cdot,\theta,\eta))$ is dense in $L_{2*}(g(\cdot,\theta,\eta))$. Indeed if $\underline{b} \in L_{2*}(g(\cdot,\theta,\eta))$ and $\underline{b} \perp \underline{T}_\eta(g(\cdot,\theta,\eta))$, then for all $\eta' \in H_{\theta\eta}$

$$\int \underline{b}(s)\underline{b}_{\eta'}(s)\, g(s,\theta,\eta)\, d\nu_\theta(s) = \int \underline{b}(s)\, g(s,\theta,\eta')\, d\nu_\theta(s) = 0.$$

Completeness implies that $\underline{b} = 0$. ∎

5.4.2. Non-i.i.d. models

Whereas for i.i.d. models there is a theory of asymptotic lower bounds which can serve as a basis for defining a reasonable efficiency concept, for non-i.i.d. models of type (5.29) the situation is much more complicated. Here the problem is not so much that the model is non-i.i.d, but rather the inclusion of infinitely many nuisance parameters. As yet there is no satisfactory concept of efficiency in such models. The content of this section is purely heuristic.

It is certainly possible to obtain lower bounds for models with infinitely many nuisance parameters in the same spirit as for i.i.d. models (cf. the theorems in Section 2.7). However, these bounds would be based on *local neighbourhoods* of the true distribution. (More precisely a neighbourhood of contiguous sequences of distributions). Heuristically, such lower bounds are sharp in the i.i.d. case, because the problem can be initially reduced to a simpler one, wherein the underlying distribution is approximately known. Indeed, θ can be estimated within $n^{-\frac{1}{2}}$-range and the nuisance parameter well enough for the purpose of estimating θ. Because of the possibility of this reduction, the difficulty of estimating θ is determined by the local problem.

In the presence of infinitely many nuisance parameters an initial reduction to a local problem is unlikely to be possible. Therefore, except in some special cases, notably when the efficient score function is independent of the nuisance parameters, we expect bounds based on local neighbourhoods to be unattainable. To be more precise, an estimator which attains the bound may exist for each local problem separately, but we can not find a single estimator which is efficient in every local problem. To

overcome this difficulty it seems that a concept of optimality should take into account the performance of the estimator at different non-contiguous sequences of underlying distributions simultaneously, so as to include characteristics of the global problem.

Definitions of efficiency for models with infinitely many (real) nuisance parameters are contained in Hasminski and Nussbaum (1984) and Nussbaum (1984), and Bickel and Klaassen (1986).

Hasminski and Nussbaum (1984) and Nussbaum (1984) establish a lower bound for a maximum risk over a certain set of sequences of parameters. However, because this set is rather extensive, differences in performance between estimators are not apparent in terms of this maximum risk. For instance, while Nussbaum (1984) establishes efficiency of the direct maximum likelihood estimator in the incidental version of the errors in variables model (also cf. Example 5.7.4), the performance of this estimator is improved upon by the estimator constructed in Sections 5.2-5.3.

Bickel and Klaassen (1986) consider optimality within a class of *regular* estimators. Their definition can be extended to our model and the estimator of Section 5.2-5.3 would then be a best regular estimator. Unfortunately this type of regularity seems rather strong, so that we only reach the conclusion that the estimator is best in a much restricted class.

We are not aware of any estimators in the literature which perform better than the one of Sections 5.2-5.3. However, we end this section with the following embarrassing result. The estimator constructed in Sections 5.2-5.3 is asymptotically inadmissible, at least in the non- i.i.d. model (5.29) and to first order.

A 'better' estimator can be constructed by the following scheme. Divide the set of observations in two parts. Next construct two independent estimators for θ as in Sections 5.2-5.3 separately, based on the two sets of observations. Finally combine the two estimators by taking the optimal linear combination. Here the weights should be chosen stochastic, based on estimates of the information in the two samples.

We have to leave it to the reader to judge the implications of this result. An explanation of the improvement is the following. In another attempt to define an efficiency concept, let us restrict ourselves to estimators for which $\{L(\sqrt{n}(T_n-\theta_n))\}$ has zero-mean normal limit points under every triangular array $\{(\theta_n,\eta'_{n1},\ldots,\eta'_{nn})\}$, under which the distribution of $(X_{n1},\ldots,X_{nn})$ is contiguous to that under $\{(\theta_n,\eta_{n1},\ldots,\eta_{nn})\}$. Furthermore

assume that

$$(5.78) \qquad \sqrt{n}(T_n-\theta) = n^{-\frac{1}{2}} \sum_{j=1}^{n} g_n(X_{nj}) + o_{P_{\theta\eta_{n1}..\eta_{nn}}}(1).$$

The important feature here is that g_n is independent of j, which has as interpretation that the asymptotic influence of the observations is symmetric in j.

It can be proved that an estimator is optimal within this class when it satisfies (5.78) with g_n equal to the *efficient score for the average density* (for this terminology cf. the discussion below (5.33)). Hence, under a local completeness assumption, the estimator constructed in Sections 5.2-5.3 is optimal in this sense.

On the other hand we have in the same manner that the improvement suggested above, is optimal in the class of estimators which satisfy

$$\sqrt{n}(T_n-\theta) = n^{-\frac{1}{2}} [\sum_{j=1}^{k} g_{n1}(X_{nj}) + \sum_{j=k+1}^{n} g_{n2}(X_{nj})] + o_{P_{\theta\eta_{n1}..\eta_{nn}}}(1),$$

where $k = k_n$, and $X_{n1},\ldots,X_{nk}$ respectively $X_{nk},\ldots,X_{nn}$ are the two sets in which the observations are split up. The improvement is explained by the fact that the latter class is wider.

Of course the improvement can be further improved by splitting the sample in four parts and so on. The general idea of repeated splitting is better adaptation of the estimator to the non-i.i.d. model. An estimator is optimal in the local problem with as centre the triangular array $\{(\theta,\eta_{n1},\ldots,\eta_{nn})\}$ if it satisfies

$$\sqrt{n}(T_n-\theta) = n^{-\frac{1}{2}} \sum_{j=1}^{n} \ell(X_{nj},\theta,\eta_{nj}) + o_{P_{\theta\eta_{n1}..\eta_{nn}}}(1).$$

In this expression the influence of the j-th observation does depend on j and one can best approximate this by splitting the set of observations in parts.

Note however, that it may not be a good idea to split the sample in too many parts. For instance if $X_{n1},\ldots,X_{nn}$ happen to be i.i.d. there is only one score function $\ell(\cdot,\theta,\eta)$ involved. The splitting scheme employs

estimates for this which are based on only part of the observations. So splitting must lead to a worse estimator in the i.i.d. case, though this is not apparent in the first order asymptotic behaviour as long as the number of parts in which we split the sample is finite.

5.5. MIXTURE MODELS, MODELS WITH INCIDENTAL PARAMETERS

Important examples of semi-parametric models belong to the class of *mixture models*. Let H be a collection of probability measures on a measurable space (Z,A) and for each $(\theta,z) \in \Theta \times Z$ let $\underline{p}(\cdot,\theta,z)$ be a density with respect to a σ-finite measure μ on (X,B). Suppose that $\underline{p}(x,\theta,z)$ is measurable as a function of (x,z) and set

$$p(x,\theta,\eta) = \int \underline{p}(x,\theta,z)\, d\eta(z). \tag{5.80}$$

A *mixture model* is defined by

$X_1, X_2, \ldots$ are i.i.d. random elements
$(\theta,\eta) \in \Theta \times H$ is unknown
X_j has density $p(\cdot,\theta,\eta)$ w.r.t. μ on (X,B)
$p(\cdot,\theta,\eta)$ takes the form (5.80).

Mixture models are sometimes called *structural models* as opposed to *functional models*, which also go under the name of *models with incidental parameters*. The latter type of model is described by

$X_1, X_2, \ldots$ are independent random elements
$(\theta, z_1, z_2 \ldots) \in \Theta \times Z^{\infty}$ is unknown
X_j has density $\underline{p}(\cdot,\theta,z_j)$ w.r.t. μ on (X,B)

Both types of models have been studied in the literature and it is difficult to say which of the two is of more practical interest. Fortunately it is not necessary to choose between the two, as both can be accomodated in a single set-up. Consider the model

for each $n=1,2,\ldots$

(5.81)
$X_{n1},X_{n2},\ldots,X_{nn}$ are independent random elements
$(\theta_n,\eta_{n1},\ldots,\eta_{nn}) \in \Theta\times H^n$ is unknown
X_{nj} has density $p(\cdot,\theta_n,\eta_{nj})$ w.r.t. μ on $(\mathcal{X},\mathcal{B})$
$p(\cdot,\theta,\eta)$ takes the form (5.80).

If H contains the degenerate distributions δ_z then the functional model is a submodel of (5.81). It suffices to note that $p(\cdot,\theta,\delta_z) = \underline{p}(\cdot,\theta,z)$.

Next suppose that there exist measurable functions $h(\cdot,\theta)$: $(\mathcal{X},\mathcal{B}) \to \mathbb{R}$ and $g(\cdot,\theta,z)$: $\mathbb{R}^m \to \mathbb{R}$ with

(5.82) $\quad \underline{p}(\cdot,\theta,z) = h(\cdot,\theta)\, g(\psi(\cdot,\theta),\theta,z).$

The model determined by (5.80)-(5.82) is a special case of (5.29). It follows that the constructions in Sections 5.2 and 5.3 pertain to both mixture models and models with incidental parameters, and lead to an asymptotically normal estimator for θ. We believe that this estimator performs reasonably well for both types of models. For important examples of the *i.i.d.* mixture model it is possible to prove *efficiency*.

Before discussing this, we mention two other methods of estimation.

First we can use the θ-component of the maximum likelihood estimator $(\hat{\theta}_{nm},\hat{\eta}_{nm})$ for the mixture model (which maximizes $\Pi_{j=1}^n p(X_j,\theta,\eta)$ over $\Theta\times H$). Kiefer and Wolfowitz (1956) have shown that $\hat{\theta}_{nm}$ is often consistent for θ. The question whether $\hat{\theta}_{nm}$ is $\sqrt{n}$-consistent is still unanswered, but it would be interesting to compare its performance with that of the adaptively constructed estimator of Sections 5.2-5.3.

A construction half way between the construction of Sections 5.2-5.3 and maximum likelihood as above is discussed in Van der Vaart (1987). Here the one-step method is applied with an estimate for the efficient score function $\tilde{\ell}(\cdot,\theta,\eta)$, which is obtained by substituting for η the distribution $\hat{\eta}_n(\theta)$ which maximizes

$$\Pi_{j=1}^n p(X_{nj},\theta,\eta)$$

over (some subset of) the set of all distributions η. For several examples this construction has been shown to lead to estimator sequences for θ which satisfy (5.39).

5.5.1. Efficiency for mixture models

For a convex set H of mixing distributions, a mixture model satisfies the convexity condition of Section 5.4.1.1. A simple sufficient condition for local completeness can therefore be obtained.

Define a measure ν_θ on $\mathbb{R}^m$ by

$$\nu_\theta(B) = \int 1\{\psi(x,\theta) \in B\}\, h(x,\theta)\, d\mu(x)$$

and set

$$g(s,\theta,\eta) = \int g(s,\theta,z)\, d\eta(z).$$

As an easily applicable special case of Theorem 5.10 we have

COROLLARY 5.11. *Let* (5.80), (5.82) *and* (5.3) *hold and suppose that the distributions with densities* $g(\cdot,\theta,z)$ *with respect to* ν_θ *are mutually absolutely continuous. If* H *is convex and*

$$\{\, g(\cdot,\theta,z)d\nu_\theta\colon\ \delta_z \in H,\ \int g^2(s,\theta,z)/g(s,\theta,\eta)\, d\nu_\theta(s) < \infty \,\}$$

is complete, then $\psi(X,\theta)$ *is locally complete at* (θ,η). □

In many cases it is possible to improve upon Corollary 5.11 (and Theorem 5.10) and to show *strong* local completeness of the sufficient statistic. Basically we have strong local completeness of $\psi(X,\theta)$ at (θ,η) if the mixing distribution is completely unknown and $\{g(\cdot,\theta,\eta)d\nu_\theta\colon\ z \in \text{support }(\eta)\}$ is complete.

For a careful treatment of our examples we need to make this rigorous with an adapted completeness concept, L_2*-completeness.*

DEFINITION 5.12. *A set of probability distributions* $\mathcal{P}$ *on a measurable space* $(X,\mathcal{B})$ *is called* L_2*-complete if* $b \in L_2(P)$ *and* $\int b\, dP = 0$ *for all* $P \in \mathcal{P}$, *implies that* $b = 0$ P*-a.e. for all* $P \in \mathcal{P}$. □

Let $T(\eta,H)$ denote a tangent cone for η in the set of measures H on the measurable space $(Z,\mathcal{A})$.

THEOREM 5.13. *Let* (5.80) *and* (5.82) *hold and suppose that for measurable functions* $\underline{\ell}(x,\theta,z)$: $(X\times Z,\mathcal{B}\times A) \to \mathbb{R}^k$, *every* $h \in \mathbb{R}^k$ *and* $t \to 0$

$$\int\int [t^{-1}(\underline{p}^{\frac{1}{2}}(x,\theta+th,z)-\underline{p}^{\frac{1}{2}}(x,\theta,z))-\tfrac{1}{2}h'\underline{\ell}(x,\theta,z)\underline{p}^{\frac{1}{2}}(x,\theta,z)]^2 d\mu(x)d\eta(z)\to 0.$$

Then (5.3) *holds with*

$$\ell(x,\theta,\eta) = p^{-1}(x,\theta,\eta) \int \underline{\ell}(x,\theta,z)\ \underline{p}(x,\theta,z)\ d\eta(z)$$

Furthermore assume that $\{\underline{g}(\cdot,\theta,z)d\nu_\theta\colon z \in A\}$ *is* L_2*-complete for every* $A \in \mathcal{A}$ *with* $\int_A d\eta = 1$. *If* $T(\eta,H) = L_{2*}(\eta)$, *then* $\psi(X,\theta)$ *is strongly locally complete at* (θ,η). *That is, we may choose as a tangent cone*

$$(5.83)\qquad T(p(\cdot,\theta,\eta)) = \{h'\ell(\cdot,\theta,\eta) + \underline{b}(\psi(\cdot,\theta))\colon h \in \mathbb{R}^k,\ \underline{b} \in L_{2*}(g(\cdot,\theta,\eta))\},$$

and for any $h \in \mathbb{R}^k$ *and* $\underline{b} \in L_{2*}(g(\cdot,\theta,\eta))$ *there exists* $\{\eta_t\} \subset H$ *satisfying* (5.13). □

PROOF. Given $\underline{c} \in L_{2*}(\eta)$ choose $\{\eta_t\} \subset H$ $(t > 0)$ such that

$$\int\int \left[\ t^{-1}((d\eta_t)^{\frac{1}{2}}-(d\eta)^{\frac{1}{2}}) - \tfrac{1}{2}\underline{c}(d\eta)^{\frac{1}{2}}\ \right]^2 \to 0.$$

Assume without loss of generality that η_t and η have densities k_t and k with respect to a probability measure τ. Then

$$\int\int [t^{-1}(\underline{p}^{\frac{1}{2}}(x,\theta+th,z)k_t^{\frac{1}{2}}(z) - \underline{p}^{\frac{1}{2}}(x,\theta,z)k^{\frac{1}{2}}(z))$$

$$- \tfrac{1}{2}(h'\underline{\ell}(x,\theta,z)+\underline{c}(z))\ \underline{p}^{\frac{1}{2}}(x,\theta,z)k^{\frac{1}{2}}(z)]^2\ d\mu(x)d\tau(z)$$

$$\le 3\int\int[t^{-1}(\underline{p}^{\frac{1}{2}}(x,\theta+th,z)-\underline{p}^{\frac{1}{2}}(x,\theta,z))-\tfrac{1}{2}h'\underline{\ell}(x,\theta,z)\underline{p}^{\frac{1}{2}}(x,\theta,z)]^2 d\mu(x)d\eta(z)$$

$$+ 3\int\int [t^{-1}(k_t^{\frac{1}{2}}(z)-k^{\frac{1}{2}}(z)) - \tfrac{1}{2}\underline{c}(z)k^{\frac{1}{2}}(z)]^2\ \underline{p}(x,\theta+th,z)\ d\mu(x)d\tau(z)$$

$$+ 3/4\int\int [\underline{c}(z)k^{\frac{1}{2}}(z)]^2\ [\underline{p}^{\frac{1}{2}}(x,\theta+th,z)-\underline{p}^{\frac{1}{2}}(x,\theta,z)]^2\ d\mu(x)d\tau(z)$$

$$\le o(1) + t^{-1}\int\int [\underline{p}^{\frac{1}{2}}(x,\theta+th,z)-\underline{p}^{\frac{1}{2}}(x,\theta,z)]^2\ d\mu(x)d\eta(z)$$

$$+ 2\int \underline{c}^2(z)\, 1\{|\underline{c}(z)|\geq t^{-\frac{1}{2}}\}\, d\eta(z) \to 0.$$

By Lemma 5.18 conclude that

$$\int [t^{-1}(p^{\frac{1}{2}}(x,\theta+th,\eta_t)-p^{\frac{1}{2}}(x,\theta,\eta)) - \tfrac{1}{2}(h'\ell(x,\theta,\eta)+A\underline{c}(\psi(x,\theta)))\, p^{\frac{1}{2}}(x,\theta,\eta)]^2\, d\mu(x)$$

converges to zero, where

$$\ell(x,\theta,\eta) = p^{-1}(x,\theta,\eta) \int \underline{\ell}(x,\theta,z)\, \underline{p}(x,\theta,z)\, d\eta(z)$$

$$A\underline{c}(s) = g^{-1}(s,\theta,\eta) \int \underline{c}(z)\, \underline{g}(s,\theta,z)\, d\eta(z).$$

Setting $\eta_t = \eta$ and $\underline{c} = 0$ we obtain (5.3).

Next we prove that the linear space $\{A\underline{c}(\cdot)\colon \underline{c} \in L_{2*}(\eta)\}$ is dense in $L_{2*}(g(\cdot,\theta,\eta))$. Indeed suppose that $\underline{b} \in L_{2*}(g(\cdot,\theta,\eta))$ and $\underline{b} \perp A\underline{c}$ for all $\underline{c} \in L_{2*}(\eta)$. Then

$$\int\int \underline{b}(s)\, 1\{g(s,\theta,\eta)>0\}\, \underline{g}(s,\theta,z)\, d\nu_\theta(s)\, \underline{c}(z)\, d\eta(z) = 0.$$

Hence

$$\int \underline{b}(s)\, 1\{g(s,\theta,\eta)>0\}\, \underline{g}(s,\theta,z)\, d\nu_\theta(s) = 0 \qquad \eta\text{-a.a. } z.$$

By L_2-completeness

$$\underline{b}(s)\, 1\{g(s,\theta,\eta)>0\} = 0 \qquad g(\cdot,\theta,z)\text{-a.e., } \eta\text{-a.a. } z.$$

Hence $\underline{b} = 0$ $g(\cdot,\theta,\eta)$-a.e..

So far we have shown that the set of $g \in L_2(p(\cdot,\theta,\eta))$ for which (5.13) holds, is dense in $T(p(\cdot,\theta,\eta))$ given by (5.83). Finally by a 'diagonalization' scheme we can for any $g \in T(p(\cdot.\theta,\eta))$ find $\{\eta_t\} \subset H$ such that (5.13) holds.

Let $\underline{b} \in L_{2*}(g(\cdot,\theta,\eta))$ be arbitrary. Then by the above argument wthere exists $\{\underline{b}_n\} \subset L_{2*}(g(\cdot,\theta,\eta))$ (of the form Ac_n) with $\underline{b}_n \to \underline{b}$ and such for every $n = 1,2,\ldots$ there is $\{\eta_{nt}\} \subset H$ such that

$$r_{nt} = \int [t^{-1}(p^{\frac{1}{2}}(x,\theta+th,\eta_{nt})-p^{\frac{1}{2}}(x,\theta,\eta)) - \tfrac{1}{2}(h\dot{\ell}(x,\theta,\eta)+\underline{b}_n(\psi(x,\theta)))\, p^{\frac{1}{2}}(x,\theta,\eta)]^2 d\mu(x) \to 0$$

as $t \downarrow 0$. Choose a sequence $\{t_n\}$ with $t_n \downarrow 0$ such that $r_{nt} < n^{-1}$ if $t \le t_n$. Next let η_t be η_{nt} if $t_{n+1} < t \le t_n$. Then $\{\eta_t\}$ satisfies (5.7). ∎

Important examples of mixture models are generated by the exponential family. Suppose that $g(\cdot,\theta,z)$ in (5.82) takes the form

$$g(s,\theta,z) = c(z,\theta)\, d(s,\theta)\, e^{s'ze(\theta)} \qquad (s,z \in \mathbb{R}^m) \tag{5.84}$$

Let $Z(\theta)$ be the set of z-values for which the family is defined, i.e.

$$Z(\theta) = \{z \in \mathbb{R}^m: \int d(s,\theta) \exp(s'z\, e(\theta))\, d\nu_\theta(s) < \infty\}.$$

Next let H be a set of probability distributions on $Z = \cap \{Z(\theta): \theta \in \Theta\}$.

As preparation for the next lemma we introduce the notion of continuation set. Recall that a function $f: G \to \mathbb{C}$ defined on an open subset G of $\mathbb{C}^m$, is said to be *analytic* if it is analytic in each of its m arguments separately. We shall call a set $B \subset \mathbb{R}^m$ a *continuation set* if for every convex open set $C \subset \mathbb{R}^m$ containing $\bar{B}$, any analytic function defined on $G = \{\zeta \in \mathbb{C}^m: \text{Re } \zeta \in C\}$, which is constant on B, is necessarily constant on G. In case m = 1, a set B is a continuation set if and only if it has a limit point. To our knowledge, there is no simple necessary and sufficient condition for a set to be a continuation set for m > 1, which is the reason to introduce the concept. A simple sufficient condition for $B \subset \mathbb{R}^m$ to be a continuation set is that B has nonempty interior (as a subset of $\mathbb{R}^m$), but this can be relaxed considerably. Applying the above mentioned result for m = 1 to the coordinate functions, one for instance sees that the set $\{(z_1,z_2) = (p+q^{-1},p^{-1}): p=1,2,..,\ q=1,2,...\} \cup \{(0,0)\}$ is a continuation set in $\mathbb{R}^2$.

Recall that the support of a probability distribution on $Z \subset \mathbb{R}^m$ is the smallest closed set with probability one.

LEMMA 5.14. *Let* (5.84) *hold with* $e(\theta) \neq 0$ *and suppose that the support of* η *contains a closed continuation set within the interior of* Z. *Then* $\{g(\cdot,\theta,z)d\nu_\theta\colon z \in A\}$ *is* L_2*-complete for every* $A \in \mathcal{A}$ *with* $\int_A d\eta = 1$. □

PROOF. Let $\int_A d\eta(z) = 1$ and let $b\colon \mathbb{R}^m \to \mathbb{R}$ be measurable and satisfy

$$\int b^2(s)\, g(s,\theta,z)\, d\nu_\theta(s) < \infty \qquad \text{all } z \in A$$

$$\int b(s)\, g(s,\theta,z)\, d\nu_\theta(s) = 0 \qquad \text{all } z \in A. \tag{5.85}$$

Set

$$Z_b(\theta) = \{z \in Z\colon \int |b(s)|\, g(s,\theta,z)\, d\nu_\theta(s) < \infty\}.$$

Since $\text{support}(\eta) \subset \bar{A}$, we are guaranteed a closed continuation set $B \subset \text{Int } Z \cap \bar{A}$. We first show that that $B \subset \text{Int } Z_b(\theta)$. Indeed, given $z_0 \in B$, there exists a sequence $\{z_j\} \subset A$ with $z_j \to z_0$. For any z and z_j

$$\int |b(s)|\, e^{s'ze(\theta)}\, c(z,\theta)\, d(s,\theta)\, d\nu_\theta(s)$$

$$\leq \left[\int b^2(s)\, g(s,\theta,z_j)\, d\nu_\theta(s)\right]^{\frac{1}{2}} \left[\int e^{4s'(z-z_0)e(\theta)}\, g(s,\theta,z_j)\, d\nu_\theta(s)\right]^{\frac{1}{4}}$$

$$\left[\int e^{4s'(z_0-z_j)e(\theta)}\, g(s,\theta,z_j)\, d\nu_\theta(s)\right]^{\frac{1}{4}} c(z,\theta)/c(z_j,\theta).$$

But, as $z_0 \in \text{Int } Z$, this is finite for sufficiently large j and small $\|z-z_0\|$.

It is well-known that the function $\zeta \to \int b(s)\, g(s,\theta,\zeta)\, d\nu_\theta(s)$, is analytic on $G = \{\zeta \in \mathbb{C}^m\colon \text{Re } \zeta \in \text{Int } Z_b(\theta)\}$. By continuity we see that (5.85) must hold for all $z \in B$. Next, because B is a continuation set, we conclude that

$$\int b(s)\, g(s,\theta,\zeta)\, d\nu_\theta(s) = 0 \qquad \text{all } \zeta \in G.$$

Hence, there exists $u \in \text{Int } Z_b(\theta)\cdot e(\theta)$ such that for all $v \in \mathbb{R}$

$$\int e^{is'v}\, b^+(s)e^{s'u} d(s,\theta)\, d\nu_\theta(s) = \int e^{is'v}\, b^-(s)e^{s'u} d(s,\theta)\, d\nu_\theta(s).$$

By uniqueness of Fourier transforms the finite measures given by

$\tau^+(B) = \int b^+(s)e^{s'u}d(s,\theta)\ d\nu_\theta(s)$ and $\tau^-(B) = \int b^-(s)e^{s'u}d(s,\theta)\ d\nu_\theta(s)$ respectively, are equal. Hence

$$b^+ = b^- \qquad \text{a.e.} \quad [g(\cdot,\theta,z)d\nu_\theta]. \ \blacksquare$$

5.6. ESTIMATION OF SOME OTHER FUNCTIONALS

In this section we restrict ourselves to the i.i.d. model described in Section 5.1.1. Hence $X_1,\ldots,X_n$ are i.i.d. random elements in $(\mathcal{X},\mathcal{B})$ with density $p(\cdot,\theta,\eta)$ satisfying (5.1)-(5.2). For convenience of notation let X be another independent copy of X_1.

In the foregoing sections we have discussed estimation of θ and constructed an efficient estimator for θ in a variety of cases. Under local completeness many other functionals can be estimated efficiently as well. We shall not prove this in great generality, but content ourselves with estimating linear functionals of the form

$$\kappa(p(\cdot,\theta,\eta)) = \int f(x)\ p(x,\theta,\eta)\ d\mu(x)\ ,$$

where $f\colon (\mathcal{X},\mathcal{B}) \to \mathbb{R}$ is a given measurable function.

κ is differentiable under a slight regularity condition on f. Let

$$B(\theta,\eta,\varepsilon) = \{(\theta',\eta') \in \Theta\times H\colon \int |p(x,\theta',\eta')-p(x,\theta,\eta)|\ d\mu(x) < \varepsilon\}$$

and suppose that for some $\varepsilon > 0$

$$\sup\ \{\ E_{\theta',\eta'}f^2(X)\colon (\theta',\eta') \in B(\theta,\eta,\varepsilon)\ \} < \infty.$$

Then it follows from Lemma 5.21 that κ is differentiable at $p(\cdot,\theta,\eta)$ with influence function

$$\dot{\kappa}(\cdot,p(\cdot,\theta,\eta)) = f(\cdot).$$

Under local completeness the efficient influence function is given by

(5.86) $$\tilde{\kappa}(\cdot,p(\cdot,\theta,\eta)) = c(\theta,\eta)'\tilde{I}^{-1}(\theta,\eta)\tilde{\ell}(\cdot,\theta,\eta)$$

$$+ E_\theta(f(X)\,|\,\psi(X,\theta)=\psi(\cdot,\theta)) - E_{\theta\eta}f(X),$$

where (cf. Lemma 2.4)

$$c(\theta,\eta) = E_{\theta\eta}f(X)\tilde{\ell}(X,\theta,\eta).$$

A definition of an efficient estimator for κ may be motivated as follows. Local completeness requires that, for fixed θ, the distribution $G(\theta,\eta)$ of $\psi(X,\theta)$ can be approached along one-dimensional submodels from any direction. In consequence, asymptotic bounds for the present model are the same as for the situation where $G(\theta,\eta)$ is completely unknown. Thus, for given θ, the empirical distribution $\hat{G}_n(\theta)$ of $\psi(X_1,\theta),\ldots,\psi(X_n,\theta)$ is an efficient estimator for $G(\theta,\eta)$. ($\hat{G}_n(\theta)$ is the measure putting mass n^{-1} in each of the $\psi(X_j,\theta)$).

Now

$$\kappa(p(\cdot,\theta,\eta)) = \int E_\theta(f(X)\,|\,\psi(X,\theta) = s)\; dG(\theta,\eta)(s).$$

We are lead to consider estimators for κ of the form

(5.87) $$\int E_{\hat{\theta}_n}(f(X)\,|\,\psi(X,\hat{\theta}_n) = s)\; d\hat{G}_n(\hat{\theta}_n)(s).$$

Of course we should not expect to obtain an efficient estimator for κ in this manner, unless $\hat{\theta}_n$ is efficient for θ. Under local completeness $\hat{\theta}_n$ is efficient for θ if (cf. (5.73))

(5.88) $$\sqrt{n}(\hat{\theta}_n-\theta) = n^{-\frac{1}{2}}\Sigma_{j=1}^n\, \tilde{I}^{-1}(\theta,\eta)\,\tilde{\ell}(X_j,\theta,\eta) + o_{P_{\theta\eta}}(1)\ ,$$

so that efficient $\hat{\theta}_n$ generally exist (cf. Sections 5.2-5.3).

For technical reasons we take a different route. We wish to employ a *discretized* $\sqrt{n}$-consistent estimator $\bar{\theta}_n$ in (5.87), which, unfortunately, is necessarily inefficient. However, the inefficiency of $\bar{\theta}_n$ can be remedied by the addition of a correction term. With $\hat{\theta}_n$ efficient for θ we set

(5.89) $$T_n = \hat{c}_n(\bar{\theta}_n)'(\hat{\theta}_n-\bar{\theta}_n) + \int E_{\bar{\theta}_n}(f(X)\,|\,\psi(X,\bar{\theta}_n) = s)\; d\hat{G}_n(\bar{\theta}_n)(s)\ ,$$

where the vector $\hat{c}_n(\theta)$ is defined by

$$(\hat{c}_n(\theta))_j = \sqrt{n}\Big[\int E_{\theta+n^{-\frac{1}{2}}e_j}(f(X)\,|\,\psi(X,\theta+n^{-\frac{1}{2}}e_j) = s)\; d\hat{G}_n(\theta+n^{-\frac{1}{2}}e_j)(s)$$

$$- \int E_\theta(f(X)\,|\,\psi(X,\theta) = s)\; d\hat{G}_n(\theta)(s)\Big].$$

As before there is no other justification for discretization than mathematical convenience. The price paid for simple proofs and weak conditions is the term $\hat{c}_n(\bar{\theta}_n)'(\hat{\theta}_n-\bar{\theta}_n)$, which admittedly is not very pleasant.

THEOREM 5.15. *Let* (5.1)-(5.3) *and* (5.41) *hold and suppose that* $f\colon (X,\mathcal{B}) \to \mathbb{R}$ *is measurable with*

$$\text{(5.90)}\qquad \limsup_{t\to 0} E_{\theta+t,\eta} f^2(X) < \infty.$$

$$\text{(5.91)}\qquad \int [E_{\theta+t}(f(X)\,|\,\psi(X,\theta+t) = s)\; p^{\frac{1}{2}}(x,\theta+t,\eta) - E_\theta(f(X)\,|\,\psi(X,\theta) = s)\; p^{\frac{1}{2}}(x,\theta,\eta)]^2 d\mu(x) \to 0.$$

If $\hat{\theta}_n$ *satisfies* (5.88), *then* T_n *given by* (5.89) *satisfies (cf.* (5.86)*)*

$$\sqrt{n}(T_n - E_{\theta\eta} f(X)) = n^{-\frac{1}{2}} \Sigma_{j=1}^n \tilde{\kappa}(X_j, p(\cdot,\theta,\eta)) + o_{P_{\theta\eta}}(1).$$

Under local completeness this implies that $\{T_n\}$ *is efficient for* κ *at* $p(\cdot,\theta,\eta)$ *(provided that* κ *is differentiable).* □

PROOF. Because of the discretized nature of $\bar{\theta}_n$, we have by the same argument as in the proof of Theorem 5.3 that it suffices to prove for any sequence $\{\theta_n\} \subset \mathbb{R}^k$ with $\|\theta_n-\theta\| = O(n^{-\frac{1}{2}})$

$$\text{(5.92)}\qquad \sqrt{n}\Big[\hat{c}_n(\theta_n)'(\hat{\theta}_n-\theta_n) + \int E_{\theta_n}(f(X)\,|\,\psi(X,\theta_n)=s)\; d\hat{G}_n(\theta_n)(s) - E_{\theta\eta} f(X)\Big] = n^{-\frac{1}{2}} \Sigma_{j=1}^n \tilde{\kappa}(X_j, p(\cdot,\theta,\eta)) + o_{P_{\theta\eta}}(1).$$

By Lemma 5.23, (5.41) and (5.90)

$$\sqrt{n}\,(E_{\theta_n\eta}f(X) - E_{\theta\eta}f(X)) = \sqrt{n}(\theta_n-\theta)'E_{\theta\eta}f(X)\dot{\ell}(X,\theta,\eta) + o(1). \tag{5.93}$$

By Proposition A.10, (5.41) and (5.93) we have

$$n^{-\frac{1}{2}}\Sigma_{j=1}^{n}\,[\,E_{\theta_n}(f(X_j)|\psi(X_j,\theta_n)) - E_{\theta}(f(X_j)|\psi(X_j,\theta))\,]$$

$$= \sqrt{n}(\theta_n-\theta)'[\,-E_{\theta\eta}(E_{\theta}(f(X)|\psi(X,\theta))\dot{\ell}(X,\theta,\eta) + E_{\theta\eta}f(X)\dot{\ell}(X,\theta,\eta)\,] + o_{P_{\theta\eta}}(1) \tag{5.94}$$

$$= \sqrt{n}(\theta_n-\theta)'c(\theta,\eta) + o_{P_{\theta\eta}}(1).$$

As a consequence of (5.94)

$$\hat{c}_n(\theta_n) - c(\theta,\eta) \overset{P_{\theta\eta}}{\to} 0\ . \tag{5.95}$$

The theorem follows from combination of (5.92), (5.94)-(5.95), (5.86) and (5.88). ■

It is possible to construct efficient estimator sequences for many other classes of functionals κ, such as M or L functionals, in the same manner. The following route may be attractive to do this in general.

Taking f equal to the indicator of a suitable set $C \in \mathcal{B}$ we get by the above procedure anefficient estimator sequence $T_n(C)$ for $P_{\theta\eta}(X \in C)$. Next we consider estimation of the distribution of X, seen as an element of the space B(C) as defined in 3.6.2, for a suitable collection of sets $\mathcal{C} \subset \mathcal{B}$. It follows from Theorem 5.15 that the marginals of the B-valued process

$$\{\ T_n(C):\ C \in \mathcal{C}\ \}$$

are efficient estimators for the corresponding marginals of the distribution $\{\ P_{\theta\eta}(X \in C):\ C \in \mathcal{C}\ \}$. Then if we prove tightness of the process

$$\{\ \sqrt{n}(T_n(C)-P_{\theta\eta}(C)):\ C \in \mathcal{C}\ \}$$

we have efficiency of the estimator $\{ T_n(C): C \in \mathcal{C} \}$ for the distribution $\{ P_{\theta\eta}(X \in C): C \in \mathcal{C} \}$ of X, seen as an element of B. Next Theorem 4.9 may be applied to obtain efficient estimators for other functionals.

For the purpose of proving tightness, discretization of $\bar{\theta}_n$ is useful. Indeed, it suffices to show tightness of the B-valued process

$$\int P_{\theta_n}(X \in C | \psi(X,\theta_n) = s) \, d[\sqrt{n}\,(\hat{G}_n(\theta_n) - G(\theta_n,\eta))\,](s) ,$$

for every sequence $\{\theta_n\} \subset \mathbb{R}^k$ with $\|\theta_n - \theta\| = O(n^{-\frac{1}{2}})$. But this is the empirical process of $\psi(X_1,\theta_n),\ldots,\psi(X_n,\theta_n)$ indexed by the sets of functions

$$\mathcal{F}_n = \{ f(s) = P_{\theta_n}(X \in C | \psi(X,\theta_n) = s): C \in \mathcal{C} \}.$$

Using a uniform version of a central limit theorem for empirical processes indexed by sets of functions, it is possible to give simple sufficient conditions for tightness.

5.7. EXAMPLES

Because of space limitations we only discuss a selection of examples and not all of them in depth. In particular we give only three examples from the rich class of mixture models described in Section 5.5. See e.g. Heckman and Singer (1984) and references cited there for some applications of this type of models in practice. Also see Lindsay (1983, 1985).

We do not try to be fully consistent in notation. The observed variables X may be replaced by pairs (X,Y) and what was called θ before may be called (μ,Σ) below and so on.

5.7.1. Symmetric location in $\mathbb{R}$

Let H be a set of probability densities η with respect to Lebesgue measure λ on $\mathbb{R}$, absolutely continuous, symmetric about zero, and with finite and positive Fisher information for location

$$I_1(\eta) = \int (\eta'/\eta)^2 \, \eta \, d\lambda.$$

Let $p(x,\theta,\eta) = \eta(x-\theta)$. Then (5.1)-(5.3) are satisfied with the sufficient statistic $\psi(X,\theta) = |X-\theta|$, which has density $g(\cdot,\theta,\eta) = 2\eta(\cdot)1_{(0,\infty)}(\cdot)$ with respect to λ. It is straightforward to check the conditions for Theorems 5.3 and Corollary 5.9 for the i.i.d. model of Section 5.1.1. (In Theorem 5.16 below we prove a stronger result). Hence we obtain an asymptotically normal estimator for the centre of symmetry θ, satisfying (5.73). We have

$$E_\theta(\dot{\ell}(X,\theta,\eta) \mid |X-\theta| = s) = E_\theta(-\eta'/\eta(X-\theta) \mid |X-\theta| = s) = 0.$$

It follows that the estimator is efficient regardless of the size of the set H (cf. (5.74)). This model has been studied by many authors and our only contribution is to analyse the model in terms of a sufficient statistic.

Let us now consider the non-i.i.d. version (5.29). The following extends Bickel and Klaassen (1986). Set

$$\bar{\eta}_n = n^{-1}\Sigma_{j=1}^n \eta_{nj}.$$

The following theorem gives sufficient conditions for applicability of Theorem 5.3 and Corollary 5.9.

THEOREM 5.16. *Suppose that for* $\theta_n = O(n^{-\frac{1}{2}})$ *and every* $\varepsilon > 0$

$$n^{-1}\Sigma_{j=1}^n \int [\eta_{nj}'/\eta_{nj}^{\frac{1}{2}}(x+\theta_n) - \eta_{nj}'/\eta_{nj}^{\frac{1}{2}}(x)]^2 \, d\lambda(x) \to 0$$

$$n^{-1}\Sigma_{j=1}^n \int [\eta_{nj}'/\eta_{nj}^{\frac{1}{2}}(x)]^2 \, d\lambda(x) = O(1)$$

$$n^{-1}\Sigma_{j=1}^n \int [\eta_{nj}'/\eta_{nj}^{\frac{1}{2}}(x)]^2 \, 1\{|\eta_{nj}'/\eta_{nj}(x)| \geq \varepsilon\sqrt{n}\} \, d\lambda(x) \to 0.$$

Furthermore assume the existence of an absolutely continuous density η *with* $I_1(\eta) \in (0,\infty)$, *such that*

(5.96) $\bar{\eta}_n \to \eta$ *weakly as measures on* $\mathbb{R}$

(5.97) $I_1(\bar{\eta}_n) \to I_1(\eta)$.

Then there exists an estimator satisfying (5.39). □

PROOF. It suffices to check (5.41)-(5.49), (5.59)-(5.60) and (5.65)-(5.68), with the notation of Corollary 5.9. Here (5.42)-(5.43) and (5.59) are true by assumption. For (5.47)-(5.48), (5.60) and (5.68) note that

$$\tilde{\ell}(x,\theta,\eta) = -\eta'/\eta(x-\theta) = \eta'/\eta(|x-\theta|)\ \tilde{\psi}(x,\theta),$$

where

$$\tilde{\psi}(x,\theta) = -\mathrm{sgn}(x-\theta).$$

For (5.41), we use that $\eta_{nj}^{\frac{1}{2}}$ is absolutely continuous with derivative $\frac{1}{2}\eta'_{nj}/\eta_{nj}^{\frac{1}{2}}$, so that for $\theta_n = O(n^{-\frac{1}{2}})$

$$n^{-1}\ \Sigma_{j=1}^{n}\ \int\ [\theta_n^{-1}(\eta_{nj}^{\frac{1}{2}}(x-\theta_n)-\eta_{nj}^{\frac{1}{2}}(x)) + \tfrac{1}{2}\eta'_{nj}/\eta_{nj}^{\frac{1}{2}}(x)]^2\ d\lambda(x)$$

$$= \tfrac{1}{4}n^{-1}\Sigma_{j=1}^{n}\ \int\ [{}_0\int^1 -\eta'_{nj}/\eta_{nj}^{\frac{1}{2}}(x-\theta_n u)du + \eta'_{nj}/\eta_{nj}^{\frac{1}{2}}(x)]^2\ d\lambda(x)$$

$$\leq \tfrac{1}{4}\ \sup_{u\in[0,1]}\ n^{-1}\Sigma_{j=1}^{n}\ \int\ [\eta'_{nj}/\eta_{nj}^{\frac{1}{2}}(x-\theta_n u)\ -\ \eta'_{nj}/\eta_{nj}^{\frac{1}{2}}(x)]^2\ d\lambda(x).$$

Next, as is shown in Bickel and Klaassen (1986), (5.96)-(5.97) ensure that

$$\int\ [\bar{\eta}'_n/\bar{\eta}_n^{\frac{1}{2}}(x)\ -\ \eta'/\eta^{\frac{1}{2}}(x)]^2\ d\lambda(x) \to 0.$$

$$\int\ |\ \bar{\eta}_n\ -\ \eta\ |\ d\lambda \to 0.$$

Relations (5.44)-(5.46) and (5.65)-(5.67) are simple consequences of this and the other assumptions.

Finally, a $\sqrt{n}$-consistent estimator $\hat{\theta}_n$ can be defined by

$$\Sigma_{j=1}^{n}\ (1-2F)(X_{nj}-\hat{\theta}_n) = 0,$$

where $F(x) = (1 + e^{-x})^{-1}$ is the logistic distribution function. Indeed we have

$$P(\ \sqrt{n}(\hat{\theta}_n-\theta) \leq x\ ) = P(\ \Sigma_{j=1}^{n}\ (1-2F)(X_{nj}-\theta-xn^{-\frac{1}{2}}) \geq 0\ )$$

$$= P(\Sigma_{j=1}^{n}\ [(1-2F)(X_{nj}-\theta) + 2n^{-\frac{1}{2}}f(X_{nj}-\theta) - 2x^2n^{-1}f'(X_{nj}-\tau_n(x))] \geq 0\ )$$

$$\to N(0,\ \int\ (1-2F)^2\ \eta\ d\lambda\ )(\ -\infty,\ 2x\ \int f\eta d\lambda\ ].\ \blacksquare$$

5.7.2. Symmetric location in higher dimensions

A natural generalization of the foregoing example to k-dimensional observations is obtained by assuming that X has a density with respect to Lebesgue measure on $\mathbb{R}^k$, satisfying

$$p(x,\theta,\eta) = \eta(x-\theta) = \eta(\theta-x) \qquad (x,\theta \in \mathbb{R}^k).$$

There are several good choices of a sufficient statistic $\psi(X,\theta)$ in this model, all reducing $\mathbb{R}^k$ to a half space. For instance, define a map $t: \mathbb{R}^k \to \mathbb{R}^k$ by

$$t(s) = \begin{matrix} s \\ -s \end{matrix} \quad \text{if} \quad \begin{matrix} s^1 \geq 0 \\ s^1 < 0 \end{matrix} .$$

Then $\psi(X,\theta) = t(X-\theta)$ is sufficient for η.

Let $\Theta = \mathbb{R}^k$ and assume that η is absolutely continuous on $\mathbb{R}^k$, in the sense that there exists a function $\nabla\eta: \mathbb{R}^k \to \mathbb{R}^k$ such that

$$\eta(x+h) - \eta(x) = {}_0\int^1 h'\nabla\eta(x+uh)\, du , \qquad (x,h \in \mathbb{R}^k) .$$

Moreover assume $\int \|\nabla\eta/\eta^{\frac{1}{2}}(x)\|^2\, dx < \infty$. Then it is easily shown that

$$\int [\eta^{\frac{1}{2}}(x+h) - \eta^{\frac{1}{2}}(x) - h'\nabla\eta/\eta\ \eta^{\frac{1}{2}}(x)]^2\, dx = o(\|h\|^2) .$$

This implies (5.3) with

$$\ell(x,\theta,\eta) = -\nabla\eta/\eta(x-\theta) .$$

Now

$$E_\theta(\ell(X,\theta,\eta) \mid \psi(X,\theta) = s) = -\tfrac{1}{2}[\nabla\eta/\eta(s) + \nabla\eta/\eta(-s)] = 0 \qquad \text{a.e..}$$

Hence

$$\tilde{\ell}(x,\theta,\eta) = -\nabla\eta/\eta(x-\theta) = \nabla\eta/\eta(t(x-\theta))\ \tilde{\psi}(x,\theta) ,$$

where

$$\psi(x,\theta) = -\operatorname{sgn}(x^1-\theta^1) .$$

Just as in the one-dimensional case, it follows that the estimator sequence

which is constructed in this chapter, is efficient regardless the size of the set of possible densities η (provided $\eta(s) = \eta(-s)$).

I thank J. Fabius for attracting my attention to this example.

5.7.3. Elliptic distributions.

Let $X_1,\ldots,X_n$ be i.i.d. random elements in $\mathbb{R}^k$ with a density

$$p(x,\mu,\Sigma,\eta) = \eta(\ (x-\mu)'\Sigma(x-\mu)\ )\ \det \Sigma^{\frac{1}{2}}$$

with respect to Lebesgue measure. Here $\mu \in \mathbb{R}^k$ and Σ is an unknown symmetric, positive definite matrix with determinant 1, and $\eta(\|s\|^2)$ is an unknown density with respect to Lebesgue measure on $\mathbb{R}^k$.

It is clear that this class of densities is contained in the class of densities of 5.7.2, where it was shown that μ can be estimated adaptively in the sense of Bickel (1982). Hence for the purpose of estimating the location parameter, one does not gain in first order efficiency by assuming the present *elliptic* structure.

To estimate the pair $\theta = (\mu,\Sigma)$ (viewed as an element of an open set in $\mathbb{R}^{k+\frac{1}{2}k(k+1)-1}$, we can use $\psi(x,\mu,\Sigma) = (x-\mu)'\Sigma(x-\mu)$ as a sufficient statistic. It has density $g(s,\eta) = \frac{1}{2}c_k\eta(s)s^{\frac{1}{2}k-1}$ on (o,∞), for constants c_k.

Calculation of scores is not entirely obvious in this model. For a further discussion we refer to Bickel (1982).

5.7.4. Two sample location scale

Let H consist of probability densities with respect to Lebesgue measure on $\mathbb{R}$, absolutely continuous and with finite and positive Fisher information for both location and scale, i.e. $I_l(\eta) \in (0,\infty)$ and

$$I_s(\eta) = \int (1+x\eta'/\eta(x))^2\eta(x)\ d\lambda(x) \in (0,\infty).$$

Let $p(x,y,\mu,\sigma,\eta) = \eta(x)\ \sigma^{-1}\eta(\sigma^{-1}(y-\mu))$. Then we have a version of the two sample problem, where the two samples have been paired so as to fit with an i.i.d model. (A more natural set-up is considered in van der Vaart (1986a)). Choose as a sufficient statistic the ordered pair

$$(\ X\wedge\sigma^{-1}(Y-\mu)\ ,\ X\vee\sigma^{-1}(Y-\mu)\ ).$$

It can be checked that (5.1)-(5.3) hold with $g(s_1,s_2) = 2\eta(s_1)\eta(s_2)1\{s_1<s_2\}$ and

$$\ell(x,y,\mu,\sigma,\eta) = -\sigma^{-1}\ (\ \eta'/\eta(\sigma^{-1}(y-\mu))\ ,\ 1+\sigma^{-1}(y-\mu)\eta'/\eta(\sigma^{-1}(y-\mu))\ )'.$$

Hence the decomposition (5.47) holds with

$$\tilde{\ell}(x,y,\mu,\sigma,\eta) = \tfrac{1}{2}\sigma^{-1}\ (\ \eta'/\eta(x) - \eta'/\eta(\sigma^{-1}(y-\mu)),$$

$$x\eta'/\eta(x) - \sigma^{-1}(y-\mu)\eta'/\eta(\sigma^{-1}(y-\mu))\ )'$$

$$\tilde{H}(x,y,\mu,\sigma) = 0$$

$$\tilde{\psi}(x,y,\mu,\sigma) = \tfrac{1}{2}\sigma^{-1}\ [\ \begin{bmatrix} 1 & -1 \\ x & -(y-\mu)/\sigma \end{bmatrix} 1\{x\sigma<y-\mu\} + \begin{bmatrix} -1 & 1 \\ -(y-\mu)/\sigma & x \end{bmatrix} 1\{x\sigma>y-\mu\}\ ]$$

$$Q(s,\mu,\eta) = \nabla g/g(s_1,s_2) = (\eta'/\eta(s_1),\eta'/\eta(s_2))'\ 1\{s_1<s_2\}.$$

Hence, letting $\|(\alpha_{ij})\|^2 = \Sigma\Sigma\ \alpha_{ij}^2$,

$$E_{\mu\sigma}(\ \|\tilde{\psi}(X,Y,\mu,\sigma)\|^2 |\ \psi(X,Y,\mu,\sigma) = s) = \tfrac{1}{4}\sigma^{-2}\ (2+s_1^2+s_2^2).$$

Assume that $\tilde{I}(\theta,\eta)$ is nonsingular and a $\sqrt{n}$-consistent estimator given. It is straightforward to check the conditions of Theorem 5.3 and Corollary 5.9 for the i.i.d version of the model, so we are assured of the existence of an asymptotically normal estimator satisfying (5.73). As for efficiency: $\psi(X,Y,\mu,\sigma)$ is not locally complete in this model, even if H is the set of all densities η as above. Scores for η take the form

$$\underline{b}(X) + \underline{b}(\sigma^{-1}(Y-\mu))$$

($\underline{b} \in L_{2*}(\eta)$), which does not span the class of *all* zero-mean functions of the sufficient statistic in $L_2(p(\cdot,\mu,\sigma,\eta))$. Nevertheless (5.74) holds, for instance if H contains with each η also the location-scale family $\{\ \tau^{-1}\eta(\tau^{-1}(s-\nu))\colon\ \tau > 0,\ \nu \in \mathbb{R}\ \}$.

We do not discuss the non-i.i.d. case.

5.7.5. A paired hazard model

Let H be a class of probability distributions on $(0,\infty) \subset \mathbb{R}$ and let $p(\cdot,\theta,\eta)$ satisfy (5.80) with the following density with respect to Lebesgue measure on $\mathbb{R}^2$

$$p(x,y,\theta,z) = ze^{-zx}\ \theta ze^{-\theta zy}\ 1\{x>0,y>0\}.$$

In the functional form of this model we have a sequence of pairs (X_j,Y_j) of independent exponentially distributed random variables with hazard rates z_j and θz_j respectively and the problem is to estimate the ratio θ of the hazard rates. This model is considered in Lindsay (1985), who proposes to estimate θ by partial adaptation. The fully adaptive estimator sequence constructed here improves his proposal, and hence also the best invariant estimator defined as the solution to

$$\Sigma_{j=1}^{n} \frac{X_{nj}-\theta Y_{nj}}{X_{nj}+\theta Y_{nj}} = 0\ .$$

This is true for both the structural and functional form of the model.

For the i.i.d. mixture model, (5.1)-(5.2) is satisfied with as sufficient statistic $\psi(X,Y,\theta) = X+\theta Y$ which has density $g(s,\eta) = \int z^2 s\ e^{-zs}\ 1\{s>0\}\ d\eta(z)$ with respect to Lebesgue measure. No further conditions are needed to ensure applicability of Theorem 5.3 and Corollary 5.9 to this case. (Below we give a careful proof of a stronger result). By Theorem 5.10 or Corollary 5.11 we have efficiency of the resulting estimator T_n, under a variety of conditions on H, for instance efficiency in all (θ,η) when H is the set of all probability distributions on $\mathbb{R}$.

For the latter case Theorem 5.13 combined with Lemma 5.14 shows immediately that T_n is also efficient in the stronger sense that the linear space spanned by its efficient influence function $\tilde{\ell}(\cdot,\theta,\eta)$ is contained in the tangent cone, in all (θ,η) for which the support of η contains a limit point. In fact this result can be strengthened by an ad hoc argument. Set

$$\underset{\sim}{\ell}(x,y,\theta,z) = \theta^{-1} - zy.$$

It is proved below that (5.3) holds for every $(\theta,\eta) \in \Theta\times H$, with

$$\dot{\ell}(x,y,\theta,\eta) = p(x,y,\theta,\eta)^{-1} \int \underline{\dot{\ell}}(x,y,\theta,z)\underline{p}(x,y,\theta,z)\, d\eta(z).$$

Hence

$$\tilde{\dot{\ell}}(x,y,\theta,\eta) = p(x,y,\theta,\eta)^{-1} \int [\underline{\dot{\ell}}(x,y,\theta,z)$$

$$- E_\theta(\underline{\dot{\ell}}(X,Y,\theta,z)\,|X+\theta Y = x+\theta y)]\ \underline{p}(x,y,\theta,z)\, d\eta(z)$$

$$= p(x,y,\theta,\eta)^{-1} \int (2\theta)^{-1} z(x-\theta y)\ \underline{p}(x,y,\theta,z)\, d\eta(z).$$

For a subset B of $\mathbb{R}$ and $\alpha \in \mathbb{R}$ write αB for the set $\{\alpha b\colon b \in B\}$. Define measures η_t on $\mathbb{R}$ by $\eta_t(B) = \eta((1-\frac{1}{2}t\theta^{-1})^{-1}B)$ $(|t| < 2\theta)$. It can be checked (Lemma 5.19 is helpful) that

$$\int\int [t^{-1}(p^{\frac{1}{2}}(x,y,\theta+t,\eta_t)-p^{\frac{1}{2}}(x,y,\theta,\eta))-\tfrac{1}{2}\tilde{\dot{\ell}}(x,y,\theta,\eta)p^{\frac{1}{2}}(x,y,\theta,\eta)]^2 dxdy \to 0.$$

Hence if for $\eta \in H$, the scale family $\{\sigma^{-1}\eta(\sigma^{-1}s)\colon \sigma > 0\}$ also belongs to H, then a tangent cone $T(p(\cdot,\theta,\eta))$ may be chosen such that $\operatorname{lin}\{\tilde{\dot{\ell}}(\cdot,\theta,\eta)\}$ is contained in $T(p(\cdot,\theta,\eta))$, implying efficiency of T_n in the strong sense.

Next we turn to the non-i.i.d. model given by (5.29). Set

$$\bar{\eta}_n = n^{-1}\Sigma_{j=1}^n\ \eta_{nj}\ .$$

For applicability of Theorems 5.3 and 5.7, it suffices that no mass of $\bar{\eta}_n$ escapes to either zero or infinity.

THEOREM 5.17. *Suppose that the sequence of distributions* $\{\bar{\eta}_n\}$ *is tight in such a way that every limit point* η *has* $\eta(0,\infty) = 1$. *Then there exists an estimator sequence* $\{T_n\}$ *satisfying* (5.39). □

PROOF. It suffices to check (5.41)-(5.49) and (5.59)-(5.64) with the replacements of Corollary 5.9. We have

$$\int\int [t^{-1}(\underline{p}^{\frac{1}{2}}(x,y,\theta+t,z)-\underline{p}^{\frac{1}{2}}(x,y,\theta,z)) - \tfrac{1}{2}\underline{\dot{\ell}}(x,y,\theta,z)\underline{p}^{\frac{1}{2}}(x,y,\theta,z))]^2 dxdy$$

$$= \int_0^\infty [t^{-1}((\theta+t)^{\frac{1}{2}}e^{-\frac{1}{2}(\theta+t)y}-\theta^{\frac{1}{2}}e^{-\frac{1}{2}\theta y}) - \tfrac{1}{2}(\theta^{-1}-y)\theta^{\frac{1}{2}}e^{-\frac{1}{2}\theta y}]^2\, dy \to 0.$$

$$\iint \underline{\ell}^2(x,y,\theta,z)\ \underline{p}(x,y,\theta,z)\ dxdy = {}_0\int^{\infty} (\theta^{-1}-y)^2\theta e^{-\theta y}\ dy\ < \infty$$

$$\iint \underline{\ell}^2(x,y,\theta_n,z)\ 1\{|\underline{\ell}(x,y,\theta_n,z)|\geq\varepsilon\sqrt{n}\}\ \underline{p}(x,y,\theta_n,z)\ dxdy$$

$$= {}_0\int^{\infty}(\theta_n^{-1}-y)^2 1\{|\theta_n-y|\geq\varepsilon\sqrt{n}\}\ \theta_n e^{-\frac{1}{2}\theta_n y}\ dy \to 0.$$

Apply Lemmas 5.18-5.19 and 5.21 to see that (5.41)-(5.44) are satisfied with

$$\ell(x,y,\theta,\eta) = p^{-1}(x,y,\theta,\eta)\ \int \underline{\ell}(x,y,\theta,z)\ \underline{p}(x,y,\theta,z)\ d\eta(z).$$

Set

$$\underline{\tilde{\ell}}(x,y,\theta,z) = \underline{\ell}(x,y,\theta,z) - E_\theta(\underline{\ell}(X,Y,\theta,z)|\ X+\theta Y = x+\theta y\ )$$

$$= (2\theta)^{-1}\ z\ (x-\theta y).$$

Then uniformly in z

$$\iint [\underline{\tilde{\ell}}(x,y,\theta_n,z)\underline{p}^{\frac{1}{2}}(x,y,\theta_n,z) - \underline{\tilde{\ell}}(x,y,\theta,z)\underline{p}^{\frac{1}{2}}(x,y,\theta,z)]^2\ dxdy \to 0.$$

Apply Lemma 5.19 to see the validity of (5.45). Here

$$\tilde{\ell}(x,y,\theta,\eta) = p^{-1}(x,y,\theta,\eta)\ \int \underline{\tilde{\ell}}(x,y,\theta,z)\ \underline{p}(x,y,\theta,z)\ d\eta(z)$$

$$= (2\theta(x+\theta y))^{-1}(x-\theta y) + Q(x+\theta y,\eta)\ (2\theta)^{-1}(\theta y-x),$$

where

$$Q(s,\eta) = \int (s^{-1}-z)\underline{g}(s,z)\ d\eta(z)/\ g(s,\eta) = g'(s,\eta)\ /\ g(s,\eta).$$

Hence

$$\tilde{\ell}_n(x,y,\theta) = (2\theta(x+\theta y))^{-1}(x-\theta y) + Q(x+\theta y,\bar{\eta}_n)\ (2\theta)^{-1}(\theta y-x)\ . \tag{5.98}$$

Relation (5.47) has been verified. Next we note that $L_\theta(\ X-\theta Y|\ X+\theta Y = s\ )$ is the uniform distribution on $[-s,s] \subset \mathbb{R}$, implying (5.48). Furthermore

$$\beta^2(s,\theta) = E_\theta(\ \|\tilde{\psi}(X,Y,\theta)\|^2|\ X+\theta Y = s) = (12\theta^2)^{-1}s^2\ ,$$

so that (5.60) holds.

To show (5.46) and (5.61)-(5.64), assume without loss of generality

that $\bar{\eta}_n \to \eta$ weakly. Then for every s

$$g(s,\bar{\eta}_n) = \int z^2 s\, e^{-zs}\, 1\{s>0\}\, d\bar{\eta}_n(z) \to g(s,\eta) \tag{5.99}$$

$$g'(s,\bar{\eta}_n) = \int (s^{-1}-z)\, z^2 s\, e^{-zs} 1\{s>0\}\, d\bar{\eta}_n(z) \to g'(s,\eta). \tag{5.100}$$

By the Cauchy-Schwarz inequality

$$\begin{aligned} Q^2(s,\bar{\eta}_n)g(s,\bar{\eta}_n)\beta^2(s,\theta_n) &\le (12\theta_n)^{-2}\int [g'(s,z)/g^{\frac{1}{2}}(s,z)s]^2\, d\bar{\eta}_n(z) \\ &= (12\theta_n)^{-2} \int zf(sz)\, d\bar{\eta}_n(z), \end{aligned} \tag{5.101}$$

where

$$f(s) = (1-s)^2 s e^{-s}\, 1\{s>0\}.$$

Now for every s

$$\int zf(zs)\, d\bar{\eta}_n(z) \to \int zf(zs)\, d\eta(z)\ .$$

Furthermore by Fubini's theorem

$${}_0\int^\infty |\int zf(zs)\, d\bar{\eta}_n(z)|ds = {}_0\int^\infty f(s)\, ds = {}_0\int^\infty |\int zf(zs)\, d\eta(z)|ds.$$

By a convergence lemma (cf. Hewitt and Stromberg (1965), Th.13.47)

$${}_0\int^\infty |\int zf(zs)\, d\bar{\eta}_n(z) - \int zf(zs)\, d\eta(z)|\, ds \to 0\ . \tag{5.102}$$

From (5.101)-(5.102) we conclude that $\{Q^2(s,\bar{\eta}_n)g(s,\bar{\eta}_n)\beta^2(s,\theta_n)\colon n=1,2,..\}$ is equi-integrable with respect to Lebesgue measure. Combining this with (5.98)-(5.100) we obtain

$$\begin{aligned} &\iint \ell_n^2(x,y,\theta)\, \bar{p}_n(x,y,\theta)\, dxdy = \int E_\theta(\ell_n^2(X,Y,\theta)|X+\theta Y=s)\, g(s,\bar{\eta}_n)\, ds \\ &= \int (2\theta)^{-2}\, (s^{-1}-Q(s,\bar{\eta}_n))^2\, E_\theta((X-\theta Y)^2|X+\theta Y = s)\, g(s,\bar{\eta}_n)\, ds \\ &= \int (s^{-1}-Q(s,\bar{\eta}_n))^2\, \beta^2(s,\theta)\, g(s,\bar{\eta}_n)\, ds \\ &\to \int (s^{-1}-g'/g(s,\eta))^2\, \beta^2(s,\theta)\, g(s,\eta)\, ds\ > 0\ . \end{aligned}$$

This shows (5.46). By similar arguments

$$(5.103)\quad \int [g'/g^{\frac{1}{2}}(s,\bar{\eta}_n)\beta(s,\theta_n) - g'/g^{\frac{1}{2}}(s,\eta)\,\beta(s,\theta)]^2\,ds \to 0\ .$$

Relations (5.62)-(5.64) follow from (5.101) and (5.99) (cf Lemma 5.8).

Finally we consider (5.49). For $x,y > 0$ the function $\theta \to (x+\theta y)^{-1}(x-\theta y)$ decreases strictly on $[0,\infty)$, from 1 to -1. Hence we can define an estimator $\hat{\theta}_n$ as the unique solution to

$$(5.104)\quad \Sigma_{j=1}^n \frac{X_{nj}-\theta Y_{nj}}{X_{nj}+\theta Y_{nj}} = 0.$$

We have

$$(5.105)\quad E_{\theta_0\eta}\frac{X_{nj}-\theta Y_{nj}}{X_{nj}+\theta Y_{nj}} = 1 - 2\theta(\theta-\theta_0)^{-1} - 2\theta\theta_0(\theta-\theta_0)^{-2}\log(\theta_0/\theta)$$

$$= \theta_0/\theta - 1 + o(\theta_0/\theta - 1) \qquad \text{as } \theta_0/\theta \to 1.$$

Furthermore $P_{\theta\eta}(\, X_{nj}/Y_{nj} \le t\,) = (t+\theta)^{-1}t\ 1\{t > 0\}$, independent of η.

Standard arguments show that $\sqrt{n}(\hat{\theta}_n-\theta)$ is asymptotically normal. ∎

We note that we used tightness of $\{\bar{\eta}_n\}$ on $(0,\infty)$ only to show feasibility of estimating the location scores (cf. Section 5.3). Without the tightness it is still possible to estimate θ $\sqrt{n}$-consistently. Indeed the distribution of the estimator defined by (5.104) is independent of $\{(\eta_{n1},\ldots,\eta_{nn})\}$.

Under slightly stronger conditions on $\{\bar{\eta}_n\}$ van der Vaart (1987) carries through a one step-construction with an estimator for the efficient score function $\tilde{\ell}(x,y,\theta,\eta)$ which is based on an approximate maximum likelihood estimator for η. For fixed θ let $\hat{\eta}_n(\theta)$ satisfy

$$(5.106)\quad \Pi_{j=1}^n p(X_j,Y_j,\theta,\hat{\eta}_n(\theta)) = \sup_{\eta\in H} \Pi_{j=1}^n p(X_j,Y_j,\theta,\eta)\ ,$$

where H is the set of all probability distributions on $(0,\infty)$. Then $\tilde{\ell}(x,y,\theta,\hat{\eta}_n(\theta))$ may be used in the definition of T_n given in (5.55), instead of $\hat{\ell}_n(x,y,\theta)$ based on a kernel estimator as above. This is true for both

the incidental and the structural version of the model.

5.7.6. Generalizations of Example 5.7.5.

The class of distributions in 5.7.5 can be embedded in a larger set of distributions without making estimation of θ asymptotically more difficult. We discuss two cases.

5.7.6.1.

The treatment of the paired exponential model given in (5.7.5) depends on the exponential distribution only in so far, as this implies that $X+\theta Y$ is a sufficient statistic for the nuisance parameter and $L_\theta((X,Y) \mid X+\theta Y = s)$ is uniform on $\{(x,y) \in \mathbb{R}^2: x > 0, y > 0, x+\theta y = s\}$. The construction in 5.7.5 therefore applies equally well to the set of distributions of (X,Y) given by the density

$$(5.107) \quad p(x,y,\theta,g) = \theta\,(x+\theta y)^{-1}\, g(x+\theta y)\, 1\{x > 0,\ y > 0\}\ ,$$

where g is an unknown absolutely continuous density on $(0,\infty)$, satisfying $\int (1+sg'/g(s))^2\, g(s)\, ds < \infty$. Moreover estimation of θ in this larger model is not more difficult in terms of efficient estimation than in 5.7.5 at any point of the mixture model (where the information of the two models is comparable). This follows, since any score for the nuisance parameter in the present model is a function of the sufficient statistic $X+\theta Y$, so that the projection of the score for θ on the set of nuisance scores is still its conditional expectation given $X+\theta Y$.

Of course a similar generalization of the model is also possible in other examples. For instance in 5.7.4 the independence structure is of little importance. The following generalization is more remarkable.

5.7.6.2.

Let $(X_1,Y_1),\ldots,(X_n,Y_n)$ be i.i.d. with density

$$(5.108) \quad p(x,y,\theta,\eta) = \eta(x,\theta y)\,\theta$$

with respect to Lebesgue measure on $\{(x,y) \in \mathbb{R}^2: x > 0, y > 0\}$. Here η is symmetric in its coordinates: $\eta(s_1,s_2) = \eta(s_2,s_1)$. Moreover we assume that it is absolutely continuous (as in 5.7.2)with gradient $\nabla\eta = (\eta_1,\eta_2)$ such

that

$$(5.109)\qquad {}_0\int^{\infty}{}_0\int^{\infty} (1 + s_2\eta_2/\eta(s_1,s_2))^2\ \eta(s_1,s_2)\ ds_1 ds_2 < \infty\ .$$

In this model $\psi(X,Y,\theta) = (X\wedge\theta Y,\ X\vee\theta Y)$ is sufficient for η. It has density $g(s_1,s_2) = 2\eta(s_1,s_2)1\{s_1<s_2\}$ and

$$L_\theta((X,\theta Y)\mid \psi(X,Y,\theta) = s) = \tfrac{1}{2}\delta_{(s_1,s_2)} + \tfrac{1}{2}\delta_{(s_2,s_1)}\ .$$

We have

$$\ell(x,y,\theta,\eta) = \theta^{-1}\ (1 + \theta y\ \eta_2/\eta(x,\theta y))$$

and, using $\eta_1(s_1,s_2) = \eta_2(s_2,s_1)$,

$$(5.110)\qquad E_\theta(\ell(X,Y,\theta,\eta)\mid\psi(X,Y,\theta)=s) = \tfrac{1}{2}\theta^{-1}(2 + s_1\eta_1/\eta(s_1,s_2) + s_2\eta_2/\eta(s_1,s_2)).$$

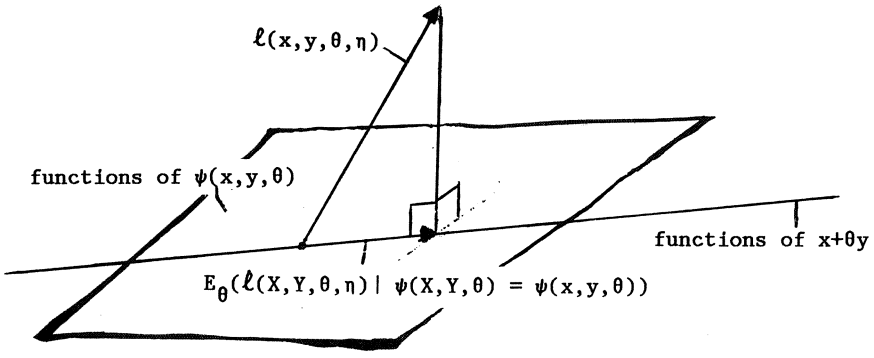
If η happens to have the special form of 5.7.5, i.e.
$\eta(x,\theta y)\theta = \int \theta z^2 e^{-(x+\theta y)z}\ d\gamma(z)$, then

$$\eta_1(\psi(x,y,\theta)) = \eta_2(\psi(x,y,\theta)) = -\int z^3 e^{-(x+\theta y)z}\ d\gamma(z) \qquad \text{a.e.}$$

Hence

$$E_\theta(\ell(X,Y,\theta,\eta)\mid \psi(X,Y,\theta)) = E_\theta(\ell(X,Y,\theta,\eta)\mid X+\theta Y)\ .$$

This proves that the efficient information for θ in the present model is not smaller, when evaluated at a mixture distribution as in 5.7.5: in this case the orthogonal projection of the score function for θ on the larger class of functions of $\psi(x,y,\theta)$ happens to be in the smaller class of functions of $x+\theta y$. We can draw the following picture:



It is hard to imagine, though, that for the model of 5.7.5 the construction based on the present sufficient statistic $\psi(X,Y,\theta)$ would not be inferior to the construction based on $X+\theta Y$, in some important sense; albeit not first order efficiency. Intuitively the best estimate for θ is obtained from using the best estimate of the nuisance parameter η (or the density $g(\cdot,\eta)$). If we base the latter on a sufficient statistic, this ought to be as minimal as possible.

5.7.7. Errors in variables

Let H be a set of probability distributions on $\mathbb{R}^k$ and let $p(\cdot,\mu,\Sigma,\eta)$ satisfy (5.80) with $p(\cdot,\mu,\Sigma,z)$ equal to the density of a $N_{k+1}((z',z'\mu)',\Sigma)$ distribution. Here $\mu \in \mathbb{R}^k$ and Σ is a positive definite matrix.

In the functional version of this model we have an unknown sequence of vectors z_j in $\mathbb{R}^k$ and i.i.d. $N_{k+1}(0,\Sigma)$ distributed error terms $(e_j,f_j)'$ and we observe the pairs

$$\begin{aligned} X_j &= z_j + e_j \\ Y_j &= z_j'\mu + f_j \end{aligned} \tag{5.11}$$

Hence we have a linear regression of Y_j on a vector z_j which is observed with error. This *errors in variables model* is also referred to as a *linear functional relationship*. (See Nussbaum (1984) and Bickel and Ritov (1987), and references cited there).

The i.i.d. mixture model case satisfies (5.1)-(5.2) with sufficient statistic (letting I denote the (k×k) identity matrix)

$$\psi(X,Y,\mu,\Sigma) = (I,\mu)\ \Sigma^{-1}\ (X,Y)'.$$

The distribution of $\psi(X,Y,\mu,\Sigma)$ equals

$$\int N_k((I,\mu)\ \Sigma^{-1}(z',z'\mu)',\ (I,\mu)\ \Sigma^{-1}(I,\mu)')\ d\eta(z).$$

We do not explore simple sufficient conditions on the set of distributions H or the parameter (μ,Σ), which would imply identifiability of the parameter and ensure applicability of Theorems 5.3 and 5.7. Constructions of efficient estimators in simpler versions of the model are given in Bickel and Ritov (1987) and van der Vaart (1986a).

Preliminary results show that estimation of the efficient score function by insertion of a (restricted) maximum likelihood estimator, as suggested at the end of 5.7.5, may be possible in this model too.

5.7.8. Neyman-Scott model

Let H be a set of probability distributions on $\mathbb{R}$, $\theta \in \Theta = (0,\infty)$ and let $p(\cdot,\theta,\eta)$ satisfy (5.80) with $p(\cdot,\theta,z)$ equal to the density of a $N_2((z,z)',\theta I)$ distribution. The model satisfies (5.1)-(5.2) with sufficient statistic $\psi(X,Y,\theta) = X+Y$ which has density

$$g(s,\theta,\eta) = \int (2\theta)^{-\frac{1}{2}}\ \phi((2\theta)^{-\frac{1}{2}}(s-2z))\ d\eta(z)$$

with respect to Lebesgue measure on $\mathbb{R}$.

In this case we have a decomposition (5.25) with $\tilde{\psi}$ equal to zero. Thus it is unnecessary to estimate a score function $\nabla g/g$. Theorem 5.3 gives the one step version of the *conditional maximum likelihood estimator* $\hat{\theta}_n$ which is obtained by maximizing

$$\Pi_{j=1}^{n}\ h(X_j,Y_j,\theta).$$

(cf. Andersen (1970)). This terminology becomes clear if $h(x,y,\theta)$ is interpreted as the conditional density of (X,Y) in (x,y) given that X+Y equals x+y. (This interpretation is not without conceptual difficulties, though. Perhaps *partial likelihood estimator* would be a preferable name). We have

$$\hat{\theta}_n = (2n)^{-1}\Sigma_{j=1}^{n} (X_j - Y_j)^2.$$

This example became famous because the *direct* maximum likelihood estimator $\hat{\theta}_{mn}$ for the functional version of the model, obtained by maximizing

$$\Pi_{j=1}^{n} \underline{p}(X_j, Y_j, \theta, z_j)$$

over all $(\theta, z_1, z_2, \ldots, z_n) \in \mathbb{R}^{n+1}$ is inconsistent. (Indeed $\hat{\theta}_{mn} = \frac{1}{2}\hat{\theta}_n \to \frac{1}{2}\theta$). In contrast to this, $\hat{\theta}_n$ is not only consistent, but also efficient in the i.i.d. mixture model, for instance if H is the set of all distributions on $\mathbb{R}$ (cf. Theorem 5.10 and Corollary 5.11. Here at (θ,η) for which the support of η contains a limit point, $\hat{\theta}_n$ is, up to first order asymptotic behaviour, the only LAM estimator for θ. Our efficiency statement does not rule out the possibility that at other (θ,η) the performance of $\hat{\theta}_n$ can be improved in the sense of the discussion of Section 2.6.(ii). We do not know of any results in this direction.

Mixture models of type (5.80), (5.82) with a sufficient statistic $\psi(X,\theta) = \psi(X)$ which does not depend on θ, are considered in Pfanzagl (1982), who shows efficiency of the conditional maximum likelihood estimator in these models.

5.8. TECHNICAL RESULTS

5.8.1. Some useful lemmas concerning mixture distributions.

For mixture models it may be difficult to check the differentiability condition imposed on the densities, or the continuity required of score functions. The lemmas in this section may be helpful.

For $n = 1,2\ldots$ and $j = 1,2,\ldots,n$, η_{nj} are probability distributions on measurable spaces $(\mathcal{Z}_{nj}, \mathcal{A}_{nj})$. Given $z \in \mathcal{Z}_{nj}$, $\underline{Q}_{nj}(\cdot,z)$ and $\underline{P}_{nj}(\cdot,z)$ are probability distributions on measurable spaces $(\mathcal{X}_{nj}, \mathcal{B}_{nj})$ with densities $\underline{q}_{nj}(\cdot,z)$ and $\underline{p}_{nj}(\cdot,z)$ with respect to σ-finite measures μ_{nj}. We suppose that $\underline{q}_{nj}(x,z)$ and $\underline{p}_{nj}(x,z)$ are measurable functions of (x,z) and are interested in the mixture distributions Q_{nj} and P_{nj} on $(\mathcal{X}_{nj}, \mathcal{B}_{nj})$ with densities with respect to μ_{nj} given by

$$q_{nj}(x) = \int \underline{q}_{nj}(x,z)\, d\eta_{nj}(z)$$

$$p_{nj}(x) = \int \underline{p}_{nj}(x,z)\, d\eta_{nj}(z).$$

We write $\underline{P}_{nj}\times\eta_{nj}$ for the probability distribution with density $\underline{p}_{nj}(x,z)$ with respect to $\mu_{nj}\times\eta_{nj}$ on $(\mathcal{X}_{nj}\times\mathcal{Z}_{nj},\mathcal{B}_{nj}\times\mathcal{A}_{nj})$ and define operators $A_{nj}\colon L_2(\underline{P}_{nj}\times\eta_{nj}) \to L_2(P_{nj})$ by

$$A_{nj}g(x) = \int g(x,z)\underline{p}_{nj}(x,z)\, d\eta_{nj}(z) \;/\; p_{nj}(x).$$

Note that if (X,Z) is a $(\mathcal{X}_{nj}\times\mathcal{Z}_{nj},\mathcal{B}_{nj}\times\mathcal{A}_{nj})$-valued random element with distribution $\underline{P}_{nj}\times\eta_{nj}$ then p_{nj} is the marginal density of X and

$$A_{nj}g(x) = E_{\underline{P}_{nj}\times\eta_{nj}}(g(X,Z) \mid X = x)\ .$$

Analogously we define $\underline{Q}_{nj}\times\eta_{nj}$ and $B_{nj}\colon L_2(\underline{Q}_{nj}\times\eta_{nj}) \to L_2(Q_{nj})$. Thus

$$B_{nj}g(x) = \int g(x,z)\underline{q}_{nj}(x,z)\, d\eta_{nj}(z) \;/\; q_{nj}(x)\ .$$

LEMMA 5.18. *Suppose that for a triangular array of functions* $g_{nj} \in L_2(\underline{P}_{nj}\times\eta_{nj})$

$$(5.112)\quad n^{-1}\Sigma_{j=1}^{n} \iint g_{nj}^2\, d\underline{P}_{nj}d\eta_{nj} = O(1)$$

$$(5.113)\quad n^{-1}\Sigma_{j=1}^{n} \iint g_{nj}^2\, 1\{|g_{nj}| \geq \varepsilon\sqrt{n}\}\, d\underline{P}_{nj}d\eta_{nj} \to 0 \qquad \textit{for every } \varepsilon > 0.$$

$$(5.114)\quad n^{-1}\Sigma_{j=1}^{n} \iint [\sqrt{n}(q_{nj}^{\frac{1}{2}} - p_{nj}^{\frac{1}{2}}) - \tfrac{1}{2}g_{nj}p_{nj}^{\frac{1}{2}}]^2\, d\mu_{nj}d\eta_{nj} \to 0.$$

Then

$$n^{-1}\Sigma_{j=1}^{n} \int (A_{nj}g_{nj})^2\, dP_{nj} = O(1)$$

$$n^{-1}\Sigma_{j=1}^{n} \int (A_{nj}g_{nj})^2\, 1\{|A_{nj}g_{nj}| \geq \varepsilon\sqrt{n}\}\, dP_{nj} \to 0 \qquad \textit{for every } \varepsilon > 0$$

$$n^{-1}\Sigma_{j=1}^{n} \int [\sqrt{n}(q_{nj}^{\frac{1}{2}} - p_{nj}^{\frac{1}{2}}) - \tfrac{1}{2}(A_{nj}g_{nj})p_{nj}^{\frac{1}{2}}]^2\, d\mu_{nj} \to 0. \ \square$$

PROOF. This lemma is a special case of Popositions A.11-A.12 (appendix). ■

LEMMA 5.19. *Let* (5.112)-(5.114) *hold and suppose that for triangular arrays* $(k_{n1},\dots,k_{nn}) \in \otimes_{j=1}^{n} L_{2*}(\underline{Q}_{nj}\times\eta_{nj})$ *and* $(\ell_{n1},\dots,\ell_{nn}) \in \otimes_{j=1}^{n} L_{2*}(\underline{P}_{nj}\times\eta_{nj})$

$$n^{-1}\Sigma_{j=1}^{n} \int\int [k_{nj}\underline{q}_{nj}^{\frac{1}{2}} - \ell_{nj}\underline{p}_{nj}^{\frac{1}{2}}]^2 \, d\mu_{nj}d\eta_{nj} \to 0$$

$$n^{-1}\Sigma_{j=1}^{n} \int\int \ell_{nj}^2 \, d\underline{P}_{nj}d\eta_{nj} = O(1)$$

$$n^{-1}\Sigma_{j=1}^{n} \int\int k_{nj}^2 \, 1\{|k_{nj}| \geq \varepsilon\sqrt{n}\} \, d\underline{Q}_{nj}d\eta_{nj} \to 0 \qquad \textit{for every } \varepsilon > 0.$$

Then

$$(5.115)\quad n^{-1}\Sigma_{j=1}^{n} \int [(B_{nj}k_{nj})q_{nj}^{\frac{1}{2}} - (A_{nj}\ell_{nj})p_{nj}^{\frac{1}{2}}]^2 \, d\mu_{nj} \to 0. \ \square$$

PROOF. This is a special case of Proposition A.13 (appendix). ■

5.8.2. Technical results.

LEMMA 5.20. *Let* $(\Omega_{nj},\mathcal{A}_{nj})$ *be measurable spaces,* $Y_{nj}\colon (\Omega_{nj},\mathcal{A}_{nj}) \to \mathbb{R}$ *be measurable, and* $\mathcal{A}'_{nj} \subset \mathcal{A}_{nj}$ *be sub* σ*-algebra's* $(n=1,2,\dots,\ j=1,2,\dots,n)$. *Suppose*

$$(5.116)\quad n^{-1}\Sigma_{j=1}^{n} EY_{nj}^2 = O(1)$$

and

$$(5.117)\quad n^{-1}\Sigma_{j=1}^{n} E\, Y_{nj}^2 1\{|Y_{nj}| \geq \varepsilon\sqrt{n}\} \to 0 \qquad \textit{for every } \varepsilon > 0.$$

Then

$$n^{-1}\Sigma_{j=1}^{n} E\,(Y_{nj}-E(Y_{nj}|\mathcal{A}'_{nj}))^2 = O(1)$$

and

$$n^{-1}\Sigma_{j=1}^{n} E\,(Y_{nj}-E(Y_{nj}|\mathcal{A}'_{nj}))^2 \, 1\{|Y_{nj}-E(Y_{nj}|\mathcal{A}'_{nj})| \geq \varepsilon\sqrt{n}\} \to 0$$

for every $\varepsilon > 0$. $\square$

PROOF. The first assertion is obvious. To prove the second, for given $\varepsilon > 0$ set

$$C_{nj} = \{\ |E(Y_{nj}|\mathcal{A}'_{nj})| \geq \varepsilon\sqrt{n}\ \}$$

$$D_{nj} = \{\ |Y_{nj}-E(Y_{nj}|\mathcal{A}'_{nj})| \geq \varepsilon\sqrt{n}\ \}.$$

Then by (5.117) for any $\varepsilon_0 > 0$

$$n^{-1}\Sigma_{j=1}^n E\,(E(Y_{nj}|A'_{nj}))^2\,1\{C_{nj}\} \le n^{-1}\Sigma_{j=1}^n E\,E(Y_{nj}^2|A'_{nj})\,1\{C_{nj}\}$$

$$= n^{-1}\Sigma_{j=1}^n E\,Y_{nj}^2 1\{C_{nj}\} \le n^{-1}\Sigma_{j=1}^n \varepsilon_0^2 n\,P(C_{nj}) + o(1)$$

$$\le \varepsilon_0^2\varepsilon^{-2}n^{-1}\Sigma_{j=1}^n E\,(E(Y_{nj}|A'_{nj}))^2 + o(1) \le \varepsilon_0^2\varepsilon^{-2}n^{-1}\Sigma_{j=1}^n EY_{nj}^2 + o(1).$$

Hence for any $\varepsilon > 0$

$$(5.118)\quad n^{-1}\Sigma_{j=1}^n E\,(E(Y_{nj}|A'_{nj}))^2\,1\{|E(Y_{nj}|A'_{nj})| \ge \varepsilon\sqrt{n}\} \to 0.$$

Next by (5.117)-(5.118)

$$n^{-1}\Sigma_{j=1}^n E\,(Y_{nj}-E(Y_{nj}|A'_{nj}))^2\,1\{D_{nj}\}$$

$$\le 2n^{-1}\Sigma_{j=1}^n E\,Y_{nj}^2 1\{D_{nj}\} + 2n^{-1}\Sigma_{j=1}^n E\,(E(Y_{nj}|A'_{nj}))^2\,1\{D_{nj}\}$$

$$\le 4\varepsilon_0^2\Sigma_{j=1}^n P(D_{nj}) + o(1) \le 4\varepsilon_0^2\varepsilon^{-2}n^{-1}\Sigma_{j=1}^n EY_{nj}^2 + o(1). \blacksquare$$

LEMMA 5.21. *Let* $\{P_t\}$ $(t > 0)$ *and* P *be probability measures on a measurable space* $(X,\mathcal{B})$ *with*

$$(5.119)\quad \int [t^{-1}((dP_t)^{\frac12}-(dP)^{\frac12}) - \tfrac12 g\,(dP)^{\frac12}]^2 \to 0 \qquad (t\downarrow 0).$$

Let $f \in L_2(P)$ *and satisfy*

$$(5.120)\quad \limsup_{t\downarrow 0} \int f^2\,dP_t < \infty.$$

Then

$$t^{-1}\Big(\int f\,dP_t - \int f\,dP\Big) \to \int fg\,dP. \ \square$$

PROOF. Since we may restrict ourselves to countable sequences $\{P_{tn}\}$ $(t_n\downarrow 0)$ it is no loss of generality to assume that P_t and P have densities p_t and p with respect to a σ-finite measure μ. We have

$$|\int f\,[t^{-1}(p_t-p) - g\,p]\,d\mu|$$

(5.121)

$$\leq \int |\,f[p_t^{\frac12}+p^{\frac12}]\,[t^{-1}(p_t^{\frac12}-p^{\frac12}) - \tfrac12 g p^{\frac12}]\,|\,d\mu + \tfrac12\int |fgp^{\frac12}[p_t^{\frac12}-p^{\frac12}]|\,d\mu.$$

Here the first integral is smaller than

$$[2\int |f^2[p_t+p]|\,d\mu \int [t^{-1}(p_t^{\frac12}-p^{\frac12}) - \tfrac12 g p^{\frac12}]^2\,d\mu\,]^{\frac12}$$

and the second can be bounded by

$$[\,t^{-1}\int g^2 dP \int (p_t^{\frac12}-p^{\frac12})^2 d\mu\,]^{\frac12} + [\int f^2 (p_t^{\frac12}-p^{\frac12})^2 d\mu \int g^2 1\{|f|>t^{-\frac12}\} dP\,]^{\frac12}.\blacksquare$$

5.8.3. Proofs

PROOF OF LEMMA 5.4. From (5.42) and the Cauchy-Schwarz inequality we infer that for all $h \in \mathbb{R}^k$

$$\int (h'\dot{\ell}_n(x,\theta))^2\,\bar{p}_n(x,\theta)\,d\mu(x) = O(1).$$

But then also

$$\int (h'\dot{\ell}_n(x,\theta))^2\,\bar{p}_n(x,\theta)\,d\mu(x) = O(1).$$

By (5.45)

(5.122) $$\int (h'\dot{\ell}_n(x,\theta_n))^2\,\bar{p}_n(x,\theta_n)\,d\mu(x) = O(1).$$

For c and d in $\mathbb{R}^n$ we have

$$|\langle c,d\rangle - \|c\|\|d\|\,| = |\langle c,d-c\rangle + \langle c,c\rangle^{\frac12}(\langle c,c\rangle^{\frac12}-\langle d,d\rangle^{\frac12})| \leq 2\|c\|\|d-c\|,$$

Applying this and the Cauchy-Schwarz inequality we see

$$|n^{-1}\Sigma_{j=1}^n \int \|\dot{\ell}_n(x,\theta_n)p^{\frac12}(x,\theta_n,\eta_{nj}) - \dot{\ell}_n(x,\theta)p^{\frac12}(x,\theta,\eta_{nj})\|^2\,d\mu(x)$$

$$- \int \|\dot{\ell}_n(x,\theta_n)\bar{p}_n^{\frac12}(x,\theta_n) - \dot{\ell}_n(x,\theta)\bar{p}_n^{\frac12}(x,\theta)\|^2\,d\mu(x)|$$

$$= 2 \left| \int \dot{\ell}_n(x,\theta_n)'\dot{\ell}_n(x,\theta) \left[n^{-1}\Sigma_{j=1}^n p^{\frac{1}{2}}(x,\theta_n,\eta_{nj})p^{\frac{1}{2}}(x,\theta,\eta_{nj}) - \bar{p}_n^{\frac{1}{2}}(x,\theta_n)\bar{p}_n^{\frac{1}{2}}(x,\theta)\right] d\mu(x)\right|$$

$$\leq 4\int \|\dot{\ell}_n(x,\theta_n)\varepsilon\sqrt{n}\ \bar{p}_n^{\frac{1}{2}}(x,\theta_n) \left[n^{-1}\Sigma_{j=1}^n (p^{\frac{1}{2}}(x,\theta_n,\eta_{nj})-p^{\frac{1}{2}}(x,\theta,\eta_{nj}))^2\right]^{\frac{1}{2}} d\mu(x)$$

$$+ \int \|\dot{\ell}_n(x,\theta_n)\|\|\dot{\ell}_n(x,\theta)\| 1\{\|\dot{\ell}_n(x,\theta)\|>\varepsilon\sqrt{n}\}\ 4\bar{p}_n^{\frac{1}{2}}(x,\theta_n)\bar{p}_n^{\frac{1}{2}}(x,\theta)\ d\mu(x)$$

$$\leq \varepsilon O(1) + o(1),$$

by (5.122), (5.41)-(5.42) and (5.44). Combination with (5.45) shows

$$(5.123) \quad n^{-1}\Sigma_{j=1}^n \int \left[h'\dot{\ell}_n(x,\theta_n)p^{\frac{1}{2}}(x,\theta_n,\eta_{nj}) - h'\dot{\ell}_n(x,\theta)p^{\frac{1}{2}}(x,\theta,\eta_{nj})\right]^2 d\mu(x) \to 0.$$

Finally, in view of (5.41)-(5.44), (5.122)-(5.123) we can apply Proposition A.10 to conclude that

$$n^{-\frac{1}{2}} \Sigma_{j=1}^n \left(h'\dot{\ell}_n(X_{nj},\theta_n) - h'\dot{\ell}_n(X_{nj},\theta)\right)$$

$$+ n^{-1}\Sigma_{j=1}^n \int h'\dot{\ell}_n(x,\theta)\ \sqrt{n}(\theta_n-\theta)'\dot{\ell}(x,\theta,\eta_{nj})\ p(x,\theta,\eta_{nj})\ d\mu(x) \to 0,$$

in $P_{\theta_n\eta_{n1}\cdots\eta_{nn}}$ - probability. This is equivalent to the assertion of the lemma (cf. (5.40), (5.31) and (5.32)). ■

PROOF OF THEOREM 5.7. We use the following notation

$$\bar{g}_{n\sigma}(s) = \int \bar{g}_n(s-\sigma y)\ \omega(y)\ dy$$

$$\nabla\bar{g}_{n\sigma}(s) = \int \nabla\bar{g}_n(s-\sigma y)\ \omega(y)\ dy$$

$$A_n = \{s \in S:\ \alpha_n < \beta_n(s) < b_n,\ \sup\{|\beta_n(s+\sigma_n y)-\beta_n(s)|:\ \|y\|\leq 1\} < \gamma_n,$$

$$\|s-\partial S\| > \varepsilon_n\}.$$

Furthermore we omit the index n whenever convenient.

As $\sigma\varepsilon^{-1} \to 0$ we have for sufficiently large n that $s \in A_n$ implies that $s-\sigma_n y \in S$ for all y in the support of ω. Hence for $s \in A_n$ and $n \to \infty$

$$(5.124)\qquad \int |\bar g_n(s-\sigma y)-\bar g_n(s)|\,\omega(y)\,dy = \sigma\int \Big|\int_0^1 \nabla\bar g_n(s-\sigma uy)'y\,du\Big|\,\omega(y)\,dy.$$

For $s \in A_n$ and $\|u\| \le 1$

$$(5.125)\qquad |\beta_n(s-\sigma u)| \ge |\beta_n(s)| - \sup\{|\beta_n(s)-\beta_n(s-\sigma y)|: \|y\| \le 1\} \ge \alpha_n-\gamma_n.$$

Hence for $n \to \infty$

$$\begin{aligned} &A_n\int |\bar g_{n\sigma}(s)-\bar g_n(s)|\,ds \le A_n\iint |\bar g_n(s-\sigma y)-\bar g_n(s)|\,\omega(y)\,dyds \\ &(5.126)\qquad \le \sigma(\alpha_n-\gamma_n)^{-1}\int_{A_n}\int_0^1 \|\nabla\bar g_n\beta_n(s-\sigma uy)\|\,duds\,\|y\|\omega(y)\,dy \\ &\le \sigma(\alpha-\gamma)^{-1}\{\int\|\nabla\bar g_n/\bar g_n^{\frac12}\beta_n(s)\|^2 ds\int\bar g_n(s)ds\}^{\frac12}\int\|y\|\omega(y)dy = O(\sigma\alpha^{-1}) \to 0. \end{aligned}$$

Next we show through a long sequence of inequalities that

$$(5.127)\qquad A_n\int \Big\| \nabla\bar g_n/\bar g_n^{\frac12}\beta_n(s) - \nabla\bar g_{n\sigma}\beta_n\bar g_{n\sigma}^{\frac12}/(\bar g_{n\sigma}(s)+\delta)(s)\Big\|^2 \to 0.$$

First, by (5.61)

$$\begin{aligned} &\int \Big\| \nabla\bar g_n/\bar g_n^{\frac12}\beta_n(s) - \int\nabla\bar g_n/\bar g_n^{\frac12}\beta_n(s-\sigma y)\omega(y)dy\Big\|^2 ds \\ &(5.128)\qquad \le \int\omega(y)dy\,\sup\Big\{\int\|\nabla\bar g_n/\bar g_n^{\frac12}\beta_n(s) - \nabla\bar g_n/\bar g_n^{\frac12}\beta_n(s-\sigma y)\|^2 ds:\ \|y\|\le1\Big\} \to 0. \end{aligned}$$

Second, setting $F_n = \{s:\ \|\nabla\bar g_n/\bar g_n\beta_n(s)\| \le M_n\}$ where $M_n \to \infty$ at a rate to be fixed later,

$$\begin{aligned} &A_n\int \Big\| \int\nabla\bar g_n\beta_n(s-\sigma y)\omega(y)dy/\bar g_{n\sigma}^{\frac12}(s) - \int\nabla\bar g_n/\bar g_n^{\frac12}\beta_n(s-\sigma y)\omega(y)dy\Big\|^2 ds \\ &\le 3\int\Big[\Big\|\int\nabla\bar g_n\beta_n 1_{F_n^c}(s-\sigma y)\omega(y)dy\Big\|^2/\bar g_{n\sigma}(s)+\Big\|\int\nabla\bar g_n/\bar g_n^{\frac12}\beta_n 1_{F_n^c}(s-\sigma y)\omega(y)dy\Big\|^2\Big]ds \\ &\qquad + 3A_n\int\Big\|\int\nabla\bar g_n/\bar g_n^{\frac12}\beta_n 1_{F_n}(s-\sigma y)[\bar g_n^{\frac12}(s-\sigma y)/\bar g_{n\sigma}^{\frac12}(s)-1]\omega(y)dy\Big\|^2 ds \\ &(5.129)\qquad \le 6\int\|\nabla\bar g_n/\bar g_n^{\frac12}\beta_n 1_{F_n^c}(s)\|^2 ds + 3A_n\int M_n^2\bar g_{n\sigma}(s)\int[\bar g_n^{\frac12}(s-\sigma y)/\bar g_{n\sigma}^{\frac12}(s)-1]^2\omega(y)dyds \end{aligned}$$

$$\leq o(1) + 3M_n^2 A_n \int\int |\bar{g}_n(s-\sigma y) - \bar{g}_{n\sigma}(s)| \, \omega(y)dyds \qquad \text{(cf.(5.63))}$$

$$\leq o(1) + 3M_n^2 A_n \int [\int |\bar{g}_n(s-\sigma y) - \bar{g}_n(s)| \, \omega(y)dy + |\bar{g}_n(s) - \bar{g}_{n\sigma}(s)|] \, ds \to 0$$

if $M_n^2 \sigma_n \alpha_n^{-1} \to 0$ (cf. (5.126)). Third for suitably chosen $M_n \uparrow \infty$, $p_n \uparrow \infty$ and $\eta_n \downarrow 0$

$$\int \| \int \nabla \bar{g}_n \beta_n(s-\sigma y)\omega(y)dy / \bar{g}_{n\sigma}^{\frac{1}{2}}(s) - \int \nabla \bar{g}_n \beta_n(s-\sigma y)\omega(y)dy \; \bar{g}_{n\sigma}^{\frac{1}{2}}/(\bar{g}_{n\sigma}+\delta)(s) \|^2 \, ds$$

$$\leq \int \int \|\nabla \bar{g}_n / \bar{g}_n^{\frac{1}{2}} \beta_n(s-\sigma y)\|^2 \omega(y)dy \; \delta^2 (\bar{g}_{n\sigma}(s)+\delta)^{-2} \, ds$$

$$\text{(5.130)} \quad \leq \int \int \|\nabla \bar{g}_n / \bar{g}_n^{\frac{1}{2}} \beta_n(s-\sigma y)\|^2 \omega(y)dy \, 1_{\{\bar{g}_{n\sigma}(s) \leq \eta_n\}} \, ds + \delta^2 \eta^{-2} \int \|\nabla \bar{g}_n / \bar{g}_n^{\frac{1}{2}} \beta_n(s)\|^2 ds$$

$$\leq \int\int \|\nabla \bar{g}_n / \bar{g}_n^{\frac{1}{2}} \beta_n(s-\sigma y)\|^2 \omega(y)dy 1_{\{|s| \geq p_n\}} ds + \int\int \|\nabla \bar{g}_n / \bar{g}_n^{\frac{1}{2}} \beta_n 1_{F_n^c}(s-\sigma y)\|^2 \omega(y)dyds$$

$$\int M_n^2 \bar{g}_{n\sigma}(s) \, 1_{\{\bar{g}_{n\sigma}(s) \leq \eta_n, \; |s| \leq p_n\}} \, ds + O(\delta^2 \eta^{-2})$$

$$\leq o(1) + 2p_n M_n^2 \eta_n + O(\delta^2 \eta^{-2}) \to 0.$$

Finally

$$A_n \int \| \nabla \bar{g}_n \beta_n(s) - \int \nabla \bar{g}_n \beta_n(s-\sigma y)\omega(y)dy \|^2 \, \bar{g}_{n\sigma}/(\bar{g}_{n\sigma}+\delta)^2(s) \, ds$$

$$\text{(5.131)} \quad \leq \gamma_n^2 A_n \int [\int \|\nabla \bar{g}_n(s-\sigma y)\| \omega(y) \, dy]^2 \, \bar{g}_{n\sigma}^{-1}(s) \, ds$$

$$\leq \gamma_n^2 (\alpha_n - \gamma_n)^{-2} \int\int \|\nabla \bar{g}_n / \bar{g}_n^{\frac{1}{2}} \beta_n(s-\sigma y)\|^2 \omega(y) \, dyds = O(\gamma_n^2 \alpha_n^{-2}) \to 0.$$

(cf. (5.125)). Relation (5.127) follows from (5.128)-(5.131).

Let P_n and E_n denote probability and expectation under $g_{n1}, \ldots, g_{nn}$. There exist constants M_1 and M_2 depending on ω, such that for sufficiently large n and $s \in A_n$

$$E_n \| \nabla \hat{g}_{n\sigma}^j(s) - \nabla \bar{g}_{n\sigma}(s) \|^2$$

$$= E_n \| n^{-1} \Sigma_{i \neq j} \sigma^{-m-1} \{\nabla \omega(\sigma^{-1}(s-V_{ni})) - E_n \nabla \omega(\sigma^{-1}(s-V_{ni}))\} \|^2$$

$$+ n^{-2} \sigma^{-2m-2} \| E_n \nabla \omega(\sigma^{-1}(s-V_{nj})) \|^2$$

$$\leq n^{-1}\sigma^{-2m-2}M_2$$

and, analogously,

$$E_n \mid \hat{g}^j_{n\sigma}(s) - \bar{g}_{n\sigma}(s) \mid^2 \leq n^{-1}\sigma^{-2m}M_1.$$

Hence for sufficiently large n and $s \in A_n$

$$E_n \| \nabla\hat{g}^j_{n\sigma}/(\hat{g}^j_{n\sigma}+\delta)(s) - \nabla\bar{g}_{n\sigma}/(\bar{g}_{n\sigma}+\delta)(s) \|^2$$

$$\leq 2E_n[\|(\bar{g}_{n\sigma}-\hat{g}^j_{n\sigma})/(\hat{g}^j_{n\sigma}+\delta)\nabla\bar{g}_{n\sigma}/(\bar{g}_{n\sigma}+\delta)(s)\|^2 + \|(\nabla\hat{g}^j_{n\sigma}-\nabla\bar{g}_{n\sigma})/((\hat{g}^j_{n\sigma}+\delta)(s)\|^2] \tag{5.132}$$

$$\leq 2\delta^{-2}n^{-1}\sigma^{-2m}M_1 \| \nabla\bar{g}_{n\sigma}/(\bar{g}_{n\sigma}+\delta)(s) \|^2 + 2\delta^{-2}n^{-1}\sigma^{-2m-2}M_2 .$$

Consequently

$$C_n \int \max_{j=1..n} E_n \| \hat{Q}^j_n \bar{g}_n^{-\frac{1}{2}}(s) - \nabla\bar{g}_{n\sigma}\bar{g}_{n\sigma}^{-\frac{1}{2}}/(\bar{g}_{n\sigma}+\delta)(s) \|^2 \beta_n^2(s)\, ds$$

$$\leq 2c_n^2 \int_{A_n} [\bar{g}_n^{-\frac{1}{2}}(s)-\bar{g}_{n\sigma}^{-\frac{1}{2}}(s)]^2\, ds + b_n^2\, 4\delta^{-2}n^{-1}\sigma^{-2m-2}M_2 \tag{5.133}$$

$$+ 4\delta^{-2}n^{-1}\sigma^{-2m}M_1 \int_{A_n} \|\nabla\bar{g}_{n\sigma}/(\bar{g}_{n\sigma}+\delta)(s)\|^2 \beta_n^2(s)\, \bar{g}_{n\sigma}(s)\, ds \to 0,$$

by (5.126), (5.127) and (5.62).

By (6.60) and λ-a.a. s there exist constants c_s such that

$$\gamma_n^{-1} \sup \{ |\beta_n(s+\sigma y)-\beta_n(s)| : \|y\|\leq 1 \} \leq \gamma_n^{-1}k(\sigma_n)c_s \to 0. \tag{5.134}$$

We conclude that the function on the left-hand side of (5.134) converges λ-a.e. -and hence in λ-measure- to zero (cf. Bauer(1981), Th. 2.11.6). As a consequence of this, (5.62) and (5.64) we have

$$\int_{A_n^c} \| \nabla\bar{g}_n\beta_n/\bar{g}_n^{\frac{1}{2}}(s) \|^2\, ds \to 0. \tag{5.135}$$

Let $H_n = \{s \in S: \bar{g}_{n\sigma}(s) > \eta_n\}$ and $G_n = \{s \in S: \|\nabla\bar{g}_{n\sigma}\beta_n/(\bar{g}_{n\sigma}+\delta)\| \geq \frac{1}{2}c_n\}$. By (5.128)

$$(5.136)\qquad A_n \cap H_n^c \int \|\nabla \bar{g}_n / \bar{g}_n^{\frac{1}{2}} \beta_n(s)\|^2 ds = A_n \cap H_n^c \int \|\int \nabla \bar{g}_n / \bar{g}_n^{\frac{1}{2}} \beta_n(s-\sigma y) \omega(y) dy\|^2 ds + o(1) \to 0$$

(cf. the last part of (5.130)). Next

$$\lambda(G_n \cap H_n \cap A_n) \le \lambda(\{s \in A_n : \|\nabla \bar{g}_{n\sigma} \beta_n \bar{g}_{n\sigma}^{\frac{1}{2}} / (\bar{g}_{n\sigma} + \delta)\| \ge \tfrac{1}{2} c_n \eta_n^{\frac{1}{2}}\})$$

$$\le 4 c_n^{-2} \eta_n^{-1} \int_{A_n} \|\nabla \bar{g}_{n\sigma} \beta_n \bar{g}_{n\sigma}^{\frac{1}{2}} / (\bar{g}_{n\sigma} + \delta)\|^2 \, ds = O(c_n^{-2} \eta_n^{-1})$$

by (5.127) and (5.62). For $\eta_n \downarrow 0$ sufficiently slowly we see by (5.136), (5.137) and (5.62)

$$(5.138)\qquad A_n \cap G_n \int \|\nabla \bar{g}_n / \bar{g}_n^{\frac{1}{2}} \beta_n(s)\|^2 ds = A_n \cap G_n \cap H_n \int \|\nabla \bar{g}_n / \bar{g}_n^{\frac{1}{2}} \beta_n(s)\|^2 ds + o(1) \to 0.$$

Finally by (5.136) and (5.138)

$$E_n \, A_n - C_n \int \|\nabla \bar{g}_n / \bar{g}_n^{\frac{1}{2}} \beta_n(s)\|^2 \, ds$$

$$= \int_{A_n} \|\nabla \bar{g}_n / \bar{g}_n^{\frac{1}{2}} \beta_n(s)\|^2 \, P_n(\|\nabla \hat{g}_{n\sigma}^j \beta_n(s)\| \ge c_n (\hat{g}_{n\sigma}^j(s) + \delta)) \, ds$$

$$= o(1) + \int_{A_n \cap H_n} \|\nabla \bar{g}_n / \bar{g}_n^{\frac{1}{2}} \beta_n(s)\|^2$$

$$\cdot \, P_n(\beta_n \|\nabla \hat{g}_{n\sigma}^j / \hat{g}_{n\sigma}^j + \delta) - \nabla \bar{g}_{n\sigma} / (\bar{g}_{n\sigma} + \delta)\|(s) \ge \tfrac{1}{2} c_n) \, ds.$$

We show that this converges to zero. By (5.132) $1_{\{A_n \cap H_n\}}$ times the probability in the right hand side of (5.139) is dominated, uniformly in s, by

$$(5.140)\qquad 1_{A_n}(s) \, [8 c_n^{-2} \, \delta^{-2} n^{-1} \sigma^{-2m} M_1 \, \eta_n^{-1} \|\nabla \bar{g}_{n\sigma} \beta_n \bar{g}_{n\sigma}^{\frac{1}{2}} / (\bar{g}_{n\sigma} + \delta)\|^2 + o(1)] \, .$$

From (5.140) we conclude that the integrand in (5.139) converges to zero in λ-measure if $\eta_n \downarrow 0$ at a sufficiently slow rate. By (5.62) it also converges to zero in mean.

Together with (5.135) this shows

(5.141) $$C_n^c \int \|\nabla \bar{g}_n / \bar{g}_n^{\frac{1}{2}} \beta_n\|^2 \, ds \to 0.$$

Combination of (5.127), (5.133) and (5.141) yields the theorem. ■

PROOF OF LEMMA 5.8. (5.61) and (5.62) are immediate consequences of (5.65)-(5.66). Let $A_n = \{s: \|\nabla \bar{g}_n / \bar{g}_n^{\frac{1}{2}} \beta_n\| \geq b_n\}$, where $b_n \to \infty$. Then

$$A_n \int \bar{g}_n(s) \, ds \leq b_n^{-2} \int \|\nabla \bar{g}_n / \bar{g}_n^{\frac{1}{2}} \beta_n(s)\|^2 \, ds \to 0.$$

Combination with (5.65), (5.66) and (5.67) yields

$$A_n \int \|\nabla \bar{g}_n / \bar{g}_n^{\frac{1}{2}} \beta_n(s)\|^2 \, ds = A_n \int \|\nabla g / g^{\frac{1}{2}} \beta(s)\|^2 \, ds + o(1) \to 0.$$

We finally prove (5.64). For any $\varepsilon > 0$ and sufficiently large n

$$\{s: \beta_n(s) \not\in [h_n, b_n]\} \subset \{s: \beta(s) \not\in [h_n + \varepsilon, b_n - \varepsilon]\} \cup \{s: |b_n(s) - \beta(s)| \geq \varepsilon\}$$

$$\subset \{s: \beta(s) \not\in [2\varepsilon, \varepsilon^{-1}]\} \cup \{s: |\beta_n(s) - \beta(s)| \geq \varepsilon\}.$$

Combination with (5.65), (5.66) and (5.68) shows

$$\int \|\nabla \bar{g}_n / \bar{g}_n^{\frac{1}{2}} \beta_n(s)\|^2 \, 1_{\{s: \beta_n(s) \not\in [h_n, b_n]\}} \, ds$$

$$\int \|\nabla \bar{g} / \bar{g}^{\frac{1}{2}} \beta(s)\|^2 \, 1_{\{s: \beta(s) \not\in [2\varepsilon, \varepsilon^{-1}]\}} \, ds + o(1).$$

But this converges to zero by dominated convergence as $\varepsilon \downarrow 0$. ■

APPENDIX

CONTIGUITY AND DIFFERENTIABILITY IN QUADRATIC MEAN

In this appendix we derive three types of results. In Section A.1 we discuss some basic facts concerning contiguity. Next in Section A.2 we discuss local asymptotic normality and differentiability of scores. Finally in Section A.3 we show that the properties of local asymptotic normality and differentiability of scores are retained under measurable transformations of the observations.

A.1. CONTIGUITY

At many points in this manuscript we have referred to *contiguity* or more deviously to *contiguity arguments.* In this section we present a self-contained discussion of contiguity. All of this is well-known, the idea of contiguity going back to LeCam (1960,1969). Below we follow the treatment given in Helmers et al. (1976) (in dutch!), which itself is based on Roussas (1968) and Witting and Nölle (1970). Also see Oosterhoff and van Zwet (1979)). There is a slight addition, because we formulate convergence in distribution in terms of Pollard (1984) (cf. Chapter 4).

Let P_n and Q_n be probability measures on measurable spaces $(\mathcal{X}_n,\mathcal{B}_n)$ $(n=1,2\ldots)$ with densities p_n and q_n, respectively, with respect to a σ-finite measure μ_n dominating P_n+Q_n.

Appendix

DEFINITION A.1. *The sequences of probability measures* $\{P_n\}$ *and* $\{Q_n\}$ *are called contiguous if for any sequence* $\{A_n\}$ *with* $A_n \in \mathcal{B}_n$

$$P_n(A_n) \to 0 \qquad \textit{iff} \qquad Q_n(A_n) \to 0. \ \square$$

Contiguity of sequences of measures can be characterized by the asymptotic behaviour of their *log likelihood ratios*.

DEFINITION A.2. *The log likelihood ratio of* Q_n *with respect to* P_n *is the measurable map* $\Lambda_n(Q_n,P_n)\colon (X_n,\mathcal{B}_n) \to (\bar{\mathbb{R}},\bar{\mathcal{B}})$ *given by*

$$\Lambda_n(Q_n,P_n) = \log \frac{q_n}{p_n} .$$

Here $\log a/b$ *is* $-\infty$ *if* $a = 0 < b$, ∞ *if* $b = 0 < a$, *and* 0 *if* $a = b = 0$. $\square$

The measurable map $\Lambda_n(Q_n,P_n)$ induces probability measures $\underline{L}_{P_n}(\Lambda_n(Q_n,P_n))$ and $\underline{L}_{Q_n}(\Lambda_n(Q_n,P_n))$ on $\bar{\mathbb{R}} = [-\infty,\infty]$. We let $L_{P_n}(\Lambda_n(Q_n,P_n))$ and $L_{Q_n}(\Lambda_n(Q_n,P_n))$ be the possibly *defective* restrictions of these measures to $\mathbb{R}$. We say that a sequence of defective probability measures $\{L_n\}$ on $\mathbb{R}$ is *tight* if to any $\varepsilon > 0$ there exists a compact set $K_\varepsilon \subset \mathbb{R}$ and $N_\varepsilon \in \mathbb{N}$ with $L_n(K_\varepsilon) \geq 1-\varepsilon$ for all $n \geq N_\varepsilon$. We say that it *converges weakly* to a *proper* probability measure L if $\int f\, dL_n \to \int f\, dL$ for all bounded, continuous $f\colon \mathbb{R} \to \mathbb{R}$. More generally we define:

DEFINITION A.3. *Let* $\{L_n\}$ *be a sequence of possibly defective probability measures on a measurable space* $(Y,\mathcal{A})$, *which is provided with a topology* τ. *Then* $\{L_n\}$ *converges weakly on* $(Y,\mathcal{A},\tau)$ *to a probability measure* L *iff*

$$\int f\, dL_n \to \int f\, dL,$$

for all bounded, τ*-continuous,* $\mathcal{A}$*-measurable functions* $f\colon Y \to \mathbb{R}$. $\square$

PROPOSITION A.4. *The following statements are equivalent:*

(i) $\{P_n\}$ *and* $\{Q_n\}$ *are contiguous*

(ii) $\{L_{P_n}(\Lambda_n(Q_n,P_n))\}$ *and* $\{L_{Q_n}(\Lambda_n(Q_n,P_n))\}$ *are tight*

(iii) $\{L_{P_n}(\Lambda_n(Q_n,P_n))\}$ *is tight with every limit point* L_0 *satisfying*

$\int \exp(\lambda)\, dL_0(\lambda) = 1$. □

PROOF. (i)⇔ (ii). Write $\{Q_n\}\blacktriangleright\{P_n\}$ if for every sequence $\{A_n\}$ with $A_n \in \mathcal{B}_n$, we have that $Q_n(A_n) \to 0$ implies $P_n(A_n) \to 0$. We first show

(A.1) $\quad \{Q_n\}\blacktriangleright\{P_n\}$ iff $P_n(|\Lambda_n(Q_n,P_n)| \geq c_n) \to 0$ whenever $c_n \to \infty$.

Indeed we have

$$P_n(|\Lambda_n(Q_n,P_n)| \geq c_n) = \int 1\{p_n \leq \exp(-c_n)q_n \text{ or } q_n \leq \exp(-c_n)p_n\}\, p_n d\mu_n$$

(A.2)

$$\leq \exp(-c_n) + P_n(q_n \leq \exp(-c_n)p_n) \to 0,$$

if $c_n \to \infty$ and $\{Q_n\}\blacktriangleright\{P_n\}$, because

$$Q_n(q_n \leq \exp(-c_n)p_n) \leq \exp(-c_n) \to 0.$$

Conversely, suppose that the right hand side of (A.1) holds and let $Q_n(A_n) \to 0$. Then for suitable $c_n \to \infty$

$$P_n(A_n) \leq P_n(|\Lambda_n(Q_n,P_n)| \geq c_n) + \int_{A_n} \exp(c_n)\, q_n d\mu_n$$

(A.3)

$$= o(1) + \exp(c_n)\, Q_n(A_n) \to 0.$$

Next note that the right hand side of (A.1) is equivalent to tightness of $\{L_{P_n}(\Lambda_n(Q_n,P_n))\}$. Finally complete the proof of equivalence of (i) and (ii) by using symmetry and $\Lambda_n(P_n,Q_n) = -\Lambda_n(Q_n,P_n)$.

(ii)⇒(iii). Let L_0 be a limit point of $\{L_{P_n}(\Lambda_n(Q_n,P_n))\}$. By tightness of $\{L_{Q_n}(\Lambda_n(Q_n,P_n))\}$ there exists a subsequence of $\{n\}$ (denoted $\{n\}$) and a probability measure L' with

$$L_{P_n}(\Lambda_n(Q_n,P_n)) \to L_0 \qquad \text{and} \qquad L_{Q_n}(\Lambda_n(Q_n,P_n)) \to L'.$$

Hence for any continuous function h: $\mathbb{R} \to \mathbb{R}$ with compact support (setting $h(-\infty) = h(\infty) = 0$)

$$\begin{aligned} &\int_{\mathbb{R}} h(\lambda)e^{\lambda}\, dL_0(\lambda) = \lim_{n\to\infty} \int_{\mathbb{R}} h(\lambda)e^{\lambda}\, dL_{P_n}(\Lambda_n(Q_n,P_n))(\lambda) \\ &= \lim_{n\to\infty} \int_{\bar{\mathbb{R}}} h(\lambda)e^{\lambda}\, dL_{P_n}(\Lambda_n(Q_n,P_n))(\lambda) \\ &= \lim_{n\to\infty} \int h(\Lambda_n(Q_n,P_n)(x))\, \exp(\Lambda_n(Q_n,P_n)(x))\, dP_n(x) \\ &= \lim_{n\to\infty} \int h(\Lambda_n(Q_n,P_n)(x))\, dQ_n(x) = \int_{\mathbb{R}} h(\lambda)\, dL'(\lambda). \end{aligned} \tag{A.4}$$

Now choose a sequence $\{h_k\}$ with $0 \le h_k \le 1$ and increasing to 1, to see by monotone convergence

$$\int e^{\lambda}\, dL_0(\lambda) = \int dL'(\lambda) = 1. \tag{A.5}$$

(iii)⇒(ii). For any subsequence of {n} there is a further subsequence (denoted {n}) and a probability measure L_0 on $\mathbb{R}$ with $\{L_{P_n}(\Lambda_n(Q_n,P_n))\} \to L_0$ and

$$\int e^{\lambda}\, dL_0(\lambda) = 1. \tag{A.6}$$

Hence for any h as above

$$\begin{aligned} &\int h(\lambda)\, dL_{Q_n}(\Lambda_n(Q_n,P_n))(\lambda) \\ &= \int h(\Lambda_n(Q_n,P_n)(x))\, \exp(\Lambda_n(Q_n,P_n)(x))\, dP_n(x) \to \int h(\lambda)e^{\lambda}\, dL_0(\lambda). \end{aligned}$$

By (A.6) there exists h with $0 \le h \le 1$ and $\int h(\lambda)e^{\lambda}\, dL_0(\lambda) \ge 1-\varepsilon$. Then for sufficiently large n

$$Q_n(\, \Lambda_n(Q_n,P_n) \in \text{support}(h)\,) \ge 1-2\varepsilon. \quad \blacksquare$$

COROLLARY A.5. *Suppose that* $L_{P_n}(\Lambda_n(Q_n,P_n)) \to N(\mu,\sigma^2)$. *Then* $\{P_n\}$ *and* $\{Q_n\}$ *are contiguous if and only if* $\mu = -\frac{1}{2}\sigma^2$. □

PROOF. Use (iii) of Proposition A.4. ■

In the following propositions $L_{P_n}(Y_n,\Lambda_n(Q_n,P_n))$ is the restriction to $\mathcal{Y}\times\mathbb{R}$ of the probability distribution induced on $\mathcal{Y}\times\bar{\mathbb{R}}$ by the map $(Y_n,\Lambda_n(Q_n,P_n))$.

PROPOSITION A.6. *Let* $Y_n\colon (\mathcal{X}_n,\mathcal{B}_n) \to (\mathcal{Y},\mathcal{A})$ *be measurable maps, let* $\{P_n\}$ *and* $\{Q_n\}$ *be contiguous and suppose that*

$$L_{P_n}(Y_n,\Lambda_n(Q_n,P_n)) \to L \qquad \text{on } (\mathcal{Y}\times\mathbb{R},\mathcal{A}\times\mathcal{B},\tau\times|\cdot|).$$

Then

$$L_{Q_n}(Y_n) \to L' \qquad \text{on } (\mathcal{Y},\mathcal{A},\tau),$$

where L' *is a probability measure on* $(\mathcal{Y},\mathcal{A})$ *defined by*

$$L'(A) = \int_{A\times\mathbb{R}} e^{\lambda}\, dL(y,\lambda). \qquad (A.7)$$ □

PROOF. Let $f\colon \mathcal{Y} \to \mathbb{R}$ be bounded, τ-continuous and $\mathcal{A}$-measurable. We must show that

$$E_{Q_n} f(Y_n) \to \int_{\mathcal{Y}\times\mathbb{R}} f(y)\, e^{\lambda}\, dL(y,\lambda).$$

Write Λ_n for $\Lambda_n(Q_n,P_n)$ and define a probability measure L_0 on $\mathbb{R}$ by $L_0(B) = L(\mathcal{Y}\times B)$. Then $L_{P_n}(\Lambda_n) \to L_0$. By Proposition A.4(iii)

$$\int_{\mathcal{Y}\times\mathbb{R}} e^{\lambda}\, dL(y,\lambda) = \int_{\mathbb{R}} e^{\lambda}\, dL_0(\lambda) = 1.$$

Furthermore by (ii) of the same proposition $\{L_{Q_n}(\Lambda_n)\}$ is tight. Hence for every $\varepsilon > 0$ there exists $N_\varepsilon \in \mathbb{N}$ and a compact set $K_\varepsilon \subset \mathbb{R}$ with

$$\int_{\mathcal{Y}\times K_\varepsilon^c} e^{\lambda}\, dL(y,\lambda) \le \varepsilon \qquad (A.8)$$

(A.9) $\qquad Q_n(\Lambda_n \in K_\varepsilon^c) \le \varepsilon \qquad\qquad n \ge N_\varepsilon.$

Choose a continuous function $h: \mathbb{R} \to \mathbb{R}$ with compact support and $0 \le 1_{K_\varepsilon} \le h \le 1$. Set $h(-\infty) = h(\infty) = 0$. We have

$$|E_{Q_n} f(Y_n) - \gamma_{\times\mathbb{R}} \int f(y)\, e^{\lambda}\, dL(y,\lambda)|$$

$$\le |E_{Q_n} f(Y_n)\,(1-h(\Lambda_n))| + |E_{Q_n} f(Y_n) h(\Lambda_n) - \gamma_{\times\mathbb{R}} \int f(y)\, h(\lambda) e^{\lambda}\, dL(y,\lambda)|$$

$$+ \gamma_{\times\mathbb{R}} \int f(y)\,(h(\lambda)-1) e^{\lambda}\, dL(y,\lambda)|.$$

By (A.8)-(A.9) and boundedness of f, it suffices to show that the middle term converges to zero. Since the function given by $(y,\lambda) \to f(y)h(\lambda)e^{\lambda}$, from $\gamma_{\times\mathbb{R}}$ to $\mathbb{R}$, is bounded, $\tau_{\times}|\cdot|$-continuous and $\mathcal{A}\times\mathcal{B}$-measurable

$$E_{Q_n} f(Y_n) h(\Lambda_n) = \int f(Y_n(x))\, h(\Lambda_n(x))\, 1\{x: |\Lambda_n(x)| < \infty\}\, dQ_n(x)$$

$$= \int f(Y_n(x))\, h(\Lambda_n(x)) \exp \Lambda_n(x)\, dP_n(x)$$

$$= \gamma_{\times\mathbb{R}} \int f(y) h(\lambda) e^{\lambda}\, dL_{P_n}(Y_n, \Lambda_n)(y,\lambda) \to \gamma_{\times\mathbb{R}} \int f(y) h(\lambda) e^{\lambda}\, dL(y,\lambda). \blacksquare$$

The following proposition is known as the third lemma of Le Cam.

COROLLARY A.7. *Let* $Y_n: (X_n, \mathcal{B}_n) \to \mathbb{R}^k$ *be measurable maps and assume that*

(A.10) $\qquad L_{P_n}(Y_n, \Lambda_n(Q_n,P_n)) \to N_{k+1}\left[\begin{pmatrix}\mu\\ -\frac{1}{2}\sigma^2\end{pmatrix}, \begin{pmatrix}\Sigma & \tau\\ \tau' & \sigma^2\end{pmatrix}\right]$ *on* $\mathbb{R}^{k+1}$.

Then $\{P_n\}$ *and* $\{Q_n\}$ *are contiguous and*

$$L_{Q_n}(Y_n) \to N_k(\mu+\tau, \Sigma) \qquad \text{on } \mathbb{R}^k. \ \square$$

PROOF. Contiguity follows from Corollary A.5. Next by Proposition A.6 we

have $L_{Q_n}(Y_n) \to L'$, where L' satisfies (A.7), with L equal to the normal distribution at the right hand side of (A.10). The characteristic function of L' equals

$$\phi_{L'}(s) = \int \exp(is'y+\lambda)\, dL(y,\lambda) = \phi_L(s,-i)$$

$$= \exp\left(is'\mu - \tfrac{1}{2}\sigma^2 - \tfrac{1}{2}[(s',-i)\begin{bmatrix}\Sigma & \tau \\ \tau' & \sigma^2\end{bmatrix}\begin{bmatrix}s \\ -i\end{bmatrix}]\right) = \exp(is'(\mu+\tau) - \tfrac{1}{2}s'\Sigma s). \blacksquare$$

A.2. LAN, DIFFERENTIABILITY IN QUADRATIC MEAN AND RELATED RESULTS

Local asymptotic normality is the key property underlying the theory of Chapters 2 and 3 as well as part of later chapters. For models with independent observations local asymptotic normality is implied by *differentiability in quadratic mean* of the square root of the densities. This well-known fact, which is due to LeCam, is proved in an abstract form in Proposition A.8. Next Proposition A.9 gives a sufficient condition for differentiability in quadratic mean of the roots of the densities in terms of L_2-derivatives of the densities themselves. This proposition is useful in some applications, including mixture models. The last proposition in this section, Proposition A.10, has important applications as a statement on the behaviour of the derivatives, or projections thereof. It allows the construction of one-step estimators under a weak continuity condition on the score functions (cf. Lemma 5.4).

For $n = 1,2,\ldots$ and $j = 1,2,\ldots,n$ let P_{nj} and Q_{nj} be probability measures on measurable spaces $(\mathcal{X}_{nj},\mathcal{B}_{nj})$ with densities p_{nj} and q_{nj} with respect to a σ-finite measure μ_{nj} dominating $P_{nj}+Q_{nj}$. Let $\Lambda_n(Q_n,P_n)$ be the log likelihood ratio of the *product* measures $P_n = \otimes_{j=1}^n P_{nj}$ and $Q_n = \otimes_{j=1}^n Q_{nj}$ on the product space $(\mathcal{X}_n,\mathcal{B}_n) = \otimes_{j=1}^n(\mathcal{X}_{nj},\mathcal{B}_{nj})$. Elements of $(\mathcal{X}_n,\mathcal{B}_n)$ are denoted $(X_{n1},\ldots,X_{nn})$. Of course we may think of $(X_{n1},\ldots,X_{nn})$ as a random element in $(\mathcal{X}_n,\mathcal{B}_n)$ defined on some probability space. Note that $\Lambda_n(Q_n,P_n)$ is a function of $(X_{n1},\ldots,X_{nn})$.

Appendix

PROPOSITION A.8. *Suppose that*

$$\text{(A.11)} \qquad n^{-1} \Sigma_{j=1}^{n} \int [\sqrt{n}(q_{nj}^{\frac{1}{2}} - p_{nj}^{\frac{1}{2}}) - \tfrac{1}{2} g_{nj} p_{nj}^{\frac{1}{2}}]^2 \, d\mu_{nj} \to 0,$$

for a triangular array of measurable real functions $(g_{n1}, \ldots, g_{nn})$ *with*

$$\text{(A.12)} \qquad n^{-\frac{1}{2}} \Sigma_{j=1}^{n} \int g_{nj} \, dP_{nj} = o(1)$$

$$\text{(A.13)} \qquad \sigma_n^2 = n^{-1} \Sigma_{j=1}^{n} \int g_{nj}^2 \, dP_{nj} = O(1)$$

$$\text{(A.14)} \qquad n^{-1} \Sigma_{j=1}^{n} \int g_{nj}^2 \, 1\{|g_{nj}| \geq \varepsilon\sqrt{n}\} \, dP_{nj} = o(1) \qquad \textit{for every } \varepsilon > 0.$$

Then in P_n*-probability*

$$\Lambda_n(Q_n, P_n) - n^{-\frac{1}{2}} \Sigma_{j=1}^{n} g_{nj}(X_{nj}) + \tfrac{1}{2} \sigma_n^2 \to 0. \ \square$$

PROOF. Write E_n and σ_n^2 for expectation and variance under P_n and omit X_{nj} in such expressions as $q_{nj}(X_{nj})$, $g_{nj}(X_{nj})$, etc. Set

$$W_{nj} = 2[q_{nj}^{\frac{1}{2}} / p_{nj}^{\frac{1}{2}} - 1].$$

By (A.11)

$$\text{(A.15)} \qquad \sigma_n^2 (\Sigma_{j=1}^{n} W_{nj} - n^{-\frac{1}{2}} \Sigma_{j=1}^{n} g_{nj}) \leq n^{-1} \Sigma_{j=1}^{n} E_n(\sqrt{n} W_{nj} - g_{nj})^2 \to 0.$$

Furthermore

$$E_n(\Sigma_{j=1}^{n} W_{nj}) = 2\Sigma_{j=1}^{n}(\int q_{nj}^{\frac{1}{2}} p_{nj}^{\frac{1}{2}} - 1) = -\Sigma_{j=1}^{n} \int (q_{nj}^{\frac{1}{2}} - p_{nj}^{\frac{1}{2}})^2 d\mu_{nj} = -\tfrac{1}{4}\sigma_n^2 + o(1).$$

Combining this with (A.12) and (A.15) we see

$$\text{(A.16)} \qquad \Sigma_{j=1}^{n} W_{nj} - n^{-\frac{1}{2}} \Sigma_{j=1}^{n} g_{nj} + \tfrac{1}{4}\sigma_n^2 \overset{P_n}{\to} 0.$$

From the right hand side of (A.15) and (A.13) deduce

$$n^{-1} \Sigma_{j=1}^{n} E_n |n W_{nj}^2 - g_{nj}^2| \to 0.$$

Setting $W_{nj}^2 = n^{-1}g_{nj}^2 + n^{-1}A_{nj}$ we can express this alternatively by

(A.17) $\quad n^{-1}\Sigma_{j=1}^n E_n|A_{nj}| \to 0.$

Together with (A.14) this gives

$$P_n(\max_{j=1,n} |W_{nj}| \geq \varepsilon\sqrt{2}) \leq \Sigma_{j=1}^n P_n(|W_{nj}| \geq \varepsilon\sqrt{2})$$

(A.18) $$\leq \Sigma_{j=1}^n P_n(n^{-1}g_{nj}^2 \geq \varepsilon^2) + \Sigma_{j=1}^n P_n(n^{-1}|A_{nj}| \geq \varepsilon^2)$$

$$\leq \varepsilon^{-2}n^{-1}\Sigma_{j=1}^n E_n g_{nj}^2 1\{|g_{nj}| \geq \varepsilon\sqrt{n}\} + \varepsilon^{-2}n^{-1}\Sigma_{j=1}^n E_n|A_{nj}| \to 0.$$

Next since $\log(1+x) = x - \frac{1}{2}x^2 + x^2R(x)$, where $R(x) \to 0$ as $x \to 0$,

(A.19) $\quad \Lambda_n(Q_n,P_n) = 2\Sigma_{j=1}^n \log(1+\frac{1}{2}W_{nj}) = \Sigma_{j=1}^n[W_{nj} - \frac{1}{4}W_{nj}^2 + \frac{1}{2}W_{nj}^2 R(\frac{1}{2}W_{nj})] + o_{P_n}(1).$

By (A.14) there exist $\varepsilon_n \downarrow 0$ such that

$$n^{-1}\Sigma_{j=1}^n E_n\, g_{nj}^2\, 1\{|g_{nj}| \geq \varepsilon_n\sqrt{n}\} \to 0.$$

Then

(A.20) $\quad n^{-1}\Sigma_{j=1}^n E_n g_{nj}^2 - n^{-1}\Sigma_{j=1}^n E_n g_{nj}^2 1\{|g_{nj}| \leq \varepsilon_n\sqrt{n}\} \to 0$

(A.21) $\quad \sigma_n^2(\, n^{-1}\Sigma_{j=1}^n\, g_{nj}^2\, 1\{|g_{nj}| \leq \varepsilon_n\sqrt{n}\}\,) = O(\varepsilon_n^2) \to 0.$

Using (A.17), (A.21) and (A.20) we see

(A.22) $\quad \Sigma_{j=1}^n W_{nj}^2 - n^{-1}\Sigma_{j=1}^n E_n g_{nj}^2 \xrightarrow{P} 0.$

Combine (A.13), (A.18)-(A.19) and (A.22) to conclude

$$\Lambda_n(Q_n,P_n) - \Sigma_{j=1}^n W_{nj} + \tfrac{1}{4}\sigma_n^2 \xrightarrow{P} 0.$$

Together with (A.16) this proves the proposition. ■

Appendix

Though differentiability as in Proposition A.8 is theoretically convenient, in some applications it is easier to establish local asymptotic normality by a stronger differentiability property.

PROPOSITION A.9. *Suppose that for a triangular array of measurable functions* $(g_{n1},\ldots,g_{nn})$ *satisfying* (A.13)-(A.14), *we have*

$$n^{-1}\,\Sigma_{j=1}^{n}\int\,[\,\sqrt{n}\,p_{nj}^{-1}(q_{nj}-p_{nj})\,-\,g_{nj}\,]^2\,dP_{nj}\to 0$$

and

$$\Sigma_{j=1}^{n}\int 1\{x:\,p_{nj}(x)=0\}\,dQ_{nj}\to 0.$$

Then (A.11) *holds.* □

PROOF. We have

$$n^{-1}\Sigma_{j=1}^{n}\int\,[\sqrt{n}(q_{nj}^{\frac12}-p_{nj}^{\frac12})\,-\,\tfrac12 g_{nj}p_{nj}^{\frac12}]^2\,d\mu_{nj}$$

$$=\,n^{-1}\Sigma_{j=1}^{n}\int[\sqrt{n}p_{nj}^{-\frac12}(q_{nj}^{\frac12}-p_{nj}^{\frac12})\,-\,\tfrac12 g_{nj}]^2\,dP_{nj}\,+\,\Sigma_{j=1}^{n}\int 1\{x:\,p_{nj}(x)=0\}\,dQ_{nj}.$$

It therefore suffices to show

$$\Sigma_{j=1}^{n}\int\,[p_{nj}^{-\frac12}(q_{nj}^{\frac12}-p_{nj}^{\frac12})\,-\,\tfrac12 p_{nj}^{-1}(q_{nj}-p_{nj})]^2\,dP_{nj}\to 0. \tag{A.23}$$

Set $b_{nj}=p_{nj}^{-1}\,(q_{nj}-p_{nj})$ and let $\varepsilon_n\downarrow 0$ such that

$$n^{-1}\,\Sigma_{j=1}^{n}\int\,g_{nj}^2\,1\{\,|g_{nj}|\,\ge\,\varepsilon_n^2\sqrt{n}\}\,dP_{nj}\to 0.$$

Then

$$\Sigma_{j=1}^{n}\int\,b_{nj}^2\,1\{\,|b_{nj}|\ge\varepsilon_n\}\,dP_{nj}\,\le\,2n^{-1}\Sigma_{j=1}^{n}\int\,g_{nj}^2\,1\{\,|b_{nj}|\ge\varepsilon_n\}\,dP_{nj}\,+\,o(1)$$

$$\le\,2n^{-1}\Sigma_{j=1}^{n}\varepsilon_n^4\,n\,P_{nj}(\,|b_{nj}|\ge\varepsilon_n)\,+\,o(1)\,\le\,2\Sigma_{j=1}^{n}\varepsilon_n^2\int\,b_{nj}^2\,dP_{nj}\,+\,o(1)\to 0.$$

Now apply Lemma A.14 (below) with $a=1$ and $c=0$ to see that

$$|\,(1+b_{nj})^{\frac12}\,-\,1\,-\,\tfrac12 b_{nj}\,1\{\,|b_{nj}|\,<\,\varepsilon_n\}\,|^2$$

$$\le\,3(1-\varepsilon_n)^{-1}b_{nj}^2\,1\{\,|b_{nj}|\,\ge\,\varepsilon_n\}\,+\,((1-\varepsilon_n)^{-\frac12}-1)^2\,b_{nj}^2\,1\{\,|b_{nj}|\,<\,\varepsilon_n\}.$$

Hence the left hand side of (A.23) is smaller than

$$2(\tfrac{1}{4}+3(1-\varepsilon_n)^{-1})\ \Sigma_{j=1}^n \int b_{nj}^2\ 1\{|b_{nj}| \geq \varepsilon_n\}\ dP_{nj}$$

$$+\ 2((1-\varepsilon_n)^{-\frac{1}{2}}-1)^2\ \Sigma_{j=1}^n \int b_{nj}^2\ 1\{|b_{nj}| < \varepsilon_n\}\ dP_{nj} \to 0. \ \blacksquare$$

This section is concluded with a proposition which in a disguised form gives a sufficient condition for a frequently used differentiability property of score functions. For instance, for efficient score functions $\mathcal{L}(\cdot,\theta,\eta)$ in a semi-parametric model one expects for $\sqrt{n}(\theta_n-\theta) \to h$

$$n^{-\frac{1}{2}}\ \Sigma_{j=1}^n (\mathcal{L}(X_j,\theta_n,\eta) - \mathcal{L}(X_j,\theta,\eta)) \approx h\ n^{-1}\Sigma_{j=1}^n\ \partial/\partial\theta \mathcal{L}(X_j,\theta,\eta)$$

$$P_\theta \to hE_\theta \partial/\partial\theta \mathcal{L}(X_1,\theta,\eta) \approx -hE_\theta \mathcal{L}^2(X_1,\theta,\eta)\ .$$

It is quite remarkable that this is in fact true under an L_2- continuity condition on $\mathcal{L}(\cdot,\theta,\eta)$ only, without making reference to a derivative. Bickel (1982) proves the proposition below for score functions in parametric models. Proposition A.10 applies to general elements of L_2-spaces, in particular to efficient score functions. For the application as above only continuity in θ of the efficient score function $\mathcal{L}(\cdot,\theta,\eta)$ is required, η being fixed.

PROPOSITION A.10. *Let the conditions of Proposition A.8 hold and suppose that for triangular arrays of measurable functions* $(k_{n1},\ldots,k_{nn}) \in \otimes_{j=1}^n L_2(Q_{nj})$ *and* $(\ell_{n1},\ldots,\ell_{nn}) \in \otimes_{j=1}^n L_2(P_{nj})$

(A.24) $$n^{-1}\ \Sigma_{j=1}^n \int [k_{nj}q_{nj}^{\frac{1}{2}} - \ell_{nj}p_{nj}^{\frac{1}{2}}]^2\ d\mu_{nj} \to 0$$

(A.25) $$n^{-1}\ \Sigma_{j=1}^n \int \ell_{nj}^2\ dP_{nj} = O(1)$$

(A.26) $$n^{-1}\ \Sigma_{j=1}^n \int k_{nj}^2\ 1\{|k_{nj}| \geq \varepsilon\sqrt{n}\}\ dQ_{nj} \to 0 \qquad \textit{for every } \varepsilon > 0.$$

Then

$$n^{-\frac{1}{2}}\ \Sigma_{j=1}^n\ (k_{nj}(X_{nj}) - \ell_{nj}(X_{nj}))$$

$$+\ n^{-1}\Sigma_{j=1}^n \int \ell_{nj}g_{nj}dP_{nj} - n^{-\frac{1}{2}}\Sigma_{j=1}^n (\int k_{nj}dQ_{nj} - \int \ell_{nj}dP_{nj}) \overset{P_n}{\to} 0. \ \square$$

PROOF. Write E_n and σ_n^2 for expectation and variance under P_n and omit X_{nj} in such expressions as $q_{nj}(X_{nj})$, $g_{nj}(X_{nj})$, etc. Set

$$\tilde{I}_n = n^{-1}\Sigma_{j=1}^n \int \ell_{nj} g_{nj}\, dP_{nj}$$

By (A.24)

$$\sigma_n^2\Big(n^{-\frac{1}{2}}\Sigma_{j=1}^n \left(k_{nj} q_{nj}^{\frac{1}{2}}/p_{nj}^{\frac{1}{2}} - \ell_{nj}\right)\Big) \to 0. \tag{A.28}$$

By (A.24),(A.25) and (A.11), (A.13)

$$E_n\left(n^{-\frac{1}{2}}\Sigma_{j=1}^n k_{nj} q_{nj}^{\frac{1}{2}}/p_{nj}^{\frac{1}{2}}\right) = n^{-\frac{1}{2}}\Sigma_{j=1}^n \int k_{nj} q_{nj}^{\frac{1}{2}}(p_{nj}^{\frac{1}{2}} - q_{nj}^{\frac{1}{2}})\, d\mu_{nj} + n^{-\frac{1}{2}}\Sigma_{j=1}^n \int k_{nj}\, dQ_{nj}$$

$$= -\tfrac{1}{2}n^{-1}\Sigma_{j=1}^n \int \ell_{nj} p_{nj}^{\frac{1}{2}}\, g_{nj} p_{nj}^{\frac{1}{2}}\, d\mu_{nj} + n^{-\frac{1}{2}}\Sigma_{j=1}^n \int k_{nj}\, dQ_{nj} + o(1) \tag{A.29}$$

$$= -\tfrac{1}{2}\tilde{I}_n + n^{-\frac{1}{2}}\Sigma_{j=1}^n \int k_{nj}\, dQ_{nj} + o(1).$$

Combining (A.28)-(A.29) we see that it suffices to show

$$n^{-\frac{1}{2}}\Sigma_{j=1}^n k_{nj}(1 - q_{nj}^{\frac{1}{2}}/p_{nj}^{\frac{1}{2}}) + \tfrac{1}{2}\tilde{I}_n \overset{P_n}{\to} 0. \tag{A.30}$$

Define n-fold product measures P_n' by the densities $\Pi_{j=1}^n q_{nj}^{\frac{1}{2}}(x_j) p_{nj}^{\frac{1}{2}}(x_j) c_{nj}$, where

$$c_{nj}^{-1} = \int q_{nj}^{\frac{1}{2}} p_{nj}^{\frac{1}{2}}\, d\mu_{nj} = 1 - \tfrac{1}{2}\int (q_{nj}^{\frac{1}{2}} - p_{nj}^{\frac{1}{2}})^2\, d\mu_{nj} = 1 - \tfrac{1}{2}\alpha_{nj},$$

say. By (A.11), (A.13)

$$\Sigma_{j=1}^n \alpha_{nj} = \tfrac{1}{4}\sigma_n^2 + o(1)$$

and by (A.11) and (A.14)

$$\max_{j=1..n} |\alpha_{nj}| \le 2n^{-1}\Sigma_{j=1}^n \int [\sqrt{n}(q_{nj}^{\frac{1}{2}} - p_{nj}^{\frac{1}{2}}) - \tfrac{1}{2} g_{nj} p_{nj}^{\frac{1}{2}}]^2 d\mu_{nj} + \tfrac{1}{2}n^{-1}\max_{j=1..n} \int g_{nj}^2\, dP_{nj} \to 0.$$

Therefore

$$\log \Pi_{j=1}^n c_{nj} = -\Sigma_{j=1}^n \log(1 - \tfrac{1}{2}\alpha_{nj}) = 1/8\ \sigma_n^2 + o(1)$$

and

$$1 \le \max_{j=1,n} c_{nj} \to 1. \tag{A.31}$$

By Proposition A.8

$$A_n(P_n', P_n) = \tfrac{1}{2} n^{-\frac{1}{2}} \Sigma_{j=1}^n g_{nj} - 1/8\, \sigma_n^2 + o_{P_n}(1).$$

By Corollary A.5 $\{P_n'\}$ and $\{P_n\}$ are contiguous and we may now finish the proof by showing convergence to zero in (A.30) in P_n'-probability. First by (A.31), (A.11), (A.24) and (A.25)

$$E_n' \left| n^{-\frac{1}{2}} \Sigma_{j=1}^n k_{nj}(1 - q_{nj}^{\frac{1}{2}}/p_{nj}^{\frac{1}{2}}) + \tfrac{1}{2} n^{-1} \Sigma_{j=1}^n k_{nj} g_{nj} \right|$$

$$\le n^{-1} \Sigma_{j=1}^n \int |k_{nj} q_{nj}^{\frac{1}{2}}| \left| \sqrt{n}(p_{nj}^{\frac{1}{2}} - q_{nj}^{\frac{1}{2}}) + \tfrac{1}{2} g_{nj} p_{nj}^{\frac{1}{2}} \right| c_{nj}\, d\mu_{nj} \to 0.$$

Hence it suffices to show

$$n^{-1} \Sigma_{j=1}^n k_{nj} g_{nj} - \tilde{I}_n \overset{P_n'}{\to} 0. \tag{A.32}$$

Now

$$n^{-1} \Sigma_{j=1}^n \int |k_{nj} g_{nj} q_{nj}^{\frac{1}{2}} p_{nj}^{\frac{1}{2}} c_{nj} - \ell_{nj} g_{nj} p_{nj}|\, d\mu_{nj} \to 0.$$

Hence, writing E_n' and $\sigma_n'^2$ for expectation and variance under P_n', we have

$$E_n' \left(n^{-1} \Sigma_{j=1}^n k_{nj} g_{nj} - \tilde{I}_n \right) \to 0. \tag{A.33}$$

Furthermore for any $\varepsilon > 0$

$$\begin{aligned} &\sigma_n'^2 \Big(n^{-1} \Sigma_{j=1}^n k_{nj} 1\{|k_{nj}| \le \varepsilon \sqrt{n}\}\; g_{nj} 1\{|g_{nj}| \le \varepsilon \sqrt{n}\} \Big) \\ &\le n^{-1} \varepsilon^2 \Sigma_{j=1}^n E_n' |k_{nj} g_{nj}| = \varepsilon^2 O(1). \end{aligned} \tag{A.34}$$

Finally by (A.25)-(A.26) and the Cauchy Schwarz inequality

$$E_n' \left| n^{-1} \Sigma_{j=1}^n k_{nj} g_{nj}\, 1\{|k_{nj}| \ge \varepsilon \sqrt{n} \text{ or } |g_{nj}| \ge \varepsilon \sqrt{n}\} \right|$$

$$\text{(A.35)} \qquad \leq n^{-1} \max c_{nj} \Big[\Big(\Sigma_{j=1}^{n} \int k_{nj}^2 1\{|k_{nj}| \geq \varepsilon\sqrt{n}\}\, dQ_{nj}\Big)^{\frac{1}{2}} \Big(\Sigma_{j=1}^{n} \int g_{nj}^2\, dP_{nj}\Big)^{\frac{1}{2}}$$

$$+ \Big(\Sigma_{j=1}^{n} \int k_{nj}^2\, dQ_{nj}\Big)^{\frac{1}{2}} \Big(\Sigma_{j=1}^{n} \int g_{nj}^2 1\{|g_{nj}| \geq \varepsilon\sqrt{n}\}\, dP_{nj}\Big)^{\frac{1}{2}} \Big] \to 0.$$

(A.33)-(A.35) yield (A.32), concluding the proof of (A.30). ∎

A.3. DIFFERENTIABILITY AND CONTINUITY UNDER MEASURABLE TRANSFORMATIONS

In the situation of Section A.2 let $t_{nj}: (\mathcal{X}_{nj}, \mathcal{B}_{nj}) \to (\mathcal{T}_{nj}, \mathcal{C}_{nj})$ be measurable maps and let $T_{nj} = t_{nj}(X_{nj})$. Consider the situation that instead of $(X_{n1}, \ldots, X_{nn})$, one observes the vector $(T_{n1}, \ldots, T_{nn})$. Clearly this involves a loss in 'information'. Information in the 'original' sample $(X_{n1}, \ldots, X_{nn})$ can be measured by the score functions $(g_{n1}, \ldots, g_{nn})$. Intuitively, the information in the transformed sample can be measured by the *transformed scores* $(\underline{g}_{n1}, \ldots, \underline{g}_{nn})$, where

$$\text{(A.36)} \qquad \underline{g}_{nj}(t) = E_{P_{nj}}\big(g_{nj}(X_{nj}) \mid T_{nj} = t \big) \qquad \text{a.e..}$$

The information loss is then expressed by

$$E_{P_{nj}} \underline{g}_{nj}^2(T_{nj}) \leq E_{P_{nj}} g_{nj}^2(X_{nj}) \ .$$

The above can be given a rigorous interpretation and proof. Versions of the propositions in this section can be found in the appendix of Bickel, Klaassen, Ritov and Wellner (198?) and more generally in LeCam and Yang (1986). The proofs below are based on the proofs in Van der Vaart (1987a), which were written out for the special case of mixture models.

Let $\underline{P}_{nj} = t_{nj}(P_{nj})$ adn $\underline{Q}_{nj} = t_{nj}(Q_{nj})$ be the image measures of P_{nj} and Q_{nj} on $(\mathcal{T}_{nj}, \mathcal{C}_{nj})$, respectively, and let $\underline{p}_{nj}$ and $\underline{q}_{nj}$ be their densities with respect to some arbitrary σ-finite measure $\underline{\mu}_{nj}$ on $(\mathcal{T}_{nj}, \mathcal{C}_{nj})$. Of course, if X_{nj} is a random element in $(\mathcal{X}_{nj}, \mathcal{B}_{nj})$ with distribution P_{nj} (or Q_{nj}), then $T_{nj} = t_{nj}(X_{nj})$ is distributed according to $\underline{P}_{nj}$ ($\underline{Q}_{nj}$). Our aim in this section is to show that (A.11)-(A.14) and (A.24)-(A.26) carry over from the original to the transformed densities and score functions.

PROPOSITION A.11. *Suppose that* (A.12)-(A.14) *hold. Then*

$$\text{(A.37)} \qquad n^{-1}\Sigma_{j=1}^{n} \int \underline{g}_{nj}\, d\underline{P}_{nj} = o(1)$$

$$\text{(A.38)} \qquad n^{-1}\Sigma_{j=1}^{n} \int \underline{g}_{nj}^2\, d\underline{P}_{nj} = O(1)$$

$$\text{(A.39)} \qquad n^{-1}\Sigma_{j=1}^{n} \int \underline{g}_{nj}^2\, 1\{|\underline{g}_{nj}| \geq \varepsilon\sqrt{n}\}\, d\underline{P}_{nj} \to 0 \qquad \textit{for every } \varepsilon > 0 \ . \ \square$$

PROOF. The first two assertions follow directly from the properties of a conditional expectation. By (A.14) there exists $\varepsilon_n \downarrow 0$ such that

$$\text{(A.40)} \qquad n^{-1}\Sigma_{j=1}^{n} \int g_{nj}^2\, 1\{|g_{nj}| \geq \varepsilon_n\sqrt{n}\}\, dP_{nj} \to 0 \ .$$

Let $C_{nj} = \{t: |\underline{g}_{nj}(t)| \geq \varepsilon\sqrt{n}\}$. We have

$$n^{-1}\Sigma_{j=1}^{n} \int_{C_{nj}} \underline{g}_{nj}^2\, d\underline{P}_{nj} \leq n^{-1}\Sigma_{j=1}^{n} E_{P_{nj}} E_{P_{nj}}(g_{nj}^2 | T_{nj}) 1_{C_{nj}}(T_{nj}) \ .$$

By (A.40) this is smaller than

$$\varepsilon_n^2 \Sigma_{j=1}^{n} \underline{P}_{nj}(C_{nj}) + o(1) \leq \varepsilon_n^2 \varepsilon^{-2}\, n^{-1}\Sigma_{j=1}^{n} \int \underline{g}_{nj}^2\, d\underline{P}_{nj} + o(1). \ \blacksquare$$

PROPOSITION A.12. *Suppose that* (A.11)-(A.14) *hold. Then*

$$\text{(A.41)} \qquad n^{-1}\Sigma_{j=1}^{n} \int [\sqrt{n}(\underline{q}_{nj}^{\frac{1}{2}} - \underline{p}_{nj}^{\frac{1}{2}}) - \tfrac{1}{2}\underline{g}_{nj}\underline{p}_{nj}^{\frac{1}{2}}]^2\, d\underline{\mu}_{nj} \to 0. \ \square$$

PROOF. Let $\nu_{nj} = \underline{Q}_{nj} + \underline{P}_{nj}$. First note that the left hand side of (A.41) is equal to

$$n^{-1}\Sigma_{j=1}^{n} \int [\sqrt{n}((\frac{d\underline{Q}_{nj}}{d\nu_{nj}})^{\frac{1}{2}} - (\frac{d\underline{P}_{nj}}{d\nu_{nj}})^{\frac{1}{2}}) - \tfrac{1}{2}\underline{g}_{nj}(\frac{d\underline{P}_{nj}}{d\nu_{nj}})^{\frac{1}{2}}]^2\, d\nu_{nj} \to 0 \ .$$

Indeed, ν_{nj} is absolutely continuous with respect to $\underline{\mu}_{nj}$ and

$$\underline{q}_{nj} = \frac{d\underline{Q}_{nj}}{d\nu_{nj}} \frac{d\nu_{nj}}{d\underline{\mu}_{nj}} \ ;$$

and similarly for $\underline{p}_{nj}$. It follows that (A.41) is true for *any* σ-finite measures $\underline{\mu}_{nj}$, if it is true for one choice of such measures.

By the same argument we know that (A.11) is true for $\mu_{nj} = Q_{nj}+P_{nj}$. Clearly, $\underline{\mu}_{nj}$ defined by $\underline{\mu}_{nj} = t_{nj}(\mu_{nj})$ is finite, hence certainly σ-finite. We now prove (A.41) for this choice of $\underline{\mu}_{nj}$. For this $\underline{\mu}_{nj}$ we can write

$$\text{(A.42)} \qquad \underline{p}_{nj}(t) = E_{\mu_{nj}}(p_{nj}(X_{nj}) \mid T_{nj} = t) \qquad \text{a.e.}$$

$$\text{(A.43)} \qquad \underline{q}_{nj}(t) = E_{\mu_{nj}}(q_{nj}(X_{nj}) \mid T_{nj} = t) \qquad \text{a.e.}.^{1)}$$

Let $\varepsilon_n \downarrow 0$ such that (A.40) holds. We omit X_{nj} and T_{nj} in expressions as $q_{nj}(X_{nj})$, $g_{nj}(X_{nj})$, $\underline{p}_{nj}(T_{nj})$ etc.. Set

$$\bar{g}_{nj} = g_{nj}\, 1\{|g_{nj}| \le \varepsilon_n\sqrt{n}\}$$

$$\text{(A.44)} \qquad u_{nj} = \sqrt{n}(q_{nj}^{\frac{1}{2}} - p_{nj}^{\frac{1}{2}}) - \tfrac{1}{2}\bar{g}_{nj}p_{nj}^{\frac{1}{2}}.$$

Then by (A.14) and (A.11)

$$\text{(A.45)} \qquad n^{-1}\Sigma_{j=1}^{n}\int (\underline{g}_{nj} - \bar{\underline{g}}_{nj}))^2 \, d\underline{P}_{nj} \le n^{-1}\Sigma_{j=1}^{n}\int (g_{nj} - \bar{g}_{nj})^2 \, dP_{nj} \to 0$$

$$\text{(A.46)} \qquad n^{-1}\Sigma_{j=1}^{n}\int u_{nj}^2 \, d\mu_{nj} \to 0 \ .$$

Next by (A.44) and (A.42)-(A.43)

$$\underline{q}_{nj} = \underline{p}_{nj} + n^{-\frac{1}{2}}E_{\mu_{nj}}(\bar{g}_{nj}p_{nj} \mid T_{nj}) + n^{-1}E_{\mu_{nj}}(u_{nj}^2 \mid T_{nj})$$

$$\text{(A.47)} \qquad + n^{-\frac{1}{2}}E_{\mu_{nj}}(2u_{nj}p_{nj}^{\frac{1}{2}} + n^{-\frac{1}{2}}u_{nj}\bar{g}_{nj}p_{nj}^{\frac{1}{2}} + n^{-\frac{1}{2}}\tfrac{1}{4}\bar{g}_{nj}^2 p_{nj} \mid T_{nj}) \ .$$

$$= \underline{p}_{nj} + \underline{b}_{nj} + \underline{c}_{nj} + \underline{d}_{nj} \quad \text{(say)}.$$

---

1) *Formally* $\underline{p}_{nj}$ *is defined as a measurable, real function on* $(\mathcal{T}_{nj}, \mathcal{C}_{nj})$ *such that* $\int \underline{p}_{nj}(t_{nj}(x))1_C(t_{nj}(x))\, d\mu_{nj}(x) = \int p_{nj}(x)1_C(t_{nj}(x))\, d\mu_{nj}(x)$ *for any* $C \in \mathcal{C}_{nj}$, *and* $\underline{q}_{nj}$ *is defined similarly.*

We have for any g_{nj}

$$(A.48) \qquad E_{\mu_{nj}}(g_{nj}p_{nj}|T_{nj}) = E_{P_{nj}}(g_{nj}|T_{nj})\, E_{\mu_{nj}}(p_{nj}|T_{nj}) = \underline{g}_{nj}\, \underline{p}_{nj} \ .$$

Apply (A.47)-(A.48) together with Lemma A.14 below, to see, for $\underline{p}_{nj} > 0$

$$(A.49) \qquad \begin{aligned} &[\underline{q}_{nj}^{\frac{1}{2}} - \underline{p}_{nj}^{\frac{1}{2}} - n^{-\frac{1}{2}}\, \tfrac{1}{2}\bar{g}_{nj}\, \underline{p}_{nj}^{\frac{1}{2}}]^2 \\ &\le 3(1-\varepsilon_n)^{-1}\underline{d}_{nj}^2/\underline{p}_{nj} + 3\underline{c}_{nj} + ((1-\varepsilon_n)^{-\frac{1}{2}}-1)^2\, \underline{b}_{nj}^2/\underline{p}_{nj} \ . \end{aligned}$$

Now using repeatedly that $|\bar{g}_{nj}| \le \varepsilon_n\sqrt{n}$ and the Cauchy-Schwarz inequality,

$$(A.50) \qquad \int n\underline{d}_{nj}^2/\underline{p}_{nj}\, d\underline{\mu}_{nj} \le 12\int u_{nj}^2\, d\mu_{nj} + 3\varepsilon_n^2 \int u_{nj}^2\, d\mu_{nj} + \varepsilon_n^2 \int \bar{g}_{nj}^2 p_{nj}\, d\mu_{nj}$$

$$(A.51) \qquad \int n\underline{b}_{nj}^2/\, \underline{p}_{nj}\, d\underline{\mu}_{nj} \le \int \bar{g}_{nj}^2 p_{nj}\, d\mu_{nj} \ .$$

By (A.13), (A.45)-(A.51)

$$n^{-1}\Sigma_{j=1}^n \int [\sqrt{n}(\underline{q}_{nj}^{\frac{1}{2}} - \underline{p}_{nj}^{\frac{1}{2}}) - \tfrac{1}{2}\underline{g}_{nj}\underline{p}_{nj}^{\frac{1}{2}}]^2\, 1\{\underline{p}_{nj} > 0\}\, d\underline{\mu}_{nj} \to 0 \ .$$

Let $A_{nj} = \{\underline{p}_{nj} = 0\}$. Then $P_{nj}(t_{nj}^{-1}(A_{nj})) = \underline{P}_{nj}(A_{nj}) = 0$. Thus

$$\mu_{nj}(\, t_{nj}^{-1}(A_{nj}) \cap \{p_{nj} > 0\}\,) = 0 \ .$$

Finally

$$\begin{aligned} &n^{-1}\Sigma_{j=1}^n \int [\sqrt{n}(\underline{q}_{nj}^{\frac{1}{2}} - \underline{p}_{nj}^{\frac{1}{2}}) - \tfrac{1}{2}\underline{g}_{nj}\underline{p}_{nj}^{\frac{1}{2}}]^2\, 1\{\underline{p}_{nj} = 0\} d\underline{\mu}_{nj} \\ &= \Sigma_{j=1}^n\, \underline{Q}_{nj}(\, A_{nj}\,) = \Sigma_{j=1}^n\, Q_{nj}(\, t_{nj}^{-1}(A_{nj})\,) \\ &\le \Sigma_{j=1}^n\, Q_{nj}(\, t_{nj}^{-1}(A_{nj}) \cap \{p_{nj} = 0\}\,) \le \Sigma_{j=1}^n\, Q_{nj}(\, \{p_{nj} = 0\}\,) \to 0 \ , \end{aligned}$$

by (A.11). ■

Finally we show that the continuity property (A.24) also carries over to the transformed score functions. For triangular arrays $(k_{n1},\ldots,k_{nn})$ and

$(\ell_{n1},\ldots,\ell_{nn})$ as in Proposition A.10, set

$$\underline{k}_{nj}(t) = E_{Q_{nj}}(k_{nj}(X_{nj}) \mid T_{nj} = t) \qquad \text{a.e.}$$

$$\underline{\ell}_{nj}(t) = E_{P_{nj}}(\ell_{nj}(X_{nj}) \mid T_{nj} = t) \qquad \text{a.e..}$$

PROPOSITION A.13. *Let* (A.11)-(A.14) *and* (A.24)-(A.26) *hold. Then*

$$n^{-1}\Sigma_{j=1}^{n} \int [\underline{k}_{nj}q_{nj}^{\frac{1}{2}} - \underline{\ell}_{nj}p_{nj}^{\frac{1}{2}}]^2 \, d\underline{\mu}_{nj} \to 0 \ .\square$$

PROOF. Again we may assume without loss of generality that $\underline{\mu}_{nj} = t_{nj}(Q_{nj}+P_{nj})$. For fixed $\varepsilon > 0$ we set $\bar{k}_{nj} = k_{nj}1\{|k_{nj}| \le \varepsilon\sqrt{n}\}$. Using (A.48) and a similar relation for Q, dropping the indices, and reading zero for 0/0, we have

$$|\underline{\bar{k}}\underline{q}^{\frac{1}{2}} - \underline{\ell}\underline{p}^{\frac{1}{2}}| = |E_\mu\{\bar{k}q|T\}\ \underline{q}^{-\frac{1}{2}} - E_\mu\{\ell p|T\}\ \underline{p}^{-\frac{1}{2}}|$$

$$\le |E_\mu\{\bar{k}q^{\frac{1}{2}}\ (q^{\frac{1}{2}}-p^{\frac{1}{2}})|T\}\ \underline{q}^{-\frac{1}{2}}| + |E_\mu\{\bar{k}q^{\frac{1}{2}}p^{\frac{1}{2}}|T\}\ (\underline{q}^{-\frac{1}{2}}-\underline{p}^{-\frac{1}{2}})|$$

$$+ |E_\mu\{(\bar{k}q^{\frac{1}{2}}-\ell p^{\frac{1}{2}})\ p^{\frac{1}{2}}|T\}\ \underline{p}^{-\frac{1}{2}}|$$

$$\le \varepsilon\sqrt{n}\,|[E_\mu\{(q^{\frac{1}{2}}-p^{\frac{1}{2}})^2|T\}]^{\frac{1}{2}} + \varepsilon\sqrt{n}\ |\underline{q}^{\frac{1}{2}}-\underline{p}^{\frac{1}{2}}| + [E_\mu\{(\bar{k}q^{\frac{1}{2}}-\ell p^{\frac{1}{2}})^2|T\}]^{\frac{1}{2}} \ .$$

Thus we obtain by (A.11) and (A.13), (A.38) and (A.41), and (A.24) and (A.26)

$$n^{-1}\Sigma_{j=1}^{n} \int [\underline{\bar{k}}_{nj}q_{nj}^{\frac{1}{2}} - \underline{\ell}_{nj}p_{nj}^{\frac{1}{2}}]^2 \, d\underline{\mu}_{nj} \le \varepsilon^2 O(1) + o(1) \ .$$

Finish the proof by using (A.26) to infer that

$$n^{-1}\Sigma_{j=1}^{n} \int [\underline{\bar{k}}_{nj} - \underline{k}_{nj}]^2 \, d\underline{Q}_{nj} \le n^{-1}\Sigma_{j=1}^{n} \int [\bar{k}_{nj} - k_{nj}]^2 \, dQ_{nj} \to 0. \ \blacksquare$$

LEMMA A.14. *Let* $\{a,b,c,d\} \subset \mathbb{R}$ *with* $a > 0$, $|ba^{-1}| \le \varepsilon < 1$, $c \ge 0$ *and* $a+b+c+d \ge 0$. *Then*

$$((a+b+c+d)^{\frac{1}{2}} - a^{\frac{1}{2}} - \tfrac{1}{2}ba^{-\frac{1}{2}})^2 \le \frac{3d^2}{a(1-\varepsilon)} + 3c + ((1-\varepsilon)^{-\frac{1}{2}}-1)^2 \frac{b^2}{a}. \quad \square$$

PROOF. We use the following inequalities

$$|(x+y)^{\frac{1}{2}} - x^{\frac{1}{2}}| \le |y|x^{-\frac{1}{2}} \qquad (x > 0,\ x+y \ge 0) \tag{A.52}$$

$$(x^{\frac{1}{2}} - y^{\frac{1}{2}})^2 \le |x - y| \qquad (x \ge 0,\ y \ge 0) \tag{A.53}$$

$$|(x+y)^{\frac{1}{2}} - x^{\frac{1}{2}} - \tfrac{1}{2}yx^{-\frac{1}{2}}| \le \tfrac{1}{2}|y|x^{-\frac{1}{2}}((1-\varepsilon)^{-\frac{1}{2}}-1) \qquad (x >0,\ |yx^{-1}| \le \varepsilon < 1). \tag{A.54}$$

Here the first and the third inequality can be deduced from

$$|(1+z)^{\frac{1}{2}} - (1+\tfrac{1}{2}z)| \le \tfrac{1}{2}z^{-\frac{1}{2}} \qquad (z \ge -1)$$

and

$$|(1+z)^{\frac{1}{2}} - (1+\tfrac{1}{2}z)| = \left| \int_0^1 \tfrac{1}{2}(1+uz)^{-\frac{1}{2}}z \, du - \tfrac{1}{2}z \right|$$

$$\le \tfrac{1}{2}|z| \sup_{0\le u\le 1} |(1+uz)^{-\frac{1}{2}} - 1| \le \tfrac{1}{2}|z|\,((1-\varepsilon)^{-\frac{1}{2}}-1) \qquad (|z| \le \varepsilon),$$

respectively. The lemma follows from (A.52)-(A.54) and

$$|(a+b+c+d)^{\frac{1}{2}} - a^{\frac{1}{2}} - \tfrac{1}{2}ba^{-\frac{1}{2}}|^2 \le 3|(a+b+c+d)^{\frac{1}{2}} - (a+b+c)^{\frac{1}{2}}|^2$$

$$+ 3|(a+b+c)^{\frac{1}{2}} - (a+b)^{\frac{1}{2}}|^2 + 3|(a+b)^{\frac{1}{2}} - a^{\frac{1}{2}} - \tfrac{1}{2}ba^{-\frac{1}{2}}|^2. \quad \blacksquare$$

REFERENCES

Andersen, E.B. (1970). Asymptotic properties of conditional maximum likelihood estimators. *Journal Royal Statist. Soc. B*, 283-301.

Andersen, N.T., Dobric, V. (1987). The central limit theorem for stochastic processes. *Ann. Probab. 15*, 164-177.

Anderson, R. (1955). The integral of a symmetric unimodal function. *Proc. Amer. Math. Soc. 6*, 170-176.

Araujo, A., Giné, E. (1980). *The Central Limit Theorem for Real and Banach Valued Random Variables*, John Wiley, New York.

Bauer, H. (1981). *Probability Theory and Elements of Measure Theory*, Academic Press, London.

Begun, J.M., Hall, W.J., Huang, W.M. and Wellner J. (1983). Information and asymptotic efficiency in parametric-nonparametric models. *Ann. Statist. 11*, 432-452.

Beran, R. (1974). Asymptotically efficient adaptive rank estimates in location models. *Ann. Statist. 2*, 63-74.

Beran, R. (1977). Estimating a distribution function. *Ann. Statist. 5*, 400-404.

Bickel P.J. (1982). On adaptive estimation. *Ann. Statist. 10*, 647-671.

Bickel P.J., Klaassen C.A.J. (1986). Empirical Bayes estimation in functional and structural models, and uniformly adaptive estimation of location. *Adv. Appl. Math. 7*, 55-69.

Bickel P.J., Ritov, Y. (1987). Efficient Estimation in the Errors in Variables Model. *Ann. Statist. 15*, 513-540.

Bickel, P.J., Klaassen, C.A.J., Ritov, Y., Wellner, J.A.. (198?). *Efficient and Adaptive Inference in Semi-Parametric Models.*

Billingsley, P. (1968), *Convergence of Probability Measures*, Wiley, New York.

Breslow, N., Crowley, J. (1974). A large sample study of the life table and product limit estimates under random censorship. *Ann. Statist. 2*, 437-453.

Droste, W., Wefelmeyer, W. (1984). On Hájek's convolution theorem. *Statist. Decisions 2*, 131-144.

Dudley, R.M. (1966). Weak convergence of measures on nonseparable metric spaces and empirical measures on euclidean spaces. *Illinois J. Math. 10*, 109-126.

Dudley, R.M. (1967). Measures on non-separable metric spaces. *Illinois J. Math. 11*, 449-453.

Dudley, R.M. (1984). A course on empirical processes. P.L. Hennequin (ed.) *Ecole d'Eté de Probabilités de Saint-Flour XII-1982*, Lecture Notes Math. 1097, Springer-Verlag, Berlin, 2-142.

Eeden, C. van (1970). Efficiency-robust estimation of location. *Ann. Math. Statist. 41*, 172-181.

Fernholz, L.T. (1983). *Von Mises Calculus for Statistical Functionals*, Lecture Notes in Statistics 19, Springer Verlag, New York.

Gaenssler, P. (1983). *Empirical Processes*, Institute of Mathematical Statistics, Hayward, California.

Gihman, I.I., Skorohod, A.V. (1974). *The Theory of Stochastic Processes I*, Springer Verlag, Berlin.

Gill, R.D. (1980). *Censoring and Stochastic Integrals*, MC Tract 124. Math. Centrum, Amsterdam.

Gill, R.D. (1986). *Non- and Semi-Parametric Maximum Likelihood Estimators and the von Mises Method*, Report MS-R8604, CWI, Amsterdam.

Gill, R.D., Johansen, S. (1987). *Product Integrals and Counting Processes*. Report MS-R8707, CWI, Amsterdam.

Hájek, J. (1970). A characterization of limiting distributions of regular estimators. *Z. Wahrsch. Verw. Gebiete 14*, 323-330.

Hájek, J. (1972). Local asymptotic minimax and admissibility in estimation. *Proc. Sixth Berkeley Symp. Math. Statist. Probab. 1*, University of California Press, Berkeley, 175-194.

Hasminski, R.Z., Nussbaum, M. (1984). An asymptotic minimax bound in a regression model with an increasing number of nuisance parameters. P. Mandl, M. Huskova (eds.). *Proc. Third Prague Symp. As. Statistics*, Elsevier, Amsterdam, 275-283.

Heckman, J. Singer, B. (1984). A method for minimizing the impact of distributional assumptions in economic studies for duration data. *Econometrica 52*, 271-320.

Helmers, R., Oosterhoff, J., Ruymgaart, F.H., van Zuylen, M.C.A. (1976). *Asymptotische Methoden in de Toetsingstheorie*, MC Syllabus 22, Mathematical Centre, Amsterdam.

Hewitt, E., Stromberg, K. (1965), *Real and Abstract Analysis*, Springer Verlag, Berlin.

Ibragimow, I.A., Has'minskii, R.Z. (1981). *Statistical Estimation; Asymptotic Theory*, Springer Verlag, New York.

Jameson, G.J.O. (1974). *Topology and Normed Spaces*, Chapman and Hall, London.

Keiding, N., Gill, R.D. (1987). *Random Truncation Models and Markov Processes*. Report MS-R8702, CWI, Amsterdam.

Kiefer J., Wolfowitz, J. (1956). Consistency of the maximum likelihood estimator in the presence of infinitely many nuisance parameters. *Ann. Math. Statist. 27*, 887-906.

Klaassen, C.A.J., van der Vaart, A.W., van Zwet, W.R. (1987). On estimating a parameter and its score function II. To appear in S. Gupta (ed.) *Statist. Decision Theory and Rel. Topics IV*.

Klaassen, C.A.J., van Zwet, W.R. (1985). On estimating a parameter and its score function. L.M. Le Cam, R.A. Olshen (eds.). *Proc. Berkeley Conference in Honor of Jerzy Neyman and Jack Kiefer II*, Wadsworth, Belmont, 827-839.

Koshevnik, Yu.A., Levit, B.Ya (1976). On a nonparametric analogue of the information matrix. *Theory Prob. Appl. 21*, 738-753.

Le Cam, L. (1960). Locally asymptotically normal families of distributions. *Univ. California Publ. Statist. 3*, University of California Press, 37-98.

LeCam, L. (1969). *Théorie Asymptotique de la Décision Statistique*, Les Presses de l'Université de Montréal, Montréal.

LeCam, L. (1972). Limits of experiments. *Proc. Sixth Berkeley Symp. Math. Statist. Probab. 1*, University of California Press, Berkeley, 245-261.

Le Cam, L. (1986). *Asymptotic Methods in Statistical Decision Theory*, Springer-Verlag, New York.

Le Cam, L., Yang, G.L. (1986). *On the Preservation of Local Asymptotic Normality under Information Loss.* Tech. Report 53, Depart. Statist., University of California, Berkeley.

Lindsay, B.G. (1983). Efficiency of the conditional score in a mixture setting. *Ann. Statist. 11*, 486-497.

Lindsay, B.G. (1985). Using empirical partially Bayes inference for increased efficiency. *Ann. Statist. 13*, 914-931.

Millar, P.W. (1983). The minimax principle in asymptotic statistical theory. *Ecole d'Eté de Probabilités de Saint-Flour XI-1981*. Lecture Notes Math. 976, Springer Verlag, Berlin, 67-267.

Millar, P.W. (1985). Nonparametric applications of an infinite dimensional convolution theorem. *Z. Wahrsch. Th. Verw. Gebiete 68*, 545-556.

Novinger, W.P. (1972), Mean convergence in L^p spaces. *Proc. Amer. Math. Soc. 34*, 627-628.

Nussbaum, M. (1984). An asymptotic minimax risk bound for estimation of a linear functional relationship. *J. Multivariate Anal. 14*, 300-314.

Oosterhoff, J., van Zwet W.R. (1979). A note on contiguity and Hellinger distance. J. Jureckova (ed.). *Contributions to Statistics (J. Hajek Memorial Volume)*, Academia, Prague, 157-166.

Parthasarathy, K.R. (1967). *Probability Measures on Metric Spaces*, Academic Press, New York and London.

Pfanzagl, J., (1982) (with W. Wefelmeyer). *Contributions to a General Asymptotic Statistical Theory*, Lecture Notes in Statistics 13, Springer Verlag, New York.

Pfanzagl, J., (1985) (with W. Wefelmeyer). *Asymptotic Expansions for General Statistical Models.* Lecture Notes in Statistics 31, Springer Verlag, New York.

Pollard, D. (1984). *Convergence of Stochastic Processes*, Springer-Verlag, New York.

Prokhorov, Yu.V. (1956). Convergence of random processes and limit theorems in probability theory. *Theory Probab. Appl. 1*, 157-214.

Reeds, J.A. (1976). *On the Definition of von Mises Functionals*, Research Report S-44, Dept. of Statistics, University of Harvard.

Roussas, G.G. (1972). *Contiguity of Probability Measures; Some Applications in Statistics*, Cambridge Univ. Press, Cambridge.

Rudin, W. (1966). *Real and Complex Analysis*, McGraw-Hill Inc., New York.

Rudin, W. (1973). *Functional Analysis*, McGraw-Hill, New York.

Sacks, J. (1963). Generalized Bayes solutions in estimation problems. *Ann. Math. Statist. 3*, 751-768.

Stein, Charles (1956). Efficient nonparametric testing and estimation. *Proc. Third Berkeley Symp. Math. Statist. Probab. 1*, University of California Press, Berkeley, 187-195.

Stone, C. (1975). Adaptive maximum likelihood estimation of a location parameter. *Ann. Statist. 3*, 267-284.

Takeuchi, K. (1971). A uniformly asymptotically efficient estimator of a location parameter. *J. Amer. Statist. Assoc. 66*, 292-301.

Vaart, A.W. van der (1986a). *Estimating a Real Parameter in a Class of Semi-Parametric Models*, Report 86-9, Dep. Math., Univ. Leiden, Leiden.

Vaart, A.W. van der (1986b). *On the Asymptotic Information Bound*, Report 86-13, Dep. Math., Univ. Leiden, Leiden.

Vaart, A.W. van der (1987a). *Statistical Estimation in Large Parameter Spaces.* Thesis, Univ. of Leiden.

Vaart, A.W. van der (1987). *Efficient Estimation in the Paired Exponential Model.* Tech. Report. Free Univ. Amsterdam.

Wellner, J.A. (1982). Asymptotic optimality of the product limit estimator. *Ann. Statist. 10*, 595-602.

Wellner, J.A. (1985). Semiparametric models: progress and problems. *Newsletter 9*, CWI, Amsterdam, 1-24.

Witting, H., Nölle G. (1970). *Angewandte Mathematische Statistik*, Teubner, Stuttgart.

SUBJECT INDEX

Anderson's lemma 23,72
P-bridge 76
censoring 104
complete
  L_2- 140
  locally - 118
  strongly locally - 118
conditional maximum likelihood 163
cone 17
($\sqrt{n}$-)consistency 120,124
contiguity 175..
continuation set 143
convergence in distribution 89..,176
convolution 54..
convolution theorem 32..,54..
  generalized - 21..
covariance function 56
covariance operator 74
cumulative hazard function 105
cylinder measure 50
cylinder σ-algebra 50
differentiable functional 18,44,49
differentiability in quadratic mean 181..
dual space 49
efficiency 94..
  marginal 95
elliptic distribution 153
errors in variables 162
estimator sequence
  best regular 36,95
  B'-regular 53
  LAM 36
  $\|\cdot\|$-regular 64
  Π-regular 66
  regular 22,53

Skorohod-regular 70
weakly regular 22,53
functional model 138..
gradient 19,49
canonical 19,49
Hadamard differentiable 92
- tangentially to 92
incidental parameter 138..
influence function 19,49
efficient 19,49
kernel operator 74
LAM 15,36
LAM theorem 26..,58..
LAN 181..,43
linear functional relationship 162
local asymptotic minimax theorem 26..,58..
local asymptotic normality 43,181..
local asymtotic normal model 43
loss function
bowlshaped 26
cylinder 58
subconvex 60
marginal distributions 50
mixture model 138..
Neyman-Scott model 163
paired hazards model 155
parametric model 1
product integral 105
product limit estimator 110
projection σ-field 66,71
relatively compact 91
score (function) 7
efficient - (for θ) 19
efficient - for average density 122
semi-parametric model 19
separable (probability measure) 91,14
separates points 51

standard conditions 94
structural model 138..
subconvex 60
subdistribution function 104..
empirical 109
support 143
symmetric location 10,111,149
in $\mathbb{R}^k$ 111,149
tangent cone 17
maximal 18
third lemma of Le Cam 180
tight
probability measure 14,91
sequence of - 91
$\|\cdot\|$-,τ- 14
uniqueness set 44
weak convergence 89..

SYMBOL INDEX

$T(P)$: tangent cone 17

$T_m(P)$: maximal tangent cone 18

$(dP)^{\frac{1}{2}}$: square root of measure 17

$\dot{\kappa}(\cdot,P)$, $\dot{\kappa}_{b^*}(\cdot,P)$: gradient, influence function 19,49

$\tilde{\kappa}_i(\cdot,P)$, $\tilde{\kappa}_{b^*}(\cdot,P)$: canonical grad., efficient influence function 19,49

$\mathcal{L}(\cdot,\theta,\eta)$: efficient score 19

$\Lambda_n(Q,P)$: log likelihood ratio of Q and P 176

$L_2(P)$: square integrable functions 13

$L_{2^*}(P)$: square integrable functions with zero-mean 13

$\|\cdot\|_P$: norm in $L_2(P)$ 13

$\langle\cdot,\cdot\rangle_P$: inner product in $L_2(P)$ 13

lin C : linear span 13

$\bar{C}$: closure

C[a,b] : continuous real functions 14

D[a,b] : cadlag functions 66

B(T) : uniformly bounded real functions 70

UC(T,ρ) : uniformly continuous real functions 85

$B^*, B_\tau^*, B(T)^*$.. : dual spaces 49

Π : linear space spanned by projections 66,71

π_t : coordinate projection 66,71

$\|\cdot\|_\infty$: supremum norm 14

J : Skorohod topology 70

$\tau(B')$: weak topology generated by B' 14

$\mathcal{U}(\tau)$, $\mathcal{U}(d)$, $\mathcal{U}(\|\cdot\|)$, $\mathcal{U}(\tau(B'))$: Borel σ-field 14

$\mathcal{U}$(d-balls) : σ-field generated by closed balls 19

$\mathcal{U}(B')$: σ-field generated by B' 50

$\mathcal{U}(\Pi)$: projection σ-field 66,71

(α_{ij}) : matrix 14

$(A)_{ij}$: element of matrix A 14

δ_z : measure degenerated at a point 14

$N(\mu,\Sigma)$, $N_k(\mu,\Sigma)$: normal measure on $\mathbb{R}^k$

$b'(M)$, $\pi_{u_1\ldots u_m}(L)$: induced measures 14

$\mathcal{L}(G)$: law of the random element G 14

of product measures 21

$\mathcal{R}(B',N)$: loss functions 58

$d_1 v \ldots v d_k$, $\|\cdot\|_\infty v \|\cdot\|_\infty$, $\|\cdot\|_\infty v |\cdot|$: measure or norm on product space 98

MATHEMATICAL CENTRE TRACTS

1 T. van der Walt. *Fixed and almost fixed points.* 1963.

2 A.R. Bloemena. *Sampling from a graph.* 1964.

3 G. de Leve. *Generalized Markovian decision processes, part I: model and method.* 1964.

4 G. de Leve. *Generalized Markovian decision processes, part II: probabilistic background.* 1964.

5 G. de Leve, H.C. Tijms, P.J. Weeda. *Generalized Markovian decision processes, applications.* 1970.

6 M.A. Maurice. *Compact ordered spaces.* 1964.

7 W.R. van Zwet. *Convex transformations of random variables.* 1964.

8 J.A. Zonneveld. *Automatic numerical integration.* 1964.

9 P.C. Baayen. *Universal morphisms.* 1964.

10 E.M. de Jager. *Applications of distributions in mathematical physics.* 1964.

11 A.B. Paalman-de Miranda. *Topological semigroups.* 1964.

12 J.A.Th.M. van Berckel, H. Brandt Corstius, R.J. Mokken, A. van Wijngaarden. *Formal properties of newspaper Dutch.* 1965.

13 H.A. Lauwerier. *Asymptotic expansions.* 1966, out of print; replaced by MCT 54.

14 H.A. Lauwerier. *Calculus of variations in mathematical physics.* 1966.

15 R. Doornbos. *Slippage tests.* 1966.

16 J.W. de Bakker. *Formal definition of programming languages with an application to the definition of ALGOL 60.* 1967.

17 R.P. van de Riet. *Formula manipulation in ALGOL 60, part 1.* 1968.

18 R.P. van de Riet. *Formula manipulation in ALGOL 60, part 2.* 1968.

19 J. van der Slot. *Some properties related to compactness.* 1968.

20 P.J. van der Houwen. *Finite difference methods for solving partial differential equations.* 1968.

21 E. Wattel. *The compactness operator in set theory and topology.* 1968.

22 T.J. Dekker. *ALGOL 60 procedures in numerical algebra, part 1.* 1968.

23 T.J. Dekker, W. Hoffmann. *ALGOL 60 procedures in numerical algebra, part 2.* 1968.

24 J.W. de Bakker. *Recursive procedures.* 1971.

25 E.R. Paërl. *Representations of the Lorentz group and projective geometry.* 1969.

26 European Meeting 1968. *Selected statistical papers, part I.* 1968.

27 European Meeting 1968. *Selected statistical papers, part II.* 1968.

28 J. Oosterhoff. *Combination of one-sided statistical tests.* 1969.

29 J. Verhoeff. *Error detecting decimal codes.* 1969.

30 H. Brandt Corstius. *Exercises in computational linguistics.* 1970.

31 W. Molenaar. *Approximations to the Poisson, binomial and hypergeometric distribution functions.* 1970.

32 L. de Haan. *On regular variation and its application to the weak convergence of sample extremes.* 1970.

33 F.W. Steutel. *Preservation of infinite divisibility under mixing and related topics.* 1970.

34 I. Juhász, A. Verbeek, N.S. Kroonenberg. *Cardinal functions in topology.* 1971.

35 M.H. van Emden. *An analysis of complexity.* 1971.

36 J. Grasman. *On the birth of boundary layers.* 1971.

37 J.W. de Bakker, G.A. Blaauw, A.J.W. Duijvestijn, E.W. Dijkstra, P.J. van der Houwen, G.A.M. Kamsteeg-Kemper, F.E.J. Kruseman Aretz, W.L. van der Poel, J.P. Schaap-Kruseman, M.V. Wilkes, G. Zoutendijk. *MC-25 Informatica Symposium.* 1971.

38 W.A. Verloren van Themaat. *Automatic analysis of Dutch compound words.* 1972.

39 H. Bavinck. *Jacobi series and approximation.* 1972.

40 H.C. Tijms. *Analysis of (s,S) inventory models.* 1972.

41 A. Verbeek. *Superextensions of topological spaces.* 1972.

42 W. Vervaat. *Success epochs in Bernoulli trials (with applications in number theory).* 1972.

43 F.H. Ruymgaart. *Asymptotic theory of rank tests for independence.* 1973.

44 H. Bart. *Meromorphic operator valued functions.* 1973.

45 A.A. Balkema. *Monotone transformations and limit laws.* 1973.

46 R.P. van de Riet. *ABC ALGOL, a portable language for formula manipulation systems, part 1: the language.* 1973.

47 R.P. van de Riet. *ABC ALGOL, a portable language for formula manipulation systems, part 2: the compiler.* 1973.

48 F.E.J. Kruseman Aretz, P.J.W. ten Hagen, H.L. Oudshoorn. *An ALGOL 60 compiler in ALGOL 60, text of the MC-compiler for the EL-X8.* 1973.

49 H. Kok. *Connected orderable spaces.* 1974.

50 A. van Wijngaarden, B.J. Mailloux, J.E.L. Peck, C.H.A. Koster, M. Sintzoff, C.H. Lindsey, L.G.L.T. Meertens, R.G. Fisker (eds.). *Revised report on the algorithmic language ALGOL 68.* 1976.

51 A. Hordijk. *Dynamic programming and Markov potential theory.* 1974.

52 P.C. Baayen (ed.). *Topological structures.* 1974.

53 M.J. Faber. *Metrizability in generalized ordered spaces.* 1974.

54 H.A. Lauwerier. *Asymptotic analysis, part 1.* 1974.

55 M. Hall, Jr., J.H. van Lint (eds.). *Combinatorics, part 1: theory of designs, finite geometry and coding theory.* 1974.

56 M. Hall, Jr., J.H. van Lint (eds.). *Combinatorics, part 2: graph theory, foundations, partitions and combinatorial geometry.* 1974.

57 M. Hall, Jr., J.H. van Lint (eds.). *Combinatorics, part 3: combinatorial group theory.* 1974.

58 W. Albers. *Asymptotic expansions and the deficiency concept in statistics.* 1975.

59 J.L. Mijnheer. *Sample path properties of stable processes.* 1975.

60 F. Göbel. *Queueing models involving buffers.* 1975.

63 J.W. de Bakker (ed.). *Foundations of computer science.* 1975.

64 W.J. de Schipper. *Symmetric closed categories.* 1975.

65 J. de Vries. *Topological transformation groups, 1: a categorical approach.* 1975.

66 H.G.J. Pijls. *Logically convex algebras in spectral theory and eigenfunction expansions.* 1976.

68 P.P.N. de Groen. *Singularly perturbed differential operators of second order.* 1976.

69 J.K. Lenstra. *Sequencing by enumerative methods.* 1977.

70 W.P. de Roever, Jr. *Recursive program schemes: semantics and proof theory.* 1976.

71 J.A.E.E. van Nunen. *Contracting Markov decision processes.* 1976.

72 J.K.M. Jansen. *Simple periodic and non-periodic Lamé functions and their applications in the theory of conical waveguides.* 1977.

73 D.M.R. Leivant. *Absoluteness of intuitionistic logic.* 1979.

74 H.J.J. te Riele. *A theoretical and computational study of generalized aliquot sequences.* 1976.

75 A.E. Brouwer. *Treelike spaces and related connected topological spaces.* 1977.

76 M. Rem. *Associons and the closure statement.* 1976.

77 W.C.M. Kallenberg. *Asymptotic optimality of likelihood ratio tests in exponential families.* 1978.

78 E. de Jonge, A.C.M. van Rooij. *Introduction to Riesz spaces.* 1977.

79 M.C.A. van Zuijlen. *Emperical distributions and rank statistics.* 1977.

80 P.W. Hemker. *A numerical study of stiff two-point boundary problems.* 1977.

81 K.R. Apt, J.W. de Bakker (eds.). *Foundations of computer science II, part 1.* 1976.

82 K.R. Apt, J.W. de Bakker (eds.). *Foundations of computer science II, part 2.* 1976.

83 L.S. van Benthem Jutting. *Checking Landau's "Grundlagen" in the AUTOMATH system.* 1979.

84 H.L.L. Busard. *The translation of the elements of Euclid from the Arabic into Latin by Hermann of Carinthia (?), books vii-xii.* 1977.

85 J. van Mill. *Supercompactness and Wallman spaces.* 1977.

86 S.G. van der Meulen, M. Veldhorst. *Torrix I, a programming system for operations on vectors and matrices over arbitrary fields and of variable size.* 1978.

88 A. Schrijver. *Matroids and linking systems.* 1977.

89 J.W. de Roever. *Complex Fourier transformation and analytic functionals with unbounded carriers.* 1978.

90 L.P.J. Groenewegen. *Characterization of optimal strategies in dynamic games.* 1981.

91 J.M. Geysel. *Transcendence in fields of positive characteristic.* 1979.

92 P.J. Weeda. *Finite generalized Markov programming.* 1979.

93 H.C. Tijms, J. Wessels (eds.). *Markov decision theory.* 1977.

94 A. Bijlsma. *Simultaneous approximations in transcendental number theory.* 1978.

95 K.M. van Hee. *Bayesian control of Markov chains.* 1978.

96 P.M.B. Vitányi. *Lindenmayer systems: structure, languages, and growth functions.* 1980.

97 A. Federgruen. *Markovian control problems; functional equations and algorithms.* 1984.

98 R. Geel. *Singular perturbations of hyperbolic type.* 1978.

99 J.K. Lenstra, A.H.G. Rinnooy Kan, P. van Emde Boas (eds.). *Interfaces between computer science and operations research.* 1978.

100 P.C. Baayen, D. van Dulst, J. Oosterhoff (eds.). *Proceedings bicentennial congress of the Wiskundig Genootschap, part 1.* 1979.

101 P.C. Baayen, D. van Dulst, J. Oosterhoff (eds.). *Proceedings bicentennial congress of the Wiskundig Genootschap, part 2.* 1979.

102 D. van Dulst. *Reflexive and superreflexive Banach spaces.* 1978.

103 K. van Harn. *Classifying infinitely divisible distributions by functional equations.* 1978.

104 J.M. van Wouwe. *Go-spaces and generalizations of metrizability.* 1979.

105 R. Helmers. *Edgeworth expansions for linear combinations of order statistics.* 1982.

106 A. Schrijver (ed.). *Packing and covering in combinatorics.* 1979.

107 C. den Heijer. *The numerical solution of nonlinear operator equations by imbedding methods.* 1979.

108 J.W. de Bakker, J. van Leeuwen (eds.). *Foundations of computer science III, part 1.* 1979.

109 J.W. de Bakker, J. van Leeuwen (eds.). *Foundations of computer science III, part 2.* 1979.

110 J.C. van Vliet. *ALGOL 68 transput, part I: historical review and discussion of the implementation model.* 1979.

111 J.C. van Vliet. *ALGOL 68 transput, part II: an implementation model.* 1979.

112 H.C.P. Berbee. *Random walks with stationary increments and renewal theory.* 1979.

113 T.A.B. Snijders. *Asymptotic optimality theory for testing problems with restricted alternatives.* 1979.

114 A.J.E.M. Janssen. *Application of the Wigner distribution to harmonic analysis of generalized stochastic processes.* 1979.

115 P.C. Baayen, J. van Mill (eds.). *Topological structures II, part 1.* 1979.

116 P.C. Baayen, J. van Mill (eds.). *Topological structures II, part 2.* 1979.

117 P.J.M. Kallenberg. *Branching processes with continuous state space.* 1979.

118 P. Groeneboom. *Large deviations and asymptotic efficiencies.* 1980.

119 F.J. Peters. *Sparse matrices and substructures, with a novel implementation of finite element algorithms.* 1980.

120 W.P.M. de Ruyter. *On the asymptotic analysis of large-scale ocean circulation.* 1980.

121 W.H. Haemers. *Eigenvalue techniques in design and graph theory.* 1980.

122 J.C.P. Bus. *Numerical solution of systems of nonlinear equations.* 1980.

123 I. Yuhász. *Cardinal functions in topology - ten years later.* 1980.

124 R.D. Gill. *Censoring and stochastic integrals.* 1980.

125 R. Eising. *2-D systems, an algebraic approach.* 1980.

126 G. van der Hoek. *Reduction methods in nonlinear programming.* 1980.

127 J.W. Klop. *Combinatory reduction systems.* 1980.

128 A.J.J. Talman. *Variable dimension fixed point algorithms and triangulations.* 1980.

129 G. van der Laan. *Simplicial fixed point algorithms.* 1980.

130 P.J.W. ten Hagen, T. Hagen, P. Klint, H. Noot, H.J. Sint, A.H. Veen. *ILP: intermediate language for pictures.* 1980.

131 R.J.R. Back. *Correctness preserving program refinements: proof theory and applications.* 1980.

132 H.M. Mulder. *The interval function of a graph.* 1980.

133 C.A.J. Klaassen. *Statistical performance of location estimators.* 1981.

134 J.C. van Vliet, H. Wupper (eds.). *Proceedings international conference on ALGOL 68.* 1981.

135 J.A.G. Groenendijk, T.M.V. Janssen, M.J.B. Stokhof (eds.). *Formal methods in the study of language, part I.* 1981.

136 J.A.G. Groenendijk, T.M.V. Janssen, M.J.B. Stokhof (eds.). *Formal methods in the study of language, part II.* 1981.

137 J. Telgen. *Redundancy and linear programs.* 1981.

138 H.A. Lauwerier. *Mathematical models of epidemics.* 1981.

139 J. van der Wal. *Stochastic dynamic programming, successive approximations and nearly optimal strategies for Markov decision processes and Markov games.* 1981.

140 J.H. van Geldrop. *A mathematical theory of pure exchange economies without the no-critical-point hypothesis.* 1981.

141 G.E. Welters. *Abel-Jacobi isogenies for certain types of Fano threefolds.* 1981.

142 H.R. Bennett, D.J. Lutzer (eds.). *Topology and order structures, part 1.* 1981.

143 J.M. Schumacher. *Dynamic feedback in finite- and infinite-dimensional linear systems.* 1981.

144 P. Eijgenraam. *The solution of initial value problems using interval arithmetic; formulation and analysis of an algorithm.* 1981.

145 A.J. Brentjes. *Multi-dimensional continued fraction algorithms.* 1981.

146 C.V.M. van der Mee. *Semigroup and factorization methods in transport theory.* 1981.

147 H.H. Tigelaar. *Identification and informative sample size.* 1982.

148 L.C.M. Kallenberg. *Linear programming and finite Markovian control problems.* 1983.

149 C.B. Huijsmans, M.A. Kaashoek, W.A.J. Luxemburg, W.K. Vietsch (eds.). *From A to Z, proceedings of a symposium in honour of A.C. Zaanen.* 1982.

150 M. Veldhorst. *An analysis of sparse matrix storage schemes.* 1982.

151 R.J.M.M. Does. *Higher order asymptotics for simple linear rank statistics.* 1982.

152 G.F. van der Hoeven. *Projections of lawless sequences.* 1982.

153 J.P.C. Blanc. *Application of the theory of boundary value problems in the analysis of a queueing model with paired services.* 1982.

154 H.W. Lenstra, Jr., R. Tijdeman (eds.). *Computational methods in number theory, part I.* 1982.

155 H.W. Lenstra, Jr., R. Tijdeman (eds.). *Computational methods in number theory, part II.* 1982.

156 P.M.G. Apers. *Query processing and data allocation in distributed database systems.* 1983.

157 H.A.W.M. Kneppers. *The covariant classification of two-dimensional smooth commutative formal groups over an algebraically closed field of positive characteristic.* 1983.

158 J.W. de Bakker, J. van Leeuwen (eds.). *Foundations of computer science IV, distributed systems, part 1.* 1983.

159 J.W. de Bakker, J. van Leeuwen (eds.). *Foundations of computer science IV, distributed systems, part 2.* 1983.

160 A. Rezus. *Abstract AUTOMATH.* 1983.

161 G.F. Helminck. *Eisenstein series on the metaplectic group, an algebraic approach.* 1983.

162 J.J. Dik. *Tests for preference.* 1983.

163 H. Schippers. *Multiple grid methods for equations of the second kind with applications in fluid mechanics.* 1983.

164 F.A. van der Duyn Schouten. *Markov decision processes with continuous time parameter.* 1983.

165 P.C.T. van der Hoeven. *On point processes.* 1983.

166 H.B.M. Jonkers. *Abstraction, specification and implementation techniques, with an application to garbage collection.* 1983.

167 W.H.M. Zijm. *Nonnegative matrices in dynamic programming.* 1983.

168 J.H. Evertse. *Upper bounds for the numbers of solutions of diophantine equations.* 1983.

169 H.R. Bennett, D.J. Lutzer (eds.). *Topology and order structures, part 2.* 1983.

CWI TRACTS

1 D.H.J. Epema. *Surfaces with canonical hyperplane sections.* 1984.

2 J.J. Dijkstra. *Fake topological Hilbert spaces and characterizations of dimension in terms of negligibility.* 1984.

3 A.J. van der Schaft. *System theoretic descriptions of physical systems.* 1984.

4 J. Koene. *Minimal cost flow in processing networks, a primal approach.* 1984.

5 B. Hoogenboom. *Intertwining functions on compact Lie groups.* 1984.

6 A.P.W. Böhm. *Dataflow computation.* 1984.

7 A. Blokhuis. *Few-distance sets.* 1984.

8 M.H. van Hoorn. *Algorithms and approximations for queueing systems.* 1984.

9 C.P.J. Koymans. *Models of the lambda calculus.* 1984.

10 C.G. van der Laan, N.M. Temme. *Calculation of special functions: the gamma function, the exponential integrals and error-like functions.* 1984.

11 N.M. van Dijk. *Controlled Markov processes; time-discretization.* 1984.

12 W.H. Hundsdorfer. *The numerical solution of nonlinear stiff initial value problems: an analysis of one step methods.* 1985.

13 D. Grune. *On the design of ALEPH.* 1985.

14 J.G.F. Thiemann. *Analytic spaces and dynamic programming: a measure theoretic approach.* 1985.

15 F.J. van der Linden. *Euclidean rings with two infinite primes.* 1985.

16 R.J.P. Groothuizen. *Mixed elliptic-hyperbolic partial differential operators: a case-study in Fourier integral operators.* 1985.

17 H.M.M. ten Eikelder. *Symmetries for dynamical and Hamiltonian systems.* 1985.

18 A.D.M. Kester. *Some large deviation results in statistics.* 1985.

19 T.M.V. Janssen. *Foundations and applications of Montague grammar, part 1: Philosophy, framework, computer science.* 1986.

20 B.F. Schriever. *Order dependence.* 1986.

21 D.P. van der Vecht. *Inequalities for stopped Brownian motion.* 1986.

22 J.C.S.P. van der Woude. *Topological dynamix.* 1986.

23 A.F. Monna. *Methods, concepts and ideas in mathematics: aspects of an evolution.* 1986.

24 J.C.M. Baeten. *Filters and ultrafilters over definable subsets of admissible ordinals.* 1986.

25 A.W.J. Kolen. *Tree network and planar rectilinear location theory.* 1986.

26 A.H. Veen. *The misconstrued semicolon: Reconciling imperative languages and dataflow machines.* 1986.

27 A.J.M. van Engelen. *Homogeneous zero-dimensional absolute Borel sets.* 1986.

28 T.M.V. Janssen. *Foundations and applications of Montague grammar, part 2: Applications to natural language.* 1986.

29 H.L. Trentelman. *Almost invariant subspaces and high gain feedback.* 1986.

30 A.G. de Kok. *Production-inventory control models: approximations and algorithms.* 1987.

31 E.E.M. van Berkum. *Optimal paired comparison designs for factorial experiments.* 1987.

32 J.H.J. Einmahl. *Multivariate empirical processes.* 1987.

33 O.J. Vrieze. *Stochastic games with finite state and action spaces.* 1987.

34 P.H.M. Kersten. *Infinitesimal symmetries: a computational approach.* 1987.

35 M.L. Eaton. *Lectures on topics in probability inequalities.* 1987.

36 A.H.P. van der Burgh, R.M.M. Mattheij (eds.). *Proceedings of the first international conference on industrial and applied mathematics (ICIAM 87).* 1987.

37 L. Stougie. *Design and analysis of algorithms for stochastic integer programming.* 1987.

38 J.B.G. Frenk. *On Banach algebras, renewal measures and regenerative processes.* 1987.

39 H.J.M. Peters, O.J. Vrieze (eds.). *Surveys in game theory and related topics.* 1987.

40 J.L. Geluk, L. de Haan. *Regular variation, extensions and Tauberian theorems.* 1987.

41 Sape J. Mullender (ed.). *The Amoeba distributed operating system: Selected papers 1984-1987.* 1987.

42 P.R.J. Asveld, A. Nijholt (eds.). *Essays on concepts, formalisms, and tools.* 1987.

43 H.L. Bodlaender. *Distributed computing: structure and complexity.* 1987.

44 A.W. van der Vaart. *Statistical estimation in large parameter spaces.* 1988.