International Statistical Review (1991), **59**, 1, pp. 97–121. Printed in Great Britain

An Asymptotic Representation Theorem

Aad van der Vaart

Faculteit Wiskunde en Informatica, Vrije Universiteit, De Boelelaan 1081, 1081 HV Amsterdam, The Netherlands

Summary

Let a sequence of statistical experiments converge to a limit experiment in the sense of Le Cam (1972). Furthermore assume that a sequence of statistics possesses a limit distribution under every statistical parameter. Then its set of limit distributions is also the set of distributions of some randomized estimator in the limit experiment. This is a simple way of saying that the limit experiment is a 'lower bound' for the converging sequence of experiments. Moreover, convolution and minimax theorems can be obtained as corollaries.

Key words: Asymptotic representation theorem; Convolution theorem; Experiment; Minimax theorem.

1 Introduction

This paper discusses limit theorems that give bounds on the asymptotic performance of estimators and tests The results are based on Le Cam's (1964, 1972) notion of (weak) convergence of experiments and the main result is a version of Proposition 8 of Le Cam (1972). However, functional analytic notions, such as *weak*-$*$ convergence or transitions, are avoided in favour of classical statistical notions as weak convergence and randomized estimators. It is proposed to use limit *distributions* rather than risks as the criterium to judge asymptotic performance. This makes it possible to formulate some consequences of Le Cam's theory of experiments for limit theorems in a fairly simple manner.

This review paper is partly motivated by recent developments in semi-parametric statistics. Following Begun et al. (1983), Pfanzagl (1982) and Millar (1983, 1985) results on asymptotic efficiency have been successfully obtained for a wealth of semi-parametric models. For an overview see e.g. Bickel et al. (1990). Most of these results concern 'regular' functionals of the parameter, the ones that belong to the $\sqrt{n}$ domain. In the terminology of this paper they concern situations where the appropriate limit experiment is a Gaussian location model with the expectation vector ranging over a linear space (the 'tangent space'). Now there are many examples of 'non-regular' functionals for which a similar approach is impossible, or at best not useful. See e.g. the discussion in van der Vaart (1991). In such cases one may hope to base asymptotic lower bound theorems on either approximations by non-Gaussian experiments, or by Gaussian experiments with a restricted parameter space. (Some examples appear in Groeneboom (1987), Barndorff-Nielsen, James & Leigh (1988) and van der Vaart (1989).) To gain insight into these types of models the proposal of Le Cam (1972) of 'passing to the limit first and then arguing the case for the limiting problem' may be very appropriate. (The question as to how to pass to which limit needs considerable attention. This would be a matter of extending the idea of a 'least favourable submodel' to nonregular cases.) The present paper is aimed at presenting Le Cam's approach in an easily accessible manner, reminiscent of the manner in which results for the regular case are usually stated.

The key result is an *asymptotic representation theorem*. Suppose that a sequence (or net) of estimators has a limit distribution under every parameter. Then there exists a randomized estimator in the limit experiment which has precisely the set of limit distributions as its set of distributions. Thus if one considers the asymptotic behaviour of a sequence of estimators as completely determined by its limit distributions, then in the limit experiment one can always do at least as good as along the sequence.

The representation theorem leads to a simple proof of the Hájek–Le Cam *convolution theorem*. Suppose that a standardized estimator sequence has the same limit distribution under every parameter; i.e. it is *regular*. Then the matching randomized estimator in the limit experiment must be equivariant in law. Thus if the limit experiment has the property that the law of every equivariant-in-law randomized estimator is a certain convolution, then the limit law of a regular estimator sequence must be a certain convolution, too. A shift experiment on Euclidean space is an example having this property.

Asymptotic minimax theorems can be obtained as a consequence of the representation theorem too. Initial versions are restricted to estimator sequences which are tight under every parameter separately. The liminf of the maximum risk is then larger than the minimax risk in the limit experiment, computed over all randomized estimators. For a large class of loss functions it is easy to see that the tightness condition may be dropped.

Most of the terminology of the paper is taken from estimation theory. However, the representation theorem applies equally well to *tests*. Suppose a sequence of power functions converges pointwise to a limit. Then this limit function is necessarily a power function of a test in the limit experiment. Also, if the corresponding sequence of tests is asymptotically of level α, then the matching test in the limit experiment is of level α too. As a consequence the power envelope function of the limit experiment is an upper bound for the pointwise limsup of a sequence of power functions of level α tests.

The organization of the paper is as follows. Convergence of experiments is introduced in § 2. The asymptotic representation theorem is stated in § 3. Some of its corollaries are stated and proved in §§ 4–7. Next follow some examples of limit experiments, without proofs, in § 8. Finally, after discussing some more technical results in § 9, the proof of the representation theorem is given in § 10.

Some special notation is as follows. On any metric space $\mathscr{B}_o$ denotes the Borel σ-field: the smallest σ-field containing the open sets. If $\mathbf{D}$ is a metric space $C_b(\mathbf{D})$ denotes the set of continuous and bounded functions $f:\mathbf{D}\to\mathbb{R}$. For a finite set I, $|I|$ denotes its number of elements, and $\mathbb{R}^I=\mathbb{R}^{|I|}$. Given an arbitrary real function T on a probability space $(\mathscr{X}, \mathscr{B}, P)$, E^*T denotes outer integral:

$$E^*T=\inf\{EY: Y\geq T, Y:(\mathscr{X}, \mathscr{B})\to\mathbb{R}, \text{ measurable}\}.$$

Furthermore, E_* denotes inner integral, P^* is outer and P_* inner probability. Finally ε_v is the probability distribution degenerated at v.

It turns out that *nets* are useful in many arguments. For completeness let us briefly recall their definition and properties. A *directed set* A is a partially ordered set such that given α_1 and α_2 in A, there exists an α_3 in A with $\alpha_1\leq\alpha_3$ and $\alpha_2\leq\alpha_3$. A *net* $\{x_\alpha:\alpha\in A\}$ is a subset of a set X indexed by a directed set A. A net $\{x_\alpha\}$ in a topological space converges to a point x, if for every open neighbourhood G of x, there is an α_0 such that $x_\alpha\in G$ for every $\alpha\geq\alpha_0$. A *subnet* $\{x_{\alpha(\beta)}:\beta\in B\}$ is a subset of $\{x_\alpha:\alpha\in A\}$ which is a net itself, of course, and moreover has the property that for every $\alpha_0\in A$ there is a $\beta_0\in B$ with $\alpha(\beta)\geq\alpha_0$ for every $\beta\geq\beta_0$ (that is $\alpha(\beta)$ is 'eventually past' α_0 for every α_0). (The map $\beta\to\alpha(\beta)$ need not be one-to-one (nor onto).) A sequence is clearly a net and a subsequence a subnet, but a subnet of a sequence is not necessarily a subsequence. It holds that:

(i) a set F is closed if and only if it contains the limits of every convergent net $\{x_\alpha : \alpha \in A\} \subset F$,
(ii) a set K is compact if and only if every net $\{x_\alpha : \alpha \in A\} \subset K$ has a converging subnet with limit in K.

See e.g. Kelley (1955) for a proof of these assertions. If the topological space does not satisfy the 'first axiom of countability' (there is not for every point a countable base for the neighbourhoods of that point), then the assertions can be false if 'net' is replaced by 'sequence'. Such topological spaces will be encountered frequently in the paper. For instance, when the product of uncountably many metric spaces is equipped with the product topology.

2 Convergence of experiments

An *experiment* $\mathscr{E}$ will here be understood to be a collection of probability measures $\{P_h : h \in H\}$ on a measurable space $(\mathscr{X}, \mathscr{B})$, the *sample space*. The *parameter set* H may be arbitrary. The experiment $\mathscr{E}$ is *dominated* if there exists a σ-finite measure μ on $(\mathscr{X}, \mathscr{B})$ with $P_h \ll \mu$ for all $h \in H$.

There are several equivalent ways of saying that a net (or a sequence) of experiments. $\mathscr{E}_\alpha = (\mathscr{X}_\alpha, \mathscr{B}_\alpha, P_{\alpha,h} : h \in H)$ converges to a limit experiment $\mathscr{E} = (\mathscr{X}, \mathscr{B}, P_h : h \in H)$ in the sense introduced by Le Cam. Note that it is assumed that there is a common parameter set H, but that the sample space may be different each time. Le Cam (1964, 1972, 1986) initially defines convergence in terms of a distance measure between experiments. Though this 'distance' is fundamental in the theory of comparison of experiments, it is, perhaps, somewhat too involved for the present paper. A definition that is easy to work with in applications is in terms of the *likelihood ratio processes*.

Say that $\mathscr{E}_\alpha$ *converges to* $\mathscr{E}$ if

$$\mathscr{L}_{h_0}\left(\left(\frac{dP_{\alpha,h}}{dP_{\alpha,h_0}}\right)_{h \in I}\right) \to \mathscr{L}_{h_0}\left(\left(\frac{dP_h}{dP_{h_0}}\right)_{h \in I}\right), \tag{2.1}$$

for every finite subset $I \subset H$ and every $h_0 \in H$. Here the likelihood ratio processes can be formed as follows. One takes densities $p_{\alpha,h}$ and p_{α,h_0} of $P_{\alpha,h}$ and P_{α,h_0} with respect to some measure (for instance a σ-finite measure that dominates both), and sets

$$\frac{dP_{\alpha,h}}{dP_{\alpha,h_0}} = \frac{p_{\alpha,h}}{p_{\alpha,h_0}}.$$

Similarly for $\mathscr{E}$. It can be seen that the quotient that is formed in this manner is up to P_{α,h_0}-equivalence independent of the choice of the dominating measure. (In fact, the quotient equals the Radon–Nikodym density of the absolute continuous part of $P_{\alpha,h}$ with respect to P_{α,h_0}.) Furthermore, (2.1) is the usual weak convergence of Borel laws on $\mathbb{R}^I$. In words (2.1) may be expressed as: marginal weak convergence of the likelihood ratio processes.

Note that for $\mathscr{E}_\alpha$ to converge to $\mathscr{E}$ it is required that (2.1) holds for *every* h_0. Often pairs of nets $\{P_{\alpha,h}\}$ and $\{P_{\alpha,h_0}\}$ will be contiguous. Then it suffices to check (2.1) for just one fixed $h_0 \in H$. More precisely, if one has one-sided contiguity,

$$\{P_{\alpha,h}\} \lhd \{P_{\alpha,h_0}\}, \tag{2.2}$$

for every $h \in H$ and (2.1) holds, then $\mathscr{E}_\alpha \to \mathscr{E}$.

Another way of defining convergence is in terms of *canonical measures*. Given a finite subset $I \subset H$ set $\mu_{\alpha,I} = \sum_{h \in I} P_{\alpha,h}$ and let μ_I be the corresponding quantity for $\mathscr{E}$. Now $\mathscr{E}_\alpha \to \mathscr{E}$ if and only if

$$\mathscr{L}_{\mu_{\alpha,I}}\left(\left(\frac{dP_{\alpha,h}}{d\mu_{\alpha,I}}\right)_{h \in I}\right) \to \mathscr{L}_{\mu_I}\left(\left(\frac{dP_h}{d\mu_I}\right)_{h \in I}\right), \tag{2.3}$$

for every finite subset $I \subset H$. The measures in (2.3), the canonical or 'standard measures' of the experiments, give measure $|I|$ to the unit simplex $\mathbf{S}_I = \{x \in \mathbb{R}^I : x \geqslant 0, \sum x_h = 1\}$, and the convergence is the usual weak convergence of measures on $\mathbb{R}^I$. As is also clear in Millar (1983), this definition is suitable for elementary proofs of relatively deep results. We use it here for the proof of the main result of the paper, Theorem 3.1.

A useful manner for both finding the form of a limit experiment and proving convergence to it, is based on a notion of 'asymptotic sufficiency'. Fix h_0. Suppose one has

$$\frac{dP_{\alpha,h}}{dP_{\alpha,h_0}} = g_h(\Delta_\alpha) + o_{P_{\alpha,h_0}}(1), \quad \text{every } h \in H, \tag{2.4}$$

where $\Delta_\alpha : \mathscr{X}_\alpha \to S$ are Borel measurable maps into a metric space S with

$$\mathscr{L}_h(\Delta_\alpha) \to P_h, \quad \text{every } h \in H. \tag{2.5}$$

(Both Δ_α and g_h may depend on the fixed h_0.) Intuitively, (2.4) suggests that the maps Δ_α are 'asymptotically sufficient'. Therefore one may hope that $\mathscr{E} = (S, \mathscr{B}_o, P_h : h \in H)$ is a version of a limit experiment. This is not true in general, but, for instance, if (2.2) holds and the map $y \to g_h(y)$ is continuous for every h, it is.

It should be noted that the measurable spaces $(\mathscr{X}_\alpha, \mathscr{B}_\alpha)$ do not play a role in (2.1) or (2.3), except that they carry the $P_{\alpha,h}$ which define the likelihood ratio processes or canonical measures. Thus a limit experiment is clearly not unique. Two experiments with the same system of canonical measures (or, equivalently, for which the likelihood process are equal in law) are called *equivalent* or 'of the same type'. An important example of equivalent experiments arises from sufficiency. The experiment consisting of observing a statistic which is sufficient in another experiment $\mathscr{E}$, is equivalent to this experiment $\mathscr{E}$. This follows readily from the factorization theorem. Thus through sufficiency it is possible to replace a 'complicated' limit experiment by a simpler one.

The assertions in this section are not proved here. See e.g. Le Cam (1986), or Strasser (1985, § 60).

3 Asymptotic Representation Theorem

Let X_α be an *observation* in the αth experiment $\mathscr{E}_\alpha = (\mathscr{X}_\alpha, \mathscr{B}_\alpha, P_{\alpha,h} : h \in H)$, that is a random element in $(\mathscr{X}_\alpha, \mathscr{B}_\alpha)$ with law $P_{\alpha,h}$. Consider a net of statistics $T_\alpha = T_\alpha(X_\alpha)$ which converges in law under every parameter:

$$T_\alpha \overset{h}{\Rightarrow} Q_h, \quad \text{every } h \in H. \tag{3.1a}$$

For this to make sense we suppose that every $T_\alpha : \mathscr{X}_\alpha \to \mathbf{D}$ is a map with values in a metric space $\mathbf{D}$ and that every Q_h is a probability measure on the Borel σ-field $\mathscr{D}_o$ on $\mathbf{D}$. Then (3.1a) is taken to mean:

$$E_h^* f(T_\alpha) \to \int f \, dQ_h, \quad \text{every } f \in C_b(\mathbf{D}), \quad \text{every } h \in H.$$

In addition to (3.1a) we assume that there exists a complete, separable subset $\mathbf{D}_0$ of $\mathbf{D}$ such that

$$Q_h(\mathbf{D}_0) = 1, \quad \text{every } h \in H. \tag{3.1b}$$

Perhaps the case of greatest interest is $\mathbf{D} = \mathbb{R}^m$. Other popular examples of 'decision spaces' $\mathbf{D}$ are the spaces of all continuous, or all cadlag functions on an interval, as natural spaces that contain distribution functions. These can both be equipped with the uniform norm, or in the case of the latter, one of the Skorohod metrics. More recently, there is an interest in the space of all bounded functions on some set, equipped with the uniform norm.

A number of these decision spaces, including $\mathbb{R}^m$, are complete and separable. If this is the case and, moreover, every T_α is Borel measurable, then (3.1) is just the usual weak convergence of the net of induced Borel laws $\mathscr{L}_h(T_\alpha)$.

Actually, (3.1) also includes a more general situation, since no measurability condition has been imposed on T_α. The given expectation is an outer expectation:

$$E_h^* f(T_\alpha) := \inf\left\{\int g \, dP_{\alpha,h} \mid g : (\mathscr{X}_\alpha, \mathscr{B}_\alpha) \to \mathbb{R} \text{ measurable}, \, g \geq f(T_\alpha)\right\}.$$

At first reading one shouldn't worry about the possible non-measurability. The more general situation is included here, because some applications require a decision space $\mathbf{D}$ which is non-separable. In this case statistics of interest are often not Borel measurable and hence do not have Borel laws. Then the extension using the generalized weak convergence (3.1a) based on outer expectations, is appropriate. This concept of 'convergence in law without having laws' is due to Hoffman-Jorgensen (1984) (as reported in Dudley (1985) and Andersen & Dobric (1987)) and generalizes both the classical theory of weak convergence as in Billingsley (1968) and the theory of weak convergence of measures defined on the σ-field generated by the closed balls due to Dudley (1966) (as expounded by Gaenssler (1983) and Pollard (1984)). Though most of the properties of weak convergence of Borel laws go through for the generalized weak convergence, proper handling of the outer expectations requires some care. See e.g. van der Vaart & Wellner (1989) for a review. Alternatively, one can often extend results to the non-measurable case by using a device of Le Cam (1989), who shows that, for a converging net of possibly non-measurable maps, there is always a net of measurable maps that is 'close' and converges to the same limit.

The following theorem is the main result of the paper. It can be described as a 'concretization' of Proposition 8 of Le Cam (1972) and is related to Proposition 1 on page 100 and Theorem 2 on page 111 of Le Cam (1986).

THEOREM 3.1. *Suppose that the net of experiments $\mathscr{E}_\alpha$ converges to a dominated experiment $\mathscr{E}$ and that* (3.1) *holds. Then there exists a randomized estimator t in $\mathscr{E}$ such that*

$$Q_h = \mathscr{L}_h(t(X, U)), \quad \textit{every } h \in H. \tag{3.2}$$

Here a *randomized estimator t* is a measurable map $t : (\mathscr{X} \times [0, 1], \mathscr{B} \times \mathscr{B}_o) \to (\mathbf{D}, \mathscr{D}_o)$. Furthermore, X in (3.2) is an 'observation' in $\mathscr{E}$ and satisfies $\mathscr{L}_h(X) = P_h$ for every $h \in H$, while U is a random variable with a uniform distribution on $[0, 1]$ which is independent from X. Thus an 'estimate' $t(X, U)$ is based on an observation in $\mathscr{E}$ and some external randomization.

Theorem 3.1 expresses that any set of limit laws $\{Q_h : h \in H\}$ is the set of laws of a randomized estimator in the limit experiment. If the asymptotic performance of an estimator sequence is considered completely determined by its set of limit laws, then the implication of Theorem 3.1 for a theory of asymptotic lower bounds is clear: whatever meaning is given to 'good', no estimator sequence satisfying (3.1) can in the limit be better than the *best* randomized estimator in the limit experiment.

Next, every method of making 'good' precise and an analysis of the limit experiment leads to a concrete lower bound theorem. In §§ 5 and 6 we consider the classical notions of equivariance and minimaxity, which are popular in asymptotic theory, in particular since the papers of Hájek (1970, 1972).

Note. Condition (3.1b) recurs regularly in the paper, so that it is good to know that it is often satisfied as soon as $Q_{h_0}(\mathbf{D}_0)=1$ for just one $h_0 \in H$. In fact, this is true whenever all nets $\{P_{\alpha,h}\}$ and $\{P_{\alpha,h_0}\}$ are contiguous. Call a net of maps $T_\alpha : \mathscr{X}_\alpha \to \mathbf{D}$ *asymptotically tight* under h if for every $\varepsilon > 0$ there exists a compact subset $K_\varepsilon \subset \mathbf{D}$ with

$$\liminf_\alpha P_{\alpha,h^*}(d(T_\alpha, K_\varepsilon) < \delta) > 1-\varepsilon,$$

for every $\delta > 0$. The following lemma is a consequence of Le Cam's third Lemma.

LEMMA. *Suppose that the net $\{P_{\alpha,h}\}$ is contiguous with respect to $\{P_{\alpha,h_0}\}$ and that $T_\alpha \Rightarrow Q_{h_0}$ under h_0 where $Q_{h_0}(\mathbf{D}_0)=1$ for some complete, separable subset $\mathbf{D}_0$ of $\mathbf{D}$. Then $\{T_\alpha\}$ is asymptotically tight and relatively compact under h and every limit point Q_h satisfies $Q_h(\mathbf{D}_0)=1$.*

A proof of Theorem 3.1 is deferred to § 10.

4 Standardized Statistics

In estimation problems the T_α of § 3 is often a standardized statistic used to estimate a functional $\kappa_\alpha(P_{\alpha,h})$. In this section Theorem 3.1 is reformulated for this case.

Let H be an arbitrary set (but in this section typically $\mathbb{R}^m$ or a Hilbert space) and let $\kappa_\alpha(P_{\alpha,h})$ be a quantity in a normed space $\mathbf{D}$ for every $h \in H$. Let $T_\alpha = T_\alpha(X_\alpha)$ be a net of $\mathbf{D}$-valued statistics such that for probability measures L_h on the Borel sets of $\mathbf{D}$

$$R_\alpha(T_\alpha - \kappa_\alpha(P_{\alpha,h})) \overset{h}{\Rightarrow} L_h, \quad \text{every } h \in H, \tag{4.1a}$$

for linear operators $R_\alpha : \mathbf{D} \to \mathbf{D}$ (often just real numbers). Furthermore, assume that every L_h concentrates on a complete, separable subspace $\mathbf{D}_0$ of $\mathbf{D}$:

$$L_h(\mathbf{D}_0) = 1, \quad \text{every } h \in H. \tag{4.1b}$$

Finally assume that κ_α is 'differentiable in the limit' in the sense that for some $h_0 \in H$

$$R_\alpha(\kappa_\alpha(P_{\alpha,h}) - \kappa_\alpha(P_{\alpha,h_0})) \to \kappa'(h) - \kappa'(h_0), \quad \text{every } h \in H \tag{4.2}$$

for some map $\kappa' : H \to \mathbf{D}_0$. (Usually this will be linear and continuous, as a 'true' derivative should be.)

The following theorem says that 'in the limit' the problem of estimating $\kappa_\alpha(P_{\alpha,h})$ is no easier than estimating $\kappa'(h)$ in the limit experiment.

THEOREM 4.1. *Let $\mathscr{E}_\alpha$ be a net of experiments converging to a dominated experiment $\mathscr{E}$ and let* (4.1)–(4.2) *hold. Then there exists a randomized estimator t in $\mathscr{E}$ such that*

$$L_h = \mathscr{L}_h(t(X, U) - \kappa'(h)), \quad \text{every } h \in H.$$

Proof. Make the substitutions

$$T_\alpha \leftrightarrow R_\alpha(T_\alpha - \kappa_\alpha(P_{\alpha,h_0})), \quad Q_h(.) \leftrightarrow L_h(. - \kappa'(h) + \kappa'(h_0)), \quad t(X, U) \leftrightarrow t(X, U) + \kappa'(h_0)$$

in Theorem 3.1. □

5 Asymptotic Convolution Theorem

In the set-up of § 4, call a sequence of statistics T_n *regular* if the distributions L_h in (4.1) are equal to a single, fixed probability distribution L. It follows immediately from Theorem 4.1 that, for any limit distribution L of a regular sequence of statistics, there exists a randomized estimator t in $\mathscr{E}$ such that

$$L = \mathscr{L}_h(t(X, U) - \kappa'(h)) = \mathscr{L}_0(t(X, U)), \quad \text{every } h \in H. \tag{5.1}$$

In other words, L is the distribution of a randomized estimator in the limit experiment, which as an estimator for $\kappa'(h)$, is '(shift) equivariant in law'. Clearly, in most situations this will severely limit the possible limit distributions a regular (or perhaps better: 'asymptotically equivariant') estimator sequence can have. For some special types of limit experiments any such L even has a nice characterization as a certain convolution. The following result is a version of the Hájek–Le Cam convolution theorem. For the case that the limit experiment is a Gaussian shift this is due to Hájek (1970). The general case is due to Le Cam (1972).

THEOREM 5.1. *Let $H = \mathbb{R}^m$. Assume that the sequence of experiments $\mathscr{E}_n$ converges to a 'shift experiment' $\mathscr{E} = (\mathbb{R}^m, \mathscr{B}_o^m, \mathscr{L}(V + \Sigma h): h \in H)$, where $\mathscr{L}(V)$ is a fixed Lebesgue dominated probability measure on $\mathbb{R}^m$ and Σ is a nonsingular matrix. Let* (4.2) *hold with κ' a linear map. Then a limit distribution L of a regular sequence of statistics can be written as*

$$L = \mathscr{L}(\kappa'(\Sigma^{-1}V) + W),$$

where W is a Borel measurable random element in $\mathbf{D}_0$ which is independent of V.

Proof. For simplicity of notation assume that Σ equals the identity matrix. Let t be a randomized estimator as in (5.1). Assume without loss of generality that t takes its values in $\mathbf{D}_0$. Rewrite (5.1) as

$$L = \mathscr{L}(t(V + h, U) - \kappa'(h)), \quad \text{every } h \in \mathbb{R}^m, \tag{5.2}$$

where V is independent of U.

For $p = 1, 2, \ldots$ let θ_p be independent of (U, V) and uniformly distributed on $[-p, p] \subset \mathbb{R}^m$. By (5.2) and linearity of κ'

$$\begin{aligned} L &= \mathscr{L}(t(V + \theta_p, U) - \kappa'(\theta_p)) \\ &= \mathscr{L}(t(V + \theta_p, U) - \kappa'(V + \theta_p) + \kappa'(V)). \end{aligned}$$

Here $V + \theta_p$ and V are asymptotically independent as $p \to \infty$. Indeed,

$$\begin{aligned} &|P(V + \theta_p \in A, V \in B) - P(V + \theta_p \in A)P(V \in B)| \\ &\quad \leq |P(V + \theta_p \in A, V \in B) - P(\theta_p \in A)P(V \in B)| \\ &\quad\quad + |P(\theta_p \in A) - P(V + \theta_p \in A)|\, P(V \in B) \\ &\quad \leq 2 \int |P(\theta_p + v \in A) - P(\theta_p \in A)|\, dP^V(v) \\ &\quad \leq 4P(|V| > \sqrt{p}) + 2 \sup_{|v| \leq \sqrt{p}} |P(\theta_p + v \in A) - P(\theta_p \in A)| \to 0, \end{aligned}$$

uniformly in Borel sets A and B. In consequence, with $\tilde{W}_p = t(V + \theta_p, U) - \kappa'(V + \theta_p)$,

$$|P(\tilde{W}_p \in C, V \in D) - P(\tilde{W}_p \in C)P(V \in D)| \to 0,$$

uniformly in sets $C \in \mathscr{D}_o$, $D \in \mathscr{B}_o$. Let W_p have the same law as $\tilde{W}_p$ but be independent of V. Conclude that $\|\mathscr{L}(W_p, V) - \mathscr{L}(\tilde{W}_p, V)\| \to 0$. Thus

$$\mathscr{L}(W_p + \kappa'(V)) \to \mathscr{L}(\tilde{W}_p + \kappa'(V)) = L,$$

also in total variation norm. Since L is tight, there exists for every $\varepsilon > 0$ a compact $K_\varepsilon \subset \mathbf{D}_0$ with

$$\liminf_{p\to\infty} P(W_p + \kappa'(V) \in K_\varepsilon) \geq 1 - \varepsilon.$$

There is also a compact $K \subset \mathbb{R}^m$ with $P(V \in K) \geq 1 - \varepsilon$. But then

$$\liminf_{p\to\infty} P(W_p \in K_\varepsilon - \kappa'(K)) \geq 1 - 2\varepsilon,$$

so that $\{\mathscr{L}(W_p)\}$ is asymptotically tight. Every limit point $\mathscr{L}(W, V)$ of $\mathscr{L}(W_p, V)$ satisfies

$$L = \mathscr{L}(W + \kappa'(V)). \qquad \square$$

The random element W in Theorem 5.1 is usually interpreted as (a translate of) a 'noise factor', that ought to be zero for an 'optimal' regular sequence of statistics. Hence a 'best' regular sequence of statistics is asymptotically distributed as (a translate of) $\kappa'(\Sigma^{-1}V)$. The simplest way to motivate this is to think in terms of variance: the variance of the limit distribution L equals the sum of the variances of $\kappa'(\Sigma^{-1}V)$ and W. Hence for quadratic loss the best possible limit distribution is $\mathscr{L}(\kappa'(\Sigma^{-1}(V - EV)))$. This argument extends to other loss functions. Note, however, that in general the appropriate centering depends on the loss function.

Note. If a shift experiment $\mathscr{E}$ as in Theorem 5.1 is dominated, then it is automatically dominated by Lebesgue measure. This explains the assumption that $\mathscr{L}(V)$ is Lebesgue absolutely continuous, which otherwise is not used in the proof.

Note. There are cases wherein the assumption that H equals the whole of $\mathbb{R}^m$ may be relaxed. For instance, let $\mathbf{D} = \mathbb{R}^k$ and suppose that the distribution of X is such that the function

$$h \to E_h e^{is'(t(X,U) - \kappa'(h))}$$

is defined and analytic for $h \in \mathbb{C}^m$, for every $s \in \mathbb{R}^k$ and every randomized estimator t. Regularity of a sequence of statistics for all h in a set H implies that its limit distribution L satisfies

$$\int e^{it'y}\, dL(y) = E_h e^{is'(t(X,U) - \kappa'(h))}, \quad \text{every } h \in H,$$

for some randomized estimator t. Then, if H is rich enough to ensure uniqueness of analytic continuation, (5.1) is obtained once again for all $h \in \mathbb{R}^m$. This extension applies for instance to Gaussian shifts. By a similar argument one sees that it always suffices that H is dense in $\mathbb{R}^m$.

It appears to be unknown whether the Euclidean space $\mathbb{R}^m$ in Theorem 5.1, which plays the role of both sample space and parameter space, can be replaced by a more general, infinitely dimensional space. In fact, only results for the situation where V has a Gaussian distribution have been obtained so far (e.g. Millar (1985) and the extensive literature on semi-parametric models). The following theorem is a somewhat special result in this direction, though it is general enough to cover the known Gaussian situation.

THEOREM 5.2. *Let H be a linear space and let $\tau: H \to \mathbb{R}^\infty$ be a one-to-one linear map of which the range contains all unit vectors $e_i = (0, \ldots, 0, 1, 0, \ldots)$. Let $V_1, V_2, \ldots$ be independent random variables with Lebesgue dominated distributions on $\mathbb{R}$. Assume that the sequence of experiments $\mathscr{E}_n$ converges to $\mathscr{E} = (\mathbb{R}^\infty, \mathscr{B}_o^\infty, \mathscr{L}(V + \tau h): h \in H)$. Let* (4.2) *hold. Suppose that the series*

$$\kappa'(\tau^{-1}V) = \sum_{i=1}^{\infty} V_i \kappa'(\tau^{-1}e_i)$$

converges in law. Then a limit distribution L of a regular sequence of statistics can be written as

$$L = \mathscr{L}(\kappa'(\tau^{-1}V) + W), \tag{5.3}$$

where W is a Borel measurable random element in $\mathbf{D}_0$ which is independent of V. If there exists a regular sequence of statistics and every V_i is symmetrically distributed about zero, then almost sure convergence of the series is automatically satisfied.

Proof. Fix m. By assumption the range of τ contains $\mathbb{R}^m \times \{0\} \times \{0\} \times \ldots$. The sequence of experiments

$$\mathscr{E}_n^{(m)} = (\mathscr{X}_n, \mathscr{B}_n, P_{n,\tau^{-1}(\alpha,0,0,\ldots)}: \alpha \in \mathbb{R}^m)$$

converges to

$$\mathscr{E}^{(m)} = (\mathbb{R}^\infty, \mathscr{B}_o^\infty, \mathscr{L}(V + (\alpha, 0, 0, \ldots)): \alpha \in \mathbb{R}^m).$$

In the latter experiment the first m components of the observation are sufficient for α. Thus one also has that

$$\mathscr{E}_n^{(m)} \to (\mathbb{R}^m, \mathscr{B}_o^m, \mathscr{L}(V^{(m)} + \alpha): \alpha \in \mathbb{R}^m),$$

where $V^{(m)} = (V_1, \ldots, V_m)$.

Let L be the limit distribution of a regular sequence of statistics in $\mathscr{E}_n$. This sequence of statistics is certainly regular for the sequence $\mathscr{E}_n^{(m)}$. Furthermore,

$$R_n(\kappa_n(P_{n,\tau^{-1}(\alpha,0,0,\ldots)}) - \kappa_n(P_{n,0})) \to \kappa'(\tau^{-1}(\alpha, 0, 0, \ldots)), \quad \text{every } \alpha \in \mathbb{R}^m.$$

Set

$$Z^{(m)} = \kappa'(\tau^{-1}(V^{(m)}, 0, 0, \ldots)) = \sum_{i=1}^{m} V_i \kappa'(\tau^{-1}e_i).$$

By Theorem 5.1 there exists a random element $W^{(m)}$ that is independent of $V^{(m)}$ such that

$$L = \mathscr{L}(Z^{(m)} + W^{(m)}). \tag{5.4}$$

If the series $Z = \kappa'(\tau^{-1}V)$ converges, then $\mathscr{L}(Z^{(m)}) \to \mathscr{L}(Z)$. By Prohorov's theorem the sequence of Borel laws $\{\mathscr{L}(Z^{(m)})\}$ on $\mathbf{D}_0$ is tight. This can be combined with (5.4) to show tightness of the sequence $\{\mathscr{L}(W^{(m)})\}$. Any limit point of $\mathscr{L}(Z^{(m)}, W^{(m)})$ satisfies (5.3).

Next suppose that $V_1, V_2, \ldots$ are symmetrically distributed about zero. Then $Z^{(m)}$ has the same distribution as $-Z^{(m)}$. Fix $\varepsilon > 0$. Since L is tight, there exists a compact $K_\varepsilon \subset \mathbf{D}_0$ such that

$$L(K_\varepsilon) = P(Z^{(m)} + W^{(m)} \in K_\varepsilon) \geq 1 - \varepsilon.$$

By symmetry $P(-Z^{(m)} + W^{(m)} \in K_\varepsilon) \geq 1 - \varepsilon$. But then

$$P(Z^{(m)} \in \tfrac{1}{2}(K_\varepsilon - K_\varepsilon)) \geq 1 - 2\varepsilon.$$

Since $\frac{1}{2}(K_\varepsilon - K_\varepsilon)$ is compact, the sequence $\mathscr{L}(Z^{(m)})$ is tight. As before conclude by combination with (5.4) that $\{\mathscr{L}(W^{(m)})\}$ is tight. Every limit point $\mathscr{L}(Z, W, V)$ of the sequence $\mathscr{L}(Z^{(m)}, W^{(m)}, V)$ has that Z and V are independent and $\mathscr{L}(Z+W)=L$. Thus the series of independent random elements

$$W+\sum_{i=1}^{\infty} V_i\kappa'(\tau^{-1}e_i)$$

converges in law (to L). But then it converges almost surely too by Theorem 2.10 in Araujo & Giné (1980, p. 105). □

6 Asymptotic Minimax Theorem

Consider again the situation of § 4. Let $l_h:\mathbf{D}\to[0,\infty)$ be fixed, lower semi-continuous functions. For instance, let $l_h(y)=\|y-h\|^2$ or $l_h(y)=w(\|y-h\|)$ for a fixed, continuous, non-negative, non-decreasing function w on $[0,\infty)$. Let T_n be a sequence of statistics that are Borel measurable, or at least satisfy the following 'asymptotic measurability condition': for every $f\in C_b(\mathbf{D})$

$$E_{h}^{*}f(R_n(T_n-\kappa_n(P_{n,h})))-E_{h,*}f(R_n(T_n-\kappa_n(P_{n,h})))\to 0,\quad \text{every } h\in H. \tag{6.1}$$

Condition (6.1) is certainly satisfied if each T_n is Borel measurable, since then the expressions in (6.1) are identically zero for every n. It is also satisfied by any sequence that converges weakly to a limit in the generalized sense introduced in § 3.

THEOREM 6.1. *Let the sequence of experiments $\mathscr{E}_n$ converge to a dominated experiment $\mathscr{E}$. Let* (4.2) *hold. In the case that $\mathbf{D}=\mathbb{R}^k$ and $l_h(y)\le\liminf_{|y|\to\infty} l_h(y)$ for every y and h one has for every measurable sequence of estimators T_n that*

$$\sup_I \liminf_{n\to\infty}\sup_{h\in I} E_{h,*}\, l_h(R_n(T_n-\kappa_n(P_{n,h})))\ge \inf_t \sup_{h\in H} E_h l_h(t(X,U)-\kappa'(h)), \tag{6.2}$$

where the first supremum is taken over all finite subsets I of H and the infimum is taken over all randomized estimators t in $\mathscr{E}$.

For general $\mathbf{D}$ (6.2) *holds for every sequence T_n that satisfies* (6.1) *and for which the sequence $R_n(T_n-\kappa_n(P_{n,h}))$ is asymptotically tight under every $h\in H$ with limit points concentrating on a fixed complete, separable subset $\mathbf{D}_0$ of $\mathbf{D}$.*

Proof. We first prove the second assertion. Partially order the finite subsets of H by inclusion: $I_1\le I_2$ iff $I_1\subset I_2$. Call the left side of (6.2) 'risk'. Assume without loss of generality that it is finite. There exists a subnet $\{n_I: I\subset H, \text{ finite}\}$ such that

$$\text{risk}=\lim_I \sup_{h\in I} E_{h,*}\, l_h(R_{n_I}(T_{n_I}-\kappa_{n_I}(P_{n_I,h}))).$$

For instance, take n_I any natural number with $n_I\ge|I|$ and

$$|r(n_I, I)-\liminf_{n\to\infty} r(n, I)|<|I|^{-1},$$

where

$$r(n, I)=\sup_{h\in I} E_{h,*}\, l_h(R_n(T_n-\kappa_n(P_{n,h}))).$$

By the tightness and measurability condition this subnet has a further subnet such that for probability measures L_h concentrating on a fixed separable complete subset of $\mathbf{D}$

$$R_\alpha(T_\alpha - \kappa_\alpha(P_{\alpha,h})) \overset{h}{\Rightarrow} L_h, \quad \text{every } h \in H,$$

where we write α for $n_{I(\alpha)}$. This follows basically from a combination of Prohorov's theorem and Tychonov's theorem. (For each h tightness of the standardized sequence of estimators implies its weak pre-compactness; next the sequence indexed by h, $(\{R_n(T_n - \kappa_n(P_{n,h}))\})_{h\in H}$, is pre-compact in the product of the weak topologies, and hence has a converging subnet. For a formal statement see Theorem 1 in the appendix.)

This means that (4.1) is satisfied along this subnet. By Theorem 4.1 there exists a randomized estimator t in $\mathscr{E}$ such that

$$L_h = \mathscr{L}_h(t(X, U) - \kappa'(h)), \quad \text{every } h \in H. \tag{6.3}$$

By lower semi-continuity of $y \to l_h(y)$,

$$\liminf_\alpha E_{h,*}\, l_h(R_\alpha(T_\alpha - \kappa_\alpha(P_{\alpha,h}))) \wedge M \geqslant \int l_h(y) \wedge M \, dL_h(y) \tag{6.4}$$

for every M and h. In consequence

$$\text{risk} = \lim_\alpha \sup_{h \in I(\alpha)} E_{h,*}\, l_h(R_\alpha(T_\alpha - \kappa_\alpha(P_{\alpha,h}))) \geqslant \sup_{h\in H} \int l_h(y)\, dL_h(y). \tag{6.5}$$

By (6.3) the right hand side of (6.4) is larger or equal to the right hand side of (6.2).

For the proof of the first assertion let $\bar{\mathbf{D}}$ be the one-point compactification of $\mathbb{R}^k$. Extend $y \to l_h(y)$ to a lower continuous function on $\bar{\mathbf{D}}$ by setting $l_h(\infty) = \liminf_{|y| \to \infty} l_h(y)$. The sequence of laws of the estimators, seen as maps in the compact space $\bar{\mathbf{D}}$ is trivially tight for every h. Thus by Theorem 6.1 the inequality (6.2) holds with the infimum taken over all randomized estimators t with values in $\bar{\mathbf{D}}$. But by the condition on l_h the risk $E_h l_h(t(X, U) - \kappa'(h))$ decreases if t is replaced by, for example, $t1_{t\neq\infty}$. Thus taking the infimum may be restricted to randomized estimators with values in $\mathbb{R}^k$, without changing the minimax risk. □

It is noted by Le Cam (1986, pp. 109–110) that the minimax theorem can be obtained as a corollary of a stronger (and more complicated) theorem. The following is a concrete, limit version of his Theorem 1 on page 109.

THEOREM 6.2. *Let the sequence of experiments $\mathscr{E}_n$ converge to a dominated experiment $\mathscr{E}$. Let (4.2) hold for $\mathbb{R}^k$-valued functionals $\kappa_n(P_{n,h})$. Let $l_h : \mathbb{R}^k \to [0, \infty)$ be lower semi-continuous functions with $l_h(y) \leqslant \liminf_{|y|\to\infty} l_h(y)$ for every y and h. Suppose that $R : H \to [0, \infty)$ is such that, for every randomized estimator $t(X, U)$ in $\mathscr{E}$ with values in $\mathbb{R}^k$, there is a parameter h with $R(h) < E_h l_h(t(X, U) - \kappa'(h))$. Then there is a probability measure μ with finite support on H with*

$$\int R\, d\mu < \liminf_{n\to\infty} \int E_{h,*}\, l_h(R_n(T_n - \kappa_n(P_{n,h})))\, d\mu(h) \tag{6.6}$$

for every measurable sequence of estimators T_n.

Proof. Let μ be an arbitrary probability measure μ of finite support on H. By a similar method as used in the proof of the previous theorem one can obtain a lower bound for the average (rather than maximum) risk of the following form. The right side of (6.6) is not smaller than

$$\inf_t \int E_h l_h(t(X, U) - \kappa'(h))\, d\mu(h),$$

where the infimum is taken over all randomized estimators t in $\mathscr{E}$ with values in $\mathbb{R}^k$ (or its one-point compactification). Next apply the non-asymptotic minimax theorem (allowing changing the order of infimum and supremum in the identity: maximum Bayes risk equals minimax risk) to obtain that for every function R with the stated properties there is a probability measure μ with finite support such that

$$\int R\,d\mu < \inf_t \int E_h l_h(t(X, U) - \kappa'(h))\,d\mu(h).$$

See for instance Theorem 1 of Le Cam (1986, p. 16). (Under the present conditions his $\chi(\mu)$ reduces to the right side of the last displayed equation.) □

Note. A constant function R satisfies the condition of Theorem 6.2 if and only if R is smaller than the minimax risk in $\mathscr{E}$. Furthermore, the left side of (6.2) is bounded below by the right side of (6.6). Combination of these facts shows that Theorem 6.2 is a strengthening of Theorem 6.1.

7 Testing

A *test* in an experiment $\mathscr{E}$ is a Borel measurable map $\phi : \mathscr{X} \to [0, 1]$. The corresponding *power function* is the map $\pi : H \to [0, 1]$ given by

$$\pi(h) = E_h\phi(X), \quad h \in H. \tag{7.1}$$

Let ϕ_α be a net of tests in experiments $\mathscr{E}_\alpha = (\mathscr{X}_\alpha, \mathscr{B}_\alpha, P_{\alpha,h} : h \in H)$. Suppose that the corresponding net of power functions π_α converges pointwise to some function π

$$\pi(h) := \lim_\alpha \pi_\alpha(h), \quad \text{every } h \in H. \tag{7.2}$$

THEOREM 7.1. *Suppose that the net of experiments $\mathscr{E}_\alpha$ converges to a dominated experiment $\mathscr{E}$. Let* (7.2) *hold. Then there exists a test in $\mathscr{E}$ with power function π.*

Proof. Since $[0, 1]$ is compact, there exists by Theorem 1 in the Appendix a subnet, indexed by $\alpha(\beta)$, such that

$$\mathscr{L}_h(\phi_{\alpha(\beta)}(X_{\alpha(\beta)})) \xrightarrow{\beta} Q_h, \quad \text{every } h \in H, \tag{7.3}$$

where the Q_h are Borel measures on $[0, 1]$. By Theorem 3.1 there exists a randomized estimator t satisfying (3.3). Clearly

$$\pi(h) = \lim_\beta E_h\phi_{\alpha(\beta)}(X_{\alpha(\beta)}) = \int y\,dQ_h(y) = E_h t(X, U).$$

But then $\phi(x) = E(t(X, U) \mid X = x)$ is a test in $\mathscr{E}$ with power function π. □

Usually one chooses to find a 'best' test within the class of tests of level α_0. The next theorem says that the asymptotic power of any sequence of level α_0 tests is bounded from above by the power envelope function of level α_0 of $\mathscr{E}$.

THEOREM 7.2. *Suppose that the sequence of experiments $\mathscr{E}_n = (\mathscr{X}_n, \mathscr{B}_n, P_{n,h} : h \in H)$ converges to a dominated experiment $\mathscr{E}$. Let ϕ_n be tests in $\mathscr{E}_n$ such that*

$$\limsup_{n\to\infty} E_h\phi_n(X_n) \le \alpha_0, \quad \textit{every } h \in H_0 \subset H, \tag{7.4}$$

for some $\alpha_0 \in (0, 1)$. *Then*

$$\limsup_{n\to\infty} E_h\phi_n(X_n) \leq \sup_{\phi\in\Phi_{\alpha_0}} E_h\phi(X), \quad \textit{every } h \in H - H_0, \tag{7.5}$$

where Φ_{α_0} *is the collection of level* α_0 *tests for testing the null hypothesis* H_0 *in* $\mathscr{E}$.

Proof. Fix $h \in H - H_0$. First choose a subsequence of $\{n\}$ along which the limsup in (7.5) is taken for this particular h. Next, extract a further subnet along which (7.2) is satisfied. This is possibly, since $[0, 1]^H$ is compact in the product topology. By Theorem 7.1 there exists a test ϕ in $\mathscr{E}$ for which

$$\lim_{\beta} E_h\phi_{n(\beta)}(X_{n(\beta)}) = E_h\phi(X), \quad \text{every } h \in H.$$

Combination with (7.4) gives that ϕ is a level α_0 test for H_0. Hence (7.5). □

8 Some Examples

This section contains only a few examples, for illustration. Many more examples can be found in the literature. For instance, there are many concrete examples of the semi-parametric set-up of § 8.2. Also, asymptotic expansions for the likelihood processes of a number of classical models are given in e.g. Ibragimov & Has'minskii (1981). From these the form of a limit experiment can usually be guessed.

8.1 Smooth Parametric Models

Let the nth experiment consist of the observation of an independent identically distributed sample $X_1, \ldots, X_n$ from a density p_θ with respect to some σ-finite measure μ on the sample space $(\mathscr{X}, \mathscr{B})$. Assume that $\theta \in \Theta$, an open subset of $\mathbb{R}^k$ and that the map $\theta \to p_\theta$ is smooth in the sense of existence of a measurable, vector-valued function $\dot{l}_\theta$ such that

$$\int [t^{-1}(p_{\theta+th}^{\frac{1}{2}} - p_\theta^{\frac{1}{2}}) - \tfrac{1}{2}h'\dot{l}_\theta p_\theta^{\frac{1}{2}}]^2\, d\mu \to 0, \quad \text{as } t \to 0, \tag{8.1}$$

for every $h \in \mathbb{R}^k$. In nice cases $\dot{l}_\theta(x)$ will be the vector of partial derivatives $\partial/\partial\theta_i \log p_\theta(x)$. Simple sufficient conditions for the differentiability in quadratic mean (8.1) in terms of this ordinary pointwise derivative can be found in the appendix of Hájek (1972) and in Le Cam (1970). The smoothness of $\theta \to p_\theta$ imposed here is roughly what one also needs for the Cramér–Rao bound.

Fix θ. Let $I_\theta := \int \dot{l}_\theta \dot{l}_\theta' p_\theta\, d\mu$ be the *Fisher information matrix*. Under (8.1) there is convergence of 'local' experiments:

$$\mathscr{E}_{n,\theta} := \Big(\mathscr{X}^n, \mathscr{B}^n, \bigotimes_{j=1}^{n} P_{\theta+h/\sqrt{n}} : h \in \mathbb{R}^k\Big) \to \mathscr{E}_\theta := (\mathbb{R}^k, \mathscr{B}_o^k, N_k(I_\theta h, I_\theta) : h \in \mathbb{R}^k), \tag{8.2}$$

where $P_{\theta+h/\sqrt{n}}$ is arbitrary if $\theta + h/\sqrt{n} \notin \Theta$. (Since Θ is open, $\theta + h/\sqrt{n}$ will be contained in Θ for sufficiently large n, for every fixed h.)

The convergence (8.2) follows easily from the log likelihood expansion

$$\log \prod_{j=1}^{n} \frac{p_{\theta+h/\sqrt{n}}}{p_\theta}(X_j) = h'n^{-\frac{1}{2}} \sum_{j=1}^{n} \dot{l}_\theta(X_j) - \tfrac{1}{2}h'I_\theta h + o_{P_\theta}(1). \tag{8.3}$$

In turn, for a sufficiently smooth map $\theta \to \log p_\theta(x)$, (8.3) follows from a Taylor expansion. The matrix I_θ in the quadratic term would appear as the limit in probability of

$$-\left(n^{-1}\sum_{j=1}^{n}\frac{\partial^2}{\partial\theta_i\,\partial\theta_j}\log p_\theta(X_j)\right).$$

Validity of (8.3) under the weak condition (8.1) is due to Le Cam. See Le Cam (1970) for a detailed discussion.

Of course, in the limit experiment $\mathscr{E}_\theta$ a 'good' estimator for a continuous, linear functional $\kappa'(h)$ is $\kappa'(I_\theta^{-1}X)$, where X is a $N_k(I_\theta h, I_\theta)$ distributed observation. In particular, this estimator is 'best equivariant-in-law' and minimax for a large class of loss functions. For the functional $\kappa(P_\theta)=\theta$ one has $\kappa'(h)=h$. Hence one obtains the classical result that the normal distribution $N_k(0, I_\theta^{-1})$, the distribution of $I_\theta^{-1}X$, is in some senses the best possible limit distribution of a sequence of estimators of θ. It is known that under certain conditions the maximum likelihood estimators have this limit distribution, so that in some senses these are asymptotically optimal. Note, however, that the estimator $I_\theta^{-1}X$ is in general not 'best of all': for many loss functions it is inadmissible if $k \geq 3$ by Stein's (1956) and subsequent results on shrinkage estimation.

8.2 Smooth Functionals in i.i.d. Models; Tangent Spaces

Let $\mathscr{P}$ be a set of probability measures on a measurable space $(\mathscr{X}, \mathscr{B})$ and let the nth experiment consist of the observation of an independent identically distributed sample $X_1, \ldots, X_n$ from some $P \in \mathscr{P}$. Just as in § 8.1 we consider a sequence of 'local experiments' centered at a fixed $P \in \mathscr{P}$.

Let $\mathscr{P}(P)$ be a set of maps $t \to P_{t,h}$ ('paths') from an interval $(0, \varepsilon) \subset \mathbb{R}$ to $\mathscr{P}$ such that

$$\int\left[t^{-1}(dP_{t,h}^{\frac{1}{2}} - dP^{\frac{1}{2}}) - \tfrac{1}{2}h\,dP^{\frac{1}{2}}\right]^2 \to 0, \quad \text{as } t \downarrow 0, \tag{8.4}$$

for some measurable function $h:\mathscr{X}\to\mathbb{R}$. Every path in $\mathscr{P}(P)$ is a 'one-dimensional parametric submodel'. The 'local' difficulty of the given, possibly infinite dimensional, experiment will be assessed by considering every one-dimensional submodel in the classical way. Note that h is roughly the score function $\partial/\partial t_{|t=0}\log dP_{t,h}$ of the model $\{P_{t,h}: t \in (0, \varepsilon)\}$. Of course, the size of the set of different one-dimensional submodels and of the set of 'scores' h depends on the size of $\mathscr{P}$. The larger the set of scores, the harder the statistical problem will be.

It follows from (8.4) that $\int h\,dP = 0$ and that $\int h^2\,dP < \infty$. Denote the set of all measurable h with the second property by $\mathscr{L}_2(P)$. This is a pre-Hilbert space with respect to the inner product $\langle h_1, h_2\rangle = \int h_1 h_2\,dP$. Write $\|h\| = \langle h, h\rangle^{\frac{1}{2}}$ for its semi-norm. As usual let $L_2(P)$ be the set of equivalence classes of a.e. equal h in $\mathscr{L}_2(P)$. This is a Hilbert space with respect to the given inner product. Note that for a score h, its norm $\|h\|^2$ is the 'Fisher information' (about t) in the corresponding submodel $t \to P_{t,h}$.

Let $T(P) \subset L_2(P)$ be the set of all scores h obtained as a 'derivative' in the sense of (8.4). If the path $t \to P_{at}$ is in $\mathscr{P}(P)$ for every $a \geq 0$ whenever $t \to P_t$ is, then $T(P)$ is a cone: $ah \in T(P)$ whenever $a \geq 0$ and $h \in T(P)$. Usually $\mathscr{P}(P)$ can be and is chosen such that $T(P)$ is even a linear space, the *tangent space*. Just as in § 8.1, (8.4) implies asymptotic normality of the log likelihood ratios:

$$\log\prod_{j=1}^{n}\frac{dP_{1/\sqrt{n},h}}{dP}(X_j) = n^{-\frac{1}{2}}\sum_{j=1}^{n}h(X_j) - \tfrac{1}{2}\|h\|^2 + o_P(1). \tag{8.5}$$

Using Kolmogorov's extension theorem one can construct on a suitable probability space $(\Omega, \mathcal{A}, Q_0)$ a Gaussian process $\{\Delta_h : h \in L_2(P)\}$ with

$$E_{Q_0}\Delta_h = 0, \quad E_{Q_0}\Delta_{h_1}\Delta_{h_2} = \langle h_1, h_2\rangle. \tag{8.6}$$

Thus the Δ_h are random variables such that every finite set of them has a multivariate normal distribution with parameters specified by (8.6). Now define probability measures Q_h on $(\Omega, \mathcal{A})$ through their densities with respect to Q_0:

$$\frac{dQ_h}{dQ_0}(\omega) = e^{\Delta_h(\omega) - \frac{1}{2}\|h\|^2}. \tag{8.7}$$

Comparing (8.5) and (8.7), one sees that

$$\mathcal{E}_{n,P} := \left(\mathcal{X}^n, \mathcal{B}^n, \bigotimes_{j=1}^{n} P_{1/\sqrt{n},h} : h \in T(P)\right) \to \mathcal{E}_P := (\Omega, \mathcal{A}, Q_h : h \in T(P)).$$

The limit experiment $\mathcal{E}_P$ can also be described in several different forms. A very nice and simple one is as follows. For simplicity, assume that $T(P)$ is separable. Let $h_1, h_2, \ldots$ be an orthonormal base of a subspace of $L_2(P)$ that contains the closure of lin $T(P)$. (If $T(P)$ is finitely dimensional, one could actually take a finite number of h_i, but to make it not too easy, think of $T(P)$ as infinitely dimensional.) Then

$$h = \sum_{i=1}^{\infty} \langle h, h_i\rangle h_i,$$

and by (8.6)

$$\Delta_h = \sum_{i=1}^{\infty} \langle h, h_i\rangle \Delta_{h_i}$$

(where the convergence of the series can equivalently be taken in quadratic mean or almost sure sense). Substituting this in (8.7) one immediately sees that the vector $(\Delta_{h_1}, \Delta_{h_2}, \ldots)$ is sufficient for h in $\mathcal{E}_P$. But then $\mathcal{E}_P$ is equivalent to the experiment that consists of observing this vector. It is straightforward to compute that under Q_h, Δ_{h_1}, $\Delta_{h_2}, \ldots$ are independent, normally distributed random variables with means $\langle h, h_i\rangle$ and variances equal to 1. Therefore, it is also true that

$$\mathcal{E}_{n,P} \to (\mathbb{R}^\infty, \mathcal{B}_o^\infty, N_\infty((\langle h, h_i\rangle), I) : h \in T(P)). \tag{8.8}$$

In the set-up of this section one is typically interested in estimating a functional $\kappa(P)$ taking values in a normed space $\mathbf{D}$, that is differentiable in the sense of existence of a linear map κ'_P: lin $T(P) \to \mathbf{D}$ such that

$$t^{-1}(\kappa(P_{t,h}) - \kappa(P)) \to \kappa'_P(h),$$

for every path. This clearly implies

$$\sqrt{(n)}(\kappa(P_{1/\sqrt{n},h}) - \kappa(P)) \to \kappa'_P(h),$$

as in (4.2). Thus by Theorem 4.1 in the limit as $n \to \infty$ the difficulty of estimating $\kappa(P)$ is 'bounded from below' by the difficulty of estimating $\kappa'_P(h)$ in $\mathcal{E}_P$. Linearity and continuity of κ'_P gives that

$$\kappa'_P(h) = \sum_{i=1}^{\infty} \langle h, h_i\rangle \kappa'_P(h_i).$$

For the situation that $T(P)$ is a linear space one can next show that

$$\sum_{i=1}^{\infty} \Delta_{h_i} \kappa'_P(h_i)$$

is both a best equivariant-in-law and a minimax estimator for a wide class of loss functions, provided this series converges. Thus in the latter case the distribution of this series is in some senses the best possible limit distribution for an estimator sequence of $\kappa(P)$ based on $X_1, \ldots, X_n$. (For a real valued functional the series converges if and only if κ'_P is continuous. It then has a $N(0, \|\kappa'_P\|^2)$ distribution. In general continuity of κ'_P is necessary, but not sufficient for convergence.)

Another representation of the limit experiment is the one where one observes a Brownian P-bridge with drift. To motivate this, consider the *empirical distribution*

$$\hat{P}_n = n^{-1} \sum_{j=1}^{n} \varepsilon_{X_j},$$

i.e. the random measure putting mass n^{-1} at each of the observations X_j. Given a subset $F \subset \mathscr{L}_2(P)$ one can consider both P and $\hat{P}_n$ as elements of the space $l^\infty(F)$ of all functions $z: F \to \mathbb{R}$ with $\|z\| := \sup_{f \in F} |z(f)| < \infty$, through

$$P(f) = \int f\, dP, \quad \hat{P}_n(f) = \int f\, d\hat{P}_n = n^{-1} \sum_{j=1}^{n} f(X_j).$$

The set F is called a *P-Donsker class* if

$$\sqrt{(n)}(\hat{P}_n - P) \overset{P}{\Rightarrow} B_P, \quad \text{in } l^\infty(F), \tag{8.9}$$

where B_P is *F-indexed Brownian P-Bridge*, i.e. a tight, Borel measurable random element in $l^\infty(F)$ such that $(B_P(f_1), \ldots, B_P(f_k))$ is multivariate Gaussian for every finite subset $\{f_1, \ldots, f_k\} \subset F$, with

$$\mathbf{E} B_P(f) = 0, \quad \text{every } f \in F$$

$$\mathbf{E} B_P(f_1) B_P(F_2) = \int f_1 f_2\, dP - \int f_1\, dP \int f_2\, dP, \quad \text{every } f_1, f_2 \in F. \tag{8.10}$$

One can show by contiguity arguments that

$$\sqrt{(n)}(\hat{P}_n - P) \overset{h}{\Rightarrow} B_P(.) + \langle ., h \rangle, \tag{8.11}$$

where $\overset{h}{\Rightarrow}$ means weak convergence in $l^\infty(F)$ under $P_{1/\sqrt{n}, h}$.

Provided that F is not too small, the values of $X_1, \ldots, X_n$ can be regained from $\hat{P}_n$. Furthermore, in the local experiment $\mathscr{E}_{n,P}$, P is fixed, and hence can be considered known. Therefore $\sqrt{n}(\hat{P}_n - P)$ is 'sufficient' for $\mathscr{E}_{n,P}$. Then, if (8.11) holds, one may hope that the experiment consisting of observing the limit variable $B_P + \langle ., h \rangle$ if h is the true parameter, is a limit experiment for the sequence $\mathscr{E}_{n,P}$. This is roughly true.

It is necessary that the set F is large enough, but it shouldn't be too large, as otherwise (8.9) fails. An appropriate condition is that F is a P-Donsker class such that

$$\operatorname{lin} F \supset T(P). \tag{8.12}$$

Then

$$\mathscr{E}_{n,P} \to (l^\infty(F), \mathscr{B}_0, \mathscr{L}(B_P + \langle ., h \rangle) : h \in T(P)).$$

In the case that $(\mathscr{X}, \mathscr{B})$ equals the real line with Borel σ-field, an appropriate

P-Donsker class is

$$F = \{1_{(-\infty,u]} : u \in \mathbb{R}\}.$$

Then the limit experiment consists of observing

$$B \circ P(-\infty, .] + \int_{-\infty}^{\cdot} h\, dP,$$

where B is classical Brownian Bridge on the unit interval.

The set-up of this section follows Koshevnik & Levit (1976), Pfanzagl (1982) and van der Vaart (1989). These authors do not consider limit experiments, but give direct proofs of lower bound theorems for this special example. In semi-parametric models the class $\mathscr{P}$ is parametrised by a pair of a finite dimensional and an infinite dimensional parameter and one defines the 'tangent space' $T(P)$ in terms of the partial derivatives of P with respect to the parameters. There are many examples of such models. See Begun et al. (1983), van der Vaart (1991) and Bickel et al. (1990). Millar (1983) gives representations of the limit experiment in terms of 'abstract Wiener spaces'.

Rereading this section one can see that the independent identically distributed structure only matters in that it enables one to use a nice form of differentiability (8.4) together with the 'local parameter space' $T(P)$. However, what really matters is the Hilbert space structure of the parameter set and (8.5). The following is therefore an easy generalization. Let H be a subset of a Hilbert space (with inner product $\langle .,. \rangle$), with $0 \in H$ for simplicity. Let $\mathscr{E}_n = (\mathscr{X}_n, \mathscr{B}_n, P_{n,h} : h \in H)$ be experiments such that

$$\log \frac{dP_{n,h}}{dP_{n,0}}(X_n) = \Delta_{n,h}(X_n) - \tfrac{1}{2}\|h\|^2 + o_{P_{n,0}}(1),$$

$$\mathscr{L}_0(\Delta_{n,h_1}, \ldots, \Delta_{n,h_k}) \to \mathscr{L}_0(\Delta_{h_1}, \ldots, \Delta_{h_k}), \quad \text{every } h_1, \ldots, h_k \in H,$$

where $\{\Delta_h : h \in \operatorname{lin} H\}$ is a Gaussian process (on $(\Omega, \mathscr{A}, Q_0)$) satisfying (8.6). Then, with notation as before,

$$\mathscr{E}_n \to \mathscr{E} = (\Omega, \mathscr{A}, Q_h : h \in H).$$

Here the limit experiment $\mathscr{E}$ can be replaced by a simpler representation, as before.

This generalization may be straightforward, it is also extremely useful. It applies to many examples where one observes stochastic processes, such as time series and the counting processes of lifetime analysis.

8.3 Local Asymptotic Mixed Normality

Let Θ be an open subset of $\mathbb{R}^m$. For $\theta \in \Theta$ and $h \in \mathbb{R}^m$ let $P_{\theta,h}$ be the distribution of a random vector $(\Delta_\theta, \Sigma_\theta)$ satisfying $\mathscr{L}_h(\Delta_\theta \mid \Sigma_\theta = \sigma) = N(\sigma h, \sigma)$ and $\mathscr{L}(\Sigma_\theta)$ not dependent on h and concentrating on the set of positive definite matrices $\mathscr{M}$. Let $(X_n, \mathscr{B}_n, P_{n,\theta} : \theta \in \Theta)$ be experiments and suppose that for a sequence of norming constants $R_n \to \infty$, sequences of nonsingular matrices $S_{n,\theta}$, and every θ

$$\log \frac{dP_{n,\theta+S_{n,\theta}h/R_n}}{dP_{n,\theta}} = h'\Delta_{n,\theta} - \tfrac{1}{2}h'\Sigma_{n,\theta}h + o_{P_{n,\theta}}(1), \tag{8.13}$$

for measurable maps $(\Delta_{n,\theta}, \Sigma_{n,\theta})$ with

$$\mathscr{L}_\theta(\Delta_{n,\theta}, \Sigma_{n,\theta}) \to \mathscr{L}_0(\Delta_\theta, \Sigma_\theta). \tag{8.14}$$

Such a sequence of experiments is called *locally asymptotically mixed normal*. The (local) limit experiment is the one where one observes the pair $(\Delta_\theta, \Sigma_\theta)$ with law $P_{\theta,h}$. One has

$$\mathscr{E}_{n,\theta} := (\mathscr{X}_n, \mathscr{B}_n, P_{n,\theta+S_{n,\theta}h/R_n} : h \in \mathbb{R}^m) \to \mathscr{E}_\theta := (\mathbb{R}^m \times \mathscr{M}, \mathscr{B}_o, P_{\theta,h} : h \in \mathbb{R}^m).$$

The expansion (8.13) can be thought of as a Taylor expansion. The convergence (8.14) would usually follow from a central limit theorem for martingales.

Local asymptotic mixed normality is studied by Jeganathan (1981). Concrete examples are discussed in Basawa & Scott (1980). Also see Hall & Heyde (1980).

8.4 *Binomial Distribution with Small Probability*

Suppose one records the number of successes X_n in a sequence of n Bernoulli trials with success probability p. Set $P_{n,p} = \mathscr{L}_p(X_n)$. The usual asymptotic approximation for this model views this experiment as the nth in the sequence of experiments $(\mathbb{R}, \mathscr{B}_o, P_{n,p_0+h/\sqrt{n}} : h \in \mathbb{R})$, for a given, fixed p_0. The resulting situation is a special case of the one of § 8.1. Alternatively, if there is reason to believe that the true p is close to zero, one might rather embed the experiment in a sequence of experiments with p converging to zero.

Let X be a random variable distributed according to the Poisson distribution Q_h given through the density $dQ_h(x) = e^{-h}h^x/x!$ with respect to counting measure on the nonnegative integers. Then

$$\mathscr{E}_n := (\mathbb{R}, \mathscr{B}_o, P_{n,h/n} : h > 0) \to \mathscr{E} := (\mathbb{R}, \mathscr{B}_o, Q_h : h > 0).$$

8.5 *Uniform Distribution*

Fix two numbers σ and τ with $\sigma < \tau$. The nth experiment consists of observing a sample $X_1, \ldots, X_n$ from a uniform distribution on $[\sigma, \tau]$. As in § 8.1 convergence of localized experiments will be considered, though now the 'rate' is n rather than $\sqrt{n}$.

In the local experiment

$$\left(\mathbb{R}^n, \mathscr{B}_o^n, \bigotimes_{j=1}^{n} U[\sigma + s/n, \tau - t/n] : (s, t) \in \mathbb{R}^2\right)$$

the pair formed by $V_n = n(X_{(1)} - \sigma)$ and $W_n = -n(X_{(n)} - \tau)$ is sufficient for (s, t) and satisfies

$$\mathscr{L}_{s,t}(V_n, W_n) \to \mathscr{L}_{s,t}(V, W), \quad \text{every } (s, t) \in \mathbb{R}^2, \tag{8.15}$$

for (independent) random variables V and W with joint Lebesgue density on $\mathbb{R}^2$

$$(\tau - \sigma)^{-2} e^{-(\tau-\sigma)^{-1}(v-s)} e^{-(\tau-\sigma)^{-1}(w-t)} 1_{v>s} 1_{w>t}.$$

It is therefore no surprise that

$$\mathscr{E}_n := (\mathbb{R}^n, \mathscr{B}_o^n, P_{n,(s,t)}^{\sigma,\tau} : (s, t) \in \mathbb{R}^2) \to \mathscr{E} := (\mathbb{R}^2, \mathscr{B}_0^2, \mathscr{L}_{s,t}(V, W) : (s, t) \in \mathbb{R}^2).$$

8.6 *Densities with Jump from Zero*

Let the probability density p on $\mathbb{R}$ be differentiable at every $x \neq 0$ with derivative satisfying $\int |p'(x)|\, dx < \infty$. Furthermore assume that $p(0+) = \lim_{x \downarrow 0} p(x) > 0$, whereas

$p(0-)=\lim_{x\uparrow 0}p(x)=0$. Let the nth experiment $\mathscr{E}_n$ consist of the observation of an independent identically distributed sample $X_1,\ldots,X_n$ from the density $x\to p(x-h/n)$ where $h\in\mathbb{R}$.

Ibragimov & Has'minskii (1981, Ch. V) show that for every $h_0<h$

$$\prod_{j=1}^{n}\frac{p(X_j-h/n)}{p(X_j-h_0/n)}=e^{(h-h_0)p(0+)}1_{Z_n>h}+o_{P_{n,h_0}}(1), \tag{8.16}$$

where $Z_n=n\min\{X_j:X_j>h/n\}$ satisfies

$$\mathscr{L}_{h_0}(Z_n)\to\mathscr{L}_{h_0}(Z), \tag{8.17}$$

for Z with a shifted exponential distribution with Lebesgue density $p(0+)e^{-p(0+)(z-h)}1_{z>h}$. This can be seen to imply

$$\left(\mathbb{R}^n,\mathscr{B}_o^n,\bigotimes_{j=1}^{n}p(x_j-h/n):h\in\mathbb{R}\right)\to(\mathbb{R},\mathscr{B}_o,\mathscr{L}_h(Z):h\in\mathbb{R}).$$

8.7 *Densities with Jumps*

A limit experiment need not always be simple. In this section it consists of observing a number of Poisson processes.

Let p be a probability density with respect to Lebesgue measure on the real line that is differentiable at every x except possibly at the points $a_1,a_2,\ldots,a_r$. Assume that the derivative satisfies $\int|p'(x)|\,dx<\infty$ and that p has finite left and right limits $p(a_i-)$ and $p(a_i+)$ at every a_i. The nth experiment $\mathscr{E}_n$ consists of the observation of a sample $X_1,\ldots,X_n$ from the density $x\to p(x-h/n)$, where h ranges over $\mathbb{R}$.

For $i=1,\ldots,r$ and every h let $\mu_{i,h}$ be continuous (piecewise linear) functions with $\mu_{i,h}(0)=0$ and derivative satisfying

$$\mu'_{i,h}(t)=\begin{cases}p(a_i-) & \text{if } t<h,\\ p(a_i+) & \text{if } t>h.\end{cases} \tag{8.18}$$

Let $N_1,\ldots,N_r$ be r independent two-sided, standard Poisson processes, i.e. every $\{N_i(t):t\in\mathbb{R}\}$ is a counting process with independent increments, $N_i(0)\equiv 0$, and $N_i(t)-N_i(s)$ has a Poisson distribution with mean $t-s$ for every $s<t$ and i. Set P_h equal to the distribution on $D(-\infty,\infty)^r$ of the multivariate counting process $(N_1\circ\mu_{1,h},\ldots,N_r\circ\mu_{r,h})$. Write $P_{n,h}$ for the distribution of $(X_1,\ldots,X_n)$ under h. Using results of Ibragimov & Has'minskii (1981, Ch. V), one can show that

$$\mathscr{E}_n=(\mathbb{R}^n,\mathscr{B}_o^n,P_{n,h}:h\in\mathbb{R})\to(D(-\infty,\infty)^r,\mathscr{B}_o^r,P_h:h\in\mathbb{R}). \tag{8.19}$$

9 Randomized Estimators and Compactness

The results of this section are technical and are preparation for the proof of Theorem 3.1 in § 10.

An alternative notion of randomized estimator is that of a *Markov kernel* $(x,A)\to\delta_x(A)=\delta(x,A)$ from $(\mathscr{X},\mathscr{B})$ into $(\mathbf{D},\mathscr{D}_o)$:

(i) for every fixed $x\in\mathscr{X}$, δ_x is a probability measure on $(\mathbf{D},\mathscr{D}_o)$,
(ii) for every fixed $A\in\mathscr{D}_o$ the map $x\to\delta_x(A)$ is measurable from $(\mathscr{X},\mathscr{B})$ into $\mathbb{R}$.

It will first be shown that for Polish decision spaces this notion is equivalent to that of a randomized estimator as introduced in § 3. The following result is similar to that of Wald & Wolfowitz (1951).

LEMMA 9.1. *Let* $\mathbf{D}$ *be a complete, separable, metric space with Borel* σ*-field* $\mathcal{D}_o$. *Let* δ *be a Markov kernel from* $(\mathcal{X}, \mathcal{B})$ *into* $(\mathbf{D}, \mathcal{D}_o)$. *Then there exists a measurable map* $t:(\mathcal{X}\times[0,1], \mathcal{B}\times\mathcal{B}_o)\to(\mathbf{D}, \mathcal{D}_o)$ *with*

$$\delta(x, A) = P(t(x, U)\in A), \quad \text{every } A\in\mathcal{D}_o, \tag{9.1}$$

where the random variable U *has a uniform distribution on* $[0, 1]$. *Conversely, given a map* t *as above,* (9.1) *defines a Markov kernel* δ *from* $(\mathcal{X}, \mathcal{B})$ *into* $(\mathbf{D}, \mathcal{D}_o)$.

Proof. For the second assertion one only needs to note that by Fubini's theorem the maps

$$x \to P(t(x, U)\in A) = \int 1_A(t(x, u))\,du$$

are measurable for every measurable set A.

There exist a Borel set $B\subset\mathbb{R}$ and a bijection $\phi:\mathbf{D}\to B$ which is a Borel isomorphism, i.e. both ϕ and its inverse ϕ^{-1} are Borel measurable. (See e.g. Dudley (1989). If $\mathbf{D}$ is uncountable, then one can even take $B=\mathbb{R}$; if $\mathbf{D}$ is countable the statement is trivial.) It is straightforward to see that this implies that it is sufficient to prove the theorem for $\mathbf{D}=\mathbb{R}$.

For simplicity of notation, let (X, Y) be a random element in $\mathcal{X}\times\mathbb{R}$ with

$$\mathcal{L}(Y\mid X=x)=\delta_x$$

for every x. Let $H(.\mid x)$ be the quantile function of $\mathcal{L}(Y\mid X=x)$. Then $H(u\mid x)$ is right continuous in u for every fixed x and measurable in x for every fixed u, because

$$\{H(u\mid x)\leq a\}=\{u\leq P(Y\leq a\mid X=x)\}.$$

Thus H is measurable in its two arguments. Set $t(x, u)=H(u\mid x)$. Then by the quantile transformation

$$\mathcal{L}(t(x, U))=\mathcal{L}(Y\mid X=x)=\delta_x. \qquad \square$$

The second result of this section is concerned with compactness of the set of randomized estimators under certain weak topologies. In a statistical context this goes back to Le Cam (1955).

THEOREM 9.2. *Let* $\mathbf{D}$ *be a compact, metric space with Borel* σ*-field* $\mathcal{D}_o$ *and let* $\mathcal{P}$ *be a set of probability measures on a measurable space* $(\mathcal{X}, \mathcal{B})$, *dominated by a* σ*-finite measure. Then the set of all Markov kernels* δ *from* $(\mathcal{X}, \mathcal{B})$ *into* $(\mathbf{D}, \mathcal{D}_o)$ *is compact under the weak topology generated by the collection of all maps of the form*

$$(f, P)\to\iint f(y)\delta(x, dy)\,dP(x), \tag{9.2}$$

where $f\in C_b(\mathbf{D})$ *and* $P\in\mathcal{P}$.

Proof. Set $\Gamma=C_b(\mathbf{D})$ and let $\mathbf{L}$ be the vector space of all signed measures on $(\mathcal{X}, \mathcal{B})$ that are dominated by a σ-finite measure that dominates $\mathcal{P}$. For $Q\in\mathbf{L}$ write $\|Q\|$ for the total variation norm. The space

$$\mathbf{Z}=\prod_{(f,Q)\in\Gamma\times\mathbf{L}}[-\|f\|_\infty\|Q\|, \|f\|_\infty\|Q\|]$$

is by definition the set of all functions $z:\Gamma \times \mathbf{L} \to \mathbb{R}$ with $|z(f, Q)| \leqslant \|f\|_\infty \|Q\|$ for every $f \in \Gamma$ and $Q \in \mathbf{L}$. Every Markov kernel δ induces an element $z_\delta \in \mathbf{Z}$ through

$$z_\delta(f, Q) = \int\int f(y)\delta(x, dy)\, dQ(x).$$

Let δ_α be a net of Markov kernels. By Tychonov's theorem $\mathbf{Z}$ is compact in the product topology. Hence the net z_{δ_α} has a subnet that converges to some $z \in \mathbf{Z}$, in the sense that

$$z_{\delta_{\alpha(\beta)}}(f, Q) \to z(f, Q) \tag{9.3}$$

for every $(f, Q) \in \Gamma \times \mathbf{L}$. From the properties of the Markov kernels z_δ and (9.3) one sees that z satisfies:

(a) z is bilinear,
(b) $z(f, Q) \geqslant 0$ if $f \geqslant 0$ and $Q \geqslant 0$,
(c) $z(1, Q) = Q(\mathscr{X})$.

Moreover, since $z \in \mathbf{Z}$:

(d) $|z(f, Q)| \leqslant \|f\|_\infty \|Q\|$.

But then there exists a Markov kernel δ such that $z = z_\delta$ (see e.g. Strasser (1985, Theorem 6.11)). For this δ one certainly has that $\delta_{\alpha(\beta)} \to \delta$ in the weak topology generated by the maps (9.2). Thus every net δ_α has a converging subnet, whence the compactness. □

10 Proof of Theorem 3.1

It will be shown that there exists a Markov kernel δ from $(\mathscr{X}, \mathscr{B})$ into $(\mathbf{D}, \mathscr{D}_o)$ such that

$$Q_h = \int \delta_x \, dP_h(x), \quad \text{every } h \in H. \tag{10.1}$$

It is clear from (3.1b) that t and δ in (3.2) and (10.1) can without loss of generality be assumed to take values in $\mathbf{D}_0$ or be concentrated on $\mathbf{D}_0$, respectively. Therefore the assertions (3.2) and (10.1) are equivalent by Lemma 9.1.

(i) Assume first that H consists of finitely many elements. Set

$$\mu_\alpha = \frac{1}{|H|} \sum_{h \in H} P_{\alpha,h}, \quad V_\alpha = \left(\frac{dP_{\alpha,h}}{d\mu_\alpha}\right)_{h \in H},$$

and let μ be the corresponding object in the limit experiment. By (3.1) and (2.3), (T_α, V_α) converges marginally under μ_α to tight limit laws on $(\mathbf{D}, \mathscr{D}_o)$ and $\mathbb{R}^H$, respectively. (Tightness follows from (3.1b).) By an argument using Prohorov's theorem (apart from the possible nonmeasurability, it is that marginal tightness implies joint tightness, and hence the existence of a converging subnet; see Theorem 2 in the Appendix) there exists a subnet of $\{\alpha\}$ (abusing notation denoted $\{\alpha\}$) such that

$$(T_\alpha, V_\alpha) \Rightarrow (T, V), \quad \text{under } \mu_\alpha, \tag{10.2}$$

where (T, V) is a random element in $(\mathbf{D} \times \mathbb{R}^H, \mathscr{D}_o \times \mathscr{B}_o^H)$ and

$$\mathscr{L}(V) = \mathscr{L}_\mu\left(\left(\frac{dP_h}{d\mu}\right)_{h \in H}\right),$$

by (2.3). Without loss of generality assume that T takes its values in $\mathbf{D}_0$ and V in $\tilde{\mathbf{S}}_H := \{x \in \mathbb{R}^H : x \geqslant 0, \sum x_h = |H|^{-1}\}$.

Given $f \in C_b(\mathbf{D})$ and $h \in H$, the function $(t, v) \to f(t)v_h$ is contained in $C_b(\mathbf{D} \times \tilde{\mathbf{S}}_H)$. Hence

$$E_h^* f(T_\alpha) = E_{\mu_\alpha}^* f(T_\alpha) \frac{dP_{\alpha,h}}{d\mu_\alpha} \to Ef(T)V_h.$$

Combination with (3.1) gives $\int f\, dQ_h = Ef(T)V_h$. Thus for $A \in \mathcal{D}_o$

$$Q_h(A) = E1_A(T)V_h. \tag{10.3}$$

Set

$$\delta(x, A) = P\left(T \in A \mid V = \left(\frac{dP_h}{d\mu}(x)\right)_{h \in H}\right),$$

where $P(T \in . \mid V = v)$ is a regular version of the conditional law of T given $V = v$. (This exists, because T takes its values in the Polish space $\mathbf{D}_0$.) Now

$$\begin{aligned}\int \delta_x(A)\, dP_h(x) &= \int \delta_x(A) \frac{dP_h}{d\mu}(x)\, d\mu(x)\\ &= EP(T \in A \mid V)V_h = E1_A(T)V_h.\end{aligned} \tag{10.4}$$

Combination of (10.3) and (10.4) completes the proof for the case that $|H|$ is finite.

(ii) Now let H be arbitrary. By (i) there exists for every finite subset I of H a Markov kernel δ^I such that

$$Q_h = \int \delta_x^I\, dP_h(x), \quad \text{every } h \in I. \tag{10.5}$$

Moreover, δ^I can be constructed such that $\delta_x^I(\mathbf{D}_0) = 1$ for every x. Since $\mathbf{D}_0$ is Polish, it has a metrizable compactification $\bar{\mathbf{D}}_0$ such that $\mathbf{D}_0$ is a Borel set in $\bar{\mathbf{D}}_0$. (If $\mathbf{D} = \mathbf{D}_0 = \mathbb{R}^k$ one can simply take $\bar{\mathbf{D}}_0$ equal to the usual compactification $\bar{\mathbb{R}}^k$. In general, $\mathbf{D}_0$ can be homeomorphically identified with a G_δ subset of a countable product of unit intervals [0, 1] (Dudley, 1989). Next one can take $\bar{\mathbf{D}}_0$ equal to its closure there.)

View every δ^I as a Markov kernel from $(\mathcal{X}, \mathcal{B})$ into $(\bar{\mathbf{D}}_0, \mathcal{B}_o)$. By Theorem 9.2 the set of all such Markov kernels is compact with respect to the weak topology generated by the maps of the form

$$\delta \to \iint f(y)\delta(x, dy)\, dP_h(x), \quad f \in C_b(\bar{\mathbf{D}}_0), h \in H.$$

Direct the finite subsets of H by $I \leqslant J$ if $I \subset J$. Because of the compactness, the net $\{\delta^I : I \subset H, \text{ finite}\}$ has a subnet $\{\delta^{I(\beta)} : \beta \in B\}$ converging to a Markov kernel δ from $(\mathcal{X}, \mathcal{B})$ into $(\bar{\mathbf{D}}_0, \mathcal{B}_o)$, in the sense that

$$\iint f(y)\delta^{I(\beta)}(x, dy)\, dP_h(x) \to \iint f(y)\delta(x, dy)\, dP_h(x),$$

for every $f \in C_b(\bar{\mathbf{D}}_0)$ and $h \in H$. Thus

$$\int \delta_x^{I(\beta)}\, dP_h(x) \Rightarrow \int \delta_x\, dP_h(x), \tag{10.6}$$

for every $h \in H$.

By the definition of a subnet, there exists for any given finite $I \subset H$ a $\beta_0 \in B$ such that $I \subset I(\beta)$ for every $\beta \geqslant \beta_0$. But then, by (10.5), the left hand side of (10.6) equals Q_h for every $\beta \geq \beta_0$ and $h \in I$. Conclude that

$$Q_h = \int \delta_x \, dP_h(x), \quad \text{every } h \in H. \tag{10.7}$$

More precisely, this is true for Q_h seen as a measure on $(\bar{\mathbf{D}}_0, \mathscr{B}_o)$. By (10.7)

$$1 = Q_h(\mathbf{D}_0) = \int \delta_x(\mathbf{D}_0) \, dP_h(x), \quad \text{every } h \in H.$$

Thus $P_h(x : \delta_x(\mathbf{D}_0) < 1) = 0$ for every $h \in H$. Therefore, if necessary, δ can be redefined on $\{x : \delta_x(\mathbf{D}_0) < 1\}$ so as to correspond to a Markov kernel into $(\mathbf{D}_0, \mathscr{B}_o)$, without affecting the validity of (10.7). The resulting Markov kernel can be viewed as a Markov kernel into $(\mathbf{D}, \mathscr{D}_o)$ too, of course. □

Note. The second part of this proof can also be based on a different idea of compactifying. Since $\mathbf{D}_0$ is Polish there exists a metric on $\mathbf{D}_0$ (possibly inducing a topology different from the given one) for which $\mathbf{D}_0$ is compact and for which the Borel σ-field equals the given Borel σ-field (Dudley, 1989). One can next apply Theorem 9.2 to $\mathbf{D}_0$ with this metric. This manner of compactifying gives a slightly smoother proof, but is based on a somewhat deeper result. In particular, existence of a metric as stated is not trivial for the case that $\mathbf{D}_0 = \mathbb{R}^k$, whereas existence of $\bar{\mathbb{R}}^k$ is.

Acknowledgements

I thank Maryse Loranger for inspiring discussions and Lucien Le Cam for several of the references. The research was partially supported by the Netherlands Organization for the Advancement of Pure Research (Z.W.O.) and ONR Contract N00014-80-C-0163 and NSF Grant DMS-8723011.

Appendix: Two Weak Convergence Results

This appendix contains two results on weak convergence, neither of which is surprising. However, since they are used a couple of times in proofs of results in the paper, it is useful to have precise statements for reference. One wouldn't find these results in the literature, as they are concerned with weak convergence of possibly nonmeasurable maps under a *set* of possible probability laws, rather than under one given law. Proofs are as usual, except for the use of Tychonov's theorem ('product of compacts is compact'), and hence basically omitted.

Let $(\mathscr{X}_\alpha, \mathscr{B}_\alpha, P_{n,h} : h \in H)$ be a net of experiments and $T_\alpha : \mathscr{X}_\alpha \to \mathbf{D}$ maps in a metric space $\mathbf{D}$. Say that the net T_α converges in distribution under h to a Borel measure Q_h on $\mathbf{D}$ if

$$E_h^* f(T_\alpha) \to \int f \, dQ_h, \quad \text{every } f \in C_b(\mathbf{D}).$$

Denote this by

$$T_\alpha \overset{h}{\Rightarrow} Q_h.$$

Call the net T_α *asymptotically tight* under h if and only if for every $\varepsilon > 0$ there exists a compact subset $K_\varepsilon \subset \mathbf{D}$ such that, with $K_\varepsilon^\delta = \{y : d(y, K_\varepsilon) < \delta\}$

$$\liminf_\alpha P_{\alpha,h,*}(T_\alpha \in K_\varepsilon^\delta) \geq 1 - \varepsilon.$$

The first result is a version of Prohorov's theorem.

THEOREM 1. *Suppose that T_α is asymptotically tight under every $h \in H$ and satisfies*

$$E_{hJ}^* f(T_\alpha) - E_{h^*} f(T_\alpha) \to 0, \quad \text{every } h \in H.$$

Then there exists a subnet and tight Borel measures Q_h on $\mathbf{D}$ such that

$$T_{\alpha(\beta)} \overset{h}{\Rightarrow} Q_h, \quad \text{every } h \in H.$$

Proof. Let $F = C_b(\mathbf{D})$. Consider $\{E_{hJ}^* f(T_\alpha)\}_{h\in H, f\in F}$ as a net in the product space $\prod_{h\in H}\prod_{f\in F}[-\|f\|_\infty, \|f\|_\infty]$. By Tychonov's theorem the latter space is compact in the product topology. Therefore the given net has a converging subnet. This is equivalent to the existence of a subnet and numbers $L_h(f) \in [-\|f\|_\infty, \|f\|_\infty]$ such that

$$E_{hJ}^* f(T_{\alpha(\beta)}) \overset{\beta}{\to} L_h(f), \quad \text{every } h \in H, f \in C_b(\mathbf{D}).$$

Now fix some h. By Prohorov's theorem the subnet $T_{\alpha(\beta)}$ has a further subnet that converges weakly under h to some Borel measure Q_h. Taken together this gives that $L_h(f) = \int f \, dQ_h$. But then $T_{\alpha(\beta)} \Rightarrow Q_h$ under h along the first extracted subnet. □

The second result shows that, given marginal convergence of the components of a vector, one can extract a subnet that converges jointly.

THEOREM 2. *Let T_α and T'_α be nets of maps from $\mathscr{X}_\alpha$ into metric spaces $\mathbf{D}$ and $\mathbf{D}'$, respectively, such that*

$$T_\alpha \overset{h}{\Rightarrow} Q_h, \quad T'_\alpha \overset{h}{\Rightarrow} Q'_h, \quad \text{every } h \in H.$$

where Q_h and Q'_h are tight Borel measures on $\mathbf{D}$ and $\mathbf{D}'$. Then there exists a subnet, indexed by $\alpha(\beta)$, and tight Borel measures M_h on $\mathbf{D} \times \mathbf{D}'$ with

$$(T_{\alpha(\beta)}, T'_{\alpha(\beta)}) \overset{h}{\Rightarrow} M_h \quad \text{every } h \in H.$$

Proof. As in the proof of Corollary 1.1 of van der Vaart & Wellner (1989)

$$E_{hJ}^* f(T_\alpha, T'_\alpha) - E_{h^*} f(T_\alpha, T'_\alpha) \to 0, \quad \text{every } h \in H, f \in C_b(\mathbf{D} \times \mathbf{D}').$$

Moreover, the net (T_α, T'_α) is asymptotically tight in $\mathbf{D} \times \mathbf{D}'$. Now apply Theorem 1. □

References

Andersen, N.T. & Dobric, V. (1987). The central limit theorem for stochastic processes. *Ann. Prob.* **15,** 164–177.

Araujo, A. & Giné, E. (1980). *The Central Limit Theorem for Real and Banach Valued Random Variables.* New York: Wiley.

Barndorff-Nielsen, O.E., James, I.R. & Leigh, G.M. (1988). Asymptotic inference for a semi-parametric estimator of natural mortality. Unpublished manuscript.

Basawa, I.V. & Scott, D.J. (1980). *Asymptotic Optimal Inference for Nonergodic Models,* Lecture Notes in Statistics **17**. New York: Springer-Verlag.

Begun, J.M., Hall, W.J., Huang, W.M. & Wellner, J.A. (1983). Information and asymptotic efficiency in parametric-nonparametric models. *Ann. Statist.* **11,** 432–452.

Bickel, P.J., Klaassen, C.A.J., Ritov, Y. & Wellner, J.A. (1990). Efficient and Adaptive Inference in Semi-Parametric Models. Unpublished manuscript.

Billingsley, P. (1968). *Convergence of Probability Measures.* New York: Wiley.

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. (1985). An extended Wichura theorem, definition of Donsker class, and weighted empirical distributions, Lecture Notes in Mathematics **1153,** pp. 141–178.

Dudley, R.M. (1989). *Real Analysis and Probability*. Pacific Grove, Calif.: Wadsworth & Brooks/Cole.
Gaenssler, P. (1983). *Empirical Processes*. Hayward, Calif.: Institute of Mathematical Statistics.
Groeneboom, P. (1987). Asymptotics for incomplete censored observations. Unpublished manuscript.
Hájek, J. (1970). A characterization of limiting distributions of regular estimators. *Z. Wahr. Th. Verw. Geb.* **14,** 323–330.
Hájek, J. (1972). Local asymptotic minimax and admissibility in estimation. In *Proc. 6th Berkeley Symp. Math. Statist. Prob.* **1,** pp. 175–194. Berkeley: University of California Press.
Hall, P. & Heyde, C.C. (1980). *Martingale Limit Theory and its Application*. New York: Academic Press.
Hoffmann-Jorgensen, J. (1984). Unpublished manuscript.
Ibragimov, I.A. & Has'minskii, R.Z. (1981). *Statistical Estimation: Asymptotic Theory*. New York: Springer.
Jeganathan, P. (1981). On a decomposition of the limit distribution of a sequence of estimators. *Sankhyā* A**43,** 26–36.
Kelley, J.L. (1955). *General Topology*. Princeton: Van Nostrand.
Koshevnik, Y.A. & Levit, B.Y. (1976). On a nonparametric analogue of the information matrix. *Theory Prob. Applic.* **21,** 738–753.
Le Cam, L. (1955). An extension of Wald's theory of statistical decision functions. *Ann. Math. Statist.* **26,** 69–81.
Le Cam, L. (1964). Sufficiency and asymptotic sufficiency. *Ann. Math. Statist.* **35,** 1419–1455.
Le Cam, L. (1970). On the assumptions used to prove asymptotic normality of maximum likelihood estimates. *Ann. Math. Statist.* **41,** 802–828.
Le Cam, L. (1972). Limits of experiments. In *Proc. 6th Berkeley Symp. Math. Statist. Prob.* **1,** pp. 245–261. Berkeley: University of California Press.
Le Cam, L. (1986). *Asymptotic Methods in Statistical Decision Theory*. New York: Springer Verlag.
Le Cam, L. (1989). On measurability and convergence in distribution. Technical Report 211. Dept. Statist., Univ. of California, Berkeley.
Millar, P. W. (1983). The minimax principle in asymptotic statistical theory. *École d'Été de Probabilités de Saint-Flour XI, 1981,* Lecture Notes in Mathematics **976,** pp. 67–267. New York: Springer-Verlag.
Millar, P. W. (1985). Nonparametric applications of an infinite dimensional convolution theorem. *Z. Wahr. Th. Verw. Geb.* **68,** 545–556.
Pfanzagl, J. (1982). *Contributions to a General Asymptotic Statistical Theory,* Lecture Notes in Statistics **13**. New York: Springer Verlag.
Pollard, D. (1984). *Convergence of Stochastic Processes*. New York: Springer Verlag.
Stein, C. (1956). Inadmissibility of the usual estimator for the mean of a normal distribution. In *Proc. 3rd Berkeley Symp. Math. Statist. Prob.* **1,** pp. 197–206. Berkeley: University of California Press.
Strasser, H. (1985). *Mathematical Theory of Statistics*. Berlin: Walter de Gruyter.
Vaart A. W. van der (1989). On the asymptotic information bound. *Ann. Statist.* **17,** 1487–1500.
Vaart, A. W. van der (1991). On differentiable functionals. *Ann. Statist.* **19**. To appear.
Vaart, A.W. van der & Wellner, J.A. (1989). Convolution and asymptotic minimax theorems for non-measurable 'estimators'. Unpublished report.
Wald, A. & Wolfowitz, J. (1951). Two methods of randomization in statistics and the theory of games. *Ann. Math. Statist.* **53,** 581–586.

Résumé

Soit une suite d'expériences stochastiques convergeant vers une expérience limite au sens de Le Cam (1972). Supposons aussi qu'une suite de statistiques a une distribution limite pour toute valeur possible du paramètre, alors l'ensemble des distributions limites est l'ensemble des distributions d'un estimateur randomisé dans l'expérience limite. Ceci exprime de façon simple que l'expérience limite est une 'borne inférieure' pour la suite convergeante d'expériences. De plus, l'on peut obtenir comme corollaires des résultats minimax et des théorèmes de convolution.

[*Received October* 1989, *accepted September* 1990]