JOURNAL OF MULTIVARIATE ANALYSIS 24, 237–251 (1988)

A Characterization of Translation-Invariant Experiments Admitting Adaptive Estimates

GEORG CH. PFLUG

Justus-Liebig University, Giessen, West Germany

AND

WILFRIED GROSSMANN

University of Vienna, Vienna, Austria

Communicated by P. R. Krishnaiah

Statistical experiments possess the property of adaptivity, if the ignorance of a nuisance parameter does not cause any loss in efficiency. In order to include a large variety of cases, the efficiency is measured in terms of minimax bounds. It is shown that a necessary and sufficient condition for adaptivity of translation invariant experiments is that the parameter of interest and the nuisance parameter are a.s. independent w.r.t. the posterior distributions.

1. Introduction

Many statistical models are of the following type: They contain a finite dimensional parameter of interest $t \in \mathbb{R}^k$ and a (finite or infinite dimensional) nuisance parameter u. It is known that under some circumstances the ignorance of the nuisance parameter u causes no loss in efficiency for the estimation of t, i.e., the asymptotic properties of the best estimates are the same both for fixed known u and for unknown u. The term "adaptivity" is used to describe such a situation and the pertaining estimates for t are called "adaptive." Thus an experiment has the property of adaptivity if at least one adaptive estimate exists.

Received June 25, 1986; revised October 2, 1986.

AMS 1980 subject classifications: 62E20, 62G05, 62G20.

Key words and phrases: adaptivity, minimax bounds, Pitman estimates, semi-parametric models.

0047-259X/88 $3.00

The notion of adaptivity depends on the used measure of efficiency. If the asymptotic distribution of the estimates is normal, then the most important and widely used measure of efficiency is the asymptotic covariance matrix, since by Anderson's lemma minimizing the covariance matrix means at the same time minimizing the risk for a large class of loss functions. For non-normal asymptotic distributions, however, asymptotic variances need not exist and if they do they are only related to the risk of the quadratic loss. Therefore the concept of adaptivity will be based on the comparison of minimax risks for a variety of loss functions:

Let $\{P_{t,u}\colon t \in T = \mathbb{R}^k,\ u \in U\}$ be a family of probability measures on a probability space $(\Omega, \mathfrak{A})$. The parameter $t \in \mathbb{R}^k$ is that of interest and U is an arbitrary set of nuisance parameters. Let $W(\cdot)$ be a nonnegative loss function on $\mathbb{R}^k$ and $\tau(\omega, dy)$ a (randomized) estimate of t, i.e., a transition from Ω into the set of all probability measures in $\mathbb{R}^k$. The set of all these estimates is denoted by $\mathscr{T}$. The risk of τ is

$$r_W(\tau, t, u) := \int_\Omega \int_{\mathbb{R}^k} W(t-y)\, \tau(\omega, dy)\, dP_{t,u}(\omega).$$

(1.1) DEFINITION. The experiment $(\Omega, \mathfrak{A}, \{P_{t,u}\colon t \in T, u \in U\})$ fulfills condition (A) (adaptivity) with respect to the loss W if for every finite subset U_1 of U

$$\text{(A)} \quad \inf_{\tau^* \in \mathscr{T}} \sup_{u \in U_1} \left[\sup_{t \in T} r_W(\tau^*, t, u) - \inf_{\tau \in \mathscr{T}} \sup_{t \in T} r_W(\tau, t, u) \right] = 0.$$

Since the left-hand side of (A) is obviously always nonnegative, this notion of adaptivity may be loosely characterized as: There is an estimate τ^* treating u as unknown, which achieves for every u the same minimax bound as those procedures τ which make use of the knowledge of u.

It is easy to conclude from (A) the validity of the following condition

(B) For every finite subset $U_1 \subseteq U$,

$$\inf_{\tau \in \mathscr{T}} \sup_{u \in U_1} \sup_{t \in T} r_W(\tau, t, u) = \sup_{u \in U_1} \inf_{\tau \in \mathscr{T}} \sup_{t \in T} r_W(\tau, t, u).$$

Condition (B) says that if adaptivity holds then the minimax bound for the problem of unknown u is not larger than the worst minimax bound for all subproblems with known fixed u. Trivially the sign $\geqq$ in (B) is valid for any model.

It is an important problem to characterize experiments satisfying the condition (A). There are, however, many examples for which (A) is fulfilled only in an asymptotic manner: Let $(\mathscr{E}_n)$ be a sequence of experiments:

$\mathscr{E}_n = (\Omega_n, \mathfrak{A}_n, \{P^n_{t,u}\colon t \in T, u \in U\})$ and $\mathscr{T}_n$ the sets of randomized estimates with risk $r_{n,W}(\cdot, \cdot, \cdot)$.

(1.2) DEFINITION. The sequence of experiments $\{\mathscr{E}_n\}$ fulfills condition (A_n) if for every finite subset U_1 of U

(A_n)

$$\limsup_n \inf_{\tau_n^* \in \mathscr{T}_n} \sup_{u \in U_1} [\sup_{t \in T} r_{n,W}(\tau_n^*, t, u) - \inf_{\tau_n \in \mathscr{T}_n} \sup_{t \in T} r_{n,W}(\tau_n, t, u)] = 0.$$

Under some mild conditions it is shown in the Appendix that if $\{\mathscr{E}_n\}$ fulfills the condition (A_n) and $\mathscr{E}_n$ converges to a limiting experiment $\mathscr{E}$ (in the sense of Le Cam [4]) then $\mathscr{E}$ fulfills condition (A). Thus a characterization of experiments possessing the property of adaptivity is at the same time a characterization of limiting experiments of sequences having the asymptotic adaptivity property.

Notice however, that there is no converse of the theorem in the Appendix. As a counterexample take

$$P_{t,u,n} = \begin{cases} N\left(\begin{pmatrix} nt \\ nu \end{pmatrix}, \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}\right) & \text{if } 2/n \leq t, u \leqslant 3/n \\ N\left(\begin{pmatrix} t \\ u \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}\right) & \text{if } t, u \text{ not in } [1/n, 4/n] \end{cases}$$

and for the remaining points such that the continuity assumption is fulfilled. Then $\mathscr{E}_n$ converges to a limit $\mathscr{E}$ which has property (A) but $\mathscr{E}_n$ does not have property (A_n).

The most important limiting experiment in statistics is that of a normal shift model. Suppose that the set of nuisance parameters is $\mathbb{R}^l$. The normal shift experiment is the collection of normal distributions in $\mathbb{R}^{k+l}$ with mean $\binom{t}{u}$ and fixed covariance matrix

$$I^{-1} = \begin{pmatrix} I_{11} & I_{12} \\ I_{21} & I_{22} \end{pmatrix}^{-1}$$

$$P_{t,u} = N\left(\begin{pmatrix} t \\ u \end{pmatrix}, \begin{pmatrix} I_{11} & I_{12} \\ I_{21} & I_{22} \end{pmatrix}^{-1}\right).$$

It is a direct consequence of Stein's result [9] that this experiment satisfies condition (A) if $I_{12} = 0$.

One might conjecture that this is due to the fact that if $I_{12} = 0$ then the experiment has a decomposition in a product form

$$P_{t,u} = P_t \otimes P_u,$$

where $P_t = N(t, I_{11})$ and $P_u = N(u, I_{22})$.

Another look at the problem may lead to the conjecture that the important property for (A) is that

$$(\sqrt{dP_{t,u}} - \sqrt{dP_{t,v}}) \text{ is orthogonal to } (\sqrt{dP_{t,u}} - \sqrt{dP_{s,u}}) \text{ in } L^2(\Omega)$$

for all t, u, s, v. Our main theorem shows, however, that both mentioned properties are not essential. The basic requirement for adaptivity for translation invariant experiments is the IPD-Property (see Section 2).

Stein's condition ($I_{12}=0$) was generalized to the case of arbitrary nuisance parameter sets U by Bickel [3], Fabian and Hannan [4], Begun *et al.* [2].

The latter authors give also an explicit expression for the loss in efficiency for experiments which do not have the adaptivity property ("projection formula"). All cited papers treat the local asymptotic normal (LAN) situation. Shick [8] studied adaptivity in the local asymptotic mixed normal (LAMN) case.

2. The Theorem

(2.1) *Assumptions on the Experiment* $\mathscr{E}$

(i) $\mathscr{E}$ is dominated by some σ-finite measure ν. Let $f(t, u, \omega) := (dP_{t,u}/d\nu)(\omega)$.

(ii) For each $u \in U$, $t \mapsto P_{t,u}$ is continuous for the variational distance.

(iii) For every $u \in U$ the experiment, $\mathscr{E}_u = (\Omega, \mathfrak{A}, \{P_{t,u} : t \in \mathbb{R}^k\})$ is measurable, integrable of order 1, and translation invariant.

By integrability of order 1 it is meant that

$$\int \max(1, \|t\|) f(t, u, \omega)\, dt < \infty \ \ \nu\text{-a.e.} \qquad \text{for every} \quad u \in U.$$

Translation invariance is the property that for every fixed u the Hellinger transforms

$$H(t_1, ..., t_m, \alpha_1, ..., \alpha_m, u) = \int [f(t_1, u, \omega)]^{\alpha_1} \cdots [f(t_m, u, \omega)]^{\alpha_m}\, d\nu(\omega),$$

$m \in \mathbb{N}$; $t_1, ..., t_m \in T$; $\sum \alpha_i = 1$; $\alpha_i \geqslant 0$; depend only on the differences $t_2 - t_1, ..., t_m - t_1$. The Hellinger transforms characterize the experiment and their pointwise convergence is equivalent to the weak convergence of the corresponding experiments (see Le Cam [5]). Translation invariance is a typical property of limit experiments.

(2.2) *Assumption on the Set $\mathfrak{W}$ of Loss Functions*

Let $\mathfrak{W}$ be the set of all nonnegative loss functions defined on $\mathbb{R}^k$ satisfying

(i) W is lower semicontinuous and $W(0)=0$,

(ii) W is quasiconvex, i.e., $\{W \leqslant a\}$ is convex for all $a \in \mathbb{R}^+$,

(iii) W is inf-compact, i.e., $\{W \leqslant a\}$ is compact for all $a < \sup W$.

(iv) there are constants c_1, c_2 (possibly depending on W), such that

$$|W(t)| \leqslant c_1 + c_2 \|t\|.$$

Assumptions (ii)–(iv) guarantee that for every Lebesgue-density g on $\mathbb{R}^k$ with $\int \|t\| \, g(t)\, dt < \infty$ the set

$$\arg\min_s \int W(t-s)\, g(t)\, dt$$
$$= \left\{ s : \int W(t-s)\, g(t)\, dt = \inf_r \int W(t-r)\, g(t)\, dt \right\}$$

is nonvoid and compact. It is thus possible to define a Pitman estimate $\tilde{T}_u(\omega)$ for every u by

$$\tilde{T}_u(\omega) = \inf \arg\min_s \int W(t-s)\, f(t \mid u, \omega)\, dt. \tag{1}$$

Here

$$f(t \mid u, \omega) = \frac{f(t, u, \omega)}{\int f(s, u, \omega)\, ds} \tag{2}$$

are posterior densities which exist according to (2.1)(iii) and the infimum for a compact set $S \subseteq \mathbb{R}^k$ is recursively defined by: Let $x_1 = \inf\{ y_1$ such that $\exists y_2, ..., y_k$ with $(y_1, y_2, ..., y_k) \in S\}$ then

$$\inf(S) = (x_1, \inf\{(y_2, ..., y_k) : (x_1, y_2, ..., y_k) \in S\}).$$

By the application of the inf operation, the uniqueness of the the Pitman estimate is ensured. Moreover, the Pitman estimate is always measurable, since the arg min is determined by the values at points with rational coordinates.

If the experiment $\mathscr{E}$ is translation invariant, then the Pitman estimate $\tilde{T}_u$ is also translation invariant in the sense that its distribution $\mathfrak{L}(\tilde{T}_u \mid \cdot)$ satisfies

$$\mathfrak{L}(\tilde{T}_u \mid P_{t,u}) = \mathfrak{L}(\tilde{T}_u - h \mid P_{t+h,u})$$

for all $t, h \in \mathbb{R}^k$, $u \in U$. Clearly a Markov-kernel $\tilde{\tau}_u(\omega, dy)$ is associated with $\tilde{T}_u$ by setting

$$\tilde{\tau}_u(\omega, A) = \begin{cases} 1 & \text{if } \tilde{T}_u(\omega) \in A \\ 0 & \text{otherwise.} \end{cases} \tag{3}$$

Pitman estimates for translation invariant experiments are admissible and enjoy the following minimax property which was proved by Strasser [10].

(2.3) THEOREM. *Let the assumptions* (2.1) *and* (2.2) *be fulfilled. The Pitman estimate* $\tilde{\tau}_u$ *defined by* (1) *has the following properties for fixed* u:

(i) $r_W(\tilde{\tau}_u, t, u) \equiv c(u)$ *(constant)*

(ii) $r_W(\tilde{\tau}_u, t, u) = \inf_\tau \sup_t r_W(\tau, t, u)$

(iii) $c(u) = \int_\Omega \inf_s \int_{\mathbb{R}^k} W(t-s) f(t \mid u, \omega)\, dt\, f(\bar{t}, u, \omega)\, d\nu(\omega)$ *for all* $\bar{t} \in \mathbb{R}^k$.

Proof. The proof is contained in [10, Proposition (1.6)]. The method for showing (ii) and (iii) uses the restricted posterior densities

$$f_\alpha(t \mid u, \omega) = \frac{f(t, u, \omega)}{\int_{D_\alpha} f(s, u, \omega)\, ds} \cdot 1_{D_\alpha}(t),$$

where $D_\alpha := \{t \mid \|t\| \leqslant \alpha\}$. It is shown that for any estimate

$$\begin{aligned}
&\sup_t r_W(\tau, t, u) \\
&\quad\geqslant \limsup_{\alpha \to \infty} \left\{ \int_\Omega \int_{\mathbb{R}^k} W(t-y)\, \tau(\omega, dy)\, f_\alpha(t \mid u, \omega)\, dt \right. \\
&\quad\quad \left. \cdot \frac{1}{\lambda(D_\alpha)} \int_{D_\alpha} f(r, u, \omega)\, dr\, d\nu(\omega) \right\} \\
&\quad\geqslant \limsup_{\alpha \to \infty} \int_\Omega \inf_y \int_{\mathbb{R}^k} W(t-y)\, f_\alpha(t \mid u, \omega)\, dt \\
&\quad\quad \cdot \frac{1}{\lambda(D_\alpha)} \int_{D_\alpha} f(r, u, \omega)\, dr\, d\nu(\omega) \\
&\quad\geqslant \int_\Omega \inf_y \int_{\mathbb{R}^k} W(t-y) f(t \mid u, \omega)\, dt \cdot f(\bar{t}, u, \omega)\, d\nu(\omega)
\end{aligned}$$

for all $\bar{t} \in \mathbb{R}^k$.

The following slightly more general proposition may be proved along exactly the same lines. Its proof is therefore omitted.

(2.4) PROPOSITION. *Let u_1 and u_2 be two arbitrary elements of U. Then*

$$\inf_{\tau} \sup_{t} \int_{\Omega} \int_{\mathbb{R}^k} W(t-y)\, \tau(\omega, dy)\, \tfrac{1}{2}[f(t, u_1, \omega) + f(t, u_2, \omega)]\, d\nu(\omega)$$

$$\geqslant \int_{\Omega} \inf_{y} \int_{\mathbb{R}^k} W(t-y)\, \tfrac{1}{2}[f(t \mid u_1, \omega)\, f(\bar{t}, u_1, \omega) + f(t \mid u_2, \omega)\, f(\bar{t}, u_2, \omega)]\, dt\, d\nu(\omega)$$

for all $\bar{t} \in \mathbb{R}^k$.

We are now ready to state our main theorem. It makes use of a special structure of an experiment, called the IPD-property.

(2.5) DEFINITION. Let $\mathscr{E}$ be an experiment fulfilling assumption (2.1) and $F(t \mid u, \omega)$ be the posterior distributions defined by

$$F(t \mid u, \omega) = \int_{-\infty}^{t} f(s \mid u, \omega)\, ds.$$

$\mathscr{E}$ has *independent posterior distributions* (IPD) if there is a family of posterior distributions $F(t \mid \omega)$ such that

$$F(t \mid u, \omega) = F(t \mid \omega) \qquad \nu\text{-a.e.}$$

for all $u \in U$.

Furthermore we consider a denumerable subset $\mathfrak{W}_0$ of $\mathfrak{W}$, which is large enough to guarantee that an experiment, which has the adaptivity property with respect to this subset, also exhibits adaptivity for the larger class $\mathfrak{W}$:

Let $\tilde{W}$ be a symmetric, strictly convex, twice boundedly differentiable loss function from $\mathfrak{W}$. A possible choice is, e.g., $\tilde{W}(t) = \sqrt{1 + \|t\|^2} - 1$. The set $\mathfrak{W}_0$ is defined as

$$\mathfrak{W}_0 = \{ W(t) \mid W(t) = \tilde{W}(t) + \lambda(t'y - b)\, 1_{\{t'y > b\}},\quad 0 \leqslant \lambda \in \mathbb{Q},\, b \in \mathbb{Q},\, y \in \mathbb{Q}^k \}, \tag{5}$$

where $\mathbb{Q}$ is the set of all rationals.

We are now ready to state our main theorem.

(2.6) THEOREM. *Suppose that the experiment $\mathscr{E}$ satisfies* (2.2). *Then the following statements are all equivalent.*

(i) *$\mathscr{E}$ has the adaptivity property for all $W \in \mathfrak{W}$*

(ii) *$\mathscr{E}$ has the adaptivity property for all $W \in \mathfrak{W}_0$*

(iii) *$\mathscr{E}$ has the IPD-property.*

Proof. Since trivially (i) ⇒ (ii),, we have to show that (ii) ⇒ (iii) and (iii) ⇒ (i).

(ii) ⇒ (iii) Since ν is σ-finite, it is the denumerable sum of finite measures and we may w.l.o.g. assume for the subsequent considerations that ν is finite itself. Let $A_u = \{\omega \mid \int f(t, u, \omega)\, dt > 0\}$. Choose u_1 such that $\nu(A_{u_1}) > \frac{1}{2} \sup_u \nu(A_u)$ and then $\{u_i\}$, $i > 1$, by

$$\nu\left(A_{u_{i+1}} \setminus \bigcup_{k=1}^{i} A_{u_k}\right) > \frac{1}{2} \sup_u \nu\left(A_u \setminus \bigcup_{k=1}^{i} A_{u_k}\right).$$

By this construction $\nu(A_u \setminus \bigcup_i A_{u_i}) = 0$ for all $u \in U$. Let $B_i = A_{u_i} \setminus \bigcup_{j<i} A_{u_j}$ and define the U-valued random variable $V(\omega) = \sum_i u_i 1_{B_i}(\omega)$.

It will be shown that the posterior distributions satisfy

$$F(t \mid u, \omega) \equiv F(t \mid V(\omega), \omega) \qquad \nu\text{-a.e.} \tag{6}$$

for all $u \in U$, which is the IPD property of Definition (2.5). Let v_1 be an element of U and v_2 an arbitrary element of $\{u_i\}$. Define for $W \in \mathfrak{W}_0$

$$S_i(\omega, W) = \arg\min_s \int_{\mathbb{R}^k} W(t-s) f(t \mid v_i, \omega)\, dt$$

and

$$M_i(\omega, W) = \inf_s \int_{\mathbb{R}^k} W(t-s) f(t \mid v_i, \omega)\, dt.$$

Since all loss functions $W \in \mathfrak{W}_0$ are strictly convex, the minimal points $S_i(\omega, W)$ are uniquely defined. Let $B = \{\omega \mid S_1(\omega, W) \neq S_2(\omega, W)\}$ and

$$C = A_{v_1} \cap A_{v_2} = \left\{\omega \;\middle|\; \int f(t, v_i, \omega)\, dt > 0;\, i = 1, 2\right\}.$$

Let further $D_\alpha := \{t \mid \|t\| \leqslant \alpha\}$ with volume $\lambda(D_\alpha)$,

$$h_\alpha(u, \omega) = \frac{1}{\lambda(D_\alpha)} \int_{D_\alpha} f(t, u, \omega)\, dt$$

and

$$C_\alpha = \{\omega \mid h_\alpha(v_i, \omega) > 0;\, i = 1, 2\}.$$

Since $C = \bigcup_{\alpha>0} C_\alpha$ we may show that $\nu(B \cap C) = 0$ by showing that $\nu(B \cap C_\alpha) = 0$ for all α. Suppose the contrary. Then

$$a_\alpha(\omega) := \frac{1}{2} \inf_s \sum_{i=1}^{2} \left[\int W(t-s)\, f(t \mid v_i, \omega)\, dt - M_i(\omega, W)\right] h_\alpha(v_i, \omega) > 0$$

on $B \cap C_\alpha$ and therefore $\int a_\alpha(\omega)\, d\nu(\omega) > 0$. By the adaptivity property (A), Theorem (2.3), and Proposition (2.4) we get

$$\begin{aligned}
0 &= \inf_{\tau^*} \max_{i=1,2} \Bigg[\sup_t \int_\Omega \int_{\mathbb{R}^k} W(y-t)\, \tau^*(\omega, dy) f(t, v_i, \omega)\, d\nu(\omega) \\
&\quad - \inf_\tau \sup_t \int_\Omega \int_{\mathbb{R}^k} W(y-t)\, \tau(\omega, dy) f(t, v_i, \omega)\, d\nu(\omega) \Bigg] \\
&\geqslant \inf_{\tau^*} \sup_t \int_\Omega \int_{\mathbb{R}^k} W(y-t)\, \tau^*(\omega, dy)\, \tfrac{1}{2}[f(t, v_1, \omega) + f(t, v_2, \omega)]\, d\nu(\omega) \\
&\quad - \frac{1}{2} \sum_{i=1}^{2} \left[\int_\Omega \inf_s \int_{\mathbb{R}^k} W(t-s) f(t \mid v_i, \omega)\, dt\, f(\bar{t}, v_i, \omega) \right] d\nu(\omega) \\
&\geqslant \frac{1}{2} \int_\Omega \inf_y \int_{\mathbb{R}^k} W(t-y) \\
&\quad \times \left[\sum_{i=1}^{2} (f(t \mid v_i, \omega) - M_i(\omega, W)) \cdot f(\bar{t}, v_i, \omega) \right] dt\, d\nu(\omega)
\end{aligned}$$

for all $\bar{t} \in \mathbb{R}$. Taking the integral $(1/\lambda(D_\alpha)) \int_{D_\alpha} (\cdot)\, d\bar{t}$ on both sides of the inequality we get

$$0 \geqslant \tfrac{1}{2} \int a_\alpha(\omega)\, d\nu(\omega)$$

for all α. This contradiction establishes $\nu(B \cap C) = 0$ and hence that $S_1(\omega, W) = S_2(\omega, W)$ a.e. on C.

Thus a.e. equality is true for all $W \in \mathfrak{W}_0$. Since $\mathfrak{W}_0$ is denumerable it holds a.e. simultaneously for all W. Hence we may infer from Lemma (4.2) that

$$F(t \mid v_1, \omega) = F(t \mid v_2, \omega)$$

a.e. on $C = A_{v_1} \cap A_{v_2}$. Thus

$$F(t \mid u, \omega) = F(t \mid u_i, \omega) \qquad \text{a.e. on } A_u \cap A_{u_i},$$

which implies that

$$F(t \mid u, \omega) = F(t \mid V(\omega), \omega) \qquad \text{a.e. on } A_u.$$

This proves (6) and hence the first part of the theorem is shown.

(ii) $\Rightarrow$ (iii) Assume that the experiment has independent posterior distributions

$$F(t \mid u, \omega) = F(t \mid \omega).$$

Define the estimate

$$\tilde{T}(\omega) = \inf \arg\min_{s} \int W(t-s)\, F(dt \mid \omega)$$

which is for all $u \in U$ a.s. equal to the Pitman estimate

$$\tilde{T}_u(\omega) = \inf \arg\min_{s} \int_{\mathbb{R}^k} W(t-s) f(t \mid u, \omega)\, dt.$$

Hence, by virtue of Theorem (2.3),

$$\begin{aligned} 0 \leqslant \inf_{\tau^*} \sup_{u} \Bigg\{ & \sup_{t} \int_{\Omega} \int_{\mathbb{R}^k} W(t-y)\, \tau^*(\omega, dy) f(t, u, \omega)\, d\nu(\omega) \\ & - \int_{\Omega} \left[\inf_{s} \int_{\mathbb{R}^k} W(t-s) f(t \mid u, \omega)\, dt \right] f(0, u, \omega)\, d\nu(\omega) \Bigg\} \\ \leqslant \sup_{u} & \left[\sup_{t} \int W(t - \tilde{T}(\omega)) f(t, u, \omega)\, d\nu(\omega) - r_W(\tilde{T}, t, u) \right] = 0, \end{aligned}$$

which implies that $\mathscr{E}$ is adaptive.

(2.7) *Remark.* The equality of the posterior distributions does, in general, not imply the pointwise equality of the posterior densities. If, however, the following condition is satisfied

$$\nu\{\omega \mid t \mapsto f(t, u, \omega) \text{ is discontinuous at } \bar{t}\} = 0$$

for all $\bar{t} \in \mathbb{R}^k$ and all $u \in U$, then the IPD property may be formulated in terms of the posterior densities: The experiment has the IPD property iff the posterior densities have a product representation

$$f(t, u \mid \omega) = g(t, \omega) \cdot h(u, \omega) \qquad \nu\text{-a.e.}$$

This condition can often be easily checked.

3. Examples

(3.1) *LAN Experiments*

The asymptotic experiment belonging to a LAN-family is the gaussian location model described earlier: If (t, u) is the parameter vector in $\mathbb{R}^{k+l}$, the normal shift model may be defined on $\Omega = \mathbb{R}^{k+l}$ as probability space by

$$P_{t,u} = N\left(\begin{pmatrix} t \\ u \end{pmatrix}, \begin{pmatrix} I_{11} & I_{12} \\ I_{21} & I_{22} \end{pmatrix}^{-1} \right).$$

The posterior density at $\omega = (y_1, y_2)$; $y_1 \in \mathbb{R}^k$, $y_2 \in \mathbb{R}^l$ is

$$f(t \mid u, (y_1, y_2)) = |I_{11}|^{1/2} (2\pi)^{-k/2} \exp[-\tfrac{1}{2}(t - y_1 + I_{11}^{-1} I_{12}(y_2 - u))' \times I_{11}(t - y_1 + I_{11}^{-1} I_{12}(y_2 - u))].$$

The IPD property is satisfied iff $I_{12} = 0$. This is in accordance with Stein's result. More generally, if U is a Hilbert space and $P_{t,u}$ is a gaussian translation family on $H = \mathbb{R}^k \times U$ with covariance operator S then IPD is satisfied if the subspaces $\mathbb{R}^k$ and U are orthogonal with respect to the inner product induced by S (see Begun *et al.* [2]). Moreover, by Theorem (2.3) the minimax bounds for this experiment are:

(i) for known, fixed u,

$$\inf_{\tau} \sup_{t} r_W(\tau, t, u) = \int W(\cdot)\, dN(0, I_{11}^{-1});$$

(ii) for unknown u,

$$\inf_{\tau} \sup_{t,u} r_W(\tau, t, u) = \int W(\cdot)\, dN(0, (I_{11} - I_{12} I_{22}^{-1} I_{21})^{-1}).$$

Thus our definition of adaptivity (1.1) coincides with the usual one (Bickel [3], Begun *et al.* [2], etc.), which is based on the limiting risk for LAN-families.

(3.2) *LAMN Experiments*

Many limiting experiments for dependent data are of a mixed normal form (local asymptotic mixed normal families). For a couple of examples see Basawa and Scott [1]. These experiments can be represented on a measurable space $\mathbb{R}^k \times \mathbb{R}^l \times \Omega_0$. Let μ be a probability measure on Ω_0. The density of $P_{t,u}$ with respect to $\lambda_k \otimes \lambda_l \otimes \mu$ (with λ_k being the Lebesgue-measures on $\mathbb{R}^k$) are defined as

$$\frac{dP_{t,u}}{d(\lambda_k \otimes \lambda_l \otimes \mu)} = (2\pi)^{-(k+l)/2} \det(I(\omega)) \cdot \exp\left(-\frac{1}{2}\begin{pmatrix} t \\ u \end{pmatrix} \begin{pmatrix} I_{11}(\omega) & I_{12}(\omega) \\ I_{21}(\omega) & I_{22}(\omega) \end{pmatrix} \begin{pmatrix} t \\ u \end{pmatrix}\right),$$

where

$$\omega \mapsto I(\omega) = \begin{pmatrix} I_{11}(\omega) & I_{12}(\omega) \\ I_{21}(\omega) & I_{22}(\omega) \end{pmatrix}$$

is a measurable mapping from Ω_0 into the set of nonnegative definite matrices on $\mathbb{R}^k \times \mathbb{R}^l$. LAN-families are included in this model by setting $I(\omega) \equiv$ constant. By our main theorem, a mixed normal experiment exhibits the property of adaptivity, iff $I_{12}(\omega) = 0$ μ-a.e. (see Schick [8] for a related result).

(3.3) *Poisson Experiments*

Poisson-type experiments occur as limiting experiments for location parameter models with densities having jumps (see Ibragimov and Has'minskii [6]). They can be described as follows: Let S_{k+l} be the unit sphere in $\mathbb{R}^{k+l}$ and $\bar{\mathbb{R}} = \mathbb{R} \cup \{\infty\}$. Let μ be a probability measure on $S_{k+l} \times \bar{\mathbb{R}}$ and Ω a probability space rich enough to carry a Poisson process with intensity λ and a sequence of i.i.d. random variables $\{s_i(\omega), y_i(\omega)\}$ with distribution μ.

The probability measure on Ω is denoted by P^* and the jump times of the Poisson process by $\{l_i(\omega)\}$. The experiment $\mathscr{E} = (\Omega, \mathfrak{A}, \{P_{t,u} : t \in \mathbb{R}^k, u \in \mathbb{R}^l\})$ is given by

$$\frac{dP_{t,u}}{dP^*}(\omega) = \exp\left[\sum_{\{i:\, (t,u)' \cdot s_i(\omega) > l_i(\omega)\}} y_i(\omega) - \lambda \int_{S_{k+l}} \int_{\mathbb{R}} (t,u)' \cdot s \, d\mu(s,y)\right],$$

see Pflug [7]). This is a translation invariant experiment. It enjoys the IPD property iff the measure μ is concentrated on $((S_k \times \{0\}^l) \cup (\{0\}^k \times S_l)) \times \bar{\mathbb{R}}$, i.e., on orthogonal parts of S_{k+l}. An an application it is easily seen that a location/scale parameter model of uniform densities of the form

$$g(x, t, u) = \frac{1}{2u} 1_{[t-u,\, t+u]}(x) \qquad x, t, u \in \mathbb{R}$$

does not have the asymptotic adaptivity property. In contrast the model

$$g(x, t, u) = \frac{1}{(u-t)} 1_{[t,u]}(x)$$

has this property for t, u being the nuisance.

Appendix

In this appendix, a relation between the conditions (A) and (A_n) is shown and an auxiliary lemma is proved. Let W be a loss function from $\mathfrak{W}$. By $\mathfrak{V}$ we denote the set

$$\mathfrak{V} = \{V \in \mathfrak{W} \mid V \text{ bounded and continuous, } V \leqslant W\}.$$

Since W is lower semicontinuous, $W = \sup_{V \in \mathfrak{V}} V$. By $\mathfrak{K}$ we denote the family of compact sets in $\mathbb{R}^k$.

(A.1) THEOREM. *Let $\mathscr{E}_n = (\Omega_n, \mathfrak{A}_n, \{P^n_{t,n} : t \in \mathbb{R}^k, u \in U\})$ be a sequence of experiments. If*

(i) *$\mathscr{E}_n$ converges to $\mathscr{E} = (\Omega, \mathfrak{A}, \{P_{t,u} : t \in \mathbb{R}^k, u \in U\})$ in the sense of LeCam,*

(ii) *for all $V \in \mathfrak{V}$, all finite sets $U_1 \subseteq U$ and all $\varepsilon > 0$ there is a $K_\varepsilon \in \mathfrak{K}$ such that for all compact $K \supset K_\varepsilon$*

$$\limsup_n \inf_{\tau_n^*} \sup_{u \in U_1} [\sup_{t \in K} r_{n,V}(\tau_n^*, t, u) - \inf_{\tau_n} \sup_{t \in K} r_{n,V}(\tau_{n,V}(\tau_n, t, u)] \leqslant \varepsilon,$$

(iii) *for all $V \in \mathfrak{V}$, all $K \in \mathfrak{K}$, and all $u \in U$,*

$$\limsup_n \inf_{\tau_n} \sup_{t \in K} r_{n,V}(\tau_n, t, u) \leqslant \inf_\tau \sup_t r_W(\tau, t, u)$$

then the condition (A) *is fulfilled for the limiting experiment $\mathscr{E}$.*

Proof. Let $R(u) = \inf_\tau \sup_t r_W(\tau, t, u)$ and

$$g_n(V, K, u) = \inf_{\tau_n} \sup_{t \in K} r_{n,V}(\tau_n, t, u).$$

Define $g(V, K, u)$ in a similar manner. We know that

$$R(u) = \sup_{V \in \mathfrak{V}} \sup_{K \in \mathfrak{K}} g(V, K, u).$$

Suppose that (A) does not hold. Then there exists a $K_1 \in \mathfrak{K}$ and a $V \in \mathfrak{V}$ such that

$$\inf_{\tau^*} \sup_{u \in U_1} [\sup_{t \in K} r_V(\tau^*, t, u) - R(u)] > \delta > 0. \tag{7}$$

Choose now $\varepsilon = \delta/4$. There is a $K \supset K_1$ and a $V \geqslant V_1$ such that

$$g(V, K, u) \geqslant R(u) - \varepsilon.$$

for all $u \in U_1$. The famous minimax theorem for weakly convergent experiments (Le Cam [5] or Strasser [11, Corollary 62.6]) implies that

$$\liminf_n g_n(V, K, u) \geqslant g(V, K, u) \geqslant R(u) - \varepsilon. \tag{8}$$

On the other hand, by (iii)

$$\limsup_n g_n(V, K, u) \leqslant R(u). \tag{9}$$

By (ii) there is a sequence of estimates $\{\tau_n^*\}$ such that

$$\limsup_n \sup_{u \in U_1} [\sup_{t \in K} r_{n,V}(\tau_n^*, t, u) - g_n(V, K, u)] \leqslant 2\varepsilon,$$

which implies that for sufficiently large n and all $u \in U_1$,

$$r_{n,V}(\tau_n^*, t, u) - g_n(V, K, u) < 3\varepsilon, \qquad t \in K. \tag{10}$$

The sequence $\{\tau_n^*\}$ has a cluster point τ^* (see Le Cam [5] or Strasser [11]), which by (8), (9), and (10) satisfies

$$r_V(\tau^*, t, u) - R(u) \leqslant 4\varepsilon$$

for all $u \in U_1$, $t \in K$. Thus

$$\inf_{\tau^*} \sup_{u \in U_1} [\sup_{t \in K} r_W(\tau^*, t, u) - R(u)] \leqslant 4\varepsilon = \delta,$$

which is a contradiction to (7).

(A.2) LEMMA. *Let F be a continuous distribution function on $\mathbb{R}^k$, such that $\int W(t)\, dF(t) < \infty$ for all $W \in \mathfrak{W}_0$. The function*

$$W \mapsto \arg\min_s \int W(t-s)\, dF(t), \qquad W \in \mathfrak{W}_0$$

determines F uniquely.

Proof. Let $F_y(x) = \int_{\{t'y \leqslant x\}} dF(t)$. The "projections" F_y, $y \in \mathbb{Q}^k$ determine F uniquely. This well-known fact is usually called the Cramér–Wold device. Let

$$W_{\lambda, y, b}(t) = \tilde{W}(t) + \lambda(t'y - b)\, 1_{\{t'y > b\}}$$

and

$$s(\lambda, y, b) = \arg\min_s \int W_{\lambda, y, b}(t-s)\, dF(t).$$

$s(\lambda, y, b)$ is uniquely determined, since $W_{\lambda, y, b}$ is strictly convex, continuous in λ, and satisfies $s(0, y, b) \equiv s_0$. Since $s \mapsto \int W_{\lambda, y, b}(t-s)\, dF(t)$ is differentiable, one sees that $s(\lambda, y, b)$ must satisfy

$$\int \nabla \tilde{W}(t - s(\lambda, y, b))\, dF(t) = -\lambda \cdot [1 - F_y(s(\lambda, y, b)' \cdot y + b)]\, y$$

with ∇ denoting the gradient. Let $C = \int \nabla^2 W(t - s_0)\, dF(t)$ with ∇^2 being the Hessian matrix. By the boundedness and continuity of $\nabla^2 \tilde{W}$ and F it follows that the limit $\dot{s}(0, y, b) := \lim_{\lambda \downarrow 0} \lambda^{-1}(s(\lambda, y, b) - s_0)$ exists and

$$C \cdot \dot{s}(0, y, b) = (1 - F_y(s_0' y + b)) \cdot y.$$

By considering the limit $b \to -\infty$ and letting y run through $\mathbb{R}^k$, we see that C is determined by $s(\lambda, y, b)$ for all $\lambda, b \in Q$; $y \in \mathbb{R}^k$. By the same argument $F_y(s_0' y + b)$ is determined for all $b \in Q$ and so is $F(\cdot)$.

REFERENCES

[1] BASAWA, I. V., AND SCOTT, D. J. (1983). *Asymptotic Optimal Inference for Non-Ergodic Models*. Lecture Notes in Statistics Vol. 17, (Springer), New York/Berlin.

[2] BEGUN, J. M., HALL, W. J., HUANG, W.-M., AND WELLNER, J. A. (1983). Information and asymptotic efficiency in parametric-nonparametric models. *Ann. Statist.* **11** No. 2, 432–452.

[3] BICKEL, P. J. (1982). On adaptive estimation. *Ann Statist.* **10** 647–671.

[4] FABIAN, V., AND HANNAN, J. (1982). On estimation and adaptive estimation for locally asymptotically normal families. *Z. Wahrsch. Verw. Gebiete* **59** 459–478.

[5] LE CAM, L. (1972). Limits of experiments. In *Proceedings Sixth Berkeley Symposium, Vol. I*, pp. 245–261. Univ. California Press, Berkeley.

[6] IBRAGIMOV, I. A., AND HAS'MINSKII, R. Z. (1981). *Statistical Estimation. Asymptotic Theory.* Appl. of Mathematics, Vol. 16, Springer-Verlag, New York.

[7] PFLUG, G. (1982). The limiting log-likelihood process for discontinuous multiparameter density families. In *Probability and Statistical Inference*, pp. 287–295, Reidel, Dordrecht.

[8] SCHICK, A. (1985). On estimation in LAMN families when there are nuisance parameters present. Preprint, SUNY Binghamton, submitted.

[9] STEIN, CH. (1956). Efficient nonparametric testing and estimation. In *Proceedings, Third Berkeley Symp., Vol. 1*, pp. 187–195, Univ. of California Press, Berkeley.

[10] STRASSER, H. (1982). Local asymptotic minimax properties of Pitman estimates, *Z. Wahrsch. Verw. Gebiete* **60** 223–247.

[11] STRASSER, H. (1985). *Mathematical Theory of Statistics.* (Statistical experiments and asymptotic decision theory.) de Gruyter, Berlin/New York.