PROBABILITY
AND
MATHEMATICAL STATISTICS

Vol. 10, Fasc. 2 (1989), p. 179-189

STATISTICAL MODELS DEFINED BY SUFFICIENCY

BY

LUDGER RÜSCHENDORF (MÜNSTER)

Abstract. We conctruct statistical models by the assumption that certain statistics or σ-algebras are sufficient for the model. This point of view has been developed especially by Lauritzen and has led in connection with the notion of partial sufficiency to very interesting characterization results of de Finetti type. In this paper we investigate some consequences for the estimation and test theory of models of this type.

1. Introduction. A general formulation of de Finetti's theorem is the following:

If

$$(\Omega, \mathscr{A}) = \bigotimes_{i=1}^{\infty} (X, \mathscr{B})$$

is a product of Polish spaces and $P \in M^1(\Omega, \mathscr{A})$ is an exchangeable probability measure on $(\Omega, \mathscr{A})$, then conditionally on the exchangeable σ-algebra $\mathscr{A}_0 \subset \mathscr{A}$, P is i.i.d. and $\mathscr{A}_0$ is sufficient for the set $\mathscr{P}$ of all exchangeable probability measures.

The extreme points of $\mathscr{P}$ are given by the infinite products $Q^{(\infty)}$, $Q \in M^1(X, \mathscr{B})$ and there holds a corresponding integral representation. Extensions of this result, including a broad field of applications, have been proved under the notion of partial sufficiency; the role of $\mathscr{A}_0$ then is taken by the partial exchangeable σ-field. The importance of the sufficiency aspect has been pointed out in full generality by Dynkin [4]. Lauritzen [6] pointed out the relevance of the sufficiency and partial sufficiency aspect for the construction of statistical models leading especially to a general characterization of exponential families. Diaconis and Freedman [2], [3] showed that many important characterization results can be established in this frame work, so e.g. the characterization of orthogonally invariant distributions on $(\Omega, \mathscr{A}) = (\boldsymbol{R}^{\infty}, \mathscr{B}^{\infty})$ as mixtures of i.i.d. $N(0, \sigma^2)$ sequences.

Let now $(\Omega, \mathscr{A})$ be any measure space, let $\mathscr{A}_0 \subset \mathscr{A}$ be a sub σ-algebra and K be a Markov-kernel from $(\Omega, \mathscr{A}_0)$ to $(\Omega, \mathscr{A})$, $K_\omega := K(\omega, \cdot)$, $\omega \in \Omega$. Define

$$(1) \qquad \mathscr{P}(K) := \{P \in M^1(\Omega, \mathscr{A});\ K = P(\cdot|\mathscr{A}_0)[P]\}$$

the set of all probability measures on $(\Omega, \mathscr{A})$ such that K is a regular conditional distribution of P on $\mathscr{A}_0$. If T: $(\Omega, \mathscr{A}) \to (Y, \mathscr{B})$ is a surjective mapping, $\{t\} \in \mathscr{B}$ for all $t \in Y$, and if K is a Markov kernel from $(Y, \mathscr{B})$ to $(\Omega, \mathscr{A})$, then with $K_t := K(t, \cdot)$ and $\mathscr{A}_0 = \sigma(T) \subset \mathscr{A}$, the σ-algebra generated by T, we use the notation in (1) also for the factorized kernel and define in this case

$$(2) \qquad \mathscr{P}(K) := \{P \in M^1(\Omega, \mathscr{A});\ K_t = P(\cdot|T = t)[P^T]\}.$$

By definition $\mathscr{P}(K)$ is the largest class $\mathscr{P}$ of probability measures on $(\Omega, \mathscr{A})$, such that $\mathscr{A}_0$ resp. T are sufficient for $\mathscr{P}$ and K is a conditional distribution given $\mathscr{A}_0$ resp. T. $\mathscr{P}(K)$ is called the *maximal family generated by* K by Lauritzen [6] considering the discrete case. It seems to be not a severe restriction from the point of view of possible applications to restrict to (regular) conditional distributions.

In section 2 we introduce some subsets of $\mathscr{P}(K)$ and extend some basic relations for these models due to Lauritzen [6] in the discrete case. Furthermore, we determine the total variation distance between two different maximal families and describe the tangent cone of $\mathscr{P}(K)$. In section 3 we derive some basic results for the testing theory of these models and establish in section 4 some construction methods for optimal MVU estimators in models related to $\mathscr{P}(K)$. The main point in these results is to establish the relation to optimality within the more simple conditional models.

2. Models admitting sufficient statistics. If K is a Markov kernel from $(X, \mathscr{A}_0)$ to $(X, \mathscr{A})$, then a natural question is, whether K is the conditional distribution of any P given $\mathscr{A}_0$. The following lemma uses a well known property of conditional expectations for the characterization. There are several related characterizations of conditional expectation operators by positivity and contraction properties in the literature (cf. Neveu [8]).

LEMMA 1. *$\mathscr{P}(K) \neq \emptyset \Leftrightarrow \exists P_0 \in M^1(X, \mathscr{A}_0)$ such that for $f \in \mathscr{B}_b(\mathscr{A})$, $g \in \mathscr{B}_b(\mathscr{A}_0)$ — the bounded $\mathscr{A}$ resp. $\mathscr{A}_0$ measurable functions —*

$$(3) \qquad K(fg) = gKf\,[P_0].$$

Proof. "$\Rightarrow$" If $P \in \mathscr{P}(K)$, then define $P_0 := P/\mathscr{A}_0$. Since

$$Kf(\omega) = \int f(x)\,K_\omega(dx) \in B_b(A_0),$$

(3) is a well known property of conditional expectations.

"$\Leftarrow$" Define $P := KP_0$, i.e. $P(A) = \int K_\omega(A) P_0(d\omega)$, $A \in \mathscr{A}$. By (3), $K^2 f = K(Kf) = (Kf)(K1) = Kf$, i.e. $K^2 = K$ is a projection. Therefore, $E_p fg = \int K(fg)\,dP_0 = \int g(Kf)\,dP_0 = \int g(K^2 f)\,dP_0 = \int K(gKf)\,dP_0 = \int g(Kf)\,dP$, by definition of P. This implies that $Kf = E_p(f|\mathscr{A}_0)$.

We assume in the following that each occurring Markov kernel K satisfies (3) for all $P_0 \in M^1(\Omega, \mathscr{A}_0)$ and call kernels K with this property *full conditional kernels*. For full conditional kernels $\mathscr{A}_0 = \{A \in \mathscr{A};\ 1_A = K1_A\}$; under (3) this equality holds $P = KP_0$ almost surely and we would have to restrict in the following to the subclass of all elements of $\mathscr{P}(K)$ dominated by $P = KP_0$. If K is a Markov kernel from $(Y, \mathscr{B})$ to $(\Omega, \mathscr{A})$, $T:(\Omega, \mathscr{A}) \to (Y, \mathscr{B})$ as in the introduction, then K is called a *full conditional kernel* (*given* T) iff, for all $t \in Y$, $K_t\{T = t\} = 1$.

For $M \subset M^1(\Omega, \mathscr{A}_0)$ define now

$$\mathscr{P}(K, M) := \{P_\mu := K_\mu;\ \mu \in M\}. \tag{4}$$

PROPOSITION 2. *If* $M \subset M^1(\Omega, \mathscr{A}_0)$, *then*:

1. $\mathscr{A}_0$ *is sufficient for* $\mathscr{P}(K, M)$.
2. $\mathscr{P}(K) = \mathscr{P}(K, M^1(\Omega, \mathscr{A}_0))$.
3. $M^1(\Omega, \mathscr{A}_0) \xrightarrow[\mu \to P_\mu = K\mu]{} \mathscr{P}(K)$ *is bijective*; *the inverse mapping is the retraction* $P \to P/\mathscr{A}_0$.
4. $\mathscr{P}(K) = \{P \in M^1(\Omega, \mathscr{A});\ P = KP\}$.

Proof. 1. In the proof of Lemma 1 it was shown that $K = P_\mu(\cdot|\mathscr{A}_0)$ for $\mu \in M$, i.e. the sufficiency of $\mathscr{A}_0$ for $\mathscr{P}(K, M)$.

2, 3, 4. If $P \in \mathscr{P}(K)$, then $K = P(\cdot|\mathscr{A}_0)$ and therefore, with $P_0 := P/\mathscr{A}_0$, $P = KP_0 = KP$, i.e. $P \in \mathscr{P}(K, M^1(\Omega, \mathscr{A}_0))$. This implies: $\mathscr{P}(K) \subset \mathscr{P}(K, M^1(\Omega, \mathscr{A}_0))$ and $\mathscr{P}(K) \subset \{P \in M^1(\Omega, \mathscr{A});\ P = KP\}$. The converse inclusions follows from 1. If $\mu \in M^1(\Omega, \mathscr{A}_0)$ and $A \in \mathscr{A}_0$, then

$$P_\mu(A) = \int K_\omega(A)\,\mu(d\omega) = \int 1_A\,\mu(d\omega) = \mu(A),$$

i.e. $P_\mu \mathscr{A}_0 = \mu$. Therefore, $P_\mu = K_\mu = KP_\mu$ and $\mu \to P_\mu$ is bijective.

Remarks. (a) A class $\mathscr{P}$ of distributions on $(\Omega, \mathscr{A})$ admits $\mathscr{A}_0$ as a sufficient σ-algebra, if and only if $\mathscr{P} = \mathscr{P}(K, M)$ with $M = \mathscr{P}/\mathscr{A}_0$ and $K = P(\cdot|\mathscr{A}_0)$, assuming the existence of a full conditional kernel.

(b) An analogue to Proposition 2 in the situation with a statistic T and a factorized kernel is obvious.

(c) $\mathscr{P}(K)$ is a simplex. Each element in $\mathscr{P}(K)$ has a unique integral representation

$$P(A) = \int K_\omega(A) P_0(d\omega),\ P_0 := P/\mathscr{A}_0, \quad A \in \mathscr{A}, \tag{5}$$

and $\operatorname{ex}(\mathscr{P}(K)) \subset \{K_\omega;\ \omega \in \Omega\}$. $P \in \operatorname{ex}(\mathscr{P}(K)) \Leftrightarrow P\{\omega \in \Omega; K_\omega = P\} = 1$ (cf. Dynkin [4], Th. 3.1).

The following proposition is obvious but important in the statistical analysis.

PROPOSITION 3. *If $M \subset M^1(\Omega, \mathscr{A}_0)$ is (boundedly) complete, then $\mathscr{A}_0$ is (boundedly) complete and sufficient for $\mathscr{P}(K, M)$. Especially, $\mathscr{A}_0$ is a minimal-sufficient σ-algebra in this case.*

The following examples list up some typical situations.

EXAMPLES

1. *Exchangeable distributions.* Let $(\Omega, \mathscr{A}) = (X, \mathscr{B})^k$, γ_k the permutation group operating on Ω, $\mathscr{A}_0 = \{A \in \mathscr{B}^k;\ \pi A = A,\ \forall \pi \in \gamma_k\}$ the σ-algebra of symmetric (exchangeable) sets and $\mathscr{P} := \{P \in M^1(\Omega, \mathscr{A});\ P^\pi = P,\ \forall \pi \in \gamma_k\}$ the set of all symmetric (exchangeable) distributions on $(\Omega, \mathscr{A})$.

For $f \in \mathscr{L}_b(\Omega, \mathscr{A})$ define

$$Kf(x) := \frac{1}{k!} \sum_{\pi \in \gamma_k} f(\pi x), \qquad x \in \Omega.$$

Then $\mathscr{P} = \mathscr{P}(K)$ and, especially, by Proposition 3, $\mathscr{A}_0$ is a complete and sufficient σ-algebra for $\mathscr{P}$. For $x = (x_1, \ldots, x_k) \in \Omega$ holds

$$K_x = \frac{1}{k!} \sum_{\pi \in \gamma_k} \varepsilon_{\pi(x)},$$

any symmetric distribution P has the integral representation

$$P = \int K_x P(dx) \tag{6}$$

and $\{K_x;\ x \in \Omega\}$ are the extreme points of $\mathscr{P}$.

It is intuitively obvious and can be made precise that the l-dimensional marginals of K_x are close to the products

$$\left(\frac{1}{k} \sum_{j=1}^{k} \varepsilon_{x_j}\right)^{(l)},$$

w.r.t. variation distance, e.g. for $l \ll k$, and, therefore, by (6) the l-dimensional marginals of any $P \in \mathscr{P}$ are close to mixtures of i.i.d.-probabilities (cf. Diaconis and Freedman [2]) for $l \ll k$.

If $\mathscr{P}_1 = \{P^{(k)};\ P \in M^1(X, \mathscr{B})\} \subset \mathscr{P}$ are the i.i.d.-products and $\mathscr{P}_2 = \{P_\alpha;\ P_\alpha = \int P^k \alpha(dP),\ \alpha \in M^1(M^1(X, \mathscr{B}))\}$ are the mixtures of products (also called *PDM*– (positive dependent by mixture) *distributions*), then $\mathscr{P}_1 \sim \mathscr{P}_2$ in the sense of domination and again $\mathscr{A}_0$ is complete and sufficient for $\mathscr{P}_1, \mathscr{P}_2$. Many further subfamilies $\mathscr{P}(K, M) \subset \mathscr{P}$ are known for which $\mathscr{A}_0$ still is complete (cf. [7]).

If $(X, \mathscr{B}) = (R^1, \mathscr{B}^1)$, then $\mathscr{A}_0$ is generated by the order statistic

$$T: R^k \to E := \{x \in R^k; x_1 \leqslant \ldots \leqslant x_k\},\ T(x) := x_{()} = (x_{(1)}, \ldots, x_{(k)}).$$

With

$$E = \bigcup_{i=1}^{k} E_i,\ E_i{:} = \{x \in E;\ x_1 = \ldots = x_i < x_{i+1} < \ldots < x_k\}$$

and $\Omega_i{:} = T_k^{-1}(E_i)$ any $P \in \mathscr{P}$ has a decomposition

$$P = \sum_{i=1}^{k} P_i, \quad P_i{:} = P/\Omega_i$$

and

(7) $$P^T = \sum_{i=1}^{k} \frac{k!}{i!} P_i/E_i.$$

Conversely, the maping

(8) $$M^1(E) \to \mathscr{P}, \quad P = \sum P_i \to \sum_{i=1}^{k} S_i(P_i),$$

where $S_i(P_i)$ is the symmetrization on E_i, is bijective. Especially, a submodel $\mathscr{P}_1 = \mathscr{P}(K, M) \subset \mathscr{P}$ is symmetrically complete if and only if $\mathscr{P}_1/E_i$ is complete, $1 \leqslant i \leqslant k$. This remark generalizes a result of Bell and Smith [1] concerning the set of all "continuous" probability measures in $\mathscr{P}$.

2. *Invariant distributions.* Example 1 extends immediately to the class $\mathscr{P}$ of all distributions invariant w.r.t. more general groups G of transformations on $(\Omega, \mathscr{A})$. For amenable groups it can be shown that the σ-algebra $\mathscr{A}_0 = \bigcap_{g \in G} \overline{I(g)}$, $\overline{I(g)}$ the $\mathscr{P}$-completion of all g-invariant sets, is sufficient for $\mathscr{P}$. If G admits a normalized Haar-measure, then K_x is the normalized Haar--measure on Gx and $\mathscr{A}_0$ is equivalent to the σ-algebra of G-invariant sets.

If e.g. $G = O_k$ is the orthogonal group on $\mathbf{R}^k$, then $T{:}\ \mathbf{R}^k \to [0, \infty)$, $Tx = |x|^2 = \sum x_i^2$, the squared length, is maximalinvariant w.r.t. G and K_t is the Lebesgue-measure on the surface of the ball of radius t. For $M \subset M^1([0, \infty), \mathscr{B}[0, \infty))$ T is sufficient and complete for the submodel $\mathscr{P}(K, M)$ if and only if M is complete. If $\mathscr{P}_1 = \{N_{0,\sigma^2}^{(k)};\ \sigma^2 > 0\} \subset \mathscr{P}$, then from a basic result in exponential families T is complete and sufficient for $\mathscr{P}_1$ or, equivalently, $\mathscr{P}_1^T = M$, the scale family generated by a $\cdot \chi_k^2$-distribution, is complete.

3. *Families generated by i.i.d.-models.* (a) If $\mathscr{P}_1 = \{\mathscr{B}(1, \theta)^{(k)};\ \theta \in [0, 1]\}$ is a Bernoulli-experiment of order k on $(\Omega, \mathscr{A}) = (\{0, 1\}^k, \mathscr{P}(\{0, 1\}^k))$, then $T{:}\ \Omega \to Y{:} = \{0, \ldots, k\}$,

$$T(x){:} = \sum_{i=1}^{k} x_i$$

is sufficient and complete for $\mathscr{P}_1$ and, for $t \in Y$,

$$K_t(x) = \frac{1}{\binom{k}{t}} 1_{\{T=t\}}(x).$$

The generated maximal family is the set of all distributions

$$P_\mu(x) := \frac{\mu(T(x))}{\binom{k}{T(x)}}, \quad \mu \in M^1(Y) \tag{9}$$

(cf. Lauritzen [6], p. 20). The l-dimensional distributions of $K_{T(x)}$ are "close" to $\mathscr{B}(1, T(x))^{(l)}$ for $l \ll k$ and, therefore, $\mathscr{P}(K)$ is "close" to the mixtures of $\mathscr{P}_1$ concerning low dimensional marginals. $M_1 := \{\mathscr{B}(k, \theta); \theta \in [0, 1]\}$ characterizes the Bernoulli-experiment $\mathscr{P}_1$ in $\mathscr{P}(K)$. The set M of all possible distributions $p = (p_1, \ldots, p_k)$ of T w.r.t. mixtures of $\mathscr{P}_1$ can be described as

$$p_i = (-1)^{k-i} \Delta^{k-i} c_i, \tag{10}$$

where $(c_0, c_1, \ldots)$ is any completely monotone sequence, Δ the difference operator (cf. Feller [5], p. 224–227).

(b) If $(\Omega, \mathscr{A}) = (\mathbf{R}^k, \mathscr{B}^k)$, $T(x) = (\sum x_i, \sum x_i^2) = (T_1(x), (T_2(x))$, then T is complete and sufficient for

$$\mathscr{P}_1 = \{N^{(k)}_{\mu,\sigma^2}; \mu \in \mathbf{R}^1, \sigma^2 \geqslant 0\}$$

and the conditional distribution K_t is the uniform distribution $\lambda_{k,t}$ on the $(k-2)$-sphere $\{T = t\}$ in $\mathbf{R}^k$. Again the l-dimensional marginals of $K_{T(x)}$ are close to

$$N\left(\frac{1}{k} T_1(x), \frac{1}{k} T_2(x)\right)^{(l)}$$

for $l \ll k$, implying that the l-dimensional marginals of any $P \in \mathscr{P}(K)$ are close to the mixtures

$$\int N\left(\frac{1}{k} T_1(x), \frac{1}{k} T_2(x)\right)^{(l)} P(dx). \tag{11}$$

Let K, R be two full kernels from $(X, \mathscr{A}_0)$ to $(X, \mathscr{A})$ and let d_v denote the half total variation $d_v(P, Q) = \|P - Q\|$.

PROPOSITION 4. (a) $d_v(\mathscr{P}(K), \mathscr{P}(R)) = \inf_{\omega \in \Omega} d_v(K_\omega, R_\omega)$, *if* $\{\omega\} \in \mathscr{A}_0$ *for* $\omega \in \Omega$.

(b) *If* $T: \Omega \to Y$, $K = (K_t)$, $R = (R_t)$ *are full kernels from* $(Y, \mathscr{B})$ *to* $(X, \mathscr{A})$, *then*

$$d_v(\mathscr{P}(K), \mathscr{P}(R)) = \inf_{t \in Y} d_v(K_t, P_t).$$

Proof. We consider only case (b). For the half total variation distance we use the representation

$$d_v(P, Q) = \inf\{W(x \neq y);\ W \in M(P, Q)\},$$

where $M(P, Q)$ denotes the set of all probability measures on $\Omega \times \Omega$ with marginals P, Q. For $P \in \mathscr{P}(K)$, $Q \in \mathscr{P}(R)$ holds

$$W(x \neq y) = \int W(x \neq y) \mid Tx = t,\ Ty = s)\, dW^{T,T}(t, s) \geqslant \inf_{t,s} W(x \neq y \mid Tx = t,\ Ty = s). \tag{12}$$

Since $W(\cdot \mid T \circ \pi_1 = t, T \circ \pi_2 = s) \in M(K_t, R_s)$, it follows, for $t \neq s$, $W(x \neq y \mid Tx = t,\ Ty = s) = 1$ implying that

$$W(x \neq y) \geqslant \inf_t W(x \neq y \mid Tx = t,\ Ty = t) \geqslant \inf_t \|K_t - R_t\|.$$

Therefore,

$$d_v(P, Q) \geqslant \inf_t \|K_t - R_t\| \quad \text{for any pair } P, Q,$$

implying

$$d_v(\mathscr{P}(K), \mathscr{P}(R)) \geqslant \inf_t \|K_t - R_t\|.$$

Since $K_t \in \mathscr{P}(K)$, $R_t \in \mathscr{P}(R)$ for all t, the converse is obvious.

Proposition 4 remains true for the distance between $\mathscr{P}(K, M)$, $\mathscr{P}(R, M)$ under topological assumptions, , if $t \to K_t(A)$, $t \to R_t(A)$ is continuous for all A and if for all $t \in Y$ there exists a sequence $(\mu_n) \subset M$ such that $\mu_n \xrightarrow{\mathscr{D}} \varepsilon_{\{t\}}$.

PROPOSITION 5. *If* $P, Q \in P(K)$ *with* $P_0 := P^T$, $Q_0 := Q^T$, *then* $Q \ll P \Leftrightarrow Q_0 \ll P_0$ *and* $dQ/dP = h_0 \circ T$ *with* $h_0 := dQ_0/dP_0$.

Proof. If $Q_0 = h_0 P_0$, then

$$Q = \int K_t Q_0(dt) = \int K_t h_0(t) P_0(dt) = \int K_{T(x)} h_0 \circ T(x) P(dx) = (h_0 \circ T) P.$$

Conversely, if $Q \ll P$, then $Q_0 = Q^T \ll P^T = P_0$.

Remarks. (a) If $(\Omega, \mathscr{A}) = (X, \mathscr{B})^{(\infty)}$ is an infinite product of Polish spaces, $\mathscr{P}$ is the class of exchangeable distributions and $\mathscr{A}_0$ is the σ-algebra of permutation symmetric sets, then for $Q, P \in \mathscr{P}$, $Q \ll P \Leftrightarrow Q_0 = Q/\mathscr{A}_0 \ll P/\mathscr{A}_0 = P_0$, i.e. iff the mixing measures are continuous.

(b) If $P \in \mathscr{P} = \mathscr{P}(K)$ and $(P_t)_{t \in R^1}$ is a P continuous path in $\mathscr{P}$ with $P_0 = P$ and tangent vector h, i.e. $dP_t/dP = 1 + th + tr_t$, $h \in L^2(P)$, $\int r_t^2\, dP \to 0$ for $t \to 0$, then by Proposition 5, the tangent cone in $P \in \mathscr{P}$ is given by

$$\begin{aligned} T(P, \mathscr{P}) &:= \{h \in L^2(P);\ h \text{ is a tangent vector}\} \\ &= \{h = \psi \circ T \in L^2(P);\ \textstyle\int h\, dP = 0\}. \end{aligned} \tag{13}$$

This tangent cone gives the local description of the model $\mathscr{P}$ in P and is of interest in finite and in asymptotic estimation theory (cf. [9]). If e.g. $\mathscr{P}$ are the exchangeable distributions on R^k, then $T(P, \mathscr{P})$ is equal to the set of all symmetric elements of $L^2(P)$ with $\int h dP = 0$. For the subclass $\mathscr{P}_1 = \{P^{(k)}; P \in M^1(X, \mathscr{B})\}$ the tangent cone

$$T(P^{(k)}, \mathscr{P}_1) = \{\sum_{i=1}^{k} h(x_i);\ h \in L^2(P),\ \int h dP = 0\}$$

is essentially smaller.

3. Construction of optimal tests. Let $K = (K_t)$, $R = (R_t)$ be full kernels, $\mathscr{P}(K)$, $\mathscr{P}(R)$ be the corresponding maximal families and consider testing the hypotheses $\mathscr{P}(K)$, $\mathscr{P}(R)$ at level $\alpha \in [0, 1]$.

PROPOSITION 6. *Let φ_t^* be the NP-test at level α for $\{K_t\}$, $\{R_t\}$, $t \in Y$ and $\varphi^*(X) := \varphi^*_{T(x)}(x)$. Then φ^* is a UMP-test at level α for $\mathscr{P}(K)$, $\mathscr{P}(R)$.*

Proof. If one chooses a constant randomization for the NP-test φ_t^*, then it is clear that φ^* is measurable. Let $\varphi \in \Phi_\alpha(\mathscr{P}(K))$; then $K_t \in \mathscr{P}(K)$ implies that $\int \varphi dK_t \leqslant \alpha$ and, therefore, $\int \varphi dR_t \leqslant \int \varphi_t^* dR_t = \int \varphi^* dR_t$. This implies, for any $Q = RQ_0 \in \mathscr{P}(R)$,

$$\int \varphi\, dQ = \int (\int \varphi\, dR_t)\, Q_0(dt) \leqslant \int (\int \varphi^*\, dR_t)\, Q_0(dt) = \int \varphi^*\, dQ.$$

Remark. If $(Y, \mathscr{B})$ is a topological space, if $t \to \int \varphi dK_t$ is lower semicontinuous for all test functions φ and if $M \subset M^1(Y, \mathscr{B})$ is a subset such that for all $t \in Y$ there exists a sequence $(\mu_n) \subset M$ converging weakly to $\varepsilon_{\{t\}}$, then the conclusion of Proposition 6 remains true for testing the submodels $\mathscr{P}(K, M)$ against $\mathscr{P}(R)$.

Let $\mathscr{K}_0$, $\mathscr{K}_1$ be sets of full Markov kernels from $(Y, \mathscr{B})$ to $(X, \mathscr{A})$ and define

$$\mathscr{P}(\mathscr{K}_i) := \bigcup_{K \in \mathscr{K}_i} \mathscr{P}(K), \quad i = 0, 1. \tag{14}$$

The following generalization of Proposition 6 is obvious.

COROLLARY 1. *Let φ_t^* be a UMP-test at level α for $\mathscr{K}_{0,t} := \{K_t; K \in \mathscr{K}_0\}$, $\mathscr{K}_{1,t} := \{K_t; K \in \mathscr{K}_1\}$, $t \in Y$. If there exists a measurable version φ^* w.r.t. $\mathscr{P}(\mathscr{K}_0)$, $\mathscr{P}(\mathscr{K}_1)$ of $\varphi^*_{T(x)}(x)$, then φ^* is a UMP-test at level α for $\mathscr{P}(\mathscr{K}_0)$, $\mathscr{P}(\mathscr{K}_1)$.*

In other words, Corollary 1 says that conditionally UMP-tests are UMP.

PROPOSITION 7. *If φ_t^* is a maximin test at level α for $\mathscr{K}_{0,t}$, $\mathscr{K}_{1,t}$ and if φ^* is a measurable version $\varphi^*_{T(x)}(x)$, then φ^* is a maximin test for $\mathscr{P}(\mathscr{K}_0)$, $\mathscr{P}(\mathscr{K}_1)$.*

Proof. If $\varphi \in \Phi_\alpha(\mathscr{P}(\mathscr{K}_0))$, then also $\varphi \in \Phi_\alpha(\mathscr{K}_{0,t})$, implying that

$$\inf_{Q \in \mathscr{P}(\mathscr{K}_1)} E_Q \varphi \leqslant \inf_{K_t \in \mathscr{K}_{1,t}} \int \varphi dK_t \leqslant \inf_{K_t \in \mathscr{K}_{1,t}} \int \varphi^* dK_t \quad \text{for all } t \in Y.$$

Therefore,

$$\inf_{Q\in\mathscr{P}(\mathscr{K}_1)} E_Q\varphi = \inf_{K\in\mathscr{K}_1,\mu\in M^1(Y,\mathscr{B})} \int(\int\varphi\, dK_t)\,\mu(dt) = \inf_{K\in\mathscr{K}_1}\inf_{t\in Y}\int\varphi\, dK_t$$

$$\leqslant \inf_{K\in\mathscr{K}_1}\inf_{t\in Y}\int\varphi^*\, dK_t = \inf_{Q\in\mathscr{P}(\mathscr{K}_1)} E_Q\varphi^*.$$

Similarly, for further optimality criteria like Bayes tests and minimax tests, it can be shown that testing theory can be reduced to the construction of optimal tests for the conditional distributions.

4. MVU-Estimation of real functionals. Let, for $M \subset M^1(Y, \mathscr{B})$ and $\mathscr{K}$ a set of full kernels, $D_0(\mathscr{P}(\mathscr{K}, M))$ denote the unbiased estimators of zero w.r.t. $\mathscr{P}(\mathscr{K}, M) := \bigcup_{K\in\mathscr{K}} \mathscr{P}(K, M)$. The following characterization is obvious.

LEMMA 8. (a) $D_0(\mathscr{P}(K)) = \{f\in L^1(\mathscr{P}(K)), \int f\, dK_t = 0, \forall t\in Y\}$.

(b) *If* $M\subset M^1(Y,\mathscr{B})$ *is complete, then* $D_0(\mathscr{P}(K, M)) = \{f\in L^1(\mathscr{P}(K, M));\ \int f dK_t = 0\,[M]\}$.

(c) *If* $f\in L^1(\mathscr{P}(\mathscr{K}))$, *then* $f\in D_0(\mathscr{P}(\mathscr{K}))\Leftrightarrow f\in D_0(K_t)$, $\forall t\in Y$.

(d) *If* $f = \psi\circ T\in D_0(\mathscr{P}(K))$, *then* $f = 0$.

From the completeness and sufficiency of T for $\mathscr{P}(K, M)$, for M complete one obtains

PROPOSITION 9. *If* $M\subset M^1(Y,\mathscr{B})$ *is complete and* $f\in L^1(\mathscr{P}(K, M))$, *then*: f *is UMVU (for its expectation)*$\Leftrightarrow f = \psi\circ T[\mathscr{P}(K, M)]$ *for some* $\psi\in L^2(M)$.

Let now $\mathscr{K}$ be a set of full kernels, $M\subset M^1(Y,\mathscr{B})$ and $\mathscr{P} = \mathscr{P}(\mathscr{K}, M)$. Any element $P\in\mathscr{P}$ can be identified with a pair $(K,\mu)\in\mathscr{K}\times M$. Let $g\colon \mathscr{K}\to\mathbf{R}^1$ be a functional we want to estimate, and let $D_g = D_g(\mathscr{P})$ denote the unbiased estimators of g. If M is complete, then optimal estimators in the conditional models $\mathscr{K}_t$, $t\in Y$, are optimal w.r.t. $\mathscr{P}$.

PROPOSITION 10. *If* $\mathscr{P} = \mathscr{P}(\mathscr{K}, M)$, *if* M *is complete and* $g\colon\mathscr{K}\to\mathbf{R}^1$, *then*

(a) $f\in D_g\Rightarrow f\in D_g(K_t)$ *for* M *a.a.* $t\in Y$.

(b) *If* $f^*\in D_g$ *is UMVU w.r.t.* K_t *for* M *a.a.* $t\in Y$, *then* f^* *is UMVU w.r.t.* $\mathscr{P} = \mathscr{P}(\mathscr{K}, M)$.

Proof. (a) If for all $Q\in\mathscr{P}(K, M)$, $Q = (K, Q_0)$, $E_Q f = \int(\int f dK_t)\, Q_0(dt) = g(K)$, then, by completeness of M, $\int f\, dK_t = g(K)\,[M]$.

(b) follows from (a).

If M is not complete and $\mathscr{P} = \mathscr{P}(\mathscr{K}, M)$, then typically there will be no UMVU-estimators.

PROPOSITION 11. *Let* $g\colon\mathscr{K}\to\mathbf{R}^1$. *Then*:

(a) *If* f *is MVU for* g *in* $Q = (K, Q_0)\in\mathscr{P}$, *then* $\int f\, dK_t = g(K)\,[Q_0]$.

(b) *If* f^* *is MVU for* g *in* K_t *w.r.t.* $\mathscr{K}_t$ *for* Q_0 *a.a.* $t\in Y$, *then* f^* *is MVU for* g *in* $Q = (K, Q_0)$ *w.r.t.* $\mathscr{P}$.

(c) *If* f^* *is UMVU for* g *w.r.t* $\mathscr{K}_t$, $\forall t \in Y$, *then* f^* *is UMVU for* g *w.r.t.* $\mathscr{P}$.

(d) *If, conversely,* f^* *is UMVU for* g *w.r.t.* $\mathscr{P}$, *if* $f_t^*(K, \cdot)$ *is MVU in* K_t *w.r.t.* $\mathscr{K}_t$ *for all* $t \in Y$ *and if there exists a measurable version of* $f^*_{T(x)}(K, x)$, *then* f^* *is UMVU for* g *w.r.t.* $\mathscr{K}_t$.

Proof. (a) If f is MVU for g in $Q = (K, Q_0)$, then let $h := f - E_Q(f|T) + g(K)$.

$h \in D_g$ since, for $P = (R, P_0) \in \mathscr{P}$ holds,

$$E_p h = E_p f - E_p E_Q(f|T) + g(K) = g(R) - \int(\int f dK_t) P_0(dt) + g(K)$$
$$= g(R) - \int f d(KP_0) + g(K) = g(R).$$

If $Q\{E_Q(f|T) \neq g(K)\} > 0$, then $E_Q(h - g(K))^2 < E_Q(f - g(K))^2$ in contradiction to the assumption. Therefore, $\int f dK_t = g(K)[Q_0]$.

(b), (c) are consequences of (a).

(d) By (b) is $f^*_{T(x)}(K, x)$ MVU for g in $Q = (K, Q_0)$ w.r.t. $\mathscr{P}$ and, therefore, by the uniqueness of optimal estimators $f^*(x) = f^*_{T(x)}(K, x)[Q]$.

From Propositions 10 and 11 the construction of MVU estimators of functions $g = g(K)$ can be reduced to the construction of conditional minimum variance unbiased estimators (CMVU) in the models $\mathscr{K}_t$, $t \in Y$. We finally establish an independence property of UMVU estimators. If M is complete, then T is complete and sufficient for $\mathscr{P}(K, M)$, $\forall K \in \mathscr{K}$. If S is distribution free for $\mathscr{P}(K, M)$, $\forall K \in \mathscr{K}$, then, by Basu's theorem, S, T are stochastically independent w.r.t. $\mathscr{P}$. Without the assumption of completeness of M there is the following independence property of statistics which are ancillary on $\mathscr{P}(K, M)$.

Proposition 12. *Let* $\mathscr{P} = \mathscr{P}(\mathscr{K}, M)$, *let a statistics* S *be distribution free on* $\mathscr{P}(K, M)$, $\forall K \in \mathscr{K}$ *and sufficient for* $\mathscr{P}(\mathscr{K}, Q_0)$, $\forall Q_0 \in M$. *If* $d^* = \psi \circ S$ *is a bounded UMVU for* $g = g(K)$ *for some function* ψ, *then* d^* *and* T *are stochastically independent w.r.t.* $\mathscr{P}$.

Proof. Let $\mathscr{A}_0$ denote the σ-algebra generated by the bounded UMVU-estimators w.r.t. $\mathscr{P}$. A well-known theorem due to Bahadur implies that any element of $L^2(\mathscr{A}_0, \mathscr{P})$ is a UMVU-estimator w.r.t. $\mathscr{P}$. Therefore, for any function h such that $h \circ d^* \in L^2(\mathscr{P})$ holds that $h \circ d^*$ is a UMVU w.r.t. $\mathscr{P}$ and $h \circ d^*$ is again distribution free on $\mathscr{P}(K, M)$, $\forall K \in \mathscr{K}$, and, therefore, estimates a function $g = g(K)$. By Proposition 11, (a),

$$E_Q(h \circ d^* | T = t) = \int h \circ d^* \, dK_t = g(K)[M] \text{ for all } Q \in \mathscr{P}(K, M),$$

implying that d^*, T are stochastically independent w.r.t. $\mathscr{P}(K, M)$, $\forall K \in \mathscr{K}$.

By a well known argument in connection with the covariance method the boundedness of $d^* = \psi \circ S$ can be replaced by the assumption that all moments of d^* exist and determine the distribution of d^*.

If, conversely, d^* and T are stochastically independent w.r.t. $\mathscr{P}$, then $E_Q d^* = E_Q(d^*|T = t) = \int d^* dK_t = g(Q)$ i.e. $g(Q) = g(K)$, so d^* estimates a function of K.

Example. Let $\mathscr{P}^n := \{P^n;\ P \in M^1(R^1, \mathscr{B}^1),\ P \text{ symmetric around zero}\}$.

$\mathscr{P}^n$ is invariant w.r.t. two groups, the permutation group γ_n and the sign group with corresponding maximal invariants $T_1(x) = |x|$, $x \in \boldsymbol{R}^1$, and $T_2(x) = x_{()}$, $x \in \boldsymbol{R}^n$. A minimal sufficient statistic for $\mathscr{P}^n$ is $T(x) := |x|_{()}$, $x \in \boldsymbol{R}^n$, the order statistic of $|x_1|, \ldots, |x_n|$ and $K_t = P^n(\cdot \mid T = t)$ is the uniform distribution on $\{T = t\}$. With $M := (\mathscr{P}^n)^T$, $\mathscr{P} = \mathscr{P}(K, M)$ and M is complete since $\mathscr{P}^{T_1}$ is complete and, therefore, the order statistic T_2 is complete for $(\mathscr{P}^{T_1})^n$. Equivalently, $T = S\circ(T_1, \ldots, T_1)$ is complete for $\mathscr{P}^n$. This implies that any function $h = \psi \circ T \in L^2(\mathscr{P}^n)$ is a UMVU. Especially, the invariant U-statistics

$$(15) \qquad U(x) = \frac{1}{2^n \binom{n}{k}} \sum_{\varepsilon \in \{-1,1\}^n} \sum_{\{i_1, \ldots, i_k\} \subset \{1, \ldots, n\}} f(\varepsilon_{i_1} x_{i_1}, \ldots, \varepsilon_{i_k} x_{i_k})$$

are UMVU estimators w.r.t. $\mathscr{P}^n$.

Let for $\theta \in [0, 1]$ and $t = (t_1, \ldots, t_n)$, $0 \leqslant t_1 \leqslant \ldots \leqslant t_n$

$$K_{\theta,t} = \frac{1}{n!\,2^n} \sum_{\pi \in \gamma_n} \bigotimes_{i=1}^{n} (\theta \varepsilon_{t_{\pi(i)}} + (1-\theta)\varepsilon_{-t_{\pi(i)}}).$$

By Proposition 6 the conditional test φ_t is a UMP test for $\mathscr{P}(K_{1/2}) = \mathscr{P}(K)$ against $\mathscr{P}(K_\theta)$, $\theta \geqslant \frac{1}{2}$.Clearly, the conditional test is the sign-test and is independent of θ. This implies that the sign test is UMP for (the generated models)

$$\mathscr{P}_0 = \bigcup_{\theta \leqslant 1/2} \mathscr{P}(K_\theta) \text{ against } \mathscr{P}_1 = \bigcup_{\theta > 1/2} \mathscr{P}(K_\theta).$$

REFERENCES

[1] C. B. Bell and P. J. Smith, *Some nonparametric tests for the multivariate goodness of fit, multisample, independence and symmetry problems.* In: *Multivariate Analysis,* ed. by P. R. Kirshnaiah, Academic Press, 1969.

[2] D. Diaconis and D. Freedman, *Finite exachangeable sequences,* Ann. Prob. 8 (1980), p. 745–764.

[3] — *Partial exchangeability and sufficiency.* In: Proceedings of the Indian Statistical Institute, Congress on "Statistics: Applications and New Directions", 1981, p. 205–236.

[4] E. B. Dynkin, *Sufficient statistics and extreme points,* Ann. Prob. 6 (1978), p. 705–730.

[5] W. Feller, *An Introduction to Probability Theory and its Applications,* Second ed., Wiley, 1971.

[6] S. L. Lauritzen, *Statistical Models as Extremal Families,* Aalborg University Press, 1982.

[7] A. Mandelbaum and L. Rüschendorf, *Complete and symmetrically complete families of distributions,* Ann. Statist. 15 (1987), p. 1229–1244.

[8] J. Neveu, *Martingales à temps discret,* Masson, 1972.

[9] L. Rüschendorf, *Unbiased estimation of von Mises functionals,* Statistics and Probab. Letters 5 (1987), p. 287–292.

University of Münster

Received on 7. 8. 1987