The Annals of Statistics
1996, Vol. 24, No. 3, 1265–1282

COMPLETE ORDER STATISTICS IN PARAMETRIC MODELS

BY L. MATTNER

Universität Hamburg

For a given statistical model $\mathscr{P}$ it may happen that the order statistic is complete for each IID model based on $\mathscr{P}$. After reviewing known relevant results for large nonparametric models and pointing out generalizations to small nonparametric models, we essentially prove that this happens generically even in smooth parametric models.

As a consequence it may be argued that any statistic depending symmetrically on the observations can be regarded as an optimal unbiased estimator of its expectation.

In particular, the sample mean $\overline{X}_n$ is generically an optimal unbiased estimator, but, as it turns out, also generically asymptotically inefficient.

1. Introduction and results.

1.1. *Completeness in statistical theory*. The existence of a complete and sufficient statistic in a given statistical model may greatly simplify the application of various statistical theories, such as unbiased estimation, median unbiased estimation, unbiased tests, and conditional inference [see, e.g., Lehmann (1983, 1986), Pfanzagl (1979, 1994), Rüschendorf (1987) and Lehmann and Scholz (1992)]. Hence it is unfortunate that so far only few techniques have proved successful for deducing completeness.

1.2. *Contents of this paper*. In this paper we consider IID models only, with the emphasis on sample sizes $n \geq 2$, thus excluding from the discussion completeness results for, for example, linear models [see Anderson (1962)], sampling from finite populations [see Lehmann (1983), page 209] or full shift models [Isenbeck and Rüschendorf (1992) and Mattner (1992, 1993)]. In an IID model it may happen that the order statistic is complete. Subsection 1.4 recalls known relevant results for large nonparametric models (Theorems 1 and 2) and points out generalizations to smaller nonparametric models (Theorems 3, 4 and 5). In subsection 1.5 we state as Theorems 6 and 7 the main result of this paper: the order statistic is generically complete even in smooth parametric models. Section 2 discusses two somewhat interrelated applications, the first more practical, the second purely theoretical. Section 3 contains remarks on generalizations and the history of Theorems 1 to 7.

Received March 1994; revised March 1995.

AMS 1991 *subject classifications*. 62B05, 62F10.

Key words and phrases. Asymptotic efficiency, contamination model, IID model, minimal sufficiency, nonparametric neighborhoods, optimal unbiased estimation, symmetrical completeness, UMVU, unimodality.

Finally, Section 4 presents proofs of the new results of this section, Theorems 3 to 7 and a corollary.

1.3. *IID models.* Let $n \in \mathbf{N}$ and let $(\mathscr{X}^n, \mathscr{A}^n, \mathscr{P}^n)$ be the IID model based on a given model $(\mathscr{X}, \mathscr{A}, \mathscr{P})$. Here $\mathscr{X}$ is the sample space for one observation, $\mathscr{A}$ is a σ-algebra on $\mathscr{X}$ and $\mathscr{P}$ is a nonempty set of probability measures on $(\mathscr{X}, \mathscr{A})$. $(\mathscr{X}^n, \mathscr{A}^n) := (\times_{\nu=1}^n \mathscr{X}, \otimes_{\nu=1}^n \mathscr{A})$ is the n-fold product of the measurable space $(\mathscr{X}, \mathscr{A})$ and

$$\mathscr{P}^n = \{P^n : P \in \mathscr{P}\}$$

is the set of all product measures

$$P^n := \bigotimes^n P := \bigotimes_{\nu=1}^n P,$$

with identical marginals belonging to $\mathscr{P}$.

1.4. *Complete order statistics in nonparametric models.* It may happen that the order statistic (i.e., the observations ignoring their order) is complete and sufficient. More precisely, let $\mathscr{A}^n_{\mathrm{sym}}$ denote the sub-σ-algebra of $\mathscr{A}^n$ consisting of all permutation-invariant events. If $\mathscr{X} = \mathbf{R}$, then $\mathscr{A}^n_{\mathrm{sym}}$ is just the σ-algebra generated by the usual order statistic. In any case, a real-valued function on $\mathscr{X}^n$ is measurable with respect to $\mathscr{A}^n_{\mathrm{sym}}$ iff it is measurable with respect to $\mathscr{A}^n$ and permutation invariant. $\mathscr{A}^n_{\mathrm{sym}}$ is always sufficient for $\mathscr{P}^n$, regardless of $\mathscr{P}$. $\mathscr{A}^n_{\mathrm{sym}}$ is called complete for $\mathscr{P}^n$, according to the general definition of a complete σ-algebra, if for every real-valued $\mathscr{A}^n_{\mathrm{sym}}$-measurable function h the relation

$$\int h \, dP^n = 0 \qquad (P \in \mathscr{P}), \tag{1}$$

implies

$$h = 0 \quad [\mathscr{P}^n]. \tag{2}$$

[The statement in (2) means: for every $P^n \in \mathscr{P}^n$, we have $h = 0$ P^n-almost surely.] In this case it is natural to say that the order statistic is complete for $\mathscr{P}^n$ (even without giving any general definition of "order statistic"). Alternatively, we may say that $\mathscr{P}^n$ is symmetrically complete, and this is the terminology used in this paper. (In the literature we also find the somewhat misleading terminology of calling $\mathscr{P}$, rather than $\mathscr{P}^n$, symmetrically complete.) The two basic symmetric completeness results are the following.

THEOREM 1 (Convex models). *Let $\mathscr{P}$ be a complete and convex model on $(\mathscr{X}, \mathscr{A})$. Then $\mathscr{P}^n$ is symmetrically complete for every sample size n.*

THEOREM 2 (Uniform densities). *Let μ be a nonnegative measure without atoms on $(\mathscr{X}, \mathscr{A})$ and let $\mathscr{R}$ be a ring generating $\mathscr{A}$. Then each IID model based on*

$$\mathscr{P} = \left\{ \frac{1_B}{\mu(B)} \cdot \mu : B \in \mathscr{R}, 0 < \mu(B) < \infty \right\}$$

is symmetrically complete.

The history of Theorem 1 is sketched in Remark 1 in Section 3. Theorem 2 is due to Fraser (1954).

Although Theorems 1 and 2 are somewhat incomparable, it seems that Theorem 1 yields symmetric completeness results for more interesting models. As an example we mention the following easy, but apparently so far unnoticed, consequence of Theorem 1, which does not seem to be easily deducible from Theorem 2 (see also Remark 2 below).

THEOREM 3 (Unimodal densities). *Each IID model based on the set of all probability measures on the real line with unimodal Lebesgue densities is symmetrically complete.*

Symmetric completeness results are usually thought of as referring to "large" nonparametric models, containing many "unrealistic" probability measures. Hence statistical optimality results, for example, in unbiased estimation, based on symmetric completeness may appear to rest on doubtful assumptions. This picture changes somewhat if we observe the following two variations of Theorem 1. Similar results have been obtained before in Rüschendorf (1988), page 298.

THEOREM 4 (Contamination models). *Let* $\mathscr{P}_0$ *be an arbitrary model on* $(\mathscr{X}, \mathscr{A})$. *For each* $P_0 \in \mathscr{P}_0$ *let* $\mathscr{Q}(P_0)$ *be a complete and convex model on* $(\mathscr{X}, \mathscr{A})$ *satisfying* $P_0 \ll \mathscr{Q}(P_0)$. *Then, for every* $\varepsilon \in (0, 1]$, *each IID model based on*

$$\mathscr{P}_\varepsilon = \{(1-t) \cdot P_0 + t \cdot Q \colon P_0 \in \mathscr{P}_0, \, Q \in \mathscr{Q}(P_0), \, t \in [0, \varepsilon]\}$$

is symmetrically complete.

THEOREM 5 (Nonparametric neighborhoods). *Let* $\mathscr{P}_0$ *be an arbitrary model on* $(\mathscr{X}, \mathscr{A})$. *Then, for every* $\varepsilon > 0$, *each IID model based on*

$$\mathscr{P}_\varepsilon = \left\{ P \colon P \ll P_0 \text{ and } \sup \left| \frac{dP}{dP_0} - 1 \right| \le \varepsilon [P_0] \text{ for some } P_0 \in \mathscr{P}_0 \right\}$$

is symmetrically complete.

1.5. *Complete order statistics in parametric models.* So far the existence of a complete sufficient statistic or σ-algebra for parametric IID models seems to have been established only either for exponential families or in certain so-called irregular cases. Disregarding the latter, this appears to have led to the belief that the applicability of the theory of optimal unbiased estimation in the parametric case is essentially restricted to exponential families (see, e.g., statements in Pfanzagl [(1980), page 3], Lehmann [(1983), pages 115 and 163] and Witting [(1985), page 304]). The main and perhaps somewhat surprising results of this paper, stated below, imply that this belief is erroneous: generically, the order statistic is complete. (In particular: existence of a complete and sufficient statistic is the rule, not the exception.) As a consequence, symmetric functions are generically optimal unbiased estimators of their expectation. (For simplicity, we have just stated slightly more than what is rigorously stated and proved below: under reasonable conditions, we prove, e.g., only generic quadratic completeness instead of generic completeness.)

We need some notation in order to state our results precisely.

Let $\mathscr{L}^{\infty}(\mathscr{A})$ denote the Banach space of all bounded, real-valued $\mathscr{A}$-measurable functions on $\mathscr{X}$, the norm being the supremum norm simply denoted by $\|\cdot\|$.

Let $\mathscr{G}$ denote the set of all infinitely often differentiable functions $G = (g_\vartheta\colon \vartheta \in \Theta)$ defined on an interval Θ and taking values g_ϑ in $\mathscr{L}^{\infty}(\mathscr{A})$, which satisfy

$$\|G^{(k)}\| := \sup_{\vartheta \in \Theta} \|g_\vartheta^{(k)}\| < \infty \qquad (k \in \mathbf{N}_0),$$

where $g_\vartheta^{(k)}$ denotes the kth derivative with respect to ϑ, and

$$\|G\| = \sup_{\vartheta \in \Theta} \|g_\vartheta\| < \tfrac{1}{2}. \tag{3}$$

$\mathscr{G}$ is a metric space with metric d defined by

$$d(G, \tilde{G}) = \sum_{l=0}^{\infty} 2^{-l-1} \min\bigl(1, \|G^{(l)} - \tilde{G}^{(l)}\|\bigr). \tag{4}$$

For any parametrized model $\mathscr{P} = (P_\vartheta\colon \vartheta \in \Theta)$ and any $G \in \mathscr{G}$, put

$$\mathscr{P}_G := (P_{G,\vartheta}\colon \vartheta \in \Theta) := \left(\frac{(1 + g_\vartheta) P_\vartheta}{\int_{\mathscr{X}} (1 + g_\vartheta)\, dP_\vartheta} \colon \vartheta \in \Theta \right).$$

We are interested in the symmetric completeness properties of IID models based on $\mathscr{P}_G$. First we consider bounded completeness.

If the implication "(1) $\Rightarrow$ (2)" holds for every bounded $\mathscr{A}^n_{\mathrm{sym}}$-measurable function h, then we call the model $\mathscr{P}^n$ occurring in (2) symmetrically boundedly complete.

THEOREM 6 (Generic bounded completeness). *Let $\mathscr{P} = (P_\vartheta\colon \vartheta \in \Theta)$ be a parametrized statistical model over the sample space $(\mathscr{X}, \mathscr{A})$. Suppose*:

(i) *$\mathscr{A}$ is countably generated.*
(ii) *Θ is a nondegenerate interval on the real line.*
(iii) *$\mathscr{P}$ is infinitely often differentiable with respect to total variation distance.*

Then

(5) $\{G \in \mathscr{G} : \mathscr{P}^n_G$ *symmetrically boundedly complete for every sample size* $n\}$

contains a dense G_δ-subset of $\mathscr{G}$.

If $p \in [1, \infty]$ and if the implication "(1) $\Rightarrow$ (2)" holds for every $\mathscr{A}^n_{\mathrm{sym}}$-measurable function h such that $|h|^p$ is integrable with respect to every $P^n \in \mathscr{P}^n$, then we call the model $\mathscr{P}^n$ occurring in (2) symmetrically p-complete for sample size n.

THEOREM 7 (Generic p-completeness). *Let $p \in [1, \infty)$ and let $q = p/(p - 1)$ denote the exponent conjugate to p. Let $\mathscr{P}$ be as in Theorem* 6 *but with* (iii) *replaced by the following more restrictive assumption*:

(iii$_p$) *$\mathscr{P}$ is homogeneous and for every $\vartheta_0 \in \Theta$ there exists an $\alpha > 0$ such that*

$$\Theta \cap [\vartheta_0 - \alpha, \vartheta_0 + \alpha] \ni \vartheta \mapsto \frac{dP_\vartheta}{dP\vartheta} \in L^q(P_{\vartheta_0})$$

is infinitely often differentiable with respect to the $L^q(P_{\vartheta_0})$-norm.

Then

$$\{G \in \mathscr{G} : \mathscr{P}_G^n \text{ symmetrically } p\text{-complete for every sample size } n\}$$

contains a dense G_δ-subset of $\mathscr{G}$.

Finally, if $\mathscr{P}^n$ is symmetrically p-complete for every $p > 1$, then we call $\mathscr{P}^n$ symmetrically $(1+)$-complete.

COROLLARY. *If $\mathscr{P}$ satisfies the assumptions of Theorem* 7 *for every $p > 1$, then*

$$\{G \in \mathscr{G} : \mathscr{P}_G^n \text{ symmetrically } (1+)\text{-complete for every sample size } n\}$$

contains a dense G_δ-subset of $\mathscr{G}$.

The reason for stating the corollary is the following: for $p > 1$, assumption (iii$_p$) of Theorem 7 is true for every exponential family with open parameter space, whereas for $p = 1$, there are many counterexamples.

Note that, for example, Theorem 7 can roughly be stated as follows: If $\mathscr{P}$ is a sufficiently smooth parametric model (with "sufficiently" depending on $p \in [1, \infty)$), then not only the neighborhood model $\mathscr{P}_\varepsilon$ from Theorem 5 gives rise to symmetrically p-complete IID models, even most smooth parametric submodels $\mathscr{P}_G$ of $\mathscr{P}_\varepsilon$ do. See subsection 2.1 below for a more concrete discussion.

Theorems 6, 7 and the corollary are existence results: no models with the stated completeness properties are actually constructed in their proofs [unless one insists that the standard proof of Baire's theorem is constructive; see the discussion in Oxtoby (1980), pages 2, 7 and 46]. In fact, no such models seem to be known, at least not in the statistical literature. So the main implication of our results seems to be the following: if we take any smooth parametric model which is not too special (e.g., not an exponential family, not a group family, ...), then we should not be surprised if it has the completeness properties under discussion. However, to prove that it actually has will typically be difficult, as is witnessed by the sparsity of known parametric completeness results in general. This situation might be compared with the state of knowledge about, say, transcendental numbers: there are many, but few are recognized as such.

2. Applications.

2.1. *Estimation of a Poissonian variance*. Consider the problem of estimating the variance of an unknown probability distribution P which is believed to be well described by a member of

$$\mathscr{P}_0 = \{P_\lambda \colon \lambda \in (\lambda_1, \lambda_2)\},$$

where P_λ denotes the Poisson distribution with expectation = variance = λ and $0 < \lambda_1 < \lambda_2 < \infty$. Let $X_1, \dots, X_n$ denote independent observations according to P. What does the theory of unbiased estimation say for our task? Consider three cases.

(a) If $\mathscr{P}_0$ is assumed as an exactly true model, then $\overline{X}_n = n^{-1}\sum_{i=1}^n X_i$ is optimal unbiased for $\mathrm{Var}(X_1)$.

(b) If not $\mathscr{P}_0$ but rather the associated model $\mathscr{P}_\varepsilon$ as in Theorem 5 with some $\varepsilon > 0$ is taken as a true model, then $S_n^2 = (n-1)^{-1}\sum_{i=1}^n (X_i - \overline{X}_n)^2$ is optimal unbiased for $\mathrm{Var}(X_1)$.

(c) If we like to think in exact but not precisely known parametric models, then the corollary to Theorem 7 is relevant for us. It says that most models $\mathscr{P}_G$ close to $\mathscr{P}_0$ are $(1+)$-symmetrically complete. If we use just symmetrical 2-completeness, then S_n^2 is in particular UMVU for $\mathrm{Var}(X_1)$ under most models $\mathscr{P}_G$. We also note that, by a theorem of Plachky (1993), $\overline{X}_n$ is not UMVU for $\mathrm{Var}(X_1)$ under any model $\mathscr{P}_G$ containing at least one non-Poissonian distribution.

We observe here a remarkable discontinuity property of optimal unbiased estimators with respect to the underlying model which has apparently been neglected in the literature: for example, in the notation of case (b), we have $\mathscr{P}_\varepsilon \downarrow \mathscr{P}_0$ as $\varepsilon \downarrow 0$, but the "limit" S_n^2 of the optimal unbiased estimators in the model $\mathscr{P}_\varepsilon$ differs markedly from the optimal unbiased estimator $\overline{X}_n$ in the model $\mathscr{P}_0$. The former has, for instance, a variance $\lambda/n + 2\lambda^2/(n-1)$, more than $(1 + 2\lambda)$ times as large as the variance λ/n of the latter. This is quite different from the classical case of the discontinuity property, n IID observations according to $\mathscr{P}_\varepsilon = \{N(a, \sigma^2)\colon \ \sigma^2 > 0,\ |a| \le \varepsilon\}$, where the difference between the respective optimal unbiased estimators S_n^2 and $(1/n)\sum_1^n X_i^2$, measured in terms of their variances $2\sigma^2/(n-1)$ and $2\sigma^2/n$ under $\mathscr{P}_0$, becomes small for n large.

Clearly, the optimality results in cases (b) and (c) do not imply unconditionally that S_n^2 should be considered as a good estimator or superior to $\overline{X}_n$. For example, we may assess the quality of both estimators according to their mean squared errors $\mathrm{MSE}(\overline{X}_n)$ and $\mathrm{MSE}(S_n^2)$ in the model $\mathscr{P}_\varepsilon$ of case (b). Then we may check that, for fixed n and sufficiently small ε, these quantities are close to corresponding quantities $\mathrm{MSE}_\lambda(\overline{X}_n)$ and $\mathrm{MSE}_\lambda(S_n^2)$ in the model

$\mathscr{P}_0$. Hence MSE($\overline{X}_n$) may be appreciably smaller than MSE(S_n^2). However, it is not obvious how to get reasonable quantitative criteria for choosing between the two estimators.

2.2. *Asymptotically inefficient sequences of UMVU estimators.* Assume that $\mathscr{P}$ is a model on the real line satisfying the assumption of the corollary to Theorem 7 and, in addition,

$$E_{P_\vartheta}[X^2] \le M \qquad (\vartheta \in \Theta), \tag{6}$$

$$\left\| \frac{d}{d\vartheta} P_\vartheta \right\|_V \le M \qquad (\vartheta \in \Theta), \tag{7}$$

for some finite M, where $\|\cdot\|_V$ denotes the total variation norm and X is the identity on $\mathbf{R}$, and, for the sake of simplicity,

$$E_{P_\vartheta}[X] = \vartheta \qquad (\vartheta \in \Theta), \tag{8}$$

as well as the assumptions used to construct asymptotically efficient sequences of approximate maximum likelihood estimators for ϑ given in Lehmann (1983), pages 422, 415 and 406 [note that (6) and (8) imply the existence of an $\sqrt{n}$-consistent estimator sequence for ϑ]. Assume further that $\mathscr{P}$ is either dominated by Lebesgue measure or concentrated on the integers but not on only one or two points.

For example, $\mathscr{P}$ could be any of the usual one-parameter exponential families [with appropriate parameter space and suitably parametrized such that (8) holds], except for subfamilies of the binomial family {Bin(1, p): $0 < p < 1$}. A specific possible choice for $\mathscr{P}$ is the Poisson family of 2.1.

Let $\mathscr{P}_G$ with $G \in \mathscr{G}$ be a model with $\mathscr{P}^n$ symmetrically $(1 +)$-complete for every sample size n. We claim that:

(i) $\mathscr{P}_G$ is not an exponential family in the identity.

(ii) If G is taken sufficiently close to 0 with respect to the metric in $\mathscr{G}$, then $\mathscr{P}_G$ satisfies all the assumptions made above for $\mathscr{P}$ except that (8) is replaced by

$$\gamma(\vartheta) := E_{P_{G,\vartheta}}[X] \qquad (\vartheta \in \Theta), \tag{9}$$

has an everywhere strictly positive derivative.

It will then follow from (i) that X does not everywhere achieve the Cramér–Rao bound for estimating $\gamma(\vartheta)$ with one observation [see, e.g., Müller-Funk, Pukelsheim and Witting (1989)]. Hence, by linearity, the estimator sequence ($\overline{X}_n$: $n \in \mathbf{N}$) is not everywhere asymptotically efficient in the sense of asymptotically achieving the Cramér–Rao bound for estimating $\gamma(\vartheta)$. Since asymptotically efficient estimator sequences for $\gamma(\vartheta)$ exist [reparametrize $P_{G,\vartheta}$ through $\gamma(\vartheta)$ and use ($\overline{X}_n$: $n \in \mathbf{N}$) as an $\sqrt{n}$-consistent estimator sequence], we see that ($\overline{X}_n$: $n \in \mathbf{N}$) is asymptotically first-order inadmissible for estimating $\gamma(\vartheta)$ although, by symmetric $(1 +)$-completeness of $\mathscr{P}_G^n$, $\overline{X}_n$ is optimal unbiased within the class of all $(1+)$-integrable estimators (hence in particular UMVU) for every n.

Earlier examples of models such that $(\overline{X}_n\colon n \in \mathbf{N})$ is asymptotically inadmissible although $\overline{X}_n$ is optimal unbiased for every n have been given by Portnoy (1977) and Pfanzagl (1993). The above argument shows that this phenomenon is generic for sufficiently smooth parametric models, if optimality refers to the class of $(1+)$-integrable estimators.

To prove the above claim (i), assume the contrary. Let $n \in \mathbf{N}$ and let $g = g(X_1, \dots, X_n)$ be a quadratically integrable statistic depending symmetrically on the X_ν. By symmetric 2-completeness, g is UMVU for its expectation. Since, by assumption, $\sum_{\nu=1}^n X_\nu$ is sufficient, an application of the Rao–Blackwell theorem yields the existence of some Borel function h satisfying

$$g = h\left(\sum_{\nu=1}^{n} X_\nu\right) \quad [\mathscr{P}^n]. \tag{10}$$

In the case that $\mathscr{P}$ is concentrated on the integers, we may choose integers k_1, k_2, k_3 with $k_1 < k_2 < k_3$ such that $P_\vartheta(\{k_1\}) \cdot P_\vartheta(\{k_2\}) \cdot P_\vartheta(\{k_3\}) > 0$ for some (hence every) $\vartheta \in \Theta$. Put $n := k_3 - k_1$, $n_1 := k_2 - k_1$ and $n_2 := n - n_1 = k_3 - k_2$. Then $nk_2 = n_1 k_3 + n_2 k_3$. Put $g := \sum_{\nu=1}^n 1(X_\nu = k_2)$. Then the events $\{\sum_{\nu=1}^n X_\nu = nk_2,\ g = n\}$ and $\{\sum_{\nu=1}^n X_\nu = nk_2,\ g = 0\}$ both have positive P_ϑ^n-probability, contradicting (10).

In the case that $\mathscr{P}$ is dominated by Lebesgue measure, we take $n = 2$ and $g(X_1, X_2) = |X_1 - X_2|$. Then (10) implies that

$$\{(x, y) \in \mathbf{R}^2\colon |x - y| = h(x + y)\}$$

has positive two-dimensional Lebesgue measure. By a change of variables, the same must be true for

$$\{(s, t) \in \mathbf{R}^2\colon |t| = h(s)\},$$

a union of two graphs, which is impossible.

To see that claim (ii) is true, first observe that

$$\gamma(\vartheta) = \int X \frac{1 + g_\vartheta}{\int (1 + g_\vartheta)\, dP_\vartheta}\, dP_\vartheta.$$

Hence

$$\begin{aligned} \gamma'(\vartheta) = {} & \int X \left(\frac{d}{d\vartheta} \frac{1 + g_\vartheta}{\int (1 + g_\vartheta)\, dP_\vartheta} \right) dP_\vartheta \\ & + \int X \frac{1 + g_\vartheta}{\int (1 + g_\vartheta)\, dP_\vartheta}\, d\left(\frac{d}{d\vartheta} P_\vartheta \right). \end{aligned} \tag{11}$$

Obviously, if G is close to 0 with respect to the metric in $\mathscr{G}$, then $(1 + g_\vartheta)/\int (1 + g_\vartheta)\, dP_\vartheta$ is close to 1 and $(d/d\vartheta)[(1 + g_\vartheta)/\int (1 + g_\vartheta)\, dP_\vartheta]$ is close to 0, in both cases with respect to the supremum norm in $\mathscr{L}^\infty(\mathscr{A})$ and uniformly in

ϑ. Hence, using (6), the first integral in (11) is close to 0 uniformly in ϑ, and, using (6) and (7), the second integral in (11) is close to

$$\int X \, d\left(\frac{d}{d\vartheta} P_\vartheta\right) = 1,$$

the latter equality holding in view of (8). Hence the statement involving the derivative of γ is true.

Checking the other statements concerning $\mathscr{P}_G$ made in claim (ii) is easy.

3. Remarks. This section contains remarks on possible and nonpossible generalizations and the history of the theorems in Section 1. The numbers of the remarks are the same as the numbers of the corresponding theorems.

REMARK 1. A proof of Theorem 1, paralleling the classical one given by Halmos (1946) for the special case where $\mathscr{P}$ consists of all discrete measures, may be found in Pfanzagl (1994), page 21. A more general result [suggested by Lehmann (1959), page 152] is given in Mandelbaum and Rüschendorf (1987), Theorem 7′, with a more complicated proof. Both proofs rely on a result of Landers and Rogge (1976).

REMARK 2. Contrary to statements made in Bell, Blackwell and Breiman (1960) and in Mandelbaum and Rüschendorf (1987), page 1239, the word "ring" in Theorem 2 cannot be replaced by "strong semi-algebra" and, in particular, not by "strong semi-ring" [the terminology used here being as in Billingsley (1986), pages 164 and 170]. (Note that if Theorem 2 were true with "strong semi-ring" instead of "ring," then Theorem 3 would trivially follow by considering the strong semi-ring consisting of the empty set and all left-open and right-closed intervals.) A counterexample is given by taking $(\mathscr{X}, \mathscr{A}, \mu)$ as the interval $(0, 1]$ with Lebesgue measure on its Borel σ-algebra generated by the strong semi-algebra consisting of the empty set and all left-open and right-closed subintervals: for sample size 3, an example of a function not obeying the definition "(1) $\Rightarrow$ (2)" is the difference of the sample mean and the sample median, since the latter two statistics have, by symmetry, identical expectations under every $P \in \mathscr{P}$.

On the other hand, Theorem 2 as stated is a slight extension of results stated in the literature in that no assumption of σ-finiteness is imposed on μ. [Fraser (1954) considered finite measures μ only.] To see that this extension is valid, use Theorem 2 for finite μ and apply it, for every $B \in \mathscr{R}$ with $0 < \mu(B) < \infty$, to $\tilde{\mathscr{X}} := B$, $\tilde{\mathscr{A}} := \{A \cap B\colon A \in \mathscr{A}\}$, $\tilde{\mathscr{R}} := \{A \cap B\colon A \in \mathscr{R}\}$, $\tilde{\mu} := \mu(\cdot \cap B)$ and $\tilde{\mathscr{P}}$ defined in terms of $\tilde{\mu}$ as $\mathscr{P}$ has been defined in terms of μ. Since $\tilde{\mathscr{R}}$ is a ring generating $\tilde{\mathscr{A}}$, we see, for any $\mathscr{A}_{\mathrm{sym}}$-measurable function, that

$$\int h \, dP^n = 0 \qquad (P \in \mathscr{P}),$$

implies, via $\int h|_{B^n}\, d\tilde{P}^n = 0$, $\tilde{P} \in \tilde{\mathscr{P}}$, in particular, $h|_{B^n} = 0$ [$\tilde{\mu}^n$]. Hence

$$h = 0 \quad \left[\left(\frac{1_B}{\mu(B)}\mu\right)^n\right] \qquad (0 < \mu(B) < \infty),$$

holds, which is the desired conclusion. [This argument is almost given in Heyer (1982), page 137.]

REMARK 3. This has obvious discrete and multivariate analogs. Further, it suffices to consider either step functions or smooth functions only as densities.

REMARK 4. More generally, one can write $\varepsilon = \varepsilon(P_0)$ in Theorem 4, and it suffices in fact to have $\varepsilon(P_0) > 0$ on a subset of $\mathscr{P}_0$ having the same null sets as $\mathscr{P}_0$.

REMARK 5. This could alternatively have been formulated in terms of the metric defined for probability measures P and Q by

$$d(P, Q) := \sup\left\{\frac{|P(A) - Q(A)|}{(1/2)(P(A) + Q(A))} : A \in \mathscr{A}\right\}.$$

In any case it is clear that $\mathscr{P}_\varepsilon$ is a rather small neighborhood, much smaller than, for example, a total variation neighborhood.

A possible extension is to replace the constant ε by a function $\varepsilon(\cdot) > 0$.

REMARK 6. It introduces only further notational complexity, not real difficulties, to extend Theorem 6 and its proof to parameter spaces in higher dimensions.

REMARK 7. The above remark concerning higher dimensions would also apply to Theorem 7. But a moment of reflection shows that, for example, the obvious two-dimensional analog follows, in view of the homogeneity condition, already from the stated one-dimensional version.

4. Proofs.

PROOF OF THEOREM 3. First fix $x_0 \in \mathbf{R}$ and consider the family of all unimodal probability measures which have a mode at x_0. This family is obviously convex and, by considering uniform distributions on $[x_0, x_0 + h]$ and $[x_0 - h, x_0]$, easily seen to be complete. Hence Theorem 1 yields the symmetric completeness of each corresponding IID model. The theorem as stated (i.e., without a mode specified) follows by observing that any union of symmetrically complete models is symmetrically complete. □

PROOF OF THEOREM 4. For fixed $P_0 \in \mathscr{P}_0$, the model

$$\mathscr{P}_\varepsilon(P_0) := \{(1-t)P_0 + tQ: Q \in \mathscr{Q}(P_0), t \in [0, \varepsilon]\}$$

is easily seen to be convex and complete. Hence Theorem 1 yields that $\mathscr{P}_\varepsilon^n(P_0)$ is symmetrically complete for every sample size n, for each P_0, and so must be $\mathscr{P}_\varepsilon^n = \bigcup_{P_0 \in \mathscr{P}_0} \mathscr{P}_\varepsilon^n(P_0)$. □

PROOF OF THEOREM 5. For $P_0 \in \mathscr{P}_0$ put

$$\mathscr{P}_\varepsilon(P_0) := \left\{ fP_0 : \sup_{x \in \mathscr{X}} |f(x) - 1| \le \varepsilon \right\}.$$

Each $\mathscr{P}_\varepsilon(P_0)$ is obviously convex. To prove completeness of each $\mathscr{P}_\varepsilon(P_0)$, fix P_0 and assume that $h: (\mathscr{X}, \mathscr{A}) \mapsto (\mathbf{R}^1, \mathscr{B}^1)$ satisfies $\int h \, dP = 0$ for every $P \in \mathscr{P}_\varepsilon(P_0)$. Then, for every $A \in \mathscr{A}$, we may put

$$f_A = \frac{1 + \varepsilon \cdot 1_A}{1 + \varepsilon \cdot P_0(A)}$$

and easily check that $f_A P_0 \in \mathscr{P}_\varepsilon(P_0)$. This implies $\int h \cdot f_A \, dP_0 = 0$ and hence, since $\int h \, dP_0 = 0$, $\int_A h \, dP_0 = 0$. Since A was arbitrary, we get $h = 0\,[P_0]$ and thus $h = 0\,[\mathscr{P}_\varepsilon(P_0)]$. Hence symmetric completeness of $\mathscr{P}_\varepsilon^n = \bigcup_{P_0 \in \mathscr{P}_0} \mathscr{P}_\varepsilon^n(P_0)$ follows again from Theorem 1, for each n. □

PROOF OF THEOREM 6. We denote the total variation norm of signed measures by $\|\cdot\|_V$ and the supremum norm of bounded real- or $\mathscr{L}^\infty(\mathscr{A})$-valued functions simply by $\|\cdot\|$. We write $\otimes^n Q$ for the n-fold product of a signed measure Q. Denote the set in (5) by $\mathscr{C}$.

Step 1. $\mathscr{G}$ is an open subset of the metric space $\overline{\overline{\mathscr{G}}}$ defined with "$\le \frac{1}{2}$" instead of "$< \frac{1}{2}$" in (3) and with metric d as in (4). $\overline{\overline{\mathscr{G}}}$ is easily seen to be complete. Hence $\mathscr{G}$ is a Baire space, that is, a topological space such that every countable intersection of open dense subsets is dense.

Strategy of this proof. We will exhibit a countable family of sets, consisting of the sets $\mathscr{C}_{n, \vartheta_0, \varphi, m}$ introduced in step 6, such that their intersection is contained in $\mathscr{C}$ (steps 2, 5 and 6), and each of them is open (step 7) and dense (step 8). Steps 3 and 4 are preparations, necessitated by the countability requirement, for performing step 5, the goal of which is to simplify condition (14) in the definition of $\mathscr{C}_n$ to $\int h \, d \otimes^n \varphi P_{\vartheta_0} = 0$, with n, ϑ_0 and φ fixed. The aim of step 6 is to make steps 7 and 8 possible.

The idea underlying the crucial but unfortunately somewhat messy step 8 may be described as follows. We would like to have

$$\tilde{g}_\vartheta = \varphi \cdot (\vartheta - \vartheta_0)^{k+1} \qquad (\vartheta \text{ near } \vartheta_0),$$

so that

$$\begin{aligned}\bigotimes^n \varphi P_{\vartheta_0} &= \left(\frac{1}{\beta^{(k+1)n}}\Delta_\beta^{(k+1)n} \bigotimes^n (1+\tilde{g}_\vartheta)P_{\vartheta_0}\right)_{\vartheta=\vartheta_0}\\ &\approx \left(\frac{1}{\beta^{(k+1)n}}\Delta_\beta^{(k+1)n} \bigotimes^n (1+\tilde{g}_\vartheta)P_{\vartheta}\right)_{\vartheta=\vartheta_0}\\ &\in \operatorname{span}\left\{\bigotimes^n (1+g_\vartheta)P_\vartheta : \vartheta\in\Theta\right\},\end{aligned}$$

where Δ_β denotes differencing with respect to ϑ, with sufficiently small increment β, and "$\approx$" should hold by smoothness of $\mathscr{P}$. The actual definition (20) of $\tilde{g}_\vartheta$ contains, of course, g_ϑ, with suitable smooth cutoff functions, such that (22) can be proved while the above argument can still be made rigorous; that is, (23) is true. The mess comes in when we start choosing the constants appropriately.

Step 2. We have $\mathscr{C} = \bigcap_{n\in\mathbf{N}}\mathscr{C}_n$, where

$\mathscr{C}_n := \{G\in\mathscr{G} : \mathscr{P}_G^n$ symmetrically boundedly complete for sample size $n\}$.

Step 3. Choose a countable set $\Theta_0 \subset \operatorname{int}\Theta$ such that

$$\left\{\bigotimes^n P_{\vartheta_0} : \vartheta_0\in\Theta_0\right\} \equiv \left\{\bigotimes^n P_\vartheta : \theta\in\Theta\right\}$$

in the sense that both of the above sets of probability measures have the same null sets. This is possible since $\{\otimes^n P_\vartheta : \vartheta\in\Theta\}$ is separable with respect to total variation distance, hence dominated [compare Witting (1985), page 139].

Step 4. For each $\vartheta_0\in\Theta_0$ we choose a countable set $D_{\vartheta_0}\subset\mathscr{L}^\infty(\mathscr{A})$ with $1\in D_{\vartheta_0}$,

$$\varphi\ge 0,\qquad \|\varphi\|\le 1,\qquad \varphi\in D_{\vartheta_0},$$

and such that the IID model based on the family

$$\{\varphi P_{\vartheta_0} : \varphi\in D_{\vartheta_0}\} \tag{12}$$

is symmetrically complete for sample size n. [The definitions of "completeness" and "symmetrical completeness" as well as the following argument are in no way affected by the possible non-normedness of the finite measures φP_{ϑ_0} occurring in (12).]

For example, we may take a countable generator $\mathscr{E}$ of $\mathscr{A}$ which contains $\mathscr{X}$ and is closed with respect to finite intersections and put

$$D_{\vartheta_0} = \left\{\sum_{k=1}^N r_k 1_{E_k} : N\in\mathbf{N},\ E_k\in\mathscr{E},\ r_k\ge 0 \text{ rational and } \sum r_k = 1\right\},$$

independently of ϑ_0. Then the family in (12) is complete and weakly convex in the sense of Mandelbaum and Rüschendorf (1987), page 1233, and hence symmetric completeness for every sample size follows, for example, from Mandelbaum and Rüschendorf (1987), page 1239, Theorem 7.

Step 5. We have

$$\mathscr{C}_n \supset \bigcap_{\vartheta_0 \in \Theta_0} \bigcap_{\varphi \in D_{\vartheta_0}} \mathscr{C}_{n, \vartheta_0, \varphi},$$

where

$$\mathscr{C}_{n, \vartheta_0, \varphi} = \left\{ G \in \mathscr{G}: h \in \mathscr{L}^\infty(\mathscr{A}^n_{\text{sym}}), \int h \, d \bigotimes^n (1 + g_\vartheta) P_\vartheta = 0 \; (\vartheta \in \Theta) \right. \\ \left. \Rightarrow \int h \, d \bigotimes^n \varphi P_{\vartheta_0} = 0 \right\}.$$

Proof. Let $G \in \mathscr{C}_{n, \vartheta_0, \varphi}$ for each ϑ_0 and φ and assume that

$$h \in \mathscr{L}^\infty(\mathscr{A}^n_{\text{sym}}), \qquad \int h \, d \bigotimes^n P_{G, \vartheta} = 0 \qquad (\vartheta \in \Theta). \tag{13}$$

By the definition of $\mathscr{C}_{n, \vartheta_0, \varphi}$ we have, for each $\vartheta_0 \in \Theta_0$,

$$\int h \, d \bigotimes^n \varphi P_{\vartheta_0} = 0 \qquad (\varphi \in D_{\vartheta_0}).$$

The symmetric completeness of the IID model based on the family in (12) and $1 \in D_{\vartheta_0}$ imply

$$h = 0 \quad \left[\bigotimes^n P_{\vartheta_0} \right].$$

Since $\vartheta_0 \in \Theta_0$ was arbitrary, step 3 implies

$$h = 0 \quad \left[\bigotimes^n P_\vartheta \right] \qquad (\vartheta \in \Theta),$$

and thus

$$h = 0 \quad \left[\bigotimes^n P_{G, \vartheta} \right] \qquad (\vartheta \in \Theta). \tag{14}$$

Hence $G \in \mathscr{C}_n$.

Step 6. For each $\vartheta_0 \in \Theta_0$ and $\varphi \in D_{\vartheta_0}$,

$$\mathscr{C}_{n, \vartheta_0, \varphi} \supset \bigcap_{m \in \mathbf{N}} \mathscr{C}_{n, \vartheta_0, \varphi, m},$$

where

$$\mathscr{C}_{n, \vartheta_0, \varphi, m} := \left\{ G \in \mathscr{G}: \left\| \bigotimes^n \varphi P_{\vartheta_0} - Q \right\|_V < \frac{1}{m} \right. \\ \left. \text{for some } Q \in \operatorname{span}\left\{ \bigotimes^n (1 + g_\vartheta) P_\vartheta : \vartheta \in \Theta \right\} \right\}.$$

Proof. Let $G \in \bigcap_{m \in \mathbf{N}} \mathscr{C}_{n, \vartheta_0, \varphi, m}$ and assume

$$h \in \mathscr{L}^\infty(\mathscr{A}^n_{\mathrm{sym}}), \qquad \int h \, d \bigotimes^n (1 + g_\vartheta) P_\vartheta = 0 \qquad (\vartheta \in \Theta).$$

Let $m \in \mathbf{N}$ and choose $Q \in \operatorname{span}\{\otimes^n (1 + g_\vartheta) P_\vartheta \colon \vartheta \in \Theta\}$ with

$$\left\| \bigotimes^n \varphi P_{\vartheta_0} - Q \right\|_V < \frac{1}{m}. \tag{15}$$

Then

$$\left| \int h \bigotimes^n \varphi \, d \otimes P_{\vartheta_0} \right| = \left| \int h \, d\left(\bigotimes^n \varphi P_{\vartheta_0} - Q \right) \right| \le \|h\| \frac{1}{m}. \tag{16}$$

Since m was arbitrary, we conclude that $\int h \, d \otimes^n \varphi P_{\vartheta_0} = 0$. Hence $G \in \mathscr{C}_{n, \vartheta_0, \varphi}$.

Step 7. For each ϑ_0, φ, m, the set $\mathscr{C}_{n, \vartheta_0, \varphi, m}$ is open in $\mathscr{G}$.

Proof. Let $G \in \mathscr{C}_{n, \vartheta_0, \varphi, m}$. Then

$$\eta := \left\| \bigotimes^n \varphi P_{\vartheta_0} - \sum_{k=1}^N \alpha_k \bigotimes^n (1 + g_{\vartheta_k}) P_{\vartheta_k} \right\|_V < \frac{1}{m}$$

for some $N \in \mathbf{N}$, $\alpha_1, \dots, \alpha_k \in \mathbf{R}$ and $\vartheta_1, \dots, \vartheta_k \in \Theta$. Put

$$\delta := \frac{1/m - \eta}{n(3/2)^{n-1} \sum_1^N |\alpha_k|}.$$

Then $\mathscr{C}_{n, \vartheta_0, \varphi, m}$ contains the open ball of radius δ around G.

Step 8. For each ϑ_0, φ, m, the set $\mathscr{C}_{n, \vartheta_0, \varphi, m}$ is dense in $\mathscr{G}$.

Proof. Let $G \in \mathscr{G}$ and $\varepsilon > 0$. We may assume that $\varepsilon < \frac{1}{2} - \|G\|$. Choose a positive integer k such that $\sum_{j=k+1}^\infty 2^{-j-1} < \varepsilon/2$. Let $h \in \mathscr{C}^\infty(\mathbf{R})$ be a function identically 1 on the interval $[0, (k+1)n]$ and identically 0 outside $[-1, (k+2)n]$. Put

$$C_1 := \max_{0 \le l \le k} \|h^{(l)}\|$$

and

$$\delta := \frac{\varepsilon}{2C_1(k+1)!(2(k+2)n)^{k+1}}.$$

Put

$$R_\vartheta := P_\vartheta - P_{\vartheta_0} \qquad (\vartheta \in \Theta),$$

and choose $\alpha > 0$ such that $[\vartheta_0, \vartheta_0 + \alpha] \subset \Theta$ (possible, since $\vartheta_0 \in \operatorname{int} \Theta$ by step 3).

During this step, we write as shorthand

$$t = \vartheta - \vartheta_0$$

and let D denote differentiation with respect to ϑ.

Put

$$C_2 := \sup_{\substack{\vartheta\in[\vartheta_0,\vartheta_0+\alpha],\\ n_1+n_2+n_3=n}} \left\| D^{(k+1)n}\left(\bigotimes^{n_1} \varphi P_{\vartheta_0} t^{k+1} \otimes \bigotimes^{n_2}\left(1+\sum_{l=0}^{k} g_{\vartheta_0}^{(l)} t^l\right) P_\vartheta \otimes \bigotimes^{n_3} \varphi t^{k+1} R_\vartheta \right)\right\|_V . \tag{17}$$

Choose $\beta > 0$ and $< \min(\delta, 1/(k+1)n)$ such that $I := [\vartheta_0, \vartheta_0 + (k+1)n\beta] \subset [\vartheta_0, \vartheta_0 + \alpha]$,

$$\left| g_\vartheta^{(i)} - \sum_{l=0}^{k-i} g_{\vartheta_0}^{(i+l)} \frac{t^l}{l!} \right| \le \frac{\varepsilon |t|^{k-i}}{C_1 2^{k+2}((k+2)n)^k}, \qquad |t| \le \beta(k+2)n,\ i = 0,\dots,k, \tag{18}$$

and

$$2^{(k+1)n} \frac{((k+1)n)^{(k+1)n}}{((k+1)n)!} \sup_{\vartheta\in I} \|R_\vartheta\|_V < \frac{1}{m3^n}, \qquad \frac{\beta C_2}{((k+1)n)!}\delta < \frac{1}{m3^n}. \tag{19}$$

Define $\tilde{G} = (\tilde{g}_\vartheta \colon \vartheta \in \Theta)$ by

$$\tilde{g}_\vartheta := \left(\sum_{l=0}^{k} g_{\vartheta_0}^{(l)} \frac{t^l}{l!} + \frac{\delta}{\beta}\varphi t^{k+1} \right) h\left(\frac{t}{\beta}\right) + g_\vartheta\left(1 - h\left(\frac{t}{\beta}\right)\right) \quad (\vartheta \in \Theta). \tag{20}$$

Then

$$g_\vartheta - \tilde{g}_\vartheta = \left(g_\vartheta - \sum_{l=0}^{k} g_{\vartheta_0}^{(l)} \frac{t^l}{l!} \right) h\left(\frac{t}{\beta}\right) - \frac{\delta}{\beta}\varphi t^{k+1} h\left(\frac{t}{\beta}\right) \tag{21}$$

is identically 0 for $|t| \ge \beta(k+2)n$, whereas for $|t| \le \beta(k+2)n$ an application of the Leibniz formula to both terms on the right-hand side in (21) yields together with (18), for $j = 0,\dots,k$,

$$\begin{aligned}\| D^j(g_\vartheta - \tilde{g}_\vartheta)\| &\le \sum_{i=0}^{j} \binom{j}{i} \frac{\varepsilon |t|^{k-i}}{C_1 2^{k+2}((k+2)n)^k} \frac{C_1}{\beta^{j-i}} \\ &\quad + \frac{\delta}{\beta}\|\varphi\| \sum_{i=0}^{j} \binom{j}{i} \frac{(k+1)!}{(k+1-i)!} |t|^{k+1-i} \frac{C_1}{\beta^{j-i}} \\ &\le \frac{\varepsilon}{4} + \frac{\varepsilon}{4} = \frac{\varepsilon}{2}\end{aligned}$$

[for the next to the last step, first replace j by k, then $|t|/\beta$ by $(k+2)n$, $k-i$ by k, and $(k+1-i)!$ by 1, then remember that $\|\varphi\| \le 1$ and look at the definition of δ]. Hence

$$\|G^{(j)} - \tilde{G}^{(j)}\| \le \frac{\varepsilon}{2} \qquad (j = 0, \dots, k),$$

so that $\tilde{G} \in \mathscr{G}$ and

$$d(G, \tilde{G}) \le \varepsilon. \tag{22}$$

It remains to prove that $\tilde{G} \in \mathscr{C}_{n, \vartheta_0, \varphi, m}$. To this end, put

$$Q := \frac{1}{((k+1)n)!\,\delta^n \beta^{kn}} \left(\Delta_\beta^{(k+1)n} \bigotimes^n (1 + \tilde{g}_\vartheta) P_\vartheta \right)_{\vartheta = \vartheta_0}.$$

Here Δ_β denotes the difference operator defined by $\Delta_\beta S_\vartheta := S_{\vartheta+\beta} - S_\vartheta$. Then $Q \in \operatorname{span}\{\otimes^n (1 + \tilde{g}_\vartheta) P_\vartheta \colon \vartheta \in \Theta\}$ and we claim that

$$\left\| \bigotimes^n \varphi P_{\vartheta_0} - Q \right\|_V < \frac{1}{m}. \tag{23}$$

To prove (23), note that, for $\vartheta \in [\vartheta_0, \vartheta_0 + (k+1)n\beta]$,

$$\begin{aligned} (1 + \tilde{g}_\vartheta) P_\vartheta &= \left(1 + \frac{\delta}{\beta} \cdot \varphi \cdot t^{k+1} + \sum_{l=0}^{k} g_{\vartheta_0}^{(l)} t^l \right) (P_{\vartheta_0} + R_\vartheta) \\ &= \frac{\delta}{\beta} \varphi P_{\vartheta_0} t^{k+1} + \left(1 + \sum_{l=0}^{k} g_{\vartheta_0}^{(l)} t^l \right) P_\vartheta + \frac{\delta}{\beta} \varphi t^{k+1} R_\vartheta. \end{aligned}$$

Now the tensor product

$$\bigotimes^n \left(\frac{\delta}{\beta} \varphi P_{\vartheta_0} t^{k+1} + \left(1 + \sum_{l=0}^{k} g_{\vartheta_0}^{(l)} t^l \right) P_\vartheta + \frac{\delta}{\beta} \varphi t^{k+1} R_\vartheta \right)$$

occurring in the definition of Q can be computed in an obvious way as a sum of 3^n tensor products. One of these products is

$$\frac{\delta^n}{\beta^n} t^{(k+1)n} \bigotimes^n \varphi P_{\vartheta_0}.$$

Its contribution to Q is precisely $\otimes^n \varphi P_{\vartheta_0}$.

There are $2^n - 1$ further products containing only the factors $(\delta/\beta)\varphi P_{\vartheta_0} t^{k+1}$ and $(\delta/\beta)\varphi \cdot t^{k+1} R_\vartheta$, and the latter at least once. Using the fact that the operator norm of $\Delta_\beta^{(k+1)n}$ with respect to the supremum norm is

$2^{(k+1)n}$, the contribution to Q of each of these products is seen to be bounded in norm by

$$\begin{aligned}&\frac{1}{((k+1)n)!\,\delta^n\beta^{kn}}\left(\frac{\delta}{\beta}\right)^n 2^{(k+1)n}((k+1)n)^{(k+1)n}\beta^{(k+1)n}\sup_{\vartheta\in I}\|R_\vartheta\|_V\\ &\quad= 2^{(k+1)n}\frac{((k+1)n)^{(k+1)n}}{((k+1)n)!}\sup_{\vartheta\in I}\|R_\vartheta\|_V < \frac{1}{m3^n}.\end{aligned}\tag{24}$$

Each of the remaining $3^n - 2^n$ products contains at least once the factor $(1 + \sum_{l=0}^{k} g_{\vartheta_0}^{(l)} t^l)P_\vartheta$. By applying the inequality connecting the norm of a difference operator with the norm of the corresponding differential operator [see, e.g., Dieudonné (1960), Section 8.12, Problem 4] and by looking at (17), the contribution to Q of each of these products is seen to be bounded in norm by

$$\frac{1}{((k+1)n)!\,\delta^n\beta^{kn}}\left(\frac{\delta}{\beta}\right)^{n-1}\beta^{(k+1)n}C_2 = \frac{\beta C_2}{((k+1)n)!\,\delta} < \frac{1}{m3^n}.$$

Hence (23) is true.

Step 9. By steps 2, 5, 6, 7 and 8 and the Baire property of $\mathscr{G}$, the theorem is proved. □

PROOF OF THEOREM 7. This is very similar to the proof of Theorem 6, so it suffices to indicate the necessary changes.

Step 1. Same as for Theorem 6.

Step 2. Replace "boundedly complete" in the definitions of $\mathscr{C}$ and $\mathscr{C}_n$ by "p-complete."

Steps 3 *and* 4. Same as for Theorem 6.

Step 5. Replace "$h \in \mathscr{L}^\infty(\mathscr{A}^n_{\mathrm{sym}})$" in the definition of $\mathscr{C}_{n,\vartheta_0,\varphi}$ and in (13) by "$h \in \bigcap_{\vartheta\in\Theta}\mathscr{L}^n_{\mathrm{sym}}(\otimes^n P_\vartheta)$" (with the obvious meaning of the subscript "sym").

Step 6. Replace $\mathscr{L}^\infty(\mathscr{A}^n_{\mathrm{sym}})$ by $\mathscr{L}^p_{\mathrm{sym}}(\{\otimes^n P_\vartheta\colon \vartheta\in\Theta\})$ and $\|h\|$ by $\|h\|_{L^p(\otimes^n P_{\vartheta_0})}$. Also replace $\|\cdot\|_V$ by $\|\cdot\|_{L^q(\otimes^n P_{\vartheta_0})}$ in the definition of $\mathscr{C}_{n,\vartheta_0,\varphi,m}$ and in (16).

Step 7. Replace $\|\cdot\|_V$ by $\|\cdot\|_{L^q(\otimes^n P_{\vartheta_0})}$.

Step 8. Replace $\|\cdot\|_V$ by $\|\cdot\|_{L^q(\otimes^n P_{\vartheta_0})}$ in (17) and (23) and choose the α one line before (17) also in accordance with assumption (iii$_p$), so that C_2 is well defined and finite. Also replace $\|\cdot\|_V$ by $\|\cdot\|_{L^q(P_{\vartheta_0})}$ in (19) and (24).

Step 9. Same as for Theorem 6. □

PROOF OF THE COROLLARY. An obvious consequence of the Baire property of $\mathscr{G}$, since the set in the corollary may be written as a countable intersection of the sets occurring in Theorem 7. □

Acknowledgments. I would like to thank J. Pfanzagl and D. Plachky for sending me copies of their papers from 1993 prior to publication, and

J. Pfanzagl, L. Rüschendorf and the referees for helpful remarks on earlier versions of this paper.

REFERENCES

ANDERSON, T. W. (1962). Least squares and best unbiased estimation. *Ann. Math. Statist.* **33** 266–272.

BELL, C. B., BLACKWELL, D. and BREIMAN, L. (1960). On the completeness of order statistics. *Ann. Math. Statist.* **31** 794–797.

BILLINGSLEY, P. (1986). *Probability and Measure*, 2nd ed. Wiley, New York.

DIEUDONNÉ, J. (1960). *Foundations of Modern Analysis*. Academic Press, New York.

FRASER, D. A. S. (1954). Completeness of order statistics. *Canad. J. Math.* **6** 42–45.

HALMOS, P. R. (1946). The theory of unbiased estimation. *Ann. Math. Statist.* **17** 34–43.

HEYER, H. (1982). *Theory of Statistical Experiments*. Springer, New York.

ISENBECK, M. and RÜSCHENDORF, L. (1992). Completeness in location families. *Probab. Math. Statist.* **13** 321–343.

LANDERS, D. and ROGGE, L. (1976). A note on completeness. *Scand. J. Statist.* **3** 139.

LEHMANN, E. L. (1959). *Testing Statistical Hypothesis*, 1st ed. Wiley, New York.

LEHMANN, E. L. (1983). *Theory of Point Estimation*. Wiley, New York.

LEHMANN, E. L. (1986). *Testing Statistical Hypothesis*, 2nd ed. Wiley, New York.

LEHMANN, E. L. and SCHOLZ, F. W. (1992). Ancillarity. In *Current Issues in Statistical Inference, Essays in Honor of D. Basu*. IMS, Hayward, CA.

MANDELBAUM, A. and RÜSCHENDORF, L. (1987). Complete and symmetrically complete families of distributions. *Ann. Statist.* **15** 1229–1244.

MATTNER, L. (1992). Completeness of location families, translated moments, and uniqueness of charges. *Probab. Theory Related Fields* **92** 137–149.

MATTNER, L. (1993). Some incomplete but boundedly complete location families. *Ann. Statist.* **21** 2158–2162.

MÜLLER-FUNK, U., PUKELSHEIM, F. and WITTING, H. (1989). On the attainment of the Cramér–Rao bound in $\mathbb{L}_r$-differentiable families of distributions. *Ann. Statist.* **17** 1742–1748.

OXTOBY, J. C. (1980). *Measure and Category*, 2nd ed. Springer, New York.

PFANZAGL, J. (1979). On optimal median unbiased estimators in the presence of nuisance parameters. *Ann. Statist.* **7** 187–193.

PFANZAGL, J. (1980). Asymptotic expansions in parametric statistical theory. In *Developments in Statistics* (P. K. Krishnaiah, ed.) **3** 1–97. Academic Press, New York.

PFANZAGL, J. (1993). Sequences of optimal unbiased estimators need not be asymptotically optimal. *Scand. J. Statist.* **20** 73–76.

PFANZAGL, J. (1994). *Parametric Statistical Theory*. de Gruyter, Berlin.

PLACHKY, D. (1993). An estimation-theoretical characterization of the Poisson distribution. *Statist. Decisions* 3 175–178.

PORTNOY, S. (1977). Asymptotic efficiency of minimum variance unbiased estimators. *Ann. Statist.* **5** 522–529.

RÜSCHENDORF, L. (1987). Estimation in the presence of nuisance parameters. In *Contributions to Stochastics* (W. Sendler, ed.) 190–201. Physica, Heidelberg.

RÜSCHENDORF, L. (1988). Unbiased estimation and local structure. In *Probability Theory and Mathematical Statistics with Applications* (W. Grossmann, J. Magyorodi, I. Vicze and W. Wertz, eds.) 295–306. Reidel, Dordrecht.

WITTING, H. (1985). *Mathematische Statistik I*. Teubner, Stuttgart.

UNIVERSITÄT HAMBURG
INSTITUT FÜR MATHEMATISCHE STOCHASTIK
BUNDESSTRASSE 55
D-20146 HAMBURG
GERMANY