The Annals of Statistics
1986, Vol. 14, No. 2, 665–678

THE STATISTICAL INFORMATION CONTAINED IN ADDITIONAL OBSERVATIONS[1]

By Enno Mammen

Universität Heidelberg

Let $\mathscr{E}^n$ be a statistical experiment based on n i.i.d. observations. We compare $\mathscr{E}^n$ with $\mathscr{E}^{n+r_n}$. The gain of information due to the r_n additional observations is measured by the deficiency distance $\Delta(\mathscr{E}^n, \mathscr{E}^{n+r_n})$, i.e., the maximum diminution of the risk functions. We show that under general dimensionality conditions $\Delta(\mathscr{E}^n, \mathscr{E}^{n+r_n})$ is of order r_n/n. Further the behavior of Δ is studied and compared for asymptotically Gaussian experiments. We show that the information gain increases logarithmically. The Gaussian and the binomial family turn out to be—in some sense—opposite extreme cases, with the increase of information asymptotically minimal in the Gaussian case and maximal in the binomial.

1. Introduction. When considering a complicated statistical model it may be useful to construct another model which is close to the original one but statistically easier to handle. The analysis of the second model may make the essential structure of the first model better understandable and help to construct suitable statistical procedures for a decision problem. The usual way to get such an approximating model is to imbed the original one into a sequence of models and to expand the log-likelihood function. Because one is more interested in approximations than in limit theorems it is necessary to estimate the closeness of the two models. A natural quantity for comparing two models or—in more common use of language—two experiments is the deficiency distance of Le Cam (1964). It is based on the comparison of risk functions available in the two experiments. We recall its definition.

Let $\mathscr{E} := (\mathscr{X}, \mathscr{A}, (P_\theta\colon \theta \in \Theta))$ and $\mathscr{F} := (\mathscr{Y}, \mathscr{B}, (Q_\theta\colon \theta \in \Theta))$ be two experiments with the same parameter set Θ, i.e., two families of probability measures $(P_\theta\colon \theta \in \Theta)$ and $(Q_\theta\colon \theta \in \Theta)$ defined on measurable spaces $(\mathscr{X}, \mathscr{A})$ and $(\mathscr{Y}, \mathscr{B})$, respectively. $\mathscr{E}$ is called ε-deficient relative to $\mathscr{F}$ ($\varepsilon > 0$), if for every finite decision space $(\mathscr{T}, \mathscr{S})$, for every bounded loss function $L\colon \Theta \times \mathscr{T} \to \mathbb{R}$ and for every decision rule σ in $\mathscr{F}$ there exists a decision rule ρ in $\mathscr{E}$ such that for every $\theta \in \Theta$ the following inequality between the risk functions is valid:

$$\begin{aligned}(1.1)\qquad &\int_{\mathscr{X}}\int_{\mathscr{T}} L(\theta, t)\rho(x, dt)P_\theta(dx)\\ &\leq \int_{\mathscr{Y}}\int_{\mathscr{T}} L(\theta, t)\sigma(x, dt)Q_\theta(dx) + \varepsilon \sup_{\theta \in \Theta,\, t \in \mathscr{T}} |L(\theta, t)|.\end{aligned}$$

The deficiency $\delta(\mathscr{E}, \mathscr{F})$ of $\mathscr{E}$ with respect to $\mathscr{F}$ is the smallest $\varepsilon > 0$ for which $\mathscr{E}$

Received July 1984; revised July 1985.

[1]This work has been supported by the Deutsche Forschungsgemeinschaft.

AMS 1980 *subject classification*. Primary 62B15.

Key words and phrases. Experiments, deficiency, information, additional observations.

is ε-deficient with respect to $\mathscr{F}$. The deficiency distance is the symmetrical quantity $\Delta(\mathscr{E}, \mathscr{F}) := \delta(\mathscr{E}, \mathscr{F}) \vee \delta(\mathscr{F}, \mathscr{E})$. It defines a pseudo-distance between experiments. We cite two other characterizations of the deficiency. For a detailed motivation and discussion see Le Cam (1964).

(i) *The randomisation criterion.*

$$\delta(\mathscr{E}, \mathscr{F}) = \inf_K \sup_{\theta \in \Theta} \|KP_\theta - Q_\theta\|, \tag{1.2}$$

where the infimum is taken over all transitions which map the band $L(\mathscr{E})$, generated by $(P_\theta\colon \theta \in \Theta)$, into the band $L(\mathscr{F})$, generated by $(Q_\theta\colon \theta \in \Theta)$. A transition K is a positive norm one linear map [i.e., $K\mu^+ \geq 0$, $\|K\mu^+\| = \|\mu^+\|$, $K(a\mu + b\nu) = aK\mu + bK\nu$ for $\mu, \nu \in L(\mathscr{E})$ and $a, b \in \mathbb{R}$]. If $(P_\theta\colon \theta \in \Theta)$ is dominated, $\mathscr{Y}$ is a Borel subset of a complete separable metric space and $\mathscr{B}$ is the class of Borel subsets of $\mathscr{Y}$ then it is sufficient to take the infimum in (1.2) over all Markov kernels from $(\mathscr{X}, \mathscr{A})$ to $(\mathscr{Y}, \mathscr{B})$.

(ii) *The Bayes criterion.*

$$\delta(\mathscr{E}, \mathscr{F}) = \sup\{\rho_{\mathscr{E}}(\pi, L, D) - \rho_{\mathscr{F}}(\pi, L, D)\}, \tag{1.3}$$

where the supremum is taken over all prior measures π with finite support, finite decision spaces D, and loss functions L bounded in absolute value by 1. $\rho_{\mathscr{E}}(\pi, L, D)$ resp. $\rho_{\mathscr{F}}(\pi, L, D)$ is the corresponding Bayes risk in $\mathscr{E}$ resp. $\mathscr{F}$.

Unfortunately, the deficiency distance is very difficult to calculate in general. But for translation experiments it suffices to take the infimum in (1.2) over invariant kernels. This can be used to calculate the deficiency explicitly in some cases. For instance, Torgersen (1972) showed that

$$\begin{aligned} \Delta(\mathscr{E}_R^n, \mathscr{E}_R^{n+r}) &= \int_0^1 |nx^{n-1} - (n+r)x^{n+r-1}|\,dx \\ &= (2/e)(r/n) + o(1/n) \approx 0.73(r/n) \end{aligned} \tag{1.4}$$

if $\mathscr{E}_r^n$ is the experiment of taking n i.i.d. observations from a rectangular distribution on $[0, \theta]$ for $\theta > 0$. He also showed that

$$\begin{aligned} \Delta(\mathscr{E}_E^n, \mathscr{E}_E^{n+r}) &= \|\chi^2_{n+r} - \chi^2_{n,1+r/n}\| \\ &= \sqrt{2/\pi e}\,(r/n) + o(1/n) \approx 0.48(r/n) \end{aligned} \tag{1.5}$$

if $\mathscr{E}_E^n$ is the experiment of observing n times an exponentially distributed variable. $\chi^2_{n,1+r/n}$ denotes the distribution of $(1 + r/n)X$, if X is distributed according to χ^2_n. Another example is

$$\Delta(\mathscr{G}^n, \mathscr{G}^{n+r}) = \|N(0, I_k/n) - N(0, I_k/(n+r))\| \tag{1.6}$$

for the Gaussian shift experiment $\mathscr{G} := (N(\theta, \Sigma)\colon \theta \in \mathbb{R}^k)$. I_k is the $k \times k$ identity matrix and Σ a positive definite $k \times k$ matrix.

For $k = 1$ (1.6) yields

$$\Delta(\mathscr{G}^n, \mathscr{G}^{n+r}) = \sqrt{2/\pi e}\, r/n + o(1/n). \tag{1.7}$$

The results in Le Cam (1964) and Torgersen (1972) on invariance are also used in Swensen (1980) to compute deficiencies between linear models.

For one-dimensional exponential families $\mathscr{E}$ Helgeland (1982) has calculated lower and upper asymptotic bounds for $\Delta(\mathscr{E}^n, \mathscr{E}^{n+r_n})$:

$$\begin{aligned} \sqrt{2/\pi e} &\le \liminf_{n\to\infty} \frac{n}{r_n} \Delta(\mathscr{E}^n, \mathscr{E}^{n+r_n}) \\ &\le \limsup_{n\to\infty} \frac{n}{r_n} \Delta(\mathscr{E}^n, \mathscr{E}^{n+r_n}) \\ &\le 2\sqrt{2/\pi e}, \end{aligned} \tag{1.8}$$

provided $r_n \le n^\beta$ for some fixed $\beta < 1$.

For finite parameter sets the behaviour of the products $\mathscr{E}^n$ of an experiment $\mathscr{E}$ has been studied by Torgersen (1981). Using the deficiency he compares $\mathscr{E}^n$ with the totally informative experiment and the least informative experiment. In the context of robust statistics the deficiency distance can be used to measure how much the assumed model differs from the true model. Müller (1980/81) gives an estimate of the deficiency distance between two different models, in terms of the bounded Lipschitz distance between the probability measures. This estimate is of the correct order as the Lipschitz distance tends to zero.

Some of our examples deal with measuring the distance between an experiment $\mathscr{E} = (\mathscr{X}, \mathscr{A}, (P_\theta\colon \theta \in \Theta))$ and some subexperiment $\mathscr{E}' = (\mathscr{X}, \mathscr{A}', (P_\theta|_{\mathscr{A}'}\colon \theta \in \Theta))$ (i.e., $\mathscr{A}' \subset \mathscr{A}$). A quantity $\eta(\mathscr{E}', \mathscr{E})$ called insufficiency has been introduced for this situation by Le Cam (1974). The insufficiency seems to be more tractable than the deficiency. It measures how much the P_θ's must be modified in order for $\mathscr{A}'$ to be sufficient. Under certain regularity conditions on the experiments $\mathscr{E}$, $\mathscr{E}'$ the insufficiency $\eta(\mathscr{E}', \mathscr{E})$ can be defined by

$$\eta(\mathscr{E}', \mathscr{E}) = \inf \sup_{\theta\in\Theta} \|P_\theta - P'_\theta\|,$$

where the infimum is taken over all families of measures $(P'_\theta\colon \theta \in \Theta)$ for which $\mathscr{A}'$ is sufficient and P'_θ and P_θ agree on $\mathscr{A}'$. For a general definition see Le Cam (1974).

The notion of insufficiency can be used to measure the information contained in additional observations. Le Cam (1974) gives the following general estimate:

$$\eta(\mathscr{E}^n, \mathscr{E}^{n+r}) \le \sqrt{8\beta}\sqrt{r/n} \tag{1.9}$$

with $\beta := \inf_{\hat\theta} \sup_{\theta\in\Theta} n E_\theta H^2(P_\theta, P_{\hat\theta})$. Here the infimum has to be taken over all randomised estimates $\hat\theta$ of θ in the experiment $\mathscr{E}^n$, and $H(\cdot,\cdot)$ is the Hellinger distance, i.e., $H^2(P, Q) := \int(\sqrt{dP} - \sqrt{dQ})^2$ for two probability measures. Using a dimensionality condition Le Cam has shown that β is bounded [see Birgé (1983)]:

$$\beta \le 6.5D + 5.5 \tag{1.10}$$

with a dimensionality constant D defined as follows:

(1.11) Consider $h(\theta, \tau) := H(P_\theta, P_\tau)$ as a pseudo-distance on θ. Then D is the smallest number such that, for every $\delta > 0$, every subset of Θ with diameter δ can be covered by 2^D sets of diameter $\delta/2$.

(1.9) and (1.10) yield $\eta(\mathscr{E}^n, \mathscr{E}^{n+1}) = O(1/\sqrt{n})$ for finite dimensional experiments. For k-dimensional Gaussian experiments $\mathscr{G}$ this is the right order [Le Cam (1974)]. Because of $\eta(\mathscr{E}, \mathscr{F}) \geq \delta(\mathscr{E}, \mathscr{F})$ the insufficiency can be used to calculate upper bounds for the deficiency. For instance one gets $\delta(\mathscr{E}^n, \mathscr{E}^{n+1}) = O(1/\sqrt{n})$ for finite dimensional $\mathscr{E}$.

As in the most given examples, in this paper we are concerned only with the calculation of the deficiency for the special case of comparing an experiment for a different number of observations. First we will show that $\delta(\mathscr{E}^n, \mathscr{E}^{n+1}) = O(1/n)$ for finite dimensional $\mathscr{E}$. This improves the above mentioned result which was based on the calculation of the insufficiency. In the rest of the paper the increase of information $\delta(\mathscr{E}^n, \mathscr{E}^{n+1})$ will be studied for asymptotically Gaussian experiments. Then $\delta(\mathscr{E}^n, \mathscr{E}^{an})$ converges for $n \to \infty$ to a constant depending only on a. Thus the information increases, as it were, logarithmically. For one additional observation the increase of information turns out to be asymptotically minimal for Gaussian experiments and—among exponential families—maximal for a binomial family. Further investigations of $\delta(\mathscr{E}^n, \mathscr{E}^{n+r})$ for exponential families can be found in Mammen (1983).

The following section formally presents our results and mentions some of the main elements of their proofs. Detailed proofs of these results are contained in Section 3.

2. Results. To prove his upper bounds in (1.8) Helgeland (1982) made use of the randomisation criterion (1.2). He constructed a kernel as follows. First estimate the parameter θ, then generate a random variable distributed according to the estimated measure and mix this variable randomly among the observations drawn first. This idea can also be used in the more general situation when in an experiment $\mathscr{E} = (\mathscr{X}, \mathscr{A}, (P_\theta\colon \theta \in \Theta))$ estimators exist which are $\sqrt{n}$-consistent in the following sense.

(2.1) There exist positive constants γ and B such that for every n there is an estimate $\hat{\theta}_n$ in $\mathscr{E}^n$ with (i) $E_\theta \exp(\gamma n h^2(\theta, \hat{\theta}_n)) \leq B$ and (ii) $x \to P_{\hat{\theta}_n(x)}(A)$ measurable for $A \in \mathscr{A}$.

Here $h(\theta, \tau) := H(P_\theta, P_\tau)$ is the pseudo-distance on Θ induced by the Hellinger distance.

THEOREM 1. *Let $\mathscr{E}$ be an experiment satisfying* (2.1). *Then there exists a constant C such that*

$$\Delta(\mathscr{E}^n, \mathscr{E}^{n+r}) \leq C\frac{r}{n}. \tag{2.2}$$

In particular, if $\mathscr{E}$ is an experiment which is finite dimensional in the sense of (1.11) then (2.1) and therefore Theorem 1 holds. Then the constant C depends only on the dimension D. This can be seen from results of Birgé, who proves (2.1) for finite dimensional experiments [see Dacunha-Castelle (1978) and Birgé (1983), where slightly different dimensionality conditions are used].

As a first step for calculating limit experiments Le Cam (1968) associates products of experiments $\mathscr{E}^n$ with Poisson experiments $\mathscr{P}^n$. The Poisson experiment is defined according to the following rule: first observe a Poisson variable N with mean value n. Then carry out the experiment $\mathscr{E}^N$. When applied to this situation Theorem 1 yields the following estimate:

COROLLARY. *For experiments $\mathscr{E}$ fulfilling* (2.1) *one has*

$$\Delta(\mathscr{E}^n, \mathscr{P}^n) = O(1/\sqrt{n}). \tag{2.3}$$

Now we discuss the case when the products $\mathscr{E}^n$ of an experiment $\mathscr{E} = (\mathscr{X}, \mathscr{A}, (P_\theta\colon\ \theta \in \Theta))$ can be approximated locally by a Gaussian experiment. More precisely we assume that $\Theta \subset \mathbb{R}^k$ and suppose that there is a $\theta_0 \in \Theta$ such that the following condition holds.

(2.4) There exists a positive-definite $k \times k$ matrix Σ, such that for all $c > 0$:

$$\Delta(\mathscr{E}^n_{n,c}, \mathscr{G}^n_{n,c,\Sigma}) \to 0 \quad \text{for } n \to \infty,$$

where $\mathscr{E}_{n,c} := (\mathscr{X}, \mathscr{A}, (P_\theta\colon\ \|\theta - \theta_0\| \le c/\sqrt{n}))$ and $\mathscr{G}_{n,c,\Sigma} := (\mathbb{R}^k, \mathscr{B}(\mathbb{R}^k), (N(\theta, \Sigma)\colon \|\theta - \theta_0\| \le c/\sqrt{n}))$.

Here $\mathscr{B}(\mathbb{R}^k)$ denotes the Borel-σ-algebra of $\mathbb{R}^k$. Sufficient conditions for (2.4) can be found in Le Cam (1968). For instance the following holds: Assume $\mathscr{E} := (P_\theta\colon \theta \in \Theta)$ is an experiment with $\Theta \subset \mathbb{R}^k$. Let θ_0 be an element of the interior of Θ, such that for θ in a neighborhood of θ_0, the measures P_θ can be dominated by a finite measure m. Further assume that the function $\xi(\theta) := \sqrt{dP_\theta/dm}$ mapping Θ into $L^2(m)$ is Frechet-differentiable at θ_0 with derivative $\dot{\xi}(\theta_0) \in L^2(m)^k$, and that the matrix $\Gamma := \int \dot{\xi}(\theta_0)\dot{\xi}(\theta_0)^T\,dm$ is positive definite. Then (2.4) holds with $\Sigma := \Gamma$. For more general conditions for local Gaussian approximation see Le Cam (1985).

Under further conditions local Gaussian approximations can be pieced together to a global approximation of $\mathscr{E}^n$ by a heteroscedastic Gaussian experiment. This was discussed in Le Cam (1975). For $\Theta \subset \mathbb{R}^k$ we state such an approximation.

(2.5) There exist positive definite $k \times k$ matrices $\Gamma(\theta)$ depending continuously on θ, such that $\Delta(\mathscr{E}^n, \hat{\mathscr{G}}^n) \to 0$ for $n \to \infty$, where $\tilde{\mathscr{G}} = (\mathbb{R}^k, \mathscr{B}(\mathbb{R}^k), (N(\theta, \Gamma^{-1}(\theta))\colon \theta \in \Theta))$.

It can be shown that

$$\Delta(\tilde{\mathscr{G}}^n, \tilde{\mathscr{G}}^{n+r_n}) = \Delta(\mathscr{G}^n, \mathscr{G}^{n+r_n}) + o(1),$$

where $\mathscr{G}$ is a k-dimensional homoscedastic Gaussian experiment. This proves the following theorem.

THEOREM 2. *Assume* $\mathscr{E} = (\mathscr{X}, \mathscr{A}, (P_\theta\colon \theta \in \Theta))$ *is an experiment with* $\Theta \subset \mathbb{R}^k$. *Assume further* (2.5). *Put* $\mathscr{G} := (\mathbb{R}^k, \mathscr{B}(\mathbb{R}^k), (N(\theta, I_k)\colon \theta \in \mathbb{R}^k))$. *Then for every sequence of integers* (r_n) *the following holds*:

$$\Delta(\mathscr{E}^n, \mathscr{E}^{n+r_n}) = \Delta(\mathscr{G}^n, \mathscr{G}^{n+r_n}) + o(1). \tag{2.6}$$

The statement of Theorem 2 is interesting only in the case of a large number of additional observations when r_n is of order n. This is because (2.6) holds trivially if $r_n = o(n)$, since then $\Delta(\mathscr{E}^n, \mathscr{E}^{n+r_n}) = o(1)$ and $\Delta(\mathscr{G}^n, \mathscr{G}^{n+r_n}) = o(1)$ as can be seen by (1.6) and Theorem 1 using the fact that $\mathscr{E}$ is finite-dimensional in the sense of (1.11).

In the special case where the number of additional observations is proportional to n $[r_n = [an] := \sup\{k \in \mathbb{N}\colon k \le an\}$ for a constant $a]$ one obtains by (1.6)

$$\begin{aligned} \Delta(\mathscr{E}^n, \mathscr{E}^{[an]}) &= \Delta(\mathscr{G}^n, \mathscr{G}^{[an]}) + o(1) \\ &= \| N(0, aI_k) - N(0, I_k) \| + o(1). \end{aligned} \tag{2.7}$$

Here the increase of information depends asymptotically only on the dimension k and the constant a. Thus the information increases, as it were, logarithmically in the number of observations.

Now we show that for a small number of additional observations the gain of information is asymptotically minimal in the case of Gaussian experiments.

THEOREM 3. *Let* $\mathscr{E} = (\mathscr{X}, \mathscr{A}, (P_\theta\colon \theta \in \Theta))$ *be an experiment with* $\Theta \subset \mathbb{R}^k$. *Assume that* $\mathscr{E}$ *can be approximated locally by homoscedastic Gaussian experiments in the sense of* (2.4) *at a point* θ_0 *contained in the interior of* Θ. *Then for all sequences* (r_n) *with* $r_n = o(n)$ *the following holds.*

$$\liminf_{n \to \infty} \frac{n}{r_n} \Delta(\mathscr{E}^n, \mathscr{E}^{n+r_n}) \ge \lim_{n \to \infty} \frac{n}{r_n} \Delta(\mathscr{G}^n, \mathscr{G}^{n+r_n}). \tag{2.8}$$

Theorem 3 is a generalization of the lower bound of Helgeland (1.8). The proof consists of the following simple arguments: First, for r fixed, $\Delta(\mathscr{E}^n, \mathscr{E}^{n+r})$ decreases in n (later observations are less informative). Asymptotically a large number of additional observations in the experiment $\mathscr{E}$ is not less informative than in a Gaussian experiment. Finally in Gaussian experiments the information increases almost additively. Combining these arguments one obtains the following asymptotic inequalities, where l has to be chosen suitably depending on n:

$$\begin{aligned} \frac{n}{r} \Delta(\mathscr{E}^n, \mathscr{E}^{n+r}) &\ge \frac{n}{lr} \Delta(\mathscr{E}^n, \mathscr{E}^{n+lr}) \\ &\ge \frac{n}{lr} \Delta(\mathscr{G}^n, \mathscr{G}^{n+lr}) + o(1) \\ &= \frac{n}{r} \Delta(\mathscr{G}^n, \mathscr{G}^{n+r}) + o(1). \end{aligned}$$

The next theorem shows that the upper bound in (1.8) is sharp and is attained by the binomial family.

THEOREM 4. *Assume* $0 < a < \frac{1}{2} < b < 1$. *Let* $\mathscr{E}_B := (\{0,1\}, 2^{\{0,1\}}, (Q_p\colon p \in (a, b))$ *be a Bernoulli experiment*:

$$Q_p(\{x\}) = px + (1-p)(1-x) \quad \text{for } x \in \{0,1\}.$$

Then

$$\Delta(\mathscr{E}_B^n, \mathscr{E}_B^{n+1}) = 2\Delta(\mathscr{G}^n, \mathscr{G}^{n+1}) + o(1/n). \tag{2.9}$$

Further investigations of $\Delta(\mathscr{E}^n, \mathscr{E}^{n+r})$ for exponential families can be found in Mammen (1983). There the inequality of Helgeland (1.8) is generalized to arbitrary finite dimensional exponential families. As in the one-dimensional case the increase of information is asymptotically at most twice as much as in the Gaussian case. Furthermore, it turns out that one has to distinguish two cases. If the measures of the exponential family are lattice distributions the gain of information is asymptotically strictly larger than in the Gaussian case. For strongly nonlattice distributions (i.e., measures fulfilling Cramér's condition) the information gain increases exactly as in the Gaussian case, asymptotically as $n \to \infty$. The proof of these results is based on Edgeworth expansions of Bayes risks which old uniformly over all Bayes decision problems with bounded loss function.

3. Proofs.

PROOF OF THEOREM 1. It suffices to prove Theorem 1 for $r = 1$. We construct a Markov kernel K from $(\mathscr{X}^n, \mathscr{A}^n)$ to $(\mathscr{X}^{n+1}, \mathscr{A}^{n+1})$. Let $m_n \le n$ be a sequence of natural numbers with

$$m_n/(n+1-m_n) \le \gamma, \tag{3.1}$$

$$n/m_n = O(1). \tag{3.2}$$

According to (2.1) there exists an estimate $\tilde{\theta}_n$ in $\mathscr{E}^n$ depending only on the first $n+1-m_n$ observations with

$$E_\theta \exp\big(\gamma(n+1-m_n)h^2(\theta, \tilde{\theta}_n)\big) \le B. \tag{3.3}$$

Using (3.1) one gets

$$E_\theta \exp\big(m_n h^2(\theta, \tilde{\theta}_n)\big) \le B. \tag{3.4}$$

The kernel K is defined as follows:

$$K = (1/m) \sum_{1 \le i \le m} \delta_{x_1} \times \cdots \times \delta_{x_{n-m+i}} \times P_{\tilde{\theta}_n} \times \delta_{x_{n-m+i+1}} \times \cdots \times \delta_{x_n}. \tag{3.5}$$

Here the index n of m is dropped. The randomization criterion (1.2) yields

$$\Delta(\mathscr{E}^n, \mathscr{E}^{n+1}) \leq \sup_{\theta\in\Theta}\|KP_\theta^n - P_\theta^{n+1}\| \tag{3.6}$$
$$= \sup_{\theta\in\Theta}\int\left\|1/m\sum_{1\leq i\leq m}P_\theta^{i-1}\times P_{\tilde{\theta}_n}\times P_\theta^{m-i} - P_\theta^m\right\| d\mathscr{L}(\tilde{\theta}_n|P_\theta^n).$$

To complete the proof we use the following lemma.

LEMMA 1. *For two probability measures the following holds*:

$$\left\|1/m\sum_{1\leq i\leq m}P^{i-1}\times Q\times P^{m-i} - P^m\right\|^2 \tag{3.7}$$
$$\leq \exp\big((m-2)H^2(P,Q)\big)4H^2(P,Q)\big[m^{-1} + H^2(P,Q))\big].$$

For example it follows from Lemma 1 that given two sequences of measures (P_n) and (Q_n) with $H(P_n, Q_n) = O(1/\sqrt{n})$, then $\|1/n\Sigma_{1\leq i\leq n}P_n^{i-1}\times Q_n\times P_n^{n-\iota} - P_n^n\|$ is of order $1/n$, whereas $\|P_n^{n-1}\times Q_n - P_n^n\| = \|P_n - Q_n\|$ may be eventually of order $1/\sqrt{n}$. Thus Lemma 1 underlines the importance of randomly mixing the generated variable in the construction of $\tilde{K}$.

APPLICATION OF LEMMA 1. Using Lemma 1 one gets from (3.6)

$$\Delta(\mathscr{E}^n, \mathscr{E}^{n+1})$$
$$\leq \sup_{\theta\in\theta}E_\theta\exp\left(\frac{m}{2}h^2(\theta,\tilde{\theta}_n)\right)\left(4h^2(\theta,\tilde{\theta}_n)/m + 4h^4(\theta,\tilde{\theta}_n)\right)^{1/2}.$$

An application of the Cauchy–Schwarz inequality, (3.4), and (3.2) finishes the proof.

PROOF OF LEMMA 1. Set $M := (P+Q)/2$, $g(x) := dP/dM(x)$, and $h(x) := 2 - g(x) = dQ/dM(x)$. The following holds:

$$\left\|1/m\sum_{1\leq i\leq m}P^{i-1}\times Q\times P^{m-i} - P^m\right\|^2$$
$$= \left(\int\left|1/m\sum_{1\leq i\leq m}g(x_1)\cdots g(x_{i-1})\right.\right.$$
$$\left.\left.\times\{h(x_i)-g(x_i)\}g(x_{i+1})\cdots g(x_m)\right|\prod_{1\leq i\leq m}M(dx_i)\right)^2$$
$$\leq \int\left(1/m\sum_{1\leq i\leq m}g(x_1)\cdots g(x_{i-1})\{h(x_i)-g(x_i)\}g(x_{i+1})\cdots g(x_m)\right)^2$$
$$\times\prod_{1\leq i\leq m}M(dx_i)$$
$$= 1/m\int(h-g)^2\,dM\left(\int g^2\,dM\right)^{m-1}$$
$$+\frac{m-1}{m}\left(\int(h-g)g\,dM\right)^2\left(\int g^2\,dM\right)^{m-2}.$$

The following estimates yield the lemma:

$$
\begin{aligned}
\int (h-g)^2\,dM &= 2\int (dP - dQ)^2/(dP+dQ) \\
&= 2\int \left(\sqrt{dP} - \sqrt{dQ}\right)^2\left(\sqrt{dP} + \sqrt{dQ}\right)^2/(dP+dQ) \qquad (3.8)\\
&\le 4H^2(P,Q),
\end{aligned}
$$

$$
\begin{aligned}
\left|\int (h-g)g\,dM\right| &= \left|\int (h-g)g\,dM - \tfrac{1}{2}\int (h-g)(h+g)\,dM\right| \\
&= \tfrac{1}{2}\int (h-g)^2\,dM \qquad (3.9)\\
&\le 2H^2(P,Q),
\end{aligned}
$$

$$
\begin{aligned}
\int g^2\,dM &= \tfrac{1}{4}\int ((g-h)+(g+h))^2\,dM \\
&= \tfrac{1}{4}\int (g-h)^2\,dM + 1 \qquad (3.10)\\
&\le 1 + H^2(P,Q),
\end{aligned}
$$

$$
\begin{aligned}
\left(\int g^2\,dM\right)^m &\le \left(1 + H^2(P,Q)\right)^m \qquad (3.11)\\
&\le \exp\left(mH^2(P,Q)\right).
\end{aligned}
$$

Proof of the Corollary. Assume N is a Poisson variable with $EN = n$. The following holds:

$$
\begin{aligned}
\Delta(\mathscr{E}^n, \mathscr{P}^n) &\le \delta(\mathscr{E}^n, \mathscr{P}^n) + \delta(\mathscr{P}^n, \mathscr{E}^n) \\
&\le \sum_{k\ge 0} P(N=k)\Delta(\mathscr{E}^n, \mathscr{E}^k) \\
&\le 2P(N < n/2) + \sum_{k \ge n/2} P(N=k)\Delta(\mathscr{E}^n, \mathscr{E}^k).
\end{aligned}
$$

Clearly, the first term is of order $o(1/\sqrt{n})$. To treat the second term we use the following lemma.

Lemma 2. *For every experiment $\mathscr{E}$ and $r \in \mathbb{N}$ the deficiency $\Delta(\mathscr{E}^n, \mathscr{E}^{n+r})$ is monotonically decreasing in n.*

Application of Lemma 2. One gets

$$
\begin{aligned}
\sum_{k\ge n/2} P(N=k)\Delta(\mathscr{E}^n, \mathscr{E}^k) &\le \sum_{k\ge n/2} P(N=k)\,|n-k|\,\Delta(\mathscr{E}^{[n/2]}, \mathscr{E}^{[n/2]+1}) \\
&\le E|N-n|\,O(1/n) = O(1/\sqrt{n}).
\end{aligned}
$$

PROOF OF LEMMA 2. Using the randomization criterion (1.2) one gets

$$\begin{aligned}\Delta(\mathscr{E}^{n+1}, \mathscr{E}^{n+1+r}) &= \delta(\mathscr{E}^{n+1}, \mathscr{E}^{n+1+r}) \\ &= \inf_K \sup_{\theta\in\Theta} \|KP_\theta^{n+1} - P_\theta^{n+1+r}\| \\ &\le \inf_L \sup_{\theta\in\Theta} \|(LP_\theta^n) \times P - P_\theta^{n+r} \times P\| \\ &= \inf_L \sup_{\theta\in\Theta} \|LP_\theta^n - P_\theta^{n+r}\| \\ &= \Delta(\mathscr{E}^n, \mathscr{E}^{n+r}).\end{aligned}$$

Here the infimum is taken over all transitions which map the band $L(\mathscr{E}^{n+1})$ into $L(\mathscr{E}^{n+1+r})$ or $L(\mathscr{E}^n)$ into $L(\mathscr{E}^{n+r})$, respectively.

PROOF OF THEOREM 2. It suffices to show

$$\Delta(\tilde{\mathscr{G}}^n, \tilde{\mathscr{G}}^{n+r_n}) = \Delta(\mathscr{G}^n, \mathscr{G}^{n+r_n}) + o(1). \tag{3.12}$$

Firstly the following holds:

$$\begin{aligned}\Delta(\tilde{\mathscr{G}}^n, \tilde{\mathscr{G}}^{n+r_n}) &= \inf_K \sup_{\theta\in\Theta} \left\| KN(\theta, n^{-1}\Gamma^{-1}(\theta)) - N\big(\theta, (n+r_n)^{-1}\Gamma^{-1}(\theta)\big)\right\| \\ &\le \sup_{\theta\in\Theta} \left\| N\big(\Gamma^{1/2}(\theta)\theta, n^{-1}I_k\big) - N\big(\Gamma^{1/2}(\theta)\theta, (n+r_n)^{-1}I_k\big)\right\| \\ &\le \Delta(\mathscr{G}^n, \mathscr{G}^{n+r_n}).\end{aligned} \tag{3.13}$$

(The infimum has to been taken over all kernels from $(\mathbb{R}^k, \mathscr{B}(\mathbb{R}^k))$ to $(\mathbb{R}^k, \mathscr{B}(\mathbb{R}^k))$.

To prove the other direction we use the fact that asymptotically it suffices to calculate the deficiencies for local subexperiments of $\mathscr{G}^n$ and $\mathscr{G}^{n+r_n}$:

$$\begin{aligned}&\Delta(\mathscr{G}^n, \mathscr{G}^{n+r_n}) \\ &\quad = \sup_{\tau\in\mathbb{R}^k} \inf_K \sup\Big\{\left\| KN(\theta, n^{-1}I_k) - N\big(\theta, (n+r_n)^{-1}I_k\big)\right\|: \\ &\qquad\qquad \|\theta - \tau\| \le n^{\alpha-1/2}\Big\} + o(1)\end{aligned} \tag{3.14}$$

for $\alpha > 0$. (3.14) can be deduced from the existence of an estimator whose probability being outside an $n^{\alpha-1/2}$-neighbourhood of the true parameter decreases exponentially in n. For a detailed proof see Theorem 1 in Le Cam (1975). Using (3.14) one gets

$$\begin{aligned}\Delta(\mathscr{G}^n, \mathscr{G}^{n+r_n}) &= \inf_K \sup\Big\{\left\| KN(\theta, n^{-1}I_k) \right. \\ &\qquad \left. - N\big(\theta, (n+r_n)^{-1}I_k\big)\right\| : \|\theta\| \le n^{\alpha-1/2}\Big\} + o(1) \\ &= \inf_K \sup\Big\{\left\| KN(\theta, n^{-1}\Gamma^{-1}(0)) \right. \\ &\qquad \left. - N\big(\theta, (n+r_n)^{-1}\Gamma^{-1}(0)\big)\right\| : \|\Gamma^{1/2}(0)\theta\| \le n^{\alpha-1/2}\Big\} + o(1) \\ &= \inf_K \sup\Big\{\left\| KN(\theta, n^{-1}\Gamma^{-1}(\theta)) \right. \\ &\qquad \left. - N\big(\theta, (n+r_n)^{-1}\Gamma^{-1}(\theta)\big)\right\| : \|\Gamma^{1/2}(0)\theta\| \le n^{\alpha-1/2}\Big\} + o(1) \\ &\le \Delta(\tilde{\mathscr{G}}^n, \tilde{\mathscr{G}}^{n+r_n}).\end{aligned} \tag{3.15}$$

Proof of Theorem 3. Without loss of generality we assume $r_n \equiv 1$. Using Lemma 2 and the triangle inequality one gets for a constant $a > 1$

$$\begin{aligned}\Delta(\mathscr{E}^n, \mathscr{E}^{[an]}) &\leq \Delta(\mathscr{E}^n, \mathscr{E}^{n+1}) + \cdots + \Delta(\mathscr{E}^{[an]-1}, \mathscr{E}^{[an]}) \\ &\leq ([an] - n)\Delta(\mathscr{E}^n, \mathscr{E}^{n+1}).\end{aligned} \tag{3.16}$$

This gives

$$n\Delta(\mathscr{E}^n, \mathscr{E}^{n+1}) \geq (a-1)^{-1}\Delta(\mathscr{E}^n, \mathscr{E}^{[an]}). \tag{3.17}$$

According to the assumptions there exists a sequence $(d_n)_{n \geq 1}$ in $\mathbb{R}^+$ such that

$$d_n \to \infty \quad \text{for } n \to \infty \text{ and } d_n/\sqrt{n} \text{ is monotone decreasing}, \tag{3.18}$$

$$\Delta\left(\mathscr{E}^n_{n, d_n}, \mathscr{G}^n_{n, d_n, \Sigma}\right) \to 0 \quad \text{as } n \to \infty. \tag{3.19}$$

Setting $m := [an]$ and $c_n := \sqrt{n/m}\, d_m$ one gets from (3.19):

$$\Delta\left(\mathscr{E}^m_{n, c_n}, \mathscr{G}^m_{n, c_n, \Sigma}\right) = \Delta\left(\mathscr{E}^m_{m, d_m}, \mathscr{G}^m_{m, d_m, \Sigma}\right) \to 0 \quad \text{for } n \to \infty. \tag{3.20}$$

Further $c_n/\sqrt{n} = d_m/\sqrt{m} \leq d_n/\sqrt{n}$ implies

$$\Delta\left(\mathscr{E}^n_{n, c_n}, \mathscr{G}^n_{n, c_n, \Sigma}\right) \to 0 \quad \text{for } n \to \infty. \tag{3.21}$$

With $\Theta_n := \{\theta \in \Theta: \|\theta - \theta_0\| \leq c_n/\sqrt{n}\}$ (3.20) and (3.21) entail

$$\begin{aligned}\Delta(\mathscr{E}^n, \mathscr{E}^m) &\geq \Delta\left(\mathscr{E}^n_{n, c_n}, \mathscr{E}^m_{n, c_n}\right) \\ &= \Delta\left(\mathscr{G}^n_{n, c_n, \Sigma}, \mathscr{G}^m_{n, c_n, \Sigma}\right) + o(1) \\ &= \Delta\big((N(\theta, \Sigma)^n: \theta \in \Theta_n), (N(\theta, \Sigma)^m: \theta \in \Theta_n)\big) + o(1) \\ &= \Delta\big((N(\theta, \Sigma): \|\theta\| \leq c_n), (N(\theta, (n/m)\Sigma): \|\theta\| \leq c_n)\big) + o(1) \\ &= \Delta\big((N(\theta, \Sigma): \|\theta\| \leq c_n), (N(\theta, a^{-1}\Sigma): \|\theta\| \leq c_n)\big) + o(1).\end{aligned} \tag{3.22}$$

Using the Bayes criterion (1.3) this gives

$$= \Delta\big((N(\theta, \Sigma): \theta \in \mathbb{R}^k), (N(\theta, a^{-1}\Sigma): \theta \in \mathbb{R}^k)\big) + o(1).$$

According to Torgersen (1972) [see also (1.6)] this is asymptotically equal to

$$\begin{aligned}&= \|N(0, \Sigma) - N(0, a^{-1}\Sigma)\| + o(1) \\ &= \|N(0, I_k) - N(0, aI_k)\| + o(1).\end{aligned}$$

Putting (3.17) and (3.22) together one gets

$$n\Delta(\mathscr{E}^n, \mathscr{E}^{n+1}) \geq (a-1)^{-1}\|N(0, I_k) - N(0, aI_k)\| + o(1). \tag{3.23}$$

For $a \to 1$ the last term converges to $\lim_{n \to \infty} n\Delta(\mathscr{G}^n, \mathscr{G}^{n+1})$ as can be seen by using (1.6). This completes the proof.

Proof of Theorem 4. According to (1.7), (1.8) it suffices to prove:

$$\lim_{n \to \infty} n\Delta\left(\mathscr{E}^n_B, \mathscr{E}^{n+1}_B\right) \geq 2\sqrt{2/\pi e}\,. \tag{3.24}$$

Put $\theta = \ln(p/(1-p))$, $h = \sqrt{n}\,\theta$, $c(\theta) = \ln(2(1-p))$. The Bernoulli distribution has the following density with respect to the uniform distribution on $\{0, 1\}$:

$$\exp(\theta x + c(\theta)).$$

One can write

$$c(\theta) - c(0) = -\frac{h}{2\sqrt{n}} - \frac{h^2}{8n} + \Delta\left(\frac{h}{\sqrt{n}}\right),$$

where $\Delta(h/\sqrt{n}) = c^{(3)}(\theta')(h/\sqrt{n})^3/6$ for some θ' between 0 and θ. For proving his asymptotic lower bound (1.8) Helgeland (1982) considers the following Bayes decision problem: Let $0 < \alpha < \frac{1}{6}$ and $c_n = n^{\alpha}$.

Given a prior distribution having density wrt Lebesgue measure

$$\gamma_n \exp\{-n\Delta(h/\sqrt{n}) - h^2/2\kappa^2\} I_{(-c_n, c_n)}(h), \tag{3.25}$$

construct a confidence interval of length $2l = 2(1/4 + 1/\kappa^2)^{-1/2}$. The loss function is -1 or 1 according to the true parameter falls into the confidence interval or not.

We modify this decision problem slightly. Consider the prior measure λ_n having density (3.25) with respect to the measure:

$$\nu_n(A) := \#(A \cap (a_n\mathbb{Z} + b_n)),$$
$$a_n = \left(\frac{n+1}{4n} + \frac{1}{\kappa^2}\right)^{-1} \frac{1}{\sqrt{n}},$$
$$b_n = -a_n(n+1)/2.$$

Further suppose that the confidence intervals to be constructed are closed with length $2l_n = [2l/a_n]a_n$.

Arguing as in Helgeland (1982) one can show that for $m = n$ and $m = n + 1$ the a posteriori distribution function $H_{mn}(t|X^m)$ in $\mathscr{E}^m$ fulfills the following approximation ($X^m = (X_1, \ldots, X_m)$ is the vector of the first m observations):

$$E_{P^m\lambda_n}\left(\sup_t |F_{mn}(t|X^m) - H_{mn}(t|X^m)|\right) = o(1/n), \tag{3.26}$$

where

$$F_{mn}(t|X^m) = \int_{-\infty}^{t} \frac{1}{\sigma_{mn}} \phi\left(\frac{x - \mu_{mn}}{\sigma_{mn}}\right) \nu_n(dx)\, a'_{mn},$$
$$\sigma_{mn}^2 = \left(\frac{m}{4n} + \frac{1}{\kappa^2}\right)^{-1},$$
$$\mu_{mn} = \sigma_{mn}^2 \frac{1}{\sqrt{n}} \sum_{i=1}^{m} (x_i - \tfrac{1}{2}),$$
$$a'_{mn} = \left(\int_{-\infty}^{+\infty} \frac{1}{\sigma_{mn}} \phi\left(\frac{x - \mu_{mn}}{\sigma_{mn}}\right) \nu_n(dx)\right)^{-1}.$$

The main idea of the proof is that the constants a_n, b_n are chosen so that μ_{n+1n} is an element of the support of the prior measure λ_n (and ν_n) and the point μ_{nn} lies close to the midpoint of two neighbouring points of the support of λ_n (see (3.32)). (3.26) can be used to construct confidence intervals C_n, C_{n+1} having asymptotically minimal Bayes risk. For $m = n + 1$ choose the interval $\mu_{n+1n} \pm l_n$ and for $m = n$ choose $\tilde{\mu}_{nn} \pm l_n$, where $\tilde{\mu}_{nn}$ is the point in $a_n\mathbb{Z} + b_n$ closest to μ_{nn}. According to (3.26) the corresponding Bayes risks differ from the minimal

Bayes risk by $o(1/n)$. Using (1.3) one gets an asymptotic lower bound for the deficiency:

$$\Delta(\mathscr{E}_B^n, \mathscr{E}_B^{n+1}) \geq \rho_{nn} - \rho_{n+1n} + o(1/n), \tag{3.27}$$

where for $m = n$, $m = n + 1$, resp.

$$\rho_{mn} = E_{P^m\lambda_n}\left(1 - 2\int_{C_m} \frac{1}{\sigma_{mn}}\phi\left(\frac{x - \mu_{mn}}{\sigma_{mn}}\right)\nu_n(dx)a'_{mn}\right).$$

Put $A_n := \{x^n\colon |\Sigma_{1\leq i\leq n}(x_i - 1/2)| \leq n^{2\alpha+1/2}\}$. Then $E_\theta X_i - \frac{1}{2} = O(n^{\alpha-1/2})$ uniformly for θ in the support of λ_n and therefore

$$P^n\lambda_n(A_n^c) = o(1/n). \tag{3.28}$$

Uniformly for $X^{n+1} \in A_n \times \{0,1\}$ one gets $\sigma_{n+1n} - \sigma_{nn} = O(1/n)$, $\mu_{n+1n} - \mu_{nn} = O(n^{-1/2})$. This implies

$$a'_{nn}/a'_{n+1n} = 1 + O(n^{-3/2}), \tag{3.29}$$

$$a'_{nn}/a_n = 1 + O(n^{-1}). \tag{3.30}$$

Using (3.28)–(3.30) one gets

$$\rho_{nn} - \rho_{n+1n} = I_{1n} + I_{2n} + o(1/n), \tag{3.31}$$

where

$$I_{1n} = 2\int_{-l_n}^{l_n} \phi\left(-\frac{1}{\sigma_{nn}}\phi\left(\frac{x}{\sigma_{nn}}\right) + \frac{1}{\sigma_{n+1n}}\phi\left(\frac{x}{\sigma_{n+1n}}\right)\right)\tilde{\nu}_n(dx),$$

$$I_{2n} = 2E_{P^n\lambda_n}\int_{-l_n}^{l_n} \frac{1}{\sigma_{nn}}\left(\phi\left(\frac{x}{\sigma_{nn}}\right) - \phi\left(\frac{x - \mu_{nn} + \tilde{\mu}_{nn}}{\sigma_{nn}}\right)\right)\tilde{\nu}_n(dx).$$

$\tilde{\nu}_n$ is the following normalised counting measure:

$$\tilde{\nu}_n(A) := a_n\#(A \cap a_n\mathbb{Z}).$$

As in Helgeland (1982) one shows that

$$I_{1n} = \left(1 + \frac{4}{\kappa^2}\right)^{-1}\frac{1}{n}\sqrt{\frac{2}{\pi e}} + o\left(\frac{1}{n}\right). \tag{3.32}$$

Further, for n large enough, the following holds in A_n with $\xi = 1$ or $\xi = -1$:

$$\begin{aligned}\mu_{nn} - \tilde{\mu}_{nn} &= \sigma_{nn}^2 n^{-1/2}\sum_{i=1}^n \left(X_i - \tfrac{1}{2}\right) - \sigma_{n+1n}^2 n^{-1/2}\sum_{i=1}^n \left(X_i - \tfrac{1}{2}\right) - \xi\sigma_{n+1n}^2 n^{-1/2}/2\\ &= -\xi a_n/2 + O(n^{2\alpha-1}).\end{aligned}$$

This can be used to evaluate I_{2n}. Put $\delta_n := (\tilde{\mu}_{nn} - \mu_{nn})/\sigma_{nn}$ and $J_n := (-l_n/\sigma_{nn}, l_n/\sigma_{nn})$ and $\nu'_n(\cdot) := \tilde{\nu}_n((\cdot)\sigma_{nn})$. Then

$$\begin{aligned}I_{2n} &= 2E_{P^n\lambda_n}\int_{J_n} \phi(x) - \phi(x + \delta_n)\nu'_n(dx)\\ &= 2E_{Q_n}\int_{J_n} \phi(x)\{x\delta_n + (1 - x^2)\delta_n^2/2\}\\ &\quad -\phi(x + \delta_n(x))\{3(x + \delta_n(x))\\ &\qquad - (x + \delta_n(x))^3\}\delta_n^3/6\,\nu'_n(dx) + o(1/n),\end{aligned} \tag{3.33}$$

where Q_n is the restriction of $P^n\lambda_n$ on A_n and $\delta_n(\cdot)$ is a function with $|\delta_n(\cdot)| \le |\delta_n|$.

Evaluating the integrals in (3.33) one gets

$$
\begin{aligned}
I_{2n} &= E_{Q_n}\delta_n^2 \int_{J_n} \phi(x)(1 - x^2)\, dx + o(1/n) \\
&= a_n^2/4\, \sigma_{nn}^{-2} [y\phi(y)]_{-1}^{1} + o(1/n) \qquad (3.34)\\
&= \left(1 + \frac{4}{\kappa^2}\right)^{-1} \frac{1}{n} \sqrt{\frac{2}{\pi e}} + o(1/n).
\end{aligned}
$$

Since κ^2 can be chosen arbitrarily large, (3.24) follows from (3.27), (3.31), (3.32), and (3.34).

Acknowledgment. The author wishes to thank the referees for many helpful suggestions.

REFERENCES

BIRGÉ, L. (1983). Approximation dans les espaces métriques et théorie de l'estimation. *Z. Wahrsch. verw. Gebiete* **65** 181–237.

DACUNHA-CASTELLE, D. (1978). Vitesse de convergence pour certains problemes statistiques. In *Ecole d'Eté de Probabilités de Saint Flour VII-1977.* Springer, Berlin.

HELGELAND, J. (1982). Additional observations and statistical information in the case of 1-parameter exponential distributions. *Z. Wahrsch. verw. Gebiete* **59** 77–100.

LE CAM, L. (1964). Sufficiency and approximate sufficiency. *Ann. Math. Statist.* **35** 1419–1455.

LE CAM, L. (1968). Théorie asymptotique de la décision statistique. Les Presses de l'Université de Montréal.

LE CAM, L. (1974). On the information contained in additional observations. *Ann. Statist.* **2** 630–649.

LE CAM, L. (1975). On local and global properties in the theory of asymptotic normality of experiments. In *Stochastic Processes and Related Topics* (M. L. Puri, ed.). Academic, London.

LE CAM, L. (1985). Sur l'approximation de familles de mesures par des familles Gaussiennes. *Ann. Inst. Henri Poincaré* **21** 225–287.

MAMMEN, E. (1983). Die statistische Information zusätzlicher Beobachtungen. Inaugural-Dissertation, Heidelberg.

MÜLLER, D. W. (1980 / 81). The increase of risk due to inaccurate models. *Symposia Mathematica* **25** 73–84.

SWENSEN, A. R. (1980). Deficiencies in linear normal experiments. *Ann. Statist.* **8** 1142–1155.

TORGERSEN, E. N. (1972). Comparison of translation experiments. *Ann. Math. Statist.* **43** 1383–1399.

TORGERSEN, E. N. (1981). Measures of information based on comparison with total information and with total ignorance. *Ann. Statist.* **9** 638–657.

UNIVERSITÄT HEIDELBERG
IM NEUENHEIMER FELD 294
6900 HEIDELBERG 1
FEDERAL REPUBLIC OF GERMANY