Z. Wahrscheinlichkeitstheorie verw. Gebiete
59, 77–100 (1982)

Zeitschrift für
Wahrscheinlichkeitstheorie
und verwandte Gebiete

Additional Observations and Statistical Information in the Case of 1-Parameter Exponential Distributions

Jon Helgeland

Norwegian Computing Center, Oslo, Norway

Summary. We study the increase in statistical information obtained by adding independent observations, as measured by the LeCam-deficiency δ. The main object of our study is the case where the observations follow a 1-parameter exponential law. We show that when the parameter set is a compact, non-degenerate interval and r is a fixed integer, then

$$\sqrt{2/\pi e} \leqq \liminf \frac{n}{r} \delta(\mathscr{E}^n, \mathscr{E}^{n+r}) \leqq \limsup \frac{n}{r} \delta(\mathscr{E}^n, \mathscr{E}^{n+r}) \leqq 2\sqrt{2/\pi e}$$

where $\delta(\mathscr{E}^n, \mathscr{E}^{n+r})$ is the deficiency of $\mathscr{E}^n$ with respect to $\mathscr{E}^{n+r}$, and $\mathscr{E}^n$ is the experiment consisting in taking n independent observations from $\mathscr{E}$.

1. Introduction

We define an *experiment* as a pair $((\mathscr{X}, \mathscr{A}), (P_\theta: \theta \in \Theta))$ where $(\mathscr{X}, \mathscr{A})$ is a measurable space, $\{P_\theta\}$ is a family of probability measures over $(\mathscr{X}, \mathscr{A})$ indexed by some set Θ, the parameter space.

In order to compare experiments with respect to "content of statistical information" we use the concept of deficiency introduced by LeCam (1964):

Let $\mathscr{E} = (\mathscr{X}, \mathscr{A}, P_\theta: \theta \in \Theta)$, $\mathscr{F} = (\mathscr{Y}, \mathscr{B}, Q_\theta: \theta \in \Theta)$ be experiments with a common parameter space Θ, and let $\varepsilon: \Theta \to [0, \infty\rangle$. We say that $\mathscr{E}$ is *ε-deficient* relative to $\mathscr{F}$ if for any decision space $(T, \mathscr{S})$ where $\mathscr{S}$ is finite, any bounded loss function L: $\Theta \times T \to \mathbb{R}$ and any decision rule σ (to $(T, \mathscr{S})$) in $\mathscr{F}$, there exists a decision rule ρ in $\mathscr{E}$ (to $(T, \mathscr{S})$) so that

$$P_\theta \rho L_\theta \leqq Q_\theta \sigma L_\theta + \varepsilon_\theta \|L_\theta\|, \qquad \forall \theta \tag{1}$$

where $\|L_\theta\| = \sup_t |L_\theta(t)|$.

In (1) we may replace $\|L_\theta\|$ by $\|L\|$ and we may confine ourselves to non-negative L if we replace "ε_θ" in (1) by "$\frac{1}{2}\varepsilon_\theta$". If $\mathscr{E}$ is 0-deficient rel. $\mathscr{F}$, we say that $\mathscr{E}$ is *more informative* than $\mathscr{F}$ (written $\mathscr{E} \geqq \mathscr{F}$) and if both $\mathscr{E} \geqq \mathscr{F}$ and

0044-3719/82/0059/0077/$04.80

$\mathscr{F} \geqq \mathscr{E}$, $\mathscr{E}$ and $\mathscr{F}$ are said to be *equivalent* (written $\mathscr{E} \sim \mathscr{F}$). The infimum over all constants $\varepsilon > 0$ such that $\mathscr{E}$ is ε-deficient relative to $\mathscr{F}$ is written $\delta(\mathscr{E}, \mathscr{F})$ and is called the *deficiency* of $\mathscr{E}$ relative to $\mathscr{F}$. The *Δ-distance* between $\mathscr{E}$ and $\mathscr{F}$ is defined by $\Delta(\mathscr{E}, \mathscr{F}) = \delta(\mathscr{E}, \mathscr{F}) \vee \delta(\mathscr{F}, \mathscr{E})$. The class of experiments which are equivalent to an experiment $\mathscr{E}$, is called the *experiment type* of $\mathscr{E}$. The class of all experiment types (strictly the class of suitably chosen representatives) form a set $\mathbb{E}$, and $(\mathbb{E}, \Delta)$ becomes a complete metric space (LeCam (1974a)).

If $\mathscr{F} = (\mathscr{X}, \mathscr{A}, P_\theta; \theta \in \Theta)$ and $\mathscr{E} = (\mathscr{X}, \mathscr{B}, P_\theta | \mathscr{B}; \theta \in \Theta)$ where $\mathscr{B}$ is a sub-σ-algebra of $\mathscr{A}$ and $P_\theta | \mathscr{B}$ is the restriction of P_θ to $\mathscr{B}$; then obviously $\mathscr{E} \leqq \mathscr{F}$. One measure of the loss of information when observing only $\mathscr{B}$-measurable events is $\delta(\mathscr{E}, \mathscr{F})$, another is the *insufficiency* (LeCam (1974b)). It is defined by

$$\eta(\mathscr{E}, \mathscr{F}) = \inf_{\{P_\theta^*\}} \sup_\theta \|P_\theta^* - P_\theta\|$$

where the infimum is taken over all families $\{P_\theta^*\}_{\theta \in \Theta}$ such that $P_\theta^* | \mathscr{B} = P_\theta | \mathscr{B}$ and $\mathscr{B}$ is sufficient for $\{P_\theta^*\}$; $\| \cdot \|$ is the total variation norm.

The concept of deficiency has several interpretations, which are natural ways of formally defining loss of information. We mention here the following theorems (LeCam (1974)):

(i) *Let* $\mathscr{E} = (\mathscr{X}, \mathscr{A}, P_\theta: \theta \in \Theta)$, $\mathscr{F} = (\mathscr{Y}, \mathscr{B}, Q_\theta: \theta \in \Theta)$, $\varepsilon: \Theta \to [0, \infty\rangle$.

Assume $\mathscr{E}$ is dominated. Then $\mathscr{E}$ is ε-deficient relative to $\mathscr{F}$ if and only if to every decision space $(T, \mathscr{S})$ which is a Borel-subset of a Polish space with the restricted Borel-σ-algebra and to every decision rule σ in $\mathscr{F}$, there is a decision rule ρ in $\mathscr{E}$ such that $\|P_\theta \rho - Q_\theta \sigma\| \leqq \varepsilon_\theta$, $\forall \theta$.

(ii) *The Markov kernel criterion:*

Let $\mathscr{E}, \mathscr{F}$ be as above. Assume that $\mathscr{Y}$ is a Borel-subset of a Polish space and $\mathscr{B}$ is the restricted Borel-σ-algebra. Then $\mathscr{E}$ is ε-deficient relative to $\mathscr{F}$ if and only if there exists a Markov kernel M: $\mathscr{B} \times \mathscr{X} \to [0, 1]$ *such that* $\|P_\theta M - Q_\theta\| \leqq \varepsilon_\theta$, $\forall \theta$. (A *Polish space* is a complete separable metric space equipped with its Borel-σ-algebra. A *Markov kernel* is a mapping M: $\mathscr{B} \times \mathscr{X} \to [0, 1]$ such that

(a) $M(\cdot | x)$ is a probability measure for every $x \in \mathscr{X}$

(b) $M(B | \cdot)$ is measurable for every $B \in \mathscr{B}$.)

Assume $\mathscr{E}, \mathscr{F}, \varepsilon, T, \mathscr{S}$ are as in (i), and further that $P_{(\cdot)}$, $Q_{(\cdot)}$ are Markov kernels from $(\Theta, \mathscr{V})$ where $\mathscr{V}$ is some σ-algebra over Θ. Let L be a bounded and $\mathscr{V} \times \mathscr{S}$-measurable loss function. Then both $\theta \curvearrowright P_\theta \rho L_\theta$ and $\theta \curvearrowright Q_\theta \sigma L_\theta$ are bounded and $\mathscr{V}$-measurable for all decision rules ρ and σ, and we may define Bayes risk by

$$b_\lambda^{\mathscr{E}} = \inf_\rho \lambda P \rho L$$

where λ is a probability measure over $(\Theta, \mathscr{V})$. For all constants $\varepsilon > \delta(\mathscr{E}, \mathscr{F})$ we have, for all σ in $\mathscr{F}$:

For some ρ in $\mathscr{E}$

$$P_\theta \rho L_\theta \leqq Q_\theta \sigma L_\theta + \varepsilon \|L\|, \qquad \forall \theta$$
$$\Rightarrow b_\lambda^{\mathscr{E}} \leqq \lambda Q \sigma L + \varepsilon \|L\|.$$

Then

$$\delta(\mathscr{E}, \mathscr{F}) \geqq \sup_\sigma \frac{1}{\|L\|}(b_\lambda^{\mathscr{E}} - \lambda Q \sigma L) = \frac{1}{\|L\|}(b_\lambda^{\mathscr{E}} - b_\lambda^{\mathscr{F}}). \tag{2}$$

There is a connection between CE-sufficiency ("conditional expectation"-sufficiency, i.e. sufficiency in the sense of Halmos and Savage) and deficiency, due to Bahadur:

If

$$\mathscr{E} = (\mathscr{X}, \mathscr{B}, P_\theta | \mathscr{B}; \theta \in \Theta) \text{ and } \mathscr{F} = (\mathscr{X}, \mathscr{A}, P_\theta; \theta \in \Theta)$$

where $\mathscr{B}$ *is a sub-σ-algebra of* $\mathscr{A}$, *then*

(i) $\mathscr{B}$ is *CE-sufficient for* $\mathscr{F}$

implies

(ii) $\delta(\mathscr{E}, \mathscr{F}) = 0$.

If $\mathscr{E}$ *is dominated, then* (ii) $\Rightarrow$ (i).

In the following we will consider experiments of the form

$$\mathscr{E}^n = (\mathscr{X}^n, \mathscr{A}^n, P_\theta^n; \theta \in \Theta)$$

where $\mathscr{E} = (\mathscr{X}, \mathscr{A}, P_\theta; \theta \in \Theta)$ i.e. $\mathscr{E}^n$ is n independent replications of $\mathscr{E}$. It is obvious that $\mathscr{E}^n \leqq \mathscr{E}^m$ when $n \leqq m$, and a natural question arises: How much more informative than $\mathscr{E}^n$ is $\mathscr{E}^m$ – what is $\delta(\mathscr{E}^n, \mathscr{E}^m)$? Aside from the theoretical interest, knowing $\delta(\mathscr{E}^n, \mathscr{E}^m)$ may possibly be useful in the planning of replicated experiments when the exact nature of the decision problem is not determined on beforehand. Let $K(\mathscr{E})$ denote the "cost" of performing $\mathscr{E}$ and L some loss function. Then the "total risk function" under the decision rule ρ is $R_{\mathscr{E}}(\theta) = P_\theta \rho L_\theta + K(\mathscr{E})$. Suppose that $\|L\| \leqq 1$. We then prefer $\mathscr{E}^n$ to $\mathscr{E}^{n+1}$ when $\delta(\mathscr{E}^n, \mathscr{E}^{n+1}) \leqq K(\mathscr{E}^{n+1}) - K(\mathscr{E}^n)$, and $\mathscr{E}^{n+1}$ to $\mathscr{E}^n$ when $\delta(\mathscr{E}^n, \mathscr{E}^{n+1}) \geqq K(\mathscr{E}^{n+1}) - K(\mathscr{E}^n)$. That $\mathscr{E}^n$ is better than $\mathscr{E}^m$ in the above sense means that: To any "total risk function" $R_{\mathscr{E}^m}$ there exists a $R_{\mathscr{E}^n}$ (which is the risk for the same decision problem) such that

$$R_{\mathscr{E}^n} \leqq R_{\mathscr{E}^m}.$$

Example 1. Let $\mathscr{E}$ consist in observing $X \sim \mathrm{N}(\theta, \sigma)$ where σ is known. Then (Torgersen (1972))

$$\delta(\mathscr{E}^n, \mathscr{E}^{n+1}) \sim \frac{1}{n}\sqrt{2/\pi e}.$$

If we let $K(\mathscr{E}^n) = k_0 + n k_1$, then $n_0 = \sqrt{2/\pi e}/k_1$ is the optimal sample size in the above sense.

Intuitively one may expect that $\mathscr{E}^n$ gets very informative as $n \to \infty$, and that one additional observation gets more and more unimportant. When Θ is finite, then $\Delta(\mathscr{E}^n, \mathscr{M}_a) \to 0$, where $\mathscr{M}_a$ is the experiment where θ itself is observed without uncertainty. In fact,

$$\sqrt[n]{\delta(\mathscr{E}^n, \mathscr{M}_a)} \to c(\mathscr{E})$$

where

$$c(\mathscr{E}) = \max_{\theta_1 \neq \theta_2} \inf_{0<t<1} \int dP_{\theta_1}^{1-t} dP_{\theta_2}^t.$$

(If $\theta_1 \neq \theta_2 \Rightarrow P_{\theta_1} \neq P_{\theta_2}$, then $c(\mathscr{E}) < 1$.) If Θ is countably infinite, then

$$\mathscr{E}^n \to \mathscr{M}_a \Rightarrow \delta(\mathscr{E}^n, \mathscr{M}_a) \leqq c\,\rho^n$$

for some $c > 0$ and $\rho < 1$. However, we need not have convergence at all, e.g. if $\{P_\theta\}$ has a limit point for setwise convergence then

$$\delta(\mathscr{E}^n, \mathscr{M}_a) \equiv 2.$$

If Θ is uncountable and $\mathscr{E}$ is dominated, then $\delta(\mathscr{E}, \mathscr{M}_a) = 2$ always. These results are from Torgersen (1976).

Now let $\mathscr{E}$ be an experiment with uncountable Θ such that $\theta \curvearrowright P_\theta$ is $(1-1)$. Since the restriction $\mathscr{E}^n | F$ of $\mathscr{E}^n$ to finite subsets $F \subset \Theta$ must converge to $\mathscr{M}_a | F$, $\mathscr{M}_a$ is the only possible Δ-limit for $\{\mathscr{E}^n\}$. If now $\mathscr{E}$ is dominated, $\Delta(\mathscr{E}^n, \mathscr{E}^m) \underset{n,m\to\infty}{\nrightarrow} 0$ since $(\mathbb{E}, \Delta)$ is complete. This implies that $\sum_{k=0}^{\infty} \delta(\mathscr{E}^{n+k}, \mathscr{E}^{n+k+1}) \underset{n\to\infty}{\nrightarrow} 0$ and furthermore that

$$\liminf(n^{-\alpha}/\delta(\mathscr{E}^n, \mathscr{E}^{n+1})) = 0$$

for all $\alpha > 1$.

The insufficiency $\eta(\mathscr{E}^n, \mathscr{E}^{n+1})$ may be used to study $\delta(\mathscr{E}^n, \mathscr{E}^{n+1})$ since $\eta(\cdot) \geqq \delta(\cdot)$, but the approximation may be poor: If $\mathscr{E}$ consists in observing $X \sim \mathrm{N}(\theta, 1)$ (Example 1) then

$$\eta(\mathscr{E}^n, \mathscr{E}^{n+1}) \geqq \frac{1}{2\pi} e^{-\frac{1}{4n}} \frac{1}{\sqrt{n}}$$
$$\Rightarrow \delta(\mathscr{E}^n, \mathscr{E}^{n+1}) = o(\eta(\mathscr{E}^n, \mathscr{E}^{n+1})).$$

LeCam (1974b) has shown this, and also the following result:

For all $n, k \geqq 0$

$$\eta(\mathscr{E}^n, \mathscr{E}^{n+k}) \geqq \sqrt{2D_n} \sqrt{\frac{k}{n}}$$

where D_n *is a dimensionality constant for* Θ.

D_n is determined in the following way: The Hellinger distance $H(H^2(P,Q) = \int (\sqrt{dP} - \sqrt{dQ})^2$ for probability measures $P, Q)$ induces a metric on Θ: $h(\theta, \theta') = H(P_\theta, P_{\theta'})$. Put $a_\nu = 2^{-(\nu+10)/2}$, $b_\nu = 2^{-\nu/2}$; $\nu = 1, 2, \ldots$.

For finite $S \subset \Theta$, $\operatorname{diam} S \leqq b_{\nu-1}$, let $\{A_i\}$ be a finite covering of S by sets of diameter not exceeding a_ν. Say that indices i, j are "distant" if

$$\sup\{h(\theta, \theta') : \theta \in A_i, \theta' \in A_j\} > b_\nu.$$

For each i, let C_i' be the number of indices distant from i, and let $C_S' = \sup_i C_i'$. Choose $\{A_i\}$ such that C' is minimal, and put $c(v) = \sup_S C_S'$ where the supremum is taken over finite $S \subset \Theta$ such that $\operatorname{diam} S \leqq b_{v-1}$. Let $K_n = 1 \vee \sup\{c(v): 2^v \leqq n\}$ and put $D_n = 16 \log 6 K_n$.

LeCam also gives an example of an $\mathscr{E}$ such that $\delta(\mathscr{E}^n, \mathscr{E}^{n+1}) \nrightarrow 0$:

Example 2. Let $(\mathscr{X}, \mathscr{A}, \lambda)$ be $[0,1]$ equipped with Lebesgue-measure λ, and $\Theta = \{0, 1, 2, \ldots\}$.

Let $I_{k\theta}$ be the indicator function of the interval $[2^{-\theta}(2^k+1), 2^{-\theta}(2^k+2)]$. Assume that

$$\frac{dP_\theta}{d\lambda}(x) = \sum_{k=0}^{\theta-1} 2 I_{k\theta}(x) \quad \text{for } \theta \geqq 1,$$
$$P_0 = \lambda.$$

Let $\mathscr{E} = (\mathscr{X}, \mathscr{A}, P_\theta; \theta \in \Theta)$. Then $\delta(\mathscr{E}^n, \mathscr{E}^{n+1}) \geqq 1$, $\forall n$. In fact, for large enough k, let $m = k^3 2^n$. Then $\lim_{m \to \infty} \delta(\mathscr{E}|_{\Theta_m}^n, \mathscr{E}|_{\Theta_m}^{n+1}) \geqq 1$ where $\Theta_m = \{1, 2, \ldots, m+1\}$ and $\mathscr{E}|_{\Theta_m}$ denotes the restriction of $\mathscr{E}$ to Θ_m.

Torgersen treats the case where $\mathscr{E}$ is a translation experiment, and mentions the following examples:

Example 1 (Continued). Let $\mathscr{E}$ consist in observation of $X \sim \mathrm{N}_k(\theta, \Sigma)$ where Σ is known and positive definite, and the unknown parameter $\theta \in \Theta = \mathbb{R}^k$. Then

$$\delta(\mathscr{E}^n, \mathscr{E}^{n+r}) \sim \frac{2k\,\Gamma_k'(k)\,r}{n}$$

where Γ_k is the cumulative distribution function of the χ_k^2-distribution.

Example 3. Let $\mathscr{E}$ consist in observation of $X \sim \mathrm{R}\langle 0, \theta]$, $\theta \in \Theta = \langle 0, \infty \rangle$. Then

$$\delta(\mathscr{E}^n, \mathscr{E}^{n+r}) \sim \frac{2}{e}\frac{r}{n}.$$

In the light of these results, it seems reasonable to guess that

$$\delta(\mathscr{E}^n, \mathscr{E}^{n+1}) = \frac{c}{n}(1 + o(1))$$

for Θ uncountable and $\mathscr{E}$ "nice". Our main result is that in the 1-parameter exponential case with Θ a nondegenerate compact interval

$$\sqrt{2/\pi e} \leqq \liminf(n\,\delta(\mathscr{E}^n, \mathscr{E}^{n+1})) \leqq \limsup(n\,\delta(\mathscr{E}^n, \mathscr{E}^{n+1})) \leqq 2\sqrt{2/\pi e}.$$

We will be referring to wellknown results about the exponential experiments, which can be found in e.g. Lehmann (1959).

2. Upper Bounds for $\delta(\mathscr{E}^n, \mathscr{E}^{n+1})$. Multinomial and General Experiments

In this section we will first consider the experiments $\mathscr{E}^n$ consisting in observation of the i.i.d. variables $B_1, \dots, B_n$, where B_i assumes the values $1, \dots, s$ with probabilities $\theta_1, \dots, \theta_s$. The parameter $\theta = (\theta_1, \dots, \theta_s) \in \Theta$ which is the standard simplex in $\mathbb{R}^s \{x \in [0,1]^s: \Sigma x_i = 1\}$. By sufficiency we get $\mathscr{E}^n \sim \mathscr{F}_n$ where $\mathscr{F}_n$ consists in observation of the multinomially distributed variable $Z_n = (X_1, \dots, X_s)$. X_j is the number of B_i, $i = 1, \dots, n$ with value j.

The Markov kernel criterion is a useful tool for finding upper limits for deficiencies. In our situation a Markov kernel may be found in the following way:

B_{n+1} takes on the value i with probability θ_i. An estimate for θ_i based on Z_n is $\hat{\theta}_i = X_i/n$. We may predict B_{n+1} by letting the predicted value $\hat{B}_{n+1}$ equal i with probability $\hat{\theta}_i$. This corresponds to using the Markov kernel

$$m(y|x) = \frac{1}{n} x_i, \quad \text{when } y_j = x_j + \delta_{ij}, \ \forall_j$$
$$= 0 \quad \text{otherwise.}$$

(Here δ_{ij} is the Kronecker-delta.)

Let P_θ be the law $\mathscr{L}_\theta(Z_n)$ of Z_n under θ and let Q_θ be $\mathscr{L}_\theta(Z_{n+1})$. Then $P_\theta M$ has the density w.r.t. counting measure

$$f_\theta(y) = \sum_x m(y|x) P_\theta(\{x\})$$
$$= \sum_{i: y_i \neq 0} \frac{y_i - 1}{n} \frac{n!}{y_1! \dots (y_i - 1)! \dots y_s!} \theta_1^{y_1} \dots \theta_i^{y_i - 1} \dots \theta_s^{y_s}.$$

If we write q_θ for the density of Q_θ we have

$$\|P_\theta M - Q_\theta\| = \sum_y |f_\theta(y) - q_\theta(y)| = \sum_{y: q_\theta(y) \neq 0} \left| \frac{f_\theta(y)}{q_\theta(y)} - 1 \right| q_\theta(y).$$

This may be written as an expectation (where $(Y_1, \dots, Y_{n+1})$ has the distribution Q_θ):

$$\|P_\theta M - Q_\theta\| = E_\theta \left| 1 - \sum_{S_\theta} \frac{Y_i(Y_i - 1)}{n(n+1)\theta_i} \right| \tag{3}$$

where the summation is over the set S_θ of $i \in \{1, \dots, n+1\}$ such that $\theta_i \neq 0$.

The r.h.s. of (3) is

$$E_\theta \left| \sum_{S_\theta} \frac{Y_i}{n+1} \left(1 - \frac{Y_i - 1}{n\theta_i}\right) \right|$$
$$\leq E_\theta \left| \sum_{S_\theta} \frac{Y_i}{n+1} \left(1 - \frac{Y_i}{(n+1)\theta_i}\right) \right| + E_\theta \left| \sum_{S_\theta} \frac{Y_i}{(n+1)\theta_i} \left(\frac{Y_i}{n+1} - \frac{Y_i - 1}{n}\right) \right|. \tag{4}$$

The last member of (4) is

$$E_\theta\left|\sum_{S_\theta}\frac{Y_i(n+1-Y_i)}{n(n+1)^2\theta_i}\right|=\sum_{S_\theta}E_\theta\frac{Y_i(n+1-Y_i)}{n(n+1)^2\theta_i}$$
$$=\sum_{S_\theta}\frac{1-\theta_i}{n+1}\leqq\frac{s-1}{n+1}.$$

The first member is

$$E_\theta\left|\sum_{S_\theta}\left(\frac{Y_i}{n+1}-\theta_i\right)\left(1-\frac{Y_i}{(n+1)\theta_i}\right)\right|\leqq\sum_{S_\theta}E_\theta\left|\frac{Y_i}{n+1}-\theta_i\right|\left|1-\frac{Y_i}{(n+1)\theta_i}\right|$$
$$\leqq\sum_{S_\theta}\left[E_\theta\left(\frac{Y_i}{n+1}-\theta_i\right)^2E_\theta\left(1-\frac{Y_i}{(n+1)\theta_i}\right)^2\right]^{\frac{1}{2}}=\sum_{S_\theta}\left[\frac{\theta_i(1-\theta_i)}{n+1}\cdot\frac{1-\theta_i}{(n+1)\theta_i}\right]^{\frac{1}{2}}$$
$$=\sum_{S_\theta}\frac{1-\theta_i}{n+1}\leqq\frac{s-1}{n+1}.$$

It follows that in the multinomial case

$$\delta(\mathscr{E}^n,\mathscr{E}^{n+1})\leqq 2\frac{s-1}{n+1}. \tag{5}$$

This must also hold for all experiments $\mathscr{E}$ where the σ-algebra has at most 2^s elements. This condition is equivalent to the σ-algebra being generated by at most s atoms (see Neveu (1965)). The indicator variables for these atoms have a s-nomial distribution, and since every function from the sample space must be a function of these indicator variables, they must be Δ-sufficient for the original experiment.

One may attempt to approximate more general experiments by multinomial ones in order to extend the deficiency result above. However, we have the following example, which shows that letting s increase introduces complications:

Example 2 (Continued). $\mathscr{E}|_{\Theta_m}$ has a sufficient σ-algebra $\mathscr{C}$ generated by the partition $B_m=\{[0,2^{-m}\rangle,\ [1\cdot 2^{-m},\ 2\cdot 2^{-m}\rangle,\ldots\}$ since $p_\theta(x)$ only depends on x through $I_{[0,2^{-m}\rangle},\ldots$. Then $\operatorname{card}(\mathscr{C})=2^{2^m}$, so that

$$\delta(\mathscr{E}^n|\Theta_m,\mathscr{E}^{n+1}|\Theta_m)\leq\delta(\mathscr{F}_m^n,\mathscr{F}_m^{n+1})$$

where $\mathscr{F}_m$ is the 2^m-nomial experiment. Since $\delta(\mathscr{E}^n|_{\Theta_m},\mathscr{E}^n|_{\Theta_{m+1}})\underset{m\to\infty}{\to}1$, we see that if $\mathscr{E}_s$ is s-nomial, then

$$\sup_s\delta(\mathscr{E}_s^n,\mathscr{E}_s^{n+1})\geqq 1.$$

The above calculations (leading to (5)) were first carried out for $s=2$, and Torgersen noted the validity in the general case.

The idea behind the Markov kernel method for the multinomial case can be applied to more general experiments.

Let $\mathscr{E}=(\mathscr{X},\mathscr{A},P_\theta;\theta\in\Theta)$ where $\{P_\theta\}$ is a homogenous family dominated by some σ-finite measure μ. Let $f_\theta=dP_\theta/d\mu$, and let $X_1,\ldots,X_n$ denote the observations from $\mathscr{E}^n$. We will now construct a Markov-kernel from $\mathscr{E}^n$ to $\mathscr{E}^{n+1}$, in the following intuitive way: We first estimate a density $\hat{f}$ for P_θ and draw $\hat{X}$ randomly according to this density. We then draw an $I \in \{1, \ldots, n+1\}$, and use $X_1, \ldots, X_{1-1}, \hat{X}, X_{I+1}, \ldots, X_n$ as a new set of observations. The last step "distributes the error among the components" of $\mathscr{E}^{n+1}$. This method is an analogue of the method for the multinomial case, except that in general we cannot use reduction by sufficiency.

Let us denote the following condition by **A**:

(i) *$\mathscr{B}$ contains all singletons $\{x\}$, $x \in \mathscr{X}$*

(ii) *There exists a nonnegative function*

$$\hat{f}\colon \mathscr{X}^{n+1} \to \mathbb{R}$$

which is simultaneously measurable and such that

$$\int_{\mathscr{X}} \hat{f}(y; x)\, d\mu(y) = 1 \qquad \textit{for all } x \in \mathscr{X}^n.$$

Define the Markov kernels

$$M_r\colon \mathscr{B}^{n+1} \times \mathscr{X}^n \to [0, 1]$$

by

$$M_r(\cdot|x) = \delta_{x_1} \times \ldots \times \delta_{x_{r-1}} \times \hat{M}(\cdot|x) \times \delta_{x_r} \times \ldots \times \delta_{x_n} \text{ where } \delta_t$$

is the one-point (Dirac) measure in $t \in \mathscr{X}$, $x = (x_1, \ldots, x_n)$ and

$$\hat{M}(A|x) = \int_A \hat{f}(y; x)\, d\mu(y), \qquad A \in \mathscr{B}.$$

We see that $M_r(\mathscr{X}^{n+1}|x) = 1$, $\forall x$ and that, for all $B \in \mathscr{B}^{n+1}$

$$M_r(B|x) = \int I_B(x_1, \ldots, x_{r-1}, y, x_r, \ldots, x_n)\, \hat{f}(y; x)\, d\mu(y)$$

which is measurable in x by the Tonelli theorem. Put

$$M = \frac{1}{n+1} \sum_{r=1}^{n+1} M_r$$

which obviously is a Markov kernel.

If $R = R_1 \times \ldots \times R_{n+1}$ is a rectangle in $\mathscr{B}^{n+1}$, we get

$$\begin{aligned} P_\theta M_r(R) &= \int_{\mathscr{X}^n} \delta_{x_1}(R_1) \ldots \Big(\int_{R_r} \hat{f}(y; x)\, d\mu(y)\Big) \ldots \delta_{x_n}(R_{n+1})\, dP_\theta^n(x) \\ &= \int_R f_\theta(y_1) \ldots \hat{f}(y_r; y_1, \ldots, y_{r-1}, y_{r+1}, \ldots, y_{n+1}) \ldots f_\theta(y_{n+1})\, d\mu^{n+1} \end{aligned}$$

by Tonelli's theorem. It follows that

$$\frac{dP_\theta^n M}{d\mu^{n+1}}(y) = \frac{1}{n+1} \sum_{r=1}^{n+1} \frac{\hat{f}(y_r; y_1, \ldots, y_{r-1}, y_{r+1}, \ldots, y_{n+1})}{f_\theta(y_r)} \prod_1^{n+1} f_\theta(y_i)$$

and that

$$\|P_\theta^n M - P_\theta^{n+1}\| = E_\theta \left| \frac{1}{n+1} \sum_1^{n+1} \frac{\hat{f}(Y_r; Y_1, \ldots, Y_{r-1}, Y_{r+1}, \ldots, Y_{n+1})}{f_\theta(Y_r)} - 1 \right|$$

where $(Y_1, \ldots, Y_{n+1})$ is distributed according to P_θ^{n+1}.

We have thus proved

Lemma 1. *If $\mathscr{E}$ is an experiment and $\hat{f}$ a function, satisfying condition* **A**, *then*

$$\delta(\mathscr{E}^n, \mathscr{E}^{n+1}) \leqq \sup_{\theta \in \Theta} E_\theta \left| \frac{1}{n+1} \sum_1^{n+1} \frac{\hat{f}(Y_r; Y_1, \ldots, Y_{r-1}, Y_{r+1}, \ldots, Y_{n+1})}{f_\theta(Y_r)} - 1 \right| \tag{6}$$

where $Y_1, Y_2, \ldots$ are i.i.d. with law P_θ.

3. Upper Bound for $\delta(\mathscr{E}^n, \mathscr{E}^{n+1})$ when $\{P_\theta\}$ is a 1-Parameter Exponential Family

Let $\mathscr{E} = ((\mathscr{X}, \mathscr{A}), (P_\theta: \theta \in \Theta))$ where $\Theta \subset \mathbb{R}$ and

$$\frac{dP_\theta}{d\mu} = A(\theta)\, e^{\theta T} h \tag{7}$$

where μ is some σ-finite measure on $(\mathscr{X}, \mathscr{A})$, T and $h \geqq 0$ are random variables and $A: \Theta \to \mathbb{R}$. The set of θ's such that (7) defines a probability measure for a suitable A, is the natural parameter space of $\{P_\theta\}$, and must be an interval I. In the interior I^0 of I, the function A is analytic. For all θ, $A(\theta) > 0$, and we can without loss of generality assume $0 \in I$ and write

$$\frac{dP_\theta}{dP_0} = e^{c(\theta) + \theta T}, \qquad \theta \in \Theta.$$

We can now formulate the following result:

Proposition 1. *Let $\mathscr{E} = ((\mathscr{X}, \mathscr{A}), (P_\theta: \theta \in \Theta))$ where*

$$\frac{dP_\theta}{dP_0} = e^{c(\theta) + \theta T}, \qquad \theta \in \Theta \subset \mathbb{R}.$$

Let Θ be a bounded set, and assume that an endpoint θ_0 of the natural parameter set is a limit point of Θ only if c has continuous one-sided derivatives up to and including 4th order in θ_0, and that $c''(\theta_0) \neq 0$. Then

$$\limsup (n\, \delta(\mathscr{E}^n, \mathscr{E}^{n+1})) \leqq 2\sqrt{2/\pi e}.$$

Since δ is a pseudometric and thus obeys the triangle inequality, we get the following trivial

Corollary 1. *Under the conditions of Proposition* 1 *and if r is a fixed integer, then*

$$\limsup \frac{n}{r} \delta(\mathscr{E}^n, \mathscr{E}^{n+r}) \leqq 2\sqrt{2/\pi e}.$$

Examples. The conditions above are fulfilled when $\mathscr{E}$ consists in observation of:

(i) $X \sim \text{bin}(1, p)$, $p \in [p_0, p_1]$ where $0 < p_0 \leqq p_1 < 1$.

(ii) $X \sim \text{Po}(\lambda)$, $\lambda \in \Lambda$ where Λ is bounded away from 0 and ∞.

(iii) $X \sim \text{N}(\xi, \sigma)$, with σ known, $\xi \in \Theta$ which is bounded.

The exact deficiency is (Torgersen (1972))

$$\delta(\mathscr{E}^n, \mathscr{E}^{n+1}) \sim \sqrt{2/\pi e}/n \qquad (\simeq 0.48/n)$$

and this holds even for unbounded Θ. It is seen that our method gives a bound that is 2 times too large, but with correct rate, and we have to assume an unnecessary boundedness condition for Θ.

Proof of the Proposition. We use Lemma 1 in the proof. The estimated density $\hat{f}(y; x)$ is obtained in the natural way from the maximum likelihood estimate for θ. To facilitate the use of maximum likelihood estimation we first reparametrize the experiment. The expression (6) is then simplified by a Taylor expansion and evaluated.

We may assume that Θ is a compact interval. Furthermore, T is sufficient for $\mathscr{E}$, so if $\mathscr{F}$ consists in observation of T, then $\delta(\mathscr{F}^n, \mathscr{F}^{n+1}) = \delta(\mathscr{E}^n, \mathscr{E}^{n+1})$. We can accordingly assume that $(\mathscr{X}, \mathscr{A}) = (\mathbb{R}, \mathscr{B})$ and put $f_\theta(t) = (dP_\theta/dP_0)(t) = \exp(c(\theta) + \theta t)$, $\theta \in \Theta$. For $\theta \in I^0$ we have $E_\theta T = -c'(\theta)$, $\text{var}_\theta T = -c''(\theta)$. If $c''(\theta) = 0$ for some θ, then all P_θ must be concentrated in 0. In that case $\mathscr{E} \sim \mathscr{M}_i$ (the totally non-informative experiment) and obviously $\mathscr{E}^n \sim \mathscr{E}^{n+1}$. Assume therefore that $c''(\theta) < 0$ for $\theta \in I^0$. If $I^0 = \emptyset$, then Θ is just one point, so that $\mathscr{E}^n \sim \mathscr{E}^{n+1}$, so we may assume that $I^0 \neq \emptyset$. It is convenient to reparametrize the experiment as follows: Define $\xi\colon I^0 \to \mathbb{R}$ by $\xi(\theta) = -c'(\theta) = E_\theta T$. Then ξ is a diffeomorphism from I^0 onto its image J^0, and can be extended to an open interval $I' \supset \Theta$ if Θ contains an endpoint θ_0 of I as indicated in the proposition. Since the deficiency between experiments stays unchanged under $(1-1)$-transformations of the parameter set, we can view $\mathscr{E}$ as an experiment over N where N is the image of Θ under ξ and thus a compact interval. Put $\tau = \xi^{-1}$, $\omega = c \circ \xi^{-1}$, defined on an open interval J' such that $N \subset J'$. We can thus assume that $\mathscr{E}$ is given by the densities

$$f_\xi(t) = \frac{dP_\xi(t)}{dP_{\xi_0}(t)} = e^{\omega(\xi) + \tau(\xi)t}, \qquad \xi \in N.$$

For $\xi \in J^0$,

$$E_\xi T = \xi = -\frac{\omega'(\xi)}{\tau'(\xi)}$$

and

$$\text{var}_\xi T = -c''(\tau(\xi)) = \frac{(-c'(\tau(\xi)))'}{\tau'(\xi)} = \frac{1}{\tau'(\xi)}.$$

ω and τ are analytic in J^0, and if $\xi_0=\xi(\theta_0)$ is an endpoint of $J=\xi(I)$, then, since $\xi^{(3)}$ is continuous in θ_0 and $\xi'(\theta_0)\neq 0$, τ and ω must have continuous 3-order derivatives in ξ_0. If $c^{(4)}$ is continuous in θ_0, then $A=\exp\circ c$ must be too, but for $\theta\in I^0$, $A^{(4)}(\theta)=\int T^4 e^{\theta T}dP_0=A(\theta)E_\theta T^4$, so that $E_\theta T^4$ is bounded near θ_0. Fatou's lemma then gives

$$E_{\theta_0}T^4\leqq\liminf_{\theta\to\theta_0}E_\theta T^4<\infty$$

so that $E_\theta|T|^r$ is bounded when $\theta\to\theta_0$ for $r\leqq 4$. Since $\theta\curvearrowright|T|^r e^{\theta T}$ is convex in θ, we have for θ between θ_0 and θ_1, $\theta_1\in I^0$,

$$|T|^r e^{\theta T}\leqq|T|^r e^{\theta_0 T}\vee|T|^r e^{\theta_1 T}.$$

It follows from Lebesgue's dominated convergence theorem that

$$\int|T|^r e^{\theta T}dP_0\underset{\theta\to\theta_0}{\to}\int|T|^r e^{\theta_0 T}dP_0$$

which entails that $E_\theta|T|^r$ is continuous in θ_0 for $r\leqq 4$.

We may assume that $\mathscr{E}^n$ consists in observing the first n from the i.i.d. sequence

$T_1, T_2, \ldots$

where T_i is distributed according to P_θ. This common probability space simplifies the notation in the proof.

A reasonable estimator for ξ is $\hat\xi_n=\hat\xi_n(T_1,\ldots,T_n)=\frac{1}{n}\sum_1^n T_i$. We see that $E_\xi\hat\xi_n=\xi$ and $\operatorname{var}_\xi\hat\xi_n=(n\tau'(\xi))^{-1}$ for all $\xi\in N$. If $N=[a,b]$, let

$$\begin{aligned}\tilde\xi_n=\tilde\xi_n(T_1,\ldots,T_n)&=\hat\xi_n, &&\hat\xi_n\in N\\ &=a, &&\hat\xi_n<a\\ &=b, &&\hat\xi_n>b.\end{aligned}$$

To use Lemma 1 we put

$$\hat f(t;t_1,\ldots,t_n)=f_{\tilde\xi_n}(t)$$

which obviously is measurable in $(t,t_1,\ldots,t_n)$. Let

$$\phi_\xi(t)=(\log f_\xi)'(t)=\tau'(\xi)(t-\xi)$$
$$\psi_\xi(t)=f_\xi''(t)/f_\xi(t).$$

If ξ, $\xi+\Delta\in N$, then

$$\frac{f_{\xi+\Delta}-f_\xi}{f_\xi}=\Delta\phi_\xi+\tfrac{1}{2}\Delta^2\psi_\xi+\Delta^3 B_{\xi,\Delta} \tag{8}$$

where $B_{\xi,\Delta}=\frac{1}{6}f_{\xi'}^{(3)}/f_\xi$ for some ξ' between ξ and $\xi+\Delta$. Obviously, $B_{\xi,\Delta}(t)$ is measurable in (t,Δ). We see that

$$\hat{\xi}_n = \xi + (n\tau'(\xi))^{-1} \sum_1^n \phi_\xi(T_i).$$

Put

$$\hat{\xi}_{ni} = \hat{\xi}_n(T_1, \ldots, T_{i-1}, T_{i+1}, \ldots, T_{n+1})$$
$$\tilde{\xi}_{ni} = \tilde{\xi}_n(T_1, \ldots, T_{i-1}, T_{i+1}, \ldots, T_{n+1})$$
$$\bar{\phi}_{ni} = \frac{1}{n} \sum_{\substack{j=1 \\ j \neq i}}^{n+1} \phi_\xi(T_j)$$
$$\Delta_{ni} = \hat{\xi}_{ni} - \xi = \bar{\phi}_{ni}/\tau'(\xi).$$

Let $\varepsilon > 0$, and let $N' = N \cap \langle \xi - \varepsilon, \xi + \varepsilon \rangle$.

The expectation in the r.h.s. of inequality (6) from Lemma 1 can be written

$$E_\xi \left| \frac{1}{n+1} \sum_1^{n+1} f_\xi(T_i)^{-1} (f_{\tilde{\xi}_{ni}}(T_i) - f_\xi(T_i))(I_{N'}(\hat{\xi}_{ni}) + 1 - I_{N'}(\hat{\xi}_{ni})) \right|. \tag{9}$$

Introducing (8), we see that the above expression (9) is less than or equal to (we suppress the dependency upon ξ and n and write $\phi_i = \phi(T_i)$, $\psi_i = \psi(T_i)$ where convenient)

$$\begin{aligned}
&E \left| \frac{1}{n+1} \sum_1^{n+1} \phi(T_i) \frac{\bar{\phi}_i}{\tau'} \right| + E \left| \frac{1}{n+1} \sum_1^{n+1} \psi(T_i) \left(\frac{\bar{\phi}_i}{\tau'} \right)^2 \right| \\
&\quad + E \left| \frac{1}{n+1} \sum_1^{n+1} B_{\Delta_i}(T_i) \Delta_i^3 I_{N'}(\hat{\xi}_i) \right| \\
&\quad + E \left| \frac{1}{n+1} \sum_1^{n+1} (f_{\tilde{\xi}_i}(T_i) f_\xi(T_i)^{-1} - 1)(1 - I_{N'}(\hat{\xi}_i)) \right| \\
&\quad = A_{n1}(\xi) + \ldots + A_{n4}(\xi) = A_1 + \ldots + A_4.
\end{aligned}$$

Let $\xi_n \to \xi$ in N. If we can show that

$$nA_{ni}(\xi_n) \to g_i(\xi), \qquad i = 1, 2$$

then the convergence must be uniform, so that

$$\sup_{\xi \in N} nA_{ni}(\xi) \to \sup_{\xi \in N} g_i(\xi).$$

Assume therefore that A_{n1} and A_{n2} are evaluated under the parameter value ξ_n:

$$nA_1 = nA_{n1}(\xi_n) = \frac{1}{\tau'} E|B_n|$$

where

$$B_n = \frac{1}{n+1} \sum_{i=1}^{n+1} \sum_{\substack{j=1 \\ j \neq i}}^{n+1} \phi_i \phi_j$$

$$= \frac{1}{n+1} \left(\sum_{i=1}^{n+1} \sum_{j=1}^{n+1} \phi_i \phi_j - \sum_{i=1}^{n+1} \phi_i^2 \right)$$

$$= \left(\frac{1}{\sqrt{n+1}} \sum_i \phi_i \right)^2 - \frac{1}{n+1} \sum_i \phi_i^2 = U_n^2 - V_n. \tag{10}$$

Now

$$|B_n| \leqq C_n = U_n^2 + V_n$$

$$EC_n = \frac{1}{n+1} E(\sum_i \sum_j \phi_i \phi_j) + \frac{1}{n+1} E(\sum_i \phi_i^2) = 2\tau'(\xi_n).$$

We recall that $\xi \curvearrowright E_\xi |T|^r$ is continuous and therefore bounded for $r \leqq 4$.
Then

$$\phi_{\xi_n}(T)/\tau'(\xi_n) = T - \xi_n$$

has zero expectation and bounded 3. order moment. It follows from Lyapunov's theorem that

$$U_n \xrightarrow{\mathscr{L}} U \tag{11}$$

where U is $N(0, (\tau'(\xi))^{\frac{1}{2}})$.

$E_\xi \phi_\xi^2$ is continuous in ξ, and $\mathrm{var}_\xi \phi_\xi^2$ is bounded. Then obviously

$$\frac{1}{n+1} \sum_1^{n+1} (\phi_i^2 - E\phi_i^2) \xrightarrow{P} 0$$

which entails

$$V_n \xrightarrow{P} E_\xi \phi_\xi^2 = \tau'(\xi).$$

We see that

$$E(U^2 + \tau'(\xi)) = 2\tau'(\xi) = \lim EC_n.$$

This implies (see Loéve (1963), 11.4.A) that C_n is uniformly integrable in the sequence of distributions of (U_n, V_n). Then $|B_n|$ is also uniformly integrable, so that

$$E|B_n| \to E|U^2 - \tau'(\xi)| = \tau'(\xi) \int_{\infty}^{\infty} |x^2 - 1| \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2} dx$$

$$= \tau'(\xi)(2/\pi e)^{\frac{1}{2}}.$$

We have thus showed that

$$\limsup_n \sup_\xi n A_1(\xi) \leqq \sqrt{2/\pi e}.$$

To show that $nA_{n2}(\xi_n)\to 0$, we need to know that

$$\frac{1}{n}\sum_1^{n+1} g(\psi_i)\xrightarrow{P} E_\xi\, g(\psi_i) \tag{12}$$

and

$$n^{-\frac{r}{2}}\frac{1}{n}\sum_1^{n+1} g(\psi_i)\,\phi_i^r\xrightarrow{P} 0 \qquad \text{for } r=1,2 \tag{13}$$

where $g(x)$ denotes either x or $|x|$.

We have $\psi_\xi(t)=\phi_\xi^2(t)+\tau''(\xi)(t-\xi)-\tau'(\xi)$ so that

$$|\psi_\xi(T)|\leqq \tau'(\xi)^2(T-2\,\xi\,T+\xi^2)+|\tau''(\xi)|(|T|+|\xi|)+\tau'(\xi)$$

which has continuous expectation w.r.t. ξ under P_ξ.

Since $g(\psi_{\xi_n})\to g(\psi_\xi)$ pointwise, it follows from the (generalized) Lebesgue dominated convergence theorem that $\xi \curvearrowright E_\xi\, g(\psi_\xi)$ is also continuous. Also, $\operatorname{var}_{\xi_n} g(\psi_{\xi_n})$ must be bounded, so that

$$\frac{1}{n}\sum_1^{n+1}(g(\psi_i)-E_{\xi_n} g(\psi_i))\xrightarrow{P} 0$$

which proves (12). By the general Markov inequality we get that, for all $\varepsilon>0$,

$$P(|n^{-1-\frac{r}{2}}\sum_1^{n+1} g(\psi_i)\,\phi_i^r|>\varepsilon)\leqq \frac{1}{\varepsilon}E\,|n^{-1-\frac{r}{2}}\sum_1^{n+1} g(\psi_i)\,\phi_i^r|$$

$$\leqq \frac{1}{\varepsilon}n^{-\frac{r}{2}}E\,|g(\psi_i)\,\phi_i^r|\to 0$$

since the last expectation is bounded, and (13) is verified.

Let

$$D_n=\frac{n}{n+1}\sum_1^{n+1}\psi_i\left(\frac{\bar\phi_i}{\tau'}\right)^2$$

and

$$F_n=\frac{n}{n+1}\sum_1^{n+1}|\psi_i|\left(\frac{\bar\phi_i}{\tau'}\right)^2.$$

Using the notation introduced above, we may write D_n and F_n as

$$\frac{n}{n+1}(\tau')^{-2}\left\{\left(\frac{1}{n+1}\sum g(\psi_i)\right)(\sqrt{n+1}\,\bar\phi)^2\right.$$
$$\left.-2(\sqrt{n+1}\,\bar\phi)\left(\frac{1}{n\sqrt{n+1}}\sum g(\psi_i)\phi_i\right)+\frac{1}{n^2}\sum g(\psi_i)\,\phi_i^2\right\} \tag{14}$$

with

$$\bar\phi=n^{-1}\sum_1^{n+1}\phi_i.$$

We see that $nA_{n2}=E|D_n|$, $EF_n=E|\psi_i|E(U_n/\tau')^2$, where U_n is given by (10), so that

$$EF_n=E_{\xi_n}|\psi_{i,\xi_n}|/\tau'(\xi_n)\to E_\xi|\psi_\xi|/\tau'(\xi).$$

Reasoning as before we see that, because of (11), (12), (13) and (14)

$$F_n \xrightarrow{\mathscr{L}} U^2 E_\xi|\psi_\xi|/\tau'(\xi)^2$$

which has expectation $E|\psi|/\tau'$, and

$$D_n \xrightarrow{P} E_\xi \psi_\xi = 0.$$

This implies that $E|D_n|\to 0$.

Now drop the assumption that $\xi_n\to\xi$.

We have

$$nA_{n3}(\xi)\leqq nE|B_{\Delta_i}(T_i)|\,|\Delta_i|^3 I_{\langle -\varepsilon,\varepsilon\rangle}(\Delta_i).$$

Since τ is $(1-1)$ and $\theta\curvearrowright\exp(\theta T)$ is convex,

$$e^{\tau(\xi')T}\leqq e^{\tau(\xi_1)T}+e^{\tau(\xi_2)T}$$

for $\xi'\in[\xi_1,\xi_2]=N'$. Also,

$$|f_{\xi'}^{(3)}|=f_{\xi'}\,|\phi_{\xi'}^3+3\,\phi_{\xi'}\,\phi'_{\xi'}+\phi''_{\xi'}|. \tag{15}$$

Because ϕ, ϕ' and ϕ'' are linear in T with coefficients continuous in ξ, the second factor of (15) must be bounded by $M(|T|^3+1)$ for all choices of $\xi\in N$. If we put

$$H_\xi=H_\xi(T)=M(|T|^3+1)\frac{e^{\tau(\xi_1)T}+e^{\tau(\xi_2)T}}{e^{\tau(\xi)T+\omega(\xi)}},$$

we see that

$$I_{N'}(\xi')\,|f_{\xi'}^{(3)}|/f_\xi\leqq H_\xi$$

and that

$$E_\xi H_\xi\leqq M'(E_{\xi_1}|T|^3+E_{\xi_2}|T|^3)$$

which is bounded on N.

Now $H_\xi(T_i)$ is independent of Δ_i, so

$$nA_3\leqq n(EH)(E|\Delta_i|^3)\leqq\frac{1}{n}(EH)(E|T-\xi|^3)\Rightarrow\sup_\xi nA_{n3}(\xi)\to 0.$$

To conclude the proof of the proposition, we note that

$$
\begin{aligned}
A_4 &= E(1-I_{N'}(\hat{\xi}_i)) + E(f_{\tilde{\xi}_i}(T_i) f_{\xi}(T_i)^{-1}(1-I_{N'}(\hat{\xi}_i))) \\
&= P(\hat{\xi}_i \notin N') + \int_{[\hat{\xi}_i \notin N']} f_{\tilde{\xi}_i}(t_i) \prod_{j \neq i} f_{\xi}(t_j)\, dP \\
&= 2P(\hat{\xi}_i \notin N') \leqq 2P(|\Delta_i| \geqq \varepsilon) \leqq 2\varepsilon^{-4} E(\Delta_i^4) \leqq 2\varepsilon^{-4} n^{-3} E|T-\xi|^4.
\end{aligned}
$$

Then obviously sup $n A_{n4}(\xi) \to 0$.

4. Lower Bounds for $\delta(\mathscr{E}^n, \mathscr{E}^{n+1})$

Let $\mathscr{E}$, $\mathscr{F}$ be experiments over Θ and let λ be a prior distribution on Θ. Under certain regularity conditions we may interpret $\delta(\mathscr{E}, \mathscr{F})$ as the maximal difference in achievable Bayes-risk. To bound the deficiency from below, we shall use inequality (2) for a certain decision problem. In some important special cases this bound actually coincides with the deficiency. However, it may be of some interest to note that this bound may be derived from an alternative measure of information with a certain intuitive appeal. This gives another interpretation of the deficiency (in some cases):

An informative experiment must give rise to posterior distributions that are "concentrated" (on the average). A measure of this property is the concentration function (see e.g. Hengartner and Theodorescu (1973)):

Let μ be a Borel measure on $\mathbb{R}$. The *concentration function* of μ is defined as

$$
\begin{aligned}
Q_\mu(l) &= \sup_{x \in \mathbb{R}} \mu[x-l/2, x+l/2]; \qquad l \geqq 0 \\
&= 0, \qquad l < 0,
\end{aligned}
$$

i.e. $Q_\mu(l)$ is the "maximal concentration of μ on a closed interval of length l". Acdording to the above reference Q_μ is a right-continuous distribution function and the supremum is achieved, in say $x_0(l)$. Now choose $l_n \downarrow l$ and $r_n \in \mathbb{Q}$,

$$
r_n \to x_0(l) \text{ such that } \bigcap_n \left[r_n - \frac{l_n}{2},\ r_n + \frac{l_n}{2}\right] = \left[x_0(l) - \frac{l}{2}, x_0(l) + \frac{l}{2}\right].
$$

Then

$$
\begin{aligned}
Q_\mu(l) &= \mu\left[x_0(l) - \frac{l}{2}, x_0(l) + \frac{l}{2}\right] = \lim \mu\left[r_n - \frac{l_n}{2}, r_n + \frac{l_n}{2}\right] \\
&\leqq \lim R_\mu(l_n) \leqq \lim Q_\mu(l_n) = Q_\mu(l)
\end{aligned}
$$

where

$$
R_\mu(l) = \sup_{r \in \mathbb{Q}} \mu\left[r - \frac{l}{2}, r + \frac{l}{2}\right].
$$

If now $\mu(\cdot|x)$ is a $(\mathscr{X}, \mathscr{A})$-measurable Borel probability measure, then for a fixed $l > 0$, $Q_{\mu(\cdot|x)}(l) = \lim R_{\mu(\cdot|x)}(l_n)$ which is $\mathscr{A}$-measurable since $R_{\mu(\cdot|x)}(l_n)$ must be.

Assume that $\mathscr{E}$ and $\mathscr{F}$ are experiments with common parameter set $\Theta \subset \mathbb{R}$, consisting in observing the r.v.'s X and Y respectively. Let λ be a prior distribution on Θ and let the concentration functions of the posterior distributions be $Q_{\mathscr{E}}(l|X)$ and $Q_{\mathscr{F}}(l|Y)$. Then a plausible measure of information distance would be

$$\sup_l (E Q_{\mathscr{F}}(l|Y) - E Q_{\mathscr{E}}(l|X)). \tag{16}$$

Returning to the task of finding lower bounds for deficiencies, we will need the following observation:

Let $\mathscr{E} = (\mathscr{X}, \mathscr{A}, P_\theta : \theta \in \Theta)$ where Θ is a Borel subset of $\mathbb{R}$, and $\theta \curvearrowright P_\theta(A)$ is measurable for all A. Let the decision space $(T, \mathscr{S})$ be the set of closed intervals of $\mathbb{R}$ with length l (with the obvious σ-algebra induced from $\mathbb{R}^2$). Let the loss-function be

$$\begin{aligned} L_\theta(t) &= -1, \quad \theta \in t \\ &= 1, \quad \theta \notin t \end{aligned}$$

and let λ be a prior distribution, with $\lambda(\cdot|x)$ as posterior distribution. Then the posterior Bayes-risk equals $1 - 2Q_{\lambda(\cdot|x)}(l)$ and the Bayes-risk

$$b_\lambda = 1 - 2\lambda P Q_{\lambda(\cdot|x)}(l). \tag{17}$$

This is seen as follows:

Let ρ be a decision-rule. We can, according to Loéve (1963; 27.2 B) specify $\lambda(\cdot|x)$ as an $\mathscr{A}$-measurable measure over Θ, where

$$\lambda P \rho L = \int (\int L_\theta(t)\, \lambda(d\theta|x))(\lambda P \times \rho)(dx \times dt),$$

but

$$\inf_{t \in T} \int L_\theta(t)\, \lambda(d\theta|x) = 1 - 2Q_{\lambda(\cdot|x)}(l)$$

so that

$$b_\lambda = \int (1 - 2Q_{\lambda(\cdot|x)}(l))\, \lambda P(dx).$$

If we insert (17) into inequality (2), we get a lower bound for deficiencies which coincides with 2 times the measure (16). Also, we can utilize the well-known fact that posterior distributions often are asymptotically normal to find estimates of this lower bound.

5. Lower Bound for $\delta(\mathscr{E}^n, \mathscr{E}^{n+r})$ when $\mathscr{E}$ is a 1-Parameter Exponential Experiment

We will use the technique outlined in the preceding section to prove:

Proposition 2. *Let $\mathscr{E} = (\mathscr{X}, \mathscr{A}, P_\theta : \theta \in \Theta)$ and let*

$$\frac{dP_\theta}{d\mu} = h\, e^{c(\theta) + \theta T}; \qquad \theta \in \Theta$$

where μ is σ-finite, $h \geqq 0$ and T are random variables, and $\Theta \subset \mathbb{R}$ contains a non-degenerate interval. Let r_n be a sequence of integers such that $0 < r_n \leqq n^\beta$ for $0 < \beta < 1$. If θ is identifiable or equivalently, T is not a.s. constant, then

$$\liminf_{n \to \infty} \frac{n}{r_n} \delta(\mathscr{E}^n, \mathscr{E}^{n+r_n}) \geqq \sqrt{2/\pi e}.$$

Otherwise, $\delta(\mathscr{E}^n, \mathscr{E}^{n+r_n}) = 0$.

Under the hypothesis of the above proposition, with θ identifiable, and when in addition $\Theta \subset K \subset I^0$ where K is compact and I the natural parameter space of $\{P_\theta\}$, then liminf and limsup of $\frac{n}{r}\delta(\mathscr{E}^n, \mathscr{E}^{n+r})$ must lie in the interval

$$[\sqrt{2/\pi e}, 2\sqrt{2/\pi e}]$$

(Here r is fixed). The lower bound is known to be sharp, cf. Example 1.

Proof of the Proposition. We choose a suitable sequence of prior distributions and compute the difference in Bayes risk between $\mathscr{E}^n$ and $\mathscr{E}^{n+r}$ for the special decision problem of the previous section. This is done by approximating the posterior distributions by normal distributions, and showing that the error thus introduced is of order $o(r/n)$.

If θ is non-identifiable (i.e. $\theta \neq \theta' \nRightarrow P_\theta \neq P_{\theta'}$) then T must be a.s. constant so that $\mathscr{E}$ is the totally non-informative experiment $P_\theta \equiv P$, so that $\mathscr{E}^n \sim \mathscr{E}^{n+1}$.

In the case of identifiability we may assume without loss of generality that $0 \in \Theta^0$. Introduce the new parameter h (not to be confused with the r.v. h appearing in $dP_\theta/d\mu$) in $\mathscr{E}^m$, $m = 1, 2, \ldots$ by

$$\theta = h/\sqrt{n}.$$

Then

$$\frac{dP_\theta}{dP_0} = \exp\{c(\theta) - c(0) + \theta T\}.$$

We may write

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

where $\Delta(h/\sqrt{n}) = c^{(3)}(\theta')(h/\sqrt{n})^3/6$ for sufficiently small h, for some θ' between 0 and θ. This makes

$$\frac{dP_h^m}{dP_0^m} = \exp\left\{-\frac{m h^2}{2n}(-c''(0)) + \frac{h}{\sqrt{n}} \sum_1^m (T_i + c'(0)) + m\Delta\left(\frac{h}{\sqrt{n}}\right)\right\}.$$

Let the prior distribution λ_n have density w.r.t. Lebesgue-measure

$$\gamma_n \exp\left\{-n\Delta\left(\frac{h}{\sqrt{n}}\right) - \frac{h^2}{2\kappa^2}\right\} I_{[-c_n, c_n]}(h)$$

where

$$c_n = c\,n^q, \qquad c>0 \quad \text{and} \quad 0<q<1/6.$$

We see that $n^{-\frac{1}{2}}[-c_n, c_n] \subset \Theta^0$ for all $n \geqq N$ for some N. It is easy to see that the posterior distribution function $H_n(t|X_m)$ in $\mathscr{E}^m$ (where $X_m = (X_{m1}, \ldots, X_{mm})$ is the vector of observations from $\mathscr{E}^m$) is given by

$$H_{mn}(t|X_m) = C_{nm}(X_m) \int_{-c_n}^{t} \exp\left\{-\frac{(h-\mu_{mn})^2}{2\sigma_{mn}^2} + (m-n)\Delta\left(\frac{h}{\sqrt{n}}\right)\right\} dh$$

for $|t| \leqq c_n$, with

$$\sigma_{mn}^2 = \left(\frac{m}{n}\frac{1}{\tau^2} + \frac{1}{\kappa^2}\right)^{-1}, \qquad \tau^2 = \frac{1}{-c''(0)}$$

$$\mu_{mn} = \sigma_{mn}^2 \frac{1}{\sqrt{n}} \sum_{1}^{m} (T_i - \xi), \qquad \xi = -c'(0).$$

We may as before regard $T_1, T_2, \ldots$ as being defined on the same probability space. However, their distribution depends on n (through λ_n). We proceed to prove that H_{mn} is approximately normal.

First, note that $c^{(3)}$ is continuous and bounded so that

$$n\Delta\left(\frac{h}{\sqrt{n}}\right) = \tfrac{1}{6} c^{(3)}(\theta') \frac{h^3}{\sqrt{n}} \to 0 \tag{18}$$

uniformly in $|h| \leqq c_n$.

By the way, this shows that λ_n converges weakly to the $\mathrm{N}(0, \kappa)$-distribution.

We will need the following

Lemma 2. *Let $f, g \in L_1(\mathbb{R})$ be non-negative and such that the L_1-norms $\|f\|$, $\|g\| > 0$.*

Then

$$\left\|\frac{f}{\|f\|} - \frac{fg}{\|fg\|}\right\| \leqq 2\frac{\|f - fg\|}{\|f\| \vee \|fg\|} \leqq 2\frac{\|f - fg\|}{\|f\|}.$$

Proof of the Lemma. Assume first that $\|f\| \geqq \|fg\|$. Then

$$\left\|\frac{f}{\|f\|} - \frac{fg}{\|fg\|}\right\| = \left\|\left(\frac{f}{\|f\|} - \frac{fg}{\|f\|}\right) + \left(\frac{fg}{\|f\|} - \frac{fg}{\|fg\|}\right)\right\| \leqq \frac{\|f - fg\|}{\|f\|}$$

$$+ \|fg\| \left|\frac{1}{\|f\|} - \frac{1}{\|fg\|}\right| \leqq \frac{\|f - fg\|}{\|f\|} + \left|\frac{\|fg\| - \|f\|}{\|f\|}\right| \leqq 2\frac{\|f - fg\|}{\|f\|}.$$

The case $\|f\| < \|fg\|$ is treated analogously.

Let, for fixed X_m, the L_1-functions f_{mn}, g_{mn} be given by

$$f_{mn}(h)=\frac{1}{\sqrt{2\pi}\,\sigma_{mn}}\exp\left\{-\frac{(h-\mu_{mn})^2}{2\sigma_{mn}^2}\right\}$$

$$g_{mn}(h)=\exp\left\{(m-n)\,\Delta\left(\frac{h}{\sqrt{n}}\right)\right\} I_{[-c_n,c_n]}(h).$$

Applying the above lemma to f_{mn}, g_{mn} we see that the difference between the distribution functions H_{mn} and

$$F_{mn}(t|X_m)=\int_{-\infty}^{t} f_{mn}$$

satisfies

$$|F_{mn}(t|X_m)-H_{mn}(t|X_m)|\leqq\int_{-\infty}^{t}\left|f_{mn}-\frac{f_{mn}\,g_{mn}}{\|f_{mn}\,g_{mn}\|}\right|\leqq\left\|f_{mn}-\frac{f_{mn}\,g_{mn}}{\|f_{mn}\,g_{mn}\|}\right\|$$
$$\leqq 2A_{mn}+2B_{mn},$$

where

$$A_{mn}=\int_{-c_n}^{c_n}|e^{(m-n)\Delta(\frac{h}{\sqrt{n}})}-1|\,dF_{mn}(h|X_m)$$

$$B_{mn}=F_{mn}(-c_n|X_m)+1-F_{mn}(c_n|X_m).$$

Writing

$$K_n=n\sup_{|h|\leqq c_n}\left|\Delta\left(\frac{h}{\sqrt{n}}\right)\right|$$

and using the relation

$$|e^u-1|=|u|\,e^v,\qquad |v|\leqq|u|$$

we find that when $m\leqq n$, A_{mn} is bounded by

$$2\frac{m-n}{n}K_n\exp\left\{\frac{m-n}{n}K_n\right\}.$$

When $m=n$, $A_{mn}=0$ and when $m=n+r_n$, (18) entails that

$$E_{\lambda_n p^m}\left(\frac{n}{r_n}A_{mn}\right)\to 0. \tag{19}$$

It will be proved later that also

$$E_{\lambda_n p^m}\left(\frac{n}{r_n}B_{mn}\right)\to 0. \tag{20}$$

Let Q_{mn} and R_{mn} be the concentration functions of H_{mn} and F_{mn} respectively. Then

$$\sup_l|Q_{mn}(l|X_m)-R_{mn}(l|X_m)|\leqq 2\sup_t|H_{mn}(t|X_m)-F_{mn}(t|X_m)|$$

for all X_m. It is obvious that F_{mn} will achieve maximal mass over closed intervals of length 21 in the interval

$$J_{mn}=[\mu_{mn}-l, \mu_{mn}+l]$$

so that

$$R_{mn}(2l)=\int_{J_{mn}} f_{mn}=2\Phi\left(\frac{l}{\sigma_{mn}}\right)-1.$$

Here Φ is the cumulative normal distribution function

$$\Phi(t)=\int_{-\infty}^{t}\phi, \qquad \phi(t)=(2\pi)^{-\frac{1}{2}}e^{-\frac{1}{2}t^2}.$$

It is easily seen that when $m=n+r$

$$\sigma_{nn}^2, \sigma_{mn}^2 \to \frac{1}{\alpha}=\left(\frac{1}{\tau^2}+\frac{1}{\kappa^2}\right)^{-1} \tag{21}$$

and that (Taylor's formula)

$$\frac{1}{\sigma_{mn}}-\frac{1}{\sigma_{nn}}=\frac{1}{2\sqrt{\alpha_n}\,\tau^2}\,\frac{r}{n}$$

where α_n is between $\tau^{-2}mn^{-1}+\kappa^{-2}$ and α, so that $\alpha_n\to\alpha$. Accordingly, by the mean value theorem

$$R_{mn}(2l)-R_{nn}(2l)=2\frac{r}{n}\cdot\frac{l\phi(\beta_n)}{2\tau^2\sqrt{\alpha_n}}$$

where $l\sigma_{nn}^{-1}\leqq\beta_n\leqq l\sigma_{mn}^{-1}$, so that

$$\beta_n\to l\sqrt{\alpha}$$

and

$$\frac{n}{r}(R_{mn}(2l)-R_{nn}(2l))\to\frac{\sqrt{\alpha}\,l\phi(l\sqrt{\alpha})}{\tau^2\alpha}=\frac{l\sqrt{\alpha}\,\phi(l\sqrt{\alpha})}{\left(1+\frac{\tau^2}{\kappa^2}\right)}.$$

The function $t\phi(t)$ achieves its maximum for $t=1$. Also, we may choose κ arbitrarily large. This entails that for $l=\alpha^{-\frac{1}{2}}$

$$\sup_{\kappa}\limsup_{n}\frac{n}{r}(R_{mn}(2l)-R_{nn}(2l))\geqq(2\pi e)^{-\frac{1}{2}}.$$

By the inequality (2) and by (17) we get, with $m=n+r$

$$\frac{n}{r}\delta(\mathscr{E}^n, \mathscr{E}^{n+r}) \geqq 2\frac{n}{r} E_{\lambda_n P^m}[Q_{mn}(2l|X_m) - Q_{nn}(2l|X_n)]$$

$$\geqq \frac{n}{r}[R_{mn}(2l|X_m) - R_{nn}(2l|X_n)] - \frac{n}{r} E|Q_{mn}(2l|X_m) - R_{mn}(2l|X_m)|$$

$$-\frac{n}{r} E|Q_{nn}(2l|X_n) - R_{nn}(2l|X_n)|.$$

The two last members tend to zero by (19) and (20). Thus, if we can show (20), the proposition will be proved.

We may write $F_{mn}(-c_n|X_m)$ and $1-F_{mn}(c_n|X_m)$ as

$$\Phi(\sigma_{mn}^{-1}(W_{mn} - c_n))$$

where

$$W_{mn} = u\sum_1^m (T_i - \xi)\,\sigma_{mn}^2$$

and u denotes $n^{-\frac{1}{2}}$ or $-n^{-\frac{1}{2}}$.

Assume that m is either n or $n+r_n$. Put

$$X_n = n\,\Phi(\sigma_{mn}^{-1}(W_{mn} - c_n)).$$

Now $0 \leqq X_n \leqq n$ so that for all $\varepsilon > 0$

$$E X_n = E X_n I_{[X_n \leqq \varepsilon]} + E X_n I_{[X_n > \varepsilon]} \leqq \varepsilon + n P(X_n > \varepsilon)$$

and

$$P(X_n > \varepsilon) = P\left(\sum_1^m V_i \geqq \left(c_n + \sigma_{mn}\,\Phi^{-1}\left(\frac{\varepsilon}{n}\right)\right)\sigma_{mn}^{-2}\right) \tag{22}$$

where

$$V_i = u(T_i - \xi).$$

It is seen that we may use a "large deviation"-type argument. By the general Markov inequality applied to the distribution P_θ^m of $T_1, T_2, \ldots$ given θ, we get

$$P_\theta\left(\sum_1^m V_i > d_n\right) \leqq e^{-d_n} E_\theta\left[\exp\left(\sum_1^m V_i\right)\right].$$

$V_1, \ldots, V_m$ are now independent, so that the last expectation is

$$(E_\theta\, e^V)^m.$$

We have

$$E_\theta\, u\,T = -u\,c'(\theta)$$

$$E_\theta\, e^{uT} = \int e^{c(\theta) + (\theta+u)T}\, dP_0 = e^{c(\theta) - c(\theta+u)}$$

for all θ in a suitable neighbourhood of 0, for sufficiently large n.

Now, by using Taylor's formula twice

$$c(\theta)-c(\theta+u)=-uc'(\theta)-\tfrac{1}{2}u^2c''(\theta') \\ =-uc'(0)-u\theta c''(0)-\tfrac{1}{2}u\theta^2c^{(3)}(\theta'')-\tfrac{1}{2}u^2c''(\theta')$$

where $|\theta'|, |\theta''|\leqq c_n n^{-\frac{1}{2}}$. Then

$$(E_\theta e^V)^m=\exp(\pm mn^{-1}h)\exp\{-\tfrac{1}{2}(mn^{-1}c''(\theta')\pm mn^{-3/2}h^2c^{(3)}(\theta''))\}.$$

The second factor above is easily seen to be bounded when $|h|\leqq c_n$. It follows from (18) that for suitable K'

$$\int\exp(\pm mn^{-1}h)\lambda_n(dh)\leqq K'\int_{-\infty}^{\infty}\exp(2|h|-\tfrac{1}{2}h^2)\,dh$$

which is finite. Accordingly,

$$\lambda_n P_h^m\left(\sum_1^m V_i>d_n\right)\leqq K e^{-d_n}. \tag{23}$$

Since the normal distribution has moments of any order,

$$|x|^{2/q}\Phi(x)\xrightarrow[x\to-\infty]{}0.$$

For all $\varepsilon>0$, there is an $M\in\langle 0,\infty\rangle$ such that for $x<-M$

$$\Phi(x)<\varepsilon|x|^{-\frac{2}{q}}$$
$$\Rightarrow\ -\Phi^{-1}(\varepsilon|x|^{-\frac{2}{q}})<|x|.$$

Putting $x=-n^{q/2}$ we see that for large enough n

$$\left|\phi^{-1}\left(\frac{\varepsilon}{n}\right)\right|<n^{q/2}.$$

Together with (21) this implies that

$$\sigma_{mn}\phi^{-1}\left(\frac{\varepsilon}{n}\right)=o(c_n)\qquad\text{as } n\to\infty.$$

By (22) and (23) we now see that

$$E_{\lambda_n P^m}(nB_{mn})\to 0$$

which finishes the proof.

1. Comments

As mentioned before, we may expect that $\delta(\mathscr{E}^n,\mathscr{E}^{n+1})\sim c/n$ for a wide class of experiments $\mathscr{E}$, and it would be natural to try to extend our results. One direction which is likely to be successful is to multiparameter exponential families. Another is the class of experiments fulfilling certain "Cramér-type" regularity conditions. To establish our upper bound we have essentially used

(i) that the density can be expanded in a Taylor formula where the coefficients have bounded moments up to a certain order.

(ii) The existence of a "nice" estimator $\hat{\xi}$ such that

$$\hat{\xi}-\xi \simeq \frac{1}{\sqrt{n}} \sum_{1}^{n} \frac{\partial \log f}{\partial \xi}(T_i).$$

In rather general situations, similar estimators exist, e.g. the maximum likelihood estimator.

The proof for the lower bound also essentially uses (i). Torgersen has suggested that when $\mathscr{E}$ is a translation experiment, it may be simple to establish (i) and (ii). In that case, we may hope to avoid the boundedness condition for Θ.

Of course, an interesting question is whether our upper bound can be improved. This seems to call for a new method of proof.

Acknowledgement. This paper is a cand. real. thesis written under the guidance of Erik N. Torgersen, and the author is grateful to him for suggesting the problem, simplifying the proofs and for his advice throughout.

References

1. Ferguson, R.: Mathematical Statistics. New York-London: Academic Press 1967
2. Hengartner, W., Theodorescu, R.: Concentration Functions. New York-London: Academic Press 1973
3. LeCam, L.: Sufficiency and approximate sufficiency. Ann. Math. Statist. **35**, 1419–1455 (1964)
4. LeCam, L.: Notes on asymptotic methods in statistical decision theory. Montreal: Centre de Recherches Math., Univ. de Montréal 1974a
5. LeCam, L.: On the information contained in additional observations. Ann. Statist. **2**, 630–649 (1974b)
6. Lehmann, E.: Testing statistical hypotheses. New York: Wiley 1959.
7. Loéve, M.: Probability theory. 3rd Edition. Princeton: Van Nostrand 1963
8. Neveu, J.: Mathematical foundations of the calculations of probability. San Francisco: Holden-Day 1965
9. Torgersen, E.: Comparison of translation experiments. Ann. Math. Statist. **43**, 1383–1399 (1972)
10. Torgersen, E.: Asymptotic behaviour of powers of dichotomies. Stat. res. rep., Univ. of Oslo (1974)
11. Torgersen, E.: Deviations from total information and from total ignorance as measures of information. Stat. res. rep., Univ. of Oslo (1976)

Received December 17, 1979; in revised form September 7, 1981