STATISTICAL RESEARCH REPORT June 1979
Institute of Mathematics
University of Oslo

ADDITIONAL OBSERVATIONS AND STATISTICAL INFORMATION IN THE CASE OF 1-PARAMETER EXPONENTIAL DISTRIBUTIONS

by

Jon Helgeland

Contents

Summary

We study the increase in information by replication of experiments E, which are of 1-parameter exponential type. We show that when the parameter space is a compact, non-degenerated interval, then $\sqrt{\frac{2}{\pi e}} \leq \underline{\lim}\, n\delta(E^n, E^{n+1}) \leq \overline{\lim}\, n\delta(\cdot) \leq 2$.

1. Introduction

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

In order to compare experiments w.r.t. "content of statistical information" we use the concept of deficiencies (introduced by L. LeCam, [3]):

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

(*) $$P_\theta \rho L_\theta \leq Q_\theta \sigma L_\theta + \varepsilon_\theta \, ||L_\theta|| \quad , \quad \forall \theta$$

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

In (*) we may replace $||L_\theta||$ by $||L||$ and we may confine ourselves to non-negative L if we replace "ε_θ" in (*) by "$\frac{1}{2}\varepsilon_\theta$". If E is 0-deficient rel. F, we say that E is more informative than F .(written $E \geq F$) and if both $E \geq F$ and $F \geq E$, E and F are said to be equivalent (written $E \sim F$). The infimum over all constants $\varepsilon > 0$ such that E is ε-deficient rel. F is written $\delta(E, F)$ and is called the deficiency of E

rel. F. The Δ-distance between E and F is defined by $\Delta(E,F) = \delta(E,F) \vee \delta(E,F)$. The class of experiments which are equivalent to an experiment E, is called the experiment type of E. We may define the set of all experiment types $\mathbb{E}$, and $(\mathbb{E},\Delta)$ becomes a complete metric space ([13]).

If $F = (\chi,A,P_\theta;\theta\in\Theta)$ and $E = (\chi,B,P_\theta|B;\theta\in\Theta)$ where B is a sub-σ-algebra of A and $P_\theta|B$ is the restriction of P_θ to B, then obviously $E \leq F$. One measure of the loss of information when observing only B-measurable events is $\delta(E,F)$, another is the insufficiency (LeCam[4]) which is given by

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

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

The concept of deficiency has several interpretations that each are natural ways of formally defining loss of information. We mention here the following theorems (LeCam [3])

(i) Let $E = (\chi,A,P_\theta:\theta\in\Theta)$, $F = (Y,B,Q_\theta:\theta\in\Theta)$ $\varepsilon : \Theta \to [0,\infty>$.

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

(ii) <u>The Markov kernel criterion</u>:

Let E,F be as above. Assume that Y is a Borel-subset of a Polish space and B is the restricted Borel-σ-algebra. Then E is ε-deficient rel. F if and only if there exists a Markov

kernel $M : \mathcal{B} \times \chi \to [0,1]$ such that $\|P_\theta M - Q_\theta\| \leq \varepsilon_\theta$, θ.
(A Polish space is a complete separable metric space equipped with its Borel-σ-algebra, a Markov kernel is a mapping $M : \mathcal{B} \times \chi \to [0,1]$ such that

(a) $M(\circ|x)$ is a probability measure for every $x \in \chi$

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

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

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

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

(*) $$P_\theta \rho L_\theta \leq Q_\theta \sigma L_\theta + \varepsilon \|L\| , \quad \forall \theta$$

$$\Rightarrow \; b_\lambda^{\mathcal{E}} \leq \lambda Q \sigma L + \varepsilon \|L\| .$$ Then

$$\delta(\mathcal{E}, \mathcal{F}) \geq \frac{1}{\|L\|} (b_\lambda^{\mathcal{E}} - \lambda Q \sigma L)$$

$$\geq \frac{1}{\|L\|} (b_\lambda^{\mathcal{E}} - b_\lambda^{\mathcal{F}}) .$$

There is a connection between CE-sufficiency ("conditional expectation"-sufficiency), i.e. sufficiency in the sense of Halmos and Savage) and deficiency: (Bahadur see [12]).

If $\mathcal{E} = (\chi, \mathcal{B}, P_\theta | \mathcal{B}; \theta \in \Theta)$ and

$\mathcal{F} = (\chi, A, P_\theta; \theta \in \Theta)$

where $\mathcal{B}$ is a sub-σ-algebra of A then:

(i) $\mathcal{B}$ CE-sufficient for $\mathcal{F}$, implies

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

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

In the following we will consider experiments of the form

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

where $\mathcal{E} = (\chi, A, P_\theta^n; \theta \in \Theta)$

i.e. $\mathcal{E}^n$ is n independent replications of $\mathcal{E}$. It is obvious that $\mathcal{E}^n \leq \mathcal{E}^m$ when $n \leq m$, and a natural question arises: How much more informative than $\mathcal{E}^n$ is $\mathcal{E}^m$? This may be of interest in e.g. planning of (replicated) experiments when the exact nature of the decision problem is not determined on beforehand. Let $K_{\mathcal{E}}$ denote the "cost" of performing $\mathcal{E}$, L the loss function. Then the risk function is, under the decision rule ρ : $R_{\mathcal{E}}(\theta) = P_\theta \rho L_\theta + K_{\mathcal{E}}$. Suppose that $||L|| \leq$. Then we prefer $\mathcal{E}^n$ to $\mathcal{E}^{n+1}$ when $\delta(\mathcal{E}^n, \mathcal{E}^{n+1}) \leq K_{\mathcal{E}^{n+1}} - K_{\mathcal{E}^n}$, and $\mathcal{E}^{n+1}$ to $\mathcal{E}^n$ when

$$\delta(\mathcal{E}^n, \mathcal{E}^{n+1}) \geq K_{\mathcal{E}^{n+1}} - K_{\mathcal{E}^n} .$$

That $\mathcal{E}^n$ is better than $\mathcal{E}^m$ in the above sense means that: To any risk function $R_{\mathcal{E}^m}$ there exists a $R_{\mathcal{E}^n}$ (which is the risk in the same decision problem) such that

$$R_{\mathcal{E}^m} \geq R_{\mathcal{E}^n} .$$

Example 1.1. Let $\mathcal{E}$ consist in observing $X \sim N(\theta,\sigma)$ where σ is known. Then (Torgersen [9])

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

If we let $K_{\mathcal{E}^n} = k_o + nk_1$, then $n_o = \frac{\sqrt{\frac{2}{\pi e}}}{k}$ is the optimal sample size in the above sense.

Intuitively one may expect that $\mathcal{E}^n$ gets very informative as $n \to \infty$, and that one additional observation gets more and more unimportant. In fact, when Θ is finite, then

$\Delta(\mathcal{E}^n,\mathcal{M}_a) \to 0$, when $\mathcal{M}_a$ is the experiment

where θ itself is observed without uncertainty, and

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

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

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

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

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

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

If Θ is uncountable and $\mathcal{E}$ is dominated, then always $\delta(\mathcal{E},\mathcal{M}_a) = 2$. These results are from Torgersen [11] .

Let now $\mathcal{E}$ be an experiment with arbitrary Θ such that $\theta \mapsto P_\theta$ is (1-1). Since the restriction $\mathcal{E}^n|F$ of $\mathcal{E}^n$ to finite subsets $F \subset \Theta$ must converge to $\mathcal{M}_a|F$, $\mathcal{M}_a$ is the only possible Δ-limit for $\{\mathcal{E}^n\}$. If now $\mathcal{E}$ is dominated, $\Delta(\mathcal{E}^n,\mathcal{E}^m) \underset{n,m\to\infty}{\not\to} 0$ since $(\mathbb{E},\Delta)$ is complete. This implies that $\sum_{k=0}^{\infty} \delta(\mathcal{E}^{n+k},\mathcal{E}^{n+k+1}) \underset{n\to\infty}{\not\to} 0$ and furthermore that

$$n^{-\alpha} = o(\delta(\mathcal{E}^n,\mathcal{E}^{n+1}))$$

for all $\alpha > 1$.

The insufficiency $\eta(\mathcal{E}^n,\mathcal{E}^{n+1})$ may be used to study $\delta(\mathcal{E}^n,\mathcal{E}^{n+1})$ since always $\eta(\circ) \geq \delta(\circ)$, but the approximation may be poor: If $\mathcal{E}$ consists in observing $X \sim N(\theta,1)$ (Example 1.1) then

$$\eta(\mathcal{E}^n,\mathcal{E}^{n+1}) \geq \frac{1}{2\pi} e^{-\frac{1}{4n}} \frac{1}{\sqrt{n}}$$

$$\Rightarrow \delta(\mathcal{E}^n,\mathcal{E}^{n+1}) = o(\eta(\mathcal{E}^n,\mathcal{E}^{n+1})).$$

This, and the following result are shown by LeCam [4]: for all $n,k \geq 0$,

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

where D_n is a dimensionality constant for Θ, given by: The Hellinger distance H $(H^2(P,Q) = \int(\sqrt{dP} - \sqrt{dQ})^2$ for probability measure $P,Q)$ induces a metric on Θ: $h(\theta,\theta') = H(P_\theta,P_{\theta'})$. Put $a_\nu = \sqrt{\frac{1}{2^{\nu+10}}}$, $b_\nu = \sqrt{\frac{1}{2^\nu}}$, $\nu = 0,1,2,\ldots$

For finite $S \subset \Theta$, diam $S \leq b_{\nu-1}$, 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

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

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 $c(\nu) = \sup_S C'_S$ where the supremum is taken over finite $S \subset \Theta$ such that diam $S \leq b_{\nu-1}$. Let $K_n = 1 \vee \sup_{2^\nu \leq n} c(\nu)$ and put $D_n = 16 \log 6 K_n$. LeCam also gives an example $\mathcal{E}$ such that $\delta(\mathcal{E}^n, \mathcal{E}^{n+1}) \nrightarrow 0$:

Example 1.2. Let (χ, A, λ) be $[0,1]$ equipped with Lebesgue-measure λ, let $\Theta = \{0,1,2,\ldots\}$. Let P_θ be given by

$$\frac{dP_\theta}{d\lambda}(x) = \sum_{k=o}^{\theta-1} 2\, I_{\left[\frac{2^k+1}{2^\theta}, \frac{2^k+2}{2^\theta}\right]}(x) \;, \text{ for } \theta \geq 1$$

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

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

Example 1.1. (Continued).

(i) Let $\mathcal{E}$ consist in observation of $X \sim N_k(\xi, \Sigma)$ where Σ is known, positive definite, ξ unknown vector. Then

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

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

(ii) Let $\mathcal{E}$ consist in observation of $X \sim R\,\langle 0,\theta]$, $\theta \in \Theta = \langle 0,\infty\rangle$. Then

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

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

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

for Θ uncountable and $\mathcal{E}$ "nice". We will show that in the 1-parameter exponential case, with Θ a nondegenerate compact interval

$$\sqrt{\frac{2}{\pi e}} \leq \underline{\lim}\, n\, \delta(\mathcal{E}^n,\mathcal{E}^{n+1}) \leq \overline{\lim}\, n\, \delta(\mathcal{E}^n,\mathcal{E}^{n+1}) \leq 2.$$

We will be referring to wellknown results about these experiments, see [5,8].

<u>About the notation</u>: We will (usually) employ lower indices to index experiments, and upper indices to index components of vectors.

$(X^1,\ldots,\check{X}^i,\ldots,X^n)$ means

$(X^1,\ldots,X^{i-1},X^{i+1},\ldots,X^n)$ and

$$X^{(c)} = \begin{cases} X, & |X| \leq c \\ 0 & \text{otherwise} \end{cases}$$

$X_n \xrightarrow{\mathcal{L}_{P_n}}$ means

$\mathcal{L}(X_n|P_n) \xrightarrow{w}$.

The symbols $P_o(\lambda)$, bin (n,p), $N(\xi,\sigma)$, χ^2_k denote respectively the Poisson, binomial, normal (with variance σ^2), and chi-square (with k-degrees of freedom) - distributions.

2. Multinomial experiments

In this section we will consider the experiments $\mathcal{E}^n$ consisting in observation of the i.i.d. variables $Y_1,\dots,Y_n$, where Y_i assumes the values $1,\dots,s$ with probabilities $\theta_1,\dots,\theta_s$, $\theta \in \Theta = K_s$ which is the standard simplex in $\mathbb{R}^s$ (i.e. $\{x \in [0,1]^s : \Sigma x_i = 1\}$). By sufficiency we get $\mathcal{E}^n \sim \tilde{\mathcal{E}}^n$ where $\tilde{\mathcal{E}}^n$ consists in observation of the s-nomial variable $X_n = (X_n^1,\dots,X_n^s)$.

2.1. Upper bound for $\delta(\mathcal{E}^n,\mathcal{E}^{n+1})$.

The Markov kernel criterion provides a tool for finding upper bounds for deficiencies. In our case, we may define a Markov kernel M thus: Y_{n+1} assumes the value ν with probability θ_ν, we may predict this value by letting $\hat{Y}_{n+1} = \nu$ with probability $\hat{\theta}_\nu = \frac{1}{n} X_n^\nu$. This means

$$m(y|x) = \begin{cases} X^\nu/n & ; \quad y = x + e^\nu \\ 0 & \text{otherwise} \end{cases}$$

where $e^\nu = \{0,\dots,1,\dots,0\}$, for $x \in \{0,1,\dots,n\}^s$, $\Sigma x^\nu = n$ and

$y \in \{0,1,\dots,n+1\}^s$, $\Sigma y^\nu = n+1$.

Let $P_\theta = \mathcal{L}_\theta(X_n)$, $Q_\theta = \mathcal{L}_\theta(X_{n+1})$. Then $P_\theta M$ has density

$$f_\theta(y) = \sum_x m(y|x) P_\theta(x) = \sum_{y=x+e^\nu} m(y|x) P_\theta(x)$$

$$= \sum_{\nu : y^\nu \neq 0} \frac{y^\nu - 1}{n} \frac{n!}{y^1! \dots (y^\nu - 1)! \dots y^s!} (\theta^1)^{y^1} \dots (\theta^\nu)^{y^\nu - 1} \dots (\theta^s)^{y_s}$$

Then (q is the density of Q)

$$\|P_\theta M - Q_\theta\| = \sum_y |f_\theta(y) - q_\theta(y)|$$

$$= \sum_y \left| \frac{f_\theta(y)}{q_\theta(y)} - 1 \right| q_\theta(y)$$

$$= E_{Q_\theta} \left| 1 - \sum_{\nu : \theta^\nu Y^\nu \neq 0} \frac{Y^\nu}{(n+1)\theta^\nu} \circ \frac{Y^\nu - 1}{n} \right| = E_{Q_\theta} \left| \sum_{\theta^\nu \neq 0} \frac{Y^\nu}{n+1} \left(1 - \frac{Y^\nu - 1}{n\theta^\nu}\right) \right|$$

$$\leq E_{Q_\theta}\Big|\sum_{\theta^\nu \neq o} \frac{Y^\nu}{n+1}\Big(1 - \frac{Y^\nu}{\theta^\nu(n+1)}\Big)\Big|$$

$$+ E_{Q_\theta}\Big|\sum_{\theta^\nu \neq o} \frac{Y^\nu}{(n+1)\theta^\nu}\Big(\frac{Y^\nu}{n+1} - \frac{Y^\nu - 1}{n}\Big)\Big|.$$

The last membrum is

$$E_{Q_\theta}\Big|\sum_{\theta^\nu \neq o} \frac{Y^\nu(n+1-Y^\nu)}{\theta^\nu n(n+1)^2}\Big| = \sum_{\theta^\nu \neq o} E_{Q_\theta} \frac{Y^\nu(n+1-Y^\nu)}{n(n+1)^2\theta^\nu}$$

$$= \sum_{\theta^\nu \neq o} \frac{1-\theta^\nu}{n+1} \leq \frac{s-1}{n+1}.$$

The first membrum is

$$E_{Q_\theta}\Big|\sum_{\theta^\nu \neq o} \Big(\frac{Y^\nu}{n+1} - \theta^\nu\Big)\Big(1 - \frac{Y^\nu}{\theta^\nu(n+1)}\Big)\Big|$$

$$\leq \sum_{\theta^\nu \neq o} E_{Q_\theta}\Big|\frac{Y}{n+1} - \theta^\nu\Big|\Big|1 - \frac{Y^\nu}{\theta^\nu(n+1)}\Big|$$

$$\leq \sum_{\theta^\nu \neq o} \Big[E_{Q_\theta}(\circ)^2 E_{Q_\theta}(\cdot)^2\Big]^{\frac{1}{2}}$$

$$= \sum_{\theta^\nu \neq o} \Big[\frac{\theta^\nu(1-\theta^\nu)}{n+1} \cdot \frac{1-\theta^\nu}{(n+1)\theta^\nu}\Big]^{\frac{1}{2}} = \sum_{\theta^\nu \neq o} \frac{1-\theta^\nu}{n+1} \leq \frac{s-1}{n+1}.$$

It follows that

$$\delta(\mathcal{E}^n, \mathcal{E}^{n+1}) \leq 2\,\frac{s-1}{n+1}.$$

This must also hold for all experiments $\mathcal{E}$ where the σ-algebra has at most 2^s elements. One may attempt to approximate more general experiments by multinomial ones in order to extend these results. However, we have the following:

Example 1.2. (Continued.)

$\mathcal{E}|_{\Theta_m}$ has a sufficient σ-algebra $\tilde{\mathcal{B}}$ generated by the partition $B_m = \{[0,1/2^m>,[1/_2m,2/_2m>,\ldots\}$ since $p_\theta(x)$ only depends on x through $I_{[0,1/_2m>},\ldots$. Then $\mathrm{card}(\tilde{\mathcal{B}}) = 2^{2^m}$, so that

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

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

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

The above calculations were first carried out in the binomial case, and Torgersen noted the validity in the general case.

3. 1-parameter exponential distributions

3.1. An upper bound for $\delta(\mathcal{E}^n,\mathcal{E}^{n+1})$ in a general case.

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

Formally, let us assume: $\{P_\theta\}$ homogeneous, and $\mathcal{B}$ contains all the singletons $\{x\}$, $x \in \chi$, and there exists a $\tilde{f} : \chi^n \to L_1(\mu)$ so that the function

(*) $(x^1, \dots, x^n, y) \mapsto \tilde{f}(x^1, \dots, x^n)(y)$

is simultaneously measurable and

$\int \tilde{f}(x)(y)d\mu(y) = 1$ for all $x \in \chi^n$.

Define the following Markov kernels

$$\tilde{M}^r_n : \mathcal{B}^{n+1} \times \chi^n \to [0,1] \; ; \; \tilde{M}^r_n(\circ|x) = \delta_{x^1} \times \dots \times \delta_{x^{r-1}} \times \tilde{M}(\circ|x) \times \delta_{x^r} \times \dots \times \delta_{x^n}$$

where δ_x is the one-point (Dirac) measure in x, and $\tilde{M}(A|x) = \int_A \tilde{f}(x)(y)d\mu(y)$. We see that $\tilde{M}^r_n(\chi^n|x) = 1$, $\forall x$ and that for all $A \in \mathcal{B}^{n+1}$

$$\tilde{M}^r_n(A|x) = \int I_A(x^1, \dots, x^{r-1}, y, x^r, \dots, x^n) f(x)(y)d\mu(y)$$

which is measurable in x by the Torelli theorem. Put $M_n = \frac{1}{n+1} \sum_{r=1}^{n+1} \tilde{M}^r_n$ (obviously a Markov kernel). When $R \in \mathcal{B}^{n+1}$ is a rectangle, then (π^i is the i-th projection)

$$P^n_\theta \tilde{M}^r_n(R) = \int_{\chi^n} \delta_{x^1}(\pi^1 R) \dots \left(\int_{\pi^r R} \tilde{f}(x)(y)d\mu(y) \right) \dots \delta_{x^n}(\pi^{n+1} R)dP^n_\theta(x)$$

$$= \int_R f_\theta(y^1) \dots \tilde{f}(y^1 \dots \overset{\vee}{y^r} \dots y^{n+1})(y^r) \dots f_\theta(y^{n+1})d\mu^{n+1}$$

by Tonelli's theorem. It follows immediately that

$$\frac{dP^n_\theta M_n}{d\mu^{n+1}}(y) = \frac{1}{n+1} \sum_{r=1}^{n+1} \frac{\tilde{f}(y^1 \dots \overset{\vee}{y^r} \dots y^{n+1})(y^r)}{f_\theta(y^r)} \prod_1^{n+1} f_\theta(y^i)$$

and that

$$\| P^n_\theta M_n - P^{n+1}_\theta \| = E_\theta \left| \frac{1}{n+1} \sum_1^{n+1} \frac{\tilde{f}(Y^1 \dots \overset{\vee}{Y^r} \dots Y^{n+1})(Y^r)}{f_\theta(Y^r)} - 1 \right|$$

where the expectation is taken w.r.t. P_θ^{n+1}. By the Markov kernel criterion, we now get:

<u>Lemma 3.1.1</u>. If $\mathcal{E}$ is an experiment satisfying condition (*), then

$$\delta(\mathcal{E}^n,\mathcal{E}^{n+1}) \leq \sup_{\theta\in\Theta} E_\theta\left|\frac{1}{n+1}\sum_1^{n+1}\frac{\tilde{f}(Y^1,\ldots,\overset{\vee}{Y}{}^r,\ldots,Y^{n+1})(Y^r)}{f_\theta(Y^r)}-1\right|$$

where the Y^i are i.i.d. $\sim P_\theta$.

<u>3.2</u> Upper bound for $\delta(\mathcal{E}^n,\mathcal{E}^{n+1})$ when $\{P_\theta\}$ is an 1-parameter exponential family.

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

$$\frac{dP_\theta}{d\mu} = A(\theta)e^{\theta T}\,h \tag{1}$$

where μ is some σ-finite measure on $(\mathcal{X},\mathcal{A})$, T and $h\geq 0$ random variables and $A:\Theta\to\mathbb{R}$. The set of θ's such that (1) defines, for a suitable A, a probability measure, is the natural parameter space of $\{P_\theta\}$, and this is an interval I. In the interior of I, I°, A is analytic. For all θ, $A(\theta)>0$, and we can without loss of generality assume $0\in I$ and write

$$\frac{dP_\theta}{dP_0} = e^{c(\theta)+\theta T}\quad,\quad 0\in\Theta.$$

We can now formulate the following result:

<u>Proposition 3.2.1</u> Let $\mathcal{E} = ((\mathcal{X},\mathcal{A}),(P_\theta:\theta\in\Theta))$ where

$$\frac{dP_\theta}{dP_{\theta_0}} = e^{c(\theta)+\theta T}\quad,\ \theta\in\Theta\subset\mathbb{R}.$$

Let Θ be a bounded set, and assume that an endpoint θ_1 of the natural parameter set is a limit point of Θ only if c has continuous one-sided derivatives up to 4. order in θ_1, and $c''(\theta_0)\neq 0$. Then

$$\overline{\lim}\ n\,\delta(\mathcal{E}^n,\mathcal{E}^{n+1}) \leq 2.$$

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

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

(ii) $X \sim P_0(\lambda)$, $\lambda \in \Lambda$ where Λ is bounded away from 0 and ∞.

(iii) $X \sim N(\xi,\sigma)$, with σ known, $\xi \in \Theta$ which is bounded. The exact deficiency is [9],

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

and this holds even for unbounded Θ. It is seen that our method gives a bound that is 4 times too large, but with correct rate, and we have to assume an unnecessary boundedness condition for Θ.

Proof of the proposition: We may assume that Θ is a compact interval. Furthermore, T is sufficient for $\mathcal{E}$, so if $\tilde{\mathcal{E}}$ consist in observation of T, then $\delta(\tilde{\mathcal{E}}^n,\tilde{\mathcal{E}}^{n+1}) = \delta(\mathcal{E}^n,\mathcal{E}^{n+1})$. We can accordingly assume that $(\chi,A) = (\mathbb{R},\mathcal{B})$ and put $f_\theta(t) = \frac{dP_\theta}{dP_0}(t) = \exp(c(\theta)+\theta t)$, $\theta \in \Theta$. For $\theta \in I^o$ 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 $\mathcal{E} \sim \mathcal{M}_i$ (the totally non-informative experiment) and obviously $\mathcal{E}^n \sim \mathcal{E}^{n+1}$. Assume therefore that $c''(\theta) < 0$ for $\theta \in I^o$. If $I^o = \emptyset$, then Θ is just one point, so that $\mathcal{E}^n \sim \mathcal{E}^{n+1}$, so we may assume that $I^o \neq \emptyset$. In the course of the proof we shall have to construct an estimator for the unknown parameter, and to this end it is convenient to reparametrize the experiment as follows: Define $\xi : I^o \to \mathbb{R}$ by $\xi(\theta) = -c'(\theta) = E_\theta T$. Then ξ is a diffeomorphism from I^o onto its image J^o, 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 $\mathcal{E}$ as an experiment over N where N is the image of Θ under ξ and thus a compact interval. Put $\tau_1 = \xi^{-1}$, $\tau_o = c \circ \xi^{-1}$, defined on an open interval J' such that $N \subset J'$. We can thus assume that $\mathcal{E}$ is given by the densities

$$f_\xi(t) = \frac{dP_\xi(t)}{dP_{\xi_0}(t)} = e^{\tau_0(\xi)+\tau_1(\xi)t} \quad , \ \xi \in N$$

w.r.t. Lebesgue measure.

For $\xi \in J^0$, $E_\xi T = \xi = - \frac{\tau_0'(\xi)}{\tau_1'(\xi)}$

and $\operatorname{var}_\xi T = -c''(\tau_1(\xi)) = \frac{(-c'(\tau_1(\xi)))'}{\tau_1'(\xi)} = \frac{1}{\tau_1'(\xi)}$.

τ_0 and τ_1 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_1 and τ_0 must have continous 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 $E_{\theta_0} T^4 \leq \underline{\lim}_{\theta \to \theta_0} E_{\theta_0} T^4 < \infty$): $E_\theta |T|^r$ is bounded when $\theta \to \theta_0$ for $r \leq 4$. Since $\theta \mapsto |T|^r e^{\theta T}$ is convex in θ, we have for θ between θ_0 and θ_1, $\theta_1 \in I^0$,

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

It follows from Lebesgues' dominated convergence theorem that $\int |T|^r e^{\theta T} dP_0 \underset{\theta \to \theta_0}{\longrightarrow} \int |T|^r e^{\theta_0 T} dP_0$ which entails that $E_\theta |T|^r$ is continous in θ_o for $r \leq 4$.

Let $T_n = (T_n^1, \ldots, T_n^n)$ be the observations from $\mathcal{E}^n$. Then $\hat{\xi}_n = \bar{T}_n$ is a reasonable estimator for ξ, and $E_\xi \hat{\xi}_n = \xi$

$\mathrm{var}_\xi \hat{\xi}_n = \dfrac{1}{n\tau_1'(\xi)}$ for all $\xi \in N$. Now put, if $N = [a,b]$,

$$\tilde{\xi}_n = \begin{cases} \hat{\xi}_n & , \hat{\xi}_n \in N \\ a & , \hat{\xi}_n < a \\ b & , \hat{\xi}_n > b \end{cases} .$$

We will now use lemma 3.1.1 and put $f(t^1,\dots,t^n)(t) = f_{\hat{\xi}_n}(t)$, which obviously is measurable in $(t^1,\dots,t^n,t)$. Let

$$\phi_\xi(t) = (\ln f_\xi)'(t) = \tau'(\xi)(t-\xi)$$

$$\tilde{\phi}_\xi(t) = f''_\xi(t)/f_\xi(t) .$$

If $\xi, \xi+\Delta \in N$, so

$$\frac{f_{\xi+\Delta}-f_\xi}{f_\xi} = \Delta\phi_\xi + \tfrac{1}{2}\Delta^2\tilde{\phi}_\xi + \Delta^3 B_{\xi,\Delta}$$

where $B_{\xi,\Delta} = \dfrac{1}{6}\dfrac{f^{(3)}_{\xi'}}{f_\xi}$ for some ξ' between ξ and $\xi+\Delta$.

We see that $\hat{\xi}(t^1,\dots,t^n) = \xi + \dfrac{1}{n}\sum_1^n \dfrac{\phi_\xi(t^i)}{\tau_1'(\xi)}$.

Put

$$\hat{\xi}_n(T^1_{n+1},\dots,\check{T}^i_{n+1},\dots,T^{n+1}_{n+1}) = \hat{\xi}^i_n$$

$$\tilde{\xi}_n(T^1_{n+1},\dots,\check{T}^i_{n+1},\dots,T^{n+1}_{n+1}) = \tilde{\xi}^i_n$$

$$\frac{1}{n}\sum_{\substack{j\neq i\\ j=1,\dots,n+1}} \phi_\xi(T^j_{n+1}) = \bar{\phi}^i_{n,\xi}$$

$\Delta^i_n = \hat{\xi}^i_n - \xi = \dfrac{1}{\tau_1'(\xi)}\,\bar{\phi}^i_{n,\xi}$, and let $\varepsilon > 0$.

Then the expression from 3.1.1 becomes

$$E_\xi\left|\frac{1}{n+1}\sum_1^{n+1}\frac{\tilde{f}(T^1_{n+1},\dots,\check{T}^i_{n+1},\dots,T^{n+1}_{n+1})(T^i)}{f_\xi(T^i)} - 1\right| .$$

Let $N_\varepsilon = N \cap \langle \xi-\varepsilon, \xi+\varepsilon \rangle$

$$= E_\xi \left| \frac{1}{n+1} \sum_1^{n+1} \frac{f_{\tilde{\xi}_n^i}(T_{n+1}^i) - f_\xi(T_{n+1}^i)}{f_\xi(T_{n+1}^i)} I_{N_\varepsilon}(\hat{\xi}_n^i) \right.$$

$$\left. + \frac{1}{n+1} \sum_1^{n+1} \left(\frac{f_{\tilde{\xi}_n^i}(T_{n+1}^i)}{f_\xi(T_{n+1}^i)} \right) I_{N_\varepsilon^c}(\hat{\xi}_n^i) \right|$$

In the first membrum we can replace $\hat{\xi}$ by $\tilde{\xi}$, and we supress the index ξ:

$$\leq E \left| \frac{1}{n+1} \sum_1^{n+1} \left[\phi(T_{n+1}^i) \frac{\bar{\phi}_n^i}{\tau_1'} + \frac{1}{n} \tilde{\phi}(T_{n+1}^i) \left(\frac{\bar{\phi}_n^i}{\tau_1'} \right)^2 \right] \right|$$

$$+ E \left| \frac{1}{n+1} \sum_1^{n+1} B_{\Delta_n^i}(T_{n+1}^i)(\Delta_n^i)^3 I_{N_\varepsilon}(\hat{\xi}_n^i) \right|$$

$$+ E \left| \frac{1}{n+1} \sum_1^{n+1} \left(\frac{f_{\tilde{\xi}_n'}(T_{n+1}^i)}{f(T_{n+1}^i)} + 1 \right) I_{N^c}(\hat{\xi}_n^i) \right| = A^1 + A^2 + A^3 .$$

$$A^3 = E I_{N_\varepsilon^c}(\hat{\xi}_n^i) + E \frac{f_{\tilde{\xi}_n^i}(T_{n+1}^i)}{f(T_{n+1}^i)} I_{N_\varepsilon^c}(\hat{\xi}_n^i)$$

$$= P(\hat{\xi}_n^i \notin N_\varepsilon) + \int_{\hat{\xi}_n^i \notin N_\varepsilon} f_{\tilde{\xi}_n^i}(t^i) \prod_{j \neq i} f(t^j) dP^{n+1}$$

$$= 2P(\hat{\xi}_n^i \notin N_\varepsilon) \leq 2P(|\Delta_n^i| \geq \varepsilon) \leq 2 \frac{E|\Delta_n^i|^4}{\varepsilon^4} \leq \frac{2}{\varepsilon^4} \frac{1}{n^3} E|T-\xi|^4 .$$

Since $E_\xi |T|^4$ is continuous and N is compact, $\sup_\xi n A_\xi^3 \underset{n\to\infty}{\longrightarrow} 0$

$$n A^2 \leq nE|B_{\Delta_n^i}(T_{n+1}^i)| |\Delta_n^i|^3 I_{\langle -\varepsilon, \varepsilon \rangle}(|\Delta_n^i|) .$$

We have

$$|f_{\xi'}'''| = f_{\xi'} |\phi_{\xi'}^3 + 3\phi_{\xi'}\phi_{\xi'}' + \phi_{\xi'}''| \quad , \quad \xi' \in N_\varepsilon .$$

Since τ_1 is (1-1) and $\theta \mapsto e^{\theta T}$ is convex, we have

$$e^{\tau_1(\xi)T} \leq e^{\tau_1(\xi_1)T} + e^{\tau_1(\xi_2)T}$$

where

$$\xi' \in [\xi_1, \xi_2] = N_\varepsilon .$$

Since ϕ, ϕ' and ϕ'' are linear in T with continuous coefficients, the second factor above is bounded by

$$M(|T|^3 + |T|^2 + |T| + 1),$$

for all choices of $\xi \in N$.

If we put $H_\xi = M \dfrac{e^{\tau_1(\xi_1)T} + e^{\tau_1(\xi_2)T}}{e^{\tau_1(\xi)T + \tau_0(\xi)}} (|T|^3 + |T|^2 + |T| + 1)$,

we see that

$$H_\xi \geq \frac{|f'''_{\xi'}|}{f_\xi} I_{N_\varepsilon}(\xi') \quad , \text{ and that}$$

$$E_\xi H_\xi \leq M'(E_{\xi_1}(|T|^3 + |T|^2 + 1) + E_{\xi_2}(|T|^3 + |T|^2 + |T| + 1))$$

which is bounded on N.

This implies that (H_ξ and Δ are independent)

$$nA_\xi^2 \leq nE_\xi H_\xi E_\xi |\Delta_n^i|^3 \leq \frac{1}{n} E_\xi H_\xi E_\xi |T-\xi|^3 \Rightarrow \sup_\xi nA_\xi^2 \to 0 .$$

The following will become useful when dealing with A^1 :

<u>Lemma 3.2.2.</u> (See [6], 11.4.A.)

If F_n, F are d.F.'s on $\mathbb{R}$ and $g : \mathbb{R} \to \mathbb{R}$ is continuous, $g \geq 0$, $F_n \overset{w}{\to} F$, then:

$$\int g dF_n \to \int g dF \Longleftrightarrow g \text{ uniformly integrable in } (F_n).$$

Lemma 3.2.3. (See [7], 5.2.1.)

Let k be a compact metric space, $f_n, f \in C(K)$. If f_n <u>converges continuously</u> to f (i.e. $x_n \to x \Rightarrow f_n(x_n) \to f(x)$) then $f_n \to f$ uniformly.

We now put

$$C_\xi^1 = E_\xi \left| \frac{1}{n+1} \sum_1^{n+1} \phi\,(T_{n+1}^i)\, \frac{\bar{\phi}_n^i}{\tau_1'} \right|$$

$$C_\xi^2 = E_\xi \left| \frac{1}{n+1} \sum_1^{n+1} \tilde{\phi}\,(T_{n+1}^i) \left(\frac{\bar{\phi}_n^i}{\tau_1'}\right)^2 \right|$$

$$C_\xi^1 = E_\xi \left| \frac{1}{n(n+1)\tau_1'} \left[\sum_1^{n+1} \phi(T_{n+1}^i)^2 - \sum_1^{n+1} \phi(T_{n+1}^i)^2 \right] \right|$$

$$\leq E|\circ| + E|\circ| = \frac{2}{n} E_\xi \phi(T_{n+1}^i)^2/\tau_1'$$

$$= \frac{2}{n}\,.$$

Let now $\xi_n \to \xi$. If we can show that $n\,c_{\xi_n}^2 \to 0$, it follows from lemma 3.2.3 that $\sup_\xi n\,c_\xi^2 \to 0$.

and since $A^1 \leq C^1 + C^2$, the proposition will be proved. We first show the following assertions:

(i) $$\frac{1}{\sqrt{n}} \sum_1^{n+1} \frac{1}{\tau_1'(\xi_n)} \phi_{\xi_n}(T_{n+1}^i) \xrightarrow{L_{\xi_n}} N\left(0, \frac{1}{\sqrt{\tau_1'(\xi)}}\right)$$

(ii) $$\frac{1}{n} \sum_1^{n+1} |\tilde{\phi}_{\xi_n}(T_{n+1}^i)| \xrightarrow{L_{\xi_n}} E_\xi|\tilde{\phi}_\xi(T)|$$

(iii) $$\frac{1}{\sqrt{n}(n+1)} \sum_1^{n+1} |\tilde{\phi}_{\xi_n}(T_{n+1}^i)|\,\phi_{\xi_n}(T_{n+1}^i) \xrightarrow{L_{\xi_n}} 0$$

(iv) $$\frac{1}{n(n+1)} \sum_1^{n+1} |\tilde{\phi}_{\xi_n}(T_{n+1}^i)|\,\phi_{\xi_n}(T_{n+1}^i) \xrightarrow{L_{\xi_n}} 0.$$

In (iii) and (iv) we can replace $|\tilde{\phi}|$ by ϕ.

We recall that $\xi \mapsto E_\xi |T|^r$ is continuous, and therefore bounded, for $r \leq 4$.

Now $$\phi_{\xi_n}(T^i)/\tau_1'(\xi_n) = T^i - \xi_n$$

which has zero expectation and bounded 3. order moment, so (i) follows from Lyapunyov's theorem.

We have $\tilde{\phi}_\xi(t) = \phi_\xi^2(t) + \tau_1''(\xi)(t-\xi) - \tau_1'(\xi)$ so that

$|\tilde{\phi}_\xi| \leq (\tau_1'(\xi))^2(T^2-2\xi T+\xi^2) + |\tau''(\xi)|(|T|+|\xi|) + |\tau'(\xi)|$

which has continuous expectation under P_ξ. Since $\tilde{\phi}_{\xi_n} \to \tilde{\phi}_\xi$ pointwise, it follows from the (generalized) Lebesgue dominated convergence theorem, that $E_\xi|\tilde{\phi}|$ is also continuous. Also, $\text{var}_{\xi_n}|\tilde{\phi}_{\xi_n}|$ must be bounded, so that

$$\frac{1}{n}\sum_1^{n+1}(|\tilde{\phi}_{\xi_n}(\tau^i)| - E_{\xi_n}|\tilde{\phi}_{\xi_n}(T)|) \xrightarrow{P_{\xi_n}} 0$$ and (ii) is proved.

To prove (iii) and (iv), we note that the summands have bounded P_{ξ_n}-expectations, and by the general Markov inequality, we get for all $\varepsilon > 0$

$$P_{\xi_n}\left[\frac{1}{\sqrt{n(n+1)}}\sum_1^{n+1}\left|\tilde{\phi}_{\xi_n}(T^i)\right|(\phi_{\xi_n}(T^i))^r > \varepsilon\right] \leq$$

$$\frac{1}{\varepsilon}E_{\xi_n}(\cdot) \leq \frac{1}{\varepsilon}\frac{1}{\sqrt{n}}E_{\xi_n}\left|\tilde{\phi}_{\xi_n}(T)\right|\left|\phi_\xi(T)^r\right| \to 0$$

for $r=1,2$. Let

$$Y_n = \frac{n}{n+1}\sum_1^{n+1}\tilde{\phi}_{\xi_n}(T^i)\left(\frac{\phi^{\pm}_{n,\xi_n}}{\tau_1'(\xi_n)}\right)^2, \quad |Y_n| \leq \frac{n}{n+1}\sum_1^{n+1}|\tilde{\phi}(T^i)|\left(\frac{\bar{\phi}_n^i}{\tau_1'}\right)^2 = Z_n$$

$$= \left(\frac{1}{n+1}\sum_1^{n+1}|\tilde{\phi}(T^i)|\right)\left[\frac{1}{\sqrt{n}}\sum_1^{n+1}\frac{\phi(T^i)}{\tau_1'}\right]^2$$

$$- 2\left(\frac{1}{\sqrt{n}}\sum_1^{n+1}\frac{\phi(T^i)}{\tau_1'}\right)\frac{1}{\sqrt{n}(n+1)\tau_1'}\sum_1^{n+1}|\tilde{\phi}(T^i)|\phi(T^i)$$

$$+ \frac{1}{(\tau_1')^2}\frac{1}{n(n+1)}\sum_1^{n+1}|\tilde{\phi}(T^i)|\phi(T^i)^2 .$$

We see that $Z_n \xrightarrow{L_{\xi_n}} Z \cdot E_\xi|\tilde{\phi}_\xi|/\tau_1'(\xi)$ where $Z \sim \chi_1^2$. Now

$$E\,Z_n = E|\tilde{\phi}(T^i)|E\left(\frac{1}{\sqrt{n}}\sum_{j\neq i}\frac{\phi(T^i)}{\tau_1'}\right)^2$$

$$= E\,\frac{E_{\xi_n}|\tilde{\phi}_{\xi_n}(T^i)|}{\tau_1'(\xi_n)} \underset{n\to\infty}{\longrightarrow} E\,Z\,\frac{E_\xi|\phi_\xi|}{\tau_1'(\xi)}$$

so that Z_n is uniformly integrable in P_{ξ_n}. This must also hold for $|Y_n|$, and since $E_{\xi_n}\tilde{\phi}_{\xi_n}(T) = 0$, we must have

$$\frac{1}{n+1}\sum_1^{n+1}\tilde{\phi}_{\xi_n}(T^i) \xrightarrow{P_{\xi_n}} 0 \Rightarrow Y_n \xrightarrow{P_{\xi_n}} 0 \Rightarrow E_{\xi_n}Y_n = n\,c_{\xi_n}^2 \to 0.$$

Remark: A trivial corollary is that under the conditions in proposition 3.2.1,

$$\overline{\lim_{n\to\infty}}\ \frac{n}{r}\,\delta(E^n,E^{n+r}) \leq 2$$

for fixed $r \geq 1$.

3.3 Lower bounds for $\delta(E^n,E^{n+1})$.

Let E,F be experiments over Θ, and let λ be a prior distribution on Θ. Under certain regularity conditions we may interpret $\delta(E,F)$ as the maximal difference in achievable Bayes-risk. In this situation there is another way of "measuring" the "information content" of an experiment; we examine the posterior distributions, and an experiment that gives "concentrated" posterior distributions must obviously be an informative one.

Let us define:

If μ is a measure on $(\mathbb{R},\mathcal{B})$ then the concentration function (see [2]) is

$$Q_\mu(\ell) = \begin{cases} \sup\limits_{x\in\mathbb{R}} \mu[x-\frac{\ell}{2},x+\frac{\ell}{2}] & ; \quad \ell \geq 0 \\ 0 & ; \quad \ell < 0 \end{cases}$$

i.e. $Q_\mu(\ell)$ is the "maximal concentration of μ on a closed interval of length ℓ". According to ([2], 1.1.4 and 1.1.5) Q_μ is a right-continuous distribution function and the supremum is achieved, in say $x_o(\ell)$. Now choose $\ell_n \downarrow \ell$ and $r_n \in \mathbb{Q}$, $r_n \to x_o(\ell)$ such that $\bigcap_n [r_n - \frac{\ell_n}{2}, r_n + \frac{\ell_n}{2}] = [x_o(\ell) - \frac{\ell}{2}, x_o(\ell) + \frac{\ell}{2}]$.

Then

$$Q_\mu(\ell) = \mu[x_o(\ell) - \tfrac{\ell}{2}, x_o(\ell) + \tfrac{\ell}{2}] = \lim \mu[r_n - \tfrac{\ell_n}{2}, r_n + \tfrac{\ell_n}{2}]$$

$$\leq \lim \tilde{Q}_\mu(\ell_n) \leq \lim Q_\mu(\ell_n) = Q_\mu(\ell) \quad \text{where}$$

$$\tilde{Q}_\mu(\ell) = \sup_{r \in Q} \mu[r - \tfrac{\ell}{2}, r + \tfrac{\ell}{2}].$$

If now $\mu(\circ|x)$ is a (X,A)-measurable Borel probability measure, then for a fixed $\ell > 0$, $Q_{\mu(\circ|x)}(\ell) = \lim \tilde{Q}_{\mu(\circ|x)}(\ell_n)$ which is A-measurable since $\tilde{Q}_{\mu(\circ|x)}(\ell_n)$ must be.

Let $E = (X, A, P_\theta : \theta \in \Theta)$ where $\Theta \in \mathcal{B}$, and all $\theta \mapsto P_\theta(A)$ measurable. Let the decision space (T, S) be closed intervals of length ℓ (with the obvious σ-algebra induced from $\mathbb{R}^2$). Let the loss-function be

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

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

This is seen as follows:

Let ρ be a decision-rule. We can, according to ([6], 27.2.B) specify $\lambda(\circ|x)$ as a A-measurable measure over Θ, where

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

but

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

so that

$$b_\lambda = \int (1 - 2Q_{\lambda(\circ|x)}(\ell))\lambda P(dx).$$

3.4. Lower bound for $\delta(\mathcal{E}^n,\mathcal{E}^{n+r})$ when $\mathcal{E}$ is a 1-parameter exponential experiment

In this section we will use posterior concentration functions to prove:

Proposition 3.4.1. Let $\mathcal{E} = ((X,A),P_\theta : \theta \in \Theta)$ where

$$\frac{dP_\theta}{d\mu}(x) = e^{c(\theta)+\theta T(x)}h(x) \quad ; \quad \theta \in \Theta \subset \mathbb{R}$$

(for suitable σ-finite μ, $h \geq 0$ and T random variables) and Θ contains a non-degenerate interval. If θ is identifiable (i.e. T is not a.s. constant) and $r_n \leq n^\beta$, $0 < \beta < \frac{1}{2}$, then

$$\underline{\lim}_{n\to\infty} \frac{n}{r_n}\, \delta(\mathcal{E}^n,\mathcal{E}^{n+r_n}) \geq \sqrt{\frac{2}{\pi e}} \ .$$

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

An immediate corollary of proposition 3.2.1 is

Corollary 3.4.2. If, in addition to the conditions of proposition 3.4.1, $\Theta \subset K \subset I^o$ where K is a compact, I the natural parameter space of $(P_\theta)_{\theta\subset\Theta}$, θ identifiable, then

$$\sqrt{\frac{2}{\pi e}}\,(1+o(1)) \leq n\,\delta(\mathcal{E}^n,\mathcal{E}^{n+1}) \leq 2(1+o(1)).$$

Examples:

(i) If $\mathcal{E}^n$ consists in observing $X \sim \text{bin}(n,p) : p \in [0,1]$, we have

$$\sqrt{\frac{2}{\pi e}}\,\frac{1}{n}(1+o(1)) \leq \delta(\mathcal{E}^n,\mathcal{E}^{n+1}) \leq \frac{2}{n}.$$

(ii) If $\mathcal{E}$ consists in observing $X \sim N(\xi,1)$, $\xi \in \Theta$ which has non-empty interior, then

$$\delta(\mathcal{E}^n,\mathcal{E}^{n+1}) \simeq \sqrt{\frac{2}{\pi e}}\,\frac{1}{n}(1+o(1))$$

(here Θ may be unbounded).

Proof of the proposition: If θ is non-identifiable, then E^n is the totally non-informative experiment, so that $E^n \sim E^{n+1}$. In the other case we can assume without loss of generality that $0 \in \Theta^o$.

Then
$$\frac{dP^m_\theta}{dP^m_o} = e^{m(c(\theta)-c(0))+\theta \sum_1^m T_i} .$$

Introduce the new parameter h by
$$\theta = \frac{h}{\sqrt{n}} .$$

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

where $\Delta(\frac{h}{\sqrt{n}}) = \frac{1}{6} c''(\tilde{\theta})(\frac{h}{\sqrt{n}})^3$ for sufficiently small $|\frac{h}{\sqrt{n}}|$, for some $\tilde{\theta}$ between θ_o and θ,):

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

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

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

where $c_n = cn^q$; $c > 0$ and $0 < q < \frac{1}{6}$, and such that $\frac{1}{\sqrt{n}}[-c_n, c_n] \subset \Theta^o$ for all $n \geq N$ for some N. It is easy to see that the posterior distribution

$H_n(\ell | X_m)$ in E^m (where $X_m = (X^1_m, \ldots, X^m_m)$) is given by

$$c_{nm}(X_m) \int_{-c_n}^{\ell} \exp\left\{-\frac{(h-\mu_{mn})^2}{2\sigma^2_{mn}} + (m-n)\Delta(\frac{h}{\sqrt{n}})\right\} dh$$

for $|t| \leq c_n$, where

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

$$\mu_{mn} = \sigma^2_{mn} \frac{1}{\sqrt{n}} \sum_1^m (T_i - \xi), \quad \xi = -c'(0) .$$

Let (for fixed X_m) $f_m, g_m \in L_1([-c_n, c_n])$ be

$$f_m(h) = \exp\left\{-\frac{(h-\mu_{mn})^2}{2\sigma^2_{mn}}\right\}$$

$$g_m(h) = \exp\left\{ (m-n)\Delta(\tfrac{h}{\sqrt{n}})\right\}.$$

Let $\|\circ\|$ denote the L_1-norm. Then $\|f_m\|$, $\|g_m\| > 0$ and

$$\left\| \frac{f_m}{\|f_m\|} - \frac{f_m g_m}{\|f_m g_m\|} \right\| \leq 2 \frac{\|f_m - f_m g_m\|}{\|f_m\| \vee \|f_m g_m\|} \leq 2 \int_{-c_n}^{c_n} |g_m - 1| \frac{f_m}{\|f_m\|} .$$

This is seen as follows: Assume first that $\|f_m\| \geq \|f_m g_m\|$. Then

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

$$\leq \frac{\|f-fg\|}{\|f\|} + \|fg\| \left|\frac{1}{\|f\|} - \frac{1}{\|fg\|}\right| \leq \frac{\|f-fg\|}{\|f\|} + \left|\frac{\|fg\| - \|f\|}{\|f\|}\right|$$

$$\leq \frac{\|f-fg\|}{\|f\|} + \frac{\|fg-g\|}{\|f\|} .$$

The case $\|f_m\| < \|f_m g_m\|$ is treated in the same way. Furthermore,

$$\frac{\|f-fg\|}{\|f\| \vee \|fg\|} \leq \frac{\|f-fg\|}{\|f\|} = \int \frac{f|1-g|}{\int f} .$$

The above inequality entails that the difference between the distribution functions

$$H_n(k|X_m) \quad \text{and} \quad F_n(k|X_m) = \int_{-c_n}^{k} \frac{f_m}{\|f_m\|} ,$$

is at most

$$2\int_{-c_n}^{c_n} \left| e^{(m-n)\Delta(\frac{h}{\sqrt{n}})} - 1 \right| dF_n(h|X_m)$$

$$\leq 2e^{\delta(\frac{c_n}{\sqrt{n}})^3} \int_{-c_n}^{c_n} \delta \left|\frac{h^3}{n\sqrt{n}}\right| dF_n(h|X_m)$$

$$\leq (\frac{c_n}{\sqrt{n}})^3 (2\delta e^{\delta(\frac{c_n}{\sqrt{n}})^3})$$

where $\delta = \frac{1}{6} \sup_{|t| \leq \frac{c_n}{\sqrt{n}}} |c'''(t)| |m-n| < \infty$.

Let $Q_m(\circ|X_m)$, $Q_m'(\circ|X_m)$ be the concentration functions of $H_n(\circ|X_m)$ and $F_n(\circ|X_m)$. Then, for all ℓ and X_m:

$$|Q_m'(\ell|X_m) - Q_m(\ell|X_m)| \leq$$

$$2 \sup_{|t| \leq c_n} |H_n(t|X_m) - F_n(t|X_m)| \leq (\frac{c_n}{\sqrt{n}})^3 K_n$$

where $K_n = 2\delta e^{\delta(\frac{c_n}{\sqrt{n}})^3}$. Now obviously $\delta \leq D n^{\beta}$, so that

$$\delta(\frac{c_n}{\sqrt{n}})^3 = O(1)\ n^{3q+\beta-\frac{3}{2}} .$$

We see that by choosing q suitably small, we get

$$3q + \beta - \frac{3}{2} < -1 \Rightarrow \delta(\frac{c_n}{\sqrt{n}})^3 = o(\frac{1}{n}) \Rightarrow K_n = o(\frac{1}{n})$$

It is easy to see that $F_n(\circ|X_m)$ will achieve maximal concentration over closed intervals of length 2ℓ in the interval J_m, where

$$J_m = \begin{cases} [c_n - 2\ell, c_n] , & \text{if } \mu_{mn} + \ell > c_n \\ [-c_n, -c_n + 2\ell] , & \text{if } \mu_{mn} < -c_n + \ell \\ \mu_{mn} + [-\ell, \ell] , & \text{otherwise.} \end{cases}$$

By substituting $\ell = \dfrac{h-\mu_{mn}}{\sigma_{mn}}$ we obtain

$$Q_m'(2\ell \mid X_m) = \int_{J'_m} \phi \Big/ \int_{I_m} \phi \quad , \text{ with}$$

$$\phi(\ell) = \frac{1}{\sqrt{2\pi}} \exp(-\frac{\ell^2}{2}) \quad , \text{ and}$$

$$J_m' = \frac{1}{\sigma_{mn}} (J_m - \mu_{mn})$$

$$I_m = \frac{1}{\sigma_{mn}} ([-c_n, c_n] - \mu_{mn}).$$

Let $Y_{mn} = \begin{cases} 1, & |\mu_{mn}|, \ |\mu_{nn}| \leq c_n/2 \\ 0 & \text{otherwise.} \end{cases}$

For sufficiently large n, $Y_{mn} = 1$ must entail $Q_m'(2\ell \mid X_m) - Q_n'(2\ell \mid X_n) \geq 0$, so that

$$E_{\lambda_n p_m}(Q_m'(2\ell \mid X_m) - Q_n'(2\ell \mid X_n)) \geq E(\circ)Y_{mn} - E(\circ)^{-}(1-Y_{mn}),$$

and

$$E(\circ)Y_{mn} \geq E\left(\int_{J'_m}\phi - \int_{J'_{nn}}\phi \Big/ \int_{I_n}\phi \right) \quad Y_{mn} \ .$$

When $Y_{mn} = 1$, we have

$$\frac{1}{\int_{I_n}\phi} \leq \frac{1}{2\Phi(\frac{c_n}{2\sigma_{mn}})-1} = 1 + o(\frac{1}{n^2})$$

where $\Phi(x) = \int_{-\infty}^{x} \phi$.

This is because Φ has moments of arbitrary order, so that $x^r(1-\Phi(x)) \xrightarrow[x\to\infty]{} 0, \forall r$, and

$$n^2\left(\frac{1}{2\Phi(\frac{c_n}{2\sigma_{mn}})-1} -1\right) = 2\,\frac{n^2(1-\Phi(\frac{c_n}{2\sigma_{mn}}))}{2\Phi(\frac{c_n}{2\sigma_{mn}})-1},$$

but $c_n/2\sigma_{mn} \sim k\,n^q$, $q > 0$.

This implies

$$E(\circ)Y_{mn} \geq E\left(\int_{-\ell/\sigma_{mn}}^{\ell/\sigma_{mn}} \phi - (1+o(1/n^2))\int_{-\ell/\sigma_{mn}}^{\ell/\sigma_{mn}} \phi\right) Y_{mn}$$

It is easy to see that

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

where α_n is between $\frac{1}{\tau^2}\frac{m}{n} + \frac{1}{\kappa^2}$ and $\frac{1}{\tau^2} + \frac{1}{\kappa^2}$. Accordingly,

$$\int_{-\ell/\sigma_{mn}}^{\ell/\sigma_{mn}} \phi - \int_{-\ell/\sigma_{nn}}^{\ell/\sigma_{nn}} \phi = \frac{1}{\sqrt{\alpha_n}}\;\frac{\ell}{\tau^2}\;\frac{r}{n}\,\phi(\beta_n)\;, \text{ where}$$

$$\frac{\ell}{\sigma_{nn}} \leq \beta_n \leq \frac{\ell}{\sigma_{mn}}\;, \text{ and}$$

$$E(\circ)Y_{mn} \geq \frac{\ell}{\tau^2\sqrt{\alpha_n}}\;\frac{r}{n}\,\phi(\beta_n)\,\lambda_n P^m(Y_{mn}=1) + o(\frac{1}{n^2}).$$

Since $m \sim n$, $\sqrt{\alpha}_n \to \alpha = \frac{1}{\tau^2} + \frac{1}{\kappa^2}$ and $\beta_n \to \ell\sqrt{\alpha}$. For sufficiently large n, $\frac{1}{\sqrt{\alpha}} < \frac{c_n}{2}$, so that we may choose $\ell\sqrt{\alpha} = 1$, and obtain

$$\frac{n}{r}\left(\int_{-\ell/\sigma_{mn}}^{\ell/\sigma_{mn}} \phi - \int_{-\ell/\sigma_{nn}}^{\ell/\sigma_{nn}} \phi\right) \to \frac{1}{\sqrt{2\pi e}} \circ \frac{1}{1+\frac{\tau^2}{\kappa^2}}$$

if $\lambda_n P^m(Y_{mn}=1) \to 1$. Also, $E(\circ)^- Y_{mn} \leq \lambda_n P^m(Y_{mn} \neq 1)$.

From the remark on p.3, it follows that

$$\frac{1}{2}\,\frac{n}{r}\,\delta(E^n, E^{n+r}) \geq \frac{n}{r}\,E_{\lambda_n P^m}(Q_m(2\ell|X_m) - Q_n(2\ell|X_m))$$

$$\geq \frac{n}{r}\,E(Q_m'(\circ)-Q_n(\circ)) - \frac{n}{r}\,E|Q_m(\circ)-Q_m'(\circ)|$$

the last membrum is less than $K_n(\frac{c_n}{\sqrt{n}})^3 \frac{n}{r} = o(1)$ as $n \to \infty$.
If we can show that

$$\frac{n}{r}\lambda_n P^m(Y_{mn} \neq 1) \to 0, \quad \text{then}$$

$$\underline{\lim}\, \frac{n}{r}\,\delta(E^n, E^{n+r}) \geq \frac{1}{\sqrt{2\pi e}}\,\frac{1}{1+\frac{\tau^2}{\kappa^2}},$$

but since κ is arbitrary, the proposition follows.

We will now use the following result, which is a consequence of Chebychev's inequality (and also of the Chernoff root-theorem for large deviations (see [10])):

$$P((X_1+\ldots+X_n)/n>a)^{1/n} \leq \inf_{t\geq o} E_P\, e^{t(X-a)}$$

where $X_1,\ldots,X_n$ are i.i.d.
Put

$$X_i = \pm\sigma^2_{mn}(T^i_m - \xi), \quad a = a_n = \alpha\frac{c_n - \ell}{m}\sqrt{n}, \quad \alpha \in \langle 0,1\rangle.$$

Then

$$E_{P_h}\, e^{t(X-a)} = \int \exp\{(\pm t\sigma^2 + h/\sqrt{n})T \mp t\sigma^2\xi - ta + c(h/\sqrt{n}) - c(0)\}dP_o$$

$$= \exp\{c(h/\sqrt{n}) \mp t\xi\sigma^2 - ta - c(h/\sqrt{n} \pm t\sigma^2)\} \quad \text{whenever } h/\sqrt{n} \pm t\sigma^2 \text{ is small enough}$$

$= \exp f(a,t,h)$. Now

$$f(a,t,h) = -ta \pm t\,\frac{\sigma^2 h}{\tau^2\sqrt{n}} + t^2\frac{\sigma^4}{2\tau^2} + \Delta(h/\sqrt{n}) - \Delta(h/\sqrt{n} \pm t\sigma^2).$$

Now $a_n > |h|\,\frac{\sigma^2}{\tau^2\sqrt{n}}$ when

$$a_n = \lambda\,\frac{c_n - \ell}{m}\sqrt{n} > c_n\,\frac{\sigma^2_{mn}}{\tau^2\sqrt{n}} \iff \lambda(1-\frac{\ell}{c_n}) > \frac{1}{1+\frac{\tau^2}{\kappa^2}\frac{n}{m}}$$

The left side converges to λ, and the right to $\frac{1}{1+\frac{\tau^2}{\kappa^2}} < 1$, so the inequality holds for all large enough n, with

$$\frac{1}{1+\frac{\tau^2}{\kappa^2}} < \lambda < 1.$$

Put $t_n = \frac{\tau^2}{\sigma^4}(a_n \pm \frac{h}{\sqrt{n}}\frac{\sigma^2}{\tau^2})$. This minimizes the quadratic part of f and obviously t_n is eventually in $[0,t_o]$ for all $t_o > 0$ (uniformly in h). This implies that

$$\inf_{0 \leq t \leq t_o} f(a_n,t,h) \leq f(a_n,t_n,h)$$

$$= -\frac{\tau^2}{2\sigma^2}(a_n \pm \frac{h\sigma^2}{\sqrt{n}\tau^2}) + (\Delta(\frac{h}{\sqrt{n}}) - \Delta(\frac{h}{\sqrt{n}} + t_n\sigma^2)).$$

From the cited inequality it follows that, for $n \geq N$ which is independent of h ;

$$P_h^m(\pm\mu_{mn} > c_n - \ell) \leq \exp\{-mf(a_n,t_n,h)\} \Rightarrow \lambda_n P^m(\pm\mu_{mn} > c_n - \ell)$$

$$\leq \gamma_n \int_{-c_n}^{c_n} \exp\{(m-n)\Delta(\frac{h}{\sqrt{n}}) - \frac{h^2}{2\kappa^2} - m\Delta(\frac{h}{\sqrt{n}} + t_n\sigma_{mn}^2) - m\frac{\tau^2}{2\sigma^4}(a_n \mp \frac{\sigma^2}{\tau^2}\frac{h}{\sqrt{n}})^2\}dh$$

$$= \gamma_n \exp\{-\frac{1}{2\sigma^2}(\frac{m\tau^2 a^2}{\sigma^2} - \frac{a^2m^2}{n}) \circ \int_{-c_n}^{c_n} \exp\{-\frac{1}{2\sigma^2}(h \mp \frac{am}{\sqrt{n}})^2\}\exp C_n(h)dh$$

where

$$|C_n| \leq (m+r)K_n \left(\frac{c_n}{\sqrt{n}}\right)^3.$$

which is bounded. It is thus seen that the integral is bounded. Accordingly,

$$n\lambda_n P^m(\pm\mu_{mn} > c_n - \ell) \leq O(1) \circ \exp\{\ln n - \frac{a_n^2 m}{2\sigma_{mn}^2}(\frac{\tau^2}{\sigma_{mn}^2} - \frac{m}{n})\}.$$

Now

$$\frac{t^2}{\sigma^2_{mn}} - \frac{m}{n} \to \frac{\tau^2}{\kappa^2} > 0 \quad , \quad \sigma^2_{mn} \to \sigma^2 > 0 \quad , \text{ and}$$

$$m\,a^2_n = \lambda^2 \frac{(c_n - \ell)^2}{m} n \sim \lambda^2 c^2_n = \lambda^2 c^2 n^{2q} \quad , \quad q > 0$$

$$\Rightarrow \quad \ln n - \frac{a^2_n m}{2\sigma^2_{mn}} \left(\frac{\tau^2}{\sigma^2_{mn}} - \frac{m}{n}\right) \to -\infty$$

$$\Rightarrow \quad n\,\lambda_n P^m(|\mu_{mn}| > c_n - \ell) \to 0.$$

Q.E.D.

<u>Remark</u>: We might suspect that deficiencies are determined by decision problems of little practical interest, and that accordly they are unrealistic measures of "loss of information". Take as an example the experiments $\mathcal{E}^n$ consisting in observation of $X \sim \text{bin}(n,p)$, and let our problem be that of estimating p. For a quadratic loss function it is easily seen (see e.g. [1]) that the difference in minimaxrisk between $\mathcal{E}^n$ and $\mathcal{E}^{n+1}$ is $O(\frac{1}{n^2}) = o(\delta(\mathcal{E}^n,\mathcal{E}^{n+1}))$. However, if we use the loss function

$$L_\theta(\mathscr{k}) = \begin{cases} -1 & , \ |\mathscr{k}-\theta| \leq \frac{\ell}{\sqrt{n}} \\ 1 & , \ \text{otherwise} \end{cases}$$

we obtain the deficiency as difference in Bayes-risk (with the prior distribution being approximately $N(\theta_o, \frac{\kappa}{\sqrt{n}}))$, as follows from the above proof.

4. Some conjectures

As mentioned before, we may expect that $\delta(E^n, E^{n+1}) \sim \frac{c}{n}$ for a wide class of experiments 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.

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

$$\hat{\xi}-\xi = \frac{1}{\sqrt{n}} \sum_1^n \frac{\partial \ln f}{\partial \xi} (T_i)$$

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

The proof for the lower bound also essentially uses (i).

A case where we may expect to establish (i) and (ii) is when E is a general translation experiment.

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

References

1. Ferguson, R. (1967): "Mathematical Statistics", Academic Press, New York & London.
2. Hengartner, W. and Theodoresen, R. (1973): "Concentration Functions", Academic Press, New York & London.
3. LeCam, L. (1964): "Sufficiency and approximate sufficiency", Ann. Math. Stat. 35, p. 1419-1455.
4. LeCam,L. (1974): "On the information contained in additional observations". Ann. Stat. 2, p. 630-649.
5. Lehmann, E. (1959): "Testing statistical hypotheses". Wiley, New York.
6. Loeve, M. (1955): "Probability theory". Van Nostrand, Princeton.
7. Roussas, G. (1972): "Contiguity of probability measures". Univ. Press, Cambridge.
8. Sverdrup, E. (1971): "Testing statistical hypotheses", Statistical memoirs, Univ. of Oslo.
9. Torgersen, E. (1972): "Comparison of translation experiments". Ann. Math. Stat. 43, p. 1383-1399.
10. Torgersen, E. (1974): "Asymptotic behaviour of powers of dichotomies". Stat. res. rep., Univ. of Oslo.
11. Torgersen, E. (1976): "Deviations from total information and from total ignorance as measures of information". Stat. res. rep., Univ. of Oslo.
12. Torgersen, E. (1976): "Comparison of statistical experiments". Scand. Journ. of Stat. 3.
13. Le Cam, L. (1974): "Notes on asymptotic methods in statistical decision theory". Centre de Recherches Math., Univ. de Montréal.