The Annals of Statistics
1981, Vol. 9, No. 3, 638–657

MEASURES OF INFORMATION BASED ON COMPARISON WITH TOTAL INFORMATION AND WITH TOTAL IGNORANCE

BY ERIK N. TORGERSEN

University of California, Berkeley and University of Oslo

Two measures of content of information in statistical experiments are considered. They are both based on Le Cam's notion of deficiency of one experiment with respect to another.

The measures are:

(i) The deficiency of the given experiment with respect to a totally informative experiment.

(ii) The deficiency of a totally uninformative experiment with respect to the given experiment.

We shall here discuss the interpretations of such measures, establish inequalities for them and related quantities, and study their behaviour under replications.

If the parameter set is finite then closed expressions for exponential rates of convergence, as the number of replications increase, are given. In particular the exponential rate of the minimax probability of not covering the true values of the parameter by an r-point confidence set is expressed in terms of Hellinger transforms. If convergence to the totally informative experiment takes place at all, then the speed of convergence is necessary exponential. Examples are given indicating various possibilities.

1. Introduction. In this paper we shall consider two measures of the contents of information in statistical experiments. They are both based on Le Cam's [14] notion of a deficiency of one experiment with respect to another.

Let us agree, following Le Cam [14], to use the notation $\delta(\mathscr{E}, \mathscr{F})$ for the deficiency of an experiment $\mathscr{E}$ with respect to an experiment $\mathscr{F}$. This deficiency is defined for pairs $(\mathscr{E}, \mathscr{F})$ of experiments having the same parameter set Θ. $\delta(\mathscr{E}, \mathscr{F})$ is a function of two variables $\mathscr{E}$ and $\mathscr{F}$. It provides a partial answer to the question: What do we lose, in the sense of risk, by basing ourselves on $\mathscr{E}$ rather than on $\mathscr{F}$ under the least favorable conditions for this comparison?

The deficiency $\delta(\mathscr{E}, \mathscr{F})$ is monotonically increasing in $\mathscr{F}$ and monotonically decreasing in $\mathscr{E}$. It is also convex in each variable separately. Convexity is then defined in terms of mixtures of experiments [30].

Deficiencies, or the related distances, may be considered as measures of contents of information in many situations. It may be used, see Lindqvist [19], to measure the loss of memory of the initial State X_o in the tail $(X_t, X_{t+1}, \cdots)$ of a Markov chain $(X_o, X_1, \cdots)$. Swensen [22] has investigated the problem of measuring the value of a potential additional variable in a regression model. It is also possible to construct local measures of information based on deficiencies [24] but we shall not dwell on this here.

In order to investigate the properties of such measures it is tempting to consider, in spite of their artificiality, distances to experiments which are either totally informative or totally uninformative. A totally informative experiment is an experiment $(P_\theta; \theta \in \Theta)$ such that P_{θ_1} is P_{θ_2} singular when $\theta_1 \neq \theta_2$. As any two totally informative experiments are

Received June, 1977; revised April, 1978; July, 1980.

AMS 1970 *subject classifications.* Primary 62B15; secondary 62C05.

Key words and phrases. Risk, replications, exponential convergence, confidence sets, translation experiments.

equivalent, we shall use the symbol $\mathcal{M}_a$ to denote any of them. To fix ideas we might, if we so prefer, let $\mathcal{M}_a$ denote the experiment $(\delta_\theta; \theta \in \Theta)$ where δ_θ is the one point distribution in θ. Intuitively $\mathcal{M}_a$ is the experiment consisting of observing the underlying theory θ itself.

A totally uninformative experiment is an experiment $(P_\theta; \theta \in \Theta)$ where P_θ does not depend on θ. Clearly any two noninformative experiments are also equivalent and we shall reserve the notation $\mathcal{M}_i$ for any of them.

Any experiment $\mathcal{E}$ is obviously at least as informative as $\mathcal{M}_i$ and at most as informative as $\mathcal{M}_a$.

In this paper we shall consider the numbers $\delta(\mathcal{M}_i, \mathcal{E})$ and $\delta(\mathcal{E}, \mathcal{M}_a)$ as measures of the content of information in an experiment $\mathcal{E}$. The deficiency $\delta(\mathcal{M}_i, \mathcal{E})$ will usually be written $\delta_i(\mathcal{E})$ while the deficiency $\delta(\mathcal{E}, \mathcal{M}_a)$ will usually be written $\delta_a(\mathcal{E})$.

A small value of $\delta_a(\mathcal{E})$ suggests that an observation of $\mathcal{E}$, provided that it is properly used, is almost as good as knowing the unknown parameter. A large value, on the other hand, tells us that there are decision problems such that any decision procedure is risky for some of the underlying theories. A small value of $\delta_i(\mathcal{E})$ tells us that the chance mechanism governing the random outcome is almost independent of the various explanatory theories in Θ. If, on the other hand, this distance is large, then there are situations where an observation of $\mathcal{E}$ is helpful.

The values of these deficiencies are often large for all experiments $\mathcal{E}$ under consideration. This reflects the fact that it may be much too ambitious to compare with total information and much too modest to compare with no information.

Having made this limitation, we shall see that these deficiencies have interesting properties.

The contents of the paper are as follows. A few of the basic definitions and results from the theory of comparison of experiments are summarized in Section 2. It is shown in Section 2 that $\frac{1}{2}\,\delta(\mathcal{E}, \mathcal{M}_a)$ is the minimax probability of an incorrect guess of the true value of θ. Similarly the deficiency $\delta(\mathcal{M}_i, \mathcal{E})$ is the minimax risk for the problem of guessing the true value of θ, when no observations are available and the loss is measured by statistical distance. If we restrict attention to testing problems then the corresponding deficiency reduces to the half diameter of $\mathcal{E}$ for statistical distance.

The case of dichotomies, i.e., the case where θ has two elements, is investigated in Section 3. We begin by slightly completing previously known comparison criteria. In the case of a finite sample space and one distribution being uniform, the results generalize some of the basic inequalities from the theory of majorization.

The remainder of the paper is devoted to replicated experiments. Let $\mathcal{E}^n$ denote the experiment obtained by combining n independent replications of $\mathcal{E}$. All limits, if not otherwise stated, are taken as $n \to \infty$. How do these quantities behave under replications? As is well known, it follows from the weak law of large numbers that $\mathcal{E}^n \to \mathcal{M}_a$ provided that $P_{\theta_1} \neq P_{\theta_2}$ when $\theta_1 \neq \theta_2$ and that Θ is finite. If Θ is infinite, however, then one is tempted to conclude that "normally" $\delta_a(\mathcal{E}^n) \equiv_n 2$. There are nevertheless interesting experiments with Θ infinite where $\delta_a(\mathcal{E}^n) \to 0$.

We shall see in Section 5 that $\delta_a(\mathcal{E}^n)^{1/n}$ converges, for any experiment $\mathcal{E}$, to a constant $\sigma(\mathcal{E})$ in $[0, 1]$. This implies that if $\delta_a(\mathcal{E}^n) \to 0$ then the speed of convergence is exponential. The rate of convergence is not determined by $\delta_a(\mathcal{E})$ alone, since there are experiments $\mathcal{E}$ such that $\delta_a(\mathcal{E})$ has the maximal value 2 although $\delta_a(\mathcal{E}^n)$ converges rapidly to zero.

Now Chernoff [6] proved, when $\Theta = \{1, 2\}$, that the nth root of the minimum Bayes probability of error converges to the minimum of the Hellinger transform. It follows that $\sigma(\mathcal{E})$ and this minimum are the same number when $\mathcal{E}$ is a dichotomy.

Because pairwise equivalence for ordered experiments imply equivalence, one might hope that comparison with respect to $\mathcal{M}_i$ and $\mathcal{M}_a$ may, to some extent, be expressed in terms of the dichotomies defined by restrictions to pairs. If Θ is infinite, then — as shown in Section 5 — this does not hold in general. If Θ is finite, however, then — as is shown in Section 4 — the approximations are readily expressed in terms of dichotomies. As in

Section 3 we also get the exponential rate of convergence of many other functionals. $\sigma(\mathscr{E})$ does not, however, define the minimax probability of not covering the true value of θ with a r-point confidence set when $r \geqq 2$. Generalizing Chernoff's result we obtain the exponential rates of convergence of these minimax probabilities (and of the corresponding minimum Bayes probabilities).

The paper concludes with an example showing that dramatic improvement may be obtained by adding a single replication. This brings us close to a related topic — the relative amount of information in additional observations. We refer the reader to Helgeland [12], Le Cam [15], Swensen [22] and Torgersen [25] for some results in this direction.

2. Notations and basic facts. An experiment $\mathscr{E}$ will be defined as a family of probability measures on a common measurable space. This measurable space and the index set of the family are called, respectively, *the sample space of* $\mathscr{E}$ and *the parameter set of* $\mathscr{E}$. Thus an experiment $\mathscr{E}$ with sample space $(\chi, \mathscr{A})$ and parameter set Θ is a family $(P_\theta; \theta \in \Theta)$ of probability measures on $(\chi, \mathscr{A})$. This experiment may be denoted by $(P_\theta; \theta \in \Theta)$ or by $(\chi, \mathscr{A}; P_\theta; \theta \in \Theta)$. Expositions of the theory of the comparison of experiments may be found in Blackwell and Girshick [4], Heyer [13], Le Cam [17] and Torgersen [28].

New experiments may be derived from old ones by various devices. If Θ_0 is a subset of Θ and $\mathscr{E} = (P_\theta; \theta \in \Theta)$ is an experiment with parameter set Θ, then $\mathscr{E}_{\Theta_0}$ denotes the restriction $(P_\theta; \theta \in \Theta_0)$ of $\mathscr{E}$ to Θ_0. The restrictions of $\mathscr{M}_a$ and $\mathscr{M}_i$ to Θ_0 will, however, usually be denoted by $\mathscr{M}_a$ and $\mathscr{M}_i$, respectively.

If $\mathscr{E}_i = (\chi_i, \mathscr{A}_i, P_{\theta,i}; \theta \in \Theta)$; $1 \leqq i \leqq n$ are experiments, then $(\prod_{i=1}^n (\chi_i, \mathscr{A}_i), \prod_{i=1}^n P_{\theta,i}; \theta \in \Theta)$ is called the *product* of $\mathscr{E}_1, \mathscr{E}_2, \cdots, \mathscr{E}_n$ and it is denoted $\mathscr{E}_1 \times \cdots \times \mathscr{E}_n$ or $\prod \mathscr{E}_i$. If $\mathscr{E}_1 = \mathscr{E}_2 = \cdots = \mathscr{E}_n = \mathscr{E}$ then we may write $\mathscr{E}^n$ instead of $\mathscr{E}_1 \times \cdots \times \mathscr{E}_n$. Experiments $\mathscr{E}^n$; $n = 1, 2, \cdots$ are called *replicated experiments.*

Let $\mathscr{E} = (P_\theta; \theta \in \Theta)$ be an experiment. A random variable X will be called an *observation of* $\mathscr{E}$ if the distribution of X under θ is P_θ. Note that $(X_1, X_2, \cdots, X_n)$ is an observation of $\mathscr{E}_1 \times \cdots \times \mathscr{E}_n$ if $X_1, \cdots, X_n$ are independent and X_i; $i = 1, \cdots, n$ is an observation of $\mathscr{E}_i$.

If Π is a probability distribution on $\{1, \cdots, r]$ and $\mathscr{E}_1, \cdots, \mathscr{E}_r$ are experiments with the same parameter set, then (see [30]), the Π-mixture, $\sum_1^r \Pi_i \mathscr{E}_i$ of $\mathscr{E}_1, \cdots, \mathscr{E}_r$ is the experiment obtained by first observing a random index I with distribution Π, and then carrying out the experiment $\mathscr{E}_I$.

NOTATION. Before proceeding, we make some remarks on our use of the symbols $\#$, $\wedge$, $\vee$, $\| \ \|$, Λ and $\mathscr{L}$. $\#$ is short for "the number of elements in". The notations $\wedge$ and $\vee$ are used on several occasions for inf and sup. If, in particular, μ_t; $t \in T$ are measures then $\wedge_t \mu_t$ and $\vee_t \mu_t$ are, respectively, notations for $\inf_t \mu_t$ and $\sup_t \mu_t$ for the family $\{\mu_t; t \in T\}$ with respect to the setwise ordering of measures. If μ is a measure, then $\|\mu\|$ denotes the total variation of μ: i.e., $\|\mu\| = \sup\{\int f \, d\mu: -1 \leqq f \leqq 1\}$. We will reserve the letter Λ as the notation for the set of all prior distributions with finite support. The letter $\mathscr{L}$ will be used as notation for "distribution of". Thus, for example, $\mathscr{L}_P(f)$ is the distribution of f under P.

If $\mathscr{E} = (P_\theta; \theta \in \Theta)$ and Θ is finite then the distribution on the set of prior distributions on Θ induced from $\sum_\theta P_\theta$ by the map $(dP_\theta/d \sum P_\theta; \theta \in \Theta)$ is called the standard measure of $\mathscr{E}$. This measure characterizes $\mathscr{E}$ up to equivalence. We refer the reader to Le Cam [17] for further information on these measures.

Important functionals of experiments may be defined as follows. Let ϕ be a homogenous and measurable function on $[0, \infty[^\Theta$ and suppose $\mathscr{E} = (P_\theta; \theta \in \Theta)$ is dominated by the σ-finite measure σ. Then we may put $\phi(\mathscr{E}) = \int \phi(dP_\theta; \theta \in \Theta) = \int \phi(dP_\theta/d\sigma; \theta \in \Theta) \, d\sigma$.

If P and Q are probability measures then $\int |dP - dQ|$, $\int (\sqrt{dP} - \sqrt{dQ})^2$ and $\int \sqrt{dP \, dQ}$ are respectively: the statistical distance, the squared Hellinger distance, $D^2(P.Q.)$ and the affinity $\gamma(P, Q)$ between them.

The *Hellinger transform* of $\mathscr{E} = (P_\theta; \theta \in \Theta)$ may be defined as the map $H_\mathscr{E}$ which associates with each prior distribution t having finite support the number $H_\mathscr{E}(t) =$

$\int \Pi_\theta dP_\theta^{t_\theta}$. *Thus* $\gamma(P, Q) = H_{\mathscr{E}}(\frac{1}{2}, \frac{1}{2})$, where $\mathscr{E} = (P, Q)$. The Hellinger transform converts products of experiments into products of functions i.e.:

$$H_{\mathscr{E}_1 \times \cdots \times \mathscr{E}_n} = H_{\mathscr{E}_1} \cdot H_{\mathscr{E}_2, \ldots}, H_{\mathscr{E}_n}. \tag{2.1}$$

Let $\mathscr{E} = ((\chi, \mathscr{A}), (P_\theta, \theta \in \Theta))$ and $\mathscr{F} = ((\mathscr{Y}, \mathscr{B}), (Q_\theta; \theta \in \Theta))$ be two experiments with the same parameter set Θ and let $\theta \rightsquigarrow \varepsilon_\theta$ be a nonnegative function on Θ. We shall say, following Le Cam [14], that $\mathscr{E}$ *is* ε*-deficient relative to* $\mathscr{F}$ if to each finite decision space D, every family W_θ; $\theta \in \Theta$ of loss functions on D and every risk function r obtainable in $\mathscr{F}$, there is a risk function r' obtainable in $\mathscr{E}$ so that:

$$r'(\theta) \leqq r(\theta) + \varepsilon_\theta \sup_d |\, W_\theta(d)\,|; \qquad \theta \in \Theta \tag{2.2}$$

Restricting attention to decision spaces D where $\#D \leqq k$ we obtain the definition of ε-deficiency for k-decision problems; see [28]. If decision rules are defined as in Le Cam [14] then ε-deficiency (for k-decision problems) for all finite subsets of Θ implies and is implied by ε-deficiency. If $\mathscr{E}$ is 0-deficient relative to $\mathscr{F}$ (for k-decision problems), then we shall say that $\mathscr{E}$ is more informative than $\mathscr{F}$ (for k-decision problems) and write this $\mathscr{E} \geqq \mathscr{F}$ ($\mathscr{E} \geqq_k \mathscr{F}$).

If $\mathscr{E} \geqq \mathscr{F}$ and $\mathscr{F} \geqq \mathscr{E}$ then we shall say that $\mathscr{E}$ and $\mathscr{F}$ are equivalent and write this $\mathscr{E} \sim \mathscr{F}$. The *deficiency* of the experiment $\mathscr{E}$ with respect to the experiment $\mathscr{F}$ is the greatest lower bound of all constants $\varepsilon \geq 0$ such that $\mathscr{E}$ is ε-deficient with respect to $\mathscr{F}$. This number will be denoted by $\delta(\mathscr{E}, \mathscr{F})$. The deficiency is not symmetric and, henceforth, not a proper distance. A *distance* Δ for experiments is obtained by putting

$$\Delta(\mathscr{E}, \mathscr{F}) = \max\{\delta(\mathscr{E}, \mathscr{F}), \delta(\mathscr{F}, \mathscr{E})\}; \tag{2.3}$$

see Le Cam [14]. Similarly we may define deficiencies δ_k and distances Δ_k based on k-decision problems.

According to Le Cam's randomization criterion, theorem 3 in [14], $\mathscr{E} = (\chi, \mathscr{A}, P_\theta; \theta \in \Theta)$ is ε-deficient with respect to $\mathscr{F} = (\mathscr{Y}, \mathscr{B}, Q_\theta; \theta \in \Theta)$ if and only if there is a transition M from the band generated by $(P_\theta; \theta \in \Theta)$ to the band generated by $(Q_\theta; \theta \in \Theta)$ so that

$$\| P_\theta M - Q_\theta \| \leqq \varepsilon_\theta; \theta \in \Theta. \tag{2.4}$$

If $\mathscr{E}$ is dominated, or more generally coherent, then M may be represented as a conditional probability of $\mathscr{B}$ given $\mathscr{A}$ which may be regularized to a proper Markov kernel from $(\chi, \mathscr{A})$ to $(\mathscr{Y}, \mathscr{B})$ when $(\mathscr{Y}, \mathscr{B})$ is Euclidean.

If Θ is finite then, using the notations in [23], Γ denotes the class of sub linear functions γ on R^θ such that

$$\gamma(e^\theta) = \gamma(-e^\theta); \theta \in \Theta \qquad \text{and} \qquad \textstyle\sum_\theta \gamma(e^\theta) = 1.$$

Here e^θ, for each θ, is the θth unit vector $i\, R^\Theta$, i.e., $e^\theta = (0, \cdots 1, \cdots 0)$. The subclass of Γ consisting of those functions in Γ which are maximums of k-linear functionals will be denoted by Γ_k. A function γ will be called superlinear if $-\gamma$ is sublinear. The deficiencies are then (the sublinear function criterion) given by

$$\delta_{(k)}(\mathscr{E}, \mathscr{F}) = \sup_{\gamma \in \Gamma_{(k)}} [\gamma(\mathscr{F}) - \gamma(\mathscr{E})]. \tag{2.5}$$

Let us apply these results to deficiencies $\delta_i(\mathscr{E})$ and $\delta_a(\mathscr{E})$ for an experiment $\mathscr{E} = (\chi, \mathscr{A}, P_\theta; \theta \in \Theta)$.

The randomization criterion yields directly that.

$$\delta_i(\mathscr{E}) = \min_Q \sup_\theta \| P_\theta - Q \| \qquad \text{and} \qquad \delta_a(\mathscr{E}) = 2 \min_M \sup_\theta P_\theta M(\{\theta\}^c), \tag{2.6}$$

where Q runs through all probability distributions on $\mathscr{A}$, while M runs through all transitions from the band generated by the P_θ's to the band of discrete finite measures on Θ.

It follows that $\delta_t(\mathscr{E})$ is the minimax risk in the estimation problem where no observations are available and loss is measured by statistical distance. Furthermore

$$\tfrac{1}{2}\sup_{\theta_1,\theta_2}\|P_{\theta_1}-P_{\theta_2}\| \leqq \delta_t(\mathscr{E}) \leqq \sup_{\theta_1,\theta_2}\|P_{\theta_1}-P_{\theta_2}\|, \tag{2.7}$$

where the left-hand side is $\delta_2(\mathscr{M}_t, \mathscr{E})$, by [23].

Similarly $\frac{1}{2}\delta_a(\mathscr{E})$ is the minimax risk in the problem of estimating θ on the basis of $\mathscr{E}$ when the loss is 1 or 0 as the estimator hits or not. If $\mathscr{E}$ is dominated then $(P_\theta M;\ \theta\in\Theta)$ is also dominated, so that $\inf_\theta P_\theta M(\{\theta\}) = 0$ when Θ is uncountable. Hence $\delta_a(\mathscr{E}) = 2$ in this case. More generally $\mathscr{E}$ is ε-deficient with respect to $\mathscr{M}_a$ if and only if $P_\theta M(\{\theta\}) \geqq 1-\varepsilon_\theta/2;\ \theta\in\Theta$ for some transition M. If Θ is finite and each P_θ is atomless then, by Dvoretzky et al's extension [9] of Lyapunov's theorem [20], this is equivalent to the existency of a measurable partition $A_\theta;\ \theta\in\Theta$ of χ such that $P_\theta(A_\theta)\geqq 1-\varepsilon_\theta/2$.

If λ is a prior distribution on Θ with countable support then the minimum Bayes risk in the last estimation problem will be denoted by $b(\lambda\,|\,\mathscr{E})$. It is easily seen that

$$b(\lambda\,|\,\mathscr{E}) = 1 - \|\vee_\theta \lambda_\theta P_\theta\| \tag{2.8}$$

and the map $\lambda \sim\!\!\rightarrow b(\lambda\,|\,\mathscr{E})$ on Λ defines, by Morse and Sacksteder [21], $\mathscr{E}$ up to equivalence. If Θ is countable then each decision rule δ may be represented as a random distribution which to each x assigns the distribution $\delta.(x)$ on Θ. Suppose in addition that each P_θ has density f_θ with respect to the measure σ. Then δ achieves minimum Bayes risk if and only if, for σ almost all x, $\delta.(x)$ is supported by $\{\theta;\ \lambda_\theta f_\theta(x) = \max_{\theta'}\lambda_{\theta'}f_{\theta'}(x)\}$. In particular any maximum likelihood estimator of θ achieves minimum Bayes risk when θ is finite and λ is uniform. It follows from straight-forward minimax theory that

$$\tfrac{1}{2}\delta_a(\mathscr{E}) = \sup_{\lambda\in\Lambda} b(\lambda\,|\,\mathscr{E}) = 1 - \inf_{\lambda\in\Lambda}\|\vee_\theta\lambda_\theta P_\theta\|. \tag{2.9}$$

The prior distribution λ (with countable support) will be called least favorable if it is least favorable in this estimation problem, i.e., if $b(\lambda\,|\,\mathscr{E}) = \frac{1}{2}\delta_a(\mathscr{E})$.

It follows from [23] that $\mathscr{E}$ is ε-deficient with respect to $\mathscr{M}_a$ for testing problems if and only if to each subset Θ_0 of Θ there corresponds a power function π in $\mathscr{E}$ so that $\pi(\theta)\leqq \varepsilon_\theta/2$ or $\geqq 1-\varepsilon_\theta/2$ as $\theta\in\Theta_0$ or $\theta\notin\Theta_0$. The sublinear function criterion (2.5) shows that

$$\delta_2(\mathscr{E}, \mathscr{M}_a) = \sup(1 - \|\,|\textstyle\sum_\theta a_\theta P_\theta\|) \tag{2.10}$$

where a runs through all finite measures a on θ with finite support and total variation $= \sum_\theta |\,a_\theta\,| = 1$.

By [23], $\delta_i(\mathscr{M}_a) = \delta_a(\mathscr{M}_t) = 2(1-m^{-1})$ where $m = \#\Theta$. Hence, by the triangular inequality for deficiencies

$$2 - 2/m \leqq \delta_t(\mathscr{E}) + \delta_a(\mathscr{E}). \tag{2.11}$$

One might expect that $\delta_a(\mathscr{E})$ is small (large) when $\delta_i(\mathscr{E})$ is large (small) and conversely. If Θ is finite, then this may be made precise by inequalities; see [29]. If Θ is infinite, however, then it may easily happen that $\delta_i(\mathscr{E}^n)\to 2$ while $\delta_a(\mathscr{E}^n)\geqq 1$ for all n.

3. Replicated dichotomies. We shall in this section assume that $\Theta = \{1, 2\}$ i.e., that our experiments are dichotomies. In this case deficiencies may, [23], be expressed in terms of testing problems only. We shall need:

THEOREM 3.1. *Suppose $\mathscr{E} = (P_1, P_2)$ and $\mathscr{F} = (Q_1, Q_2)$. Denote by $\beta(\alpha\,|\,\mathscr{E})$ and $\beta(\alpha\,|\,\mathscr{F})$ the powers of the most powerful level α tests for $\theta = 1$ against $\theta = 2$ in $\mathscr{E}$ and $\mathscr{F}$. Also denote by $b(\lambda\,|\,\mathscr{E})$ and $b(\lambda\,|\,\mathscr{F})$ the minimum Bayes risks for the same problem for $0-1$ loss and prior (λ_1, λ_2) in $\mathscr{E}$ and $\mathscr{F}$.*

Then each of the following conditions are equivalent:

(i) $$\beta\left(\alpha + \frac{\varepsilon_1}{2} \mid \mathscr{E}\right) + \frac{\varepsilon_2}{2} \geqq \beta(\alpha \mid \mathscr{F}) \qquad ; \alpha \geqq 0$$

(ii) $$[\lambda_1\varepsilon_1 + \lambda_2\varepsilon_2]/2 \geqq b(\lambda \mid \mathscr{E}) - b(\lambda, \mathscr{F}) \qquad ; 0 \leqq \lambda \leqq 1$$

(iii) $$\int \gamma(dP_2/dP_1) \leqq \int \gamma(dQ_2/dQ_1)\, dQ_1 + \frac{\varepsilon_1}{2}[\gamma(\infty) - \gamma(0)] + \frac{\varepsilon_2}{2}\gamma'(0)$$

for each concave function γ *on* $[0, \infty[$ *such that* $\gamma(x)/x \to 0$ *as* $x \to \infty$.

(iv) $$\frac{\varepsilon_1}{2}[\phi'(1) - (\phi(1) - \phi(0))] + \frac{\varepsilon_2}{2}[(\phi(1) - \phi(0)) - \phi'(0)]$$
$$\geqq \int \phi(dQ_2/d(Q_1 + Q_2))d(Q_1 + Q_2) - \int \phi(dP_2/d(P_1 + P_2))d(P_1 + P_2)$$

for any convex function ϕ *on* $[0, 1]$.

(v) $\mathscr{E}$ *is* $(\varepsilon_1, \varepsilon_2)$ *deficient with respect to* $\mathscr{F}$.

REMARK. If μ and ν are finite measures then $d\mu/d\nu$ are the Radon-Nikodym derivative of the ν continuous part of μ with respect to ν.

PROOF. The equivalence of (i), (ii) and (v) follows from [23] while the implications (iv) $\Rightarrow$ (iii) $\Rightarrow$ (ii) follows by applying (iv) and (iii) to, respectively, the functions $\phi(x) = \gamma\left(\frac{x}{1-x}\right)(x-1)$ and $\gamma(x) = \min\{1 - \lambda, \lambda x\}$. It remains therefore to show that (ii) implies (iv). Suppose (ii) holds and let ϕ be convex on $[0, 1]$. Define convex functions ϕ_n; $n = 1, 2, \cdots$ by requiring that: $\phi_n(k/2n) = \phi(k/2n)$; $k = 0, 1, \cdots, 2^n$ and that ϕ_n is linear on each interval $\left[\frac{i-1}{2^n}, \frac{i}{2^n}\right]$; $i = 1, \cdots, 2^n$. Then $\phi_n \downarrow \phi$, $\phi_n'(0) \downarrow \phi'(0)$ and $\phi_n'(1) \uparrow \phi'(1)$. It follows, without loss of generality, that we may assume that the graph of ϕ consists of a finite number of line segments. Hence we may write $\phi(x) = \max\{a_i + b_i x;\ i = 1, \cdots, k\}$ where $a_1 > a_2 > \cdots > a_k$ and $b_1 < b_2 < \cdots < b_k$. It follows that: $\phi(x) = a_1 + b_1 x + \sum_{i=1}^{k} (a_{i+1} + b_{i+1}x - a_i - b_i x)^+$. The inequality is trivial when ϕ is linear. It suffices therefore to prove it for functions ϕ of the form

$$\phi(x) = A \vee Bx = Bx + A(1 - x) - \min\{A(1 - x), (B - A)x\}$$

where, since the case $A \geqq B$ is trivial, $0 < A < B$. Putting $\lambda = 1 - A/B$ we see then that the inequality reduces to (ii). □

A prior distribution λ may in this case be identified with the probability it assigns to $\{2\}$. We shall therefore in this Section write $b(\lambda \mid \mathscr{E})$ instead of $b(1 - \lambda, \lambda \mid \mathscr{E})$ and $H(t \mid \mathscr{E})$ instead of $H(1 - t, t \mid \mathscr{E})$ when $\lambda, t \in [0, 1]$. The Bayes risk function $\lambda \rightsquigarrow b(\lambda \mid \mathscr{E})$ may then be written

$$b(\lambda \mid \mathscr{E}) = 1 - \|1 - \lambda)P_1 \vee \lambda P_2\| = \|(1 - \lambda)P_1 \wedge \lambda P_2\|.$$

Hence

$$0 \leqq b(\lambda \mid \mathscr{E}) \leqq (1 - \lambda) \wedge \lambda; \qquad \lambda \in [0, 1].$$

Note that the right-hand side is $b(\lambda \mid \mathscr{M}_i)$ while the left-hand side is $b(\lambda \mid \mathscr{M}_a)$. Conversely, any concave function f on $[0, 1]$ such that $0 \leqq f(\lambda) \leqq (1 - \lambda) \wedge \lambda$ for $\lambda \in [0, 1]$ is of the form $f(\lambda) = b(\lambda \mid \mathscr{E})$, where $\mathscr{E}$ is, up to equivalence, determined by f. A particularly interesting aspect of this representation is the relation

$$b(\lambda \mid \sup_t \mathscr{E}_t) = \inf_t b(\lambda \mid \mathscr{E}_t),$$

which is valid for any family $\{\mathscr{E}_t;\ t \in T\}$ of dichotomies. Deficiencies are easily described

in terms of these functions. By Theorem 3.1

$$\delta(\mathscr{E}, \mathscr{F}) = 2 \sup_{0 \leqq \lambda \leqq 1} [b(\lambda \mid \mathscr{E}) - b(\lambda \mid \mathscr{F})]$$

for any pair $(\mathscr{E}, \mathscr{F})$ of dichotomies. In particular $\delta_i(\mathscr{E}) = 1 - 2b(\frac{1}{2}, \mathscr{E})$; see [23].

There is a simple connection between the Hellinger transform and the Bayes risk for dichotomies as follows.

THEOREM 3.2. [28]

$$H(t \mid \mathscr{E})/(1-t)t = \int_0^1 b(\lambda \mid \mathscr{E})(1-\lambda)^{t-2}\lambda^{-t-1}d\lambda = \int_0^\infty x^{-t-1} \| xP_1 \wedge P_2 \| \, dx.$$

As an application consider the problem of finding bounds for $H(t \mid \mathscr{E})$ in terms of $b(\lambda \mid \mathscr{E})$.

THEOREM 3.3. *For any pair* (λ, t) *of prior distributions.*

$$b(\lambda \mid \mathscr{E}) \leqq (1-\lambda)^{1-t}\lambda^t H(t \mid \mathscr{E})$$

and

$$H(t \mid \mathscr{E}) \leqq [1 - b(\lambda \mid \mathscr{E}) + \lambda b'(\lambda \mid \mathscr{E})]^{1-t}[b(\lambda \mid \mathscr{E}) + (1-\lambda)b'(\lambda \mid \mathscr{E})]^t$$
$$+ [b(\lambda \mid \mathscr{E}) - \lambda b'(\lambda \mid \mathscr{E})]^{1-t}[1 - b(\lambda \mid \mathscr{E}) - (1-\lambda)b'(\lambda \mid \mathscr{E})]^t,$$

where $b'(\lambda \mid \mathscr{E})$ *is any number between the left and right derivative of* $\lambda \rightsquigarrow b(\lambda \mid \mathscr{E})$ *at* λ. *Both inequalities are, for given values of* $b(\lambda \mid \mathscr{E})$, *sharp.*

PROOF. Put $\lambda = \lambda_0$, $b_0 = b(\lambda_0 \mid \mathscr{E})$, $b_0' = b'(\lambda_0 \mid \mathscr{E})$ and suppose $0 < \lambda_0 < 1$. Then $b(\cdot \mid \mathscr{C}) \leqq b(\cdot \mid \mathscr{E}) \leqq b(\cdot \mid \mathscr{D})$ where $b(\lambda \mid \mathscr{C}) = b_0[(\lambda/\lambda_0) \wedge (1-\lambda)/(1-\lambda_0)]$ and $b(\lambda \mid \mathscr{D}) = [b_0 + (\lambda - \lambda_0)b_0'] \wedge (1-\lambda) \wedge \lambda$. Hence $\mathscr{C} \geqq \mathscr{E} \geqq \mathscr{D}$ so that $H(t \mid \mathscr{C}) \leqq H(t \mid \mathscr{E}) \leqq H(t \mid \mathscr{D})$. These are the desired inequalities. □

REMARK. The first inequality implies that

$$b(\lambda \mid \mathscr{E}) \leqq ((1-\lambda) \wedge \lambda) \inf_t H(t \mid \mathscr{E}),$$

yielding the right of the inequalities:

$$[\gamma(\mathscr{E})/2]^2 \leqq \delta_a(\mathscr{E})/2 \leqq \inf_t H(t \mid \mathscr{E}),$$

while the left inequality follows by choosing λ least favorable and by putting $t = \frac{1}{2}$.

Inserting $\lambda = t = \frac{1}{2}$ we find

$$\| P_1 \wedge P_2 \| \leqq \gamma(P_1, P_2) \leqq \sqrt{\| P_1 \wedge P_2 \| \, \| P_1 \vee P_2 \|},$$

or equivalently

$$D^2(P_1, P_2) \leqq \| P_1 - P_2 \| \leqq 2\sqrt{1 + (P_1, P_2)}\, D(P_1, P_2).$$

This proves the known fact, see Le Cam [15], that the Hellinger distance is equivalent to the statistical distance.

Using the fact that

$$\lambda_1 f_1 g_1 \wedge \lambda_2 f_2 g_2 \geqq (\lambda_1 f_1 \wedge \lambda_2 f_2)(g_1 \wedge g_2) \geqq (\lambda_1 f_1 \wedge \lambda_2 f_2)(\mu_1 g_1 \wedge \mu_2 g_2)$$

whenever $\lambda_1, \lambda_2, \mu_1, \mu_2, f_1, f_2, g_1$ and g_2 are nonnegative numbers we get:

PROPOSITION 3.4. *If $\mathscr{E}$ and $\mathscr{F}$ are dichotomies, then for any pair (λ, μ) of prior distributions*

$$b(\lambda \mid \mathscr{E} \times \mathscr{F}) \geqq 2b(\lambda \mid \mathscr{E})b(\tfrac{1}{2} \mid \mathscr{F}) \geqq b(\lambda \mid \mathscr{E})b(\mu \mid \mathscr{F}).$$

In particular

$$\delta_a(\mathscr{E} \times \mathscr{F})/2 \geqq [\delta_a(\mathscr{E})/2]2b(\tfrac{1}{2} \mid \mathscr{F}) \geqq [\delta_a(\mathscr{E})/2][\delta_a(\mathscr{F})/2].$$

By proposition 3.4, $b(\lambda \mid \mathscr{E}^{m+n}) \geqq b(\lambda \mid \mathscr{E}^m)b(\lambda \mid \mathscr{E}^n)$ and $\delta_a(\mathscr{E}^{m+n})/2 \geqq [\delta_a(\mathscr{E}^m)/2][\delta_a(\mathscr{E}^n)/2]$. It follows that $\delta_a(\mathscr{E}^n)^{1/n}$, $n = 1, 2, \cdots$ and $b(\lambda \mid \mathscr{E}^n)^{1/n}$, $n = 1, 2, \cdots$ converge, as $n \to \infty$, to respectively $\sup_n[\delta_a(\mathscr{E}^n)/2]^{1/n}$ and $\sup_n b(\lambda \mid \mathscr{E}^n)^{1/n}$. Furthermore, since $b(\lambda \mid \mathscr{E}^n) \leq \delta_a(\mathscr{E}^n)/2 \leqq b(\lambda \mid \mathscr{E}^n)/[(1 - \lambda) \wedge \lambda]$ when $\lambda \in]0, 1]$, these limits are the same. Thus

THEOREM 3.5. *There is for each dichotomy $\mathscr{E}$, a constant $C(\mathscr{E})$ in $[0, 1]$ such that, for each nondegenerate prior distribution λ*

$$\lim_{n\to\infty} b(\lambda \mid \mathscr{E}^n)^{1/n} = \lim_{n\to\infty} [\delta_a(\mathscr{E}^n)/2]^{1/n} = \sup_n b(\lambda \mid \mathscr{E}^n)^{1/n}$$

$$= \sup_n [\delta_a(\mathscr{E}^n)/2]^{1/n} = C(\mathscr{E}).$$

REMARK. It follows that $C(\mathscr{E}^r) = C(\mathscr{E})^r$; $r = 1, 2, \cdots$ and that $C(\mathscr{E}) \geqq C(\mathscr{F})$ when $\mathscr{E} \leqq \mathscr{F}$. Also, by this Theorem and Proposition 3.4: $C(\mathscr{E} \times \mathscr{F}) \geqq C(\mathscr{E})C(\mathscr{F})$ for any pair $(\mathscr{E}, \mathscr{F})$ of dichotomies. Using that $1 - \delta_i(\mathscr{E}) = 2b(\tfrac{1}{2}, \mathscr{E})$ we get

COROLLARY 3.6.

$$\lim_{n\to\infty} [1 - \delta_i(\mathscr{E}^n)]^{1/n} = C(\mathscr{E}).$$

By the remark after Theorem 3.3: $[\delta_a(\mathscr{E}^n)/2]^{1/n} \leqq \inf_t H(t \mid \mathscr{E})$. The fact that the limit $C(\mathscr{E})$ of the left-hand side equals the right-hand side is a consequence of the following result of Chernoff [6].

THEOREM 3.7. *For any dichotomy $\mathscr{E}$ and any nondegenerate prior distribution λ*

$$\lim_{n\to\infty} b(\lambda \mid \mathscr{E}^n)^{1/n} = \inf_t H(t \mid \mathscr{E}).$$

Chernoff first derives the basic convergence result for large deviations,

$$\lim_{n\to\infty} P(1/n\ (X_1 + \cdots + X_n) \geqq 0)^{1/n} = \inf_{t\geqq 0} Ee^{tX},$$

which is valid for any sequences $X_1, X_2, \cdots$ of independent and identically distributed variables. Then, applying this to log likelihoods, he proves Theorem 6.7.

Chernoff's result has simple and important interpretations in terms of experiments and it is somewhat unsatisfactory that several of the steps in the various proofs are not so easily interpreted in terms of experiments. We shall now give a simple and "natural" proof of this important result which is based solely on a few basic facts about statistical experiments. The proof admits variations and is of considerable interest in itself. We shall see in the next section (Theorem 7.2) how this idea may be used to extend Chernoff's result to a more general result. The main tools of the proof are compactness of Δ convergence, [17], and a type of product homomorphism for experiments which we shall describe now.

Let $\mathscr{E} = (P_\theta; \theta \in \Theta)$ be an experiment dominated by the σ-finite measure μ. Put $f_\theta = dP_\theta/d\mu$ and consider a family $\{t^{(i)}: i \in I\}$ of prior distributions with finite supports and

such that $H(t^{(i)} \mid \mathscr{E}) > 0$, $i \in I$. For each $i \in I$ define a probability measure Q_i by $dQ_i/d\mu = H(t^{(i)} \mid \mathscr{E})^{-1} \prod_\theta f_\theta^{t_\theta^{(i)}}$. It is easily seen that the measures Q_i, do not depend on our choice of dominating measure. We shall denote the experiment $(Q_i: i \in I)$ by $(\mathscr{E}: t^{(i)}; i \in I)$. If s is a prior on I with finite support, then

$$H(s \mid (\mathscr{E}: t^{(i)}; i \in I)) = H(\textstyle\sum_i s_i t^{(i)} \mid \mathscr{E}) \prod_i H(t^{(i)} \mid \mathscr{E})^{-s_i}.$$

It follows that $(\mathscr{E}: t^{(i)}; i \in I) \sim (\mathscr{F}: t^{(i)}; i \in I)$ when $\mathscr{E} \sim \mathscr{F}$. Also, by the product rule for Hellinger transforms

$$(\mathscr{E}: t^{(i)}; i \in I) \times (\mathscr{F}: t^{(i)}; i \in I) \sim (\mathscr{E} \times \mathscr{F}: t^{(i)}; i \in I)$$

in the sense that the left side is defined if and only if the right side is defined and then equivalence holds.

Consider now a dichotomy $\mathscr{E} = (P_1, P_2) \sim \mathscr{M}_a$ and a real number a. Put $\rho(t) = e^{-at}H(t \mid \mathscr{E})$ and $\rho_0 = \inf_t \rho(t)$ and assume $\rho_0 = \rho(t_0)$ where $t_0 \in]0, 1[$. Then, when n is sufficiently large, the Hellinger transform of

$$\left(\mathscr{E}: t_0; t_0 + \frac{s}{\sqrt{n}}\right)^n \sim \left(\mathscr{E}^n: t_0; t_0 + \frac{s}{\sqrt{n}}\right)$$

is

$$\left[\rho\left(t_0 + \frac{ts}{\sqrt{n}}\right) \Big/ \rho_0^{1-t}\rho\left(t_0 + \frac{s}{\sqrt{n}}\right)^t\right]^n$$

which converges, as $n \to \infty$, to $e^{-s^2\tau^2 t(1-t)/2}$ where $\tau^2 = \rho''(t_0)/\rho_0$. This limit, as a function of t, is the Hellinger transform of the dichotomy $\mathscr{G}_s = (Q_0, Q_s)$ where Q_s, for each s, is the Normal $(s\tau, 1)$ distribution. Hence $\Delta\left(\left(\mathscr{E}^n: t_0; t_0 + \frac{s}{\sqrt{n}}\right), \mathscr{G}_s\right) \to 0$ so that the laws of likelihood ratios of $\left(\mathscr{E}^n: t_0; t_0 + \frac{s}{\sqrt{n}}\right)$ converges weakly to the corresponding laws for $\mathscr{G}_s$.

Thus.

$$\mathscr{L}\left(U_n^{s/\sqrt{n}} e^{-as\sqrt{n}} \rho\left(t_0 + \frac{s}{\sqrt{n}}\right)^{-n} \rho_0^n \mid e^{-nat_0}\rho_0^{-n} U_n^{t_0} P_1^n\right) \to \mathscr{L}(U \mid Q_0),$$

where $U_n = dP_2^n/dP_1^n$ and $U = dQ_s/dQ_0$. Hence, since $\rho\left(t_0 + \frac{s}{\sqrt{n}}\right)^{-n} \rho_0^n \to e^{-s^2\tau^2/2}$,

$$E_{P_1^n}\phi(U_n^{s/\sqrt{n}} e^{-as\sqrt{n}}) U_n^{t_0} \sim \rho_0^n e^{nat_0} E_{Q_0}\phi(e^{s^2\tau^2/2} U)$$

when ϕ is bounded and continuous a.e. Lebesgue if $\tau > 0$ and continuous in 1 if $\tau = 0$.

Choose a number $M > 1$ and put $\phi(x) = 1$ or 0 as $x \in [M^{-1}, M]$ or not.
Then

$$\begin{aligned} \| P_1^n \wedge e^{-na} P_2^n \| &\geqq \int (1 \wedge e^{-na} U_n) \phi(e^{-\sqrt{n}a} U_n^{1/\sqrt{n}}) U_n^{-t_0} U_n^{t_0}\, dP_1^n \\ &\geqq M^{-\sqrt{n}} M^{-\sqrt{n}t_0} e^{-nat_0} \int \phi(e^{-\sqrt{n}a} U_n^{1/\sqrt{n}}) U_n^{t_0}\, dP_1^n \\ &\sim M^{-\sqrt{n}(1+t_0)} \rho_0^n E_{Q_0}\phi(e^{\tau^2/2}\, dQ_1/dQ_0). \end{aligned}$$

Hence, since the last expectation is > 0, we get

$$\liminf_n \| P_1^n \wedge e^{-na} P_2^n \|^{1/n} \geqq \inf_t e^{-at} H(t \mid \mathscr{E}). \tag{3.1}$$

If $\mathscr{E} \sim \mathscr{M}_a$ then $\rho_0 = 0$ so that (3.1) holds trivially. In general the function ρ may be forced

to obtain its infimum in]0, 1[by increasing the amount of information in $\mathscr{E}$ by a negligible amount. To be more precise let N be large and put $\mathscr{G} = \left(1 - \frac{1}{N}\right) \mathscr{M}_i + \frac{1}{N} \mathscr{F} = (\bar{P}_1, \bar{P}_2)$ where $-H'(0 + \mid \mathscr{F}) = H'(1 - \mid \mathscr{F}) = \infty$. Then $t \rightsquigarrow H(t \mid \mathscr{E} \times \mathscr{G}) e^{-at}$ obtains its infimum in]0, 1[. Hence

$$\liminf_n \| P_1^n \wedge e^{-na} P_2^n \|^{1/n} \geqq \liminf_n \| (P_1 \times Q_1)^n \wedge e^{-n}(P_2 \times Q_2)^n \|$$

$$\geqq \inf_t [e^{-at} H(t \mid \mathscr{E}) H(t \mid \mathscr{G})] \geqq \left(1 - \frac{1}{N}\right) \rho_0 \to \rho_0$$

as $N \to \infty$. We used here that $H(t \mid \mathscr{G}) = \left(1 - \frac{1}{N}\right) + \frac{1}{N} H(t \mid \mathscr{F}) \geqq 1 - \frac{1}{N}$.

It follows that (3.1) holds for any dichotomy (P_1, P_2). $a = 0$ yields $\liminf_n b(\lambda \mid \mathscr{E}^n)^{1/n} \geqq \inf_t H(t \mid \mathscr{E})$ while $b(\lambda \mid \mathscr{E}^n) \leqq H(t \mid \mathscr{E})^n$ by Theorem 3.3. Hence $b(\lambda \mid \mathscr{E}^n)^{1/n} \to \inf_t H(t \mid \mathscr{E})$ and this is Chernoff's theorem. By Theorems 3.5 and 3.7

THEOREM 3.8.

$$C(\mathscr{E}) = \inf_t H(t \mid \mathscr{E}).$$

REMARK. By the remarks after Theorem 3.3, $b(\lambda \mid \mathscr{E}) \leqq [(1 - \lambda) \vee \lambda] C(\mathscr{E})$ and $\gamma[(\mathscr{E})/2]^2 \leqq \delta_a(\mathscr{E})/2 \leqq C(\mathscr{E})$.

We shall now consider a few extensions of Theorem 3.5. Let us first consider the asymptotic behavior of minimum Bayes risk in other decision problems. It is known, see for example [8] or [23], that the minimum Bayes risk may often be expressed as functionals $\psi(\mathscr{E}) = \int \psi(dP_1, dP_2)$ where the function ψ is super linear (i.e., $\psi(x + y) \geqq \psi(x) + \psi(y)$ and $\psi(tx) = t\psi(x)$ when $t \geqq 0$) or R^2. The function ψ is determined by the loss function. It follows, since the standard measure of $\mathscr{E}^n$ converges weakly to the standard measure of $\mathscr{M}_a$ when $\mathscr{E} \not\perp \mathscr{M}_i$, that $\psi(\mathscr{E}^n) \to \psi(\mathscr{M}_a)$ as $n \to \infty$ provided $\mathscr{E} \not\perp \mathscr{M}_i$. By Theorem 2 in [23]: $0 \leqq \psi(\mathscr{E}) - \psi(\mathscr{M}_a) \leqq \frac{\delta_a(\mathscr{E})}{2} [\psi(1, 0) + \psi(-1, 0) + \psi(0, 1) + \psi(0, -1)]$. It follows, by replacing $\mathscr{E}$ with $\mathscr{E}^n$ and applying Theorem 3.5 that $\limsup_n [\psi(\mathscr{E}^n) - \psi(\mathscr{M}_a)]^{1/n} \leqq C(\mathscr{E})$.

Suppose ψ is not affine on $[0, \infty[^2$. Put $\phi(x) = \psi(1 - x, x)$, $x \in [0, 1]$. Then ϕ is concave on [0, 1] and for some $x_0 \in]0, 1[$: $\phi(x_0) > (1 - x_0)\phi(0) + x_0\phi(1)$. Let χ be the function on [0, 1] which is linear on the intervals $[0, x_0]$ and $[x_0, 1]$ and which satisfies $\chi(0) = \phi(0)$, $\chi(x_0) = \phi(x_0)$, $\chi(1) = \phi(1)$. Then $\psi(\mathscr{E}) - \psi(\mathscr{M}_a) = \bar{\psi}(\mathscr{E}) - \bar{\psi}(\mathscr{M}_a)$ where $\bar{\psi}(x) = \psi(x) - x_1\psi(1, 0) - x_2\psi(0, 1)$. Thus we may as well assume that $\phi(0) = \phi(1) = 0$ and then $\psi(\mathscr{E}) - \psi(\mathscr{M}_a) = \psi(\mathscr{E}) \geqq \int \chi\left(\frac{dP_2}{d(P_1 + P_2)}\right) d(P_1 + P_2) = kb(\lambda, \mathscr{E})$ where $k = \phi(x_0)\left(\frac{1}{x_0} + \frac{1}{1 - x_0}\right)$ and $\lambda = \frac{1}{x_0}\left(\frac{1}{x_0} + \frac{1}{1 - x_0}\right)^{-1}$. Using Theorem 3.5 once more, we find the following.

THEOREM 3.9. *If ψ is super linear on R^2 and not linear on $[0, \infty[^2$ then*

$$\lim_{n \to \infty} [\psi(\mathscr{E}^n) - \psi(\mathscr{M}_a)]^{1/n} = C(\mathscr{E}).$$

REMARK. If $\psi(x) = \wedge_t \sum_{\theta=1}^2 \lambda_\theta L_\theta(t) x_\theta$ where L is the loss function, then $\psi(\mathscr{E})$ is the minimum Bayes risk for the loss function L. If $\wedge_t$ is attained, then the exceptional case is the situation where, for some t_0, $L_\theta(t_0) \leqq L_\theta(t)$ for $\theta = 1, 2$, and all t. In that case no observations are needed and the decision rule $x \rightsquigarrow t_0$ is "uniformly" optimal.

In spite of the last remark there are interesting measures of information based on

deficiencies whose exponential rates of convergence differs from $C(\mathscr{E})$.

EXAMPLE 3.10. Fix a number $\alpha \in [0, 1]$ and consider the smallest number $2\varepsilon_2$ such that the dichotomy (P_1, P_2) is $(2\alpha, 2\varepsilon_2)$ deficient with respect to $\mathscr{M}_a$. By Theorem 2.3 this is just $1 - \beta(\alpha \mid P_1, P_2)$ where $\beta(\alpha \mid P_1, P_2)$ is the power of the most powerful level α test for testing "P_1" against "P_2". It is shown in an unpublished paper of Stein (see [7] or [1] for a proof) that

$$[1 - \beta(\alpha \mid P_1^n, P_2^n)]^{1/n} \to e^{E_{P_1} \log \frac{dP_2}{dP_1}}$$

We shall for the remaining part of this section assume that $\mathscr{L}_{P_1}(\log(dP_2/dP_1))$ is nonlattice and that $\inf_t H(t \mid \mathscr{E})$ is obtained at $t_0 \in]0, 1[$. Put $\tau^2 = C(\mathscr{E})^{-1}H''(t_0 \mid \mathscr{E})$ and $A_n = C(\mathscr{E})^{-n}(2\pi\tau^2 n)^{1/2}$. Then the expansion of Efron and Truax [11] may be written

$$b(\lambda \mid \mathscr{E}^n) \sim A_n^{-1}(1 - \lambda)^{1-t_0}\lambda^{t_0}/(1 - t_0)t_0.$$

The functions $\lambda \rightsquigarrow b(\lambda \mid \mathscr{E}^n)$ and the function $\lambda \rightsquigarrow (1 - \lambda)^{1-t_0}\lambda^{t_0}$ are all concave on $[0, 1]$. It follows that the convergence is uniform in λ. Maximizing with respect to λ we find that $\delta_a(\mathscr{E}^n)/2 \sim A_n^{-1}(1 - t_0)^{-t_0}t_0^{t_0-1}$ i.e., $\delta_a(\mathscr{E}^n)/2 = b(t^0 \mid \mathscr{E}^n)(1 + o(1))$. Let μ_n be least favorable in $\mathscr{E}^n$; i.e., $b(\mu_n \mid \mathscr{E}^n) = \delta_a(\mathscr{E}^n)/2$. Then

$$A_n b(\mu_n, \mathscr{E}^n) - [(1 - \mu_n)^{1-t_0}\mu_n^{t_0}/(1 - t_0)t_0] \to 0.$$

Hence $(1 - \mu_n)^{1-t_0}\mu_n^{t_0} \to (1 - t_0)^{1-t_0}t_0^{t_0}$ so that $\mu_n \to t_0$. By a slight extension of this argument we find that $\delta_a(\mathscr{E}^n)/2 = b(\mu_n \mid \mathscr{E})(1 + o(1))$ as $n = \infty$, if and only if $\mu_n \to t_0$. This proves

THEOREM 3.11.

(i) $\delta_a(\mathscr{E}^n)/2 = (1 - t_0)^{-t_0}t_0^{t_0-1}\frac{1}{\sqrt{2\pi\tau^2}}\frac{1}{\sqrt{n}}(1 + o(1))$ *as* $n \to \infty$.

(ii) t_0 *is asymptotically least favorable in the sense that* $\delta_a(\mathscr{E}^n)/2 = b(t_0 \mid \mathscr{E}^n)(1 + o(1))$ *as* $n \to \infty$.

(iii) *More generally* $\delta_a(\mathscr{E}^n)/2 = b(\mu_n \mid \mathscr{E}^n)(1 + o(1))$ *as* $n \to \infty$ *if and only if* $\lim_n \mu_n = t_0$.

The prior t_0 which minimizes $t \rightsquigarrow H(t \mid \mathscr{E})$ is, by Theorem 3.11, asymptotically least favorable.

Let U be the distribution on $]0, \infty[$ with density $x \rightsquigarrow x^{-t_0}$ with respect to Lebesgue measure. Note that the expansion of Efron and Truax may be written,

$$\lim_{n\to\infty} \int [(1 - \lambda) \wedge \lambda x]A_n K_n(dx) = \int [(1 - \lambda) \wedge \lambda x]U(dx)$$

where $K_n = \mathscr{L}_{P_1^n}(dP_2^n/dP_1^n)$. It follows that $\lim_{n\to\infty} \int \phi(x)A_n K_n(dx) = \int \phi(x)U(dx)$ for any function ϕ on $[0, \infty[$ which is a linear combination of functions $x \rightsquigarrow (1 - \lambda) \wedge \lambda x; \lambda \in [0, 1]$. It is not difficult to see that a function ϕ is a linear combination of functions $x \rightsquigarrow (1 - \lambda) \wedge \lambda x; \lambda \in [0, 1]$, if and only if ϕ is polygonal, $\phi(0) = 0$ and $\phi(x) = \lim_{x\to\infty} \phi(x)$ when x is sufficiently large. Hence, by the theory of weak convergence of measures,

THEOREM 3.12.

$$\int \phi(dP_2^n/dP_1^n)dP_1^n = \left[\int \phi(x)U_{t_0}(dx)\right]\frac{1}{\sqrt{2\pi\tau^2}}\frac{1}{\sqrt{n}}C(\mathscr{E})^n(1 + o(1))$$

as $n \to \infty$ *for any bounded function* ϕ *on* $[0, \infty[$ *which is continuous* a.e. *Lebesgue and such that* $\sup_{x>0} |\phi(x)/x| < \infty$. *Or equivalently*

$$\int \rho\left(\frac{dP_2^n}{dP_1^n + dP_2^n}\right) d(P_1^n + P_2^n) = \int_0^1 \rho(x) \frac{(1+x)^{t_0}}{x^{1+t_0}} dx \frac{1}{\sqrt{2\pi\tau^2}} \frac{1}{\sqrt{n}} C(\mathscr{E})^n (1 + o(1))$$

as $n \to \infty$ *for any function* ρ *on* [0, 1] *which is continuous* a.e. *Lebesgue and such that* $\sup_{1>x>0} |\rho(x)/x| < \infty$ *and* $\sup_{0<x<1} |\rho(x)/(1-x)| < \infty$.

If ψ is sublinear or superlinear on R^2 then $\rho(x) = \psi(1-x, x)$; $x \in [0, 1]$ satisfies the requirements of the Theorem. Thus, by specializing to functions $x \rightsquigarrow \wedge_t \sum_{\theta=1}^2 L_\theta(t) x_\theta$, we find asymptotic expressions for minimum Bayes risk in various decision problems.

4. Replicated experiments when the parameter set is finite. How fast does the content of information in n replicates of an experiment $\mathscr{E}$ increase as $n \uparrow \infty$? In this section we shall investigate this question when Θ is finite. In view of the fact that pairwise sufficiency implies sufficiency, it is not too surprising that, up to a first approximation, the problem may be reduced to the same problem for dichotomies. A few crude, but for our purposes sufficient, inequalities are collected in

PROPOSITION 4.1.

(i) *Suppose* $\lambda(\Theta_0) > 0$ *and let* λ^0 *be the conditional distribution on* Θ_0 *given "*$\theta \in \Theta_0$*." Then*

$$\lambda(\Theta_0) b(\lambda^0 \mid \mathscr{E}_{\Theta_0}) \leqq b(\lambda \mid \mathscr{E}).$$

(ii) *If* λ *is nondegenerate then*

$$2b(\lambda \mid \mathscr{E}) \leqq \sum_{\theta_1 \neq \theta_2} (\lambda_{\theta_1} + \lambda_{\theta_2}) b\left(\frac{\lambda_{\theta_1}}{\lambda_{\theta_1} + \lambda_{\theta_2}}, \frac{\lambda_{\theta_2}}{\lambda_{\theta_2} + \lambda_{\theta_1}} \,\middle|\, \mathscr{E}_{\{\theta_1, \theta_2\}}\right).$$

(iii) *Suppose* Θ *has* $m < \infty$ *elements. Then*

$$\delta_\alpha(\mathscr{E}) \leqq \min\left\{2mb\left(\frac{1}{m}, \cdots, \frac{1}{m} \,\middle|\, \mathscr{E}\right), (m-1)\max_{\theta_1 \neq \theta_2} \delta_\alpha(\mathscr{E}_{\{\theta_1, \theta_2\}})\right\}.$$

PROOF. (i) and part of (iii) follows from the inequalities $\sum_{\Theta_0} \lambda(\theta) r(\theta) \leqq \sum \lambda(\theta) r(\theta) \leqq \sum r(\theta)$ which are valid for any nonnegative risk function r. (ii) follows from the inequality

$$\sum \lambda_\theta f_\theta - \vee \lambda_\theta f_\theta \leqq \tfrac{1}{2} \sum_{\theta_1 \neq \theta_2} \lambda_{\theta_1} f_{\theta_2} \wedge \lambda_{\theta_2} f_{\theta_2}$$

valid for any nonnegative numbers f_θ, $\theta \in \Theta$. (iii) follows now from (ii) and (2.9). □

It will be assumed throughout this section that the parameter set Θ is finite. The θth unit vector e^θ in R^Θ is defined by $e^\theta(\theta') = 1$ or 0 as $\theta' = \theta$ or $\theta' \neq \theta$. We extend the definition of the constant $C(\mathscr{E})$ in Section 3 by defining

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

Thus

$$C(\mathscr{E}) = \max_{\theta_1 \neq \theta_2} C(\mathscr{E}_{\{\theta_1, \theta_2\}}).$$

Consider now an experiment $\mathscr{E} = (P_\theta; \theta \in \Theta)$. Let ψ be sublinear on R^Θ and let F be a nonempty subset of Θ. Then, by sublinearity $\psi(z) = \psi(\sum z_\theta e^\theta) \leqq \psi(\sum_F z_\theta e^\theta) + \sum_{F^c} z_\theta \psi(e^\theta)$. Let S denote the standard measure of $\mathscr{E}$. Then

$$\psi(\mathscr{M}_a) - \psi(\mathscr{E}) = \sum_\theta \psi(e^\theta) - \int \psi \, dS = \int [\sum z_\theta \psi(e^\theta) - \psi(\sum z_\theta e^\theta)] S(dz)$$

$$\geqq \int [\textstyle\sum_F z_\theta \psi(e^\theta) - \psi(\sum_F z_\theta e^\theta)] S(dz) = \psi_F(\mathscr{M}_a) - \psi_F(\mathscr{E})$$

where $\psi_F(z) = \psi(\sum_F z_\theta e^\theta)$. If in particular, $F = \{\theta_1, \theta_2\}$ then $\psi(\mathscr{M}_a) - \psi(\mathscr{E}) \geqq \psi_{\{\theta_1,\theta_2\}}(\mathscr{M}_a) - \psi_{\{\theta_1,\theta_2\}}(\mathscr{E})$. Substituting $\mathscr{E}^n$ for $\mathscr{E}$ and applying Theorem 3.9 we find, provided $\psi_{\{\theta_1,\theta_2\}}$ is not affine on $[0, \infty[^\Theta$, that $\liminf_n [\psi(\mathscr{M}_a) - \psi(\mathscr{E}^n)]^{1/n} \geqq C(\mathscr{E}_{\{\theta_1,\theta_2\}})$. Suppose now that this provision is satisfied for all two point sets $\{\theta_1, \theta_2\}$. Then

$$\liminf_n [\psi(\mathscr{M}_a) - \psi(\mathscr{E}^n)]^{1/n} \geqq C(\mathscr{E}).$$

The provision above is obviously satisfied for any function $z \rightsquigarrow \vee_\theta \lambda_\theta z_\theta$ where λ is a prior distribution on Θ such that $\lambda_\theta > 0$ for all θ. Suppose λ satisfies this condition. Then the above results imply that

$$\liminf_n b(\lambda \mid \mathscr{E}^n)^{1/n} \geqq C(\mathscr{E}).$$

By Proposition 4.1 and Theorem 3.5

$$\limsup_n b(\lambda \mid \mathscr{E}^n)^{1/n} \leqq \max_{\theta_1 \neq \theta_2} C(\mathscr{E}_{\{\theta_1,\theta_2\}}) = C(\mathscr{E}).$$

It follows that $b(\lambda \mid \mathscr{E}^n)^{1/n} \to C(\mathscr{E})$ as $n \to \infty$. Hence, by (2.9) and Proposition 4.1

$$C(\mathscr{E}) = \lim_n b(\text{uniform} \mid \mathscr{E}^n)^{1/n} \leqq \liminf_n [\delta_a(\mathscr{E}^n)/2]^{1/n}$$

$$\leqq \limsup_n [\delta_a(\mathscr{E}^n)/2]^{1/n} \leqq \limsup_n [(\#\Theta) b(\text{uniform} \mid \mathscr{E}^n)]^{1/n} = C(\mathscr{E}),$$

so that

$$\lim_n \delta_a(\mathscr{E}^n)^{1/n} = C(\mathscr{E}).$$

By the sublinear function criterion

$$\limsup_n [\psi(\mathscr{M}_a) - \psi(\mathscr{E}^n)]^{1/n} \leqq \limsup_n [\textstyle\sum_\theta \tfrac{1}{2}[\psi(e^\theta) + \psi(-e^\theta)] \delta_a(\mathscr{E}^n)]^{1/n}$$

$$\leqq \limsup_n \delta_a(\mathscr{E}^n)^{1/n} = C(\mathscr{E})$$

for any sublinear function ψ on R^Θ. Altogether we have proved the following.

THEOREM 4.2. *Let $\mathscr{E} = (P_\theta; \theta \in \Theta)$ be an experiment with finite parameter set. Then*
(i) $\lim_{n\to\infty} \delta_a(\mathscr{E}^n)^{1/n} = C(\mathscr{E})$.
(ii) $\lim_{n\to\infty} b(\lambda \mid \mathscr{E}^n)^{1/n} = C(\mathscr{E})$ *provided $\lambda_\theta > 0$ for all θ.*
(iii) $\limsup_{n\to\infty} [\psi(\mathscr{M}_a) - \psi(\mathscr{E}^n)]^{1/n} \leqq C(\mathscr{E})$ *for any sublinear function ψ.*
(iv) $\lim_{n\to\infty} [\psi(\mathscr{M}_a) - \psi(\mathscr{E}^n)]^{1/n} = C(\mathscr{E})$ *for any sublinear function ψ on R^Θ such that none of the maps $z \rightsquigarrow \psi(z_{\theta_1} e^{\theta_1} + Z_{\theta_2} e^{\theta_2})$; $\theta_1 \neq \theta_2$ are linear on $[0, \infty[^\Theta$.*

REMARK. Using the inequalities mentioned after (2.11) it may be shown that $[2 - 2/m - \delta_i(\mathscr{E}^n)]^{1/n} \to C(\mathscr{E})$.

If $\psi(x) = \vee_t \sum_\theta \lambda_\theta U_\theta(t) x_\theta$ where T is a decision space and U is the utility function, then (iv) describes the exponential rate of convergence to $\sum_\theta \lambda_\theta \vee_\theta U_\theta(t)$ of maximum Bayes utility. If the $\vee_t$ in the expression for ψ is attained and $\lambda_\theta > 0$ when $\theta \in \Theta$, then the exceptional case is precisely the situation where for some two point set $\{\theta_1, \theta_2\}$ no observations are needed when it is known that $\theta \in \{\theta_1, \theta_2\}$. Thus (iv) is not applicable to expressions like $\| \wedge_\theta \lambda_\theta P_\theta^n \| = \| \vee_\theta (-\lambda_\theta) P_\theta^n \|$ when $\#\Theta \geqq 3$ and, in fact, $\| \wedge_\theta \lambda_\theta P_\theta^n \| \leqq \prod \lambda_\theta^{t_\theta} H_\theta(t)^n$ for any pair (λ, t) of prior distributions on Θ.

Although Theorem 4.2 yields the exact rate of exponential convergence in many situations, there are situations of interest where the condition in (iv) is not satisfied. Consider, for example, the problem of catching θ with an r-point confidence set. Then the minimax probability of not covering the true value is

$$\kappa_r(\mathscr{E}) = 1 - \inf_\lambda \| \vee_U \textstyle\sum_{\theta\in U} \lambda_\theta P_\theta \|,$$

where U runs through all r-point subsets of Θ. Clearly $\kappa_1 = \delta(\mathscr{E}, \mathscr{M}_a)/2$ and it is easily seen that κ_r is monotonically decreasing in r.

The minimum Bayes probability of not covering θ for the prior λ is $1 - \| \vee_U \sum_{\theta\in U} \lambda_\theta P_\theta \| = \| \wedge_U \sum_{\theta\in U^c} \lambda_\theta P_\theta \|$. This minimum is achieved by the confidence set $x \rightsquigarrow U_x$ if and only if the class of sets U such that $\sum_{\theta\in U} \lambda_\theta f_\theta(x) = \vee_U \sum_{\theta\in U} \lambda_\theta f_\theta(x)$ has the probability 1 for $\sum_\theta \lambda_\theta P_\theta$ almost all x. Here $f_\theta = dP_\theta / d\sum_\theta P_\theta$, $\theta \in \Theta$.

Let us briefly consider the asymptotic behavior of these quantities.

THEOREM 4.3. *Suppose Θ is finite and put $m = \#\Theta$. Define for each experiment and each integer $r \in \{1, 2, \cdots, m-1\}$ the quantity $\kappa_r(\mathscr{E})$ as above. Then*

$$\lim_{n\to\infty} \kappa_r(\mathscr{E}^n)^{1/n} = \max_W \inf_{t\in\Lambda_W} \int \prod_W (dP_\theta)^{t_\theta}$$

where

(i) *W runs through all $(r+1)$-point subset of Θ and*

(ii) *Λ_W, for each W, is the set of all prior distributions on Θ which are supported by W.*

Furthermore, the nth root of the minimum Bayes probability in $\mathscr{E}^n$ of not covering θ for the prior λ converges to the same limit, provided $\lambda_\theta > 0$ for all $\theta \in \Theta$.

REMARK. Suppose the prior distribution λ assigns positive mass to each $\theta \in \Theta$. Then any sequence $U_1, U_2, \cdots$ of confidence sets such that U_n, for each n, is an optimal Bayes procedure based on $\mathscr{E}^n$, achieves the optimal rate.

Putting $m = 2$ and $r = 1$ we see that the last statement generalizes Chernoff's result, Theorem 3.7. The proof of Theorem 4.3 is based on Theorem 4.4 and Proposition 4.5 below.

THEOREM 4.4. *Let μ_θ; $\theta \in \Theta$ be a finite family of finite nonnegative measures on a common measurable space. Then*

$$\lim_{n\to\infty} \| \wedge_\theta \mu_\theta^n \|^{1/n} = \sup_n \| \wedge_\theta \mu_\theta^n \|^{1/n} = \inf_{t\in\Lambda} \int \prod_\theta (d\mu_\theta)^{t_\theta}$$

REMARK. The statement of the Theorem is clearly equivalent to the following.

Let $\mathscr{E} = (P_\theta; \theta \in \Theta)$ be an experiment with finite parameter set Θ and let a_θ; $\theta \in \Theta$ be a family of nonnegative numbers. Then

$$\lim_{n\to\infty} \| \wedge_\theta a_\theta^n P_\theta^n \|^{1/n} = \sup_n \| \wedge_\theta a_\theta^n P_\theta^n \|^{1/n} = \inf_{t\in\Lambda} \prod_\theta a_\theta^{t_\theta} H(t \mid \mathscr{E}).$$

PROOF. As $\wedge_\theta f_\theta \leqq \prod_\theta f_\theta^{t_\theta}$, $f \in [0, \infty[^\Theta$, $t \in \Lambda$ we get

$$\| \wedge_\theta \mu_\theta^n \| \leqq \int \prod_\theta d(\mu_\theta^n)^{t_\theta} = \left(\int \prod_\theta d\mu_\theta^{t_\theta} \right)^n.$$

It remains therefore to show that $\liminf_n \| \wedge_\theta \mu_\theta^n \|^{1/n} \geqq \inf_t \int \prod_\theta d\mu_\theta^{t_\theta}$. Note that both sides of this inequality remains unchanged if each μ_θ is replaced by its $\wedge_\theta \mu_\theta$ absolutely continuous component. This reduces the problem to showing that

$$\liminf \| \wedge_{i=1}^m e^{-na_i} P_i^n \|^{1/n} \geqq \inf_t e^{-(a,t)} H(t \mid P_1, \cdots, P_m) \tag{4.1}$$

for any homogeneous experiment $\mathscr{E} = (P_1, \cdots, P_m)$ and any $a \in R^m$ such that $a_1 = 0$. If $m = 2$ then (4.1) is (3.1) in our proof of Chernoff's result in the previous section and the

general case may be proved quite similarly. Again we may assume that the inf is obtained at a $t^0 \in \Lambda^0$ and we may also assume that $\mathscr{L}(\log(dP_i/dP_1), i = 2, \cdots, m \mid P_1)$ is nonsingular. By considering the Hellinger transforms it follows again that $(\mathscr{E}^n: t^0 + s'/\sqrt{n}, \cdots t^0 + s^m/\sqrt{n})$ converges to an experiment $\mathscr{G} = (Q_1, \cdots, Q_m)$, where (see Example 4.6) Q_i, $i = 1, 2, \cdots, m$, are the $(m-1)$-variate multinormal distributions with means $(s_2^i, s_3^i, \cdots, s_m^i)$ and covariance matrix σ^{-1} given by $\sigma_{\alpha\beta} = \rho_0^{-1} \dfrac{\partial^2}{\partial t_\alpha \partial t_\beta} \rho \mid_{t=t_0}$, $\rho(t) = e^{-(a,t)} H(t \mid \mathscr{E})$ and $\rho_0 = \inf_t \rho(t)$. If $U_{ni} = dP_i^n/dP_1^n$ and $U_i = dQ_i/dQ_1$ then this implies that

$$E_{P_1^n}\phi(U_{ni}^{1/\sqrt{n}} e^{-\sqrt{n}a_i}; i = 2, \cdots, m) \prod_2^m U_{ni}^{t_i^0} \sim \rho_0^n e^{nat^0} E_{Q_1}\phi(e^{1/2\sigma_{ii}} U_i, i = 2, \cdots, m)$$

when ϕ is bounded and continuous a.e. If $\phi(x_2, \cdots, x_m) = 1$ or 0 as $(x_2, \cdots, x_m) \in [M^{-1}, M]^{m-1}$ or not then we get $\| \wedge_i e^{-na_i} P_i^n \| \geqq \gamma_n$ where $\gamma_n \sim M^{-\sqrt{n}(2-t_1^0)} \rho_0^n E_{Q_1}\phi(e^{\sigma_{ii}/2} U_i, i = 2, \cdot, m)$ and this yield the desired inequality. □

PROPOSITION 4.5. *For any experiment* $\mathscr{E} = (P_1, P_2, \cdots, P_m)$, *if* $1 \leq r < m$ *then*

$$\sum_{B_{r+1}} \binom{m-1}{r}^{-1} \kappa_r(\mathscr{E}_{\{i_1, i_2, \cdots, i_{r+1}\}}) \leqq \kappa_r(\mathscr{E}) \leqq \sum_{B_{r+1}} \kappa_r(\mathscr{E}_{\{i_1, i_2, \cdots, i_{r+1}\}}),$$

where $B_{r+1} = \{(i_1, \cdots, i_{r+1}): i_1 < i_2 < \cdots < i_{r+1}\}$.

PROOF. We may write $\kappa_r(\mathscr{E}) = \sup_\lambda \| \wedge_U \sum_{U^c} \lambda_\theta P_\theta \|$. Let $y_1 \leqq y_2 \leqq y_3 \leqq \cdots \leqq y_m$. Then $\wedge_U \sum_{U^c} y_\theta = y_1 + \cdots y_s$, where $s = m - r$. The inequalities follow now immediately from the identity $\sum \{y_{i_1}: i_1 < i_2 < \cdots i_{r+1}\} = \binom{m-1}{r} y_1 + \binom{m-2}{r} y_2 + \cdots + \binom{r}{r} y_s$. □

PROOF OF THEOREM 4.3. The Theorem follows by applying Proposition 4.5 to $\mathscr{E}^n$, taking the nth root, letting $n \to \infty$ and using Theorem 4.4. □

EXAMPLE 4.6. Let P_i, $i = 1, \cdots, m$, be the p-variate normal distributions with means ξ^i and nonsingular covariance matrix M. Then $H(t \mid P_1, \cdots, P_m) = e^{-1/4Q}$ where $Q = \sum t_i t_j (\xi^i - \xi^j)' M^{-1} (\xi^i - \xi^j)$. If, in particular, M is the identity matrix and $\xi^1, \cdots, \xi^m$ is an orthonormal basis for R^m then

$$\lim_{n\to\infty} \frac{1}{n} \log \kappa_r(P_1^n, P_2^n, \cdots, P_m^n) = -\frac{r}{2(r+1)}.$$

If the map $\theta \rightsquigarrow P_\theta$ is not $1-1$ then $C(\mathscr{E}) = 1$ and $\delta(\mathscr{E}^n, \mathscr{M}_a) \geqq 1$ for all n. The obvious way out is to replace the parameter set Θ by the set $\{P_\theta; \theta \in \Theta\}$. We omit the details.

5. The general case. It follows, since deficiencies decrease by taking restrictions, that $\liminf_n \delta_a(\mathscr{E}^n)^{1/n} \geqq C(\mathscr{E})$ for any experiment $\mathscr{E} = (P_\theta; \theta \in \Theta)$ where $C(\mathscr{E}) = \sup_{\theta_1 \neq \theta_2} C(\mathscr{E}_{\{\theta_1, \theta_2\}})$. Although providing a lower bound, $C(\mathscr{E})$ alone does not determine the exponential rate of convergence when Θ is infinite. We shall, nevertheless, see that $\delta_a(\mathscr{E}^n)^{1/n}$ always converges as $n \to \infty$ to the quantity $\sigma(\mathscr{E}) = \max(C(\mathscr{E}), \tau(\mathscr{E}))$ where $\tau(\mathscr{E}) = \inf_n (\delta_a(\mathscr{E}^n)/2)^{1/n}$. In order to establish this and related results we shall have to search for estimators $\hat{\theta}_n$ making $P_\theta^n(\hat{\theta}_n \neq \theta)$ small.

Suppose that an estimator $\hat{\theta} = \hat{\theta}(X)$ of θ based on one observation X of $\mathscr{E} = (\chi, \mathscr{A}, P_\theta; \theta \in \Theta)$ is available. Let $X_1, \cdots, X_n$ be n independent observations of X and put $\hat{\theta}_i = \hat{\theta}(X_i)$; $i = 1, 2, \cdots n$. Restrict the search to the subset U of Θ consisting of those θ' whose relative frequency $h_n(\theta')$ in $(\hat{\theta}_1, \hat{\theta}_2, \cdots, \hat{\theta}_n)$ exceed a fixed number $\xi \in]0, 1]$. Then pick an estimator $s_n = s_n(X_1, \cdots, X_n)$ within the subset of U consisting of those θ' which maximizes the likelihood *within* U. Thus s_n is a restricted maximum likelihood estimator.

To be more precise we may proceed as follows. Equip Θ with the σ-algebra generated

by the one-point sets and suppose $\hat{\theta}$ is measurable. Put $g_{\theta,F} = dP_\theta^n / d\sum_{\theta\in F} P_\theta^n$ whenever F is a finite subset of Θ containing θ. Choose for each nonempty finite subset F of Θ a measurable estimator $t_F = t_F(X_1, \cdots, X_n) \in F$ such that $g_{t_F,F} = \vee_{\theta\in F} g_{\theta,F}$. We may then take $s_n = t_U$ provided $U \neq \emptyset$. Put $s_n = \theta_0$ where θ_0 is a fixed element of Θ when $U = \emptyset$. Then s_n is measurable with respect to the completion of $\mathscr{A}^n$ for P_θ^n for each θ and

PROPOSITION 5.1. *Put* $r = r(\xi) = \xi^{-\xi}(1-\xi)^{\xi-1}$, $C_\theta = \sup\{C(\mathscr{E}_{\{\theta,\theta'\}}; \theta' \neq \theta\}$ *and assume* $0 < \eta \leqq \xi \leqq P_\theta(\hat{\theta} = \theta)$. *Then*

$$P_\theta^n(s_n \neq \theta) \leqq [P_\theta(\hat{\theta} \neq \theta)^{1-\xi} r]^n + \eta^{-1}[r\eta^\xi]^n + \eta^{-1}C_\theta^n.$$

PROOF. Put $F_\theta = \{\theta' : P_\theta(\hat{\theta} = \theta') \geqq \eta\}$. Then

$$P_\theta^n(s_n \neq \theta) \leqq P_\theta^n(U \not\subseteq F_\theta) + P_\theta^n(\theta \notin U) + P_\theta^n(\theta \in U \subseteq F_\theta \;\&\; s_n \neq \theta).$$

Now

$$P_\theta^n(U \not\subseteq F_\theta) = \sum{}' P_\theta^n(U \not\ni \theta') = \sum{}' P_\theta^n(h_n(\theta') \geqq \xi)$$

where $\sum' = \sum_{\theta'\notin F_\theta}$. Applying the inequality $P_\theta(Z \geqq C) \leqq \inf_{t\geqq 0} e^{-tC} E_\theta e^{tZ}$ to each term we find

$$P_\theta^n(U \not\subseteq F_\theta) = \sum{}'[P_\theta(\hat{\theta} \neq \theta')^{1-\xi} P_\theta(\hat{\theta} = \theta')^\xi r]^n \leqq \sum{}' P_\theta(\hat{\theta} = \theta') P_\theta(\hat{\theta} = \theta')^{n\xi-1} r^n \leqq \eta^{n\xi-1} r^n$$

provided $n\xi \geqq 1$. Hence $P_\theta^n(U \not\subseteq F_\theta) \leqq \eta^{n\xi-1} r^n$ in any case. Similarly

$$P_\theta^n(\theta \notin U) = P_\theta^n(h_n(\theta) < \xi)) \leqq [P_\theta(\hat{\theta} \neq \theta)^{1-\xi} P_\theta(\hat{\theta} = \theta)^\xi r]^n \leqq [P_\theta(\hat{\theta} \neq \theta)^{1-\xi} r]^n.$$

Finally, by the remark to Theorem 3.8, we find

$$P_\theta^n(\theta \in U \subseteq F_\theta, s_n \neq \theta) = \sum_{\theta'\in F_\theta - \{\theta\}} P_\theta^n(\theta \in U \subseteq F_\theta \;\&\; s_n = \theta')$$

$$\leqq \sum_{\theta'\in F_\theta - \{\theta\}} P_\theta^n(g_{\theta',F_\theta} \geqq g_{\theta,F_\theta}) \leqq \sum_{\theta'\in F_\theta - \{\theta\}} \| P_{\theta'}^n \wedge P_\theta^n \| \leqq \#(F_\theta - \{\theta\}) C_\theta^n \leqq \eta^{-1} C_\theta^n. \quad \square$$

COROLLARY 5.2. *Suppose* $\mathscr{E}^k$ *is* $\varepsilon = (\varepsilon_\theta; \theta \in \Theta)$ *deficient with respect to* $\mathscr{M}_a$ *and let* $0 < \eta \leqq \xi \leqq 1$. *Then* $\mathscr{E}^n$ *is* $(\gamma_{n\theta}; \theta \in \Theta)$ *deficient with respect to* $\mathscr{M}_a$ *where*

$$\tfrac{1}{2}\gamma_{n\theta} = [(\varepsilon_\theta/2)^{1-\xi} r]^{[n/k]} + \eta^{-1}[r\eta^\xi]^{[n/k]} + \eta^{-1} C_\theta^{[n/k]}$$

or $= 1$ *as* $\xi_\theta \leqq 1 - (\varepsilon_\theta/2)$ *or not.*

REMARK. Choosing $\eta = 1/n$ and $\xi = (\log n)^{-1/2}$ we find that there is a sequence $s_1, s_2, \cdots$ of estimators such that

$$\limsup_n P_\theta^n(s_n \neq \theta)^{1/n} \leqq (\varepsilon_\theta/2)^{1/k} \vee C_\theta.$$

Here is the main result of this section.

THEOREM 5.3.

$$\delta_a(\mathscr{E}^n)^{1/n} \to \sigma(\mathscr{E}) \qquad \text{as } n \to \infty.$$

REMARK. It follows that $\liminf_n \sup_\theta P_\theta^n(s_n \neq \theta)^{1/n} \geqq \sigma(\mathscr{E})$ for any sequence $s_n(X_1, \cdots, X_n)$; $n = 1, 2, \cdots$ of estimators of θ and that $\lim_n \sup_\theta P_\theta^n(s_n \neq \theta)^{1/n} = \sigma(\mathscr{E})$ for some sequence $(s_1, s_2, \cdots)$ of estimators.

PROOF. Clearly $\liminf_n \delta_a(\mathscr{E}^n)^{1/n} \geqq \sigma(\mathscr{E})$. It remains to show that $\limsup_n \delta_a(\mathscr{E}^n)^{1/n}$

$\leqq C(\mathscr{E}) \vee (\varepsilon_k/2)^{1/k}$ when $\varepsilon = \delta_a(\mathscr{E}^k) < 2$. This, however, follows from the Corollary by letting $\eta \to 0$ and then $\xi \to 0$ in the inequality $\limsup_n \gamma_{n\theta}^{1/n} \leqq ((\varepsilon/2)^{1-\xi} r)^{1/k} \vee (r\eta^{\xi})^{1/k} \vee C(\mathscr{E})$, $\eta < \xi < 1 - \varepsilon/2$. □

Applying the Theorem to $\mathscr{E}^k$ we find easily that $\sigma(\mathscr{E}^k) = \sigma(\mathscr{E})^k$ and clearly $\sigma(\mathscr{E}) \leqq \sigma(\mathscr{F})$ when $\mathscr{E} \geqq \mathscr{F}$.

COROLLARY 5.4. *The following conditions are equivalent for an experiment* $\mathscr{E} = (P_\theta; \theta \in \Theta)$,

(i) $\delta_a(\mathscr{E}^n) < 1$ *for some* n;

(ii) $\lim_n \delta_a(\mathscr{E}^n) = 0$;

(iii) $\delta_a(\mathscr{E}^n) \leqq c\rho^n$, $n = 1, 2, \cdots$ *for some constant* $c > 0$ *and some constant* $\rho < 1$;

(iv) $\delta_a(\mathscr{E}^n) < 2$ *for some* n *and* $\inf_{\theta_1 \neq \theta_2} \| P_{\theta_1} - P_{\theta_2} \| > 0$.

PROOF. Suppose $\delta_a(\mathscr{E}^r) < 1$. Then $\tau(\mathscr{E}) \leqq \delta_a(\mathscr{E}^r)^{1/r} < 1$ and $\sup_{\theta_1 \neq \theta_2} \delta_a(\mathscr{E}^r_{\{\theta_1,\theta_2\}}) < 1$. The last inequality is equivalent to $\inf_{\theta_1 \neq \theta_2} \delta_t(\mathscr{E}^r_{\{\theta_1,\theta_2\}}) > 0$ or $C(\mathscr{E})^r = \sup_{\theta_1 \neq \theta_2} C(\mathscr{E}^r_{\{\theta_1,\theta_2\}}) < 1$. Hence, as $\tau(\mathscr{E}) < 1$ and $C(\mathscr{E}) < 1$, $\sigma(\mathscr{E}) < 1$. Thus, by the Theorem, (i) $\Rightarrow$ (iii). The other implications of the Corollary are then straightforward. □

It follows that if $\mathscr{E}^n \to \mathscr{M}_a$ then the speed of convergence is necessarily exponential provided $\mathscr{E} \nsim \mathscr{M}_a$.

The constant 1 in (i) cannot be increased. If, for example, $\mathscr{E} = (P_1, P_2, P_3, \cdots)$, where $P_1 = P_2$ and $P_i \wedge P_j = 0$ when $i, j \geqq 2$, and $\delta_a(\mathscr{E}^n) = 1$ for all $n = 1, 2, \cdots$.

We saw in the previous section that $C(\mathscr{E})$ alone determined the rate of convergence when Θ is finite. Here is an example, showing that this (i.e., $C(\mathscr{E}) \geqq \tau(\mathscr{E})$) does not hold in general when Θ is infinite.

EXAMPLE 5.5. Suppose $\Theta = \{1, 2, \cdots\}$ and that the density of P_θ with respect to the uniform distribution P on $[0, 1]$ is f_θ where $f_\theta(x) = 2$ or $= 0$ according to whether x belongs to one of the intervals $[(k-1)/2^\theta, (k-\tfrac{1}{2})/2^\theta[$; $k = 1, \cdots, 2^\theta$ or not. Then $E_P f_{\theta_1}(X)^{1-t} f_{\theta_2}(X)^t = \tfrac{1}{2}$, $\theta_1 \neq \theta_2$, $0 < t < 1$. It follows that the dichotomies $\mathscr{E}_{\{\theta_1,\theta_2\}}$; $\theta_1 \neq \theta_2$ are all equivalent to the simple dichotomy $((\tfrac{1}{2}, 0, \tfrac{1}{2}), (0, \tfrac{1}{2}, \tfrac{1}{2}))$ and, in particular, that $C(\mathscr{E}) = \tfrac{1}{2}$. On the other hand it is not difficult to see that $P_\theta(B) \to P(B)$ as $\theta \to \infty$ for any Borel set B. By the Proposition below $\delta_a(\mathscr{E}^n) \equiv_n 2$.

PROPOSITION 5.6. $\delta_a(\xi^n) \equiv_n 2$ *provided either*

(i) Θ *contains an uncountable subset* Θ_0 *such that* $(P_\theta; \theta \in \Theta_0)$ *is dominated, or*

(ii) Θ *contains distinct elements* $\theta_1, \theta_2, \cdots$ *such that* $\lim_n P_{\theta_n}(A)$ *exists for all* $A \in \mathscr{A}$.

REMARK. Let μ be a probability measure dominating $\mathscr{E}$ and put $f_\theta = dP_\theta/d\mu$. Then condition (ii) is satisfied whenever Θ is infinite and $(f_\theta, \theta \in \Theta)$ are uniformly integrable. Thus $\delta_a(\mathscr{E}^n) \equiv_n 2$ when the sample space of $\mathscr{E}$ is finite and Θ is infinite. In spite of this $\mathscr{E}^n$ converges weakly, i.e., for restrictions to finite subparameter sets, whenever $\theta \rightsquigarrow P_\theta$ is 1 − 1.

PROOF. The sufficiency of condition (i) was noted after (2.7). We may then without loss of generality assume that $\Theta = \{1, 2, \cdots\}$ and that $P_n(A) \to P(A)$ for any event A. If M is an estimator of θ then $P_n(M = n) \leqq P_n(M \geqq N) \to 0$ as $n \to \infty$ and then $N \to \infty$. Hence $\inf_\theta P_\theta(M = \theta) = 0$ so that $\delta_a(\mathscr{E}) = 2$. The Proposition follows, since by the Vitali-Hahn-Saks Theorem, $P_n^r \to P^r$ in the same sense. □

If Θ is finite and $\delta_a(\mathscr{E})$ equals its maximal value then $\mathscr{E} \sim \mathscr{M}_i$ and replications do not yield any information. In the infinite case, however, such a discouraging start does not (see

Example 5.9) prevent information from being accumulated rapidly and with arbitrarily small $\sigma(\mathscr{E})$, as the number of replications increase.

How do distances and deficiencies for replicated experiments behave in general? It is clear by Theorem 5.3 and the inequality $\delta_a(\mathscr{E}^n) - \delta_a(\mathscr{F}^n) \leqq \delta(\mathscr{E}^n, \mathscr{F}^n) \leqq \delta_a(\mathscr{E}^n)$ that $\lim \sup_n \delta(\mathscr{E}^n, \mathscr{F}^n)^{1/n} \leqq \sigma(\mathscr{E})$ and that $\delta(\mathscr{E}^n, \mathscr{F}^n)^{1/n} \to \sigma(\mathscr{E})$ when $\sigma(\mathscr{E}) > \sigma(\mathscr{F})$. The problem of the asymptotic behavior of $\delta(\mathscr{E}^n, \mathscr{F}^n)$ when $\sigma(\mathscr{E}) < \sigma(\mathscr{F})$ is open. Our knowledge about the asymptotic behaviour of Δ distances is more complete. It follows from the considerations above that $\lim \sup_n \Delta(\mathscr{E}^n, \mathscr{F}^n)^{1/n} \leqq \sigma(\mathscr{E}) \vee \sigma(\mathscr{F})$ and that $\Delta(\mathscr{E}^n, \mathscr{F}^n)^{1/n} \to \sigma(\mathscr{E}) \vee \sigma(\mathscr{F})$ when $\sigma(\mathscr{E}) \neq \sigma(\mathscr{F})$. If Θ is a two point set then by [26] the above holds whenever $\mathscr{E} \sim \mathscr{F}$. Thus it also holds whenever Θ is finite and $\mathscr{E}_{\{\theta_1,\theta_2\}} \sim \mathscr{F}_{\{\theta_1,\theta_2\}}$ for $\theta_1 \neq \theta_2$.

EXAMPLE 5.7. Blackwell [2] considered the experiments $\mathscr{E}$ and $\mathscr{F}$ given by the matrices

$\mathscr{E}$:

θ \ x	0	1
1	1	0
2	½	½
3	0	1

and $\mathscr{F}$:

θ \ x	0	1
1	1	0
2	½	½
3	½	½

It is easily seen that $\mathscr{E}$ is pairwise more informative than $\mathscr{F}$ and that $H(t \mid \mathscr{E}) \leqq H(t \mid \mathscr{F})$ for all t. Consider, however, decision problems with safety bounds for the risk at $\theta = 2$. Then a very good performance in $\mathscr{E}$ when $\theta = 3$ will easily lead to a bad performance for $\theta = 1$ and vice versa. The experiment $\mathscr{F}$, on the other hand, is obviously not as bad in this respect.

By sufficiency $\mathscr{E}^n$ and $\mathscr{F}^n$ may be reduced respectively to

x :	0,	$*$,	n
$P_{n,1}$:	1 ,	0 ,	0
$P_{n,2}$:	$\frac{1}{2^n}$,	$1 - \frac{1}{2^{n-1}}$,	$\frac{1}{2^n}$
$P_{n,3}$:	0 ,	0 ,	1

and

x :	0 ,	n
$Q_{n,1}$:	1 ,	0
$Q_{n,2}$:	$\frac{1}{2^n}$,	$1 - \frac{1}{2^n}$
$Q_{n,3}$:	$\frac{1}{2^n}$,	$1 - \frac{1}{2^n}$

Let M be the randomization from $\{0, *, n\}$ to $\{0, n\}$ such that $M(n \mid 0) = (2^{2n} + 2^{n+1})^{-1}$, $M(n \mid *) = 1$ and $M(n \mid n) = (2^n + 1 - 2^{-n})(2^n + 2)^{-1}$. Then $\| P_{n,i}M - Q_{n,i} \| = 2(2^{2n} + 2^{n+1})^{-1}$; $i = 1, 2, 3$ so that $\delta(\mathscr{E}^n, \mathscr{F}^n) \leqq 2(2^{2n} + 2^{n+1})^{-1}$.

Now $(1, 1/2^n, 1/2^n)$ is an available power function in $\mathscr{F}^n$. Hence, by testing criterion [22], there is a power function π in $\mathscr{E}^n$ so that

$$\pi(1) \geqq 1 - \frac{\delta}{2}, \qquad \pi(2) \leqq \frac{1}{2^n} + \frac{\delta}{2} \quad \text{and} \quad \pi(3) \geqq \frac{1}{2^n} - \frac{\delta}{2}$$

where $\delta = \delta(\mathscr{E}^n, \mathscr{F}^n)$. We may therefore, by sufficiency, write

$$\frac{1}{2^n} + \frac{\delta}{2} \geqq \pi(2) = \frac{1}{2^n}(a + c) + \left(1 - \frac{1}{2^{n-1}}\right) b \geqq \frac{1}{2^n}(a + c)$$

$$= \frac{1}{2^n}(\pi(1) + \pi(3)) \geqq \frac{1}{2^n}\left[1 - \frac{\delta}{2} + \frac{1}{2^n} - \frac{\delta}{2}\right].$$

where $a, b, c \in [0, 1]$. It follows that $\delta \geqq 2(2^{2n} + 2^{n+1})^{-1}$ so that $\delta(\mathscr{E}^n, \mathscr{F}^n) = 2(2^{2n} + 2^{n+1})^{-1}$. Hence $\delta(\mathscr{E}^n, \mathscr{F}^n)^{1/n} \to \frac{1}{4}$ while $C(\mathscr{E}) = \frac{1}{2}$ and $C(\mathscr{F}) = 1$.

EXAMPLE 5.8. (Translation experiments on the integers.) For each distribution P on the integers and each integer θ, let P_θ be the right θ-translate of P. Consider the experiment $\mathscr{E}_P = (P_\theta; \theta \in \Theta)$. It was shown in [25] that

$$\delta_a(\mathscr{E}^n) = 2(1 - \textstyle\sum_{x_2,\ldots,x_n} \max_n P(x)P(x + x_2) \cdots P(x + x_n)).$$

It follows in particular that $\delta_a(\mathscr{E}^n) = 2(1 - P(a))^n$ when $P(a) \geqq P(a+1) \geqq \cdots$ and $\sum_{x=a}^{\infty} P(x) = 1$.

Furthermore it is not difficult to see that there are translation invariant maximum likelihood estimators $\hat{\theta}_n$ based on $\mathscr{E}^n$ and that any such estimator is minimax for the 0 − 1 loss estimation problem i.e., $2P_\theta(\hat{\theta}_n \neq \theta) \equiv_\theta \delta_a(\mathscr{E}^n)$. Thus $\delta_a(\mathscr{E}^n) < 2$ for all n and clearly $\inf_{\theta_1 \neq \theta_2} \| P_{\theta_2} - P_{\theta_1} \| > 0$. Hence $P_\theta(\hat{\theta}_n \neq \theta) = \delta_a(\mathscr{E}^n)/2 \leqq C\rho^n$ for some constant $C > 0$ and for some $\rho < 1$.

In spite of this there are translation experiments $\mathscr{E}_P$ with P nondegenerate such that two replications are not better than one in the sense of minimax risk. For example, if $P(x) = p\left(\frac{1-p}{3-p}\right)^{|x|}$; $x = \cdots, -1, 0, 1, \cdots$ when $p \in [0, 1[$ then X_1 alone is a translation invariant maximum likelihood estimator for θ based on two observations (X_1, X_2).

One would expect that if we are able to do extremely well with two observations, then we should be able to do at least moderately well with one observation. If Θ is finite then this follows immediately from the compactness of Δ-convergence. If the parameter set is infinite, however, then this is not necessarily true. We shall here satisfy ourselves with an example of a translation experiment where we may, on the basis of two observations, guess θ with marvelous accuracy while any estimator based on one observation is quite inaccurate. To make things more concrete we may choose the constant so that the probability of a wrong guess for a translation invariant maximum likelihood estimator based on two observations is less than 10^{-200} while, on the other hand, the probability of making a wrong guess is greater than $1\text{-}10^{-200}$ for some θ for any estimator based on one observation.

EXAMPLE 5.9. Let P be the uniform distribution on $\{1, 2, 4, \cdots, 2^{N-1}\}$. Then, by the example in Section 2 in [25],

$$\delta_a(\mathscr{E}_P^n)/2 = P_\theta^n(\hat{\theta}_n \neq \theta) = (N-1)N^{-n}, \qquad n = 1, 2, \cdots$$

where $\hat{\Theta}_n(X_1, \cdots, X_n) = \min \cap_{i=1}^n \{X_i - 1, X_i - 2, X_i - 4, \cdots, X_i - 2^{N-1}\}$. It follows that $\lim_{N\to\infty} \delta_a(\mathscr{E}_P^n) = 1$ or 0 as $n = 1$ or $n \geqq 2$.

If only one observation X_1 is available, then θ is located in the N-point set $\{1 - X_1, 2 - X_1, \cdots, 2^{n-1} - X_1\}$. If, however, another observation X_2 is available and $X_2 \neq X_1$, (this has probability $1 - 1/N$), then θ is completely known. Thus we see that the phenomenon is related to the uniqueness of dyadic expansions.

This example may be sharpened by exhibiting a family $\mathscr{E}_\varepsilon$; $\varepsilon > 0$ of experiments, necessarily not translation experiments on the integers, such that the deficiency $\delta_a(\mathscr{E}_\varepsilon) \equiv_\varepsilon 2$ while $\delta_a(\mathscr{E}_\varepsilon^n) \leqq \varepsilon^{n-1}$ for all $\varepsilon > 0$ when $n \geqq 2$.

REFERENCES

[1] BAHADUR, R. R. (1960). Asymptotic efficiency of tests and estimates. Sankhyā **22** 229–252.

[2] BLACKWELL, D. (1951). Comparison of experiments. *Proc. Second Berkeley Symposium Math. Statist. Prob.* 93–102.

[3] BLACKWELL, D. (1953). Equivalent comparisons of experiments. *Ann. Math. Statist.* **24** 265–272.

[4] BLACKWELL, D. and GIRSHICK, M. A. (1954). *Theory of Game and Statistical Decisions*. Wiley, New York.

[5] BOHNENBLUST, F., SHAPLEY, L., SHERMANN, S. (1951). Unpublished Rand Memorandum.

[6] CHERNOFF, H. (1952). A measure of asymptotic efficiency for tests of a hypothesis based on the sum of observations. *Ann. Math. Statist.* **23** 493–507.

[7] CHERNOFF, H. (1956). Large sample theory-parametric case. *Ann. Math. Statist.* **27** 1–22.

[8] DE GROOT, M. H. (1962). Uncertainty, information and sequential experiments. *Ann. Math. Statist.* **33** 404–419.

[9] DVORETZKY, A. WALD and WOLFOWITZ, J. (1951). Elimination of randomization in certain statistical decision procedures and zero-sum two person games. *Ann. Math. Statist.* **22** 1–21.

[10] EFRON, B. (1967). The power of the likelihood ratio test. *Ann. Math. Statist.* **38** 802–806.

[11] EFRON, B. and TRUAX, D. R. (1968). Deviation theory in exponential families. *Ann. Math. Statist.* **39** 1402–1424.
[12] HELGELAND, J. (1979). Additional observations and statistical information in the case of 1-parameter exponential distribution. *Statist. Res. Report.*, Univ. of Oslo.
[13] HEYER, H. (1973). *Mathematische Theorie Statistischer Experiments.* Springer, Berlin.
[14] LE CAM, L. (1964). Sufficiency and approximate sufficiency. *Ann. Math. Statist.* **35** 1419–1455.
[15] LE CAM, L. (1968). *Théorie Asymptotique de la Décision Statistique.* Les Presses de l'Université de Montréal.
[16] LE CAM, L. (1974). On the information contained in additional observations. *Ann. Statist.* **2** 630–649.
[17] LE CAM, L. (1974). Notes on asymptotic methods in statistical decision theory. Centre de Recherches Math., Univ. de Montréal.
[18] LE CAM, L. (1975). Distances between experiments. In *A Survey of Statistical Design and Linear Models* (J. N. Srivastava, ed.) 383–395. North Holland, Amsterdam.
[19] LINDQVIST, B. (1977). How fast does a Markov chain forget the initial state? A decision theoretic approach. *Scand. J. Statist.* **4** 145–152.
[20] LYAPUNOV, A. (1940). Sur les fonctions-vecteurs completément additives. *Izvestia Akad. Nauk. SSR. Ser. Mat.* **4** 465–478.
[21] MORSE, N. and SACKSTEDER, R. (1966). Statistical isomorphism. *Ann. Math. Stat.* **37** 203–214.
[22] SWENSEN, A. R. (1980). Deficiencies in linear normal experiments. *Ann. Statist.* **8** 1142–1155.
[23] TORGERSEN, E. N. (1970). Comparison of experiments when the parameter space is finite. *Z. Wahrscheinlichkeitstheorie und verw. Gebiete.* **16** 219–249.
[24] TORGERSEN, E. N. (1972a). Local comparison of experiments. Statist. Res. Report., Univ. of Oslo.
[25] TORGERSEN, E. N. (1972b). Comparison of translation experiments. *Ann. Math. Statist.* **43** 1383–1399.
[26] TORGERSEN, E. N. (1974). Asymptotic behaviour of powers of dichotomies. Statist. Res. Report., Univ. of Oslo.
[27] TORGERSEN, E. N. (1975). *Notes on comparison of statistical experiments.* Chapters 0–8. Recorded and supplemented by B. Lindqvist. Statist. Memoir., Univ. of Oslo.
[28] TORGERSEN, E. N. (1976a). Comparison of statistical experiments. *Scand. J. Statist.* **3** 186–208.
[29] TORGERSEN, E. N. (1976b). Deviations from total information and total ignorance as measures of information. *Proc. Fourth Int. Conf. Math. Statist.* Wisla, Poland.
[30] TORGERSEN, E. N. (1977). Mixtures and products of dominated experiments. *Ann. Statist.* **5** 44–64.

INSTITUTE OF MATHEMATICS
UNIVERSITY OF OSLO
P.O. BOX 1053
BLINDERN, OSLO 3
NORWAY