Z. Wahrscheinlichkeitstheorie verw. Geb. 16, 219-249 (1970)

Comparison of Experiments when the Parameter Space is Finite*

ERIK NIKOLAI TORGERSEN

The convex function criterion for "being more informative" for k-decision problems is – in Section 2 – generalized to a convex function criterion for ε-deficiency for k-decision problems. The particular case of comparison by testing problems is discussed in Section 3. A theorem of Blackwell on comparison of dichotomies is generalized and a problem on products of experiments raised by Blackwell is settled by counter-example. Pairwise comparison of experiments and minimal combinations of experiments are discussed. The problem of composing and decomposing experiments by mixtures is treated in Section 4. It is shown that any experiment with finite parameter space is a mixture of complete experiments, and the complete experiments are characterized.

Table of Contents

1. Introduction

In [12] Le Cam introduced the notion of ε-deficiency of one experiment relative to another. This generalized the concept of "being more informative" which was introduced by Bohenblust, Shapley and Sherman and may be found in Blackwell [3]. "Being more informative for k-decision problems" was introduced by Blackwell in [4]. We shall here consider the hybrid of "ε-deficiency for k-decision problems".

An experiment will here be defined as a pair $\mathscr{E}=((\chi,\mathscr{A}),(P_\theta:\theta\in\Theta))$ where $(\chi,\mathscr{A})$ is a measurable space and $(P_\theta:\theta\in\Theta)$ is a family of probability measures on $(\chi,\mathscr{A})$. The set Θ – the parameter set of $\mathscr{E}$ – will be assumed fixed, but arbitrary. An experiment $\mathscr{E}$ as defined above may be identified with an experiment $\underline{\mathscr{E}}$ in the sense of [12] by taking $\underline{\mathscr{E}}=(\Theta,E,\chi,\{P_{-\theta}\})$ where E is the set of bounded measurable functions on $(\chi,\mathscr{A})$ and $P_{-\theta}$ for each θ is given by $fP_{-\theta}=\int f(x)\,P_\theta(dx)$; $f\in E$.

* This paper is – essentially – the author's dissertation submitted in partial fulfilment of the requirements for the Ph.D. degree in Statistics at the University of California, Berkeley, June 1968. The paper was prepared with the partial support of Norges Almenvitenskapelige Forskningsråd, the University of Oslo and U.S. Army Research Grant DA-31-124-AROD-548, Da-ARO-D-31-124-G816.

Definition. 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 (and let $k\geqq 2$ be an integer).

Then we shall say that $\mathscr{E}$ is ε-deficient relative to $\mathscr{F}$ (for k-decision problems[1]) if to each decision space[2] $(D,\mathscr{S})$ where $\mathscr{S}$ is finite (where $\mathscr{S}$ contains at most 2^k sets), every bounded loss function[3] $(\theta,d)\rightsquigarrow W_\theta(d)$ on $\Theta\times D$ and every risk function r obtainable in $\mathscr{F}$ there is a risk function r' obtainable in $\mathscr{E}$ so that

$$r'(\theta)\leqq r(\theta)+\varepsilon_\theta\,\|W\|,\qquad \theta\in\Theta \tag{1}$$

where $\|W\|=\sup_{\theta,d}|W_\theta(d)|$.

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 $\mathscr{E}\geqq\mathscr{F}\ (\mathscr{E}\underset{k}{\geqq}\mathscr{F})$.

The greatest lower bound of all constants ε such that $\mathscr{E}$ is ε-deficient relative to $\mathscr{F}$ (for k-decision problems) will be denoted by $\delta(\mathscr{E},\mathscr{F})$, respectively: $\delta_k(\mathscr{E},\mathscr{F})$ and $\max[\delta(\mathscr{E},\mathscr{F}),\delta(\mathscr{F},\mathscr{E})]$ respectively: $\max[\delta_k(\mathscr{E},\mathscr{F}),\delta_k(\mathscr{F},\mathscr{E})]$ will be denoted by $\Delta(\mathscr{E},\mathscr{F})$ respectively: $\Delta_k(\mathscr{E},\mathscr{F})$.

If $\mathscr{E}$, $\mathscr{F}$ and $\mathscr{G}$ are experiments then: $0\leqq\delta_k(\mathscr{E},\mathscr{F})\leqq\delta_{k+1}(\mathscr{E},\mathscr{F})\leqq\delta(\mathscr{E},\mathscr{F})$, $\delta(\mathscr{E},\mathscr{E})=0$, $\delta_k(\mathscr{E},\mathscr{F})\uparrow\delta(\mathscr{E},\mathscr{F})$ as $k\to\infty$, and $\delta_k(\mathscr{E},\mathscr{F})\leqq\delta_k(\mathscr{E},\mathscr{G})+\delta_k(\mathscr{G},\mathscr{F})$ so that $\Delta_2,\Delta_3,\ldots,\Delta$ are all pseudometrics.

Remark 1. Equivalent definitions may be obtained by replacing $\|W\|$ in (1) by $\|W_\theta\|=\sup_d|W_\theta(d)|$ or by requiring that $W\geqq 0$ and at the same time replacing $\varepsilon_\theta\|W\|$ by $\frac{1}{2}\varepsilon_\theta\|W\|$ (or $\frac{1}{2}\varepsilon_\theta\|W_\theta\|$). This may be seen by noting that if W is a loss function then $(\theta,d)\rightsquigarrow\|W_\theta\|^{-1}W_\theta(d)$ is a loss function bounded by 1, $(\theta,d)\rightsquigarrow W_\theta(d)+\|W\|$ is a nonnegative loss function bounded by $2\|W\|$, and that $W\geqq 0$ implies that $\|\tilde W_\theta\|\leqq\|W_\theta\|$, $\theta\in\Theta$ where $\tilde W_\theta(d)=2W_\theta(d)-\|W_\theta\|$.

If $\mathscr{E}=((\chi,\mathscr{A}),(P_\theta:\theta\in\Theta))$ and $\Theta_0\subseteqq\Theta$, the restricted experiment $((\chi,\mathscr{A}),(P_\theta:\theta\in\Theta_0))$ will be denoted by $\mathscr{E}_{\Theta_0}$ or $\mathscr{E}_{\theta_1,\theta_2}$ if $\Theta_0=\{\theta_1,\theta_2\}$.

From here on – unless otherwise stated – Θ will be assumed finite. The following remark is on the significance of this restriction.

Remark 2. Let Θ be an arbitrary (not necessarily finite) set. If for each finite subset $\tilde\Theta$ and each integer $k\geqq 2$, $\mathscr{E}_{\tilde\Theta}$ is $\varepsilon|_{\tilde\Theta}$ deficient relative to $\mathscr{F}_{\tilde\Theta}$, then – provided certain regularity conditions hold – there is a randomization[4] M from $(\chi,\mathscr{A})$ to $(\mathscr{Y},\mathscr{B})$ such that $(\|\mu\|=\sup(|\int f(x)\,\mu(dx)|:-1\leqq f\leqq 1))$ $\|MP_\theta-Q_\theta\|\leqq\varepsilon_\theta$; $\theta\in\Theta$ or equivalently: to each decision space $(D,\mathscr{S})$ and each operational characteristic $\mathcal{O}(\cdot|\theta)$; $\theta\in\Theta$ available in $\mathscr{F}$ there is an operational characteristic $\mathcal{O}'(\cdot|\theta)$; $\theta\in\Theta$ available in $\mathscr{E}$ such that $\|\mathcal{O}'(\cdot|\theta)-\mathcal{O}(\cdot|\theta)\|\leqq\varepsilon_\theta$; $\theta\in\Theta$. We will refer to this result as the randomization criterion.

[1] When $k=2$: testing problems.

[2] I.e., a measurable space.

[3] It is always to be understood that $d\rightsquigarrow W_\theta(d)$ is measurable for each θ.

[4] I.e. M is a function from $\chi\times\mathscr{B}$ to $[0,1]$ which is measurable in x for fixed B and a probability measure in B for fixed x.

It follows from [12] that a sufficient condition for this to hold is that $(P_\theta: \theta \in \Theta)$ is dominated and that $\mathscr{Y}$ is a complete separable metric space (or a Borel subset of such a space) with Borel class $\mathscr{B}$. [Let π be a probability measure which is equivalent to $(P_\theta: \theta \in \Theta)$ and let $\underline{\mathscr{E}}$ and $\underline{\mathscr{F}}$ be the corresponding experiments in the sense of [12]. The L-space of $\underline{\mathscr{E}}$ is then $L_1(\chi, \mathscr{A}, \pi)$. By Theorem 3 of [12] and its proof, $\underline{\mathscr{E}}$ is ε-deficient relative to $\underline{\mathscr{F}}$ in the sense of [12]. By Proposition 7 of [12] there is a positive normalized linear operator M from $L_1(\chi, \mathscr{A}, \pi)$ to the L-space of $\mathscr{F}$ such that $\|MP_\theta - Q_\theta\| \leqq \varepsilon_\theta$; $\theta \in \Theta$. $M\pi$ and the MP_θ's are probability measures since the Q_θ's are, and the range of M is contained in $L_1(\mathscr{Y}, \mathscr{B}, M\pi)$. By Proposition V 4.4 of [16] the operator M is induced by a randomization.] If $(P_\theta: \theta \in \Theta)$ is dominated then ε-deficiency (for k-decision problems) for all finite subsets of Θ implies—by weak compactness (Proposition IV.2.3 of [16])—ε-deficiency (for k-decision problems).

The product $\prod_{t \in T} \mathscr{E}_t$ of a family $\mathscr{E}_t = ((\chi_t, \mathscr{A}_t); (P_{\theta t}: \theta \in \Theta))$; $t \in T$ of experiments is the experiment $((\prod_t \chi_t, \prod_t \mathscr{A}_t); (\prod_t P_{\theta t}: \theta \in \Theta))$. Products are then commutative and associative up to equivalence. If $\mathscr{E}_1 = \cdots = \mathscr{E}_n = \mathscr{E}$ then $\prod_{i=1}^n \mathscr{E}_i$ will be written $\mathscr{E}^n$.

Let $\mathscr{E} = ((\chi, \mathscr{A}), (P_\theta: \theta \in \Theta))$ be an experiment and let Λ be the set of probability distributions on Θ. The convex extension of $\mathscr{E}$ is defined as the experiment $\hat{\mathscr{E}} = ((\chi, \mathscr{A}); (\int P_\theta \lambda(d\theta); \lambda \in \Lambda))$.

Criteria for $\underset{k}{\geqq}$ were given by Blackwell in [4]. They are generalized in Section 3 to criteria for ε-deficiency for k-decision problems. The methods used (e.g., comparison of Bayes risks) are essentially the same as those of Blackwell in [4], the main difference being that the class of convex functions is replaced by the class of sublinear functions.

It is shown in this section that convergence may be decided by comparison of testing problems alone.

Comparison by testing problems is discussed in Section 4. In [4] Blackwell proved that for dichotomies, "being more informative", was the same as "being more informative for testing problems" and gave a simple criterion in terms of errors of the first and the second kind. It will be shown in Section 4 that this extends to ε-deficiency in a natural way. Blackwell proved in [3] that if $\mathscr{E}_1, \ldots, \mathscr{E}_n$, $\mathscr{F}_1, \ldots, \mathscr{F}_n$ are experiments such that $\mathscr{E}_i \geqq \mathscr{F}_i$; $i = 1, \ldots, n$ then $\prod_{i=1}^n \mathscr{E}_i \geqq \prod_{i=1}^n \mathscr{F}_i$ and raised the problem whether $\mathscr{E}^2 \geqq \mathscr{F}^2$ implies $\mathscr{E} \geqq \mathscr{F}$. The answer is shown to be negative and a counter-example of the apparently weaker statement "$\mathscr{E}^n \geqq \mathscr{F}^n$ from a certain n on implies $\mathscr{E} \geqq \mathscr{F}$" is given.

Since pairwise sufficiency implies sufficiency, one might ask about an analogue for ε-deficiency. It is shown in Section 4 that comparison by testing problems is equivalent to pairwise comparison of the convex extensions of the experiments. It follows from the "error of first and second kind" criterion for dichotomies that a set of dichotomies have inf and sup. It is shown by counterexample that this does not hold when $\# \Theta \geqq 3$.

If a statistician observes a random experiment according to a known (i.e., not depending on θ) distribution on a set $\{\mathscr{E}\}$ of experiments and then performs the observed experiment $\mathscr{E}$, the resulting "total experiment" performed will be called a mixture of the experiments in $\{\mathscr{E}\}$. It is shown by Birnbaum in [1] that any dichotomy is a mixture of a totally ordered set of double dichotomies. It is shown in Section 5 that any experiment is a mixture of complete experiments, and that the complete experiments are characterized as those minimal sufficient experiments whose standard measures are concentrated on the vertices of a simplex.

Some of the notations which will be used are:

$\mathscr{L}(X)=$ the law of X

$\bigvee_t X_t = \sup_t X_t$ and $\bigwedge_t X_t = \inf_t X_t$

$X^+ = X \vee 0$ and $X^- = (-X)^+$

$\mathscr{F}=$ the class of all subsets of Θ.

For each $\theta\in\Theta$ the vector $e_\theta \in R^\Theta$ is defined by:

$e_\theta(\theta')=1$ or 0 as $\theta'=\theta$ or $\theta'\neq\theta$

$e=\sum_\theta e_\theta$

$K=\{x: x\in R^\Theta; x\geqq 0 \text{ and } \sum_\theta x_\theta = 1\}$

$\# A=$ the number of elements in A if A is finite

$\# A=\infty$ if A is infinite

$dP_2/dP_1=$ the density relative to P_1 of the P_1-absolutely continuous part of P_2.

$\|f\|=\sup_x |f(x)|$ and $\|\mu\|=\sup\{|\int f(x)\,\mu(dx)|: \|f\|\leqq 1\}$.

The Lévy distance between distribution functions F and G on R^n:

$$\begin{aligned}\Lambda(F,G)=\inf\{h\colon h\geqq 0,\, F(x_1-h,\dots,x_n-h)-h\leqq G(x_1,\dots,x_n)\\ \leqq F(x_1+h,\dots,x_n+h)+h \quad \text{for all } (x_1,\dots,x_n)\in R^n\}.\end{aligned}$$

If the symbol – "(k)" – appears in a statement, then it may either be replaced throughout the statement by "k" or be deleted throughout the statement.

2. Comparison by k-Decision Problems

Let $k\geqq 2$ be an integer, put $D_k=\{1,\dots,k\}$, let $\mathscr{S}_k$ be the class of all subsets of D_k and consider $(D_k,\mathscr{S}_k)$ as a decision space. Then we have the following average risk criterion for ε-deficiency for k-decision problems.

Theorem 1. *Let $\mathscr{E}=((\chi,\mathscr{A}),(P_\theta\colon\theta\in\Theta))$ and $\mathscr{F}=((\mathscr{Y},\mathscr{B}),(Q_\theta\colon\theta\in\Theta))$ be two experiments and $\theta\rightsquigarrow\varepsilon_\theta$ a nonnegative function on Θ.*

Then $\mathscr{E}$ is ε-deficient relative to $\mathscr{F}$ for k-decision problems if and only if:

To each loss function W on $\Theta\times D_k$ and each risk function r available in $\mathscr{F}$ there is a risk function r' available in $\mathscr{E}$ so that:

$$\sum_\theta r'(\theta)\leqq\sum_\theta r(\theta)+\sum_\theta \varepsilon_\theta\|W_\theta\|;\qquad \theta\in\Theta. \tag{1}$$

Remark. As in Remark 1 after the definition we may restrict attention to non negative W's provided ε_θ is replaced by $\varepsilon_\theta/2$. In this case (1) may be replaced by the minimax criterion $\bigvee_\theta r'(\theta) \leqq \bigvee_\theta (r(\theta)+(\varepsilon_\theta/2)\,\|W_\theta\|)$.

We may – in (1) – restrict attention to W's such that

$$\bigvee_d W_\theta(d) + \bigwedge_d W_\theta(d) = 0; \qquad \theta\in\Theta, \qquad \text{since } \tilde{W}_\theta(d) = W_\theta(d) - \bigvee_d W_\theta(d) - \bigwedge_d W_\theta(d) \quad (2)$$

satisfies this condition.

Proof of the Theorem. Let σ be any decision procedure in $\mathscr{F}$. By assumption there is – for each W – a decision procedure ρ in $\mathscr{E}$ so that:

$$\sum_\theta (\rho P_\theta W_\theta - \sigma Q_\theta W_\theta - \varepsilon_\theta \|W_\theta\|) \leqq 0 \quad \text{i.e.} \quad \sup_{\|W\|\leqq 1} \min_\rho \sum_\theta \leqq 0.$$

It follows, by weak compactness, – since $\sum_\theta$ is affine in ρ and concave in W – that sup and min may be interchanged – i.e. ρ may be chosen independently of W. This implies $\|\rho P_\theta - \sigma Q_\theta\| \leqq \varepsilon_\theta$; $\theta\in\Theta$. □

Another proof of the randomization criterion described in Remark 2 after the definition follows from this proof. (ρ may always be modified so that $\sum_\theta \rho P_\theta \ll \sum_\theta \sigma Q_\theta$.)

The criterion given in Theorem 1 may also be expressed as a convex function criterion. Let Ψ_k, $k=1,2,\ldots$ be the set of functions on R^Θ which are pointwise maximum of k-linear functionals on R^Θ. Let Ψ be the class of all sublinear functionals on R^Θ; i.e. the class of all functions ψ such that $\psi(tx)=t\psi(x)$ and $\psi(x+y)\leqq \psi(x)+\psi(y)$ when $t\geqq 0$ and $x,y\in R^\Theta$. Clearly $\Psi_1 \subseteqq \Psi_2 \subseteqq \cdots \subseteqq \Psi$ and each $\psi\in\Psi$ may be written as $\lim\uparrow\psi_k$ where $\psi_k\in\Psi_k$, $k=1,2,\ldots$.

Theorem 2. *Let* $\mathscr{E}=((\chi,\mathscr{A}),(P_\theta\colon\theta\in\Theta))$ *and* $\mathscr{F}=((\mathscr{Y},\mathscr{B}),(Q_\theta\colon\theta\in\Theta))$ *be two experiments. Put*

$$f=\left(\frac{dP_\theta}{d\sum_\theta P_\theta}\colon \theta\in\Theta\right), \qquad g=\left(\frac{dQ_\theta}{d\sum_\theta Q_\theta}\colon \theta\in\Theta\right)$$

and let $\theta \rightsquigarrow \varepsilon_\theta$ *be a non-negative function on* Θ. *Then* $\mathscr{E}$ *is* ε*-deficient relative to* $\mathscr{F}$ *(for k-decision problems) if and only if:*

$$\int \psi\circ f\, d\sum_\theta P_\theta \geqq \int \psi\circ g\, d\sum_\theta Q_\theta - \sum_\theta \varepsilon_\theta \psi(e_\theta) \quad (3)$$

for each

$$\psi\in\Psi(\Psi_k)$$

such that

$$\psi(-e_\theta)=\psi(e_\theta); \qquad \theta\in\Theta. \quad (4)$$

Remark 1. If the P_θ's are given by densities $h_\theta\colon\theta\in\Theta$ relative to some positive measure μ and $\psi\in\Psi$, then it follows from the positive homogenity of ψ that $\int\psi\circ f\, d\sum_\theta P_\theta = \int \psi\circ h\, d\mu$ where h *denotes the map:* $x\rightsquigarrow(h_\theta(x)\colon\theta\in\Theta)$.

Remark 2. Equivalent conditions may be obtained by

(i) Requiring that ψ be monotonically increasing (decreasing) dropping the requirement (4) and replacing (3) with

$$\int \psi \circ (-f)\, d\sum_\theta P_\theta \geqq \int \psi \circ (-g)\, d\sum_\theta Q_\theta - \sum_\theta \tfrac{1}{2}\varepsilon_\theta \psi(e_\theta)$$

$$\Big(\int \psi \circ f\, d\sum_\theta P_\theta \geqq \int \psi \circ g\, d\sum_\theta Q_\theta - \sum_\theta \tfrac{1}{2}\varepsilon_\theta \psi(-e_\theta)\Big).$$

(ii) Dropping the requirement (4) and replacing (3) with

$$\int \psi \circ f\, d\sum_\theta P_\theta \geqq \int \psi \circ g\, d\sum_\theta Q_\theta - \sum_\theta \tfrac{1}{2}\varepsilon_\theta (\psi(-e_\theta) + \psi(e_\theta)).$$

(ii) follows by noting that if $\psi \in \Psi$ then

$$x \rightsquigarrow \psi(x) - \sum_\theta \frac{\psi(e_\theta) - \psi(-e_\theta)}{2} x_\theta$$

satisfies (4), and (i) may be deduced from (ii) using the fact that $x \rightsquigarrow \psi(x) - \sum_\theta \psi(e_\theta)\, x_\theta$ is monotonically decreasing in x when $\psi \in \Psi_k$. Since $\psi(e_\theta) \leqq 0$, $\theta \in \Theta$ for each monotonically decreasing $\psi \in \Psi$, (i) implies (ii). Note that the set of ψ's which satisfies (ii) is a convex cone and that ψ satisfies (ii) provided $\psi = \tilde{\psi}$ on K and $\psi(-e_\theta) \geqq \tilde{\psi}(-e_\theta)$; $\theta \in \Theta$ where $\tilde{\psi}$ satisfies (ii).

It may be shown that $\mathscr{E}$ being ε-deficient relative to $\mathscr{F}$ for k-decision problems does not imply (3) for all $\psi \in \Psi_k$.

Proof of the Theorem. Any $\psi \in \Psi_k$ is a maximum of k linear functionals and may therefore be written:

$$\psi(x) = \bigvee_d \Big(\sum_\theta - W_\theta(d)\, x_\theta\Big); \qquad x \in R^\Theta$$

for some constants $W_\theta(d)$; $\theta \in \Theta$, $d \in D$. Since $\psi(-e_\theta) = \bigvee_d W_\theta(d)$ and $\psi(e_\theta) = \bigvee_d - W_\theta(d)$, the condition (4) for ψ is equivalent to the condition (2) (in the remark after Theorem 1) for W. If the condition holds, then $\psi(e_\theta) = \|W_\theta\|$.

Now

$$-\#(\Theta)^{-1} \int \psi \circ f\, d\sum_\theta P_\theta = \#(\Theta)^{-1} \int \bigwedge_d \Big(\sum_\theta W_\theta(d) f_\theta\Big)\, d\sum_\theta P_\theta$$

and

$$-\#(\Theta)^{-1} \int \psi \circ g\, d\sum_\theta Q_\theta$$

are the Bayes risks relative to W for the uniform distribution on Θ. Hence the theorem follows from the remark after Theorem 1. □

Corollary 3. *Let $\Gamma_{(k)}$ be the set of functions $\gamma \in \Psi_{(k)}$ such that $\gamma(-e_\theta) = \gamma(e_\theta)$; $\theta \in \Theta$ and $\sum_\theta \gamma(e_\theta) = 1$. Then $\Delta_{(k)}(\mathscr{E}, \mathscr{F})$ may be written:*

$$\Delta_{(k)}(\mathscr{E}, \mathscr{F}) = \sup_{\gamma \in \Gamma_{(k)}} \Big| \int \gamma \circ f\, d\sum_\theta P_\theta - \int \gamma \circ g\, d\sum_\theta Q_\theta \Big|.$$

$\delta(\mathscr{E}, \mathscr{F})$ clearly attains its maximum when the P_θ's are all equal and the Q_θ's have disjoint supports. In this case $\int \psi \circ f\, d\sum_\theta P_\theta = \psi(e)$ and $\int \psi \circ g\, d\sum_\theta Q_\theta = \sum_\theta \psi(e_\theta)$.

Define γ by

$$\gamma(x)=\frac{2}{n}\bigvee_{\theta} x_\theta-\frac{1}{n}\sum_{\theta} x_\theta \quad \text{where} \quad n=\#\Theta.$$

Then $\gamma\in\Gamma$ and

$$\int\gamma\circ g\,d\sum_{\theta} Q_\theta-\int\gamma\circ f\,d\sum_{\theta} P_\theta=2-\frac{2}{n} \qquad \text{so that } \Delta(\mathscr{E},\mathscr{F})\geqq 2-\frac{2}{n}.$$

(This γ corresponds to the estimation problem $(D,\mathscr{S},W)$ where $D=\Theta$ and $W_\theta(d)=1$ or 0 as $d\neq\theta$ or $d=\theta$.)

On the other hand

$$\bigvee_i \left\|P_i-\frac{1}{n}\sum_i P_i\right\|\leqq 2-\frac{2}{n}$$

for any family $P_1,\ldots,P_n$ of probability measures on $\mathscr{A}$. Hence

$$\delta(\mathscr{E},\mathscr{F})=2-\frac{2}{n}.$$

It will follow from Corollary 6 that

$$\delta_k(\mathscr{E},\mathscr{F})=2-\frac{2}{k\wedge n}.$$

Hence

$$\delta_k(\mathscr{E},\mathscr{F})\leqq 2-\frac{2}{k\wedge\#\Theta}$$

for any pair $\mathscr{E},\mathscr{F}$ of experiments and "=" is obtained for $\mathscr{E}$ and $\mathscr{F}$ specified as above. It may be shown that $\delta_2(\mathscr{E},\mathscr{F})=1\Leftrightarrow$ there are prior distributions λ and μ such that

$$\sum_\theta\lambda(\theta)P_\theta=\sum_\theta\mu(\theta)P_\theta \quad\text{and}\quad \sum_\theta\lambda(\theta)Q_\theta\wedge\sum_\theta\mu(\theta)Q_\theta=0,$$

and that

$$\delta(\mathscr{E},\mathscr{F})=2-\frac{2}{n}\Leftrightarrow P_\theta$$

does not depend on θ and $Q_\theta\wedge Q_\eta=0$ when $\theta\neq\eta$.

Corollary 4. *Let $\mathscr{E}_i$ be ε_i-deficient relative to $\mathscr{F}_i$ (for k-decision problems); $i=1,\ldots,n$. Then $\prod\mathscr{E}_i$ is $\sum\varepsilon_i$-deficient relative to $\prod\mathscr{F}_i$ (for k-decision problems) (ε_i; $i=1,\ldots,n$ are nonnegative functions on Θ).*

Proof. It suffices to consider the case $n=2$. The proof is based on the fact that if $\psi\in\Psi(\Psi_k)$ and c_θ: $\theta\in\Theta$ are constants, then $x\rightsquigarrow\psi(c_\theta x_\theta\colon\theta\in\Theta)$ belongs to $\Psi(\Psi_k)$. Put $\mathscr{E}_i=((\chi_i,\mathscr{A}_i),(P_{\theta i}\colon\theta\in\Theta))$ and $\mathscr{F}_i=((\mathscr{Y}_i,\mathscr{B}_i),(Q_{\theta i}\colon\theta\in\Theta))$; $i=1,2$ and let $\psi\in\Psi(\Psi_k)$. The notations which we have used so far for two experiments $\mathscr{E}$ and $\mathscr{F}$ will be adapted to this situation by using i as a subscript. Thus:

$$\frac{d(P_{\theta 1}\times P_{\theta 2})}{d[(\sum_\theta P_{\theta 1})\times(\sum_\theta P_{\theta 2})]}=f_{\theta 1}\times f_{\theta 2},$$

$$\frac{d(Q_{\theta 1}\times Q_{\theta 2})}{d[(\sum_\theta Q_{\theta 1})\times(\sum_\theta Q_{\theta 2})]}=g_{\theta 1}\times g_{\theta 2}.$$

By Fubini's theorem:

$$\int \psi(f_{\theta 1} \times f_{\theta 2}; \theta \in \Theta)\, d\big[(\sum_\theta P_{\theta 1}) \times (\sum_\theta P_{\theta 2})\big]$$
$$= \int \sum_\theta P_{\theta 2}(dx_2) \int \psi(f_{\theta 1} \cdot f_{\theta 2}(x_2); \theta \in \Theta)\, d \sum_\theta P_{\theta 1}$$
$$\geqq \int \sum_\theta P_{\theta 2}(dx_2) \big[\int \psi(g_{\theta 1} \cdot f_{\theta 2}(x_2); \theta \in \Theta)\, d \sum_\theta Q_{\theta 1}$$
$$- \sum_\theta \tfrac{1}{2} \varepsilon_{\theta 1} (\psi(-e_\theta) + \psi(e_\theta))\, f_{\theta 2}(x_2)\big]$$
$$= \int \sum_\theta Q_{\theta 1}(dy_1) \int \psi(g_{\theta 1}(y_1)\, f_{\theta 2}; \theta \in \Theta)\, d \sum_\theta P_{\theta 2} - \sum_\theta \tfrac{1}{2} \varepsilon_{\theta 1} (\psi(-e_\theta) + \psi(e_\theta))$$
$$\geqq \int \sum_\theta Q_{\theta 1}(dy_1) \big[\int \psi(g_{\theta 1}(y_1)\, g_{\theta 2}; \theta \in \Theta)\, d \sum_\theta Q_{\theta 2}$$
$$- \sum \tfrac{1}{2} \varepsilon_{\theta 2} (\psi(-e_\theta) + \psi(e_\theta))\, g_\theta(y_1)\big] - \sum_\theta \tfrac{1}{2} \varepsilon_{\theta 1} (\psi(-e_\theta) + \psi(e_\theta))$$
$$= \int \psi(g_{\theta 1} \times g_{\theta 2}; \theta \in \Theta)\, d\big[(\sum_\theta Q_{\theta 1}) \times (\sum_\theta Q_{\theta 2})\big] - \sum_\theta \frac{\varepsilon_{\theta 1} + \varepsilon_{\theta 2}}{2} (\psi(-e_\theta) + \psi(e_\theta)). \quad \square$$

Remark 3. It follows that:

$$\delta_{(k)}\left(\prod_{i=1}^n \mathscr{E}_i, \prod_{i=1}^n \mathscr{F}_i\right) \leqq \sum_{i=1}^n \delta_{(k)}(\mathscr{E}_i, \mathscr{F}_i)$$

and

$$\Delta_{(k)}\left(\prod_{i=1}^n \mathscr{E}_i, \prod_{i=1}^n \mathscr{F}_i\right) \leqq \sum_{i=1}^n \Delta_{(k)}(\mathscr{E}_i, \mathscr{F}_i).$$

In particular if $\mathscr{E}_i \geqq \mathscr{F}_i$; $i = 1, \ldots, n$, then $\prod_{i=1}^n \mathscr{E}_i \geqq \prod_{i=1}^n \mathscr{F}_i$. This was proved by Blackwell in [3]. For equal factors this implies that "$\mathscr{E} \geqq \mathscr{F} \Rightarrow \mathscr{E}^n \geqq \mathscr{F}^n$". Blackwell asked in [3] if the converse was true (for $n = 2$). We shall see in the next section that the answer is no.

It may be shown, however, that $\mathscr{E}^n \sim \mathscr{F}^n \Rightarrow \mathscr{E} \sim \mathscr{F}$. [The Laplace transform of an experiment $\mathscr{E}$ is defined by:

$$L_{\mathscr{E}}(t) = \int \Big(\prod_\theta f_\theta^{t_\theta}\Big)\, d \sum_\theta P_\theta; \qquad t \in K. \tag{5}$$

If $\mathscr{E} \geqq \mathscr{F}$ then $L_{\mathscr{E}} \leqq L_{\mathscr{F}}$ (since $x \rightsquigarrow \prod_\theta x_\theta^{t_\theta}$ is concave for $t \in K$) — the converse, however, is not true. An experiment is — up to equivalence — determined by its Laplace transform and we have

$$L_{\prod_{i=1}^n \mathscr{E}_i} = \prod_{i=1}^n L_{\mathscr{E}_i}. \tag{6}$$

It follows that it would be useful to have a criterion for $\geqq$ in terms of them.]

Corollary 6 below generalize results in [4] to the case of ε-deficiency. The method of proof is essentially that of [4].

Corollary 5. *If $\mathscr{B}$ contains at most 2^k sets then $\mathscr{E}$ is ε-deficient relative to $\mathscr{F}$ if and only if $\mathscr{E}$ is ε-deficient relative to $\mathscr{F}$ for k-decision problems.*

Proof. The "only if" follows as in the proof (by choosing σ as the identity map) of Theorem 1. □

Corollary 6. *$\mathscr{E}$ is ε-deficient relative to $\mathscr{F}$ for k-decision problems if and only if $\mathscr{E}$ is ε-deficient relative to every experiment $((\mathscr{Y},\tilde{\mathscr{B}}),(Q_{\theta\tilde{\mathscr{B}}}\colon\theta\in\Theta))$ where $\tilde{\mathscr{B}}$ is a sub σ-algebra of $\mathscr{B}$ containing at the most 2^k sets.*

Proof. It follows directly from Corollary 5 that the condition is necessary, so suppose the condition of the corollary holds. Let $\psi\in\Psi_k$. Then we may write $\psi=\bigvee_{i=1}^{k}L_i$ where $L_1,\dots,L_k\in\Psi_1$. Let $(B_1,\dots,B_k)$ be a $\mathscr{B}$-measurable partition of $\mathscr{Y}$ such that $\psi\circ g=\sum L_i\circ g\, I_{B_i}$. Let $\tilde{\mathscr{B}}$ be the algebra generated by this partition, put

$$\tilde{\mathscr{F}}=((\mathscr{Y},\tilde{\mathscr{B}}),(Q_{\theta\tilde{\mathscr{B}}}\colon\theta\in\Theta))$$

and define

$$\tilde{g}=\left(\frac{dQ_{\theta\tilde{\mathscr{B}}}}{d\sum_\theta Q_{\theta\tilde{\mathscr{B}}}}\colon\theta\in\Theta\right).$$

Then, since $\tilde{g}$ is $\tilde{\mathscr{B}}$-measurable:

$$\int\psi\circ f\,d\sum_\theta P_\theta\geqq\int\psi\circ\tilde{g}\,d\sum_\theta Q_{\theta\mathscr{B}}-\sum_\theta\varepsilon_\theta\frac{\psi(e_\theta)+\psi(-e_\theta)}{2}.$$

But

$$\begin{aligned}\int\psi\circ\tilde{g}\,d\sum_\theta Q_{\theta\tilde{\mathscr{B}}}&=\sum_i\int_{B_i}\psi\circ\tilde{g}\,d\sum_\theta Q_{\theta\mathscr{B}}\\&=\sum_i\int_{B_i}\psi\left(\left(\frac{Q_\theta(B_i)}{\sum_\theta Q_\theta(B_i)}\colon\theta\in\Theta\right)\right)d\sum_\theta Q_\theta\\&=\sum_i\psi(Q_\theta(B_i)\colon\theta\in\Theta)\geqq\sum_i L_i(Q_\theta(B_i)\colon\theta\in\Theta)\\&=\sum_i L_i\Big(\int_{B_i}g\,d\sum_\theta Q_\theta\Big)=\sum_i\int_{B_i}L_i\circ g\,d\sum_\theta Q_\theta=\int\psi\circ g\,d\sum_\theta Q_\theta.\end{aligned}$$

So that

$$\int\psi\circ f\,d\sum_\theta P_\theta\geqq\int\psi\circ g\,d\sum_\theta Q_\theta-\sum\tfrac{1}{2}\varepsilon_\theta(\psi(e_\theta)+\psi(-e_\theta)).\quad\square$$

Let $\mathscr{F}$ be the experiment given by the Markov matrix:

$\theta \backslash x$	1	2	3	4
1	$\frac{1}{2}$	$\frac{1}{2}$	0	0
2	$\frac{1}{2}$	0	$\frac{1}{2}$	0
3	$\frac{1}{2}$	0	0	$\frac{1}{2}$

Let $\mathscr{F}_{ij}$ be the sub experiment of $\mathscr{F}$ obtained by adding the i-th and the j-th columns. It follows from Corollary 6 that $\prod\{\mathscr{F}_{ij}\colon 1\leqq i<j\leqq 4\}\underset{(3)}{\geqq}\mathscr{F}$. $\prod\{\mathscr{F}_{ij}\colon 1\leqq i<j\leqq 4\}$ is not, however, more informative than $\mathscr{F}$ since:

$$\prod\{L_{\mathscr{F}_{ij}}(\tfrac{1}{3},\tfrac{1}{3},\tfrac{1}{3})\colon 1\leqq i<j\leqq 4\}=\tfrac{3}{4}\cdot\tfrac{8}{9}>\tfrac{3}{4}=L_{\mathscr{F}}(\tfrac{1}{3},\tfrac{1}{3},\tfrac{1}{3}).$$

It has been shown by Stein, see Blackwell [4], that – in general – $\underset{k}{\geqq}$ does not imply $\underset{k+1}{\geqq}$.

Corollary 7. *$\mathscr{E} \underset{(k)}{\geqq} \mathscr{F}$ if and only if*

$$\int \psi \circ f \, d \sum_\theta P_\theta \geqq \int \psi \circ g \, d \sum_\theta Q_\theta$$

for every $\psi \in \Psi_{(k)}$.

Proof. Follows directly from Theorem 2. □

This is the same criterion as that of (3) of Theorem 9 in [4] since any linear function φ agrees on K with a homogenous linear function.

Remark 4. The convex function (or average risk) criterion may be interpreted as convex set relations as follows. For each nonnegative function $\theta \rightsquigarrow \varepsilon_\theta$ on Θ let U: $R^{\Theta \times D_k} \to R$ be the function: $W \rightsquigarrow \sum \varepsilon_\theta \|W_\theta\|$, i.e. the support function of $D = \{W: \sum_d |W_\theta(d)| \leqq \varepsilon_\theta; \theta \in \Theta\}$. For each experiment $\mathscr{E} = ((\chi \mathscr{A}), (P_\theta: \theta \in \Theta))$ let $F_{\mathscr{E}}$: $R^{\Theta \times D_k} \to R$ be the function

$$W \rightsquigarrow \int \left[V \sum_\theta - W_\theta(d) f_\theta\right] d \sum_\theta P_\theta = \bigvee_\rho - \sum_\theta P_\theta \, \rho \, W_\theta = - \# \Theta \text{ [Bayes risk]}$$

i.e. the support function of $-A_{\mathscr{E}}$ where $A_{\mathscr{E}}$ is the set of all operational characteristics. Theorem 2 may now be formulated as $F_{\mathscr{E}} + U \geqq F_{\mathscr{F}}$ or equivalently $A_{\mathscr{E}} - D \supseteqq A_{\mathscr{F}}$. These relations are, however, direct consequences of the definition of ε-deficiency. Another proof of the randomization criterion follows from this observation.

Finally some remarks on convergence for the pseudometrics $\Delta_2, \ldots, \Delta$. If $\mathscr{E} = ((\chi, \mathscr{A}), (P_\theta: \theta \in \Theta))$ is an experiment, the standard experiment [3, p. 94] $\mathscr{E}'$ of $\mathscr{E}$ is the experiment $((K, \text{Borel class}), (P_\theta f^{-1}: \theta \in \Theta))$. Since f is a sufficient statistic in $\mathscr{E}$, $\Delta(\mathscr{E}, \mathscr{E}') = 0$. Moreover $\mathscr{E}'$ is its own standard experiment and an experiment $((K, \text{Borel class}), (Q_\theta: \theta \in \Theta))$ is a standard experiment if and only if $x \rightsquigarrow x_\theta$ is a version of $dQ_\theta / d \sum_\theta Q_\theta$ for each $\theta \in \Theta$. This shows that a standard experiment is uniquely determined by its standard measure $\sum_\theta Q_\theta$. A positive measure S on the Borel subsets of K is a standard measure if and only if $\int x S(dx) = e$. We shall see later (Proposition 18) that if $\mathscr{E}$ and $\mathscr{F}$ are standard experiments and $\Delta_2(\mathscr{E}, \mathscr{F}) = 0$ then $\mathscr{E} = \mathscr{F}$. Assume for a moment that this has been shown. It follows that if $\mathscr{E}$ and $\mathscr{F}$ are any experiments then $\Delta_2(\mathscr{E}, \mathscr{F}) = 0 \Rightarrow \Delta(\mathscr{E}, \mathscr{F}) = 0$. So that the equivalence relations induced by $\Delta_2, \ldots, \Delta$ are all the same.

Let $\mathscr{M}$ denote the set of all standard measures. The pseudometrics $\Delta_2, \ldots, \Delta$ define metrics on $\mathscr{M}$ which – by abuse of notations – again will be denoted by $\Delta_2, \ldots, \Delta$. Another metric on $\mathscr{M}$ may be obtained by using the Levy-distance Λ on the set of standard measures. It has been shown in [13] that Δ and Λ are equivalent. Since $\Delta_k \leqq \Delta$ and $\mathscr{M}$ is compact for Λ this implies that the metrics $\Delta_2, \ldots, \Delta$ and Λ are all equivalent. This may also be concluded from Corollary 3 as follows. Let $L(\mathscr{E}, \mathscr{F})$ denote the Prohorov distance (based upon the norm $x \rightsquigarrow \bigvee_\theta |x_\theta|$) between the normalized standard measures of $\mathscr{E}$ and $\mathscr{F}$. By Theorem 11 in Strassen's paper [18], there is a probability measure R on $K \times K$ with marginals

$[\#\Theta]^{-1}\sum_\theta P_\theta$ and $[\#\Theta]^{-1}\sum_\theta Q_\theta$ so that $R(D)\leqq L(\mathscr{E},\mathscr{F})$ *where*

$$D=\{(x,y):\ \bigvee_\theta |x_\theta-y_\theta|\geqq L(\mathscr{E},\mathscr{F})\}.$$

Let $\gamma\in\Gamma$. Then – using the inequality, $\gamma(x)-\gamma(y)\leqq\bigvee_\theta|x_\theta-y_\theta|$, – we get:

$$[\#\Theta]^{-1}[\int\gamma\, d\sum_\theta P_\theta-\int\gamma\, d\sum_\theta Q_\theta]=\int[\gamma(x)-\gamma(y)]\,R(d(x,y))\leqq\int\bigvee_\theta|x_\theta-y_\theta|\,R(d(x,y))$$

$$=\int_D+\int_{D^c}\leqq R(D)+L(\mathscr{E},\mathscr{F})\,R(D^c)\leqq 2L(\mathscr{E},\mathscr{F}).$$

It follows from Corollary 3 that:

$$\Delta(\mathscr{E},\mathscr{F})\leqq 2\,\#\Theta\, L(\mathscr{E},\mathscr{F}).$$

The equivalence of the metrics Δ and L (or equivalently Δ and Λ) follows now by a standard compactness argument.

We formulate this as:

Proposition 8. *The equivalence relations induced by $\Delta_2, \Delta_3, \dots, \Delta$ are the same. Their restrictions to the set of standard experiments define metrics which all are equivalent to the Levy distance for standard measures.*

Example 9. (The factorization theorem.) Let $\mathscr{E}=((\chi,\mathscr{A}),(P_\theta:\ \theta\in\Theta))$ be an experiment and let $\mathscr{F}=((\chi,\mathscr{B}), P_{\theta\mathscr{B}}:\ \theta\in\Theta))$ be the subexperiment of $\mathscr{E}$ determined by the sub σ-algebra $\mathscr{B}$ of $\mathscr{A}$. Put $\pi=\frac{1}{\#\Theta}\sum_\theta P_\theta$. Then $g_\theta=E^{\mathscr{B}}f_\theta$; $\theta\in\Theta$. Hence $\Delta(\mathscr{E},\mathscr{F})=0\Leftrightarrow\mathscr{L}_\pi(f)=\mathscr{L}_\pi(g)\Leftrightarrow f=g$ a.e. π, using the fact that two real random variables on the same probability space, having the same distribution, are equal a.e., provided one of them is the conditional expectation of the other relative to some σ-algebra. [If the random variables are X and $E^{\mathscr{B}}X$ and $EX^2<\infty$ then

$$E(X-E^{\mathscr{B}}X)^2=EX^2-E(E^{\mathscr{B}}X)^2=0 \quad \text{since } \mathscr{L}(X)=\mathscr{L}(E^{\mathscr{B}}X).$$

More generally, suppose that the random variables are X and $E^{\mathscr{B}}X$ and $E|X|<\infty$. Let φ be a real valued continuous and convex function defined on an interval I so that $X\in I$ a.s. By assumption $E\varphi(E^{\mathscr{B}}X)=E\varphi(X)$. Since $E^{\mathscr{B}}\varphi(X)\geqq\varphi(E^{\mathscr{B}}X)$ (Jensen's inequality) this implies

$$E^{\mathscr{B}}\varphi(X)=\varphi(E^{\mathscr{B}}X) \quad \text{a.s.}$$

so that

$$\mathscr{L}(E^{\mathscr{B}}\varphi(X))=\mathscr{L}(\varphi(E^{\mathscr{B}}X))=\mathscr{L}(\varphi(X)).$$

In particular

$$\mathscr{L}(E^{\mathscr{B}}X^{\pm})=\mathscr{L}(X^{\pm}).$$

It follows that we may – without loss of generality – assume $X\geqq 0$. Since $t\rightsquigarrow-\sqrt{t}$ is convex on $[0,\infty[$ we have

$$\mathscr{L}(E^{\mathscr{B}}\sqrt{X})=\mathscr{L}(\sqrt{X}).$$

Hence – since $E(\sqrt{X})^2<\infty$ – $E^{\mathscr{B}}\sqrt{X}=\sqrt{X}$ a.s. so that

$$E^{\mathscr{B}}X\underset{\text{a.s.}}{=}E^{\mathscr{B}}(E^{\mathscr{B}}\sqrt{X})^2\underset{\text{a.s.}}{=}(E^{\mathscr{B}}\sqrt{X})^2\underset{\text{a.s.}}{=}X.]$$

More generally – let $\mathscr{E}_n = ((\chi_n, \mathscr{A}_n), (P_{\theta n}: \theta \in \Theta))$, $n = 1, 2, \ldots$ be a sequence of experiments, and for each n let $\mathscr{F}_n$ be the subexperiment determined by a sub σ-algebra $\mathscr{B}_n$ of $\mathscr{A}_n$. Put $f_{\theta n} = \frac{dP_{\theta n}}{d \sum_\theta P_{\theta n}}$, $g_{\theta n} = \frac{dP_{\theta n \mathscr{B}_n}}{d \sum_\theta P_{\theta n \mathscr{B}_n}}$; $\theta \in \Theta$, $n = 1, 2, \ldots$ and put $\pi_n = \frac{1}{\# \Theta} \sum_\theta P_{\theta n}$; $n = 1, 2, \ldots$. Then – using the uniform boundedness of $\{f_n\}$ and $\{g_n\}$ and an asymptotic version of the fact mentioned above – we get:

$$\lim_n \Delta(\mathscr{E}_n, \mathscr{F}_n) = 0 \Leftrightarrow \lim_n \Lambda(\mathscr{L}_{\pi_n}(f_n), \mathscr{L}_{\pi_n}(g_n)) = 0 \Leftrightarrow \lim_n \|P_{\theta n} - g_{\theta n} \pi_n\| = 0, \qquad \theta \in \Theta.$$

3. Comparison by Testing Problems[5]

Theorem 2 applied to the case $k = 2$ yields:

Theorem 10. *$\mathscr{E}$ is ε-deficient relative to $\mathscr{F}$ for testing problems if and only if*

$$\left\|\sum_\theta a_\theta P_\theta\right\| \geqq \left\|\sum_\theta a_\theta Q_\theta\right\| - \sum_\theta \varepsilon_\theta |a_\theta| \tag{1}$$

for each vector $a \in R^\Theta$.

Proof. By the identity $a \vee b = \frac{1}{2}(a + b + |a - b|)$ any $\psi \in \Psi_2$ may be written in the form $L_1 + |L_2|$ where $L_1, L_2 \in \Psi_1$. Hence by Theorem 2 it suffices to require that $\int |L \circ f| \, d \sum_\theta P_\theta \geqq \int |L \circ g| \, d \sum_\theta Q_\theta - \sum_\theta \varepsilon_\theta |L(e_\theta)|$ for each $L \in \Psi_1$. By writing $L: x \rightsquigarrow \sum_\theta a_\theta x_\theta$ we obtain the above criterion. □

Theorem 10 has a geometric interpretation as follows. Let $\mathscr{E} = ((\chi, \mathscr{A}), (P_\theta: \theta \in \Theta))$ be an experiment. The set of all critical functions in $\mathscr{E}$ (i.e., the measurable functions from χ to $[0, 1]$) will be denoted by $\mathscr{C}_\mathscr{E}$ and $V_\mathscr{E}$ shall denote the subset of $[0, 1]^\Theta$ consisting of all vectors of the form $((\int \delta \, dP_\theta): \theta \in \Theta)$ where $\delta \in \mathscr{C}_\mathscr{E}$, i.e., $V_\mathscr{E}$ is the set of available power functions. Finally put for $x, y \in R^\Theta$, $I_{[x, y]} = \{z: x \leqq z \leqq y\}$. Then we have:

Corollary 11. *$\mathscr{E}$ is ε-deficient relative to $\mathscr{F}$ for testing problems if and only if*

$$V_\mathscr{E} + \tfrac{1}{2} I_{[-\varepsilon, \varepsilon]} \supseteqq V_\mathscr{F}.$$

Proof. If μ is a finite measure on a measurable space $(\chi, \mathscr{A})$ then $\|\mu\| = [2 \sup \mu(\delta)] - \mu(\chi)$ where sup is taken over all measurable functions from χ to $[0, 1]$. It follows that the support function $H_\mathscr{E}$ of $V_\mathscr{E}$ is given by:

$$H_\mathscr{E}(a) = \sup_{\delta \in \mathscr{C}_\mathscr{E}} \sum_\theta P_\theta(\delta) \, a_\theta = \frac{\left\|\sum_\theta a_\theta P_\theta\right\| + \sum_\theta a_\theta}{2}; \qquad a \in R^\Theta$$

and that the support function $H_\mathscr{F}$ of $V_\mathscr{F}$ is:

$$H_\mathscr{F}(a) = \frac{\left\|\sum_\theta a_\theta Q_\theta\right\| + \sum_\theta a_\theta}{2}; \qquad a \in R^\Theta$$

[5] Throughout this section, we will use the notations of Theorem 2 for experiments $\mathscr{E}$ and $\mathscr{F}$.

while the support function of $\frac{1}{2} I_{[-\varepsilon, +\varepsilon]}$ is:

$$H(a)=\tfrac{1}{2}\sum_\theta |a_\theta|\,\varepsilon_\theta; \qquad \theta\in R^\Theta.$$

The inequality of the corollary now follows since $H_{\mathscr{E}}+H\geqq H_{\mathscr{F}}$. □

In terms of power functions this may be expressed as:

Corollary 12. *$\mathscr{E}$ is ε-deficient relative to $\mathscr{F}$ for testing problems if and only if for each testing problem H: $\theta\in\Theta_0$ against K: $\theta\notin\Theta_0$ and each power function $\Pi_{\mathscr{F}}$ available in $\mathscr{F}$ there is a powerfunction $\Pi_{\mathscr{E}}$ available in $\mathscr{E}$ such that:*

$$\Pi_{\mathscr{E}}(\theta)\leqq\Pi_{\mathscr{F}}(\theta)+\tfrac{1}{2}\varepsilon_\theta;\quad \theta\in\Theta_0, \qquad \Pi_{\mathscr{E}}(\theta)\geqq\Pi_{\mathscr{F}}(\theta)-\tfrac{1}{2}\varepsilon_\theta;\quad \theta\notin\Theta_0.$$

Remark. This corollary shows that for comparison of experiments by testing-problems it suffices to consider loss functions which are indicator *functions.*

Let h be the Hausdorff-distance for compact subsets of R^Θ for norm $x\rightsquigarrow\bigvee_\theta x=\|x\|$ i.e., the distance between two compact subsets C and D of R^Θ is:

$$h(C,D)=\sup_{x\in C}\text{distance}\,(x,D).$$

Then we have:

Corollary 13.

$$\Delta_2(\mathscr{E},\mathscr{F})=2h(V_{\mathscr{E}},V_{\mathscr{F}}).$$

Proof. Let $x\in V_{\mathscr{E}}$. Then by Corollary 12 there is a $y\in V_{\mathscr{F}}$ such that $\|x-y\|\leqq\frac{1}{2}\Delta_2(\mathscr{E},\mathscr{F})$. Hence

$$\text{distance}\,(x,V_{\mathscr{F}})\leqq\tfrac{1}{2}\Delta_2(\mathscr{E},\mathscr{F})$$

so that

$$2h(V_{\mathscr{E}},V_{\mathscr{F}})\leqq\Delta_2(\mathscr{E},\mathscr{F}).$$

By the definition of h there is a $y\in V_{\mathscr{F}}$ such that

$$\|x-y\|=\text{distance}\,(x,V_{\mathscr{F}})\leqq h(V_{\mathscr{E}},V_{\mathscr{F}}).$$

Hence

$$V_{\mathscr{E}}\subseteqq V_{\mathscr{F}}+I_{[-h(V_{\mathscr{E}},V_{\mathscr{F}}),\,h(V_{\mathscr{E}},V_{\mathscr{F}})]}.$$

It follows from Corollary 11 that $\delta_2(\mathscr{E},\mathscr{F})\leqq 2h(V_{\mathscr{E}},V_{\mathscr{F}})$. Similarily $\delta_2(\mathscr{F},\mathscr{E})\leqq 2h(V_{\mathscr{E}},V_{\mathscr{F}})$ such that $\Delta_2(\mathscr{E},\mathscr{F})\leqq 2h(V_{\mathscr{E}},V_{\mathscr{F}})$. □

Remark. Clearly any set $V_{\mathscr{E}}$ belongs to the class $\mathscr{T}$ of subsets of $[0,1]^\Theta$ which are symmetric about $(\frac{1}{2},\ldots,\frac{1}{2})$, compact, convex and containing 0; or equivalently any function $H_{\mathscr{E}}$ belongs to the class $\mathscr{H}$ of functions $\psi\in\Psi$ such that $a\rightsquigarrow 2\psi(a)-\sum_\theta a_\theta$ is symmetric about 0 and

$$\Big(\sum_\theta a_\theta\Big)^+\leqq\psi(a)\leqq\sum_\theta a_\theta^+;\qquad a\in R^\Theta.$$

If $\#\Theta=2$ then $\mathscr{T}$ is precisely the set of $V_{\mathscr{E}}$'s. If $\#\Theta>2$ this is no longer true, since – as we shall see – the set of standardexperiments is not a lattice – for the ordering "$\underset{2}{\geqq}$" – while $\mathscr{T}$ is a *lattice for* "contains".

A simple necessary condition for $\mathscr{E} \underset{2}{\geqq} \mathscr{F}$ is the following:

Corollary 14 (of Corollary 7). *Let $\mathscr{E}$ and $\mathscr{F}$ be experiments with standard-measures $S_{\mathscr{E}}$ and $S_{\mathscr{F}}$ respectively. Suppose $\mathscr{E} \underset{2}{\geqq} \mathscr{F}$. Then the support*[6] *of $S_{\mathscr{F}}$ is contained in the convex hull of the support of $S_{\mathscr{E}}$.*

Proof. Let C be the convex hull of the support of $S_{\mathscr{E}}$. Let $L \in \Psi_1$ and $a \in R$ be such that $C \subseteq [L \leqq a]$. By Corollary 7 – and the remark after:

$$\int (L-a)^+ S_{\mathscr{F}}(dx) \leqq \int (L-a)^+ S_{\mathscr{E}}(dx) = 0.$$

Hence $S_{\mathscr{F}}(L>a)=0$. Since C^c may be covered by a countable class of sets $[L>a]$ such that $C \subseteq [L \leqq a]$ we have $S_{\mathscr{F}}(C^c)=0$. □

It is proved in [4] that when $\# \Theta = 2$, then "being more informative for testing problems" was equivalent to "being more informative". The following theorem generalizes this to ε-deficiency.

Theorem 15. *Let $\mathscr{E}$ and $\mathscr{F}$ be dichotomies. Then $\mathscr{E}$ is ε-deficient relative to $\mathscr{F}$ if and only if $\mathscr{E}$ is ε-deficient relative to $\mathscr{F}$ for testing problems.*

Proof. Write $\Theta = \{1, 2\}$. Since only the "if" needs proof, suppose that $\mathscr{E}$ is ε-deficient relative to $\mathscr{F}$ for testing problems and let $\psi \in \Psi_k$. By definition (of Ψ_k) there are constants $a_1, \ldots, a_k$ and $b_1, \ldots, b_k$ such that $\psi(x_1, x_2) = \bigvee_{i=1}^{k} (a_i x_1 + b_i x_2)$. By rearranging we may assume that there is a s so that $\psi(1, y) = \bigvee_{i=1}^{s} (a_i + b_i y)$ when y is >0, where the representation on the right is minimal in the sense that for each $i \leqq s$ there is a $y>0$ so that $a_i + b_i y > \bigvee \{a_j + b_j y : j \neq i, 1 \leqq j \leqq s\}$. Then the numbers $b_1, b_2, \ldots, b_s$ are all distinct, and we may – without loss of generality – assume that $b_1 < b_2 < \cdots < b_s$. It follows that $a_1 > a_2 > \cdots > a_s$ and that $\psi(x) =$ or $\geqq$ $a_1 x_1 + b_1 x_2 + \sum_{i \geqq 2} (a_i x_1 + b_i x_2 - a_{i-1} x_1 - b_{i-1} x_2)^+$ as $x \in K$ or $x \in \{-e_\theta;\ \theta \in \Theta\}$. The theorem now follows by Remark 2 after Theorem 2. □

Remark. An alternative proof was given in [19]. It was shown there, that to any dichotomy $\mathscr{F} = ((\mathscr{Y}, \mathscr{B}), (Q_i;\ i=1,2))$ one might construct another $\mathscr{F}^* = ((\mathscr{Y}, \mathscr{B}), Q_i^*;\ i=1,2))$ which represents the minimum of all dichtomies $\mathscr{E}$ (regardless of sample space) which are $(\varepsilon_i:\ i=1,2)$ deficient relative to $\mathscr{F}$. $\mathscr{F}^*$ was constructed such that $Q_1^*/\varepsilon_1 + Q_2^*/\varepsilon_2 = Q_1/\varepsilon_1 + Q_2/\varepsilon_2$ (provided $\varepsilon_1, \varepsilon_2 > 0$) and $\|Q_i^* - Q_i\| \leqq \varepsilon_i$, $i=1,2$ (with "=" provided $\varepsilon_1 + \varepsilon_2 \leqq \|Q_1 - Q_2\|$).

Corollary 16. *Put*

$$\beta_{\mathscr{E}}(\alpha) = \sup\{P_2(\delta) : \delta \in \mathscr{C}_{\mathscr{E}}, P_1(\delta) \leqq \alpha\}$$

and

$$\beta_{\mathscr{F}}(\alpha) = \sup\{Q_2(\varphi) : \varphi \in \mathscr{C}_{\mathscr{F}}, Q_1(\varphi) \leqq \alpha\}; \qquad \alpha \geqq 0.$$

Then $\mathscr{E}$ is $(\varepsilon_1, \varepsilon_2)$ deficient relative to $\mathscr{F}$ if and only if $\beta_{\mathscr{E}}(\alpha + \varepsilon_1/2) + \varepsilon_2/2 \geqq \beta_{\mathscr{F}}(\alpha)$; $\alpha \geqq 0$. In particular $\frac{1}{2}\Delta_2(\mathscr{E}, \mathscr{F})$ is the Levy diagonal distance between $\beta_{\mathscr{E}}$ and $\beta_{\mathscr{F}}$ considered as distribution functions.

---

[6] More precisely: the smallest closed support.

Remark. If $\mathscr{E}$ is a dichotomy – then $V_{\mathscr{E}}$ may be any compact convex subset of $[0,1]^{\Theta}$ which is symmetric about $(\frac{1}{2},\frac{1}{2})$ and contains $(0,0)$ – i.e., the restriction of $\beta_{\mathscr{E}}$ to $[0,1]$ may be any concave function β from $[0,1]$ to $[0,1]$ such that $\beta(0+)=\beta(0)$ and $\beta(1)=1$. The functions $\beta_{\mathscr{E}}$ are distribution functions on $[0,1]$ and the probability measures they define will – by abuse of notations – again be denoted by $\beta_{\mathscr{E}}$. Note that the Lebesgue measure λ on $[0,1]$ represents the minimum information experiment. We shall later apply the fact that

$$\mathscr{E}\sim(([0,1], \text{Borel class}), (\lambda, \beta_{\mathscr{E}})).$$

Example 17. For each $(\xi,\eta)\in[0,1]^2$ let $\mathscr{E}_{\xi,\eta}$ be the dichotomy defined by the Markov matrix:

$\theta \backslash x$	0	1	2
1	0	$1-\xi$	ξ
2	η	$1-\eta$	0

The Laplace transform of this experiment is: $(t_1,t_2)\rightsquigarrow(1-\xi)^{t_1}(1-\eta)^{t_2}$ such that:

$$\prod_{i=1}^{n}\mathscr{E}_{\xi_i,\eta_i}\sim\mathscr{E}_{1-\prod_i(1-\xi_i),\,1-\prod_i(1-\eta_i)}\,.$$

In particular:

$$\mathscr{E}^{n}_{\xi,\xi}\sim\mathscr{E}_{1-(1-\xi)^n,\,1-(1-\xi)^n}\,.$$

For each $\sigma>0$ let $\mathscr{F}_\sigma$ denote the dichotomy $((]-\infty,+\infty[, \text{Borel class}), (N(0,\sigma), N(1,\sigma)))$. Then clearly $\mathscr{F}_\sigma^n\sim\mathscr{F}\sigma/\sqrt{n}$.

By the Neyman Pearson fundamental lemma:

$$\beta_{\mathscr{E}_{\xi,\eta}}(\alpha)=\left[\frac{1-\xi}{1-\eta}\alpha+\xi\right]\wedge 1;\qquad \alpha\in[0,1]$$

while

$$\beta_{\mathscr{F}_\sigma}(\alpha)=\Phi\left(\frac{1}{\sigma}+\Phi^{-1}(\alpha)\right);\qquad \alpha\in[0,1]$$

where

$$\Phi(x)=\int_{-\infty}^{x}\frac{1}{\sqrt{2\pi}}e^{-\frac{t^2}{2}}dt;\qquad x\in R.$$

It follows from the symmetry of $\mathscr{E}_{\xi,\xi}$ and $\mathscr{F}_\sigma$ that the graphs of $\beta_{\mathscr{E}_{\xi,\xi}}$ and $\beta_{\mathscr{F}_\sigma}$ are both symmetric about the line $\{(x,y): x+y=1\}$. Hence – since the graph of $\beta_{\mathscr{E}_{\xi,\xi}}$ is essentially linear with direction 1:1 – it suffices to compare the intersections with this line. $\alpha+\beta_{\mathscr{E}_{\xi,\xi}}(\alpha)=1$ and $\alpha+\beta_{\mathscr{F}_\sigma}(\alpha)=1$ have the unique solutions $\alpha=\frac{1-\xi}{2}$ and $\alpha=\Phi\left(-\frac{1}{2\sigma}\right)$ respectively. Hence

$$\mathscr{E}_{\xi,\xi}\geqq\mathscr{F}_\sigma\Leftrightarrow\frac{1-\xi}{2}\leqq\Phi\left(-\frac{1}{2\sigma}\right)$$

while

$$\delta(\mathscr{E}_{\xi,\xi},\mathscr{F}_\sigma)=2\left[\frac{1-\xi}{2}-\Phi\left(-\frac{1}{2\sigma}\right)\right]^{+}$$

so that

$$\delta(\mathscr{E}^n_{\xi,\xi}, \mathscr{F}^n_\sigma) = 2\left[\frac{(1-\xi)^n}{2} - \Phi\left(-\frac{\sqrt{n}}{2\sigma}\right)\right]^+.$$

It follows from the inequality [9, p. 166]

$$\Phi'(x)\left(\frac{1}{x}-\frac{1}{x^3}\right) < 1-\Phi(x) < \Phi'(x)\frac{1}{x}; \qquad x>0$$

that

$$\mathscr{E}^n_{\xi,\xi} \geqq \mathscr{F}^n_\sigma \quad \text{from a certain } n \text{ on} \Leftrightarrow,$$

$$\mathscr{E}^n_{\xi,\xi} \geqq \mathscr{F}^n_\sigma \quad \text{for some } n \Leftrightarrow 1-\xi < e^{-\frac{1}{8\sigma^2}}.$$

[Using the fact that $\mathscr{E}^m \geqq \mathscr{F}^m, \mathscr{E}^n \geqq \mathscr{F}^n \Rightarrow \mathscr{E}^{m+n} \geqq \mathscr{F}^{m+n}$.]

However $1-\xi < e^{-\frac{1}{8\sigma^2}}$ does not imply $\mathscr{E}_{\xi,\xi} \geqq \mathscr{F}_\sigma$ since $\delta(\mathscr{E}_{\xi,\xi}, \mathscr{F}_\sigma)$ for fixed σ may obtain values—under the condition $1-\xi < e^{-\frac{1}{8\sigma^2}}$—as close to the positive number $\left[\frac{1}{2}e^{-\frac{1}{8\sigma^2}} - \Phi\left(-\frac{1}{2\sigma}\right)\right]$ as we wish. This answers a problem raised by Blackwell in [3], and raises other problems such as the problem of finding conditions for $\mathscr{E}^n \geqq \mathscr{F}^n$ from a certain n on. Here are some remarks in this direction.

If $\mathscr{E}$ is a double dichotomy then $\mathscr{E}^n \geqq \mathscr{F}^n$ implies $\mathscr{E} \geqq \mathscr{F}$. This may be proved by first proving it in the case where $\mathscr{F}$ is also a double dichotomy [14, p. 112] and then applying Corollary 6. Let $\mathscr{M}_1$ and $\mathscr{M}_0$ denote the double dichotomies:

$\theta \backslash x$	0	1
1	0	1
2	1	0

and

$\theta \backslash x$	0	1
1	0	1
2	0	1

respectively.

As $n \to \infty$—by the weak law of large numbers—$\mathscr{E}^n \to \mathscr{M}_1$ provided $\mathscr{E} \not\sim \mathscr{M}_0$. Some rough estimates of the speed of convergence may be obtained as follows: Suppose $\mathscr{M}_1 > \mathscr{E} > \mathscr{M}_0$. Then there is a $\xi \in]0,1[$ and a $p \in]0, \frac{1}{2}[$ such that

$$\mathscr{E}_{\xi,\xi} \geqq \mathscr{E} \geqq \mathscr{D}_p \quad \text{where } \mathscr{D}_p \text{ is the double dichotomy}$$

$\theta \backslash x$	0	1
1	$1-p$	p
2	p	$1-p$

Hence

$$\delta(\mathscr{E}^n_{\xi,\xi}, \mathscr{M}_1) \leqq \delta(\mathscr{E}^n, \mathscr{M}_1) \leqq \delta(\mathscr{D}^n_p, \mathscr{M}_1).$$

Now

$\delta(\mathscr{E}^n_{\xi,\xi}, \mathscr{M}_1) = (1-\xi)^n$ and it may be shown from a consideration of the mass in the tail of the binomial distribution that

$$\delta(\mathscr{D}^n_p, \mathscr{M}_1) \leqq 2\left(4p(1-p)\right)^{n/2}.$$

It may, by a direct generalization of these ideas, be shown that—for any finite $\Theta - \delta(\mathscr{E}^n$, maximal informative experiment$) \to 0$ exponentially provided $\theta \rightsquigarrow P_\theta$ is 1-1 and $\min\{\|P_{\theta'} - P_{\theta''}\| : \theta' \neq \theta''\} < 2$. This is in strong contrast to the examples on semigroups of translation experiments treated in [19]. In these examples $\delta(\mathscr{E}^n, \mathscr{E}^{n+a})$, for fixed a, was of the form [constant $+o(1)]/n$ as $n \to \infty$.

Asymptotic comparison of powers of dichotomies by comparison of Bayes risks may be based on the paper [8] by Chernoff. It is shown there that for any dichotomy $\mathscr{E} = ((\chi, \mathscr{A}), (P_1, P_2))$:

$$\lim_n \sqrt[n]{\inf_{0 \leq \alpha \leq 1} (1 - \beta_{\mathscr{E}^n}(\alpha) + \lambda\alpha)} = \inf_t E_{P_1} \left(\frac{dP_2}{dP_1}\right)^t.$$

If $\mathscr{E}$ is ε-deficient relative to $\mathscr{F}$ for testing problems then for each pair $(\theta_1, \theta_2) \in \Theta \times \Theta$, the dichotomy $\mathscr{E}_{\theta_1, \theta_2}$ is $(\varepsilon_{\theta_1}, \varepsilon_{\theta_2})$ deficient relative to $\mathscr{F}_{\theta_1, \theta_2}$. The converse, however, is not true. In fact, as we now shall see, even pairwise equivalence does not imply equivalence.

Let $\mathscr{E} = ((\chi, \mathscr{A}), (P_1, P_2, P_3))$ and $\mathscr{F} = ((\chi, \mathscr{A}), (Q_1, Q_2, Q_3))$ where $\chi = \{1, 2, 3, 4\}$, $\mathscr{A}$ = class of all subsets, $Q_1 = P_1$, $Q_2 = P_2$ and P_1, P_2, P_3 and Q_3 is given by the Markov-matrix.

$\theta \backslash x$	1	2	3	4
P_1	$\frac{1}{2}$	$\frac{1}{2}$	0	0
P_2	$\frac{1}{8}$	$\frac{3}{8}$	$\frac{1}{8}$	$\frac{3}{8}$
P_3	$\frac{1}{6}$	$\frac{2}{6}$	$\frac{2}{6}$	$\frac{1}{6}$
Q_3	$\frac{2}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{2}{6}$

Clearly $\mathscr{E}_{12} = \mathscr{F}_{12}$ and $\mathscr{E}_{13} \sim \mathscr{F}_{13}$ $(\mathscr{E}_{23} \sim \mathscr{F}_{23})$ since $\mathscr{E}_{13}(\mathscr{E}_{23})$ may be obtained from $\mathscr{F}_{13}(\mathscr{F}_{23})$ by a permutation of the columns. Hence $\mathscr{E}$ and $\mathscr{F}$ are pairwise equivalent. However, $\Delta_2(\mathscr{E}, \mathscr{F}) \geq \frac{1}{9}$ since $\|P_1 - P_2 + P_3\| = \frac{17}{12}$ and $\|P_1 - P_2 + Q_3\| = \frac{13}{12}$.

If $\mathscr{E}$ and $\mathscr{F}$ are comparable then—since pairwise sufficiency implies sufficiency—pairwise equivalence implies equivalence. Another condition for equivalence follows from:

Proposition 18. *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 and $\theta \rightsquigarrow \varepsilon_\theta$ a nonnegative function on Θ. Let Λ denote the set of probability distributions on Θ. For each $\lambda \in \Lambda$ put $P_\lambda = \int P_\theta \lambda(d\theta)$, $Q_\lambda = \int Q_\theta \lambda(d\theta)$ and $\varepsilon_\lambda = \int \varepsilon_\theta \lambda(d\theta)$. Let $\hat{\mathscr{E}}$ and $\hat{\mathscr{F}}$ be the convex extensions $((\chi, \mathscr{A}), (P_\lambda : \lambda \in \Lambda))$ and $((\mathscr{Y}, \mathscr{B}), (Q_\lambda : \lambda \in \Lambda))$ of $\mathscr{E}$ and $\mathscr{F}$ respectively.*

Then $\mathscr{E}$ is ε-deficient relative to $\mathscr{F}$ for testing problems if and only if $\hat{\mathscr{E}}_{\lambda_1 \lambda_2}$ is $(\varepsilon_{\lambda_1}, \varepsilon_{\lambda_2})$ deficient relative to $\hat{\mathscr{F}}_{\lambda_1 \lambda_2}$ for each pair $(\lambda_1, \lambda_2) \in \Lambda \times \Lambda$.

Proof. Only the "if" needs proof. Suppose $\mathscr{E}_{\lambda_1, \lambda_2}$ is $(\varepsilon_{\lambda_1}, \varepsilon_{\lambda_2})$ deficient relative to $\mathscr{F}_{\lambda_1, \lambda_2}$ when $\lambda_1, \lambda_2 \in \Lambda$. Let $a \in R^\Theta$. We must show that:

$$\left\|\sum_\theta a_\theta P_\theta\right\| \geq \left\|\sum_\theta a_\theta Q_\theta\right\| - \sum |a_\theta| \varepsilon_\theta. \tag{2}$$

This inequality is trivial if $\sum_\theta a_\theta^+ = 0$ or $\sum_\theta a_\theta^- = 0$. Therefore, suppose $\sum_\theta a_\theta^+ > 0$ and $\sum_\theta a_\theta^- > 0$. Let $\lambda_1, \lambda_2 \in \Lambda$ be given by:

$$\lambda_1(\{\theta\}) = \frac{a_\theta^+}{\sum_\theta a_\theta^+}, \qquad \lambda_2(\{\theta\}) = \frac{a_\theta^-}{\sum_\theta a_\theta^-}; \qquad \theta \in \Theta.$$

By assumption

$$\|b_1 P_{\lambda_1} + b_2 P_{\lambda_2}\| \geqq \|b_1 Q_{\lambda_1} + b_2 Q_{\lambda_2}\| - \varepsilon_{\lambda_1} |b_1| - \varepsilon_{\lambda_2} |b_2|$$

for any numbers b_1 and b_2. (2) now follows by putting $b_1 = \sum_\theta a_\theta^+$ and $b_2 = -\sum_\theta a_\theta^-$.

Proposition 19. (Notations as in Theorem 2.) *The following conditions are equivalent:*

(i) $\Delta_2(\mathscr{E}, \mathscr{F}) = 0$.

(ii) $\Delta_k(\mathscr{E}, \mathscr{F}) = 0$ *for some* k.

(iii) $\|\sum_\theta a_\theta P_\theta\| = \|\sum_\theta a_\theta Q_\theta\|$; $a \in R^\Theta$ *i.e., the map* $P_\theta \rightsquigarrow Q_\theta$ *is well defined and may be extended to an isometry between the linear space generated by the* P_θ*'s and the linear space generated by the* Q_θ*'s.*

(iv) $\hat{\mathscr{E}}$ *is pairwise equivalent to* $\hat{\mathscr{F}}$.

(v) $\mathscr{E}$ *and* $\mathscr{F}$ *have the same standard experiments.*

(vi) $\Delta(\mathscr{E}, \mathscr{F}) = 0$.

Proof. 1. (v) $\Rightarrow$ (vi) $\Rightarrow$ (ii) $\Rightarrow$ (i) $\substack{\nearrow \text{(iii)} \\ \searrow \text{(iv)}}$. (v) $\Rightarrow$ (vi) is a consequence of the sufficiency of the statistic f in $\mathscr{E}$ and the sufficiency of the statistic g in $\mathscr{F}$. (vi) $\Rightarrow$ (ii) $\Rightarrow$ (i) is clear since $\Delta_2 \leqq \Delta_k \leqq \Delta$. (i) $\Leftrightarrow$ (iii) and (i) $\Leftrightarrow$ (iv) follows from Theorem 10 and Proposition 18 respectively.

2. It remains to show that (iv) $\Rightarrow$ (v). Assume first that $\mathscr{E}$ and $\mathscr{F}$ are dichotomies and let us use the notations of Theorem 15 and Corollary 16. Then (iv) implies (vi). Let $F_\mathscr{E} = \mathscr{L}_{P_1}(dP_2/dP_1)$ and $F_\mathscr{F} = \mathscr{L}_{Q_1}(dQ_2/dQ_1)$. Since $\Delta(\mathscr{E}, \mathscr{F}) = 0$, $\beta_\mathscr{E} = \beta_\mathscr{F}$. So that

$$\beta_\mathscr{E}(\alpha) = 1 - \int_\alpha^1 F_\mathscr{E}^{-1}(1-p)\, dp = \beta_\mathscr{F}(\alpha) = 1 - \int_\alpha^1 F_\mathscr{F}^{-1}(1-p)\, dp, \qquad 0 \leqq \alpha \leqq 1.$$

Hence $F_\mathscr{E} = F_\mathscr{F}$. Let us return to the general situation. Let λ_0 be the uniform distribution on Θ and let λ be any element of Λ. By assumption $\hat{\mathscr{E}}_{\lambda_0, \lambda} \sim \hat{\mathscr{F}}_{\lambda_0, \lambda}$. As we have just seen, this implies $\mathscr{L}_{P_{\lambda_0}}(dP_\lambda/dP_{\lambda_0}) = \mathscr{L}_{Q_{\lambda_0}}(dQ_\lambda/dQ_{\lambda_0})$ i.e.,

$$\mathscr{L}_{\frac{1}{n}\sum_\theta P_\theta}\left(\sum_\theta \lambda(\{\theta\}) f_\theta\right) = \mathscr{L}_{\frac{1}{n}\sum_\theta Q_\theta}\left(\sum_\theta \lambda(\{\theta\}) g_\theta\right)$$

where $n = \# \Theta$. Since this holds for any $\lambda \in \Lambda$ we have:

$$\mathscr{L}_{\frac{1}{n}\sum_\theta P_\theta}\left(\sum_\theta a_\theta f_\theta\right) = \mathscr{L}_{\frac{1}{n}\sum_\theta Q_\theta}\left(\sum_\theta a_\theta g_\theta\right)$$

for any $a \geqq 0$. This implies

$$E_{\frac{1}{n}\sum_\theta P_\theta} e^{\sum_\theta a_\theta f_\theta} \equiv E_{\frac{1}{n}\sum_\theta Q_\theta} e^{\sum_\theta a_\theta g_\theta}.$$

By the theory of exponential families [14, p. 50] this implies that $\mathscr{E}$ and $\mathscr{F}$ have the same standard experiments. □

Remark. (iii) $\Leftrightarrow$ (vi) follows from the corollary of Proposition 12 in [12] and (v) $\Leftrightarrow$ (vi) is a special case of the result in [13] referred to on p. 228. It was shown by Morse and Sacksteder [15] that $\mathscr{E} \sim \mathscr{F} \Leftrightarrow \|\bigvee_\theta a_\theta P_\theta\| = \|\bigvee_\theta a_\theta Q_\theta\|$; $a \in K$.

If $\# \Theta = 2$, then the class of sets $V_{\mathscr{E}}$ is a lattice for "contains". It follows that the set of standard experiments is a lattice for $\geqq$ when $\# \Theta = 2$. That this is not true when $\# \Theta \geqq 3$ may be seen from the following proposition which is of interest in itself.

Proposition 20. *Let $\mathscr{E} = ((\chi, \mathscr{A}), (P_\theta\colon \theta \in \Theta))$ and $\mathscr{F} = ((\mathscr{Y}, \mathscr{B}), (Q_\theta\colon \theta \in \Theta))$ be experiments such that χ and $\mathscr{Y}$ are complete separable metric spaces (or Borel sub sets of such spaces) with Borel classes $\mathscr{A}$ and $\mathscr{B}$ respectively, and let $\theta \rightsquigarrow \varepsilon_\theta$ be a non-negative function on Θ. Suppose there exist experiments $\mathscr{E}' = ((\mathscr{Z}, \mathscr{C}), (P'_\theta\colon \theta \in \Theta))$ and $\mathscr{F}' = ((\mathscr{Z}, \mathscr{C}), (Q'_\theta\colon \theta \in \Theta))$ such that $\mathscr{E}' \geqq \mathscr{E}$, $\mathscr{F}' \geqq \mathscr{F}$ and $\|P'_\theta - Q'_\theta\| \leqq \varepsilon_\theta$; $\theta \in \Theta$.*
Then there are experiments $\tilde{\mathscr{E}} = ((\chi \times \mathscr{Y}, \mathscr{A} \times \mathscr{B}), (\tilde{P}_\theta\colon \theta \in \Theta))$ and $\tilde{\mathscr{F}} = ((\chi \times \mathscr{Y}, \mathscr{A} \times \mathscr{B}), (\tilde{Q}_\theta\colon \theta \in \Theta))$ such that $\mathscr{E}' \geqq \tilde{\mathscr{E}} \geqq \mathscr{E}$, $\mathscr{F}' \geqq \tilde{\mathscr{F}} \geqq \mathscr{F}$ and $\|\tilde{P}_\theta - \tilde{Q}_\theta\| \leqq \varepsilon_\theta$, $\tilde{P}_\theta \pi_1^{-1} = P_\theta$, $\tilde{Q}_\theta \pi_2^{-1} = Q_\theta$; $\theta \in \Theta$ where π_1 and π_2 are the projections on χ and $\mathscr{Y}$ respectively.

Remark. Loosely speaking, the theorem says that "minimal combinations" of experiments $\mathscr{E}$ and $\mathscr{F}$ may be constructed with $(\chi \times \mathscr{Y}, \mathscr{A} \times \mathscr{B})$ as sample space.

Proof of the Proposition. By assumption there are Markov kernels M and N such that $P_\theta = M P'_\theta$ and $Q_\theta = N Q'_\theta$. Let $M \times N$ from $\mathscr{Z}$ to $\chi \times \mathscr{Y}$ be defined by:

$$(M \times N)(A \times B | z) = M(A|z)\, N(B|z), \qquad A \in \mathscr{A},\ B \in \mathscr{B},\ z \in \mathscr{Z}.$$

Put $\tilde{P}_\theta = (M \times N) P'_\theta$ and $\tilde{Q}_\theta = (M \times N) Q'_\theta$. Then

$$\begin{aligned}\tilde{P}_\theta \pi_1^{-1}(A) &= \tilde{P}_\theta(A \times \mathscr{Y}) = \int M(A|z)\, N(\mathscr{Y}|z)\, P'_\theta(dz) \\ &= (M P'_\theta)(A) = P_\theta(A); \qquad A \in \mathscr{A}.\end{aligned}$$

Similarly $\tilde{Q}_\theta \pi_2^{-1} = Q_\theta$. Hence $\tilde{\mathscr{E}}$ and $\tilde{\mathscr{F}}$ contains $\mathscr{E}$ and $\mathscr{F}$ as subexperiments such that $\tilde{\mathscr{E}} \geqq \mathscr{E}$ and $\tilde{\mathscr{F}} \geqq \mathscr{F}$. By definition $\tilde{\mathscr{E}} \leqq \mathscr{E}'$ and $\tilde{\mathscr{F}} \leqq \mathscr{F}'$, and $\|\tilde{P}_\theta - \tilde{Q}_\theta\| \leqq \varepsilon_\theta$; $\theta \in \Theta$ since $\|M \times N\| = 1$. □

Let $\mathscr{E}$ and $\mathscr{F}$ be experiments and let $\mathscr{G}$ be an experiment such that:

(i) $\mathscr{G} \underset{(k)}{\geqq} \mathscr{E}, \mathscr{F}$.

(ii) If $\mathscr{H}$ is an experiment such that $\mathscr{H} \underset{(k)}{\geqq} \mathscr{E}, \mathscr{F}$, then $\mathscr{H} \underset{(k)}{\geqq} \mathscr{G}$.

Then, since $\mathscr{G}$ is unique up to equivalence, we may denote it by $\mathscr{E} \underset{(k)}{\vee} \mathscr{F}$.

Corollary 21. *Let $\mathscr{E} = ((\chi, \mathscr{A}), (P_\theta\colon \theta \in \Theta))$ and $\mathscr{F} = ((\mathscr{Y}, \mathscr{B}), (Q_\theta\colon \theta \in \Theta))$ be two experiments such that χ and $\mathscr{Y}$ are complete separable metric spaces (or Borel subsets of such spaces) with Borel classes $\mathscr{A}$ and $\mathscr{B}$ respectively, and suppose $\mathscr{E} \vee \mathscr{F}$ exists. Then there exists a family $(R_\theta\colon \theta \in \Theta)$ of probability measures on $\mathscr{A} \times \mathscr{B}$ such that:*

$$\mathscr{E} \vee \mathscr{F} \sim ((\chi \times \mathscr{Y}, \mathscr{A} \times \mathscr{B}), (R_\theta\colon \theta \in \Theta))$$

and

$$R_\theta \text{ has marginals } P_\theta \text{ and } Q_\theta; \qquad \theta \in \Theta.$$

Proof. The corollary follows by applying Proposition 20 to $\mathscr{E}' = \mathscr{F}' = \mathscr{E} \vee \mathscr{F}$. □

Corollary 22. *Let $\mathscr{E}$ and $\mathscr{F}$ be two experiments such that $\mathscr{E} \underset{(k)}{\vee} \mathscr{F}$ exists. Then $(\mathscr{E} \underset{(k)}{\vee} \mathscr{F})_{\theta_1,\theta_2} \sim \mathscr{E}_{\theta_1,\theta_2} \vee \mathscr{F}_{\theta_1,\theta_2}$ for each pair $(\theta_1,\theta_2) \in \Theta \times \Theta$.*

Proof. Without loss of generality we may assume that $\mathscr{E}$ and $\mathscr{F}$ are standard experiments. Put $\mathscr{G} = \mathscr{E} \underset{(k)}{\vee} \mathscr{F}$. Clearly $\mathscr{G}_{\theta_1,\theta_2} \underset{(k)}{\geqq} \mathscr{E}_{\theta_1,\theta_2}, \mathscr{F}_{\theta_1,\theta_2}$ so that $\mathscr{G}_{\theta_1,\theta_2} \underset{(k)}{\geqq} \mathscr{E}_{\theta_1,\theta_2} \vee \mathscr{F}_{\theta_1,\theta_2}$. By Corollary 21 we have $\mathscr{E}_{\theta_1,\theta_2} \vee \mathscr{F}_{\theta_1,\theta_2} \sim ((\chi \times \mathscr{Y}, \mathscr{A} \times \mathscr{B}), (R_{\theta_1}, R_{\theta_2}))$ where R_{θ_i} has marginals P_{θ_i} and Q_{θ_i}; $i=1,2$. Put $R_\theta = P_\theta \times Q_\theta$ when $\theta \neq \theta_1, \theta_2$ and $\mathscr{H} = ((\chi \times \mathscr{Y}, \mathscr{A} \times \mathscr{B}), (R_\theta : \theta \in \Theta))$. Then $\mathscr{H}$ contains $\mathscr{E}$ and $\mathscr{F}$ as subexperiments such that $\mathscr{H} \geqq \mathscr{E}, \mathscr{F}$. Hence $\mathscr{H} \underset{(k)}{\geqq} \mathscr{G}$ so that

$$\mathscr{E}_{\theta_1,\theta_2} \vee \mathscr{F}_{\theta_1,\theta_2} = \mathscr{H}_{\theta_1,\theta_2} \underset{(k)}{\geqq} \mathscr{G}_{\theta_1,\theta_2}. \quad \square$$

Remark. Similarly it may be proved that if $\mathscr{E} \vee \mathscr{F}$ exists then $(\mathscr{E} \vee \mathscr{F})_{\Theta_0} \sim \mathscr{E}_{\Theta_0} \vee \mathscr{F}_{\Theta_0}$ for each nonempty subset Θ_0 of Θ.

Corollary 23. *If $\mathscr{E}$ and $\mathscr{F}$ are pairwise equivalent then $\mathscr{E} \vee \mathscr{F}$ exists if and only if $\mathscr{E} \sim \mathscr{F}$.*

Proof. Only the "only if" needs a proof, so suppose $\mathscr{E}$ and $\mathscr{F}$ are pairwise equivalent and that $\mathscr{E} \vee \mathscr{F}$ exists. Without loss of generality we may assume that $\mathscr{E}$ and $\mathscr{F}$ are standard experiments. By Corollary 22

$$(\mathscr{E} \vee \mathscr{F})_{\theta_1 \theta_2} = \mathscr{E}_{\theta_1 \theta_2} \vee \mathscr{F}_{\theta_1 \theta_2} = \mathscr{E}_{\theta_1 \theta_2}.$$

Hence, since $\mathscr{E} \vee \mathscr{F} \geqq \mathscr{F}$, $\mathscr{E} \vee \mathscr{F} \sim \mathscr{E}$. Similarly $\mathscr{E} \vee \mathscr{F} \sim \mathscr{F}$. $\square$

Remark. It follows that $\mathscr{E} \vee \mathscr{F}$ does not exists for the experiments $\mathscr{E}$ and $\mathscr{F}$ described just before Proposition 18.

Example 24. Let $\mathscr{E}$ and $\mathscr{F}$ be the double dichotomies given by the matrices

$\theta \backslash x$	0	1
1	$1-p_1$	p_1
2	$1-p_2$	p_2

and

$\theta \backslash y$	0	1
1	$1-q_1$	q_1
2	$1-q_2$	q_2

respectively.

It follows from Corollary 21 that $\mathscr{E} \vee \mathscr{F}$ has a version of the form:

$x \backslash y$	0	1
0	$1-p_1-q_1+\lambda_1$	$q_1-\lambda_1$
1	$p_1-\lambda_1$	λ_1

$x \backslash y$	0	1
0	$1-p_2-q_2+\lambda_2$	$q_2-\lambda_2$
1	$p_2-\lambda_2$	λ_2

and it may be shown that λ_1 and λ_2 are uniquely determined provided at least one of the experiments is different from the minimum information experiment.

It may be seen from this that $\mathscr{E} \underset{(k)}{\vee} \mathscr{F}$ does not exist in general when $\# \Theta \geqq 3$, since the existence of $\mathscr{E} \underset{(k)}{\vee} \mathscr{F}$ imposes conditions on the dichotomies $\mathscr{E}_{\theta_1,\theta_2}$ and $\mathscr{F}_{\theta_1,\theta_2}$; $\theta_1, \theta_2 \in \Theta$ which – in general – are inconsistent.

The minimum (maximum) of a set of dichotomies may be represented by $\inf \beta_{\mathscr{E}}$ ($\sup H_{\mathscr{E}}^*$) where inf (sup) is taken over the set. This follows from the fact that the set of possible β's is closed under pointwise inf's while the set of possible H's is closed under pointwise sup's.

If $\mathcal{G}=((\chi\times\mathcal{Y},\mathcal{A}\times\mathcal{B}),(R_\theta\colon\theta\in\Theta))$ is an experiment such that for each θ, R_θ has marginals P_θ and Q_θ, and $\delta(\mathcal{E},\mathcal{G})\leqq\varepsilon$ and $\delta(\mathcal{F},\mathcal{G})\leqq\varepsilon$, then $\Delta(\mathcal{E},\mathcal{F})\leqq\varepsilon$. Le Cam has raised the problem of whether the converse is true, i.e., if $\Delta(\mathcal{E},\mathcal{F})\leqq\varepsilon$ implies the existence of an $\mathcal{G}$ of the above form such that $\delta(\mathcal{E},\mathcal{G})\leqq\varepsilon$ and $\delta(\mathcal{F},\mathcal{G})\leqq\varepsilon$. By Proposition 20 the problem is equivalent to the problem of finding a $\mathcal{G}\geqq\mathcal{E},\mathcal{F}$ such that $\delta(\mathcal{E},\mathcal{G})$, $\delta(\mathcal{F},\mathcal{G})\leqq\varepsilon$. In the case of dichotomies it is easily checked by Corollary 16 that $\mathcal{G}=\mathcal{E}\vee\mathcal{F}$ has these properties. [If $\mathcal{E}\vee\mathcal{F}$ exists and $\mathcal{G}$ exists, then we may always take $\mathcal{G}=\mathcal{E}\vee\mathcal{F}$.]

If $\mathcal{E}=((\chi,\mathcal{A}),(P_\theta\colon\theta\in\Theta))$ and $\mathcal{F}=((\chi,\mathcal{A}),(Q_\theta\colon\theta\in\Theta))$ where $\|P_\theta-Q_\theta\|\leqq\varepsilon$; $\theta\in\Theta$, then $\mathcal{G}$ may be constructed directly by choosing R_θ such that $\|R_\theta-P_\theta T^{-1}\|=\|R_\theta-Q_\theta T^{-1}\|=\|P_\theta-Q_\theta\|$ where T is the map $x\rightsquigarrow(x,x)$ from χ to $\chi\times\chi$. [Let P and Q be probability measures on $\mathcal{A}$. Define R on $\mathcal{A}$ by $R=(P\wedge Q)T^{-1}+2\|P-Q\|^{-1}[(P-Q)^+\times(P-Q)^-]$ when $P\neq Q$, $R=PT^{-1}$ when $P=Q$. Then

$$R\pi_1^{-1}=P,\quad R\pi_2^{-1}=Q,\quad \|R-PT^{-1}\|=\|R-QT^{-1}\|=\|P-Q\|.$$

This construction shows also that $1-\frac{1}{2}\|P-Q\|$ is the largest number of the form $R^*(\{(x,x)\colon x\in\chi\})$ where R is a probability measure on $\mathcal{A}\times\mathcal{A}$ with marginals P and Q.]

One might ask if $\Delta(\mathcal{E},\mathcal{F})\leqq\varepsilon$ implies the existence of experiments $\mathcal{E}'=((\mathcal{Z},\mathcal{C}),(P'_\theta\colon\theta\in\Theta))\sim\mathcal{E}$ and $\mathcal{F}'=((\mathcal{Z},\mathcal{C}),(Q'_\theta\colon\theta\in\Theta))\sim\mathcal{F}$ such that $\|P'_\theta-Q'_\theta\|\leqq\varepsilon$; $\theta\in\Theta$. By Proposition 20 we may always assume $\mathcal{Z}=\chi\times\mathcal{Y}$, $\mathcal{C}=\mathcal{A}\times\mathcal{B}$ and that the χ marginal of P'_θ is P_θ while the $\mathcal{Y}$ marginal of Q'_θ is Q_θ. Unfortunately $\mathcal{E}'$ and $\mathcal{F}'$ do not exist in general; as the following example will show.

Example 25. Let $\mathcal{E}$ and $\mathcal{F}$ be given by the matrices

$\theta \backslash x$	0	1
1	1	0
2	$\frac{9}{10}$	$\frac{1}{10}$

and

$\theta \backslash y$	0	1
1	$\frac{3}{5}$	$\frac{2}{5}$
2	$\frac{1}{2}$	$\frac{1}{2}$

respectively.

It follows – by simple calculations – from Corollary 16 that $\delta(\mathcal{E},\mathcal{F})=\frac{4}{95}$ while $\delta(\mathcal{F},\mathcal{E})=\frac{4}{45}$. So that $\varepsilon=\Delta(\mathcal{E},\mathcal{F})=\frac{4}{45}$.

Suppose now that there were versions $\mathcal{E}'$ and $\mathcal{F}'$ with the above properties. We may assume that $\mathcal{E}'$ and $\mathcal{F}'$ are given by the matrices

P'_1:

$x \backslash y$	0	1	$\sum$
0	a	$1-a$	1
1	0	0	0

P'_2:

$x \backslash y$	0	1	$\sum$
0	u	$\frac{9}{10}-u$	$\frac{9}{10}$
1	v	$\frac{1}{10}-v$	$\frac{1}{10}$

Q'_1:

$x \backslash y$	0	1
0	$\frac{3}{5}-s$	$\frac{2}{5}-t$
1	s	t
$\sum$	$\frac{3}{5}$	$\frac{2}{5}$

Q'_2:

$x \backslash y$	0	1
0	$\frac{1}{2}-\xi$	$\frac{1}{2}-\eta$
1	ξ	η
$\sum$	$\frac{1}{2}$	$\frac{1}{2}$

By assumption, the χ projection is sufficient in $\mathscr{E}'$ while the $\mathscr{Y}$ projection is sufficient in $\mathscr{F}'$. It follows that $\frac{a}{a+u}=\frac{10}{19}$, $\frac{s}{s+\xi}=\frac{6}{11}$ and $\frac{t}{t+\eta}=\frac{4}{9}$. Hence $u=\frac{9}{10}a$, $s=\frac{6}{5}\xi$ and $t=\frac{4}{5}\eta$. The inequalities $\|P_1'-Q_1'\|\leqq\varepsilon$ and $\|P_2'-Q_2'\|\leqq\varepsilon$ may then be written respectively:

$$|a+\tfrac{6}{5}\xi-\tfrac{3}{5}|+|1-a+\tfrac{4}{5}\eta-\tfrac{2}{5}|+\tfrac{6}{5}\xi+\tfrac{4}{5}\eta\leqq\tfrac{4}{45} \tag{3}$$

and

$$|\tfrac{9}{10}a+\xi-\tfrac{1}{2}|+|\tfrac{9}{10}-\tfrac{9}{10}a+\eta-\tfrac{1}{2}|+|v-\xi|+|\tfrac{1}{10}-v-\eta|\leqq\tfrac{4}{45} \tag{4}$$

(3) implies

$$|a+\tfrac{6}{5}\xi-\tfrac{3}{5}+1-a+\tfrac{4}{5}\eta-\tfrac{2}{5}|+\tfrac{6}{5}\xi+\tfrac{4}{5}\eta\leqq\tfrac{4}{45}$$

i.e.

$$\tfrac{6}{5}\xi+\tfrac{4}{5}\eta\leqq\tfrac{2}{45}. \tag{5}$$

Hence $\frac{4}{5}\xi+\frac{4}{5}\eta\leqq\frac{2}{45}$. So that

$$\xi+\eta\leqq\tfrac{1}{18}. \tag{6}$$

In the same way we get from (4) $|\frac{1}{10}-(\xi+\eta)|\leqq\frac{2}{45}$, which by (6) yields $\xi+\eta=\frac{1}{18}$, which together with (5) gives $\xi=0, \eta=\frac{1}{18}$. Hence $|v-\xi|+|\frac{1}{10}-v-\eta|\geqq\frac{2}{45}$. So that (4) implies

$$|\tfrac{9}{10}a-\tfrac{1}{2}|+|\tfrac{9}{10}-\tfrac{9}{10}a-\tfrac{4}{9}|\leqq\tfrac{2}{45} \tag{7}$$

while (3) now may be written:

$$|a-\tfrac{3}{5}|+|1-a-\tfrac{16}{45}|\leqq\tfrac{2}{45}. \tag{8}$$

Now we have arrived at a contradiction since (8) implies $a\geqq\frac{3}{5}$ while (7) implies $a<\frac{3}{5}$.

4. Mixtures of Experiments

It has been shown by Birnbaum in [1] that any dichotomy is equivalent to a mixture of double dichotomies. This will be generalized to an arbitrary finite Θ, and an alternative proof of the decomposition in [1] – based upon the existence of a certain ancillary statistic – will be given. Most of the results in this section are "experiment-interpretations" of well known results from the theory of convex sets.

Definition. Let $(T, \mathscr{S}, \pi)$ be a probability space and $\mathscr{E}_t$; $t\in T$ a family of experiments with standard measures S_t; $t\in T$. If $t \rightsquigarrow S_t(E)$ is measurable for each Borel set E then any experiment with standard measure $\int S_t\pi(dt)$ will be denoted by $\int \mathscr{E}_t\pi(dt)$, and called a π-mixture of $\mathscr{E}_t$: $t\in T$.

The motivation for this definition is:

Proposition 26. *Let $\mathscr{E}_t=((\chi, \mathscr{A}), (P_{\theta t}: \theta\in\Theta))$; $t\in T$ be a family of experiments and $(T, \mathscr{S}, \pi)$ a probability space such that $t\rightsquigarrow P_{\theta t}(A)$ is measurable; $A\in\mathscr{A}$, $\theta\in\Theta$. Suppose $\mathscr{A}$ is separable. Then:*

$$\int \mathscr{E}_t\pi(dt)=((T\times\chi, \mathscr{S}\times\mathscr{A}), (Q_\theta: \theta\in\Theta)) \tag{1}$$

where

$$Q_\theta(S\times A)=\int_S P_{\theta t}(A)\,\pi(dt);\qquad S\in\mathcal{S},\ A\in\mathcal{A};\qquad \theta\in\Theta.$$

Proof. Let $\mathcal{E}$ be the right hand side of (1) and let S and S_t; $t\in T$ be the standard measures of $\mathcal{E}$ and $\mathcal{E}_t$; $t\in T$ respectively. We must show that $S=\int S_t\,\pi(dt)$. By the martingale convergence theorem and separability $dP_{\theta t}/d\sum_\theta P_{\theta t}$ may be specified so that it is jointly measurable relative to $\mathcal{S}\times\mathcal{A}$.

It follows that for each Borel set E and each θ, $t\rightsquigarrow P_{\theta t}((X_{\theta t}\colon\theta\in\Theta)\in E)$ is measurable and

$$\int S_t(E)\,\pi(dt)=\int\big[\sum_\theta P_{\theta t}((X_{\theta t}\colon\theta\in\Theta)\in E)\big]\,\pi(dt)$$

$$=\sum_\theta Q_\theta\big(\{(t,x)\colon(X_{\theta t}(x)\colon\theta\in\Theta)\in E\}\big).$$

The theorem now follows, since $(t,x)\rightsquigarrow X_{\theta t}(x)$ is a version of $dQ_\theta/d\sum_\theta Q_\theta$. □

In the situation described in the above theorem the statistic $(t,x)\rightsquigarrow t$ alone provides no information on θ since its distribution does not depend on θ. This does not mean, however, that $(t,x)\rightsquigarrow x$ contains all information, i.e., that $(t,x)\rightsquigarrow x$ is sufficient. ($(t,x)\rightsquigarrow x$ is sufficient if—for example—the $\mathcal{E}_t$'s are standard experiments.)

Example 27. For each $t\in T$ let β_t be a concave function from $[0,1]$ to $[0,1]$ such that $\beta_t(0^+)=\beta_t(0)$, $\beta_t(1)=1$. For each t, consider the experiment $\mathcal{E}_t=(([0,1],\text{Borel class}),(\lambda,\beta_t))$ where λ is the Lebesgue measure on $[0,1]$ while β_t—by abuse of notation—is the probability measure with distribution function β_t. Suppose $t\rightsquigarrow\beta_t(x)$ is measurable for each $x\in[0,1]$. Then we can define the experiment

$$\mathcal{E}=((T\times[0,1];\mathcal{S}\times\text{Borel sets}),(Q_\theta\colon\theta\in\Theta))\text{ by:}$$

$$Q_1(S\times A)=\int_S\lambda(A)\,\pi(dt)=(\pi\times\lambda)(A\times S),$$

$$Q_2(A\times S)=\int_S\beta_t(A)\,\pi(dt);\qquad S\in\mathcal{S},\ A\text{ is Borel}.$$

Let $\tilde{\mathcal{E}}$ be the subexperiment of $\mathcal{E}$, obtained by restriction to the statistic $(t,x)\rightsquigarrow x$. Then

$$\beta_{\tilde{\mathcal{E}}}(\alpha)=\int\beta_t(\alpha)\,\pi(dt)$$

while

$$\beta_{\mathcal{E}}(\alpha)=\sup\int\beta_t(\alpha_t)\,\pi(dt)$$

where sup is taken over all measurable functions $t\rightsquigarrow\alpha_t$ from T to $[0,1]$ such that $\int\alpha_t\,\pi(dt)=\alpha$.

If $\mathcal{E}_t$ is ε_t-deficient relative to $\mathcal{F}_t$ (for k-decision problems) for each t and $\int\mathcal{E}_t\,\pi(dt)$, $\int\varepsilon_t\,\pi(dt)$ and $\int\mathcal{F}_t\,\pi(dt)$ exists, then it follows from Theorem 2 that $\int\mathcal{E}_t\,\pi(dt)$ is $\int\varepsilon_t\,\pi(dt)$-deficient relative to $\int\mathcal{F}_t\,\pi(dt)$ (for k-decision problems). In particular $\mathcal{E}_t\geqq\mathcal{F}_t$; $t\in T$ implies $\int\mathcal{E}_t\,\pi(dt)\geqq\int\mathcal{F}_t\,\pi(dt)$. The following proposition—which is a direct application of a theorem of Strassen [18, p. 423]—goes in the converse direction.

Proposition 28. *Let $\mathscr{E} \geqq \mathscr{F}$ be two experiments, $(T, \mathscr{S}\, \pi)$ a measurable space and $\mathscr{F}_t$; $t \in T$ a family of experiments such that $\mathscr{F} = \int \mathscr{F}_t\, \pi(dt)$. Then there exists experiments $\mathscr{E}_t$; $t \in T$ such that*

(i) $\mathscr{E} = \int \mathscr{E}_t\, \pi(dt)$.

(ii) $\mathscr{E}_t \geqq \mathscr{F}_t$.

Proof. Consider the Banach-space $C(K)$ of continuous functions on K (with sup-norm). Let Φ denote the set of continuous concave functions on K. Let U, S and S_t; $t \in T$ be the standard measures of $\mathscr{E}$, $\mathscr{F}$ and $\mathscr{F}_t$; $t \in T$ respectively. For each $f \in C(K)$ and each Borel measure λ, define the functional p_λ on $C(K)$ by[7]:

$$p_\lambda(f) = \inf\{\lambda(\varphi) \colon \varphi \in \Phi; \varphi \geqq f\}.$$

Then p_λ is sublinear and

$$p_\lambda(f) = \lambda(\inf\{\varphi \colon \varphi \geqq f\}).$$

By assumption $p_s = \int p_{s_t}\, \pi(dt)$ and $U \leqq p_s$. It follows from Theorem 1 in [18] that there exist linear functionals μ_t; $t \in T$ on $C(K)$ such that – for each $f \in C(K)$ – $t \rightsquigarrow \mu_t(f)$ is measurable,

$$\mu_t(f) \leqq p_{s_t}(f); \qquad t \in T$$

and

$$U(f) = \int \mu_t(f)\, \pi(dt).$$

It remains to show that the μ_t's are standard measures. Let $t \in T$. If $f \in C(K)$, $f \leqq 0$, then

$$\mu_t(f) \leqq p_{s_t}(f) \leqq p_{s_t}(0) = 0.$$

It follows that $\mu_t \geqq 0$ such that μ_t is a positive Borel measure. Let L be linear on K. Then since $L, -L \in \Phi$, $\mu_t(L) = S_t(L)$. In particular $\int x_\theta\, \mu_t(dx) = 1, \theta \in \Theta$. Hence μ_t is a standard measure.

The set of standard measures is a convex subset of the linear space of bounded measures on the Borel subsets of K. The next proposition will give us the extreme points of this set:

Proposition 29. *Let $t \in R^\Theta$. Consider the set $\mathscr{M}_t$ of probability measures S on the Borel class in R^Θ such that $\int x_\theta\, S(dx) = t_\theta$, $\theta \in \Theta$. This set is convex and S is extreme if and only if it is supported by the vertices of a simplex. If S is extreme, then it is uniquely determined by its support*[8].

Proof. 1. Suppose S is concentrated on the vertices of a simplex $\langle a^1, \ldots, a^k \rangle$. Put $S(\{a^i\}) = \alpha_i$; $i = 1, \ldots, k$ such that $\sum_i \alpha_i = 1$ and $t = \int x\, S(dx) = \sum \alpha_i\, a^i$. Suppose $S = (1-\theta)\, Q + \theta\, R$ where $Q, R \in \mathscr{M}_t$ and $0 < \theta < 1$. Then $S \gg Q$ such that Q is supported by $\{a^1, \ldots, a^k\}$. Put $Q(\{a^i\}) = \alpha_i'$; $i = 1, \ldots, k$. Then

$$t = \int x\, Q(dx) = \sum \alpha_i'\, a^i.$$

[7] When convenient, $\int f(x)\, \mu(dx)$ will be written $\mu(f)$.

[8] More precisely: smallest closed support.

Hence — since $0=t-t=\sum_i(\alpha_i-\alpha_i')\,a_i$ and $\sum_i(\alpha_i-\alpha_i')=0$

$$\alpha_i=\alpha_i', \quad i=1,\dots,k \quad \text{i.e.} \quad Q=S.$$

Similarly $R=S$. It follows that S is extreme.

2. Suppose S is extreme. Let $p^1,\dots,p^k$ be points of increase[9] for S. We may then construct a measurable partition $\{V_1,\dots,V_k\}$ of R^Θ such that V_i is a neighborhood of p^i; $i=1,\dots,k$. Put $\lambda_i=S(V_i)$. Then $\lambda_1,\dots,\lambda_k>0$ and $\sum\lambda_i=1$. Let $f_i=\lambda_i^{-1}I_{V_i}$, such that $\int f_i\,dS=1$. $S\in\mathcal{M}_t$ implies

$$t=\int x\,S(dx)=\sum_i\lambda_i\int x f_i(x)\,S(dx)=\sum_i\lambda_i v^i$$

where

$$v^i=\int x f_i(x)\,S(dx).$$

Let $\mu_1,\dots,\mu_k\geqq 0$ be constants such that $\sum\mu_i=1$ and $\sum_i\mu_i v^i=t$. Then

$$\int x\left(\sum\mu_i f_i(x)\right)S(dx)=\sum\mu_i v^i=t$$

and

$$\int\left(\sum\mu_i f_i(x)\right)S(dx)=1.$$

It follows that $Q=(\sum\mu_i f_i)\,S\in\mathcal{M}_t$. Let $a>1$ be a constant such that $\sum\mu_i f_i\leqq a$. Put $\theta=\frac{a-1}{a}$ and $R=\frac{1}{\theta}[S-(1-\theta)Q]$. Then $S=(1-\theta)Q+\theta R$, $0<\theta<1$, $R,Q\in\mathcal{M}_t$. Since S is extreme, $R=Q$, such that $S=Q$. Hence $\sum\mu_i f_i=1$ a.e. S. Since $S(V_i)>0$ there is a $x\in V_i$ such that:

$$1=\sum\mu_i f_i(x)=\mu_i\lambda_i^{-1}.$$

Hence

$$\mu_i=\lambda_i; \quad i=1,\dots,k.$$

We have shown that

$$\mu_1,\dots,\mu_k\geqq 0, \quad \sum\mu_i=1, \quad \sum\mu_i v^i=t\Leftrightarrow\mu_i=a_i, \quad i=1,\dots,k.$$

This implies that $v^1,\dots,v^k$ are gometrically independent. Consequently $k\leqq\#\Theta+1$. It follows that S is supported by a k' point set $\{q^1,\dots,q^{k'}\}$ where $k'\leqq\#\Theta+1$. By the same reasoning, $q^1,\dots,q^{k'}$ are geometrically independent. Let $\lambda_1',\dots,\lambda_{k'}'$ be the weight assigned to $q^1,\dots,q^{k'}$ by S. Then we have

$$e=\int x\,S(dx)=\sum_i\lambda_i' q^i.$$

It follows that $\lambda_1',\dots,\lambda_{k'}'$ — and consequently S — is determined by $q^1,\dots,q^{k'}$. □

Remark. A permutation π of Θ induces a permutation of R^Θ: $(x_\theta:\theta\in\Theta)\rightsquigarrow(x_{\theta\pi};\theta\in\Theta)$ which again will be denoted by π. Let G be a group of permutations π of Θ and consider the set $\mathcal{M}_{t,G}$ of probability measures in $\mathcal{M}_t$ which are invariant under G. By a modification of the proof of Proposition 29 it may be shown that

[9] A point $p\in R^\Theta$ is called a point of increase for a positive Borel measure S if $S(V)>0$ for any measurable neighborhood of p.

S is extreme in $\mathcal{M}_{t,G}$ if and only if S has a smallest support F of the form $F = \sum_{i=1}^{k} F_i$ where $F_1, \ldots, F_k$ are k distinct orbits for G (in R^{Θ}) such that the vectors $\#(F_i)^{-1} \sum_{F_i} x$; $i = 1, \ldots, k$ are geometrically independent. If G is transitive on Θ, then by geometrical independence, $k \leqq 2$. If G is transitive and $\#(\Theta) S$ is a standard measure then $k = 1$.

Corollary 30. *Any standard measure whose support is the set of vertices of a simplex is extreme. To any simplex which contains* $\#(\Theta)^{-1}(1, \ldots, 1)$ *and is contained in K there corresponds an extreme standard measure which is supported by the set of vertices of the simplex. This correspondence is one to one and onto between the set of simplexes which contains* $\#(\Theta)^{-1}(1, \ldots, 1)$ *and is contained in K on the one hand, and the set of extreme standard measures on the other.*

Corollary 31. *Any experiment* $\mathcal{E}$ *is a mixture of experiments whose standard experiments are extreme.*

Proof. Let S be the standard experiment of $\mathcal{E}$ and consider the set of standard measures as a subset of the set of bounded measures on the Borel class of K, topologized by the Levy distance. Let $\mathscr{V}$ denote the set of extreme standard measures. Then by a theorem of Choquet [17, p. 19] there is a probability measure π on the Borel class in $\mathscr{V}$ such that $S = \int V \pi(dV)$. □

Remark 1. If B is a support of S then

$$\int V(B)\, \pi(dV) = 1$$

such that

$$\pi(\{V: V(B) = 1\}) = 1.$$

Remark 2. Let G be a group of permutations of Θ. In addition to the terminology used in the remark after Proposition 29 we introduce the following notations: If $\mathcal{E} = ((\chi, \mathcal{A}); (P_\theta; \theta \in \Theta))$ and π is a permutation of Θ then

$$\pi \mathcal{E} \overset{\text{def}}{=} ((\chi, \mathcal{A}), (P_{\theta\pi}; \theta \in \Theta)).$$

Then

$$\pi_1(\pi_2 \mathcal{E}) = (\pi_1 \pi_2) \mathcal{E}, \qquad \delta_{(k)}(\mathcal{E}, \mathcal{F}) = \delta_{(k)}(\pi \mathcal{E}, \pi \mathcal{F}).$$

If S is the standard measure of $\mathcal{E}$ then $S\pi^{-1}$ is the standard measure of $\pi\mathcal{E}$. For each experiment $\mathcal{E}$ put $G(\mathcal{E}) = \{\pi; \pi\mathcal{E} \sim \mathcal{E}\}$. Then $G(\mathcal{E})$ is a group and

$$\mathcal{E} \sim \mathcal{F} \rightarrow G(\mathcal{E}) = G(\mathcal{F}).$$

Consider the set of standard measures S of experiments $\mathcal{E}$ such that $G(\mathcal{E}) \supseteqq G$. The extreme elements of this set are — up to a factor $\#\Theta$ — described in the remark after Proposition 29. Since $\#(G)^{-1} \sum_{\pi} S\pi^{-1}$ is invariant under G for any standard measure S, it follows that any experiment is a "component" of an "invariant" experiment. Again by the theorem of Choquet, any invariant experiment is a mixture of experiments whose standard measures are extreme points for the set of standard measures of "invariant" experiments.

In particular if $\Theta=\{1,\dots,n\}$ and G is the cyclic group generated by the permutation $1\rightsquigarrow 2\rightsquigarrow 3\rightsquigarrow\cdots\rightsquigarrow n\rightsquigarrow 1$, then (by restriction to "discrete" experiments) we get Lemmas 1 and 2 of [2].

Corollary 32. *Let $\mathscr{E}=((\chi,\mathscr{A}),(P_\theta\colon\theta\in\Theta))$ be an experiment. Then:*

(i) *The standard measure of $\mathscr{E}$ is extreme provided $\mathscr{E}$ is boundedly complete.*

(ii) *$\mathscr{E}$ is complete provided $\mathscr{E}$ is extreme and $\mathscr{A}$ is minimal sufficient.*

Remark. It follows that 1) $\mathscr{E}$ is extreme and $\mathscr{A}$ is minimal sufficient if and only if it is boundedly complete. 2) $\mathscr{E}$ is complete if and only if it is boundedly complete. (This does not hold – in general – when $\#\Theta=\infty$ – see [14, p. 152].) 3) $\mathscr{A}$ has a complete and sufficient sub σ-algebra if and only if $\mathscr{E}$ is extreme.

Proof of the Corollary. Let S be the standard measure of $\mathscr{E}$.

1. Suppose $\mathscr{E}$ is boundedly complete. Let U and V be standard measures such that $S=\frac{1}{2}U+\frac{1}{2}V$. Put $u=dU/dS$. Then

$$e=\int x\,U(dx)=\int x\,u(x)\,S(dx)$$

so that

$$\int u\circ f\,dP_\theta\equiv 1.$$

By assumption $u\circ f=1$ a.e. $\sum_\theta P_\theta$ such that $u=1$ a.e. S.

Hence $U=V=S$. It follows that S is extreme.

2. Suppose $\mathscr{E}$ is extreme and $\mathscr{A}$ is minimal sufficient. Let h be $\sum_\theta P_\theta$ integrable and such that $\int h\,dP_\theta\equiv 0$. We may – since $\mathscr{A}$ is minimal sufficient – assume h is of the form $g\circ f$. Then $\int g(x)\,S(dx)=0$. By geometrical independence of the vertices in the simplex corresponding to S, $g=0$ a.e. S. Hence; $h=0$ a.e. $\sum_\theta P_\theta$.

Proposition 33. *Let $\mathscr{E}$ be complete. Then the following conditions are equivalent:*

(i) $\mathscr{E}\geqq\mathscr{F}$.

(ii) $\mathscr{E}\underset{2}{\geqq}\mathscr{F}$.

(iii) *The support of $\mathscr{F}$'s standard measure is contained in the simplex which corresponds to $\mathscr{E}$.*

Proof. 1. (i) ⇒ (ii) ⇒ (iii) follows from Corollary 14.

2. (iii) ⇒ (i).

Let $S_{\mathscr{E}}$ and $S_{\mathscr{F}}$ be the standard measures of $\mathscr{E}$ and $\mathscr{F}$ respectively, and let $C=\langle v^1,\dots,v^k\rangle$ be the simplex corresponding to $\mathscr{E}$. Each $y\in C$ may be written in the form

$$y=\sum_{j=1}^{k} m(y|j)\,v^j \tag{2}$$

where $m(y|1),\dots,m(y|k)$ are uniquely determined by y and hence affine in y. For each Borel set B put

$$\hat{m}(B|j)=\frac{1}{\lambda_j}\int_B m(y|j)\,S_{\mathscr{F}}(dy);\qquad j=1,\dots,k$$

where λ_j is the weight assigned to $\{v_j\}$ by $S_{\mathscr{E}}$; $j=1,\ldots,k$. It follows from (2) that

$$e=\int_C y\, S_{\mathscr{F}}(dy)=\sum_{j=1}^{k}\left[\int \frac{1}{\lambda_j}\, m(y|j)\, S_{\mathscr{F}}(dy)\right]\lambda_j v_j.$$

Hence, since $v^1,\ldots,v^k$ are geometrically independent

$$\int \frac{1}{\lambda_j}\, m(y|j)\, S_{\mathscr{F}}(dy)=1; \qquad j=1,\ldots,k$$

thus $\hat{m}$ defines a randomization from $\{v^1,\ldots,v^k\}$ to C. Let $(P_\theta\colon \theta\in\Theta)$ and $(Q_\theta\colon \theta\in\Theta)$ define the standard experiments of $\mathscr{E}$ and $\mathscr{F}$ respectively. Then

$$\begin{aligned}(\hat{m}P_\theta)(B)&=\sum_j \hat{m}(B|j)\, P_\theta(\{j\})=\sum_j \int_B m(y|j)\, v_\theta^j\, S_{\mathscr{F}}(dy)\\ &=\int_B y_\theta\, S_{\mathscr{F}}(dy)=Q_\theta(B); \qquad \theta\in\Theta. \quad \square\end{aligned}$$

Let m be a randomization from Θ to some finite set $\{1,\ldots,N\}$. Then m defines an experiment $\mathscr{E}_m=(\{1,\ldots,N\}$, (class of all subsets), $(P_\theta\colon \theta\in\Theta))$ with standard measure S_m where $P_\theta(\{i\})=m(\{i\}|\theta)$; $i=1,\ldots,N$; $\theta\in\Theta$.

The set $\mathscr{M}$ of all randomizations from Θ to Θ may be identified with a compact subset of $R^{\#\Theta^2}$. The open sub set $\mathscr{M}_0$ of $\mathscr{M}$ consisting of the non singular randomizations, exhausts the set of extreme experiments, and the correspondence is – since a permutation of columns does not change the equivalence class – $\#\Theta!$ to 1. It follows easily that any experiment may be represented (though not uniquely) as a random selection of points in $\mathscr{M}_0$. We summarize this as:

Theorem 34. *To any experiment $\mathscr{E}$ there is a probability measure σ on the Borel class on $\mathscr{M}_0$ such that:*

$$\mathscr{E}\sim((\mathscr{M}_0\times\Theta, \text{Borel class}\times\mathscr{F}), (Q_\theta\colon \theta\in\Theta))$$

where

$$Q_\theta(S\times F)=\int_S p_{\theta m}(F)\,\sigma(dm); \qquad S\in\text{Borel class},\; F\in\mathscr{F}.$$

With the notations of the last theorem, let W be the projection on T from $(T\times\Theta)$. Then W has the following properties relative to $((\mathscr{M}_0\times\Theta, \text{Borel class}\times\mathscr{F}), (Q_\theta\colon \theta\in\Theta))$:

(i) It is ancillary, that is, its distribution does not depend on θ.

(ii) It is $\#\Theta$ to one.

If conversely an experiment admits a statistic W such that (i) and (ii) hold, then the experiment may under regularity conditions – be decomposed into experiments whose sample spaces are $(\Theta,\mathscr{F})$. It has been shown by Birnbaum [1] that if $\#\Theta=2$, then $\mathscr{E}$ may always be decomposed into double dichotomies such that the set of double dichotomies which appears in the decomposition is totally ordered. It will be shown below that this may be deduced from properties of a "natural" two to one valued ancillary statistic. We use the notations of Corollary 16 and the following remark. The ancillarity is established by:

Proposition 35. *Let P and Q be two probability measures on an interval $[a, b]$ of the extended real line such that $P(\{x\}) = Q(\{x\})$; $x \in [a, b]$. Then the random variable $x \rightsquigarrow P[a, x] - Q[a, x]$ has the same law under P as under Q.*

Proposition 36. *Let $\mathscr{E}$ be a dichotomy such that $\beta_{\mathscr{E}}(0) = 0$ and let us consider the representation $\mathscr{F} = (([0, 1], \mathscr{B}), (\Lambda, \beta_{\mathscr{E}}))$ and the statistic T: $x \rightsquigarrow \beta_{\mathscr{E}}(x) - x$ in $\mathscr{F}$. Then T is ancillary and the conditional distribution given T decomposes $\mathscr{E}$ into double dichotomies $\{\mathscr{D}_t\}$ such that $\mathscr{D}_{t_1} \leqq \mathscr{D}_{t_2}$ when $t_1 > t_2$.*

Remark 1. The set up may be generalized to cover the case $\beta_{\mathscr{E}}(0) > 0$ as well.

Remark 2. It follows from Proposition 36 and the remark above that any distribution function concentrated on $[0, 1]$ and having expectation $\frac{1}{2}$ (i.e., essentially the $\frac{1}{2}$ standard measure of a dichotomy) is a "totally ordered[10]" mixture of two-point distributions on $[0, 1]$ with expectations $\frac{1}{2}$. Blackwell and Dubins [6] have shown – by another approach – that the analogous result holds for distributions on the line. The decomposition is essentially unique, and is related to Frechet's maximal distribution [10, p. 162] as we now shall see.

Let P be a probability distribution on the real line such that $\int x\, P(dx) = 0$, and for each pair a, b, such that $(a, b) = (0, 0)$ or $a < 0 < b$, let $\delta_{a,b}$ be the two-point distribution with support $\{a, b\}$ and expectation 0. Put $C = \int x^+ P(dx)$ and let F_1 and F_2 be the probability distribution functions on the line corresponding to the probability measures with densities $x \rightsquigarrow C^{-1} x^-$ and $x \rightsquigarrow C^{-1} x^+$ respectively with respect to P.

Then:

$$P = P(\{0\})\, \delta_{0,0} + C \int_0^1 \frac{F_1^{-1}(1-p) - F_2^{-1}(p)}{F_1^{-1}(1-p)\, F_2^{-1}(p)} \left[\delta_{F_1^{-1}(1-p),\, F_2^{-1}(p)}\right] dp. \tag{3}$$

[The distribution in the plane induced from the rectangular distribution on $[0, 1]$ by the map $p \rightsquigarrow (-F_1^{-1}(1-p), F_2^{-1}(p))$ is the Frechet maximal distribution with marginals given by $x \rightsquigarrow 1 - F_1(-x-)$ and F_2.] The decomposition of dichotomies follows directly from (3).

All "totally ordered" decompositions of P are "equivalent" to (3). To see this, note first that the Frechet maximal distributions are precisely those probability distributions which possess supports which are totally ordered for the product order on $R \times R$. [An indication of a proof of this: Let π be a probability distribution in the plane, and suppose π has a totally ordered support. Then the minimal closed support F of π is totally ordered. Since F is totally ordered:

$$\pi(]-\infty, x[\times R) = \pi(]-\infty, x[\times]-\infty, y])$$

and

$$\pi(R \times]-\infty, y[) = \pi(]-\infty, x] \times]-\infty, y]) \quad \text{when } x, y \in F.$$

It may be deduced from this that

$$\pi(]-\infty, x] \cap]-\infty, y]) = \pi(]-\infty, x] \times R) \wedge \pi(R \times]-\infty, y]) \quad \text{for all } (x, y) \in R^2.]$$

[10] The two-point distribution in $\{a, b\}$ is greater than the two-point distribution in $\{c, d\}$ for the ordering considered, if $[a, b] \supseteq [c, d]$.

We may — without loss of generality — assume $P(\{0\})=0$. Suppose we have a "totally ordered[11]" decomposition:

$$P=\int_{a<0<b} \delta_{a,b}\, \pi(d(a,b)) \tag{4}$$

where π has a totally ordered support B. It follows from (4) that:

$$\int x^- f(x)\, P(dx)=\int f(a)\frac{ab}{a-b}\, \pi(d(a,b))$$

and

$$\int x^+ f(x)\, P(dx)=\int f(b)\frac{ab}{a-b}\, \pi(d(a,b)) \quad \text{when } f\geqq 0.$$

This may be written:

$$\int x^- f(x)\, P(dx)=\int f(a)\, \sigma_1(da)$$

and

$$\int x^+ f(x)\, P(dx)=\int f(b)\, \sigma_2(db)$$

where σ_1 and σ_2 are the projections (into the coordinate spaces) of σ where σ is given by:

$$\left.\frac{d\sigma}{d\pi}\right|_{a,b}=\frac{ab}{a-b} \quad \text{when } a<0<b.$$

It follows that $C^{-1}\sigma$ has marginals F_1 and F_2. Let $U(x,y)=(-x,y)$; $x, y\in R$. Then $(C^{-1}\sigma)\, U^{-1}$ has marginals $x \rightsquigarrow 1-F_1(-x-)$ and F_2, and $(C^{-1}\sigma)\, U^{-1}(U[B])=1$. Since $U[B]$ is totally ordered for the product order, $(C^{-1}\sigma)\, U^{-1}$ is a Frechet maximal distribution, and is consequently determined by its marginals. It follows that π is unique.

Remark 3. When $\#\Theta\geqq 3$, standard measures which do not permit "totally ordered" decompositions into experiments with extreme standard measures, may be constructed, by the remark after Corollary 31.

Acknowledgments. I wish to express my gratitude to Professor David Blackwell and Professor Lucien LeCam for encouragements and many helpful discussions.

References

1. Birnbaum, A.: On the foundations of statistical inference: Binary experiments. Ann. math. Statistics **32**, 414–435 (1961).
2. — On the foundations of statistical inference II, New York University, Institute of Mathematical Sciences. IMM-NYU. **275** (1960).
3. Blackwell, D.: Comparison of experiments. Proc. Second Berkeley Sympos. math. Statist. Probab. 93–102 (1951).
4. — Equivalent comparisons of experiments. Ann. math. Statistics **24**, 265–272 (1953).
5. — Girshick, M. A.: The theory of games and statistical decisions. New York: John Wiley and Sons 1954.
6. Personal communication from D. Blackwell.
7. Boll, C.: Comparison of experiments in the infinite case. Ph. D. thesis Stanford University. (1955).

[11] The ordering is then:

$$(x_1, y_1)\leqq(x_2, y_2) \overset{\text{definition}}{\Longleftrightarrow} x_2\leqq x_1 \quad \text{and} \quad y_1\leqq y_2.$$

8. Chernoff, H.: A measure of asymptotic efficiency for tests of a hypothesis based on the sum of observations. Ann. math. Statistics **23**, 493–507 (1952).
9. Feller, W.: An introduction to probability theory and its applications, vol. I. New York: Wiley 1960.
10. – An introduction to probability theory and its applications, vol. II. New York: Wiley 1966.
11. Heyer, Herbert: Erschöpftheit und Invarianz beim Vergleich von Experimenten. Z. Wahrscheinlichkeitstheorie verw. Geb. **12**, 21–55 (1969).
12. LeCam, L.: Sufficiency and approximate sufficiency. Ann. math. Statistics **35**, 1419–1455 (1964).
13. – Unpublished manuscript.
14. Lehmann, E. L.: Testing statistical hypotheses. New York: Wiley 1959.
15. Morse, N., Sacksteder, R.: Statistical isomorphism. Ann. math. Statistics **37**, 203–214 (1966).
16. Neveu, J.: Mathematical foundations of the calculus of probability. San Francisco: Holden-Day 1965.
17. Phelps, R.: Lectures on Choquet's theorem. New York: D. Van Nostrand 1966.
18. Strassen, V.: The existence of probability measures with given marginals. Ann. math. Statistics **36**, 423–439 (1965).
19. Torgersen, E. N.: Unpublished manuscripts. Abstracts. Ann. math. Statistics **40**, 727, 1155, 1516, 2219 (1969).

Dr. E. N. Torgersen
Department of Mathematics
University of Oslo
Blindern, Oslo 3
Norway

(Received April 23, 1969)