Z. Wahrscheinlichkeitstheorie verw. Gebiete
65, 367–384 (1984)

Zeitschrift für
Wahrscheinlichkeitstheorie
und verwandte Gebiete

Simultaneously Least Favorable Experiments

Part I: Upper Standard Functionals and Sufficiency*

Andreas Buja

Department of Statistics, GN-22, University of Washington,
Seattle, WA 98195, USA

Summary. A concept of worst-case-sufficiency is defined, generalizing Le Cam's approximate sufficiency. Instead of using total variation norm, as did Le Cam (1964), neighborhoods are described by upper expectations. A corresponding version of the theorem of Le Cam-Blackwell-Sherman-Stein is proved in the case of finite parameter space. As a main tool serve standard experiments and their upper limits, here to be called upper standard functionals. A characterization of simultaneously least favorable experiments dominated by a family of upper expectations is proved. It says that least favorable experiments exist if and only if the upper standard functional acts additively on a cone of concave functions.

Contents

1. Introduction

The starting point of the present work has been the Huber-Strassen theory of robust testing and least favorable pairs. It is shown there that under suitable

* Part I is essentially from the author's dissertation submitted in partial fulfilment for the Ph.D. degree in Mathematics at the Swiss Federal Institute of Technology

assumptions, one can construct a minimax test statistic for testing a convex composite hypothesis $\mathcal{Q}_0$ against a convex composite alternative $\mathcal{Q}_1$, and that there exists a pair $(\mathcal{Q}_0, \mathcal{Q}_1)$ in $\mathcal{Q}_0 \times \mathcal{Q}_1$ which is simultaneously least favorable among the pairs in $\mathcal{Q}_0 \times \mathcal{Q}_1$ for all testing problems, e.g. for all levels or for all a priori weights. First, Huber [10, 11] proved such results for neighbourhoods $\mathcal{Q}_{0/1}$ of distributions $\tilde{Q}_{0/1}$, taken either in ε-contamination or in total variation norm, thus robustifying the Neyman-Pearson tests of $\tilde{Q}_0$ against $\tilde{Q}_1$. Huber-Strassen [12] treated the general case of $\mathcal{Q}_{0/1}$ being dominated by 2-alternating capacities. In this way the sets $\mathcal{Q}_{0/1}$ could be described without referring to metrics and topological neighbourhoods. The formalization by capacities originated from Strassen [20].

One could ask more generally whether there exist experiments $(Q_1, Q_2, \ldots, Q_n) \in \mathcal{Q}_1 \times \mathcal{Q}_2 \times \ldots \mathcal{Q}_n$ which are simultaneously least favorable for a class of loss functions $W = (W_\theta(t))$. Confining ourselves to a given a priori distribution and to sets $\mathcal{Q}_\theta$ dominated by upper expectations v_θ ($Q_\theta \in \mathcal{Q}_\theta \Leftrightarrow Q_\theta(f) \leqq v_\theta(f)\ \forall f$), we give a simple necessary and sufficient condition. In what follows, an indexed family (v_θ) of upper expectations will be called an "approximate model".

In order to achieve this characterization of least favorable experiments, we will use sufficiency theory as a tool. According to Blackwell [3, 4], an experiment (P_θ) is sufficient for another experiment (Q_θ) (possibly on a different sample space), if there is a randomization which maps P_θ onto Q_θ for all parameters θ. The more familiar Halmos-Savage definition of sufficiency applies to a statistic rather than a pair of experiments. However, a statistic T is sufficient under the experiment (Q_θ) in the sense of Halmos-Savage, iff the distributions of T, i.e. the experiment $P_\theta = \mathcal{L}(T|Q_\theta)$ is sufficient for (Q_θ) in the sense of Blackwell (assuming suitable topological conditions). For a proof see e.g. Heyer [9]. We will always use the Blackwell definition unless otherwise stated. If we weaken this definition by requesting the existence of a randomization which maps the former experiment only into a neighbourhood of the latter, we get the concept of approximate sufficiency, as used by Le Cam [14], who described neighbourhoods in terms of total variation norm. In order to consider least favorable experiments in more general convex sets, we introduce the following concept:

An experiment (P_θ) is called *worst-case-sufficient* for an approximate model (v_θ), if there is a randomization which maps P_θ onto a distribution Q_θ dominated by v_θ, for all θ. The choice of name is motivated by the trivial fact that for any decision problem the image experiment (Q_θ) is not worse than the worst case under (v_θ).

Similar to Blackwell sufficiency and Le Cam's approximate sufficiency, we will characterize worst-case-sufficiency in terms of minimal Bayes risks. An early version (purely analytical and not yet related to statistics) of such a theorem was given by Hardy, Littlewood and Polya [7]. Sherman [18] generalized it, while Blackwell [3, 4] and Stein [19] proved it in the context of sufficiency. Cartier, Fell and Meyer [5] and Strassen [21] embedded it in the Choquet theory of measures on convex compact spaces. Finally, Le Cam [14] proved it for approximate sufficiency. Here, the notion of a standard measure

(Blackwell [3]) becomes indispensable. It generalizes the notion of a probability ratio to the case when the finite parameter space has more than two points. In this way, all decision theoretic information of an experiment is condensed in one measure on the unit simplex of $\mathbb{R}^{\Theta}$. Minimal Bayes risks may be represented as the values of the standard measure on concave functions. Corresponding to the idea of a standard measure for an experiment (Q_θ), we introduce the upper standard functional for a family (v_θ) of upper expectations (i.e. for an approximate model), allowing us to represent minimal upper Bayes risks. As special upper expectations, the upper standard functionals are only subadditive.

By showing that the upper standard functional of an approximate model acts additively on a certain cone of concave functions (corresponding to the set of loss functions in question), we prove the necessary and sufficient condition for the existence of simultaneously least favorable experiments. The proof is simply by constructing a least favorable standard measure, which yields the asserted experiment via an application of the generalized Le Cam-Blackwell-Sherman-Stein Theorem.

Acknowledgements. I would like to express my gratitude to P.J. Huber (Harvard) for presenting me this matter as a dissertation theme, to F. Hampel (ETH) for valuable discussions and for accepting this work, and to L. Le Cam (Berkeley) for suggesting improvements and providing me with references.

2. Upper Expectations

We assemble some facts about upper expectations and capacities alternating of order 2. Let Y be a Polish space and $\mathcal{A}_Y$ its Borel algebra. Further let $\mathcal{C}^b(Y)$ be the space of bounded continuous functions and $\mathcal{L}_\infty(Y)$ the space of bounded $\mathcal{A}_Y$-measurable functions on Y. Probability measures are considered as linear, positive, normalized and σ-continuous functionals on $\mathcal{L}_\infty(Y)$. They are endowed with the weak topology, i.e. the topology of pointwise convergence on bounded continuous functions on Y.

Proposition 2.1. *Let $\mathcal{Q}$ be a tight set of probability measures Q on Y. Define the upper expectation v of $\mathcal{Q}$ by*

$$v(g)=\sup\{Q(g)|Q\in\mathcal{Q}\} \quad \text{for all } g\in\mathcal{L}_\infty(Y).$$

v has the following properties:

a) $v(g_1+g_2)\leqq v(g_1)+v(g_2)$, $v(cg)=cv(g)$ *for* $c\in\mathbb{R}^+$ *and* $g_1,g_2,g\in\mathcal{L}_\infty(Y)$,
b) $v(g_1)\leqq v(g_2)$ *for* $g_1\leqq g_2$ *in* $\mathcal{L}_\infty(Y)$,
c) $v(c)$ *for* $c\in\mathbb{R}$,
d) $v(g_n)\downarrow v(g)$ *for* $g_n\in\mathcal{C}^b(Y)$, $g_n\downarrow g$ (*hence g is upper semicontinuous*),
e) $v(g)=\sup\{v(g^*)|g^*$ *upper semicontinuous, bounded,* $\leqq g\}$ *for all* $g\in\mathcal{L}_\infty(Y)$.

Proposition 2.2. *Conversely, if v is a functional on $\mathcal{L}_\infty(Y)$, satisfying* a) *to* e), *define*

$$\mathcal{Q}=\{Q|Q \text{ probability}, \leqq v \text{ on } \mathcal{L}_\infty(Y)\}$$

Then we have:

f) *Any linear functional on* $\mathscr{C}^b(Y)$, *dominated by* $v|_{\mathscr{C}^b(Y)}$, *is positive, normalized,* σ*-continuous and hence extendable to a unique element of* $\mathscr{Q}$.

g) $\mathscr{Q}$ *is convex and weakly compact.*

h) v *is the upper expectation of* $\mathscr{Q}$, *being attained for every bounded upper semicontinuous function.*

If we say that v dominates Q, we simply mean that the inequality $Q(f) \leqq v(f)$ holds for all functions in question. (This has of course nothing to do with the domination concept of the Radon-Nikodym theorem.) It should be noticed that we allow for all bounded measurable functions f in the definition of the set $\mathscr{Q}$. In statement f), we consider all bounded continuous functions. It is required to allow functions of arbitrary sign, not only nonnegative ones, to make sure that the dominated functionals are positive and normalized.

The motivation of why to use compactly generated upper expectations is provided by statement f). There is no other way to ensure σ-continuity of the dominated linear functionals. One derives σ-continuity from the continuity assumption d), which is equivalent to tightness of the set $\mathscr{Q}$.

One could wonder about introducing probability measures as functionals on $\mathscr{L}_\infty(Y)$, since we use the weak topology only. The reason is that we will have to deal intensively with randomizations and compositions thereof (Sect. 3). The range of a randomization is contained in $\mathscr{L}_\infty(Y)$, but not necessarily in $\mathscr{C}^b(Y)$.

Proof of Proposition 2.2.: ad f): Let Q^* be a linear functional on $\mathscr{C}^b(Y)$, dominated by v. Then Q^* is positive by b): for $f \leqq 0$, we have $Q^*(f) \leqq v(f) \leqq v(0) = 0$. That Q^* is normalized is seen as follows: $Q^*(-1) \leqq v(-1) = -1$ and $Q^*(1) \leqq v(1) = 1$, hence $Q^*(1) = 1$. From d) and positivity of Q^*, we conclude that Q^* is σ-continuous. The Daniell-Stone procedure provides an extension Q, which is a regular probability. To assure that Q is an element of $\mathscr{Q}$, we have to show that it is dominated by v on all measurable functions. Assumption d) implies that this is true for all upper semicontinuous functions. Since Q is regular, there exists to any function $f \in \mathscr{L}_\infty(Y)$ a nondecreasing sequence of semicontinuous functions f_n which approximate the expectation $Q(f)$ from below: $f_n \leqq f$, $Q(f_n) \uparrow Q(f)$. From this we see that Q is dominated by v on f as well: $Q(f) = \sup_n Q(f_n) \leqq \sup_n v(f_n) \leqq v(f)$.

ad g): The preceding steps have shown that, by regularity, a probability measure is dominated by v on the measurable functions iff it is dominated on the continuous functions only. Thus the set $\mathscr{Q}$ is weakly closed. Tightness is equivalent to d). Hence $\mathscr{Q}$ is weakly compact. Convexity is trivial.

ad h): The proof is in three steps.

(*) For every nonzero $g \in \mathscr{C}^b(Y)$, there is a $Q \in \mathscr{Q}$ satisfying $Q(g) = v(g)$: Define a linear form Q on $\mathbb{R} \cdot g$ by $Q(cg) = cv(g)$. Using subadditivity of v, Hahn-Banach yields an extension of Q to $\mathscr{C}^b(Y)$, dominated by v. Then f) provides the asserted probability Q.

(**) The same holds for g bounded upper semicontinuous: There is a sequence $g_n \downarrow g$, $g_n \in \mathscr{C}^b(Y)$. We have

$$\sup\{Q(g_n) | Q \in \mathscr{Q}\} \downarrow \sup\{Q(g) | Q \in \mathscr{Q}\}$$

since $Q \mapsto Q(g_n)$ is continuous on the compact set $\mathscr{Q}$. From (*), we conclude $v(g) = \sup\{Q(g) | Q \in \mathscr{Q}\}$. The function $Q \mapsto Q(g)$ is upper semicontinuous on the compact $\mathscr{Q}$, hence the supremum is attained.

(***) v is the upper expectation of $\mathscr{Q}$: This follows from (**) and from e). □

In general, an upper expectation v is not uniquely determined by its upper probability $v(B) = v(I_B)$ on $\mathscr{A}_Y$. This however is true if v is generated by the Choquet integral of a capacity alternating of order 2. Consider the following properties of a setfunction w on $\mathscr{A}_Y$:

i) $w(\phi) = 0$, $w(Y) = 1$
j) $w(A) \leqq w(B)$ for $A \subset B$ in $\mathscr{A}_Y$
k) $w(A_n) \uparrow w(A)$ for $A_n \uparrow A$ in $\mathscr{A}_Y$
l) $W(F_n) \downarrow w(F)$ for $F_n \downarrow F$ closed sets of Y
m) $w(A \cap B) + w(A \cup B) \leqq w(A) + w(B)$ for A, B in $\mathscr{A}_Y$

w is called a (normalized) capacity, if it satisfies i) to l). It is 2-alternating, if it satisfies m). For any positive, monotone setfunction w, the Choquet integral $\tilde{w}$ is defined by

$$\tilde{w}(g) = \int_0^\infty w[g > t]\, dt = \int_0^\infty w[g \geqq t]\, dt \qquad \text{for } g \in \mathscr{L}_\infty^+(Y).$$

Since we intend to make of $\tilde{w}$ an upper expectation, we need an extension to the whole space $\mathscr{L}_\infty$, i.e. to functions g of arbitrary sign:

Proposition 2.3. *Let w be a normalized, monotone setfunction on $\mathscr{A}_Y$ (properties* i) *and* j)). *Then the Choquet integral of w satisfies:*

$$\tilde{w}(cg) = c\tilde{w}(g), \qquad \tilde{w}(g + c) = \tilde{w}(g) + c \quad \text{for } c \in \mathbb{R}^+, \quad g \in \mathscr{L}_\infty^+(Y).$$

By the latter equality, $\tilde{w}$ is extendable to $\mathscr{L}_\infty(Y)$, putting simply:

$$\tilde{w}(g - c) = \tilde{w}(g) - c \quad \text{for } c \in \mathbb{R}^+, \quad g \in \mathscr{L}_\infty^+(Y)$$

This definition is independent of the special representation of $g - c$.

Proposition 2.4. *Let w be a normalized, monotone, 2-alternating setfunction* (i), j), m)) *on $\mathscr{A}_Y$. Then w is a capacity* (k), l)) *iff $\tilde{w}$ satisfies properties* 2.1 d) *and* e), *i.e. $\tilde{w}$ is continuous on decreasing sequences of continuous functions, and its values on measurable functions may be approximated from below by upper semicontinuous functions.*

This follows immediately from the wellknown fact that a 2-alternating capacity is regular from below:

$$w(A) = \sup\{w(F) | F \subset A, F \text{ closed}\} \qquad \text{for all } A \in \mathscr{A}_Y$$

(Choquet's theorem of capacitability).

Proposition 2.5. *Let w be a normalized, monotone setfunction on $\mathscr{A}_Y$. Then w is 2-alternating iff $\tilde{w}$ is subadditive* (a) $\Leftrightarrow$ m)).

The if-part is proved by a simple application of the Choquet integral to $l_A + l_B = l_{A\cup B} + l_{A\cap B}$. The converse direction is more involved. Here is a straight forward proof. (For the original proof, see Choquet ([6], p. 287).)

- First let g_1 and g_2 be indicators: $g_1 = l_A$, $g_2 = l_B$. Subadditivity reduces to 2-alternation of w, since $\tilde{w}(l_A + l_B) = w(A \cup B) + w(A \cap B)$.
- Now let g_1 be a sum of indicators, which we may assume nonincreasing: $g_1 = \sum_{1 \leqq j \leqq J} l_{A_j}$, $A_j \supset A_{j+1}$. In order to make an induction, put $g_1^* = \sum_{1 \leqq j \leqq J-1} l_{A_j}$. By definition of the Choquet integral:

$$\tilde{w}(g_1 + l_B) = \sum_{1 \leqq j \leqq J} w((A_{j-1} \cap B) \cup A_j) + w(A_J \cap B) \quad (\text{put } A_0 = Y)$$

$$\tilde{w}(g_1^* + l_B) = \sum_{1 \leqq j \leqq J-1} w((A_{j-1} \cap B) \cup A_j) + w(A_j \cap B)$$

2-alternation yields:

$$w((A_{J-1} \cap B) \cup A_J) + w(A_J \cap B) \leqq w(A_{J-1} \cap B) + w(A_J).$$

Thus we have

$$\tilde{w}(g_1 + l_B) \leqq w(g_1^* + l_B) + w(A_J).$$

By induction:

$$\tilde{w}(g_1^* + l_B) \leqq w(g_1^*) + w(B).$$

Making use of

$$\tilde{w}(g_1) = \tilde{w}(g_1^*) + w(A_J),$$

we get

$$\tilde{w}(g_1 + l_B) \leqq \tilde{w}(g_1) + w(B).$$

- A second induction will cover the case $g_2 = \sum_{1 \leqq k \leqq K} l_{B_k}$. Again, assume $B_k \supset B_{k+1}$, and put $g_2^* = \sum_{1 \leqq k \leqq K-1} l_{B_k}$. Then $\tilde{w}(g_1 + g_2) \leqq \tilde{w}(g_1 + g_2^*) + w(B_k)$ by the preceding induction. By assumption, $\tilde{w}(g_1 + g_2^*) \leqq \tilde{w}(g_1) + \tilde{w}(g_2^*)$. Subadditivity follows because of $\tilde{w}(g_2) = \tilde{w}(g_2^*) + w(B_k)$.
- The next step consists of a passage to the limit via the usual approximation $g = \sup_n \sum_{1 \leqq i \leqq 2^n} \frac{1}{2^n} l_{\left[g > \frac{i}{2^n}\right]}$.
- At last, subadditivity extends from $\mathscr{L}_\infty^+$ to $\mathscr{L}_\infty$ by 2.3. □

Independently and earlier already, Prof. Le Cam found the same proof without publishing it. We summarize the preceding two lemmas as follows:

Proposition 2.6. *The Choquet integral $\tilde{w}$ is an upper expectation iff w is a 2-alternating capacity.*

A full account of the relationship between upper expectations and upper probabilities can be found in Wolf [23].

Later on, we will need the following proposition:

Proposition 2.7. *Let m be a positive finite measure on $[0,\infty[$, and let F_+ resp. F_- be its cumulative functions:*

$$F_+(t)=m([0,t]) \quad \text{resp.} \quad F_-(t)=m([0,t[)$$

Then we have:

$$\tilde{w}(F_+(g))=\int w[g\geqq t]\,m(dt),$$
$$\tilde{w}(F_-(g))=\int w[g>t]\,m(dt).$$

This may be seen by a change of variables. Proposition 2.7 says essentially, that $\tilde{w}$ is additive on the cone of monotone functions of g.

3. Randomizations

Since we are dealing only with finite parameter space, we shall not follow Le Cam's, but the conventional framework, indicated by σ-additivity and corresponding regularity conditions. A first step in this direction has been done in the preceding section with the restriction to Polish spaces and to weakly compact sets of probability measures. The following is a modified adaptation of Le Cam ([14], Sect. 3). Since the set of all transition probabilities lacks convenient topological properties, we are forced to embed them in a larger class of transformations. Let X and Y be Polish spaces. Then we define:

An (ordinary) randomization M from X to Y is a linear map $g\mapsto M(g)$, $\mathscr{L}_\infty(Y)\to\mathscr{L}_\infty(X)$, which is

positive: $M(g)\geqq 0$ if $g\geqq 0$
normalized: $M(l_Y)=l_X$
σ-continuous: $M(g_n)\downarrow 0$ if $g_n\downarrow 0$.

Given a probability P on X, we define a P-generalized randomization as a map $M\colon \mathscr{L}_\infty(Y)\to\mathscr{L}_\infty(X)$, for which linearity, positivity and normalization holds almost surely, the exceptional sets of probability zero depending on the functions used. Note that σ-continuity is not contained in this definition.

A restricted randomization will be a randomization from X to Y of the form $M=\sum_{1\leqq i\leqq n} f_i\cdot\delta_{y_i}$, where

$$f_i\in\mathscr{C}^b_+(X),\ \sum_{1\leqq i\leqq n} f_i=l_X \quad \text{and} \quad y_i\in Y.$$

Ordinary randomizations are exactly those linear maps which are induced by transition probabilities. Hence the notation $M(g)_x=\int M(x,dy)\cdot g(y)$ makes sense. Clearly, we have the inclusions "restricted"$\subset$"ordinary"$\subset$"P-generalized".

The following lemma shows how generalized randomizations may be substituted in our context:

Lemma 3.1. *If v is a compactly generated upper expectation (as will always be assumed) on Y, such that $PM_0\leqq v$ for a P-generalized randomization M_0, then there exists also an ordinary randomization M, such that $PM\leqq v$ and $M(g)=M_0(g)$ P-a.s. for all $g\in\mathscr{C}^b(Y)$.*

Proof. Put $M_1 = M_0|_{\mathscr{C}^b(Y)}$. PM_1 is a linear, positive, normalized, σ-continuous functional on $\mathscr{C}^b(Y)$, since $PM_1 \leqq v|_{\mathscr{C}^b(Y)}$. PM_1 may be extended to a probability Q on Y. By 2.2f), we have $Q \leqq v$. Let π be the natural projection $\mathscr{L}_\infty(X) \to L_\infty(X,P)$. Then $\pi M_1 \colon \mathscr{C}^b(Y) \to L_\infty(X,P)$ may be extended to a map $M_2 \colon L_\infty(Y,Q) \to L_\infty(X,P)$, since $P(|M_1(g_n)|) \to 0$ if $Q(|g_n|) \to 0$ and $g_n \in \mathscr{C}^b(Y)$, and since $\mathscr{C}^b(Y)$ is dense in $L_\infty(Y,Q)$ with respect to L_1-norm. M_2 inherits σ-continuity from Q. Neveu ([16], Prop. V4.4) yields an ordinary randomization M, such that $M(g) = M_0(g)$ P-a.s. for all $g \in \mathscr{C}^b(Y)$. This insures $PM = Q \leqq v$. □

We shall now see what makes P-generalized randomizations so useful. Denote their set by $\mathscr{T}_P(X,Y)$. Endow $\mathscr{T}_P(X,Y)$ with the topology of pointwise convergence, where $\mathscr{L}_\infty(X)$ inherits the $\sigma(L_\infty, L_1)$-topology of $L_\infty(X,P)$, for which it won't be separated. Then we have:

Lemma 3.2. *The linear forms* $M \mapsto \int M(g) \cdot f dP$ *are continuous on* $\mathscr{T}_P(X,Y)$, *for all* $g \in \mathscr{L}_\infty(Y)$ *and* $f \in \mathscr{L}_1(X,P)$.

Proof. These forms even induce the topology. □

Lemma 3.3. $\mathscr{T}_P(X,Y)$ *is convex and quasicompact.*

Proof. Convexity is clear. Quasicompactness follows from an application of Tychonoff and from the $\sigma(L_\infty, L_1)$-compactness of the closed unit ball of $L_\infty(X,P)$. □

The following is a weak version of a very strong density result of Le Cam (Theorem 1 [14]). However, we are content with Proposition 3.4 in its present form. It can be proved with more elementary tools than Le Cam's theorem.

Proposition 3.4. *The restricted randomizations form a dense subset of* $\mathscr{T}_P(X,Y)$.

Proof. Given $T \in \mathscr{T}_P(X,Y)$, $g_1, \ldots, g_m \in \mathscr{L}_\infty(Y)$, $f_1, \ldots, f_m \in \mathscr{L}_1(X,P)$ and $\varepsilon > 0$, one has to show the existence of $T^* = \sum_{1 \leqq i \leqq n} a_i \cdot \delta_{y_i}$ (where $a_i \in \mathscr{C}^b_+(X)$) such that $|\int [T(g_j) - T^*(g_j)] \cdot f_k dP| < \varepsilon$ for $j = 1, \ldots, m, k = 1, \ldots, n$. Since $\mathscr{C}^b(X)$ is dense in L_1-spaces, it is enough to find $a_i \in \mathscr{L}^+_\infty(X)$ instead of $\mathscr{C}^b_+(X)$. Further, it is enough to consider measurable step functions $g_1, \ldots, g_m$, because $\mathscr{T}_P(X,Y)$ is equicontinuous for (ess) sup norms on $\mathscr{L}_\infty(X)$ and $\mathscr{L}_\infty(Y)$. Hence we may assume $g_j = \sum_k b_{jk} \cdot l_{B_k}$, b_{jk} constants, (B_k) a finite measurable partition of Y, $B_k \neq \emptyset$. Select $y_k \in B_k$ for all k, and let $T^* = \sum_k T(l_{B_k}) \cdot \delta_{y_k}$. Then we have $T^*(g_j) = T(g_j)$. □

Now we turn to conditional expectations. Again, let P be a probability on X, and M an ordinary randomization from X to Y. Denote by fdP the (signed) measure

$$f_0 \mapsto \int f_0 \cdot f dP = f dP(f_0).$$

Under the map M, we obtain a signed measure $(fdP)M$, which is absolutely continuous with respect to PM. Thus the map

$$f \mapsto \frac{d(fdP)M}{dPM}, \qquad L_\infty(X,P) \to L_\infty(Y,PM)$$

is well defined (Radon-Nikodym) and linear, positive, normalized, σ-continuous. A regular version thereof, i.e. a randomization which induces this map, does exist and will be called a conditional expectation given M under P, denoted $E_P^M(y, dx)$. It is immediately seen:

Lemma 3.5. *If $Q = PM$, then $QE_P^M = P$, and $(f dP)M = (E_P^M f)\, dQ$.*

In other words: E_P^M sends Q back onto P. If f is a density of a P-continuous measure with respect to P, then $E_P^M(f)$ is a density of the image measure with respect to Q. The latter equation might look more familiar in the following form:

$$g \cdot E_P^M(f)\, dQ = M(g) \cdot f dP.$$

4. Models and Sufficiency

We give the essential definitions and assumptions. The parameter set Θ is assumed finite and fixed. Sample spaces X, Y are always Polish.

A family (P_θ) of probabilities on X, indexed by $\theta \in \Theta$, will be called a model or an experiment on X. Models on Y will be denoted (Q_θ). An *approximate model* on Y shall be an indexed family (v_θ) of compactly generated upper expectations (not necessarily induced by 2-alternating capacities). In contrast, we call (Q_θ) an exact model.

A model (P_θ) on X is called sufficient for the model (Q_θ) on Y, if there is a randomization M from X to Y, such that $Q_\theta = P_\theta M$ for all $\theta \in \Theta$ (Blackwell [4]). (P_θ) is called *worst-case-sufficient* for the approximate model (v_θ) on Y, if there exists M such that $P_\theta M \leqq v_\theta$. This concept describes the situation that one experiment (P_θ) is better than the worst case dominated by (v_θ).

Note that it doesn't matter whether we formulate sufficiency with ordinary or P-generalized randomizations, where P is a probability with respect to which all the P_θ are absolutely continuous:

Proposition 4.1. *If there is a P-generalized randomization M_0, such that $P_\theta M_0 \leqq v_\theta$ for all θ, then there is also an ordinary one with the same property.*

Proof. M_0 is also $\bar{P}$-generalized $\left(\bar{P} = \frac{1}{|\Theta|} \sum_\theta P_\theta\right)$. Since $PM_0 \leqq \frac{1}{|\Theta|} \sum_\theta v_\theta$, there is by 3.1 an ordinary randomization M such that $M(g) = M_0(g)$ $\bar{P}$-a.s. $(g \in \mathscr{C}^b(Y))$. From this and 2.2f), it follows $P_\theta M \leqq v_\theta$. □

5. Standard Experiments

Standard experiments will be our technical main tools. They were first used by Blackwell [3]. A further reference is Torgersen [22]. Notations and facts, introduced in this section, will be of constant use throughout the remainder.

Let K be the unit simplex of $\mathbb{R}^\Theta$: $K = \{(z_\theta) | z_\theta \geqq 0, \sum_\theta z_\theta = 1\}$. To every model (Q_θ) on a sample space Y, let q_θ be a density of Q_θ with respect to $\sum_\theta Q_\theta$, such

that $\sum_\theta q_\theta = l_Y$. Further, let $q=(q_\theta)$ be the vector of densities, which we may consider as a map q: $Y \to K$. If $S_\theta^{(Q)}$ is the distribution of q under Q_θ, we call the family $(S_\theta^{(Q)})$ the standard model or standard experiment of (Q_θ). $S_\theta^{(Q)}$ is a probability on K, and we have $S_\theta^{(Q)} = Q_\theta T^q$, where T^q is the natural "randomization" $h \mapsto h \circ q$, $\mathscr{L}_\infty(K) \to \mathscr{L}_\infty(Y)$ from Y to K, associated with the point map q.

The standard measure $S^{(Q)}$ of the model (Q_θ) is defined as the distribution of q under $\bar{Q} = \frac{1}{|Q|}\sum_\theta Q_\theta$, i.e. $S^{(Q)} = \bar{Q}T^q$. Trivially, $S^{(Q)}$ is also given by $S^{(Q)} = \frac{1}{|\Theta|}\sum_\theta S_\theta^{(Q)}$. Conversely, a standard measure determines uniquely its standard model by $S_\theta^{(Q)} = |\Theta| \cdot z_\theta \, dS^{(Q)}$, where z_θ is the θ-component of $z \in K$. From normalization of the $S^{(Q)}$, it follows $\int z_\theta \, dS^{(Q)} = \frac{1}{|\Theta|}$. Since the projections $z \mapsto z_\theta$ form a linear base of all affine functions on K, we have: $\int a \, dS^{(Q)} = a(e)$ for all affine functions a on K $\left(e = \frac{1}{|\Theta|}(l, \ldots, l) \in K\right)$. In other words: $S^{(Q)}$ represents the point e of K.

Generally, any positive measure S on K, which represents e, is called a standard measure. (It is automatically a probability.) And the family (S_θ), where $S_\theta = |\Theta| \cdot z_\theta \, dS$, is called a standard model. Any standard model is "its own" standard model, since the identity map on K is a vector of densities of (S_θ) with respect to $\sum_\theta S_\theta = |\Theta| \cdot S$.

A motivation for introducing standard experiments is given by the fact that $(S_\theta^{(Q)})$ is sufficient for (Q_θ). Indeed, the statistic $q=(q_\theta)$ is sufficient in the sense of Halmos-Savage, i.e. there exists a conditional expectation E^q given q independently of θ. The map E^q is a randomization from K to Y, which sends $(S_\theta^{(Q)})$ back onto (Q_θ). Hence $(S^{(Q)})$ is sufficient for (Q_θ) in the sense of Blackwell. Another proof follows directly from criterion b) of Theorem 7.2.

To any approximate model (v_θ) on Y, we may construct an approximate standard model $(s_\theta^{(v)})$ simply by $s_\theta^{(v)}(h) = \sup\{S_\theta^{(Q)}(h) | (Q_\theta) \leqq (v_\theta)\}$. It is equivalent to define first the upper standard functional $s^{(v)}(h) = \sup\{S^{(Q)}(h) | (Q_\theta) \leqq (v_\theta)\}$, and then the approximate standard model $s_\theta^{(v)}(h) = \sup\{S_\theta(h) | S \leqq s^{(v)}\}$. Generally, we call (upper) *standard functional* any upper expectation s on K, such that

Definition 5.1. *$s(h+a) = s(h) + a(e)$ for all $h \in \mathscr{L}_\infty(K)$ and all affine a.*

In connection with standard measures and standard functionals, those randomizations D play a special role, which leave the affine functions fixed: $D(a) = a$ for all affine a's or equivalently $D(z_\theta) = z_\theta$ for all θ. They are called dilations. If S is a standard measure, so is SD. We will use the following property of dilations:

Lemma 5.2. *If k is a continuous concave function on K, then we have: $D(k) \leqq k$.*

This is seen from $k = \inf\{a | a \text{ affine} \geqq k\}$ and $D(\inf) \leqq \inf D$ (Jensen). □

It should be noticed that the concept of a standard measure generalizes directly the wellknown information measures for pairs of distributions. A *f*-

information is defined by

$$I_f(Q_1, Q_2) = \int f\left(\frac{q_2}{q_1}\right) dQ_1 + \lim_{h \uparrow \infty} \frac{f(h)}{h} \cdot Q_2[q_1 = 0]$$

where f is a continuous convex function defined on the nonnegative numbers. Most commonly, one uses $f = -\log$, for which one obtains the wellknown additivity property on multiple independent observations (Kullback [13]). The following remark relates both notions:

Proposition 5.3. *The convex functions f on the nonnegative numbers and the concave functions k on the unit simplex K for $\Theta = \{1, 2\}$, are in a $1-1$-correspondence by*

$$k(z) = -2 \cdot f\left(\frac{z_2}{z_1}\right) \cdot z_1.$$

Corresponding f's and k's satisfy:

$$S^{(Q)}(k) = -I_f(Q_1, Q_2).$$

Because of the change of sign, the map $(Q_1, Q_2) \mapsto S^{(Q)}(k)$ is a measure of uninformativity.

6. Minimal Bayes Risks

We introduce the usual decision theoretic concepts. A loss function is an indexed family (W_θ) of functions $W_\theta \in \mathscr{L}_\infty(T)$, where T is the decision space, assumed to be Polish. A procedure is a randomization σ from a sample space Y to a decision space T. Given a model (Q_θ) on Y, a loss function (W_θ) on T and a procedure σ from Y to T, we have the risk function $\theta \mapsto Q_\theta \sigma(W_\theta)$. In general, a prior distribution would be any probability on Θ, but we shall deal only with the uniform distribution. Hence, Bayes risks are defined for us by

$$R((Q_\theta), \sigma, (W_\theta)) = \frac{1}{|\Theta|} \sum_\theta Q_\theta \sigma(W_\theta).$$

Minimizing Bayes risks, we are lead to a fact which yields another motivation for the use of standard experiments.

Proposition 6.1. *To any loss function (W_θ), we construct a concave continuous function k on K by*

$$k(z) = \inf_{t \in T} \sum_\theta W_\theta(t) \cdot z_\theta.$$

Then we have:

$$\inf_\sigma R((Q), \sigma, (W)) = S^{(Q)}(k).$$

I.e. minimal Bayes risks may be expressed by standard measures.

Proof.

$$\inf_{\sigma} R((Q), \sigma, (W)) = \inf_{\sigma} \int [\sum_{\theta} \sigma(W_\theta) \cdot q_\theta(y)] \, \bar{Q}(dy)$$
$$= \int [\inf_{t} \sum_{\theta} W_\theta(t) \cdot q_\theta(y)] \, \bar{Q}(dy) = \int k(q(y)) \, \bar{Q}(dy) = S^{(Q)}(k).$$

In this chain, only the second equality is problematic. It is easily seen to hold in the case of step functions W_θ taking on only finitely many values. Approximating uniformly a general loss function by step functions the equality follows. □

A vector $q(y) = (q_\theta(y))$ may be interpreted as a posterior distribution given the observation y under the uniform prior distribution. Then it is well known that a Bayes procedure consists of deciding such that the posterior expected loss $\sum_{\theta} W_\theta(t) \cdot q_\theta(y)$ is minimized. This is the essential content of the preceding proof. The value of k at $z = (q_\theta(y))$ is thus the minimal posterior expected loss given the observation y.

It makes sense to consider Bayes risks of exact models (Q_θ) for $\bar{Q}$-generalized, ordinary and restricted procedures. But we have:

Lemma 6.2. *The value* $\inf_{\sigma} R((Q), \sigma, (W))$ *is the same, if we let* σ *vary among* $\bar{Q}$*-generalized or ordinary or restricted procedures.*

To see this, note that $\sigma \mapsto R((Q), \sigma, (W)) = \sum_{\theta} \int \sigma(W_\theta) \cdot q_\theta \, d\bar{Q}$ is continuous on the set $\mathscr{T}_{\bar{Q}}(Y, T)$ of $\bar{Q}$-generalized procedures by 3.2. Further, the restricted procedures form a dense subset by 3.4. □

Upper risks are similarly defined by substituting the exact model by an approximate one:

$$R((v), \sigma, (W)) = \frac{1}{|\Theta|} \sum_{\theta} v_\theta \, \sigma(W_\theta).$$

But only ordinary and restricted procedures make sense. Again, we have a representation in terms of standard functionals:

Proposition 6.3. $\inf_{\sigma} R((v), \sigma, (W)) = s^{(v)}(k)$.

The proof will need an application of the minimax theorem. First, let σ, where it occurs, vary only among restricted procedures $\mathscr{L}_\infty(T) \to \mathscr{C}^b(Y)$. This restriction will be dropped afterwards.

$$\inf_{\sigma} R((v), \sigma, (W)) = \inf_{\sigma} \sup_{(Q)} R((Q), \sigma, (W))$$

where (Q_θ) varies among all experiments dominated by (v_θ). The Bayes risk is an affine function of σ, and the restricted procedures σ form a convex set. Further, the models (Q_θ) under (v_θ) form a convex, compact set in the $|\Theta|$-fold product topology of the weak topology. The Bayes risk is an affine, continuous function of (Q_θ), since the σ's are assumed to be restricted:

$\sigma(W_\theta)\in\mathscr{C}^b(Y)$. Thus, the minimax theorem may be applied:

$$\inf_\sigma R((v),\sigma,(W))=\sup_{(Q)}\inf_\sigma R((Q),\sigma,(W)).$$

From 6.1 and 6.2, we conclude that the right side equals $\sup\{S^{(Q)}(k)|(Q_\theta)\leqq(v_\theta)\}$ and this is by definition $s^{(v)}(k)$. Dropping the restriction on σ, we have so far proved:

$$\inf_\sigma R((v),\sigma,(W))\leqq s^{(v)}(k).$$

The converse inequality is trivial, since it holds always

$$\inf_\sigma\sup_{(Q)}\geqq\sup_{(Q)}\inf_\sigma. \quad \square$$

As a byproduct, we note the analogue to 6.2 as far as it makes sense:

Lemma 6.4. $\inf_\sigma R((v),\sigma,(W))$ *has the same value if we let* σ *vary among restricted or among ordinary procedures.*

7. The General Theorem of Le Cam-Blackwell-Sherman-Stein

Theorem 7.1. *The following statements are equivalent:*

a) (P_θ) *on* X *is worst-case-sufficient for* (v_θ) *on* Y.

b) *There exists a model* (Q_θ) *on* Y, *dominated by* (v_θ), *and a randomization* N *from* Y *to* X, *such that* $\bar{P}=\bar{Q}N$ *and* $N(p_\theta)=q_\theta$ ($\bar{Q}$-a.s.).

c) *There exists a model* (Q_θ) *on* Y, *dominated by* (v_θ), *and a dilation* D *on* K, *such that* $S^{(P)}=S^{(Q)}D$.

d) *For any concave continuous function* k (*which is the infimum of only finitely many affine functions*), *we have* $S^{(P)}(k)\leqq s^{(v)}(k)$.

e) *To any (finite) decision space* T, *to any loss function* (W_θ) *on* T, *to any (restricted) procedure* ρ *from* Y *to* T, *and to any* $\varepsilon>0$, *there exists a procedure* σ *from* X *to* T, *such that*

$$R((P),\rho,(W))\leqq R((v),\sigma,(W))+\varepsilon.$$

Remarks. In d) and e), we may add the contents of the parentheses. Doing so, we get weaker statements, which are also equivalent.

Statement b) is a kind of Neyman-criterion, while statement c) is its analogue for standard experiments.

The randomization M, which is provided by a), and the randomization N of b) are essentially conditional expectations of each other, as will be seen from the proof.

The proof will be cyclic. The main step e) $\Rightarrow$ a) contains essentially a variant of the arguments of Le Cam (1964, p. 1473f).

a) $\Rightarrow$ b): By a), there is a randomization M from X to Y such that $Q_\theta=P_\theta M\leqq v_\theta$. Let N be a conditional expectation given M under $\bar{P}=\frac{1}{|\Theta|}\sum_\theta P_\theta$. Then we have for $\bar{Q}=\frac{1}{|\Theta|}\sum_\theta Q_\theta$: $\bar{P}=\bar{Q}N$ and $N(p_\theta)=q_\theta$ ($\bar{Q}$-a.s.) by 3.5.

b) $\Rightarrow$ c): Let E_Q^q be a conditional expectation given q under $\bar{Q}$. Denote by T^p the "randomization" $h \mapsto h \circ p = T^p(h)$, $\mathscr{L}_\infty(K) \to \mathscr{L}_\infty(X)$. Let $D_0 = E_Q^q N T^p$, N granted by b). Then we have $S^{(Q)}D = S^{(P)}$ and $D_0(z_\theta) = z_\theta$ ($S^{(Q)}$-a.s.), since $|\Theta| \cdot z_\theta$ is a density of $S_\theta^{(P)}$ with respect to $S^{(P)}$ and it is also a density of $S_\theta^{(Q)}$ with respect to $S^{(Q)}$. The equality $D_0(z_\theta) = z_\theta$ holds strictly on a set $A \in \mathscr{A}_K$ of $S^{(Q)}$-probability 1. Put $D(h) = l_A \cdot D_0(h) + l_{A^c} \cdot h$. This is the dilation we need.

c) $\Rightarrow$ d): This follows from 5.2. We have $D(k) \leqq k$ for concave continuous functions k, hence $S^{(P)}(k) \leqq S^{(Q)}(k)$ for a model (Q_θ) dominated by (v_θ).

d) $\Rightarrow$ e): In its strong form, e) says:

$$\inf_\rho R((P), \rho, (W)) = \inf_\sigma R((v), \sigma, (W)).$$

By Sect. 6, this is equivalent to $S^{(P)}(k) \leqq s^{(v)}(k)$, where the concave function k is built from the decision space T and the loss function (W_θ) as indicated by 6.1. Clearly, finite decision spaces yield concave functions which are the infimum of only a finite number of affine functions. Thus the weak and the strong form of e) follow from the respective variants of d).

e) $\Rightarrow$ a): Assume e) in its weak form. First we drop the restriction to finite decision spaces. Let T be any Polish space and (W_θ) a loss function on T. Given a restricted procedure $\sigma = \sum_{1 \leqq i \leqq n} f_i \cdot \delta_{t_i}$ from Y to T, consider $T^* = \{t_1, \ldots, t_n\}$ as a finite decision space and σ as a procedure from Y to T^*. Then e) may be applied to T^*, so that we have the inequality:

$$\inf_\rho R((P), \rho, (W)) \leqq \inf_\sigma R((v), \sigma, (W)). \qquad (*)$$

By 6.4, it does not matter whether we let σ vary among restricted or ordinary procedures. Hence we have e) in its strong form. Now, specialize $T = Y$ and consider the ordinary procedure $\sigma = \mathrm{id} | \mathscr{L}_\infty(Y)$ from Y to Y. (Since ρ are randomizations from X to Y from now on, we write M instead of ρ.) The above inequality (*) yields:

$$\inf_M \sum_\theta P_\theta M(W_\theta) \leqq \sum_\theta v_\theta(W_\theta),$$

where M may be $\bar{P}$-generalized. If we put $U(M, (W)) = \sum_\theta [v_\theta - P_\theta M](W_\theta)$, we have $\inf_{(W)} \sup_M U(M, (W)) \geqq 0$, since the inequality holds for all loss functions on Y.

We should like to apply the minimax theorem. The set $\mathscr{T}_{\bar{P}}(X, Y)$ of all $\bar{P}$-generalized randomizations is quasicompact in a topology which makes $M \mapsto U(M, (W))$ continuous. $\mathscr{T}_{\bar{P}}(X, Y)$ is convex and U is an affine function of M. Further, the families (W_θ) form a convex set and U is a convex function of (W_θ). Thus, the minimax theorem yields the existence of a $\bar{P}$-generalized randomization M_0, such that $U(M_0, (W)) \geqq 0$ for all (W_θ). In other words: $v_\theta(g) \geqq P_\theta M_0(g)$ for all $\theta \in \Theta$ and for all $g \in \mathscr{L}_\infty(Y)$. By 3.1, there is an ordinary randomization M, such that $M(g) = M_0(g)$ ($\bar{P}$-a.s.) for all $g \in \mathscr{C}^b(Y)$. Thus we have also $P_\theta M \leqq v_\theta$. □

There is nothing essential in the restriction to a finite parameter space Θ. It is possible to prove a variant of the theorem for arbitrary Θ, but the conventional framework, dealing with σ-continuous functionals and randomizations, would carry over only to experiments (P_θ) which are absolutely continuous with respect to some σ-finite measure. Instead, one should drop σ-continuity as is done in the framework of Le Cam [14]. Further, one should replace the standard measures by conical measures (see Le Cam [15]).

For easier reference, we state also the specialization of the above theorem to exact models, leading to the classical theorem of Blackwell-Sherman-Stein:

Corollary 7.2. *The following statements are equivalent:*

a) (P_θ) *is sufficient for* (Q_θ).

b) *There is a randomization* N *from* Y *to* X, *such that* $\bar{P}=\bar{Q}N$ *and* $N(p_\theta)=q_\theta$ ($\bar{Q}$-a.s.).

c) *There is a dilation* D *on* K, *such that* $S^{(P)}=S^{(Q)}D$

d) $S^{(P)}(k)\leqq S^{(Q)}(k)$ *for all concave continuous functions* k *on* K.

e) $\inf\limits_{\rho} R((P),\rho,(W))\leqq\inf\limits_{\sigma} R((Q),\sigma,(W))$ *for all decision spaces* T *and all loss functions* (W_θ) *on* T.

In d) and e), the same restrictions are allowed as in 7.1.

8. Least Favorable Experiments

We will apply the theorem of the preceding section to characterize least favorable experiments. The condition which is necessary and sufficient for the existence of such experiments will be formulated in terms of standard measures. So it is convenient to adapt a restricted definition of being least favorable using only Bayes risks under the uniform prior distribution:

Let $\mathscr{W}$ be a family of loss functions (W_θ) on any decision spaces. We call an experiment (Q_θ) under (v_θ) least favorable for the family $\mathscr{W}$, if:

$$\inf_{\sigma} R((Q),\sigma,(W))=\inf_{\sigma} R((v),\sigma,(W)) \quad \text{for } (W_\theta) \text{ in } \mathscr{W}.$$

Actually, we shall work with the following equivalent condition:

$S^{(Q)}(k)=s^{(v)}(k)$ for all concave functions k corresponding to loss functions in $\mathscr{W}$ (see Sect. 6).

Since all loss functions generating the same k are equivalent, we may also speak of being least favorable for k.

An application of 7.2 is the following proposition:

Proposition 8.1. *A model* (Q_θ) *under* (v_θ) *is least favorable for all loss functions iff it is "least sufficient" in the sense that all models dominated by* (v_θ) *are sufficient for* (Q_θ).

Being least favorable means $S^{(Q)}(k)=\sup S^{(Q')}(k)$, where (Q'_θ) varies among all models under (v_θ). From Theorem 7.2 a) $\Leftrightarrow$ d) follows the assertion. □

The situation described in the preceding proposition is rarely found. It applies typically to the Huber-Strassen case of testing experiments.

Now let (Q_θ) be an arbitrary model under (v_θ) and put:

$$\mathscr{K} = \{k \mid S^{(Q)}(k) = s^{(v)}(k),\ k \text{ concave, continuous on } K\}.$$

The set $\mathscr{K}$ is a closed convex cone:

a) $k \in \mathscr{K} \Rightarrow c \cdot k \in \mathscr{K}$ $(c \in \mathbb{R}^+)$

b) $k_1, k_2 \in \mathscr{K} \Rightarrow k_1 + k_2 \in \mathscr{K}$

c) $k_i \in \mathscr{K}$, $k_i \to k \Rightarrow k \in \mathscr{K}$.

Thus, if (Q_θ) is least favorable for some concave k, the same holds for the elements of the closed convex coen they generate.

$s^{(v)}$ is additive on such a cone $\mathscr{K}$, since it coincides there with $S^{(Q)}$. It turns out that this is also a sufficient condition for a least favorable experiment to exist:

Theorem 8.2. *Let $\mathscr{K}$ be the closed convex cone generated by the concave functions k corresponding to the elements of a family $\mathscr{W}$ of loss functions. The following are equivalent statements:*

a) *There exists a model (Q_θ) under (v_θ) which is least favorable for all loss functions in $\mathscr{W}$.*

b) *The upper standard functional $s^{(v)}$ is additive on the cone $\mathscr{K}$.*

To prove the nontrivial implication b) $\Rightarrow$ a), we proceede as follows: Below, we show that it is possible to construct a standard measure S under $s^{(v)}$, which equals $s^{(v)}$ on the cone $\mathscr{K}$. S determines a unique standard model (S_θ) with standard measure S (see Sect. 5). Since S is dominated by $s^{(v)}$, (S_θ) is worst-case-sufficient for (v_θ) by Theorem 7.1 d) $\Rightarrow$ a). I.e. there is a randomization M from K to Y, such that $Q_\theta \leqq v_\theta$ for $Q_\theta = P_\theta M$. By the chain of inequalities $s^{(v)}(k) = S(k) \leqq S^{(Q)}(k) \leqq s^{(v)}(k)$ for $k \in \mathscr{K}$, it follows that (Q_θ) is least favorable on $\mathscr{K}$.

There remains to show the existence of S. Define a linear functional S on the linear space $\mathscr{K} - \mathscr{K}$ by

$$S(k_1 - k_2) = s^{(v)}(k_1) - s^{(v)}(k_2).$$

Additivity of $s^{(v)}$ on $\mathscr{K}$ implies that this definition is independent of the special representation $k_1 - k_2$. Subadditivity of $s^{(v)}$ on $\mathscr{K} - \mathscr{K}$ implies that S is dominated by $s^{(v)}|_{\mathscr{K}-\mathscr{K}}$. Hahn-Banach yields an extension of S to $\mathscr{C}(K)$, which is also dominated by $s^{(v)}|_{\mathscr{C}(K)}$. A further extension to $\mathscr{L}_\infty(K)$ (see 2.2f)) provides a probability S under $s^{(v)}$, which must be a standard measure since $S \leqq s^{(v)}$. □

The existence proof for S is related to a wellknown argument in the theory of measures on compact convex spaces, see Cartier-Fell-Meyer ([5], p. 441).

Theorem 8.2 is essentially a simultaneous minimax theorem. The case of $\mathscr{W}$ containing only one single loss function could be proved directly by means of the classical minimax theorem:

Corollary 8.3. *To any single loss function there exists a least favorable experiment under an approximate model.*

Statement 8.2 is a structure theorem which solves the problem of least favorability in full generality. However, for practical purposes one is content with a specialization which follows from 8.3 already. Usually a family of loss functions $\mathscr{W}$ respectively the corresponding cone $\mathscr{K}$ is given in a parametrized form. For this let A be a compact metric space and assume that the set $\mathscr{K}$ is given by a parametrization $\mathscr{K}=\{k^\alpha|\alpha\in A\}$. Assume that the map $(\alpha,z)\mapsto k^\alpha(z)$ is continuous on $A\times K$. For a finite measure $\lambda(d\alpha)$ on A it makes sense to consider the mixture $k^\lambda(z)=\int k^\alpha(z)\,\lambda(d\alpha)$. Then we have:

Theorem 8.4. *There exists a simultaneously least favorable experiment for* $\mathscr{K}=\{k^\alpha|\alpha\in A\}$ *under* (v_θ) *iff*

$$s^{(v)}(k^\lambda)=\int s^{(v)}(k^\alpha)\,\lambda(d\alpha)$$

for a finite measure λ *on* A *satisfying support* $(\lambda)=A$.

Proof. By 8.3 there exists (Q_θ) under (v_θ) which is least favorable for k^λ only. It follows:

$$\begin{aligned} s^{(v)}(k^\lambda)=S^{(Q)}(k^\lambda)&=\int S^{(Q)}(k^\alpha)\,\lambda(d\alpha)\\ &\leqq\int s^{(v)}(k^\alpha)\,\lambda(d\alpha)=s^{(v)}(k^\lambda). \end{aligned}$$

From this, from continuity of $\alpha\mapsto S^{(Q)}(k^\alpha)$, and from support $(\lambda)=A$, follows $S^{(Q)}(k^\alpha)=s^{(v)}(k^\alpha)\,\forall\alpha\in A$. □

References

1. Bednarski, T.: Minimax procedures and capacities. Dissertation, University of California, Berkeley (1976)
2. Bednarski, T.: Binary experiments, minimax tests and two-alternating capacities. Preprint, Inst. Math., Polish Acad. Sci. (1978)
3. Blackwell, D.: Comparison of experiments. Proc. Second Berkeley Sympos. Math. Statist. probability Univ. Calif. 93–102 (1951). University of California Press
4. Blackwell, D.: Equivalent comparisons of experiments. Ann. Math. Statist. **24**, 265–272 (1953)
5. Cartier, P., Fell, J.M.G., Meyer, P.A.: Comparaison des mesures portées par un ensemble convex compact. Bull. Soc. Math. France **92**, 435–445 (1964)
6. Choquet, G.: Theory of capacities. Ann. Inst. Fourier **5**, 131–292 (1953/1954)
7. Hardy, G.H., Littlewood, J.E., Polya, G.: Inequalities. Cambridge: Cambridge University Press 1934
8. Heyer, H.: Erschöpftheit und Invarianz beim Vergleich von Experimenten. Z. Wahrscheinlichkeitstheorie verw. Gebiete **12**, 21–55 (1969)
9. Heyer, H.: Mathematische Theorie statistischer Experimente. Berlin-Heidelberg-New York: Springer 1973
10. Huber, P.J.: A robust version of the probability ratio test. Ann. Math. Statist. **36**, 1753–1758 (1965)
11. Huber, P.J.: Robust confidence limits. Z. Wahrscheinlichkeitstheorie verw. Gebiete **10**, 269–278 (1968)
12. Huber, P.J., Strassen, V.: Minimax tests and the Neyman-Pearson lemma for capacities. Ann. Statist. **1**, 2, 251–263 (1973)
13. Kullback, S.: Information theory and statistics. New York: Dover Publications 1968
14. Le Cam, L.: Sufficiency and approximate sufficiency. Ann. Math. Statist. **35**, 1419–1499 (1964)
15. Le Cam, L.: Notes on asymptotic methods in statistical decision theory. Centre de Recherches Mathématiques. Université de Montréal (1974)

16. Neveu, J.: Mathematical foundations of the calculus of probability. San Francisco: Holden Day 1965
17. Parthasarathy, K.R.: Probability measures on metric spaces. New York-San Francisco-London: Academic Press 1967
18. Sherman, S.: On a theorem of Hardy, Littlewood, Polya and Blackwell. Proc. Nat. Acad. Sci., U.S.A., **37**, 826–831 (1951)
19. Stein, C.: Notes on the comparison of experiments. (mimeographed), University of Chicago (1951)
20. Strassen, V.: Meßfehler und Information. Z. Wahrscheinlichkeitstheorie verw. Gebiete **2**, 273–305 (1964)
21. Strassen, V.: The existence of probability measures with given marginals. Ann. Math. Statist. **36**, 423–439 (1965)
22. Togersen, E.N.: Comparison of experiments when the parameter space is finite. Z. Wahrscheinlichkeitstheorie verw. Gebiete **16**, 219–249 (1970)
23. Wolf, G.: Obere und untere Wahrscheinlichkeiten. Dissertation, Swiss Federal Institute of Technology, Zürich (1977)

Received June 9, 1980; in revised form May 23, 1981