STATISTICAL MEMOIRS
Institute of Mathematics
University of Oslo

No. 1.
February 1975

NOTES ON COMPARISON OF STATISTICAL EXPERIMENTS

Chapters 0-8

Lectures, 1973-74

by

Erik N. Torgersen

Recorded and supplemented

by

Bo Lindqvist

Corrections to

STATISTICAL MEMOIRS No 1 1975.

Page	Line	Old version	New version
1.15	10 ↓	$\sup_{x \in C_1} l(x) > \inf_{y \in C_2} l(y)$	$\inf_{x \in C_1} l(x) > \sup_{y \in C_2} l(y)$
2.15	9 ↓	implies $[M \geq \tau]_b \neq \emptyset$	implies $\bigcap_b [M \geq \tau]_b \neq \emptyset$
2.17	8 ↓	$I = \{i : \tilde{y}_i > -\infty\}$	$I = \{i : \tilde{y}_i > -\infty$ for all $\tilde{y} \in S-T$ such that $\Sigma l_i \tilde{y}_i > \tau\}$
2.22	6 ↑	$\text{limsup}_\alpha M(a_\alpha, b) < t$	$\text{limsup}_\alpha M(a_\alpha, b) \leq t$
2.27	10 ↓ 12 ↓ 15 ↓	theorem 2.42	theorem 2.43
4.10	3 ↓	[ ]	[19]
5. 1	9 ↓	$Q = \sum_\theta P_\theta$	$Q = \sum_\theta Q_\theta$
5. 3	6 ↑	$\Psi(\frac{\tilde{f}_1}{\Sigma \tilde{f}_\theta}, \ldots, \frac{\tilde{f}_\theta}{\Sigma \tilde{f}_\theta})$	$\Psi(\frac{\tilde{f}_1}{\Sigma \tilde{f}_\theta}, \ldots, \frac{\tilde{f}_s}{\Sigma \tilde{f}_\theta})$
5. 8	4 ↑	$P_\theta \rho L_\theta$ converges to	$P_\theta \rho_c L_\theta$ converges to
5.10	6 ↓	$Q_\theta \rho L_\theta$	$Q_\theta \sigma L_\theta$
5.23	2 ↓	subsets of F	subsets of Θ
6.12	6 ↓	$(\frac{13}{24}, \frac{11}{24}, \frac{5}{24}, -\frac{55}{24})$	$(\frac{13}{24}, \frac{11}{24}, \frac{5}{24}, -\frac{5}{24})$
7. 4	3 ↓	theorem 8.5	theorem 7.5

Corrections to

STATISTICAL MEMOIRS No 1 1975.

Page	Line	Old version	New version
7.7	11 ↑	$p(f_n \| x)$	$\rho(f_n \| x)$
8.6	1 ↓	and 8.2	and 5.28
8.6	10 ↑	$\tilde{f}_\theta = dP_{\theta\mathcal{B}} / d\pi$	$\tilde{f}_\theta = dP_{\theta\mathcal{B}} / d\pi_{\mathcal{B}}$
8.7	2 ↓ 3 ↓		Clearly, for any $s \in R$, there is a Borel-set $V_s \subseteq R^2$ such that $f_\theta \leq s \iff (\frac{f_\theta}{1+f_\theta}, \frac{1}{1+f_\theta}) \in V_s$
8.11	5 ↑	$= \frac{g_\theta}{h\sum_\theta c(\theta) g_\theta}$	$= \frac{g_0}{\sum_\theta c(\theta) g_\theta}$
8.15	1 ↑	E_0 ...	(8) E_θ...
8.16	6 ↓ 8 9 ↓	(7)	(8)
8.16	8 ↓	σ-algebra	λ- system
A. 2	10 ↑	be a 1-1 mapping	be an increasing, continuous mapping
A. 3	11 ↑	$\lim \Phi(X_{F_i}) = \alpha$	$\lim E\,\Phi(X_{F_i}) = \alpha$
B. 2	2 ↑	(λ1) and (λ2)	(λ1)
B.13	2 ↓	corollary 17.	corollary 16.
B.14	4 ↓	prop. 19 (i)	prop. 18 (i)
C. 4	1 ↓	φ_A	$\varphi(A)$
C. 4	10 ↓	a sequence of	a monotone limit of

Contents

0. INTRODUCTION TO CHAPTERS 1-8.

A non-sequential statistical experiment (or a sequential experiment with given stopping rule) consists of three parts. The first is a listing of the possible outcomes. This is the sample space of the experiment. The next part, the parameter set, is a listing of possible explaining theories. Finally there is the correspondence which to each explaining theory assigns the chance mechanism governing the random outcome.

These parts will, respectively, be formalized as: A measurable space (say $(\chi,\mathcal{A})$) , a set Θ and a map (say $\theta \to P_\theta$) from this set to the set of probability measures on the measurable space. Combining the notations used in the three parantheses above we may write the experiment in the form:

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

Most papers on experiments have treated intrinsic problems of experiments, i.e. the structure $(\chi,\mathcal{A},P_\theta : \theta \in \Theta)$ is assumed given and we investigate various derived or related structures. We may here think of the general theories of sufficiency, completeness, and invariance or on particular decision problems. The theory of sufficiency, however, indicates the need of a theory where the objects are themselves experiments. This become quite clear if we consider the equivalence (under various regularity conditions) of the "conditional expectation" definition of sufficiency and definitions (or criterions) for sufficiency in terms of risk functions.

A theory of experiments should be a theory of the statistical information carried by the experiments. Otherwise stated: an experiment should be identified with the statistical information it contains. It appears, however, difficult to provide a reasonable and explicit definition of statistical information. We avoid this obstacle by asking the fundamental question:

When does an experiment contain more information than another?

Our tasks is then to

(i) Define "more informative than"

and

(ii) Provide criterions for "more informative than".

From a decision theoretical point of view the following definition is natural:

Let $\mathcal{E}$ and $\mathcal{F}$ be two experiments having the same parameter set Θ. Then we shall say that $\mathcal{E}$ is more informative than $\mathcal{F}$ if to any decision problem and any risk function which is obtainable in $\mathcal{F}$ corresponds an everywhere smaller risk function obtainable in $\mathcal{E}$.

In this way we arrive at the partial ordering "being more informative than" for experiments having the same parameter set Θ. The concept of sufficiency corresponds to the case where $\mathcal{E}$ is a sub experiment of $\mathcal{F}$.

It is with this kind of a definition, to be expected that the ordering is not total. In fact we may, in general, expect that two experiments having the same parameter set Θ are not comparable with respect to this ordering. Thus we are led to the following generalization of the fundamental question:

How much do we loose, under the worst possible circumstances by using $\mathcal{E}$ instead of $\mathcal{F}$?

An answer to this problem may, as we shall see, be given by a non negative number; the deficiency of $\mathcal{E}$ with respect to $\mathcal{F}$.

Closely associated to the notion of deficiency is a distance for experiments or, equivalently, for the (undefined) amounts of information carried by the experiments.

Finally we may restrict attention to certain types of decision problems. This lead to deficiencies and distances relative to the relevant type of decision problems.

The main results of these 8 chapters are the various bounds and criterions for deficiencies.

Here is an outline of the content.

Chapters 1-2 and appendices A-C contain various mathematical tools, which are useful for the general theory, not only for these first 8 chapters. Chapter 3 is a short introduction to some of the main concepts of statistical decision theory.

Our investigation of statistical experiments begin in chapter 4. The formal definitions of the deficiency δ and the definiency for k-decision problems δ_k , $k = 1,2,\ldots$ are given here. Closely related are the distances $\Delta, \Delta_1, \Delta_2, \ldots$.

Using a minimax argument we derive in chapter 5 three criterions for deficiency. The first is a Baye's risk criterion, the second is in terms of operational characteristics (performance functions) and the last is in terms of sublinear functions and the functionals they define for experiments. It is shown that any experiment is equivalent (i.e. Δ-equivalent) to a certain experiment having the set of prior distributions as sample space. This experiment is called the standard experiment of the given experiment. Any standard experiment is uniquely defined by a certain probability measure on the set of prior distributions, the standard probability measures. It may be shown that the distances $\Delta_2, \Delta_3, \ldots$ and Δ all define metrics on the set of standard probability measures which all yield the usual weak* topology for standard probability measures. This imply, by a standard compactness argument, that the metrics are all equivalent and equivalent to, for example, the Paul Levy diagonal distance. The results in chapter 5 described so far are all derived under the assumption of a finite parameter set. It is finally shown how problems on general parameter sets may, provided the experiments are dominated, be reduced to the case of finite parameter sets.

In some situations, general comparison may be reduced to comparison by testing problems. This is, in particular, true for dichotomies and in the case of sufficiency. Convergence, in the case of a finite parameter set, may always be decided by testing problems. Some of the basic results on comparison by testing problems are derived in chapter 6.

A very useful and reasonable criterion for deficiency is the Markov kernel criterion. This criterion, which is closely related to the operational characteristics criterion of chapter 5, is the main topic of chapter 7. We restrict attention to dominated experiments having, essentially, Euclidean sample spaces. The last condition might easily have been avoided provided we had replaced our Markov kernels by Markov operators.

As stated above, the notion of "being more informative" for experiments, generalizes the notion of sufficiency. Actually this may, as is shown in chapter 8, be turned around. Chapter 8 provides an introduction to the theory of sufficiency. In particular we show, for dominated experiments, the equivalence of "risk sufficiency" and "conditional expectation" sufficiency. Some of the main results on pairwise sufficiency, minimal sufficiency and completeness are derived.

A review of some of our own research reports on the subject is given in appendix D.

These notes, i.e. chapters 1-8, are written in order to

(i) bring together, for easy reference, various background material (chapters 1-3) needed in introductory courses on comparison of experiments, decision theory and related subjects

(ii) provide a short introduction (chapters 4-8) to some of the basic results on comparison of statistical experiments.

No new ideas or results are given here, - except for some of Lindqvist's interesting examples. References are given only sparingly and then somewhat biased towards our own interests. Most of the basic results (and references) are, however, contained in Blackwell [1], [2] and in Le Cam [21]. We refer the reader to Sion [13a] for references on minimax theorems. The notion of ϵ-deficiency of one experiment relative to another was given by Le Cam in [7]. This generalized the concept of "being more informative" which was introduced by Bohnenblust, Shapley and Sherman and may be found in Blackwell [1]. Standard experiments and standard measures were used by Blackwell in [1]. Blackwell introduced also, in his paper [2], comparison for k decision problems. The hybrid of "ϵ-deficiency for k-decision problems" were treated by the author in [16].

We refer the reader to Le Cam [21] and Heyer [6] for historical remarks and further reading.

Finally we want to express our gratitude to Ruth Backer and Margrethe Bjerkeskaug who typed these notes.

Oslo January 1975

Erik N. Torgersen

1. CONVEXITY

NOTATIONS AND TERMINOLOGY

R^k is the set of k-tuples $x = (x_1,\ldots,x_k)$ where $x_1,\ldots,x_k$ are real numbers. The elements of R^k are called vectors or points.

The inner product of two vectors x and y is defined by

$$\langle x,y\rangle = \sum_{i=1}^{k} x_i y_i .$$

The norm of a vector x is given by $\|x\| = \langle x,x\rangle^{\frac{1}{2}}$.

The line segment between two points x and y is the set

$$[x,y] = \{(1-t)x + ty : t \in [0,1]\} .$$

The linear span of a set of vectors A is by definition the set of all linear combinations of vectors in A and is denoted $[A]$. It is the minimal linear space which contains A .

All sets occurring in this chapter are assumed to be in R^k , unless otherwise stated.

DEFINITION 1.1

A set C is said to be convex if

$x,y \in C \Rightarrow [x,y] \subseteq C$.

A set A is said to be affine if

$x,y \in A \Rightarrow (1-t)x + ty \in A$ for all $t \in R$.

PROPOSITION 1.2

(i) The collection of convex sets in R^k is closed under arbitrary intersections.

(ii) The collection of affine sets in R^k is closed under arbitrary intersections.

(iii) R^k is both convex and affine.

Proof: The statements are easy consequences of the definitions.

PROPOSITION 1.3

Given a set S, there exists a convex set K with the following properties.

(i) $S \subseteq K$

(ii) if C is convex and $S \subseteq C$, then $K \subseteq C$.

K is thus the minimal convex set which contains S. The set K is denoted $\langle S \rangle$ and is called the convex hull of S.

Proof: Put $K = \cap \{C : C \text{ convex}, S \subseteq C\}$.

PROPOSITION 1.4

Given a set S, there exists a minimal affine set A which contains S. The set A is denoted aff S and is called the affine hull of S.

Proof: Put $A = \cap \{B : B \text{ affine}, S \subseteq B\}$.

PROPOSITION 1.5

Let C be a convex set and let $x_1, \ldots, x_r \in C$. If $t_1, \ldots, t_r$ are non-negative numbers such that $\sum_{i=1}^{r} t_i = 1$, then $\sum_{i=1}^{r} t_i x_i \in C$. We call $\Sigma t_i x_i$ a convex combination of the points $x_1, \ldots, x_r$.

Proof: We use induction on r. If $r = 2$, the statement follows from the definition of convexity. Assume now that it is proved for $r = p-1$. Let $x_1,\dots,x_p \in C$ and $t_1,\dots,t_p \geq 0$, $\sum_{i=1}^{p} t_i = 1$. We have $t_i > 0$ for at least one i, and we may assume $t_p > 0$. Further, we may assume $t_p < 1$. (If $t_p = 1$, then the statement is trivial.)

By the induction hypothesis, $y = \sum_{i=1}^{p-1} \frac{t_i}{1-t_p} x_i \in C$, since $\sum_{i=1}^{p-1} \frac{t_i}{1-t_p} = 1$. But $\sum_{i=1}^{p} t_i x_i = (1-t_p)y + t_p x_p \in C$, since $y, x_p \in C$. This proves our statement.

PROPOSITION 1.6

Let A be an affine set and $x_1,\dots,x_r \in A$. If $t_1,\dots,t_r$ are real numbers such that $\sum_{i=1}^{r} t_i = 1$, then $\sum_{i=1}^{r} t_i x_i \in A$. We call $\sum_{i=1}^{r} t_i x_i$ an affine combination of the points $x_1,\dots,x_r$.

Proof: Analogous to prop. 1.5.

PROPOSITION 1.7

Let S be an arbitrary set in R^k.

Then

$$\text{(i)}\ \langle S\rangle = \{\sum_{i=1}^{r} t_i x_i : x_1,\dots,x_r \in S, t_1,\dots,t_r \geq 0, \sum_{i=1}^{r} t_i = 1, r=1,2,\dots\}$$

$$\text{(ii)}\ \mathrm{aff}S = \{\sum_{i=1}^{r} t_i x_i : x_1,\dots,x_r \in S, t_1,\dots,t_r \in R, \sum_{i=1}^{r} t_i = 1, r=1,2,\dots\}$$

i.e. $\langle S\rangle$ is the set of convex combinations and aff S is the set of affine combinations of points in S.

Proof: (i) Denote the set of convex combination of points in S , C_S . C_S is obviously a convex set containing S , and hence $\langle S\rangle \subseteq C_S$. The opposite inclusion follows from prop. 1.5.

The proof of (ii) is analogous.

PROPOSITION 1.8

Let A be an affine set and $x \in A$. Then $A-x = \{y-x:y \in A\}$ is a linear subspace of R^k .

Proof: Let $a-x, b-x \in A-x$. If α,β are real numbers, then

$$\alpha(a-x)+\beta(b-x) = \alpha a+\beta b+(1-\alpha-\beta)x-x \in A-x$$

since $\alpha a+\beta b+(1-\alpha-\beta)x$ is an affine combination of points in A .

PROPOSITION 1.9

Every affine set is a translation of a linear space, i.e. if A is an affine set, we can write $A = x+V$, where x is an arbitrary point of A and V is a linear subspace of R^k . Moreover, the space V is uniquely determined by A .

Proof: Let $x \in A$ and define $V = A-x$. The first part then follows from prop. 1.8.

Assume now $x_1,x_2 \in A$ and put $V_1 = A-x_1$, $V_2 = A-x_2$. To show uniqueness we have to show that $V_1 = V_2$.

Let $v_1 \in V_1$. Then $v_1 = a-x_1$ for an $a \in A$.

Writing $v_1 = (a-x_1+x_2)-x_2$ we see that $v_1 \in V_2$. Hence $V_1 \subseteq V_2$. By symmetry, $V_2 \subseteq V_1$ and the proof is complete.

PROPOSITION 1.10

Given a set S and a point $s_o \in S$.
Then $\text{aff } S = s_o+[S-s_o]$.

Proof: By prop. 1.8, $\text{aff } S-s_o$ is a linear space, and since $S-s_o \subseteq \text{aff } S-s_o$, it follows by the minimal property of $[S-s_o]$ that $[S-s_o] \subseteq \text{aff } S-s_o$.

Let now $a \in \text{aff } S$, i.e. $a = \sum_{i=1}^{r} t_i x_i$, where $\Sigma t_i = 1$, $x_1,\dots,x_r \in S$. We can write $a = s_o + \sum_{i=1}^{r} t_i(x_i-s_o)$, which shows that $a \in s_o+[S-s_o]$.

DEFINITION 1.11

Let A be an affine set and assume $A = x+V$ for an $x \in A$. We define the dimension of A, denoted $\dim A$, by $\dim A=\dim V$. (V is uniquely determined by prop. 1.9, so $\dim A$ is well-defined.)

PROPOSITION 1.12

Let S be an arbitrary set. Then

(i) $0 \in \text{aff } S \Rightarrow \dim \text{aff } S = \dim [S]$

(ii) $0 \notin \text{aff } S \Rightarrow \dim \text{aff } S = \dim [S]-1$.

Proof: (i) Assume $0 \in \text{aff } S$. We may then assume $0 \in S$, since neither $\text{aff } S$ nor $[S]$ is influenced by this assumption. From prop. 1.10 we now get $\text{aff } S = [S]$.

(ii) Assume $0 \notin \text{aff } S$ and let $s_o \in S$. We contend that $[S] = [s_o] \oplus [S-s_o]$ (direct sum). Clearly, $[S] = [s_o]+[S-s_o]$. It remains to show that $[s_o] \cap [S-s_o] = \{0\}$. Suppose

$s_o \in [S-s_o]$. Then by definition of $[S-s_o]$,
$s_o = \sum_{i=1}^{r} t_i(s_i-s_o)$ for $t_1,\dots,t_r \in R$ and $s_1,\dots,s_r \in S$, and hence $0 = (1+\Sigma t_i)s_o - \Sigma t_i s_i$. This shows that 0 is an affine combination of points in S , i.e. $0 \in \text{aff } S$, which gives a contradiction. Hence $[s_o] \cap [S-s_o] = \{0\}$, and $\dim[S] = \dim[s_o]+\dim[S-s_o] = 1+\dim[S-s_o]$.

DEFINITION 1.13

If $x \in R^k$ and $\epsilon > 0$, the open ball with center at x and radius ϵ is defined to be the set

$$N(x,\epsilon) = \{y : \|y-x\| < \epsilon\}$$

A point x is said to be an interior point of a set S if there exists an $\epsilon > 0$ such that $N(x,\epsilon) \subseteq S$. The set of interior points of S is called the interior of S and is denoted S^o . Clearly, $S^o \subseteq S$. We say that S is open if $S^o = S$.

The closure of S is the set of points x in R^k such that $N(x,\epsilon) \cap S \neq \emptyset$ for all $\epsilon > 0$. It is denoted $\bar{S}$. Clearly, $S \subseteq \bar{S}$. S is said to be closed if $S = \bar{S}$.

The boundary of S is defined by $\bar{S}-S^o$.

The following propositions are stated without proof:

PROPOSITION 1.14

If S is an arbitrary set, then

$$S^o = \cup\{G : G \text{ open}, \quad G \subseteq S\}$$

$$\bar{S} = \cap\{F : F \text{ closed}, \quad F \supseteq S\}$$

PROPOSITION 1.15

Let S be an arbitrary set and x a point in R^k. Then $x \in \bar{S}$ if and only if there is a sequence $\{x_n\}$ in S such that $x = \lim x_n$.

DEFINITION 1.16

Let A be a set. We now introduce the concept "relatively A". If $a \in A$, we define the open ball relatively A with center at a and radius ϵ to be the set $N(a,\epsilon) \cap A$ (where $N(a,\epsilon)$ is defined in 1.13). By use of the open ball relatively A we may now generalize the concepts of 1.13.

For example, let A be a set and $S \subseteq A$. $x \in S$ is said to be an interior point of S relatively A if there is an $\epsilon > 0$ such that $N(a,\epsilon) \cap A \subseteq S$.

PROPOSITION 1.17

Let C be a convex set, and let $x \in C^o$, $y \in C$. Then $[x,y]-\{y\} \subseteq C^o$ (where $[x,y] = \{(1-t)x+ty : t \in [0,1]\}$).

Proof: Since $x \in C^o$, there is an $\epsilon > 0$ such that $N(x,\epsilon) \subseteq C$.

Let $z = (1-\theta)x+\theta y$ for a $\theta \in]0,1[$.

We have to prove that $z \in C^o$. We contend that $N(z,\epsilon(1-\theta)) \subseteq C$.

Let $z' \in N(z,\epsilon(1-\theta))$ and choose $x' = \frac{1}{1-\theta}(z'-\theta y)$. We find that $x-x' = \frac{z-z'}{1-\theta}$, which implies $\|x-x'\| = \frac{\|z-z'\|}{1-\theta} < \epsilon$.

Hence $x' \in N(x,\epsilon)$ so $x' \in C$.

Since now $z' = (1-\theta)x'+\theta y$, it follows that $z' \in C$ by convexity of C . Hence $N(z,\epsilon(1-\theta)) \subseteq C$ and the proof is complete.

COROLLARY 1.18

If C is convex, then C^o is convex.

PROPOSITION 1.19

If C is convex, then $\bar{C}$ is convex.

Proof: Let $x,y \in \bar{C}$ and put $z = (1-t)x+ty$ $(0 \leq t \leq 1)$.
By prop. 1.15 there are sequences $\{x_n\}$ and $\{y_n\}$ in C such that $x_n \to x$, $y_n \to y$.
We define $z_n = (1-t)x_n+ty_n$.
Now $\{z_n\}$ is a sequence in C , and hence $z \in \bar{C}$, since $z_n \to z$. (prop. 1.15).

PROPOSITION 1.20

Assume $C \subseteq A$, where A is an affine set and C is convex. Let x be an interior point of C relatively A and $y \in C$. Then all points $z \in [x,y]$, $z \neq y$, are contained in the interior of C relatively A .

Proof: Analogous to 1.17.

LEMMA 1.21

Assume $C \subseteq A$, A is affine and C is convex. If 0 is an interior point of $\bar{C}$ relatively A , then $0 \in C$.

Proof: Since A is closed, $\bar{C} \subseteq A$. Hence $0 \in A$ and A is a linear space. Let $\{a^1,\dots,a^h\}$ be a basis of A. By the assumptions there is an $\epsilon > 0$ such that $N(0,\epsilon) \cap A \subseteq \bar{C}$. We can find $\eta > 0$ such that

$\eta a^1,\dots,\eta a^h$, $-\eta \sum_{i=1}^{h} a^i \in N(0,\epsilon)$, and we may therefore choose our basis of A so that $a^1,\dots,a^h$, $-\sum_{i=1}^{h} a^i \in \bar{C}$. By prop. 1.15 there are sequences $\{x_n^i\}$ $i = 1,\dots,h$, $\{y_n\}$ such that $x_n^i \to a^i$, $y_n \to -\sum_{i=1}^{h} a^i$, x_n^i , $y_n \in C$.

For n sufficiently large, $x_n^1,\dots,x_n^h$ are linearly independent and thus constitute a basis of A. Hence there are numbers t_n^i so that

$$y_n = t_n^1 x_n^1 + \dots + t_n^h x_n^h .$$

Since $y_n \to -\Sigma a^i$, it follows that $t_n^i \to -1$ as $n \to \infty$. Hence for sufficiently large n , $t_n^i < 0$, $i = 1,\dots,h$. We have $y_n - \sum_{i=1}^{h} t_n^i x_n^i = 0$. Division by $1 - \sum_{i=1}^{h} t_n^i$ (which is now greater than 1) shows that 0 may be written as a convex combination of points in C. It follows that $0 \in C$.

PROPOSITION 1.22

Assume $C \subseteq A$, C is convex and A is affine. Then $(\bar{C})^o = C^o$, where ^{o} denotes interior relatively A.

Proof: Let $x \in (\bar{C})^o$. Obviously $C-x$ is convex and $A-x$ is affine, and $C-x \subseteq A-x$. Since $x \in (\bar{C})^o$, $0 \in (\overline{C-x})^o$. (^{o} now denotes interior relatively $A-x$. We remark that the topological properties of a set are not affected by a translation.)

The preceding lemma yields $0 \in C-x$, i.e. $x \in C$ and hence $(\bar{C})^o \subseteq C$. Since $(\bar{C})^o$ is open (relatively A) we have $(\bar{C})^o \subseteq C^o$ by prop. 1.14 . The opposite inclusion follows from $\bar{C} \supseteq C$.

DEFINITION 1.23

A linear map $l:\mathbb{R}^k \to R$ is called a linear functional on R^k . The set of all functionals on R^k is called the dual space of R^k and is denoted R^{k*} .

From linear algebra we have the following theorem:

THEOREM 1.24

A map l on R^k is a linear functional if and only if there exists a vector $v \in R^k$ such that $l(x) = \langle v,x\rangle$ for all $x \in R^k$.

The vector v is uniquely determined by l .

Remark: If $l(x) = \langle v,x\rangle$ is a linear functional, we say that v represents l . In the following we will often use the same symbol for the functional l and the vector in R^k which represents l .

THEOREM 1.25

Let C be a convex set, $p \notin \bar{C}$. Then there is a linear functional l on R^k such that $\inf_{x\in C} l(x) > l(p)$.

Proof: Define $d(p,C) = \inf_{x\in C}\|x-p\|$. There is a sequence $\{x_n\}$ in C such that $\|x_n-p\| \to d(p,C)$. We have

$\|x_n\| \leq \|x_n-p\| + \|p\|$. $\{\|x_n-p\|\}$ is a convergent sequence and thus bounded. Hence $\{x_n\}$ is bounded and contains a convergent subsequence $\{x_{n_j}\}$.

Assume $x_{n_j} \to x_o$. Then $x_o \in \bar{C}$ and clearly $\|x_o-p\| = d(p,C)$.

Let $x \in C$ and $t \in [0.1]$. Then

$tx+(1-t)x_o \in \bar{C}$ and hence

$\|tx+(1-t)x_o-p\| \geq \|x_o-p\|$. This is equivalent to

$\|x_o-p+t(x-x_o)\|^2 \geq \|x_o-p\|^2$

or $\|x_o-p\|^2+2t\langle x_o-p,x-x_o\rangle+t^2\|x-x_o\|^2 \geq \|x_o-p\|^2$.

Dividing by t and letting $t \to 0$ we get

$\langle x_o-p,x-x_o\rangle \geq 0$ which gives

$\langle x_o-p,x\rangle \geq \langle x_o-p,p\rangle+\|x_o-p\|^2$.

Define $l \in R^{k*}$ by $l(x) = \langle x_o-p,x\rangle$.

Then for $x \in C$, $l(x) \geq l(p)+\|x_o-p\|^2 > l(p)$.

THEOREM 1.26

Let C be convex, $p \notin C^o$. Then there exists $l \in R^{k*}$, $l \neq 0$, such that $l(x) \geq l(p)$ for all $x \in C$.

Proof: If $p \notin \bar{C}$, we can use theorem 1.25. We may therefore assume $p \in \bar{C}$. By prop. 1.22, $p \notin (\bar{C})^o$. Hence p is a boundary point of $\bar{C}$ and there is a sequence $x_n \to p$ with $x_n \notin \bar{C}$. By theorem 1.25 we can for each n find a functional l_n such that

(1) $l_n(x) > l_n(x_n)$ for all $x \in \bar{C}$.

By norming, we may assume $||l_n|| = 1$. Since the unit sphere in R^k is compact, there is a subsequence l_{n_j} which converges, say to l , where $||l|| = 1$.

Since $l_n(x_n) = \sum_{i=1}^{k} l_n{}^i x_n{}^i$ we see that

$$l_{n_j}(x_{n_j}) \to \sum_{i=1}^{k} l^i p^i = l(p) .$$

Hence, as a consequence of (1), $l(x) \geq l(p)$ for all $x \in C$.

THEOREM 1.27

Let $C \subseteq A$, C convex and A affine.

Let $p \in A$ and assume p is not an interior point of C relatively A .

Then there is $l \in R^{k*}$ such that

(i) $l(x) \geq l(p)$ for all $x \in C$.

(ii) l is not constant on A .

Proof: We may assume $p \in \bar{C}$. (If $p \notin \bar{C}$, then theorem 1.25 may be applied.) As in the preceding theorem, we observe that p is not an interior point of $\bar{C}$ relatively A . Thus we can find $x_n \in A-\bar{C}$, $x_n \to p$, and by theorem 1.25 we can find $l_n \in R^{k*}$ so that $l_n(x) > l_n(x_n)$ for all $x \in C$.

We decompose l_n such that $l_n = u_n+v_n$ where $u_n \in A-p$, $v_n \perp A-p$. Then

$$l_n(x) = \langle l_n,x\rangle = \langle u_n,x\rangle + \langle v_n,x\rangle$$
$$= \langle u_n,x\rangle + \langle v_n,x-p\rangle + \langle v_n,p\rangle .$$

If $x \in A$, then $l_n(x) = \langle u_n,x\rangle+\langle v_n,p\rangle$.
Since l_n is not constant on A, $u_n \neq 0$. By norming, we may assume $\|u_n\| = 1$. There is now a subsequence $u_{n_j} \to u$, where $u \in A-p$ and $\|u\| = 1$.
Define $l \in R^{k*}$ by $l(x) = \langle u,x\rangle$. Let $x \in C$.
From $l_n(x) > l_n(x_n)$ it then follows that $\langle u_{n_j},x\rangle > \langle u_{n_j},x_{n_j}\rangle$.
Hence, by letting $n_j \to \infty$, $l(x) \geq l(p)$ and (i) is proved.
We can write $u = a-p$ for an $a \in A$. Now
$l(a)-l(p) = l(a-p) = l(u) = \|u\|^2 = 1$, and thus l is not constant on A.

DEFINITION 1.28

A hyperplane in R^k is an affine set of dimension $k-1$.

PROPOSITION 1.29

A set A is a hyperplane if and only if there is $l \in R^{k*}$, $l \neq 0$, and a number r such that $A = \{x:l(x) = r\}$.

Proof: Assume A is a hyperplane. Then by prop. 1.9,
$A = a+V$, where $a \in A$ and $\dim V = \dim A = k-1$. Let $w \in V^{\perp}$, $w \neq 0$ and define $l(x) = \langle w,x\rangle$. Put $r = \langle w,a\rangle$. Then
$l(x) = r \iff \langle w,x\rangle = \langle w,a\rangle \iff \langle w,x-a\rangle = 0 \iff x-a \in W^{\perp} = V$
$\iff x \in a+V = A$.
The "only if" part of the prop. is then proved. Let now $l \in R^{k*}$, $l \neq 0$, and a number r be given. Put $A = \{x:l(x)=r\}$.
Clearly $A \neq \emptyset$, since l is linear. Let $a \in A$. Then
$A = a+\mathrm{Ker}(l)$, which follows from the equivalences
$l(x) = r \iff l(x-a) = 0 \iff x-a \in \mathrm{Ker}\ l$.
Finally, $\dim A = \dim \mathrm{Ker}(l) = k-1$ and the proof is complete.

<u>Remark</u>: Let $l \neq 0$ be a linear functional on R^k and let $A = \{x:l(x) = r\}$ be a hyperplane. A divides the space R^k into closed half-spaces $\{x:l(x) \leq r\}$ and $\{x:l(x) \geq r\}$, lying on either side of the hyperplane A. If we replace the inequality signs by strict inequalities, we obtain the open half-spaces determined by A.

<u>THEOREM 1.30 (weak separation of convex sets)</u>

Let C_1 and C_2 be disjoint convex sets. Then there is $l \in R^{k*}$ such that $l \neq 0$ and $l(x) \geq l(y)$ for all $x \in C_1$, $y \in C_2$.

<u>Proof</u>: Define $C_1-C_2 = \{x-y : x \in C_1, y \in C_2\}$.
Clearly, C_1-C_2 is convex and $0 \notin C_1-C_2$. By theorem 1.26 we can find $l \in R^{k*}$, $l \neq 0$, so that

$l(z) \geq l(0) = 0$ for $z \in C_1-C_2$. Let now $x \in C_1$, $y \in C_2$.
It follows that

$l(x-y) \geq 0$ or $l(x) \geq l(y)$.

<u>Remark</u>: Put $r = \inf\limits_{x \in C_1} l(x)$. Then the hyperplane $A = \{x:l(x) = r\}$ is said to <u>separate</u> the sets C_1 and C_2, in the sense that each of the two closed half-spaces determined by A contains one of the sets C_1 and C_2.

<u>THEOREM 1.31 (weak separation)</u>

Let C_1 and C_2 be disjoint convex subsets of an affine set A. Then there exists $l \in R^{k*}$ such that

(i) $l(x) \geq l(y)$ for all $x \in C_1$, $y \in C_2$.

(ii) l is not constant on A.

Proof: Choose $a \in A$. The set $a+C_1-C_2 = \{a+x-y : x \in C_1 , y \in C_2\}$ is then a convex subset of A (prop. 1.6). Moreover, $a \notin a+C_1-C_2$. The theorem now follows from theorem 1.27 (analogous to the proof of theorem 1.30).

THEOREM 1.32 (strong separation)

Let C_1 and C_2 be disjoint convex sets. Assume C_1 is compact (i.e. closed and bounded) and C_2 is closed. Then there is $l \in R^{k*}$ such that

$$\sup_{x \in C_1} l(x) > \inf_{y \in C_2} l(y) .$$

Proof: Define $d(C_1,C_2) = \inf\{\|x-y\| : x \in C_1 , y \in C_2\}$. There are sequences $\{x_n\}$ in C_1 and $\{y_n\}$ in C_2 such that $\|x_n-y_n\| \to d(C_1,C_2)$.

$\{x_n\}$ is bounded since C_1 is bounded.

We have $\|y_n\| \leq \|y_n-x_n\|+\|x_n\|$.

$\{\|y_n-x_n\|\}$ is a convergent sequence and thus bounded. Hence $\{y_n\}$ is bounded.

Consequently there are subsequences

$x_{n_j} \to v$ and $y_{n_j} \to w$, with $v \in C_1$, $w \in C_2$ since C_1 and C_2 are closed.

Since the norm is a continuous operation, $d(C_1,C_2) = \|v-w\|$.

Put $l(x) = \langle v-w,x \rangle$.

We assert that $x \in C_1 \Rightarrow l(x) \geq l(v)$

$$x \in C_2 \Rightarrow l(x) \leq l(w)$$

We prove the first inequality, the other one follows in a similar way.

Let $x \in C_1$. Then

$$tx+(1-t)v \in C_1 \quad \text{for all} \quad t \in [0.1]$$

Hence $\|tx+(1-t)v-w\| \geq \|v-w\|$

or $\|v-w+t(x-v)\|^2 \geq \|v-w\|^2$

This is equivalent to

$$\|v-w\|^2+2t\langle v-w,x-v\rangle+t^2\|x-v\|^2 \geq \|v-w\|^2$$

Dividing by t and letting $t \to 0$ we get

$$\langle v-w,x-v\rangle \geq 0$$

or $\langle v-w,x\rangle \geq \langle v-w,v\rangle$

which is the same as $l(x) \geq l(v)$.

Moreover,

$$l(v)-l(w) = l(v-w) = \|v-w\|^2 > 0$$

The theorem follows.

Remark: Let $l(w) < r < l(v)$. Then C_1 and C_2 are contained in the open halfspaces determined by the hyperplane $A = \{x : l(x) = r\}$.

LEMMA 1.33

Let a be a real number. Let X be a real random variable such that $X > a$ a.s: (almost surely) and such that EX exists. Then $EX > a$.

THEOREM 1.34

Let C be a convex set. Let $X = (X_1,\dots,X_k)$ be a random vector such that $X \in C$ a.s. Assume further that $E|X_i| < \infty$ $i = 1,\dots,k$ and define $EX = (EX_1,\dots,EX_k)$.
Then $EX \in C$.

Proof: We may assume $C \subseteq A$, where A is an affine set of dimension h . We show the theorem by induction on h .
(i) $h = 1$. Let $c_0, c_1 \in C$, $c_0 \neq c_1$. Then

$$A = \{(1-t)c_0+tc_1 \ : \ t \in R\}$$

Consequently we may define a real random variable T by

$$X = (1-T)c_0+Tc_1$$

Let $I = \{t \ : \ (1-t)c_0+tc_1 \in C\}$.
I is a convex subset of R and is thus an interval. Since $P(T \in I) = 1$, lemma 1.33 gives $ET \in I$. (The existence of ET follows from the definition of T .)
We have $EX = (1-ET)c_0+ETc_1$ and hence $EX \in C$.

(ii) Suppose the theorem is proved for $\dim A = 1,2,\dots,h-1$ and suppose $\dim A = h$. Assume $EX \notin C$. Then there is $l \in R^{k*}$ such that

$$l(x) \geq l(EX) \quad \text{for all} \quad x \in C$$

and l is not constant on A .
With probability 1 we thus have $l(X) \geq l(EX)$. By taking expectation we see that equality holds, i.e. $l(X) = l(EX)$ a.s.
Let now $A' = \{x \ : \ x \in A, l(x) = l(EX)\}$
A' is affine, and since l is not constant on A , A' is a proper subset of A . Hence $\dim A' < \dim A$.

Now $X \in A' \cap C$ a.s.

Since $A' \cap C$ is a convex subset of A', it follows from the induction hypothesis that $EX \in A' \cap C \subseteq C$ which gives a contradiction. Hence $EX \in C$.

EXAMPLE 1.35

Assume $X \geq 0$ a.s. Let $0 < r < s$ and assume $EX^s < \infty$.

It follows easily that $EX^r < \infty$.

We will use the preceding theorem to prove that

$$(EX^r)^{\frac{1}{r}} \leq (EX^s)^{\frac{1}{s}}$$

Proof: Let $Y = X^r$, $Z = X^s$. We shall show that

$(EY)^{\frac{1}{r}} \leq (EZ)^{\frac{1}{s}}$.

Consider the set $A = \{(y,z) : y \geq 0,\ z \geq 0,\ y^{1/r} \leq z^{1/s}\}$.

A is a convex set, since the curve defined by $z = y^{s/r}$ is a convex curve. Now $(Y,Z) \in A$ a.s. and theorem 1.34 gives $(EY,EZ) \in A$.

DEFINITION 1.36

Let C be a convex set. A function $f : C \to \mathbb{R}$ is said to be convex if for all $x_1, x_2 \in C$, $\theta \in [0,1]$ we have

$$f((1-\theta)x_1+\theta x_2) \leq (1-\theta)f(x_1)+\theta f(x_2) .$$

f is said to be concave if $\geq$ holds.

PROPOSITION 1.37

A function $f : C \to R$ is convex if and only if the set $D = \{(x,y) : x \in C,\ y \in R,\ f(x) \leq y\}$ is a convex subset of R^{k+1}.

Proof: Assume f is convex. Let (x_1,y_1), $(x_2,y_2) \in D$ and $t \in [0,1]$. Then

$(1-t)(x_1,y_1)+t(x_2,y_2) = ((1-t)x_1+tx_2,(1-t)y_1+ty_2) \in D$

since $f((1-t)x_1+tx_2) \leq (1-t)f(x_1)+tf(x_2) \leq (1-t)y_1+ty_2$.

Hence D is convex.

Assume now D is convex. Let $x_1,x_2 \in C$ and $t \in [0,1]$. Clearly $(x_1,f(x_1))$, $(x_2,f(x_2)) \in D$. Hence

$(1-t)(x_1,f(x_1))+t(x_2,f(x_2)) \in D$ which immediately gives

$$f((1-t)x_1+tx_2) \leq (1-t)f(x_1)+tf(x_2) .$$

The proof is now complete.

THEOREM 1.38 (Jensen's Inequality)

Let C be a convex set and f a real convex function on C. Let X be a random vector such that $X \in C$ a.s. and assume that EX exists.

Then $f(EX) \leq Ef(X)$.

Proof: Assume $Ef(X)$ is finite. Let D be given as in prop. 1.37. Then $(X,f(X)) \in D$ a.s. By theorem 1.34 we get $(EX,Ef(X)) \in D$ which yields the desired inequality.

If $Ef(X) = \infty$, the inequality is trivial. We now show that $Ef(X)$ is either finite or ∞. Since f is a convex function, it may be proved from prop. 1.37 and the separation theorems for convex sets, that there is a linear function l on R^k such that $f(x) \geq l(x)+c$ for all $x \in C$ (c is a constant). It follows that $f^-(x) \leq |-l(x)-c|$. Since EX exists, $El(X) = l(EX)$ is finite and hence $Ef^-(X) < \infty$.

The last part of this chapter is devoted to sub-linear functionals, which will be of great importance in the theory of comparison of experiments.

DEFINITION 1.39

A function $\psi : R^k \to R$ is called a sub-linear functional on R^k if

(i) $\psi(x + y) \leq \psi(x) + \psi(y)$ whenever $x,y \in R^k$

(ii) $\psi(tx) = t\psi(x)$ for all $x \in R^k$, $t \geq 0$.

(The property (i) is called subadditivity, the property (ii) is called positive homogenity.)

Examples:

The norm $\|x\|$ will define a sub-linear functional, as well as each linear functional on R^k . The following statement is easily proved:

PROPOSITION 1.40

Let ψ_1 and ψ_2 be sub-linear functionals on R^k and let $c \geq 0$. Then $\psi_1 \vee \psi_2$, $\psi_1 + \psi_2$, $c\psi_1$ are all sub-linear functionals on R^k . ($\vee$ denotes maximum.)

In particular, if $l_1,\ldots,l_r$ are linear functionals, then $\psi = l_1\vee\ldots\vee l_r$ is a sub-linear functional.

DEFINITION 1.41

Ψ is the set of all sublinear functionals (on R^k).

Ψ_r is the subset of Ψ containing the functionals which may be defined as a maximum of r linear functionals.

We observe that $\Psi_1 \subseteq \Psi_2 \subseteq \ldots \subseteq \Psi$.

PROPOSITION 1.42

Let $\psi \in \Psi$. Then

(i) $\psi(0) = 0$

(ii) ψ is uniformly continuous on R^k .

(iii) ψ is a convex function on R^k .

Proof: (i) follows since $\psi(tx) = t\psi(x)$ for $t \geq 0$.

Let $x,y \in R^k$. Then

$$\begin{aligned}\psi(x) - \psi(y) &= \psi(x - y + y) - \psi(y) \\ &\leq \psi(x - y) + \psi(y) - \psi(y) \\ &= \psi(\sum_{i=1}^{k} (x_i - y_i)e_i) \\ &\leq \sum_{i=1}^{k} \psi((x_i - y_i)e_i) \\ &= \sum_{i:x_i \neq y_i} |x_i - y_i| \psi(\frac{x_i - y_i}{|x_i - y_i|} e_i) \\ &\leq \sum_{i=1}^{k} |x_i - y_i| \psi(e_i) \vee \psi(-e_i)\end{aligned}$$

By interchanging x and y we get

$$|\psi(x) - \psi(y)| \leq \sum_{i=1}^{k} |x_i - y_i| \psi(e_i) \vee \psi(-e_i)$$

and (ii) follows.

Let $x,y \in R^k$, $0 \leq \theta \leq 1$. Then

$$\begin{aligned}\psi((1 - \theta)x + \theta y) &\leq \psi((1 - \theta)x) + \psi(\theta y) \\ &= (1 - \theta)\psi(x) + \theta\psi(y)\end{aligned}$$

which proves (iii).

PROPOSITION 1.43

Let $\psi : R^k \to R$. Then $\psi \in \Psi$ if and only if ψ is positive homogenous (i.e. satisfies (ii) of Def. 1.39) and convex.

Proof: It is enough to prove that a positive homogenous and convex function ψ is sub-additive. Let $x,y \in R^k$. Then

$$\psi(x+y) = 2\psi(\tfrac{1}{2}x + \tfrac{1}{2}y)$$
$$\leq 2(\tfrac{1}{2}\psi(x) + \tfrac{1}{2}\psi(y)) = \psi(x) + \psi(y) .$$

PROPOSITION 1.44

Let $K \subseteq R^k$ be a compact convex set and define a functional ψ_K by $\psi_K(x) = \sup\limits_{y \in K} \langle x,y \rangle$. Then $\psi_K \in \Psi$. ψ_K is called the support function of K.

Proof: Let $x_1,x_2 \in R^k$. Then

$$\psi_K(x_1 + x_2) = \sup_{y\in K} \langle x_1 + x_2,y\rangle \leq \sup_{y\in K}\langle x_1,y\rangle + \sup_{y\in K}\langle x_2,y\rangle$$
$$= \psi_K(x_1) + \psi_K(x_2)$$

If $x \in R^k$ and $t \geq 0$, then $\psi_K(tx) = \sup\limits_{y\in K} \langle tx,y\rangle$

$$= t \sup_{y\in K} \langle x,y\rangle = t\psi_K(x) .$$

LEMMA 1.45

Let $a_1,\ldots,a_r \in R^k$. Then

$$\psi_{\langle a_1,\ldots,a_r\rangle}(x) = \bigvee_{i=1}^{r} \langle x,a_i\rangle \in \Psi_r$$

Proof: Let $y \in \langle a_1,\ldots,a_r\rangle$. By prop. 1.7,

$$y = \sum_{i=1}^{r} t_i a_i \ , \quad t_1,\ldots,t_r \geq 0 \ , \quad \Sigma t_i = 1 .$$

Now, $\langle x,y\rangle = \langle x,\Sigma t_i a_i\rangle = \Sigma t_i\langle x,a_i\rangle \leq V\langle x,a_i\rangle$, which implies (by a suitable choice of $t_1,\ldots,t_r$) that

$$\sup_{y\in\langle a_1,\ldots,a_r\rangle} \langle x,y\rangle = \bigvee_{i=1}^{r} \langle x,a_i\rangle .$$

COROLLARY 1.46

$\psi \in \Psi_r$ if and only if $\psi = \psi_K$, where K is the convex hull of a finite subset of R^k with at most r points.

DEFINITION 1.47

Let C be a convex set. A point $a \in C$ is called an extreme point of C if there do not exist two points $a_1, a_2 \in C$ such that $a = \frac{1}{2}(a_1 + a_2)$.

LEMMA 1.48

If a compact convex set has only finitely many extreme points, then it is the convex hull of its extreme points.

Proof: See e.g. Blackwell and Girschick p. 38.

COROLLARY 1.49

$\psi \in \Psi_r$ if and only if $\psi = \psi_K$, where K is a compact convex set with at most r extreme points.

We shall now prove that each $\psi \in \Psi$ is a supremum of linear functionals and is of the form $\psi = \psi_K$ for a compact convex set K.

LEMMA 1.50

Let C be a convex set which contains an open set and $f : C \to R$ a convex function. Let $x^o \in C^o$. Then there is a point $c \in R^k$ such that $f(x) \geq f(x^o) + \langle x - x^o, c\rangle$ for all $x \in C$.

Proof: Let $D = \{(x,y) : x \in C\ ,\ y \in R\ ,\ f(x) \leq y\}$. D is a convex subset of R^{k+1} by prop. 1.37. Now $(x^o, f(x^o)) \in D \setminus D^o$. By theorem 1.26 there is a $l \in R^{(k+1)*}$, $l \neq 0$, such that $l(x^o, f(x^o)) \leq l(x,y)$ whenever $(x,y) \in D$. We may write

$$l(x,y) = by + \langle c,x\rangle \quad \text{for some} \quad b \in R \, , \quad c \in R^k$$

and thus we have

$$bf(x^o) + \langle c,x^o\rangle \leq by + \langle c,x\rangle \quad \text{for all} \quad (x,y) \in D \, .$$

By letting $y \to \infty$ we conclude that $b \geq 0$ (this is necessary for the inequality to hold). We also observe that $b \neq 0$, since $b = 0$ would imply that x^o be a boundary point of C. For $x \in C$ we now have

$$bf(x^o) + \langle c,x^o\rangle \leq bf(x) + \langle c,x\rangle$$

and hence for each $x \in C$,

$$f(x) \geq a + \langle \tilde{c},x-x^o\rangle \quad \text{for some} \quad a \in R \, , \quad \tilde{c} \in R^k \, .$$

Since equality sign holds when $x = x^o$, we have $a = f(x^o)$ and the lemma follows.

LEMMA 1.51

Let $\psi \in \Psi$ and $x^o \in R^k$. Then there exists $c \in R^k$ such that $\psi(x) \geq \langle c,x\rangle$ for all $x \in R^k$ and equality sign holds for $x = x^o$.

Proof: ψ is convex by prop. 1.42. By lemma 1.50,

(1) $\quad \psi(x) \geq a + \langle c,x\rangle \, , \quad a \in R \, , \quad c \in R^k$

with equality sign for $x = x^o$.

If $t > 0$, then $\psi(tx) \geq a + \langle c,tx\rangle$ and by the positive homogenity,

(2) $\quad \psi(x) \geq \frac{a}{t} + \langle c,x\rangle$

By letting $t \to \infty$, we get

$$\psi(x) \geq \langle c,x\rangle \quad \text{for all} \quad x \in R^k \, .$$

It remains to prove that

(3) $\quad \psi(x^o) = \langle c,x^o\rangle$

Assume that $\psi(x^o) > \langle c,x^o\rangle$. Then by (1), $a > 0$. By letting $t \downarrow 0$ in (2), we observe, however, that we must have $a \leq 0$. This contradiction proves (3).

PROPOSITION 1.52

Let $\psi \in \Psi$. Then $\psi = \lim \psi_r$ (pointwise) for a non-decreasing sequence $\{\psi_r\}$ with $\psi_r \in \Psi_r$.

Proof: It is enough to prove that each $\psi \in \Psi$ may be expressed as a supremum of countably many linear functionals. By lemma 1.51, for each $y \in R^k$ there exists $c(y) \in R^k$ such that

$\psi(x) \geq \langle c(y),x \rangle$ for all $x \in R^k$ and

$\psi(y) = \langle c(y),y \rangle$. Consequently,

(3) $\psi(x) = \sup_{y \in R^k} \langle c(y),x \rangle$ for all $x \in R^k$.

Let S be a countable, dense subset of R^k (e.g. S = the set of points with rational coordinates). We claim that

(4) $\psi(x) = \sup_{y \in S} \langle c(y),x \rangle$ for all $x \in R^k$

If $x \in S$, then $\psi(x) = \langle c(x),x \rangle$, so (4) holds. The functions on the right sides of (3) and (4) are both continuous in x . Since a continuous mapping on a metric space is determined by its values on a dense subset, it follows that (4) must hold for any $x \in R^k$. This completes the proof.

PROPOSITION 1.53

Let $\psi \in \Psi$. Then $\psi = \psi_K$ for a compact convex set K , i.e. ψ is the support function of K .

Proof: Define $K = \{y : \langle y,x \rangle \leq \psi(x)$ for all $x \in R^k\}$

(i) $K \neq \emptyset$: This is a consequence of lemma 1.51.

(ii) K is convex and closed: For each fixed $x \in R^k$, $\langle y,x \rangle \leq \psi(x)$ defines a closed half-space. It follows that K is an intersection of closed half-spaces and thus closed and convex.

(iii) K is bounded: Let $y \in K$. Then

$$y_i = \langle y,e_i\rangle \leq \psi(e_i)$$

$$-y_i = \langle y,-e_i\rangle \leq \psi(-e_i) \quad \text{and hence}$$

$$-\psi(-e_i) \leq y_i \leq \psi(e_i) \quad \text{and boundedness follows.}$$

(iv) $\psi_K = \psi$: Let $x^o \in R^k$.

By lemma 1.51 there exists $y^o \in R^k$ such that

$$\psi(x) \geq \langle y^o,x\rangle \quad \text{for all} \quad x \in R^k$$

and equality holding for $x = x^o$.

Now $y^o \in K$ and hence

$$\psi_K(x^o) = \sup_{y \in K} \langle x^o,y\rangle \geq \langle x^o,y^o\rangle = \psi(x^o)$$

On the other hand

$$\psi_K(x^o) = \sup_{y \in K} \langle x^o,y\rangle \leq \psi(x^o) \quad \text{by the definition of} \quad K .$$

This completes the proof.

PROPOSITION 1.54

Let K_1,K_2 be compact, convex sets. Then $K_1 \subseteq K_2$ if and only if $\psi_{K_1}(x) \leq \psi_{K_2}(x)$ for all x.

Proof: The "only if"-part is trivial by 1.44. Assume now that $\psi_{K_1}(x) \leq \psi_{K_2}(x)$ for all x, and suppose $K_1 \not\subseteq K_2$. Then there is a point $z \in K_1$, $z \notin K_2$. Since K_2 is closed, it follows from theorem 1.25 that there exists $c \neq 0$ such that

$$\langle c,z\rangle > \sup_{y \in K_2} \langle c,y\rangle = \psi_{K_2}(c)$$

Hence $\langle c,z\rangle > \psi_{K_1}(c) \geq \langle c,z\rangle$. The last inequality holds since $z \in K_1$. Our contradiction proves that $K_1 \subseteq K_2$.

COROLLARY 1.55

$K_1 = K_2$ if and only if $\psi_{K_1} = \psi_{K_2}$. Thus there is a 1-1 correspondence between compact convex sets in R^k and sublinear functionals on R^k.

DEFINITION 1.56

If A,B are sets in R^k and λ a constant we define

$$A + B = \{a + b : a \in A , b \in B\}$$
$$\lambda A = \{\lambda a : a \in A\}$$

PROPOSITION 1.57

Let K_1, K_2 be compact convex sets and $\lambda \geq 0$. Then

$$\psi_{K_1+K_2} = \psi_{K_1} + \psi_{K_2}$$

$$\psi_{\lambda K_1} = \lambda \psi_{K_1}$$

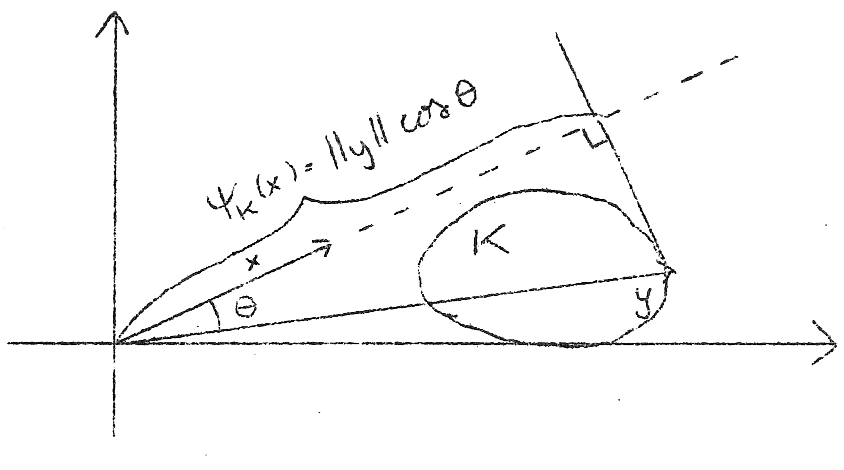
Proof: Easy consequences of the definition of ψ_K.

EXAMPLE 1.58

We assume $k = 2$. If K is a compact convex set, we define

$$\psi_K(x) = \sup_{y \in K} \langle x,y \rangle$$

We know that $\langle x,y\rangle = \|x\| \cdot \|y\| \cos \theta$, where θ is the angle between x and y. Now, if $\|x\| = 1$, $\langle x,y\rangle = \|y\| \cos \theta$ and $\psi_K(x) = \sup_{y \in K} \|y\| \cos \theta$ and $\psi_K(x)$ may be found geometrically as follows from the figure.



EXAMPLE 1.59

Let $K \subseteq R^k$ be defined by $K = \{y : \max_i |y_i| \leq 1\}$. We claim that $\psi_K(x) = \sum_i |x_i|$

This is a consequence of the following inequalities, which hold for $x \in R^k$, $y \in K$

$$\langle x,y \rangle = \sum_i x_i y_i \leq \sum_i |x_i| \ |y_i| \leq \sum_i |x_i|$$

Now, $\langle x,y \rangle = \sum_i |x_i|$ if $y_i = \text{sign } x_i$, $i = 1,\ldots,k$.

Define $K' = \{y : \sum_i |y_i| \leq 1\}$. Then $\psi_{K'}(x) = \max_i |x_i|$.

This follows from the inequalities

$$\langle x,y \rangle = \sum_i x_i y_i \leq \sum_i |x_i| \ |y_i| \leq \max_i |x_i|$$

and the fact that equality may be obtained by a suitable choice of y .

We observe the duality between ψ_K and $\psi_{K'}$ (and hence by K and K').

Many problems concerning compact convex sets or sublinear functionals may be treated by considering the dual sets or functionals.

2. GAME THEORY

DEFINITION 2.1

A two-person zero-sum game is a triple $\Gamma = (A,B,M)$ where A and B are arbitrary sets and M is a function from $A \times B$ to $[-\infty,\infty]$. The game involves two players, player I and player II. The elements of A and B are called the (pure) strategies of player I and player II, respectively. We assume that the players choose their strategies independently of each other and simultaneously. If player I uses the strategy $a \in A$ and player II uses the strategy $b \in B$, then player II "pays" player I an amount $M(a,b)$. The function M is called the pay-off function of Γ. Obviously, the sum of gain and loss is zero, as is indicated in the term "two-person zero-sum".

EXAMPLES 2.2

(a) Roulette: The two players in the case of roulette (as in many other hazard games) are the bank and the gambler. The bank has 37 strategies: one of the numbers $0,1,\ldots,36$ is chosen with equal probability by a roulette wheel.
By placing jetons on the roulette-table, the gambler chooses certain combinations of the possible outcomes. This defines the strategies of the gambler.
The payoff may be defined to be the loss of the gambler, i.e. the difference between the gambled money and the amount payed out by the bank.

(b) Many statistical problems may be regarded as two-person zero-sum games. "Nature" then takes the role of player I, choosing a parameter $\theta \in \Theta$, where Θ is the space of parameters. Without knowing the choice of nature, the statistician makes a decision d , which may for example be an estimate of θ . As a consequence of the choices, the statistician "looses" an amount $M(\theta,d)$. We call M the loss-function. In the problem of estimating a real parameter θ , the pay-off may be given by $(\theta-d)^2$ (quadratic loss). In practical statistics, observations are available for the statistician. The strategies are then decision procedures δ which determine to each possible outcome X of an experiment, which decision $\delta(X)$ to make. The pay-off is defined to be the expected loss and is called the risk-function.

(c) Finite games. Assume that each player has a finite number of strategies, i.e. A and B are finite sets. We may write $A = \{a_1,\dots,a_m\}$, $B = \{b_1,\dots,b_n\}$. If we define $m_{ij} = M(a_i,b_j)$ $i = 1,\dots,m$ $j = 1,\dots,n$, then the pay-off function of our game is given by the matrix $\{m_{ij}\}$ with elements in $[-\infty,\infty]$. Conversely, to each finite matrix $\{m_{ij}\}$ with elements in $[-\infty,\infty]$ there corresponds a finite game. m_{ij} is then the pay-off when player I uses strategy number i and player II uses strategy number j .

(d) Mr. Smith takes the train from the city to a suburban railwaystation every day. Some days he arrives home at 3 o'clock, other days at 4 o'clock. His daughter Ann wants to meet him at the station, but she is allowed to go there only once a day. If she meets her father at the station at 3, he gives her 10 cents. If she meets him at 4, she gets 20 cents. But if she fails to meet him at the time she is at the station, she gets of course nothing. If we call Ann player I and Mr. Smith player II, each player has two strategies: 3 o'clock and 4 o'clock. In accordance with example (c), the pay-off is given by the 2 × 2-matrix $\begin{pmatrix} 10 & 0 \\ 0 & 20 \end{pmatrix}$

In the following we consider a given game $\Gamma = (A,B,M)$. Each player of course wants to maximize his gain. Player I is then interested in the behaviour of $M(a_0,b)$ as a function of b for each fixed $a_0 \in A$, whence player II is interested in $M(a,b_0)$ as a function of a .

DEFINITION 2.3

Let $a_1, a_2 \in A$. We say that *a_1 dominates a_2* if $M(a_1,b) \geq M(a_2,b)$ for all $b \in B$.
If, more generally, $A_1 \subseteq A$ and $A_2 \subseteq A$ we say that *A_1 dominates A_2* if to each $a_2 \in A_2$ there exists $a_1 \in A_1$ such that a_1 dominates a_2 . If A_1 dominates A , we say that A_1 is *essentially complete*.
Let $b_1, b_2 \in B$. We say that *b_1 dominates b_2* if $M(a,b_1) \leq M(a,b_2)$ for all $a \in A$. If $B_1, B_2 \subseteq B$ we say that *B_1 dominates B_2* if for all $b_2 \in B_2$ there is $b_1 \in B_1$ such that b_1 dominates b_2 .

DEFINITION 2.4

For each $a \in A$ we define $M_I(a) = \inf\limits_{b \in B} M(a,b)$

For each $b \in B$ we define $M_{II}(b) = \sup\limits_{a \in A} M(a,b)$

Using the strategy $a \in A$, player I is certain to receive an amount of at least $M_I(a)$, and he is not guaranteed any larger amount. $M_I(a)$ is thus a measure for the "goodness" of each strategy $a \in A$ and M_I defines an ordering of the strategies of player I.

$M_{II}(b)$ is the maximum loss of player II using strategy b, and M_{II} defines an ordering of the strategies of player II.

DEFINITION 2.5

We define $\underline{V}(\Gamma) = \sup_{a \in A} M_I(a)$

and $\overline{V}(\Gamma) = \inf_{b \in B} M_{II}(b)$

$\underline{V}(\Gamma)$ is called the lower value of Γ,

$\overline{V}(\Gamma)$ is called the upper value of Γ.

When no confusion can arise, we denote the lower value of Γ by $\underline{V}$ and the upper value by $\overline{V}$.

DEFINITION 2.6

Assume that there is an $a_o \in A$ such that $M_I(a_o) = \underline{V}$. a_o is then the strategy that maximizes the minimal gain of player I and is called a maximin strategy of player I.

Similarly, player II may minimize his maximal loss by using a strategy b_o such that $M_{II}(b_o) = \overline{V}$. b_o is then called a minimax strategy for player II.

If A is an infinite set it may happen that $\sup_{a \in A} M_I(a)$ will not be attained by a certain a_o. Then we may find strategies $a \in A$ such that $M_I(a)$ lies arbitrarily near $\underline{V}$.

Similarly for B.

PROPOSITION 2.7

For all $a \in A$, $b \in B$ we have

$$M_I(a) \leq \underline{V} \leq \overline{V} \leq M_{II}(b) .$$

Proof: Let $a' \in A$, $b' \in B$.

We have $M_I(a') = \inf_{b \in B} M(a',b) \leq M(a',b') \leq \sup_{a \in A} M(a,b') = M_{II}(b')$

It follows that

$$M_I(a') \leq \inf_{b \in B} M_{II}(b) = \overline{V} .$$

By taking supremum over a' we get $\underline{V} \leq \overline{V}$. The proposition follows.

DEFINITION 2.8

A game Γ is said to have a value $V(\Gamma)$ if $\underline{V}(\Gamma)=\overline{V}(\Gamma)=V(\Gamma)$. We often write V instead of $V(\Gamma)$.

By using a maximin strategy (or a strategy that is approximately a maximin strategy) player I is guaranteed a gain of at least $\underline{V}$. On the other hand, since player II may reduce his loss to $\overline{V}$, player I is not guaranteed more than $\overline{V}$.

If now Γ has a value, i.e. $\underline{V} = \overline{V} = V$, and player I uses a strategy a_o for which $M_I(a_o) = V$ (or $M_I(a_o)$ is approximately equal to V) then a_o is an unimprovable strategy for player I, i.e. he is guaranteed an amount of V , and no other strategy can guarantee him more.

Similarly for player II.

EXAMPLE 2.9

In example 2.2(d) we find that $\underline{V} = 0$, $\overline{V} = 10$. Thus this game has no value. Ann is not guaranteed any money and Mr. Smith may reduce his expense to 10 cents.

PROPOSITION 2.10

A game Γ has a value and $a_0 \in A$ is a maximin strategy for player I and $b_0 \in B$ is a minimax strategy for player II if and only if

(1) $M(a_0,b) \geq M(a_0,b_0) \geq M(a,b_0)$ for all $a \in A$, $b \in B$.

If one of these conditionsis satisfied, then $V(\Gamma) = M(a_0,b_0)$.

Proof: Assume (1). Then $\inf_b M(a_0,b) \geq \sup_a M(a,b_0)$ which is the same as $M_I(a_0) \geq M_{II}(b_0)$.
By proposition 2.7 $M_I(a_0) = M_{II}(b_0)$ and $\underline{V} = \overline{V}$.
Assume now that $\underline{V} = \overline{V} = V$ and a_0, b_0 are maximin and minimax strategies, respectively. Then

$$M(a_0,b) \geq M_I(a_0) = V = M_{II}(b_0) \geq M(a,b_0) .$$

Choose $a = a_0$, $b = b_0$. Then $V = M(a_0,b_0)$ and (1) follows.

EXAMPLE 2.11

Consider a finite game given by a matrix $\{m_{ij}\}$. The inequality (1) of prop. 2.10 shows that the value of the game (if it exists) is an element of the matrix that is a minimal element of its row and a maximal element of its column. We call such element a saddle point of the matrix $\{m_{ij}\}$. The number of the row then defines the maximin strategy of player I, while the number of the column defines the minimax strategy of player II.

If $\{m_{ij}\}$ is given by $\begin{pmatrix} -1 & -2 & 3 & 0 \\ 2 & -1 & 0 & 6 \\ 1 & -4 & 5 & -6 \end{pmatrix}$

we observe that -1 is a saddle point. Hence $V = -1$ and a_2 is a maximin strategy of player I, b_2 is a minimax strategy of player II.

DEFINITION 2.12

Let $\Gamma = (A,B,M)$ be a game. We say that A is concave (relatively Γ) if to each pair $a_1, a_2 \in A$ and each $\theta \in [0,1]$ there exists $a \in A$ such that

(2) $M(a,b) \geq (1-\theta)M(a_1,b)+\theta M(a_2,b)$ for all $b \in B$.

If equality sign holds in (2) for all $b \in B$, then A is said to be affine (relatively Γ).

Interpretation of concavity:

Assume player I has a choice between the strategies a_1 and a_2 , and that he chooses a_2 with probability θ . The expression on the right side of (2) is then the expected gain, given that player II uses strategy b . If A is concave, then player I has a pure strategy $a \in A$ which gives at least as large gain.

Remark: Since M is assumed to be an extended real function, the undefined case $\infty - \infty$
may occur in (2). This may be avoided if we for each $b \in B$ restrict the function $M(\circ,b)$ to take at most one of the values $+\infty$ and $-\infty$. However, in the sequel we will study games where M is assumed to take at most one of the infinite values.

Similar comments may be done in connection with definition 2.14 below.

PROPOSITION 2.13

Let $\Gamma = (A,B,M)$ be a game and assume A is concave. Let $a_1,\dots,a_r \in A$ and $\theta_1\dots,\theta_r \geq 0$, $\Sigma\theta_i = 1$. Then there exists $a \in A$ such that $M(a,b) \geq \sum_{i=1}^{r} \theta_i M(a_i,b)$ for all $b \in B$. If A is affine, then equality sign holds.

Proof: Induction.

DEFINITION 2.14

Let $\Gamma = (A,B,M)$ be a game. We say that B is convex (relatively Γ) if to each pair $b_1,b_2 \in B$ and each $\theta \in [0,1]$ there exists $b \in B$ such that

(3) $M(a,b) \leq (1-\theta)M(a,b_1)+\theta M(a,b_2)$ for all $a \in A$.

If equality sign holds in (3) for all $a \in A$, we say that B is affine (relatively Γ).

Convexity of B may be interpreted in a similar way as concavity of A . We have the following analogue to prop. 2.13:

PROPOSITION 2.15

Let $\Gamma = (A,B,M)$ be a game and assume B is convex. Let $b_1,\dots,b_r \in B$ and $\theta_1,\dots,\theta_r \geq 0$, $\Sigma\theta_i = 1$. Then there exists $b \in B$ such that $M(a,b) \leq \sum_{i=1}^{r} \theta_i M(a,b_i)$ for all $a \in A$.
If B is affine, then equality sign holds.

DEFINITION 2.16

A game $\Gamma = (A,B,M)$ is said to be concave-convex if A is concave and B is convex (relatively Γ).

Remark: A sufficient condition for A to be concave rel. Γ is that A is a convex subset of an Euclidean space R^k and $M(\circ,b)$ is a real concave function on A for each $b \in B$. Given $a_1,a_2 \in A$ and $\theta \in [0,1]$ according to definition 2.12, we may put $a = (1-\theta)a_1+\theta a_2$ to satisfy (2).
Similarly, B is convex rel. Γ if B is a convex subset of an Euclidean space R^m and $M(a,\circ)$ is a real convex function on B for each $a \in A$.
That the above conditions are not necessary, becomes clear from the following example.

EXAMPLE 2.17

Consider a finite game Γ given by the matrix $\begin{pmatrix} -2 & 4 \\ 0 & 1 \end{pmatrix}$

An easy computation shows that Γ is concave-convex. Since 0 is a saddle point of the matrix, Γ has a value $V = 0$. It will be shown later that under certain conditions every concave-convex game has a value (theorem 2.35).

DEFINITION 2.18

Let X be an arbitrary set. A probability distribution over X with finite support is a non-negative real function p defined over X such that $p(x) = 0$ except for a finite number of $x \in X$ and $\sum_{x \in X} p(x) = 1$.

DEFINITION 2.19

Let $\Gamma = (A,B,M)$ be a game. The randomization of Γ is the game $\Gamma^* = (A^*,B^*,M^*)$ defined by the following:

A^* is the set of probability distributions over A with finite support.

B^* is the set of probability distributions over B with finite support.

M^* is defined on $A^* \times B^*$ by

$$(4) \quad M^*(a^*,b^*) = \sum_{a,b} M(a,b)a^*(a)b^*(b) \quad \text{for} \quad a^* \in A^* \text{ , } b^* \in B^* \text{ .}$$

Remark: The sum occurring in (4) involves finitely many terms different from zero. Thus no difficulties regarding convergence and interchange of summations will arise. We note, however, that the sum in (4) is well-defined only if M takes only one of the values $+\infty$ and $-\infty$.

Assume player I (in the game Γ^*) uses the strategy $a^* \in A^*$. In terms of the game Γ , this may be interpreted as if player I chooses a strategy $a \in A$ according to the probability distribution a^* . If we adopt a similar interpretation of the strategy $b^* \in B^*$, then it is seen from (4) that $M^*(a^*,b^*)$ is the expected gain of player I in the game Γ .

In 2.1 we introduced the notion pure strategies for the elements of A and B . The elements of A^* and B^* are called the mixed strategies of the game Γ .

If we identify the strategy $a \in A$ with the strategy $a^* \in A^*$ such that $a^*(a) = 1$, $a^*(a') = 0$ for $a' \neq a$ then A may be considered as a subset of A^* . Similarly we may assume $B \subseteq B^*$. We use the symbol a^* for the elements of A^* and the symbol a for elements of A considered as elements of A^* . Similarly for the strategies of player II.

Obviously $M^*(a,b) = M(a,b)$ for all $a \in A$, $b \in B$.

PROPOSITION 2.20

The randomization $\Gamma^* = (A^*,B^*,M^*)$ of the game $\Gamma = (A,B,M)$ is a concave-convex game. In fact, A^* and B^* are both affine relatively Γ^* .

Proof: We show that A^* is affine. The proof that B^* is affine is similar.

Let $a^{1*},a^{2*} \in A^*$ and let $\theta \in [0,1]$. Define a^* by

$$a^*(a) = (1-\theta)a^{1*}(a)+\theta a^{2*}(a) \quad \text{for} \quad a \in A .$$

Clearly, $a^* \in A^*$.

Let b^* be an arbitrary element of B^* . Then

$$M^*(a^*,b^*) = \sum_{a,b} M(a,b)a^*(a)b^*(b)$$

$$= \sum_{a,b} M(a,b)[(1-\theta)a^{1*}(a)+\theta a^{2*}(a)]b^*(b)$$

$$= (1-\theta)M^*(a^{1*},b^*)+\theta M^*(a^{2*},b^*) .$$

The affinity of A^* now follows from definition 2.12.

Let now Γ and Γ^* be as in definition 2.19.

PROPOSITION 2.21

(i) $M_I^*(a^*) = \inf_{b \in B} M^*(a^*,b)$

(ii) $M_{II}^*(b^*) = \sup_{a \in A} M^*(a,b^*)$

Proof: Since for all $a^* \in A^*$, $b^* \in B^*$,

$$M^*(a^*,b^*) = \sum_b [\sum_a M(a,b)a^*(a)]b^*(b) = \sum_b M^*(a^*,b)b^*(b)$$

$$\geq \inf_b M^*(a^*,b)$$

it follows that $M_I^*(a^*) \geq \inf_b M^*(a^*,b)$.

On the other hand, since $B \subseteq B^*$

$$M_I^*(a^*) = \inf_{b^*} M^*(a^*,b^*) \leq \inf_b M^*(a^*,b) .$$

Part (i) of the prop. follows. The proof of part (ii) is similar.

COROLLARY 2.22

(i) For any $a \in A$, $M_I^*(a) = M_I(a)$

(ii) For any $b \in B$, $M_{II}^*(b) = M_{II}(b)$.

Proof: The corollary follows from the preceding proposition and the fact that $M^*(a,b) = M(a,b)$ for all $a \in A$, $b \in B$.

COROLLARY 2.23

$\underline{V}(\Gamma) \leq \underline{V}(\Gamma^*) \leq \overline{V}(\Gamma^*) \leq \overline{V}(\Gamma)$.

Proof: From corollary 2.22 and the fact that $A \subseteq A^*$ it follows that

$$\underline{V}(\Gamma^*) = \sup_{a^*} M_I^*(a^*) \geq \sup_{a} M_I^*(a) = \sup_{a} M_I(a) = \underline{V}(\Gamma) .$$

In a similar way we get $\overline{V}(\Gamma^*) \leq \overline{V}(\Gamma)$.

COROLLARY 2.24

If Γ has a value $V(\Gamma)$, then Γ^* has a value $V(\Gamma^*)$ and $V(\Gamma^*) = V(\Gamma)$.

Proof: Easy consequence of corollary 2.23.

EXAMPLE 2.25

Let Γ be the game given in example 2.2 (d). We shall construct the game Γ^* . Ann's strategies are a_1 = "3 o'clock" , a_2 = "4 o'clock". Mr. Smith's strategies are b_1 = "3 o'clock", b_2 = "4 o'clock".

Each element a^* of A^* , being a probability distribution over $\{a_1,a_2\}$, determines a number $\theta = a^*(a_2)$.

Conversely, to each number $\theta \in [0,1]$ there corresponds a strategy $a^* \in A^*$ such that $a^*(a_2) = \theta$. We may therefore identify the set A^* and the interval $[0,1]$. The strategy a_1 of A^* now corresponds to $\theta = 0$, a_2 corresponds to $\theta = 1$.

Similarly, B^* and $[0,1]$ are identified, b_1 corresponds to 0 and b_2 to 1 .

By the definition of M^* we compute

$$M^*(\theta,\eta) = 10(1-\theta)(1-\eta)+20\theta\eta = 30\theta\eta-10\theta-10\eta+10 .$$

By prop. 2.21 (i)

$$M_I^*(\theta) = \min[M^*(\theta,0),M^*(\theta,1)] = \begin{cases} 20\theta & \text{if } 0 \leq \theta \leq \frac{1}{3} \\ 10-10\theta & \text{if } \frac{1}{3} \leq \theta \leq 1 \end{cases}$$

This gives us

$$\underline{V}(\Gamma^*) = \sup_{\theta} M_I^*(\theta) = M_I^*(\tfrac{1}{3}) = \frac{20}{3} = 6\tfrac{1}{3}$$

An analogous computation (using prop. 2.21 (ii)) shows that

$$\overline{V}(\Gamma^*) = M_{II}^*(\tfrac{1}{3}) = \frac{20}{3} = 6\tfrac{1}{3}$$

Thus the randomized game Γ^* has a value, $V(\Gamma^*) = 6\frac{1}{3}$. $\theta = \frac{1}{3}$, $\eta = \frac{1}{3}$ are, respectively, maximin and minimax strategies. Thus, the game Γ , Ann and Mr. Smith ought to choose 3 o'clock with probability 2/3 and 4 o'clock with probability 1/3.

Remark. It is enough to consider a game Γ from player I's point of view. When $\Gamma = (A,B,M)$ is given, we may namely derive the game $\tilde{\Gamma} = (B,A,\tilde{M})$, where $\tilde{M}$ is given by $\tilde{M}(b,a) = -M(a,b)$, $a \in A$, $b \in B$. Then, clearly,
$\underline{V}(\tilde{\Gamma}) = -\overline{V}(\Gamma)$, $\overline{V}(\tilde{\Gamma}) = -\underline{V}(\Gamma)$.
Hence $\tilde{\Gamma}$ has a value if and only if Γ has a value, and we observe that $\tilde{\tilde{\Gamma}} = \Gamma$.

Let now $\Gamma = (A,B,M)$ be a game. We will find conditions under which $\underline{V} = \overline{V}$.

NOTATION.

$[M \geq \tau]_b = \{a : M(a,b) \geq \tau\} \subseteq A$, $\tau \in R$.

THEOREM 2.26

The following conditions are equivalent:

(i) $\underline{V} = \overline{V}$

(ii) for each $\tau < \overline{V}$ we have $\bigcap_b [M \geq \tau]_b \neq \emptyset$.

Proof:

(ii) => (i). Let $\tau < \overline{V}$. Then there is an $a \in A$ such that $M(a,b) \geq \tau$ for all $b \in B$. Hence $M_I(a) = \inf\limits_b M(a,b) \geq \tau$. Since $\underline{V} = \sup\limits_a M_I(a)$, it follows that $\underline{V} \geq \tau$. We may choose τ arbitrary near $\overline{V}$, so $\underline{V} \geq \overline{V}$ and hence $\underline{V} = \overline{V}$.

(i) => (ii). Let $\tau < \overline{V}$. Then $\tau < \underline{V} = \sup\limits_a M_I(a)$.

There is thus an $a \in A$ such that

$\tau < M_I(a) = \inf\limits_b M(a,b) \leq M(a,b)$ for all b.

Hence $a \in [M \geq \tau]_b$ for all $b \in B$, which implies $[M \geq \tau]_b \neq \emptyset$.

THEOREM 2.27 (FUNDAMENTAL THEOREM ABOUT CONCAVE-CONVEX GAMES)

Let $\Gamma = (A,B,M)$ be a concave-convex game, where $-\infty \leq M < \infty$. Let $b_1,\ldots,b_m \in B$ and assume that $M_{II}(b_i) = \infty$ implies $M(a,b_i) > -\infty$ for all $a \in A$. (If $\overline{V} < \infty$ or M is finite, this is no restriction.) Choose $\tau < \overline{V}$. Then

$$\bigcap_{i=1}^{m} [M \geq \tau]_{b_i} \neq \emptyset .$$

Proof: Define $S = \{(M(a,b_1),\ldots,M(a,b_m)) : a \in A\}$.

Then $S \subseteq [-\infty,\infty[^m$. Put $H = [\tau,\infty[^m$.

Suppose $\bigcap\limits_{i=1}^{m} [M \geq \tau]_{b_i} = \emptyset$. Then $S \cap H = \emptyset$.

Let $T = \{y : y \in [-\infty,\infty[^m$ and $y \leq x$ for some $x \in S\}$

($y \leq x$ is defined componentwise).

Clearly $S \subseteq T \subseteq [-\infty,\infty[^m$.

We state two lemmas:

Lemma 1:

Let $y^1,\ldots,y^n \in T$, $\xi_1,\ldots,\xi_n \geq 0$, $\Sigma\, \xi_i = 1$.

Then $\sum\limits_{i=1}^{n} \xi_i y^i \in T$.

Proof: Let $y^i \leq x^i$ for $x^i \in S$, $i = 1,\ldots,n$. Then $\Sigma\, \xi_i y^i \leq \Sigma\, \xi_i x^i$. We may write $x_j{}^i = M(a_i, b_j)$ so that $\Sigma\, \xi_i x_j{}^i = \Sigma\, \xi_i M(a_i, b_j)$. Since A is concave, it now follows from prop. 2.13 that there exists $a \in A$ such that

$\Sigma\, \xi_i M(a_i, b_j) \leq M(a, b_j) = z_j$ for all j .

Hence $\Sigma\, \xi_i x^i \leq z \in S$ and the lemma follows.

Lemma 2:

Let $T' = T \cap R^m$. Then T' is a convex non-empty subset of R^m . Moreover, $T' \cap H = \emptyset$.

Proof: Assume $T \cap R^m = \emptyset$, i.e. each vector in S has at least one component which is $-\infty$. Then

$\sum\limits_{i=1}^{m} M(a, b_i) = -\infty$ for all $a \in A$. Since B is convex (relatively Γ) there is $\tilde{b} \in B$ such that

$M(a, \tilde{b}) \leq \frac{1}{m} \sum\limits_{i=1}^{m} M(a, b_i) = -\infty$ for all a .

Hence $M_{II}(\tilde{b}) = -\infty$, which implies $\overline{V} = -\infty$.

This contradicts the fact that $\tau < \overline{V}$. Thus $T' \neq \emptyset$.

T is convex because of lemma 1, so T' is convex. The last part of the lemma follows from the definition of T and the fact that $S \cap H = \emptyset$.

We now proceed with the proof of the theorem. Since H and T' by lemma 2 are disjoint convex subsets of R^m , it follows from theorem 1.30 that there are numbers $l_1,\ldots,l_m$, not all equal to 0 , so that $\Sigma l_i x_i \geq \Sigma l_i y_i$ whenever $x \in H$, $y \in T'$.

If we fix $y \in T'$ and let some $x_i \to \infty$, we observe that necessarily $l_i \geq 0$ for $i = 1,\ldots,m$.

We may assume $\Sigma l_i = 1$.

Since $(\tau,\dots,\tau) \in H$ we have $\Sigma l_i y_i \leq \tau$ for any $y \in T'$. Assume now $\Sigma l_i y_i \leq \tau$ for all $y \in S$. Then there exists b such that

$$M(a,b) \leq \Sigma l_i M(a,b_i) \leq \tau \quad \text{for all} \quad a \in A ,$$

which implies $M_{II}(b) \leq \tau$ and hence $\overline{V} \leq \tau$, which is a contradiction.

We may therefore find $\tilde{y} \in S-T'$ such that $\Sigma l_i \tilde{y}_i > \tau$. Since $\tilde{y} \in S-T'$, $\tilde{y}_i = -\infty$ for some i. Put $I = \{i : \tilde{y}_i > -\infty\}$.

We observe that, since $i \notin I$ (i.e. $\tilde{y}_i = -\infty$) implies $l_i = 0$, $I \neq \emptyset$, and $l_i \neq 0$ for at least one i.

Let $0 < p < 1$ and put $p_i = l_i p$ for all $i \in I$.

Obviously $\sum\limits_{i \in I} p_i = p$.

Define $p_i = \frac{1-p}{m-\#(I)}$, $i \notin I$. Then $\sum\limits_{i \notin I} p_i = 1-p$ and hence $\sum\limits_{i=1}^{m} p_i = 1$, $p_i \geq 0$.

By the convexity of B, there is $b_p \in B$ such that for all $a \in A$: $M(a,b_p) \leq \sum\limits_{i=1}^{m} p_i M(a,b_i) = p \sum\limits_{i \in I} l_i M(a,b_i)$

$$+ \frac{1-p}{m-\#(I)} \sum_{i \notin I} M(a,b_i) .$$

If $\sum\limits_i l_i M(a,b_i) \leq \tau$, this yields

$$M(a,b_p) \leq p\tau + \frac{1-p}{m-\#(I)} \sum_{i \notin I} M(a,b_i) \leq p\tau + \frac{1-p}{m-\#(I)} \sum_{i \notin I} M_{II}(b_i) .$$

If $\Sigma l_i M(a,b_i) > \tau$, then there is some i such that $l_i = 0$ and $M(a,b_i) = -\infty$, i.e. $i \notin I$. Hence $M(a,b_p) = -\infty$.

It follows that for all $a \in A$,

$$M(a,b_p) \leq p\tau + \frac{1-p}{m-\#(I)} \sum_{i \notin I} M_{II}(b_i) .$$

Hence

$$(5) \quad \tau < \overline{V} \leq M_{II}(b_p) \leq p\tau + \frac{1-p}{m-\#(I)} \sum_{i \notin I} M_{II}(b_i)$$

Let $i \notin I$.

Then $\tilde{y}_i = -\infty$. If we write $\tilde{y} = (M(\tilde{a},b_1),\ldots,M(\tilde{a},b_m))$ then $M(\tilde{a},b_i) = -\infty$. Hence, by the assumptions in the theorem, $M_{II}(b_i) < \infty$. Therefore, by letting $p \to 1$ in (5), we get $\tau < \overline{V} \leq \tau$ which is a contradiction.

THEOREM 2.28

Let $\Gamma = (A,B,M)$ be a concave-convex game such that $-\infty \leq M < \infty$ and such that M is finite or $\overline{V} < \infty$. Assume there is a sequence $\{b_n\}$ in B such that

$$(6) \quad \inf_b M(a,b) = \inf_{i=1,2,\ldots} M(a,b_i) \quad \text{for all } a \in A .$$

Assume further that to each sequence $\{a_n\}$ in A there exists $a \in A$ such that

$\liminf_n M(a_n,b) \leq M(a,b)$ for all $b \in B$.

Then Γ has a value.

Proof: By theorem 2.27 we have, for $\tau < \overline{V}$,

$$\bigcap_{i=1}^{m} [M \geq \tau]_{b_i} \neq \emptyset \qquad m = 1,2,\ldots$$

Choose $a_m \in \bigcap_{i=1}^{m} [M \geq \tau]_{b_i}$, $m = 1,2,\ldots$.

By the assumption, there is $a \in A$ such that $\liminf_m M(a_m,b) \leq M(a,b)$ for all $b \in B$.

The proof is complete if we can show that $M(a,b) \geq \tau$ for all $b \in B$. By (6) it suffices to show that $M(a,b_i) \geq \tau$ for $i = 1,2,\ldots$. For $m \geq i$ we have $M(a_m,b_i) \geq \tau$. It follows that $M(a,b_i) \geq \liminf_m M(a_m,b_i) \geq \tau$.

COROLLARY 2.29

Let $\Gamma = (A,B,M)$ be a game satisfying the condition in the remark succeeding def. 2.16. Assume further that M is continuous in each variable and that A is closed and bounded. Then Γ has a value.

Proof: From the theory of metric spaces we know that B has a countable dense subset (which may here be taken as the set of points in B with rational coordinates). By the continuity of $M(a,\circ)$ for each fixed $a \in A$, statement (6) of the theorem is now satisfied.

Let $\{a_n\}$ be a sequence in A. Since A is closed and bounded, $\{a_n\}$ has a subsequence $\{a_{n_j}\}$ which converges to a point $a \in A$. Now, since $M(\circ,b)$ is continuous, $\liminf_n M(a_n,b) \leq \lim M(a_{n_j},b) = M(a,b)$ for each $b \in B$. Thus the assumptions of theorem 2.28 are satisfied, so that Γ has a value.

To find weaker conditions under which a game has a value, we will use general topology. An introduction to the theory of topological spaces, compactness etc. may be found in Royden: Real Analysis.

DEFINITION 2.30

Let X be a topological space. A collection $\mathfrak{F}$ of sets in X is said to have the finite intersection property (f.i.p) if any finite subcollection of $\mathfrak{F}$ has a nonempty intersection.

Remark:

Theorem 2.27 states that the family of subsets of A, $[M \geq \tau]_b$, $b \in B$, has the f.i.p.
We have the following theorem concerning f.i.p.

THEOREM 2.31

A topological space X is compact if and only if every collection of closed sets with the f.i.p. has a nonempty intersection.

Proof: Royden: Ch. 9 prop. 1.

DEFINITION 2.32

An extended real-valued function f on a topological space is called upper semicontinuous if $-\infty \leq f < \infty$, and for each $\alpha \in R$, the set $\{x : f(x) < \alpha\}$ is open.
We remark that the last condition is equivalent to $\{x : f(x) \geq \alpha\}$ is a closed set for each $\alpha \in R$.

LEMMA 2.33

Let $\{f_i\}$, $i \in I$ be a family of upper semicontinuous functions defined on a topological space. Then $f = \inf_{i \in I} f_i$ is upper semicontinuous.

Proof: Clearly $-\infty \leq f < \infty$. Moreover $\{x : f(x) < \alpha\} = \bigcup_{i \in I} \{x : f_i(x) < \alpha\}$ is a union of open sets and thus open.

LEMMA 2.34

Let f be an upper semicontinuous function defined on a compact topological space X. Then f assumes its maximum; i.e. there exists $x_0 \in X$ such that $\sup_{x \in X} f(x) = f(x_0)$

Proof: Royden Ch. 9 prop. 10.

THEOREM 2.35

Let $\Gamma = (A,B,M)$ be a concave-convex game; $-\infty \leq M < \infty$. Assume there is a topology on A such that A is compact and $M(\cdot,b)$ is upper semicontinuous for each $b \in B$. Then the game Γ has a value and player I has a maximin strategy.

Proof: Let $\tau < \overline{V}$. By def. 2.32, the sets $[M \geq \tau]_b$ are closed sets in A and hence compact, since A is compact. By theorem 2.27, the family $[M \geq \tau]_b$, $b \in B$, has f.i.p. and so by theorem 2.31 $\bigcap_b [M \geq \tau]_b \neq \emptyset$. Theorem 2.26 now shows that Γ has a value.

Since $M_I(a) = \inf_b M(a,b)$, it follows from lemma 2.33 that M_I is upper semicontinuous, and by lemma 2.34 we may find $a_0 \in A$ such that $M_I(a_0) = \sup_a M_I(a) = \underline{V}$. Hence a_0 is a maximin strategy.

Assume $\Gamma = (A,B,M)$ is a concave-convex game, such that $-\infty \leq M < \infty$. We will now find conditions on A and M to assure the existence of a topology possessing the properties listed in theorem 2.35.

We will need the concept of a <u>net</u>, which is a generalization of a sequence. The definition of and some properties of nets may be found in Royden Ch. 8.

It suffices to consider the coarsest topology on A for which the functions $M(\cdot,b)$, $b \in B$ are upper semicontinuous on A. This is the topology T on A generated by the sets $O_{b,t} = \{a \in A : M(a,b) < t\}$, $b \in B$, $t \in R$. (Sets of this form will have to be open by definition 2.32.)

<u>LEMMA 2.36</u>

A net $\{a_\alpha\}$ in A converges to a point $a \in A$ in the topology T if and only if

(7) $\limsup_\alpha M(a_\alpha,b) \leq M(a,b)$ for all b.

<u>Proof</u>: "only if": Assume $O_{b,t}$ is a neighbourhood of a. Then $M(a,b) < t$. There is α_o such that $\alpha > \alpha_o \Rightarrow a_\alpha \in O_{b,t}$ or equivalently $\alpha > \alpha_o \Rightarrow M(a_\alpha,b) < t$. It follows that $\limsup_\alpha M(a_\alpha,b) < t$ for all t and b such that $M(a,b) < t$. Consequently $\limsup_\alpha M(a_\alpha,b) \leq M(a,b)$ for all b.

"if": The collection of finite intersections of $O_{b,t}$-sets is a base for T. Consider such a set $O_{b_1,t_1} \cap \ldots \cap O_{b_r,t_r}$ which contains the point a.

Then $M(a,b_i) < t_i \quad i = 1,\ldots,r$.

By (7), for each i we have,

$$\limsup_\alpha M(a_\alpha,b_i) \leq M(a,b_i) < t_i$$

It then follows (from the definition of limsup) that there is α_o such that $\alpha > \alpha_o \Rightarrow M(a_\alpha,b_i) < t_i$ for each i, i.e. $\alpha > \alpha_o \Rightarrow a_\alpha \in O_{b_1,t_1} \cap \ldots \cap O_{b_r,t_r}$.

PROPOSITION 2.37

A is compact in the topology T if and only if to each net $\{a_\alpha\}$ in A there exists $a \in A$ such that

$$\liminf_\alpha M(a_\alpha,b) \leq M(a,b) \quad \text{for all} \quad b \in B .$$

Proof: "only if": Assume A is compact. Then there is a subnet $\{a_\beta\}$ of $\{a_\alpha\}$ which converges to a point $a \in A$. i.e. by lemma 2.36, $\limsup_\beta M(a_\beta,b) \leq M(a,b)$ for all b. It follows that

$$\liminf_\alpha M(a_\alpha,b) \leq \liminf_\beta M(a_\beta,b) \leq \limsup_\beta M(a_\beta,b) \leq M(a,b) .$$

The first inequality holds since $\{a_\beta\}$ is a subnet of $\{a_\alpha\}$.

"if": Let $\{a_\alpha\}$ be a net in A. The set $\overline{R} = [-\infty,\infty]$ is a compact set. Consequently $\overline{R}^B$ (the set of functions from B into $\overline{R}$) is compact by Tychonoff's theorem (Royden Ch. 9 Th. 19). For all $a \in A$, $M(a,\circ) \in \overline{R}^B$, so $M(a_\alpha,\circ)$ defines a net in $\overline{R}^B$. By the compactness of $\overline{R}^B$ there is a subnet $\{a_\beta\}$ such that $M(a_\beta,\circ)$ converges in $\overline{R}^B$. Hence $\lim_\beta M(a_\beta,b)$ exists for all $b \in B$. By the assumptions in the theorem, there exists $a \in A$ such that

$$\liminf_\beta M(a_\beta,b) \leq M(a,b) \quad \text{for all} \quad b .$$

But then

$$\limsup_\beta M(a_\beta,b) = \liminf_\beta M(a_\beta,b) \leq M(a,b)$$

i.e. a_β converges to the point a in the topology T (lemma 2.36) . Hence A is compact, since each net in A contains a convergent subnet.

COROLLARY 2.38

Let $\Gamma = (A,B,M)$ be a concave-convex game with $-\infty \leq M < \infty$. Assume that to each net $\{a_\alpha\}$ in A there exists $a \in A$ such that

$$\liminf_\alpha M(a_\alpha,b) \leq M(a,b) \quad \text{for all} \quad b \in B \text{ .}$$

Then Γ has a value and player I has a maximin strategy.

Remark: The statement that Γ has a value and player I has a maximin strategy $a_o \in A$, is equivalent to the following:

$$\inf_b \sup_a M(a,b) = \sup_a \inf_b M(a,b) = \inf_b M(a_o,b)$$

EXAMPLE 2.39

Let $\Gamma = (A,B,M)$ be a game such that $A = \{a_1,\ldots,a_m\}$, $-\infty < M < \infty$. We will prove that the randomized game Γ^* has a value. (in example 2.25 we showed this for a simple 2×2 game.) An element a^* of A^* may be identified with an ordered m-tuple $(\theta_1,\ldots,\theta_m)$ where each $\theta_i \geq 0$, $\Sigma\, \theta_i = 1$, such that $a^*(a_i) = \theta_i$, $i = 1,\ldots,m$.
If we induce the standard topology on R^m on A^* , then A^* is compact (closed and bounded). Equation (4) in 2.19 shows that for each fixed $b^* \in B^*$, $M^*(a^*,b^*)$ is a linear combination

of $\theta_1,\ldots,\theta_m$. Hence $M^*(\cdot,b^*)$ is a continuous function on A^* for each fixed b^*. Since Γ^* is concave-convex (prop. 2.20), theorem 2.35 gives the desired result.

DEFINITION 2.40

Let $a_o \in A$, $b \in B$. We say that the strategy a_o is optimal for b if $M(a_o,b) \geq M(a,b)$ for all $a \in A$.
Let $\epsilon > 0$. We say that a_o is ϵ-optimal for b if $M(a_o,b) \geq M(a,b)-\epsilon$ for a l $a \in A$.

LEMMA 2.41

Let $\Gamma = (A,B,M)$ be a game with value V, $-\infty < V < \infty$. Assume that a_o is a maximin strategy for player I, i.e. $M_I(a_o) = V$. Then for each $\epsilon > 0$ there exists a strategy $b_\epsilon \in B$ such that a_o is ϵ-optimal for b_ϵ.

Proof: $V = \inf\limits_b M_{II}(b)$. Let $\epsilon > 0$. Then we can find b_ϵ such that

$$M_{II}(b_\epsilon)-\epsilon \leq V .$$

Now $M(a_o,b_\epsilon) \geq M_I(a_o) = V$ which implies
$M(a_o,b_\epsilon) \geq M_{II}(b_\epsilon)-\epsilon \geq M(a,b_\epsilon)-\epsilon$ for any $a \in A$.

DEFINITION 2.42

Denote by $\tilde{A}$ the set of strategies for player I that are ϵ-optimal for a strategy b_ϵ in B for all $\epsilon > 0$.

THEOREM 2.43

Let $\Gamma = (A,B,M)$ be a concave-convex game with finite pay-off function M. Assume (i): There is a topology on A in which A is a compact space and the functions $M(\circ,b)$ on A are upper semicontinuous for each $b \in B$.

(ii): B is affine relatively Γ.

The condition (i) is, as is proved earlier, equivalent to

(i)': given any net $\{a_\alpha\}$ in A there exists $a \in A$ such that

$$\liminf_\alpha M(a_\alpha,b) \leq M(a,b) \quad \text{for any} \quad b \in B .$$

Then $\tilde{A}$ dominates A (see def. 2.3).

<u>Proof</u>: Let $\hat{a} \in A$. We define the game

$$\hat{\Gamma} = (A,B,\hat{M}) \quad \text{so that}$$

$$\hat{M}(a,b) = M(a,b)-M(\hat{a},b) \qquad a \in A , \; b \in B .$$

A straightforward computation, using the definition of concavity and affinity, shows that A is concave and B is affine relatively $\hat{\Gamma}$. Thus the game $\hat{\Gamma}$ is concave-convex. By theorem 2.34 the game $\hat{\Gamma}$ has a value $\hat{V}$ and player I has a maximin strategy $\tilde{a} \in A$. Since $\hat{M}_I(\hat{a}) = 0$, we have $\hat{V} \geq 0$. Furthermore,

$\hat{V} = \hat{M}_I(\tilde{a}) = \inf\limits_b \hat{M}(\tilde{a},b) < \infty$ since M is finite. Hence

$0 \leq \hat{V} < \infty$.

We assert that $\tilde{a}$ dominates $\hat{a}$ in the game Γ:

Since $\hat{V} = \hat{M}_I(\tilde{a}) \geq 0$, it follows that $\hat{M}(\tilde{a},b) \geq 0$ for any $b \in B$, and consequently $M(\tilde{a},b) \geq M(\hat{a},b)$ for any $b \in B$ by the definition of $\hat{M}$.

The theorem is proved if we can show that $\tilde{a} \in \tilde{A}$.

Given $\epsilon > 0$, there exists by lemma 2.40 $b_\epsilon \in B$ such that

$$\hat{M}(\tilde{a},b_\epsilon) \geq \hat{M}(a,b_\epsilon)-\epsilon \quad \text{for any} \quad a \in A .$$

Since now M is finite, we can add $M(\hat{a},b_\epsilon)$ to each side of the inequality and get $M(\tilde{a},b_\epsilon) \geq M(a,b_\epsilon)-\epsilon$ for any $a \in A$, proving that $\tilde{a} \in \tilde{A}$.

DEFINITION 2.44

Let $a_o \in A$. a_o is said to be admissible if

$$M(a,b) \geq M(a_o,b) \quad \text{for all} \quad b \in B$$

implies $M(a,b) = M(a_o,b)$ for all $b \in B$.

COROLLARY 2.45

Let the assumptions be as in theorem 2.42. If a_o is admissible, then $a_o \in \tilde{A}$.

Proof: By theorem 2.42, there is $\tilde{a} \in \tilde{A}$ such that for any $b \in B$, $M(\tilde{a},b) \geq M(a_o,b)$. This implies that $M(\tilde{a},b) = M(a_o,b)$ for any $b \in B$, and then clearly $a_o \in \tilde{A}$.

The following example shows that the condition in theorem 2.42 that Γ be concave-convex cannot be ommitted.

EXAMPLE 2.46

Let $\Gamma = (A,B,M)$ be given by

$A = [-1,1]$, $B = \{-1,1\}$, $M(a,b) = ab$.

That B is not convex (and hence not affine) relatively Γ follows easily from definition 2.14. We will show that $\tilde{A}$ (as defined in 2.42) is the set $\{-1,1\}$.

Let $\epsilon > 0$. A strategy $a_o \in A$ is then ϵ-optimal for $b \in B$ if

$$a_o b \geq ab - \epsilon \quad \text{for all} \quad a \in A .$$

If $b = 1$, then this is equivalent to $a_o \geq 1-\epsilon$.
If $b = -1$, we have $a_o \leq -1+\epsilon$.
Consequently, the only strategies in A that are ϵ-optimal for some $b \in B$ for any $\epsilon > 0$ are $a_o = 1$ and $a_o = -1$. These strategies are, in fact, by the definition optimal for respectively $b = 1$ and $b = -1$. Now $\tilde{A} = \{-1,1\}$. We will show that $\tilde{A}$ does not dominate A . Let $a = 0$. A strategy a_o dominates $a = 0$ if

$$a_o b \geq 0 \quad \text{for all} \quad b \in B \text{ , i.e. if}$$

$a_o \geq 0$ and $- a_o \geq 0$ which implies $a_o = 0$. Hence no strategy in $\tilde{A}$ dominates $a = 0$.

3. BASIC ELEMENTS OF DECISION THEORY

DEFINITION 3.1

An experiment $\mathcal{E}$ is given by $\mathcal{E} = (\chi, \mathcal{A}; P_\theta : \theta \in \Theta)$ where $(\chi, \mathcal{A})$ is a measurable space and $\{P_\theta : \theta \in \Theta\}$ is a family of probability measures on $(\chi, \mathcal{A})$. $(\chi, \mathcal{A})$ is the sample space and Θ is the parameter set.

Remark: $\mathcal{E}$ may also be called a statistical model.

A statistician begins his study of a phenomenon by building up a mathematical model which is believed to "explain" what happens. This model is given in the form of an experiment as defined above.

The next step is to perform an experiment and to make certain decisions on the basis of the observations. A decision will in general be a statement about the "true" parameter Θ. In the theory of tests, we consider a null-hypothesis of the form $H : \theta \in \Theta_0$ where Θ_0 is a subset of Θ. Two decisions are possible: to reject or to accept the hypothesis (eventually to reject or to "say nothing"). If a real function $g(\theta)$ of the "true" parameter is to be estimated, then the set of possible decisions will be a subset of the real line.

In addition to the set of decisions, the statistician will need a rule which to each observed result tells him which decision to make. Such a rule will be called a decision-rule.

DEFINITION 3.2

A decision space is a measurable space $(T, \mathcal{S})$. The elements of T are called decisions. In the case of finite T, $\mathcal{S}$

is generally taken as the family of all subsets of T.

DEFINITION 3.3

Let $\mathcal{E} = (\chi, \mathcal{A}, P_\theta : \theta \in \Theta)$ be an experiment. A decision-rule δ is a Markov-kernel (see def. 12 of appendix B) $\delta(\circ|\circ) : \mathcal{S} \times \chi \to R$

Remark: A decision-rule defines for each $x \in \chi$ a probability distribution on the set of decisions $(T, \mathcal{S})$. The statistician chooses a decision according to this distribution. Thus we have a randomized decision-rule. However, it may in many situations seem more satisfactorily to have a decision-rule which to each observed $x \in \chi$ defines exactly which decision to make. Such a rule is called a non-randomized decision-rule. We have the following definition:

DEFINITION 3.4

A decision-rule is said to be non-randomized if there is a function $\psi : \chi \to T$ such that
$\delta(S|x) = I_S(\psi(x))$ for all $S \in \mathcal{S}, x \in \chi$

i.e. $\delta(S|x) = \begin{cases} 1 \text{ if } & \psi(x) \in S \\ 0 \text{ if } & \psi(x) \notin S \end{cases}$

Accordingly, $\delta(\circ|x)$ is the probability distribution giving mass 1 to the set $\{\psi(x)\} \subset T$, provided $\{\psi(x)\} \in \mathcal{S}$.
We remark that $\mathcal{S}$ in most cases is chosen so that each one point set of T is measurable.
Since $\delta(S|x) = I_{\psi^{-1}(S)}(x)$; we have $\psi^{-1}(S) \in \mathcal{A}$ whenever $S \in \mathcal{S}$. (This follows since $\delta(S|\circ)$ is required to be $\mathcal{A}$-measurable). Hence ψ is always measurable. $\psi(x)$ may be interpreted as "the decision to take when x is observed".

EXAMPLE 3.5

In a k-decision problem, $(T, \mathcal{S})$ is given by $T = \{1, 2, \ldots, k\}$, $\mathcal{S}$ is the family of all subsets of T. A decision rule δ is thus completely defined by the values $\delta(\{t\}|x)$, $t \in T$, $x \in \chi$. We will in common write $\delta(t|x)$ instead of $\delta(\{t\}|x)$. Note that $\sum_{t=1}^{k} \delta(t|x) = 1$ for all $x \in \chi$. If δ is non-randomized, then for each $x \in \chi$, $\delta(t|x) = 1$ for some $t \in T$.

EXAMPLE 3.6

(a) Let $\Theta_0 \subset \Theta$ and suppose we want to test the hypothesis $H : \theta \in \Theta_0$ against $\theta \in \Theta - \Theta_0$. Our decision space will then consist of two elements: "accept H" and "reject H", which we may identify with the numbers 0 and 1, respectively.

Thus $T = \{0,1\}$ and our procedure is determined by

$\delta(1|x) = \Pr$ (reject | x is observed) $= \varphi(x)$

$\delta(0|x) = 1 - \varphi(x)$

The test δ is non-randomized if δ takes only the values 0 and 1. The set $W = \{x \in \chi : \varphi(x) = 1\}$ is then the rejection region and $U = \{x \in \chi : \varphi(x) = 0\}$ is the acceptance region of our test.

(b) Suppose now $\Theta \subset R$ and we want to estimate the "true" parameter $\theta \in \Theta$. Now T is chosen as a subset of R and $\mathcal{S}$ may be taken as the family of Borel-sets in T. The decision procedure $\delta(S|x)$ determines to each observed x a probability distribution over T. A non-randomized decision rule is now seen to be given by a measurable function $\psi : \chi \to R$, which is the same as what we are accustomed to call an estimator.

REMARK 3.7

The decision-rule in definition 3.3 is a randomization "after x", in the sense that the statistician first observes x and then chooses a probability distribution over T. Another way of randomizing is the following: Let D_o be the set of non-randomized decision-rules and let $D_o{}^*$ be a set of probability measures over D_o.

We may let the strategies of the statistician be the elements of $D_o{}^*$, i.e. the statistician chooses a non-randomized decision-rule according to a probability distribution over D_o. Thus we have a randomization "before x".

Assume, for example, that χ is finite with M elements, T is finite with N elements. Let D be the set of randomized decision rules. Then

$$\dim D = M(N-1)$$

$$\# D_o = N^M$$

$$\dim D_o{}^* = N^M-1$$

Comparing dim D and dim $D_o{}^*$, we observe that randomizing "before x" gives rise to more strategies than randomizing "after x". In quite general situations, however, the two ways of randomizing are equivalent. See [20].

DEFINITION 3.8

Let δ be a decision-rule. The operational characteristic (abbreviated O.C.) of δ is the function

$$OC_\delta : \mathcal{S} \times \Theta \to R \quad \text{given by}$$

$$OC_\delta(S|\theta) = \int \delta(S|x) P_\theta(dx) = P_\theta \delta(S)$$

(see appendix B, def. 17).

For each $\theta \in \Theta$, $P_\theta\delta$ is a probability distribution on $(T,\mathcal{S})$, which may be interpreted as "the expected decision when the true parameter is θ".

EXAMPLE 3.9

Consider example 3.6 (a). We have $OC_\delta(1|\theta) = \int\delta(1|x)P_\theta(dx) = \int\varphi(x)P_\theta(dx)$ which is seen to be the power function of the test δ. If δ is non-randomized, then $OC_\delta(1|\theta) = P_\theta(W)$, where W is the rejection region. In the non-randomized situation of example 3.6 (b), we get $OC_\delta(S|\theta) = P_\theta(\psi \in S)$, which is the probability distribution of ψ.

We make the following interesting observation: The power function of a test "corresponds to" the probability distribution of an estimator. Thus the power of a test is a more "fundamental" concept than the variance of an estimator.

Decision theory may be considered as a two-person game, "nature" being player I and the statistician being player II. Their strategies are, respectively, the parameter set Θ and the set of decision-rules. The pay-off function of this game is called the risk-function and is constructed from a loss-function. These concepts are defined in the following:

DEFINITION 3.10

Let $\mathcal{E} = (\chi,\mathcal{A},P_\theta : \theta \in \Theta)$ be an experiment and $(T,\mathcal{S})$ a decision space. A loss-function is a real function

$$L_\theta(t)\,;\ \theta \in \Theta\,,\quad t \in T$$

which is measurable in t for each $\theta \in \Theta$.

The loss-function $L_\theta(t)$ may be interpreted to define the loss (or "penalty") by taking decision t when θ is the "true" parameter. Thus, if t is precisely the right decision to make when θ is the parameter, then it may seem reascnable to let $L_\theta(t) = 0$.

<u>DEFINITION 3.11</u>

Let $\mathcal{E}$ and $(T, \mathcal{S})$ be given as in definition 9, and let $L_\theta(t)$; $\theta \in \Theta$, $t \in T$ be a <u>loss-function</u>. Let δ be a decision-rule. The <u>risk-function</u> r_δ of δ is given by

$$r_\delta(\theta) = \int L_\theta(t) OC_\delta(dt|\theta) \quad , \quad \theta \in \Theta$$

provided the integral exists.

By appendix B, we may write

$$r_\delta(\theta) = P_\theta \delta L_\theta$$

<u>EXAMPLE 3.12</u>

Consider again example 3.6 (a). A loss-function may be given as follows:

$$L_0(\theta) = \begin{cases} 0 & \text{when } \theta \in \Theta_0 \\ a & \text{when } \theta \in \Theta - \Theta_0 \end{cases}$$

$$L_1(\theta) = \begin{cases} b & \text{when } \theta \in \Theta_0 \\ 0 & \text{when } \theta \in \Theta - \Theta_0 \end{cases}$$

Now, for $\theta \in \Theta_0$:

$$r_\delta(\theta) = b OC_\delta(1|\theta) = b \int \varphi dP_\theta$$

For $\theta \in \Theta - \Theta_0$,

$$r_\delta(\theta) = a OC_\delta(0|\theta) = a(1 - \int \varphi dP_\theta)$$

A frequently used loss-function in the situation of example 3.6 (b), is

$$L_\theta(t) = C(t-\theta)^2 \text{ where } C \text{ is a constant.}$$

Consider now the two-person zero-sum game (Θ,D,r), where Θ is the parameter set, D is the set of randomized decision rules and r is a risk-function. The statistician (player II), trying to find the "best" decision-rule, of course wishes to minimize the risk. We shall mention here two useful principles.

DEFINITION 3.13 (THE MINIMAX PRINCIPLE)

A decision rule $\delta_o \in D$ is said to be minimax if

$$\sup_\theta r_{\delta_o}(\theta) = \inf_\delta \sup_\theta r_\delta(\theta)$$

For further reflexions about minimax rules, we refer to chapter 2.

DEFINITION 3.14 (THE BAYES PRINCIPLE)

The Bayes principle involves the notion of a distribution on the parameter space Θ called a prior distribution.
By the Bayes risk of a decision rule $\delta \in D$ with respect to a prior distribution Λ we shall mean the quantity

$$r(\Lambda,\delta) = Er_\delta(T) \quad \text{(provided the expectation exists)}$$

where T is a random variable over Θ distributed according to Λ. A decision rule $\delta_o \in D$ is said to be Bayes w.r.t. the prior distribution Λ if $r(\Lambda,\delta_o) = \inf_\delta r(\Lambda,\delta)$
The quantity $\inf_\delta r(\Lambda,\delta)$ is called minimum Bayes risk relative to the prior distribution Λ.

A rigorous treatment of statistical decision theory is found in Ferguson [4]. Some special topics are treated in [15].

4. DEFICIENCIES

In this chapter we will give the basic definitions of ε-deficiency between two experiments with the same parameter set Θ .

DEFINITION 4.1

Let $\mathcal{E} = (\chi, \mathcal{A}, P_\theta : \theta \in \Theta)$ and

$\mathcal{F} = (\mathcal{Y}, \mathcal{B}, Q_\theta : \theta \in \Theta)$ be experiments (see def. 3.1) with the same parameter set Θ .

Let ε be a real function defined on Θ with values $\varepsilon_\theta \geq 0$ for all $\theta \in \Theta$.

We shall say that $\mathcal{E}$ is ε-deficient relative to $\mathcal{F}$ for k-decision problems if to each decision space $(T, \mathcal{S})$ where $\mathcal{S}$ contains 2^k sets and to each bounded loss function $\{L_\theta(t) : \theta \in \Theta , \ t \in T\}$ and to each decision-rule σ in $\mathcal{F}$ (relative to $(T, \mathcal{S})$) there exists a decision-rule ρ in $\mathcal{E}$ (relative to $(T, \mathcal{S})$) such that

(1) $P_\theta \rho L_\theta \leq Q_\theta \sigma L_\theta + \varepsilon_\theta \|L_\theta\|$ for all $\theta \in \Theta$ where

$$\|L_\theta\| = \max_t |L_\theta(t)|$$

Remark: A 2-decision problem will in the sequel be called a testing problem. We remark that in the definition above, T need not be a finite set. A k-decision problem arises e.g. when T contains k elements and $\mathcal{S}$ is the set of all subsets of T , or when $\mathcal{S}$ is generated by a finite partition of T in k parts. In the latter case, T itself may be infinite. It is seen that a finite σ-algebra always contains 2^k sets, for some natural number k .

L_θ is required to be $\mathcal{A}$ -measurable for each $\theta \in \Theta$. Thus for fixed θ , $L_\theta(t)$ may take only a finite number of values (at most k).

(1) may be replaced by

(2) $P_\theta \rho L_\theta \leq Q_\theta \sigma L_\theta + \epsilon_\theta \|L\|$ for all $\theta \in \Theta$, where

$$\|L\| = \max_\theta \|L_\theta\| = \max_{\theta,t} |L_\theta(t)|$$

Clearly (1) implies (2). Assume then that (2) holds and let L be a loss-function. Then L' defined by $L'_\theta(t) = L_\theta(t)/\|L_\theta\|$ is also a loss-function and $\|L'\| = 1$. Substituting L' into (2) yields (1).

PROPOSITION 4.2

In definition 4.1, we may restrict ourselves to consider only decision spaces $(T,\mathcal{A})$ with $T = \{1,2,\ldots,k\}$ and $\mathcal{A}$ consisting of all subsets of T .

Proof: Suppose the conditions in def. 4.1 are satisfied, but only for decision spaces of the above type.

Let $(T',\mathcal{A}')$ be an arbitrary decision space with $\#\mathcal{A}' = 2^k$. Then $\mathcal{A}'$ is generated by a finite partition of T' containing k sets, say $\Gamma = \{T_1,\ldots,T_k\}$ (See e.g. [10] Prop I.2.1).

The idea of the proof is now to identify the element $i \in T$ with the i-th component T_i of the partition of $T', i = 1,\ldots,k$. The proposition will then follow from the fact that every $\mathcal{A}'$-measurable function on T' is constant on each T_i . The reader is recommended to work out the details of the proof.

DEFINITION 4.3

Let $\mathcal{E}$, $\mathfrak{F}$ and ϵ be as in def. 4.1. We shall say that $\mathcal{E}$ is ϵ-deficient relative to $\mathfrak{F}$ if $\mathcal{E}$ is ϵ-deficient relative to $\mathfrak{F}$ for k-decision problems for $k = 1,2,3,\ldots..$

Remark: The concept of ϵ-deficiency of one experiment to another was introduced by Le Cam in [7]. This generalized the concept of "being more informative" which was introduced by Bohenblust, Shapley and Sherman and may be found in Blackwell [1]. "Being more informative for k-decision problems" was introduced by Blackwell [2]. In terms of ϵ-deficiency these concepts are defined as follows:

DEFINITION 4.4

We shall say that $\mathcal{E}$ is more informative than $\mathfrak{F}$ (for k-decision problems) and write this $\mathcal{E} \geq \mathfrak{F}$ ($\mathcal{E} \underset{k}{\geq} \mathfrak{F}$) if $\mathcal{E}$ is 0-deficient relative to $\mathfrak{F}$ (for k-decision problems).

Remark: "0-deficiency" means ϵ-deficiency when $\epsilon_\theta = 0$ for all $\theta \in \Theta$.

DEFINITION 4.5

If $\mathcal{E} \geq \mathfrak{F}$ ($\mathcal{E} \underset{k}{\geq} \mathfrak{F}$) and $\mathfrak{F} \geq \mathcal{E}$ ($\mathfrak{F} \underset{k}{\geq} \mathcal{E}$) , then we shall say that $\mathcal{E}$ and $\mathfrak{F}$ are equivalent experiments (with respect to k-decision problems) and write this $\mathcal{E} \sim \mathfrak{F}$ ($\mathcal{E} \underset{k}{\sim} \mathfrak{F}$) .

What is the intuitive interpretation of ϵ-deficiency? Assume that the statistician may observe values from one of the experiments $\mathcal{E}$ and $\mathfrak{F}$, but not from both. Which should he choose? If $\mathcal{E}$ is more informative than $\mathfrak{F}$, then regardless which decision rule he chooses in $\mathcal{F}$, there is a decision rule in $\mathcal{E}$ which yields lower (or equal) risk. Consequently the experiment $\mathcal{E}$ should be preferred.

If $\mathcal{E}$ is ε-deficient relative to $\mathcal{F}$ and ε_θ is small for each θ , then to each decision rule in $\mathcal{F}$ there may be found a decision rule in $\mathcal{E}$ which is almost as "good" as the first. Thus only a small amount of information will get lost if we observe $\mathcal{E}$ instead of $\mathcal{F}$.

The function ε may be called a tolerance function.

PROPOSITION 4.6

If $\mathcal{E}$ is ε-deficient relative to $\mathcal{F}$ for $(k+1)$-decision problems, then $\mathcal{E}$ is ε-deficient relative to $\mathcal{F}$ for k-decision problems.

Proof: Set $T_k = \{1,\dots,k\}$. Let $L_\theta(i), i = 1,\dots,k, \theta \in \Theta$ be a loss-function and let σ be a decision-rule in $\mathcal{F}$ relative to T_k . We shall construct a decision-rule ρ in $\mathcal{E}$ such that

(3) $P_\theta \rho L_\theta \leq Q_\theta \sigma L_\theta + \varepsilon_\theta \|L_\theta\| ; \theta \in \Theta$.

Set $L'_\theta(k+1) = L_\theta(k)$,

$L_\theta'(i) = L_\theta(i),\ i = 1,\dots,k;\ \theta \in \Theta$

σ is well-defined by the values $\sigma(i|y),\ i \in T_k,\ y \in \mathcal{Y}$.

By defining $\sigma(k+1|y) = 0$ for all $y \in \mathcal{Y}$, σ may be considered as a decision rule in $\mathcal{F}$ relative to T_{k+1} .

By assumption, there is a rule $\tilde{\rho}$ in $\mathcal{E}$ relative to T_{k+1} such that

(4) $P_\theta \tilde{\rho} L'_\theta \leq Q_\theta \sigma L_\theta' + \varepsilon_\theta \|L_\theta'\|$

Define $\rho(i|x) = \tilde{\rho}(i|x)$ if $i < k$

and $\rho(k|x) = \tilde{\rho}(k|x) + \tilde{\rho}(k+1|x)$ for all $x \in \chi$

Clearly $\sum_{i=1}^{k} \rho(i|x) = 1$, so ρ is a decision-rule in $\mathcal{E}$ relative to T_k .

It remains to prove (3). For each $\theta \in \Theta$,

$$P_\theta \tilde{\rho} L'_\theta = \int \sum_{i=1}^{k+1} L'_\theta(i)\tilde{\rho}(i|x)P_\theta(dx) = \int \sum_{i=1}^{k} L_\theta(i)\rho(i|x)P_\theta(dx) = P_\theta \rho L_\theta$$

and $Q_\theta \sigma L_\theta{}' = Q_\theta \sigma L_\theta$ since $\sigma(k+1|y) = 0$ for all $y \in \mathcal{Y}$.

Finally $\|L_\theta{}'\| = \|L_\theta\|$ so (3) follows from (4)

PROPOSITION 4.7

If $\epsilon_\theta \geq 2$ for all $\theta \in \Theta$, then $\mathcal{E}$ is ϵ-deficient relative to $\mathcal{F}$.

Proof: If L , σ and ρ are given, then for $\theta \in \Theta$

$$|P_\theta \rho L_\theta - Q_\theta \sigma L_\theta| \leq |P_\theta \rho L_\theta| + |Q_\theta \sigma L_\theta| \leq \|P_\theta \rho\| \, \|L_\theta\| + \|Q_\theta \sigma\| \, \|L_\theta\|$$

$$= \|P_\theta\| \, \|L_\theta\| + \|Q_\theta\| \, \|L_\theta\| = 2\|L_\theta\|$$

by prop. 18 of appendix B.

LEMMA 4.8

Let $\mathcal{E}$ and $\mathcal{F}$ be given as before, and let $\mathcal{G} = (\mathcal{Z}, \mathcal{C};\ H_\theta : \theta \in \Theta)$ be another experiment. Assume that $\mathcal{E}$ is ϵ-deficient relative to $\mathcal{F}$ (for k-decision problems) and that $\mathcal{F}$ is η-deficient relative to $\mathcal{G}$ (for k-decision problems). Then $\mathcal{E}$ is $(\epsilon + \eta)$-deficient relative to $\mathcal{G}$ (for k-decision problems).

Proof: Let $T_k = \{1,\ldots,k\}$, let L_θ be a loss-function and τ a decision-rule in $\mathcal{G}$. By assumption there is a decision-rule σ in $\mathcal{F}$ such that $Q_\theta \sigma L_\theta \leq H_\theta \tau L_\theta + \eta_\theta \|L_\theta\|;\ \theta \in \Theta$

Further more, there is a rule ρ in $\mathcal{E}$ with

$$P_\theta \rho L_\theta \leq Q_\theta \sigma L_\theta + \epsilon_\theta \|L_\theta\| \ ; \ \theta \in \Theta$$

Thus we have

$P_\theta \rho L_\theta \leq H_\theta \tau L_\theta + (\epsilon_\theta + \eta_\theta)\|L_\theta\|$; $\theta \in \Theta$

and we are done.

LEMMA 4.9

If $\mathcal{E}$ is ϵ-deficient relative to $\mathcal{F}$ (for k-decision problems), then $\mathcal{E}$ is η-deficient relative to $\mathcal{F}$ (for k-decision problems) if $\eta_\theta \geq \epsilon_\theta$; $\theta \in \Theta$

Proof: Obvious.

DEFINITION 4.10

The deficiency of $\mathcal{E}$ relative to $\mathcal{F}$ (for k-decision problems) is defined as infimum of all constants $\epsilon \geq 0$ such that $\mathcal{E}$ is ϵ-deficient relative to $\mathcal{F}$ (for k-decision problems). It is denoted $\delta(\mathcal{E},\mathcal{F})$ $(\delta_k(\mathcal{E},\mathcal{F}))$.

Remark: In the above definition, we consider constant functions ϵ defined on Θ . The deficiency between $\mathcal{E}$ and $\mathcal{F}$ may be interpreted as a measure for the maximal loss of information by observing $\mathcal{E}$ instead of $\mathcal{F}$. By prop. 4.7, $\delta(\mathcal{E},\mathcal{F}) \leq 2$, $\delta_k(\mathcal{E},\mathcal{F}) \leq 2$.

Note that in general $\delta(\mathcal{E},\mathcal{F}) \neq \delta(\mathcal{F},\mathcal{E})$. We now define a concept of distance between experiments.

DEFINITION 4.11

The distance between $\mathcal{E}$ and $\mathcal{F}$ (w.r.t. k-decision problems) is defined as $\Delta(\mathcal{E},\mathcal{F}) = \delta(\mathcal{E},\mathcal{F}) \vee \delta(\mathcal{F},\mathcal{E})$

$(\Delta_k(\mathcal{E},\mathcal{F}) = \delta_k(\mathcal{E},\mathcal{F}) \vee \delta_k(\mathcal{F},\mathcal{E}))$

Remark: The Δ-distance was introduced by Le Cam in [7].

In the sequel, $\delta_{(k)}$, $\Delta_{(k)}$, $\underset{(k)}{\geq}$, $\underset{(k)}{\sim}$ occuring in a statement, will signify that the conclusions hold for δ, Δ, $\geq$, $\sim$ as well as for δ_k, Δ_k, $\underset{k}{\geq}$, $\underset{k}{\sim}$

PROPOSITION 4.12

Let $\mathcal{E}$, $\mathcal{F}$ and $\mathcal{G}$ be experiments. Then

(i) $\delta_{(k)}(\mathcal{E},\mathcal{E}) = 0$, $\delta_{(k)}(\mathcal{E},\mathcal{F}) \geq 0$

(ii) $\delta_{(k)}(\mathcal{E},\mathcal{G}) \leq \delta_{(k)}(\mathcal{E},\mathcal{F}) + \delta_{(k)}(\mathcal{F},\mathcal{G})$.

Proof: (i) is trivial.

As for (ii), choose $\epsilon > \delta_{(k)}(\mathcal{E},\mathcal{F})$,

$$\eta > \delta_{(k)}(\mathcal{F},\mathcal{G}) .$$

By lemma 4.8 and 4.9, $\mathcal{E}$ is $(\epsilon+\eta)$-deficient relative to $\mathcal{G}$ (for k-dec. problems) so

$$\delta(\mathcal{E},\mathcal{G}) \leq \epsilon + \eta$$

Letting $\epsilon \downarrow \delta(\mathcal{E},\mathcal{F})$, $\eta \downarrow \delta(\mathcal{F},\mathcal{G})$, (ii) follows.

PROPOSITION 4.13

(i) $\Delta_{(k)}(\mathcal{E},\mathcal{E}) = 0$, $\Delta_{(k)}(\mathcal{E},\mathcal{F}) \geq 0$

(ii) $\Delta_{(k)}(\mathcal{E},\mathcal{F}) = \Delta_{(k)}(\mathcal{F},\mathcal{E})$

(iii) $\Delta_{(k)}(\mathcal{E},\mathcal{G}) \leq \Delta_{(k)}(\mathcal{E},\mathcal{F}) + \Delta_{(k)}(\mathcal{F},\mathcal{G})$

Proof: (i) and (ii) are easy consequences of the definitions.

Furthermore, by prop. 4.12,

$$\begin{aligned}\Delta_{(k)}(\mathcal{E},\mathcal{G}) &= \delta_{(k)}(\mathcal{E},\mathcal{G}) \vee \delta_{(k)}(\mathcal{G},\mathcal{E})\\ &\leq [\delta_{(k)}(\mathcal{E},\mathcal{F}) + \delta_{(k)}(\mathcal{F},\mathcal{G})] \vee [\delta_{(k)}(\mathcal{G},\mathcal{F}) + \delta_{(k)}(\mathcal{F},\mathcal{E})]\\ &\leq [\delta_{(k)}(\mathcal{E},\mathcal{F}) \vee \delta_{(k)}(\mathcal{F},\mathcal{E})] + [\delta_{(k)}(\mathcal{F},\mathcal{G}) \vee \delta_{(k)}(\mathcal{G},\mathcal{F})]\\ &= \Delta_{(k)}(\mathcal{E},\mathcal{F}) + \Delta_{(k)}(\mathcal{F},\mathcal{G}) \quad \text{which proves (iii)}\end{aligned}$$

Remark: This proposition shows that $\Delta_{(k)}$ has the properties of a semi-metric. However, mathematically we are not permitted to talk about "the set of experiments". In chapter 5, we will consider certain equivalence classes of experiments, which will be seen to constitute a metric space.

PROPOSITION 4.14

(i) $\delta_k(\mathcal{E},\mathcal{F}) \uparrow \delta(\mathcal{E},\mathcal{F})$ as $k \to \infty$

(ii) $\Delta_k(\mathcal{E},\mathcal{F}) \uparrow \Delta(\mathcal{E},\mathcal{F})$ as $k \to \infty$

Proof: That $\delta_k(\mathcal{E},\mathcal{F}) \leq \delta_{k+1}(\mathcal{E},\mathcal{F})$, $k = 1,2,\ldots$ is a consequence of prop. 4.6. Assume $\delta_k(\mathcal{E},\mathcal{F}) \uparrow c$, ($c \leq 2$ by prop. 4.6). Let $\varepsilon > c$. Now, $\mathcal{E}$ is ε-deficient relative to $\mathcal{F}$ for k-decision problems for each $k = 1,2,\ldots,$ and hence $\delta(\mathcal{E},\mathcal{F}) \leq \varepsilon$.

Letting $\varepsilon \downarrow c$ we get $\delta(\mathcal{E},\mathcal{F}) \leq c$. It remains to prove that $\delta(\mathcal{E},\mathcal{F}) \geq c$. Choose $\eta > \delta(\mathcal{E},\mathcal{F})$. It follows that $\delta_k(\mathcal{E},\mathcal{F}) \leq \eta$ for all k, so $c \leq \eta$. By letting $\eta \downarrow \delta(\mathcal{E},\mathcal{F})$, this implies $\delta(\mathcal{E},\mathcal{F}) \geq c$.

(ii) follows easily from (i).

PROPOSITION 4.15

$\delta_1(\mathcal{E},\mathcal{F}) = \Delta_1(\mathcal{E},\mathcal{F}) = 0$

Proof: Let the decision space be $(T,\mathcal{A})$, where $T = \{1\}$, $\mathcal{A} = \{\emptyset,T\}$. Each decision-rule σ in $\mathcal{F}$, being a Markov-kernel, has the property $\sigma(1|y) = 1$ for all $y \in \mathcal{Y}$. Thus, if L_θ is a loss-function,

$$Q_\theta\sigma L_\theta = \int[\int L_\theta(t)\sigma(dt|y)]Q_\theta(dy) = \int L_\theta(1)Q_\theta(dy) = L_\theta(1)$$

Similarly, for any decision-rule ρ in $\mathcal{E}$, $P_\theta\rho L_\theta = L_\theta(1)$.

It is seen that $\mathcal{E}$ is 0-deficient relative to $\mathcal{F}$ and conversely.

Remark: It follows that any two experiments are equivalent for 1-decision problems. However, this is a trivial decision-problem, since there is only one possible decision to take.

REMARK 4.16

By 4.5 it follows that

$$(5) \quad \mathcal{E} \underset{(k)}{\sim} \mathcal{F} \Rightarrow \Delta_{(k)}(\mathcal{E},\mathcal{F}) = 0$$

It will be proved in chapter 5 that equivalence holds in (5).

DEFINITION 4.17

If $\mathcal{E}$ and $\mathcal{F}$ are experiments as given in 4.1, then the product of $\mathcal{E}$ and $\mathcal{F}$, denoted $\mathcal{E} \times \mathcal{F}$ is the experiment

$(\chi \times \mathcal{Y}, \mathcal{A} \times \mathcal{B}; P_\theta \times Q_\theta; \theta \in \Theta)$

where $\mathcal{A} \times \mathcal{B}$ is the product σ-algebra on $\chi \times \mathcal{Y}$ and $P_\theta \times Q_\theta$ are product measures.

Remark: The above definition may easily be generalized to products of arbitrary families $\{\mathcal{E}_t : t \in T\}$ of experiments. We remark that if $\mathcal{E}$ is the experiment of observing a random variable X, and $\mathcal{F}$ is the experiment of observing a random variable Y which is independent of X, then $\mathcal{E} \times \mathcal{F}$ is the experiment obtained by observing the pair (X,Y).

REMARK 4.18

Many of the results on comparison of experiments may without difficulties be generalized to situations where the basic measures are only required to be finite (signed) measures. Such "experiments" are called pseudo experiments. They occur, for example, in the theory of local comparison of experiments. They will, however, not be treated in this book. We refer to [ ].

5. CRITERIONS FOR DEFICIENCIES

We shall in this chapter mainly consider experiments with finite parameter set. Throughout the chapter, $\mathcal{E}$ and $\mathcal{F}$ will be experiments as defined in 4.1. Unless otherwise stated, we assume that $\Theta = \{1,\ldots,s\}$. It is clear that any finite parameter set $\{\theta_1,\ldots,\theta_s\}$ may be identified with the set $\{1,\ldots,s\}$.

DEFINITION 5.1

We let $P = \sum_\theta P_\theta$, $Q = \sum_\theta P_\theta$. Then P and Q are finite positive measures and clearly $P_\theta \ll P$, $Q_\theta \ll Q$ for all $\theta \in \Theta$. By Radon-Nikodym's theorem we can define

$f_\theta = dP_\theta/dP$, $g_\theta = dQ_\theta/dQ$, $\theta \in \Theta$

Finally, let $f : \chi \to R^s$, $g : \chi \to R^s$ be defined by

$f = (f_1,\ldots,f_s)$, $g = (g_1,\ldots,g_s)$

PROPOSITION 5.2

We may assume $f_\theta , g_\theta \geq 0$ and

$$\sum_\theta f_\theta(x) = 1 \quad \text{for all} \quad x \in \chi$$

$$\sum_\theta g_\theta(y) = 1 \quad \text{for all} \quad y \in \mathcal{Y}$$

Proof: By R-N's theorem, the above conditions are valid almost everywhere w.r.t. P (respectively Q). By redefining the f's and g's on a set of measure 0 , the conditions will hold everywhere.

REMARK 5.3

For each $x \in \chi$, $(f_1(x),\ldots,f_s(x))$ defines a distribution over Θ . This distribution is the posterior distribution given x , when the prior distribution is the uniform distribution over Θ . (Similarly for the g's).

We now define what we shall mean by $\psi(\mathcal{E})$ when ψ is a positive homogenous functional. We begin with some motivation: Let $(\chi,\mathcal{A})$ be a measurable space and P,Q be probability measures on $(\chi,\mathcal{A})$. To our experiment we may assign the following quantities:

$\int\sqrt{dPdQ}$ (the affinity of P and Q).

$\int\sqrt{|dP^2 - dQ^2|}$ (the Hellinger-distance between P and Q).

$\int dP \vee dQ$

The above expressions are all of the form $\int\psi(dP,dQ)$ where ψ is a positive homogenous real function (i.e. $\psi(tx,ty) = t\psi(x,y)$ for $t \geq 0$, $x,y \in R$) .

We define $\int\psi(dP,dQ)$ in the following way:

Let μ be a non-negative measure on $\mathcal{A}$ such that $P,Q << \mu$.

Define f and g by $P(A) = \int_A f d\mu$, $Q(A) = \int_A g d\mu$ $(A \in \mathcal{A})$.

Set $\int\psi(dP,dQ) = \int\psi(f,g)d\mu$.

As will be seen later, this definition is independent of our choice of μ . It is often convenient to let $\mu = P + Q$.

As an example, the affinity of P and Q equals $\int\sqrt{fg}\,d\mu$, which is large if f and g are large "together", i.e. P and Q assigns large mass to the same sets (intuitively speaking).

DEFINITION 5.4

Let $\mathcal{E}$ be an experiment and let ψ be a positive homogenous function defined on R^S ,

(i.e. $\psi(tx) = t\psi(x)$ for $t \geq 0$, $x \in R^S$) .

Then

$$\psi(\mathcal{E}) = \int\psi(dP_1,\ldots,dP_s) \quad \text{is defined by}$$

$$\psi(\mathcal{E}) = \int\psi(\tilde{f}_1,\ldots,\tilde{f}_s)d\mu$$

where μ is a σ-finite non-negative measure, $P_\theta << \mu$ and $\tilde{f}_\theta = dP_\theta/d\mu \quad \theta = 1,\ldots,s$

PROPOSITION 5.5

$\psi(\mathcal{E})$ is well-defined by def. 5.4, i.e. the quantity $\psi(\mathcal{E})$ is independent of our choice of μ.

In fact

$$\psi(\mathcal{E}) = \int\psi(f_1,\ldots,f_s)dP$$

where $f_1,\ldots,f_s$ and P are given in 5.1.

Proof: Let μ be given as in def. 5.4.

Set $\tilde{f}_\theta = dP_\theta/d\mu \quad \theta = 1,\ldots,s$

Now

$$f_\theta = \frac{dP_\theta}{d\Sigma P_\theta} = \frac{dP_\theta/d\mu}{d\Sigma P_\theta/d\mu} = \frac{\tilde{f}_\theta}{\sum\limits_\theta \tilde{f}_\theta}$$

Hence, by the positive homogenity of ψ,

$$\int\psi(\tilde{f}_1,\ldots,\tilde{f}_s)d\mu$$

$$= \int\psi(\frac{\tilde{f}_1}{\Sigma\tilde{f}_\theta},\ldots,\frac{\tilde{f}_\theta}{\Sigma\tilde{f}_\theta})\sum_\theta\tilde{f}_\theta d\mu$$

$$= \int\psi(f_1,\ldots,f_s)\sum_\theta dP_\theta = \int\psi(f_1,\ldots,f_s)dP .$$

The proposition follows.

EXAMPLE 5.6

Let $\chi = \{1,\ldots,r\}$, $\mathcal{O}\!\mathcal{l}$ = class of subsets of χ. Each P_i will then be given by a vector $(p_{i1},\ldots,p_{ir})$ with

$p_{ij} \geq 0$, $\sum_j p_{ij} = 1$. The experiment $\mathcal{E}$ is thus completely determined by the Markov matrix

$$P_{\mathcal{E}} = \begin{pmatrix} p_{11} & \cdots & p_{1r} \\ p_{21} & \cdots & p_{2r} \\ \vdots & & \\ p_{s1} & \cdots & p_{sr} \end{pmatrix}$$

Let μ be the <u>counting measure</u> on χ (i.e. for any $A \subseteq \chi$, $\mu(A)$ = number of elements in A). Clearly each $P_i \ll \mu$ and $\tilde{f}_i(j) = p_{ij}$ $\quad i = 1,\ldots,s$; $j = 1,\ldots,r$

Thus, by definition 5.4,

$$\begin{aligned}\psi(\mathcal{E}) &= \int \psi(\tilde{f}_1,\ldots,\tilde{f}_s)d\mu \\ &= \sum_{j=1}^{r} \psi(\tilde{f}_1(j),\ldots,\tilde{f}_s(j)) \\ &= \sum_{j=1}^{r} \psi(p_{1j},\ldots,p_{sj})\end{aligned}$$

We now investigate the connection between $\psi(\mathcal{E})$ and Bayes risks.

<u>PROPOSITION 5.7</u>

Let $\psi \in \Psi_k$ and let $T_k = \{1,\ldots,k\}$ be a decision space. Then there is a loss function $\{L_\theta(t) : \theta \in \Theta$, $t \in T\}$ such that $\psi(\mathcal{E})$ equals $-s$ times the minimum Bayes risk relative to the uniform distribution on Θ . Conversely, to each decision space T_k and each loss function L on $\Theta \times T_k$ there corresponds a sublinear functional $\psi \in \Psi_k$ such that $\psi(\mathcal{E})$ has the above property.

Proof: If $\psi \in \Psi_k$, then $\psi(x) = \bigvee_{t=1}^{k} \sum_{\theta=1}^{s} a_{\theta,t} x_\theta$ for some coefficients $a_{\theta,t}$.

We define the loss function L by

$$L_\theta(t) = -a_{\theta,t} \; ; \quad \theta \in \Theta \; , \quad t \in T \; .$$

By definition

$$\psi(\mathcal{E}) = \int \psi(f_1, \ldots, f_s) dP$$

$$= \int \bigvee_{t=1}^{k} \sum_{\theta=1}^{s} (-L_\theta(t)) f_\theta(x) P(dx)$$

Now, for any decision-rule ρ,

$$(1) \quad \psi(\mathcal{E}) \geq \int \sum_{t=1}^{k} \sum_{\theta=1}^{s} (-L_\theta(t)) f_\theta(x) \rho(t|x) P(dx)$$

$$= - \sum_{\theta t}\sum L_\theta(t) \int \rho(t|x) P_\theta(dx)$$

$$= - \sum_{\theta t}\sum L_\theta(t)(P_\theta \rho)(t) = - \sum_\theta P_\theta \rho L_\theta$$

Furthermore,

$$\frac{1}{s} \sum_\theta P_\theta \rho L_\theta = \frac{1}{s} \sum_\theta r_\rho(\theta) = r(\Lambda_0, \rho)$$

is seen to be the Bayes risk of ρ w.r.t. the uniform distribution Λ_0 on Θ (see 3.13).

By (1), $r(\Lambda_0, \rho) \geq -\frac{1}{s}\psi(\mathcal{E})$ for all ρ. However, equality may be obtained in (1) if ρ for each x assigns mass 1 to the t for which maximum occurs. We leave to the reader to verify the measurability of this ρ.

It follows that $\inf_\rho r(\Lambda_0, \rho) = -\frac{1}{s}\psi(\mathcal{E})$ and the first part of the proposition follows.

The last part follows by defining

$$\psi(x) = \bigvee_{t=1}^{k} \sum_{\theta=1}^{s} (-L_\theta(t)) x_\theta \quad \text{for all} \quad x \in R^s \; .$$

We are now in position to state and prove *the fundamental theorem on comparison of experiments with finite parameter set.* We let our decision space be $T_k = \{1,\dots,k\}$ and we let $\mathcal{S}_k$ be the family of all subsets of T_k. By e_θ we shall mean the θ-th unity coordinate vector of R^s, $\theta = 1,\dots,s$.

THEOREM 5.8

The following conditions are all equivalent:

(i) $\mathcal{E}$ is ε-deficient relative to $\mathcal{F}$ for k-decision problems.

(ii) To each decision-rule σ in $\mathcal{F}$ (relative to T_k) and to each loss function $\{L_\theta(t) : \theta \in \Theta, \ t \in T_k\}$ there corresponds a decision-rule ρ in $\mathcal{E}$ (relative to T_k) so that
$$\sum_\theta P_\theta \rho L_\theta \leq \sum_\theta Q_\theta \sigma L_\theta + \sum_\theta \varepsilon_\theta \|L_\theta\| .$$

(iii) To each decision-rule σ in $\mathcal{F}$ (relative to T_k) there corresponds a decision-rule ρ in $\mathcal{E}$ (relative to T_k) so that
$$\|P_\theta \rho - Q_\theta \sigma\| \leq \varepsilon_\theta \quad \text{for all } \theta \in \Theta .$$

(iv) $\psi(\mathcal{E}) \geq \psi(\mathcal{F}) - \sum_\theta \varepsilon_\theta [\psi(e_\theta) \vee \psi(-e_\theta)]$ for any (s-dimensional) $\psi \in \Psi_k$.

Remark: The criterion (ii) is called the *average risk criterion* for ε-deficiency for k-decision problems. Dividing all terms by s, we observe that (ii) is a statement about Bayes risks relative to the uniform distribution on Θ. The criterion (iii) may be called the criterion for *comparison by operational characteristics* (see 3.8).

The criterion (iv) is called the ψ-criterion. Equivalent formulations of the ψ-criterion will be considered in 5.13.

In order to prove the theorem, we will need the following lemmas:

LEMMA 5.9

Define a two-person zero-sum game Γ by $\Gamma = (\mathcal{L}, \mathcal{D}, M)$ where $\mathcal{L} = \{L : \|L\| \leq 1\}$ is the set of loss functions on $\Theta \times T_k$ bounded by ± 1, $\mathcal{D}$ is the set of decision-rules in $\mathcal{E}$ relative to T_k, and the pay-off function M is defined on $\mathcal{L} \times \mathcal{D}$ by $M(L,\rho) = \sum_{\theta=1}^{s} (P_\theta \rho L_\theta - Q_\theta \sigma L_\theta - \epsilon_\theta \|L_\theta\|)$ for some fixed σ and ϵ.

Then Γ is concave-convex (see 2.16).

Proof: For each fixed ρ, $M(L,\rho)$ is a concave function of L. This follows since $P_\theta \rho L_\theta$ and $Q_\theta \sigma L_\theta$ are linear in L and $- \|L_\theta\|$ is a concave function of L. Thus, since $\mathcal{L}$ is a convex set, $\mathcal{L}$ is seen to be concave relative to Γ (def. 2.12).

Furthermore, $\mathcal{D}$ is convex (in fact affine) relative to Γ since $M(L,\rho)$ is a linear function in ρ for fixed L and $\mathcal{D}$ is a convex set.

The lemma follows.

LEMMA 5.10

There exists a topology on $\mathcal{D}$ for which $\mathcal{D}$ is compact and for which $M(L,\rho)$ is continous in ρ for each fixed L.

Proof: Consider first the set $\mathcal{E}$ of all measurable functions from χ to $[0,1]$. For each bounded and measurable function h on χ (i.e. for each $h \in L\infty(P)$) we define a functional F_h on $\mathcal{E}$ by

$$F_h(\delta) = \int h(x)\delta(x)P(dx) \; ; \quad \delta \in \mathcal{E}$$

Furnish $\mathcal{E}$ with the coarsest topology for which all the functionals F_h are continuous. We prove now that $\mathcal{E}$ is compact in this topology. It is enough to prove that each net $\{\delta_\alpha\}$ contains a convergent subnet. Since $0 \leq \delta_\alpha \leq 1$, each net $\{\delta_\alpha\}$ is uniformly integrable. Hence, by the weak compactness lemma (appendix C) there is a δ and a subnet $\{\delta_\beta\}$ such that

$$F_h(\delta_\beta) \to \int h(x)\delta(x)P(dx) \quad \text{for each} \quad h \in L_\infty(P) \; .$$

Clearly δ may be taken as a member of $\mathcal{E}$. Hence $\delta_\beta \to \delta$ in our topology, which proves that $\mathcal{E}$ is compact.

Each decision-rule $\rho \in \mathcal{D}$ may be identified with the vector-valued function

$$(\rho(1|\cdot),\ldots,\rho(k|\cdot)) \in \mathcal{E}^k \; , \quad \text{i.e.} \quad \mathcal{D} \subseteq \mathcal{E}^k$$

$\mathcal{D}$ is closed, since $\mathcal{D}$ is the subset of $\mathcal{E}^k$ with component-functions with sum 1. Hence $\mathcal{D}$ is a closed subset of the compact set $\mathcal{E}^k$ and thus compact itself.

As regards the last assertion of the lemma, it is enough to prove that for each θ, $P_\theta \rho L_\theta$ is continous as a function of ρ.

Now, $$P_\theta \rho L_\theta = \int \sum_t L_\theta(t)\rho(t|x)P_\theta(dx)$$

$$= \sum_t L_\theta(t) \int \rho(t|x) f_\theta(x) P(dx)$$

Since $|f_\theta(x)| \leq 1$, and by the definition of the topology on $\mathcal{D}$, the weak compactness lemma asserts that for each $\{\rho_\alpha\}$ converging to ρ, $P_\theta \rho_\alpha L_\theta$ converges to $P_\theta \rho L_\theta$. Continuity follows.

Proof of theorem 5.8:

(i) => (ii) is trivial.

(ii) => (i): Assume (ii) holds, and let σ be a given decision-rule in $\mathcal{F}$.

Then

(2) $\sup\limits_{L:\|L\|\leq 1} \inf\limits_{\rho} \sum\limits_{\theta}(P_\theta \rho L_\theta - Q_\theta \sigma L_\theta - \epsilon_\theta \|L_\theta\|) \leq 0$

which is equivalent to

(3) $\underline{V} = \sup\limits_{L} \inf\limits_{\rho} M(L,\rho) \leq 0$.

We will prove that $\underline{V} = \overline{V}$ and that player II has a minimax strategy ρ_o . In order to do this, we shall make use of theorem 2.35. This theorem, giving conditions for the existence of a maximin strategy for player I, may of course in a straightforward manner be extended to a theorem concerning player II. We note that $|M(L,\rho)| < \infty$ for all L and ρ . By lemma 5.9 and lemma 5.10, theorem 2.35 yield what we want and hence there exists a $\rho_o \in \mathcal{D}$ such that by (3)

$\sup\limits_{L:\|L\|\leq 1} M(L,\rho_o) = \overline{V} = \underline{V} \leq 0$. Hence

(4) $\sum\limits_{\theta} P_\theta \rho_o L_\theta \leq \sum\limits_{\theta} Q_\theta \sigma L_\theta + \sum\limits_{\theta} \epsilon_\theta \|L_\theta\|$

for any L such that $\|L\| \leq 1$.

Dividing by $\|L\|$, we conclude (4) holds for any finite loss function. We are now in position to deduce (i).

Choose $\theta_o \in \Theta$ arbitrary, let $\{L_\theta(t) : \theta \in \Theta , \quad t \in T_k\}$ be a loss-function and let σ be a decision-rule in $\mathcal{F}$.

Define L' by $L'_\theta(t) = \begin{cases} L_{\theta_o}(t) & \text{if } \theta = \theta_o \\ 0 & \text{if } \theta \neq \theta_o \end{cases}$

Then, by the previous results, there exists a decision-rule ρ_o in $\mathcal{E}$ such that (by (4) with L replaced by L')

(5) $P_{\theta_o} \rho_o L_{\theta_o} \leq Q_{\theta_o} \sigma L_{\theta_o} + \epsilon_{\theta_o} \|L_{\theta_o}\|$.

(ii) => (iii): Choose $\theta_o \in \Theta$. Let σ be a decision-rule in $\mathcal{F}$.

By (5), there is a ρ_o in $\mathcal{E}$ such that $(P_{\theta_o} \rho_o - Q_{\theta_o} \sigma)(L_{\theta_o}) \leq \epsilon_{\theta_o}$.

for each L such that $\|L_{\theta_o}\| \leq 1$.

Hence, by def. 10 of appendix B,

$\|P_{\theta_o}\rho_o - Q_{\theta_o}\sigma\| \leq \epsilon_{\theta_o}$ and we are done.

(iii) => (i): Let L and σ be given. By (iii) there is a ρ such that for each θ , $\|P_\theta\rho - Q_\theta\sigma\| \leq \epsilon_\theta$, which implies that $|P_\theta\rho L_\theta - Q_\theta\rho L_\theta| \leq \|P_\theta\rho - Q_\theta\rho\|\,\|L_\theta\| \leq \epsilon_\theta\|L_\theta\|$ which implies (i).

(ii) => (iv): Let $\psi \in \Psi_k$. Let the loss function L be given as in prop. 5.7. By (ii)

(6) $\min\limits_{\rho} \sum\limits_{\theta} P_\theta\rho L_\theta \leq \sum\limits_{\theta} Q_\theta\sigma L_\theta + \sum\limits_{\theta}\epsilon_\theta\|L_\theta\|$ for each σ which implies

(7) $\min\limits_{\rho} \sum\limits_{\theta} P_\theta\rho L_0 \leq \min\limits_{\sigma} \sum\limits_{\theta} Q_\theta\sigma L_\theta + \sum\limits_{\theta}\epsilon_\theta\|L_0\|$ or, by prop. 5.7

(multiplying each term by -1)

(8) $\psi(\mathcal{E}) \geq \psi(\mathcal{F}) - \sum\limits_{\theta}\epsilon_\theta\|L_\theta\|$

By the proof of prop. 5.7,

$\psi(e_\theta) = \bigvee\limits_{t}(-L_\theta(t))$, $\psi(-e_\theta) = \bigvee\limits_{t} L_0(t)$ so

$\psi(e_0) \vee \psi(-e_\theta) = \bigvee\limits_{t}|L_\theta(t)| = \|L_0\|$ (iv) follows.

Since (8), (7), (6) and (ii) are in fact equivalent, (iv) => (ii) follows as well. The proof of our fundamental theorem is thus complete.

COROLLARY 5.11

(i) $\delta_{(k)}(\mathcal{E},\mathcal{F}) = 0 <=> \psi(\mathcal{E}) \geq \psi(\mathcal{F})$ for all $\psi \in \Psi_{(k)}$

(ii) $\Delta_{(k)}(\mathcal{E},\mathcal{F}) = 0 <=> \psi(\mathcal{E}) = \psi(\mathcal{F})$ for all $\psi \in \Psi_{(k)}$

Proof: It suffices to prove (i). The version concerning δ_k is a direct consequence of 5.8 (iv). In fact, if

$\delta_k(\mathcal{E},\mathcal{F}) = 0$, and $\psi \in \Psi_k$ then for any $\varepsilon > 0$, $\psi(\mathcal{E}) \geq \psi(\mathcal{F}) - \varepsilon\sum_\theta[\psi(e_\theta) \vee \psi(-e_\theta)]$. Hence $\psi(\mathcal{E}) \geq \psi(\mathcal{F})$.

The opposite implication is trivial.

Assume now that $\delta(\mathcal{E},\mathcal{F}) = 0$. Then $\delta_k(\mathcal{E},\mathcal{F}) = 0$ for all k , by prop. 4.14 and hence $\psi(\mathcal{E}) \geq \psi(\mathcal{F})$ for any $\psi \in \Psi_1 \cup \Psi_2 \cup \ldots\ldots\ldots$

Let $\psi \in \Psi$. Then $\psi = \lim \psi_k$ where $\psi_1 \leq \psi_2 \leq \ldots\ldots\ldots$ and $\psi_k \in \Psi_k$ for all k (prop. 1.52).

By the monotone convergence theorem we have $\psi_k(\mathcal{E}) \to \psi(\mathcal{E})$, $\psi_k(\mathcal{F}) \to \psi(\mathcal{F})$, so from $\psi_k(\mathcal{E}) \geq \psi_k(\mathcal{F})$ it follows that $\psi(\mathcal{E}) \geq \psi(\mathcal{F})$.

Finally, assume that $\psi(\mathcal{E}) \geq \psi(\mathcal{F})$ for all $\psi \in \Psi$. In particular $\psi(\mathcal{E}) \geq \psi(\mathcal{F})$ for all $\psi \in \Psi_k$, $k = 1,2,\ldots$. Hence $\delta_k(\mathcal{E},\mathcal{F}) = 0$ for all k , so that $\delta(\mathcal{E},\mathcal{F}) = 0$ by prop. 4.14.

LEMMA 5.12

Let $\psi_1,\psi_2 \in \Psi$ and let $c \geq 0$. Then

$$(\psi_1+\psi_2)(\mathcal{E}) = \psi_1(\mathcal{E}) + \psi_2(\mathcal{E})$$

$$(c\psi_1)(\mathcal{E}) = c\psi_1(\mathcal{E})$$

Proof: Direct consequences of the definition of $\psi(\mathcal{E})$.

COROLLARY 5.13

The following conditions are equivalent:

(i) $\mathcal{E}$ is ε-deficient relative to $\mathcal{F}$ for k-decision problems.

(ii) $\psi(\mathcal{E}) \geq \psi(\mathcal{F}) - \frac{1}{2}\sum_\theta \varepsilon_\theta(\psi(e_\theta) + \psi(-e_\theta))$ for any $\psi \in \Psi_k$.

(iii) $\psi(\mathcal{E}) \geq \psi(\mathcal{F}) - \sum_\theta \varepsilon_\theta\psi(e_\theta)$ for each $\psi \in \Psi_k$ such that $\psi(-e_\theta) = \psi(e_\theta)$; $\theta \in \Theta$.

Proof: Clearly (ii) => (i) (Theorem 5.8 (iv)). Suppose now 5.8(iv) holds. Let $\psi \in \Psi_k$ and define $\psi' = \psi + l$ where l is the linear functional given by

$$l(x) = \tfrac{1}{2}\sum_\theta [\psi(-e_\theta) - \psi(e_\theta)]x_\theta \ .$$

It is seen that $\psi' \in \Psi_k$. Since $\Delta_1(\mathcal{E},\mathcal{F}) = 0$, $l(\mathcal{E}) = l(\mathcal{F})$ by 5.11. Hence, 5.8(iv) yields together with lemma 5.12

$$\psi(\mathcal{E}) \geq \psi(\mathcal{F}) - \sum_\theta \varepsilon_\theta[\psi'(e_\theta) \vee \psi'(-e_\theta)]$$

But $\psi'(e_\theta) = \psi(e_\theta) + l(e_\theta) = \tfrac{1}{2}[\psi(-e_\theta) + \psi(e_\theta)]$,

and $\psi'(-e_\theta) = \tfrac{1}{2}[\psi(-e_\theta) + \psi(e_\theta)]$ so (ii) holds.

As is easily verified, (ii) => (iii). Assume now (iii), and let ψ and ψ' be as before. Since $\psi'(e_\theta) = \psi'(-e_\theta)$, (ii) is derived from (iii).

COROLLARY 5.14

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

$$\Delta_{(k)}(\mathcal{E},\mathcal{F}) = \sup_{\psi \in \Gamma_{(k)}} |\psi(\mathcal{E}) - \psi(\mathcal{F})|$$

Proof: In 5.13 (iii) there is no restriction to consider only ψ for which $\sum_\theta \psi(e_\theta) = 1$. Thus the present corollary follows from 5.13 (iii) (by use of prop. 1.52).

EXAMPLE 5.15

Consider again example 5.6. Let $\chi = \mathcal{Y} = \{1,\dots,r\}$ and let $P_{\mathcal{E}} = (p_{ij})$, $Q_{\mathcal{F}} = (q_{ij})$ be the matrices defining $\mathcal{E}$ and $\mathcal{F}$. Let $\psi \in \Gamma$ (see cor. 5.14). Then by prop. 1.42,

$$\begin{aligned} |\psi(\mathcal{E}) - \psi(\mathcal{F})| &= \left|\sum_{j=1}^{r} \psi(p_{1j},\ldots,p_{sj}) - \sum_{j=1}^{r} \psi(q_{1j},\ldots,q_{sj})\right| \\ &\leq \sum_{j=1}^{r} |\psi(p_{1j},\ldots,p_{sj}) - \psi(q_{1j},\ldots,q_{sj})| \\ &\leq \sum_{j=1}^{r} \sum_{i=1}^{s} |p_{ij} - q_{ij}|\psi(e_i) \\ &= \sum_{i=1}^{s} \psi(e_i) \sum_{j=1}^{r} |p_{ij} - q_{ij}| \\ &\leq \max_{i} \sum_{j=1}^{r} |p_{ij} - q_{ij}| \ . \end{aligned}$$

The last inequality follows from the fact that $\sum_{i=1}^{s} \psi(e_i) = 1$

Hence, by corollary 5.14,

(9) $\Delta(\mathcal{E},\mathcal{F}) \leq \max_{i} \sum_{j=1}^{r} |p_{ij} - q_{ij}|$

Thus we have obtained an upper bound for the Δ-distance. The ψ-criterion is, however, in most cases a useful tool for finding lower bounds. By cor. 5.14, $\Delta(\mathcal{E},\mathcal{F}) \geq |\psi(\mathcal{E}) - \psi(\mathcal{F})|$ for any $\psi \in \Gamma$. We shall use this property in the next example.

EXAMPLE 5.16

Let the situation be as in the examples 5.6 and 5.15. Assume that $r = s$ and $\chi = \mathcal{Y} = \{1,\ldots,s\}$. Let $\mathcal{E}$ be the experiment where the true value of θ is observed. Thus $\mathcal{E}$ is given by $P_{\mathcal{E}}$ = identity matrix. This experiment contains all information about θ .

Let $\mathcal{F}$ be an experiment given by a Markov matrix $Q_{\mathcal{F}}$ where all rows are equal to $(\frac{1}{s},\ldots,\frac{1}{s})$. Intuitively, $\mathcal{F}$ gives no information about θ .

We shall find an upper bound for $\Delta(\mathcal{E},\mathcal{F})$ by applying to the inequality (9).

For any i , $\sum_{j=1}^{r} |p_{ij} - q_{ij}| = |p_{ii} - q_{ii}| + \sum_{j\neq i} |p_{ij} - q_{ij}|$

$$= 1 - \frac{1}{s} + \sum_{j\neq i} q_{ij} = 2(1 - \frac{1}{s}) = 2 - \frac{2}{s}$$

Define now ψ by

$$\psi(x) = \frac{2}{s} \bigvee_{\theta} x_\theta - \frac{1}{s}\sum_{\theta} x_\theta .$$

Then $\psi \in \Psi$, $\psi(e_\theta) = \psi(-e_\theta) = \frac{1}{s}$, $\sum_{\theta}\psi(e_\theta) = 1$ so $\psi \in \Gamma$.

By ex. 5.6 $\psi(\mathcal{E}) = \sum_{j=1}^{s} \psi(e_\theta) = 1$.

$\psi(\mathcal{F}) = \sum_{j=1}^{s} \psi(\frac{1}{s},\dots,\frac{1}{s}) = s\psi(\frac{1}{s},\dots,\frac{1}{s}) = \psi(1,\dots,1) = \frac{2}{s} - 1$.

Hence $\psi(\mathcal{E}) - \psi(\mathcal{F}) = 2 - \frac{2}{s}$ so $\Delta(\mathcal{E},\mathcal{F}) \geq 2 - \frac{2}{s}$.

Consequently, $\Delta(\mathcal{E},\mathcal{F}) = 2 - \frac{2}{s}$.

Since $\delta(\mathcal{E},\mathcal{F}) = 0$, it follows that $\delta(\mathcal{F},\mathcal{E}) = 2 - \frac{2}{s}$.

Let now $\mathcal{G}$ and $\mathcal{H}$ be arbitrary experiments. Then, since $\mathcal{E}$ is the maximal informative experiment, $\delta(\mathcal{E},\mathcal{G}) = \delta(\mathcal{E},\mathcal{H}) = 0$ and since $\mathcal{F}$ is the minimal informative experiment, $\delta(\mathcal{G},\mathcal{F}) = \delta(\mathcal{H},\mathcal{F}) = 0$.

Hence, $\delta(\mathcal{G},\mathcal{H}) \leq \delta(\mathcal{G},\mathcal{F}) + \delta(\mathcal{F},\mathcal{E}) + \delta(\mathcal{E},\mathcal{H}) = 2 - \frac{2}{s}$

and $\delta(\mathcal{H},\mathcal{G}) \leq 2 - \frac{2}{s}$

so $\Delta(\mathcal{G},\mathcal{H}) \leq 2 - \frac{2}{s}$.

(Compare this with prop. 4.7)

EXAMPLE 5.17

We shall give another example of the use of (9).

Consider a Markov-chain $X_0, X_1, \dots$ with state space $\{1,2\}$ and transition matrix

$$P = \begin{pmatrix} 1-\alpha & \alpha \\ \beta & 1-\beta \end{pmatrix} \quad 0 \leq \alpha \leq 1 , \quad 0 \leq \beta \leq 1 , \quad \alpha + \beta \neq 0,1,2$$

Let the initial state X_o be our unknown parameter θ (i.e. $\Theta = \{1,2\}$) and let $\mathcal{E}_n$ be the experiment obtained by observing X_n . $\mathcal{E}_n$ is then given by the matrix

$$P^n = \frac{1}{\alpha+\beta}\begin{pmatrix} \beta & \alpha \\ \beta & \alpha \end{pmatrix} + \frac{1}{\alpha+\beta}\begin{pmatrix} \alpha & -\alpha \\ -\beta & \beta \end{pmatrix}(1-\alpha-\beta)^n$$

Let $\mathcal{E}_\infty$ be the experiment given by the matrix

$$A = \lim P^n = \frac{1}{\alpha+\beta}\begin{pmatrix} \beta & \alpha \\ \beta & \alpha \end{pmatrix} .$$

Then, by (9)

$$\Delta(\mathcal{E}_n, \mathcal{E}_\infty) \leq \max \left\{\frac{2\alpha}{\alpha+\beta}, \frac{2\beta}{\alpha+\beta}\right\} |1-\alpha-\beta|^n = \frac{2(\alpha\vee\beta)}{\alpha+\beta} |1-\alpha-\beta|^n$$

Thus $\mathcal{E}_n$ converges to $\mathcal{E}_\infty$ with exponential speed. It will be shown in 7.11 that $\Delta(\mathcal{E}_n, \mathcal{E}_\infty) = |1-\alpha-\beta|^n$.

We state without proof a proposition on product experiments. The proof, which may be found in [16], makes use of the ψ-criterion and Fubini's theorem.

PROPOSITION 5.18

Let $\mathcal{E}_1$ be ε-deficient relative to $\mathcal{F}_1$ (for k-decision problems) and $\mathcal{E}_2$ be ε_2-deficient relative to $\mathcal{F}_2$ (for k-decision problems). Then $\mathcal{E}_1 \times \mathcal{E}_2$ is $(\varepsilon_1 + \varepsilon_2)$ - deficient relative to $\mathcal{F}_1 \times \mathcal{F}_2$ (for k-decision problems).

Remark: The statement is easily generalized to finite products of experiments. It follows that

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

and $\Delta_{(k)}\left(\prod_{i=1}^{n} \mathcal{E}_i, \prod_{i=1}^{n} \mathcal{F}_i\right) \leq \sum_{i=1}^{n} \Delta_{(k)}(\mathcal{E}_i, \mathcal{F}_i)$

In particular, if

$$\mathcal{E}_i \geq \mathcal{F}_i \ ; \quad i = 1,\ldots,n \ , \text{ then}$$

$$\prod_{i=1}^{n} \mathcal{E}_i \geq \prod_{i=1}^{n} \mathcal{F}_i \ .$$

This was proved by Blackwell in [1].

We let $\mathcal{E}^n$ denote the experiment $\prod_{i=1}^{n} \mathcal{E}$ (If $\mathcal{E}$ is the experiment of observing a random variable X, then $\mathcal{E}^n$ is the experiment of observing independent, identically distributed variables $X_1,\ldots,X_n$ with the same distribution as X).

It follows that $\mathcal{E} \geq \mathcal{F} \Rightarrow \mathcal{E}^n \geq \mathcal{F}^{-n}$.

That the converse is not true, is shown in [16]. It is shown, however, that $\mathcal{E}^n \sim \mathcal{F}^n \Rightarrow \mathcal{E} \sim \mathcal{F}$.

As is noted in 4.13, the class of experiments will not constitute a set in the strong mathematical sense. We shall now, however, show that we may let equivalent experiments be represented by a standard experiment, the class of which will define a set and constitute a metric space under the metrics $\Delta, \Delta_2, \Delta_3, \ldots\ldots\ldots$

Standard experiments were introduced by Blackwell in [1].

NOTATION

Let K be the subset of R^S defined by

$$K = \left\{x : x \in R^S, \quad x_\theta \geq 0 \text{ for all } \theta \ , \quad \sum_{\theta=1}^{S} x_\theta = 1\right\}$$

DEFINITION 5.19

A standard experiment is an experiment of the form $(K, \mathcal{B}, S_\theta : \theta \in \Theta)$ where $\mathcal{B}$ is the class of Borel-subsets of K and $dS_\theta / d\sum_\theta S_\theta$ is equal a.e. to the function φ_θ on K

defined by $\varphi_\theta(x) = x_\theta$; $\theta = 1,\ldots,s$.

A *standard measure S* is a positive measure defined on $(K,\mathcal{B})$ such that $\int x_\theta S(dx) = 1$ for all $\theta \in \Theta$.

PROPOSITION 5.20

Let $(K,\mathcal{B},S_\theta : \theta \in \Theta)$ be a standard experiment. Then $\sum_\theta S_\theta$ is a standard measure. Conversely, to each standard measure S on $(K,\mathcal{B})$ there corresponds a unique standard experiment $(K,\mathcal{B}; S_\theta : \theta = \Theta)$ such that $S = \sum_\theta S_\theta$.

Proof: Clearly $dS_\theta = x_\theta d\Sigma S_\theta$, so $\int x_\theta d\Sigma S_\theta = \int dS_\theta = 1$. This proves the first assertion.

Let now S be a standard measure on $(K,\mathcal{B})$ and define for each $\theta \in \Theta$, S_θ by $S_\theta(B) = \int_B x_\theta S(dx)$; $B \in \mathcal{B}$. By definition of S , $S_\theta(K) = 1$. Furthermore,

$\sum_\theta S_\theta(B) = \sum_\theta \int_B x_\theta S(dx) = \int_B S(dx) = S(B)$; $B \in \mathcal{B}$, so $S = \sum_\theta S_\theta$.

Uniqueness follows from the fact that $dS_\theta/dS = x_\theta$ a.e. must hold for each θ .

Remark: If $\mathcal{E}$ is a standard experiment with standard measure S , then it is seen that $\psi(\mathcal{E}) = \int \psi dS$.

PROPOSITION 5.21

Let $\mathcal{E}$ be an experiment. Then the experiment $\hat{\mathcal{E}} = (K,\mathcal{B},S_\theta : \theta \in \Theta)$ where $S_\theta = P_\theta f^{-1}$; $\theta \in \Theta$ (f is given in 5.1) defines a standard experiment.

Remark: f is a function on χ which takes values in K . Hence $P_\theta f^{-1}$ is well-defined as a measure on $\mathcal{B}$ such that $P_\theta f^{-1}(B) = P_\theta(f^{-1}(B))$.

Proof: We prove that $dS_\theta/d\sum_\theta S_\theta = x_\theta$ a.e. for all θ.

Set $S = Pf^{-1}$. Clearly $\sum_\theta S_\theta = \sum_\theta P_\theta f^{-1} = Pf^{-1} = S$.

For any $B \in \mathcal{B}$,

$$S_\theta(B) = P_\theta(f^{-1}(B)) = \int_{f^{-1}(B)} f_\theta(x)P(dx)$$

$$= \int f_\theta(x)I_B(f(x))P(dx) = \int x_\theta I_B(x)Pf^{-1}(dx)$$

$$= \int_B x_\theta Pf^{-1}(dx)$$

where we have made use of the substitution formula for integrals.

Hence $dS_\theta/dS = x_\theta$ a.e. and the proof is complete.

DEFINITION 5.22

Let $\mathcal{E}$ be an experiment. The the standard experiment of $\mathcal{E}$ is the standard experiment $\hat{\mathcal{E}}$ defined in the preceding proposition. By proposition 5.20 and 5.21 the standard experiment of $\mathcal{E}$ may be defined as the (unique) standard experiment with standard measure Pf^{-1}.

PROPOSITION 5.23

Let $\mathcal{E}$ be an experiment. Then

(i) $\hat{\hat{\mathcal{E}}} = \hat{\mathcal{E}}$

(ii) $\Delta_{(k)}(\mathcal{E}, \hat{\mathcal{E}}) = 0$

Proof: (i) follows from the fact that for a standard experiment, f (as defined in 5.1) maps each point of K into itself.

By corollary 5.11, (ii) is equivalent to $\psi(\mathcal{E}) = \psi(\hat{\mathcal{E}})$ for all $\psi \in \Psi_{(k)}$.

But

$$\psi(\mathcal{E}) = \int\psi(f(x))P(dx) = \int\psi(x)Pf^{-1}(x) = \psi(\hat{\mathcal{E}}) \text{ .}$$

THEOREM 5.24

Let $\mathcal{E}$ and $\mathcal{F}$ be experiments. Then

$$\Delta(\mathcal{E},\mathcal{F}) = 0 \iff \hat{\mathcal{E}} = \hat{\mathcal{F}} \text{ .}$$

Proof: (<=) Assume $\hat{\mathcal{E}} = \hat{\mathcal{F}}$. By prop. 5.23 (ii), $\Delta(\mathcal{E},\mathcal{F}) \leq \Delta(\mathcal{E},\hat{\mathcal{E}}) + \Delta(\hat{\mathcal{E}},\hat{\mathcal{F}}) + \Delta(\hat{\mathcal{F}},\mathcal{F}) = 0$ so $\Delta(\mathcal{E},\mathcal{F}) = 0$.

(=>) Assume $\Delta(\mathcal{E},\mathcal{F}) = 0$. Then $\Delta(\hat{\mathcal{E}},\hat{\mathcal{F}}) = 0$, since $\Delta(\hat{\mathcal{E}},\hat{\mathcal{F}}) \leq \Delta(\hat{\mathcal{E}},\mathcal{E}) + \Delta(\mathcal{E},\mathcal{F}) + \Delta(\mathcal{F},\hat{\mathcal{F}})$.

Assume that $\hat{\mathcal{E}}$ and $\hat{\mathcal{F}}$ have standard measures S and T , respectively. We have to prove that $S = T$.

By 5.11,

(10) $\int\psi(x)T(dx) = \int\psi(x)S(dx)$ for all $\psi \in \Psi$.

Since S and T are Borel-measures, it is enough to prove that

$$\int h dT = \int h dS$$

for each continuous function h on K .

We shall do this by applying to Stone-Weierstrass theorem (see e.g. [12] Ch. 7).

Let V be the set of all functions on K which are of the form $\psi_1 - \psi_2$ where ψ_1 and ψ_2 are sublinear functionals on R^S . By (10),

(11) $\int v dT = \int v dS$ for all $v \in V$.

Clearly,

(12) $u,v \in V \Rightarrow u+v \in V$

and $v \in V$, $c \in R \Rightarrow cv \in V$.

Define ψ by $\psi(x) = \sum_\theta x_\theta$. Then $\psi(x) = 1$ for all $x \in K$, which shows that V contains each constant function on K.

Let $\overline{V}$ be the uniform closure of V, i.e. the set of functions on K which are limits of uniformly convergent sequences of members of V.

We will prove that $\overline{V}$ is an algebra of continous functions.

By (12),

$u,v \in \overline{V} \Rightarrow u+v \in \overline{V}$ and $v \in \overline{V}$, $c \in R \Rightarrow cv \in \overline{V}$.

It remains to prove that $u,v \in \overline{V} \Rightarrow uv \in \overline{V}$. Since $4uv = (u+v)^2 - (u-v)^2$, it suffices to prove that $v \in \overline{V}$ implies $v^2 \in \overline{V}$.

From the identity $|\psi_1 - \psi_2| = 2\psi_1 \vee \psi_2 - (\psi_1 + \psi_2)$ it follows that $v \in V \Rightarrow |v| \in V$. Hence $v \in \overline{V} \Rightarrow |v| \in \overline{V}$ which in turn implies that $u,v \in \overline{V} \Rightarrow u \vee v \in \overline{V}$.

Let now $v \in \overline{V}$. For each $x \in K$,

$$v(x)^2 = \sup_a \, [2av(x) - a^2]$$

where sup is taken over all real numbers a. Since R is separabel, it suffices to take the supremum over a countable dense subset $\{a_1, a_2, \ldots\ldots\}$ of R. Hence

$$v^2 = \bigvee_{i=1}^{\infty} [2a_i v - a_i^2]$$

Since for each i the function in the brackets is a member of $\overline{V}$, it follows that $v^2 = \lim_{n\to\infty} v_n$ where

$$v_n = \bigvee_{i=1}^{n} [2a_i v - a_i^2] \in \overline{V} .$$

v^2 is continuous since v is, so by Dinis theorem (theorem 7.13 of [12]), the convergence $v_n \to v^2$ is uniform. Hence $v^2 \in \overline{V}$.

Obviously $\overline{V}$ separates points and vanishes at no point of K, so by Stone-Weierstrass approximation theorem, $\overline{V}$ is exactly the set of all continuous functions on K. From (11) it is seen that

$$\int v dT = \int v dS \quad \text{for any} \quad v \in \overline{V} ,$$

which completes the proof.

COROLLARY 5.25

The set of standard experiments is a metric space with metric Δ.

Proof: If $\mathcal{E}$ and $\mathcal{F}$ are standard experiments, then by theorem 5.23

$$\Delta(\mathcal{E},\mathcal{F}) = 0 \iff \mathcal{E} = \mathcal{F}$$

By prop. 4.13,

$$\Delta(\mathcal{E},\mathcal{F}) = \Delta(\mathcal{F},\mathcal{E})$$
$$\Delta(\mathcal{E},\mathcal{G}) \leq \Delta(\mathcal{E},\mathcal{F}) + \Delta(\mathcal{F},\mathcal{G})$$

REMARK 5.26

We shall see later (theorem 6.10) that if $\mathcal{E}$ and $\mathcal{F}$ are experiments, then $\Delta_2(\mathcal{E},\mathcal{F}) = 0 \Rightarrow \Delta(\mathcal{E},\mathcal{F}) = 0$.
Hence $\Delta_2(\mathcal{E},\mathcal{F}) = 0 \iff \Delta_3(\mathcal{E},\mathcal{F}) = 0 \iff \dots \iff \Delta(\mathcal{E},\mathcal{F}) = 0$
since $\Delta_2(\mathcal{E},\mathcal{F}) \leq \Delta_3(\mathcal{E},\mathcal{F}) \leq \dots \leq \Delta(\mathcal{E},\mathcal{F})$.

Consequently $\Delta_2, \Delta_3, \dots$ will also define metrics on the set of standard experiments. It may be shown that they are all equivalent and equivalent to Δ. Furthermore, it has been shown that Δ is equivalent to the Lévy-distance Λ. (We note that the Lévy-distance between distribution functions F and G on R^n is given by

$$\begin{aligned}\Lambda(F,G) = \inf \{h : h \geq 0,\ & F(x_1 - h,\dots,x_n - h) - h \\ & \leq G(x_1,\dots,x_n) \leq F(x_1 + h,\dots,x_n + h) + h \\ & \text{for all } (x_1,\dots,x_n) \in R^n\}).\end{aligned}$$

The assertions are proved in [16].

Assume now that the parameter set Θ is not necessarily finite. It turns out that problems on infinite parameter sets may occasionally be reduced to problems on finite parameter sets. This problem is treated in the following theorem. First we need some notation.

NOTATIONS

Let $\mathcal{E} = (\chi, \mathcal{A}; P_\theta : \theta \in \Theta)$ be an experiment and let F be a subset of Θ. We denote by $\mathcal{E}_F$ the experiment $(\chi, \mathcal{A}; P_\theta : \theta \in F)$.

If f is a function defined on Θ, then $f|_F$ is the restriction of f to the subset F of Θ.

THEOREM 5.27

Let $\mathcal{E} = (\chi, \mathcal{A}; P_\theta : \theta \in \Theta)$ and $\mathcal{F} = (\mathcal{Y}, \mathcal{B}; Q_\theta : \theta \in \Theta)$ be experiments (with arbitrary parameter set Θ). Assume further that $\mathcal{E}$ is dominated (see def. 10 of appendix A) and let ϵ be a non-negative function on Θ.

Then $\mathcal{E}$ is ϵ-deficient relative to $\mathcal{F}$ (for k-decision problems) if and only if $\mathcal{E}_F$ is $\epsilon|_F$-deficient relative to $\mathcal{F}_F$ (for k-decision problems) for all finite, non-empty subsets $F \subseteq \Theta$.

Proof: The "only if"-part is trivial by def. 4.1. Assume therefore that the condition holds. Let $T_k = \{1, \ldots, k\}$ and let $\{L_\theta(t) : \theta \in \Theta, \ t \in T_k\}$ be a bounded loss function. Let σ be a decision-rule in $\mathcal{F}$. By def. 4.1, for each finite, non-empty $F \subset \Theta$ there is a decision-rule ρ_F in $\mathcal{E}$ such that

(12) $P_\theta \rho_F L_\theta \le Q_\theta \sigma L_\theta + \epsilon_\theta \|L_\theta\|$ for all $\underline{\theta \in F}$.

Let S be the family of finite subsets of F . S is easily seen to be a <u>directed set</u> if we define

$F_1 \le F_2 \iff_{\text{def}} F_1 \subseteq F_2$, and hence $\{\rho_F : F \in S\}$ defines a <u>net</u> (generalized sequence).

Next, since $\mathcal{E}$ is dominated, there is a probability measure π on $(\chi,\mathcal{A})$ such that $P_\theta << \pi$ for all $\theta \in \Theta$. Let $h_\theta = dP_\theta/d\pi$

For each $t \in T_k$ and each F , $\|\rho_F(t|\circ)\|_\infty \le 1$.

Hence, by Remark 4. of appendix C, there is a subnet $\{\rho_{F'}(t|\circ) : F' \in S'\}$ and a function $\rho(t|\circ)$ such that

(13) $\int \rho_{F'}(t|x)h_\theta(x)d\pi \to \int \rho(t|x)h_\theta(x)\ \pi$ for each $t \in T_k$, $\theta \in \Theta$.

Since $\sum_{t=1}^{k} \rho_F(t|x) = 1$ for all F and x , it is seen that ρ may be chosen such that $\sum_{t=1}^{k} \rho(t|x) = 1$ for all x and $0 \le \rho(t|x) \le 1$ for all $t \in T_k$, $x \in \chi$. Hence ρ is a decision-rule in $\mathcal{E}$.

It remains to prove that

$P_\theta \rho L_\theta \le Q_\theta \sigma L_\theta + \epsilon_\theta \|L_\theta\|$ for <u>all</u> $\theta \in \Theta$.

Let $\theta_o \in \Theta$. Then $\theta_o \in F_o'$ for some $F_o' \in S'$ and hence by (12)

(14) $P_{\theta_o} \rho_{F'} L_{\theta_o} \le Q_{\theta_o} \sigma L_{\theta_o} + \epsilon_{\theta_o} \|L_{\theta_o}\|$ for any $F' \in S'$

such that $F' \ge F_o'$.

But $P_{\theta_o} \rho_{F'} L_{\theta_o} = \sum_{t=1}^{k} L_{\theta_o}(t) \int \rho_{F'}(t|\circ) h_{\theta_o} d\pi$ which by (13) converges to $P_{\theta_o} \rho L_{\theta_o}$.

Hence (14) implies that $P_{\theta_o} \rho L_{\theta_o} \le Q_{\theta_o} \sigma L_{\theta_o} + \epsilon_{\theta_o} \|L_{\theta_o}\|$ and the proof is complete.

COROLLARY 5.28

Assume that $\mathcal{E}$ and $\mathcal{F}$ are both dominated. Then $\mathcal{E} \underset{(k)}{\sim} \mathcal{F}$ if and only if $\mathcal{E}_F \underset{(k)}{\sim} \mathcal{F}_F$ for all finite, non-empty subsets F of Θ.

Moreover, $\mathcal{E} \underset{2}{\sim} \mathcal{F} \iff \mathcal{E} \underset{3}{\sim} \mathcal{F} \iff \ldots \iff \mathcal{E} \sim \mathcal{F}$.

Proof: The first assertion follows directly from the preceding theorem. The last part follows from Remark 5.26. In fact, the result given in 5.26 implies that $\mathcal{E}_F \underset{2}{\sim} \mathcal{F}_F \Rightarrow \mathcal{E}_F \sim \mathcal{F}_F$ for any finite subset $F \subseteq \Theta$. Hence by the above result $\mathcal{E} \underset{2}{\sim} \mathcal{F} \Rightarrow \mathcal{E} \sim \mathcal{F}$.

In example 6.13 we show that it is not enough to require $\mathcal{E}_F \sim \mathcal{F}_F$ only for all strict subsets F of Θ.

COROLLARY 5.29

Let $\mathcal{E}$, $\mathcal{F}$ and ϵ be as in theorem 5.27. Then $\mathcal{E}$ is ϵ-deficient relative to $\mathcal{F}$ for k-decision problems if and only if to each decision-rule σ in $\mathcal{F}$ relative to T_k, there is a decision-rule ρ in $\mathcal{E}$ relative to T_k such that

$$\|P_\theta \rho - Q_\theta \sigma\| \leq \epsilon_\theta \quad \text{for all} \quad \theta \in \Theta .$$

Proof: In the proof of theorem 5.27, the decision-rule ρ was constructed independent of the loss-function L. Hence for all bounded loss-functions L and all $\theta \in \Theta$,

$$(P_\theta \rho - Q_\theta \sigma)(L_\theta) \leq \epsilon_\theta \|L_\theta\| ,$$

which implies $\|P_\theta \rho - Q_\theta \sigma\| \leq \epsilon_\theta$ for all θ.

6. COMPARISON BY TESTING PROBLEMS

(See § 3 of Torgersen [16]).

Let $\mathcal{E}$ and $\mathcal{F}$ be defined as in 4.1 and let $\Theta = \{1,\ldots,s\}$. Theorem 5.13 applied to the case $k = 2$ yields:

THEOREM 6.1

$\mathcal{E}$ is ε-deficient relative to $\mathcal{F}$ for testing problems if and only if

$$\|\sum_\theta a_\theta P_\theta\| \geq \|\sum_\theta a_\theta Q_\theta\| - \sum_\theta \varepsilon_\theta |a_\theta| \quad \text{for each vector} \quad a \in R^s \ .$$

($\| \ \|$ is defined in def. 10 of appendix B).

Proof: Each $\psi \in \Psi_2$ is of the form $\psi = l_1 \vee l_2$ for some linear functionals l_1, l_2 on R^s . By the identity

$$l_1 \vee l_2 = \tfrac{1}{2}(l_1 + l_2 + |l_2 - l_1|) \ ,$$

any $\psi \in \Psi_2$ may be written in the form $L_1 + |L_2|$ where $L_1, L_2 \in \Psi_1$. Since $L(\mathcal{E}) = L(\mathcal{F})$ for each $L \in \Psi_1$, theorem 5.13 (iii) states that it suffices to require

$$\psi(\mathcal{E}) \geq \psi(\mathcal{F}) - \sum_\theta \varepsilon_\theta \psi(e_\theta)$$

whenever $\psi = |L|$ with $L \in \Psi_1$.

L may be written $L(x) = \sum_\theta a_\theta x_\theta$. Hence, if $\psi = |L|$,

$$\psi(\mathcal{E}) = \int |\sum_\theta a_\theta f_\theta| dP$$

$$= \int \left|\frac{d\Sigma a_\theta P_\theta}{dP}\right| dP = \|\sum_\theta a_\theta P_\theta\| \ .$$

(This is seen by splitting up the integrand in its positive and negative part).

Similarly, $\psi(\mathcal{F}) = \|\sum_\theta a_\theta Q_\theta\|$.

The theorem follows since $\psi(e_\theta) = |a_\theta|$; $\theta \in \Theta$.

Theorem 6.1 has a geometric interpretation as follows. Let $\mathcal{E}$ be an experiment. The set of all test functions in $\mathcal{E}$ (i.e. measurable functions from χ to $[0,1]$) will be denoted by $\mathcal{C}_{\mathcal{E}}$ and $V_{\mathcal{E}}$ shall denote the subset of $[0,1]^s$ consisting of all vectors of the form

$$\left(\int \delta dP_1,\ldots,\int \delta dP_s\right) \quad \text{where} \quad \delta \in \mathcal{C}_{\mathcal{E}} \quad ,$$

i.e. $V_{\mathcal{E}}$ is the set of available power functions. Finally, put for

$x,y \in R^s$, $I_{[x,y]} = \{z : x_i \leq z_i \leq y_i \ , \quad i = 1,\ldots,s\}$

Then we have:

COROLLARY 6.2

$\mathcal{E}$ is ϵ-deficient relative to $\mathcal{F}$ for testing problems if and only if

$$V_{\mathcal{E}} + \tfrac{1}{2}\, I_{[-\epsilon,\epsilon]} \supseteq V_{\mathcal{F}}$$

Remark: If $A,B \subseteq R^s$, we define $A + B = \{x+y : x \in A \ , \ y \in B\}$

Proof: By prop. 1.44, the support function of $V_{\mathcal{E}}$ is given by

$$H_{\mathcal{E}}(a) = \sup_{y \in V_{\mathcal{E}}} \langle a,y \rangle = \sup_{\delta \in \mathcal{C}_{\mathcal{E}}} \sum_{\theta=1} a_\theta \int \delta dP_\theta = \sup_{\delta \in \mathcal{C}_{\mathcal{E}}} \int \delta d\sum_\theta a_\theta P_\theta$$

If μ is a finite signed measure on a measurable space $(\chi,\mathcal{O}\!\mathcal{l})$, then

$$\|\mu\| = \sup_{\|f\| \leq 1} \int f d\mu = 2\left[\sup_{\|f\| \leq 1} \int \frac{f+1}{2} d\mu - \tfrac{1}{2}\mu(\chi)\right] = 2 \sup_{0 \leq \delta \leq 1} \int \delta d\mu - \mu(\chi)$$

Hence $\displaystyle \sup_{\delta} \int \delta d\mu = \frac{\|\mu\| + \mu(\chi)}{2}$.

If we apply this to $\mu = \sum_\theta a_\theta P_\theta$, then

$$H_{\mathcal{E}}(a) = \frac{\|\sum a_\theta P_\theta\| + \sum_\theta a_\theta}{2} \quad \text{for all} \quad a \in R^s .$$

Similarly, the support function of $V_{\mathcal{F}}$ is given by

$$H_{\mathcal{F}}(a) = \frac{\|\sum_\theta a_\theta Q_\theta\| + \sum_\theta a_\theta}{2} \quad ; \quad a \in R^s \ .$$

Finally, we shall derive the support function H of $\frac{1}{2}I_{[-\varepsilon,\varepsilon]}$. For any $a \in R^s$, $y \in \frac{1}{2}I_{[-\varepsilon,\varepsilon]}$ we have

$$\langle a,y\rangle = \sum_\theta a_\theta y_\theta \leq \tfrac{1}{2}\sum_\theta |a_\theta|\varepsilon_\theta$$

where equality is obtained by choosing

$$y_\theta = \tfrac{1}{2}\varepsilon_\theta \ \text{sign} \ a_\theta \ : \ \theta = 1,\dots,s \ .$$

Hence $H(a) = \frac{1}{2}\sum_\theta |a_\theta|\varepsilon_\theta$; $a \in R^s$.

Theorem 6.1 now states that $\mathcal{E}$ is ε-def. relative to $\mathcal{F}$ if and only if

$$H_{\mathcal{E}}(a) + H(a) \geq H_{\mathcal{F}}(a) \quad \text{for all} \quad a \in R^s \ ,$$

which by prop. 1.54 and 1.57 is equivalent to the statement of the corollary.

COROLLARY 6.3

$\mathcal{E}$ is ε-deficient relative to $\mathcal{F}$ for testing problems if and only if for each testing problem $H : \theta \in \Theta_0$ against $K : \theta \in \Theta - \Theta_0$ and each power function $\beta_{\mathcal{F}}$ available in $\mathcal{F}$ there is a power function $\beta_{\mathcal{E}}$ available in $\mathcal{E}$ such that

$$\beta_{\mathcal{E}}(\theta) \leq \beta_{\mathcal{F}}(\theta) + \tfrac{1}{2}\varepsilon_\theta \ ; \quad \theta \in \Theta_0 \ ,$$

$$\beta_{\mathcal{E}}(\theta) \geq \beta_{\mathcal{F}}(\theta) - \tfrac{1}{2}\varepsilon_\theta \ ; \quad \theta \in \Theta - \Theta_0 \ .$$

Proof: Let $\beta_{\mathcal{F}}$ be a power function in $\mathcal{F}$. Then there is a $\delta \in \mathcal{C}_{\mathcal{F}}$ such that

$y_\delta = (\int \delta dQ_1, \dots, \int \delta dQ_s) \in V_{\mathcal{F}}$, where $\beta_{\mathcal{F}}(\theta) = \int \delta dQ_\theta$; $\theta = 1,\dots,s$.

By corollary 6.2, $y_\delta \in V_{\mathcal{E}} + \frac{1}{2}I_{[-\varepsilon,\varepsilon]}$. Hence there is a $\delta' \in \mathcal{C}_{\mathcal{E}}$ such that for some $a_1,\dots,a_s$ where $|a_i| \leq 1$,

$$(\int \delta dQ_1,\dots,\int \delta dQ_s) = (\int \delta' dP_1,\dots,\int \delta' dP_s) + \tfrac{1}{2}(a_1\epsilon_1,\dots,a_s\epsilon_s) .$$

Define $\beta_{\mathcal{E}}$ by $\beta_{\mathcal{E}}(\theta) = \int \delta' dP_\theta$; $\theta = 1,\dots,s$.

Then $|\beta_{\mathcal{E}}(\theta) - \beta_{\mathcal{F}}(\theta)| = \frac{1}{2}|a_\theta|\epsilon_\theta \leq \frac{1}{2}\epsilon_\theta$ for all θ .

This statement clearly implies the corollary.

REMARK 6.4

$V_{\mathcal{E}}$ is seen to have the following properties $V_{\mathcal{E}} \subseteq [0,1]^s$, $V_{\mathcal{E}}$ is symmetric about $(\frac{1}{2},\dots,\frac{1}{2})$ (This follows since $1-\delta \in \mathcal{E}_{\mathcal{E}}$ whenever $\delta \in \mathcal{E}_{\mathcal{E}}$), V is compact and convex and $0 \in V$. It may be shown that if $s = 2$, then every set with the above properties is a$V_{\mathcal{E}}$. When $s > 2$, however, this is no longer true.

EXAMPLE 6.5

Let $\mathcal{E}$ be an experiment such that $\chi = \{1,2,3\}$, $\Theta = \{1,2,3\}$. Then $\mathcal{E}$ is determined by a Markov-matrix

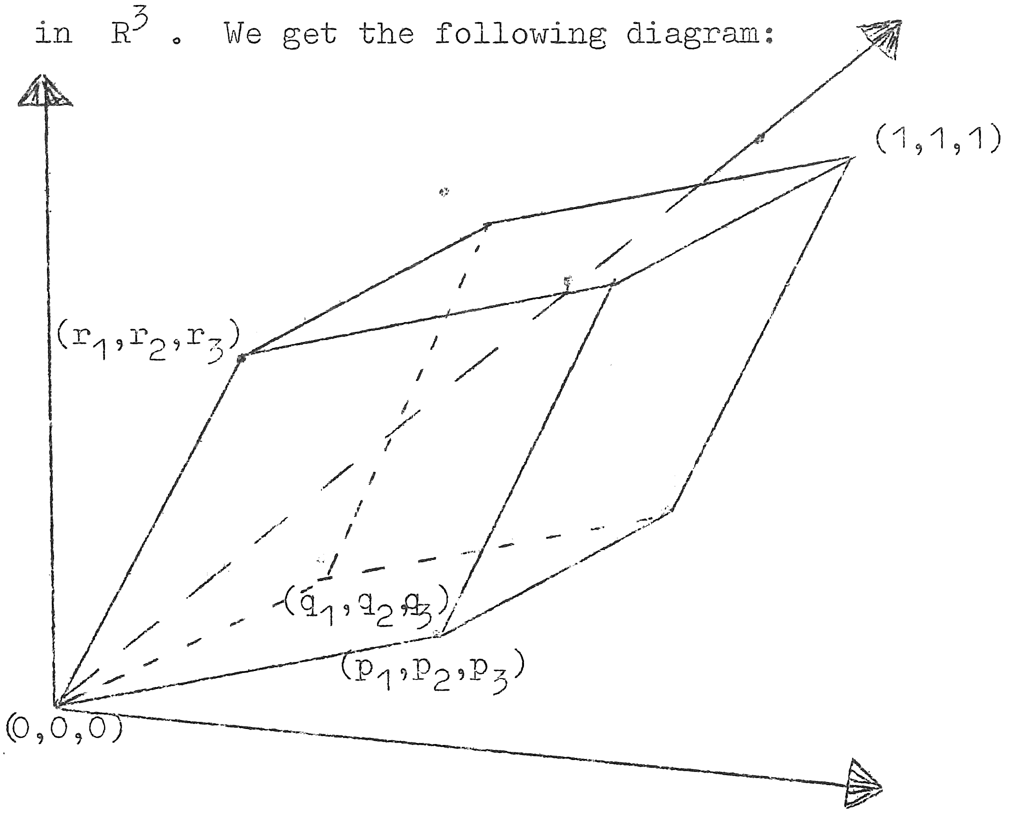
$$\begin{array}{c|ccc} \theta \backslash x & 1 & 2 & 3 \\ \hline 1 & p_1 & q_1 & r_1 \\ 2 & p_2 & q_2 & r_2 \\ 3 & p_3 & q_3 & r_3 \end{array}$$

A test function δ on χ is given by a triple $(\delta_1,\delta_2,\delta_3)$ such that $0 \leq \delta_i \leq 1$; $i = 1,2,3$. Hence $V_{\mathcal{E}}$ is the subset of R^3 consisting of all points of the form

$$(p_1\delta_1 + q_1\delta_2 + r_1\delta_3,\ p_2\delta_1 + q_2\delta_2 + r_2\delta_3,\ p_3\delta_1 + q_3\delta_2 + r_3\delta_3) .$$

It may be shown that $V_{\mathcal{E}}$ is the convex hull of the eight points obtained when δ is non-randomized (i.e. $\delta_i = 0$ or 1 for all i).

By the symmetri property of $V_{\mathcal{E}}$'s , $V_{\mathcal{E}}$ is a parallel-epiped in R^3 . We get the following diagram:



$V_{\mathcal{E}}$ is the parallel-epiped spanned by the vectors (p_1,p_2,p_3) , (q_1,q_2,q_3) and (r_1,r_2,r_3) . If $\mathcal{F}$ is another experiment with $\mathcal{Y} = (1,2,3)$, $\Theta = \{1,2,3\}$,

$$Q_{\mathcal{F}} = \begin{pmatrix} \alpha_1 & \beta_1 & \gamma_1 \\ \alpha_2 & \beta_2 & \gamma_2 \\ \alpha_3 & \beta_3 & \gamma_3 \end{pmatrix}$$

then we may sketch $V_{\mathcal{F}}$ in the same diagram and apply corollary 6.2 in order to compute deficiencies.

It is seen that $\mathcal{E} \underset{2}{\geq} \mathcal{F}$ if and only if

$(\alpha_1,\alpha_2,\alpha_3)$, $(\beta_1,\beta_2,\beta_3)$, $(\gamma_1,\gamma_2,\gamma_3) \in V_{\mathcal{E}}$.

A similar example, with $\chi = \{1,2\}$, $\Theta = \{1,2\}$ is given in Blackwell and Girshick [3].

DEFINITION 6.6

An experiment $\mathcal{E}$ is called a dichotomy if $\#\Theta = 2$ (i.e. $s = 2$ in our terminology).

THEOREM 6.7

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

Proof: Since only the "if"-part needs proof, suppose that $\mathcal{E}$ is ϵ-deficient relative to $\mathcal{F}$ for 2-decision problems and let $\psi \in \Psi_k$. By definition of Ψ_k there are constants $a_1,\dots,a_k$ and $b_1,\dots,b_k$ such that

$$\psi(x_1,x_2) = \bigvee_{i=1}^{k} (a_i x_1 + b_i x_2) .$$

Set $l_i(x_1,x_2) = a_i x_1 + b_i x_2$.

By rearranging we may assume that there is a r so that

$$(2) \quad \psi(1,x_2) = \bigvee_{i=1}^{r} l_i(1,x_2)$$

when $x_2 > 0$, where the representation on the right is minimal in the sense that for each $i \leq r$ there is a $x_2 > 0$ so that $l_i(1,x_2) > l_j(1,x_2)$ for all $j \neq i$.

The functionals $l_i(1,x_2) = a_i + b_i x_2$ will define lines in the plane with slope b_i. From the representation (2) it is thus seen that $b_1,\dots,b_r$ are all distinct (equal b_i's would correspond to parallel lines). Thus we may assume that $b_1 < b_2 < \dots < b_r$. It follows that $a_1 > a_2 > \dots > a_r$ (this is easily seen from a diagram).

Furthermore, for any $x_2 > 0$, it is seen that

$$\psi(1,x_2) = l_1(1,x_2) + [l_2(1,x_2) - l_1(1,x_2)]^+ \\ + \ldots\ldots + [l_r(1,x_2) - l_{r-1}(1, x_2)]^+$$

(where $a^+ = a \vee 0$; $a \in R$).

Hence if $x_1 > 0$, by the positive homogenity of ψ ,

(3) $$\psi(x_1,x_2) = l_1(x_1,x_2) + [l_2(x_1,x_2) - l_1(x_1,x_2)]^+ \\ + \ldots\ldots + [l_r(x_1,x_2) - l_{r-1}(x_1,x_2)]^+$$

Let $\tilde{\psi}$ be the sublinear functional which equals the right side of (3) for <u>all</u> x_1,x_2 .

$\tilde{\psi}$ is a sum of functionals which are maximum of two linear functionals. Hence $\tilde{\psi} \in \Psi_2$, and

(4) $\psi(x_1,x_2) = \tilde{\psi}(x_1,x_2)$ for all $x_1,x_2 \geq 0$.

We assert that

(5) $\psi(-e_i) \geq \tilde{\psi}(-e_i)$; $i = 1,2$.

In fact,

$$\tilde{\psi}(-e_1) = \tilde{\psi}(-1,0) = -a_1 + (a_1 - a_2)^+ + \ldots\ldots + (a_{r-1} - a_r)^+ \\ = -a_1 + a_1 - a_2 + \ldots\ldots + a_{r-1} - a_r = - a_r$$

whence $\psi(-e_1) = \bigvee_{i=1}^{s} (-a_i) \geq -a_r$.

Similarly,

$$\tilde{\psi}(-e_2) = \tilde{\psi}(0,-1) = -b_1 + (b_1 - b_2)^+ + \ldots\ldots + (b_{r-1} - b_r)^+ \\ = -b_1 + 0 + \ldots\ldots + 0 = -b_1 ,$$

whence $\psi(-e_2) = \bigvee_{i=1}^{s} (-b_i) \geq -b_1$.

By assumption and by condition (ii) of 5.13,

$$\tilde{\psi}(\xi) \geq \tilde{\psi}(\zeta) - \epsilon_1 \frac{\tilde{\psi}(e_1) + \tilde{\psi}(-e_1)}{2} - \epsilon_2 \frac{\tilde{\psi}(e_2) + \tilde{\psi}(-e_2)}{2}$$

By (4)

$$\psi(\mathcal{E}) = \int\psi(f_1,f_2)dP = \int\tilde{\psi}(f_1,f_2)dP = \tilde{\psi}(\mathcal{E})$$

so by (5)

$$\psi(\mathcal{E}) \geq \psi(\mathcal{F}) - \epsilon_1\frac{\psi(e_1) + \tilde{\psi}(-e_1)}{2} - \epsilon_2\frac{\psi(e_2) + \tilde{\psi}(-e_2)}{2}$$

$$\geq \psi(\mathcal{F}) - \epsilon_1\frac{\psi(e_1) + \psi(-e_1)}{2} - \epsilon_2\frac{\psi(e_2) + \psi(-e_2)}{2}$$

and the theorem follows from 5.13.

REMARK 6.8

If $\mathcal{E}$ and $\mathcal{F}$ are dichotomies, then theorem 6.1, corollary 6.2 and 6.3 give criterions for ϵ-deficiency.

Theorem 6.7 states in particular that if $\mathcal{E}$ and $\mathcal{F}$ are dichotomies, then

$$\Delta_2(\mathcal{E},\mathcal{F}) = 0 \Rightarrow \Delta(\mathcal{E},\mathcal{F}) = 0 .$$

We shall now prove this statement in the case of general (finite) parameter set Θ.

The proof involves standard experiments (ch. 5) and we need the succeeding lemma:

LEMMA 6.9

Let P and Q be Borel probability measures on R^k such that $P(A) = Q(A)$ whenever A is a halfspace of R^k (for definition, see 1.29). Then $P = Q$.

Proof: We shall let $\mathcal{L}(X)$ denote the distribution of the random variable X. Assume that X has distribution P and Y has distribution Q. We shall prove that $\mathcal{L}(X) = \mathcal{L}(Y)$. Let $X_1,\dots,X_k$ and $Y_1,\dots,Y_k$ denote the coordinates of X and Y, respectively. By assumption, $P(a_1X_1 + \dots + a_kX_k \leq b) = Q(a_1Y_1 + \dots + a_kY_k \leq b)$ for any real numbers $a_1,\dots,a_k,b$. Hence $\mathcal{L}(a_1X_1 + \dots + a_kX_k) = \mathcal{L}(a_1Y_1 + \dots + a_kY_k)$

for any $a_1,\dots,a_k \in R$. Let φ_X and φ_Y denote the characteristic functions of X and Y. Then

$$\varphi_X(a_1,\dots,a_r) = Ee^{i\Sigma a_j X_j} = Ee^{i\Sigma a_j Y_j} = \varphi_Y(a_1,\dots,a_r)$$

for all $a_1,\dots,a_r$.

Hence $\mathcal{L}(X) = \mathcal{L}(Y)$ by the uniqueness theorem for characteristic functions.

THEOREM 6.10

Let $\mathcal{E}$ and $\mathcal{F}$ be experiments (with finite parameter set).

Then $\Delta_2(\mathcal{E},\mathcal{F}) = 0 \Rightarrow \Delta(\mathcal{E},\mathcal{F}) = 0$.

Proof: Since $\Delta_2(\hat{\mathcal{E}},\hat{\mathcal{F}}) \leq \Delta_2(\hat{\mathcal{E}},\mathcal{E}) + \Delta_2(\mathcal{E},\mathcal{F}) + \Delta_2(\mathcal{F},\hat{\mathcal{F}})$ it follows from prop. 5.24 that we may assume that $\mathcal{E}$ and $\mathcal{F}$ are standard experiments. Suppose $\Delta_2(\mathcal{E},\mathcal{F}) = 0$ and let S and T be the standard measures of $\mathcal{E}$ and $\mathcal{F}$, respectively.

Now, $\Delta_2(\mathcal{E},\mathcal{F}) = 0 \Leftrightarrow \int_S \psi(x)S(dx) = \int \psi(x)T(dx)$ for all $\psi \in \Psi_2$

Define ψ by $\psi(x) = (\sum_{\theta=1} a_\theta x_\theta)^+$ for all $x \in R^s$.

Then $\int (\Sigma a_\theta x_\theta)^+ S(dx) = \int (\Sigma a_\theta x_\theta)^+ T(dx)$.

We shall differentiate w.r.t. a_{θ_o}. Let $h \neq 0$. Then

$$\int \frac{(\sum_\theta a_\theta x_\theta + hx_{\theta_o})^+ - (\sum_\theta a_\theta x_\theta)^+}{h} S(dx)$$

$$= \int \frac{(\sum_\theta a_\theta x_\theta + hx_{\theta_o})^+ - (\sum_\theta a_\theta x_\theta)^+}{h} T(dx) .$$

Since the integrand is dominated by $|x_{\theta_o}|$ when $h \to 0$, we may apply the dominated convergence theorem.

Let therefore $h \to 0$ and consider the following cases:

$\Sigma a_\theta x_\theta > 0$: Then $\sum_\theta a_\theta x_\theta + hx_{\theta_o} > 0$ for h sufficiently small, so the integrand converges to x_{θ_o}.

$\Sigma a_\theta x_\theta < 0$: For h sufficiently small, $\Sigma a_\theta x_\theta + hx_{\theta_o} < 0$ and the integrand converges to 0.

$\Sigma a_\theta x_\theta = 0$: The integrand is then equal to x_{θ_o} if $h > 0$,

0 if $h < 0$.

Let now $h < 0$, $h \to 0$.

Then, if we take the limit,

$$\int_{\underset{\theta}{\Sigma} a_\theta x_\theta > 0} x_{\theta_o} S(dx) = \int_{\underset{\theta}{\Sigma} a_\theta x_\theta > 0} x_{\theta_o} T(dx)$$

Since $dS_{\theta_o} = x_{\theta_o} dS$, this is equivalent to

$$(6)\ S_{\theta_o}(\underset{\theta}{\Sigma} a_\theta x_\theta > 0) = T_{\theta_o}(\underset{\theta}{\Sigma} a_\theta x_\theta > 0) .$$

We remark that S_{θ_o} and T_{θ_o} assigns positive mass only to subsets of $K = \{x : \Sigma x_\theta = 1\}$. Hence, for any b,

$$S_{\theta_o}(\underset{\theta}{\Sigma} a_\theta x_\theta > b) = S_{\theta_o}(\underset{\theta}{\Sigma} a_\theta x_\theta > b\Sigma x_\theta)$$

$$= S_{\theta_o}(\underset{\theta}{\Sigma}(a_\theta - b)x_\theta > 0) .$$

By (6), since $a_1,\ldots,a_s$ are arbitrarily chosen,

$$S_{\theta_o}(\underset{\theta}{\Sigma} a_\theta x_\theta > b) = T_{\theta_o}(\underset{\theta}{\Sigma} a_\theta x_\theta > b) \text{ for any } a_1,\ldots,a_s,b .$$

Hence $S_{\theta_o} = T_{\theta_o}$ by lemma 4.9, and since θ_o was arbitrary, $S = T$.

Finally, this implies $\Delta(\mathcal{E},\mathcal{F}) = 0$. (theorem 5.24).

<u>COROLLARY 6.11</u>

$\Delta(\mathcal{E},\mathcal{F}) = 0$ if and only if

$\|\underset{\theta}{\Sigma} a_\theta P_\theta\| = \|\underset{\theta}{\Sigma} a_\theta Q_\theta\|$ for all $a_1,\ldots,a_s$.

<u>Proof:</u> Easy consequence of theorem 6.1 and 6.10.

<u>Remark:</u> If $\mathcal{E}$ and $\mathcal{F}$ are equivalent, then it is seen that the

normed linear spaces spanned by $P_1,\ldots,P_s$ and $Q_1,\ldots,Q_s$ are isometric by the isometry $\Sigma a_\theta P_\theta \rightsquigarrow \Sigma a_\theta Q_\theta$.

EXAMPLE 6.12

Consider again example 5.6. We have

$$(7)\ \|\sum_\theta a_\theta P_\theta\| = \sum_{j=1}^{r} \left| \sum_{\theta=1}^{s} a_\theta p_{\theta j} \right| .$$

If $\mathcal{F}$ is the experiment obtained from $\mathcal{E}$ by a permutation of columns in $P_{\mathcal{E}}$, then (7) clearly implies that $\mathcal{E} \sim \mathcal{F}$. Hence the inequality (9) of example 5.15 may possibly be improved by permuting the columns of $Q_{\mathcal{F}}$.

EXAMPLE 6.13 (Taken from [16]).

This example shows that in corollary 5.28 it is not enough to require $\mathcal{E}_F \sim \mathcal{F}_F$ for all F which are strict subsets of Θ (i.e. $F \subseteq \Theta$, $F \neq \Theta$). Let $\Theta = \{1,2,3\}$. Define $\mathcal{E} = (\chi, \mathcal{A} ; P_1,P_2,P_3)$, $\mathcal{F} = (\chi, \mathcal{A} ; Q_1,Q_2,Q_3)$ where $\chi = \{1,2,3,4\}$, $\mathcal{A}$ = class of subsets, $Q_1 = P_1$, $Q_2 = P_2$ and P_1,P_2,P_3 and Q_3 .are given by the Markov-matrix

$$\begin{array}{c|cccc}
\theta \diagdown x & 1 & 2 & 3 & 4 \\
\hline
P_1 & \frac{1}{2} & \frac{1}{2} & 0 & 0 \\
P_2 & \frac{1}{8} & \frac{3}{8} & \frac{1}{8} & \frac{3}{8} \\
P_3 & \frac{1}{6} & \frac{2}{6} & \frac{2}{6} & \frac{1}{6} \\
Q_3 & \frac{2}{6} & \frac{1}{6} & \frac{1}{6} & \frac{2}{6}
\end{array}$$

Rather than $\mathcal{E}_{\{1,2\}}$ we shall write $\mathcal{E}_{12}$ etc. Clearly $\mathcal{E}_{12} = \mathcal{F}_{12}$ and $\mathcal{E}_{13} \sim \mathcal{F}_{13}$ $(\mathcal{E}_{23} \sim \mathcal{F}_{23})$ since $\mathcal{E}_{13}$ $(\mathcal{E}_{23})$ may be obtained from $\mathcal{F}_{13}$ $(\mathcal{F}_{23})$ by a per-

mutation of the columns (example 6.12).
Hence $\mathcal{E}_F \sim \mathcal{F}_F$ for any strict subsets $F \subseteq \Theta$
However, from corollary 6.11 we see that $\mathcal{E} \not\sim \mathcal{F}$, since

$$\|P_1 - P_2 + P_3\| = \frac{17}{12} \quad \text{and} \quad \|P_1 - P_2 + Q_3\| = \frac{13}{12}$$

(The measure $P_1 - P_2 + P_3$ is given by the vector $(\frac{13}{24}, \frac{11}{24}, \frac{5}{24}, - \frac{55}{24})$ and hence $\|P_1 - P_2 + P_3\| = \frac{13}{24} + \frac{11}{24} + \frac{5}{24} + \frac{5}{24} = \frac{17}{12}$

$\|P_1 - P_2 + Q_3\|$ is found in a similar way).

7. THE MARKOV KERNEL CRITERION

Corollary 5.29 gives a criterion for comparison of experiments by operational characteristics. We considered there only decision spaces of the form $T_k = \{1,\dots,k\}$. In this chapter we shall investigate situations where the decision space has a more general structure.

The following proposition tells us, in the case of experiments, that certain decision spaces are abundant for comparison by operational characteristics.

PROPOSITION 7.1

Let $\mathcal{E} = (\chi,\mathcal{A};\ P_\theta : \theta \in \Theta)$ and $\mathcal{F} = (\mathcal{Y},\mathcal{B};\ Q_\theta : \theta \in \Theta)$ be experiments and let ϵ be a non-negative function on Θ . We shall say that a decision space $(T,\mathcal{S})$ is admissible if to each decision-rule σ in $\mathcal{F}$ relative to $(T,\mathcal{S})$ there is a decision-rule ρ in $\mathcal{E}$ relative to $(T,\mathcal{S})$ such that

$$\|P_\theta\rho - Q_\theta\sigma\| \leq \epsilon_\theta \quad \text{for all} \quad \theta \in \Theta .$$

Then:

(i) If $(T,\mathcal{S})$ is admissible and $S_o \in \mathcal{S}$, $S_o \neq \emptyset$, then $(S_o, \mathcal{S} \wedge S_o)$ is admissible.

(ii) If $(T,\mathcal{S})$ is admissible and $(T',\mathcal{S}')$ is another decision space such that there exists a bi-measurable bijection $T \to T'$, then $(T',\mathcal{S}')$ is admissible.

Remark: $\mathcal{S} \wedge S_o$ is the σ-algebra $\{S \cap S_o : S \in \mathcal{S}\}$.

Proof: (ii) is clear, so suppose that $(T,\mathcal{S})$ is admissible and $\emptyset \neq S_o \in \mathcal{S}$. Let Γ be a probability measure on $(S_o, \mathcal{S} \wedge S_o)$.

Define a Markov-kernel $\gamma : (\mathcal{S} \wedge S_o) \times T \to [0,1]$.

$\gamma(S|t) = I_S(t)$ if $t \in S_o$, $S \in \mathcal{S} \wedge S_o$.

$\gamma(S|t) = \Gamma(S)$ if $t \notin S_o$, $S \in \mathcal{S} \wedge S_o$.

Let V be a probability measure on $(T,\mathcal{S})$ such that $V(S_o) = 1$.

Let $S \in \mathcal{S} \wedge S_o$. Then:

(1) $(V\gamma)(S) = \int\gamma(S|t)V(dt) = \int_{S_o} I_S(t)V(dt) = V(S \cap S_o) = V(S)$

Define a Markov-kernel

$\tilde{\gamma} : \mathcal{S} \times S_o \to [0,1]$ by $\tilde{\gamma}(S|t) = I_S(t)$; $t \in S_o$, $S \in \mathcal{S}$.

If W is a probability measure on $\mathcal{S} \wedge S_o$, then for any $S \in \mathcal{S}$

(2) $W\tilde{\gamma}(S) = \int\tilde{\gamma}(S|t)W(dt) = \int_{S_o} I_S(t)W(dt) = W(S \cap S_o)$.

Let now σ be a decision-rule in $\mathcal{F}$ relative to $(S_o, \mathcal{S} \wedge S_o)$. Then $\sigma\tilde{\gamma}$ is a decision-rule in $\mathcal{F}$ relative to $(T,\mathcal{S})$, such that by (2), $(\sigma\tilde{\gamma})(S|y) = \sigma(S \cap S_o|y)$ for all $S \in \mathcal{S}$, $y \in \mathcal{Y}$. By assumption, there is a decision-rule ρ in $\mathcal{E}$ relative to $(T,\mathcal{S})$ such that

$$\|P_\theta\rho - Q_\theta\sigma\tilde{\gamma}\| \leq \epsilon_\theta \quad ; \quad \theta \in \Theta .$$

By (1), $(\sigma\tilde{\gamma})\gamma = \sigma$. Finally, $\rho\gamma$ is seen to be a decision-rule in $\mathcal{E}$ relative to $(S_o, \mathcal{S} \wedge S_o)$. Now, for any $\theta \in \Theta$,

$$\|P_\theta\rho\gamma - Q_\theta\sigma\| = \|P_\theta\rho\gamma - Q_\theta\sigma\tilde{\gamma}\gamma\| \leq \|P_\theta\rho - Q_\theta\sigma\tilde{\gamma}\| \, \|\gamma\| \leq \epsilon_\theta$$

since $\|\gamma\| = 1$.

This completes the proof.

Remark: In 4.6 we proved that ϵ-deficiency for $(k + 1)$-decision problems implies ϵ-deficiency for k-decision problems. This is a special case of the present proposition.

DEFINITION 7.2

A Polish space is a measurable space $(T,\mathcal{B})$ where T (together with a metric d) is a separabel complete metric space and $\mathcal{B}$ is the smallest σ-algebra which contains all open subsets of T. (The measurable sets are called Borel-sets).

The Polish spaces may be ordered with respect to cardinality into three classes:

(i) T is finite.

(ii) T is infinite countable.

(iii) card T = card $[0,1]$.

LEMMA 7.3

Let T be a compact metric space (e.g. $T = [0,1]$) and let $C(T)$ be the set of continuous real functions on T, provided with the metric dist $(f,g) = \sup_{t\in T}|f(t) - g(t)|$. Then $C(T)$ is separabel and there exists a dense, countable subset $\mathcal{H}$ of $C(T)$ such that if

$$r \in Q\ ,\quad f,g \in \mathcal{H} \quad \text{then}$$
$$r,\ |f|,\ f+g,\ rf \in \mathcal{H}$$

Proof: Since $C(T)$ is separabel, there exists a countable dense subset $\mathcal{U}_0 \subseteq C(T)$. We may assume that $0 \in \mathcal{U}_0$. We shall now recussively define countable sets $\mathcal{U}_1,\mathcal{U}_2,\ldots.$ such that $\mathcal{U}_0 \subseteq \mathcal{U}_1 \subseteq \ldots. \subseteq C(T)$ and let $\mathcal{H} = \bigcup_{i=0}^{\infty} \mathcal{U}_i$.

$\mathcal{H}$ is clearly dense and countable. Assume that $\mathcal{U}_i$ is defined. Then we put

$$\mathcal{U}_{i+1} = \{r_1f_1 + r_2f_2 + r_3 + f_3^+ : f_1,f_2,f_3 \in \mathcal{U}_i,\ r_1,r_2,r_3 \in Q\}$$

It is easily verified that $\mathcal{U}_0 \subseteq \mathcal{U}_1 \subseteq \mathcal{U}_2 \subseteq \ldots\ldots$ and that $\mathcal{H}$ has the properties listed in the lemma.

In the proof of theorem 8.5 we will need the famous <u>Riesz Representation Theorem</u> which is stated below. For a proof, we refer to theorem 2.14 of [13].

<u>THEOREM 7.4</u>

Let X be a locally compact Hausdorff space and let Λ be a non-negative linear functional on $C_c(X)$ (the set of continous real functions on X with compact support). Then there exists a σ-algebra $\mathfrak{M}$ in X which contains all Borel sets in X, and there exists a unique non-negative measure μ on $\mathfrak{M}$ which represents Λ in the sense that

$$\Lambda f = \int f d\mu \quad \text{for every} \quad f \in C_c(X) .$$

<u>THEOREM 7.5</u>

Let $\mathcal{E} = (\mathcal{X}, \mathcal{A}; P_\theta : \theta \in \Theta)$ and $\mathcal{F} = (\mathcal{Y}, \mathcal{B}; Q_\theta : \theta \in \Theta)$ be experiments and let ϵ be a non-negative function on Θ. Assume further that $\mathcal{E}$ is dominated. Then

$\mathcal{E}$ is ϵ-deficient relative to $\mathcal{F}$

if and only if to each decision space $(T, \mathcal{S})$ where T is a Borel-subset (i.e. measurable set) of a Polish space and $\mathcal{S}$ is the class of Borel-subsets of T, and to each decision-rule σ in $\mathcal{F}$ (relative to $(T, \mathcal{S})$) there corresponds a decision-rule ρ in $\mathcal{E}$ (relative to $(T, \mathcal{S})$) such that

$$\|P_\theta \rho - Q_\theta \sigma\| \le \epsilon_\theta \quad \text{for all} \quad \theta \in \Theta .$$

<u>Proof</u>: The "if"-part needs no proof, since the finite decision

spaces are included in the condition.

"only if": By prop. 7.1(i) it is enough to consider the Polish space itself. Assume therefore that T is a compact space (e.g. $T = [0,1]$). Let $\{t_1, t_2, \dots\}$ be a countable dense subset of T and define $T_k = \{t_1, \dots, t_k\}$; $k = 1,2,\dots$ Let $\Pi = \sum_{\theta \in \Theta_0} c_\theta P_\theta$ be a probability measure on $(\chi, \mathcal{A})$ such that $\Theta_0 \subseteq \Theta$ is countable and $\Pi \gg P_\theta$ for all $\theta \in \Theta$ (see theorem 12 of appendix A). Let $\mathcal{H}$ be given as in lemma 7.3. For each fixed k , we define a "projection" f_k from T to T_k . If $t \in T$, we define $f_k(t) = t_i$ where t_i is uniquely determined by the inequalities

$$d(t,t_1), \dots, d(t,t_{i-1}) > d(t,t_i)$$

$$d(t,t_{i+1}), \dots, d(t,t_k) \geq d(t,t_i)$$

(Intuitively, we let t_i be the member of T_k which minimizes the distance from t to t_i).

f_k is measurable, since it is determined by a set of inequalities between continuous functions.

Since $\{t_1, t_2, \dots\}$ is dense in T , $d(f_k(t),t) \downarrow 0$ when $k \to \infty$ (the convergence is clearly monotone). Now $d(f_k(t),t) = \text{dist}\,(T_k,t)$ which is known to be continuous in t . Thus, by Dini's lemma (theorem 7.13 of [12]) $d(f_k(t),t) \to 0$ <u>uniformly</u> in t . Let now σ be a decision-rule in $\mathcal{F}$ relative to $(T, \mathcal{S})$. For each k we define a decision-rule σ_k in $\mathcal{F}$ relative to T_k by

$$\sigma_k(t|^\circ) = \sigma(f_k^{-1}(\{t\})|\circ) \; ; \quad t \in T_k \; .$$

By assumption, there is for each k a decision-rule ρ_k in $\mathcal{E}$ relative to T_k such that

(3) $\|P_\theta \rho_k - Q_\theta \sigma_k\| \leq \epsilon_\theta$ for all $\theta \in \Theta$ (corollary 5.29).

The rest of the proof will be devoted to constructing a decision rule ρ on the basis of the ρ_k 's such that

$\|P_0\rho - Q_0\sigma\| \le \epsilon_0$ for all $0 \in \Theta$.

Let $f \in \mathcal{H}$. Then for each k

$$\rho_k(f|\circ) = \int f(t)\rho_k(dt|\circ) = \sum_{i=1}^{k} \rho_k(t_i|\circ)f(t_i)$$

defines a function from χ to R. f is bounded, since it is a continuous function on a compact space.

Hence, for fixed f, the sequence $\{\rho_k(f|\circ)\}$ is uniformly bounded and uniformly integrable with respect to the probability measure Π.

By appendix C, there is a subsequence $\{\rho_{k'}(f|\circ)\}$ which converges weakly to a function $\rho(f|\circ)$ (with respect to the probability space $(\chi, \mathcal{O}, \bar{\Pi})$).

Since $\mathcal{H}$ is countable we may apply Cantor's diagonal process to obtain a subsequence $\{\rho_{k''}\}$ such that $\rho_{k''}(f|\circ)$ converges weakly (Π) to a function $\rho(f|\circ)$ for each $f \in \mathcal{H}$.

Let $f, g \in \mathcal{H}$, $r \in Q$. Then ρ has the following properties:

(i) $\rho(f + g|\circ) = \rho(f|\circ) + \rho(g|\circ)$ a.e. $[\Pi]$

(ii) $\rho(rf|\circ) = r\rho(f|\circ)$ a.e. $[\Pi]$

(iii) $\rho(1|\circ) = 1$ a.e. $[\Pi]$

(iv) $\rho(f|\circ) \ge 0$ a.e. $[\Pi]$ whenever $f \ge 0$.

We prove the first assertion. The others follow in a similar way:

By definition,

$$\rho_{k''}(f + g|\circ) = \rho_{k''}(f|\circ) + \rho_{k''}(g|\circ)$$

The left side converges weakly to $\rho(f + g|\circ)$; the right side converges weakly to $\rho(f|\circ) + \rho(g|\circ)$. Since for each f, $\rho(f|\circ)$ is determined almost everywhere $[\Pi]$, (i) follows.

Since $\mathcal{H}$ is countable and Q is countable, the subset N of χ where (i) - (iv) fail to hold for some f,g or r has Π-measure zero. Hence by redefining ρ on N, ρ may be modified so that (i) - (iv) are valid everywhere.

We shall now define $\rho(f|\circ)$ for arbitrary $f \in C(T)$.

We assert that $|\rho(f|x)| \leq \|f\|$; $f \in \mathcal{H}$, $x \in \chi$. Choose $r \in Q$ such that

$$-r < -\|f\| \leq f \leq \|f\| < r .$$

Then $\rho(f|x) \leq \rho(r|x) = r\rho(1|x) = r$ and similarly, $\rho(f|x) \geq -r$.

Hence $|\rho(f|x)| \leq r$ for all $r > \|f\|$.

The assertion follows by letting $r \to \|f\|$.

Consequently, the mapping $f \to \rho(f|\circ)$ is a contraction. In particular, each Cauchy-sequence in $\mathcal{H}$ will be mapped into a Cauchy-sequence of real functions on χ.

Let now $g \in C(T)$. Then there is a sequence $\{f_n\}$ in $\mathcal{H}$ such that $f_n \to g$.

We define $\rho(g|x) = \lim_{n \to \infty} \rho(f_n|x)$ for each $x \in \chi$.

The limit exists, since $\{p(f_n|x)\}$ by the above remark is a Cauchy-sequence in R. A straightforward verification also shows that the limit defining $\rho(g|\circ)$ is independent of our choice of $\{f_n\}$. Next, by the properties of limits of functions, (i) - (iv) are seen to hold for arbitrary $f,g \in C(T)$, $r \in R$.

Hence, for each $x \in \chi$, $\rho(\circ|x)$ is a non-negative linear functional on $C(T)$. By theorem 7.4, $\rho(\circ|x)$ may be represented by a non-negative measure $\overline{\rho}(\circ|x)$ on a σ-algebra $\mathcal{M} \supseteq \mathcal{S}$. $\overline{\rho}(\circ|x)$ is a probability measure by property (iii).

Since $\rho(f|\circ)$ is a measurable function on χ for each

$f \in C(T)$, it follows that $\bar{\rho}(S|\circ)$ is measurable for each $S \in \mathcal{S}$. Hence $\bar{\rho}$ is a decision-rule in $\mathcal{E}$ relative to $(T,\mathcal{S})$ and for any $f \in C(T)$,

$$\rho(f|\circ) = \int f(t)\bar{\rho}(dt|\circ) .$$

We may without ambiguity write ρ instead of $\bar{\rho}$.

It remains to prove that

$|P_0\rho f - Q_0\sigma f| \leq \epsilon_0\|f\|$ for all $f \in \mathcal{H}$.

Since any member of $C(T)$ and hence any measurable function on T may be approximated by members of $\mathcal{H}$, this will imply that

$|P_0\rho f - Q_0\sigma f| \leq \epsilon_0\|f\|$ for all bounded measurable functions f on T and hence by def. 10 of appendix B,

$$\|P_0\rho - Q_0\sigma\| \leq \epsilon_0 .$$

For any $f \in \mathcal{H}$, $0 \in \Theta$,

$$|P_0\rho f - Q_0\sigma f| \leq |P_0\rho f - P_0\rho_{k''}f|$$
$$+ |P_0\rho_{k''}f - Q_0\sigma_{k''}f| + |Q_\theta\sigma_{k''}f - Q_0\sigma f|$$

The second term on the right hand side is by (3) $\leq \epsilon_0\|f\|$.

Hence it suffices to prove that the two remaining terms tend to zero as $k'' \to \infty$.

Put $h_0 = dP_0/d\Pi$. Then

$$|P_0\rho f - P_0\rho_{k''}f| = |\int\rho(f|x)P_\theta(dx) - \int\rho_{k''}(f|x)P_\theta(dx)|$$
$$= |\int\rho(f|x)h_0(x)\Pi(dx) - \int\rho_{k''}(f|x)h_0(x)\Pi(dx)|$$

which tends to zero by weak compactness (consider the definition of ρ).

For each y , $\sigma_k(\circ|y) = \sigma(\circ|y)f_k^{-1}$
(by the usual notation for induced measures).

Hence, by the well-known formula

$\int f\circ X dP = \int f dPX^{-1}$, we get

$(\sigma_k f)(y) = \int f(t)\sigma_k(dt|y) = \int f(f_k(t))\sigma(dt|y)$

Finally,

$$|Q_\theta\sigma_{k''}f - Q_\theta\sigma f| = |\int(\sigma_{k''}f - \sigma f)dQ_\theta|$$
$$\leq \|\sigma_{k''}f - \sigma f\| = \|\int(f(f_{k''}(t)) - f(t))\sigma(dt|\circ)\|$$
$$\leq \sup_t |f(f_{k''}(t)) - f(t)|$$

which tends to zero when $k'' \to \infty$ since f is uniformly continuous and $d(f_k(t),t) \to 0$ uniformly in t . The proof is now complete.

THEOREM 7.6 (THE MARKOV KERNEL CRITERION)

Let $\mathcal{E} = (\chi,\mathcal{A}; P_\theta : \theta \in \Theta)$ and $\mathcal{F} = (\mathcal{Y},\mathcal{B}; Q_\theta : \theta \in \Theta)$ be experiments and let ϵ be a non-negative function on Θ . Assume further that $\mathcal{E}$ is dominated and that $\mathcal{Y}$ is a Borel-subset of a complete, separabel metric space and $\mathcal{B}$ is the class of Borel-subsets of $\mathcal{Y}$.

Then $\mathcal{E}$ is ϵ-deficient relative to $\mathcal{F}$ if and only if there exists a Markov-kernel $M : \mathcal{B} \times \chi \to [0,1]$ such that

$$\|P_\theta M - Q_\theta\| \leq \epsilon_\theta \quad \text{for all } \theta \in \Theta .$$

Proof: "if": Let $(T,\mathcal{S})$ be an arbitrary decision space and let σ be a decision-rule in $\mathcal{F}$ relative to $(T,\mathcal{S})$. Put $\rho = M\sigma$. Then ρ is a decision-rule in $\mathcal{E}$ relative to $(T,\mathcal{S})$ and for any $\theta \in \Theta$

$\|P_\theta\rho - Q_\theta\sigma\| = \|P_\theta M\sigma - Q_\theta\sigma\| \leq \|P_\theta M - Q_\theta\| \leq \epsilon_\theta$.

"only if": Put $(T,\mathcal{S}) = (\mathcal{Y},\mathcal{B})$ and let $\sigma(S|y) = I_S(y)$ for all $S \in \mathcal{S}$, $y \in \mathcal{Y}$. Then $Q_\theta\sigma = Q_\theta$.

By theorem 7.5 there is a Markov-kernel $M : \mathcal{S} \times \chi \to [0,1]$, i.e. $M : \mathcal{B} \times \chi \to [0,1]$ such that $\|P_\theta M - Q_\theta\sigma\| \leq \epsilon_\theta$.

The theorem follows.

REMARK 7.7

We note that in the proof of the "if"-part, we did not make use of the given structure of $(\mathcal{Y},\mathcal{B})$. However, the conclusion follwed from corollary 5.29, so we needed the requirement that $\mathcal{E}$ be dominated. It is proved in appendix B of [19] that if there exists a σ-finite measure μ on $(\mathcal{Y},\mathcal{B})$ such that $\mu >> Q_\theta$ for all $\theta \in \Theta$, then M may be chosen so that $\mu >> P_\theta M$ for all $\theta \in \Theta$.

COROLLARY 7.8

Let $\mathcal{E}$ and $\mathcal{F}$ be given as in theorem 7.6. Then $\mathcal{E} \geq \mathcal{F}$ if and only if there exists a Markov-kernel M on $\mathcal{B} \times \chi$ such that

(4) $P_\theta M = Q_\theta$ for all $\theta \in \Theta$.

Remark: Statement (4) asserts that, if we observe the result of experiment $\mathcal{E}$ and, when $x \in \chi$ is observed we select $y \in \mathcal{Y}$ according to the distribution $M(\cdot|x)$ on $(\mathcal{Y},\mathcal{B})$, then the resulting experiment is (in some sense) identical with the experiment $\mathcal{F}$. In other words, (4) states that $\mathcal{F}$ may be duplicated from $\mathcal{E}$ with the aid of e.g. a table of random numbers.

Corollary 7.8 thus implies that $\mathcal{E}$ is more informative than $\mathcal{F}$ if and only if $\mathcal{F}$ may be duplicated from $\mathcal{E}$.

EXAMPLE 7.9

Consider again example 5.6 and let $\mathcal{F} = (\mathcal{Y},\mathcal{B}; Q_\theta : \theta \in \Theta)$ be an experiment where $\mathcal{Y} = \{1,\ldots,k\}$, $\mathcal{B}$ = class of subsets of $\mathcal{Y}$, $\Theta = \{1,\ldots,s\}$ and the Q_θ 's are given by the Markov-matrix

$$Q_{\mathcal{F}} = \begin{pmatrix} q_{11} & \cdots & q_{1k} \\ \vdots & & \vdots \\ q_{s1} & \cdots & q_{sk} \end{pmatrix}$$

A Markov-kernel $M : \mathcal{B} \times \chi \to [0,1]$ is now given by a $(r \times k)$ Markov-matrix $M = (m_{ij})$ where

$$M(\{j\}|i) = m_{ij} \ ; \quad i = 1,\ldots,r \ , \quad j = 1,\ldots,k \ .$$

For any $\theta = 1,\ldots,s$

$$P_\theta M = \Big(\sum_{i=1}^{r} p_{\theta i} m_{i1}, \ldots, \sum_{i=1}^{r} p_{\theta i} m_{ik}\Big)$$

If $A = (a_{ij})$ is a $(m \times n)$-matrix, we define the <u>norm</u> of A, denoted $\|A\|$, by

$$\|A\| = \max_i \sum_j |a_{ij}|$$

Hence, it is seen that if $\mathcal{E}$ and $\mathcal{F}$ are experiments as given above, then

$$\delta(\mathcal{E},\mathcal{F}) = \inf_M \|P_{\mathcal{E}} M - Q_{\mathcal{F}}\|$$

where infimum is taken over all $(r \times k)$-Markov matrices M. Hence, for any such M,

(5) $\delta(\mathcal{E},\mathcal{F}) \leq \|P_{\mathcal{E}} M - Q_{\mathcal{F}}\|$

Thus the Markov kernel criterion is useful in order to achieve <u>upper bounds</u> for $\delta(\mathcal{E},\mathcal{F})$ or $\Delta(\mathcal{E},\mathcal{F})$. (We remember from example 5.15 that the ψ-criterion gives rise to lower bounds). In particular, if $k = r$ and M is the $(r \times r)$-identity matrix, then by (5),

$$\Delta(\mathcal{E},\mathcal{F}) \leq \|P_{\mathcal{E}} - Q_{\mathcal{F}}\| = \max_i \sum_{j=1}^{r} |p_{ij} - q_{ij}|$$

This is the same result as was obtained in (9) of example 5.15.

Finally, it follows that $\mathcal{E} \geq \mathcal{F}$ if and only if there exists a $(r \times k)$ Markov-matrix so that

$P_{\mathcal{E}} M = Q_{\mathcal{F}}$.

EXAMPLE 7.10

There is given a population of 10 members. It is known that 5 of the members possess the property A, and that 5 possess the property B, but the number of members having both properties is not known. Furthermore, the members that have property A are known and may easily be selected from the others. Two sampling plans are proposed in order to obtain information about the number of members having both properties.

(a) 3 of the 5 members with property A are chosen at random and the number X having property B is noted.

(b) 3 of the 10 members of the population are chosen at random and the number Y having both properties is noted.

The sampling plans (a) and (b) may be considered as finite experiments $\mathcal{E}$ and $\mathcal{F}$ where $\Theta = \{0,1,2,3,4,5\}$ and the parameter $\theta \in \Theta$ is the number of members having both properties A and B.

For fixed $\theta \in \Theta$, X and Y are hypergeometric distributed, so the Markov-matrices defining $\mathcal{E}$ and $\mathcal{F}$ are easily found:

$$P_{\mathcal{E}} = \begin{array}{c|cccc} \theta \backslash X & 0 & 1 & 2 & 3 \\ \hline 0 & 1 & 0 & 0 & 0 \\ 1 & \frac{2}{5} & \frac{3}{5} & 0 & 0 \\ 2 & \frac{1}{10} & \frac{6}{10} & \frac{3}{10} & 0 \\ 3 & 0 & \frac{3}{10} & \frac{6}{10} & \frac{1}{10} \\ 4 & 0 & 0 & \frac{3}{5} & \frac{2}{5} \\ 5 & 0 & 0 & 0 & 1 \end{array}$$

$$Q_{\mathcal{F}} = \begin{array}{c|cccc} \theta \backslash Y & 0 & 1 & 2 & 3 \\ \hline 0 & 1 & 0 & 0 & 0 \\ 1 & \frac{7}{10} & \frac{3}{10} & 0 & 0 \\ 2 & \frac{7}{15} & \frac{7}{15} & \frac{1}{15} & 0 \\ 3 & \frac{35}{120} & \frac{63}{120} & \frac{21}{120} & \frac{1}{120} \\ 4 & \frac{5}{30} & \frac{15}{30} & \frac{9}{30} & \frac{1}{30} \\ 5 & \frac{1}{12} & \frac{5}{12} & \frac{5}{12} & \frac{1}{12} \end{array}$$

If we put $M = \begin{pmatrix} 1 & 0 & 0 & 0 \\ \frac{1}{2} & \frac{1}{2} & 0 & 0 \\ \frac{2}{9} & \frac{5}{9} & \frac{2}{9} & 0 \\ \frac{1}{12} & \frac{5}{12} & \frac{5}{12} & \frac{1}{12} \end{pmatrix}$

then a simple computation shows that

$$P_{\mathcal{E}} M = Q_{\mathcal{F}} \ .$$

Hence $\mathcal{E}$ is more informative than $\mathcal{F}$, and we should prefer the sampling plan (a) to the plan (b). This is reasonable from the fact that the plan (a) takes into account the prior information ; of knowing the elements having property A.

<u>EXAMPLE 7.11</u>

We return to example 5.17.

Clearly $\mathcal{E}_n \geq \mathcal{E}_\infty$, so $\delta(\mathcal{E}_n, \mathcal{E}_\infty) = 0$.

By example 7.9,

$$\delta(\mathcal{E}_\infty, \mathcal{E}_n) = \inf_M \|AM - P^n\|$$

Let now $M = \begin{pmatrix} 1-a & a \\ b & 1-b \end{pmatrix}$ and put $\rho = 1-\alpha-\beta$

Then

$$AM - P^n = \frac{1}{\alpha+\beta}\begin{pmatrix} -\alpha & \alpha \\ \beta & -\beta \end{pmatrix}\rho^n + \frac{1}{\alpha+\beta}\begin{pmatrix} -a\beta+b\alpha & a\beta-b\alpha \\ -a\beta+b\alpha & a\beta-b\alpha \end{pmatrix}$$

Hence

$$(6)\ \|AM - P^n\| = \frac{2}{\alpha+\beta}\{|-\alpha\rho^n - a\beta + b\alpha| \vee |-\beta\rho^n + a\beta - b\alpha|\}$$

We minimize the expression (6) with respect to $\gamma = a\beta - b\alpha$. By examining the graphs of the functions (of γ) $|-\alpha\rho^n - \gamma|$ and $|-\beta\rho^n + \gamma|$ it is seen that minimum occurs when $\gamma = \frac{\beta-\alpha}{2}\rho^n$. Next, it is seen that numbers a,b with $0 \leq a,b \leq 1$ may be found such that $a\beta - b\alpha = \frac{\beta-\alpha}{2}\rho^n$. Substituting the actual M into (6) yields

$\|AM - P^n\| = |\rho^n|$.

Hence $\delta(\mathcal{E}_\infty, \mathcal{E}_n) = |\rho|^n = |1-\alpha-\beta|^n$

Further examples on the use of the Markov kernel criterion may be found in [5] and [18].

8. SUFFICIENCY

We now turn to the concept of sufficiency. We shall give the classical definition of sufficiency (CE-sufficiency) and the definition of sufficiency in terms of equivalent experiments (Δ - sufficiency), as well as investigate the relation between them. It will be proved that Δ - sufficiency is equivalent to CE-sufficiency if the experiment is dominated.

It is assumed that the reader is familiar with the concept of conditional expectation given a sub-σ-algebra. We will, however, give the definition. For a rigorous treatment, we refer to [9]. An introduction to the concept of sufficiency is given in [6] and [14].

In this chapter, we will consider experiments with arbitrary parameter set Θ.

DEFINITION 8.1

Let $(\chi,\mathcal{A},P)$ be a probability space and let $\mathcal{B}$ be a sub-σ-algebra of $\mathcal{A}$. Let X be an $\mathcal{A}$-measurable bounded, or non-negative real function on χ. Then the conditional expectation of X given $\mathcal{B}$ is defined as the unique (a.e.[P]) $\mathcal{B}$-measurable function $E^{\mathcal{B}}X$ such that

$$\int_B E^{\mathcal{B}} X dP = \int_B X dP \quad \text{for all } B \in \mathcal{B} .$$

(the existence and uniqueness follow from Radon-Nikodym's theorem).

If $X = I_A$, we may replace "expectation of X " by "probability of A " and write $P^{\mathcal{B}}(A)$ instead of $E^{\mathcal{B}} I_A$.

DEFINITION 8.2

Let $\mathcal{E} = (\chi, \mathcal{A}; P_\theta : \theta \in \Theta)$ be an experiment. Then a sub-σ-algebra $\mathcal{B}$ of $\mathcal{A}$ is said to be CE-sufficient for $\mathcal{E}$ if corresponding to each $A \in \mathcal{A}$ there exists a $\mathcal{B}$-measurable function Y_A such that

$$P_\theta^{\mathcal{B}}(A) = Y_A \quad \text{a.e.} \quad [P_\theta] \quad \text{for all} \quad \theta \in \Theta .$$

This condition is equivalent to the following:
To each bounded or non-negative $\mathcal{A}$-measurable real function on χ there corresponds a $\mathcal{B}$-measurable function Y_Z such that

$$E_\theta^{\mathcal{B}} Z = Y_Z \quad \text{a.e.} \quad [P_\theta] \quad \text{for all} \quad \theta \in \Theta .$$

Remark: CE stands for Conditional Expectation. The essence of the requirements is that the conditional probability of an event A (respectively conditional expectation of a random variable Z) may be specified (almost) independent of the (unknown) parameter $\theta \in \Theta$.

NOTATION

If P is a probability-measure on a measurable space $(\chi, \mathcal{A})$, and $\mathcal{B}$ is a sub-σ-algebra of $\mathcal{A}$, then $P_{\mathcal{B}}$ denotes the restriction of P to $\mathcal{B}$.

DEFINITION 8.3

Let $\mathcal{E} = (\chi, \mathcal{A}; P_\theta : \theta \in \Theta)$ be an experiment and let $\mathcal{B}$ be a sub-σ-algebra of $\mathcal{A}$. Define $\mathcal{F} = (\chi, \mathcal{B}; P_{\theta\mathcal{B}} : \theta \in \Theta)$. Then we say that $\mathcal{B}$ is Δ-sufficient for $\mathcal{E}$ if $\mathcal{E} \sim \mathcal{F}$.

Remark: Obviously $\mathcal{E} \geq \mathcal{F}$. (This follows from the fact that $\mathcal{B}$-measurability implies $\mathcal{A}$-measurability, so we can take $\rho = \sigma$ in def. 4.1.)

Hence $\mathcal{B}$ is Δ-sufficient for $\mathcal{E}$ if and only if $\mathcal{F} \geq \mathcal{E}$.

The σ-algebra $\mathcal{A}$ is interpreted as the set of <u>events</u> relative to the experiment $\mathcal{E}$. After the experiment is performed, we may for each $A \in \mathcal{A}$ decide whether the event A has occurred or not. Suppose now that $\mathcal{B} \subseteq \mathcal{A}$ and that we may only observe which events $B \in \mathcal{B}$ that occur. This corresponds to observing the experiment $\mathcal{F}$. It turns out that if $\mathcal{B}$ is Δ-sufficient, then we lose no information by restricting attention to the events $B \in \mathcal{B}$. We say that $\mathcal{F}$ defines a reduction of $\mathcal{E}$.

<u>PROPOSITION 8.4</u>

If $\mathcal{B}$ is CE-sufficient for $\mathcal{E}$, then $\mathcal{B}$ is also Δ-sufficient, for $\mathcal{E}$.

<u>Proof:</u> Let $\mathcal{F}$ be as defined in 8.3. We must prove that $\mathcal{F} \geq \mathcal{E}$.

Let ρ be a decision-rule in $\mathcal{E}$ relative to the decision space $T_k = \{1,\ldots,k\}$. Define $\tilde{\rho}$ on $T_k \times \chi$ by

$$\tilde{\rho}(t|\circ) = E_\theta^{\mathcal{B}} \rho(t|\circ) \quad \text{for each} \quad t \in T_k \ .$$

Since $\mathcal{B}$ is CE-sufficient, $\tilde{\rho}$ may be specified independent of θ . Furthermore, $\tilde{\rho}(t|\circ)$ is $\mathcal{B}$-measurable for each t . Hence $\tilde{\rho}$ is a decision-rule in $\mathcal{F}$. If $\{L_\theta(t) \ : \ \theta \in \Theta \ ,$ $t \in T_k\}$ is a loss-function, then for any $\theta \in \Theta$

$$\begin{aligned} P_\theta \rho L_\theta &= \sum_{t=1}^{k} L_\theta(t) \int \rho(t|\circ) dP_\theta \\ &= \sum_{t=1}^{k} L_\theta(t) \int E_\theta^{\mathcal{B}} \rho(t|\circ) dP_\theta \\ &= \sum_{t=1}^{k} L_\theta(t) \int \tilde{\rho}(t|\circ) dP_\theta \ = P_{\theta\mathcal{B}} \tilde{\rho} L_\theta \ . \end{aligned}$$

Consequently $\mathcal{F} \geq \mathcal{E}$.

<u>Remark:</u> It may be shown that in general, Δ-sufficiency will not imply CE-sufficiency. We shall now prove, however, that

the implication holds if $\mathcal{E}$ is dominated.

LEMMA 8.5

Let $(\chi, \mathcal{A}; P, Q)$ be an experiment and assume $P \gg Q$.
Then $E_P^{\mathcal{B}}(dQ/dP) = dQ_{\mathcal{B}}/dP_{\mathcal{B}}$

Proof: Clearly $P_{\mathcal{B}} \gg Q_{\mathcal{B}}$ so $dQ_{\mathcal{B}}/dP_{\mathcal{B}}$ is well-defined.
By def. 8.1, for any $B \in \mathcal{B}$ we have

$$\int_B E_P^{\mathcal{B}}(dQ/dP)\,dP_{\mathcal{B}} = \int_B (dQ/dP)\,dP$$
$$= \int_B dQ = Q(B) = Q_{\mathcal{B}}(B) .$$

The lemma follows.

LEMMA 8.6

Let $(\chi, \mathcal{A}, P)$ be a probability space. Let X and Y be random variables on $(\chi, \mathcal{A}, P)$ such that $\mathcal{L}(X) = \mathcal{L}(Y)$ and $X = E^{\mathcal{B}}Y$ a.e. for a sub-σ-algebra $\mathcal{B}$ of $\mathcal{A}$. Assume further that $E|Y| < \infty$. Then $X = Y$ a.e.

Proof: Assume first that $EY^2 < \infty$

Then $E(X-Y)^2 = EE^{\mathcal{B}}(Y-X)^2 = EE^{\mathcal{B}}(Y-E^{\mathcal{B}}Y)^2$
$= EVar^{\mathcal{B}}Y = E(E^{\mathcal{B}}Y^2 - (E^{\mathcal{B}}Y)^2)$
$=E(E^{\mathcal{B}}Y^2-X^2) = EY^2 - EX^2 = 0$

since X and Y are identically distributed. Hence $X = Y$ a.s.

Reject now the assumption that $EY^2 < \infty$. Let φ be a real valued continuous, convex function defined on an interval I such that $Y \in I$ a.s.

By assumption $E\varphi(Y) = E\varphi(X)$.

By Jensen's inequality (which is valid also for conditional expectations)

(1) $E^{\mathcal{B}}\varphi(Y) \geq \varphi(E^{\mathcal{B}}Y) = \varphi(X)$ a.s.

Since $EE^{\mathcal{B}}\varphi(Y) = E\varphi(Y) = E\varphi(X)$, equality must hold in (1), so

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

Hence $\mathcal{L}(E^{\mathcal{B}}\varphi(Y)) = \mathcal{L}(\varphi(X)) = \mathcal{L}(\varphi(Y))$.

In particular

$$\mathcal{L}(E^{\mathcal{B}}Y^{\pm}) = \mathcal{L}(Y^{\pm}) .$$

It follows that we may, without loss of generality, assume $Y \geq 0$.

Since the function φ defined by $\varphi(t) = -\sqrt{t}$ is convex on $[0,\infty[$ we have

$$\mathcal{L}(E^{\mathcal{B}}\sqrt{Y}) = \mathcal{L}(\sqrt{Y})$$

Hence, since $E(\sqrt{Y})^2 = EY < \infty$ it follows from the first part of the proof that

$$E^{\mathcal{B}}\sqrt{Y} = \sqrt{Y} \quad \text{a.s.}$$

Finally,

$$X = E^{\mathcal{B}}Y = E^{\mathcal{B}}(\sqrt{Y})^2 = E^{\mathcal{B}}(E^{\mathcal{B}}\sqrt{Y})^2 = (E^{\mathcal{B}}\sqrt{Y})^2 = Y \quad \text{a.s.}$$

The second last equality sign follows since $(E^{\mathcal{B}}\sqrt{Y})^2$ is $\mathcal{B}$-measurable.

<u>LEMMA 8.7</u>

Assume that $\mathcal{B}$ is Δ-sufficient for $\mathcal{E}$ and that $\mathcal{E}$ is dominated. Let π be given as in theorem 12 of appendix A and let θ be a fixed member of Θ .

Then the experiments (dichotomies) $(\chi,\mathcal{A}; P_\theta,\pi)$ and $(\chi,\mathcal{B}; P_{\theta\mathcal{B}},\pi\;\;)$ are equivalent.

<u>Proof:</u> By corollary 6.11 it suffices to prove that

(2) $\|aP_\theta + b\pi\| = \|aP_{\theta\mathcal{B}} + b\pi_{\mathcal{B}}\|$ for all $a,b \in R$.

We have $\pi = \sum_{j=1}^{\infty} c(\theta_j)P_{\theta_j}$ for some countable subset$=\{\theta_1,\theta_2,\dots\}\subseteq\Theta$

By corollary 6.11 and 8.2, for each $n = 1,2,\ldots;$

$$\|aP_\theta + b \sum_{j=1}^{n} c(\theta_j)P_{\theta_j}\| = \|aP_{\theta\mathcal{B}} + b\sum_{j=1}^{n} c(\theta_j)P_{\theta_j\mathcal{B}}\|$$

(2) follows by letting $n \to \infty$.

PROPOSITION 8.8

Let $\mathcal{E} = (\chi, \mathcal{A}, P_\theta : \theta \in \Theta)$ be a dominated experiment and let π be a probability measure on $(\chi, \mathcal{A})$ such that

$\pi = \sum_{\theta\in\Theta_0} c(\theta)P_\theta$ for some countable subset $\Theta_0 \subseteq \Theta$ and so

that $\sum_\theta c(\theta) = 1$ and $\pi >> P_\theta$ for all $\theta \in \Theta$ (the existence of π is proved in theorem 12 of appendix A).

Let $\mathcal{B}$ be a sub-σ-algebra of $\mathcal{A}$. Then $\mathcal{B}$ is Δ-sufficient for $\mathcal{E}$ if and only if

$dP_\theta/d\pi$ may be specified $\mathcal{B}$-measurable for each $\theta \in \Theta$.

Proof: Assume first that $\mathcal{B}$ is Δ-sufficient. Fix $\theta \in \Theta$.

Consider the equivalent experiments given in lemma 8.7. Let

$f_\theta = dP_\theta/d\pi$, $\tilde{f}_\theta = dP_{\theta\mathcal{B}}/d\pi$.

By lemma 8.5, $\tilde{f}_\theta = E_\pi^{\mathcal{B}} f_\theta$.

It is readily verified that

$$(3) \quad \frac{dP_\theta}{d(P_\theta+\pi)} = \frac{f_\theta}{1+f_\theta}, \quad \frac{d\pi}{d(P_\theta+\pi)} = \frac{1}{1+f_\theta}$$

and that similar expressions hold if P_θ and π are replaced by $P_{\theta\mathcal{B}}$ and $\pi_{\mathcal{B}}$.

Since equivalent experiments have the same standard measure, it follows from def. 5.22 that for any Borel-subset $V \subseteq R^2$,

$$(4) \quad (P_\theta + \pi)\left(\left(\frac{f_\theta}{1+f_\theta}, \frac{1}{1+f_\theta}\right) \in V\right) = (P_{\theta\mathcal{B}} + \pi_{\mathcal{B}})\left(\left(\frac{\tilde{f}_\theta}{1+\tilde{f}_\theta}, \frac{1}{1+\tilde{f}_\theta}\right) \in V\right)$$

(Rather than $\left(\left(\frac{f_\theta}{1+f_\theta}, \frac{1}{1+f_\theta}\right) \in V\right)$ we should write

$\{x \in \chi : (\frac{f_\theta(x)}{1+f_\theta(x)}, \frac{1}{1+f_\theta(x)}) \in V\}$)

Clearly, for any $s \in R$,

$$f_\theta \leq s \Leftrightarrow (\frac{f_\theta}{1+f_\theta}, \frac{1}{1+f_\theta}) \in \langle -\infty, \frac{s}{1+s}] \times [\frac{1}{1+s}, \infty\rangle .$$

Thus (4) implies that

$(P_\theta + \pi)(f_\theta \leq s) = (P_{\theta\mathcal{B}} + \pi_{\mathcal{B}})(\tilde{f}_\theta \leq s)$ for all $s \in R$

or equivalently

$\int I_{\langle-\infty,s]}(f_\theta)d(P_\theta + \pi) = \int I_{\langle-\infty,s]}(\tilde{f}_\theta)d(P_{\theta\mathcal{B}} + \pi_{\mathcal{B}})$; $s \in R$.

Hence $\int\varphi(f_\theta)d(P_\theta + \pi) = \int\varphi(\tilde{f}_\theta)d(P_{\theta\mathcal{B}} + \pi_{\mathcal{B}})$ for any bounded, Borel-measurable function $\varphi : R \to R$.

By (3),

$\int\varphi(f_\theta)(1 + f_\theta)d\pi = \int\varphi(\tilde{f}_\theta)(1 + \tilde{f}_\theta)d\pi_{\mathcal{B}}$

which is equivalent to

$\int h(f_\theta)d\pi = \int h(\tilde{f}_\theta)d\pi$

for all bounded, Borel-measurable functions h .

In particular, if $h = I_C$ for a Borel-set $C \subseteq R$, then we get

$$\pi(f_\theta \in C) = \pi_{\mathcal{B}}(\tilde{f}_\theta \in C) = \pi(\tilde{f}_\theta \in C) .$$

Consequently f_θ and $\tilde{f}_\theta$ have the same distribution relative to the probability space $(\chi,\mathcal{A},\pi)$.

Since $\tilde{f}_\theta = E^{\mathcal{B}}f_\theta$ it thus follows from lemma 8.6 that

$$f_\theta = \tilde{f}_\theta \quad \text{a.e.}$$

Hence, since $\tilde{f}_\theta$ is $\mathcal{B}$-measurable, $f_\theta = dP_\theta/d\pi$ may be specified $\mathcal{B}$-measurable.

Conversely, assume that $f_\theta = dP_\theta/d\pi$ is $\mathcal{B}$-measurable for each θ . Let $\mathcal{F}$ be given as in def. 8.3. We shall prove that $\mathcal{E} \sim \mathcal{F}$. By corollary 5.28 it is enough to prove that $\psi(\mathcal{E}_F) = \psi(\mathcal{F}_F)$ for all $\psi \in \Psi$ and all finite subsets $F \subseteq \Theta$, $F \neq \emptyset$. Clearly $f_\theta = \tilde{f}_\theta$ a.e. for all θ .

Let now $F = \{\theta_1, \ldots, \theta_s\}$ and let $\psi \in \Psi$ be defined on R^s. Then, by def. 5.4,

$$\psi(\mathcal{E}_F) = \int \psi(f_{\theta_1}, \ldots, f_{\theta_s}) d\pi = \int \psi(\tilde{f}_{\theta_1}, \ldots, \tilde{f}_{\theta_s} d\pi = \psi(\mathcal{F}_F)$$

and we are done.

PROPOSITION 8.9

Let $\mathcal{E}$ be a dominated experiment and assume that $\mathcal{B}$ is Δ-sufficient for $\mathcal{E}$. Then $\mathcal{B}$ is CE-sufficient for $\mathcal{E}$.

Proof: Let π be given as in prop. 8.7 and let $f_\theta = dP_\theta/d\pi$; $\theta \in \Theta$. For each $A \in \mathcal{A}$, let

$$Y_A = \pi^{\mathcal{B}}(A) .$$

Then for any $\theta \in \Theta$, $B \in \mathcal{B}$

$$\int_B Y_A dP_\theta = \int_B \pi^{\mathcal{B}}(A) f_\theta d\pi .$$

By definition, $\int_B \pi^{\mathcal{B}}(A) d\pi = \int_B I_A d\pi$ for all $B \in \mathcal{B}$.

Since f_θ is $\mathcal{B}$-measurable (prop. 8.8), it follows that

$$\int_B \pi^{\mathcal{B}}(A) f_\theta d\pi = \int_B I_A f_\theta d\pi$$

and hence $\int_B Y_A dP_\theta = \int_B I_A f_\theta d\pi = \int_B I_A dP_\theta$.

Since Y_A is $\mathcal{B}$-measurable, it now follows from definition 8.1 that

$$Y_A = P_\theta^{\mathcal{B}}(A) \quad \text{a.e.} \quad [P_\theta] .$$

Finally, since θ was arbitrary, the proposition follows from def. 8.2.

COROLLARY 8.10

If $\mathcal{E} = (\chi, \mathcal{A}; P_\theta : \theta \in \Theta)$ is a dominated experiment, then a sub-σ-algebra $\mathcal{B}$ of $\mathcal{O}$ is CE-sufficient if and only if $\mathcal{B}$ is Δ-sufficient. Hence, in the case of dominated experiments, we may without ambiguity use the term <u>sufficiency</u> instead of CE- or Δ-sufficiency.

<u>DEFINITION 8.11</u>

Let $\mathcal{E}$ and $\mathcal{F}$ be given as in def. 8.3. Then $\mathcal{B}$ is said to be <u>pairwise sufficient</u> for $\mathcal{E}$ if

$$\mathcal{E}_{\{\theta_1,\theta_2\}} \sim \mathcal{F}_{\{\theta_1,\theta_2\}}$$

for all pairs $(\theta_1,\theta_2) \in \Theta \times \Theta$.

<u>PROPOSITION 8.12</u>

Let $\mathcal{E}$ be a dominated experiment. Then $\mathcal{B}$ is sufficient for $\mathcal{E}$ if and only if $\mathcal{B}$ is pairwise sufficient for $\mathcal{E}$.

<u>Proof:</u> It suffices to prove the "if"-part. We assume that

$$\mathcal{E}_{\{\theta_1,\theta_2\}} \sim \mathcal{F}_{\{\theta_1,\theta_2\}} \quad \text{for all} \quad (\theta_1,\theta_2) \in \Theta \times \Theta$$

and we shall prove that $\mathcal{E} \sim \mathcal{F}$.

By corollary 5.28 we may assume that Θ is finite, say $\Theta = \{1,\dots,s\}$.

Let $\pi = \frac{1}{s}\sum_{\theta=1}^{s} P_\theta$. By prop. 8.8 it is enough to prove that

$dP_\theta/d\pi$ may be specified $\mathcal{B}$-measurable for $\theta = 1,\dots,s$.

For simplicity, we let $\theta = 1$ in the proof. Define

$$\frac{dP_1}{d\frac{1}{2}(P_1+P_i)} = h_i \quad ; \quad i = 1,\dots,s \ .$$

By assumption (and prop. 8.8), each h_i may be specified $\mathcal{B}$-measurable.

Let $N = \bigcup_{i=1}^{s} [h_i = 0]$

Clearly $N \in \mathcal{B}$, and $P_1(N) = 0$ since $P_1(h_i = 0) = 0$ $i = 1,\ldots,s$.

Hence we may put $dP_1/d\pi = 0$ on N .

We consider now $N^c = \bigcap_{i=1}^{s} [h_i > 0]$.

By Radon-Nikodym,

$$\frac{d(P_1 + P_i)}{dP_1} = \frac{2}{h_i} \quad \text{on} \quad N^c$$

Hence

$$\sum_{i=1}^{s} \frac{d(P_1 + P_i)}{dP_1} = \sum_{i=1}^{s} \frac{2}{h_i} \quad \text{on} \quad N^c .$$

But the left side is equal to

$$\frac{d(sP_1 + \Sigma P_i)}{dP_1} = s + \frac{d\Sigma P_i}{dP_1} = s + s\frac{d\pi}{dP_1} .$$

Hence $\dfrac{d\pi}{dP_1} = \dfrac{1}{s}\sum_i \dfrac{2}{h_i} - 1$ on N^c

so $\dfrac{dP_1}{d\pi} = (\dfrac{1}{s}\sum_i \dfrac{2}{h_i} - 1)^{-1}$ on N^c .

Thus $\dfrac{dP_1}{d\pi}$ is $\mathcal{B}$ -measurable, since on N^c it may be written as a continuous function of $\mathcal{B}$ -measurable functions.

We sum up the results obtained so far in the following theorem:

<u>THEOREM 8.13</u>

Let $\mathcal{E} = (\chi, \mathcal{A}, P_\theta \; : \; \theta \in \Theta)$ be a <u>dominated</u> experiment. Let $\mathcal{B}$ be a sub-σ-algebra of $\mathcal{A}$ and let π be given as in prop. 8.8. Then the following conditions are equivalent:

(i) $\mathcal{B}$ is CE-sufficient.

(ii) $\mathcal{B}$ is Δ-sufficient.

(iii) $dP_\theta/d\pi$ may for each $\theta \in \Theta$ be specified $\mathcal{B}$-measurable.

(iv) $\mathcal{B}$ is pairwise sufficient.

PROPOSITION 8.14

Let $\mathcal{E}$ be an experiment and assume that μ is a σ-finite measure such that $\mu >> P_\theta$ for all $\theta \in \Theta$.

Then a sub-σ-algebra $\mathcal{B}$ of $\mathcal{A}$ is sufficient if and only if there exists a non-negative $\mathcal{A}$-measurable function h and a set $\{g_\theta : \theta \in \Theta\}$ of non-negative $\mathcal{B}$-measurable functions such that

$$dP_\theta/d\mu = hg_\theta \quad \text{for all} \quad \theta \in \Theta .$$

Proof: Assume that $\mathcal{B}$ is sufficient. Clearly $\mu >> \pi$, where π is given in prop. 8.8. Thus by the chain rule of Radon Nikodym derivatives,

$$\frac{dP_\theta}{d\mu} = \frac{d\pi}{d\mu} \circ \frac{dP_\theta}{d\pi} \quad \text{for all} \quad \theta \in \Theta .$$

Hence we may put $h = d\pi/d\mu$, which is obviously $\mathcal{A}$-measurable.

We put $g_\theta = dP_\theta/d\pi$; $\theta \in \Theta$. The g_θ 's are $\mathcal{B}$-measurable by theorem 8.13.

Assume now that the condition of the proposition holds.

Then $d\pi/d\mu = h \sum_{\theta\in\Theta_0} c(\theta)g_\theta$

and hence, for any $\theta \in \Theta$,

$$\frac{dP_\theta}{d\pi} = \frac{dP_\theta/d\mu}{d\pi/d\mu} = \frac{g_\theta}{h\sum_\theta c(\theta)g_\theta}$$

which is $\mathcal{B}$-measurable. Hence $\mathcal{B}$ is sufficient.

DEFINITION 8.15

Let $\mathcal{E} = (\chi, \mathcal{A}; P_\theta : \theta \in \Theta)$ be an experiment and let $\mathcal{B}_1$ and $\mathcal{B}_2$ be sub-σ-algebras of $\mathcal{A}$. We define an ordering $\leq$

by

$$\mathcal{B}_1 \leq \mathcal{B}_2$$

$$\Updownarrow \text{def}$$

(5) for all $B_1 \in \mathcal{B}_1$ there exists $B_2 \in \mathcal{B}_2$ such that $P_\theta(B_1 \Delta B_2) = 0$ for all $\theta \in \Theta$. (Δ means symmetric difference.)

It is easily verified that the condition (5) is equivalent to

(6) for each $\mathcal{B}_1$-measurable bounded (or non-negative) function g_1 there is a $\mathcal{B}_2$-measurable bounded (or non-negative) function g_2 such that

$E_\theta|g_1 - g_2| = 0$ for all $\theta \in \Theta$

i.e. $g_1 = g_2$ a.e. $[P_\theta]$ for all $\theta \in \Theta$.

If $\mathcal{B}_1 \leq \mathcal{B}_2$ and $\mathcal{B}_2 \leq \mathcal{B}_1$ then we say that $\mathcal{B}_1$ and $\mathcal{B}_2$ are equivalent and write this $\mathcal{B}_1 \sim \mathcal{B}_2$.

$\leq$ defines a partial ordering on the set of sub-σ-algebras of $\mathcal{O}\!\mathcal{l}$:

(i) $\mathcal{B} \leq \mathcal{B}$ for all $\mathcal{B}$.

(ii) $\mathcal{B}_1 \leq \mathcal{B}_2$ and $\mathcal{B}_2 \leq \mathcal{B}_3 \Rightarrow \mathcal{B}_1 \leq \mathcal{B}_3$.

(iii) $\mathcal{B}_1 \leq \mathcal{B}_2$ and $\mathcal{B}_2 \leq \mathcal{B}_1 \Rightarrow \mathcal{B}_1 \sim \mathcal{B}_2$.

DEFINITION 8.16

Let $\mathcal{E}$ be an experiment. A σ-algebra $\mathcal{B}_o$ is said to be minimal CE-sufficient for $\mathcal{E}$ if $\mathcal{B}_o$ is CE-sufficient for $\mathcal{E}$ and $\mathcal{B}_o \leq \mathcal{B}$ for all CE-sufficient σ-algebras. $\mathcal{B}$.

PROPOSITION 8.17

Assume that $\mathcal{E}$ is dominated and let $\mathcal{B}_o$ be the smallest

σ-algebra such that the functions
$dP_\theta/d\pi$ are measurable for all $\theta \in \Theta$.
Then $\mathcal{B}_o$ is minimal sufficient for $\mathcal{E}$.

Proof: $\mathcal{B}_o$ is sufficient by theorem 8.13. Let $\{g_\theta : \theta \in \Theta\}$ be $\mathcal{B}_o$-measurable versions of $dP_\theta/d\pi$.
Assume now that $\mathcal{B}$ is sufficient for $\mathcal{E}$ and let $\{h_\theta : \theta \in \Theta\}$ be $\mathcal{B}$-measurable versions of $dP_\theta/d\pi$. We shall prove that $\mathcal{B}_o \leq \mathcal{B}$. It follows from Radon-Nikodym's theorem that

(7) $h_\theta = g_\theta$ a.e. $[\pi]$ for all $\theta \in \Theta$.

By definition, $\mathcal{B}_o$ is the smallest σ-algebra containing all sets of the form
$A_\theta(r) = \{x : g_\theta(x) < r\}$ for some $r \in R$, $\theta \in \Theta$.
Define $B_\theta(r) = \{x : h_\theta(x) < r\}$; $r \in R$, $\theta \in \Theta$.
Then $B_\theta(r) \in \mathcal{B}$ and by (7), $\pi(A_\theta(r) \Delta B_\theta(r)) = 0$ for all r, θ
It is easy to verify that the family sets $B_o \in \mathcal{B}_o$ such that there exists $B \in \mathcal{B}$ with $\pi(B_o \Delta B) = 0$ is a σ-algebra.
Since this σ-algebra contains the sets $A_\theta(r)$, it is equal to $\mathcal{B}_o$. Hence (5) of def. 8.15 implies that $\mathcal{B}_o \leq \mathcal{B}$.

DEFINITION 8.18

Let $\mathcal{E}$ be an experiment.
A sub-σ-algebra $\mathcal{B}$ of $\mathcal{A}$ is said to be boundedly complete if for all bounded $\mathcal{B}$-measurable functions g
$E_\theta g = 0$ for all $\theta \in \Theta \Rightarrow g = 0$ a.e. $[P_\theta]$ for all $\theta \in \Theta$.

PROPOSITION 8.19

Let $\mathcal{E}$ be an experiment.
Assume that $\mathcal{B}$ is CE-sufficient and boundedly complete.
If $\mathcal{C}$ is CE-sufficient and $\mathcal{C} \leq \mathcal{B}$, then $\mathcal{B} \sim \mathcal{C}$.

Proof: It suffices to prove that $\mathcal{B} \leq \mathcal{E}$.

Let $B \in \mathcal{B}$. By def. 8.2 there exists a $\mathcal{E}$-measurable Y such that

$$P_\theta^{\mathcal{E}}(B) = Y \quad \text{a.e.} \quad [P_\theta] \quad \text{for all} \quad \theta \in \Theta .$$

Set

$$C = \{x : Y(x) = 1\} .$$ Clearly $C \in \mathcal{E}$.

Since $\mathcal{E} \leq \mathcal{B}$ there is a $\mathcal{B}$-measurable function Z such that $Z = Y$ a.e. $[P_\theta]$; $\theta \in \Theta$.

Z is bounded (a.e.) since Y is. Furthermore, for any $\theta \in \Theta$,

$$\int Z dP_\theta = \int Y dP_\theta = \int P_\theta^{\mathcal{E}}(B) dP_\theta = P_\theta(B)$$

Hence $\int (I_B - Z) dP_\theta = 0$ for all θ .

Since $\mathcal{B}$ is boundedly complete it follows that

$$I_B = Z \quad \text{a.e.} \quad [P_\theta] \quad \text{for all} \quad \theta .$$

Hence $Y = I_B$ a.e. $[P_\theta]$ for all θ

so that $P_\theta(B \Delta C) = 0$ for all θ .

COROLLARY 8.20

Let $\mathcal{E}$ be a dominated experiment.

If $\mathcal{B}$ is sufficient and boundedly complete, then $\mathcal{B}$ is minimal sufficient.

Proof: Let $\mathcal{B}_o$ be given as in prop. 8.21. Then $\mathcal{B}_o \leq \mathcal{B}$.

By prop. 8.19, $\mathcal{B}_o \sim \mathcal{B}$. The corollary follows.

We shall now see that the concept of being more informative is closely related to the concept of being sufficient for an experiment.

Let $\mathcal{E}$ and $\mathcal{F}$ be experiments satisfying the conditions of theorem 7.6. Assume further that $\mathcal{E} \geq \mathcal{F}$.

Then there exists a Markov kernel M such that $P_\theta M = Q_\theta$ for all θ. We define an experiment

$\tilde{\mathcal{E}} = (\mathcal{Z}, \mathcal{C}, \tilde{P}_\theta : \theta \in \Theta)$ where $(\mathcal{Z}, \mathcal{C}) = (\chi, \mathcal{A}) \times (\mathcal{Y}, \mathcal{B})$ (i.e. $\mathcal{Z} = \chi \times \mathcal{Y}$ and $\mathcal{C} = \mathcal{A} \times \mathcal{B}$ is the σ-algebra generated by the sets $A \times B$; $A \in \mathcal{A}$, $B \in \mathcal{B}$) and where $\tilde{P}_\theta = P_\theta \times M$, i.e. $\tilde{P}_\theta(A \times B) = \int_A M(B|x)P_\theta(dx)$

for all $A \in \mathcal{A}$, $B \in \mathcal{B}$, $\theta \in \Theta$. (See lemma 14 of Appendix B.)

We observe that

$\tilde{P}_\theta(A \times \mathcal{Y}) = \int_A M(\mathcal{Y}|x)P_\theta(dx) = P_\theta(A)$

$\tilde{P}_\theta(\chi \times B) = \int_\chi M(B|x)P_\theta(dx) = Q_\theta(B)$ for all $\theta \in \Theta$, $A \in \mathcal{A}$, $B \in \mathcal{B}$.

The experiment $\mathcal{E}$ is thus seen to be equivalent to the reduction of $\tilde{\mathcal{E}}$ obtained by replacing the σ-algebra $\mathcal{C} = \mathcal{A} \times \mathcal{B}$ by the sub-σ-algebra $\mathcal{A} \times \{\emptyset, \mathcal{Y}\}$. We denote this experiment by $\mathcal{E}^*$.

Similarly, $\mathcal{F}$ is equivalent to the reduction of $\tilde{\mathcal{E}}$ obtained by replacing $\mathcal{C}$ by $\{\emptyset, \chi\} \times \mathcal{B}$. We call it $\mathcal{F}^*$.

$\mathcal{E}^*$ and $\mathcal{F}^*$ are said to be marginals of the experiment $\tilde{\mathcal{E}}$, and we observe that M defines the conditional distribution of the second marginal, given the first marginal. We remark that M is independent of θ.

PROPOSITION 8.21

The sub-σ-algebra $\mathcal{A} \times \{\emptyset, \mathcal{Y}\}$ is CE-sufficient in $\tilde{\mathcal{E}}$.

Proof: By def. 8.2 we have to prove that to each $C \in \mathcal{A} \times \mathcal{B}$ there exists a $\mathcal{A} \times \{\emptyset, \mathcal{Y}\}$- measurable function Y_C such that $E_\theta^{\mathcal{A} \times \{\emptyset, \mathcal{Y}\}} I_C = Y_C$ a.e. $[\tilde{P}_\theta]$ for all $\theta \in \Theta$.

First, let $C = A \times B$; $A \in \mathcal{A}$, $B \in \mathcal{B}$.

Then we will have

$E_\theta^{\mathcal{A}\times\{\emptyset,\mathcal{Y}\}} I_{A\times B} = I_A M(B|\cdot)$ a.e.

which is a function of $x \in \chi$ alone and hence is $\mathcal{A} \times \{\emptyset,\mathcal{Y}\}$-measurable considered as a function on $\chi \times \mathcal{Y}$.

Moreover, it is independent of θ . Hence (7) is satisfied for all measurable sets $C = A \times B$. Since the family of measurable sets satisfying (7) will constitute a σ-algebra, it follows that (7) holds for all $C \in \mathcal{A} \times \mathcal{B}$.

COROLLARY 8.22

$\mathcal{E} \sim \tilde{\mathcal{E}}$

Remark: Loosely speaking, the relation $\mathcal{E} \geq \mathcal{F}$ says that $\mathcal{E}$ is sufficient for the experiment $\tilde{\mathcal{E}}$ having marginals $\mathcal{E}$ and $\mathcal{F}$.

PROPOSITION 8.23

Let $\mathcal{E}$ and $\mathcal{F}$ be experiments satisfying the conditions of theorem 7.6.

Assume that $\mathcal{E} \geq \mathcal{F}$ and that

$$\mathcal{E}_{\{\theta_1,\theta_2\}} \sim \mathcal{F}_{\{\theta_1,\theta_2\}} \quad \text{for all pairs } (\theta_1,\theta_2) \in \Theta \times \Theta .$$

Then $\mathcal{E} \sim \mathcal{F}$.

Proof: From $\mathcal{E} \geq \mathcal{F}$ it follows by prop. 8.21 that $\mathcal{A} \times \{\emptyset,\mathcal{Y}\}$ is sufficient for $\tilde{\mathcal{E}}$. Since $\mathcal{E}_{\{\theta_1,\theta_2\}} \sim \mathcal{F}_{\{\theta_1,\theta_2\}}$ and since $\mathcal{E} \sim \mathcal{E}^*$, $\mathcal{F} \sim \mathcal{F}^*$, it follows that

$\mathcal{E}^*_{\{\theta_1,\theta_2\}} \sim \mathcal{F}^*_{\{\theta_1,\theta_2\}}$ for all pairs θ_1,θ_2 .

Hence, since $\tilde{\mathcal{E}} \sim \mathcal{E}^*$, $\mathcal{F}^*$ is pairwise equivalent to $\tilde{\mathcal{E}}$ so the sub-σ-algebra $\{\emptyset,\chi\} \times \mathcal{B}$ is pairwise sufficient for

$\tilde{\mathcal{E}}$. Now $\tilde{\mathcal{E}}$ is dominated since $\mathcal{E} \sim \tilde{\mathcal{E}}$ and $\mathcal{E}$ is dominated. By prop. 8.13 $\{\emptyset, \chi\} \times \mathcal{B}$ is sufficient for $\tilde{\mathcal{E}}$, and hence $\mathcal{F}^* \sim \tilde{\mathcal{E}}$. But then $\mathcal{F}^* \sim \mathcal{E}$ and finally $\mathcal{F} \sim \mathcal{E}$.

Remark: It may be shown that the preceding proposition will remain valid if we remove the requirements on $(\mathcal{Y}, \mathcal{B})$.

APPENDIX A

ESSENTIAL SUPREMUM OF A FAMILY OF RANDOM VARIABLES, WITH APPLICATION TO DOMINATED EXPERIMENTS.

DEFINITION 1.

An experiment $\mathcal{E}$ is given by $\mathcal{E} = (\chi, \mathcal{A}, P_\theta : \theta \in \Theta)$ where $(\chi, \mathcal{A})$ is a measurable space and $(P_\theta : \theta \in \Theta)$ is a family of probability measures on $(\chi, \mathcal{A})$.

DEFINITION 2.

Let $(\chi, \mathcal{A}, P)$ be a probability space. A random variable (abbreviated r.v) is an $\mathcal{A}$-measurable function $X : \chi \to [-\infty, \infty]$.

Consider now a fixed probability space $(\chi, \mathcal{A}, P)$.

DEFINITION 3.

Let X, Y be r.v's. We define $X \leq Y$ a.s. (almost surely) to mean $P(X > Y) = 0$. It is seen that $\leq$ is a partial ordering on the set of r.v's on $(\chi, \mathcal{A}, P)$.

DEFINITION 4.

Let $\{X_t ; t \in T\}$ be a family of r.v's. The r.v. Y is called essential supremum for the family if

(i) $X_t \leq Y$ a.s. for all $t \in T$

(ii) $X_t \leq Z$ a.s. for all $t \in T$ implies $Y \leq Z$ a.s.

It follows from (ii) that provided the essential supremum exists, it is uniquely determined up to a P-equivalence.

We can thus write $Y = \operatorname{ess\,sup}_{t \in T} X_t$.

Remark: Assume the index set T is countable. Then $Y = \sup_{t \in T} X_t$ is proved to be measurable and the verification of (i) and (ii) is trivial. Thus an essential supremum always exists if our family of r.v's is countable.

If T is not countable, the function $Y = \sup_{t \in T} X_t$ may not be measurable and thus not a r.v. The following theorem states, however, that an essential supremum still exists.

THEOREM 5.

To each family $\{X_t : t \in T\}$ of r.v's there exists an essential supremum.

Moreover, there is a countable subset T_0 of T such that $\operatorname{ess\,sup}_{t \in T} X_t = \sup_{t \in T_0} X_t$ a.s.

Proof: Let Φ be a 1-1 mapping of $[-\infty,\infty]$ onto $[0,1]$ (Φ may for example be taken as

$$\Phi(x) = (2\pi)^{-\frac{1}{2}} \int_{-\infty}^{x} e^{-\frac{z^2}{2}} dz) .$$

Let $\mathfrak{F}$ be the set of finite, non-empty subsets of T.

For each $F \in \mathfrak{F}$ we define $X_F = \max_{t \in F} X_t$

Clearly X_F is a r.v.

Set $\alpha = \sup_F E\Phi(X_F)$. Then $\alpha \in [0,1]$. There is a sequence $\{F_n\}$ in $\mathfrak{F}$ such that $E\Phi(X_{F_n}) \uparrow \alpha$. We may without loss of generality assume $F_1 \subseteq F_2 \subseteq \ldots$ (which again implies $X_{F_1} \leq X_{F_2} \leq \ldots$).

If this is not the case, we may namely put $G_i = F_1 \cup \ldots \cup F_i$ to obtain $X_{G_i} = \max(X_{F_1},\ldots,X_{F_i}) \geq X_{F_i}$. Hence

$\alpha \geq E\Phi(X_{G_i}) \geq E\Phi(X_{F_i})$ so that $E\Phi(X_{G_i}) \uparrow \alpha$.

Let now $T_o = \bigcup_{i=1}^{\infty} F_i$. T_o is countable, so $Y = \sup_{t \in T_o} X_t$ is a r.v.

We observe that $X_{F_i} \uparrow Y$. We shall prove that $Y = \operatorname{ess\,sup}_{t \in T} X_t$.

Let t be an arbitrary, but fixed, element of T. Condition (i) of def. 4 will follow if we show that $X_t \leq Y$ a.s.

Clearly $\max(X_{F_i},X_t) = X_{F_i \cup \{t\}} \uparrow \max(Y,X_t)$.

Hence, by the monotone convergence theorem,

$$E\Phi(\max(Y,X_t)) = \lim_{i\to\infty} E\Phi(X_{F_i \cup \{t\}}) \leq \alpha .$$

On the other hand,

$$E\Phi(Y) = E\Phi(\lim X_{F_i}) = E\lim\Phi(X_{F_i}) = \lim\Phi(X_{F_i}) = \alpha$$

and hence

$$E\Phi(\max(Y,X_t)) \geq \alpha .$$

It follows that

$$E\Phi(\max(Y,X_t)) = \alpha = E\Phi(Y) \text{ , i.e.}$$

$$E[\Phi(\max(Y,X_t))-\Phi(Y)] = 0 .$$

Since the expression in the brackets is always non-negative, we have

$$\Phi(\max(Y,X_t)) = \Phi(Y) \quad \text{a.s.}$$

i.e. $\max(Y,X_t) = Y$ a.s.

which implies $X_t \leq Y$ a.s. and (i) of def. 4 is proved.

Assume now $X_t \leq Z$ a.s. for all $t \in T$.

Obviously $X_t \leq Z$ a.s. for all $t \in T_0$ which implies $Y \leq Z$ a.s.

It follows that $Y = \operatorname{ess\,sup}_{t \in T} X_t$.

LEMMA 6.

Assume $X = \operatorname{ess\,sup}_{t \in T} X_t$. Let Y be a r.v. such that $Y \geq 0$ a.s.

Then $XY = \operatorname{ess\,sup}_{t \in T}(X_t Y)$ a.s.

Proof: From $X_t \leq X$ a.s. it follows that

$X_t Y \leq XY$ a.s. for all $t \in T$.

By theorem 5 there is a countable $T_0 \subseteq T$ such that $X = \sup_{t \in T_0} X_t$ a.s. Hence $XY = \sup_{t \in T_0} X_t Y$ a.s.

Assume $X_t Y \leq Z$ a.s. for all $t \in T$.

Then $XY = \sup_{t \in T_0} X_t Y \leq Z$ a.s. and the proof is complete.

We state the celebrated theorem of Radon-Nikodym:

THEOREM 7.

Let $(\chi, \mathcal{O}\!\mathcal{l}, \mu)$ be a σ-finite measure space and let ν be a measure defined on $\mathcal{O}\!\mathcal{l}$ which is absolutely continuous w.r.t. μ. Then there is a measurable function $f : \chi \to [0,\infty]$ such that

$\nu(A) = \int_A f d\mu$ for all $A \in \mathcal{O}\!\mathcal{l}$.

If ν is σ-finite, then f may be chosen to be finite.

f is called the Radon-Nikodym derivative of ν w.r.t. μ and is denoted $d\nu/d\mu$.

DEFINITION 8.

Two measures μ and ν on a measurable space $(\mathcal{X},\mathfrak{A})$ are said to be equivalent if $\nu \ll \mu$ and $\mu \ll \nu$. We then write $\mu \sim \nu$.

The relation $\sim$ is an equivalence relation on the set of measures on $(\mathcal{X},\mathfrak{A})$, and the equivalence classes consist of the measures having the same null-sets.

LEMMA 9.

Let μ be a σ-finite measure on the measurable space $(\mathcal{X},\mathfrak{A})$, $\mu \neq 0$. Then there is a probability measure P on $(\mathcal{X},\mathfrak{A})$ such that $\mu \sim P$.

Proof: Let $\mathcal{X} = \sum_{n=1}^{\infty} \mathcal{X}_n$ (disjoint union) with $0 < \mu(\mathcal{X}_n) < \infty$ for each n. For $A \in \mathfrak{A}$ define $P(A) = \sum_{n=1}^{\infty} \frac{\mu(A\cap\mathcal{X}_n)}{2^n\mu(\mathcal{X}_n)}$

P is easily seen to have the required properties.

DEFINITION 10.

Let $\mathcal{E} = (\mathcal{X},\mathfrak{A},P_\theta : \theta \in \Theta)$ be an experiment. $\mathcal{E}$ is said to be dominated if there is a σ-finite measure μ on $\mathfrak{A}$ such that $P_\theta \ll \mu$ for all $\theta \in \Theta$. It follows from lemma 9 that μ may be assumed to be a probability measure.

THEOREM 11.

Assume $\mathcal{E}$ is a dominated experiment. Then we can find a countable subset Θ_o of Θ such that

$$P_\theta(A) = 0 \text{ for all } \theta \in \Theta_o \Rightarrow P_\theta(A) = 0 \text{ for all } \theta \in \Theta .$$

Proof: Assume $P_\theta << P$ for all $\theta \in \Theta$ and put $f_\theta = dP_\theta/dP$. $\{f_\theta : \theta \in \Theta\}$ may now be considered as a family of r.v's on the probability space $(\chi, \mathcal{A}, P)$ and we can define

$$g = \operatorname*{ess\,sup}_{\theta\in\Theta} f_\theta$$

By theorem 5, $g = \sup_{\theta\in\Theta_0} f_\theta$ a.s. for a countable subset $\Theta_0 \subseteq \Theta$.

Assume now $P_\theta(A) = 0$ for all $\theta \in \Theta_0$.

We have $P_\theta(A) = \int_A f_\theta dP = \int I_A f_\theta dP$, and hence

$$I_A f_\theta = 0 \quad \text{a.s.} \quad [P] \quad \text{for all} \quad \theta \in \Theta_0 .$$

By lemma 6

$$\operatorname*{ess\,sup}_{\theta\in\Theta} I_A f_\theta = I_A \operatorname*{ess\,sup}_{\theta\in\Theta} f_\theta = I_A \sup_{\theta\in\Theta_0} f_\theta = \sup_{\theta\in\Theta_0} I_A f_\theta = 0 \quad \text{a.s.,}$$

which implies

$$I_A f_\theta = 0 \quad \text{a.s. for all} \quad \theta \in \Theta \quad \text{and hence}$$

$$P_\theta(A) = \int I_A f_\theta dP = 0 \quad \text{for all} \quad \theta \in \Theta$$

THEOREM 12.

Let $\mathcal{E} = (\chi, \mathcal{A}, P_\theta : \theta \in \Theta)$ be a dominated experiment. Then $\mathcal{E}$ is dominated by a probability measure π given by $\pi = \sum_\theta c(\theta)P_\theta$ where $c(\theta) \geq 0$ for all $\theta \in \Theta$ and $\sum_\theta c(\theta) = 1$ (the set of θ's for which $c(\theta) > 0$ is countable).

Proof: Choose a countable subser $\Theta_0 \subseteq \Theta$ with the property given in theorem 11. Let the elements of Θ_0 be ordered in a sequence $\theta_1, \theta_2, \ldots$.

Define $\pi(A) = \sum_{n=1}^{\infty} 2^{-n} P_{\theta_n}(A)$. π is now a probability measure on $\mathfrak{A}$ and clearly $\pi(A) = 0$ implies $P_{\theta_n}(A) = 0$ for all n , which again by the choice of Θ_o implies $P_\theta(A) = 0$ for all $\theta \in \Theta$. Hence π dominates $\mathcal{P}$ and the proof is complete.

APPENDIX B

MEASURE-THEORETIC COMPLEMENTS. MARKOV-KERNELS AND ASSOCIATED BILINEAR FUNCTIONALS.

DEFINITION 1.

A family $\mathcal{C}$ of subsets of a set χ is called a π-system if $\mathcal{C}$ has the following properties:

(π1) $\emptyset \in \mathcal{C}$

(π2) $C_1, C_2 \in \mathcal{C} \Rightarrow C_1 \cap C_2 \in \mathcal{C}$

We observe that a π-system is closed under finite intersections.

Example: Let R^{∞} be the set of sequences in R. The product σ-algebra $\mathcal{B}^{\infty}$ is generated by the sets $\{(x_1, x_2, \dots) : x_1 < a_1 , x_2 < a_2, \dots, x_r < a_r\}$ (where $a_1, \dots, a_r \in R$) which constitute a π-system.

DEFINITION 2.

A family $\mathcal{D}$ of subsets of a set χ is called a λ-system if $\mathcal{D}$ has the following properties:

(λ1) $\emptyset \in \mathcal{D}$

(λ2) $D \in \mathcal{D} \Rightarrow D^c \in \mathcal{D}$

(λ3) $D_1, D_2 \in \mathcal{D}$ and $D_1 \cap D_2 = \emptyset \Rightarrow D_1 \cup D_2 \in \mathcal{D}$

(λ4) $D_1 \subseteq D_2 \subseteq \dots \Rightarrow \bigcup_{i=1}^{\infty} D_i \in \mathcal{D}$ whenever $D_1, D_2, \dots \in \mathcal{D}$

Example: Let $(\chi, \mathcal{A})$ be a measurable space and P, Q probability measures on $(\chi, \mathcal{A})$.

Let $\mathcal{D} = \{D \in \mathcal{A} : P(D) = Q(D)\}$. $\mathcal{D}$ is a λ-system.

We recall the definition of a σ-algebra:

DEFINITION 3.

A family $\mathcal{A}$ of subsets of a set χ is said to be a σ-algebra in χ if

(σ1) $\emptyset \in \mathcal{A}$

(σ2) $A \in \mathcal{A} \Rightarrow A^c \in \mathcal{A}$

(σ3) $A_n \in \mathcal{A}$ for $n = 1,2,\ldots \Rightarrow \bigcup_{n=1}^{\infty} A_n \in \mathcal{A}$

DEFINITION 4.

Let $\mathcal{H}$ be a family of subsets of a set χ.

We denote by $\sigma(\mathcal{H})$, $\pi(\mathcal{H})$, $\lambda(\mathcal{H})$ the smallest σ-algebra, π-system, λ-system (respectively) containing $\mathcal{H}$.

The existence of $\pi(\mathcal{H})$ $(\lambda(\mathcal{H}))$ follows as in the case of $\sigma(\mathcal{H})$ by taking the intersection of all π-systems (λ-systems) containing $\mathcal{H}$.

PROPOSITION 5.

A family $\mathcal{H}$ of sets is a σ-algebra if and only if $\mathcal{H}$ is both a λ-system and a π-system.

Proof: The "only if"-part is trivial.

Assume $\mathcal{H}$ satisfies the requirements of a π-system and a λ-system.

(σ1) follows from (λ1) and (λ2)

(σ2) is the same as (λ2)

It remains to show that $(\sigma 3)$ holds.

Let $A_1, A_2, \ldots$ be a sequence of sets in $\mathcal{H}$. We have $A_1 \cup A_2 = A_1 \cup (A_2 - A_1)$. Now $A_2 - A_1 = A_2 \cap A_1^c \in \mathcal{H}$ by $(\lambda 2)$ and $(\pi 2)$. Since A_1 and $A_2 - A_1$ are disjoint, $(\lambda 3)$ yields $A_1 \cup A_2 \in \mathcal{H}$. By induction we conclude that $D_n = A_1 \cup \ldots \cup A_n \in \mathcal{H}$ for $n = 1, 2, \ldots$. Clearly $D_1 \subseteq D_2 \subseteq \ldots$ and by $(\lambda 4)$ $\bigcup_{n=1}^{\infty} D_n \in \mathcal{H}$. $(\sigma 3)$ follows since $\bigcup_{n=1}^{\infty} A_n = \bigcup_{n=1}^{\infty} D_n$.

<u>LEMMA 6.</u>

Let $\mathcal{D}$ be a λ-system and $D_1, D_2 \in \mathcal{D}$.

Then $D_1 \subseteq D_2 \Rightarrow D_2 - D_1 \in \mathcal{D}$

<u>Proof:</u> $(D_2 - D_1)^c = (D_2 \cap D_1^c)^c = D_2^c \cup D_1$.
But $D_1 \subseteq D_2$ implies $D_1 \cap D_2^c = \emptyset$ and the lemma follows from $(\lambda 3)$ and $(\lambda 2)$.

<u>PROPOSITION 7.</u>

Let $\mathcal{H}$ be an arbitrary family of subsets of χ. Then $\sigma(\mathcal{H}) = \lambda(\pi(\mathcal{H}))$.

<u>Proof:</u> It suffices to prove that if a family $\mathcal{E}$ is a π-system, then $\lambda(\mathcal{E})$ is a σ-algebra. It then follows that $\lambda(\pi(\mathcal{H}))$ is a σ-algebra and consequently $\sigma(\mathcal{H}) \subseteq \lambda(\pi(\mathcal{H}))$. The opposite inclusion follows from the fact that we by forming λ- and π-systems will not get outside the σ-algebra generated by $\mathcal{H}$.
By prop. 5 it is enough to prove that $\lambda(\mathcal{E})$ is a π-system whenever $\mathcal{E}$ is a π-system.
We first prove the following result:

(1) $D \cap C \in \lambda(\mathcal{C})$ for any $C \in \mathcal{C}$ and $D \in \lambda(\mathcal{C})$.
Let $C_o \in \mathcal{C}$. Then $\mathcal{D} = \{D: D \cap C_o \in \lambda(\mathcal{C})\}$ is a λ-system containing $\mathcal{C}$. (λ1), (λ3) and (λ4) are easily verified. As for (λ2), assume $D \in \mathcal{D}$, i.e. $D \cap C_o \in \lambda(\mathcal{C})$. Then $D^c \cap C_o = C_o - D = C_o-(D \cap C_o)$. Since $D \cap C_o \subseteq C_o$, lemma 6 applies and shows that $D^c \cap C_o \in \lambda(\mathcal{C})$ and hence $D^c \in \mathcal{D}$
$\mathcal{C} \subseteq \mathcal{D}$ since $\mathcal{C}$ is a π-system. Hence $\lambda(\mathcal{C}) \subseteq \mathcal{D}$ so $D \in \lambda(\mathcal{C})$ implies $D \in \mathcal{D}$ which again implies $D \cap C_o \in \lambda(\mathcal{C})$. (1) follows. Finally, we will use (1) to prove that
$D_1 \cap D_2 \in \lambda(\mathcal{C})$ for any $D_1, D_2 \in \lambda(\mathcal{C})$.
Let $D_1 \in \lambda(\mathcal{C})$ and let $\mathcal{D}' = \{D: D \cap D_1 \in \lambda(\mathcal{C})\}$.
As in the case of $\mathcal{D}$, $\mathcal{D}'$ is seen to be a λ-system. From the above result it follows that $\mathcal{C} \subseteq \mathcal{D}'$ which implies $\lambda(\mathcal{C}) \subseteq \mathcal{D}'$.
Let now $D_2 \in \lambda(\mathcal{C})$. Then $D_2 \in \mathcal{D}'$ which is the same as $D_1 \cap D_2 \in \lambda(\mathcal{C})$. Since D_1 and D_2 are chosen arbitrarily, (π2) is satisfied and $\lambda(\mathcal{C})$ is a π-system.

We give a few examples of the application of π-systems and λ-systems to prove important measure-theoretic results.

<u>EXAMPLE 8.</u>

Let P, Q be probabilitymeasures on a measurable space $(\mathcal{X}, \mathcal{A})$. Assume $\mathcal{A}$ is generated by a π-system $\mathcal{C}$ (i.e. $\mathcal{A} = \sigma(\mathcal{C})$) and that
$P(C) = Q(C)$ for each $C \in \mathcal{C}$. Then $P = Q$.

<u>Proof</u>: Let $\mathcal{D} = \{D: P(D) = Q(D)\}$. By the example of def. 2, $\mathcal{D}$ is a λ-system containing $\mathcal{C}$. Hence $\mathcal{A} = \sigma(\mathcal{C}) = \lambda(\pi(\mathcal{C}))$ $= \lambda(\mathcal{C}) \subseteq \mathcal{D}$.

The second equality sign holds by prop. 7, the third holds since $\mathcal{E}$ is a π-system.

The statement follows.

EXAMPLE 9. (Independence.)

Let $(X, \mathcal{A}, P)$ be a probability space.

Two events $C_1, C_2 \in \mathcal{A}$ are said to be independent if $P(C_1 \cap C_2) = P(C_1)P(C_2)$. Two families $\mathcal{E}, \mathcal{D} \subseteq \mathcal{A}$ are said to be independent if $C \in \mathcal{E}$, $D \in \mathcal{D}$ $\Rightarrow P(C \cap D) = P(C)P(D)$.

Proposition: If $\mathcal{E}$ and $\mathcal{D}$ are independent, then $\lambda(\mathcal{E})$ and $\lambda(\mathcal{D})$ are independent.

Proof: Assume $\mathcal{E}$ and $\mathcal{D}$ are independent.

Let $\mathcal{F} = \{C\colon P(C \cap D) = P(C)P(D)$ for all $D \in \mathcal{D}\}$. We observe that $\mathcal{F}$ is a λ-system containing $\mathcal{E}$. Hence $\lambda(\mathcal{E}) \subseteq \mathcal{F}$, so

(2) $P(C \cap D) = P(C)P(D)$ for any $C \in \lambda(\mathcal{E})$, $D \in \mathcal{D}$.

Let now $\mathcal{H} = \{D\colon P(C \cap D) = P(C)P(D)$ for all $C \in \lambda(\mathcal{E})\}$.

$\mathcal{H}$ is a λ-system and $\mathcal{D} \subseteq \mathcal{H}$ by (2).

Hence $\lambda(\mathcal{D}) \subseteq \mathcal{H}$ and accordingly $\lambda(\mathcal{E})$ and $\lambda(\mathcal{D})$ are independent.

Remark: If $\mathcal{E}$ and $\mathcal{D}$ are π-systems, then $\sigma(\mathcal{E})$ and $\sigma(\mathcal{D})$ are independent (prop. 7).

The above results are easily generalized to arbitrary collections of families. We recall that if T is an index set, then the families $\mathcal{E}_t : t \in T$ are said to be independent if for every finite set $\{t_1, \ldots, t_n\} \subseteq T$.

$$A_1 \in \mathcal{E}_{t_1}, \ldots, A_n \in \mathcal{E}_{t_n} \Rightarrow P(\bigcap_{k=1}^{n} A_k) = \prod_{k=1}^{n} P(A_k) .$$

DEFINITION 10.

Let μ be a signed measure on a measurable space $(\chi,\mathcal{A})$. We define the norm of μ by

$$\|\mu\| = \sup_{\|f\|\leq 1} \left|\int f d\mu\right| = \sup_{\|f\|\leq 1} \int f d\mu$$

(the supremum is taken over all measurable real functions on $(\chi,\mathcal{A})$ such that $\|f\| = \sup_{x\in\chi} |f(x)| \leq 1$)

We observe that if μ is a non-negative measure, then $\|\mu\| = \mu(\chi)$.

DEFINITION 11.

Let $(\chi,\mathcal{A})$ be a measurable space.

$\mathcal{M}(\mathcal{A})$ is the set of finite signed measures defined on $\mathcal{A}$ (i.e. measures with finite norm).

$\mathfrak{F}(\mathcal{A})$ is the set of bounded measurable functions on $(\chi,\mathcal{A})$.

The spaces defined above are obviously linear spaces.

DEFINITION 12.

Let $(\chi,\mathcal{A})$, $(\mathcal{Y},\mathcal{B})$ be measurable spaces.

An $\mathcal{A}$-measurable measure on $\mathcal{B}$ is a function ρ from $\mathcal{B}\times\chi$ to R such that

(i) for each $x\in\chi$, $\rho(\circ|x)\in\mathcal{M}(\mathcal{B})$

and $\sup_{x\in\chi}\|\rho(\circ|x)\| < \infty$

(ii) for each $B\in\mathcal{B}$, $\rho(B|\circ)\in\mathfrak{F}(\mathcal{A})$.

If ρ is non-negative and satisfies $\rho(\mathcal{Y}|x) \leq 1$ for all $x \in \chi$, then ρ is called a sub-Markov kernel.

If $\rho(\circ|x)$ is a probability measure for each $x \in \chi$, then ρ is called a Markov-kernel.

Example: Let $\nu \in \mathcal{M}(\mathcal{B})$. If we define $\rho : \mathcal{B} \times \chi \to R$ by

$$\rho(B|x) = \nu(B) \ , \ B \in \mathcal{B} \ , \ x \in \chi \ ,$$

then ρ is an $\mathcal{A}$-measurable measure on $\mathcal{B}$ such that each function $\rho(B|\circ)$ is constant.

Markov kernels may be considered as conditional probabilities.

LEMMA 13.

Let $(\chi,\mathcal{A})$, $(\mathcal{Y},\mathcal{B})$ be measurable spaces and ρ an $\mathcal{A}$-measurable measure on $\mathcal{B}$.

Let $h \in \mathfrak{F}(\mathcal{A}\times\mathcal{B})$.

Define $g(x) = \int h(x,y)\rho(dy|x)$.

Then $g \in \mathfrak{F}(\mathcal{A})$.

Proof: For each $x \in \chi$, $h(x,\circ)$ is $\mathcal{B}$-measurable. The integral defining g is thus well-defined for all x.

Since h is bounded, $|h| \leq M$ for some $0 < M < \infty$. Now

$$|g(x)| = M\left|\int \frac{h(x,y)}{M}\rho(dy|x)\right| \leq M\|\rho(\circ|x)\|$$

by definition 10. The boundedness of g is now a consequence of definition 12(i). It remains to prove that g is measurable. We prove that g is measurable whenever h is an indicator function. The lemma then follows by standard extension to simple functions and monotone limits of simple functions.

The collection $\mathcal{D}$ of sets in $\mathcal{A} \times \mathcal{B}$ such that the lemma holds for $h = I_D$ (the indicator function of D) is a λ-system:

(λ1) is trivially satisfied.

(λ2) holds since

$$\int I_{D^c}(x,y)\rho(dy|x) = \int[1-I_D(x,y)]\rho(dy|x)$$

$$= \rho(\mathcal{Y}|x)-\int I_D(x,y)\rho(dy|x)$$, which is a measurable

function of x if $D \in \mathcal{D}$.

(λ3) follows since $D_1, D_2 \in \mathcal{D}$, $D_1 \cap D_2 = \emptyset$, implies $I_{D_1 \cup D_2} = I_{D_1}+I_{D_2}$.

(λ4): Assume $D_1 \subseteq D_2 \subseteq \dots$ and that

$g_n(x) = \int I_{D_n}(x,y)\rho(dy|x)$ are measurable function for

$n = 1,2,\dots$. Set $D = \bigcup_{i=1}^{\infty} D_i$.

Obviously $I_{D_n} \uparrow I_D$. Application of the monotone convergence theorem to the positive and negative part of the signed measure $\rho(\circ|x)$ yields $g_n(x) \to \int I_D(x,y)\rho(dy|x)$, $x \in \chi$

which proves that $D \in \mathcal{D}$, since the limit of a sequence of measurable functions is measurable.

Let $\mathcal{E}$ be the set of rectangles $A \times B$, $A \in \mathcal{A}$, $B \in \mathcal{B}$.

It is now enough to prove that $\mathcal{E} \subseteq \mathcal{D}$. If this is the case, then since $\mathcal{E}$ is a π-system,

$$\mathcal{A} \times \mathcal{B} \underset{\text{def}}{=} \sigma(\mathcal{E}) = \lambda(\pi(\mathcal{E})) = \lambda(\mathcal{E}) \subseteq \mathcal{D}$$

so the lemma holds for all $I_D : D \in \mathcal{A} \times \mathcal{B}$.

Let now $C = A \times B$, $A \in \mathcal{A}$, $B \in \mathcal{B}$. Define

$h(x,y) = I_C(x,y) = I_A(x)I_B(y)$. Then

$$g(x) = \int h(x,y)\rho(dy\,|x) = \int I_A(x)I_B(y)\rho(dy\,|x)$$

$$= I_A(x)\int_B \rho(dy\,|x) = I_A(x)\rho(B\,|x)$$

which is a product of two measurable functions and hence -measurable.

Thus $\mathcal{E} \subseteq \mathcal{D}$.

NOTATIONS

Let f be a measurable function on the measure space $(\chi,\mathcal{A},\mu)$. The following notations may all be used for the integral of f w.r.t. μ :

$\int f d\mu$ $\quad\int (d\mu)f$ $\quad\int f(x)\mu(dx)$

$\int (\mu(dx))f(x)$ $\quad\mu(f)$ $\quad\mu f$

$f\mu$ $\quad(f)\mu$

For $C \in \chi \times \mathcal{Y}$ we define

$C_x = \{y \in \mathcal{Y} : (x,y) \in C\}$, called the section of C w.r.t. x .

From now on, let $(\chi,\mathcal{A})$ and $(\mathcal{Y},\mathcal{B})$ be given measurable spaces.

LEMMA 14.

Let $\mu \in \mathcal{M}(\mathcal{A})$ and let ρ be an $\mathcal{A}$-measurable measure on $\mathcal{B}$. Define $\mu \times \rho$ on $\mathcal{A} \times \mathcal{B}$ by

(3) $\quad \mu \times \rho(C) = \int \rho(C_x|x)\mu(dx)$ for $C \in \mathcal{A} \times \mathcal{B}$

Then $\mu \times \rho$ is the unique measure on $\mathcal{A} \times \mathcal{B}$ such that

(4) $\quad \mu \times \rho(A \times B) = \int_A \rho(B|x)\mu(dx)$

for each rectangle $A \times B \in \mathcal{A} \times \mathcal{B}$.

<u>Remark</u>: If $\rho(\circ|x)$ is independent of x , then ρ may be considered as a measure on $\mathcal{B}$ and in this case (3) is nothing but the usual product measure on $\mathcal{A} \times \mathcal{B}$.
(4) may then be written $\mu \times \rho(A \times B) = \mu(A)\rho(B)$.

<u>Proof</u>: Let $A \times B \in \mathcal{A} \times \mathcal{B}$. Then $(A \times B)_x = \begin{cases} B & \text{if } x \in A \\ \emptyset & \text{if } x \notin A \end{cases}$.
Hence (4) follows from (3) by letting $C = A \times B$.
It now suffices to prove that $\mu \times \rho$ as defined in (3) is a finite signed measure. The uniqueness will then follow by the extension theorem for measures (see eg. Royden: Real Analysis Ch. 12.2). Since $\rho(C_x|x) = \int I_C(x,y)\rho(dy|x)$, the measurability of $\rho(C_x|x)$ follows from lemma 13.
Let $C_1, C_2, \dots$ be a sequence of disjoint sets in $\mathcal{A} \times \mathcal{B}$.
Since

$$\left(\bigcup_{i=1}^{\infty} C_i\right)_x = \bigcup_{i=1}^{\infty} (C_i)_x \quad \text{and}$$

$C_{1x}, C_{2x}, \dots$ are disjoint, we have

$$\mu \times \rho\left(\bigcup_{i=1}^{\infty} C_i\right) = \int \rho\left(\left(\bigcup_{i=1}^{\infty} C_i\right)_x \Big| x\right)\mu(dx) = \int \sum_{i=1}^{\infty} \rho(C_{ix}|x)\mu(dx) .$$

Set $f_n(x) = \sum_{i=1}^{n} \rho(C_{ix}|x)$.

Now

$$|f_n(x)| = |\sum_{i=1}^{n} \rho(C_{ix}|x)| = |\rho((\bigcup_{i=1}^{n} C_i)_x|x)| \leq \|\rho(\circ|x)\|$$
$$\leq \sup_x \|\rho(\circ|x)\|$$

so the dominated convergence theorem may be applied to give

$$\mu \times \rho(\bigcup_{i=1}^{\infty} C_i) = \int \lim_n f_n(x)\mu(dx) = \lim_n \int f_n(x)\mu(dx)$$
$$= \sum_{i=1}^{\infty} \int \rho(C_{ix}|x)\mu(dx) = \sum_{i=1}^{\infty} \mu \times \rho(C_i) .$$

PROPOSITION 15.

Let the situation be as in lemma 14 and let $h \in \mathcal{M}(\mathcal{O\!\!L} \times \mathcal{B})$. Then

(5) $\int hd(\mu \times \rho) = \int [\int h(x,y)\rho(dy|x)]\mu(dx)$.

Proof: The expression on the right is well-defined by lemma 13. By the preceding lemma, (5) holds for indicator functions. The statement follows by standard extension to simple functions and monotone limits of simple functions.

Remark: If $\rho(\circ|x)$ is independent of x , then (5) states the same as Fubini's theorem.

COROLLARY 16.

$\mu \times \rho \in \mathcal{M}(\mathcal{O\!\!L} \times \mathcal{B})$. In fact, $\|\mu \times \rho\| \leq \|\mu\| \sup_x \|\rho(\circ|x)\|$.

Proof: Let $h \in \mathcal{M}(\mathcal{O\!\!L} \times \mathcal{B})$, $\|h\| \leq 1$.

Then

$$\int hd(\mu \times \rho) = \int [\int h(x,y)\rho(dy\,|x)]\mu(dx) \leq \|\mu\| \sup_x \|\rho(\circ\,|x)\|$$

by definition 10, since for each $x \in \chi$

$$|\int h(x,y)\rho(dy\,|x)| \leq \|\rho(\circ\,|x)\|$$

(again by def. 10).

DEFINITION 17.

Let $\mu \in \mathcal{M}(\mathfrak{A})$ and let ρ be an $\mathfrak{A}$ - measurable measure on $\mathfrak{B}$. We define $\mu\rho$ on $\mathfrak{B}$ by

$$\mu\rho(B) = \mu \times \rho(\chi \times B) = \int \rho(B|x)\mu(dx) \text{ for } B \in \mathfrak{B}.$$

$\mu\rho$ is obviously a measure, since $\mu \times \rho$ is .

Remark: If $\rho(B\,|x)$ is considered as the conditional probability of the event B , given x , and μ is the probability distribution of x , then $\mu\rho$ is simply the unconditional probability.

PROPOSITION 18.

(i) Let $f \in \mathfrak{F}(\mathfrak{B})$. Then

$$(\mu\rho)(f) = \int [\int f(y)\rho(dy\,|x]\mu(dx)$$

(ii) $\|\mu\rho\| \leq \|\mu\| \circ \sup_x \|\rho(\circ\,|x)\|$. Thus $\mu\rho \in \mathcal{M}(\mathfrak{B})$.

Equality sign holds if ρ is a Markov-kernel and $\mu \geq 0$.

(iii) The mapping $\mu \to \mu\rho$ is a linear mapping $\mathcal{M}(\mathfrak{A}) \to \mathcal{M}(\mathfrak{B})$.

Proof: By the definition of $\mu\rho$, (i) clearly holds if f is an indicator function. (i) is now proved by the standard extension procedure.

The proof of the inequality in (ii) is similar to that of corollary 17.

Assume now that ρ is a Markov kernel and μ a non-negative measure. Then $\mu\rho$ is obviously a non-negative measure, so

$$\|\mu\rho\| = (\mu\rho)(\mathcal{Y}) = \int\rho(\mathcal{Y}\,|x)\mu(dx) = \int\mu(dx) = \|\mu\| \;.$$

The linearity of the mapping $\mu \to \mu\rho$ follows at once from the definition.

<u>DEFINITION 19.</u>

Let $g \in \mathfrak{F}\,(\mathcal{B})$ and let ρ be an $\mathcal{A}$-measurable measure on $\mathcal{B}$. Define for each $x \in \chi$

$$(\rho g)_x = \int g(y)\rho(dy\,|x)$$

<u>Remark</u>: If ρ is considered as a conditional probability then $(\rho g)_x$ is the conditional expectation of g , given x .

<u>PROPOSITION 20.</u>

(i) $\rho g \in \mathfrak{F}\,(\mathcal{A})$

(ii) The mapping $g \to \rho g$ is a linear mapping $\mathfrak{F}\,(\mathcal{B}) \to \mathfrak{F}\,(\mathcal{A})$.

<u>Proof</u>: (i) follows from lemma 13.

The linearity of the mapping $g \to \rho g$ is obvious by definition 19.

<u>PROPOSITION 21.</u>

Let $\mu \in \mathcal{M}\,(\mathcal{A})$, $g \in \mathfrak{F}\,(\mathcal{B})$ and ρ be an $\mathcal{A}$-measurable measure on $\mathcal{B}$. Then

$$(\mu\rho)(g) = \mu(\rho g) \;.$$

Hence parantheses may be omitted and the expression $\mu\rho g$ is well-defined as a bilinear functional in μ and g .

Proof: $\mu(\rho g) = \int(\rho g)_x\mu(dx) = \int[\int g(y)\rho(dy|x)]\mu(dx) = (\mu\rho)(g)$

by prop. 19 (i).

Remark: Let μ and ρ have the same meaning as in the remark succeeding definitions 17 and 19. Then prop. 21 states that the expectation of g may be found either by integrating g w.r.t. the unconditional probability $\mu\rho$ or by integrating the conditional expectation of g given x w.r.t. the probability distribution μ of x .

APPENDIX C

THE WEAK COMPACTNESS LEMMA.

DEFINITION 1

An indexed family $\{\delta_\alpha : \alpha \in I\}$ of real random variables on a probability space $(\chi, \mathcal{O}\!\!\!\!l, P)$ is said to be uniformly integrable if

$$\sup_{\alpha\in I} \int_{|\delta_\alpha|\geq c} |\delta_\alpha| dP \to 0 \quad \text{when} \quad c \to \infty .$$

PROPOSITION 2

The family $\{\delta_\alpha\}$ is uniformly integrable if and only if the following two conditions are satisfied:

(i) $\sup\limits_{\alpha} \int |\delta_\alpha| dP < \infty$

(ii) To any $\epsilon > 0$ there is a $\eta_\epsilon > 0$ such that

(1) $P(A) < \eta_\epsilon \Rightarrow |\int_A \delta_\alpha dP| < \epsilon$ for all $\alpha \in I$.

Proof: "only if":

For a suitable $c > 0$, $\sup\limits_{\alpha} \int_{|\delta_\alpha|\geq c} |\delta_\alpha| < 1$.

Hence, for any $\alpha \in I$,

$$\int |\delta_\alpha| = \int_{|\delta_\alpha|<c} |\delta_\alpha| + \int_{|\delta_\alpha|\geq c} |\delta_\alpha| \leq c + 1 \text{ , so (i) follows.}$$

Let $\epsilon > 0$ and choose c so that

$$\sup_{\alpha} \int_{|\delta_\alpha|\geq c} |\delta_\alpha| < \epsilon/2$$

Let $A \in \mathcal{O}\!\!\!\!l$. Then

$$|\int_A \delta_\alpha| \leq \int_A |\delta_\alpha| = \int_{A\cap\{|\delta_\alpha|<c\}} |\delta_\alpha| + \int_{A\cap\{|\delta_\alpha|\geq c\}} |\delta_\alpha| \leq cP(A) + \epsilon/2$$

Choosing $\eta_\epsilon = \frac{\epsilon}{2c}$ and $P(A) < \eta_\epsilon$ yields $|\int_A \delta_\alpha| < \epsilon$ for any $\alpha \in I$.

"if": Let $\epsilon > 0$ and choose η so that (1) holds. Then, if $P(A) < \eta$, for any $\alpha \in I$,

$$(2) \quad \int_A |\delta_\alpha| = \int_{A\cap\{\delta_\alpha \geq 0\}} |\delta_\alpha| + \int_{A\cap\{\delta_\alpha < 0\}} |\delta_\alpha|$$

$$= |\int_{A\cap\{\delta_\alpha \geq 0\}} \delta_\alpha| + |\int_{A\cap\{\delta_\alpha < 0\}} \delta_\alpha| < 2\epsilon$$

By the generalized Chebycheff's inequality,

$$P(|\delta_\alpha| \geq c) \leq \frac{E|\delta_\alpha|}{c} \leq \frac{\sup_\alpha \int|\delta_\alpha|}{c}$$

so for some $c > 0$ $P(|\delta_\alpha| \geq c) < \eta$ for all $\alpha \in I$ by (i). Hence, by letting $A = \{|\delta_\alpha| \geq c\}$ in (2), uniform integrability is proved.

THEOREM 3 (THE WEAK COMPACTNESS LEMMA)

Let $(\chi, \mathcal{A}, P)$ be a probability space and let $\{\delta_\alpha : \alpha \in I\}$ be a net (generalized sequence) which is uniformly integrable. Then there is a subnet $\{\delta_\beta\}$ and an integrabel δ so that

$$\int \delta_\beta h dP \to \int \delta h dP \quad \text{for all} \quad h \in L_\infty(P)$$

(i.e. for all (essentially) bounded measurable functions h).

Proof: It is enough to consider non-negative δ's. Let $\|h\| = \text{ess sup } h$; $h \in L_\infty(P)$.

For each $\alpha \in I$ we define a linear functional F_α on $L_\infty(P)$ by

$$F_\alpha(h) = \int \delta_\alpha h dP$$

Now, $|F_\alpha(h)| \leq \|h\| \int \delta_\alpha dP \leq c \cdot \|h\|$

where $c = \sup_\alpha \int \delta_\alpha dP < \infty$ by prop. 2.

Hence $\|F_\alpha\| \leq c$ for any $\alpha \in I$ (i.e. F_α is continous).

and $F_\alpha \in \prod_{h \in L_\infty(P)} [-c\|h\| , c\|h\|]$

which is compact by Tychonoff's theorem (Royden Ch. 9).

Hence, there is a subnet $\{F_\beta\}$ so that

(3) $F_\beta(h) \to F(h)$ for all $h \in L_\infty(P)$ for some functional F.

F is obviously linear, since it is a (pointwise) limit of linear functionals. Furthermore, $\|F\| \leq c$, since

$F(h) \in [-c\|h\| , c\|h\|]$ for each h.

Hence F is continuous.

Since the δ's are non-negative, it follows that $F \geq 0$ (i.e. $h \geq 0 \Rightarrow F(h) \geq 0$).

We will now prove that there is a $\delta \geq 0$ so that $F(h) = \int h\delta dP$ for all $h \in L_\infty(P)$. The lemma will then follow from (3).

Define a set-function φ on $\mathcal{O}\!\mathcal{l}$ by $\varphi(A) = F(I_A)$ for all $A \in \mathcal{O}\!\mathcal{l}$

Clearly $\varphi \geq 0$. By prop. 2 (ii), to any $\epsilon > 0$ there is a $\eta > 0$ such that $P(A) < \eta \Rightarrow \int \delta_\beta I_A < \epsilon$ for all β.

This is equivalent to $F_\beta(I_A) < \epsilon$ for all β, which again implies $\varphi(A) = F(I_A) < \epsilon$.

Hence

(4) $P(A) < \eta \Rightarrow \varphi(A) < \epsilon$.

This fact, together with the fact that φ is finitely additive, implies that φ is σ-additive.

Furthermore, $\varphi(\chi) = F(1) = \lim \int \delta_\beta \leq c$.

Thus φ is a finite non-negative measure on $(\chi, \mathcal{O}\!\mathcal{l})$ and by (4), $\varphi << P$.

By Radon-Nikodym's theorem there is a measurable function $\delta \geq 0$ so that

$$F(I_A) = \varphi_A = \int_A \delta dP \ ; \quad A \in \mathfrak{A} .$$

It follows that

$$\int h d\varphi = \int h \delta dP \quad \text{for any} \quad h \in L_\infty(P) .$$

Finally, we have to prove that

(5) $F(h) = \int h d\varphi$ for all $h \in L_\infty(P)$.

By the definition of φ , (5) holds for indicator functions and hence by linearity for simple functions.

Since F is continuous,

$F(\lim h) = \lim F(h)$ and thus (5) follows from the fact that each $h \in L_\infty(P)$ may be written as a sequence of simple functions.

REMARK 4

If we in addition require

$\sup_\alpha \|\delta_\alpha\|_\infty < \infty$, then the lemma will hold for all $h \in L_1(P)$

(i.e. the set of measurable functions which are integrable w.r.t. P).

This is easily proved by approximating $h \in L_1(P)$ with functions in $L_\infty(P)$.

REMARK 5

The weak compactness lemma (and the extension noted in remark 4) has an analogue where "net" is replaced by "sequence" and "subnet" is replaced by "subsequence". The proof of the sequentiel version of the weak compactness lemma may be found in [8] in the case where $\mathfrak{A}$ is separabel (i.e. $\mathfrak{A} = \sigma(\mathfrak{B})$ for a countable family $\mathfrak{B}$ of subsets of χ) . The proof in the general case is given in [17].

REMARK 6

Consider again theorem 3. We shall say that the subnet $\{\delta_\beta\}$ converges weakly to δ if the conclusion of the theorem holds. (Similarly for the sequential case.)

APPENDIX D

RESEARCH PAPERS. ABSTRACTS.

COMPARISON OF EXPERIMENTS WHEN THE PARAMETER SPACE IS FINITE.

By E.N. Torgersen.

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

The convex function criterion for "being more informative" for k-decision problems is generalized to a convex function criterion for ϵ-deficiency for k-decision problems. The particular case of comparison by testing problems is discussed. 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. It is shown that any experiment with finite parameter space is a mixture of complete experiments, and the complete experiments are characterized.

COMPARISON OF TRANSLATION EXPERIMENTS.

By E.N. Torgersen.

Ann. Math. Statist. 43, 1383 - 1399 (1972).

In this paper we treat the problem of comparison of translation experiments. The "convolution divisibility" criterion for "being more informative" by Boll (Ph. D. dissertation, Stanford Univ., 1955) is generalized to a "ϵ-convolution divisibility" criterion for ϵ-deficiency. We also generalize the "convolution divisibility" criterion of V. Strassen (Ann. Math. Statist. 36, 423, 1965) to a criterion for "ϵ-convolution divisibility".

It is shown, provided least favourable "ϵ-factors" can be found, how the deficiencies actually may be calculated. As an application we determine the increase of information - as measured by the deficiency - contained in an additional number of observations for a few experiments (rectangular, exponential, multivariate normal, one way layout). Finally we consider the problem of convergence for the pseudo distance introduced by LeCam (1964) [7]. It is shown that convergence for this distance is topologically equivalent to strong convergence of the individual probability measures up to a shift.

LOCAL COMPARISON OF EXPERIMENTS WHEN THE PARAMETER SET IS ONE DIMENSIONAL.

By E.N. Torgersen.

Statist. Research Report, Inst. of Math., Univ. of Oslo, No. 4, 1972.

This paper treats comparison of experiments within infinitesimal neighbourhoods of a fixed point θ_o in the parameter set. If δ_ϵ is the deficiency in LeCam [7] within $[\theta_o-\epsilon, \theta_o+\epsilon]$, then $\delta_\epsilon/2\epsilon \to \overset{\circ}{\delta}$ as $\epsilon \to 0$ provided strong derivatives exists. Related to $\overset{\circ}{\delta}$ is a pseudo metric $\overset{\circ}{\Delta}$. $\overset{\circ}{\delta}$ is a "deficiency" between pseudo experiments i.e. "experiments" where the basic measures are not necessarily probability measures. Some known results on experiments are extended to pseudo experiments. Various characterizations, deficiencies and pseudo distances for the relevant pseudo experiments are considered. Particularily interesting representations are: probability distributions with expectation zero (this representation converts products to convolutions), concave functions describing the relationship between size and slope for testing "$\theta = \theta_o$" against "$\theta > \theta_o$", and strongly unimodal distributions. Conditional expectation - and factorization criterions for sufficiency are given.

LOCAL COMPARISON OF EXPERIMENTS.

By E.N. Torgersen.

Statist. Research Report, Inst. of Math., Univ. of Oslo, No. 5, 1972.

In this paper we generalize most of the results in Research Report No. 4, 1972 to the case of a finite dimensional parameter set.

MIXTURE AND COMPLETENESS PROPERTIES OF DOMINATED PSEUDO EXPERIMENTS.

By E.N. Torgersen.

Statist. Research Report, Inst. of Math., Univ. of Oslo, No. 7, 1972.

In this paper we generalizes some of the results in section 4 in Torgersen [16] to the case of dominated (pseudo) experiments. Convex combinations of (pseudo) experiments are defined, and it is shown that a (pseudo) experiment has the extreme point property (for Δ_1 equivalence) if and only if it admits a boundedly complete and sufficient sub σ algebras.
Dominated models for independent observations $X_1, \ldots, X_n$ admitting boundedly (or L_p) complete and sufficient statistics, are considered. It is shown that a sub set - say $X_1, \ldots, X_m$ where $m < n$ - has the same property provided a certain regularity condition is satisfied. This condition is automatically satisfied when the observations are identically distributed. The proof - in the case of bounded completeness -

utilizes the fact that products of experiments are distributive w.r.t. mixtures. Somewhat more involved arguments are needed for L_p completeness.

COMPARISON OF LINEAR NORMAL EXPERIMENTS.

By Ole Håvard Hansen and E.N. Torgersen.

Ann. Statist., Vol. 2, No. 2, 367 - 373, 1974.

Consider independent and normally distributed random variables $X_1,\ldots,X_n$ such that $0 < \mathrm{Var}\, X_i = \sigma^2$; $i = 1,\ldots,k$ and $E(X_1,\ldots,X_n)' = A'\beta$ where A' is a known $n \times k$ matrix and $\beta = (\beta_1,\ldots,\beta_k)'$ is an unknown column matrix. [The prime denotes transposition]. The cases of known and totally unknown σ^2 are considered simultaneously. Denote the experiment obtained by observing $X_1,\ldots,X_n$ by $\mathcal{E}_A$. Let A and B be matrices of, respectively, dimensions $n_A \times k$ and $n_B \times k$. Then, if σ^2 is known, (if σ^2 is unknown) $\mathcal{E}_A$ is more informative than $\mathcal{E}_B$ if and only if $AA' - BB'$ is non negative definit (and $n_A \geq n_B + \mathrm{rank}(AA' - BB')$).

ASYMPTOTIC BEHAVIOUR OF POWERS OF DICHOTOMIES.

By E.N. Torgersen.

Statist. Research Report, Inst. of Math., Univ. of Oslo, No. 6, 1974.

Consider random variables $X, Y, \ldots$ whose distributions are known except for an unknown parameter θ belonging to a known two-point set. Let $X_1, X_2, \ldots$ and $Y_1, Y_2, \ldots$ be independent observations of, respectively, X and Y . How does the information yielded by $(X_1, X_2, \ldots, X_n)$ compare with the information yielded by $(Y_1, Y_2, \ldots, Y_n)$ when n is large? Let $\mathcal{M}a$ and $\mathcal{M}i$ denote, respectively, a totally informative and a totally uninformative experiment. Furthermore denote by Δ the distance between experiments introduced by LeCam 1964. Then, for any variable X :

$$1 - \Delta(X,\mathcal{M}i) \leq \Delta(X,\mathcal{M}a) < 2\frac{1-\Delta(X,\mathcal{M}i)}{2-\Delta(X,\mathcal{M}i)}$$

Combining this inequality with Chernoff's result on the exponential rate of asymptotique Baye's risk we find that

$$\sqrt[n]{\Delta((X_1,X_2,\ldots,X_n),(Y_1,Y_2,\ldots,Y_n))} \to \max(c_X,c_Y)$$

provided the experiments defined by X and Y are not equivalent. Here $c_X(c_Y)$ denote the greatest lower bound of the Hellinger transform of $X(Y)$.

In order to obtain inequalities for concave approximations to the kernel of the Hellinger transform, we generalized the sub linear function criterion as follows. Let $(\chi,\mathcal{A})$ and $(\mathcal{Y},\mathcal{B})$ be measurable spaces with, respectively, probability measures P_1,P_2 and Q_1,Q_2 . Suppose the dichotomy $\mathcal{E} = ((\chi,\mathcal{A}),(P_1,P_2))$ is $(\varepsilon_1,\varepsilon_2)$ deficient w.r.t. $\mathcal{F} = ((\mathcal{Y},\mathcal{B}),(Q_1,Q_2))$. Then, for any convex function φ on $[0,1]$,

$$\varepsilon_1[\varphi'(1)-(\varphi(1)-\varphi(0))] + \varepsilon_2[(\varphi(1)-\varphi(0)) - \varphi'(0)] \geq 4\int\varphi d(T-S)$$

where S is the distribution of $dP_2/d(P_1+P_2)$ w.r.t. $(P_1+P_2)/2$ and T is the distribution of $dQ_2/d(Q_1+Q_2)$ w.r.t. $(Q_1+Q_2)/2$. (It follows directly from the testing criterion for comparison that it suffices, in order to verify $(\varepsilon_1,\varepsilon_2)$ deficiency, to consider functions φ of the form: $X \to |X-\theta|$ where $\theta \in]0,1[$.)

COMPARISON OF EXPERIMENTS BY FACTORIZATION

By E.N. Torgersen

Statist. Research Report, Inst. of Math., Univ. of Oslo, No. 3, 1974.

Consider random variables $X, Y, \ldots$ whose distributions are known except for an unknown parameter θ belonging to a known finite set Θ. Identify each variable with the experiment it defines and write $X \sim Y$ if X and Y are equally informative. We give first, for given X and Y, a functional criterion for the existence of a Z, independent of X, such that $Y \sim (X,Z)$. Combining this with a result on consistent families of experiments, we prove that X has the property that any more informative Y is $\sim (X,Z)$ for some Z independent of X if and only if there is a $\tilde{X} \sim X$ such that:

(i) $\tilde{X}$ is, with probability 1, a non empty sub set of Θ.

(ii) Each θ belongs to some possible value of $\tilde{X}$.

(iii) If $U_1 \neq U_2$ are possible values of $\tilde{X}$ then $\#(U_1 \cap U_2) \leqq 1$.

(iv) If $U_{n+1} = U_1, U_2, \ldots, U_n$ are n possible values of $\tilde{X}$ such that $U_i \cap U_{i+1} \neq \emptyset$; $i = 1, \ldots, n$ then $\bigcap_i U_i \neq \emptyset$.

References.

[1] Blackwell, D. (1951). Comparison of experiments. Proc. Second Berkeley Sympos. math. Statist. Probab. 93 - 102.

[2] Blackwell, D. (1953). Equivalent comparisons of experiments. Ann. Math. Statist. 24, 265 - 272.

[3] Blackwell, D and Girshick, M.A. (1954). Theory of Games and Statistical Decisions. John Wiley & Sons.

[4] Ferguson, Th.S. (1967). Mathematical Statistics. Academic Press.

[5] Hansen, O.H. and Torgersen, E.N. (1974). Comparison of linear normal experiments. Ann. Statist. Vol. 2., No. 2, 367 - 373.

[6] Heyer, H. (1973). Mathematische Theorie statistischer Experimente. Springer-Verlag.

[7]*) Le Cam, L. (1964). Sufficiency and approximate sufficiency. Ann. Math. Statist. 35, 1419 - 1455.

[8] Lehmann, E.L. (1959). Testing statistical hypotheses. New York, Wiley.

[9] Loève, M. (1963). Probability Theory. Van Nostrand Reinhold.

[10] Neveu, J. (1965). Mathematical foundations of the calculus of probability. San Francisco. Holden-Day.

[11] Royden, H.L. (1969). Real Analysis. Second Edition. The Macmillan Company, London.

[12] Rudin, W. (1964). Principles of Mathematical Analysis. Second Edition. McGraw-Hill.

*) See also [21]

[13] Rudin, W. (1970). Real and Complex Analysis. McGraw-Hill.

[13a] Sion, M. (1958). On general minimax theorems. Pacific J. Math. Vol. 8, 171-176.

[14] Sverdrup, E. (1966). The present state of the decision theory and the Neyman-Pearson Theory. Review of the International Statistical Institute. 34: 3, 309-333.

[15] Sverdrup, E. (1969). Multiple Decision Theory. Lecture Notes Series No. 15, University of Aarhus.

[16] Torgersen, E.N. (1970). Comparison of experiments when the parameter space is finite. Z. Wahrscheinlichkeitstheorie Verw. Geb. 16, 219-249.

[17] Torgersen, E.N. (1971). The separability condition in the weak compactness lemma. Statistical Research Report No. 7, Dep. of Mathemat., Univ. of Cslo.

[18] Torgersen, E.N. (1972). Comparison of translation experiments. Ann. Math. Stat. 43, 1383-1399.

[19] Torgersen, E.N. (1972). Local comparison of experiments when the parameter set is one dimensional. Statis. Research Rep. No. 4, Dep. of Math. Univ. of Oslo.

[20] Wald, A. and Wolfowitz, J. (1951). Two methods of randomization in statistics and the theory of games. Ann. Math. 53, 581-586.

[21] Le Cam, L. (1974). Notes on asymptotic methods in statistical decision theory. Centre de Recherches Math., Univ. de Montréal.