STATISTICAL RESEARCH REPORT No 4
Institute of Mathematics 1972
University of Oslo

LOCAL COMPARISON OF EXPERIMENTS WHEN THE PARAMETER SET IS ONE DIMENSIONAL

by

Erik N. Torgersen

ABSTRACT

LOCAL COMPARISON OF EXPERIMENTS WHEN THE PARAMETER SET IS ONE DIMENSIONAL.

by

Erik N. Torgersen

University of Oslo

This paper treats comparison of experiments within infinitesimal neighbourhoods of a fixed point θ_0 in the parameter set. If δ_ϵ is the deficiency in LeCam [Ann. Math. Statist. 35 (1964), 1419-1455] within $[\theta_0-\epsilon, \theta_0+\epsilon]$, then $\delta_\epsilon/2\epsilon \to \dot{\delta}$ as $\epsilon \to 0$ provided strong derivatives exists. Related to $\dot{\delta}$ is a pseudo metric $\dot{\Delta}$. $\dot{\delta}$ 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 "$\theta = \theta_0$" against "$\theta > \theta_0$", and strongly unimodal distributions. Conditional expectation - and factorization criterions for sufficiency are given.

CONTENTS

1. Introduction.

This paper treats local comparison of experiments.

An experiment will here be defined as a pair $\mathcal{E} = ((\chi,\mathcal{A}), (P_\theta:\theta \in \Theta))$ where $(\chi,\mathcal{A})$ is a measurable space and $P_\theta:\theta \in \Theta$ is a family of probability measures on $\mathcal{A}$. If $\mathcal{E} = ((\chi,\mathcal{A}), (P_\theta:\theta \in \Theta))$ then $(\chi,\mathcal{A})$ is the sample space of $\mathcal{E}$ and Θ is the parameter set of $\mathcal{E}$.

"Local" refers to restrictions to small neighbourhoods of a fixed point θ_0 in the parameter set Θ. The emphasize in this paper will be on one dimensional parameter sets, and it will be assumed - unless otherwise stated - that the parameter set Θ is a set of real numbers.

This paper is based on results in Blackwell [1] and [2], in LeCam [7] and in Torgersen [15]. LeCam extended the concept of "being more informative", treated by Blackwell in [1] and [2], to the concept of ϵ-definiency and introduced a definiency δ and a distance Δ. It turned out, however, that the set up in [7] was not quite general enough to cover the situations encountered in this paper. For reasons, to be explaned below, we needed a theory for "experiments" where the basic measures are not necessarily probability measures. Such "experiments" will be called pseudo experiments and we refer to appendix B for complete definitions.

A theory for pseudo experiments had, with another motivation, been attempted in [14]. Some of the results in [14] are, together with a few additional results, included with proofs, in appendix B. Pseudo experiments appears in connection with local comparison as follows:

Consider two experiments $\mathcal{E}$ and $\mathcal{F}$, each having the same k-dimensional parameter set Θ. Let θ_0 be an interior point of Θ and let δ_ε be the deficiency of $\mathcal{E}$'s restriction to the ϵ-ball with center θ_0 with respect to the same restriction of $\mathcal{F}$. Then, under differentiability conditions, $\delta_\varepsilon/2\epsilon$ tends to a limit $\dot{\delta}$ as $\epsilon \to 0$. The number $\dot{\delta}$ may be interpreted as the local deficiency of $\mathcal{E}$ w.r.t. $\mathcal{F}$ in the point θ_0. $\dot{\delta}$ can - in general - not be a deficiency since it may be arbitrarily large while ordinary deficiencies are in [0,2]. It may be shown, however, that $\dot{\delta}$ is a deficiency of one pseudo experiment $\dot{\mathcal{E}}_{\theta_0}$ w.r.t. another pseudo experiment $\dot{\mathcal{F}}_{\theta_0}$. If k=1 then the pseudo experiment $\dot{\mathcal{E}}_{\theta_0}$ consists of two parts, the distribution of the observations when $\theta=\theta_0$ and the derivative, in θ_0, of this distribution. (The role of the "derivatives" $\dot{\mathcal{E}}_{\theta_0}$ resembles somewhat that of mass and momentum in mechanics.) The experiment $\dot{\mathcal{E}}_{\theta_0}$ will - when k=1 - be called the derivative of $\mathcal{E}$ in θ_0.

A symptotic local comparison is treated by LeCam in [8]. Our approach is - in the asymptotic case - different from that in [8]. While LeCam considered infinitesimal neighbourhoods of any point $\theta \in \Theta$ we restrict ourselves to infinitesimal neighbourhoods of one fixed point $\theta_0 \in \Theta$. We do not try to put the pieces together in order to get global results. Section 7 is a exception since the class of experiments treated there have the property that "local" comparisons coinsides with "everywhere local" comparison.

It will be seen from appendix B that the existence of various randomizations (a precise definition is given at the end of this section) are only proved under the assumption that some measurable spaces are Borel subsets of Polish spaces. This assumption is - when it is used - explicitely stated in the appendixes. In the text, however, the assumption is not explicitely stated. The only results whose proofs requires such an assumption are propositions 2.3, 3.1, 3.4, 4.11, 6.5, theorems 6.1, 6.2, 6.6 and corollary 6.3. It is, however, shown in appendix C that proposition 2.3, 3.1 and 3.4 have - slightly more complicated - proofs which does not depend on any assumption of this type. The same is true for proposition 4.11 provided it is reformulated so that condition (iii) is deleted.

Section by section the content of this paper is as follows.

The basic differentiability conditions are introduced in section 2. Experiments satisfying them will be called differentiable. Sufficient conditions for differentiability may - with a little rewording - be taken from II. 4.8 in Hájek abd Šidák [4]. It is shown that products of differentiable experiments are differentiable and that sub experiments of differentiable experiments are differentiable.

The concept of a derivative of an experiment is introduced in section 3. We discuss which ordered pairs of finite measures are derivatives and it is shown that the obvious necessary conditions are also sufficient. A few characterizations of the derivative are considered. It is - particular - shown that a derivative may, up to equivalence, be characterized by a probability measure having expectation zero. The probability measure is, essentially, a version of the derivative . This representation converts products into convolutions. We have in this paper, however, not considered central limit problems.

Some basic properties of derivatives are derived in section 4. It is shown how a derivative may be represented by, either a convex function on $]-\infty,\infty[$ or a concave function on $[0,1]$. The last representation is, essentially, a version of the derivative. It describes the relationship between size and slope in θ_0 for power functions of tests for testing "$\theta = \theta_0$" against "$\theta > \theta_0$". The collection of derivatives is a "lattice" for the ordering "being more informative", and maxima are represented by pointwise maxima of convex functions while minima are represented by pointwise minima of the concave functions. We consider two types of deficiencies - $\dot{\delta}$ and δ - and their related pseudometrics $\dot{\Delta}$ and Δ. δ and Δ are - mathematically - natural extensions of δ and Δ in LeCam's paper [7]. $\dot{\Delta}$ is - up to the multiplicative factor $\frac{1}{2}$ - the sup norm distance between the convex functions, and it is exactly equal to the sup norm distance between the concave functions. Various criterions for "being more informative" in the $\dot{\delta}$ sence are given. In particular we derive the factorization criterion for sufficiency for these deficiencies. A few simple conditions for symmetry are given at the end of this section.

Convergence properties of the pseudo metrics *) Δ and $\dot{\Delta}$, on the collection of derivatives are studied in section 5. Δ and $\dot{\Delta}$ are topologically equivalent. Δ does, however, generate a larger uniformity than $\dot{\Delta}$. Convergence criterions and compactness criterions are given in terms of the various representations.

---

*) Δ is, in this paper, used both as a pseudo metric on the collection of derivatives and as a pseudo metric on the collection of experiments. Which interpretation is the correct one - at any particular appearance - should be clear from the text.

It is shown that $\dot{\Delta}$ is complete while Δ is not. Using essentially the approach in [15] we obtain **criterions** for asymptotic sufficiency. A convergence onterion for random variables, of independent interest, is derived and applied to the problem of asymptotic sufficiency.

The theory in section 2-5 is, in section 6, connected with the statistical theory of information. It is shown that the deficiency within $[\theta_0-\epsilon, \theta_0+\epsilon]$ divided by 2ϵ tends to the $\dot{\delta}$ deficiency between the derivatives in θ_0 as $\epsilon \to 0$. It follows that the Δ distance within $[\theta_0-\epsilon, \theta_0+\epsilon]$ divided by 2ϵ tends to the $\dot{\Delta}$ distance between the derivatives. It is shown that the "differentiated" distance $\dot{\Delta}$ (and deficiency $\dot{\delta}$) is determined by restrictions to the two point sets $\{\theta_0-\epsilon, \theta_0+\epsilon\}$; $\epsilon > 0$; i.e. to dichotomies. Similar results are proved for the one sided intervals $[\theta_0-\epsilon, \theta_0]$ and $[\theta_0, \theta_0+\epsilon]$. Inequalities for products of experiment - similar to those in remark 3 after corollary 4 in [15] - are derived for $\dot{\delta}$ and $\dot{\Delta}$. It is shown how $\dot{\delta}$ may be expressed by local comparison of operational characteristics. The theory developed so far is compared with the theory of locally most powerful tests. Some well known facts on locally most powerful tests are - for the sake of completeness - included. We show how the deficiency $\dot{\delta}$ and the distance $\dot{\Delta}$ may be expressed in terms of locally most powerful tests. We generalize slightly - in an example - some of the theory in II.4.11 in Hájêk and Šidák [4] in order to illustrate that $\dot{\delta}$ is not fine enough to distinguish experiments such that the differences in local behaviours are small of the second order. It is shown how local comparison may be expressed in terms of powers of most powerful tests for a simple hypotheses against a simple alternative.

Necessary and sufficient conditions for local (i.e. $\dot{\Delta}$) sufficiency in terms of conditional expectations are given. The final results in section 6 are concerned with a change of parameter - in particular of scale change.

The case of differentiable translation experiments on the real line is treated in section 7. This particular case turns out to be not so particular since any differentiable experiment - which is not $\dot{\Delta}$ equivalent with a minimum information experiment - is $\dot{\Delta}$ equivalent with a strongly unimodal translation experiment. The strongly unimodal distribution is unique up to Δ equivalence, i.e. up to a shift. This result is based on a theorem of Ibragimov [6]. The first part of section 7 treats a particular class of functions. These functions are obtained by integrating the φ functions in Hájek and Šidák [4] and they are, essentially, versions of the derivative. It is shown that the Δ distance of LeCam is topologically equivalent with the $\dot{\Delta}$ distance provided we restrict attention to strongly unimodal distributions. Convergence is then implied by weak shift convergence of distributions and implies uniform shift convergence of densities. A simple sufficient condition for $\dot{\Delta}$ convergence within the class of all differentiable translation experiments is given.

Three appendixes - A, B and C are included after section 7.

Appendix A summarizes - without proofs - some of the results on translation experiments in [16].

Appendix B is a self contained introduction to some basic results on comparison of pseudo experiments.

The purpose of appendix C is - as explained above - to point out the results whose proofs depends on assumptions stating that some of the measurable spaces involved are Borel sub sets of Polish spaces.

Probabilities and more generally, measures are occasionally computed as follows:

Let $(\chi,\mathcal{A})$ be a measurable space, $\mathcal{B}$ a sub σ-algebra of $\mathcal{A}$, P a probability measure on $\mathcal{A}$ and μ a finite measure on $\mathcal{A}$ which is dominated by P. Denote by $P_{\mathcal{B}}$ and $\mu_{\mathcal{B}}$ the restrictions of, respectively, P and μ to $\mathcal{B}$. Then:

$$d\mu_{\mathcal{B}}/dP_{\mathcal{B}} = E_P^{\mathcal{B}}(d\mu/dP)$$

so that

$$\mu(B) = \int_B E_P^{\mathcal{B}}(d\mu/dP)dP \; ; \quad B \in \mathcal{B}$$

A randomization (Markov hernel) from a measurable space $(\chi,\mathcal{A})$ to a measurable space $(\mathcal{Y},\mathcal{B})$ will here be defined as a map $(x,B) \rightsquigarrow \rho(B|x)$ from $\chi \times \mathcal{B}$ to $[0,1]$ such that $\rho(B|\cdot)$ is measurable for each $B \in \mathcal{B}$ and $\rho(\cdot|x)$ is a probability measure for each $x \in \chi$. Let ρ be a randomization from $(\chi,\mathcal{A})$ to $(\mathcal{Y},\mathcal{B})$, let μ be a finite measure on $\mathcal{A}$ and let g be a bounded measurable function on $\mathcal{Y}$. Then we may define a finite measure $\mu\rho : B \rightsquigarrow \int \mu(dx)\rho(B|x)$, on $\mathcal{B}$ and a bounded measurable function $\rho g : x \rightsquigarrow \int \rho(dy|x)g(y)$, on $\mathcal{A}$. It is not difficult to see that *) $(\mu\rho)(g) = \mu(\rho g)$ and this number will therefor be written $\mu\rho g$. Finally randomizations may be composed as follows: Let ρ be a randomization from $(\chi, \mathcal{A})$ to $(\mathcal{Y}, \mathcal{B})$ and let σ be a randomization from $(\mathcal{Y}, \mathcal{B})$ to $(\mathcal{Z}, \mathcal{C})$. Then the composite, $\rho\sigma$, is the randomization: $(x, C) \leadsto \int \sigma(C|y)\rho(dy|x)$ from $(\chi, \mathcal{A})$ to $(\mathcal{Z}, \mathcal{C})$.

*) If $(\chi, \mathcal{A}, \mu)$ is a measure space and f is a function on $\mathcal{A}$ then the integral of f w.r.t. μ may be written: $\mu(f)$, $\int f d\mu$, $\int f(x)\mu(dx)$ or $\int \mu(dx)f(x)$.

2. The differentiability conditions.

All experiments considered in this paper have - unless otherwise stated - a parameter set Θ, which is a sub set of $]-\infty, +\infty[$ having an interior point θ_0. We shall say that the experiment $\mathcal{E} = (\mathcal{X}, \mathcal{A}; P_\theta : \theta \in \Theta)$ is differentiable in θ_0 if $(P_\theta - P_{\theta_0})/(\theta - \theta_0)$ converges strongly as $\theta \to \theta_0$. More precisely: $\mathcal{E}$ is differentiable in θ_0 if and only if there is a finite measure *) $\dot{P}_{\theta_0}$ so that

$$\lim_{\theta \to \theta_0} \| (P_\theta - P_{\theta_0})/(\theta - \theta_0) - \dot{P}_{\theta_0} \| = 0$$

Writing $\Gamma_{\theta_0,\theta} = (P_\theta - P_{\theta_0})/(\theta - \theta_0) - \dot{P}_{\theta_0}$ we see that the differentiability condition for $\mathcal{E}$ may be rewritten as :

$\mathcal{E}$ is differentiable in θ_0 if and only if there are finite measures $\Gamma_{\theta_0,\theta} : \theta \in \Theta$, so that $\lim_{\theta \to \theta_0} \| \Gamma_{\theta_0,\theta} \| = 0$ and

$$P_\theta = P_{\theta_0} + (\theta - \theta_0)\dot{P}_{\theta_0} + (\theta - \theta_0)\Gamma_{\theta_0,\theta}; \qquad \theta \in \Theta$$

$$\| \Gamma_{\theta_0,\theta} \|; \qquad \theta \in \Theta$$

are - by the inequality: $|\theta - \theta_0| \| \Gamma_{\theta_0,\theta} \| \leq 2 + |\theta - \theta_0| \| \dot{P}_{\theta_0} \|$ - automatically bounded.

---

*) A measure on a σ-algebra $\mathcal{A}$ of sets is here defined as a real valued σ-additive function on $\mathcal{A}$. The term, signed measure, will not be used.

Before proceeding let us demonstrate that - together - conditions (i) - (iv) below assures the strong convergence of $(P_\theta - P_{\theta_o})/(\theta-\theta_o)$ as $\theta \to \theta_o$.

(i) There exist a positive number c and a positive measure μ so that P_θ is defined and dominated by (i.e.: has densities w.r.t.) μ when $|\theta-\theta_o| \leqq c$.

(ii) There are real valued densities

$$f_\theta = dP_\theta / d\mu \quad : \quad |\theta-\theta_o| \leqq c$$

so that the maps $\theta \mapsto f_\theta(x)$ from $[\theta_o - c, \theta_o + c]$ to $[-\infty, +\infty]$ are - for μ almost all x - absolutely continuous.

(iii) For μ almost all x $\lim_{\theta\to\theta_o} (f_\theta(x) - f_{\theta_o}(x))/(\theta-\theta_o)$ exists.

(iv) $\lim_{\theta\to\theta_o} \int |\dot{f}_\theta(x)|\mu(dx) = \int |\dot{f}_{\theta_o}(x)|\mu(dx) < \infty$

where dots indicate differentiation w.r.t. θ.

These conditions, as well as the demonstration below, are adapted from II.4.8 in Hájek and Sidák [4]

Demonstration:

Let $N \in \mathcal{A}$ be a common exceptional μ-null set for (ii) and (iii). By (ii) the map $(x,\theta) \rightsquigarrow f_\theta(x)$ from $\complement N \times [\theta_o - c, \theta_o + c]$ is jointly measurable in (x,θ). It follows that the map $(x,\theta) \rightsquigarrow \hat{f}_\theta(x) \overset{\text{def}}{=} \limsup\limits_{n \to \infty} n(f_{\theta+1/n}(x) - f_\theta(x))$ is jointly measurable on $\complement N \times \,]\theta_o - c, \theta_o + c[$. By (ii) $\hat{f}_\theta(x) = \dot{f}_\theta(x)$ for almost (Lebesgue) all θ in $[\theta_o - c, \theta_o + c]$, for all $x \in \complement N$.

For any $\theta \in [\theta_o - c, \theta_o + c]$ we have:

$$\int \left| \frac{f_\theta(x) - f_{\theta_o}(x)}{\theta - \theta_o} \right| \mu(dx) = \int \frac{1}{|\theta - \theta_o|} \left| \int_{\langle\theta_o, \theta\rangle} \hat{f}_t(x)\,dt \right| \mu(dx) \quad {}^{*)}$$

$$\leq \int \frac{1}{|\theta - \theta_o|} \left[\int_{\langle\theta_o, \theta\rangle} |\hat{f}_t(x)|\,dt \right] \mu(dx)$$

$$= \text{(by Fubini)} \ \frac{1}{|\theta - \theta_o|} \int_{\langle 0, \theta\rangle} \phi(t)\,dt$$

where $\phi(\theta) = \int |\hat{f}_\theta(x)|\,\mu(dx)$; $\quad |\theta - \theta_o| \leq c$

By (iv) $\phi(\theta) \to \phi(\theta_o)$ as $\theta \to \theta_o$.

Hence $\dfrac{1}{|\theta - \theta_o|} \displaystyle\int_{\langle\theta_o, \theta\rangle} \phi(t)\,dt \to \phi(\theta_o)$ as $\theta \to \theta_o$

---

*) $\langle a,b\rangle = [a,b]$ or $[b,a]$ as $a \leq b$ or $a \geq b$.

It follows that

$$\limsup_{\theta\to\theta_o} \int \left| \frac{f_\theta(x)-f_{\theta_o}(x)}{\theta-\theta_o} \right| \mu(dx) \leq \int |\hat{f}_{\theta_o}(x)| \mu(dx)$$

$$= \int |\dot{f}_{\theta_o}(x)| \mu(dx) < \infty .$$

By Scheffe's convergence theorem [11]

$$\lim_{\theta\to\theta_o} \int \left| (f_\theta(x) - f_{\theta_o}(x))/(\theta-\theta_o) - \dot{f}_{\theta_o}(x) \right| \mu(dx) = 0$$

That is:

$$\lim_{\theta\to\theta_o} \left\| (P_\theta - P_{\theta_o})/(\theta-\theta_o) - \dot{P}_{\theta_o} \right\| = 0$$

where $\dot{P}_{\theta_o}(A) = \int_A \dot{f}_{\theta_o}(x)\mu(dx)$; $A \in \mathcal{A}$ □

<u>Example</u> (Translation experiments)

Let f be an absolutely continuous probability density on R such that $\int |f'(x)| \mu(dx) < \infty$, and let P be the probability measure with density f. The translation experiment $\mathcal{E}_P$ is defined by

$$\mathcal{E}_P = ((]-\infty,+\infty[,\mathcal{A}), \quad P_\theta : \theta \in R)$$

where $\mathcal{A}$ is the Borel class and

$$P_\theta(A) = P(A - \theta) ; \quad A \in \mathcal{A} \quad \theta \in \Theta .$$

Then $f_\theta(x) = f(x-\theta)$, $\dot{f}_\theta(x) = -f'(x-\theta)$ and it may be checked that (i) - (iv) are satisfied. Furthermore:

$$\dot{P}_\theta(A) = \dot{P}_o(A-\theta) = \int_A -f'(x-\theta)dx \; ; \quad A \in \mathcal{A}, \quad \theta \in R$$

The proposition below implies that products of such experiments are differentiable.

Proposition 2.1.

Let $\mathcal{E}_i = (X_i, \mathcal{A}_i, P_{\theta,i};\ \theta \in \Theta)$; $i = 1,2,\cdots,n$ be differentiable in θ_o. Then $\prod_{i=1}^{n} \mathcal{E}_i$ is also differentiable in θ_o and

$$\lim_{\theta\to\theta_o} \Big\| \big(\prod_i P_{\theta,i} - \prod_i P_{\theta_o,i}\big)\Big/(\theta-\theta_o) - \big(P_{\theta_o,1} \times \cdots \times P_{\theta_o,n-1} \times \dot{P}_{\theta_o,n} +$$

$$+ P_{\theta_o,1} \times \cdots \times \dot{P}_{\theta_o,n-1} \times P_{\theta_o,n} +$$

$$+ \cdots + \dot{P}_{\theta_o,1} \times \cdots \times P_{\theta_o,n-1} \times P_{\theta_o,n}\big) \Big\| = 0$$

Proof: This is just the formula for the derivative of a product, and its proof follows from the decomposition :

$$\prod_i P_{\theta,i} - \prod_i P_{\theta_o,i} = \prod_i P_{\theta,i} - \prod_{i<n} P_{\theta,i} \times P_{\theta_o,n}$$

$$+ \prod_{i<n} P_{\theta,i} \times P_{\theta_o,n} - \prod_{i<n-1} P_{\theta,i} \times P_{\theta_o,n-1} \times P_{\theta_o,n}$$

$$- - - - - -$$

$$+ \prod_{i<2} P_{\theta,i} \times P_{\theta_o,2} \times \cdots \times P_{\theta_o,n} - \prod_i P_{\theta_o,i} \,. \qquad \square$$

The next proposition implies that sub experiments of products of the translation experiments in the previous example are differentiable.

Proposition 2.2.

Let $\mathcal{E} = ((\mathcal{X},\mathcal{A}), P_\theta: \theta \in \Theta)$ be differentiable in θ_o and let $\mathcal{B}$ be a sub σ-algebra of $\mathcal{A}$, and let $P_{\theta\mathcal{B}}$ denote the restriction of P_θ to $\mathcal{B}$. Then $((\mathcal{X},\mathcal{B}), P_{\theta\mathcal{B}}: \theta \in \Theta)$ is differentiable in θ_o and

$$\lim_{\theta\to\theta_o} \| (P_{\theta\mathcal{B}} - P_{\theta_o\mathcal{B}}) / (\theta-\theta_o) - \dot{P}_{\theta_o\mathcal{B}} \| = 0$$

where $\dot{P}_{\theta_o\mathcal{B}}$ is the restriction of $\dot{P}_{\theta_o} = \lim_{\theta\to\theta_o} (P_\theta - P_{\theta_o})/(\theta-\theta_o)$ to $\mathcal{B}$.

Proof: The proof follows from the fact that the restriction of a measure μ to a sub σ-algebra has smaller total variation than μ. □

More generally we have:

Proposition 2.3.

If $\mathcal{E} \geq \mathcal{F}$ and $\mathcal{E}$ is differentiable then $\mathcal{F}$ is also differentiable.

Proof. Write $\mathcal{E} = (\mathcal{X},\mathcal{A}); P_\theta: \theta\in\Theta)$ and $\mathcal{F} = ((\mathcal{Y},\mathcal{B}), P_\theta M; \theta \in \Theta)$ where M is a randomization from $(\mathcal{X},\mathcal{A})$ to $(\mathcal{Y},\mathcal{B})$.

Then - by the continuity of M

$$\lim_{\theta\to\theta_o} (P_\theta M - P_{\theta_o} M)/(\theta-\theta_o) = [\lim_{\theta\to\theta_o} (P_\theta - P_{\theta_o})/\theta-\theta_o]M. \quad \square$$

As a corollary of propositions 2.2 and 2.3 we get :

Corollary 2.4.

The product experiment of a finite family of experiments is differentiable in θ_o if and only if all factor experiments are differentiable in θ_o.

3. Basic properties of the derivative.

We define the derivative of a differentiable experiment $\mathcal{E} = ((X,\mathcal{A}); P_\theta: \theta \in \Theta)$ as the pseudo dichotomy $\dot{\mathcal{E}}_{\theta_0} \overset{\text{definition}}{=} ((X,\mathcal{A}) P_{\theta_0}, \dot{P}_{\theta_0})$ where

$$(3.1) \qquad \dot{P}_{\theta_0}(A) = \lim_{\theta \to \theta_0} (P_\theta(A) - P_{\theta_0}(A)) / (\theta - \theta_0)$$

The next proposition tells us that the rule $\mathcal{E} \rightsquigarrow \dot{\mathcal{E}}_{\theta_0}$ is monotonic w.r.t. $\geq$ where $\geq$ is short for "being more informative than" .

Proposition 3.1.

Let $\mathcal{E} = ((X,\mathcal{A}), (P_\theta; \theta \in \Theta))$ and $\mathcal{F} = ((\mathcal{Y},\mathcal{B}), (Q_\theta; \theta \in \Theta))$ be differentiable in θ_0. Then $\dot{\mathcal{E}}_{\theta_0} \geq \dot{\mathcal{F}}_{\theta_0}$ provided $\mathcal{E} \geq \mathcal{F}$. In particular $\dot{\mathcal{E}}_{\theta_0} \sim \dot{\mathcal{F}}_{\theta_0}$ when $\mathcal{E} \sim \mathcal{F}$.

Proof: The proof is an immediate consequence of the randomization criterion. □

Which pseudo dichotomies are of the form $\dot{\mathcal{E}}_{\theta_0}$? It follows from (3.1) that $\dot{P}_{\theta_0}(X) = 0$ and that $P_{\theta_0} \gg \dot{P}_{\theta_0}$. The theorem below asserts that these conditions are - together - characteristic properties.

Theorem 3.2.

A pseudo dichotomy $((\chi,\mathcal{A})\pi,\sigma)$ is the derivative in θ_0 of some experiment $\mathcal{E}$, if and only if $\sigma(\chi) = 0$ and π is a probability measure dominating σ. If so, then $((\chi,\mathcal{A})\pi,\sigma)$ is the derivative in θ_0 of the experiment $((\chi,\mathcal{A}), P_\theta: \theta \in \Theta)$ where *)

$$P_\theta = |\pi+(\theta-\theta_0)\sigma| / \|\pi+(\theta-\theta_0)\sigma\|; \; \theta \in \Theta$$

Furthermore, these conditions imply that

$$(3.2) \qquad \lim_{\theta\to\theta_0} \|(\pi+(\theta-\theta_0)\sigma)^-\| / (\theta-\theta_0) = 0$$

and

$$(3.3) \qquad \lim_{\theta\to\theta_0} (\|\pi + (\theta-\theta_0)\sigma\| - 1) / (\theta-\theta_0) = 0 \, .$$

Remark. P_θ is well defined since $\|\pi+(\theta-\theta_0)\sigma\| \geq \pi(\chi)+(\theta-\theta_0)\sigma(\chi) = 1$.

Proof: It remains to show

$$P_{\theta_0} = \pi, \quad \lim_{\theta\to\theta_0} \|(P_\theta-P_{\theta_0}) / (\theta-\theta_0) - \sigma\| = 0$$

and that (3.2) and (3.3) hold when π is a probability measure dominating σ and $\sigma(\chi) = 0$. By substitution, $P_{\theta_0} = \pi$ and we may without loss of generality, assume that $\theta_0 = 0$.

Let s be a version of $d\sigma / d\pi$. We get - when $\theta \neq 0$ - successively :

*) If μ is a finite measure then $|\mu| = \mu \vee (-\mu)$.

$$\|(\pi+\theta\sigma)^-\| = \int_{\theta s\leq -1} |1+\theta s|\,d\pi \leq \int_{|\theta s|\geq 1} 1+|\theta s|\,d\pi \leq 2|\theta| \int_{|s|\geq\frac{1}{\theta}} |s|\,d\pi$$

and

$$\|(P_\theta-P_o)/\theta-\sigma\| = \|(|\pi+\theta\sigma|/\|\pi+\theta\sigma\|-\pi)/\theta-\sigma\|$$

$$\leq \|(|\pi+\theta\sigma|-\pi)/\theta-\sigma\| + \|(|\pi+\theta\sigma|/\|\pi+\theta\sigma\| -\pi)/\theta-(|\pi+\theta\sigma|-\pi)/\theta\|$$

$$\leq \left[\|\,|\pi+\theta\sigma|-(\pi+\theta\sigma)\| + (\|\pi+\theta\sigma\|-1)\right]|\theta|$$

$$= \left[2\|(\pi+\theta\sigma)^-\| + 2\|(\pi+\theta\sigma)^-\|\right]/|\theta|$$

Hence

(3.4) $\|(\pi+\theta\sigma)^-\| \leq 2|\theta|\ \ |\sigma|(|s| \geq 1/|\theta|)$

and

(3.5) $\|(P_\theta-P_o)/\theta-\sigma\| \leq 8|\sigma|(|s| \geq 1/|\theta|)$

In proving (3.5) we used the first of the identities :

(3.6) $\|\pi+\theta\sigma\| = 2\|(\pi+\theta\sigma)^-\| + 1 = 2\|(\pi+\theta\sigma)^+\| - 1$

(3.6) follows from the equations :

$$\|\pi+\theta\sigma\| = \|(\pi+\theta\sigma)^+\| + \|(\pi+\theta\sigma)^-\|$$

and

$$1 = (\pi+\theta\sigma)(X) = \|(\pi+\theta\sigma)^+\| - \|(\pi+\theta\sigma)^-\|$$

The proof may be completed by noting that (3.4) and (3.5) imply - since σ is finite - (3.2) and $\lim_{\theta\to\theta_0} \| (P_\theta - P_{\theta_0})/(\theta-\theta_0) - \sigma \| = 0$, while (3.3) follows from (3.2) and (3.6). □

The pseudo dichotomy $((\chi,\mathcal{A})\pi,\sigma)$ where π is a probability measure dominating σ and $\sigma(\chi) = 0$, will be denoted by $\mathcal{S}_{\pi,\sigma}$.

The standard representation of $\mathcal{S}_{\pi,\sigma}$ is of the form $\mathcal{S}_{S_1,S_2}$, where $S_1 = \pi(1/(1+|s|),\ s/(1+|s|))^{-1}$ and $S_2 = \sigma(1/(1+|s|),\ s/(1+|s|))^{-1}$. Here s is a version of $d\sigma/d\pi$. A closely associated characteristic is the standard measure $S = S_1 + |S_2|$.

Alternatively we may - since S and πs^{-1} determines each other - use πs^{-1} as a characteristic. The measure πs^{-1} will occasionally be denoted by $F_{\pi,\sigma}$.

Let $G_{\pi,\sigma}$ be the measure whose Radon Nikodym derivative w.r.t. $F_{\pi,\sigma}$ is the identity function $x \rightsquigarrow x$. It will follow from proposition 3.4 that

$$((]-\infty,+\infty[,\ \text{Borel class}),\ F_{\pi,\sigma},\ G_{\pi,\sigma})$$

is a derivative.

Furthermore - since $x \rightsquigarrow x$ is a version of $dG_{\pi,\sigma}/dF_{\pi,\sigma}$ - $\mathcal{S}_{F_{\pi,\sigma},G_{\pi,\sigma}}$ and $\mathcal{S}_{\pi,\sigma}$ are equivalent. It may be checked that $F_{\pi,\sigma}$ is, and may be any probability distribution on $]-\infty,+\infty[$ having expectation 0. We will, occasionally, write

F_{θ_o}, instead of $F_{\pi,\sigma}$ when $\mathcal{S}_{\pi,\sigma} = \mathcal{E}_{\theta_o}$. One pleasant property of this characteristic is :

Proposition 3.3.

Let $\mathcal{E}_1, \mathcal{E}_2, \ldots, \mathcal{E}_n$ be differentiable in θ_o. Then :

$$F_{\theta_o, \Pi_i \mathcal{E}_i} = F_{\theta_o, \mathcal{E}_1} * F_{\theta_o, \mathcal{E}_2} * \cdots * F_{\theta_o, \mathcal{E}_n}$$

where $*$ means convolution.

Proof: It suffices to consider the case of two experiments $\mathcal{E} = ((\chi, \mathcal{A}), P_\theta : \theta \in \Theta)$ and $\mathcal{F} = ((\mathcal{Y}, \mathcal{B}), Q_\theta : \theta \in \Theta)$. Suppose $\mathcal{E}$ and $\mathcal{F}$ are differentiable in θ_o. Using proposition 2.1 we get :

$$\begin{aligned} F_{\theta_o, \mathcal{E} \times \mathcal{F}} &= \mathcal{L}_{P_{\theta_o} \times Q_{\theta_o}}\left(d\left[P_{\theta_o} \times \dot{Q}_{\theta_o} + \dot{P}_{\theta_o} \times Q_{\theta_o}\right] \Big/ d\left[P_{\theta_o} \times Q_{\theta_o}\right]\right) \\ &= \mathcal{L}_{P_{\theta_o} \times Q_{\theta_o}}\left(1 \otimes d\dot{Q}_{\theta_o} \big/ dQ_{\theta_o} + d\dot{P}_{\theta_o} \big/ dP_{\theta_o} \otimes 1\right) \\ &= \mathcal{L}_{P_{\theta_o}}\left(d\dot{P}_{\theta_o} \big/ dP_{\theta_o}\right) * \mathcal{L}_{Q_{\theta_o}}\left(d\dot{Q}_{\theta_o} \big/ dQ_{\theta_o}\right) \\ &= F_{\theta_o, \mathcal{E}} * F_{\theta_o, \mathcal{F}}. \end{aligned}$$

□

The fact that

$((\{(x_1,x_2): x_1 > 0, x_1 + |x_2| = 1\}$, Borel class) $S_1, S_2)$ and

$((]-\infty,+\infty[$, Borel class), $F_{\pi,\sigma}, G_{\pi,\sigma})$ both are

derivatives, is a consequence of :

Proposition 3.4.

If $\mathcal{E} = ((\mathcal{Y},\mathcal{B}),\mu,\nu) \leqq \mathcal{E}_{\pi,\sigma} = ((X,\mathcal{A})\pi,\sigma)$ then $\mathcal{E}$ is also a derivative.

Proof: Let M be a randomization such that $\mu = \pi M$ and $\nu = \sigma M$. Then μ is a probability measure, $\nu(\mathcal{Y}) = \int \sigma(dx) M(\mathcal{Y}|x) = \sigma(X) = 0$ and $\nu(B) = \int \sigma(dx) M(B|x) = 0$ when $\mu(B) = \int \pi(dx) M(B|x) = 0$. □

4. Comparison of derivatives.

In this - and the next section - derivatives will be written $\mathcal{D}_{\pi,\sigma} = ((X,\mathcal{A}),\pi,\sigma)$ with or without affixes. The following notations relative to the derivative $\mathcal{D}_{\pi,\sigma} = ((X,\mathcal{A}),\pi,\sigma)$ will be used:

$$s \overset{\text{definition}}{=\!=} d\sigma/d\pi$$

$$F \overset{\text{definition}}{=\!=} \pi s^{-1}$$

$$U(\xi) \overset{\text{definition}}{=\!=} \|\xi\pi-\sigma\|;\ \xi \in \left]-\infty,+\infty\right[$$

$$V \overset{\text{definition}}{=\!=} \{(\int\delta d\pi, \int\delta d\sigma):\ 0 \leqq \delta \leqq 1\}$$

$$\beta(\alpha) \overset{\text{definition}}{=\!=} \sup\{y:\ (\alpha,y)\in V\};\ \alpha \in [0,1]$$

Affixes on $\mathcal{D},\pi,\sigma,X,\mathcal{A},s,F,U,V$ and β; when these are referring to the same derivative will be of the same type.

For two derivatives $\mathcal{D}$ and $\tilde{\mathcal{D}}$ we will write:

$\dot{\delta}(\mathcal{D},\tilde{\mathcal{D}}) \overset{\text{definition}}{=\!=}$ the smallest $\varepsilon/2$ such that $\mathcal{D}$ is $(0,\varepsilon)$ deficient w.r.t. $\tilde{\mathcal{D}}$.

$\dot{\Delta}(\mathcal{D},\tilde{\mathcal{D}}) \overset{\text{definition}}{=\!=} \max(\dot{\delta}(\mathcal{D},\tilde{\mathcal{D}}),\ \dot{\delta}(\tilde{\mathcal{D}},\mathcal{D}))$

It follows directly from the definitions that

$$0 \leqq \dot{\delta}(\mathcal{D},\tilde{\mathcal{D}}) < \infty,$$

$$\dot{\delta}(\mathcal{D},\mathcal{D}) = 0,$$

$$\dot{\delta}(\mathcal{D},\hat{\mathcal{D}}) \leqq \dot{\delta}(\mathcal{D},\tilde{\mathcal{D}}) + \dot{\delta}(\tilde{\mathcal{D}},\hat{\mathcal{D}}),$$

$\dot{\Delta}$ is a pseudo metric,

$$\delta \leqq 2\mathring{\delta}$$

and $$\Delta \leqq 2\mathring{\Delta}.$$

Let $\mathscr{E} = ((\chi,\mathcal{A}),\pi,\sigma)$ and $\tilde{\mathscr{E}} = ((\tilde{\chi},\tilde{\mathscr{B}}),\tilde{\pi},\tilde{\sigma})$ be two derivatives. Then - since π and $\tilde{\pi}$ both are probability measures and $\sigma(\chi) = \tilde{\sigma}(\tilde{\chi}) - \Delta_1(\mathscr{E},\tilde{\mathscr{E}}) = 0$, and general comparison is equivalent with comparison for testingproblems. It follows that $\mathscr{E}$ is $(\varepsilon_1,\varepsilon_2)$ deficient w.r.t. $\tilde{\mathscr{E}}$ if and only if

$$\|a_1\pi + a_2\sigma\| \geqq \|a_1\tilde{\pi} + a_2\tilde{\sigma}\| - \varepsilon_1|a_1| - \varepsilon_2|a_2|; \quad a_1,a_2 \in]-\infty,+\infty[$$

or equivalently that:

(4.1) $$U(\xi) \geqq \tilde{U}(\xi) - \varepsilon_1|\xi| - \varepsilon_2; \qquad \xi \in]-\infty,+\infty[$$

In particular

(4.2) $$\delta(\mathscr{E},\tilde{\mathscr{E}}) = \sup_\xi (\tilde{U}(\xi) - U(\xi))^+/(1+|\xi|)$$

so that

(4.3) $$\Delta(\mathscr{E},\tilde{\mathscr{E}}) = \sup_\xi \left[|\tilde{U}(\xi) - U(\xi)|/|1+|\xi|\right]$$

Similarily:

(4.4) $$\dot{\delta}(\mathscr{E},\tilde{\mathscr{E}}) = \sup_\xi (\tilde{U}(\xi) - U(\xi))^+/2$$

so that

(4.5) $$\dot{\Delta}(\mathscr{E},\tilde{\mathscr{E}}) = \sup_\xi |\tilde{U}(\xi) - U(\xi)|/2$$

It follows directly from (4.3) that U determines $\mathscr{E}$ up to equivalence. We shall later describe the class of possible U's.

Two simple lower bound for δ and Δ (and therefore for $2\dot{\delta}$ and $2\dot{\Delta}$) follows by inserting $\xi = 0$ in (4.2) and (4.3). We get:

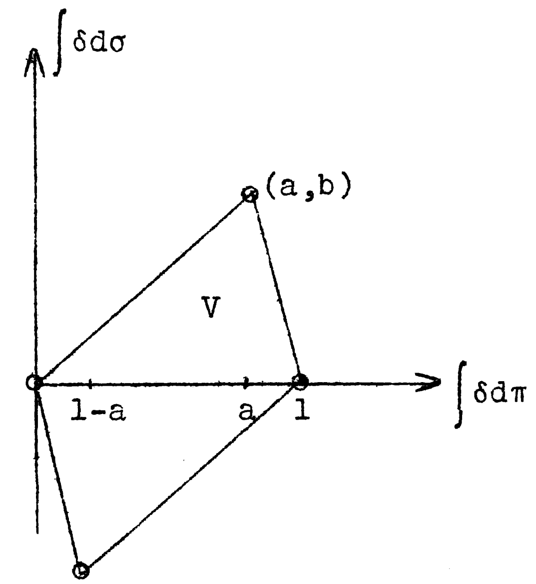
$$(4.6) \qquad \delta(\mathcal{E}, \tilde{\mathcal{E}}) \geqq \|\tilde{\sigma}\| - \|\sigma\|$$

and

$$(4.7) \qquad \Delta(\mathcal{E}, \tilde{\mathcal{E}}) \geqq \big|\|\tilde{\sigma}\| - \|\sigma\|\big|$$

The weak compactness theorem implies that V is closed - and it is easily seen that V is a compact and convex sub set of $[0,1] \times]-\infty,+\infty[$. Moreover $(0,0) \in V$ and - since $(1-\delta)$ is a test function when δ is - it is symmetric about $(\frac{1}{2},0)$.

As an example consider the case where $a \in]0,1[$ and $b > 0$ are given numbers and F assigns mass $1-a$ in $(-b)/(1-a)$ and mass a in b/a. Then V is the region bounded by the parallelogram with corners $(0,0)$, (a,b), $(1,0)$ and $(1-a,-b)$.



V is - by symmetry - determined by

$V^+ \underset{=}{\text{definition}} \{(x,y) : (x,y) \in V \,\&\, y \geqq 0\}$. The negative part $\{(x,y) : (x,y) \in V \,\&\, y \leqq 0\}$ is the reflection of V^+ w.r.t. the point $(\frac{1}{2},0)$.

If $\int \delta d\pi = 0$ then $\delta = 0$ a.e. π and - since $\pi >> \sigma$ - $\int \delta d\sigma = 0$. It follows that $(0,0)$ is the only point in V with first coordinate = 0 and that $(1,0)$ is the only point in V with first coordinate = 1. The second coordinate y of a point $(x,y) \in V$ is bounded by $\|\sigma\|/2$ in numerical value and $\delta = I_{s \geqq 0}$ give the point $(\pi(s \geqq 0), \|\sigma\|/2)$. If $\sigma = 0$, then V is the line segment $\{(\alpha,0): 0 \leqq \alpha \leqq 1\}$. V determines $\mathcal{E}$ up to equivalence since U does and

$$U(\xi) = 2H(\xi,-1)-\xi \ ; \qquad \xi \in]-\infty,+\infty[$$

where H is the support function of V.

It follows - since β obviously determines V - that $\mathcal{E}$ is, up to equivalence, determined by β. Furthermore β is concave and $\beta(0+) = \beta(0) = \beta(1-) = \beta(1) = 0$.

Conversely, let β be any concave function on $[0,1]$ such that $\beta(0) = \beta(0+) = \beta(1-) = \beta(1) = 0$. Then β is absolutely continuous with a Hahn set of the form $[0,\alpha_0]$ where $\alpha_0 \in]0,1[$ is a point where β obtains its maximum. The measure whose distribution function is β will - by abuse of notations - also be denoted by β. Let λ denote Lebesgue measure restricted to the Borel class on $[0,1]$. Then

$$(([0,1], \text{ Borel class}), \ \lambda,\beta)$$

is a derivative, and we will now show that the same procedure applied to this derivative will give us β back again, i.e.:

$$\sup(\int \delta d\beta : \quad 0 \leqq \delta \leqq 1; \ \int \delta d\lambda = \alpha) = \beta(\alpha); \quad \alpha \in [0,1].$$

We may - since this is trivial when $\alpha = 0$ or $\alpha = 1$ - assume $\alpha \in]0,1[$. Let δ be an arbitrary test function such that $\int \delta d\lambda = \alpha$. Then we have:

$$\int I_{[0,\alpha]} d\beta - \int \delta d\beta = \int_0^1 (I_{[0,\alpha]} - \delta)(\beta' - \beta'(\alpha)) d\lambda .$$

The integrand on the right hand side is - by the concavity of β - non negative whenever it is defined. It follows that $\int I_{[0,\alpha]} d\beta \geq \int \delta d\beta$ and - since $\int I_{[0,\alpha]} d\lambda = \alpha$ -, $\sup(\int \delta d\beta : 0 \leq \delta \leq 1; \int \delta d\lambda = \alpha) = \int I_{[0,\alpha]} d\beta = \beta(\alpha)$.

We have proved

Theorem 4.1.

β characterizes the derivative up to equivalence and β is and may be any concave function on $[0,1]$ such that $\beta(0+) = \beta(0) = 0 = \beta(1) = \beta(1-)$.

If β has these properties; then any derivative corresponding to β is equivalent with

$$(([0,1], \text{ Borel class}), \lambda, \beta)$$

The correspondence between V and β yields :

Corollary 4.2.

V characterizes the derivative up to equivalence and V is and may be any compact convex set contained in the strip $[0,1] \times]-\infty,+\infty[$, containing $(0,0)$ but no other point $(0,y)$, and which is symmetric w.r.t. $(\frac{1}{2},0)$.

Corollary 4.3.

A set of derivatives having the property that any derivative has a version in the set, is a lattice for the ordering $\geq$ If $\mathcal{B}$ and $\tilde{\mathcal{B}}$ are in the set, then $\mathcal{B} \wedge \tilde{\mathcal{B}}$ is represented by $\min(\beta, \tilde{\beta})$.

Let us now see how comparison of the derivatives may be expressed in terms of the β - s. Let H and $\tilde{H}$ be the support functions of V and $\tilde{V}$ respectively. The criterion for $(\varepsilon_1, \varepsilon_2)$ deficiency may now be written

$$H + \hat{H} \geq \tilde{H}$$

where $\hat{H}$ is the support function: $(a_1, a_2) \leadsto (|a_1|\varepsilon_1 + |a_2|\varepsilon_2)/2$ of $[-\varepsilon_1/2, \varepsilon_1/2] \times [-\varepsilon_2/2, \varepsilon_2/2]$. Hence:

Proposition 4.4.

$\mathcal{B}$ is $(\varepsilon_1, \varepsilon_2)$ deficient w.r.t. $\tilde{\mathcal{B}}$ if and only if

$$V + [-\varepsilon_1/2, \varepsilon_1/2] \times [-\varepsilon_2/2, \varepsilon_2/2] \supseteq \tilde{V}$$

In terms of the β - s, this may be formulated as:

Proposition 4.5.

$\mathcal{B}$ is $(\varepsilon_1, \varepsilon_2)$ deficient w.r.t. $\tilde{\mathcal{B}}$ if and only if

$$\sup\{\beta(x) : x \in [\alpha - \varepsilon_1/2, \alpha + \varepsilon_1/2]\} \geq \tilde{\beta}(\alpha) - \varepsilon_2/2; \qquad \alpha \in [0,1].$$

Proof: 1° Suppose $\mathcal{B}$ is $(\varepsilon_1, \varepsilon_2)$ deficient w.r.t. $\tilde{\mathcal{B}}$, and let $\alpha \in [0,1]$. By proposition 4.4 - since $(\alpha, \tilde{\beta}(\alpha)) \in V$ - there is a point $(x_1, x_2) \in V$ such that $|\alpha - x_1| \leq \varepsilon_1/2$ and $|\tilde{\beta}(\alpha) - x_2| \leq \varepsilon_2/2$. Hence

$$\sup\{\beta(x) : x \in [\alpha - \varepsilon_1/2, \alpha + \varepsilon_1/2]\} \geq \beta(x_1) \geq x_2 \geq \tilde{\beta}(\alpha) - \varepsilon_2/2.$$

2° Suppose $\sup\{\beta(x) : x \in [\alpha - \varepsilon_1/2,\ \alpha + \varepsilon_1/2]\} \geqq \tilde{\beta}(\alpha) - \varepsilon_2/2;\ \alpha \in [0,1]$, and consider a point $(z_1, z_2) \in \tilde{V}$ where $z_2 \geqq 0$. There is, by assumption, a x_1 in $[z_1 - \varepsilon_1/2,\ z_1 + \varepsilon_1/2]$ so that $\beta(x_1) \geqq \tilde{\beta}(z_1) - \varepsilon_2/2 \geqq z_2 - \varepsilon_2/2$. Put $x_2 = \min(\beta(x_1),\ z_2 + \varepsilon_2/2)$. Then - since $0 \leqq x_2 \leqq \beta(x_1)$ - , $(x_1, x_2) \in V$ and clearly $x_2 \in [z_2 - \varepsilon_2/2,\ z_2 + \varepsilon_2/2]$. Hence $(z_1, z_2) \in V + [-\varepsilon_1/2,\ \varepsilon_1/2] \times [-\varepsilon_2/2,\ \varepsilon_2/2]$. By symmetry this extends to any $(z_1, z_2) \in \tilde{V}$ and $(\varepsilon_1, \varepsilon_2)$ deficiency follows from proposition 4.4.

□

<u>Corollary 4.6.</u>

$\delta(\mathcal{E}, \tilde{\mathcal{E}})$ is the smallest $\varepsilon \geqq 0$ such that

$$\sup\{\beta(x) : |x - \alpha| \leqq \varepsilon/2\} \geqq \tilde{\beta}(\alpha) - \varepsilon/2; \qquad \alpha \in [0,1].$$

<u>Corollary 4.7.</u>

$$\dot{\delta}(\mathcal{E}, \tilde{\mathcal{E}}) = \sup_{\alpha}(\tilde{\beta}(\alpha) - \beta(\alpha))^{+}$$

and $$\dot{\Delta}(\mathcal{E}, \tilde{\mathcal{E}}) = \sup_{\alpha}|\tilde{\beta}(\alpha) - \beta(\alpha)| ..$$

By corollary (4.7), $\dot{\Delta}(\mathcal{E}, \tilde{\mathcal{E}})$ is simply the sup norm distance between β and $\tilde{\beta}$.

The next proposition tells us how to get U and β from F.

<u>Proposition 4.8.</u>

F determines β and U through the formulas:

$$\beta(\alpha) = \int_0^{\alpha} F^{-1}(1-p)\,dp; \quad \alpha \in [0,1]$$

and $$U(\xi) = 2\int_{-\infty}^{\xi} F(x)\,dx - \xi = 2\int_{\xi}^{\infty}(1 - F(x))\,dx + \xi; \quad \xi \in\,]-\infty, +\infty[.$$

<u>Proof:</u> 1^{o} Proof of the formula for β:

For each Borel sub set B of $[0,1]$ write $\tau(B) = \int_B F^{-1}(1-p)dp$. Then (λ,τ) defines a derivative with U function given by:

$$\xi \rightsquigarrow \|\xi\lambda-\tau\| = \int_0^1 |\xi - F^{-1}(1-p)|dp = \int_0^1 |\xi - F^{-1}(p)|dp$$

$$= \int |\xi - x| F(dx) = U(\xi).$$

It follows that (λ,τ) has the same β function as $\mathcal{E}$. Keep $\alpha \in [0,1]$ fixed and write $\delta_\alpha(p) = 0$ or 1, as $p \leqq \alpha$ or $p > \alpha$. Then $\int\delta_\alpha d\lambda = \alpha$. Hence $\beta(\alpha) \geqq \int\delta_\alpha d\tau$. If $\int\delta d\lambda = \alpha$ and $0 \leqq \delta \leqq 1$, then: $\int\delta_\alpha d\tau - \int\delta d\tau = \int(\delta_\alpha(p) - \delta(p))(F^{-1}(1-p) - F^{-1}(1-\alpha))dp \geqq 0$. It follows that δ_α is optimal; i.e. $\beta(\alpha) = \int\delta_\alpha d\tau = \int_0^\alpha F^{-1}(1-p)dp$.

2^{o} Proof of the formulas for U: In the same way as we got (3.6) we get:

$$U(\xi) = 2\|(\xi\pi - \sigma)^+\| - \xi$$

$\|(\xi\pi-\sigma)^+\|$ may - using the representation

$((]-\infty,+\infty[,\text{Borel class});F,G)$ where $G(B) = \int_B xF(dx)$; $B \in$ Borel class,

- be written :

$$\|(\xi\pi-\sigma)^+\| = \int(\xi-x)^+F(dx) = \int_{-\infty}^{\xi} F(dx).$$

This proves the first "=", and the last "=" follows from the identity:

$$\xi + \int_{\xi}^{\infty}(1-F(x)dx = \int_{-\infty}^{\xi} F(x)dx; \ \xi \in [-\infty,+\infty]. \qquad \square$$

Here is the promised description of the set of possible U-s.

Proposition 4.9.

The function U associated with the derivative $\mathcal{S}$ has the following properties :

U_1: U is convex

U_2: $\lim_{\xi\to-\infty} [U(\xi) + \xi] = \lim_{\xi\to\infty}[U(\xi) - \xi] = 0$

Conversely: any function U from $]-\infty,+\infty[$ to $]-\infty,+\infty[$ which satisfies U_1 and U_2 corresponds to a derivative $\mathcal{S}$.

Proof: 1° Suppose U is the U function associated with $\mathcal{S}$. Then U_1 follows directly from the definition, while U_2 is a consequence of proposition 4.8.

2° Let U be a function from $]-\infty,+\infty[$ to $]-\infty,+\infty[$ satisfying U_1 and U_2, and let T denote the function $\xi \leadsto [U(\xi) + \xi]/2$. Then U_1 and U_2 may be rewritten - respectively - as:

T_1: T is convex

T_2: $\lim_{\xi\to-\infty} T(\xi) = \lim_{\xi\to\infty}[T(\xi) - \xi] = 0.$

Consider numbers $\xi_1 < \xi_2$ and $\eta > 0$. By T_1:

$$T(\xi_2-\eta)-T(\xi_1-\eta) \leq T(\xi_2)-T(\xi_1) \leq T(\xi_2+\eta)-T(\xi_1+\eta)=(\xi_2-\xi_1)+[T(\xi_2+\eta)-(\xi_2+\eta)] - [T(\xi_1+\eta)-(\xi_1+\eta)].$$

$\eta \to \infty$ together with T_2 give:

$$(\S) \qquad 0 \leq T(\xi_2) - T(\xi_1) \leq \xi_2-\xi_1.$$

It follows that T is absolutely continuous on finite intervals. By the Radon Nikodym theorem there is a real valued function F so that

$$(\S\S) \qquad T(\xi_2) - T(\xi_1) = \int_{\xi_1}^{\xi_2} F(x)dx; \ \xi_1 \ \xi_2 \in]-\infty,+\infty[.$$

Here we may - and shall - by (§) - assume that $0 \leq F \leq 1$. The complement of the set $\{\xi: T'(\xi) = F(\xi)\}$ has Lebesgue measure zero and F is - by T_1 - monotonically increasing on $\{\xi: T'(\xi) = F(\xi)\}$. It follows that we may choose a Radon Nikodym derivative F which is monotonically increasing on $]-\infty,+\infty[$. Finally F may be modified on a countable set so that the final version is monotonically increasing and left continuous.

$\xi_1 \to -\infty$ in (§§) give (using T_2)

$$(\S\S\S) \qquad T(\xi) = \int_{-\infty}^{\xi} F(x)dx \ ; \qquad \xi \in]-\infty,+\infty[.$$

The convergence of this integral implies $\lim_{x\to-\infty} F(x) = 0$.

Similarily $\xi_2 = \xi$ and $\xi_1 = 0$ in (§§) yield:

$$T(\xi) - T(0) = \int_0^{\xi} F(x)dx$$

or $$T(0) - T(\xi) + \xi = \int_0^\xi (1-F(x))dx$$

$\xi \to \infty$ (using T_2) give:

(§§§§) $$T(0) = \int_0^\infty (1-F(x))dx,$$

and the convergence of this integral implies $\lim_{x\to\infty} F(x) = 1.$

Altogether we have now shown that F is a probability distribution function. (§§§) with $\xi = 0$ and (§§§§) yield

$$\int x^+ F(dx) = T(0) = \int x^- F(dx).$$

It follows that $\int xF(dx) = 0$. For each Borel set B writes $G(B) = \int_B xF(dx)$. Then $\mathcal{D} = ((]-\infty,+\infty[, \text{Borel class}) F,G)$ is a derivative and the corresponding U function is - by proposition 4.8:

$$\xi \rightsquigarrow 2\int_\infty^\xi F(x)dx - \xi = 2T(\xi) - \xi = U(\xi) .$$ □

Corollary 4.10.

Suppose $\mathcal{E}$ and $\tilde{\mathcal{E}}$ belong to a set of derivatives containing at least one version of any derivative. Then - provided $\mathcal{E}$ and $\tilde{\mathcal{E}}$ is in this set - $\mathcal{E} \vee \tilde{\mathcal{E}}$ has $\max(U,\tilde{U})$ as U function.

The ordering "$\mathcal{D} \geqq \tilde{\mathcal{E}}$" for pseudo dichotomies is defined as "$\delta(\mathcal{E},\tilde{\mathcal{E}}) = 0$". By the definition of $\mathring{\delta}$, $\mathcal{E} \geqq \tilde{\mathcal{E}}$ implies $\mathring{\delta}(\mathcal{E},\tilde{\mathcal{E}}) = 0$. Conversely, $\mathring{\delta}(\mathcal{E},\tilde{\mathcal{E}}) = 0$, implies - since $2\mathring{\delta} \geqq \delta$ - $\mathcal{E} \geqq \tilde{\mathcal{E}}$. This and other criterions for "$\geqq$" are listed in

Proposition 4.11.

The following conditions on the pair $(\mathcal{E}, \tilde{\mathcal{E}})$ of derivatives are equivalent:

(i) $\mathcal{E} \geqq \tilde{\mathcal{E}}$

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

(iii) There exists a randomization M from $(X, \mathcal{A})$ to $(\tilde{X}, \tilde{\mathcal{A}})$ so that $\pi M = \tilde{\pi}$ and $\sigma M = \tilde{\sigma}$.

(iv) $U \geqq \tilde{U}$

(v) $V \supseteqq \tilde{V}$

(vi) $\beta \geqq \tilde{\beta}$

(vii) $\int_{-\infty}^{\xi} F(x)dx \geqq \int_{-\infty}^{\xi} \tilde{F}(x)dx \; ; \qquad \xi \in \,]-\infty, +\infty[$

(viii) $\int_{\xi}^{\infty} 1-F(x)dx \geqq \int_{\xi}^{\infty} (1-\tilde{F}(x)dx \; ; \; \xi \in \,]-\infty, +\infty[$

(ix) $\int \phi dF \geqq \int \phi d\tilde{F}$ for any convex ϕ.

(x) There exists a dilatation D (i.e. D is a randomization such that $\int y\, D(dy|x) \overset{\equiv}{x} x$) so that $F = \tilde{F} D$.

Proof: We have already shown (i) <=> (ii). (i) <=> (iii) follows from the randomization criterion. (i) <=> (iv) follows from (4.2). (i) <=> (v) follows from proposition 4.4. (i) <=> (vi) follows from corollary 4.7 and (iv) <=> (vii) <=> (viii) is a consequence of proposition 4.8. Altogether we have now shown :

(i) <=> (ii) <=> .... <=> (viii).

Suppose (i). Then - by the sub linear function criterion - $\int\psi(1,x)F(dx) \geq \int\psi(1,x)\tilde{F}(dx)$ for any sub linear function ψ on $]-\infty,+\infty[^2$. This implies - since any convex function ϕ is of the form $\lim_n\uparrow\psi_n(1,x)$ where ψ_n; $n = 1,2,\cdots$ are sub linear - (ix). Conversely (ix), with ϕ's of the special type $x \mapsto \psi(1,x)$ where ψ is sub linear, implies (i). Finally (ix) <=> (x) is a consequence of theorem 2 in Strassen's paper [12]. □

The equivalence " $\mathcal{E} \sim \tilde{\mathcal{E}}$ " for pseudo dichotomies is defined as " $\Delta(\mathcal{E},\tilde{\mathcal{E}}) = 0$ ". By proposition 4.11, $\mathcal{E} \sim \tilde{\mathcal{E}}$ if and only if $\mathring{\Delta}(\mathcal{E},\tilde{\mathcal{E}}) = 0$. The particular case of sufficiency is treated in the next proposition. It will be shown that the factorization criterion is valid for **derivatives.** The argumentation is essentially that of example 9 in [15].

<u>Proposition 4.12.</u>

Let $\mathcal{E} = ((X,\mathcal{A}),\pi,\sigma)$ be a derivative and let $\tilde{\mathcal{E}}$ be the sub derivative $((X,\mathcal{B}),\pi_{\mathcal{B}},\sigma_{\mathcal{B}})$ where $\mathcal{B}$ is a sub σ algebra of $\mathcal{A}$ and the subscript $\mathcal{B}$ indicates restriction to $\mathcal{B}$.

Then $\mathcal{E} \sim \tilde{\mathcal{E}}$ if and only if $d\sigma/d\pi$ has a $\mathcal{B}$ measurable version.

<u>Proof:</u> On the probability space $(X,\mathcal{A},\pi)$ consider the variables s $(= d\sigma/d\pi)$ and $E^{\mathcal{B}}s$. Let $B \in \mathcal{B}$. Then $\int_B E^{\mathcal{B}}s\,d\pi = \int_B s\,d\pi = \sigma(B) = \sigma_{\mathcal{B}}(B)$. It follows that $E^{\mathcal{B}}s$ is a version of $d\sigma_{\mathcal{B}}/d\pi_{\mathcal{B}}$. Hence-by the discussion in section 3 - $\mathcal{E} \sim \tilde{\mathcal{E}}$ if and only if $\mathcal{L}(s|\pi) = \mathcal{L}(E^{\mathcal{B}}s|\pi)$, and this - by the argumentation in example 9 in [15]- is the case if and only if $s = E^{\mathcal{B}}s$ a.s π. □

If $\mathcal{D} = ((X,\mathcal{A}),\pi,\sigma)$ is a derivative, then $\overline{\mathcal{D}} \underset{=}{\overset{\text{definition}}{=}} ((X,\mathcal{A}),\pi,-\sigma)$ is also a derivative. A few simple properties of the correspondence $\mathcal{D} \mapsto \overline{\mathcal{D}}$ are listed in:

<u>Proposition 4.13.</u>

(i) $\overline{\overline{\mathcal{D}}} = \mathcal{D}$

(ii) $\overline{U}(\xi) = U(-\xi)$; $\xi \in]-\infty,\infty[$

(iii) $\overline{\beta}(\alpha) = \beta(1-\alpha)$; $\alpha \in [0,1]$

(iv) $\overline{V} = \{(x,y) : (1-x,y) \in V\}$

(v) $\overline{s} = - s$

(vi) $\overline{F} = \mathcal{L}(-X \mid \mathcal{L}(X) = F)$

(vii) $\dot{\delta}(\mathcal{D}_1, \mathcal{D}_2) = \dot{\delta}(\overline{\mathcal{D}_1}, \overline{\mathcal{D}_2})$

and $\dot{\Delta}(\mathcal{D}_1, \mathcal{D}_2) = \dot{\Delta}(\overline{\mathcal{D}_1}, \overline{\mathcal{D}_2})$

(viii) $\delta(\mathcal{D}_1, \mathcal{D}_2) = (\overline{\mathcal{D}_1}, \overline{\mathcal{D}_2})$

and $\Delta(\mathcal{D}_1, \mathcal{D}_2) = \Delta(\overline{\mathcal{D}_1}, \overline{\mathcal{D}_2})$

<u>Proof:</u> Follows directly from the definition. □

A derivative $\mathcal{D}$ will be called symmetric if $\dot{\Delta}(\overline{\mathcal{D}}, \mathcal{D}) = 0$.

Corollary 4.14. The following conditions are equivalent:

(i) $\mathcal{D}$ is symmetric

(ii) $\mathcal{D} \geqq \overline{\mathcal{D}}$

(iii) U is an even function

(iv) β is symmetric about $\frac{1}{2}$

(v) V is symmetric about the line "$x = \frac{1}{2}$"

(vi) F is symmetric about 0.

Proof: Straight forward. □

5 Convergence of derivatives.

The notational system in this section will be the same as in section 4. A few convergence criterions are listed in:

Theorem 5.1.

The following conditions on the derivatives $\mathcal{D}_n$, $n = 1,2,\cdots$ and $\mathcal{D}$ are equivalent:

(i) $\lim_{n\to\infty} \mathring{\Delta}(\mathcal{D}_n, \mathcal{D}) = 0$

(ii) $\lim_{n\to\infty} \Delta(\mathcal{D}_n, \mathcal{D}) = 0$

(iii) $\lim_{n\to\infty} \beta_n(\alpha) = \beta(\alpha)$; uniformly in $\alpha \in [0,1]$

(iv) $\lim_{n\to\infty} \beta_n(\alpha) = \beta(\alpha)$; $\alpha \in [0,1]$

(v) $\lim_{n\to\infty} U_n(\xi) = U(\xi)$; uniformly in $\xi \in]-\infty,+\infty[$

(vi) $\lim_{n\to\infty} U_n(\xi) = U(\xi)$; $\xi \in]-\infty,+\infty[$

(vii)*) $\lim_{n\to\infty} \Lambda(F_n, F) = 0$ and $x \mapsto x$ is uniformly integrable w.r.t. F_n; $n = 1,2,\cdots$.

Remark.

It follows from proposition 4.8 that (v) may be written,

(v') $\lim_{n\to\infty} \int_{-\infty}^{\xi} \left[F_n(x)-F(x)\right]dx = 0$; uniformly in $\xi \in]-\infty,+\infty[$

or

(v'') $\lim_{n\to\infty} \int_{\xi}^{\infty} \left[F_n(x)-F(x)\right]dx = 0$; ---"---

---

*) Λ is the Levy diagonal distance.

while (vi) may be written

(vi') $$\lim_{n\to\infty} \int_{-\infty}^{\xi} \left[F_n(x)-F(x)\right]dx = 0; \quad \xi \in]-\infty,+\infty[$$

or

(vi'') $$\lim_{n\to\infty} \int_{\xi}^{\infty} \left[F_n(x)-F(x)\right]dx = 0; \quad \text{---"---}$$

An alternative way of writing (vii) is:

(vii') $$\lim_{n\to\infty} \left[\Lambda(F_n,F) + \left|\int|x|F_n(dx) - \int|x|F(dx)\right|\right]=0.$$

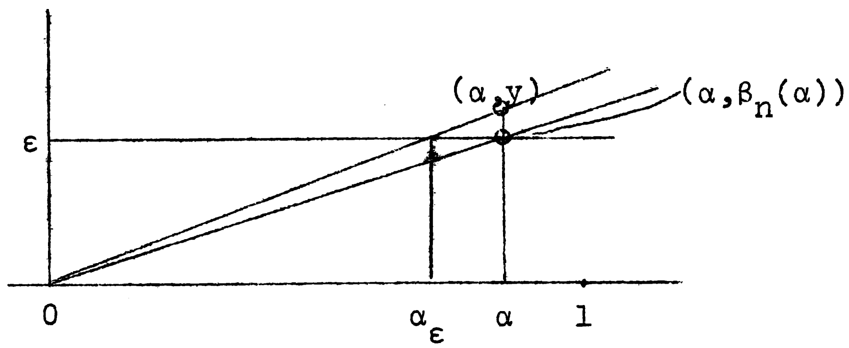
<u>Proof of the theorem:</u>

(i) <=> (iii): Follows from corollary 4.7.

(i) <=> (v) : This is a consequence of (4.5).

(iii) <=> (iv) : => is trivial, so suppose $\lim_{n\to\infty} \beta_n(\alpha) = \beta(\alpha), \alpha \in [0,1]$

Let us show that β_n; $n = 1,2,\cdots$ are equicontinuous in 1. Let $\varepsilon > 0$ be given. By the continuity of β in 1, there is a $\alpha_\varepsilon \in [\frac{1}{2},1]$, so that $\beta(\alpha_\varepsilon) < \varepsilon$. Hence - since $\beta_n(\alpha_\varepsilon) \to \beta(\alpha_\varepsilon)$ - there is a positive integer n_ε so that $\beta_n(\alpha_\varepsilon) < \varepsilon$ when $n \geqq n_\varepsilon$. Let $\alpha \in [\alpha_\varepsilon,1]$ and suppose $n \geqq n_\varepsilon$. Here is a picture of the situation:



The line through $(0,0)$ and (α_o,ε) must - by concavity - intersect the vertical through $(\alpha,0)$ in a point (α,y) where $y \geqq \beta_n(\alpha)$ (If otherwise, then $\beta_n(\alpha_\varepsilon) > \varepsilon$). Hence

$$\beta_n(\alpha) \leqq y = (\varepsilon/\alpha_\varepsilon)\alpha \leqq 2\varepsilon$$

In the same way - or by a symmetry argument - we may show that $\beta_n; n = 1,2,\cdots$ are equicontinuous in 0. It follows - by concavity - that $\beta_n; n = 1,2,\cdots$ are equicontinuous on $[0,1]$. Moreover - since β_n; $n = 1,2,\cdots$ are uniformly bounded on a set $[0,\alpha'] \cup [\alpha'',1]$ where $0 < \alpha' < \alpha'' < 1$ - β_n, $n = 1,2,\cdots$ are (by concavity again) uniformly bounded.

(iii) follows now from Ascoli's theorem.

(i) <=> (ii): => follows from the inequality $\Delta \leqq 2\dot{\Delta}$. Suppose $\lim_{n\to\infty} \Delta(\mathcal{E}_n,\mathcal{E}) = 0$. By corollary 4.6 there is - for each $\alpha \in [0,1]$ - a sequence x_n; $n = 1,2,\cdots$ so that $|x_n-\alpha| \leqq \delta(\mathcal{E},\mathcal{E}_n)/2$; $n = 1,2,\cdots$ and

$$\beta(x_n) \geqq \beta_n(\alpha) - \delta(\mathcal{E},\mathcal{E}_n)/2 \ ; \quad n = 1,2,\cdots$$

Hence

$$\limsup_{n\to\infty} \beta_n(\alpha) \leqq \beta(\alpha); \ \alpha \in [0,1]$$

It follows - as above - that $\beta_n; n = 1,2,\cdots$ are equicontinuous. Using corollary 4.6 - the other way round - we see that there is, for each $\alpha \in [0,1]$, a sequence y_n; $n = 1,2,\cdots$ so that $|y_n-\alpha| \leqq \delta(\mathcal{E}_n,\mathcal{E})/2$, $n = 1,2,\cdots$, and

$$\beta_n(y_n) \geqq \beta(\alpha) - \delta(\mathcal{E}_n,\mathcal{E})/2; \ n = 1,2,\cdots.$$

By equicontinuity

$$\beta_n(y_n) - \beta_n(\alpha) \to 0$$

Hence

$$\liminf_n \beta_n(\alpha) \geqq \beta(\alpha)$$

Altogether we have shown (iv) and we already know that (iv) => (iii) => (i).

(v) <=> (vi): => is trivial so suppose $\lim_{n\to\infty} U_n(\xi) = U(\xi)$; $\xi \in]-\infty,+\infty[$.

By proposition 4.8; $\int_{-\infty}^{\xi} F_n(x)dx \to \int_{-\infty}^{\xi} F(x)dx$; $\xi \in]-\infty,+\infty[$.

Let ξ_o by such that $\int_{-\infty}^{\xi_o} F(x)dx > 0$, and choose a n_o so that $\int_{-\infty}^{\xi_o} F_n(x)dx > 0$ when $n \geqq n_o$. Define the distribution functions M_{n_o}, M_{n_o+1}, and M by

$$M_n(\xi) = \left[\int_{-\infty}^{\xi} F_n(x)dx\right] \Big/ \left[\int_{-\infty}^{\xi_o} F_n(x)dx\right]; \qquad \xi \in]-\infty,\xi_o],$$

$$M_n(\xi) = 1 ; \qquad \xi \in]\xi_o,\infty[,$$

$$M(\xi) = \left[\int_{-\infty}^{\xi} F(x)dx\right] \Big/ \left[\int_{-\infty}^{\xi_o} F(x)dx\right]; \qquad \xi \in]-\infty,\xi_o],$$

and

$$M(\xi) = 1 ; \qquad \xi \in]\xi_o,\infty[.$$

Then $M_n: n \geqq n_o$ and M are continuous probability distribution functions and $\lim_{n\to\infty} M_n(\xi) = M(\xi); \xi \in]-\infty,\infty[$.

It follows that the convergence is uniform in ξ. Hence

$$\int_{-\infty}^{\xi} F_n(x)dx \to \int_{-\infty}^{\xi} F(x)dx \; ; \quad \text{uniformly in } \xi \in]-\infty,\xi_o]$$

i.e. $\lim_{n\to\infty} U_n(\xi) = U(\xi)$; uniformly in $\xi \in]-\infty,\xi_o]$.

Similarily it may be shown that $\lim_{n\to\infty} U_n(\xi) = U(\xi)$; uniformly in $\xi \in [\xi_1,\infty[$ when $\int_{\xi_1}^{\infty} (1-F(x)dx > 0$.

(vi) <=> (vii): <= is clear since $U_n(\xi) = \int|\xi-x|F_n(dx)$ and $U(\xi) = \int|\xi-x|F(dx)$. Suppose (vi). Then $\int|x|F_n(dx) = U_n(0) \to U(0) = \int|x|F(dx)$. It follows that F_n, $n = 1,2,\cdots$ are conditionally compact. Suppose $\Lambda(F_{n_k},F_o) \to 0$ as $k \to \infty$. By (vi) $-\int_{\xi_1}^{\xi_2} F_{n_k}(x)dx \to \int_{\xi_1}^{\xi_2} F(x)dx$ as $k \to \infty$. Hence $\int_{\xi_1}^{\xi_2} F(x)dx = \int_{\xi_1}^{\xi_2} F_o(x)dx$ when $\xi_1,\xi_2 \in]-\infty,\infty[$, so that $F = F_o$. I follows that $\Lambda(F_n,F) \to 0$ and - since $\int|x|F_n(dx) \to \int|x|F(dx)$ - $x \curvearrowright x$ is uniformly integrable w.r.t. F_n; $n = 1,2,\cdots$. □

It follows from theorem 4.1 that any continuous pointwise limit of β - s corresponds to a derivative. The set of possible functions U is, however, not a closed sub set of $C(]-\infty,+\infty[)$ with the topology of pointwise convergence.

<u>Example 5.2.</u>

Define for each $n = 1,2,\cdots$; U_n by

$$U_n(\xi) = \max\{\left[(1-\tfrac{1}{n})|\xi|+1\right],|\xi|\}; \quad \xi \in]-\infty,+\infty[.$$

By proposition 4.9, U_n represents a derivation. The continuous function $\xi \curvearrowright \lim_{n\to\infty} U_n(\xi) = |\xi| + 1$ does not, however, satisfy the criterions in proposition 4.9, - and therefore does not correspond to any derivative.

This difference in behaviour of the β's and the U's is more apparent than real, since we have

Proposition 5.3.

The following conditions on a sequence $\mathcal{S}_n$; n = 1,2, of derivatives are equivalent:

(i) There exists a derivative $\mathcal{S}$ so that $\lim_{n\to\infty} \mathring{\Delta}(\mathcal{S}_n, \mathcal{S}) = 0$.

(ii) $\lim_{n\to\infty} \beta_n(\alpha)$ exists for all $\alpha \in [0,1]$ and the function $\alpha \rightsquigarrow \lim_{n\to\infty} \beta_n(\alpha)$ is continuous in 0 and 1.

(iii) $\lim_{n\to\infty} U_n(\xi)$ exists for all $\xi \in]-\infty,+\infty[$ and the function from $[-\infty,+\infty]$ to $[0,\infty[$ which maps $-\infty$ and ∞ into 0 and any finite ξ into $\lim_{n\to\infty} U_n(\xi) - |\xi|$, is continuous on $[-\infty,+\infty]$.

Proof (i) <= (ii): Suppose $\beta(\alpha) = \lim_{n\to\infty}\beta_n(\alpha)$ exists and that β is continuous in 0 and 1. By theorem 4.1, β corresponds to a derivative $\mathcal{S}$, and by theorem 5.1, $\mathring{\Delta}(\mathcal{S}_n, \mathcal{S}) \to 0$.

(i) => (ii): Suppose $\lim_{n\to\infty} \mathring{\Delta}(\mathcal{S}_n, \mathcal{S}) = 0$. By theorem 5.1 $\lim_{n\to\infty} \beta_n(\alpha) = \beta(\alpha)$; $\alpha \in [0,1]$ and β is - by theorem 4.1 - absolutely continuous on $[0,1]$.

(i) <= (iii): Suppose $U(\xi) = \lim_{n\to\infty} U_n(\xi)$ exists for any ξ in $]-\infty,+\infty[$, and that $\lim_{|\xi|\to\infty} [U(\xi) - |\xi|] = 0$. By proposition 4.9,

U corresponds to a derivative $\mathcal{D}$ and by theorem 5.1 $\lim_{n\to\infty} \mathring{\Delta}(\mathcal{E}_n, \mathcal{E})=0.$

(i) => (iii): Suppose $\lim_{n\to\infty} \dot{\Delta}(\mathcal{E}_n, \mathcal{E}) = 0$. By theorem 5.1

$\lim_{n\to\infty} U_n(\xi) = U(\xi);\ \ \xi \in]-\infty,+\infty[$ and - by proposition 4.9 -

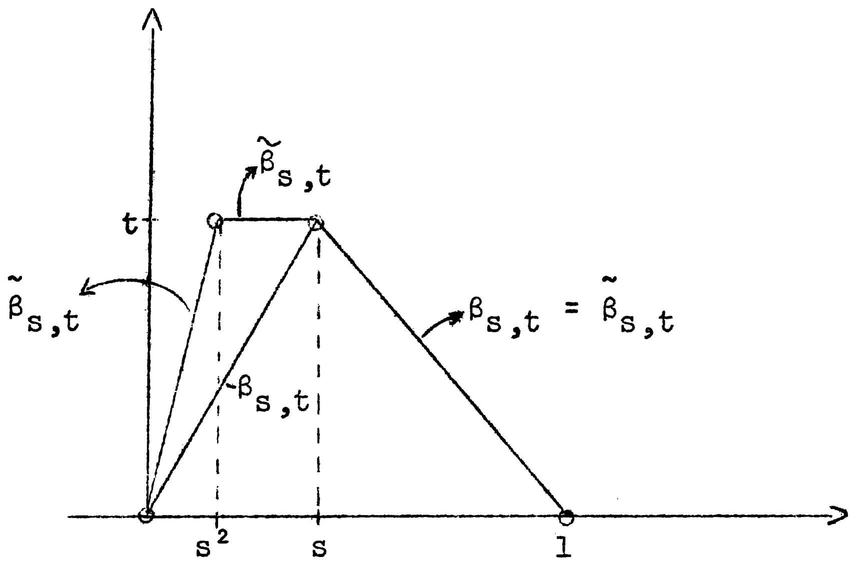
$\lim_{|\xi|\to\infty} \left[U(\xi) - |\xi|\right] = 0.$ □

It will occasionally be convenient to work with sets of experiments and related sets. We will - in such situations - always assume that we are working within a well defined set containing representations of any given experiment.

By theorem 5.1, Δ and $\mathring{\Delta}$ are topologically equivalent. Δ does, however, generate a larger uniformity than $\dot{\Delta}$.

Example 5.4.

Let $s \in]0,1[$ and $t \in [0,\infty[$. Define $\beta_{s,t}$ as the β-function whose graph consists of the line segment from $(0,0)$ to (s,t) and the line segment from (s,t) to $(1,0)$. Define $\tilde{\beta}_{s,t}$ as the β-function whose graph consists of the line segment from $(0,0)$ to (s^2,t), the line segment from (s^2,t) to (s,t) and the line segment from (s,t) to $(1,0)$. Here is a figure of the situation:



Simple calculations - using corollaries 4.6 and 4.7 - yield :

$$\Delta(\mathcal{G}_{s,t}, \tilde{\mathcal{G}}_{s,t}) = 2ts(1-s)/(t+s)$$

and

$$\mathring{\Delta}(\mathcal{G}_{s,t}, \tilde{\mathcal{G}}_{s,t}) = t(1-s)$$

It follows that $\Delta(\mathcal{G}_{s,t}, \tilde{\mathcal{G}}_{s,t}) \to 0$ whenever $s \to 0$. Nothing can, however, be interferred from "$s \to 0$" on the behaviour of $\mathring{\Delta}(\mathcal{G}_{s,t}, \tilde{\mathcal{G}}_{s,t})$.

Corollary 4.7 implies that $\mathring{\Delta}$ is complete. It may, however, easily happen that a sequence of derivatives converges in the Δ distance to a pseudo dichotomy which is not a derivative.

Example 5.5.

Define - for each $s \in]0,1[$ - the derivative $\mathcal{G}_s$ by the matrix:

$$\mathcal{G}_s : \quad \begin{matrix} \chi_s : & 0,1 \\ \pi_s : & s, 1-s \\ \sigma_s : & 1, -1 \end{matrix}$$

and define the pseudo dichotomy $\mathcal{E}$ by the matrix:

$$: \quad \begin{matrix} \chi : & 0,1 \\ \pi : & 0,1 \\ \sigma : & 1, -1 \end{matrix}$$

Then $\lim_{s \to 0} \Delta(\mathcal{G}_s, \mathcal{E}) = 0$. $\mathcal{E}$ is, however, not a derivative.

By example 5.5 the set of derivatives is not Δ-closed as a sub set of the set of all pseudo dichotomies. It follows that Δ restricted to the set of derivatives is not complete.

Δ and $\mathring{\Delta}$ determine - since they are topologically equivalent - the same class of (conditionally) compact sets. Some compactness criterions are listed in:

Theorem 5.6.

The following conditions on the set $\{\mathcal{E}_t : t \in T\}$ of derivatives are equivalent.

(i) $\{\mathcal{E}_t : t \in T\}$ is $\mathring{\Delta}$ conditionally compact

(ii) $\{\mathcal{E}_t : t \in T\}$ is Δ conditionally compact

(iii) $\{\beta_t : t \in T\}$ is equicontinuous in 0 and 1

(iv) $\lim_{|\xi| \to \infty} \left[U_t(\xi) - |\xi|\right] = 0$; uniformly in $t \in T$.

(v) $x \rightsquigarrow x$ is uniformly integrable w.r.t. $F_t : t \in T$.

Remark.

It follows from proposition 4.8 that (iv) may be written

(iv') $\lim_{\xi \to -\infty} \int_{-\infty}^{\xi} F_t(x)dx = 0$; uniformly in $t \in T$,

and

$\lim_{\xi \to \infty} \int_{\xi}^{\infty} (1-F_t(x))dx = 0$; uniformly in $t \in T$.

<u>Proof of the theorem:</u>

(i) <=> (ii) : follows directly from theorem 5.1

(i) => (iii) : follows from theorem 5.1 and Ascoli's theorem

(i) <= (iii) : Equicontinuity in 0 and 1 implies - by concavity - equicontinuity on $[0,1]$. Equicontinuity and concavity imply - since $\beta_t(0) = \beta_t(1) = 0$, $t \in T$ - that $\sup_t \sup_\alpha \beta_t(\alpha) < \infty$. (i) follows now from Ascoli's theorem.

(i) => (iv) : follows from proposition 4.9, theorem 5.1 and Ascoli's theorem.

(iv) => (i) : Suppose $\lim_{|\xi|\to\infty} \left[U_t(\xi) - |\xi|\right] = 0$, uniformly in $t \in T$. Let $\varepsilon > 0$. Then there is a $k > 0$ so that $|\xi| \geqq k \Rightarrow U_t(\xi) \leqq |\xi| + \varepsilon$. Hence :

$$\sup_\xi \left[U_t(\xi) - |\xi|\right] = U_t(0) \leqq \tfrac{1}{2}U_t(-k) + \tfrac{1}{2}U_t(k) \leqq k + \varepsilon$$

so that

$$\sup_t \sup_\xi \left[U_t(\xi) - |\xi|\right] < \infty.$$

Let $h \in [0,\varepsilon]$ and let ξ be any real number. Choose numbers ξ_1 and ξ_2 so that $\xi_1 > k,\xi$ and $\xi_2 < -k,\xi$. By convexity:

$$-3\varepsilon \leqq \left[U_t(\xi_2)+\xi_2\right] - \left[U_t(\xi_2-h)+\xi_2-h\right] - h = U_t(\xi_2)-U_t(\xi_2-h) \leqq$$

$$U_t(\xi+h)-U_t(\xi) \leqq U_t(\xi_1+h)-U_t(\xi_1) = \left[U_t(\xi_1+h)-(\xi_1+h)\right] - \left[U_t(\xi_1)-\xi_1\right] + h \leqq 3\varepsilon.$$

It follows that U_t; $t \in T$ is uniformly equicontinuous from the right on $]-\infty,+\infty[$. Similarily - or by a symmetry argument -

$U_t : t \in T$ is uniformly equicontinuous from the left on $]-\infty,+\infty[$. Define for each t in T W_t as the map from $[-\infty,+\infty]$ to $[0,\infty[$ which maps $-\infty$ and ∞ into 0 and a finite ξ into $U_t(\xi)-|\xi|$. Then $W_t : t \in T$ is uniformly equicontinuous and uniformly bounded on $[-\infty,+\infty]$. (i) now follows from Ascoli's theorem, proposition 5.3 and theorem 5.1.

(i) <=>(v) : Follows directly from theorem 5.1. □

In order to generalize proposition 4.12 to the asymptotic case, we need the following result:

Proposition 5.7.

For each n, $n = 1,2,\cdots$, let X_n be a real random variable on a probability space $(\mathcal{X}_n, \mathcal{A}_n, P_n)$. Denote expectation w.r.t. P_n by E_n. Let - for each n - $\mathcal{B}_n$ be a sub σ algebra of $\mathcal{A}_n$.

Suppose X_n; $n = 1,2,\cdots$, are uniformly integrable (i.e. $\lim_{c\to\infty} \sup_n E_n|X_n| I_{|X_n|\geq c} = 0$). Then

$$\lim_{n\to\infty} \Lambda(\mathcal{L}(X_n), \mathcal{L}(E^{\mathcal{B}_n} X_n)) = 0$$

if and only if

$$\lim_{n\to\infty} E_n|X_n - E_n^{\mathcal{B}_n} X_n| = 0.$$

Proof:

The "if" is trivial, so let us suppose that $\lim_{n\to\infty} \Lambda(\mathcal{L}(X_n), \mathcal{L}(E^{\mathcal{B}_n} X_n)) = 0$. We may - by the relative compactness of (X_n); $n = 1,2,\cdots$, assume that there is a random variable Z on some probability space so that :

$$\lim_{n\to\infty} \Lambda(\mathcal{L}(X_n), \mathcal{L}(Z)) = 0$$

Hence:

$$\lim_{n\to\infty} \Lambda(\mathcal{L}(E^{\mathcal{B}_n}X_n), \mathcal{L}(Z)) = 0.$$

1° Suppose $X_n \geq 0$; $n = 1,2,\cdots$. It suffices - since $(X_n - E^{\mathcal{B}_n}X_n)$; $n = 1,2,\cdots$, are uniformly integrable - to show that $(X_n - E^{\mathcal{B}_n}X_n) \underset{P_n}{\to} 0$. Suppose first that we have shown

$$\sqrt{X_n} - \sqrt{E^{\mathcal{B}_n}X_n} \underset{P_n}{\to} 0$$

Choose numbers $\varepsilon > 0$ and $c > 0$. Then:

$$P_n(|X_n - E_n^{\mathcal{B}_n}X_n| \geq \varepsilon)$$

$$= P_n\Big(\Big[|\sqrt{X_n} - \sqrt{E_n^{\mathcal{B}_n}X_n}|(\sqrt{X_n} + \sqrt{E_n^{\mathcal{B}_n}X_n}) \geq \varepsilon\Big] \cap \Big[(\sqrt{X_n} + \sqrt{E_n^{\mathcal{B}_n}X_n}) \geq c\Big]\Big)$$

$$+ P_n\Big(\Big[|\sqrt{X_n} - \sqrt{E_n^{\mathcal{B}_n}X_n}|(\sqrt{X_n} + \sqrt{E_n^{\mathcal{B}_n}X_n}) \geq \varepsilon\Big] \cap \Big[(\sqrt{X_n} + \sqrt{E_n^{\mathcal{B}_n}X_n}) < c\Big]\Big)$$

$$\leq P_n((\sqrt{X_n} + \sqrt{E_n^{\mathcal{B}_n}X_n}) \geq c) + P_n(|\sqrt{X_n} - \sqrt{E_n^{\mathcal{B}_n}X_n}| \geq \varepsilon/c)$$

$$\leq \sup_n E_n(\sqrt{X_n} + \sqrt{E_n^{\mathcal{B}_n}X_n})/c + P_n(|\sqrt{X_n} - \sqrt{E_n^{\mathcal{B}_n}X_n}| \geq \varepsilon/c)$$

Hence

$$\limsup_{n\to\infty} P_n(|X_n - E_n^{\mathcal{B}_n}X_n| \geq \varepsilon) \leq \sup_n E_n(\sqrt{X_n} + \sqrt{E_n^{\mathcal{B}_n}X_n})/c$$

$c \to \infty$ give - since $\sup_n E_n(\sqrt{X_n} + \sqrt{E^{\mathcal{B}_n}X_n}) < \infty$ -

$$\lim_{n\to\infty} P_n(|X_n - E_n^{\mathcal{B}_n}X_n| \geq \varepsilon) = 0.$$

It follows that we will be through if we can show that $\sqrt{X_n} - \sqrt{E_n^{\mathcal{B}_n} X_n} \underset{P_n}{\to} 0.$

Now:

$$\mathcal{L}(\sqrt{X_n}) \to \mathcal{L}(\sqrt{Z})$$

and

$$\mathcal{L}(\sqrt{E_n^{\mathcal{B}_n} X_n}) \to \mathcal{L}(\sqrt{Z}).$$

By uniform integrability:

$$E_n \sqrt{X_n} \to E\sqrt{Z}$$

and

$$E_n \sqrt{E_n^{\mathcal{B}_n} X_n} \to E\sqrt{Z}$$

Write

$$Y_n = \sqrt{E_n^{\mathcal{B}_n} X_n} - E_n^{\mathcal{B}_n} \sqrt{X_n}$$

By Jensen's inequality $Y_n \geqq 0$ a.e. P_n; $n = 1,2,\cdots,$ and

$$E_n Y_n = E_n \sqrt{E_n^{\mathcal{B}_n} X_n} - E_n \sqrt{X_n} \to 0$$

Hence

$$Y_n \underset{P_n}{\to} 0$$

so that

$$\mathcal{L}(E_n^{\mathcal{B}_n} \sqrt{X_n}) \to \mathcal{L}(\sqrt{Z}).$$

By uniform integrability again

$$E_n X_n \to EZ$$

and

$$E_n(E_n^{\mathcal{B}_n}\sqrt{X_n})^2 \to EZ$$

Hence

$$E_n(\sqrt{X_n} - E_n^{\mathcal{B}_n}\sqrt{X_n})^2 = E_n X_n - E_n(E_n^{\mathcal{B}_n}\sqrt{X_n})^2 \to 0$$

so that

$$\sqrt{X_n} - E_n^{\mathcal{B}_n}\sqrt{X_n} \underset{P_n}{\to} 0$$

It follows - since $Y_n \underset{P_n}{\to} 0$ that

$$\sqrt{X_n} - \sqrt{E_n^{\mathcal{B}_n}X_n} \underset{P_n}{\to} 0 .$$

2^o Let us return to the general case.

Clearly $\mathcal{L}(X_n^+) \to \mathcal{L}(Z^+)$ and $\mathcal{L}((E_n^{\mathcal{B}_n}X_n)^+) \to \mathcal{L}(Z^+)$. By uniform integrability

$$E_n X_n^+ - E_n(E_n^{\mathcal{B}_n}X_n)^+ \to 0$$

or

$$E_n\left[E_n^{\mathcal{B}_n}X_n^+ - (E_n^{\mathcal{B}_n}X_n)^+\right] \to 0$$

By Jensen's inequality: $E_n^{\mathcal{B}_n}X_n^+ \geqq (E_n^{\mathcal{B}_n}X_n)^+$; $n = 1,2,\cdots$.

Hence $E_n^{\mathcal{B}_n}X_n^+ - (E_n^{\mathcal{B}_n}X_n)^+ \underset{P_n}{\to} 0$

so that $\mathcal{L}(E_n^{\mathcal{B}_n}X_n^+) \to \mathcal{L}(Z^+)$

It now follows from 1^o that

$$E_n|X_n^+ - E_n^{\mathcal{B}_n} X_n^+| \to 0$$

Similarily - or by a symmetry argument -

$$E_n|X_n^- - E_n^{\mathcal{B}_n} X_n^-| \to 0$$

Hence

$$E_n|X_n - E_n^{\mathcal{B}_n} X_n| \to 0. \qquad \square$$

Proposition 4.12 may be generalized to the asymptotic case as follows:

<u>Proposition 5.8.</u>

Let $\mathcal{E}_n = ((\chi_n, \mathcal{A}_n), \pi_n, \sigma_n)$; $n = 1,2,\cdots$, be a sequence of derivatives. For each n, let $\mathcal{B}_n$ be a sub σ algebra of $\mathcal{A}_n$ and let $\tilde{\mathcal{E}}_n$ denote the sub derivative $((\chi_n, \mathcal{B}_n), \pi_n \sigma_n)$ where - by abuse of notations - π_n and σ_n are the restrictions of π_n and σ_n to $\mathcal{B}_n$. Finally let, for each n, $\hat{\sigma}_n$ be the measure on $\mathcal{A}_n$ given by

$$d\hat{\sigma}_n|d\pi_n = E_{\pi_n}(d\sigma_n|d\pi_n) .$$

Then $\hat{\mathcal{E}}_n \overset{\text{definition}}{=\!=} ((\chi_n, \mathcal{A}_n), \pi_n, \hat{\sigma}_n)$; $n = 1,2,\cdots$, are derivatives and $\hat{\mathcal{E}}_n \sim \mathcal{E}_n$; $n = 1,2,\cdots$.

Suppose $\mathcal{G}_n$, $n = 1,2,\cdots,$ are relatively compact. Then $\tilde{\mathcal{G}}_n$, $n = 1,2,\cdots,$ are also relatively compact and the following conditions are equivalent:

(i) $\lim_{n\to\infty} \mathring{\Delta}(\mathcal{G}_n, \tilde{\mathcal{G}}_n) = 0$

(ii) $\lim_{n\to\infty} \|\sigma_n - \hat{\sigma}_n\| = 0$

(iii) $\lim_{n\to\infty} \Lambda(\mathcal{L}_{\pi_n}(d\sigma_n|d\pi_n), \mathcal{L}_{\pi_n}(E_{\pi_n}^{\mathcal{S}_n}(d\sigma_n|d\pi_n))) = 0$

Remark.

(ii) may also be written:

(ii') $\lim_{n\to\infty} E_{\pi_n}|s_n - E_{\pi_n}^{\mathcal{S}_n} s_n| = 0$ where - as usual - $s_n = d\sigma_n|d\pi_n$.

$E_{\pi_n}^{\mathcal{S}_n}|s_n$ is the Random Nikodym derivative of the restriction of σ_n to $\mathcal{S}_n$, w.r.t. the restriction of π_n to $\mathcal{S}_n$. It follows that (iii) may be written

(iii') $\lim_{n\to\infty} \Lambda(F_n, \tilde{F}_n) = 0$ where - for each n - F_n and $\tilde{F}_n$ are the "F distribution" corresponding,respectively, to $\mathcal{G}_n$ and $\tilde{\mathcal{G}}_n$.

Proof of the theorem.

Let - for each n - E_n denote expectation w.r.t. π_n.

1^o $\hat{\mathcal{G}}_n$ is a derivative since $\hat{\sigma}_n << \pi_n$ and $\hat{\sigma}_n(\chi_n) = E_n E_n^{\mathcal{S}_n} s_n = E_n s_n = \sigma_n(\chi_n) = 0$. By proposition 4.12; $\hat{\mathcal{G}}_n \sim ((\chi_n, \mathcal{S}_n), \pi_n, \sigma_n^*)$ where

σ_n^* is the restriction of σ_n to $\mathcal{B}_n$. Let $B_n \in \mathcal{B}_n$. Then : $\sigma_n^*(B_n) = \hat{\sigma}_n(B_n) = E_n I_{B_n} E_n^{\mathcal{B}_n} s_n = E_n I_{B_n} s_n = \sigma_n(B_n)$. Hence $((X_n, \mathcal{B}_n), \pi_n, \sigma_n^*) = ((X_n, \mathcal{B}_n), \pi_n, \sigma_n) = \tilde{\mathcal{E}}_n$.

2° Suppose $\mathcal{E}_n$; $n = 1,2,\cdots$, are relatively compact. Then - by theorem 5.6, and since $\tilde{\beta}_n \leqq \beta_n$; $n = 1,2,\cdots$ - $\tilde{\mathcal{E}}_n$; $n = 1,2,\cdots$, are also relatively compact.

3° Suppose $\mathcal{E}_n$; $n = 1,2,\cdots$, are relatively compact.

(i) => (iii) : We may - by relative compactness - assume that $\dot{\Delta}(\mathcal{E}_n, \mathcal{E}) \to 0$, so that $\dot{\Delta}(\tilde{\mathcal{E}}_n, \mathcal{E}) \to 0$. By theorem 5.1, $\Lambda(F_n, F) \to 0$ and $\Lambda(\tilde{F}_n, F) \to 0$. Hence $\Lambda(F_n, \tilde{F}_n) \to 0$.

(iii) => (i) : We may - by relative compactness - assume that $\dot{\Delta}(\mathcal{E}_n, \mathcal{E}) \to 0$ and $\dot{\Delta}(\tilde{\mathcal{E}}_n, \tilde{\mathcal{E}})$ 0. By theorem 5.1, $\Lambda(F_n, F) \to 0$ and $\Lambda(\tilde{F}_n, \tilde{F}) \to 0$. Hence $\Lambda(F, \tilde{F}) \leqq \Lambda(F_n, F) + \Lambda(F_n, \tilde{F}_n) + \Lambda(\tilde{F}_n, \tilde{F}) \to 0$ so that $F = \tilde{F}$. It follows that $\mathcal{E} \sim \tilde{\mathcal{E}}$. Hence $\Delta(\mathcal{E}_n, \tilde{\mathcal{E}}_n) \to 0$.

(ii) <=> (iii) : Follows - since s_n; $n = 1,2,\cdots$, are (by theorem 5.6) uniformly integrable w.r.t. π_n, $n = 1,2,\cdots$ - from proposition 5.8.

□

6 Local comparison of experiments.

We associated in section 3 a derivative $\dot{\mathcal{E}}_{\theta_0}$ with any experiment which was differentiable in θ_0. A mathematical theory for the derivatives was outlined in sections 3-5. The purpose of this section is to connect the theory in sections 3-5 with the statistical theory of information.

Before proceeding, a few notational conventions. Experiments will usually be written $\mathcal{E} = ((\chi,\mathcal{A}), P_\theta:\theta\in\Theta)$ with or without affixes. It shall be assumed - unless otherwise stated - that our experiments are differentiable in θ_0. If $\mathcal{E} = ((\chi,\mathcal{A}),P_\theta:\theta\in\Theta)$, then the derivative in θ_0 will be written $\dot{\mathcal{E}}_{\theta_0} = ((\chi,\mathcal{A})P_{\theta_0},\dot{P}_{\theta_0})$. The restriction, $((\chi,\mathcal{A}), P_\theta:\theta\in\Theta_1)$ of $\mathcal{E}$, will be written $\mathcal{E}_{\Theta_1}$. In agreement with the notations in section 4 and section 5 we define :

$$s_{\theta_0} \underset{=\!=\!=}{\text{definition}} d\dot{P}_{\theta_0}/dP_{\theta_0}$$

$$F_{\theta_0} \underset{=\!=\!=}{\text{definition}} P_{\theta_0} s_{\theta_0}^{-1}$$

$$U_{\theta_0}(\xi) \underset{=\!=\!=}{\text{definition}} \|\xi P_{\theta_0} - \dot{P}_{\theta_0}\|;\quad \xi\in]-\infty,\infty[$$

$$V_{\theta_0} \underset{=\!=\!=}{\text{definition}} \{(\int\delta dP_\theta, \int\delta d\dot{P}_{\theta_0}) : 0 \leqq \delta \leqq 1\}$$

$$\beta_{\theta_0}(\alpha) = \sup\{y: (\alpha,y)\in V_{\theta_0}\};\quad \alpha\in[0,1]\}$$

Affixes on $\mathcal{E}$, χ, $\mathcal{A}$, P_θ, Θ, P_{θ_0}, $\dot{P}_{\theta_0}$, s_{θ_0}, F_{θ_0}, U_{θ_0}, V_{θ_0}, and β_{θ_0}, will - when these are referring to the same experiment - be of the same type.

For two experiments $\mathcal{E}$ and $\tilde{\mathcal{E}}$ we will write :

$$\mathring{\delta}_{\theta_0}(\mathcal{E},\tilde{\mathcal{E}}) \underset{=}{\text{definition}} \mathring{\delta}(\dot{\mathcal{E}}_{\theta_0},\dot{\tilde{\mathcal{E}}}_{\theta_0})$$

$$\mathring{\Delta}_{\theta_0}(\mathcal{E},\tilde{\mathcal{E}}) \underset{=}{\text{definition}} \mathring{\Delta}(\dot{\mathcal{E}}_{\theta_0},\dot{\tilde{\mathcal{E}}}_{\theta_0})$$

We shall now give two theorems which show that $\mathring{\delta}_{\theta_0}$ is - as the notation indicates - a sort of derivative of the deficiency δ.

<u>Theorem 6.1.</u>

Let $\mathcal{E}$ and $\tilde{\mathcal{E}}$ both be differentiable in θ_0. Then *):

$$\lim_{\varepsilon\to 0} \delta(\mathcal{E}_{[\theta_0-\varepsilon,\theta_0+\varepsilon]}, \tilde{\mathcal{E}}_{[\theta_0-\varepsilon,\theta_0+\varepsilon]})/2\varepsilon$$

$$= \lim_{\varepsilon\to 0} \delta(\mathcal{E}_{\{\theta_0-\varepsilon,\theta_0+\varepsilon\}}, \tilde{\mathcal{E}}_{\{\theta_0-\varepsilon,\theta_0+\varepsilon\}})/2\varepsilon$$

$$= \mathring{\delta}_{\theta_0}(\mathcal{E},\tilde{\mathcal{E}})$$

<u>Proof:</u> We saw in section 2 that - in a sufficiently small neighbourhood of θ_0 - we have the expansions :

$$P_\theta = P_{\theta_0} + (\theta-\theta_0)\mathring{P}_{\theta_0} + (\theta-\theta_0)\Gamma_{\theta_0,\theta}$$

and

$$\tilde{P}_\theta = \tilde{P}_{\theta_0} + (\theta-\theta_0)\tilde{P}_{\theta_0} + (\theta-\theta_0)\tilde{\Gamma}_{\theta_0,\theta}$$

where $\sup_\theta \|\Gamma_{\theta_0,\theta}\| + \|\tilde{\Gamma}_{\theta_0,\theta}\| < \infty$ and $\lim_{\theta\to\theta_0} \|\Gamma_{\theta_0,\theta}\| + \|\tilde{\Gamma}_{\theta_0,\theta}\| = 0$.

---

*) {a,b,···} is the set whose elements are a,b,··· .

(i): Let M be any randomization mapping P_{θ_o} on $\tilde{P}_{\theta_o}$, and let $0 < |\theta-\theta_o| \leq \varepsilon$. Then:

$$\|P_\theta M-\tilde{P}_\theta\| = \|P_{\theta_o}M+(\theta-\theta_o)\dot{P}_{\theta_o}M+(\theta-\theta_o)\Gamma_{\theta_o,\theta}M-\tilde{P}_{\theta_o}-(\theta-\theta_o)\dot{\tilde{P}}_{\theta_o}-(\theta-\theta_o)\tilde{\Gamma}_{\theta_o,\theta}\|$$

$$= |\theta-\theta_o| \, \|(\dot{P}_{\theta_o}M-\dot{\tilde{P}}_{\theta_o}) + (\Gamma_{\theta_o,\theta}M-\tilde{\Gamma}_{\theta_o,\theta})\|$$

$$\leq \varepsilon\|\dot{P}_{\theta_o}M-\dot{\tilde{P}}_{\theta_o}\| + \varepsilon\sup\{\|\Gamma_{\theta_o,\theta}\| + \|\tilde{\Gamma}_{\theta_o,\theta}\|: 0 < |\theta-\theta_o| \leq \varepsilon\}$$

by the randomization criterion for comparison of $\mathcal{E}_{[\theta_o-\varepsilon,\theta_o+\varepsilon]}$ and $\tilde{\mathcal{E}}_{[\theta_o-\varepsilon,\theta_o+\varepsilon]}$, and by the definition of $\dot{\delta}_{\theta_o}(\mathcal{E},\tilde{\mathcal{E}})$:

$$\delta(\mathcal{E}_{[\theta_o-\varepsilon,\theta_o+\varepsilon]}, \tilde{\mathcal{E}}_{[\theta_o-\varepsilon,\theta_o+\varepsilon]})/2\varepsilon \leq \dot{\delta}_{\theta_o}(\mathcal{E},\tilde{\mathcal{E}}) + a_\varepsilon$$

where $\lim_{\varepsilon\to 0} a_\varepsilon = 0$.

Hence:

$$\limsup_{\varepsilon\to 0} \delta(\mathcal{E}_{[\theta_o-\varepsilon,\theta_o+\varepsilon]}, \tilde{\mathcal{E}}_{[\theta_o-\varepsilon,\theta_o+\varepsilon]})/2\varepsilon \leq \dot{\delta}_{\theta_o}(\mathcal{E},\tilde{\mathcal{E}}).$$

(ii) By the testing criterion for comparison we have - for sufficiently small ε:

$$\delta(\mathcal{E}_{\{\theta_o-\varepsilon,\theta_o+\varepsilon\}}, \tilde{\mathcal{E}}_{\{\theta_o-\varepsilon,\theta_o+\varepsilon\}}) = \sup_{0<\lambda<1}\Big[\|(1-\lambda)\tilde{P}_{\theta_o+\varepsilon}-\lambda\tilde{P}_{\theta_o-\varepsilon}\|$$

$$-\|(1-\lambda)P_{\theta_o+\varepsilon}-\lambda P_{\theta_o-\varepsilon}\|\Big] = \sup_{0<\lambda<1}\Big[\|(1-2\lambda)\tilde{P}_{\theta_o}+\varepsilon\dot{\tilde{P}}_{\theta_o}+(1-\lambda)\varepsilon\tilde{\Gamma}_{\theta_o,\theta_o+\varepsilon}+\lambda\varepsilon\tilde{\Gamma}_{\theta_o,\theta_o-\varepsilon}\|$$

$$-\|(1-2\lambda)P_{\theta_o}+\varepsilon\dot{P}_{\theta_o}+(1-\lambda)\varepsilon\Gamma_{\theta_o,\theta_o+\varepsilon}+\lambda\varepsilon\Gamma_{\theta_o,\theta_o-\varepsilon}\|\Big]$$

$$= \sup_{0<\lambda<1}\Big[\|(1-2\lambda)\tilde{P}_{\theta_o}+\varepsilon\dot{\tilde{P}}_{\theta_o}\| - \|(1-2\lambda)P_{\theta_o}+\varepsilon\dot{P}_{\theta_o}\|\Big]+\varepsilon b_\varepsilon$$

where $|b_\varepsilon| \leqq \|\Gamma_{\theta_o,\theta_o+\varepsilon}\| + \|\Gamma_{\theta_o,\theta_o-\varepsilon}\| + \|\tilde{\Gamma}_{\theta_o,\theta_o+\varepsilon}\| + \|\tilde{\Gamma}_{\theta_o,\theta_o-\varepsilon}\|$

It follows that $\lim_{\varepsilon\to 0} b_\varepsilon = 0$ and :

$$\liminf_{\varepsilon\to 0} \delta(\mathcal{E}_{\{\theta_o-\varepsilon,\theta_o+\varepsilon\}}, \tilde{\mathcal{E}}_{\{\theta_o-\varepsilon,\theta_o+\varepsilon\}})/2\varepsilon$$

$$= \liminf_{\varepsilon\to 0} \sup_{0<\lambda<1} \left[\|((1-2\lambda)/\varepsilon)\tilde{P}_{\theta_o} + \dot{\tilde{P}}_{\theta_o}\| - \|(1-2\lambda)/\varepsilon\, P_{\theta_o} + \dot{P}_{\theta_o}\|\right]/2$$

Let ξ be any real number and choose $\varepsilon > 0$ so small that $[\theta_o-\varepsilon,\theta_o+\varepsilon] \subseteqq \Theta$ and $\varepsilon^{-1} > |\xi|$. It follows that $\frac{1+\xi\varepsilon}{2} \in]0,1[$ and therefore is a possible value of λ. Hence:

$$\liminf_{\varepsilon\to 0} \delta(\mathcal{E}_{\{\theta_o-\varepsilon,\theta_o+\varepsilon\}}, \tilde{\mathcal{E}}_{\{\theta_o-\varepsilon,\theta_o+\varepsilon\}})/2\varepsilon$$

$$\geqq \left[\|\xi\tilde{P}_{\theta_o} - \dot{\tilde{P}}_{\theta_o}\| - \|\xi P_{\theta_o} - \dot{P}_{\theta_o}\|\right]/2 = \left[\tilde{U}_{\theta_o}(\xi) - U_{\theta_o}(\xi)\right]/2$$

and - since deficiencies are non negative :

$$\liminf_{\varepsilon\to 0} \delta(\mathcal{E}_{\{\theta_o-\varepsilon,\theta_o+\varepsilon\}}, \tilde{\mathcal{E}}_{\{\theta_o-\varepsilon,\theta_o+\varepsilon\}})/2\varepsilon \geqq \sup_\xi \left[\tilde{U}_{\theta_o}(\xi) - U_{\theta_o}(\xi)\right]/2$$

$$= \dot{\delta}_{\theta_o}(\mathcal{E}, \tilde{\mathcal{E}}).$$

(iii) We get successively :

$$\dot{\delta}_{\theta_o}(\mathscr{E},\tilde{\mathscr{E}}) \leqq \qquad \text{(by (ii))}$$

$$\liminf_{\varepsilon\to 0} \delta(\mathscr{E}_{\{\theta_o-\varepsilon,\theta_o+\varepsilon\}},\tilde{\mathscr{E}}_{\{\theta_o-\varepsilon,\theta_o+\varepsilon\}})/2\varepsilon \leqq$$

$$\text{(since } \{\theta_o-\varepsilon,\theta_o+\varepsilon\}\subseteqq[\theta_o-\varepsilon,\theta_o+\varepsilon])$$

$$\liminf_{\varepsilon\to 0} \delta(\mathscr{E}_{[\theta_o-\varepsilon,\theta_o+\varepsilon]},\tilde{\mathscr{E}}_{[\theta_o-\varepsilon,\theta_o+\varepsilon]})/2\varepsilon \leqq$$

$$\limsup_{\varepsilon\to 0} \delta(\mathscr{E}_{[\theta_o-\varepsilon,\theta_o+\varepsilon]},\tilde{\mathscr{E}}_{[\theta_o-\varepsilon,\theta_o+\varepsilon]})/2\varepsilon \leqq \qquad \text{(by (i))}$$

$$\dot{\delta}_{\theta_o}(\mathscr{E},\tilde{\mathscr{E}}).$$

It follows that these inequalities are all equalities. Hence

$$\limsup_{\varepsilon\to 0} \delta(\mathscr{E}_{\{\theta_o-\varepsilon,\theta_o+\varepsilon\}},\tilde{\mathscr{E}}_{\{\theta_o-\varepsilon,\theta_o+\varepsilon\}})/2\varepsilon \leqq$$

$$\text{(since } \{\theta_o-\varepsilon,\theta_o+\varepsilon\}\subseteqq[\theta_o-\varepsilon,\theta_o+\varepsilon])$$

$$\limsup_{\varepsilon\to 0} \delta(\mathscr{E}_{[\theta_o-\varepsilon,\theta_o+\varepsilon]},\tilde{\mathscr{E}}_{[\theta_o-\varepsilon,\theta_o+\varepsilon]})/2\varepsilon =$$

$$\liminf_{\varepsilon\to 0} \delta(\mathscr{E}_{\{\theta_o-\varepsilon,\theta_o+\varepsilon\}},\tilde{\mathscr{E}}_{\{\theta_o-\varepsilon,\theta_o+\varepsilon\}})/2\varepsilon.$$

and this completes the proof. □

Instead of averaging over the interval $[\theta_o-\varepsilon,\theta_o+\varepsilon]$ - as we did in theorem 6.1 - we might as well use "one sided" intervals $[\theta_o-\varepsilon,\theta_o]$ or $[\theta_o,\theta_o+\varepsilon]$.

Theorem 6.2.

Let $\mathcal{E}$ and $\tilde{\mathcal{E}}$ both be differentiable in θ_o. Then:

$$\lim_{\varepsilon\to 0} \delta(\mathcal{E}_{[\theta_o-\varepsilon,\theta_o]},\tilde{\mathcal{E}}_{[\theta_o-\varepsilon,\theta_o]})/\varepsilon=\lim_{\varepsilon\to 0}\delta(\mathcal{E}_{\{\theta_o-\varepsilon,\theta_o\}},\tilde{\mathcal{E}}_{\{\theta_o-\varepsilon,\theta_o\}})/\varepsilon$$

$$= \dot{\delta}_{\theta_o}(\mathcal{E},\tilde{\mathcal{E}}),$$

and

$$\lim_{\varepsilon\to 0} \delta(\mathcal{E}_{[\theta_o,\theta_o+\varepsilon]},\tilde{\mathcal{E}}_{[\theta_o,\theta_o+\varepsilon]})/\varepsilon=\lim_{\varepsilon\to 0}(\mathcal{E}_{\{\theta_o,\theta_o+\varepsilon\}},\tilde{\mathcal{E}}_{\{\theta_o,\theta_o+\varepsilon\}})/\varepsilon$$

$$= \dot{\delta}_{\theta_o}(\mathcal{E},\tilde{\mathcal{E}}).$$

Proof: By the norm criterion for test deficiency:

$$\delta(\mathcal{E}_{\{\theta_o,\theta_o+\varepsilon\}},\tilde{\mathcal{E}}_{\{\theta_o,\theta_o+\varepsilon\}}) = \sup_\eta\left[(\|\tilde{P}_{\theta_o+\varepsilon}+\eta\tilde{P}_{\theta_o}\|-\|P_{\theta_o+\varepsilon}+\eta P_{\theta_o}\|)^+/1+|\eta|\right]$$

$$= \sup_\eta\left[(\|(1+\eta)\tilde{P}_{\theta_o}+\varepsilon\dot{\tilde{P}}_{\theta_o}\| - \|(1+\eta)P_{\theta_o}+\varepsilon\dot{P}_{\theta_o}\|)^+/1+|\eta|\right] + \varepsilon a_\varepsilon$$

where $\lim_{\varepsilon\to 0} a_\varepsilon = 0$. Inserting $\xi = -(1+\eta)/\varepsilon$ we get:

$$\delta(\mathcal{E}_{\{\theta_o,\theta_o+\varepsilon\}},\tilde{\mathcal{E}}_{\{\theta_o,\theta_o+\varepsilon\}})/\varepsilon = \sup_\xi\left[(U_{\tilde{\mathcal{E}}}(\xi) - U_{\mathcal{E}}(\xi))^+/(1+|1+\varepsilon\xi|)\right]+a_\varepsilon.$$

Let $\kappa > 0$. Then we may choose a $\xi_0 > 0$ so that

$$\left[U_{\tilde{\mathcal{E}}}(\xi) - U_{\mathcal{E}}(\xi)\right| = \left|\left[U_{\tilde{\mathcal{E}}}(\xi) - |\xi|\right] - \left[U_{\mathcal{E}}(\xi) - |\xi|\right]\right| < \kappa$$

when $|\xi| \geqq \xi_0$. Hence

$$\left|(U_{\tilde{\mathcal{E}}}(\xi) - U_{\mathcal{E}}(\xi))^+ \Big/ (1+|1+\varepsilon\xi|) - (U_{\tilde{\mathcal{E}}}(\xi) - U_{\mathcal{E}}(\xi))^+ \Big/ 2\right| \leqq \tfrac{3}{2}\kappa$$

when $|\xi| \geqq \xi_0$.

Next, choose $\varepsilon_0 > 0$ so small that $1-\varepsilon\xi_0 > 0$ and

$$\max_{|\xi| \leqq \xi_0} \left[|U_{\tilde{\mathcal{E}}}(\xi) - U_{\mathcal{E}}(\xi)| \quad \varepsilon\xi_0 \Big/ (2-2\varepsilon\xi_0)\right] < \kappa$$

when $\varepsilon \leqq \varepsilon_0$. It follows that

$$\left|(U_{\tilde{\mathcal{E}}}(\xi) - U_{\mathcal{E}}(\xi))^+ \Big/ (1+|1+\varepsilon\xi|) - (U_{\tilde{\mathcal{E}}}(\xi) - U_{\mathcal{E}}(\xi))^+ \Big/ 2\right| < \tfrac{3}{2}\kappa$$

for all ξ; provided $\varepsilon \leqq \varepsilon_0$. Hence

$$\left[\delta(\mathcal{E}_{\{\theta_0,\theta_0+\varepsilon\}}, \tilde{\mathcal{E}}_{\{\theta_0,\theta_0+\varepsilon\}}) \Big/ \varepsilon - \sup_{\xi}(U_{\tilde{\mathcal{E}}}(\xi) - U_{\mathcal{E}}(\xi))^+ \Big/ 2\right] \leqq \tfrac{3}{2}\kappa + a_\varepsilon$$

when $\varepsilon \leqq \varepsilon_0$. This implies that:

$$\lim_{\varepsilon \to 0} \delta(\mathcal{E}_{\{\theta_0,\theta_0+\varepsilon\}}, \tilde{\mathcal{E}}_{\{\theta_0,\theta_0+\varepsilon\}}) / \varepsilon = \dot{\delta}_{\theta_0}(\mathcal{E}, \tilde{\mathcal{E}}).$$

To prove the last statement it suffices to show that:

$$\limsup_{\varepsilon \to 0} \delta(\mathcal{E}_{[\theta_0,\theta_0+\varepsilon]}, \tilde{\mathcal{E}}_{[\theta_0,\theta_0+\varepsilon]}) \Big/ \varepsilon \leqq \dot{\delta}_{\theta_0}(\mathcal{E}, \tilde{\mathcal{E}}).$$

Consider any $\varepsilon > 0$ such that $[\theta_o, \theta_o+\varepsilon] \subseteq \Theta$. By the randomization criterion there is a randomization N_ε so that

$$\|N_\varepsilon P_{\theta_o} - \tilde{P}_{\theta_o}\| \leq \delta(\mathcal{E}_{\{\theta_o, \theta_o+\varepsilon\}}, \tilde{\mathcal{E}}_{\{\theta_o, \theta_o+\varepsilon\}})$$

and

$$\|N_\varepsilon P_{\theta_o+\varepsilon} - \tilde{P}_{\theta_o+\varepsilon}\| \leq \delta(\mathcal{E}_{\{\theta_o, \theta_o+\varepsilon\}}, \tilde{\mathcal{E}}_{\{\theta_o, \theta_o+\varepsilon\}})$$

To any $\theta \in [\theta_o, \theta_o+\varepsilon]$ there is a $\lambda_\theta \in [0,1]$ so that $\theta = \theta_o + \lambda_\theta\varepsilon = (1-\lambda_\theta)\theta_o + \lambda_\theta(\theta_o + \varepsilon)$.

Consider the accuracy of the approximation $(1-\lambda_\theta)P_{\theta_o}+\lambda_\theta P_{\theta_o+\varepsilon}$ of $P_\theta = P_{\theta_o}+(\theta-\theta_o)\dot{P}_{\theta_o}+(\theta-\theta_o)\Gamma_{\theta_o,\theta}$. We get:

$$P_\theta-(1-\lambda_\theta)P_{\theta_o}-\lambda_\theta P_{\theta_o+\varepsilon} = P_{\theta_o}+(\theta-\theta_o)\dot{P}_{\theta_o}+(\theta-\theta_o)\Gamma_{\theta_o,\theta}-(1-\lambda_\theta)P_{\theta_o}-$$
$$\lambda_\theta P_{\theta_o}-\lambda_\theta\varepsilon\, \dot{P}_{\theta_o}-\lambda_\theta\varepsilon\, \Gamma_{\theta_o,\theta_o+\varepsilon} = (\theta-\theta_o)\Gamma_{\theta_o,\theta} +\lambda_\theta\varepsilon\Gamma_{\theta_o,\theta_o+\varepsilon}\,.$$

Hence (by this and the analoguous expansion of $\tilde{P}_\theta$) :

$$\|N_\varepsilon P_\theta-\tilde{P}_\theta\| = \|(1-\lambda_\theta)N_\varepsilon P_{\theta_o}+\lambda_\theta N_\varepsilon P_{\theta_o+\varepsilon}+(\theta-\theta_o)N_\varepsilon\Gamma_{\theta_o,\theta}+\lambda_\theta\varepsilon N_\varepsilon\Gamma_{\theta_o,\theta_o+\varepsilon}$$
$$-(1-\lambda_\theta)\tilde{P}_{\theta_o}-\lambda_\theta\tilde{P}_{\theta_o+\varepsilon}-(\theta-\theta_o)\tilde{\Gamma}_{\theta_o,\theta}-\lambda_\theta\varepsilon\tilde{\Gamma}_{\theta_o,\theta+\varepsilon}\| \leq$$

$(1-\lambda_\theta)\|N_\varepsilon P_{\theta_o}-\tilde{P}_{\theta_o}\|+ \lambda_\theta\|N_\varepsilon P_{\theta_o+\varepsilon}-\tilde{P}_{\theta_o+\varepsilon}\| + \varepsilon a_\varepsilon$ where a_ε does not depend on θ and $\lim_{\varepsilon\to 0} a_\varepsilon = 0$. It follows that

$$\|N_\varepsilon P_\theta - \tilde{P}_\theta\| \leq \delta\,(\mathcal{E}_{\{\theta_0,\theta_0+\varepsilon\}}\ \tilde{\mathcal{E}}_{\{\theta_0,\theta_0+\varepsilon\}}) + \varepsilon a_\varepsilon$$

Hence

$$\delta(\mathcal{E}_{[\theta_0,\theta_0+\varepsilon]}, \tilde{\mathcal{E}}_{[\theta_0,\theta_0+\varepsilon]})/\varepsilon \leq \delta(\mathcal{E}_{\{\theta_0,\theta_0+\varepsilon\}}, \tilde{\mathcal{E}}_{\{\theta_0,\theta_0+\varepsilon\}})/\varepsilon + a_\varepsilon$$

so that

$$\limsup_{\varepsilon\to 0} \delta(\mathcal{E}_{[\theta_0,\theta_0+\varepsilon]}, \tilde{\mathcal{E}}_{[\theta_0,\theta_0+\varepsilon]})/\varepsilon \leq$$

$$\lim_{\varepsilon\to 0} \delta(\mathcal{E}_{\{\theta_0,\theta_0+\varepsilon\}}, \tilde{\mathcal{E}}_{\{\theta_0,\theta_0+\varepsilon\}}) = \dot{\delta}_{\theta_0}(\mathcal{E}, \tilde{\mathcal{E}}).$$

The first statement follows from the last by a symmetry argument. □

Remark. Theorems 6.1 and 6.2 imply that local comparison based upon $\dot{\delta}_{\theta_0}$ is asymptotically equivalent with δ-comparison of "statistical" dichotomies.

Corollary 6.3.

Let $\mathcal{E}$ and $\tilde{\mathcal{E}}$ both be differentiable in θ_0. Then:

$$\lim_{\varepsilon\to 0}\Delta(\mathcal{E}_{[\theta_0-\varepsilon,\theta_0+\varepsilon]}, \tilde{\mathcal{E}}_{[\theta_0-\varepsilon,\theta_0+\varepsilon]})/2\varepsilon = \lim_{\varepsilon\to 0}\Delta(\mathcal{E}_{\{\theta_0-\varepsilon,\theta_0+\varepsilon\}}, \tilde{\mathcal{E}}_{\{\theta_0-\varepsilon,\theta_0+\varepsilon\}})/2\varepsilon$$

$$= \dot{\Delta}_{\theta_0}(\mathcal{E}, \tilde{\mathcal{E}}) \quad ,$$

$$\lim_{\varepsilon\to 0}\Delta(\mathcal{E}_{[\theta_0-\varepsilon,\theta_0]}, \tilde{\mathcal{E}}_{[\theta_0-\varepsilon,\theta_0]})/\varepsilon = \lim_{\varepsilon\to 0}\Delta(\mathcal{E}_{\{\theta_0-\varepsilon,\theta_0\}}, \tilde{\mathcal{E}}_{\{\theta_0-\varepsilon,\theta_0\}})/\varepsilon$$

$$= \dot{\Delta}_{\theta_0}(\mathcal{E}, \tilde{\mathcal{E}})$$

and

$$\lim_{\varepsilon\to 0}\Delta(\mathcal{E}_{[\theta_0,\theta_0+\varepsilon]},\tilde{\mathcal{E}}_{[\theta_0,\theta_0+\varepsilon]})/\varepsilon = \lim_{\varepsilon\to 0}\Delta(\mathcal{E}_{\{\theta_0,\theta_0+\varepsilon\}},\tilde{\mathcal{E}}_{\{\theta_0,\theta_0+\varepsilon\}})/\varepsilon$$

$$\dot{\Delta}_{\theta_0}(\mathcal{E},\tilde{\mathcal{E}}).$$

Corollary 6.4.

Let $\mathcal{E}_i$, $i=1,\cdots,n$ and $\mathcal{F}_i$, $i=1,\cdots,n$ be differentiable in θ_0. Then:

$$\dot{\delta}_{\theta_0}(\prod_{i=1}^{n}\mathcal{E}_i,\prod_{i=1}^{n}\mathcal{F}_i) \leqq \sum_{i=1}^{n}\dot{\delta}_{\theta_0}(\mathcal{E}_i,\mathcal{F}_i)$$

and

$$\dot{\Delta}_{\theta_0}(\prod_{i=1}^{n}\mathcal{E}_i,\prod_{i=1}^{n}\mathcal{F}_i) \leqq \sum_{i=1}^{n}\dot{\Delta}_{\theta_0}(\mathcal{E}_i,\mathcal{F}_i).$$

Proof: By proposition 2.1; $\Pi\mathcal{E}_i$ and $\Pi\mathcal{F}_i$ are both differentiable in θ_0. The last statement follows easily from the first and the first statement is - by theorem 6.2 - a consequence of the corresponding result for δ. The inequality - $\delta(\Pi\mathcal{E}_i,\Pi\mathcal{F}_i) \leqq \sum\delta(\mathcal{E}_i,\mathcal{F}_i)$ - follows directly from corollary 4 in section 2 in [15] and remark 2 in section 1 in the same paper. □

In order to interpret $\dot{\delta}_{\theta_0}(\mathcal{E},\tilde{\mathcal{E}})$ in terms of operational characteristics, note first that for any randomization M from $(X,\mathcal{A})$ to $(\tilde{X},\tilde{\mathcal{A}})$ such that $P_{\theta_0}M = \tilde{P}_{\theta_0}$ we have:

$$\lim_{\theta\to\theta_0} (P_\theta M-\tilde{P}_\theta)/(\theta-\theta_0) = \dot{P}_{\theta_0} M-\dot{\tilde{P}}_{\theta_0}$$

We have almost proved:

Proposition 6.5.

$$2\dot{\delta}_{\theta_0}(\mathcal{E},\tilde{\mathcal{E}}) = \min_M \lim_{\theta\to\theta_0} \|P_\theta M-\tilde{P}_\theta\| / |\theta-\theta_0|$$

Proof: Since $\lim_{\theta\to\theta_0} \|\tilde{P}_\theta M-\tilde{P}_\theta\|/|\theta-\theta_0| = \infty$ when $P_{\theta_0} M \neq \tilde{P}_{\theta_0}$, we may restrict our attention to randomizations M such that $P_{\theta_0} M = \tilde{P}_{\theta_0}$. The proposition now follows directly from the definition. □

We are now ready to describe $\dot{\delta}_{\theta_0}(\mathcal{E},\tilde{\mathcal{E}})$ in terms of operational characteristics:

Theorem 6.6.

Let $(T,\mathcal{S})$ be a decision space and let $\tilde{\sigma}$ be any decision procedure from $\tilde{\mathcal{E}}$ to $(T,\mathcal{S})$. Then there is a decision procedure σ from $\mathcal{E}$ to $(T,\mathcal{S})$ so that:

(§) $$\limsup_{\theta\to\theta_0} \|P_\theta\sigma-\tilde{P}_\theta\tilde{\sigma}\|/|\theta-\theta_0| \leq 2\dot{\delta}_{\theta_0}(\mathcal{E},\tilde{\mathcal{E}})$$

It may, however, be no $\tilde{\sigma}$ from $\mathcal{E}$ to $(T,\mathcal{S})$ so that

(§§) $$\limsup_{\theta\to\theta_0} \|P_\theta\sigma-\tilde{P}_\theta\tilde{\sigma}\|/|\theta-\theta_0| < 2\dot{\delta}_{\theta_0}(\mathcal{E},\tilde{\mathcal{E}})$$

Proof: Let M be chosen so that $\|P_\theta M - \tilde{P}_\theta\| / |\theta-\theta_0| \to 2\dot{\delta}_{\theta_0}(\mathcal{E}, \tilde{\mathcal{E}})$. Then - by proposition 6.5 - $\sigma = \tilde{\sigma} \circ M$ satisfies (§). Consider now $(\tilde{X}, \tilde{\mathcal{A}})$ as a decision space and let $\tilde{\sigma}$ be the identity function. It follows from proposition 6.5 that no $\tilde{\sigma}$ satisfies (§§). □

Consider now the problem of finding tests maximizing or minimizing under side conditions, the slope of the power function in θ_0. Let $\mathcal{E}$ be differentiable in θ_0 and let $\alpha_0 \in]0,1[$ be a point where β attains its maximum. Note first that the distribution function β^+, of the measure β^+, which vanishes at 0 is given by:

$$\beta^+(\alpha) = \beta(\min(\alpha,\alpha_0)) = \begin{cases} \beta(\alpha) & \text{when } \alpha \leqq \alpha_0 \\ \max_\alpha \beta(\alpha) & \text{" } \alpha > \alpha_0 \end{cases}$$

Similarily the distribution function β^-, of the measure β^-, which vanishes at 1 is given by:

$$\beta^-(\alpha) = -\beta(\max(\alpha,\alpha_0)) = \begin{cases} -\max_\alpha \beta(\alpha) & \text{when } \alpha \leqq \alpha_0 \\ -\beta(\alpha) & \text{" } \alpha \geqq \alpha_0 \end{cases}$$

The connection between these functions and the slope problem is described in

Proposition 6.7.

The maximal slope at θ_0 among size α tests for "$\theta=\theta_0$" against "$\theta>\theta_0$" is $\beta(\alpha)$.

The maximal slope at θ_0 among tests with level of significance α for "$\theta=\theta_0$" against "$\theta>\theta_0$" is $\beta^+(\alpha)$.

The minimal slope at θ_o among size α tests for "$\theta=\theta_o$" against "$\theta<\theta_o$" is $-\beta(1-\alpha)$.

The minimal slope at θ_o among tests with level of significance α for "$\theta=\theta_o$" against "$\theta<\theta_o$" is $\beta^-(1-\alpha)$.

<u>Proof:</u> We choose - since the verifications are very similar - to prove the last statement. The minimal slope is the number:

$\min\{y: (x,y)\in V \ \&\ x \leqq \alpha\} = -\max\{-y: (1-x,-y)\in V \ \&\ 1-x \geqq 1-\alpha\}$

$= -\max\{y: (x,y)\in V \ \&\ x \geqq 1-\alpha\} = \beta^-(1-\alpha)$. □

We summarize here - for the sake of completeness - a few simple and essentially known facts on local properties of tests for "$\theta=\theta_o$" against "$\theta>\theta_o$". A level α test δ will be called locally most powerfull (LMP) if to any other level α test $\tilde{\delta}$ there corresponds a $\varepsilon_{\tilde{\delta}} > 0$ so that

$$E_\theta\delta \geqq E_\theta\tilde{\delta}; \qquad \theta_o < \theta \leqq \theta_o + \varepsilon_{\tilde{\delta}} .$$

δ will be called uniformly locally most powerfull (ULMP) if $\varepsilon_{\tilde{\delta}}$ may be chosen so that it does not depend on the particular level α test $\tilde{\delta}$. Trivially any ULMP level α test is a LMP level α test and any LMP level α test has size α and maximizes the slope at θ_o among all size α tests. We may also define the properties LMP and ULMP w.r.t. a specified class of tests.

Let c be any $1-\alpha$ fractile of F_{θ_0} *). It is easily seen that a test δ has size α and maximizes the slope at θ_0 among all tests of size α if and only if :

(i) $\delta = 1$ a.s P_{θ_0} on $[s_{\theta_0} > c]$

(ii) $\int_{s_{\theta_0}=c} \delta dP_{\theta_0} = P_{\theta_0}(s=c) - (1-\alpha)$

(iii) $\delta = 0$ a.s P_{θ_0} on $[s_{\theta_0} < c]$.

In particular test of the form $I_{[s_{\theta_0}>d]} + \gamma I_{[s_{\theta_0}=d]}$ where $\gamma \in [0,1]$ and $d \in [-\infty,+\infty]$ maximizes the slope at θ_0 among all tests having the given size.

If we require our test to be a.s P_{θ_0}, s_{θ_0} measurable, then (i) - (iii) determines - up to P_{θ_0} equivalence - δ. This is no restriction when $P_{\theta_0}(s_{\theta_0}=c) = 0$. It may be checked - provided $P_{\theta_0} >> P_\theta$ when $\theta>\theta_0$ - that a s_{θ_0} measurable test which maximizes the slope among all size α tests is LMP w.r.t. all s_{θ_0} measurable α tests. If - moreover - $P_{\theta_0}(s_{\theta_0}=c) = 0$, then such a test is a LMP level α test. Any test of the form "$s_{\theta_0} \geqq d$" is - provided $P_{\theta_0} >> P_\theta$ when $\theta \in [\theta_0,\theta_0+\varepsilon]$ - a LMP level $P_{\theta_0}(s_{\theta_0} \geqq d)$ test.

If X is finite and $\mathcal{A}$ is the class of all sub sets then

---

*) An element $a_p \in [-\infty,+\infty]$ is called p - ($p \in [0,1]$) - fractile for the probability measure P on R if $P(]-\infty,x_p[) \leqq p \leqq P(]-\infty,x])$.

there is a $\varepsilon > 0$ so that $[dP_\theta/dP_{\theta_0}]_{x_1} > [dP_\theta/dP_{\theta_0}]_{x_2}$ when $s_{\theta_0}(x_1) > s_{\theta_0}(x_2)$ and $\theta \in]\theta_0, \theta_0+\varepsilon]$. In this case the test "$s_{\theta_0} \geq d$" - where d is a constant in $[-\infty,\infty]$ - is a LMP level $P_{\theta_0}(s_{\theta_0} \geq d)$ test for testing "$\theta=\theta_0$" against $\theta \in]\theta_0,\theta_0+\varepsilon]$.

<u>Example 6.8</u> (Rank tests). This example is modelled after the theory in II.4.8 in Hájek and Šidák [4]

Consider an experiment $\mathcal{E}$ of the form $\mathcal{E} = ((]-\infty,\infty[^n$, Borel class), $P_\theta : \theta \in \Theta)$ such that:

(i) $P_\theta(\{(x_1,\cdots,x_n): x_1,\cdots,x_n$ are all different$\}) = 1; \ \theta \in \Theta$

(ii) P_{θ_0} is symmetric, i.e. $\int h(x_{\pi 1},\cdots,x_{\pi n}) P_{\theta_0}(d(x_1,\cdots,x_n))$

$= \int h \, d P_{\theta_0}$ for any permutation π of $\{1,\cdots,n\}$ and any bounded measurable function h on $]-\infty,+\infty[^n$

(iii) $\mathcal{E}$ is differentiable in θ_0.

For each $i \in \{1,\cdots,n\}$ and each $(x_1,\cdots,x_n)$ in $]-\infty,+\infty[^n$ the rank r_i of x_i in $(x_1,\cdots,x_n)$ is the number of subscripts j such that $x_j \leq x_i$. $r = (r_1,\cdots,r_n)$ is a permuation of $\{1,2,\cdots,n\}$ provided $x_1,\cdots,x_n$ are all different.

The order statistic ,ord, is the function $x \rightsquigarrow (\text{ord}_1(x), \text{ord}_2(x),\cdots,\text{ord}_n(x))$ where $\text{ord}_1(x) \leq \text{ord}_2(x) \leq \cdots \leq \text{ord}_n(x)$ are $x_1,\cdots,x_n$ arranged in increasing order. If there are no repetitions in $(x_1,\cdots,x_n)$ then $x_i = \text{ord}_{r_i(x)}(x)$; $i = 1,\cdots,n$. It is easily seen that r and ord are independent under P_{θ_o}.

For each permutation π of $\{1,2,\cdots n\}$ write

$$a(\pi) = E_{\theta_o} s_{\theta_o}(\text{ord}_{\pi(1)}(X),\cdots,\text{ord}_{\pi(n)}(X)) .$$

Then $a(r)$ is - under P_{θ_o} - a version of $E^r_{\theta_o} s_{\theta_o}$. It follows that there is a $\varepsilon > 0$ so that any test of the form $a(r) \geqq d$ or $a(r) > d$ is ULMP among all rank tests with, respectively, level $P_{\theta_o}(a(r) \geqq d)$ and $P_{\theta_o}(a(r) > d)$. If each P_{θ_o} is a product measure $P_{\theta_1} \times \cdots \times P_{\theta n}$, then - by proposition 2.1 - $a(\pi) = \sum_{i=1}^{n} a_i(\pi(i))$ where $a_i(j) = E_{\theta_o}\left[d\dot{P}_{\theta_o,i}/dP_{\theta_o,i}\right]\text{ord}_j$.

Note that it may - as in the two sample problems - happen that there are i's such that $P_{\theta,i}$ does not depend on θ. The corresponding random variables may then - by the "principle" of sufficiency - be excluded from the sample. No damage is done by that, since all information is stored in the remaining variables. The ranks, of the remaining variables, w.r.t. the non deleted variables, however, may contain no information at all. We are only pointing out the - perhaps trivial - fact, that ranks computed within a sufficient set of variables may be worthless. The ranks of the sufficient variables may - on the other hand - be "locally rank sufficient". This example generalizes somewhat the theory in II. 4.8 in [4].

The set of possible levels of tests of the form "$s_{\theta_0} \geqq d$" may not contain numbers sufficiently close to some preassigned level. It may happen that this set is $\{0,1\}$. This is the case if and only if $\beta = 0$. All experiments with $\beta = 0$ are, of course, equivalent in the $\dot{\Delta}$ sense. A more carefull analysis based upon derivatives of higher order reveals that their behaviour in small neighbourhoods of θ_0 may vary much. In particular "the local behaviour" may be very different from that of a trivial experiment where P_θ does not depend on θ. We shall see that $\beta = 0$ does not exclude the possibility of a large collection of ULMP tests.

A convenient way to express this possibility is to use the lexicographic ordering $\underset{lex}{\geqq}$ in $[-\infty,+\infty]^n$ corresponding to the coordinate wise ordering $\geqq$. More precisely: $x \underset{lex}{\geqq} y$ if and only if either $x = y$ or there is a j so that $x_i = x_j$ when $i < j$ and $x_j > y_j$. The ordering $\underset{lex}{\geqq}$ is a total ordering of $[-\infty,+\infty]^n$. If $x \underset{lex}{\geqq} y$ and $x \neq y$ then we will write $x \underset{lex}{>} y$.

<u>Proposition 6.9.</u> Consider an experiment

$\mathcal{X} = ((\chi,\mathcal{A}),\ P_{\theta,\eta};\ \theta \in [\theta_0,\theta_0+\varepsilon[,\ \eta \in \mathcal{N})$ such that

(§) $P_{\theta_0,\eta}$ does not depend on η. This measure will be denoted by P_{θ_0}.

(§§) There are $r \geqq 1$ finite measures $P_{\theta_0}^{(1)}, P_{\theta_0}^{(2)}, \dots, P_{\theta_0}^{(r)}$ so that

$$\lim_{\theta\to\theta_0} \| P_{\theta,\eta} - P_{\theta_0} - \sum_{i=1}^{r} (\theta-\theta_0)^i P_{\theta_0}^{(i)} \| / (\theta-\theta_0)^r = 0 ,$$

- uniformly in η.

(§§§) χ is finite and $\mathcal{A}$ is the class of all sub sets of χ.

For each $i = 1,\cdots,r$ and each $x \in X$ put $s^{(i)}_{\theta_o}(x) = P^{(i)}_{\theta_o}(x) / P_{\theta_o}(x)$ or $= \infty$ as $P_{\theta_o}(x) > 0$ or $P_{\theta_o}(x) = 0$. Let t_{θ_o} denote the map $x \rightsquigarrow (s^{(1)}_{\theta_o}(x),\cdots,s^{(r)}_{\theta_o}(x))$ from X to $[-\infty,\infty]^r$. Write $P_{\theta,\eta}(x)/P_{\theta_o}(x) = \infty$ if $P_{\theta_o}(x) = 0$.

Then there is a $\varepsilon > 0$ so that $P_{\theta,\eta}(x_2)/P_{\theta_o}(x_2) > P_{\theta,\eta}(x_1)/P_{\theta_o}(x_1)$ when $t_{\theta_o}(x_2) \underset{lex}{>} t_{\theta_o}(x_1)$ and $\theta \in]\theta_o,\theta_o+\varepsilon]$. Any test of the form

$$"t_{\theta_o} \underset{lex}{\geq} d"$$

where d is a constant in $[-\infty,\infty]^r$ is a UMP level $P_{\theta_o}(t_{\theta_o} \geq d)$ test for testing $"\theta = \theta_o"$ against $"\theta \in]\theta_o,\theta_o+\varepsilon]"$.

<u>Proof.</u> 1^o. Define - for each $\theta > \theta_o$ - $\psi_{\theta,\eta}(x)$ by the expansion

$$P_{\theta,\eta}(x) = P_{\theta_o}(x) + \sum_{i=1}^{r} (\theta-\theta_o)^i \, P^{(i)}_{\theta_o}(x) + (\theta-\theta_o)^r \psi_{\theta,\eta}(x).$$

In order to prove the first statement it suffices to show that to each pair $(x_1,x_2) \in X \times X$ such that $t_{\theta_o}(x_2) \underset{lex}{>} t_{\theta_o}(x_1)$, $P_{\theta_o}(x_2) > 0$ and $P_{\theta_o}(x_1) > 0$ there is an $\varepsilon > 0$ so that $P_{\theta,\eta}(x_2)/P_{\theta_o}(x_2) > P_{\theta,\eta}(x_1)/P_{\theta_o}(x_1)$ when $\theta_o < \theta \leq \theta_o+\varepsilon$. Let j be the smallest index such that $s^{(j)}_{\theta_o}(x_2) > s^{(j)}_{\theta_o}(x_1)$. Then we have;

$$[P_{\theta,\eta}(x_2)/P_{\theta_o}(x_2) - P_{\theta,\eta}(x_1)/P_{\theta_o}(x_1)]/(\theta-\theta_o)^j$$

$$= s_{\theta_o}^{(j)}(x_2) - s_{\theta_o}^{(j)}(x_1) + v(\theta,\eta)$$

where $$v(\theta,\eta) = \sum_{j<i\leq r} [s_{\theta_o}^{(i)}(x_2) - s_{\theta_o}^{(i)}(x_1)](\theta-\theta_o)^{i-j}$$

$$+ [\psi_{\theta,\eta}(x_2)/P_{\theta_o}(x_2) - \psi_{\theta,\eta}(x_1)/P_{\theta_o}(x_1)](\theta-\theta_o)^{r-j}$$

By (§§)

$$\sup_\eta |v(\theta,\eta)| \to 0 \qquad \text{as} \quad \theta \downarrow \theta_o$$

Choose $\varepsilon > 0$ so small that

$$\sup_\eta |v(\theta,\eta)| < s_{\theta_o}^{(j)}(x_2) - s_{\theta_o}^{(j)}(x_1) \quad \text{when} \quad \theta \in]\theta_o,\theta_o+\varepsilon]$$

Then $$P_{\theta,\eta}(x_2)/P_{\theta_o}(x_2) > P_{\theta,\eta}(x_1)/P_{\theta_o}(x_1) \quad \text{when} \quad \theta \in]\theta_o,\theta_o+\varepsilon].$$

2^o. We may - without loss of generality - assume that $d = t_{\theta_o}(x)$ for some x in χ. Let $\theta \in]\theta_o,\theta_o+\varepsilon]$, $\eta \in \mathcal{M}$ and let $x_o \in \chi$ maximize $P_{\theta,\eta}(x)/P_{\theta_o}(x)$ subject to the condition "$d = t_{\theta_o}(x)$". Then

$$t_{\theta_o}(x) \underset{\text{lex}}{\geq} d \quad \text{when} \quad P_{\theta,\eta}(x)/P_{\theta_o}(x) > P_{\theta,\eta}(x_o)/P_{\theta_o}(x_o)$$

and

$$t_{\theta_o}(x) \underset{\text{lex}}{<} d \quad \text{when} \quad P_{\theta,\eta}(x)/P_{\theta_o}(x) < P_{\theta,\eta}(x_o)/P_{\theta_o}(x_o).$$

It follows now from Neyman Pearson lemma that the test "$t_{\theta_o} \underset{\text{lex}}{\geq} d$" is UMP for testing "$\theta-\theta_o$" against "$\theta \in]\theta_o,\theta_o+\varepsilon]$".

□

Example 6.10. (Rank test for independence). This example is modelled after the theory in II. 4.11 in Hájek and Šidák [4].

Consider an experiment $\mathcal{H}$ of the form

$$\mathcal{H} = \left(\left(\prod_{j=1}^{n}\prod_{i=1}^{k}\,]-\infty,\infty[\,,\text{Borel class}\right), Q_{\theta,M};\ \theta\in\,]-\infty,\infty[\,,\ M\in\mathcal{M}\right)$$

such that

(i) There are non atomic probability measures $P_{\theta,i,j}$; $\theta\in\,]-\infty,\infty[$, $i=1,\cdots,k$, $j=1,\cdots,n$ so that

(i_1) $P_{0,i,j}$ does not depend on j. We will write P_i instead of $P_{0,i,j}$.

(i_2) There are finite measures $\dot{P}_{ij}$ so that

$$\lim_{\theta\to 0}\|(P_{\theta,i,j} - P_i)/\theta - \dot{P}_{ij}\| = 0, \text{ uniformly in } M\in\mathcal{M}.$$

The measure $\sum_{1\leq h<v\leq k}\prod_{i<h}P_i\times\dot{P}_{h,j}\times\prod_{h<i<v}P_i\times\dot{P}_{v,j}\times\prod_{i>v}P_i$ will be denoted by V_j.

(i_3) $P_{\theta,i,j}(\mathcal{B})$ is measurable in θ.

(ii) $\mathcal{M}$ is a collection of probability measure on $]-\infty,\infty[$ so that $t \rightsquigarrow t^2$ is uniformly integrable w.r.t. $\mathcal{M}$. If $M\in\mathcal{M}$ then μ_M and σ_M^2 denotes, respectively, the expectation in M and the variance in M.

(iii) $Q_{\theta,M} = \prod_{j=1}^{n}S_{\theta,M,j}$ where $S_{\theta,M,j} = \int\left(\prod_{i=1}^{k}P_{\theta t,i,j}\right)M(dt)$.

We will - since $S_{0,M,j}$ and $Q_{0,M}$ do not depend on M and j - write S_0 and Q_0 instead of, respectively, $S_{0,M,j}$ and $Q_{0,M}$.

$\mathcal{X}$ may be obtained by observing real random variables X_{ij}; $i = 1,\cdots,k$, $j=1,\cdots,n$ such that:

(*) There are random variables $T_1,\cdots,T_n$ so that the $k+1$ dimensional random vectors $(X_{1j}, X_{2j},\cdots X_{kj}, T_j)$ are stochastically independent. $T_1,\cdots,T_n$ may not be observable.

(**) $X_{1j}, X_{2j},\cdots, X_{kj}$ are conditionally independent given T_j.

(***) $P_{\theta T_j,i,j}$ is a conditional distribution of X_{ij} given T_j.

(****) $T_1,\cdots,T_n$ are independently and identically distributed, each T_j having the distribution M.

The joint distribution of X_{ij}, $i=1,\cdots,k$, $j=1,\cdots,n$ is - under (*), (**), (***) and (****) - $Q_{\theta,M}$.

Let x_{ij}, $i=1,\cdots,k$, $j=1,\cdots,n$ be a point in $\prod_{j=1}^{n}\prod_{i=1}^{k}]-\infty,\infty[$ such that $x_{ij_1} \neq x_{ij_2}$ when $j_1 \neq j_2$. For each i - $i=1,2,\cdots,k$ - the vector $(x_{i1},\cdots,x_{i,n})$ will be written $x_{i.}$. The rank of x_{ij} w.r.t. $x_{i.}$ and the j-th order statistic w.r.t. $x_{i.}$ will be written, respectively, r_{ij} and ord_{ij}. The symbol o_θ may - in this example - represent any quantity which converges to 0 as $\theta \to 0$ uniformly in M. Finally put $s_{ij} = d\dot{P}_{ij}/dP_i$ and $a_{ij}(\ell) = E_0 s_{ij}(\text{ord}_{i,\ell}(X_{i.}))$. We shall now show that there is an $\varepsilon > 0$ so that any test of the form

$$\text{"} \sum_{j=1}^{n} \sum_{1\leq h<v\leq k} a_{hj}(r_{hj})a_{vj}(r_{vj}) \geq \text{constant "}$$

is UMP among all rank tests for testing

$$" \theta = 0 " \text{ against } " 0 < \theta \leqq \varepsilon " \text{ at the level}$$

$$Q_o\left(\sum_{j=1}^{n} \sum_{1\leqq h<v\leqq k} a_{hj}(r_{hj})a_{vj}(r_{vj}) \geqq \text{constant}\right)$$

If $\dot{P}_{ij}$ does not depend on j, then we may write s_i and $a_i(\ell)$ instead of, respectively, s_{ij} and $a_{ij}(\ell)$. Using the formula $(\Sigma y_j)^2 = \Sigma y_j^2 + 2 \sum_{h<v} y_h y_v$ we see that these tests are - in this particular case - precisely the tests of the form:

$$" \sum_{j=1}^{n} \left(\sum_{h=1}^{k} a_h(r_{hj})\right)^2 \geqq \text{constant} "$$

Note that X_{ij}, $i=1,\cdots,k$, $j=1,\cdots,n$ are - since $Q_o = \prod_{j=1}^{n} S_o = \prod_{j=1}^{n} \prod_{i=1}^{k} P_i$ - stochastically independent when $\theta = 0$.

Let $U_{\theta,i,j}$ denote the probability measure $\int P_{\theta t,i,j}\, M(dt)$. Put $\Gamma_{\theta,i,j} = (P_{\theta,i,j} - P_i)/\theta - \dot{P}_{ij}$ or 0 as $\theta \neq 0$ or $\theta = 0$. Thus:

$$P_{\theta,i,j} = P_i + \theta\dot{P}_{ij} + \theta\Gamma_{\theta,i,j}$$

and

$$P_{\theta t,i,j} = P_i + \theta t\dot{P}_{ij} + \theta t\Gamma_{\theta t,i,j} .$$

Integration w.r.t. M gives:

$$U_{\theta,i,j} = P_i + \theta\mu_M\dot{P}_{ij} + \theta\int t\Gamma_{\theta t,i,j}M(dt) .$$

Hence - since $\int t\Gamma_{\theta t,i,j}M(dt) = o_\theta$ -

$$\prod_i U_{\theta,i,j} = \prod_i P_i$$

$$+ \theta\mu_M \sum_{h=1}^{k} \prod_{i<h} P_i \times \dot{P}_h \times \prod_{i>h} P_i$$

$$+ \theta^2\mu_M^2 V_j$$

$$+ \theta \sum_{h=1}^{k} \prod_{i<h} P_i \times \int t\Gamma_{\theta t,h,j}M(dt) \times \prod_{i>h} P_i$$

$$+ \theta^2 o_\theta$$

Similarily - using that $\int t\Gamma_{\theta t,i,j}M(dt) = o_\theta$ and $\int t^2\Gamma_{\theta t,i,j}M(dt) = o_\theta$ - we get:

$$S_{\theta,M,j} = \prod_i P_i$$

$$+ \theta\mu_M \sum_{h=1}^{k} \prod_{i<h} P_i \times \dot{P}_{hj} \times \prod_{i>h} P_i$$

$$+ \theta^2\left[\int t^2 M(dt)\right] V_j$$

$$+ \theta \sum_{h=1}^{k} \prod_{i<h} P_i \times \int t\Gamma_{\theta t,hj}M(dt) \times \prod_{i>h} P_i$$

$$+ \theta^2 o_\theta$$

It follows that:

$$S_{\theta,M,j} = \prod_i U_{\theta,i,j} + \theta^2\sigma_M^2 V_j + \theta^2 o_\theta$$

Hence - since $U_{\theta,i,j} = P_i + o_\theta(1)$ - we get:

$$Q_{\theta,M} = \prod_j \prod_i U_{\theta,i,j} + \theta^2\{\sigma_M^2 \sum_{j=1}^{n} \left[\prod_{j'<j} (\prod_i P_i)\right] \times V_j \times \left[\prod_{j'>j} (\prod_i P_i)\right]\} + \theta^2 o_\theta$$

Restricting the measures to the algebra generated by the vector of ranks $r = (r_{ij},\ i=1,\cdots,k, j=1,\cdots,n)$ we get:

$$Q_{\theta,M}(r=r^0)/Q_0(r=r_0) = 1+\theta^2\{\sigma_M^2 \sum_{j=1}^{n} \sum_{1\leq h<v\leq k} a_{hj}(r^0_{hj})a_{vj}(r^0_{vj})\}+\theta^2 o_\theta$$

[The $\left[\prod_{j'<j} (\prod_i P_i)\right] \times V_j \times \left[\prod_{j'>j} (\prod_i P_i)\right]$ measure of $[r=r^0]$ may be found by first considering the Q_0 measure of the same set and then using that

$(x_{ij},\ i=1,2,\cdots,k,\ j=1,2,\cdots,n) \rightsquigarrow \sum_{1\leq h<v\leq k} s_{hj}(x_{hj})s_{vj}(x_{hj})$ is a version of $d\{\left[\prod_{j'<j} (\prod_i P_i)\right] \times V_j \times \left[\prod_{j'>j} (\prod_i P_i)\right]\}/dQ_0$.

Using the independence of ranks and order statistics unless H_0, as in example 6.7, we get:

$$\{\left[\prod_{j'<j} (\prod_i P_i)\right] \times V_j \times \left[\prod_{j'>j} (\prod_i P_i)\right]\} \ [r=r_0]/Q_0[r=r_0]$$

$$= \sum_{1\leq h<v\leq k} a_{hj}(r^0_{hj})a_{vj}(r^0_{vj}).$$]

The β-function of the restricted experiment is obviously the 0-function on $[0,1]$. By proposition 6.8, however, there is a

$\varepsilon > 0$ so that any test of the form "$\sum_{j=1}^{n} \sum_{1 \leq h < v \leq k} a_{hj}(r_{hj})a_{vj}(r_{vj}) \geq$ constant" is a UMP level $Q_o(\sum_{j=1}^{n} \sum_{1 \leq h < v \leq k} a_{hj}(r_{hj})a_{vj}(r_{vj}) \geq \text{constant})$ test.

Let $\mathcal{E} = ((X, \mathcal{A}), P_\theta: \theta \in \Theta)$ be differentiable in θ_o. The power of any slope maximizing test of size α for testing "$\theta = \theta_o$" against "$\theta > \theta_o$" is approximately $\alpha + (\theta - \theta_o)\beta(\alpha)$ when θ is close to θ_o. We may therefore expect that the maximum power - $\beta_{\theta_o, \theta_o + \varepsilon}(\alpha)$ - among all level α tests for testing "$\theta = \theta_o$" against "$\theta = \theta_o + \varepsilon$" is - for small $\varepsilon > 0$ - approximately $\alpha + \varepsilon\beta(\alpha)$. This and other approximations are treated in the next theorem. We use the notation $\beta_{\theta_1, \theta_2}(\alpha)$ for the maximum power at θ_2 among all level α tests for testing "$\theta = \theta_1$" against "$\theta = \theta_2$". It is easily seen that $\beta_{\theta_1, \theta_2}$ and $\beta_{\theta_2, \theta_1}$ are connected through the identity:

$$\beta_{\theta_1, \theta_2}(1 - \beta_{\theta_2, \theta_1}(\alpha)) = 1 - \alpha; \qquad \alpha \leq 1 - \beta_{\theta_1, \theta_2}^{(o)} .$$

Theorem 6.11.

Denote by o_ε any quantity which converges to 0 as $\varepsilon \downarrow 0$, uniformly in $\alpha \in [0,1]$. Then we have:

(i) $$\beta_{\theta_o, \theta_o + \varepsilon}(\alpha) = \alpha + \varepsilon\beta(\alpha) + \varepsilon o_\varepsilon$$

(ii) $$\beta_{\theta_o + \varepsilon, \theta_o}(\alpha) = \alpha + \varepsilon\beta(1 - \alpha) + \varepsilon o_\varepsilon$$

(iii) $\beta_{\theta_o,\theta_o-\varepsilon}(\alpha) = \alpha + \varepsilon\beta(1-\alpha) + \varepsilon o_\varepsilon$

(iv) $\beta_{\theta_o-\varepsilon,\theta_o}(\alpha) = \alpha + \varepsilon\beta(\alpha) + \varepsilon o_\varepsilon$

(v) $\beta_{\theta_o-\varepsilon,\theta_o+\varepsilon}(\alpha) = \alpha + 2\varepsilon\beta(\alpha) + \varepsilon o_\varepsilon$

(vi) $\beta_{\theta_o+\varepsilon,\theta_o-\varepsilon}(\alpha) = \alpha + 2\varepsilon\beta(1-\alpha) + \varepsilon o_\varepsilon$

<u>Proof:</u> Write $P_\theta = P_{\theta_o} (\theta-\theta_o)\dot{P}_{\theta_o} + (\theta-\theta_o)\Gamma_{\theta_o,\theta}$.

(i) Let δ be any size α test. Then

$$P_{\theta_o+\varepsilon}(\delta) = \alpha + \varepsilon\dot{P}_{\theta_o}(\delta) + \varepsilon\Gamma_{\theta_o,\theta_o+\varepsilon}(\delta)$$

Hence $$|P_{\theta_o+\varepsilon}(\delta) - (\alpha + \varepsilon\dot{P}_{\theta_o}(\delta))| \leqq \varepsilon\|\Gamma_{\theta_o,\theta_o+\varepsilon}\|$$

so that $$|\beta_{\theta_o,\theta_o+\varepsilon}(\alpha) - (\alpha+\varepsilon\beta(\alpha))| \leqq \varepsilon\|\Gamma_{\theta_o,\theta_o+\varepsilon}\|$$

(ii) Write $\beta_{\theta_o+\varepsilon,\theta_o}(\alpha) = \alpha + \varepsilon\beta(1-\alpha) + \varepsilon v_\varepsilon(\alpha)$. We must show that $v_\varepsilon(\alpha) = o_\varepsilon$. Let δ be a size α test such that $P_{\theta_o}(\delta) = \beta_{\theta_o+\varepsilon,\theta_o}(\alpha)$. Then $\beta_{\theta_o+\varepsilon}(\alpha) - \alpha = (P_{\theta_o+\varepsilon}-P_{\theta_o})(\delta) = o_\varepsilon$.
Put $\alpha_\varepsilon = 1-\beta_{\theta_o,\theta_o+\varepsilon}(o)$ and let $\alpha \in [0,\alpha_\varepsilon]$. Then:
$1-\alpha = \beta_{\theta_o,\theta_o+\varepsilon}(1-\beta_{\theta_o+\varepsilon,\theta_o}(\alpha))$ = (by (i)) $1-\beta_{\theta_o+\varepsilon,\theta_o}(\alpha) +$
$\varepsilon\beta(1-\beta_{\theta_o+\varepsilon,\theta_o}(\alpha))+\varepsilon o_\varepsilon = 1-\alpha-\varepsilon\beta(1-\alpha)-\varepsilon v_\varepsilon(\alpha)+\varepsilon\beta(1-\alpha+o_\varepsilon)+\varepsilon o_\varepsilon$. Solving w.r.t.

v_ε we get: $v_\varepsilon(\alpha) = \beta(1-\alpha+o_\varepsilon)-\beta(1-\alpha)+o_\varepsilon = o_\varepsilon$.

It follows from (i) that:

$$\alpha_\varepsilon = 1-\beta_{\theta_o,\theta_o+\varepsilon}(o) = 1-\varepsilon o_\varepsilon. \quad \text{Let} \quad \alpha \in [\alpha_\varepsilon, 1].$$

Then: $$1 = \beta_{\theta_o+\varepsilon,\theta_o}(\alpha) = \alpha + \varepsilon\beta(1-\alpha) + \varepsilon v_\varepsilon(\alpha)$$

i.e.: $$v_\varepsilon(\alpha) = (1-\alpha)/\varepsilon + \beta(o) - \beta(1-\alpha)$$

Hence: $$|v_\varepsilon(\alpha)| \leq (1-\alpha_\varepsilon)/\varepsilon + \sup\{|\beta(\alpha_2) - \beta(\alpha_1)| : |\alpha_2-\alpha_1| \leq \varepsilon o_\varepsilon\} = o_\varepsilon$$

(iii): The proof is very similar to that for (i).

(iv): The proof is very similar to that for (ii).

(v) We have for any test δ:

$$P_{\theta_o+\varepsilon}(\delta) - 2P_{\theta_o}(\delta) + P_{\theta_o-\varepsilon}(\delta) = \varepsilon(\Gamma_{\theta_o,\theta_o+\varepsilon}(\delta) - \Gamma_{\theta_o,\theta_o-\varepsilon}(\delta)).$$

If δ is a size α test, then this may be written:

$$P_{\theta_o+\varepsilon}(\delta) = 2P_{\theta_o}(\delta) - \alpha + \varepsilon o_\varepsilon.$$

Hence - by (iv):

$$\beta_{\theta_o-\varepsilon,\theta_o+\varepsilon}(\alpha) = 2\beta_{\theta_o-\varepsilon,\theta_o}(\alpha) - \alpha + \varepsilon o_\varepsilon = \alpha + 2\varepsilon\beta(\alpha) + \varepsilon o_\varepsilon.$$

(vi) The proof is very similar to that for (v). □

Corollary 6.12.

Let $\mathcal{E} = ((X,\mathcal{A}),P_\theta : \theta \in \Theta))$ and $\tilde{\mathcal{E}} = ((\tilde{X},\tilde{\mathcal{A}}),\tilde{P}_\theta : \theta \in \Theta))$ (both be differentiable in θ_0. Denote by $\beta_{\theta_1,\theta_2}(\alpha)$ $(\tilde{\beta}_{\theta_1,\theta_2}(\alpha))$ the maximum power in $\mathcal{E}$ $(\tilde{\mathcal{E}})$ among all level α tests for testing "$\theta=\theta_1$" against "$\theta=\theta_2$". Then:

(i) $$\lim_{\varepsilon\to 0} \sup_\alpha (\tilde{\beta}_{\theta_0,\theta_0+\varepsilon}(\alpha) - \beta_{\theta_0,\theta_0+\varepsilon}(\alpha))^+/\varepsilon = \dot{\delta}_{\theta_0}(\mathcal{E},\tilde{\mathcal{E}})$$

(ii) $$\lim_{\varepsilon\to 0} \sup_\alpha (\tilde{\beta}_{\theta_0+\varepsilon,\theta_0}(\alpha) - \beta_{\theta_0+\varepsilon,\theta_0}(\alpha))^+/\varepsilon = \dot{\delta}_{\theta_0}(\mathcal{E},\tilde{\mathcal{E}})$$

(iii) $$\lim_{\varepsilon\to 0} \sup_\alpha (\tilde{\beta}_{\theta_0,\theta_0-\varepsilon}(\alpha) - \beta_{\theta_0,\theta_0-\varepsilon}(\alpha))^+/\varepsilon = \dot{\delta}_{\theta_0}(\mathcal{E},\tilde{\mathcal{E}})$$

(iv) $$\lim_{\varepsilon\to 0} \sup_\alpha (\tilde{\beta}_{\theta_0-\varepsilon,\theta_0}(\alpha) - \beta_{\theta_0-\varepsilon,\theta_0}(\alpha))^+/\varepsilon = \dot{\delta}_{\theta_0}(\mathcal{E},\tilde{\mathcal{E}})$$

(v) $$\lim_{\varepsilon\to 0} \sup_\alpha (\tilde{\beta}_{\theta_0-\varepsilon,\theta_0+\varepsilon}(\alpha) - \beta_{\theta_0-\varepsilon,\theta_0+\varepsilon}(\alpha))^+/2\varepsilon = \dot{\delta}_{\theta_0}(\mathcal{E},\tilde{\mathcal{E}})$$

(vi) $$\lim_{\varepsilon\to 0} \sup_\alpha (\tilde{\beta}_{\theta_0+\varepsilon,\theta_0-\varepsilon}(\alpha) - \beta_{\theta_0+\varepsilon,\theta_0-\varepsilon}(\alpha))^+/2\varepsilon = \dot{\delta}_{\theta_0}(\mathcal{E},\tilde{\mathcal{E}})$$

Remark: Let $\mathcal{D} = ((X,\mathcal{A}),P_1,P_2)$ and $\tilde{\mathcal{D}} = ((\tilde{X},\tilde{\mathcal{A}}),\tilde{P}_1,\tilde{P}_2)$ be two dichotomies. Let, for each $\alpha \in [0,1]$, $\gamma(\alpha)$ $(\tilde{\gamma}(\alpha))$ be the maximum power in $\mathcal{D}$ $(\tilde{\mathcal{D}})$ of any level α test for "P_1" against "P_2" ("$\tilde{P}_1$" against "$\tilde{P}_2$"). Then $2\sup_\alpha(\tilde{\gamma}(\alpha) - \gamma(\alpha))^+$ is the smallest number η such that $\mathcal{D}$ is $(0,\eta)$ deficient w.r.t. $\tilde{\mathcal{D}}$. This is a particular case of the "error of the first - and error of the second" criterion for comparison of dichotomies given in [15].

Corollary 6.13.

With the same notations as in corollary 6.12 we have:

(i) $$\lim_{\varepsilon\to 0} \sup_{\alpha} |\beta_{\theta_0,\theta_0+\varepsilon}(\alpha) - \tilde{\beta}_{\theta_0,\theta_0+\varepsilon}(\alpha)|/\varepsilon = \dot{\Delta}_{\theta_0}(\mathcal{E},\tilde{\mathcal{E}})$$

(ii) $$\lim_{\varepsilon\to 0} \sup_{\alpha} |\beta_{\theta_0+\varepsilon,\theta_0}(\alpha) - \tilde{\beta}_{\theta_0+\varepsilon,\theta_0}(\alpha)|/\varepsilon = \dot{\Delta}_{\theta_0}(\mathcal{E},\tilde{\mathcal{E}})$$

(iii) $$\lim_{\varepsilon\to 0} \sup_{\alpha} |\beta_{\theta_0,\theta_0-\varepsilon}(\alpha) - \tilde{\beta}_{\theta_0,\theta_0-\varepsilon}(\alpha)|/\varepsilon = \dot{\Delta}_{\theta_0}(\mathcal{E},\tilde{\mathcal{E}})$$

(iv) $$\lim_{\varepsilon\to 0} \sup_{\alpha} |\beta_{\theta_0-\varepsilon,\theta_0}(\alpha) - \tilde{\beta}_{\theta_0-\varepsilon,\theta_0}(\alpha)|/\varepsilon = \dot{\Delta}_{\theta_0}(\mathcal{E},\tilde{\mathcal{E}})$$

(v) $$\lim_{\varepsilon\to 0} \sup_{\alpha} |\beta_{\theta_0-\varepsilon,\theta_0+\varepsilon}(\alpha) - \tilde{\beta}_{\theta_0-\varepsilon,\theta_0+\varepsilon}(\alpha)|/2\varepsilon = \dot{\Delta}_{\theta_0}(\mathcal{E},\tilde{\mathcal{E}})$$

(vi) $$\lim_{\varepsilon\to 0} \sup_{\alpha} |\beta_{\theta_0+\varepsilon,\theta_0-\varepsilon}(\alpha) - \tilde{\beta}_{\theta_0+\varepsilon,\theta_0-\varepsilon}(\alpha)|/2\varepsilon = \dot{\Delta}_{\theta_0}(\mathcal{E},\tilde{\mathcal{E}})$$

Example 6.14. Suppose P_θ does not depend on θ, i.e. $\mathcal{E}$ is a minimum information experiment. Then $\beta_{\theta_0,\theta_0+\varepsilon}(\alpha) = \alpha$, $\alpha \in [0,1]$ so that:

$$\dot{\delta}_{\theta_0}(\mathcal{E},\tilde{\mathcal{E}}) = \lim_{\varepsilon\to 0} \sup_{\alpha} (\tilde{\beta}_{\theta_0,\theta_0+\varepsilon}(\alpha)-\alpha)/\varepsilon = \lim_{\varepsilon\to 0} \|\tilde{P}_{\theta_0+\varepsilon} - \tilde{P}_{\theta_0}\|/2\varepsilon = \|\dot{\tilde{P}}_{\theta_0}\|/2$$

It follows that $\|\dot{\tilde{P}}_{\theta_0}\|$ measures how far our experiment is away - in the $\dot{\delta}_{\theta_0}$ sense - from the "no information" experiment. This follows also directly from corollary 4.7.

It is not surprising that conditional expectations under θ is - when θ is small - close to the corresponding conditional expectations under θ_o. We shall need the following result in this direction.

__Proposition 6.15.__

Let $\mathcal{E} = ((X,\mathcal{A}), \; P_\theta \colon \theta \in \Theta)$ be differentiable in θ_o and let $\mathcal{B}$ be a sub σ algebra of $\mathcal{A}$. Let X be a bounded random variable and choose a bounded version $E_{\theta_o}^{\mathcal{B}} X$. Then:

$$\sup_\theta E_\theta |(E_\theta^{\mathcal{B}} X - E_{\theta_o}^{\mathcal{B}} X)/(\theta-\theta_o)| < \infty.$$

__Proof:__ Let $-1 \leqq h \leqq 1$ be a measurable function. Then

$$E_\theta h(E_\theta^{\mathcal{B}} X - E_{\theta_o}^{\mathcal{B}} X) = E_\theta(h\, E_\theta^{\mathcal{B}} X) - E_\theta(h\, E_{\theta_o}^{\mathcal{B}} X)$$

$$= E_\theta(E_\theta^{\mathcal{B}} hX) - E_\theta(h\, E_{\theta_o}^{\mathcal{B}} X) = E_\theta hX - E_\theta h E_{\theta_o}^{\mathcal{B}} X$$

$$= E_\theta(hX - hE_{\theta_o}^{\mathcal{B}} X) = E_{\theta_o}(hX - hE_{\theta_o}^{\mathcal{B}} X) +$$

$$+ \int (hX - hE_{\theta_o}^{\mathcal{B}} X)d(P_\theta - P_{\theta_o}) = E_{\theta_o} hX - E_{\theta_o} E_{\theta_o}^{\mathcal{B}} hX +$$

$$+ \int (hX - hE_{\theta_o}^{\mathcal{B}} X)d(P_\theta - P_{\theta_o}) = \int h(X - E_{\theta_o}^{\mathcal{B}} X)d(P_\theta - P_{\theta_o})$$

$$\leqq \sup_x \left| X(x) - [E_{\theta_o} X]_x \right| \, \|(P_\theta - P_{\theta_o})/(\theta-\theta_o)\| \; |\theta-\theta_o| .$$

Hence

$$E_\theta |E_\theta^{\mathcal{B}} X - E_{\theta_o}^{\mathcal{B}} X| \Big/ |\theta-\theta_o| \leqq \sup_x |X(x) - [E_{\theta_o}^{\mathcal{B}} X]_x|$$

$$\times \sup_\theta \|P_\theta - P_{\theta_o}\| \Big/ |\theta-\theta_o| < \infty .$$

□

Le Cam has shown ([7]) that - under regularity conditions - sufficiency for his distance Δ is equivalent with "conditional expectation" sufficiency. The next two propositions treats this problem for the $\dot{\Delta}_{\theta_0}$ distance. To simplify the writing we introduce the unpronounciable notion of $\dot{\Delta}_{\theta_0}$ sufficiency. Let $\mathcal{E} = ((X,\mathcal{A}), P_\theta : \theta \in \Theta)$ be differentiable in θ_0, let $\mathcal{B}$ be a sub σ algebra of $\mathcal{A}$, and let $\mathcal{E}_{\mathcal{B}} = ((X,\mathcal{B}), P_{\theta\mathcal{B}} : \theta \in \Theta)$ where $P_{\theta\mathcal{B}}$ - for each θ - is the restriction of P_θ to $\mathcal{B}$. Proposition 2.2 implies that $\mathcal{E}_{\mathcal{B}}$ is differentiable in θ_0. We will write that $\mathcal{B}$ is $\dot{\Delta}_{\theta_0}$ sufficient if and only if $\dot{\Delta}_{\theta_0}(\mathcal{E}_{\mathcal{B}},\mathcal{E}) = 0$.

<u>Proposition 6.16.</u>

Let $\mathcal{E} = ((X,\mathcal{A}), P_\theta : \theta \in \Theta)$ be differentiable in θ_0 and let $\mathcal{B}$ be a $\dot{\Delta}_{\theta_0}$ sufficient sub σ algebra of $\mathcal{A}$. Let X be a bounded random variable in $\mathcal{E}$ and choose a bounded version $E_{\theta_0}^{\mathcal{B}} X$. Then:

$$\lim_{\theta\to\theta_0} E_\theta |E_\theta^{\mathcal{B}} X - E_{\theta_0}^{\mathcal{B}} X| \Big/ |\theta-\theta_0| = 0$$

<u>Proof:</u> Let $-1 \leq h \leq 1$ be $\mathcal{B}$ measurable. It follows from the proof of proposition 6.15 that :

$$E_\theta h(E_\theta^{\mathcal{B}} X - E_{\theta_o}^{\mathcal{B}} X)/(\theta-\theta_o) = \int h(X - E_{\theta_o}^{\mathcal{B}} X)d(P_\theta - P_{\theta_o})/(\theta-\theta_o)$$

$$= \int h(X-E_{\theta_o}^{\mathcal{B}} X)d\dot{P}_{\theta_o} + \int h(X-E_{\theta_o}^{\mathcal{B}} X)d\left[(P_\theta - P_{\theta_o})/(\theta-\theta_o) - \dot{P}_{\theta_o}\right]$$

$$= E_{\theta_o} h(X-E_{\theta_o}^{\mathcal{B}} X)s_{\theta_o} + \int h(X-E_{\theta_o}^{\mathcal{B}} X)d\left[(P_\theta - P_{\theta_o})/(\theta-\theta_o) - \dot{P}_{\theta_o}\right]$$

$$= \text{(by proposition 4.12)} \int h(X-E_{\theta_o}^{\mathcal{B}} X)d\left[(P_\theta - P_{\theta_o})/(\theta-\theta_o) - \dot{P}_{\theta_o}\right]$$

$$\leq \sup_x \left|X(x) - \left[E_{\theta_o}^{\mathcal{B}} X\right]_x\right| \, \|(P_\theta - P_{\theta_o})/(\theta-\theta_o) - \dot{P}_{\theta_o}\| .$$

Hence

$$E_\theta|(E_\theta^{\mathcal{B}} X - E_{\theta_o}^{\mathcal{B}} X)/(\theta-\theta_o)| \leq \sup_x \left|X(x) - \left[E_{\theta_o}^{\mathcal{B}} X\right]_x\right|$$

$$\times \|(P_\theta - P_{\theta_o})/(\theta-\theta_o) - \dot{P}_{\theta_o}\| \to 0 \text{ as } \theta \to \theta_o .$$ □

Proposition 6.16 tells us that conditional expectations given a $\dot{\Delta}_{\theta_o}$ sufficient sub σ algebra does not depend too much on θ when θ is small. We will now - using proposition 6.6 - give a converse of this result.

<u>Proposition 6.17.</u>

Let $\mathcal{E} = ((X,\mathcal{A}), P_\theta: \theta \in \Theta)$ be differentiable in θ_o and let $\mathcal{B}$ be a sub σ algebra of $\mathcal{A}$. Let $\mathcal{A}_o$ be a π-system (i.e. $A_1 \cap A_2 \in \mathcal{A}_o$ when $A_1, A_2 \in \mathcal{A}_o$) generating $\mathcal{A}$.

Then $\mathcal{B}$ is $\dot{\Delta}_{\theta_o}$ sufficient provided there corresponds to any $A \in \mathcal{A}_0$ a measurable Y_A so that

$$\lim_{\theta\to\theta_0} E_\theta |P_\theta^{\mathfrak{B}}(A) - Y_A| \Big/ |\theta-\theta_0| = 0 .$$

<u>Proof:</u> Let $A \in \mathcal{A}_0$. We may assume that $0 \leq Y_A \leq 1$. By assumption: $\lim_{\theta\to\theta_0} E_\theta|P_\theta^{\mathfrak{B}}(A) - Y_A| = 0$, and by proposition 6.15: $\lim_{\theta\to\theta_0} E_\theta|P_\theta^{\mathfrak{B}}(A) - P_{\theta_0}^{\mathfrak{B}}(A)| = 0$ where $P_{\theta_0}^{\mathfrak{B}}$ is specified such that $0 \leq P_{\theta_0}^{\mathfrak{B}}(A) \leq 1$. It follows that:

$$E_{\theta_0}|P_{\theta_0}^{\mathfrak{B}}(A) - Y_A| = \lim_{\theta\to\theta_0} E_\theta|P_{\theta_0}^{\mathfrak{B}}(A) - Y_A| = 0$$

so that

$$P_{\theta_0}^{\mathfrak{B}}(A) = Y_A \text{ a.s. } \quad P_{\theta_0}$$

Using the notation $\Gamma_{\theta_0,\theta} = (P_\theta - P_{\theta_0})/(\theta-\theta_0) - \dot{P}_{\theta_0}$ we get

$$0 = \lim_{\theta\to\theta_0} \int (P_\theta^{\mathfrak{B}}(A) - Y_A)/(\theta-\theta_0)\,dP_\theta = \lim_{\theta\to\theta_0} \int (I_A - Y_A)/(\theta-\theta_0)\,dP_\theta$$

$$= \lim_{\theta\to\theta_0} \Big[\int I_A dP_{\theta_0} + (\theta-\theta_0)\int I_A d\dot{P}_{\theta_0} + (\theta-\theta_0)\int I_A d\Gamma_{\theta_0,\theta} - \int Y_A dP_{\theta_0} - (\theta-\theta_0)\int Y_A d\dot{P}_{\theta_0}$$

$$- (\theta-\theta_0)\int Y_A d\Gamma_{\theta_0,\theta}\Big] \Big/ (\theta-\theta_0)$$

$$= \lim_{\theta\to\theta_0} \Big[\int I_A dP_{\theta_0} - \int P_{\theta_0}^{\mathfrak{B}}(A) dP_{\theta_0} + (\theta-\theta_0)\int I_A d\dot{P}_{\theta_0} - (\theta-\theta_0)\int P_{\theta_0}^{\mathfrak{B}}(A) d\dot{P}_{\theta_0}$$

$$+ (\theta-\theta_0)\int I_A d\Gamma_{\theta_0,\theta} - (\theta-\theta_0)\int Y_A d\Gamma_{\theta_0,\theta}\Big] \Big/ (\theta-\theta_0)$$

$$= \lim_{\theta\to\theta_0} \Big[\Big(\int I_A d\dot{P}_{\theta_0} - \int P_{\theta_0}^{\mathfrak{B}}(A) d\dot{P}_{\theta_0}\Big) + \int (I_A - Y_A) d\Gamma_{\theta_0,\theta}\Big]$$

$$= \left(\text{since } \lim_{\theta\to\theta_0} \|\Gamma_{\theta_0,\theta}\| = 0\right) \quad \Big[\int I_A d\dot{P}_{\theta_0} - \int P_{\theta_0}^{\mathfrak{B}}(A) d\dot{P}_{\theta_0}\Big]$$

Hence

$$\int I_A d\dot{P}_{\theta_0} = \int P_{\theta_0}^{\mathcal{B}}(A) d\dot{P}_{\theta_0} ; \qquad A \in \mathcal{A}_0, \quad B \in \mathcal{B}$$

so that

$$\int \delta d\dot{P}_{\theta_0} = \int (E_{\theta_0}^{\mathcal{B}} \delta) d\dot{P}_{\theta_0}$$

for any test function δ. It follows that the maximal slope at θ_0 for tests of size α for "$\theta = \theta_0$" against "$\theta > \theta_0$" is attained by $\mathcal{B}$ measurable tests. $\dot{\Delta}_{\theta_0}$ sufficiency follows now from proposition 6.7. □

Let $\mathcal{B}$ be $\dot{\Delta}_{\theta_0}$ sufficient for $\mathcal{A}$, and let δ be any test for - say - "$\theta = \theta_0$" against "$\theta > \theta_0$". Then any version of $E_{\theta_0}^{\mathcal{B}} \delta$ which is a test function, is "differentiably" as good as δ in infinitesimal neighbourhoods of θ_0, i.e. it has the same size and the same slope as δ.

If $\mathcal{B}$ is any sub σ algebra of then - by proposition 4.12 - $\mathcal{B}$ is $\dot{\Delta}_{\theta_0}$ sufficient if and only if $d\dot{P}_{\theta_0} | dP_{\theta_0}$ is almost (P_{θ_0}) $\mathcal{B}$ measurable. It follows that any sub σ algebra $\mathcal{B}$ of $\mathcal{A}$ induced by a version of $d\dot{P}_{\theta_0} | dP_{\theta_0}$ is minimal $\dot{\Delta}_{\theta_0}$ sufficient.

Example 6.18.

Suppose $X_1, \cdots, X_n$ are independent identically distributed random variables, each having the density $f(x-\theta)$; $x \in]-\infty,\infty[$ with respect to Lebesgue measure. We will assume that f is absolutely continuous on finite intervals and that $\int |(f')(x)| dx < \infty$. By the example in section 2, this experiment is differentiable in θ_0 for any θ_0.

It was shown in [13] that the order statistic is minimal sufficient when f is meromorfic and the set of zeros (or the set of poles) satisfies a mild boundedness condition. Locally, however, (i.e. in the $\dot{\Delta}_{\theta_0}$ sense) considerable compression may be obtained since $\sum_{i=1}^{n} f'(X_i-\theta_0)/f(X_i-\theta_0)$ is $\dot{\Delta}_{\theta_0}$ minimal sufficient.

Finally some remarks on the effect of a change of parameter, and in particular of scale change. Let P be a probability distribution on R^n. A localization model $\{Q_{\theta,\sigma}: \theta \in]-\infty,\infty[, \sigma > 0\}$ may be defined by putting $Q_{\theta,\sigma}(B) = P((B-(\theta,\cdots,\theta))/\sigma)$. Suppose σ is known. Then our experiment $\{Q_{\theta,\sigma}: \theta \in]-\infty,\infty[\}$ is equivalent with the experiment $P_{\theta/\sigma}: \theta \in]-\infty,\infty[$. It follows that the scale change may be carried out in the parameter space. The local comparison of experiments $\{Q_{\theta,\sigma}: \theta \in]-\infty,\infty[\}$ for different values of σ may therefore be based on the following result.

Proposition 6.19.

Let $\mathcal{E} = (\chi, \mathcal{A}, P_\theta: \theta \in \Theta)$ be differentiable in 0 and write $\mathcal{E}^{(\sigma)} = (\chi, \mathcal{A}, P_{\theta/\sigma}: \theta \quad \sigma\Theta)$. Then $\dot{\mathcal{E}}_0^{(\sigma)} = ((\chi, \mathcal{A})(P_0, \dot{P}_0/\sigma))$

so that

$$\beta^{(\sigma)} = \beta/\sigma \quad \text{when} \quad \sigma > 0$$

and

$$\beta^{(\sigma)}(p) = -\beta(1-p)/\sigma \quad \text{when} \quad \sigma < 0$$

Proof: This is a particular case of the next proposition.

□

Proposition 6.20.

Let $\mathcal{E} = ((\chi, \quad), P_\theta: \theta \in \Theta)$ be differentiable in θ_o and let γ be a function from a subset $\mathcal{M}$ of $]-\infty,\infty[$ to Θ. Suppose γ is differentiable in $\eta_o \in \mathcal{M}$ and that $\gamma(\eta_o) = \theta_o$. Then $\mathcal{E} = ((\chi,\mathcal{A}), P_{\gamma(\eta)}: \eta \in \mathcal{M})$ is differentiable in η_o and

$$\dot{\mathcal{E}}_{\eta_o} = ((\chi,\mathcal{A}), P_{\theta_o}, \quad \gamma'(\eta_o)\dot{P}_{\theta_o})$$

so that

$$\hat{\beta}_{\eta_o} = \gamma'(\eta_o)\beta_{\theta_o} \quad \text{when} \quad \gamma'(\eta_o) \geq 0$$

and

$$\hat{\beta}_{\eta_o}(p) = -\gamma'(\eta_o)\beta_{\theta_o}(1-p);\ p \quad [0,1] \quad \text{when} \quad \gamma'(\eta_o) \leq 0$$

Proof: $$\lim_{\eta\to\eta_o} |P_{\gamma(\eta)} - P_{\gamma(\eta_o)}| \big/ (\eta-\eta_o) = \gamma'(\eta_o)\mathring{P}_{\theta_o}$$

and $P_{\gamma(\eta_o)} = P_{\theta_o}$

□

7. Local comparison of translation experiments.

Let G be a probability measure on $]-\infty,\infty[$. For any $\theta \in]-\infty,\infty[$ the θ translate G_θ of G is the distribution of $X+\theta$ when X has the distribution G. The experiment defined by $G_\theta : \theta \in \Theta =]-\infty,\infty[$ will be denoted by $\mathcal{E}_G$. Experiments of the form $\mathcal{E}_G$ will be called translation experiments. Comparison of these experiments have been treated by Boll [3], LeCam [9], Heyer [5], the author [16] and others. Some relevant results in [16] are given in appendix A.

We will in this section study $\dot{\delta}$ (and $\dot{\Delta}$) comparison of differentiable translation experiments and our first task is to describe the probability measures G for which $\mathcal{E}_G$ is differentiable. It is not necessary to specify the points θ_0 at which $\mathcal{E}_G$ is differentiable since we have the following easily proved result.

Theorem 7.1.

$\mathcal{E}_G$ is differentiable in all points θ if and only if $\mathcal{E}_G$ is differentiable at some point θ.

Proof: Straight forward. □

Henceforth we will write "differentiable" instead of "differentiable in θ_0". The differentiable translation experiments are described in:

Theorem 7.2.

$\mathscr{E}_G$ is differentiable if and only if G has a absolutely continuous density* g such that

$$\int_{-\infty}^{\infty} |g'(x)|dx < \infty .$$

Remark 1.

The almost everywhere existence of the derivative $g'(x)$ is implied by the absolute continuity of g.

Remark 2.

A continuous density g is necessarily unique. If $\mathscr{E}_G$ is differentiable then g will - unless otherwise stated - denote the (absolute) continuous density of G.

Proof of the theorem: The "if" part was (essentially taken from Hajek and Sidak [4]) treated in the example in section 2. Suppose now that $\mathscr{E}_G$ is differentiable and put $\dot{G} = \lim_{\theta \to 0} (G_\theta - G)/\theta$. The existence of this limit imply the continuity of the map $\theta \leadsto G_\theta$. If follows that $\mathscr{E}_G$ is dominated and it is known (A proof is given in [16]) that this occur if and only if G is absolutely continuous. Hence $\dot{G}$ is absolutely continuous.

*"Absolute continuity" and "density" are - if not otherwise stated - always w.r.t. Lebesgue measure.

For any x we get*:

$$\lim_{\theta\to o} [G(]-\infty, x-\theta[) - G(]-\infty, x[)]/\theta$$

$$= \lim_{\theta\to o} [G_\theta(]-\infty, x[) - G_o(]-\infty, x[)]/\theta = \dot{G}(]-\infty, x[) = \int_{-\infty}^{x} (d\dot{G}/d\mu)d\mu.$$

It follows that $g: x \rightsquigarrow -\dot{G}]-\infty, x[$ is a density for G having the required properties. □

In the following $\mathcal{G}$ will denote the set of all probability measures G such that $\mathcal{E}_G$ is differentiable. The continuous density of $G \in \mathcal{G}$ will be denoted by g. If affixes are used on G then corresponding affixes will be used on g. For any probability distribution H on $]-\infty, \infty[$ and each $p \in [0,1]$ we put $H^{-1}(p) = \inf \{x: H\,]-\infty,x[\ \geqq p\}$ and $H_*^{-1}(p) = \inf \{x: H]-\infty,x[\ >p\}$. Then $[H^{-1}(p), H_*^{-1}(p)]$ consists precisely of the p fractiles of H i.e. the elements $x \in [-\infty,\infty]$ such that $H\,]-\infty,x[\ \leqq p \leqq H\,]-\infty,x]$. In particular $H^{-1}(0) = -\infty$ and $H_*^{-1}(1) = \infty$.

To each $G \in \mathcal{G}$ we will associate the function γ_G from $[0,1]$ defined by:

$$\gamma_G(p) = g(G^{-1}(p)) \ ; \ p \in [0,1].$$

The functions $\gamma_G : G \in \mathcal{G}$ will play an important part in our investigations. We will first - and almost without statistical motivation - study some properties of these functions.

---

* μ wil in this section be reserved for Lebesgue measure. The restriction of μ to $[0,1]$ will be denoted by λ.

Note first that $g(x) = g(G^{-1}(p))$ for any p fractile x. Further properties of these functions are listed in:

Proposition 7.3

(i) $\gamma_G \geqq 0$,

$$\int_{G(x_1)}^{G(x_2)} dp/\gamma_G(p) = \mu([x_1,x_2]\cap[g>0]) \text{ when } x_1 \leqq x_2$$

and $$\int_{p_1}^{p_2} dp/\gamma_G(p) = \mu([G^{-1}(p_1),\ G^{-1}(p_2)] \cap [g > 0]) \text{ when}$$

$0 < p_1 \leqq p_2 < 1$

(ii) γ_G is absolutely continuous and

$\gamma_G'(p) = g'(G^{-1}(p))/g(G^{-1}(p))$ a.e Lebesgue.

(iii) $\gamma_G(0) = \gamma_G(1) = 0$.

Remark. By (i) $\gamma_G > 0$ a.e Lebesgue and $\int_{\epsilon}^{1-\epsilon} dp/\gamma_G(p) < \infty$ when $0 < \epsilon < \frac{1}{2}$.

Proof: (i) Let $x_1 \leqq x_2$. Then the sets $\{p: G(x_1) \leqq p \leqq G(x_2)\}$ and $\{p: x_1 \leqq G^{-1}(p) \leqq x_2\}$ are μ equivalent. Hence:

$$\int_{G(x_1)}^{G(x_1)} dp/\gamma_G(p) = \int [I_{[x_1,x_2]}(G^{-1}(p))/g(G^{-1}(p))]dp$$

$$= \int_{x_1}^{x_2} [1/g(x)]G(dx) = \mu([x_1,x_2] \cap [g > 0])$$

The last formula in (i) follows by substituting $x_1 = G^{-1}(p_1)$ and $x_2 = G^{-1}(p_2)$.

(ii) $\int |g'(G^{-1}(p))|/g(G^{-1}(p))dp = \int [|g'(x)|/g(x)]G(dx)$

$= \int |g'(x)|dx < \infty$

Absolute continuty follows now from:

$$\int_0^t [g'(G^{-1}(p))/g(G^{-1}(p))]dp = \gamma_G(t) \; ; \; t \in [0,1].$$

(iii) $\gamma_G(0) = g(G^{-1}(0)) = g(-\infty) = 0$

$\gamma_G(1) = g(G_*^{-1}(1)) = g(\infty) = 0.$ □

<u>Proposition 7.4.</u>

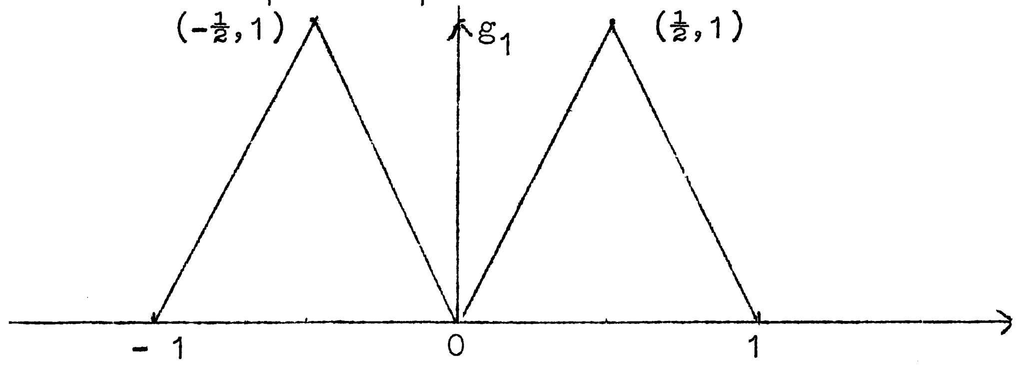
If $G_1 \in \mathcal{G}$ and G_2 is a translate of G_1, then $G_2 \in \mathcal{G}$ and $\gamma_{G_1} = \gamma_{G_2}$

<u>Proof</u>: Straight forward □

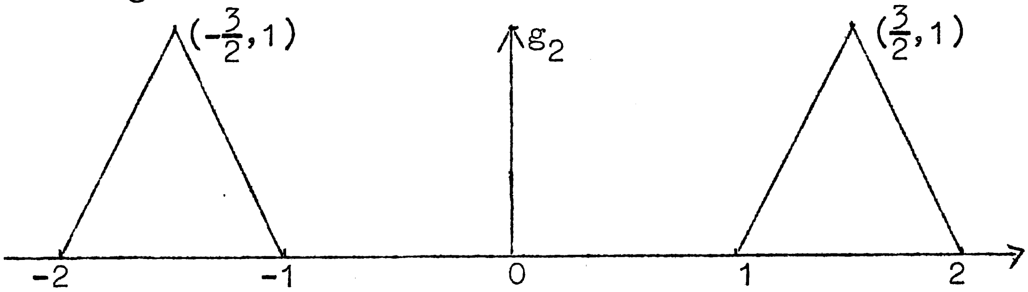
The last proposition tells us that the map $G \rightsquigarrow \gamma_G$ does not distinguish between translates of the same distribution. Here is an example of two distributions in $\mathcal{G}$ having identical γ functions, which are not translates of each other.

<u>Example 7.5.</u>

The density g_1 of G_1 is given by the triangles:



The density g_2 of G_2 is obtained from g_1 by splitting the triangles as follows:



It is easily checked that

$$\gamma_{G_1}(p) = \gamma_{G_2}(p) = \begin{cases} 2\sqrt{p} & \text{when } p \in [0,\frac{1}{4}] \\ 2\sqrt{\frac{1}{2}-p} & \text{" } p \in [\frac{1}{4},\frac{1}{2}] \\ 2\sqrt{p-\frac{1}{2}} & \text{" } p \in [\frac{1}{2},\frac{3}{4}] \\ 2\sqrt{1-p} & \text{" } p \in [\frac{3}{4},1] \end{cases}$$

(By symmetry $\gamma_{G_i}(p) = \gamma_{G_i}(1-p)$. It suffices therefore to consider $p \in [0,\frac{1}{2}]$).

G_2, however, is obviously not a translate of G_1.

A miniresult on the uniqueness problem is:

<u>Proposition 7.6.</u>

Let $G_1, G_2 \in \mathcal{G}$ and suppose $\gamma_{G_1} = \gamma_{G_2}$. Then

$$\{G_1(x) : g_1(x) > 0\} = \{G_2(x) : g_2(x) > 0\}$$

<u>Proof:</u> Let $g_1(x) > 0$ and put $p = G_1(x)$. Then $p \in]0,1[$ and $x = G_1^{-1}(p)$. Put $\tilde{x} = G_2^{-1}(p)$. Then $g_2(\tilde{x}) = \gamma_{G_2}(p) = \gamma_{G_1}(p)$ $= g_1(x) > 0$ and $G_2(\tilde{x}) = p = G_1(x)$. This proves $\subseteq$ and $\supseteq$ follows by symmetry. □

The next proposition is an imediate consequence of proposition 7.3 part (i).

Proposition 7.7.

Let $G_1, G_2 \in \mathcal{G}$ and suppose $\gamma_{G_1} = \gamma_{G_2}$. Let $I \subseteq]-\infty,\infty[$ be an interval such that

(i) $[g_1 > 0] \cap I = [g_2 > 0] \cap I$ a.e Lebesgue.

(ii) There is a $x_0 \in I$ so that $G_1(x_0) = G_2(x_0)$

Then $G_1(x) = G_2(x)$ when $x \in I$.

Proof: By proposition 7.3 we have for any $x \in I$:

$$\int_{G_1(x_0)}^{G_1(x)} dp/g_1(G_1^{-1}(p)) = \operatorname{sgn}(x-x_0)\mu([x_0,x] \cap [g_1 > 0])$$

$$= \operatorname{sgn}(x-x_0)\mu([x_0,x] \cap [g_2 > 0]) = \int_{G_2(x_0)}^{G_2(x)} dp/g_2(G_2^{-1}(p))$$

Hence $G_1(x) = G_2(x)$ when $x \in I$

□

Corollary 7.8.

Let $G_1, G_2 \in \mathcal{G}$ and $a \in]-\infty,\infty[$. Then $G_1(x) = G_2(x-a)$; $x \in]-\infty,\infty[$ if and only if

(i) $\gamma_{G_1} = \gamma_{G_2}$

(ii) $\{x: g_1(x) > 0\} = \{x: g_2(x-a) > 0\}$

(iii) There is a $x_0 \in]-\infty,\infty[$ so that $G_1(x_0) = G_2(x_0-a)$.

Proof: 1° "Only if". Suppose $G_1(x) = G_2(x-a)$; $x \in]-\infty,\infty[$. Then (ii) and (iii) follows imediately, and (i) follows from proposition 7.4.

2° "if". Suppose (i), (ii), (iii) hold. Then (i) and (ii) in proposition 7.7 hold with G_2 replaced by* $G_2 * \delta_a$ and $I =]-\infty,\infty[$ □

Proposition 7.9.

Let $G \in \mathcal{G}$. Then a subset V of $]-\infty,\infty[$ is a topological component of $[g > 0]$ if and only if V is of the form

$$V =]G_*^{-1}(p),\ G^{-1}(q)[$$

where $0 \leqq p < q \leqq 1$, $\gamma_G(p) = \gamma_G(q) = 0$ and $\gamma_G(r) > 0$ when $r \in]p,q[$. The numbers p and q are determined by V.

Proof: Straight forward. □

We are now ready to give a complete answer to the uniqueness problem.

Theorem 7.10.

Let $G_1, G_2 \in \mathcal{G}$ and let $\mathcal{C}_{G_1}$ and $\mathcal{C}_{G_2}$ denote - respectively - the class of topological components of $[g_1 > 0]$ and the class of topological components of $[g_2 > 0]$. Then $\gamma_{G_1} = \gamma_{G_2}$ if and only if there is a correspondence (1-1 and onto), <->, between $\mathcal{C}_{G_1}$ and $\mathcal{C}_{G_2}$ so that:

*) δ_a is the one point distribution in a.

(i) If $V_1, V_2 \in \mathscr{C}_{G_1}$, $W_1, W_2 \in \mathscr{C}_{G_2}$,

$V_1 <-> W_1$ and $V_2 <-> W_2$ then *

$W_1 < W_2$ provided $V_1 < V_2$

(ii) There is a map $V \rightsquigarrow t_V$ from $\mathscr{C}_{G_1}$ to

$]-\infty,\infty[$ so that $V <-> W$ imply

$W = V + t_V$ and $g_2(y) = g_1(y - t_V);\ y \in W$.

If conditions (i) and (ii) are satisfied then $G_2(y) = G_1(y-t_V)$ when $y \in W <-> V$. In particular $G_2(W) = G_1(V)$ when $W <-> V$.

Remark:

Condition (i) is simply the condition that $<->$ is order preserving, and the content of (ii) is that the restriction of G_2 to W is a translate of the restriction of G_1 to V when $V <-> W$. It follows from part 1° of the proof that $W =]G_{2*}^{-1}(p), G_2^{-1}(q)[$ if $]G_{1*}^{-1}(p), G_1^{-1}(q)[= V <-> W$ and the conditions are satisfied.

Prof 1°. Suppose $g_1(G_1^{-1}(p)) = g_2(G_2^{-1}(p));\ p \in]0,1[$.
Let $V =]G_{1*}^{-1}(p), G_1^{-1}(q)[\in \mathscr{C}_{G_1}$ and put $W =]G_{2*}^{-1}(p), G_2^{-1}(q)[$.
Then $g_2(G_{2*}^{-1}(p)) = g_1(G_{1*}^{-1}(p)) = 0 = g_1(G_1^{-1}(q)) = g_2(G_2^{-1}(q))$ and $g_2(G_2^{-1}(r)) = g_1(G_1^{-1}(r)) > 0$ when $p < r < q$. It follows from proposition 7.9 that $W \in \mathscr{C}_{G_2}$. It is easily seen that we have

---

*) If A and B are sub sets of $]-\infty,\infty[$ then "$A < B$" means that $a < b$ when $a \in A$ and $b \in B$.

established a correspondence, <->, between $\mathcal{C}_{G_1}$ and $\mathcal{C}_{G_2}$ which is 1-1, onto, and order preserving. Furthermore $G_1(V)=G_2(W)=q-p$ when $V =]G_{1*}^{-1}(p), F_1^{-1}(q)[$ and $W <-> V$. Let

$\mathcal{C}_{G_1} \ni V =]G_{1*}^{-1}(p), G_1^{-1}(q)[<->]G_{2*}^{-1}(p), G_2^{-1}(q)[= W \in \mathcal{C}_{G_2}$.

Choose a point $x_o \in V$. Then - since G_2 is continuous and strictly increasing on W - there is a y_o in $]G_{2*}^{-1}(p), G_2^{-1}(q)[$. Put $t = y_o - x_o$. The "only if" will be proved if we can show that:

$$G_{2*}^{-1}(p) = G_{1*}^{-1}(p)+t$$

$$G_2^{-1}(q) = G_1^{-1}(p)+t$$

and that $G_1(y-t) = G_2(y)$ when $y \in W$.

Let $y \in [y_o, \min\{G_2^{-1}(q), G_1^{-1}(q)+t\}]$. Then - by proposition 7.3 :

$$\int_{G_2(y_o)}^{G_2(y)} ds/g_2(G_2^{-1}(s)) = y-y_o$$

and $$\int_{G_2(y_o)}^{G_1(y-t)} ds/g_2(G_2^{-1}(s)) = \int_{G_1(x_o)}^{G_1(y-t)} ds/g_1(G_1^{-1}(s)) = y-t-x_o = y-y_o .$$

Hence $G_2(y) = G_1(y-t)$ when $y_o \leqq y \leqq \min\{G_2^{-1}(q), G_1^{-1}(q)+t\}$.

Suppose $G_2^{-1}(q) < G_1^{-1}(q)+t$. Then:

$$q-G_2(y_o)=G_2([y_o,G_2^{-1}(q)])=G_1([x_o,G_2^{-1}(q)+t])<G_1([x_o,G_1^{-1}(q)])=q-G_1(x_o)$$

Similarily $G_2^{-1}(q) > G_1^{-1}(q)+t$ imply:

$$q-G_2(y_o)=G_2([y_o,G_2^{-1}(q)])>G_2([x_o+t,G_1^{-1}(q)+t])=G_1([x_o,G_1^{-1}(q)])=q-G_1(x_o)$$

It follows that $G_2^{-1}(q) = G_1^{-1}(q)+t$ and that $G_2(y) = G_1(y-t)$ when $y \in [y_o,G_2^{-1}(q)]$. In the same way we may show that $G_{2*}^{-1}(p) = G_{1*}^{-1}(p)+t$ and that $G_2(y) = G_1(y-t)$ when $y \in [G_{2*}^{-1}(p),y_o]$.

2^o Suppose (i) and (ii) hold. Let $V \in \mathscr{C}_{G_1}$ and $W \in \mathscr{C}_{G_2}$ be such that $V \leftrightarrow W$. Then $G_2(W) = \int_W g_2(y)dy = \int_{V+t_V} g_1(y-t_V)dy =$ $= \int_V g_1(x)dx = G_1(V)$. Let $\mathscr{C}_{G_1} \ni V =]G_{1*}^{-1}(p),\ G_1^{-1}(q)[\leftrightarrow W$ and let $y \in W$. Then

$$\begin{aligned} G_2(y) &= \Sigma\{G_2(W') : W' \in \mathscr{C}_{G_2} \text{ and } W' < W\} + G_2(W \cap\ ]-\infty,y[) \\ &= \Sigma\{G_1(V') : V' \in \mathscr{C}_{G_1} \text{ and } V' < V\} + \int_{W \cap]-\infty,y[} g_1(z-t_V)dz \\ &= \Sigma\{G_1(V') : V' \in \mathscr{C}_{G_1} \text{ and } V' < V\} + \int_{V \cap]-\infty,y-t_V[} g_1(x)dx \\ &= G_1(y-t_V) \end{aligned}$$

We have so far proved the last two statements. Let $r \in\ ]0,1[$ be such that $g_1(G_1^{-1}(r)) > 0$. [This is true for almost (Lebesgue) all $r \in\ ]0,1[$.] Then $G_1^{-1}(r) \in V$ for some $V =]G_{1*}^{-1}(p),\ G_1^{-1}(q)[\ \in \mathscr{C}_{G_1}$. Put $W = V+t_V$. Then $G_1^{-1}(r)+t_V \in W$ and $G_2(G_1^{-1}(r)+t_V) = G_1(G_1^{-1}(r)) = r$ so that $G_1^{-1}(r)+t_V = G_2^{-1}(r)$ and $g_2(G_2^{-1}(r)) = g_2(G_1^{-1}(r)+t_V) = g_1(G_1^{-1}(r))$. □

Which functions are of the form γ_G with $G \in \mathscr{G}$? The next theorem provides the answer to that question. The construction in the "if" part of the proof is essentially that in the proof of lemma f in I 2.4 in Hájek and Šidák [4].

Theorem 7.11.

Let γ be a function from $[0,1]$ to $[0,\infty[$. Then there is a $G \in \mathcal{G}$ so that $\gamma = \gamma_G$ if and only if:

(i) γ is absolutely continuous

(ii) $\int_{\epsilon}^{1-\epsilon} dp/\gamma(p) < \infty$ when $0 < \epsilon < \frac{1}{2}$

(iii) $\gamma(0) = \gamma(1) = 0$

Suppose $p_0 \in]0,1[$ and that γ satisfies (i), (ii) and (iii). Then there is one and only one $G \in \mathcal{G}$ so that $G(0) = p_0$ satisfying

$$\gamma_G = \gamma$$

and having the property there is an interval I so that $[g > 0]$ is equivalent (Lebesgue) with I.

Remark.

1° Let γ be a continuous function on $[0,1]$ such that $\gamma(0) = \gamma(1) = 0$ and let $G \in \mathcal{G}$. Then $\gamma_G = \gamma$ if and only if G satisfies the differential equation $g = \gamma(G)$. Demonstration:
1° Suppose $g(x) = \gamma(G(x))$; $x \in]-\infty,\infty[$. Let $p \in]0,1[$ and put $x = G^{-1}(p)$. Then $\gamma_G(p) = g(x) = \gamma(G(x) = \gamma(p)$.

2° Suppose $\gamma_G = \gamma$. Let $g(x) > 0$ and put $p = G(x)$. Then $x = G^{-1}(p)$ so that $g(x) = \gamma_G(p) = \gamma(p) = \gamma(G(x))$. It follows - by continuity - that $g(x) = \gamma(G(x))$ when $x \in \overline{[g > 0]}$. Let $x \in \complement\overline{[g > 0]} = \overset{\circ}{\complement[g > 0]}$. Then $g(y) = 0$ for all y in some interval $]x-\epsilon, x+\epsilon[$. On this interval G is a constant p and we may assume that $0 < p < 1$ (if otherwise then: $g(x)=0=\gamma(G(x))$. The interval $[G^{-1}(p), G_*^{-1}(p)]$ is not - since it contains $]x-\epsilon, x+\epsilon[$ - degenerate. Hence $g(x) = 0 = g(G^{-1}(p)) = \gamma(p) = \gamma(G(x))$.

<u>Proof of the theorem:</u> 1^{o} Suppose γ satisfies (i), (ii) and (iii). For each $p \in [0,1]$ put $\Psi(p) = \int_{p_0}^{p} dp/\gamma(p)$. Then Ψ is continuous, strictly increasing and finite on $]0,1[$. It follows that to each $x \in]\Psi(0), \Psi(1)[$ there corresponds one and only one number $G(x) \in]0,1[$ so that

$$\Psi(G(x)) = x .$$

Extend - if necessary - G to $]-\infty,\infty[$ - by writing $G(x)=0$ when $x \leqq \Psi(0)$ and writing $G(x) = 1$ when $x \geqq \Psi(1)$. It is easily seen that G is a continuous distribution function on $]-\infty,\infty[$ which is strictly increasing on $]\Psi(0), \Psi(1)[$. $G(0) = p_0$ since $\Psi(p_0) = 0$.

Put $g = \gamma(G)$. Then g is continuous and non negative on $]-\infty,\infty[$. Furthermore $g(x) = 0$ when $x \leqq \Psi(0)$ or $x \geqq \Psi(1)$. Let x be any number in $]\Psi(0), \Psi(1)[$. Then $\int_{-\infty}^{x} g(y)dy$

$$= \int_{\Psi(0)}^{x} \gamma(G(y))dy = \int_{0<G\leqq G(x)} \gamma(G(y))\mu(dy) = \int_{0}^{G(x)} \gamma(p)(\mu G^{-1})(dp).$$ The measure

μG^{-1} is clearly non atomic on $]0,1[$ and for $0 < a < b < 1$: $\mu G^{-1}([a,b]) = \mu(\{x: a \leqq G(x) \leqq b\}) = \mu([\Psi(a), \Psi(b)]) = \Psi(b) - \Psi(a)$ $= \int_{a}^{b} dp/\gamma(p)$. Hence

$$\int_{-\infty}^{x} g(y)dy = \int_{0}^{G(x)} \gamma(p)\, dp/\gamma(p) = G(x).$$

It follows that G is absolutely continuous with density g. This in turn imply - since g is the composite $\gamma \circ G$ where γ is absolutely continuous and G is an absolutely continuous distribution function - that g is absolutely continuous on finite intervals. Let N be a Borel sub set of $]0,1[$ having Lebesgue measure 0. Then $(\mu G^{-1})(N) = \mu(\Psi[N]) = 0$ since Ψ is absolutely continuous on any interval $[\epsilon, 1-\epsilon]$ where $0 < \epsilon < \frac{1}{2}$. By (ii) $\gamma > 0$ a.e. μ on $]0,1[$. It follows that $g(x) = \gamma(G(x)) > 0$ a.e. on $]\Psi(0), \Psi(1)[$. Similarily $\gamma'(G(x))$ exist a.e. on $]\Psi(0), \Psi(1)[$ so that $g'(x) = \gamma'(G(x))g(x)$ a.e. on $]\Psi(0), \Psi(1)[$.

Hence $$\int_{-\infty}^{\infty} |g'(x)|dx = \int_{\Psi(0)}^{\Psi(1)} |\gamma'(G(x))|g(x)dx = \int_{\Psi(0)}^{\Psi(1)} |\gamma'(G)|dG$$

$$= \int^{1} |\gamma'(p)|dp < \infty .$$ This proves that $G \in \mathcal{G}$ and substituting $G(x) = p \in]0,1[$ in $g(x) = \gamma(G(x))$ yields $g(G^{-1}(p)) = \gamma(p)$. Hence $\gamma = \gamma_G$.

Let $G_1 \in \mathcal{G}$ be such that $G_1(0) = p_0$, $g_1 > 0$ is almost (Lebesgue) equal to the interval $]k_0, k_1[$, and satisfying $\gamma = \gamma_{G_1}$. Clearly $0 \in]k_0, k_1[$. By proposition 7.3 we have - for $1 > p > p_0$ -

$$G^{-1}(p) = G^{-1}(p) - G^{-1}(p_0) = \mu([G^{-1}(p_0), G^{-1}(p)]) = \int_{p_0}^{p} dp/\gamma(p)$$

$$= \mu([G_1^{-1}(p_0), G_1^{-1}(p)]) = G_1^{-1}(p) - G_1^{-1}(p_0) = G_1^{-1}(p).$$ Similarily $G^{-1}(p) = G_1^{-1}(p)$ when $p \in]0,p_0[$. It follows that $G = G_1$.

Altogether we have proved the last statement and the "if" part of the first statement. The proof is completed by noting that the "only if" part of the first statement follows from proposition 7.3.
□

Let $G \in \mathcal{G}$. The derivative of the experiment $\mathcal{E}_G$ may be represented as the ordered pair $(G, \dot{G})$. It was shown in the proof of theorem 7.2 that $g(x) = -\dot{G}\,]-\infty,x[$ so that $d\dot{G}/d\mu = -g'$.

Adapting the notations of chapter 6 we write:

$$F_G \overset{\text{def.}}{=} \mathcal{L}_G(-g'/g) \ ,$$

$$U_G(\xi) \overset{\text{def.}}{=} \|\xi G - \dot{G}\| = \int |\xi g + g'| d\mu; \ \xi \in \ ]-\infty,\infty[$$

and $$\beta_G(\alpha) \overset{\text{def.}}{=} \sup \{\dot{G}(\delta) : 0 \leqq \delta \leqq 1, \ G(\delta) = \alpha\}; \ \alpha \in [0,1]$$

The derivatives $[\dot{\mathcal{E}}_G]_{\theta_0}$, $\theta_0 \in \ ]-\infty,\infty[$ are - since translates of the same distribution are Δ equivalent - $\dot{\Delta}$ eqivalent. It follows that no ambiguity should arise by deleting the subscript θ_0 on F_G, U_G and β_G.

Rewriting the expression for U_G we get:

$$U_G(\xi) = \int_{g>0} |\xi + g'/g| g d\mu = \int |\xi + g'(x)/g(x)| G(dx) = \int |\xi + \gamma_G'| d\lambda$$

Now γ_G is the distribution function of the measure which assigns mass $\gamma_G(q) - \gamma_G(p)$ to $[p,q]$ when $0 \leqq p \leqq q \leqq 1$. This measure will - by abuse of notations - also be written γ_G. The measure γ_G is absolutely continuous w.r.t. λ. The pair $(\lambda, -\gamma_G)$ defines a derivative and the U function for this derivative maps ξ into $\|\xi\lambda + \gamma_G\| = \int |\xi + \gamma_G'| d\lambda$. We have proved:

<u>Theorem 7.12.</u>

The pair $(\lambda, -\gamma_G)$, considered as a derivative, is $\dot{\Delta}$ equivalent with $\dot{\mathcal{E}}_G$. If $G_1, G_2 \in \mathcal{G}$ then:

$$\Delta(G_1, G_2) = 0 \Rightarrow \gamma_{G_1} = \gamma_{G_2} \Rightarrow \dot{\Delta}(G_1, G_2) = 0$$

Remark.

Neither of these arrows can be reversed. An example where $\gamma_{G_1} = \gamma_{G_2}$ and $\Delta(G_1, G_1) > 0$ is provided by example 7.5. We shall later show that G_1 and G_2 easily may be chosen so that $\dot{\Delta}(G_1, G_2) = 0$ and $\gamma_{G_1} \neq \gamma_{G_2}$.

Which derivatives β [i.e. which concave functions β on $[0,1]$ with $\beta(0) = \beta(1) = 0$] are of the form β_G for some $G \in \mathcal{G}$? We begin the study of this problem with the negative result:

Theorem 7.13.

$\beta_G \neq 0$ for all $G \in \mathcal{G}$, i.e. $\beta_G(p) > 0$ when $p \in \,]0,1[$ and $G \in \mathcal{G}$.

Proof: Suppose $\beta_G = 0$. Then $\dot{G} = 0$ and this would imply that G_θ would be independent of θ and this an impossibility for countably additive probability measures G. □

The situation described in theorem 7.13 is, however, the only exception since we have:

Theorem 7.14.

Let $\beta \neq 0$ be a derivative. Then the differential equation

$$G' = \beta(1-G)$$

has a solution $G \in \mathcal{G}$ such that

$$\beta_G = \beta$$

The class of all non constant solutions of this differential equation is precisely the class of translates of G.

Proof: Conditions (i), (ii) and (iii) of theorem 7.11 are satisfied by the map $\gamma : p \rightsquigarrow \beta(1-p)$. Let $p_o \in]0,1[$ and put

$$\Psi(p) = \int_{p_o}^{p} dp/\beta(1-p) \; ; \; p \in [0,1].$$

It was shown in the proof of theorem 7.11 that there is a $G \in \mathcal{G}$ with $\gamma_G = \gamma$ satisfying $\Psi(G(x)) = x \; ; \; x \in]\Psi(0), \Psi(1)[$. Let $x \in]\Psi(0), \Psi(1)[$ and put $p = G(x)$. Then $G^{-1}(p) = x$ so that $g(x) = \gamma_G(p) = \beta(1-p)$ $= \beta(1-G(x))$. Trivially $g(x) = \beta(1-G(x))$ when $x \notin]\Psi(0), \Psi(1)[$.

By theorem 7.12:

$$U_G(\xi) = \int |\xi+\gamma_G'|d\lambda = \int_0^1 |\xi-\beta'(1-p)|dp = \int_0^1 |\xi-\beta'(p)|dp$$

Hence - by theorem 4.1 - $\beta_G = \beta$.

Let H be any nonconstant solution of the differential equation. It is easily seen that any translate of H is also a solution.

Clearly H is continuous, monotonically increasing and the range is a subinterval of $[0,1]$. Suppose $H \leqq C < 1$. There is, by assumption, a x_o so that $H(x_o) > 0$. Let $x > x_o$. Then: $0 < 1-C \leqq 1-H(x) \leqq 1-H(x_o) < 1$. It follows that there is a $k > 0$ so that $H'(x) = \beta(1-H(x)) \geqq k$ when $x \geqq x_o$. Hence $1 \geqq H(x)-H(x_o) \geqq k(x-x_o)$ when $x \geqq x_o$ and this is a contradiction since $k(x-x_o) \to \infty$ as $x \to \infty$. It follows that $H(\infty) = 1$. Similarily $H(-\infty) = 0$. It follows, since H is continuous, that we may - wihout loss of generality - assume that H is a distribution function such that $H(0) = p_o$.

Put $t_o = \inf\{x : H(x) > 0\}$ and $t_1 = \sup\{x : H(x) < 1\}$. Then $t_o < 0 < t_1$. Consider the map $x \rightsquigarrow \Psi(H(x))$ from $]t_o,t_1[$ to $]-\infty,\infty[$. The derivative is $x \rightsquigarrow \Psi'(H(x))H'(x) = = [\beta(1-H(x))]^{-1}\beta(1-H(x)) = 1$ an it maps 0 into $\Psi(H(0)) = \Psi(p_o)=0$. Hence $\Psi(H(x)) = x$ when $x \in]t_o,t_1[$. Let $x \downarrow t_o$. Then $H(x) \downarrow 0$ so that $x = \Psi(H(x)) \downarrow \Psi(0)$. Hence $t_o = \Psi(0)$. Similarily $t_1 = \Psi(1)$. It follows that $H = G$. □

The distribution functions G satisfying $G' = \beta(1-G)$ are - by theorem 7.14 - in $\mathcal{G}$ and have the further property that $p \rightsquigarrow g(G^{-1}(p)) = \beta(1-p)$ is concave on $]0,1[$. Let $\mathcal{G}_o$ be the class of probavility distributions G having a continuous density g so that $p \rightsquigarrow g(G^{-1}(p))$ is concave on $]0,1[$. Clearly $\mathcal{G}_o$ is invariant under translations. Our first result on $\mathcal{G}_o$ is:

Proposition 7.15.

If $G \in \mathcal{G}_o$ then

$$\lim_{p\to o} g(G^{-1}(p)) = \lim_{p\to 1} g(G^{-1}(p)) = 0 .$$

Proof: Put $\tau(p) = g(G^{-1}(p))$ when $p \in]0,1[$. Then $\tau > 0$ a.e. Lebesgue on $]0,1[$ so that $\tau(p) \geqq 0$ for all $p \in]0,1[$. Clearly $\tau(0+)$ and $\tau(1-)$ exist. By concavity $\tau(p) \geqq \tau(1-)p$; $p \in]0,1[$; i.e $g(G^{-1}(p)) \geqq \tau(1-)p$; $p \in]0,1[$. Inserting $p = G(x)$ we get $g(x) = g(G^{-1}(G(x)) \geqq \tau(1-)G(x)$ when $G(x) \in]0,1[$. Suppose first that $G(x) < 1$ for all x . Then $\liminf_{x\to\infty} g(x) \geqq \tau(1-)$ and this is only possible when $\tau(1-) = 0$. Suppose next that $x_o = \inf\{x : G(x) = 1\} < \infty$. Then $g(x_o) \geqq \tau(1-)$.

If $\tau(1-) > 0$ then - by continuity $g > 0$ in a neighbourhood of x_0. G is - necessarily - < 1 on this neighbourhood and this contradicts the assumption on x_0. It follows that $\tau(1-) = 0$. Similarity $\tau(0+) = 0$. □

As the notation $\mathcal{G}_0$ indicates we have:

Proposition 7.16.

$$\mathcal{G}_0 \subseteq \mathcal{G}$$

Proof: Let $G \in \mathcal{G}_0$. By proposition 7.15 $p \rightsquigarrow g(G^{-1}(p))$ is concave on $[0,1]$. Put $\beta(1-p) = g(G^{-1}(p))$ when $p \in [0,1]$. Then β is a derivative and $G'(x) = \beta(1-G(x))$. The proposition follows now from theorem 7.14. □

Proposition 7.17.

To any derivative $\beta \neq 0$ corresponds a $G \in \mathcal{G}_0$ so that $\beta_G = \beta$. G is unique up to a translation. If $G \in \mathcal{G}_0$ then $\beta_G(p) = g(G^{-1}(1-p))$

Proof: The first statement follows from theorem 7.14. Let $G \in \mathcal{G}_0$ and put $\beta(p) = g(G^{-1}(1-p))$; $p \in [0,1]$. Then β is a derivative and $\beta \neq 0$. By theorem 7.14 $\beta_G = \beta$ and this proves the last statement. Suppose $G_1 \in \mathcal{G}_0$ and that $\beta_G = \beta$. As we have seen $\beta(p) = g(G^{-1}(1-p))$ or equivalently

$$g(x) = \beta(1-G(x))$$

By theorem 7.14 again this determines G up to a translation. □

Proposition 7.17 tells us that any derivative $\beta \neq 0$ is the derivative of an experiment $\mathcal{E}_G$ with $G \in \mathcal{G}_0$ and that G is (restricted to $\mathcal{G}_0$) unique up to Δ equivalence.

<u>Proposition 7.18.</u>

Let $G \in \mathcal{G}_o$. Then

$$\int_a^b -F_G^{-1}(G(x))dx = \log g(b) - \log g(a)$$

when $a,b \in \left]\inf\{x:G(x) > 0\}, \sup\{x:G(x) < 1\}\right[$

In particular $\log g$ is concave on this interval.

<u>Proof:</u> Put $k_o = \inf\{x:G(x) > 0\}$, $k_1 = \sup\{x:G(x) < 1\}$ and $\beta(p) = g(G^{-1}(1-p))$; $p \in [0,1]$. g is absolutely continuous. Hence $\log g$ is absolutely continuous in any interval $[k_o+\varepsilon, k_1-\epsilon]$ where $\epsilon > 0$. Now $g(x) = \beta(1-G(x))$ and $\beta(\alpha) = \beta_G(\alpha) = \int_o^\alpha F_G^{-1}(1-p)dp$. It follows that $\beta'(\alpha) = F^{-1}(1-\alpha)$ when $1-\alpha$ is a point of continuity for the map $p \leadsto F_G^{-1}(p)$. Let $C = \{p: 0 < p < 1$ and F_G^{-1} is discontinuous in $p\}$. Then C is countable and $\beta'(\alpha) = F_G^{-1}(1-\alpha)$ when $1-\alpha \notin C$. G is strictly increasing - and therefor 1-1 - on $]k_o, k_1[$. Now $g'(x) = -F_G^{-1}(G(x))g(x)$ when $x \in]k_o, k_1[$ and $1-G(x) \notin C$. It folows that ${}^d/_{dx}\log g(x) = -F_G^{-1}(G(x))$ for any $x \in]k_o, k_1[$ with at most a countable set of exceptions. □

<u>Theorem 7.19.</u>

The distribution function $G \in \mathcal{G}_o$ if and only if G has a continuous density g such that $[g > 0]$ is an interval on which $\log g$ is concave.

Proof: The "only if" follows from proposition 7.18. Suppose G is a distribution function having a continuous density g such that $g > 0$ is an interval on which $\log g$ is concave. Put $t_0 = \inf\{x:g(x) > 0\}$, $t_1 = \sup\{x:g(x) > 0\}$ and $l(x) = \log g(x)$ when $x \in]t_0, t_1[$. Then $[g > 0] =]t_0, t_1[$. Let $x_0 \in]t_0, t_1[$ be such that $g(x_0) = \sup_x g(x)$. Then l - and consequently g - is monotonically increasing on $]t_0, x_0]$ and monotonically decreasing on $[x_0, t_1[$. It follows that $\lim_{x \to t_0} g(x) = \lim_{x \to t_1} g(x) = 0$ and that $g'(x)$ exist for almost (Lebesgue) all x. Put $N = \{x: g'(x) \text{ does not exist}\}$. Then $\mu(N) = 0$. Now $d/_{dp}G^{-1}(p) = [g(G^{-1}(p)]^{-1}$, $p \in]0,1[$. Furthermore $d/_{dx}\log g(x) = g'(x)$ when $x \in]t_0, t_1[-N$. Hence $d/_{dp}g(G^{-1}(p)) = g'(G^{-1}(p))/g(G^{-1}(p))$ when $G^{-1}(p) \in]k_0, k_1[-N$. G^{-1} is absolutely continuous on compact subintervals of $]0,1[$ and $g(x)=e^{l(x)}$ is absolutely continuous on compact subintervals of $]t_0, t_1[$. It follows - since G^{-1} is increasing on $]0,1[$ - that $p \leadsto g(G^{-1}(p))$ is absolutely continuous on compact sub intervals of $]0,1[$. Hence, since $d/_{dp}g(G^{-1}(p))$ is monotonically decreasing on the set $]0,1[- G(N)$, $p \leadsto g(G^{-1}(p))$ is concave on $]0,1[$. It follows that $G \in \mathcal{G}_0$. □

A probability distribution G on $]-\infty,\infty[$ is called unimodal if there is a number a (not necessarily unique) so that G is convex on $]-\infty, a[$ and concave on $]a,\infty[$. If G is unimodal and G is convex on $]-\infty, a[$ and concave on $]a,\infty[$ then the left hand derivative $(D_l G)(x)$ and the right hand derivative $(D_r G)(x)$ exists for all x and they are finite when $x \neq a$. The set, J_G, of points x such that $(D_l G)(x) > 0$ and $(D_r G)(x) > 0$ is an interval of G probability 1. Any point $x \in J_G$ is a point of increase for G.

<u>Proposition 7.20.</u>

If $G \in \mathcal{G}_0$ then G is unimodal..

<u>Proof</u>: We use the notations of the proof of theorem 7.19. It was shown there that g is monotonically increasing on $]-\infty, x_0[$ and monotonically decreasing on $]x_0, \infty[$. It follows that G is convex on $]-\infty, x_0[$ and concave on $]x_0, \infty[$. □

A probability distribution G is called strongly unimodal if the convolution $G * H$ is unimodal whenever the probability distribution H is unimodal. Any strongly unimodal probability distribution is unimodal. It has been shown by Ibragimov [6] that a non atomic unimodal distribution function is strongly unimodal if and only if $x \rightsquigarrow \log G^*(x)$ is concave on the interval $J_G = \{x: (D_l G)(x) \text{ and } (D_r G)(x) > 0\}$. Here G^* may denote any function such that - for each x - $G^*(x)$ is either the left hand derivative $(D_l G)(x)$ or the right hand derivative $(D_r G)(x)$. We will use this to prove

<u>Theorem 7.21.</u>

Let G be a non atomic probability distribution and let $\hat{G}(x)$ be a function from $]-\infty, \infty[$ such that $\hat{G}(x)$ is - for each x - an accumulation point for $[G(x+h)-G(x)]/h$ as $h \to 0$.

Then $G \in \mathcal{G}_0$ if and only if G is strongly unimodal and $\hat{G}$ is a continuous function from $]-\infty, \infty[$ to $]-\infty, \infty[$.

Proof: 1° Suppose $G \in \mathcal{G}_0$. Then $(D_l G)(x) = (D_r G)(x) = G'(x) = g(x)$ for all x and J_G is the interval $[g > 0]$. Strong unimodality follows now from Ibragimov's criterion and theorem 7.19.

2° Suppose G is strongly unimodal, nonatomic, and that $\hat{G}$ is continuous. Let a be a number so that G is convex on $]-\infty, a[$ and concave on $]a, \infty[$. It is easily seen that this - since $G(\{a\}) = 0$ - imply that G is absolutely continuous. A density $g \geqq 0$ for G may now be specified so that g is monotonically increasing on $]-\infty, a[$ and monotonically decreasing on $]a, \infty[$. Then $(D_l G)(x) = g(x-)$ and $(D_r G)(x) = g(x+)$ for all x. By the continuit of $\hat{G}$ and the piecewise monotonicity of g we get $g(x) = \hat{G}(x)$ when $x \neq a$, and we may modify - if necessary - g so that $g(a) = \hat{G}(a)$. It follows that G has a continuous density g and that $J_G = [g > 0]$. Hence - by Ibragimov's criterion and theorem 7.19 - $G \in \mathcal{G}_0$. □

Corollary 7.22.

$G \in \mathcal{G}_0$ if and only if G is strongly unimodal and has a continuous density.

Let us next consider the problem of symmetry. If G is any probability distribution on $]-\infty, \infty[$ then the distribution of $-X$ when X has the distribution G will be denoted by $\bar{G}$. It is easily seen that $\bar{\bar{G}} = G$, $G \in \mathcal{G} \Leftrightarrow \bar{G} \in \mathcal{G}$ and that $G \in \mathcal{G}_0 \Leftrightarrow \bar{G} \in \mathcal{G}_0$.

Proposition 7.23.

Let G be absolutely continuous. Then $\Delta(\mathcal{E}_G, \mathcal{E}_{\overline{G}}) = 0$ if and only if G is symmetric. In particular $\dot{\mathcal{E}}_G$ is symmetric provided G is symmetric and $G \in \mathcal{G}$. On the other hand G is symmetric provided $G \in \mathcal{G}_o$ and $\mathcal{E}_G$ is symmetric.

Proof: The first statement is an immediate consequence of the convolution criterion for Δ comparison of translation experiments. This and the fact "$\dot{\Delta}(\dot{\overline{\mathcal{E}}}_G, \dot{\mathcal{E}}_{\overline{G}}) = 0$" implies the next statement. Finally suppose $G \in \mathcal{G}_o$ and that $\dot{\mathcal{E}}_G$ is symmetric. Then - by corollary 4.4 - $\beta_G(p) = g(G^{-1}(1-p)) = g(G^{-1}(p))$. Simple calculations show that $\beta_{\overline{G}}(p) = \overline{g}(\overline{G}^{-1}(1-p)) = g(G^{-1}(p)) = \beta_G(p)$. By theorem 7.14 $\overline{G}$ is a translate of G i.e. G is symmetric. □

We include here - for the sake of completeness - a few facts (it is essentially example 1 in chapter 8 in Lehmann [10]) on monotone likelihood ratios of translation families.

Suppose $G \in \mathcal{G}_o$ and let $\theta_1 < \theta_2$. Then $g_{\theta_2}(x)/g_{\theta_1}(x) = 0$, $\exp-[\log g(-\theta_1+x)-\log g(-\theta_2+x)]$, and ∞ as $x \in [g_{\theta_1} > 0] \cap [g_{\theta_2} = 0]$, $x \in [g_{\theta_2} > 0] \cap [g_{\theta_1} > 0]$ and $x \in [g_{\theta_1} = 0] \cap [g_{\theta_2} > 0]$. It follows - by concavity - that G_{θ_2} has monotonically increasing likelihood ratio w.r.t. G_{θ_1} when $\theta_2 > \theta_1$. Hence the test $I_{[G^{-1}(1-\alpha),\infty[}$ is a UMP test for testing $\theta \leqq 0$ against $\theta > 0$, provided $\alpha > 0$. The power function of this test is

$$\theta \leadsto 1 - G(G^{-1}(1-\alpha)-\theta)$$

and the derivative in 0 of this function is $g(G^{-1}(1-\alpha))$, as it, by proposition 7.17, should be. Conversely suppose the probability distribution G has a continuous density g and that G_{θ_2} has monotonically increasing likelihood ratio w.r.t. G_{θ_1} when $\theta_2 > \theta_1$. Then $g_{\theta_2}/g_{\theta_1}$ is monotonically increasing on $[g_{\theta_1} > 0] \cup [g_{\theta_2} > 0]$. Let $g(a) > 0$, $g(b) > 0$ and $a < b$. Put $x = 0$, $x' = (b-a)/2$, $\theta = -(a+b)/2$ and $\theta' = -a$. Then $x < x'$ and $\theta < \theta'$. Hence

$$g(a)/g((a+b)/2) = g(x-\theta')/g(x-\theta) \leqq g(x'-\theta')/g(x'-\theta) =$$
$$= g(a+b)/2)/g(b).$$

Hence $g((a+b)/2) > 0$ and

$$\tfrac{1}{2}\log g(a) + \tfrac{1}{2}\log g(b) \leqq \log g(\tfrac{1}{2}(a+b))$$

It follows that $[g > 0]$ is an interval and that $\log g$ is concave on $[g > 0]$. By theorem 7.19: $G \in \mathcal{G}_0$.

<u>Example 7.24.</u> (Normal distribution)

Let $G = \Phi$ where Φ is the normal (0,1) distribution. Write $\varphi = \Phi'$. Then: $\varphi'(x)/\varphi(x) = -x$; $x \in]-\infty,\infty[$ so that $\Phi \in G_0$ and $F_\Phi = \Phi$, $U_\Phi(\xi) = \int |\xi - x| d\Phi$ and $\beta_\Phi(p) = \varphi(\Phi^{-1}(1-p))$; $p \in [0,1]$.

<u>Example 7.25.</u> (Triangular distribution)

Let G be the distribution whose density g is given by:

$$G(x) = (1-|x|)^+ \; ; \; x \in]-\infty,\infty[$$

Then $G \in \mathcal{G}_0$ and G is symmetric about 0. It suffices therefor to calculate $F_G(x)$ for $x \geqq 0$, $U_G(\xi)$ for $\xi \geqq 0$ and $\beta_G(p)$ for $p \leqq \frac{1}{2}$. Now $-g'(x)/g(x) = -(1+x)^{-1}$ or $= (1-x)^{-1}$ as $x \in]-1,0[$ or $x \in]0,1[$. It follows that

$$F_G(x) = \begin{cases} \frac{1}{2x^2} & \text{when} \quad x \leq -1 \\ \frac{1}{2} & \text{"} \quad |x| \leq 1 \\ 1-\frac{1}{2x^2} & \text{"} \quad x \geq 1 \end{cases},$$

$$U_G(\xi) = \begin{cases} 2 & \text{when} \quad |\xi| \leq 1 \\ |\xi| + |\xi|^{-1} & \text{when} \quad |\xi| \geq 1 \end{cases},$$

and $\beta_G(p) = \sqrt{2 \min\{p,1-p\}}$; $p \in [0,1]$

<u>Example 7.26.</u> (Logistic distribution)

Put $G(x) = [1 + e^{-x}]^{-1}$; $x \in]-\infty,\infty[$.

Then $G \in \mathcal{G}_0$, G is symmetric and $g(x) = e^{-x}[1+e^{-x}]^{-2}$; $x \in]-\infty,\infty[$ so that $-g'(x)/g(x) = 2e^{-x}(1+e^{-x})^{-1}-1$; $x \in]-\infty,\infty[$

Hence

$\beta_G'(p) = -g'(G^{-1}(1-p))/g(G^{-1}(1-p)) = 2p-1$; $p \in [0,1]$ so that $\beta_G(p) = p(1-p)$; $p \in [0,1]$

and

$F_G(x) = \lambda(\{y : \beta_G'(y) \leq x\}) = (1+x)/2$ when $|x| \leq 1$ i.e. F_G is the inform distribution on $[-1,1]$.

Finally

$$U_G(\xi) = 2\int_{-\infty}^{\xi} F_G(x)dx-\xi = \begin{cases} |\xi| & \text{when} \quad |\xi| \geq 1 \\ (1+\xi^2)/2 & \text{when} \quad |\xi| < 1. \end{cases}$$

Let us compare this experiment with the experiment Φ treated in example 7.24.

We get: $\beta_\Phi(1-\Phi(x)) - \beta_G(1-\Phi(x)) = \varphi(x) - \Phi(x)(1-\Phi(x))$.

The derivative of this function is $\psi(x)\varphi(x)$ where $\psi(x) = 2\Phi(x)-x-1$ so that

$\psi'(x) = 2\varphi(x)-1 < \sqrt{\frac{2}{\pi}} - 1 < 0$

It follows that $\beta_\Phi(p) - \beta_G(p)$ has maximum $1/\sqrt{2\pi} - 1/4$ for $p = 1/2$ and minimum $= 0$ at $p = 0,1$. It follows that $\dot{\delta}(\mathcal{E}_\Phi, \mathcal{E}_G) = 0$ and $\dot{\delta}(\mathcal{E}_G, \mathcal{E}_\Phi) = \dot{\Delta}(\mathcal{E}_G, \mathcal{E}_\Phi) = 1/\sqrt{2\pi} - 1/4$.

Example 7.27. (Double exponential).

Let G be given by the density $g(x) = \frac{1}{2}e^{-|x|}$; $x \in]-\infty,\infty[$. Then $g(G^{-1}(1-p)) = \min\{p,1-p\}$. It follows that

$$\beta_G(p) = \min\{p,1-p\} \;;\; p \in [0,1]$$

Now $F_G = \mathcal{L}_\lambda(\beta_G')$ so that

$$F_G(\{-1\}) = F_G(\{1\}) = 1/2$$

Hence

$$U_G(\xi) = \max(1, |\xi|) .$$

Examples 7.24-7.27 were all concerned with strongly unimodal distributions. Any experiment $\mathcal{E}_G$, however, is $\dot{\Delta}$ equivalent with some experiment $\mathcal{E}_{G_0}$ with $G_0 \in \mathcal{G}_0$. G_0 is - up to a shift - determined by G. If G is given then G_0 may be found by solving the differential equation $G_0' = \beta_G(1-G_0)$. If $G \in \mathcal{G}_0$ then G_0 is (and may be any) a shift of G. On the other hand - if G is not strongly unimodal - then we have a situation where $\gamma_G \neq \gamma_{G_0}$ while $\dot{\Delta}(\mathcal{E}_{G_1}, \mathcal{E}_{G_0}) = 0$. This proves the last assertion made in the remark after theorem 7.12.

Example 7.28. (Examples 7.5 and 7.25 continued)

Simple calculations yield

$$\beta_{G_1}(p) = \beta_{G_2}(p) = \sup\{-\int \delta(x)g_1'(x)dx : \int \delta(x)g_1(x)dx = p\} = \sqrt{8\min\{p,1-p\}}.$$

By proposition 6.19 and example 7.25: $\beta_{G_1} = \beta_{G_2} = \beta_{G_3}$ where G_3 is the triangular distribution with density:

$$g_3(x) = [1-|x|/2]^+/2 \quad : \quad x \in]-\infty,\infty[.$$

The pseudo metric $\dot{\Delta}$ defines a pseudo metric - which by abuse of the notations, also will be written $\dot{\Delta}$ - on $\mathcal{G}_o$ by:

$$\dot{\Delta}(G_1,G_2) \overset{\text{def}}{=} \dot{\Delta}(\mathcal{E}_{G_1},\mathcal{E}_{G_2}) \; ; \; G_1, G_2 \in \mathcal{G}_o$$

Any differentiable experiment (or derivative) with $\beta \neq 0$ is $\dot{\Delta}_{\theta_o}$ equivalent with an experiment $\mathcal{E}_{G_o}$ where $G_o \in \mathcal{G}_o$ is determined up to a shift. In particular; any differentiable experiment based on n observations is $\dot{\Delta}_{\theta_o}$ equivalent with an experiment $\mathcal{E}_{G_o}$ which is based on one observation.

We shall now consider convergence for the pseudo metric $\dot{\Delta}$ on $\mathcal{G}_o$. It will turn out that $\dot{\Delta}$ on $\mathcal{G}_o$ is topologically equivalent with Δ on $\mathcal{G}_o$. Various convergence criterions will be derived and as a biproduct we will get a result relating the convergence of densities to the convergence of the probability measures determined by the densities.

<u>Proposition 7.29.</u>

Let $G, G_1, G_2, \ldots \in \mathcal{G}_o$ and let $p_o \in \,]0,1[$. Suppose $G(0) = G_n(0) = p_o$; $n = 1,2 \ldots$ and that

$$\lim_{n\to\infty} \dot{\Delta}(G_n,G) = 0$$

Then

$$\lim_{n\to\infty} \sup_x |g_n(x)-g(x)| = 0$$

In particular

$$\lim_{n\to\infty} \|G_n-G\| = 0.$$

<u>Proof:</u> Write $\Psi_n(p) = \int_{p_o}^{p} dp/\beta_{G_n}(1-p)$ and $\Psi(p) = \int_{p_o}^{p} dp/\beta_G(1-p)$.

By assumption

$\sup_p |\beta_{G_n}(p)-\beta_G(p)| \to 0$ so that $^1/\beta_{G_n}(1-p) \to {^1/\beta_G}(1-p)$

uniformly on any interval $[\epsilon,1-\epsilon]$ where $\epsilon > 0$. It follows that $\Psi_n(p) \to \Psi(p)$ uniformly on any of these intervals. In particular:

$\limsup_n \Psi_n(0) \leqq \Psi(0) < \Psi(1) \leqq \liminf_n \Psi_n(1)$

Let $x \in]\Psi(0),\Psi(1)[$ and consider a sub sequence G_{n_k} ; $k=1,2...$ so that $G_{n_k}(x) \to \tau$ as $k \to \infty$. Then $\Psi_n(0) < x < \Psi_n(1)$ for n sufficiently large. If $\tau = 0$ then $G_{n_i}(x) < y$ where $y \in]0,1[$ for i sufficiently large and then:

$x = \Psi_{n_i}(G_{n_i}(x) \leqq \Psi_{n_i}(y)$ when $\Psi_{n_i}(0) < x < \Psi_{n_i}(1)$

Hence $x \leqq \Psi(y)$ for $y > 0$ so that $x \leqq \Psi(0)$; i.e. a contradiction. If $\tau = 1$ then $G_{n_i}(x) > y$ where $y \in]0,1[$ for i sufficiently large, and then:

$x = \Psi_{n_i}(G_{n_i}(x)) \geqq \Psi_{n_i}(y)$ when $\Psi_{n_i}(0) < x < \Psi_{n_i}(1)$.

Hence $x \geqq \Psi(y)$ for $y < 1$ so that $x \geqq \Psi(1)$ i.e. another contradiction. It follows that $0 < \tau < 1$ and then $0 < G_{n_i}(x) < 1$ for i sufficiently large so that $x = \Psi_{n_i}(G_{n_i}(x) \to \Psi(\tau)$, i.e. $x = \Psi(\tau)$, Hence $\tau = G(x)$. By a standard compactness argument $G_n(x) \to G(x)$ when $x \in]\Psi(0),\Psi(1)[$. This, however, imply that $G_n(x) \to G(x)$ for any $x \in]-\infty,\infty[$. In particular $G_n \to G$ weakly. Hence - since G is continuous - $\sup_x |G_n(x)-G(x)| \to 0$ so that

$g_n(x) = \beta_n(1-G_n(x)) \to \beta(1-G(x)) = g(x)$ uniformly in x. The last statement follows from Scheffés convergence theorem. □

Dropping the condition $G_n(0) = G(0) = \frac{1}{2}$, $n = 1,2,...$ we get:

Proposition 7.30.

Let $G, G_1, G_2 \ldots \in \mathcal{G}_o$ and suppose $\lim_{n\to\infty} \dot{\Delta}(G_n,G) = 0$. Then there is a sequence $\theta_1, \theta_2, \ldots$ in Θ so that $\lim_{n\to\infty} \sup_x |g_n(x-\theta_n)-g(x)| = 0$. In particular $\lim_{n\to\infty} \Delta(G_n,G) = 0$.

Proof: Choose $\eta_1\ \eta_2 \ldots$ and η so that $G_{\eta_1}(0) = G_{\eta_2}(0) = \ldots = G_\eta(0) = p_o$ where $p_o \in]0,1[$. By proposition 7.29

$$\sup_x |g_n(x-\eta_n)-g(x-\eta)| \to 0$$

so that

$$\sup_x |g_n(x-\theta_n)-g(x)| \to 0$$

where $\theta_n = \eta_n - \eta$; $n = 1,2 \ldots$ □

A result on the converse direction is:

Proposition 7.31.

Let $G, G_1, G_2, \ldots \in \mathcal{G}_o$ and suppose $\lim_{n\to\infty} \sup_x |g_n(x)-g(x)| = 0$.

Then

$$\lim_{n\to\infty} \dot{\Delta}(G_n,G) = 0$$

Proof: Trivially: $\sup_x |G_n(x)-G(x)| \to 0$. Let $p \in]0,1[$. Then $G_n(G_n^{-1}(p))-G(G_n^{-1}(p)) \to 0$ i.e. $G(G_n^{-1}(p)) \to p$ so that $G_n^{-1}(p) \to G^{-1}(p)$. Hence $g_n(G_n^{-1}(p)) \to g(G^{-1}(p))$ $p \in]0,1[$ i.e. $\beta_{G_n}(p) \to \beta_G(p)$ when $p \in [0,1]$.

The result now follows from theorem 5.1 (iv). □

Combining these results we get the convergence criterion:

Theorem 7.32.

Let $G, G_1, G, \ldots \in \mathcal{G}_o$. Then $\dot{\Delta}(G_n, G) \to 0$ if and only if there is a sequence θ_n ; $n = 1,2\ldots$ in Θ so that

$$\sup_x |g_n(x-\theta_n)-g(x)| \to 0$$

If so, then $\Delta(G_n, G) \to 0$

We shall need the following result:

Proposition 7.33.

Let $G \in \mathcal{G}_o$ and let $I = [\alpha_o, \alpha_1]$ be the closed sub interval of $]0,1[$ where β_G obtains its maximum. Put $k_o = \inf\{x: g(x) > 0\}$ and $k_1 = \sup\{x: g(x) > 0\}$. Then:

$$k_o < G^{-1}(1-\alpha_1) \leqq G^{-1}(1-\alpha_o) < k_1 \ ,$$

g is strictly increasing on $[k_o, G^{-1}(1-\alpha_1)]$,

$g = \max_p \beta_G(p) = \max_x g(x)$ on $[G^{-1}(1-\alpha_1), G^{-1}(1-\alpha_o)]$ and g is strictly decreasing on $[G^{-1}(1-\alpha_o), k_1]$.

Proof: The inequalities are obvious and the three last statements is a consequence of the differential equation $g = \beta_G(1-G)$ □

It is often difficult to obtain non trivial convergence statements on densities on the basis of weak convergence of the probability measures. If the probability measures are in $\mathcal{G}_o$, however, then quite strong condusions may be drawn.

Proposition 7.34.

Let $G, G_1, G_2 \ldots \in \mathcal{G}_o$ and suppose $\lim_{n\to\infty} G_n(x) = G(x)$ when $x \in]-\infty,\infty[$. Then $\lim_{n\to\infty} \sup_x |g_n(x)-g(x)| = 0$, provided $\max_x g_n(x) = g_n(0)$, $n > 1,2 \ldots$ and $\max_x g(x) > g(0)$. In particular g_n, $n = 1,2 \ldots$ are uniformly equicontinuous on $]-\infty,\infty[$.

Proof: Put
$$k_{nc} = \inf\{x: g_n(x) > 0\} ,$$
$$k_{n1} = \sup\{x: g_n(x) > 0\} ,$$
$$k_o = \inf\{x: g(x) > 0\}$$
and
$$k_1 = \sup\{x: g(x) > 0\}$$

Let $x \in]k_o,k_1[$. Then $G_n(x) \in]0,1[$ for n sufficiently large, and then $k_{no} < x < k_{n1}$. Hence $\liminf_n k_{no} \leqq x \leqq \limsup_n k_{n1}$.

The arbitrariness of x in $]k_o,k_1[$ implies
$$\limsup_n k_{no} \leqq k_o < k_1 \leqq \liminf_n k_{n1}$$

Let $0 < x$. Then for $0 < \epsilon < x$:
$$G_n[x-\epsilon,x] \geqq \epsilon g_n(x)$$
so that
$$\limsup_n g_n(x) \leqq \frac{G[x-\epsilon,x]}{\epsilon} \to g(x) \quad \text{as} \quad \epsilon \to 0$$

Hence $\limsup_n g_n(x) \leqq g(x)$

On the other hand:
$$G_n[x,x+\epsilon[\leqq g_n(x)\epsilon$$
so that
$$\liminf_n g_n(x) \geqq \frac{G[x,x+\epsilon[}{\epsilon} \to g(x)$$

Hence $\liminf_n g_n(x) \geqq g(x)$. Note that this argument holds for $x = 0$ also.

It follows that:

$$\lim_n g_n(x) = g(x) \quad \text{when} \quad x > 0 \quad \text{and} \quad \liminf_n g_n(0) \geqq g(0).$$

In the same way we may show that $\lim_n g_n(x) = g(x)$ when $x < 0$.

Let $0 < \epsilon < 2\epsilon < k_1$. Then - for n sufficiently large - $2\epsilon < k_{1n}$. By theorem 7.19 $\log g_n(\epsilon) = \log g_n(\frac{1}{2}\cdot 0+\frac{1}{2}\cdot\epsilon) \geqq$ $\geqq \frac{1}{2}\log g_n(0) + \frac{1}{2}\log g_n(2\epsilon)$, or equivalently:
$\log g_n(0) \leqq 2\log g_n(\epsilon)-\log g_n(2\epsilon)$ - for n sufficiently large.
Hence:

$$\log \limsup_n g_n(0) \leqq [2\log g(\epsilon)-\log g(2\epsilon)] \to \log g(0) \quad \text{as} \quad \epsilon \to 0.$$

It follows that $g_n(x) \to g(x)$; $x \in]-\infty,\infty[$.

Uniform convergence follows now from the fact that if $F, F_1, F_2, \ldots$ are probability distributions on $]-\infty,\infty[$ such that $F_n \to F$ weakly and F is continuous, then $\sup_x |F_n(x)-F(x)| \to 0$. The uniform convergence $g_n \to g$, in turn, implies uniform equicontinuity of the sequence $g_1, g_2, \ldots$. □
We shall now show that the conditions on the maxima is abundant.

<u>Theorem 7.35.</u>

Let $G, G_1, G_2, \ldots$ be strongly unimodal distributions with continuous densities $g, g_1, g_2, \ldots$. Then $\limsup_n \sup_x |g_n(x)-g(x)| = 0$ provided $G_n \to G$ weakly

---

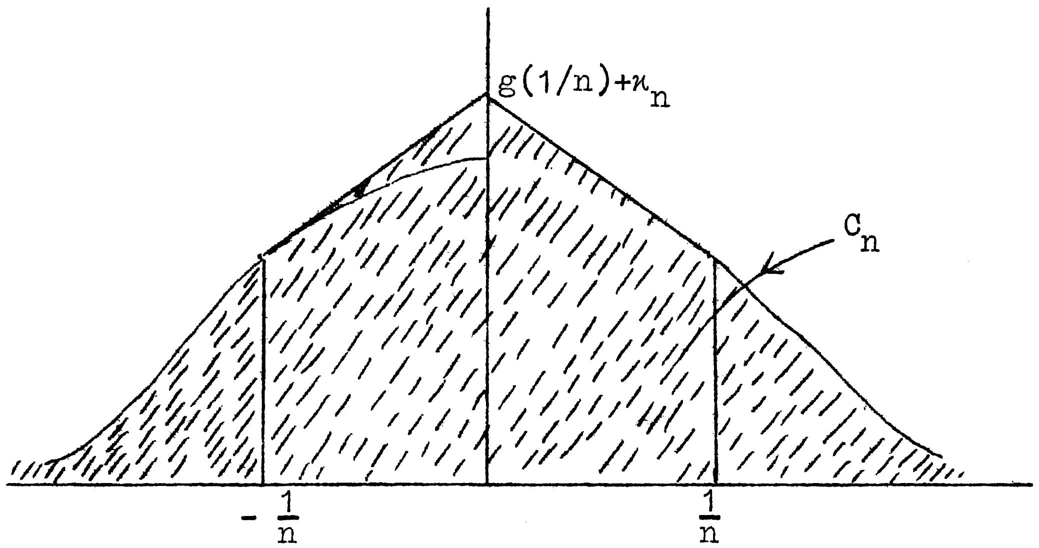
Proof: Let $g_n(a_n) \geqq g_n(x)$; $x \in]-\infty,\infty[$. Then a_n; $n = 1,2,\ldots$ is bounded and we may - without loss of generality - assume that $a_n \to a$. Then $g(a) \geqq g(x)$; $x \in]-\infty,\infty[$. It follows that $G_n * \delta_{-a_n}$; $n = 1,2 \ldots$ and $G * \delta_{-a}$ satisfies the conditions in proposition 7.34. Hence $\sup_x |g_n(x-a_n)-g(x-a)| \to 0$. By uniform equicontinuity $\sup_x |g_n(x-a_n)-g_n(x-a)| \to 0$, so that

$\sup_x |g_n(x-a)-g(x-a)| = \sup_x |g_n(x)-g(x)| \to 0$ □

Example 7.36.

Let $G \in \mathcal{G}_0$ be symmetric. Then g is even. Let $\varkappa_n$, $n = 1,2\ldots$ be any sequence of non negative numbers such that $\varkappa_n/n \to 0$ and $\liminf_n \varkappa_n > 0$ (Example: $\varkappa_n = \text{constant} \cdot n^\gamma$ where $\gamma \in [0,1[$ and constant > 0). Put $C_n = \int_{n|x|>1} g(x)dx+(2/n)g(1/n)+\varkappa_n/n$.

Then C_n is the area of shaded region of this figure:



Define - for each n - a probability density - g_n by

$$g_n(x) = \begin{cases} g(x)/C_n & \text{when } |x| \geqq 1/n \\ [g(1/n)+\varkappa_n-n\varkappa_n|x|]/C_n & \text{when } |x| \leqq 1/n \end{cases}$$

Let G_n be the distribution with density g_n. Then G_n is unimodal and symmetric, $\mathcal{G}_{G_n}$ is differentiable and:

$$g_n'(x) = \begin{cases} g'(x)/C_n & \text{when} \quad |x| > 1/n \\ -n\kappa_n \operatorname{sgn} x/C_n & \text{when} \quad 0 < |x| < 1/n \end{cases}$$

It follows that:

$$g_n(0) = (g(1/n)+\kappa_n)/C_n$$

$$\int_{-\infty}^{+\infty} |g_n'(x)|dx = C_n^{-1} \int_{n|x|>1} |g'(x)|dx + 2C_n^{-1}\kappa_n$$

and it is easily seen that:

$$\lim_n C_n = 1 \quad ,$$

$$\lim_n g_n(x) = g(x) \quad \text{when} \quad x \neq 0 \quad ,$$

$$\liminf_n g_n(0) = g(0) + \liminf_n \kappa_n > g(0) \quad ,$$

and

$$\liminf_n \int_{-\infty}^{+\infty} |g_n'(x)|dx = \int |g'(x)|dx + 2 \liminf \kappa_n > \int |g'(x)|dx$$

By Scheffé's convergence theorem: $\|G_n - G\| \to 0$. In particular $G_n \to G$ weakly. The conclusion in theorem 7.35 (or proposition 7.34) is, however, not valid here since $g_n(0) \not\to g(0)$. It follows that the condition $G_n \in \mathcal{G}_0$, n= 1,2 ... (even when $G \in \mathcal{G}_0$) in theorem 7.35 (or proposition 7.34) can not be replaced by the condition that G_n ; n = 1,2,... are unimodal.

By the convergence criterion for translation experiments [16] we have $\Delta(\mathcal{G}_n, \mathcal{G}) \to 0$. We do not, however, have $\Delta(G_n, G) \to 0$ since $\int |g_n'(x)|dx \not\to \int |g'(x)|dx$.

We will now show that the metrics $\dot{\Delta}$ and Δ are topologically equivalent on $\mathcal{G}_0$.

Theorem 7.37.

Let $G, G_1, G_2, \ldots \in \mathcal{G}_0$. Then $\lim_{n\to\infty} \dot{\Delta}(G_n,G) = 0$ if and only if $\lim_{n\to\infty} \Delta(G_n,G) \to 0$.

Proof: The "only if follows from" theorem 7.32 and the "if" follows from theorem 7.35 and theorem 7.32. □

Finally we give a necessary condition for convergence which is valid without any condition on unimodality.

Theorem 7.38.

Let $\mathcal{E}_{G_n}$; n = 1,2... and $\mathcal{E}_G$ be differentiable and suppose $g_n' \to g'$ a.e. Lebesgue. Then

$$\lim_{n\to\infty} \dot{\Delta}(G_n,G) = 0$$

provided

$$\limsup_{n\to\infty} \int |g_n'(x)|dx \leqq \int |g'(x)|dx$$

Remark Conversely; $\dot{\Delta}(G_n,G) \implies$ - by theorem 5.1 (vi) - that $\int |g_n'(x)|dx \to \int |g(x)|dx$.

Proof of the theorem: By Scheffe's convergence theorem: $\int |g_n'(x)-g'(x)|dx \to 0$. Hence $\int |g_n(x)-g(x)| = |\int_{-\infty}^{x} (g_n'(t)-g'(t))dt| \leqq \int |g_n'(t)-g'(t)|dt \to 0$, so that

$$U_{G_n}(\xi) = \int_{-\infty}^{\infty} |\xi g_n(x)+g_n'(x)|dx \to \int_{-\infty}^{\infty} |\xi g(x)+g'(x)|dx = U_G(\xi) .$$

Convergence now follows from theorem 5.1 (vi). □

Appendix A

Comparison of translation experiments.

A summary of results.

Appendix A. Comparison of translation experiments.

A translation experiment will here be defined as an experiment $\mathcal{E}_P = ((\chi, \mathcal{A})(P_\theta : \theta \in \Theta))$ where χ is a second countable locally compact topological group with Borel class $\mathcal{A}$, $\Theta = \chi$, P is a probability measure on $\mathcal{A}$ and

$$P_\theta(A) = P(A\theta^{-1}); \; A \in \mathcal{A}, \; \theta \in \Theta$$

Clearly $\mathcal{E}_P$ is uniquely defined by P. μ will always denote a right Haar measure on $(\chi, \mathcal{A})$.

It will frequently be assumed that P is absolutely continuous i.e. that $P \ll \mu$. This assumption is equivalent with each of the following conditions:

(D_1) $\mathcal{E}_P$ is dominated

(D_2) $(P_\theta : \theta \in \Theta) \sim \mu$

(D_3) $\theta \mapsto P_\theta(A)$ is continuous for each $A \in \mathcal{A}$

(D_4) $\theta \mapsto P_\theta$ is strongly continuous

We summarize here some results form [16] on the comparison of translation experiments for LeCam's deficiency δ and distance Δ [7]

Theorem A.1.

Let P and Q be probability measures on $(\chi, \mathcal{A})$ and let $\epsilon \geq 0$ be a constant.

(i) If there is a probability measure M on $\mathcal{A}$ so that

$\|M * P - Q\| \leq \epsilon$ then $\delta(\mathcal{E}_P, \mathcal{E}_Q) \leq \epsilon$

(ii) Suppose* $M(\chi)$ has an invariant mean and that $P \ll \mu$. Then $\delta(\mathcal{E}_P, \mathcal{E}_Q) \leq \epsilon$ if and only if there is a probability measure M on $\mathcal{A}$ so that $\|M * P - Q\| \leq \epsilon$.

*) $M(\chi)$ is the space of bounded measurable functions on χ.

Corollary A.2.

$\delta(\mathcal{E}_P, \mathcal{E}_Q) \leq \inf_M \|M * P - Q\|$ and "=" holds if $P \ll \mu$ and $M(\chi)$ has an invariant mean.

Theorem A.3.

Let P and Q be probability measures on $(\chi,\mathcal{A})$ and $\epsilon \geq 0$ a constant. Then there exists a probability measure M so that

$$\|M * P - Q\| \leq \epsilon$$

if and only if

$$\int f dQ \leq \sup_x \int f(xy)P(dy) + \epsilon\|f\| , \quad f \in C(\chi).$$

We introduce now the notations:

$$\delta(P,Q) = \inf_M \|M*P-Q\| = \min_M \|M*P-Q\|$$

$$= \sup_{\|f\|\leq 1} (\inf_x \int f(xy)P(dy) - \inf_x \int f(xy)Q(dy))$$

$$= \sup_{\|f\|\leq 1} (\inf_x \int f(xy)P(dy) - Q(f))$$

$$\Delta(P,Q) = \delta(P,Q) \vee \delta(Q,P) =$$

$$= \sup_{\|f\|\leq 1} |\inf_x \int f(xy)P(dy) - \inf_x \int f(xy)Q(dy)|$$

Then $\delta(P,Q) = \delta(\mathcal{E}_P, \mathcal{E}_Q)$ $(\Delta(P,Q) = \Delta(\mathcal{E}_P, \mathcal{E}_Q))$ provided P is (P and Q are) absolutely continuous.

Theorem A.4.

Let P be absolutely continuous. Then $\Delta(P_n,P) \to$ if and only if there exist elements $a_1, a_2, \ldots$ in χ so that $\|\delta_{a_n} * P_n - P\| \to 0$.

Theorem A.5.

Suppose $\Delta(P_m P_n) \to 0$ as $m,n \to \infty$. Then there is a P so that $\Delta(P_n,P) \to 0$.

Appendix B

Comparison of pseudo experiments.

Content:

B.1 Introduction

In [7] Le Cam introduced the notion of ε-deficiency of one experiment relative to another. This generalized the concept of "being more informative" which was introduced by Bohnenblust, Shapley, and Sherman and may be found in Blackwell [1]. "Being more informative for k-decision problems" was introduced by Blackwell in [2]. The hybrid of "ε-deficiency for k-decision problems" was considered by the author in [15].

An experiment will here be defined as a pair $\mathcal{E} = ((\chi, \mathcal{A}), (P_\theta : \theta \in \Theta))$ where $(\chi, \mathcal{A})$ is a measurable space and $(P_\theta : \theta \in \Theta)$ is a family of probability measures on $(\chi, \mathcal{A})$. The set Θ -- the parameter set of $\mathcal{E}$ -- will be assumed fixed, but arbitrary.

Definition. Let $\mathcal{E} = ((\chi, \mathcal{A}), (P_\theta : \theta \in \Theta))$ and $\mathcal{F} = ((\mathcal{Y}, \mathcal{B}), (Q_\theta : \theta \in \Theta))$ be two experiments with the same parameter set Θ and let $\theta \to \varepsilon_\theta$ be a non-negative function on Θ (and let $k \geq 2$ be an integer).

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

$$r'(\theta) \leq r(\theta) + \varepsilon_\theta \|W_\theta\|, \quad \theta \in \Theta \quad \text{where } \|W_\theta\| = \sup_d |W_\theta(d)|; \ \theta \in \Theta$$

---

* When $k = 2$: testing problems.

** i.e., a measurable space.

*** It is always to be understood that $d \to W_\theta(d)$ is measureable for each θ.

Let $\mathcal{E} = ((\chi, \mathcal{A}), (P_\theta : \theta \in \Theta))$ and $\mathcal{F} = ((\mathcal{Y}, \mathcal{B}), (Q_\theta : \theta \in \Theta))$ be two experiments such that:

(i) $P_\theta : \theta \in \Theta$ is dominated

(ii) $\mathcal{Y}$ is a Borel-sub set of a Polish space and $\mathcal{B}$ is the class of Borel sub sets of $\mathcal{Y}$.

It follows from theorem 3 in Le Cam's paper [7] that $\mathcal{E}$ is ϵ-deficient w.r.t. $\mathcal{F}$ if and only if there is a randomization M from $(\chi, \mathcal{A})$ to $(\mathcal{Y}, \mathcal{B})$ so that $\|P_\theta M - Q_\theta\| \leq \epsilon_\theta$; $\theta \in \Theta$. (An alternative proof of this result is given in section 3)

Many of the results on comparison of experiments generalizes without difficulties to situations where the basic measures are only required to be finite. (Here as elsewhere in this paper a measure may be "non negative", "non positive" or neither. The notion of a signed measure will not be used.)

As an example of a situation where such "experiments" naturally enter consider two experiments $\mathcal{E} = ((\chi, \mathcal{A}); \mu_\theta : \theta \in \Theta)$ and $\mathcal{F} = ((\mathcal{Y}, \mathcal{B}), \nu_\theta : \theta \in \Theta)$, a decision space $(D, \mathcal{S})$, a loss function W and two functions a and b on Θ. Then we may ask: does there to any risk function s obtainable in $\mathcal{F}$ correspond a risk function r obtainable in $\mathcal{E}$ so that $r(\theta) \leq a_\theta\, s(\theta) + b_\theta \|W_\theta\|$; $\theta \in \Theta$? It turns out - under regularity conditions - that a necessary and sufficient condition is the existence of a randomization M from $(\chi, \mathcal{A})$ to $(\mathcal{Y}, \mathcal{B})$ so that $\|P_\theta M - a_\theta Q_\theta\| \leq b_\theta$; $\theta \in \Theta$. Considering $\theta \to a_\theta\, r(\theta)$ as a "risk function" relative to the "experiment" $((\mathcal{Y}, \mathcal{B}), (a_\theta Q_\theta; \theta \in \Theta))$ we see that this is essentially the criterion of theorem 3 in Le Cam's paper [7].

In this paper measures which are not probability measures are derived from probability measures by differentiation.

A pseudo experiment $\mathcal{E}$ will here be defined as a pair $\mathcal{E} = ((\chi, \mathcal{A}), \mu_\theta : \theta \in \Theta)$ where $(\mathcal{X}, \mathcal{A})$ is a measurable space and $\mu_\theta : \theta \in \Theta$ is a family of finite measures on $(\chi, \mathcal{A})$. We will stretch the usual terminology and call $(\chi, \mathcal{A})$ the sample space of $\mathcal{E}$ and Θ the parameter set of $\mathcal{E}$. A pseudo experiment with a two point parameter set will be called a pseudo dichotomy. An experiment (A dichotomy), $\mathcal{E}$, is a pseudo experiment (dichotomy) $\mathcal{E} = ((\chi, \mathcal{A}), \mu_\theta : \theta \in \Theta)$ where the measures $\mu_\theta : \theta \in \Theta$ are probability measures.

Some of the results on pseudo experiments are quite straight forward generalizations of those in [15]. This is, in particular, the case for most of the results included in this appendix. Other results, however, do not have the generalizations which may appear natural. As an example we mention the result (proved in [15]) that two experiments are equivalent provided they are equivalent for testing problems. We shall see in the next section that equivalence for testing problems does not - in general - imply equivalence for pseudo experiments.

The definition of ϵ-deficiency is extended as follows:

<u>Definition</u>. Let $\mathcal{E} = ((\chi, \mathcal{A}), (\mu_\theta : \theta \in \Theta))$ and $\mathcal{F} = ((\mathcal{Y}, \mathcal{B}), \nu_\theta : \theta \in \Theta))$ be pseudo experiments with the same parameter set Θ and let $\epsilon_\theta ; \theta \in \Theta$ be a function from Θ to $[0,\infty]$. We shall say that $\mathcal{E}$ is ϵ-deficient w.r.t. $\mathcal{F}$ (for k-decision problems if to each measurable space $(D, \mathcal{S})$ where $\# \mathcal{S} < \infty$ (where $\# \mathcal{S} = 2^k$), to each family $W_\theta : \theta \in \Theta$ of measurable functions on D, and each randomization σ from $(\mathcal{Y}, \mathcal{B})$ to $(D, \mathcal{S})$ there is a randomization ρ from $(\chi, \mathcal{A})$ to $(D, \mathcal{S})$ so that

$$\underline{W_\theta \rho \mu_\theta \leq W_\theta \sigma \nu_\theta + \epsilon_\theta \|W_\theta\|}; \quad \theta \in \Theta .$$

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

If $\mathcal{E} \geq \mathcal{F}$ ($\mathcal{E} \underset{k}{\geq} \mathcal{F}$) and $\mathcal{F} \geq \mathcal{E}$ ($\mathcal{F} \underset{k}{\geq} \mathcal{E}$) then we shall say that $\mathcal{E}$ and $\mathcal{F}$ are equivalent (for k-decision problems) and write this $\mathcal{E} \sim \mathcal{F}$ ($\mathcal{E} \underset{k}{\sim} \mathcal{F}$) . By proposition 8 in [15] and by weak compactness $\mathcal{E} \underset{2}{\sim} \mathcal{F} \iff \mathcal{E} \underset{3}{\sim} \mathcal{F} \iff \ldots \iff \mathcal{E} \sim \mathcal{F}$ provided $\mathcal{E}$ and $\mathcal{F}$ are dominated experiments.

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

The greatest lower bound of all constants ϵ such that $\mathcal{E}$ is ϵ-deficient relative to $\mathcal{F}$ will be denoted by $\delta(\mathcal{E},\mathcal{F})$ and max $[\delta(\mathcal{E},\mathcal{F}), \delta(\mathcal{F},\mathcal{E})]$ will be denoted by $\Delta(\mathcal{E},\mathcal{F})$.

<u>Proposition B.1.1</u> Let $\mathcal{E} = ((\chi,\mathcal{A}), (\mu_\theta : \theta \in \Theta))$ and $\mathcal{F} = ((\mathcal{Y},\mathcal{B}), (\nu_\theta : \theta \in \Theta))$ be two pseudo experiments, and let ϵ be a non negative function on θ . Then $\mathcal{E}$ is ϵ-deficient w.r.t. $\mathcal{F}$ for k decision problems provided $\mathcal{E}$ is ϵ deficient w.r.t. $\mathcal{F}$ for k+1 decision problems. If $\mathcal{E}$ is ϵ-deficient w.r.t. $\mathcal{F}$ for k decision problems, then $\epsilon_\theta \geq |\mu_\theta(\chi) - \nu_\theta(\mathcal{Y})|$. $\mathcal{E}$ is $\theta \rightsquigarrow |\mu_\theta(\chi) - \nu_\theta(\mathcal{Y})|$ deficient w.r.t. $\mathcal{F}$ for 1 decision problems and $\mathcal{E}$ is $\theta \to \|\mu_\theta\| + \|\nu_\theta\|$ deficient w.r.t. $\mathcal{F}$ for <u>k - decision problems for k = 1,2,... .</u>

<u>Proof:</u> Suppose $\mathcal{E}$ is ϵ-deficient w.r.t. $\mathcal{F}$ for k+1 decision problems. Put $D_k = \{1,2,\ldots,k\}$ and $D_{k+1} = \{1,2,\ldots,k+1\}$. Let $W_\theta : \theta \in \Theta$ be a family of functions on D_k and let σ be a randomization from $(\mathcal{Y},\mathcal{B})$ to D_k . Extend W_θ to D_{k+1} by writing

$W_\theta(k+1) = W_\theta(k)$. By assumption there is a randomization $\bar{\rho}$ from $(\chi,\mathcal{A})$ to D_{k+1} so that

$$\mu_\theta \bar{\rho} W_\theta \leq \nu_\theta \sigma W_\theta + \epsilon_\theta \|W_\theta\|$$

ϵ-deficiency for k-decision problems follows now since $\mu_\theta \bar{\rho} W_\theta = \mu_\theta \rho W_\theta$ where $\rho(k|x) = \bar{\rho}(k|x) + \bar{\rho}(k+1|x)$; $x \in \chi$ and $\rho(k'|x) = \bar{\rho}(k'|x)$; $k' \leq k$, $x \in \chi$.

Suppose $\mathcal{E}$ is ϵ-deficient w.r.t. $\mathcal{F}$ for k-decision problems. Inserting $W_\theta = 1$ and $W_\theta = -1$ in the inequalities appearing in the definitions of ϵ-deficiency we get; respectively $\epsilon_\theta \geq \mu_\theta(\chi) - \nu_\theta(\mathcal{Y})$ and $\epsilon_\theta \geq \nu_\theta(\mathcal{Y}) - \mu_\theta(\chi)$. Let $(D,\mathcal{S})$ be any measurable space and let σ and ρ be randomizations to $(D,\mathcal{S})$ from: respectively; $(\chi,\mathcal{A})$ and $(\mathcal{Y},\mathcal{B})$. Finally let $\{W_\theta\}$ be any family of (real valued) measureable functions on $(D,\mathcal{S})$.
Then:

$$\mu_\theta \rho W_\theta = \nu_\theta \sigma W_\theta + \mu_\theta \rho W_\theta - \nu_\theta \sigma W_\theta \leq \nu_\theta \sigma W_\theta + (\|\mu_\theta\| + \|\nu_\theta\|) \|W_\theta\| .$$

□

If $\mathcal{E}$, $\mathcal{F}$ and $\mathcal{G}$ are pseudo experiments
then:

$$\delta_k(\mathcal{E},\mathcal{G}) \leq \delta_k(\mathcal{E},\mathcal{F}) + \delta_k(\mathcal{F},\mathcal{G}) \quad ; k = 1,2,\ldots,$$

$$\Delta_k(\mathcal{E},\mathcal{G}) \leq \Delta_k(\mathcal{E},\mathcal{F}) + \Delta_k(\mathcal{F},\mathcal{G}) \quad ; k = 1,2,\ldots,$$

$$\delta_k(\mathcal{E},\mathcal{E}) = \Delta_k(\mathcal{E},\mathcal{E}) = 0 \quad ; k = 1,2,\ldots,$$

$$\Delta_k(\mathcal{E},\mathcal{F}) = \Delta_k(\mathcal{F},\mathcal{G}) \quad ; k = 1,2,\ldots,$$

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

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

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

$$\Delta(\mathcal{E},\mathcal{G}) \leq \Delta(\mathcal{E},\mathcal{F}) + \Delta(\mathcal{F},\mathcal{G}) \ ,$$

$$\delta(\mathcal{E},\mathcal{E}) = \Delta(\mathcal{E},\mathcal{E}) = 0 \ ,$$

$$\Delta(\mathcal{E},\mathcal{F}) = \Delta(\mathcal{F},\mathcal{E})$$

$$\delta_1(\mathcal{E},\mathcal{F}) = \Delta_1(\mathcal{E},\mathcal{F}) = \sup_\theta |\mu_\theta(\chi) - \nu_\theta(\mathcal{Y})| \ ,$$

and

$$\Delta(\mathcal{E},\mathcal{F}) \leq \sup_\theta (\|\mu_\theta\| + \|\nu_\theta\|) \ .$$

B.2 Finite parameter space

All pseudo experiments considered in this section are assumed to have the same finite parameter space Θ . $(D_k, \mathcal{S}_k)$; $k = 1,2,..$ will denote the decision space where $D_k = \{1,\dots,k\}$ and $\mathcal{S}_k$ is the class of subsets of D_k . If $\mathcal{E} = ((\chi,\mathcal{A}),(\mu_\theta:\theta\in\Theta))$ and ψ is a sub linear function on R^Θ then the integral $\int\psi(d\mu_\theta/d\sum_\theta|\mu_\theta|;\ \theta\in\Theta)d\sum_\theta|\mu_\theta|$ will be denoted by $\psi(\mathcal{E})$. If $\mathcal{E} = ((\chi,\mathcal{A}),(\mu_\theta:\theta\in\Theta))$ and $\mu_\theta(A) = \int f_\theta d\tau$; $A \in \mathcal{A}; \theta\in\Theta$ for some non negative measure τ on $\mathcal{A}$ then $\psi(\mathcal{E}) = \int\psi(f_\theta;\theta\in\Theta)d\tau$ for any sub linear function ψ on R^Θ .

Let $\mathcal{E} = ((\chi,\mathcal{A}),\ (\mu_\theta:\ \theta\in\Theta))$ and $\mathcal{F} = ((\mathcal{Y},\mathcal{B}),(\nu_\theta:\theta\in\Theta))$ be two pseudo experiments, and let ϵ be a function from Θ to $[0,\infty]$.

The basic result on ϵ-deficiency is:

Theorem B.2.1

The following conditions are all equivalent:

(i) $\mathcal{E}$ is ϵ-deficient w.r.t. $\mathcal{F}$ for k-decision problems

(ii) To each randomization σ from $(\mathcal{Y},\mathcal{B})$ to $(D_k,\mathcal{S}_k)$, and to each family $W_\theta:\theta\in\Theta$ of real valued functions on D_k corresponds a randomization ρ from $(\chi,\mathcal{A})$ to $(D_k, \mathcal{S}_k)$ so that:

$$\sum_\theta \mu_\theta \rho W_\theta \leq \sum_\theta \nu_\theta \sigma W_\theta + \sum_\theta \epsilon_\theta \|W_\theta\| .$$

(iii) To each randomization σ from $(\mathcal{Y},\mathcal{B})$ to $(D_k,\mathcal{S}_k)$ corresponds a randomization ρ from $(\chi,\mathcal{A})$ to $(D_k,\mathcal{S}_k)$ so that:

$$\|\mu_\theta\rho - \nu_\theta\sigma\| \leq \epsilon_\theta \ ; \ \theta\in\Theta$$

(iv)* $\psi(\mathcal{E}) \geq \psi(\mathcal{F}) - \sum_\theta \epsilon_\theta \max\{\psi(-e_\theta), \psi(e_\theta)\}$ for any sub linear function ψ on R^Θ which is the maximum of k homogenuous linear functions.

<u>Remark</u> If $\Delta_1(\mathcal{E},\mathcal{F}) = 0$ then (iv) is equivalent with:

(iv') $\psi(\mathcal{E}) \geq \psi(\mathcal{F}) - \frac{1}{2}\sum_\theta \epsilon_\theta(\psi(e_\theta) + \psi(-e_\theta))$ for any sub linear function ψ on R^Θ which is the maximum of k homogenous linear functions.

Demonstration: Clearly (iv') implies (iv) and (iv) for $x \rightsquigarrow \psi(x) - \frac{1}{2}\sum_\theta(\psi(e_\theta) - \psi(e_\theta))x_\theta$ impliea (iv') for ψ.

Note that the set of sub linear functions ψ which satisfies (iv') is a cone.

<u>Proof of the theorem</u>:

Suppose (ii) holds and let σ be a randomization from $(\mathcal{Y},\mathcal{B})$ to $(D_k,\mathcal{S}_k)$. Then:

$$\underset{W: \|W_\theta\| \leq 1;\, \theta\in\Theta}{\text{maximum}} \ \min_\rho \sum_\theta [\mu_\theta \rho W_\theta - \nu_\theta \sigma W_\theta - \epsilon_\theta \|W_\theta\|] \leq 0 .$$

It follows by weak compactness, - since $\sum_\theta$ is affine in ρ and concave in W - that maximum and minimum may be interchanged - i.e. ρ may be chosen independently of W. This implies $\|\mu_\theta\rho - \nu_\theta\sigma\| \leq \epsilon_\theta;\ \theta \in \Theta$.

Hence (ii) ==> (iii). It follows - since (iii) ==> (i) ==> (ii) is trivial - that (i) <==> (ii) <==> (iii). Interchanging W with $-W$ in (ii) we get:

$$\max_\rho \sum_\theta \mu_\theta \rho W_\theta \geq \max_\delta \sum_\theta \nu_\theta \sigma W_\theta - \sum_\theta \epsilon_\theta \|W_\theta\|$$

---

*) for each $\theta \in \Theta$ we define the vector e_θ by: $e_\theta(\theta') = 1$ or 0 as $\theta' = \theta$ or $\theta' \neq \theta$.

and this is (iv) for $\psi : x \rightsquigarrow \max_d \sum_\theta W_\theta(d) x_\theta$. □

An immediate consequence is:

Corollary B.2.2

$\mathcal{E}$ is ϵ-deficient w.r.t. $\mathcal{F}$ if and only if $\psi(\mathcal{E}) \geq \psi(\mathcal{F}) - \sum_\theta \epsilon_\theta \max\{\psi(e_\theta), \psi(e_\theta)\}$ for any sub linear function ψ on R^Θ .

Remark

If $\Delta_1(\mathcal{E}, \mathcal{F}) = 0$ then the inequality in corollary B.2.2 may be replaced by:

$$\psi(\mathcal{E}) \geq \psi(\mathcal{F}) - \tfrac{1}{2} \sum_\theta \epsilon_\theta (\psi(-e_\theta) + \psi(e_\theta))$$

Corollary B.2.3

Suppose $\Delta_1(\mathcal{E}, \mathcal{F}) = 0$. Then $\mathcal{E}$ is ϵ-deficient w.r.t. $\mathcal{F}$ for 2 decision problems if and only if

$$\|\sum_\theta a_\theta \mu_\theta\| \geq \|\sum_\theta a_\theta \nu_\theta\| - \sum_\theta \epsilon_\theta |a_\theta|$$

for any $a \in R^\Theta$.

Proof:

It suffices, in (iv') to consider functions ψ of the form $x \to |\sum_\theta a_\theta x_\theta|$.

□

Theorem B.2.4

Suppose $\Theta = \{1,2\}, \mu_1 \geq 0, \nu_1 \geq 0$ and that $\Delta_1(\mathcal{E}, \mathcal{F}) = 0$. Then $\mathcal{E}$ is ϵ-deficient w.r.t. $\mathcal{F}$ if and only if $\mathcal{E}$ is ϵ-deficient w.r.t. $\mathcal{F}$ for 2 decision problems.

Proof:

Suppose $\mathcal{E}$ is ϵ-deficient w.r.t. $\mathcal{F}$ for 2 decision problems.

Let $a_1,\dots,a_k,b_1,\dots,b_k$ be $2k$ constants and consider $\psi : x \to \max\{a_i x_1 + b_i x_2;\ i = 1,\dots,k\}$. By rearranging we may assume that there is a s so that

$$\psi(1,x_2) = \max\{a_i + b_i x_2;\ i = 1,2,\dots,s\} \qquad \text{where}$$

the representation on the right is minimal in the sense that for each $i \leq s$ there is a $x_2 > 0$ so that $a_i + b_i x_2 > \max\{a_j + b_j x_2;\ j \neq i,\ 1 \leq j \leq s\}$. Then the numbers $b_1,b_2,\dots,b_s$ are all distinct and we may without loss of generality - assume that $b_1 < b_2 < \dots < b_s$. It follows that $a_1 > a_2 > \dots > a_s$ and that $\psi(x) =$ or $\geq a_1x_1 + b_1x_2 + \sum_{i=2}^{s}(a_ix_1+b_ix_2-a_{i-1}x_1-b_{i-1}x_2)^+$ as $x_1 \geq 0$ or $x = -e_1,-e_2$. Put $\tilde{\psi}(x) = a_1x_1 + b_1x_2 + \sum_{i=2}^{s}(a_ix_1+b_ix_2-a_{i-1}x_1-b_{i-1}x_2)^+;\ x \in R^{\Theta}$.

Then - by the remark after theorem B.2.1 :

$$\psi(\mathcal{E}) = \tilde{\psi}(\mathcal{E}) \geq \tilde{\psi}(\mathcal{F}) - \tfrac{1}{2}\sum_{\theta} \epsilon_\theta (\tilde{\psi}(e_\theta) + \tilde{\psi}(-e_\theta)) =$$

$$\psi(\mathcal{F}) - \tfrac{1}{2}\sum_{\theta} \epsilon_\theta (\psi(e_\theta) + \tilde{\psi}(-e_\theta)) \geq \psi(\mathcal{F}) - \tfrac{1}{2}\sum_{\theta}\epsilon_\theta(\psi(e_\theta) + \psi(-e_\theta)) .$$

□

<u>Definitions</u>

A standard pseudo experiment is a pseudo experiment of the form $((K,\mathcal{B}),(S_\theta : \theta \in \Theta))$ where $K = \{x : x \in R^{\Theta} \text{ and } \sum_{\theta}|x_\theta| = 1\}$, $\mathcal{B}$ is the class of Borel sub sets of K and $x \leadsto x_\theta$ is - for each θ - a version of $dS_\theta / d\sum_{\theta}|S_\theta|$.

A finite non negative measure on K will* be called a standard measure.

If $\mathcal{E} = ((\chi,\mathcal{A}),(\mu_\theta : \theta \in \Theta))$ is a pseudo experiment then the standard pseudo experiment of $\mathcal{E}$ is the standard pseudo experi-

---

*) If A is some Borel sub set of a Polish space then "a measure on A" is - if not otherwice stated - synonymus with "a measure on the class of Borel sub sets of A".

ment

$$\hat{\mathcal{E}} = ((K, \mathcal{B}), (S_\theta : \theta \in \Theta))$$

where - for each θ - S_θ is the measure on K induced by the map : $x \to [d\mu_\theta / d \sum_\theta |\mu_\theta|]_x$; $\theta \in \Theta$ from $(\chi, \mathcal{A}, \mu_\theta)$ to K. The standard measure of the pseudo experiment $\mathcal{E} = ((\chi, \mathcal{A}), (\mu_\theta : \theta \in \Theta))$ is the standard measure induced by the map : $x \to [d\mu_\theta / d \sum_\theta |\mu_\theta|]_x$; $\theta \in \Theta$ from $(\chi, \mathcal{A}, \sum_\theta |\mu_\theta|)$ to K.

The standard measure of the standard pseudo experiment $((K, \mathcal{B}), (S_\theta : \theta \in \Theta))$ is the measure $\sum_\theta |S_\theta|$ and a standard pseudo experiment is determined by its standard measure. Any standard measure is the standard measure of a standard pseudo experiment. The standard measure of a pseudo experiment $\mathcal{E}$ is also the standard measure of its standard pseudo experiment $\hat{\mathcal{E}}$. Clearly $\hat{\hat{\mathcal{E}}} = \hat{\mathcal{E}}$ and $\Delta(\mathcal{E}, \hat{\mathcal{E}}) = 0$ for any pseudo experiment $\mathcal{E}$.

<u>Theorem B.2.5</u> $\Delta(\mathcal{E}, \mathcal{F}) = 0 \iff \hat{\mathcal{E}} = \hat{\mathcal{F}}$.

<u>Proof:</u>

$\Leftarrow$ is clear so suppose $\Delta(\mathcal{E}, \mathcal{F}) = 0$. We may without loss of generality assume that $\mathcal{E}$ and $\mathcal{F}$ are standard pseudo experiments with - respectively - standard measures S and T. 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 sub linear functions on R^Θ. It is easily seen that V is a vector lattice coutaining the constants. [If ψ_1, ψ_2 are real numbers then $|\psi_1 - \psi_2| = 2 \max\{\psi_1, \psi_2\} - (\psi_1 + \psi_2)$ - thus $|f| \in V$ when $f \in V$]. It follows from the formula $f^2 = \max_a 2a(f-a) + a^2$ that the closure $\bar{V}$ of V for uniform convergence is an algebra which obviously distinguish points in K.

Hence-by the Stone-Weierstrass approximation theorem - $\bar{V} = C(K)$. Clearly $S(f) = T(f)$ for any $f \in V$. It follows that $S(f) = T(f)$ when $f \in C(K)$ i.e. $S = T$. □

Example B.2.6

Suppose $\Theta = \{1,2\}$ Define standard probability measures S and T on K by:

$$S(\{(0,1)\}) = S(\{(1,0)\}) = S(\{(-\tfrac{1}{2},-\tfrac{1}{2})\})/2 = \frac{1}{4} = T(\{(\tfrac{1}{2},\tfrac{1}{2})\})/2 = T(\{(-1,0)\}) = T(\{(0,-1)\}) .$$

Let $\mathcal{E} = ((\chi, \mathcal{A}), (\mu_1, \mu_2))$ and $\mathcal{F} = ((\mathcal{Y}, \mathcal{B}), (\nu_1, \nu_2))$ be pseudo experiments with, respectively, standard measures S and T . Then:

$$\mu_i(\mathcal{X}) = \nu_i(\mathcal{Y}) = 0 \; ; \; i = 1,2$$

and

$$\int |ax_1+bx_2| S(dx) = |a|/4 + |b|/4 + |a+b|/4 = \int |ax_1+bx_2| T(dx)$$

It follows that $\Delta_2(\mathcal{E}, \mathcal{F}) = 0$. $\mathcal{E}$ and $\mathcal{F}$ are, however, not equivalent since:

$$\int \max\{x_1, x_2, 0\} S(dx) = \tfrac{1}{2}$$

and

$$\int \max\{x_1, x_2, 0\} T(dx) = \frac{1}{4}$$

so that $\Delta_3(\mathcal{E}, \mathcal{F}) \geq \delta_3(\mathcal{E}, \mathcal{F}) \geq \frac{1}{4}$.

It follows that equivalence for testing problems does not - even for pseudo dichotomies - imply equivalence. This demonstrates that

(i) the statement obtained from theorem B.2.4 by deleting the conditions $\mu_1 \geq 0$, $\nu_1 \geq 0$ is wrong.

and

(ii) Δ in theorem B.2.5 can not - even if we restrict ourselves to pseudo dichotomies - be replaced by Δ_2 .

If we restrict ourselves to experiments, however, then the conditions $\mu_1 \geq 0$, $\nu_1 \geq 0$ in theorem B.2.4 become superfluous and it was shown in [15] that Δ_2 equivalence for experiments implied Δ equivalence.

B.3 General parameter space

Problems on infinite parameter spaces may occasionally be reduced to problems on finite parameter spaces by:

Proposition B.3.1

Let $\mathcal{E} = (\chi, \mathcal{A}), (\mu_\theta : \theta \in \Theta))$ and $\mathcal{F} = ((\mathcal{Y}, \mathcal{B}), (\nu_\theta : \theta \in \Theta))$ where $|\mu_\theta| : \theta \in \Theta$ is dominated. Let ϵ be a non-negative function on Θ. Then $\mathcal{E}$ is ϵ-deficient w.r.t. $\mathcal{F}$ (for k-decision problems) if and only if $((\chi, \mathcal{A}), (\mu_\theta : \theta \in F))$ is $(\epsilon_\theta : \theta \in F)$ deficient w.r.t. $((\mathcal{Y}, \mathcal{B}), (\nu_\theta : \theta \in F))$ (for k-decision problems) for all finite non-empty sub sets F of Θ.

Proof:

The condition is clearly sufficient so suppose that the condition holds. It suffices to do the proof in the case of k-decision problems. Let D be a k-point set and let $\mathcal{S}$ be the class of all sub sets of D. Let σ be a randomization from $(\mathcal{Y}, \mathcal{B})$ to $(D, \mathcal{S})$. By assumption there is for each finite non-empty sub set F of Θ a randomization ρ^F from $(\chi, \mathcal{A})$ to $(D, \mathcal{S})$ so that

$$\|\mu_\theta \rho^F - \nu_\theta \sigma\| \leq \epsilon_\theta \; ; \; \theta \in F .$$

Let π be any probability measure dominating $|\mu_\theta| : \theta \in \Theta$. By weak compactness there is a sub set $\rho^{F'}$ and a ρ so that $\rho^{F'} \to \rho$ weakly $[L_1(\chi, \mathcal{A}, \pi)]$. It follows that

$$\|\mu_\theta \rho - \nu_\theta \sigma\| \leq \epsilon_\theta : \theta \in \Theta . \qquad \square$$

We proved in fact a little more and this is the content of the next theorem.

<u>Theorem B.3.2</u>

Let $\mathcal{E} = ((\chi,\mathcal{A}),(\mu_\theta:\theta\in\Theta))$ and $\mathcal{F} = ((\mathcal{Y},\mathcal{B});(\nu_\theta:\theta\in\Theta))$ where $\mu_\theta:\theta\in\Theta$ is dominated. Let ϵ be a non-negative function on Θ, let $\# D = k$ and let $\mathcal{S}$ be the class of all sub sets of D.

Then $\mathcal{E}$ is ϵ-deficient w.r.t. $\mathcal{F}$ for k-decision problems if and only if to each randomization σ from $(\mathcal{Y},\mathcal{B})$ to $(D,\mathcal{S})$ there is a randomization ρ from $(\chi,\mathcal{A})$ to $(D,\mathcal{S})$ so that:

$$\|\mu_\theta\rho - \nu_\theta\sigma\| \leq \epsilon_\theta;\ \theta\in\Theta .$$

The next proposition tells us -- in the case of experiments -- that certain decision spaces are abundant for comparision by operational characteristics.

<u>Proposition B.3.3</u>

Let $\mathcal{E} = ((\chi,\mathcal{A}),(\mu_\theta:\theta\in\Theta))$ and $\mathcal{F} = ((\mathcal{Y},\mathcal{B}),(\nu_\theta:\theta\in\Theta))$ be two pseudo experiments and let $\theta \to \epsilon_\theta$ be a non-negative function on Θ. Denote by T the collection of decision spaces $(D,\mathcal{S})$ having the following property:

To each randomization σ from $(\mathcal{Y},\mathcal{B})$ to $(D,\mathcal{S})$ there is a randomization ρ from $(\chi,\mathcal{A})$ to $(D,\mathcal{S})$ so that $\|\mu_\theta\rho - \nu_\theta\sigma\| \leq \epsilon_\theta;\ \theta\in\Theta$.

Then;

(i) If $(D,\mathcal{S})$ is in T and $\emptyset \subset S_0 \in \mathcal{S}$ then $(S_0, \mathcal{S}\cap S_0)$ is in T.

(ii) If $(D,\mathcal{S})$ is in T and $(D',\mathcal{S}')$ is a measurable space such that there exists a bimeasurable bijection $D \to D'$ then $(D',\mathcal{S}')$ is in T.

Proof:

(ii) is clear, so suppose $(D,\mathcal{S})$ is in T and $\emptyset \subset S_o \in \mathcal{S}$.

Let Γ be a probability measure on $(S_o, \mathcal{S} \cap S_o)$. Define a randomization $\gamma: (D,\mathcal{S}) \to (S_o, \mathcal{S} \cap S_o)$ by:

$$\gamma(S|d) = I_S(d);\ d \in S_o\ ,\ S \in \mathcal{S} \cap S_o$$

$$\gamma(S|d) = \Gamma(S)\ ;\ d \notin S_o\ ,\ S \in \mathcal{S} \cap S_o\ .$$

Let V be any probability measure on $(D,\mathcal{S})$ such that $V(S_o)=1$. Let $S \in \mathcal{S} \cap S_o$. Then: $(V\gamma)(S) = \int \gamma(S|\cdot)dV = \int_{S_o} \gamma(S|\cdot)dV = V(S)$. It follows that $V\gamma$ is the restriction of V to $\mathcal{S} \cap S_o$. Define a randomization $\tilde{\gamma}$ from $(S_o, \mathcal{S} \cap S_o)$ to $(D,\mathcal{S})$ by:

$$\tilde{\gamma}(S|d) = I_S(d);\ S \in \mathcal{S}\ ,\ d \in S_o$$

Then

$$(\gamma\tilde{\gamma})(S|d) = I_S(d);\ S \in \mathcal{S} \cap S_o,\ d \in S_o$$ and for any probability measure W on $S_o \cap \mathcal{S}$:

$$(W\tilde{\gamma})(S) = \int_{S_o} I_S\ dW = W(S \cap S_o);\ S \in \mathcal{S}\ .$$

Let σ be any randomization from $(\mathcal{Y}, \mathcal{B})$ to $(S_o, S_o \cap \mathcal{S})$. By assumption there is a randomization ρ from $(\chi,\mathcal{A})$ to $(D,\mathcal{S})$ so that:

$$\|\mu_\theta\rho - \nu_\theta\sigma\tilde{\gamma}\| \leq \epsilon_\theta;\ \theta \in \Theta\ .$$

Hence $\|\mu_\theta\rho - \nu_\theta\sigma\| \leq \epsilon_\theta;\ \theta \in \Theta$. □

Theorem B.3.4

Suppose $(|\mu_\theta| : \theta \in \Theta)$ is dominated. Then $\mathcal{E} = ((\chi,\mathcal{A});(\mu_\theta:\theta\in\Theta))$ is ϵ-deficient w.r.t. $\mathcal{F} = ((\mathcal{Y},\mathcal{B});\ (\nu_\theta : \theta \in \Theta))$ if and only if ; to each decision space $(D,\mathcal{S})$ where D is a Borel sub set of a Polish space and $\mathcal{S}$ is the class of Borel sub sets of D and to each randomization σ from $(\mathcal{Y},\mathcal{B})$ to $(D,\mathcal{S})$, there is a randomization ρ from $(\chi,\mathcal{A})$ to $(D,\mathcal{S})$ so that

$$\|\mu_\theta\rho - \nu_\theta\sigma\| \leq \epsilon_\theta;\ \theta \in \Theta .$$

If the condition is satisfied and at least one of the measures $\nu_\theta\sigma \neq 0$, then ρ may be chosen so that $\mu_\theta\rho$ is - for each θ - in the band generated by $\nu_\theta\sigma$: $\theta \in \Theta$.

Proof:

The condition is clearly sufficient, so suppose $\mathcal{E}$ is ϵ-deficient w.r.t. $\mathcal{F}$. By proposition B.3.3 we may - without loss of generality - assume that D is compact metric. Let π be a probability measure on $(\chi,\mathcal{A})$ which is equivalent with $(|\mu_\theta|:\theta\in\Theta)$ and let $\mathcal{H}$ be a countable dense sub set of $C(D)$ such that: r rational, $f,g \in \mathcal{H} \Rightarrow r, |f|, f+g$ and $rf \in \mathcal{H}$. [We may put $\mathcal{H} = \bigcup_{i=0}^{\infty} U_i$ where U_0 is a dense countable sub set of $C(D)$ and $U_1, U_2, \ldots$ are defined recursively by:

$$U_{i+1} = \{r_1f_1 + r_2f_2 + r_3 + f_3^+ : f_1,f_2,f_3 \in U_i;\ r_1,r_2 \text{ and } r_3 \text{ are rationals}\}].$$

Let $\{d_1,d_2,\ldots\}$ be dense in D. Put $D_k = \{d_1,d_2,\ldots,d_k\}$, and let $\mathcal{S}_k$ be the class of all subsets of D_k. For each k define $f_k: D \to D$ as follows: Let $d \in D$. Consider the k numbers: distance $(d,d_1),\ldots,$ distance (d,d_k). Let i be the unique integer among $\{1,\ldots,k\}$ such that:

distance $(d,d_1),\ldots,$ distance $(d,d_{i-1}) >$ distance $(d,d_i) \leq$ distance (d,d_{i+1}), distance $(d,d_{i+2}),\ldots,$ distance (d,d_k).

Define $f_k(d) = d_i$. Clearly f_k is measurable. Let σ be a randomization from $(\mathcal{Y},\mathcal{B})$ to $(D,\mathcal{S})$. Define the randomization σ_k from $(\mathcal{Y},\mathcal{B})$ to $(D_k,\mathcal{S}_k)$ by:

$$\sigma_k(\cdot|y) = \sigma(\cdot|y)f_k^{-1} .$$

By theorem B.3.2 there is a randomization ρ_k from $(\chi,\mathcal{A})$ to $(D_k, \mathcal{S}_k)$ so that:

$$\|\mu_\theta\rho_k - \nu_\theta\sigma_k\| \leq \epsilon_\theta; \ \theta \in \Theta .$$

For each $f \in \mathcal{H}$, $\sum_{i=1}^{k}\rho_k(d_i|\cdot)f(d_i)$; $k = 1,2,...$ has a weakly $(L_1(\mathcal{X},\mathcal{A},\pi))$ convergent sub sequence. By a diagonal process (or by Tychonoff's theorem) we may obtain a sub sequence $\rho_{k''}$ so that $\sum_{i=1}^{k'}\rho_{k'}(d|\cdot)f(d_i)$ converges weakly to a function $\rho(f|\cdot)$, for each $f \in \mathcal{H}$. ρ may be modified so that:

$$\rho(f+g|\cdot) = \rho(f|\cdot) + \rho(g|\cdot); \ f,g \in \mathcal{H}$$

$$\rho(rf|\cdot) = r\rho(f|\cdot) \quad f \in \mathcal{H}$$

$$\rho(1|\cdot) = 1$$

$$\rho(f|\cdot) \geq 0 ; \quad f \in \mathcal{H}, \ f \geq 0 .$$

By continuity - there is for each $x \in \chi$ - a probability measure $\bar{\rho}(\cdot|x)$ on $\mathcal{S}$ so that $\bar{\rho}(f|x) = \rho(f|x)$; $f \in \mathcal{H}$. Since $\bar{\rho}(f|x)$ is measureable for each $f \in \mathcal{H}$, $\bar{\rho}$ defines a randomization from $(\chi,\mathcal{A})$ to $(D,\mathcal{S})$. Let $f \in \mathcal{H}$.
Then:

$$\left|\int fd(\mu_\theta\bar{\rho}) - \int fd(\nu_\theta\sigma)\right| \leq \left|\int fd(\mu_\theta\bar{\rho}) - \sum_{i=1}^{k} f(d_i)(\mu_\theta\rho_{k'})(d_i)\right| +$$

$$\left|\sum_{i=1}^{k}(\mu_\theta\rho_{k'})(d_i)f(d_i) - \sum_{i=1}^{k}(\nu_\theta\sigma_{k'})(d_i)f(d_i)\right| +$$

$$\left|\sum_{i=1}^{k}(\nu_\theta\sigma_{k'})(d_i)f(d_i) - \int fd(\nu_\theta\sigma)\right| .$$

Since, $\|\mu_\theta\rho_k - \nu_\theta\sigma_k\| \leq \epsilon_\theta$; $\theta \in \Theta$, the second term to the right of $\leq$ is $\leq \epsilon_\theta\|f\|$. Since distance $(d,f_k(d))$ = distance $(d,\{d_1,\ldots,d_k\})\downarrow 0$ and D is compact - distance $(d,f_k(d))\downarrow 0$ uniformly in d . Hence - since f is uniformly continuous - $\|f\circ f_k - f\| \to 0$. The last term may be written

$$\sum_{i=1}^{k} f(d_i)(\nu_\theta\sigma_{k'})(d_i) = \int (f\circ f_{k'})d(\nu_\theta\sigma) .$$

It follows that the last term $\to 0$.

The first term to the right of $\leq$ which may be written

$$|\int [\int f(d')\rho(dd'|\cdot) - \sum_{i=1}^{k} f(d_i)\rho_{k'}(d_i|\cdot)]d\mu_\theta|$$

tends - by weak convergence - to 0 .

It follows that

$$\|\mu_\theta\bar{\rho} - \nu_\theta\sigma\| \leq \epsilon_\theta; \ \theta \in \Theta .$$

Let us - finally - return to the general case and suppose ρ is a randomization such that $\|\mu_\theta\rho - \nu_\theta\sigma\| \leq \epsilon_\theta$; $\theta \in \Theta$. Let τ be a probability measure on $(D,\mathcal{S})$ which is equivalent with $\mu_\theta\rho$; $\theta \in \Theta$ and let for each finite measure $\varkappa$ on $\mathcal{S}$, $\varkappa'$ be the projection of $\varkappa$ on the band generated by $\nu_\theta\sigma$: $\theta \in \Theta$. Let π be a probability measure in the band generated by $\nu_\theta\sigma$: $\theta \in \Theta$. Then the map $\varphi: \varkappa \to \varkappa' + [\varkappa(D) - \varkappa'(D)]\pi$ maps $L_1(\tau)$ into $L_1(\tau\varphi)$. The restriction of φ to $L_1(\tau)$ may be represented by a randomization φ from $(D,\mathcal{S})$ to $(D,\mathcal{S})$. It follows that $\|\mu_\theta\rho\varphi - \nu_\theta\sigma\| = \|(\mu_\theta\rho - \nu_\theta\sigma)\varphi\| \leq \|\mu_\theta\rho - \nu_\theta\sigma\| \leq \epsilon_\theta$ and $\mu_\theta\rho\varphi$ is in the band generated by $\nu_\theta\sigma$; $\theta \in \Theta$. □

Corollary B.3.5

Let $\mathcal{E} = ((\chi,\mathcal{A});(\mu_\theta:\theta \in \Theta))$ and $\mathcal{F} = ((\mathcal{Y},\mathcal{B}),(\nu_\theta:\theta \in \Theta))$ be two pseudo experiments where $(|\mu_\theta|:\theta \in \Theta)$ is dominated and $\mathcal{Y}$ is a Borel sub set of a Polish space and $\mathcal{B}$ is the class of Borel sub sets of $\mathcal{Y}$. Let ϵ be a non-negative function on Θ .

Then:

(i) $\mathcal{E}$ is ϵ-deficient w.r.t. $\mathcal{F}$ if and only if there is a randomization M from $(\chi,\mathcal{A})$ to $(\mathcal{Y},\mathcal{B})$ so that:

$$\|\mu_\theta M - \nu_\theta\| \leq \epsilon_\theta; \ \theta \in \Theta$$

If the condition is satisfied and $\nu_\theta \neq 0$ for at least one θ, then M may be chosen so that $\mu_\theta M$ is - for each θ - in the band generated by ν_θ: $\theta \in \Theta$.

(ii) $\mathcal{E}$ is ϵ-deficient w.r.t. $\mathcal{F}$ if and only if to each decision space $(D, \mathcal{S})$ and to each randomization σ from $(\mathcal{Y}, \mathcal{B})$ to $(D, \mathcal{S})$ there is a randomization ρ from $(\chi, \mathcal{A})$ to $(D, \mathcal{S})$ so that:

$$\|\mu_\theta \rho - \nu_\theta \sigma\| \leq \epsilon_\theta; \ \theta \in \Theta$$

Remark.

If $\mu_\theta : \theta \in \Theta$ and $\nu_\theta : \theta \in \Theta$ are probability measures then (i) is a direct consequence of theorem 3 in LeCam's paper [7].

Proof of the Corollary.

1° Suppose $\mathcal{E}$ is ϵ-deficient w.r.t. $\mathcal{F}$. Consider the decision space $(D, \mathcal{S}) = (\mathcal{Y}, \mathcal{B})$ and the identity map σ from $\mathcal{Y}$ to $\mathcal{Y}$. By theorem 7 there is a randomization M from $(\chi, \mathcal{A})$ to $(\mathcal{Y}, \mathcal{B})$ so that

$$\|\mu_\theta M - \nu_\theta\| \leq \epsilon_\theta; \ \theta \in \Theta$$

The last statement in (i) follows from the last statement in theorem B.3.4.

2° Assume there is a randomization M from $(\chi, \mathcal{A})$ to $(\mathcal{Y}, \mathcal{B})$ so that $\|\mu_\theta M - \nu_\theta\| \leq \epsilon_\theta; \ \theta \in \Theta$. Let $(D, \mathcal{S})$ be any decision space and σ a randomization from $(\mathcal{Y}, \mathcal{B})$ to $(D, \mathcal{S})$. Then:

$$\|\mu_\theta M\sigma - \nu_\theta \sigma\| \leq \epsilon_\theta : \ \theta \in \Theta . \qquad \square$$

The next proposition generalized Corollary 6 in [15].

Proposition B.3.6

Let $\mathcal{E} = ((\chi, \mathcal{A}), (\mu_\theta : \theta \in \Theta))$ and $\mathcal{F} = ((\mathcal{Y}, \mathcal{B}), (\nu_\theta : \theta \in \Theta))$ be two pseudo experiments and let $\theta \to \epsilon_\theta$ be a non-negative function on Θ. Suppose $(|\mu_\theta| \theta \in \Theta)$ is dominated. Then $\mathcal{E}$ is ϵ-deficient w.r.t. $\mathcal{F}$ for k-decision problems if and only if $\mathcal{E}$ is ϵ-deficient w.r.t. each experiment $((\mathcal{Y}, \tilde{\mathcal{B}}), (\mu_\theta | \tilde{\mathcal{B}} : \theta \in \Theta))$ where $\tilde{\mathcal{B}} \subseteq \mathcal{B}$ and $\# \tilde{\mathcal{B}} \leq 2^k$.

Proof:

1° Suppose $\mathcal{E}$ is ϵ-deficient w.r.t. $\mathcal{F}$ for k-decision problems and that $\tilde{\mathcal{B}}$ is a sub algebra of $\mathcal{B}$ containing at most 2^k sets. Clearly $\mathcal{E}$ is ϵ-deficient w.r.t. $\tilde{\mathcal{F}} = ((\mathcal{Y}, \tilde{\mathcal{B}});$ $(Q_\theta | \tilde{\mathcal{B}} : \theta \in \Theta))$ for k-decision problems. Consider the decision space $(\mathcal{Y}, \tilde{\mathcal{B}})$ and let σ be the identity map from $(\mathcal{Y}, \mathcal{B})$ to $(\mathcal{Y}, \tilde{\mathcal{B}})$. By theorem B.3.4 there is a randomization ρ from $(\mathcal{X}, \mathcal{A})$ to $(\mathcal{Y}, \tilde{\mathcal{B}})$ so that:

$$\|\mu_\theta \rho - (\nu_\theta | \tilde{\mathcal{B}})\sigma\| \leq \epsilon_\theta : \theta \in \Theta$$

or - since $(\nu_\theta | \tilde{\mathcal{B}})\sigma = \nu_\theta | \tilde{\mathcal{B}} : \theta \in \Theta$ - :

$$\|\mu_\theta \rho - \nu_\theta | \tilde{\mathcal{B}}\| \leq \epsilon_\theta ; \ \theta \in \Theta .$$

By corollary B.3.5 this implies that $\mathcal{E}$ is ϵ-deficient w.r.t. $\tilde{\mathcal{F}}$.

2° Suppose $\mathcal{E}$ is ϵ-deficient w.r.t. each experiment $((\mathcal{Y}, \tilde{\mathcal{B}}), (\nu_\theta | \tilde{\mathcal{B}} : \theta \in \Theta))$. We may - without loss of generality - assume $\# \Theta < \infty$. The proposition now follows from theorem B.2.1 in section 2 in the same way as corollary 6 in [15] followed from theorem 2 in [15].

□

Appendix C

Arguments depending on an assumption stating that some of the measurable spaces involved are Borel sub sets of Polish spaces.

Appendix C Arguments depending on an assumption stating that some of the measurable spaces involved are Borel sub sets of Polish spaces.

The only results whose proofs depend on such assumptions are:

Proposition	2.3	Page 2.6
"	3.1	3.1
"	3.4	3.6
"	4.11	4.12
Theorem	6.1	6.2
"	6.2	6.6
Corollary	6.3	6.9
Proposition	6.5	6.11
Theorem	6.6	6.11

We shall now show how these assumptions may be avoided in proposition 2.3, proposition 3.1 and in proposition 3.4.

Proof of proposition 2.3

Let θ_n be any sequence in $\Theta - \{\theta_o\}$ such that $\theta_n \to \theta_o$. By the testing criterion - corollary B.2.3 -

$$\left\|\frac{Q_{\theta_m} - Q_{\theta_m}}{\theta_m - \theta_o} - \frac{Q_{\theta_n} - Q_{\theta_o}}{\theta_n - \theta_o}\right\| \leq \left\|\frac{P_{\theta_m} - P_{\theta_o}}{\theta_m - \theta_o} - \frac{P_{\theta_n} - P_{\theta_o}}{\theta_n - \theta_o}\right\| .$$

The right hand side of this inequality tends - since $\mathcal{E}$ is differentiable in θ_o - to zero as $m,n \to \infty$.

It follows that $\frac{Q_{\theta_n} - Q_{\theta_o}}{\theta_n - \theta_o}$; $n = 1, \ldots$ is a Cauchy sequence.

□

Proof of proposition 3.1

Let a and b be real numbers. By corollary B.2.3:

$$\left\|a \frac{P_\theta - P_{\theta_o}}{\theta - \theta_o} + b\, P_{\theta_o}\right\| \geq \left\|a \frac{Q_\theta - Q_{\theta_o}}{\theta - \theta_o} + b\, Q_{\theta_o}\right\|$$

$\theta \to \theta_o$ yields:

$$\|a\, \dot{P}_{\theta_o} + b\, P_{\theta_o}\| \geq \|a\, \dot{Q}_{\theta_o} + b\, Q_{\theta_o}\|$$

so that - by corollary B.2.3 again - $\dot{\mathcal{E}}_{\theta_o} \geq \dot{\mathcal{F}}_{\theta_o}$ □

Proof of proposition 3.4

By assumption $\delta_1(\mathcal{G}, \mathcal{G}_{\pi,\sigma}) = 0$ so that-using the formula for δ_1 in B.1 - $\mu(\mathcal{Y}) = \pi(\mathcal{X}) = 1$ and $\nu(\mathcal{Y}) = H(\chi) = 0$. Hence - by corollary B.2.3 - $\|a\mu + b\nu\| \leq \|a\pi + bH\|$ for all real numbers a and b. $a = 1$ and $b = 0$ yields $\|\mu\| \leq \|\pi\| = 1$. On the other hand $\|\mu\| \geq \mu(\mathcal{Y}) = 1$ so that $\|\mu\| = 1$. It follows that $\|\mu^+\| + \|\mu^-\| = \|\mu\| = 1 = \mu(\mathcal{Y}) = \|\mu^+\| - \|\mu^-\|$ so that $\|\mu^-\| = 0$. Hence μ is a probability measure and it remains to show that $\mu \gg \nu$. Decompose $\nu = \nu_1 + \nu_2$ where $\nu_1 \gg \mu$ and $\nu_2 \perp \mu$. Then:

$$\|\xi\mu - \nu_1\| + \|\nu_2\| = \|\xi\mu - \nu\| \leq \|\xi\pi - \sigma\|$$

so that

$$(\S) \quad \|\nu_2\| \leq [\|\xi\pi - \sigma\| - |\xi|] - [\|\xi\mu - \nu_1\| - |\xi|]$$

Put $g = d\nu_1/d\mu$. Then we may write

$$\|\xi\mu - \nu_1\| - |\xi| = \|(\xi\mu - \nu_1)^+\| + \|(\xi\mu - \nu_1)^-\| - |\xi|$$

$$= \|(\xi\mu - \nu_1)^+\| - \|(\xi\mu - \nu_1)^-\| + 2\|(\xi\mu - \nu_1)^-\| - |\xi|$$

$$= (\xi\mu - \nu_1)(\mathcal{Y}) + 2\int(\xi - g)^- d\mu - |\xi|$$

$$= \xi - \nu_1(\mathcal{Y}) + 2\int_{g>\xi} (g - \xi)\,d\mu - |\xi|$$

$$= \begin{cases} 2 \int_{g>\xi} (g-\xi)d\mu - \nu_1(\mathcal{Y}) & \text{when } \xi \geq 0 \\ 2 \int_{g>\xi} (g-\xi)d\mu - \nu_1(\mathcal{Y}) + 2\xi & \text{when } \xi \leq 0 \end{cases}$$

$$= \begin{cases} -\nu_1(\mathcal{Y}) + a_\xi & \text{when } \xi \geq 0 \quad \text{where } a_\xi \to 0 \text{ as } \xi \to \infty \\ 2\int g d\mu - \int_{g \leq \xi} g d\mu + 2\xi\, \mu(g \leq \xi) - \nu_1(\mathcal{Y}) & \text{when } \xi \leq 0 \end{cases}$$

$$= \begin{cases} -\nu_1(\mathcal{Y}) + a_\xi & \text{when } \xi \geq 0 \\ \nu_1(\mathcal{Y}) + b_\xi & \text{when } \xi \leq 0 \quad \text{where } b_\xi \to 0 \text{ as } \xi \to -\infty \end{cases}$$

Similarily - using that $\pi >> \sigma$ - we get:

$$\lim_{|\xi| \to \infty} [\|\xi\| - \sigma\| - |\xi| = 0] = 0$$

$\xi \to \infty$ in (§) yields $\|\nu_2\| \leq \nu_1(\mathcal{Y})$ while $\xi \to -\infty$ in (§) yields $\|\nu_2\| \leq -\nu_1(\mathcal{Y})$. Hence $\nu_1(\mathcal{Y}) = 0 = \|\nu_2\|$ so that $\nu << \mu$. □

The missing assumption in proposition 4.11, theorem 6.1, theorem 6.2, proposition 6.5 and theorem 6.6 is: $\tilde{\chi}$ is a Borel sub set of a Polish space and $\tilde{\mathcal{A}}$ is the class of Borel sub sets of $\tilde{\chi}$. This assumption may be avoided in proposition 4.11 by dropping condition (iii).

Finally the proof of corollary 6.3 requires not only this assumption but also the same assumption on $(\chi, \mathcal{A})$.

References

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] Boll, C. (1955). Comparison of experiments in the infinite case. Ph.D. thesis Stanford University.

[4] Hájek and Šidák (1967). Theory of rank tests. Academic Press and Academia Publishing house of the Czechoslovak academy of sciences.

[5] Heyer (1969). Erschöpfheit und Invarianz beim Vergleich von Experimenten. Z. Wahrscheinlichkeitstheorie verw. Geb. 12, 21-55.

[6] Ibragimov, J.A. (1956). On the composition of unimodal distributions (Russian). Teoriya veroyatnostey 1. 283-288.

[7] LeCam (1964). Sufficiency and approximate sufficiency. Ann. Math. Statist. 35, 1419-1455.

[8] LeCam (1968). Theorie asymptotique de la decision statistique. Seminaire de mathematique supérieures. Les presses de l'université de Montréal.

[9] LeCam (1971). Limits of experiments. Sixth Berkeley Symposium on mathematical statistics and probability.

[10] Lehmann, E.L. (1959). Testing statistical hypotheses. New York, Wiley.

[11] Scheffé, H. (1947). A usefull convergence theorem for probability distributions. Ann. Math. Statist. 18, 434-438.

[12] Strassen, V. (1965). The existence of probability measures with given marginals. Ann. Math. Statist. 36, 423-439.

[13] Torgersen, E.N. (1965). Minimalsufficiency of the order statistic in the case of translation and scale parameters. Skandinavisk Aktuarietidskrift, 16-21.

[14] Torgersen, E.N. (1969). On ε-comparison of experiments. Unpublished manuscript. Abstract. Ann. Math. Statist. (1969) 40, 2219.

[15] Torgersen, E.N. (1970). Comparison of experiments when the parameter space is finite. Z. Wahrscheinlichkeitstheorie verw. Geb. 16, 219-249.

[16] Torgersen, E.N. (To appear in the Annals of Math. Statist, approximately october, december 1972). Comparison of translation experiments.