STATISTICAL RESEARCH REPORT No 5
Institute of Mathematics 1972
University of Oslo

LOCAL COMPARISON OF EXPERIMENTS

by

Erik N. Torgersen

ABSTRACT

LOCAL COMPARISON OF EXPERIMENTS

by

Erik N. Torgersen

University of Oslo

In this paper we generalizes most of the results in Torgersen, E.N. [Local comparison of experiments when the parameter set is one dimensional. Statistical Research Report no 4, 1972 from Department of Mathematics, University of Oslo. Abstract to appear in Ann. Math. Statist.] to the case of a finite dimensional parameter set.

CONTENTS

1. Introduction

In this paper most of the results in "Local comparison of experiments when the parameter set is one dimensional", Statistical Research Report no 4, 1972 - from here on denoted by LC1 - are generalized to the case of a finite dimensional parameter set Θ . The paper is organized so that section i ; $i = 2,3,4,5,6$; here generalizes section i in LC1. We refer to section 1 in LC1 for an introduction to the type of problems treated in this paper. Occasionally the general case may be reduced to the one dimensional case by considering directional derivatives, i.e. linear combinations of partial derivatives.

Local comparison will here - as it was in LC1 - be expressed in terms of pseudo experiments; i.e. "experiments" where the basic measures are not required to be probability measures. The results on pseudo experiments which was given in LC1 are not sufficient for our needs here. Appendix B in LC1 is therefor extended and included as an appendix to this paper.

Differentiable experiments are defined in section 2. It is shown that products of differentiable experiments are differentiable and that sub experiments of differentiable experiments are differentiable.

The derivative in $\theta^{o} \in \Theta$ of a differentiable experiment consists essentially of the probability distribution in θ^{o} together with all its partial derivatives in θ^{o} . Simple necessary and sufficient conditions on a pseudo experiment for being a derivative are given in section 3. It is shown how derivatives may be identified with probability distributions on R^{Θ} with expectation $(0,\ldots,0)$. This representation converts products into convolutions.

Differentiated deficiencies and distances are treated in sections 4 and 5. The asymptotic factorization criterion for sufficiency in LC1 is generalized and various convergence - and compactness criterions are given.

The local information in products of a large number of uniformly uninformative experiments would be the object of a central limit theorem. We have not formulated such a theorem here but it will be clear from proposition 3.3, example 4.2 and theorem 5.1 that it reduces to the central limit theorem in the case of central variables with uniformly bounded first order absolute moments.

The statistical motivation for the theory developed so far is given in section 6. Let δ_ϵ be the ϵ-difenciency within an ϵ sphere around θ^0 for the norm $x \rightsquigarrow \sum_\theta |x_\theta|$. It is shown that $\delta_\epsilon/2\epsilon$ converges as $\epsilon \to 0$ to the "differentiated" deficiency $\dot{\delta}$ introduced in section 4. The conditional expectation criterions for sufficiency are generalized.

The list of references should be considered as the combined reference list for LC1 and this paper. We have kept the numbering from LC1.

2. The differentiability conditions.

All experiments considered in this paper have - unless otherwise stated - a parameter set Θ , which is a sub set of R^r having an interior point θ^o . We shall say that $\mathcal{E} = ((\chi,\mathcal{A});P_\theta : \theta \in \Theta)$ is differntiable in θ^o if there are finite measures $\dot{P}_{\theta^o,1}$, $\dot{P}_{\theta^o,2}$,... $\dot{P}_{\theta^o,r}$ so that *)

$$\lim_{\theta\to\theta_o} \|P_\theta - P_{\theta^o} - \sum_{i=1}^{r} (\theta_i-\theta_i^o)\dot{P}_{\theta^o,i}\|/\|\theta-\theta^o\| = 0$$

Writing $\Gamma_{\theta^o,\theta} = P_\theta - P_{\theta_o} - \sum_i(\theta_i-\theta_i^o)\dot{P}_{\theta^o,i}$ when $\theta \neq \theta^o$ we see that the differentiability condition for $\mathcal{E}$ may be rewritten as:

$\mathcal{E}$ is differentiable in θ^o if and only if there are finite measures $\dot{P}_{\theta^o,1}$, $\dot{P}_{\theta^o,2}$,...,$\dot{P}_{\theta^o,r}$ and $\Gamma_{\theta^o,\theta}$; $\theta \neq \theta_o$ so that $\lim_{\theta\to\theta^o} \|\Gamma_{\theta^o,\theta}\| = 0$ and $P_\theta = P_{\theta^o} + \sum_{i=1}^{r} (\theta_i-\theta_i^o)\dot{P}_{\theta^o,i} + \|\theta-\theta_o\|\Gamma_{\theta_o,\theta}$

$\|\Gamma_{\theta^o,\theta}\|$; $\theta \in \Theta$ are - by the inequality:

$\|\theta-\theta^o\| \; \|\Gamma_{\theta^o,\theta}\| \leqq 2 + \sum_{i=1}^{r} |\theta_i-\theta_i^o| \; \|\dot{P}_{\theta^o,i}\|$ - automatically bounded.

The measures $\dot{P}_{\theta^o,i}$; $i = 1,2...r$ are determined by $\mathcal{E}$ since

$$\left\| \frac{P_{\theta^o+hv_i} - P_{\theta^o}}{h} - \dot{P}_{\theta^o,i} \right\| \to 0 \quad \text{as} \quad h \to 0$$

where $v_i = \overset{(1)}{(0},0,\ldots,\overset{(i)}{1},\ldots 0,\overset{(r)}{0)}$. If $\mathcal{E}$ is differentiable and I

---

*) If $v = (v_1,\ldots,v_r) \in R^r$ then $\|v\| = (v_1{}^2+\ldots+v_r{}^2)^{\frac{1}{2}}$

is a proper sub set of $\{1,2,\dots,r\}$, then the experiment obtained from $P_\theta : \theta \in \Theta$ by putting $\theta_i = \theta_i^o$; $i \in I$ is also differentiable and the partial derivative in $(\theta_i^o ; i \notin I)$ w.r.t. θ_j is $\dot{P}_{\theta^o,j}$. The differentiability of $\mathcal{E}$ in θ^o implies the differentiability of the experiment $(\chi, \mathcal{A}, P_\theta : (\theta,\eta) \in \Theta \times R^t)$ in any point (θ^o,η) ; $\eta \in R^t$.

Suppose now that there is a measure μ so that each measure P_θ - with θ restricted to some neighbourhood of θ^o - has a density f_θ w.r.t. μ. The following conditions are - together - sufficient for differentiability in θ_o.

(i) $f_\theta(x)$ is, for each $x \in \chi$, differntiable in θ (i.e. the partial derivatives exist and are continuous). Denote by $\dot{f}_{\theta,i}(x)$ the partial derivative of $\theta \rightsquigarrow f_\theta(x)$ w.r.t. θ_i.

(ii) $\int \sup_\theta |\dot{f}_{\theta,i}(x)| \mu(dx) < \infty$

Demonstration:

Define: $\dot{P}_{\theta^o,i}(A) = \int_A \dot{f}_{\theta^o,i}(x)\mu(dx)$, $A \in \mathcal{A}$, $i = 1,\dots,r$.

Then $\|\dot{P}_{\theta^o,i}\| < \infty$ $i = 1,2\dots r$ and

$$\|P_\theta - P_{\theta^o} - \sum_{i=1}^r (\theta_i - \theta_i^o)\dot{P}_{\theta^o,i}\| / \|\theta - \theta^o\|$$

$$\leq \int |f_\theta(x) - f_{\theta^o}(x) - \sum_{i=1}^r (\theta_i - \theta_i^o)\dot{f}_{\theta^o,i}(x)| / \max_i |\theta_i - \theta_i^o|$$

$$= \int |\sum_{i=1}^r (\theta_i - \theta_i^o)(\dot{f}_{\theta^*,i}(x)) - \dot{f}_{\theta^o,i}(x))| / \max_i |\theta_i - \theta_i^o|$$

where $\theta^* = \theta^*(x)$ is on the line segment joining θ^o and θ.

Hence:

$$\|P_\theta - P_{\theta^o} - \sum_{i=1}^{r} (\theta_i-\theta_i^o)\dot{P}_{\theta^o,i}\|/\|\theta-\theta^o\| \leqq$$

$$\sum_i \int \sup\{ |\dot{f}_{\bar{\theta},i}(x) - \dot{f}_{\theta^o,i}(x)| : \|\bar{\theta}-\theta^o\| \leqq \|\theta-\theta^o\|\}\mu(dx) \to 0 \quad \text{as} \quad \theta \to \theta^o$$

by the dominated convergence criterion. These conditions are in particular satisfied when f_θ is of the form $c(\theta)h_\theta$ where $c(\theta)$ does not depend on x and h_θ satisfies (i) and (ii).

Example 2.1 (Exponential family).

Suppose the density f_θ og P_θ w.r.t. μ may be written in the form

$$f_\theta(x) = c(\theta)e^{\sum_i \theta_i T_i(x)}$$

for all θ belonging to some neighbourhood of θ^o .

Then $h_\theta(x) = e^{\sum_i \theta_i T_i(x)}$ satisfies (i) and (ii). It follows that $\mathcal{E}$ is differntiable and that

$$d\dot{P}_{\theta^o,i}/d\mu = \dot{f}_{\theta^o,i}(x) = \dot{c}_i(\theta)e^{\sum_i \theta_i T_i(x)} + c(\theta)T_i(x)e^{\sum_i \theta_i T_i(x)}$$

$$= f_{\theta^o}(T_i - E_{\theta^o}T_i) .$$

Propositions 2.2 and 2.3 below state, respectively, that products of differentiable experiment and sub experiments of differentiable experiments are differentiable.

Proposition 2.2

Let $\mathcal{E}_i = (\chi_i, \mathcal{A}_i, P_\theta^{(i)} ; \theta \in \Theta)$, $i = 1,\dots,n$ be differentiable in θ^o. Then $\prod_{i=1}^{n} \mathcal{E}_i$ is differentiable in θ^o and *)

$$\lim_{h\to o}(\prod_i P^{(i)}_{\theta^o+hv_j} - \prod P^{(i)}_{\theta^o})/h$$

$$= P^{(1)}_{\theta^o} \times\dots\times P^{(n-1)}_{\theta^o} \times \dot{P}^{(n)}_{\theta^o,j} + \dots + \dot{P}^{(1)}_{\theta^o,j} \times\dots\times P^{(n-1)}_{\theta^o} \times P^{(n)}_{\theta_o}$$

Proof: Very similar to that of proposition 2.1 in LC1 □

Proposition 2.3

Let $\mathcal{E} = (\chi, \mathcal{A} ; P_\theta : \theta \in \Theta)$ 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 $(\chi, \mathcal{B} ; P_{\theta\mathcal{B}} : \theta \in \Theta)$ is differentiable in θ^o and

$$\lim_{h\to o}(P_{\theta^o+hv_j,\mathcal{B}} - P_{\theta^o,\mathcal{B}})/h = \dot{P}_{\theta^o,j,\mathcal{B}}$$

where $\dot{P}_{\theta^o j,\mathcal{B}}$ is the restriction of $\dot{P}_{\theta^o,j}$ to $\mathcal{B}$

Proof: Very similar to that of proposition 2.2 in LC1 □

Proposition 2.3 is a particular case of:

---

*) v_j is the j-th unit vector in R^r, i.e. $v_j(i) = 1$ or 0 as $i = j$ or $i \neq j$.

Proposition 2.4

If $\mathcal{E} \underset{(2)}{\geqq} \mathcal{F}$ and $\mathcal{E}$ is differentiable in θ^o then $\mathcal{F}$ is differentiable in θ^o .

Proof: Write $\mathcal{E} = ((\chi,\mathcal{A}),P_\theta : \theta \in \Theta)$ and $\mathcal{F} = ((\mathcal{Y},\mathcal{B}), Q_\theta : \theta \in \Theta)$. It follows from proposition 2.3 in LC1 that $\| Q_{\theta^o+hv_j} - Q_{\theta^o})/h - \dot{Q}_{\theta^o,j} \| \to 0$ as $h \to 0$. By the testing criterion (theorem 10 in [15]) we have:

$\|P_\theta - P_{\theta^o} - \Sigma(\theta_i-\theta_i^o)(P_{\theta^o+hv_i} - P_{\theta^o})/h\| \geqq$ same expression in $Q_\theta : \theta \in \Theta$. $h \to 0$ yields:

$\|P_\theta - P_{\theta^o} - \Sigma(\theta_i-\theta_i^o)\dot{P}_{\theta^o,i} \| \geqq$ same expression in $Q_\theta : \theta \in \Theta$. □

Corollary 2.5

A product of experiments is differentiable in θ^o if and only each factor is differentiable in θ^o .

---

3. Basic properties of the derivative.

We define the derivative of a differentiable (in θ^o) experiment $\mathcal{E} = (\chi, \mathcal{A}, P_\theta : \theta \in \Theta)$ as the pseudo experiment

$$\dot{\mathcal{E}}_{\theta_o} \overset{\text{definition}}{=\!=} (\chi, \mathcal{A}, P_{\theta_o}, \dot{P}_{\theta_o,1}, \ldots, \dot{P}_{\theta_o,r})$$

where *) $\dot{P}_{\theta^o,i} = \lim_{h\to o}(P_{\theta_o+hv_i} - P_{\theta_o})/h$; $i = 1,2\ldots r$.

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

Proposition 3.1

Let $\mathcal{E} = ((\chi, \mathcal{A}), (P_\theta : \theta \in \Theta))$ and $\mathcal{F} = ((\mathcal{Y}, \mathcal{B}), (Q_\theta : \theta \in \Theta))$ be differentiable in θ_o . Then $\dot{\mathcal{E}}_{\theta_o} \underset{(k)}{\geqq} \dot{\mathcal{F}}_{\theta_o}$ provided $\mathcal{E} \underset{(k)}{\geqq} \mathcal{F}$.

In particular $\dot{\mathcal{E}}_{\theta_o} \sim \dot{\mathcal{F}}_{\theta_o}$ when $\mathcal{E} \sim \mathcal{F}$.

Remark. We have always $\dot{\mathcal{E}}_{\theta_o} \underset{1}{\sim} \dot{\mathcal{F}}_{\theta_o}$ and by theorem B.2.7 in the appendix $\dot{\mathcal{E}}_{\theta_o} \underset{2}{\sim} \dot{\mathcal{F}}_{\theta_o} \Leftrightarrow \dot{\mathcal{E}}_{\theta_o} \sim \dot{\mathcal{F}}_{\theta_o}$.

Proof: Suppose $\mathcal{E} \underset{(k)}{\geqq} \mathcal{F}$. Write $\dot{P}_{\theta^o,i} = \lim_{h\to o}(P_{\theta_o+hv_i} - P_{\theta_o})/h$ and $\dot{Q}_{\theta^o,i} = \lim_{h\to o}(Q_{\theta_o+hv_i} - Q_{\theta_o})/h$. Put $s_i = d\dot{P}_{\theta^o,i}/dP_{\theta^o}$ and

---

*) v_i denotes – troughout this paper – the i-th unit vector in R^r ; i.e. $v_j(j) = 1$ or 0 as $j = i$ or $j \neq i$.

$t_i = d\dot{Q}_{\theta^o,i} \Big/ d\dot{Q}_{\theta_o}$. Let $\psi \in \psi_k$ on R^{r+1} . We must show that

$$\int \psi(1,s,\ldots,s_r)dP_{\theta_o} \geqq \int \psi(1,t,\ldots,t_r)dQ_{\theta_o} .$$

Write $\psi(x) = \max_{1 \leqq i \leqq k} \sum_{j=o}^{r} a_{ij}x_j$; $x \in R^{r+1}$

Then $\int \psi(1,s,\ldots,s_r)dP_{\theta_o}$

$$= \int [\max_i(a_{io} + \sum_{j=1}^{r} a_{ij}s_j)]dP_{\theta_o}$$

$$= \|\bigvee_i(a_{io} + \sum_{j=1}^{r} a_{ij}\dot{P}_{\theta_o,j})\|$$

$$= \lim_{h \to o} \|\bigvee_i[a_{io} + \sum_{j=1}^{r} a_{ij}(P_{\theta_o}+hv_j-P_{\theta_o})/h]\|$$

It suffices therefor to show that the expression after the "lim" is $\geqq$ the same expression in $Q_\theta : \theta \in \Theta$. This - however - is an immediate consequence of the sub linear function criterion (theorem 2 in [15] or theorem B.2.1 in the appendix). □

The derivatives are characterized in

Theorem 3.2

A pseudo experiment $\mathcal{D} = ((\chi,\mathcal{A}),\pi,\sigma_1,\ldots,\sigma_r)$ is the derivative in θ^o of some differentiable experiment $\mathcal{E}$, if and only if, $\sigma_1(\chi)=\ldots=\sigma_r(\chi) = 0$ and π is a probability measure dominating $\sigma_1,\sigma_2,\ldots,\sigma_r$. If so, then $((\chi,\mathcal{A}),\pi,\sigma_1,\ldots,\sigma_r)$ is the derivative in θ^o of the experiment $\mathcal{E} = ((\chi,\mathcal{A}),P_\theta : \theta \in \Theta)$ where $\theta^o \in \Theta^o$ and

$$P_\theta = |\pi + \sum_i(\theta_i-\theta_i^o)\sigma_i|/\|\pi + \sum_i(\theta_i-\theta_i^o)\sigma_i\| ; \theta \in \Theta .$$

Furthermore these conditions imply that

$(\|\pi + \sum_i(\theta_i-\theta_i^o)\sigma_i\| - 1)/\|\theta-\theta^o\| \to 0$ as $\theta \to \theta^o$.

<u>Remark:</u> P_θ is well defined since $\|\pi + \sum_i(\theta_i-\theta_i^0)\sigma_i\| \geqq 1$.

<u>Proof:</u> We may - without loss of generality - assume that $\theta^0 = 0$. The conditions are obviously necessary so suppose $\sigma_1(\chi)=\ldots=\sigma_r(\chi)=0$ and that π is a probability measure dominating $\sigma_1,\ldots,\sigma_r$. Put $s_i = d\sigma_i/d\pi$. Then

$$\|\pi + \Sigma\theta_i s_i\| - 1 = \int |1 + \Sigma\theta_i s_i|d\pi - 1$$

$$= \int (1 + \Sigma\theta_i s_i)^+ d\pi + \int (1 + \Sigma\theta_i s_i)^- d\pi - \int (1 + \Sigma\theta_i s_i)^-$$

$$+ \int (1 + \Sigma\theta_i s_i)^- d\pi - 1 = \int (1 + \Sigma\theta_i s_i)d\pi + 2\int (1 + \Sigma\theta_i s_i)^- d\pi - 1$$

$$= 2\int (1 + \Sigma\theta_i s_i)^- d\pi \leqq 2 \int_{\Sigma\theta_i s_i < -1} (1 + \Sigma|\theta_i||s_i|d\pi$$

$$\leqq 2 \int_{\Sigma|\theta_i|\max_i|s_i|>1} [1 + (\Sigma|\theta_i|)\max_i|s_i|]d\pi$$

$$\leqq 4\,\Sigma|\theta_i| \int_{\max_i|s_i|>(\Sigma|\theta_i|)^{-1}} \max_i|s_i|d\pi \quad .$$

Hence:

$(\|\pi + \Sigma\theta_i\sigma_i\| - 1)/\Sigma|\theta_i| \to 0$ as $\Sigma|\theta_i| \to 0$

and this is equivalent with the last statement. Clearly $P_0 = \pi$. It remains to show that $\|P_\theta - \pi - \Sigma\theta_i\sigma_i\|/\|\theta\| \to 0$ as $\|\theta\| \to 0$.

We get successively:

$$\|\theta\|^{-1}\|P_\theta-\pi-\Sigma\theta_i\sigma_i\| = \|\theta\|^{-1}\|\frac{|\pi+\Sigma\theta_i\sigma_i|}{\|\pi+\Sigma\theta_i\sigma_i\|} - \pi-\Sigma\theta_i\sigma_i\|$$

$$\leqq \|\theta\|^{-1} \|\frac{|\pi+\Sigma\theta_i\sigma_i|}{\|\pi+\Sigma\theta_i\sigma_i\|} - |\pi+\Sigma\theta_i\sigma_i|\|$$

$$+ \|\theta\|^{-1} \| |\pi+\Sigma_i\theta_i\sigma_i| - \pi - \Sigma\theta_i\sigma_i\|$$

$$= \|\theta\|^{-1} \|\pi+\Sigma\theta_i\sigma_i\|(1 - \frac{1}{\|\pi+\Sigma\theta_i\sigma_i\|})$$

$$+ 2\|\theta\|^{-1} \|(\pi+\Sigma\theta_i\sigma_i)^-\|$$

$$= \|\theta\|^{-1}[\|\pi+\Sigma\theta_i\sigma_i\| - 1] + \|\theta\|^{-1}[\|\pi+\Sigma\theta_i\sigma_i\| - 1]$$

$$= 2\|\theta\|^{-1}[\|\pi+\Sigma\theta_i\sigma_i\| - 1] \to 0 \quad \text{as} \quad \|\theta\| \to 0 \qquad \square$$

The pseudo experiment $((\chi,\mathcal{A}),\pi_1\sigma,\ldots,\sigma_r)$ where $\sigma_1(\chi)=\ldots=\sigma_2(\chi) = 0$ and π is a probability measure dominating $\sigma_1,\ldots,\sigma_r$ will be written $\mathcal{E}_{\pi,\sigma}$ where σ denotes $(\sigma_1,\ldots \sigma_r)$.

The standard representation of $\mathcal{E}_{\pi,\sigma}$ is also derivative. If $s_i = d\sigma_i/d\pi$; $i = 1,2\ldots r$ then the standard representation is $\mathcal{E}_{S_o,S}$ where $S_o = \pi((1+\Sigma|s_i|)^{-1}, s_i(1+\Sigma|s_i|)^{-1}$; $i = 1,\ldots,r)^{-1}$

and $S_i = \sigma_i((1+\Sigma|s_i|)^{-1}, s_i(1+\Sigma|s_i|)^{-1}$, $i = 1,\ldots,r)^{-1}$; $i=1,\ldots r$.

A closely associated characteristic is the standard measure

$$S = S_o + \sum_{i=1}^{r} |S_i| .$$

Alternatively we may - since S and πs^{-1} - where $s = (s_1, \ldots, s_r)$ - determines each other, use πs^{-1} as a characteristic. The measure πs^{-1} will occasionally be denoted by $F_{\pi,\sigma}$.

Let - for each $i = 1,2,\ldots,r$ - $G_{\pi,\sigma,i}$ be the measure on R^r whose Radon Nikodym derivative is the i-th coordinate function $x \rightsquigarrow x_i$. It will follow from proposition 3.4 that

$$((R^r, \text{Borel class}), F_{\pi,\sigma}, G_{\pi,\sigma,1}, \ldots, G_{\pi,\sigma,r})$$

is a derivative.

Furthermore - since $x \rightsquigarrow x_i$ is a version of $dG_{\pi,\sigma,i}/dF_{\pi,\sigma}$ - this derivative is equivalent with $\mathcal{E}_{\pi,\sigma}$. It may be checked that $F_{\pi,\sigma}$ is, and may be any probability distribution on R^r such that $\int x_i F_{\pi,\sigma}(dx) = 0$, $i = 1,\ldots r$. We will occasionally write $F_{\theta_0}, \mathcal{E}$ instead of $F_{\pi,\sigma}$ when $\mathcal{E}_{\pi,\sigma} = \dot{\mathcal{E}}_{\theta_0}$. One pleasant property of this characteristic is:

<u>Proposition 3.3</u>

Let $\mathcal{E}_1, \mathcal{E}_2, \ldots, \mathcal{E}_n$ be differentiable in θ_0. Then:

$$F_{\theta_0}, \prod_i \mathcal{E}_i = F_{\theta_0}, \mathcal{E}_1 * F_{\theta_0}, \mathcal{E}_2 * \ldots * F_{\theta_0}, \mathcal{E}_n$$

where (*) means convolution.

<u>Proof</u>: The proof is very similar to that of proposition 3.3 in LC1. □

The fact that the standard representation of $\mathcal{E}_{\pi,\sigma}$ as well as $((R^r, \text{Borel class}), F_{\pi,\sigma}, G_{\pi,\sigma,1}, \ldots, G_{\pi,\sigma,r})$ are derivatives is a consequence of

Proposition 3.4

If $\mathcal{E} = ((\mathcal{Y},\mathcal{B}),\mu,\nu_1,\ldots,\nu_r) \leqq \mathcal{E}_{\pi,\sigma}$ then $\mathcal{E}$ is also a derivative.

Proof: If suffices to show that each pseudo dichotomy $((\mathcal{Y},\mathcal{B}),\mu,\nu_i)$; $i = 1,2,\ldots,r$ is a derivative and this is a consequence proposition 3.4 in LC1. □

4. Comparison of derivatives.

In this - and the next section - derivatives will be written $\mathcal{E}_{\pi,\sigma_1,\dots,\sigma_r} = ((\chi,\mathcal{A}),\pi,\sigma_1,\dots,\sigma_r)$ with or without affixes. The following notations relative to the derivative $\mathcal{E}_{\pi,\sigma_1,\dots,\sigma_r} = ((\chi,\mathcal{A}),\pi,\sigma_1,\dots,\sigma_r)$ will be used:

$S \overset{\text{def}}{=}$ the standard measure of $\mathcal{E}_{\pi\,\sigma_1\dots\sigma_r}$

$\sigma \overset{\text{def}}{=} (\sigma_1,\dots,\sigma_r)$

$s_i \overset{\text{def}}{=} d\sigma_i / d\pi$; $i = 1,2\,\dots,r$.

$s \overset{\text{def}}{=} (s_1,\dots,s_r)$

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

$N(\xi) \overset{\text{def}}{=} \|\pi+\xi_1\sigma_1+\dots+\xi_r\sigma_r\|$; $(\xi_1,\dots,\xi_r) \in R^r$

$V \overset{\text{def}}{=} \{(\int\delta d\pi,\int\delta d\sigma_1,\dots,\int\delta d\sigma_r) : 0 \leqq \delta \leqq 1\}$

$\psi(\mathcal{E}) \overset{\text{def}}{=} \int\psi(1,s_1,\dots,s_r)d\pi$ where ψ is a sub linear function on R^{r+1} .

Affixes on $\mathcal{E},\pi,\sigma_1,\dots,\sigma_r$, $\sigma,\chi,\mathcal{A},s_1,\dots,s_r$, s, F, N and V ; when these are referring to the same derivative will be of the same type.

For two derivatives $\mathcal{E}$, $\tilde{\mathcal{E}}$ we will - for each $k = 2,3,\dots$ - write:

$\dot{\delta}_k(\mathcal{E},\tilde{\mathcal{E}}) \overset{\text{def}}{=}$ the smallest $\epsilon/2$ so that $\mathcal{E}$ is $(0,\epsilon,\dots,\epsilon)$ deficient w.r.t. $\tilde{\mathcal{E}}$ for k-decision problems

and $\dot{\Delta}_k(\mathcal{E},\tilde{\mathcal{E}}) \overset{\text{def}}{=} \max(\dot{\delta}_k(\mathcal{E},\tilde{\mathcal{E}}),\ \dot{\delta}_k(\tilde{\mathcal{E}},\mathcal{E}))$.

The smallest $\epsilon/2$ so that $\mathcal{E}$ is $(0,\epsilon,\ldots,\epsilon)$ deficient w.r.t. $\tilde{\mathcal{E}}$ will be denoted by $\dot{\delta}(\mathcal{E},\tilde{\mathcal{E}})$ and we shall write $\dot{\Delta}(\mathcal{E},\tilde{\mathcal{E}}) = \max(\dot{\delta}(\mathcal{E},\tilde{\mathcal{E}}), \delta(\tilde{\mathcal{E}},\mathcal{E}))$.

It follows directly from the definition that:

$$\dot{\delta}_k \uparrow \dot{\delta} \quad \text{and} \quad \dot{\Delta}_k \uparrow \dot{\Delta} \quad \text{as} \quad k \to \infty$$

$$0 \leqq \dot{\delta}_{(k)}(\mathcal{E},\tilde{\mathcal{E}}) \leqq \dot{\Delta}_{(k)}(\mathcal{E},\tilde{\mathcal{E}}) < \infty$$

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

$$\delta_{(k)}(\mathcal{E},\hat{\mathcal{E}}) \leqq \delta_{(k)}(\mathcal{E},\tilde{\mathcal{E}}) + \delta_{(k)}(\tilde{\mathcal{E}},\hat{\mathcal{E}})$$

$\dot{\Delta}_2,\dot{\Delta}_3,\ldots$ and Δ are pseudo metrics.

$$\delta_{(k)} \leqq 2\,\dot{\delta}_{(k)}$$

and $\Delta_{(k)} \leqq 2\,\dot{\Delta}_{(k)}$

It follows directly from the definition that $\Delta_{(k)}=0 \Leftrightarrow \dot{\Delta}_{(k)}=0$ and it is a consequence of theorem B.2.7 in the appendix that $\Delta_2,\ldots,\Delta$, $\dot{\Delta}_2,\ldots,\dot{\Delta}$ induces the same equivalence "relations". Trivially $\Delta_1(\mathcal{E},\tilde{\mathcal{E}}) = \dot{\Delta}_1(\mathcal{E},\tilde{\mathcal{E}}) = 0$ for any pair $(\mathcal{E},\tilde{\mathcal{E}})$ of derivatives.

Let $\hat{\Psi}(\hat{\Psi}_k)$ denote the set of sub linear functions on R^{r+1} (which are a maximum of k linear functionals) satisfying $\psi(e_i) = \psi(-e_i)$; $i = 1,2\ldots,r$ where $e_i(j) = 1$ or 0 as $j = i$ or $j \neq i$.

Proposition 4.1

Compact expressions for the quantities $\delta_2, \Delta_2, \dot{\delta}_2, \dot{\Delta}_2, \ldots, \delta, \Delta, \dot{\delta}$ and $\dot{\Delta}$ are *)

$$\delta_2(\mathcal{E},\tilde{\mathcal{E}}) = \sup_{\xi}[(\tilde{N}(\xi)-N(\xi))^{(+)}/(1+\Sigma|\xi_i|)]$$

$$\Delta_2(\mathcal{E},\tilde{\mathcal{E}}) = \sup_{\xi}[|\tilde{N}(\xi)-N(\xi)|/(1+\Sigma|\xi_i|)]$$

$$2\dot{\delta}_2(\mathcal{E},\tilde{\mathcal{E}}) = \sup_{\xi\neq 0}[(\tilde{N}(\xi)-N(\xi))^{(+)}/\Sigma|\xi_i|]$$

$$2\dot{\Delta}_2(\mathcal{E},\tilde{\mathcal{E}}) = \sup_{\xi\neq 0}[|\tilde{N}(\xi)-N(\xi)|/\Sigma|\xi_i|]$$

$$\delta_{(k)}(\mathcal{E},\tilde{\mathcal{E}}) = \sup\ \{(\psi(\tilde{\mathcal{E}})-\psi(\mathcal{E}))^{(+)}/\sum_{i=0}^{r}\psi(e_i)\ :\ 0\neq\psi\in\hat{\Psi}_{(k)}\}$$

$$\Delta_{(k)}(\mathcal{E},\tilde{\mathcal{E}}) = \sup\ \{|\psi(\tilde{\mathcal{E}})-\psi(\mathcal{E})|/\sum_{i=0}^{r}\psi(e_i)\ :\ 0\neq\psi\in\hat{\Psi}_{(k)}\}$$

$$2\dot{\delta}_{(k)}(\mathcal{E},\tilde{\mathcal{E}}) = \sup\{(\psi(\tilde{\mathcal{E}})-\psi(\mathcal{E}))^{(+)}/\sum_{i=1}^{r}\psi(e_i):\psi\in\hat{\Psi}_{(k)};\ \sum_{i=1}^{r}\psi(e_i)>0\}$$

$$2\dot{\Delta}_{(k)}(\mathcal{E},\tilde{\mathcal{E}}) = \sup\{|\psi(\tilde{\mathcal{E}})-\psi(\mathcal{E})|/\sum_{i=1}^{r}\psi(e_i):\psi\in\hat{\Psi}_{(k)};\ \sum_{i=1}^{r}\psi(e_i)>0\}$$

Upper bounds for $\dot{\delta}_2$ is provided by: For any $\varkappa > 0$:

$$\dot{\delta}_2(\mathcal{E},\tilde{\mathcal{E}}) \leqq \max\{\tfrac{1+\varkappa}{2\varkappa}\delta_2(\mathcal{E},\tilde{\mathcal{E}}),4(\bigvee_i|\tilde{\sigma}_i|)([\max_i|\tilde{s}_i|\geqq\varkappa^{-1}])\}$$

---

*) When the symbol "(+)" appears in one of these formulas, then it may either be deleted or be replaced by "+".

Proof 1^o: The formulas with "(+)" replaced by "+" follows directly from theorem B.2.1 in the appendix. That "(+)" may be deleted in the formulas for δ_2, $\delta_{(k)}$, $\dot{\delta}_2$ and $\dot{\delta}_{(k)}$ follows from the last statement in theorem 3.2 .

2^o: Let $\Sigma|\xi_i| > \varkappa$. Then:

$$\Sigma|\xi_i| > \frac{\varkappa}{1+\varkappa}(1 + \Sigma|\xi_i|)$$

so that:

$$\frac{1}{2}\frac{(\tilde{N}(\xi)-N(\xi))^+}{\Sigma|\xi_i|} \leqq \frac{1+\varkappa}{2\varkappa}\frac{\tilde{N}(\xi)-N(\xi)}{(1+\Sigma|\xi_i|)} \leqq \frac{1+\varkappa}{2\varkappa}\delta_2(\mathcal{E},\tilde{\mathcal{E}}) .$$

It follows from the proof of theorem 3.2 that

$$(\tilde{N}(\xi)-1)/\Sigma|\xi_i| \leqq 4 \int_{\max_i|\tilde{s}_i| \geqq (\Sigma|\xi_i|)^{-1}} \max_i|\tilde{s}_i|\, d\tilde{\pi} \leqq 4 \int_{\max_i|\tilde{s}_i| \geqq \varkappa^{-1}} \max_i|\tilde{s}_i|\, d\tilde{\pi} \quad \text{when}$$

$$\Sigma|\xi_i| \leqq \varkappa .$$

Hence:

$$\frac{1}{2}\frac{(\tilde{N}(\xi)-N(\xi))^+}{\Sigma|\xi_i|} \leqq \frac{1}{2}\frac{\tilde{N}(\xi)-1}{\Sigma|\xi_i|} \leqq 4 \int_{\max_i|\tilde{s}_i| \geqq \varkappa^{-1}} \max_i|\tilde{s}_i|\,d\tilde{\pi} \quad \text{when} \quad \Sigma|\xi_i| \leqq \varkappa .$$

□

Example 4.2 Let us compare the derivatives $\mathcal{E} = ((\chi,\mathcal{A}), \pi, \sigma_1,\ldots,\sigma_r)$ with "the minimum informative" derivative $\tilde{\mathcal{E}} = ((\tilde{\chi},\tilde{\mathcal{A}}), \tilde{\pi}, 0,\ldots,0)$.

Clearly $\mathcal{E} \geqq \tilde{\mathcal{E}}$. By proposition 3.1:

$$\Delta_2(\mathcal{E},\tilde{\mathcal{E}}) \geqq \frac{\|\pi+\xi_i\sigma_i\|}{1+|\xi_i|} \to \|\sigma_i\| \quad \text{as} \quad |\xi_i| \to \infty$$

Hence

$$\Delta_2(\mathcal{E},\tilde{\mathcal{E}}) \geqq \max_i \|\sigma_i\| .$$

Let $\psi \in \hat{\Psi}$ satisfy: $\sum_i \psi(e_i) > 0$. Then

$$\psi(\mathcal{E})-\psi(\tilde{\mathcal{E}}) = \int (\psi(1,s_1,\ldots,s_r)-\psi(1,0,\ldots,0))d\pi$$

$$\leqq \int \sum_{i=1}^{r} \psi(e_i)|s_i| d\pi = \sum_{i=1}^{r} \psi(e_i)\|\sigma_i\| \leqq \left(\sum_{i=1}^{r} \psi(e_i)\right)\max_i \|\sigma_i\|$$

By proposition 4.1 again: $2\dot{\Delta}(\mathcal{E},\tilde{\mathcal{E}}) \leqq \max_i \|\sigma_i\|$. It follows that:

$$\Delta_2(\mathcal{E},\tilde{\mathcal{E}}) = 2\dot{\Delta}_2(\mathcal{E},\tilde{\mathcal{E}}) = \ldots = \Delta(\mathcal{E},\tilde{\mathcal{E}}) = 2\dot{\Delta}(\mathcal{E},\tilde{\mathcal{E}}) = \max_i \|\sigma_i\| .$$

V is a compact convex subset of the cube: $[0,1] \times \prod_i [-\|\sigma_i\|/2, \|\sigma_i\|/2]$. It is symmetric about $(\frac{1}{2},0,\ldots,0)$ and the only points with first coordinates 0 or 1 are the points $(0,0,\ldots,0)$ and $(1,0,\ldots 0)$. V determines $\mathcal{E}$ up to equivalence since (i): N does and (ii): the support function of V and N determines each other.

$\mathcal{E}$ is $(\epsilon_0,\epsilon_1,\ldots,\epsilon_r)$ deficient w.r.t. $\tilde{\mathcal{E}}$ if and only if $V + \prod_i [-\epsilon_i/2, \epsilon_i/2] \supseteqq \tilde{V}$. In particular $\delta(\mathcal{E},\tilde{\mathcal{E}})$ is the smallest ϵ so that

$$V + \prod_i [-\epsilon/2, \epsilon/2] \supseteqq \tilde{V} .$$

Similarily $2\dot{\delta}(\mathcal{E},\tilde{\mathcal{E}})$ is the smallest ϵ so that

$$V + \{0\} \times \prod_{i=1}^{r} [-\epsilon/2, \epsilon/2] \supseteqq \tilde{V} .$$

F determines N through the formula:

$$N(\xi_1,\ldots,\xi_r) = \int |1+\sum_i \xi_i x_i| F(dx) \; ; \; (\xi_1,\ldots,\xi_r) \in R^r .$$

Comparison by testing problems may be reduced to the case $r = 1$; i.e. the case of pseudo dichotomies by:

Proposition 4.2

$\mathcal{E}$ is $(\epsilon_o,\epsilon_1,\ldots,\epsilon_r)$ deficient w.r.t. $\tilde{\mathcal{E}}$, for testing problems, if and only if $(\pi,\Sigma\xi_i\sigma_i)$ is $(\epsilon_o,\Sigma|\xi_i|\epsilon_i)$ deficient w.r.t. $(\tilde{\pi},\Sigma\xi_i\tilde{\sigma}_i)$ for all $\xi \in R^r$.

In particular:

$\delta(\mathcal{E},\tilde{\mathcal{E}})$ is the smallest ϵ so that $(\pi,\Sigma\xi_i\sigma_i)$ is $(\epsilon,\epsilon\Sigma|\xi_i|)$ deficient w.r.t. $(\tilde{\pi},\Sigma\xi_i\tilde{\sigma}_i)$ for all $\xi \in R^r$.

and:

$2\dot{\delta}(\mathcal{E},\tilde{\mathcal{E}})$ is the smallest ϵ so that $(\pi,\Sigma\xi_i\sigma_i)$ is $(0,\epsilon\Sigma|\xi_i|)$ deficient w.r.t. $(\tilde{\pi},\Sigma\xi_i\tilde{\sigma}_i)$ for all $\xi \in R^r$.

Proof: It follows from the testing criterion for comparison of pseudo experiments that $\mathcal{E}$ is $\epsilon_o,\epsilon_1,\ldots,\epsilon_r$ deficient w.r.t. $\tilde{\mathcal{E}}$ if and only if $N(\xi) \geqq \tilde{N}(\xi) - (\epsilon_o + \sum_{i=1}^{r} |\xi_i|\epsilon_i)$. The criterions follows by comparing this criterion for the case $r = 1$ with the case of a general r . □

Corollary 4.3

$$\dot{\delta}_2(\mathcal{E},\tilde{\mathcal{E}}) = \sup_{\xi\neq 0} \frac{\dot{\delta}(\mathcal{E}_{\pi,\Sigma\xi_i\sigma_i})\tilde{\mathcal{E}}_{\tilde{\pi},\Sigma\xi_i\tilde{\sigma}_i})}{\Sigma|\xi_i|}$$

A few criterions for "$\geq_{\bar{k}}$" are listed in:

Proposition 4.4

Let k be one of the numbers 2,3,... . The following conditions on the pair $(\mathcal{S},\tilde{\mathcal{S}})$ of derivatives are equivalent:

(i) $\mathcal{S} \underset{k}{\geqq} \tilde{\mathcal{S}}$

(ii) $\dot{\delta}_k(\mathcal{S},\tilde{\mathcal{S}}) = 0$

(iii) $\psi(\mathcal{S}) \geqq \psi(\tilde{\mathcal{S}})$ for any function ψ on R^{r+1} which is a maximum of k linear functionals.

(iv) $\int \varphi dF \geqq \int \varphi d\tilde{F}$ for any function φ on R^r which is a maximum of k linear functions.

(v) (for $k = 2$) $N \geqq \tilde{N}$

(vi) (for $k = 2$) $V \supseteqq \tilde{V}$

Proof: (i) <=> (ii): Follows from $\delta_k \leqq 2\dot{\delta}_k$ and the definition of $\dot{\delta}_k$.

(i) <=> (iii): Follows from the sub linear function criterion.

(iii) <=> (iv): A function φ has the properties described in (iv) if and only if it is of the form $(x_1,\ldots,x_r) \rightsquigarrow \psi(1,x_1,\ldots,x_r)$ where ψ has the properties described in (iii).

It follows from the considerations above that v <=> (vi) and that v <=> (i), when $k = 2$. □

Corollary 4.5

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

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

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

(iii) $\psi(\mathcal{E}) \geqq \psi(\tilde{\mathcal{E}})$ for any sub linear function ψ on R^{r+1} .

(iv) $\int \varphi dF \geqq \int \varphi d\tilde{F}$ for any convex function φ on R^{r+1}

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

Remark: If $\tilde{\chi}$ is a Borel sub set of a Polish space and $\tilde{\mathcal{A}}$ is the class of Borel sub sets of $\tilde{\chi}$ then - by the randomization criterion - each of these conditions are equivalent with:

(vi) There is a randomization M from $(\chi, \mathcal{A})$ to $(\tilde{\chi}, \tilde{\mathcal{A}})$ so that $\pi M = \tilde{\pi}$ and $\sigma_i M = \tilde{\sigma}_i$; $i = 1,\dots,r$.

Proof of the corollary: (i) <=> (ii) <=> (iii) <=> (iv) <=> (v) follows from proposition 4.4 and theorem 2 in Strassen's paper [12]. Let φ be convex on R^r . Then there are constant a_{ij} so that

$$\varphi(x) = \lim_n \uparrow \max_{1 \leqq j \leqq n} (a_{oj} + a_{1j}x_1 + \dots + a_{rj}x_r) = \lim_n \uparrow \psi_n(1, x, \dots x_r)$$

where

$$\psi_n(x_o, x_1, \dots, x_r) = \max_{1 \leqq j \leqq n} (a_{oj}x_o + a_{1j}x_1 + \dots + a_{rj}x_r) \ .$$ □

Finally we give the factorization criterion for sufficiency

Proposition 4.6

Let $\mathcal{E} = ((\chi, \mathcal{A}), \pi, \sigma_1, \dots, \sigma_r)$ be a derivative and let $\tilde{\mathcal{E}}$ be the sub derivative $((\chi, \mathcal{B}), \pi_{\mathcal{B}} \sigma_{1\mathcal{B}}, \dots, \sigma_{r\mathcal{B}})$ 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 s may be specified $\mathcal{B}$ measurable.

Proof: It follows from proposition 4.12 in LC1 that the condition is necessary. Suppose s is $\mathcal{B}$ measurable. Then $s_i = d\sigma_{i\mathcal{B}}/d\pi_{\mathcal{B}}$ so that $\tilde{F} = \mathcal{L}_{\pi_{\mathcal{B}}}(s) = \mathcal{L}_{\pi}(s) = F$. □

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 $\mathfrak{D}_n$; $n = 1,2,\ldots$ and $\mathfrak{D}$ are equivalent

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

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

(iii) $\lim_{n\to\infty} \Delta_2(\mathcal{E}_n, \mathcal{E}) = 0$

(iv) $\lim_{n\to\infty} \dot{\Delta}_2(\mathcal{E}_n, \mathcal{E}) = 0$

(v) $\lim_{n\to\infty} \Lambda(S_n, S) = 0$

(vi) $\lim_{n\to\infty} [\Lambda(F_n, F) + |\int \sum_{i=1}^{r} |x_i|(F_n - F)(dx)|] = 0$

(vii) $\lim_{n\to\infty} N_n(\xi) = N(\xi)$; $\xi \in R^r$.

---

*) If μ, ν are finite non negative measures on some space R^k then $\Lambda(\mu,\nu)$ is the smallest number $h \geqq 0$ so that:

$$\mu\left(\prod_{i=1}^{k}]-\infty, x_i - h[\right) - h \leqq \nu\left(\prod_{i=1}^{k}]-\infty, x_i[\right) \leqq \mu\left(\prod_{i=1}^{k}]-\infty, x_i + h[\right) + h \; ;$$

$$(x_1, \ldots, x_k) \in R^k$$

Convergence for the metric Λ is weak convergence in the sence that $\Lambda(\mu_n, \mu) \to 0$ if and only if $\mu_n(f) \to \mu(f)$ when f is continuous and bounded.

<u>Remark.</u> We have not listed the criterions which follows directly from the right hand sides of the distance formulas in proposition 4.1. By this theorem and proposition 4.1 the convergence in (vii) implies that $N_n(\xi)/\Sigma|\xi_i| \to N(\xi)/\Sigma|\xi_i|$ as $n \to \infty$, uniformly in $\xi \neq 0$. Clearly the "2" in (iii) and (iv) may be replaced by any $k = 3,4,\ldots$.

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

(i) => (iv) : Follows from the inequality $\dot{\Delta}_2 \leqq \dot{\Delta}$

(iv) => (iii): " " " " $\Delta_2 \leqq 2\dot{\Delta}_2$

(v) <=> (iii) <=> (ii): Follows from theorem B.2.13 in the appendix

(v) => (vi): Suppose $S_n(h) \to S(h)$ when h is continuous and bounded on R^{r+1} . Let g be continuous and bounded on R^r .

Then $\int g dF_n = \int g(^{x}1/x_o,\ldots,^{x}r/x_o)S_{n,o}(dx)$

$= \int x_o g(^{x}1/x_o,\ldots,^{x}r/x_o)S_n(dx) = \int h(x_o,x_1,\ldots,x_r)S_n(dr)$ where

$h(x_o,x_1,\ldots x_r) = \min\{x_o,1\}g(^{x}1/x_o,\ldots,^{x}r/x_o)$ or $= 0$ as $x_o > 0$ or $x_o \leqq 0$. Then h is continuous and bounded on R^{r+1} so that $\int g dF_n = \int h dS_n \to \int h dS = \int g dF$. It follows that $\Lambda(F_n,F) \to 0$.

By the formula for Δ in proposition 4.1 we get for

$\psi(x_o,x_1,\ldots,x_r) = \sum_{i=1}^{r}|x_i|$ that $\int \sum_{i=1}^{r}|x_i|F_n(dx) = \psi(\mathcal{S}_n) \to \psi(\mathcal{S})$

$= \int \sum_{i=1}^{r}|x_i|F(dx)$.

(vi) => (v): Suppose (vi) holds. Let g be continuous and bounded on R^{r+1}. Then - since $(x_1,\dots x_r) \to$

$$g((1+\sum_{i=1}^r |x_i|)^{-1}, x_1(1+\sum_{i=1}^r |x_i|)^{-1}, \dots, x_r(1+\sum_{i=1}^r |x_i|)^{-1})(1+\sum_{i=1}^r |x_i|)$$

is uniformly integrable w.r.t. $F_1, F_2, \dots$: $\int g dS_n =$

$$\int g((1+\sum_{i=1}^r |x_i|)^{-1}, x_1(1+\sum_{i=1}^r |x_i|)^{-1}, \dots, x_r(1+\sum_{i=1}^r |x_i|)^{-1})(1+\sum_{i=1}^r |x_i|) F_n(dx)$$

$$\to \int g((1+\sum_{i=1}^r |x_i|)^{-1}, x_1(1+\sum_{i=1}^r |x_i|)^{-1}, \dots, x_r(1+\sum_{i=1}^r |x_i|)^{-1})(1+\sum_{i=1}^r |x_i|) F(dx)$$

$$= \int g dS$$

(ii) => (i). Suppose (ii). Then - as we have seen - (iii), (v) and (vi) hold. Let L_n denote the Prohorov distance between F_n and F for the metric ; $(x,y) \mapsto \max_i |x_i - y_i|$. By theorem 11 in Strassen's paper [12] there is for each n - a probability distribution Q_n on $R^r \times R^r$ with marginals F_n and F such that $Q_n(D_n) \leqq L_n$ where $D_n = \{(x,y): \max_i |x_i - y_i| \geqq L_n\}$. Let $\psi \in \hat{\Psi}$. Then $|\psi(\mathcal{G}_n) - \psi(\mathcal{G})| = |\int \psi(1,x_1,\dots,x_r) F_n(dx) - \int \psi(1,x_1,\dots,x_r) F(dx)| =$

$$|\int [\psi(1,x_1,\dots,x_r) - \psi(1,y_1,\dots,y_r)] Q_n(d(x,y))| \leqq$$

$$\int \sum_i \psi(e_i) |x_i - y_i| Q_n(d(x,y)) \leqq \sum_{i=1}^r \psi(e_i) \int \max_i |x_i - y_i| Q_n(d(x,y)) =$$

$$\sum_{i=1}^r \psi(e_i) \int_{\complement D_n} \max_i |x_i - y_i| Q_n(d(x,y)) + \sum_{i=1}^r \psi(e_i) \int_{D_n} \max_i |x_i - y_i| Q_n(d(x,y)) \leqq$$

$$\sum_{i=1}^r \psi(e_i) [L_n + \int_{D_n} \max_i |x_i - y_i| Q_n(d(x,y))] .$$

By proposition 4.1: $2\dot{\Delta}(\mathcal{Q}_n,\mathcal{Q}) \leqq L_n + \int_{D_n} \max_i |x_i - y_i| Q_n(d(x,y))$.

By (vi): $L_n \to 0$. It follows that $Q_1, Q_2, \ldots$ are relatively compact and that $Q_n(D_n) \to 0$. $(x,y) \rightsquigarrow \max_i |x_i - y_i|$ is uniformly integrable since it is dominated by $(x,y) \rightsquigarrow \sum_i |x_i| + \sum_i |y_i|$ and the latter is uniformly integrable. Hence - by uniform absolute continuity - $\dot{\Delta}(\mathcal{Q}_n,\mathcal{Q}) \to 0$.

(vi) => (vii): follows directly from proposition 4.1

(vii) => (vi). Suppose $N_n(\xi) \to N(\xi)$ for all ξ .

Then $\int |1+x_i| F_n(dx) \to \int |1+x_i| F(dx)$; $i = 1,2\ldots r$

Hence: $\limsup_n \int |x_i| F_n(dx) \leqq \limsup_n \int (1+|1+x_i|) F_n(dx)$

$= 1 + \lim_n \int |1+x_i| F(dx) < \infty$; $i = 1,2\ldots r$.

It follows that $F_1, F_2, \ldots$ are conditionally compact.

Let $\xi_0 \neq 0$. Then:

$\int |\xi_0 + \xi_1 x_1 + \ldots + \xi_r x_r| F_n(dx) = |\xi_0| N_n(\xi_0/\xi_0, \ldots \xi_r/\xi_0) \to$

$|\xi_0| N(\xi_1/\xi_0, \ldots, \xi_r/\xi_0) = \int |\xi_0 + \xi_1 x_1 + \ldots + \xi_r x_r| F(dx)$. Hence:

$\limsup_n \int |\xi_1 x_1 + \ldots + \xi_r x_r| F_n(dx) \leqq \limsup_n \int [|\xi_0| + |\xi_0 + \xi_1 x_1 + \ldots + \xi_r x_r| F_n(dx)]$

$= |\xi_0| + \int |\xi_0 + \xi_1 x_1 + \ldots + \xi_r x_r| F(dx)$. $\xi_0 \to 0$ yields:

$\limsup_n \int |\xi_1 x_1 + \ldots + \xi_r x_r| F_n(dx) \leqq \int |\xi_1 x_1 + \ldots + \xi_r x_r| F(dx)$

Similarily:

$\liminf_n \int |\xi_1 x_1 + \ldots + \xi_r x_r| F_n(dx) \geqq \int [|\xi_0 + \xi_1 x_1 + \ldots + \xi_r x_r| - |\xi_0|] F_n(dx)$

$= \int |\xi_0 + \xi_1 x_1 + \ldots + \xi_r x_r| F(dx) - |\xi_0|$. $\xi_0 \to 0$ yields:

$\liminf_n \int |\xi_1 x_1 + \ldots + \xi_r x_r| F_n(dx) \geqq \int |\xi_1 x_1 + \ldots + \xi_r x_r| F(dx)$

It follows that $\int|\xi_0+\xi_1x_1+\dots+\xi_rx_r|F_n(dx) \to \int|\xi_0+\xi_1x_1+\dots+\xi_rx_r|F(dx)$ for all $(\xi_0,\xi_1,\dots \xi_r) \in R^{r+1}$. Consider r constants $a_1,\dots,a_r$. Then $\int |\xi-(a_1x_1+\dots+a_rx_r)|F_n(dx) \to \int |\xi-(a_1x_1+\dots+a_rx_r)|F(dx)$

By theorem 5.1 in LC1:

$\Delta((F_nG_n),(F,G)) \to 0$ where

$dG_n/dF_n|_x = a_1x_1+\dots+a_rx_r = dG/dF$.

By the same theorem

$\mathcal{L}_{F_n}(a_1X_1+\dots+a_rX_r) \to \mathcal{L}_F(a_1X_1+\dots+a_rX_r)$

where $X_1,\dots,X_r$ are the coordinate functions on R^r. It follows that $\Lambda(F_n,F) \to 0$. Hence - since $\int|x_i|F_n(dx) \to \int|x_i|F(dx)$; $i = 1,2\dots r$ - $\Lambda_n(F_n,F) + |\int\sum_{i=1}^{r}|x_i|F_n(dr) - \int\sum_{i=1}^{r}|x_i|F(dx)| \to 0$. □

When does a sequence $\mathcal{E}_1, \mathcal{E}_2,\dots$ converge? A necessary and sufficient condition is given in:

<u>Proposition 5.2</u>

A sequence $\mathcal{E}_n$; $n = 1,2,\dots$ of derivatives is $\dot{\Delta}$ convergent if an only if $\lim_{n\to\infty} N_n(\xi)$ exists for all $\xi \in R^r$ and

$$\lim_{\Sigma|\xi_i|\to 0} [\lim_{n\to\infty} N_n(\xi)-1][\Sigma|\xi_i|]^{-1} = 0 .$$

<u>Proof:</u> The conditions are - by theorem 5.1 - and theorem 3.2 - necessary. Suppose now that $M(\xi) = \lim_{n\to\infty} N_n(\xi)$ exists for all $\xi \in R^r$ and that $\lim_{\Sigma|\xi_i|\to 0} \frac{M(\xi)-1}{\Sigma|\xi_i|} = 0$. It follows from theorem

5.1 that it suffices to show that M is the N function N of some derivative $\mathcal{G}$. Let $\xi_0 \neq 0$. Then:

$$\int |\xi_0+\xi_1 x_1+\dots+\xi_r x_r| F_n(dx) = |\xi_0| N_n(\xi_1/\xi_0,\dots,\xi_r/\xi_0)$$

$\rightarrow |\xi_0| M(\xi_1/\xi_0,\dots,\xi_1/\xi_0)$. Hence:

$$\limsup_n \int |\xi_1 x_1+\dots+\xi_r x_r| F_n(dx) \leqq \limsup_n \int [\,|\xi_0|+|\xi_0+\xi_1 x_1+\dots+\xi_r x_r|\,] F_n(dx)$$

$= |\xi_0|+|\xi_0| M(\xi_1/\xi_0,\dots \xi_r/\xi_0) < \infty$. It follows that F_n; $n = 1,2\dots$ are conditionally compact. We may therefore - without loss of generality assume that $\Lambda(F_n,F) \rightarrow 0$.

Similarily:

$$\liminf_n \int |\xi_1 x_1+\dots+\xi_r x_r| F_n(dx) \geqq -|\xi_0|+|\xi_0| M(\xi_1/\xi_0,\dots \xi_r/\xi_0)$$

Hence: $\limsup_n \int |\xi_1 x_1+\dots+\xi_r x_r| F_n(dx) - \liminf_n \int |\xi_1 x_1+\dots+\xi_r x_r| F_n(dx) \leqq |\xi_0|$.

It follows by letting $\xi_0 \rightarrow 0$ - that

$W(\xi_0,\xi_1,\dots,\xi_r) = \lim_n \int |\xi_0+\xi_1 x_1+\dots+\xi_r x_r| F_n(dx)$ exists for all $(\xi_0,\xi_1,\dots,\xi_r) \in R^{r+1}$. Fix r numbers $a_1,a_2,\dots,a_r$.

Then - for any number ξ -:

$\int |\xi-a_1 x_1-\dots-a_r x_r| F_n(dx) \rightarrow W(\xi,-a_1,\dots,-a_r)$ and

$W(\xi,-a_1,\dots,-a_r)-|\xi|$

$$= [M(-a_1/|\xi|,\dots, -a_r/|\xi|)-1] |(-a_1/|\xi|)+\dots+(-a_r/|\xi|)| \Sigma |a_i| \rightarrow 0$$

as $|\xi| \rightarrow \infty$. Let the measures G_n, $n = 1,2\dots$ and G be determined by: $\left[dG_n/dF_n\right]_x = \left[dG/dF\right]_x = \Sigma a_i x_i$. By proposition 5.3

in LC1 the derivative (F_n, G_n) ; $n = 1,2...$ converges as $n \to \infty$. In particular $x \rightsquigarrow |\sum_i a_i x_i|$ is uniformly integrable w.r.t. $F_1, F_2, ...$. It follows that $\int |\xi - a_1x_1 - ... - a_rx_r| F_n(dx) \to \int |\xi - a_1x_1 - ... - a_rx_r| F(dx)$ for all points $(\xi, a_1, ..., a_r) \in R^{r+1}$. This imply that $\int x_i F(dx) = 0$; $i = 1, ..., r$.

Hence F represents a derivative $\mathcal{D}$ and $M(\xi) = \lim_n \int |1+\Sigma\xi_i x_i| F_n(dx) = \int |1+\Sigma\xi_i x_i| F(dx) = N(\xi)$. By theorem 5.1 $\dot{\Delta}(\mathcal{E}_n, \mathcal{E}) \to 0$. □

<u>Corollary 5.3</u>

The pseudo metrics $\dot{\Delta}_2, \dot{\Delta}_3, ...$ and $\dot{\Delta}$ are all complete. It may - however - happen that a $\dot{\Delta}$ divergent sequence $\mathcal{E}_1, \mathcal{E}_2, ...$ is Δ convergent to a pseudo experiment $\mathcal{E}$ which is not a derivative.

<u>Proof</u>: Let $\mathcal{E}_1, \mathcal{E}_2 ...$ be a sequence of derivatives such that $\dot{\Delta}_2(\mathcal{E}_m, \mathcal{E}_n) \to 0$ as $m, n \to \infty$. By proposition 4.1: $N_m(\xi) - N_n(\xi) \to 0$ as $m, n \to \infty$ for all ξ . Hence $M(\xi) = \lim_{n\to\infty} N_n(\xi)$ exists (finite) for all $\xi \in R^r$. By proposition 4.1 again: $\frac{N_m(\xi)-1}{\Sigma|\xi_i|} - \frac{N_n(\xi)-1}{\Sigma|\xi_i|} \to 0$ as $m, n \to \infty$; uniformly in $\xi \neq 0$. It follows that $\frac{N_n(\xi)-1}{\Sigma|\xi_i|} \to \frac{M(\xi)-1}{\Sigma|\xi_i|}$ as $n \to \infty$ uniformly in $\xi \neq 0$. Hence by theorem 3.2, $\frac{M(\xi)-1}{\Sigma|\xi_i|} \to 0$ as $\Sigma|\xi_i| \to 0$, so

that - by proposition 5.2 - $\dot{\Delta}(\mathcal{G}_n, \mathcal{G}) \to 0$. It follows that $\dot{\Delta}_2, \dot{\Delta}_3, \ldots$ and $\dot{\Delta}$ are all complete. An example proving the last statement is readily furnished by example 5.4 in LC1. □

Some compactness criterions are listed in:

Theorem 5.4

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

(i) $\{\mathcal{G}_t : t \in T\}$ is $\dot{\Delta}$ conditionally compact.

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

(iii) $\sup_t \frac{N_t(\xi)-1}{\sum_i |\xi_i|} \to 0$ as $\sum_i |\xi_i| \to 0$

(iv) $x \rightsquigarrow \sum_i |x_i|$ is uniformly integrabel w.r.t. $\{F_t : t \in T\}$.

Proof: (i) <=> (ii): Follows from theorem 5.1

(ii) => (iv) . Suppose (iv) does not hold. Then there is an $\eta > 0$ so that

$$\sup_t \int_{\sum |x_i| > n} \sum |x_i| F_t(dx) > \eta \; ; \; n = 1,2,\ldots .$$

It follows that for each $n = 1,2,\ldots$ there is a $t_n \in T$ so that:

$$\int_{\sum |x_i| > n} \sum |x_i| F_{t_n}(dx) > \eta .$$

By theorem 5.1 , no subsequence of $\mathcal{G}_{t_1}, \mathcal{G}_{t_2}, \ldots$ is Δ convergent.

(iv) => (i). Suppose (iv) holds. Then $\{F_t : t \in T\}$ is conditionally Λ compact and (i) follows from theorem 5.1.

We have, so far, shown that (i) <=> (ii) <=> (iv).

(i) => (iii). Suppose (iii) does not hold:
Then there is a sequence $\xi^{(n)}$, $n = 1,2\ldots$ in R^r and a number $\alpha > 0$ so that $\sum_i |\xi_i^{(n)}| \rightsquigarrow 0$ and $\sup_t \frac{N_t(\xi^{(n)})-1}{\sum|\xi_i^{(n)}|} > \alpha$; $n = 1,2\ldots$.

It follows that for each n there is a t_n so that $\frac{N_{t_n}(\xi^{(n)})-1}{\sum_i|\xi_i^{(n)}|} > \alpha$.

By the last part of theorem 3.2 and proposition 4.1, no sub sequence of $\mathcal{E}_{t_1}, \mathcal{E}_{t_2}, \ldots$ is $\dot{\Delta}$ convergent.

(iii) => (iv): Suppose $\sup_t \frac{N_t(\xi)-1}{\sum_i|\xi_i|} \to 0$ as $\sum_i|\xi_i| \to 0$.

Let $a_1, a_2, \ldots, a_r$ be fixed. Then:

$$\int |\eta - a_1x_1 - \ldots - a_rx_r| F_t(dx) - |\eta| = |\eta|[N_t(-{}^{a_1}/\eta, \ldots, -{}^{a_r}/\eta) - 1] =$$

$$= [N_t(-{}^{a_1}/\eta, \ldots, -{}^{a_r}/\eta)-1][\Sigma|-\frac{a_i}{\eta}|]^{-1} \sum_{i=1}^r |a_i| \to 0 \quad \text{as} \quad |\eta| \to \infty ;$$

uniformly in $t \in T$. Let G_t be the measure determined by: $dG_t/dF_t|_x = \Sigma a_i x_i$. By theorem 5.6 in LC1 the set $\{(F_t, G_t); t \in T\}$ of derivatives is conditionally compact. In particular $x \rightsquigarrow |\Sigma a_i x_i|$ is uniformly integrable w.r.t. $F_t : t \in T$. This being true for all $(a_1, \ldots, a_r) \in R^r$ imply that $\sum_i |x_i|$ is uniformly integrable w.r.t. $F_t : t \in T$. □

The particular case of asymptotic $\dot{\Delta}$ sufficiency is treated in:

<u>Theorem 5.5</u>

Let $\mathcal{E}_n = ((\chi_n, \mathcal{A}_n), \pi_n, \sigma_{1n}, \sigma_{2n}, \ldots, \sigma_{r,n})$; $n = 1,2,\ldots$ be a sequence of derivatives. For each n let $\mathcal{B}_n$ be a sub σ-algebra of $\mathcal{A}_n$ and let $\tilde{\mathcal{E}}_n$ be the sub derivative $((\chi_n, \mathcal{B}_n), \pi_n, \sigma_{n1}, \ldots, \sigma_{nr})$ where - by abuse of notations - $\pi_n, \sigma_{n1}, \ldots, \sigma_{nr}$ are the restrictions of $\pi_n, \sigma_{n1}, \ldots, \sigma_{nr}$ to $\mathcal{B}_n$. Finally let, for each n, $\hat{\sigma}_{n1}, \hat{\sigma}_{n2}, \ldots, \hat{\sigma}_{nr}$ be the measures on $\mathcal{A}_n$ determined by:

$$d\hat{\sigma}_{ni}/d\pi_n = E_{\pi_n}^{\mathcal{B}_n}[d\sigma_{ni}/d\pi_n] \; ; \; i = 1, \ldots, r$$

Then $\hat{\mathcal{E}}_n \stackrel{\text{def}}{=} ((\chi_n, \mathcal{A}_n), \pi_n, \hat{\sigma}_{n1}, \ldots, \hat{\sigma}_{nr})$ are all derivatives and $\hat{\mathcal{E}}_n \sim \tilde{\mathcal{E}}_n$; $n = 1, 2, \ldots$.

Suppose $\mathcal{E}_n$; $n = 1, 2 \ldots$ are conditionally compact. Then $\tilde{\mathcal{E}}_n$; $n = 1, 2 \ldots$ are also conditionally compact and the following conditions are equivalent:

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

(ii) $\lim_{n\to\infty} \|\sigma_{ni} - \hat{\sigma}_{ni}\| = 0$; $i = 1, 2 \ldots r$

(iii) $\lim_{n\to\infty} \Lambda(F_n, \tilde{F}_n) = 0$

<u>Proof</u>: Let - for each n - E_n denote expectation w.r.t. π_n

1° $\hat{\mathcal{E}}_n$ is a derivative since $\hat{\sigma}_{ni} \ll \pi_n$; $i = 1, 2, \ldots, r$ and $\hat{\sigma}_{ni}(\chi) = 0$; $i = 1, 2 \ldots r$. It follows from proposition 4.6 that $\hat{\mathcal{E}}_n \sim \tilde{\mathcal{E}}_n$.

2^o Suppose $\mathcal{E}_n$; n = 1,2... are conditionally compact. By proposition 5.8 in LC1, the derivatives $((\chi_n, \mathcal{B}_n), \pi_n, \sigma_{ni})$; n = 1,2... are conditionally compact for each i . Hence $E^{\mathcal{S}_n} s_{ni}$; n = 1,2... are uniformly integrable w.r.t. π_n ; n = 1,2... By theorem 5.4 this - for i = 1,2...r - implies that $\tilde{\mathcal{E}}_n$; n = 1,2... are conditionally compact.

3^o Suppose now that $\mathcal{E}_n$; n = 1,2... are conditionally compact:

(i) => (ii) : follows from proposition 5.8 in LC1

(ii) => (i) : Put $\epsilon_n = \max_i \|\sigma_{ni} - \hat{\sigma}_{ni}\|$. Then

$\dot{\Delta}(\mathcal{E}_n, \tilde{\mathcal{E}}_n) = \dot{\Delta}(\mathcal{E}_n, \hat{\mathcal{E}}_n) \leqq \epsilon_n \to 0$.

(i) <=> (iii): Follows - by conditional compactness - from theorem 5.1 . □

6. Local comparison of experiments.

We proceed as in section 6 in LC1, connecting the theory of derivatives with the statistical theory of information.

It was noted in LC1 that some results *) was only proved under the assumption that some of the measurable spaces involved, were Borel sub sets of Polish spaces. We shall not use such assumptions here and it will be shown in theorem 6.5 that - with the exception of proposition 4.11, proposition 6.5 and theorem 6.6 - all other results are proved without using assumptions of this type.

A few notational conventions are: Experiments will usually be written $\mathcal{E} = ((\chi,\mathcal{A}), (P_\theta : \theta \in \Theta))$ with or without affixes. If $\mathcal{E} = ((\chi,\mathcal{A}), P_\theta : \theta \in \Theta)$ then the derivative in θ^0 will - if it exists - be written:

$$\mathcal{E}_{\theta^0} = ((\chi,\mathcal{A}), P_{\theta^0}, \dot{P}_{\theta^0,1}, \ldots, \dot{P}_{\theta^0,r})$$

The restriction $((\chi,\mathcal{A}), (P_\theta : \theta \in \Theta_0))$ of $\mathcal{E}$, will be written $\mathcal{E}_{\Theta_0}$. Affixes on $\mathcal{E}, \chi, \mathcal{A}, P_\theta, \Theta, P_{\theta^0}, \dot{P}_{\theta^0,1}, \ldots, \dot{P}_{\theta^0,r}$ will, when when these are referring to the same experiment, be of the same type.

If $\mathcal{E}$ and $\tilde{\mathcal{E}}$ are both differentiable in θ^0 then we will write:

$$\dot{\delta}_{\theta^0(k)}(\mathcal{E}, \tilde{\mathcal{E}}) \underset{\text{definition}}{=} \dot{\delta}_{(k)}(\dot{\mathcal{E}}_{\theta^0}, \dot{\tilde{\mathcal{E}}}_{\theta^0})$$

---

*) These results were listed as: proposition 4.11, theorem 6.1, theorem 6.2, corollary 6.3, proposition 6.5 and theorem 6.6. Corollary 6.4 - whose proof depends on theorem 6.2 - were, by an oversight omitted from this list.

and

$$\dot{\Delta}_{\theta^o,(k)}(\mathcal{E},\tilde{\mathcal{E}}) \underset{\equiv}{\text{definition}} \dot{\Delta}_{(k)}(\dot{\mathcal{E}}_{\theta^o},\dot{\tilde{\mathcal{E}}}_{\theta^o})$$

If θ is close to θ^o and $\mathcal{E}$ is differentiable in θ^o then P_θ is approximately affine in θ. Hence deficiencies on a sub set Θ_o "close" to θ^o should be approximately equal to deficiencies on the convex hull $\langle\Theta_o\rangle$ of Θ_o. Inequalities implying results in this direction are given in:

<u>Proposition 6.1.</u> Let $\mathcal{E}$ and $\tilde{\mathcal{E}}$ be differentiable in θ^o and consider the expansions:

$$P_\theta = P_{\theta^o} + \sum_i(\theta_i-\theta_i^o)\dot{P}_{\theta^o,i} + \|\theta-\theta^o\|\Gamma_{\theta^o,\theta}$$

and

$$\tilde{P}_\theta = \tilde{P}_{\theta^o} + \sum_i(\theta_i-\theta_i^o)\dot{\tilde{P}}_{\theta^o,i} + \|\theta-\theta^o\|\tilde{\Gamma}_{\theta^o,\theta}$$

Then any sub set Θ_o of Θ, whose convex hull $\langle\Theta_o\rangle \subseteqq \Theta$, satisfies - provided $\mathcal{E}_{\Theta_o}$ is dominated - the inequalities:

$$\delta_{(k)}(\mathcal{E}_{\Theta_o},\tilde{\mathcal{E}}_{\Theta_o}) \leqq \delta_{(k)}(\mathcal{E}_{\langle\Theta_o\rangle},\tilde{\mathcal{E}}_{\langle\Theta_o\rangle}) \leqq$$

$$\delta_{(k)}(\mathcal{E}_{\Theta_o},\tilde{\mathcal{E}}_{\Theta_o}) + 2 \text{ distance } (\Theta_o,\theta^o)[\sup\|\Gamma_{\theta^o,\theta}\| + \sup\|\tilde{\Gamma}_{\theta^o,\theta}\|]$$

where both sup's are taken over $\langle\Theta_o\rangle$.

If both $\mathcal{E}_{\Theta_o}$ and $\tilde{\mathcal{E}}_{\Theta_o}$ are dominated then δ - in these inequalities may be replaced throughout with Δ.

Proof. (ii) follows directly from (i). Hence it suffices - since $\delta_k \uparrow \delta$ and $\Theta_o \subseteq \langle\Theta_o\rangle$ to prove that

$$\delta_k(\mathcal{E}_{\langle\Theta_o\rangle},\tilde{\mathcal{E}}_{\langle\Theta_o\rangle}) \leqq t_k + 2d[\sup\|\Gamma_{\theta^o,\theta}\| + \sup\|\tilde{\Gamma}_{\theta^o,\theta}\|]$$

where $t_k = \delta_k(\mathcal{E}_{\Theta_o},\tilde{\mathcal{E}}_{\Theta_o})$ and $d = \text{distance }(\Theta_o,\theta^o)$. Let $\theta \in \langle\Theta_o\rangle$. Then there are points $\theta^1,\theta^2,\ldots,\theta^k$ in Θ_o and non negative members $c_1,c_2,\ldots,c_k$ so that $\theta = \sum_{i=1}^{k} c_i\theta^i$ and

$1 = \sum_{i=1}^{k} c_i$. Hence

$$\Sigma c_j P_{\theta^j} = P_{\theta^o} + \Sigma(\theta_i-\theta_i^o)\dot{P}_{\theta^o,i} + \sum_j c_j\|\theta^j-\theta^o\|\Gamma_{\theta^o,\theta^j}$$

so that:

$$P_\theta-\sum_j c_j P_{\theta^j} = \|\theta-\theta^o\|\Gamma_{\theta^o,\theta}-\sum_j c_j\|\theta^j-\theta^o\|\Gamma_{\theta^o,\theta^j}$$

Hence:

$$\|P_\theta-\sum_j c_j P_{\theta^j}\| \leqq \|\theta-\theta^o\|\sup\|\Gamma_{\theta^o,\theta}\|+d\sup\|\Gamma_{\theta^o,\theta}\| \leqq 2d\sup\|\Gamma_{\theta^o,\theta}\| .$$

Similarily:

$$\|\tilde{P}_\theta-\Sigma c_j\tilde{P}_{\theta^j}\| \leqq 2d\sup\|\tilde{\Gamma}_{\theta^o,\theta}\| .$$

Let $\tilde{\rho}$ be a randomization from $(\tilde{\chi},\tilde{\mathcal{A}})$ to $\{1,\ldots,k\}$. By theorem B.3.4 in the appendix there is a randomization ρ from $(\chi,\mathcal{A})$ to $\{1,2,\ldots k\}$ so that:

$$\|P_\theta\rho-\tilde{P}_\theta\tilde{\rho}\| \leqq t_k \; ; \; \theta \in \Theta^o$$

Let $\theta \in \langle\Theta_o\rangle$ and write θ - as above - as a convex combination $\sum_{j=1}^{k} c_j\theta^j$ of points $\theta^1,\dots,\theta^k$ in Θ_o. Then:

$$\|P_\theta\rho-P_\theta\tilde\rho\| \leqq \|P_\theta-\Sigma c_jP_{\theta^j}\|+\|(\Sigma c_jP_{\theta^j})\rho-(\Sigma c_j\tilde P_{\theta^j})\tilde\rho\|$$

$$+\|\Sigma c_j\tilde P_{\theta^j}-\tilde P_\theta\| \leqq 2d[\sup\|\Gamma_{\theta^o,\theta}\|+\sup\|\tilde\Gamma_{\theta^o,\theta}\|]+\Sigma c_jt_k = 2d\,[\quad]+t_k\,. \qquad \square$$

The condition of dominans does not matter very much since it is - for small neighbourhoods - "approximately" implied by differentiability. Using the construction in section 3 we get:

Proposition 6.2

Let $\mathcal{E}$ and $\tilde{\mathcal{E}}$ be differentiable in θ^o. Consider the expansions:

$$P_\theta = P_{\theta^o}+\sum_i(\theta_i-\theta_i^o)\dot P_{\theta^o,i}+\|\theta-\theta^o\|\Gamma_{\theta^o,\theta}$$

$$|P_{\theta^o}+\sum_i(\theta_i-\theta_i^o)\dot P_{\theta^o,i}|\ \|P_{\theta^o}+\sum_i(\theta_i-\theta_i^o)\dot P_{\theta^o,i}\|^{-1} =$$

$$= P_{\theta^o}+\sum_i(\theta_i-\theta_i^o)\dot P_{\theta^o,i}+\|\theta-\theta^o\|\Lambda_{\theta^o,\theta}$$

$$\tilde P_\theta = \tilde P_{\theta^o}+\sum_i(\theta_i-\theta_i^o)\dot{\tilde P}_{\theta^o,i}+\|\theta-\theta^o\|\tilde\Gamma_{\theta^o,\theta}$$

$$|\tilde P_{\theta^o}+\sum_i(\theta_i-\theta_i^o)\dot{\tilde P}_{\theta^o,i}|\ \|\tilde P_{\theta^o}+\sum_i(\theta_i-\theta_i^o)\dot{\tilde P}_{\theta^o,i}\| =$$

$$= \tilde P_{\theta^o}+\Sigma(\theta_i-\theta_i^o)\dot{\tilde P}_{\theta^o,i}+\|\theta-\theta_o\|\tilde\Lambda_{\theta^o,\theta}$$

Then any sub set Θ_0 of Θ, whose convex hull $\langle\Theta_0\rangle \subseteq \Theta$, satisfies the inequalities:

$$\delta_{(k)}(\mathcal{E}_{\Theta_0},\tilde{\mathcal{E}}_{\Theta_0}) \leqq \delta_{(k)}(\mathcal{E}_{\langle\Theta_0\rangle},\tilde{\mathcal{E}}_{\langle\Theta_0\rangle})$$

$$\leqq \delta_{(k)}(\mathcal{E}_{\Theta_0},\tilde{\mathcal{E}}_{\Theta_0}) + 2 \text{ distance } (\Theta_0,\theta^0)[\sup\|\Lambda_{\theta^0,\theta}-\Gamma_{\theta^0,\theta}\|+$$

$$+ \sup\|\Lambda_{\theta^0,\theta}\| + \sup\|\tilde{\Gamma}_{\theta^0,\theta}\|]$$

where all sup's are taken over $\langle\Theta_0\rangle$. δ may - in these inequalities - be replaced throughout by Δ provided 2 distance $(\Theta_0,\theta^0)[\sup\|\tilde{\Lambda}_{\theta^0,\theta}-\tilde{\Gamma}_{\theta^0,\theta}\| + \sup\|\tilde{\Lambda}_{\theta^0,\theta}\| + \sup\|\Gamma_{\theta^0,\theta}\|]$ is added to the expression after the last $\leqq$.

__Proof__: Write $Q_\theta = |P_{\theta^0}+\Sigma(\theta_i-\theta_i^0)\dot{P}_{\theta^0,i}|\,\|P_{\theta^0}+\Sigma(\theta_i-\theta_i^0)\dot{P}_{\theta^0,i}\|^{-1}$ and $\mathcal{F}=((\chi,\mathcal{A}),Q_\theta : \theta \in \Theta)$. Then $\dot{\mathcal{E}}_{\theta^0} = \dot{\mathcal{F}}_{\theta^0}$ and $\mathcal{F}$ is dominated. Write $d = \text{distance } (\Theta_0,\theta^0)$. Then:

$$\delta_k(\mathcal{E}_{\langle\Theta_0\rangle},\tilde{\mathcal{E}}_{\langle\Theta_0\rangle}) \leqq \delta_k(\mathcal{E}_{\langle\Theta_0\rangle},\mathcal{F}_{\langle\Theta_0\rangle}) + \delta_k(\mathcal{F}_{\langle\Theta_0\rangle},\tilde{\mathcal{E}}_{\langle\Theta_0\rangle}) \leqq$$

$$\leqq \Delta(\mathcal{E}_{\langle\Theta_0\rangle},\mathcal{F}_{\langle\Theta_0\rangle}) + \delta_k(\mathcal{F}_{\Theta_0},\tilde{\mathcal{E}}_{\Theta_0}) + 2d[\sup\|\Lambda_{\theta^0,\theta}\|+\sup\|\tilde{\Gamma}_{\theta^0,\theta}\|] \leqq$$

$$\leqq 2d[\tfrac{1}{2}\sup\|\Lambda_{\theta^0,\theta}-\Gamma_{\theta^0,\theta}\|+\sup\|\Lambda_{\theta^0,\theta}\|+\sup\|\tilde{\Gamma}_{\theta^0,\theta}\|] + \delta_k(\mathcal{F}_{\Theta_0},\mathcal{E}_{\Theta_0}) +$$

$$+ \delta_k(\mathcal{E}_{\Theta_0},\tilde{\mathcal{E}}_{\Theta_0}) \leqq \delta_k(\mathcal{E}_{\Theta_0},\tilde{\mathcal{E}}_{\Theta_0}) + 2d[\sup\|\Lambda_{\theta^0,\theta}-\Gamma_{\theta^0,\theta}\| + \sup\|\Lambda_{\theta^0,\theta}\|$$

$$+ \sup\|\tilde{\Gamma}_{\theta^0,\theta}\|]$$

The last statement follows by a symmetry argument.

Corollary 6.3.

Let $\mathscr{E}$ and $\tilde{\mathscr{E}}$ be differentiable in θ^o. Then:

(i) $[\delta_{(k)}(\mathscr{E}_{\langle\Theta_o\rangle},\tilde{\mathscr{E}}_{\langle\Theta_o\rangle})-\delta_{(k)}(\mathscr{E}_{\Theta_o},\tilde{\mathscr{E}}_{\Theta_o})][\text{distance }(\Theta^o,\theta_o)]^{-1} \to 0$ uniformly in (k) as distance $(\Theta^o,\theta_o) \to 0$

(ii) $[\Delta_{(k)}(\mathscr{E}_{\langle\Theta_o\rangle},\tilde{\mathscr{E}}_{\langle\Theta_o\rangle})-\Delta_{(k)}(\mathscr{E}_{\Theta_o},\tilde{\mathscr{E}}_{\Theta_o})][\text{distance }(\Theta_o,\theta^o)]^{-1} \to 0$ uniformly in (k) as distance $(\Theta^o,\theta_o) \to 0$

Proof: Follows immediately from proposition 6.2. □

Corollary 6.4.

Let $\mathscr{E}$ and $\tilde{\mathscr{E}}$ be differentiable in θ^o and let "lim" be short for "limit as distance $(\Theta_o,\theta^o) \to 0$". Then:

(i) $\lim \delta_{(k)}(\mathscr{E}_{\langle\Theta_o\rangle},\tilde{\mathscr{E}}_{\langle\Theta_o\rangle})[\text{distance }(\Theta_o,\theta^o)]^{-1}$ exists if and only if $\lim \delta_{(k)}(\mathscr{E}_{\Theta_o},\tilde{\mathscr{E}}_{\Theta_o})[\text{distance }(\Theta_o,\theta^o)]^{-1}$ exists and if so then these limits are equal.

(ii) $\lim \Delta_{(k)}(\mathscr{E}_{\langle\Theta_o\rangle},\tilde{\mathscr{E}}_{\langle\Theta_o\rangle})[\text{distance }(\Theta_o,\theta^o)]^{-1}$ exists if and only if $\lim \Delta_{(k)}(\mathscr{E}_{\Theta_o},\tilde{\mathscr{E}}_{\Theta_o})[\text{distance }(\Theta_o,\theta^o)]^{-1}$ exists and if so then these limits are equal.

Proof: This is a direct consequence of corollary 6.3. □

Theorem 6.5

Theorem 6.1 in LC1, theorem 6.2 in LC1, corollary 6.3 in LC1 and corollary 6.4 in LC1 hold without any assumptions on the structures of the involved sample spaces.

Proof:

(i) Proof of theorem 6.1 in LC1. Write δ_ϵ for $\delta(\mathcal{E}_{\{\theta_0-\epsilon,\theta_0+\epsilon\}},\tilde{\mathcal{E}}_{\{\theta_0-\epsilon,\theta_0+\epsilon\}})$ and write $\dot{\delta}$ for $\dot{\delta}_{\theta_0}(\mathcal{E},\tilde{\mathcal{E}})$.

If suffices - by corollary 6.4 in this paper - to show that $\delta_{\epsilon/2\epsilon} \to \dot{\delta}$. We will - to this end - apply theorem 6.11 (v) in LC1. By theorem 15 in [15]:

$$\beta_\epsilon(\alpha+\delta_{\epsilon/2})+\delta_{\epsilon/2} \geqq \tilde{\beta}_\epsilon(\alpha) \ ; \ \alpha \in [0,1]$$

where $\beta_\epsilon(\alpha)$ ($\tilde{\beta}_\epsilon(\alpha)$) is the power of the most powerful level α test for testing $P_{\theta_0-\epsilon}$ against $P_{\theta_0+\epsilon}$ ($\tilde{P}_{\theta_0-\epsilon}$ against $\tilde{P}_{\theta_0+\epsilon}$). By theorem 6.11 (v) in LC1 this may be written:

$$\delta_{\epsilon/2\epsilon} \geqq \tilde{\beta}(\alpha)-\beta(\alpha)+o_\epsilon \ ; \ \alpha \in [0,1]$$

where $o_\epsilon \to 0$ as $\epsilon \to 0$, uniformly in α. (We use the obvious fact that $\delta_\epsilon \to 0$ as $\epsilon \to 0$) It follows that

$$\liminf_{\epsilon\to 0} \delta_{\epsilon/2\epsilon} \geqq \sup_\alpha(\tilde{\beta}(\alpha)-\beta(\alpha))^+ = \dot{\delta}$$

By theorem 15 in [15] there is an $\alpha_\epsilon \in [0,1-\delta_{\epsilon/2}]$ satisfying $\beta_\epsilon(\alpha_\epsilon+\delta_{\epsilon/2})+\delta_{\epsilon/2} = \tilde{\beta}_\epsilon(\alpha_\epsilon)$.

Using theorem 6.11 (v) again this may be written:

$$\delta_{\epsilon/2\epsilon} = \tilde{\beta}_\epsilon(\alpha_\epsilon)-\beta_\epsilon(\alpha_\epsilon)+o_\epsilon$$

Hence

$$\delta_{\epsilon/2\epsilon} \leqq \dot{\delta}+o_\epsilon$$

so that

$$\limsup_{\epsilon\to 0} \delta_{\epsilon/2\epsilon} \leqq \dot{\delta}$$

Altogether we have shown that $\delta_{\epsilon/2\epsilon} \to \dot{\delta}$. Theorem 6.1 in LC1 follows now from corollary 6.4 in this paper.

(ii) Proof of theorem 6.2 in LC1. The proof is very similar to that in (i).

(iii) Corollary 6.3 in LC1 follows from (i) and (ii) above.

(iv) Proof of corollary 6.4 in LC1. If $\mathcal{E}_1,\dots,\mathcal{E}_n$ are dominated then the proof may be based on theorem 6.2 in LC1 as explained there. This may - in the general case - be applied to dominated experiments $\hat{\mathcal{E}}_1,\dots,\hat{\mathcal{E}}_n$ such that $\dot{\delta}(\mathcal{E}_i, \hat{\mathcal{E}}_i) = 0$; $i = 1,2,\dots,n$. If $\mathcal{E}_i = ((\chi_i,\mathcal{A}_i)(P_\theta^{(i)} ; \theta \in \Theta))$ then we may take

$\hat{\mathcal{E}}_i = ((\chi_i,\mathcal{A}_i)(\hat{P}_\theta^{(i)} ; \theta \in \Theta))$ where

$$\hat{P}_\theta^{(i)} = \|P_{\theta_0}^{(i)}+(\theta-\theta_0)\dot{P}_{\theta_0}^{(i)}\|^{-1} |P_{\theta_0}^{(i)}+(\theta-\theta_0)\dot{P}_{\theta_0}^{(i)}| .$$

Then $\hat{P}_{\theta_0}^{(i)} = P_{\theta_0}^{(i)}$ and $\dot{\hat{P}}_{\theta_0}^{(i)} = \dot{P}_{\theta_0}^{(i)}$

so that $\prod_i \hat{P}_{\theta_0}^{(i)} = \prod_i P_{\theta_0}^{(i)}$ and

$\left[\widehat{\prod_i \hat{P}_\theta^{(i)}}\right]_{\theta=\theta_o} = \left[\widehat{\prod_i P_\theta^{(i)}}\right]_{\theta=\theta_o}$. Hence

$\dot{\Delta}(\Pi\,\mathcal{E}_i, \Pi\,\hat{\mathcal{E}}_i) = 0$. It follows that:

$\delta(\Pi\,\mathcal{E}_i, \Pi\,\mathcal{F}_i) = \delta(\Pi\,\hat{\mathcal{E}}_i, \Pi\,\mathcal{F}_i) \leqq \Sigma\delta(\hat{\mathcal{E}}_i, \mathcal{F}_i) = \Sigma\delta(\mathcal{E}_i, \mathcal{F}_i)$. □

Generalizing theorem 6.1 to the r dimensional case we get:

Theorem 6.6

Let $\mathcal{E}$ and $\hat{\mathcal{E}}$ be differentiable in θ^o . Denote by ω_ϵ the convex set: $\{\theta : \Sigma|\theta_j - \theta_j^o| \leqq \epsilon\}$. The set of extreme points of ω_ϵ will be denoted by $\text{ext}\omega_\epsilon$; i.e.

$\text{ext}\omega_\epsilon = \{\theta^o - \epsilon v_1, \theta^o + \epsilon v_1, \theta^o - \epsilon v_2, \theta^o + \epsilon v_2, \ldots, \theta^o - \epsilon v_r, \theta^o + \epsilon v_r\}$.

Then - uniformly in (k) -

$$\lim_{\epsilon\to 0} \delta_{(k)}(\mathcal{E}_{\omega_\epsilon}, \tilde{\mathcal{E}}_{\omega_\epsilon})/2\epsilon = \lim_{\epsilon\to 0} \delta_{(k)}(\mathcal{E}_{\text{ext}\omega_\epsilon}, \tilde{\mathcal{E}}_{\text{ext}\omega_\epsilon})/2\epsilon = \dot{\delta}_{\theta^o,(k)}(\mathcal{E}, \tilde{\mathcal{E}})$$

and

$$\lim_{\epsilon\to 0} \Delta_{(k)}(\mathcal{E}_{\omega_\epsilon}, \tilde{\mathcal{E}}_{\omega_\epsilon})/2\epsilon = \lim_{\epsilon\to 0} \Delta_{(k)}(\mathcal{E}_{\text{ext}\omega_\epsilon}, \tilde{\mathcal{E}}_{\text{ext}\omega_\epsilon})/2\epsilon = \dot{\Delta}_{\theta^o,(k)}(\mathcal{E}, \tilde{\mathcal{E}})$$

Proof: If suffices (by corollary 6.3) to show that $\lim \delta_{k,\epsilon}/2\epsilon = \dot{\delta}_{\theta^o,k}(\mathcal{E}, \tilde{\mathcal{E}})$ uniformly in $k = 2,3,\ldots$ as $\epsilon \to 0$ where $\delta_{k,\epsilon} = \delta_k(\mathcal{E}_{\text{ext}\omega_\epsilon}, \tilde{\mathcal{E}}_{\text{ext}\omega_\epsilon})$. Fix a $k \in \{2,3\ \ldots\}$ and choose a number $\eta > 0$. There is, by proposition B.3.6 in the appendix for each permissible ϵ a sub algebra $\tilde{\mathcal{B}}$ of $\tilde{\mathcal{A}}$ having at most 2^k sets satisfying:

$$\delta(\mathcal{E}_{\text{ext}\omega_\epsilon}, \tilde{\mathcal{E}}_{\text{ext}\omega_\epsilon} | \tilde{\mathcal{B}}) \geqq \delta_{k,\epsilon} - \eta$$

where $|\tilde{\mathcal{B}}$ indicates restriction to $\tilde{\mathcal{B}}$. By theorem B.3.5 in the appendix - since $\tilde{\mathcal{B}}$ is finite - there is a randomization M from $(\mathcal{X}, \mathcal{A})$ to $(\tilde{\mathcal{X}}, \tilde{\mathcal{B}})$ so that:

$$P_{\theta^o}M = \tilde{P}_{\theta^o}|\tilde{\mathcal{B}} \text{ and } \max_i \|\dot{P}_{\theta^o,i}M - \dot{\tilde{P}}_{\theta^o,i}|\tilde{\mathcal{B}}\| = 2\dot{\delta}_{\theta_o}(\mathcal{E}, \tilde{\mathcal{E}}|\tilde{\mathcal{B}})$$

Expand P_θ and $\tilde{P}_\theta$ around θ^o as follows:

$$P_\theta = P_{\theta^o} + \Sigma(\theta_i - \theta_i^o)\dot{P}_{\theta^o,i} + \|\theta-\theta^o\|\Gamma_{\theta^o,\theta}$$

and

$$\tilde{P}_\theta = \tilde{P}_{\theta^o} + \sum_i(\theta_i - \theta_i^o)\dot{\tilde{P}}_{\theta^o,i} + \|\theta-\theta^o\|\tilde{\Gamma}_{\theta^o,\theta}$$

Then:

$$P_\theta M - \tilde{P}_\theta|\tilde{\mathcal{B}} = \Sigma(\theta_i - \theta_i^o)[\dot{P}_{\theta^o,i}M - \dot{\tilde{P}}_{\theta^o,i}|\tilde{\mathcal{B}}] + \|\theta-\theta^o\|[\Gamma_{\theta^o,\theta}M - \tilde{\Gamma}_{\theta^o,\theta}|\tilde{\mathcal{B}}]$$

Hence - provided $\theta \in \omega_\epsilon$:

$$\|P_\theta M - \tilde{P}_\theta|\tilde{\mathcal{B}}\| \leqq \epsilon 2\dot{\delta}_{\theta_o}(\mathcal{E}, \tilde{\mathcal{E}}|\tilde{\mathcal{B}}) + \epsilon \sup_{\omega_\epsilon}\{\|\Gamma_{\theta^o,\theta}\| + \|\tilde{\Gamma}_{\theta^o,\theta}\|\}$$

It follows that:

$$\delta(\mathcal{E}_{\text{ext}\omega_\epsilon}, \tilde{\mathcal{E}}_{\text{ext}\omega_\epsilon}|\tilde{\mathcal{B}})/2\epsilon \leqq \dot{\delta}_{\theta^o,k} + \tfrac{1}{2}\sup_{\omega_\epsilon}[\|\Gamma_{\theta^o,\theta}\| + \|\tilde{\Gamma}_{\theta^o,\theta}\|]$$

where $\dot{\delta}_{\theta^o,k} = \dot{\delta}_{\theta_o,k}(\mathcal{E}, \tilde{\mathcal{E}})$.

Hence:

$$\delta_{k,\epsilon}/2\epsilon \leqq \dot{\delta}_{\theta^o,k} + \tfrac{1}{2}\sup_{\omega_\epsilon}[\ \ ] + \eta/2\epsilon$$

$\eta \downarrow 0$ yields

(§) $$\delta_{k,\epsilon}/2\epsilon \leqq \dot{\delta}_{\theta^o,k}(\mathcal{E}, \tilde{\mathcal{E}}) + \tfrac{1}{2}\sup_{\omega_\epsilon}[\ \ ]$$

Let $\psi \in \hat{\Psi}_k(R^{r+1})$ (i.e. ψ is a maximum of k linear functionals on R^{r+1} and $\psi(1,0\ldots0) = \psi(-1,0\ldots0),\ldots,\psi(0,\ldots,1) = \psi(0,\ldots,-1)$). Put - for each point $(x_1,y_1,x_2,y_2,\ldots,x_r,y_r) \in R^{2r}$ -

$$\varphi_\epsilon(x_1,y_1,x_2,y_2,\ldots,x_r,y_r) = \psi\left(\frac{y_1+x_1}{2}, \frac{y_1-x_1}{2\epsilon}, \frac{y_2-x_2}{2\epsilon},\ldots,\frac{y_r-x_r}{2\epsilon}\right) .$$

Then $\varphi \in \Psi_k$ on R^{2r} and

$$\lim_{\epsilon\to0} \epsilon\varphi_\epsilon(x_1,y_1,\ldots,x_r,y_r) = \psi\left(0,\frac{y_1-x_1}{2},\ldots,\frac{y_r-x_r}{2}\right)$$

Put $\Sigma_\epsilon = \varphi_\epsilon(-1,0\ldots0)+\varphi_\epsilon(1,0\ldots0)+\ldots+\varphi_\epsilon(0,\ldots,-1)+\varphi_\epsilon(0,\ldots,1)$

Then $\lim_{\epsilon\to0}\epsilon\Sigma_\epsilon = 2[\psi(0,1,\ldots0)+\psi(0,0,1,\ldots,0)+\ldots+\psi(0,\ldots1)]$

By the remark after theorem B.2.1 in the appendix:

$$\delta_{k,\epsilon}/2\epsilon \geqq [\varphi(\tilde{P}_{\theta^o-\epsilon v_1},\tilde{P}_{\theta^o+\epsilon v_1},\ldots)-\varphi(P_{\theta^o-\epsilon v_1},P_{\theta^o+\epsilon v_1},\ldots,P_{\theta^o-\epsilon v_r}, P_{\theta^o+\epsilon v_r})]/\epsilon\Sigma_\epsilon$$

$$= [\psi\left(\frac{\tilde{P}_{\theta^o+\epsilon v_1}+\tilde{P}_{\theta^o-\epsilon v_1}}{2}, \frac{\tilde{P}_{\theta^o+\epsilon v_1}-\tilde{P}_{\theta^o-\epsilon v_1}}{2\epsilon}, \ldots , \frac{\tilde{P}_{\theta^o+\epsilon v_r}-\tilde{P}_{\theta^o-\epsilon v_r}}{2\epsilon}\right)$$

$$-\psi\left(\frac{P_{\theta^o+\epsilon v_1}+P_{\theta^o-\epsilon v_1}}{2}, \frac{P_{\theta^o+\epsilon v_1}-P_{\theta^o-\epsilon v_1}}{2\epsilon}, \ldots , \frac{P_{\theta^o+\epsilon v_r}-P_{\theta^o-\epsilon v_r}}{2\epsilon}\right)]/\epsilon\Sigma_\epsilon$$

$$\to [\psi(\tilde{P}_{\theta^o},\dot{\tilde{P}}_{\theta^o,1},\ldots,\dot{\tilde{P}}_{\theta^o,r})-\psi(P_{\theta^o},\dot{P}_{\theta^o,1},\ldots,\dot{P}_{\theta^o,r})]\cdot$$

$$[2\psi(0,1,\ldots 0)+\ldots+2\psi(0,\ldots,1)] \quad \text{as} \quad \epsilon \to 0 .$$

Proposition 4.1:

(§§) $\liminf_{\epsilon\to 0} \delta_{k,\epsilon}/2\epsilon \geqq \dot{\delta}_{\theta^o,k}$

Let $\varkappa > 0$ be given. Choose a $k_o \geqq 2$ so that

$\dot{\delta}_{\theta^o,k_o} \geqq \dot{\delta}_{\theta^o}-\varkappa$ where $\dot{\delta}_{\theta^o} = \dot{\delta}_{\theta^o}(\mathscr{E}, \tilde{\mathscr{E}})$.

By (§§) there is a $\epsilon_o > 0$ so that:

$$\delta_{k_o,\epsilon}/2\epsilon \geqq \dot{\delta}_{\theta^o,k_o}-\varkappa \quad \text{when} \quad \epsilon \leqq \epsilon_o$$

Let $\epsilon \leqq \epsilon_o$ and $k \geqq k_o$. Then:

$$\delta_{k,\epsilon}/2\epsilon \geqq \delta_{k_o,\epsilon}/2\epsilon \geqq \dot{\delta}_{\theta^o,k_o}-\varkappa \geqq \dot{\delta}_{\theta^o}-2\varkappa \geqq \dot{\delta}_{\theta^o,k}-2\varkappa$$

Finally choose $\epsilon_1 \in]0,\epsilon_o[$ so small that

$\delta_{k,\epsilon}/2\epsilon \geqq \dot{\delta}_{\theta^o,k}-2\varkappa$ when $k \leqq k_o$ and $\epsilon \leqq \epsilon_1$. Then $\epsilon \leqq \epsilon_1$ imply $\delta_{k,\epsilon}/2\epsilon-\dot{\delta}_{\theta^o,k} \geqq -2\varkappa$ for all k . Uniform convergence follows now from this and (§). □

<u>Corollary 6.7</u>

Let $\mathscr{E}_1,\ldots,\mathscr{E}_n$ and $\mathfrak{F}_1,\ldots,\mathfrak{F}_n$ be differentiable in θ^o . Then:

$$\dot{\delta}_{\theta^o,(k)}(\Pi\mathscr{E}_i,\Pi\mathfrak{F}_i) \leqq \Sigma\dot{\delta}_{\theta^o,(k)}(\mathscr{E}_i,\mathfrak{F}_i)$$

and

$$\dot{\Delta}_{\theta^o,(k)}(\Pi\mathscr{E}_i,\Pi\mathfrak{F}_i) \leqq \Sigma\dot{\Delta}_{\theta^o,(k)}(\mathscr{E}_i,\mathfrak{F}_i)$$

Proof: If $\mathcal{E}_1, \ldots, \mathcal{E}_n$ are dominated then the proof follows from theorem B.3.12 in the appendix. The general case now follows as in part (iv) of the proof of theorem 6.5. □

The next result implies that conditional expectations evaluated at θ's close to θ^o can not vary to much.

Proposition 6.8

Let $\mathcal{E} = ((\mathcal{X}, \mathcal{A}), (P_\theta : \theta \in \Theta))$ be differentiable in θ^o.
Let $\mathcal{B}$ be a sub σ-algebra of $\mathcal{A}$, X a bounded random variable and $E^{\mathcal{B}}_{\theta^o} X$ a bounded version. Then:

$$\sup_\theta E_\theta |E^{\mathcal{B}}_\theta X - E^{\mathcal{B}}_{\theta^o} X| / \|\theta - \theta^o\| < \infty$$

<u>Proof</u>: Let $-1 \leqq h \leqq 1$ be $\mathcal{B}$ measurable. Then:

$$E_\theta h(E_\theta^{\mathcal{B}} X - E_{\theta^o}^{\mathcal{B}} X) = E_\theta(hE_\theta^{\mathcal{B}} X) - E_\theta(hE_{\theta^o}^{\mathcal{B}} X) =$$

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

$$= \int h(X - E_{\theta^o}^{\mathcal{B}} X) d(P_\theta - P_{\theta^o}) = \int h(X - E_{\theta^o}^{\mathcal{B}} X) d[\Sigma(\theta_i - \theta_i^o)\dot{P}_{\theta^o,i} + \|\theta - \theta^o\| \Gamma_{\theta^o,\theta}]$$

$$= \sum_i (\theta_i - \theta_i^o) \int h(X - E_{\theta^o}^{\mathcal{B}} X) d\dot{P}_{\theta^o,i} + \|\theta - \theta^o\| \int h(X - E_{\theta^o}^{\mathcal{B}} X) d\Gamma_{\theta^o,\theta}$$

$$\leqq \sum_i |\theta_i - \theta_i^o| \max_i \int h(X - E_{\theta^o}^{\mathcal{B}} X) d\dot{P}_{\theta^o,i} + \|\theta - \theta^o\| \|h\| \|X - E_{\theta^o}^{\mathcal{B}} X\| \|\Gamma_{\theta^o,\theta}\|$$

$$\leqq \sqrt{r} \|\theta - \theta^o\| \|h\| \cdot \|X - E_{\theta^o}^{\mathcal{B}} X\| \max_i \|\dot{P}_{\theta^o,i}\| + \|\theta - \theta^o\| \|h\| \|X - E_{\theta^o}^{\mathcal{B}} X\| \|\Gamma_{\theta^o,\theta}\|$$

Hence

$$E_\theta |E_\theta^{\mathcal{B}} X - E_{\theta^o}^{\mathcal{B}} X| / \|\theta - \theta^o\| \leqq \|X - E_{\theta^o}^{\mathcal{B}} X\| [\max_i \|\dot{P}_{\theta^o,i}\| \sqrt{r} + \|\Gamma_{\theta^o,\theta}\|] < \infty .$$

□

$\dot{\Delta}_{\theta^o}$ sufficiency may be characterized by conditional expectations as follows

<u>Theorem 6.9</u>

Let $\mathcal{E} = ((\chi, \mathcal{A}), P_\theta : \theta \in \Theta))$ be differentiable in θ^o and let $\mathcal{B}$ be a sub σ-algebra of $\mathcal{A}$. Let $\hat{\mathcal{E}}$ be the experiment $\hat{\mathcal{E}} = ((\chi, \mathcal{B}), P_{\theta\mathcal{B}} : \theta \in \Theta)$ where - for each θ - $P_{\theta\mathcal{B}}$ is the restriction of P_θ to $\mathcal{B}$. Then :

(i) Suppose $\dot{\Delta}_{\theta^o}(\mathcal{C}, \hat{\mathcal{C}}) = 0$. Let X be a bounded random variable in $\mathcal{C}$ and $E_{\theta^o}^{\mathcal{B}} X$ a bounded version of this conditional expectation. Then:

$$\lim_{\theta \to \theta^o} E_\theta |E_\theta^{\mathcal{B}} X - E_{\theta^o}^{\mathcal{B}} X| / \|\theta - \theta^o\| \to 0 \quad \text{as} \quad \theta \to \theta^o$$

(ii) Let $\mathcal{A}_o$ be a basis for $\mathcal{A}$ which is closed under finite intersections. Suppose that to each $A \in \mathcal{A}_o$ there is a $\mathcal{B}$ measurable function Y_A so that $\lim_{\theta \to \theta^o} E_\theta |P_\theta^{\mathcal{B}}(A) - Y_A| / \|\theta - \theta^o\| = 0$. Then $\dot{\Delta}_{\theta^o}(\mathcal{C}, \hat{\mathcal{C}}) = 0$.

<u>Proof</u>: (i): Let X and $E_{\theta^o}^{\mathcal{B}} X$ be as in (i) and suppose $\dot{\Delta}_{\theta^o}(\mathcal{C}, \hat{\mathcal{C}}) = 0$. By proposition 4.6 $s_{\theta^o,i} \underset{=}{\overset{\text{definition}}{=}} d\dot{P}_{\theta^o,i}/dP_{\theta^o}$ may be chosen $\mathcal{B}$ measurable. Let $-1 \leq h \leq 1$ be $\mathcal{B}$ measurable. As in the proof of proposition 6.9 we get:

$$E_\theta h[E_\theta^{\mathcal{B}} X - E_{\theta^o}^{\mathcal{B}} X) = \Sigma(\theta_i - \theta_i^o) \int h(X - E_{\theta^o}^{\mathcal{B}} X) s_{\theta^o,i} \, dP_{\theta^o}$$

$$+ \|\theta - \theta^o\| \int h(X - E_{\theta^o}^{\mathcal{B}} X) d\Gamma_{\theta^o,\theta} = \|\theta - \theta^o\| \int h(X - E_{\theta^o}^{\mathcal{B}} X) d\Gamma_{\theta^o,\theta}$$

$$\leq \|\theta - \theta^o\| \|h\| \|X - E_{\theta^o}^{\mathcal{B}} X\| \|\Gamma_{\theta^o,\theta}\| .$$ In particular:

$$E_\theta |E_\theta^{\mathcal{B}} X - E_{\theta^o}^{\mathcal{B}} X| / \|\theta - \theta^o\| \to 0 \quad \text{as} \quad \theta \to \theta^o .$$

(ii): If suffices - by proposition 4.6 - to show that $s_{\theta^o,i} = d\dot{P}_{\theta^o,i}/dP_{\theta^o}$ may be specified $\mathcal{B}$ measurable for each $i = 1,2,\ldots,r$. Let $A \in \mathcal{A}_o$ and suppose Y_A is $\mathcal{B}$ measurable and satisfies:

$\lim_{\theta\to\theta^o} E_\theta |P_\theta^{\mathcal{B}}(A)-Y_A|/\|\theta-\theta^o\| = 0$. Then

$\lim_{h\to 0} E_{\theta^o+hv_i} |P_{\theta^o+hv_i}^{\mathcal{B}}(A)-Y_A|/|h| \to 0$ as $h \to 0$

It follows now from the one dimensional case - i.e. proposition 6.17 in LC1 - that $s_{\theta^o,i}$ may be specified $\mathcal{B}$ measurable. □

Let P be a probability distribution on R^r . For any pair (θ,σ) where $\theta \in R^r$ and $\sigma \in]0,\infty[^r$ let $Q_{\theta,\sigma}$ denote the probability distribution of $(\theta_i+\sigma_i U_i \; ; \; i = 1,2 \ldots r)$ when P is the probability distribution of $(U_i \; ; \; i = 1,2,\ldots,r)$. Put $P_\theta = Q_{\theta,1}$. The the experiment $\{Q_{\theta,\sigma} : \theta \in R^r\}$ is equivalent with the experiment $\{P_{(\theta_1/\sigma_1,\ldots,\theta_r/\sigma_r)} ; \theta \in R^n\}$, i.e. the scale change may be carried out in the parameter space. In general differentiable parameter transformations obey the chain rule for differentiation:

Proposition 6.10

Let $\eta^o \in \mathcal{H} \subseteq R^s$ and let γ be a function from $\mathcal{H}$ to Θ which is differentiable in η^o and maps η^o on θ^o . Suppose also that $\mathcal{E}$ is differentiable in θ^o and put $\mathcal{F} = ((\chi,\mathcal{A}),Q_\eta : \eta \in \mathcal{H})$, where $Q_\eta = P_{\gamma(\eta)} \; ; \; \eta \in \mathcal{H}$.

Then $\mathcal{F}$ is differentiable in η^o and:

$$\dot{Q}_{\eta^o,j} = \sum_i \dot{P}_{\theta^o,i} D_j\gamma^i(\eta_o)$$

where D_j indicates partial derivative w.r.t. η_j .

Appendix

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 $(D, \mathcal{J})$, 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-definient 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 $D_k = \{1,\ldots,k\}$ 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 $x \rightsquigarrow \psi(x) - \frac{1}{2}\sum_\theta(\psi(e_\theta) - \psi(e_\theta))x_\theta$ implies (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 $x \rightarrow |\sum_\theta a_\theta x_\theta|$. □

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 :

$$\psi(\mathcal{E}) = \tilde{\psi}(\mathcal{E}) \geq \tilde{\psi}(\mathcal{F}) - \tfrac{1}{2}\sum_{\theta}\epsilon_{\theta}(\tilde{\psi}(e_{\theta}) + \tilde{\psi}(-e_{\theta})) =$$

$$\psi(\mathcal{F}) - \tfrac{1}{2}\sum_{\theta}\epsilon_{\theta}(\psi(e_{\theta}) + \tilde{\psi}(-e_{\theta})) \geq \psi(\mathcal{F}) - \tfrac{1}{2}\sum_{\theta}\epsilon_{\theta}(\psi(e_{\theta}) + \psi(-e_{\theta})) .$$

□

Definitions

A standard pseudo experiment is a pseudo experiment of the form $((K,\mathcal{B}),(S_{\theta}: \theta \in \Theta))$ where $K = \{x:\ x \in R^{\Theta}$ and $\sum_{\theta}|x_{\theta}| = 1\}$, $\mathcal{B}$ is the class of Borel sub sets of K and $x \rightsquigarrow x_{\theta}$ 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.

The fact that Δ_2 equivalence for experiments implies equivalence is a particular case of:

<u>Theorem B.2.7</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.

Suppose there are points θ_0, θ_1 in Θ so that :

(i) $\mu_{\theta_0} \geqq 0$

(ii) $\mu_{\theta_0} >> \mu_\theta$ when $\theta \neq \theta_1$

<u>Then $\Delta(\mathcal{E},\mathcal{F}) = 0$ provided $\Delta_2(\mathcal{E},\mathcal{F}) = 0$.</u>

<u>Proof:</u>

Let $\hat{\mathcal{E}} = (S_\theta:\theta \in \Theta)$ and $\hat{\mathcal{F}} = (T_\theta:\theta \in \Theta)$ be the standard pseudo experiments of - respectively $\mathcal{E}$ and $\mathcal{F}$. Then $S = \sum_\theta |T_\theta|$ are, respectively, the standard measures of $\mathcal{E}$ and $\mathcal{F}$. Clearly (i) and (ii) hold for $\hat{\mathcal{E}}$ and $\hat{\mathcal{F}}$. Suppose $\Delta_2(\mathcal{E},\mathcal{F}) = 0.$ Then $\Delta_2(\hat{\mathcal{E}},\hat{\mathcal{F}}) = 0.$ We must show that $\Delta(\hat{\mathcal{E}},\hat{\mathcal{F}}) = 0.$ By assumption:

$$(\S) \quad \int(\Sigma a_\theta x_\theta)^+ \ S(dx) = \int(\Sigma a_\theta x_\theta)^+ \ T(dx) \ ; \quad a \in R^\Theta$$

Taking - respectively - the right hand and the left hand partial derivative w.r.t. a_θ we get:

$$(\S\S) \quad \int_{\Sigma a_\theta x_\theta > 0} x_\theta S(dx) + \int_{\Sigma a_\theta x_\theta = 0} x_\theta^+ \ S(dx) = \text{same expression in } T$$

$$(\S\S\S) \quad \int_{\Sigma a_\theta x_\theta > 0} x_\theta S(dx) + \int_{\Sigma a_\theta x_\theta = 0} x_\theta^- \ S(dx) = \text{same expression in } T$$

By Δ_2 equivalence:

$\int x_{\theta_o}^- T(dx) = \int x_{\theta_o}^- S(dx) = 0$. Hence $T_{\theta_o} \geqq 0$.

Subtraction of (§§§) from (§§) yields:

$$\int_{\Sigma a_\theta x_\theta = 0} |x_\theta| \; S(dx) = \int_{\Sigma a_\theta x_\theta = 0} |x_\theta| \; T(dx)$$

In particular

$$\int_{x_{\theta_o} = 0} |x_\theta| \; T(dx) = \int_{x_{\theta_o} = 0} |x_\theta| \; S(dx) = 0 \quad \text{when} \quad \theta \neq \theta_1 ,$$

It follows that $T_{\theta_o} \gg T_\theta$ when $\theta \neq \theta_1$.

By (§§):

$$S_{\theta_o}\big(\sum_{\theta \neq \theta_o} a_\theta (x_\theta / x_{\theta_o}) > a_{\theta_o}\big) = \text{same expression in } T$$

for all $a \in R^\Theta$. It follows that $x \rightsquigarrow x_\theta / x_{\theta_o}$; $\theta \neq \theta_o$ has the same distribution under S_{θ_o} as under T_{θ_o} ; i.e. $S_{\theta_o} = T_{\theta_o}$. Hence S and T are equal on $\{x : x_{\theta_o} > 0\}$, and we have seen that

$$x_\theta = 0 \quad \text{a.e.} \quad S + T \quad \text{on} \quad \{x : x_{\theta_o} = 0\} \quad \text{when} \quad \theta \neq \theta_1 .$$

It follows that the restrictions of S and T to $\{x : x_{\theta_o} = 0\}$ are consentrated on the two point set $\{v, w\}$ where $v_\theta = w_\theta = 0$ when $\theta \neq \theta_1$ and $v_{\theta_1} = - w_{\theta_1} = 1$. Now

$$S(\{v\}) + S(\{w\}) = S(x_{\theta_o} = 0) = \|S\| - S(x_{\theta_o} > 0)$$

$$= (\text{by } \Delta_2 \text{ equivalence}) \quad \|T\| - T(x_{\theta_o} > 0) = T(\{v\} + T(\{w\})$$

and

$$S(\{v\}) - S(\{w\}) = \int_{x_{\theta_o}=0} x_{\theta_1} \, S(dx) = \int x_{\theta_1} \, S(dx) - \int_{x_{\theta_o}>0} x_{\theta_1} \, S(dx)$$

$$= (\text{by } \Delta_1 \text{ equivalence}) \int x_{\theta_1} \, T(dx) - \int_{x_{\theta_o}>0} x_{\theta_1} \, T(dx)$$

$$= T(\{v\}) + T(\{w\}).$$

It follows that $S(\{v\}) = T(\{v\})$ and $S(\{w\}) = T(\{w\})$.

=

Example B.2.8.

Let μ_o, μ_1, μ_2, ν_o, ν_1, ν_2 be given by the matrix:

	1	2	3	4
μ_o	0	0	0	1
μ_1	0	$\frac{1}{4}$	$-\frac{1}{4}$	0
μ_2	$\frac{1}{4}$	0	$-\frac{1}{4}$	0
ν_o	0	0	0	1
ν_1	0	$-\frac{1}{4}$	$\frac{1}{4}$	0
ν_2	$-\frac{1}{4}$	0	$\frac{1}{4}$	0

Then : $\|a_o\mu_o + a_1\mu_1 + a_2\mu_2\| = |a_o| + \|a_1\mu_1 + a_2\mu_2\|$

$= |a_o| + \|a_1\nu_1 + a_2\nu_2\| = \|a_o\nu_o + a_1\nu_1 + a_2\nu_2\|$

and $\mu_i(\chi) = \nu_i(\chi)$; $i = 0,1,2$. It follows that $\Delta_2((\mu_o,\mu_1,\mu_2), (\nu_o,\nu_1,\nu_2)) = 0$. By example B.2.6, however $\Delta_3((\mu_o,\mu_1,\mu_2), (\nu_o,\nu_1,\nu_2)) \geq \Delta_3((\mu_1,\mu_2), (\nu_1,\nu_2)) > 0$.

This show that assumption (ii) can not be deleted in proposition B.2.7.

Another situation in which it is permissible to conclude Δ equivalence from Δ_2 equivalence is the case of ordered pseudo experiments. This is the content of:

<u>Theorem B.2.9</u>

Let $\mathcal{E}$ and $\mathcal{F}$ be pseudo experiments. Then $\Delta(\mathcal{E},\mathcal{F}) = 0$ if and only if $\delta(\mathcal{E},\mathcal{F}) = \Delta_2(\mathcal{E},\mathcal{F}) = 0$.

<u>Proof:</u>

The "only if" is obvious so suppose $\delta(\mathcal{E},\mathcal{F}) = \Delta_2(\mathcal{E},\mathcal{F}) = 0$. Let S and T be the standard measures of, respectively, $\mathcal{E}$ and $\mathcal{F}$. By Δ_2 equivalence $\int \psi dS = \int \psi dT$ when $\psi \in \Psi_2$. Let $\psi \in \Psi$, $1 \in \Psi_1$ and $b \geq 0$. Then $\varphi_b = 0 \vee 1 \vee b\psi \in \Psi$ and $\varphi_o = 1^+ \in \Psi_2$. Hence:

$$\int (\varphi_b - \varphi_o)/b dS = \left(\int \varphi_b dS - \int \varphi_o dS\right)/b \geq \left(\int \varphi_b dT - \int \varphi_o dT\right)/b$$
$$= \int (\varphi_b - \varphi_o)/b dT.$$

$b \downarrow 0$ yields:

$$(\dagger) \quad \int_{1 \leq 0} \psi^+ dS \geq \int_{1 \leq 0} \psi^+ dT$$

The derivation of (§§) and (§§§) in the proof of theorem B.2.7 holds without changes. By subtracting (§§§) from (§§) we get:

$$(\dagger\dagger) \quad \int_{1=0} |x_\theta| \, S(dx) = \int_{1=0} |x_\theta| \, T(dx)$$

Substituting $|x_\theta|$ for ψ in $(\dagger)$ we get:

$$\int_{1 \leq 0} |x_\theta| \, S(dx) \geq \int_{1 \leq 0} |x_\theta| \, T(dx)$$

Hence - by (††) :

(*) $\int_{1<0} |x_\theta|\ S(dx) > \int_{1<0} |x_\theta|\ T(dx)$

Replacement of 1 with -1 yields

(**) $\int_{1>0} |x_\theta|\ S(dx) > \int_{1>0} |x_\theta|\ T(dx).$

By Δ_2 equivalence the sum, $\int |x_\theta|\ S(dx)$, of the left hand sides of (††), (*) and (**) are equal to the sum, $\int |x_\theta|\ T(dx)$, of the right hand sides of the same inequalities. It follows that the inequalities in (*) and (**) may be replaced by equalities, i.e.

$$\int_{1<0} |x_\theta|\ dS = \int_{1<0} |x_\theta|\ dT \quad \text{for any } \theta \in \Theta \text{ and any } 1 \in \Psi_1.$$

Combining this with (††) we find that:

$$\int_{1\leqq 0} |x_\theta|\ dS = \int_{1\leqq 0} |x_\theta|\ dT \quad \text{for any } \theta \in \Theta \text{ and any } 1 \in \Psi_1.$$

Substituting x_θ^+ and x_θ^- in (†) we get:

$$\int_{1\leqq 0} x_\theta^+\ dS \geq \int_{1\leqq 0} x_\theta^+\ dT$$

and

$$\int_{1\leqq 0} x_\theta^-\ dS \geqq \int_{1\leqq 0} x_\theta^-\ dT.$$

Hence - since the sum of the left hand sides equals the sum of the right hand sides:

$$S_{\theta_o}^+(1\leqq 0) = T_{\theta_o}^+(1\leqq 0)$$

and $$S_{\theta_o}^-(1\leq 0) = T_{\theta_o}^-(1\leq 0) \quad \text{for any } \theta_o \in \Theta.$$

It follows that for any $a \in R^{\Theta}$:

$$S_{\theta_o}^+\left(\sum_{\theta\neq\theta_o} a_\theta(x_\theta/x_{\theta_o}) \leqq -a_{\theta_o}\right) = \text{same expression in } T$$

and

$$S_{\theta_o}^-\left(\sum_{\theta\neq\theta_o} a_\theta(x_\theta/x_{\theta_o}) \geqq -a_{\theta_o}\right) = \text{same expression in } T.$$

Hence - since $x_\theta = (x_\theta/x_{\theta_o})[1 + \sum_{\theta\neq\theta_o} |x_\theta/x_{\theta_o}|]^{-1}$ on

$\{x : x_{\theta_o} > 0 \ ; \ x \in K\}$ and $x_\theta = (x_\theta/x_{\theta_o})[-1 + \sum_{\theta\neq\theta_o} |x_\theta/x_{\theta_\emptyset}|]^{-1}$

on $\{x : x_{\theta_o} < 0 \ ; \ x \in K\}$ - :

$$S_{\theta_o}^+ = T_{\theta_o}^+ \quad \text{and} \quad S_{\theta_o}^- = T_{\theta_o}^-.$$

It follows that $S_{\theta_o} = T_{\theta_o}$ for any $\theta_o \in \Theta$. □

Somewhat surprisingly Δ_3 equivalence will always imply Δ equivalence.

Theorem B.2.10. Let $\mathcal{E}$ and $\mathcal{F}$ be speudo experiments. Then $\Delta(\mathcal{E},\mathcal{F}) = 0$ if and only if $\Delta_3(\mathcal{E},\mathcal{F}) = 0$.

Proof:

We use the notations of the proof of theorem B.2.9. If $\psi \in \Psi_1$ then $\varphi_b \in \Psi_3$. Hence (†) holds with "≧" replaced by "=" i.e.:

$$\int_{1\leqq 0} \psi^+ \, dS = \int_{1\leqq 0} \psi^+ \, dT$$

Substituting x_{θ_0} and $-x_{\theta_0}$ for ψ we get:

$$S_{\theta_0}^{+}(1 \leqq 0) = T_{\theta_0}^{+}(1 \leqq 0)$$

and
$$S_{\theta_0}^{-}(1 \leqq 0) = T_{\theta_0}^{-}(1 \leqq 0).$$

The final argument is the same as in the proof of theorem B.2.9. □

Here is the factorization criterion for Δ sufficiency:

Theorem B.2.11.

Let $\mathcal{E} = ((\chi,\mathcal{A}),(\mu_\theta : \theta \in \Theta))$ be a pseudo experiment and let $\mathcal{B}$ be a sub σ-algebra of $\mathcal{A}$. Denote by $\mathcal{F}$ the pseudo experiment $\mathcal{F} = ((\chi,\mathcal{B})),(\mu_{\theta\mathcal{B}};\ \theta \in \Theta))$ where the subscript $\mathcal{B}$ indicates restriction to $\mathcal{B}$. The following conditions are equivalent:

(i) $\Delta(\mathcal{E},\mathcal{F}) = 0$

(ii) $d\mu_\theta / d\sum_\theta |\mu_\theta|$ may - for each θ - be specified $\mathcal{B}$ measurable.

In order to prove the theorem we need:

Proposition B.2.10.

Let X be an integrable random variable on a probability space $(\chi,\mathcal{A},P)$ and let $\mathcal{B}$ be a sub σ-algebra of $\mathcal{A}$. Then $E|X| = E|E^{\mathcal{B}}X|$ if and only if there exist a $\mathcal{B}$ measurable random variable Y so that:

$$|Y| = 1 \quad \text{a.s} \quad P$$

$$XY = |X| \quad \text{a.s} \quad P$$

Proof:

Suppose the condition hold. Then $E|E^{\mathfrak{B}}X| = E|E^{\mathfrak{B}}Y|X|| = E|YE^{\mathfrak{B}}|X|| = EE^{\mathfrak{B}}|X| = E|X|$. To prove the converse, suppose $E|X| = E|E^{\mathfrak{B}}X|$. Then $E[E^{\mathfrak{B}}|X| - |E^{\mathfrak{B}}X|] = 0$ so that $E^{\mathfrak{B}}|X| = |E^{\mathfrak{B}}X|$ a.s , i.e. $E^{\mathfrak{B}}X^+ + E^{\mathfrak{B}}X^- = |E^{\mathfrak{B}}X^+ - E^{\mathfrak{B}}X^-|$ a.s. Put $Y = 1$ on $E^{\mathfrak{B}}X^- = 0$ and $Y = -1$ on $E^{\mathfrak{B}}X^- > 0$. Then

$$\int_{Y=1} X^- \, dP = \int_{Y=1} E^{\mathfrak{B}}X^- \, dP = 0 = \int_{Y=-1} E^{\mathfrak{B}}X^+ \, dP = \int_{Y=-1} X^+ \, dP.$$

It follows that $X \geqq 0$ a.s on $[Y = 1]$ and that $X \leqq 0$ a.s on $Y = -1$. □

Proof of theorem B.2.11: (ii) => (i) is clear.
Suppose $\Delta(\mathcal{E}, \mathcal{F}) = 0$. We may, without loss of generality, assume $c = \sum_\theta \|\mu_\theta\| > 0$. Let E denote expectation w.r.t. $\pi = c^{-1} \sum_\theta |\mu_\theta|$. We must show that $f_\theta = d\mu_\theta/d\pi$ may be specified $\mathfrak{B}$ measurable. Note that $\sum_\theta |f_\theta| = c$; a.s π.

By Δ equivalence $S = T$ where S and T are, respectively the standard measures of $\mathcal{E}$ and $\mathcal{F}$. Let h be bounded measurable on R^Θ. Then

$$\int h(f_\theta : \theta \in \Theta) d\pi = \int c^{-1} h(c d\mu_\theta / d\textstyle\sum_\theta |\mu_\theta| \; ; \; \theta \in \Theta) d\textstyle\sum_\theta |\mu_\theta|$$

$$= \int c^{-1} h(cx_\theta : \theta \in \Theta) dS = \int c^{-1} h(cx_\theta : \theta \in \Theta) dT$$

$$= \int c^{-1} (h(c \, d\mu_{\theta\mathfrak{B}} / d\textstyle\sum_\theta |\mu_{\theta\mathfrak{B}}| \; ; \; \theta \in \Theta) d\textstyle\sum_\theta |\mu_{\theta\mathfrak{B}}|.$$

Now $\qquad g_\theta = d\mu_{\theta\mathfrak{B}} / d\pi_{\mathfrak{B}}$.

Hence $\qquad |g_\theta| = d|\mu_{\theta\mathfrak{B}}| d\pi_{\mathfrak{B}}$

and $$\sum_\theta |g_\theta| = d\sum_\theta |\mu_{\theta\mathfrak{B}}|/d\pi_{\mathfrak{B}} .$$

By Δ_2 equivalence

$$E|g_\theta| = E|f_\theta|.$$

It follows from proposition B.2.12 that $|f_\theta| = h_\theta f_\theta$ where h_θ is $\mathfrak{B}$ measurable and $|h_\theta| = 1$. Hence $\sum_\theta |g_\theta| = \sum_\theta |E_\theta^{\mathfrak{B}} f_\theta| = \sum_\theta |E_\theta^{\mathfrak{B}} h_\theta |f_\theta|| = \sum_\theta |h_\theta| E_\theta^{\mathfrak{B}} |f_\theta| = \sum_\theta E^{\mathfrak{B}} |f_\theta| = E^{\mathfrak{B}} \sum_\theta |f_\theta| = E^{\mathfrak{B}} c = c.$

Hence $$\sum_\theta |\mu_{\theta\mathfrak{B}}| = c\pi$$ so that:

$$\int h(f_\theta : \theta \in \Theta) d\pi = \int h(g_\theta : \theta \in \Theta) d\pi.$$

In particular $\mathcal{L}(f_\theta) = \mathcal{L}(E^{\mathfrak{B}} f_\theta)$.

By proposition 5.7 in L C 1 $f_\theta = E^{\mathfrak{B}} f_\theta$ a.s π. □

Various convergence criterions are listed in:

<u>Theorem B.2.13.</u>

Let $\mathcal{E}$, , $\mathcal{E}_1$, $\mathcal{E}_2$,... be pseudo experiments with respectively, standard measures $S, S_1, S_2, \ldots$. Then * the following conditions are equivalent:

) If μ and ν are finite measures on R^Θ then the Levy distance $\Lambda(\mu,\nu)$ is defined as $\inf\{h : h > 0$ and $\mu(\Pi]-\infty, x_\theta - h[) - h \leq \nu(\prod_\theta]-\infty, x_\theta[) \leq \mu(\prod_\theta]-\infty, x_\theta + h[) + h ;$ $x \in R^\Theta\}$. Λ convergence is the same as weak convergence on (R^Θ) i.e. $\Lambda(\mu_n,\mu) = 0$ if and only if $\mu_n(f) \to \mu(f)$ when f is bounded and continuous.

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

(ii) $\lim_{n\to\infty} \Delta_3(\mathcal{E}_n, \mathcal{E}) = 0$

(iii) $\lim_{n\to\infty} \Lambda(S_n, S) = 0.$

If $\mathcal{E}$ satisfies (i) and (ii) in theorem B.2.7 then $\lim_{n\to\infty} \Delta(\mathcal{E}_n, \mathcal{E}) = 0$ if and only if $\lim_{n\to\infty} \Delta_2(\mathcal{E}_n, \mathcal{E}) = 0$. This hold also when either $\mathcal{E}_n \geq \mathcal{E}$ for all n or $\mathcal{E}_n \leq \mathcal{E}$ for all n.

Proof.

1°. Suppose $\Lambda(S_n, S) \to 0$. Let Φ be the class of restrictions ψ/K where $\psi \in \Psi$ satisfies

$$\sum_\theta \max\{\psi(-e_\theta), \psi(e_\theta)\} \leq 1.$$

Then Φ is a compact subset of $C(K)$ and S_n; $n = 1,2,\ldots$ are - since $\sup_n \|S_n\| < \infty$ - uniformly equicontinuous on Φ. By Ascoli's theorem: $\lim_n S_n(\varphi)$, uniformly in φ. Δ convergence follows now from corollary B.2.2.

2°. Suppose $\Delta_3(\mathcal{E}_n, \mathcal{E}) \to 0$. Then $\|S_n\| \to \|S\|$. It follows that S_n; $n = 1,2 \ldots$ is conditionally Λ compact. Λ convergence follows now from 1° and theorem B.2.10.

3°. The last statements follows easily from theorem B.2.7 and theorem B.2.9. □

Corollary B.2.14.

The pseudometrics $\Delta_1, \Delta_2, \ldots, \Delta$ are complete.

Proof:

Let d denote one of the pseudometrics $\Delta_2, \Delta_3, \ldots, \Delta$.

Consider a sequence $\mathcal{E}_1, \mathcal{E}_2, \ldots$ such that $d(\mathcal{E}_m, \mathcal{E}_n) \to 0$ as $m,n \to \infty$. Let S_n be the standard measure of $\mathcal{E}_n$. Then $\|S_m\| - \|S_n\| \to 0$ as $m,n \to \infty$ so that $\sup_n \|S_n\| < \infty$. It follows that there is a sub sequence $S_{n'}$ and a S so that $\Lambda(S_{n'}, S) \to 0$. Then – by theorem B.2.13 – $d(\mathcal{E}_{n'}, \mathcal{E}) \to 0$. The proof for Δ_1 is trivial. □

Corollary B.2.15.

Let $\Delta_{(k)}$ denote one of the pseudometrics $\Delta_2, \Delta_3, \ldots, \Delta$. A family $\mathcal{E}_t = ((\chi_t, \mathcal{A}_t) ; \mu_{\theta t} : \theta \in \Theta) ; t \in T$ of pseudo experiments is $\Delta_{(k)}$ conditionally compact if and only if $\sup_t \|\mu_{\theta t}\| < \infty ; \theta \in \Theta$.

Proof:

Follows directly from theorem B.2.13. □

Generalizing theorem B.2.11. to the asymptotic case we get:

Theorem B.2.16.

Let $\mathcal{E}_n = ((\chi_n, \mathcal{A}_n ; \mu_{\theta n}, \theta \in \Theta) ; n = 1,2,\ldots$ be a sequence of pseudo experiments. For each n, let $\mathcal{B}_n$ be a sub σ algebra of $\mathcal{A}_n$, and let $\mathcal{F}_n$ denote the restriction of $\mathcal{E}_n$ to $\mathcal{B}_n$. Let π_n ; $n = 1,2\ldots$ be any probability measure on $\mathcal{A}_n$ such that $\sum_\theta \|\mu_{\theta n}\| \pi_n = \sum_\theta |\mu_{\theta n}|$. Expectations w.r.t π_n will be written E_n.

The measure on $\mathcal{A}_n$ whose Radon Nikodym derivative w.r.t π_n is $E_n^{\mathcal{B}_n}[d\mu_{\theta n}/d\pi_n]$ will be denoted by $\hat{\nu}_{\theta n}$. Put $\hat{\mathcal{F}}_n = ((\chi_n, \mathcal{A}_n), \hat{\nu}_{\theta n} ; \theta \in \Theta$. Finally the standard measures of $\mathcal{E}_n$ and $\mathcal{F}_n$ will be denoted by, respectively, S_n and T_n.

Then $\Delta(\mathcal{F}_n, \tilde{\mathcal{F}}_n) = 0$; $n = 1,2,\ldots$.
Suppose $\mathcal{E}_1, \mathcal{E}_2,\ldots$ are conditionally compact. Then $\mathcal{F}_1, \tilde{\mathcal{F}}_2,\ldots$ are conditionally compact and the follwing conditions are equivalent:

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

(ii) $\lim_{n\to\infty} \Delta_2(\mathcal{E}_n, \tilde{\mathcal{F}}_n) = 0$

(iii) $\lim_{n\to\infty} \Lambda(S_n, T_n) = 0$

(iv) $\lim_{n\to\infty} \|\mu_{\theta n} - \hat{\nu}_{\theta n}\| = 0$; $\theta \in \Theta$

(v) $\lim_{n\to\infty} \Lambda\left[\mathcal{L}(d\mu_{\theta n}/d\pi_n;\ \theta \in \Theta),\ \mathcal{L}(d\hat{\nu}_{\theta n}/d\mu_n;\ \theta \in \Theta)\right] = 0$.

Remark.

It will follow from theorem B. 3.10 that (i) - provided $\mathcal{E}_1, \mathcal{E}_2,\ldots$ are conditionally compact - is equivalent with:

(i') $\lim_{n\to\infty} \Delta(\mathcal{E}_{n,\theta_1,\theta_2}, \mathcal{F}_{n,\theta_1,\theta_2}) = 0$; $\theta_1,\theta_2 \in \Theta$.

Proof of the theorem:

1°. Put $f_{\theta n} = d\mu_{\theta n}/d\pi_n$ and $g_{\theta n} = d\nu_{\theta n}/d\pi_n$ where $\nu_{\theta n}$ is the restriction of $\mu_{\theta n}$ to $\mathcal{B}_n$ and π_n - by abuse of notations - is the restriction of π_n to $\mathcal{B}_n$. Then $g_{\theta n} = E_n^{\mathcal{B}_n} f_{\theta n} = d\hat{\nu}_{\theta n}/d\pi_n$, $|f_{\theta n}| = d|\mu_{\theta n}|/d\pi_n$ and $\sum_\theta |f_{\theta n}| = \sum_\theta \|\mu_{\theta n}\|$ a.s. π_n.

Let $B_n \in \mathcal{B}_n$. Then $\hat{\nu}_{\theta n}(B_n) = \int_{B_n} g_{\theta n} d\pi_n = \nu_{\theta n}(B_n)$.

By theorem B.2.11, $\Delta(\tilde{\mathcal{F}}_n, \mathcal{F}_n) = 0$; $n = 1,2,\ldots$

2°. Suppose $\mathcal{E}_1, \mathcal{E}_2,\ldots$ are conditionally compact. Then - by corollary B.2.15 - $\sup_n\|\nu_{\theta n}\| \leq \sup_n\|\mu_{\theta n}\| < \infty$. By the same corollary ; $\mathcal{F}_1, \mathcal{F}_2, \ldots$ are conditionally compact. (i) <=> (ii) <=> (iii) follows from theorem B.2.13. (iv) <=> (v) follows by applying proposition 5.7 in LC 1 to linear combinations of densities. (iv) => (i) follows from part 1°. It remains to show that (iii) => (v). Suppose (iii) holds. Then - as we have seen - (ii) and (i) hold. We may - without loss of generality assume $\Delta(\mathcal{E}_n, \mathcal{E}) \to 0$ and $\Delta(\mathcal{F}_n, \mathcal{E}) \to 0$ where $\mathcal{E}$ has standard measure S. We have $\pi_n(|g_{\theta n}| \geq M) \leq \|\nu_{\theta n}\|/M \leq \|\mu_{\theta n}\|/M$ and $\pi_n(|f_{\theta n}| \geq M) \leq \|\mu_{\theta n}\|/M$. It follows that we may - without loss of generality - assume that $\mathcal{L}_{\pi_n}(f_n) \to P$ and that $\mathcal{L}_{\pi_n}(g_n) \to Q$ where $f_n = f_{\theta,n}$; $\theta \in \Theta$ and $g_n = g_{\theta,n}$; $\theta \in \Theta$. We must show that $P = Q$. Let $\Psi \in \psi$. Then

$$\int \psi dP = \lim_n \int \psi(f_n) d\pi_n = \lim_n \int \psi dS_n = \int \psi dS$$

$$= \lim_n \int \psi dT_n = \lim_n \int \psi(g_n) d\pi_n = \int \psi dQ.$$

Suppose $\|S\| > 0$ and let $h \in C(R^{\Theta})$. Then

$$\int h dP = \lim_n \int h(f_n) d\pi_n$$

$$= \lim_n \int h\left[(d\mu_{\theta n}/d\textstyle\sum_\theta \mu_{\theta n}) \|S_n\|; \theta \in \Theta\right] d\textstyle\sum_\theta |\mu_{\theta n}| / \|S_n\|$$

$$= \lim_n \int h(x_\theta \|S_n\|; \theta \in \Theta) dS_n/\|S_n\| = \int h(x_\theta \|S\|; \theta \in \Theta dS/\|S\|.$$

Hence $P = \mathcal{L}(X\|S\|)$ when $\mathcal{L}(X) = S/\|S\|$. In particular $P(\sum_\theta |X_\theta| = \|S\| = 1$, where $X_\theta(x) = x_\theta$; $\theta \in \Theta$, $x \in R^{\Theta}$.

Δ_2 convergence imply:

$E_n|f_{\theta n}| - E_n|g_{\theta n}| \to 0$; $\theta \in \Theta$.

It follows that

$$E_n|E^{\mathcal{S}_n} \sum_\theta |f_{\theta n}| - \sum_\theta |g_{\theta n}|| =$$

$$E_n(E^{\mathcal{S}_n} \sum_\theta |f_{\theta n}| - \sum_\theta |g_{\theta n}|) =$$

$$= \sum_\theta E_n|f_{\theta n}| - \sum_\theta E_n|g_{\theta n}| \to 0 .$$

Hence – since $\sum_\theta |f_{\theta n}| = \|S_n\|$ a.s. π_n – :

$\sum_\theta |g_{\theta n}| \to \|S\|$; π_n. It follows that

$$\mathcal{L}_Q(\sum_\theta |X_\theta|) = \lim_n \mathcal{L}_{\pi_n g_n^{-1}}(\sum_\theta |X_\theta|)$$

$$= \lim_n \mathcal{L}_{\pi_n}(\sum_\theta |g_{\theta n}|) = \text{the one point distribution}$$

in $\|S\|$, so that $Q(\sum_\theta |X_\theta| = \|S\|) = 1$.

Put $\tilde{P} = \mathcal{L}_P(X/\|S\|)$ and $\tilde{Q} = \mathcal{L}_Q(X/\|S\|)$. Then $\tilde{P}$ and $\tilde{Q}$ are standard (probability) measures and

$$\int \psi d\tilde{P} = \int \psi(x/\|S\|)P(dx) = \|S\|^{-1} \int \psi dP$$

$$= \|S\|^{-1} \int \psi dQ = \int \psi(x/\|S\|)dQ = \int \psi d\tilde{Q}.$$ Hence $\tilde{P} = \tilde{Q}$

so that $P = Q$.

If $\|S\| = 0$, then $\sum_\theta |f_{\theta n}| \to 0$; π_n, and $\sum_\theta |g_{\theta n}| \to 0$; π_n so that $P = Q =$ the one point distribution in $0 \in R^\Theta$. □

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

$$\|\mu_\theta\rho^F - \nu_\theta\sigma\| \le \epsilon_\theta \;;\; \theta \in F .$$

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

$$\|\mu_\theta\rho - \nu_\theta\sigma\| \le \epsilon_\theta : \theta \in \Theta . \qquad \square$$

We proved in fact a little more and this is the content of the next theorem.

Theorem B.3.2

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.

Proposition B.3.3

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.

---

each randomization σ from $(\mathcal{M},\mathcal{B})$ 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 $(\mathcal{M},\mathcal{B})$ to $(D,\mathcal{S})$. Define the randomization σ_k from $(\mathcal{M},\mathcal{B})$ 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 $f \in \mathcal{H}$, $\sum_{i=1}^{k}\rho_k(d_i|\cdot)f(d_i);\ k = 1,2,\ldots$ 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

$$\left| \int [\int f(d')\rho(dd'|\cdot) - \sum_{i=1}^{k} f(d_i)\rho_{k'}(d_i|\cdot)] d\mu_\theta \right|$$

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 φ: $\varkappa \to \varkappa' + [\varkappa(D) - \varkappa'(D)]\pi$ 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 $(\nu_\theta | \tilde{\mathcal{B}})\sigma = \nu_\theta | \tilde{\mathcal{B}} : \theta \in \Theta$ - :

$$\|\mu_\theta \rho - \nu_\theta | \tilde{\mathcal{B}}\| \leq \epsilon_\theta; \quad \theta \in \Theta .$$

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].

□

If $\mathcal{F}$ is a sub pseudo experiment of a pseudo experiment $\mathcal{E}$ then $\delta(\mathcal{E},\mathcal{F}) = 0$. More generally $\delta(\mathcal{E},\mathcal{F}) = 0$ provided there are pseudo experiments $\tilde{\mathcal{E}}$ and $\tilde{\mathcal{F}}$ so that $\Delta(\mathcal{E},\tilde{\mathcal{E}}) = \Delta(\mathcal{F},\tilde{\mathcal{F}}) = 0$ and $\tilde{\mathcal{F}}$ is a sub pseudo experiment of $\tilde{\mathcal{E}}$. We will now prove a result in the opposite direction.

<u>Theorem B.3.7</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. Suppose $\mathcal{E}$ is dominated, that $\mathcal{Y}$ is a Borel sub set of a Polish space and that $\mathcal{B}$ is the class of Borel sub sets of $\mathcal{Y}$. Then $\delta(\mathcal{E},\mathcal{F}) = 0$ if and only if there are pseudo experiments $\tilde{\mathcal{E}}$ and $\tilde{\mathcal{F}}$ such that $\Delta(\mathcal{E},\tilde{\mathcal{E}}) = \Delta(\mathcal{F},\tilde{\mathcal{F}}) = 0$ and $\tilde{\mathcal{F}}$ is a sub pseudo experiment of $\tilde{\mathcal{E}}$. If so, then $\tilde{\mathcal{E}}$ and $\tilde{\mathcal{F}}$ may be chosen so that:

$$\tilde{\mathcal{E}} = ((\chi \times \mathcal{Y}, \mathcal{A} \times \mathcal{B}); \lambda_\theta:\theta \in \Theta) ,$$

$\tilde{\mathcal{F}}$ is the restriction of $\tilde{\mathcal{E}}$ to $\mathcal{X} \times \mathcal{B}$, λ_θ has - for each θ - marginals, respectively, μ_θ and ν_θ . In particular $\mathcal{A} \times \mathcal{X}$ is sufficient in $\tilde{\mathcal{E}}$; i.e. the restriction of $\tilde{\mathcal{E}}$ to $\mathcal{A} \times \mathcal{X}$ is Δ equivalent with $\mathcal{E}$.

<u>Proof</u>:

It suffices to prove the "only if" so suppose $\delta(\mathcal{E},\mathcal{F}) = 0$. By corollary B.3.5 there is a randomization M from $(\chi,\mathcal{A})$ to $(\mathcal{Y},\mathcal{B})$ so that $\mu_\theta M = \nu_\theta$; $\theta \in \Theta$. Then λ_θ may be defined by:

$$\lambda_\theta(A \times B) = \int_A M(B|x)\mu_\theta(dx)$$

The proof is now completed by checking that the last statements hold with this choice of $\lambda_\theta:\theta \in \Theta$. □

Theorem B.2.7 and theorem B.3.1 yield:

Theorem B.3.8

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. Suppose that there are points θ_0,θ_1 in Θ so that:

(i) $\mu_{\theta_0} \geqq 0$

(ii) $\mu_{\theta_0} \gg \mu_\theta$ when $\theta \neq \theta_1$

Then $\Delta(\mathcal{E},\mathcal{F}) = 0$ provided $\Delta_2(\mathcal{E},\mathcal{F}) = 0$.

Theorem B.2.10 and theorem B.3.1 yield:

Thorem B.3.9

Let $\mathcal{E}$ and $\mathcal{F}$ be two pseudo experiments. Then $\Delta(\mathcal{E},\mathcal{F}) = 0$ provided $\Delta_3(\mathcal{E},\mathcal{F}) = 0$ and $\mathcal{E}$ is dominated *)

The usual criterions for sufficiency follows from:

Theorem B.3.10

Let $\mathcal{E} = ((\chi,\mathcal{A}),\mu_\theta:\theta \in \Theta)$ be a dominated experiment and let $\mathcal{B}$ be a sub σ algebra of $\mathcal{A}$. Denote by $\mathcal{F}$ the restriction of $\mathcal{E}$ to $\mathcal{B}$. Let c be a non negative function on Θ with countable support such that $\pi = \sum_\theta c(\theta)\|\mu_\theta\|$ is a probability measure

---

*) A pseudo experiment $\mathcal{E} = ((\chi,\mathcal{A}),(\mu_\theta:\theta \in \Theta))$ will be called dominated if there is a probability measure π on $\mathcal{A}$ so that $\pi \gg |\mu_\theta|$; $\theta \in \Theta$.

dominating *) $|\mu_\theta|$; $\theta \in \Theta$. Then the following conditions are equivalent:

(i) $\Delta(\mathcal{E},\mathcal{F}) = 0$

(ii) $d\mu_\theta/d\pi$ may - for each θ - be specified $\mathcal{S}$ measurable

(iii) $\Delta(\mathcal{E}_{\theta_1,\theta_2},\mathcal{F}_{\theta_1,\theta_2}) = 0$ when $\theta_1 \neq \theta_2$

<u>Proof</u>:

(ii) $\Rightarrow$ (i) $\Rightarrow$ (iii) is straight forward so it remains to prove that (iii) $\Rightarrow$ (ii). Suppose (iii) and let $\theta_o \in \Theta$. By theorem B.2.11 there is for each θ a $\mathcal{S}$ measurable version h_θ of $d\mu_{\theta_o}/d[|\mu_{\theta_o}|+|\mu_\theta|]$. Then $|h_\theta| = d|\mu_{\theta_o}|/d[|\mu_{\theta_o}|+|\mu_\theta|]$ so that $|\mu_{\theta_o}|[h_\theta=0] = 0$. h_θ may, and will, be chosen so that $|h_\theta| \leqq 1$. Put $N = \cup\{[h_\theta=0] \,;\, c(\theta) > 0\}$. Then $N \in \mathcal{S}$ and $|\mu_{\theta_o}|(N) = 0$. Put $f = [\Sigma c(\theta)(1-|h_\theta|)|h_\theta|^{-1}]^{-1}$ on $\complement N$ where Σ is over $\{\theta : c(\theta) > 0\}$. Put $f = 0$ on N . Then f is $\mathcal{S}$ measurable. Let g be a non negative $\mathcal{S}$ measurable function. Then:

$$\int_{\complement N} g\,d\pi = \Sigma c(\theta)\int_{\complement N} g\,d|\mu_\theta| = \Sigma c(\theta)\int_{\complement N} g(1-|h_\theta|)|h_\theta|^{-1}|h_\theta|\,d[|\mu_{\theta_o}|+|\mu_\theta|]$$

$$= \Sigma c(\theta)\int_{\complement N} g(1-|h_\theta|)|h_\theta|^{-1}d|\mu_{\theta_o}| = \int_{\complement N} g\Sigma c(\theta)(1-|h_\theta|)|h_\theta|^{-1}d|\mu_{\theta_o}| .$$

---

*) It is well known that a function c with these properties exists.

It follows that $d\pi/d|\mu_{\theta_0}| = \Sigma c(\theta)(1-|h_\theta|)|h_\theta|^{-1}$ on $\complement N$.

Hence - since $\pi \sim |\mu_{\theta_0}|$ on $\complement N$ - $d|\mu_{\theta_0}|/d\pi = f$ on $\complement N$.

It follows that $d|\mu_{\theta_0}|/d\pi = f$ on χ so that $d\mu_{\theta_0}/d\pi = \operatorname{sgn} h_{\theta_0} f$, and this specification is $\mathfrak{S}$ measurable. □

Corollary B.3.11

Let $\mathcal{E}$ and $\mathcal{F}$ be pseudo experiments such that $\delta(\mathcal{E},\mathcal{F}) = 0$ and $\mathcal{E}$ is dominated. Then $\Delta(\mathcal{E},\mathcal{F}) = 0$ provided $\Delta(\mathcal{E}_{\theta_1,\theta_2},\mathcal{F}_{\theta_1,\theta_2}) = 0$ when $\theta_1 \neq \theta_2$

Proof:

It follows from theorem B.3.1 that we may, without loss of generality, assume that Θ is finite. We may then - again without loss of generality - assume that $\mathcal{E}$ and $\mathcal{F}$ are standard pseudo experiments. The corollary follows now from theorem B.3.1 and theorem B.3.10. □

Let $\mathcal{E}_i = ((\chi_i, \mathcal{A}_i); \mu_{\theta i}; \theta \in \Theta)$, $i = 1,\ldots,n$ be n pseudo experiments. Their product $\mathcal{E}$ is then defined as

$$\mathcal{E} = (\prod_i(\chi_i, \mathcal{A}_i), \prod_i \mu_{\theta_i}; \theta \in \Theta)$$

and this pseudo experiment will be written as $\mathcal{E} = \prod_{i=1}^{n} \mathcal{E}_i$ or $\mathcal{E} = \mathcal{E}_1 \times \ldots \times \mathcal{E}_n$. Obviously $\prod_i \mathcal{E}_i$ is dominated provided each $\mathcal{E}_i$ is dominated.

The next theorem on product pseudo experiments generalizes corollary 4 in [15].

Theorem B.3.12

Let $\mathcal{E}_j = ((\mathcal{X}_j, \mathcal{A}_j), \mu_{\theta j}; \theta \in \Theta)$, $\mathcal{F}_j = ((\mathcal{Y}_j, \mathcal{B}_j), \nu_{\theta j}; \theta \in \Theta)$; $j = 1,2,\ldots,n$ be $2n$ pseudo experiments and let $\epsilon_1, \epsilon_2, \ldots, \epsilon_n$ be n non negative functions on Θ. Suppose $\mathcal{E}_j$ is ϵ_j deficient w.r.t. $\mathcal{F}_j$ (for k decision problems); $j = 1,\ldots,n$ and that $\mathcal{E}_1,\ldots,\mathcal{E}_n$ are dominated.

Then $\prod_j \mathcal{E}_j$ is

$$\theta \rightsquigarrow \sum_j \epsilon_j(\theta) \prod_{i<j} \|\mu_{\theta,i}\| \prod_{i>j} \|\nu_{\theta i}\|$$

deficient w.r.t. $\prod_j \mathcal{F}_j$ (for k decision problems).

Proof:

We may - by theorem B.3.1 - assume that Θ is finite. Put $\mu_i = \sum_\theta |\mu_{\theta i}|$, $\nu_i = \sum_\theta |\nu_{\theta i}|$, $\mu = \prod_i \mu_i$, $\nu = \prod_i \nu_i$, $f_{\theta i} = d\mu_{\theta i}/d\mu_i$ and $g_{\theta,i} = d\nu_{\theta i}/d\nu_i$. Define for each $i = 1,2,\ldots,n$ a pseudo experiment $\mathcal{G}_j$ by

$$\mathcal{G}_j = \prod_{i \leq j} \mathcal{E}_i \times \prod_{i>j} \mathcal{F}_i$$

Then $\mathcal{G}_o = \prod_i \mathcal{F}_i$, $\mathcal{G}_n = \prod_i \mathcal{E}_i$ and for any $\psi \in \Psi$:

$$\psi(\prod_i \mathcal{E}_i) - \psi(\prod_i \mathcal{F}_i) = \sum_{j=1}^{n} \psi(\mathcal{G}_j) - \psi(\mathcal{G}_{j-1}) .$$

Suppose each $\mathcal{E}_j$ is ϵ_j deficient w.r.t. $\mathcal{F}_j$ for k decision problems and that $\psi \in \Psi_k$. Then:

$$\psi(\mathcal{G}_j) = \int \psi(\underset{i\leqq j}{\otimes} f_{\theta i} \otimes \underset{i>j}{\otimes} g_{\theta i} \ ; \ \theta \in \Theta) d(\underset{i\leqq j}{\Pi} \mu_i \times \underset{i>j}{\Pi} \nu_i)$$

$$= \int [\int \psi(\underset{i<j}{\otimes} f_{\theta i} \otimes \underset{i>j}{\otimes} g_{\theta i} \otimes f_{\theta j}) d\mu_j] d(\underset{i<j}{\Pi} \mu_i \times \underset{i>j}{\Pi} \nu_i)$$

$$\geqq \int [\int \psi(\underset{i<j}{\otimes} f_{\theta i} \otimes \underset{i>j}{\otimes} g_{\theta i} \otimes g_{\theta j}) d\nu_j$$

$$- \sum_{\theta} \epsilon_j(\theta) \underset{i<j}{\Pi} |f_{\theta i}| \underset{i>j}{\Pi} |g_{\theta i}| \max\{\psi(e_\theta), \psi(-e_\theta)\}] d(\underset{i<j}{\Pi} \mu_i \times \underset{i>j}{\Pi} \nu_i)$$

$$= \psi(\mathcal{G}_{j-1}) - \sum_{\theta} \epsilon_j(\theta) \underset{i<j}{\Pi} \|\mu_{\theta i}\| \underset{i>j}{\Pi} \|\nu_{\theta i}\|$$

Hence

$$\psi(\underset{i}{\Pi} \mathcal{E}_i) - \psi(\underset{i}{\Pi} \mathcal{F}_i) \geqq - \sum_{\theta} \sum_{j} \epsilon_j(\theta) \underset{i<j}{\Pi} \|\mu_{\theta i}\| \underset{i>j}{\Pi} \|\nu_{\theta i}\| \qquad \square$$

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, 219-149.

[16] Torgersen, E.N. (To appear in the Annals of Math. Statist, approximately October (1972)). Comparison of translation experiments.

LC1 Torgersen, E.N. (1972). Local comparison of experiments when the parameter set is one dimensional. Statistical Research Report no. 4, 1972 from Institute of Mathematics, University of Oslo.