ISBN 82-553-0569-6 No 11
Desember 1984

PSEUDO EXPERIMENTS

AND

MAJORIZATION

by

Geir Dahl

INTRODUCTION

There are three main goals of this paper:

1. To give some descriptions of the concepts ε-deficiency and "more informative" between a certain class of pseudo dichotomies.
2. To show how some characterizations of the concept majorization can be viewed as consequences of the theory of pseudo dichotomies.
3. To show that majorization can be considered at a statistical concept and thereby give new interpretations of majorization.

Chapter I contains the statistical background and also the general theory of comparison of pseudo dichotomies. In section I.4 an important special case is presented, which is the first step in the direction of the goals 2. and 3. above.

Chapter II contains the definition of majorization and gives the most important characterizations of this concept. It is also shown how some of these characterizations are consequences of the theory in chapter I.

Chapter III treats a generalization of majorization, the so-called ε-majorization. This concept can be considered as a "nearly-majorization", and in fact many of the results show how "old results" from chapter II by simple corrections still are valid. A numerical example is also presented in order to show some geometrical ideas of ε-majorization.

Chapter IV defines a certain measure of distance between vectors by using the "sharpest" ε-majorization. An application to the construction of inequalities for convex functions is given.

Chapter V treats multi-dimensional majorization and demonstrates how the general theory of pseudo experiments gives descriptions of this concept.

I would like to thank professor Erik N. Torgersen for his interesting lectures on the topics that make the foundation for this work, and also for his help and encouragement during the work with this paper.

CONTENTS

CHAPTER I. STATISTICAL EXPERIMENTS AND PSEUDO EXPERIMENTS

I.1. Concepts and definitions.

A statistical experiment is defined as a "tuple" $\mathcal{E} = (\mathcal{X},\mathcal{A},P_\theta:\theta\in\Theta)$, where $(\mathcal{X},\mathcal{A})$ is a measurable space and $(P_\theta:\theta\in\Theta)$ is an ordered family of probability measures on $(\mathcal{X},\mathcal{A})$. We imagine that we can observe a stochastic variable with distribution P_θ, where $\theta\in\Theta$ is unknown, and with valued in the observation space $\mathcal{X}$. When Θ is a twopoint-set, $\mathcal{E}$ is called a dichotomy.

A pseudo experiment is a generalization of an experiment, by permitting arbitrary mass distributions. A pseudo experiment is therefore a "tuple" $\mathcal{E} = (\mathcal{X},\mathcal{A},\mu_\theta:\theta\in\Theta)$, where $(\mathcal{X},\mathcal{A})$ is a measurable space and $(\mu_\theta:\theta\in\Theta)$ is an ordered family of finite measures on $(\mathcal{X},\mathcal{A})$. When Θ is a twopoint-set, $\mathcal{E}$ is called a pseudo dichotomy.

A finite experiment is an experiment $(\mathcal{X},\mathcal{A},P_\theta:\theta\in\Theta)$, where both Θ and $\mathcal{X}$ are finite sets. If $\Theta = \{1,\ldots,s\}$ and $\mathcal{X} = \{1,\ldots,n\}$ are respectively parameter space and observation space in a finite experiment $\mathcal{E}$, we define the $s\times n$ matrix $P_\mathcal{E}$ by

$$(P_\mathcal{E})_{\theta j} = P_\theta(\{j\}),\ \theta = 1,\ldots,s,\ j = 1,\ldots,n.$$

We denote $P_\mathcal{E}$ the experimentmatrix of $\mathcal{E}$, and it will be a Markow-matrix (a stochastic matrix); the elements of $P_\mathcal{E}$ are non-negative and the rowsums are equal to 1.

Analogously we can define a finite pseudo experiment and the pseudo experimentmatrix.

A decision problem is a tuple $D = (\Theta, T, L)$, where Θ and T are arbitrary sets and L is an arbitrary real-valued function defined on $\Theta \times T$. Θ is once again called the parameter space, in which we know that an otherwise unknown parameter lies. T is called the decision space, and it consists of the possible decisions that can be made. $L_\theta(t)$ expresses the loss one suffers by making the decision $t \in T$ when $\theta \in \Theta$ is the underlying parameter.

Before making a decision a statistician will usually be able to get information by performing an experiment. This means that he can choose a model where he can observe a stochastic variable X with a probability distribution P_θ which depends on θ. It is therefore of interest to compare statistical experiments in order to find out how suited they are as sources of information in decision problems. It is also useful to compare pseudo experiments, for example within local comparison of experiments. Before we give the definition of ε-deficiency, which will be the starting point for comparison of pseudo experiments, we will remind of a few measure-theoretical definitions.

Let $(\mathcal{X}, \mathcal{A})$ be a measurable space and μ a measure on $(\mathcal{X}, \mathcal{A})$. Then the norm $\|\cdot\|$ is defined by

$$\|\mu\| = \sup\{\int f d\mu \mid f: \mathcal{X} \to [-1,1] \text{ is measurable}\}.$$

Here $[-1,1]$ is considered as a measurable space with the Borel-sets as the measurable sets.

Let now $(T, \mathcal{S})$ be an arbitrary measurable space. A randomization ρ from $(\mathcal{X}, \mathcal{A})$ to $(T, \mathcal{S})$ is a function

$$\rho(\cdot \mid \cdot): \mathcal{S} \times \mathcal{X} \to [0,1]$$

where $\rho(S \mid \cdot): \mathcal{X} \to [0,1]$ is $\mathcal{A}$-measurable for every $S \in \mathcal{S}$, and $\rho(\cdot \mid x)$

is a probability measure on $(T,\mathcal{S})$ for each $x\in\mathcal{X}$. A randomization is also called a Markow-kernel.

If ρ is a randomization from $(\mathcal{X},\mathcal{A})$ to $(T,\mathcal{S})$ and μ is a finite measure on $(\mathcal{X},\mathcal{A})$, a finite measure $\mu\rho$ on $(T,\mathcal{S})$ is induced by defining

$$(\mu\rho)(S) = \int\rho(S|x)\mu(dx); \quad S\in\mathcal{S}.$$

If $(Z,\mathcal{C})$ is an arbitrary measurable space, ν is a finite measure on $(Z,\mathcal{C})$ and f a real-valued, $\mathcal{C}$-measurable function on Z, we often use the notation νf for the integral $\int f(z)\nu(dz)$. In this notation a generalized version of Fubini's theorem on interated integration in the foregoing situation will be:

$$(\mu\rho)L = \mu(\rho L),$$

where L is a real-valued, $\mathcal{S}$-measurable function on T. Therefore we can, without danger of confusion, use the notation $\mu\rho L$ for this integral.

We now have the formal background for defining ε-deficiency.

DEFINITION I.1.1.

Let $\mathcal{E} = (\mathcal{X},\mathcal{A},\mu_\theta:\theta\in\Theta)$ and $\mathcal{F} = (\mathcal{Y},\mathcal{B},\nu_\theta:\theta\in\Theta)$ be two pseudo experiments with the same parameter space Θ, and let $\varepsilon_\theta:\theta\in\Theta$ be a function from Θ to $[0,\infty]$. We then say that $\mathcal{E}$ is ε-deficient with respect to $\mathcal{F}$ (for k-decision problems) if there to every measurable space $(T,\mathcal{S})$ where $\#\mathcal{S}<\infty$ $(\#\mathcal{S} = 2^k)$ and to every family L_θ, $\theta\in\Theta$ of measurable functions on T, and to every randomization σ from $(\mathcal{Y},\mathcal{B})$ to $(T,\mathcal{S})$ is a randomization ρ from $(\mathcal{X},\mathcal{A})$ to $(T,\mathcal{S})$ such that

(I.1.1) $$\mu_\theta\rho L_\theta < \nu_\theta\sigma L_\theta + \varepsilon_\theta\|L_\theta\|, \ \forall\theta\in\Theta$$

When we hereafter discuss two pseudo experiments in the same connection, it will be implied that they have the same parameter space.

It is important to realize that (I.1.1), when $\mathcal{E}$ and $\mathcal{F}$ are experiments, gives us an inequality between risk functions. When $\mathcal{E} = (\mathcal{X},\mathcal{A},\mu_\theta:\theta\in\Theta)$ is an experiment and ρ is a randomization from $(\mathcal{X},\mathcal{A})$ to $(T,\mathcal{S})$, then ρ is also called a decision rule. If we consider L as a loss function, we see that

$$\mu_\theta\rho L_\theta = \int[\int L_\theta(t)\rho(dt|x)]\mu_\theta(dx)$$

is the risk (expected loss) by using the decision rule ρ in $\mathcal{E}$ when θ is the underlying parameter. The inequality (I.1.1) then tells us how much additional risk we may have to face by choosing $\mathcal{E}$ in stead of $\mathcal{F}$.

Let $\mathcal{E} = (\mathcal{X},\mathcal{A},\mu_\theta:\theta\in\Theta)$ be a pseudo experiment, where $\Theta = \{1,\dots,s\}$. Let further $D = (\Theta,T,L)$ be a decision problem. Then we define the risk set $R_{\mathcal{E}}^D$ in $\mathcal{E}$ relative to D by

$$R_{\mathcal{E}}^D = \{(r_{\mathcal{E}}^D(1,\delta),\dots,r_{\mathcal{E}}^D(s,\delta))\,|\,\delta \text{ is a randomization from } (\mathcal{X},\mathcal{A}) \text{ to } (T,\mathcal{S})\},$$

where $r_{\mathcal{E}}^D(\theta,\delta) = \mu_\theta\delta L_\theta$.

If $\mathcal{E}$ is 0-deficient with respect to $\mathcal{F}$ (for k-decision problems), we write $\mathcal{E} > \mathcal{F}$, or alternatively $\mathcal{F} < \mathcal{E}$ ($\mathcal{E} \underset{k}{>} \mathcal{F}$, or alternatively $\mathcal{F} \underset{k}{<} \mathcal{E}$), and in this case we say that $\mathcal{E}$ is more informative than $\mathcal{F}$ (for k-decision problems). When $\mathcal{E} > \mathcal{F}$ and $\mathcal{F} > \mathcal{E}$ ($\mathcal{E} \underset{k}{>} \mathcal{F}$ and $\mathcal{F} \underset{k}{>} \mathcal{E}$), we say that $\mathcal{E}$ and $\mathcal{F}$ are equivalent (for k-decision problems), and in this case we write $\mathcal{E} \sim \mathcal{F}$ ($\mathcal{E} \underset{k}{\sim} \mathcal{F}$).

In order to measure the maximum loss one can suffer by choosing one pseudo experiment in stead of another, we can use the following concepts:

$$\delta_{(k)}(\mathcal{E},\mathcal{F}) = \inf\{\varepsilon\in[0,\infty] \mid \mathcal{E} \text{ is } \varepsilon\text{-deficient with respect to } \mathcal{F} \text{ (for k-decision problems)}\}$$

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

When $\mathcal{E}$ and $\mathcal{F}$ are two pseudo dichotomies with the same parameter space, we define

$$\dot{\delta}_{(k)}(\mathcal{E},\mathcal{F}) = \tfrac{1}{2}\inf\{\varepsilon\in[0,\infty] \mid \mathcal{E} \text{ is } (0,\varepsilon)\text{-deficient with respect to } \mathcal{F} \text{ (for k-decision problems)}\}$$

$$\dot{\Delta}_{(k)}(\mathcal{E},\mathcal{F}) = \dot{\delta}_{(k)}(\mathcal{E},\mathcal{F})\vee\dot{\delta}_{(k)}(\mathcal{F},\mathcal{E}).$$

We denote $\delta(\mathcal{E},\mathcal{F})$ the deficiency between $\mathcal{E}$ and $\mathcal{F}$ and $\dot{\delta}(\mathcal{E},\mathcal{F})$ is denoted the dot-deficiency between $\mathcal{E}$ and $\mathcal{F}$.
If $\mathcal{E} = (\mathcal{X},\mathcal{A},\mu_\theta:\theta\in\Theta)$ and $\mathcal{F} = (\mathcal{Y},\mathcal{B},\nu_\theta:\theta\in\Theta)$ are two pseudo experiments, then $\mathcal{E}\times\mathcal{F}$ denotes the pseudo experiment

$$(\mathcal{X}\times\mathcal{Y},\mathcal{A}\times\mathcal{B},\mu_\theta\times\nu_\theta:\theta\in\Theta),$$

where $\mathcal{A}\times\mathcal{B}$ is the product sigma-algebra on $\mathcal{X}\times\mathcal{Y}$ and $\mu_\theta\times\nu_\theta$, $\theta\in\Theta$, are the product measures.

If $\mathcal{E}$ consists in observing a stochastic variable X and $\mathcal{F}$ consists in observing a stochastic variable Y which is independent of X, then $\mathcal{E}\times\mathcal{F}$ will be observing the pair (X,Y).

This definition of $\mathcal{E}\times\mathcal{F}$ can easily be extended to products of a finite number of pseudo experiments $\mathcal{E}_1,\ldots,\mathcal{E}_N$. If $\mathcal{E}_1 = \ldots = \mathcal{E}_N = \mathcal{E}$, we write $\mathcal{E}^N$ for this product pseudo experiment $\mathcal{E}\times\ldots\times\mathcal{E}$.

We will close this section by presenting the notation that will be used in this work.

When $a,b \in R$ (where R denotes the set of real numbers), $a \vee b$ and $a \wedge b$ denote respectively $\max\{a,b\}$ and $\min\{a,b\}$, and we also write a^+ for $a \vee 0$.

If $f: \mathcal{X} \to R$ is a real-valued function defined on a set $\mathcal{X}$, we define

$$\|f\| = \sup_{x \in \mathcal{X}} |f(x)|$$

If $x = (x_1,\ldots,x_n)$ and $y = (y_1,\ldots,y_n)$ are vectors in R^n, we use the notation $\langle x,y \rangle$ for the usual Euclidian scalar-product of x and y:

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

$x_{[j]}$ and $x_{(j)}$ are respectively the j-th greatest and the j-th smallest component of x. We let $\mathcal{D}^n$ be the set $\{(x_1,\ldots,x_n) \in R^n | x_1 \geq, \ldots \geq x_n\}$. When $y \in R^n$, K_y is defined as the convex hull of the set of all permutations of y. K_n denotes the set of all probability vectors in R^n ($K_n = \{(x_1,\ldots,x_n) \in R^n | x_i \geq 0;$ $i = 1,\ldots,n$ and $\sum_{i=1}^{n} x_i = 1\}$). We also define $d_0(x,y) = \sum_{i=1}^{n} |x_i - y_i|$, which is a metric on R^n, while $\|\cdot\|$ defined by $\|x\|_0 = \sum_{i=1}^{n} |x_i|$ is the induced norm. We let e denote the vector $(1,\ldots,1) \in R^n$ and H_α the set $\{(x_1,\ldots,x_n) \in R^n | \sum_{i=1}^{n} x_i = \alpha\}$, where $\alpha \in R$. The dimension n will here always be understood from the context.

When ν is a measure on a measurable space, $|\nu|$ denotes the total variation measure of ν.

If $A \subset R^n$ is a set, $\langle A \rangle$ denotes the convex hull of A. We also use the abbreviation $\langle a,b \rangle$ for $\langle \{a,b\} \rangle$, when $a,b \in R^2$, which is the line segment in R^2 between a and b. (This abbreviation will be used in example III.2.14 only, and it can therefore not be mixed up with the notation for the scalar product.)

$\mathcal{M}_{n,m}$ will denote the set of all stochastic matrices (): Markow-matrices) of dimension $n \times m$, and $\mathcal{M}^D_{n,m}$ is the set of all doubly-stochastic matrices of dimension $n \times m$.

When $a,b \in R^n$, $\mathcal{E}_{a,b}$ will denote the finite pseudo experiment which has a pseudo experimentmatrix

$$P_{\mathcal{E}_{a,b}} = \begin{pmatrix} a_1 \cdots\cdots a_n \\ b_1 \cdots\cdots b_n \end{pmatrix}$$

More accurately we define

$$\mathcal{E}_{a,b} = (\{1,\ldots,n\},\ \mathcal{P}(\{1,\ldots,n\}),\ \mu_1,\mu_2),$$

where $\mu_1(\{j\}) = a_j$ and $\mu_2(\{j\}) = b_j$; $j = 1,\ldots,n$.

I.2. SOME MAIN RESULTS ON COMPARISON OF PSEUDO EXPERIMENTS

It is an immidiate consequence of Definition I.1.1 that $\mathcal{E}$ is ε-deficient with respect to $\mathcal{F}$ for k-decision problems whenever $\mathcal{E}$ is ε-deficient with respect to $\mathcal{F}$ for $(k+1)$-decision problems. Furthermore $\mathcal{E}$ is ε-deficient with respect to $\mathcal{F}$ if and only if $\mathcal{E}$ is ε-deficient with respect to $\mathcal{F}$ for k-decision problems for $k = 1,2,\ldots$. When $\mathcal{E}$ and $\mathcal{F}$ are experiments $\Delta,(\mathcal{E},\mathcal{F}) = 0$ will hold, while this isn't necessarily true for pseudo experiments.

Since this work mainly will treat pseudo dichotomies and dichotomies, it is important to note the relations between ε-deficiency and ε-deficiency for k-decision problems in these cases.

PROPOSITION I.2.1.

Let $\mathcal{E} = (\mathcal{X},\mathcal{A},\mu_1,\mu_2)$ and $\mathcal{F} = (\mathcal{Y},\mathcal{B},\nu_1,\nu_2)$ be two pseudo dichtomies where $\mu_1 > 0$, $\nu_1 > 0$ and $\Delta,(\mathcal{E},\mathcal{F}) = 0$.

Then $\mathcal{E}$ is ε-deficient with reespect to $\mathcal{F}$ if and only if $\mathcal{E}$ is ε-deficient with respect to $\mathcal{F}$ for 2-decision problems (testing problems).

PROOF: See Theorem B.2.4. in Reference [4]. ⊔

Since the conditions in this proposition are easily seen to be satisfied in the case of dichotomies, we know that there is an equivalence between ε-deficiency and ε-deficiency for testing problems in this situation.

Let now $\mathcal{E} = (\mathcal{X},\mathcal{A},\mu_\theta:\theta\in\Theta)$ be a pseudo experiment where Θ is finite, say $\# \Theta = s$. We then let $\Psi_k^{(s)}$ denote the set of all maximum of k linear functionals on R^s, while $\Psi^{(s)}$ denotes the set of all sublinear functionals on R^s. We then define, for $\psi\in\Psi^{(s)}$,

$$\psi(\mathcal{E}) = \int\psi(d\mu_\theta|d\sum_\theta|\mu_\theta|:\theta\in\Theta)d\sum_\theta|\mu_\theta|$$

where $d\mu_\theta|d\Sigma_\theta|\mu_\theta|$ is "the" Radon-Nikodym derivative of μ_θ with respect to $\Sigma_\theta|\mu_\theta|$. If τ is a non-negative measure on $(\mathcal{X},\mathcal{A})$ which dominates $\mu_\theta:\theta\in\Theta$, then the following equation will hold

$$\psi(\mathcal{E}) = \int\psi(f_\theta:\theta\in\Theta)d\tau$$

where $f_\theta = d\mu_\theta|d\tau$.

We furthermore define $T_k = \{1,\ldots,k\}$ and $\mathcal{S}_k = \mathcal{P}(T_k)$, and we can then formulate the main result on comparison of pseudo experiments.

THEOREM I.2.2.

Let $\mathcal{E} = (\mathcal{X},\mathcal{A},\mu_\theta:\theta\in\Theta)$ and $\mathcal{F} = (\mathcal{Y},\mathcal{B},\nu_\theta:\theta\in\Theta)$ be two pseudo experiments with the same parameter set Θ, where $\# \Theta = s$. Then the statements (i)-(iv) below will be equivalent.

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

(ii) For every randomization σ from $(\mathcal{Y},\mathcal{B})$ to $(T_k,\mathcal{S}_k)$, and for every family L_θ , $\theta\in\Theta$ of real-valued functions on T_k, there is a randomization ρ from $(\mathcal{X},\mathcal{A})$ to $(T_k,\mathcal{S}_k)$ such that

$$\sum_\theta \mu_\theta\rho L_\theta \leq \sum_\theta \nu_\theta\sigma L_\theta + \sum_\theta \varepsilon_\theta \| L_\theta \| .$$

(iii) For every randomization σ from $(\mathcal{Y},\mathcal{B})$ to $(T_k,\mathcal{S}_k)$, there is a randomization ρ from $(\mathcal{X},\mathcal{A})$ to $(T_k,\mathcal{S}_k)$ such that

$$\| \mu_\theta\rho-\nu_\theta\sigma \| \leq \varepsilon_\theta , \qquad \forall\theta\in\Theta$$

(iv) $$\psi(\mathcal{E}) \geq \psi(\mathcal{F})-\sum_\theta \varepsilon_\theta(\psi(-e_\theta)\vee\psi(e_\theta)), \qquad \forall\psi\in\Psi_k^{(s)}$$

PROOF: See Theorem B.2.1. in reference [4]. □

Some comments will now be given in connection with this main result.

First of all we see that (ii) above is very closely connected to a statement around minimum Bayes risk in the case of experiments. In fact we have the following result:

PROPOSITION I.2.3.

Let $\mathcal{E}$ and $\mathcal{F}$ be defined as in Theorem I.2.2. Then (v) below will be equivalent with (i)-(iv) in Theorem I.2.2:

(v) For every a priori distribution λ on Θ and every family L_θ of real-valued functions on T_k , the following inequality will hold

$$B(\lambda|\mathcal{E}) \leq B(\lambda|\mathcal{F}) + \sum_\theta \varepsilon_\theta \lambda_\theta \| L_\theta \| \tag{I.2.1}$$

where $B(\lambda|\mathcal{E}) = \inf\{\Sigma_\theta \lambda_\theta \mu_\theta \rho L_\theta | \rho$ is a randomization from $(\mathcal{X},\mathcal{A})$ to $(T_k,\mathcal{J}_k)\}$

PROOF: Assume that (ii) of Theorem I.2.2. holds, and let λ be an a priori distribution (; a probability distribution) on Θ. We let all the subsets of Θ be measurable. Let furthermore L_θ, $\theta \in \Theta$ be a family of real-valued functions on T_k. According to (ii) we now know that there to every randomization σ from $(\mathcal{Y},\mathcal{B})$ to $(T_k,\mathcal{J}_k)$ corresponds a randomization ρ such that

$$\sum_\theta \mu_\theta \rho(\lambda_\theta L_\theta) \leq \sum_\theta \nu_\sigma \sigma(\lambda_\theta L_\sigma) + \sum_\sigma \varepsilon_\sigma \| \lambda_\theta L_\theta \|$$

This is (ii) applied to the loss-function $(\theta,t) \to \lambda_\theta L_\theta(t)$. Consequently

$$\sum_\theta \lambda_\theta \mu_\theta \rho L_\theta \leq \sum_\sigma \lambda_\theta \nu_\theta \sigma L_\theta + \sum_\theta \lambda_\theta \varepsilon_\theta \| L_\theta \|$$

since λ_θ , $\theta \in \Theta$ are non-negative constants.

From the definition of $B(\lambda|\mathcal{E})$ we see that

$$B(\lambda|\mathcal{E}) \leq \sum_\theta \lambda_\theta \mu_\theta \rho L_\theta$$

which means that

$$B(\lambda|\mathcal{E}) \leq \sum_\theta \lambda_\theta \nu_\theta \sigma L_\theta + \sum_\theta \lambda_\theta \varepsilon_\theta \| L_\theta \| .$$

By taking infimum over all randomizations σ from $(\mathcal{Y},\mathcal{B})$ to $(T_k,\mathcal{J}_k)$ we get

$$B(\lambda|\mathcal{E}) \le B(\lambda|\mathcal{F}) + \sum_\theta \lambda_\theta \varepsilon_\theta \|L_\sigma\|$$

and the implication from (ii) to (v) has been shown.

Assume now that (v) holds, and let L_θ , $\theta \in \Theta$ be a family of real-valued functions on T_k , and let λ be the uniform probability distribution on Θ, $\lambda_\theta = \frac{1}{s}$, $\theta = 1,\ldots,s$. From (v) it follows that

$$B(\lambda|\mathcal{E}) \le B(\lambda|\mathcal{F}) + \frac{1}{s} \sum_\theta \varepsilon_\theta \|L_\theta\|$$

which means that

$$\inf_\rho \sum_\theta \mu_\theta \rho L_\theta \le \inf_\sigma \sum_\theta \nu_\theta \sigma L_\theta + \sum_\sigma \varepsilon_\theta \|L_\theta\|$$

For every loss-function L and every randomization σ from $(\mathcal{Y},\mathcal{B})$ to $(T_k,\mathcal{S}_k)$ we then know that

$$\inf_\rho \sum_\theta \mu_\theta \rho L_\theta \le \sum_\theta \nu_\theta \sigma L_\theta + \sum_\sigma \varepsilon_\theta \|L_\theta\|$$

But this infimum will be attained for a suitable randomization ρ from $(\mathcal{X},\mathcal{A})$ to $(T_k,\mathcal{S}_k)$. This can be seen analogously to Lemma 5.10 in reference [5], by using weak compactness and Tychnoff's theorem on product topologies. Consequently

$$\sum_\theta \mu_\theta \rho L_\theta \le \sum_\theta \nu_\theta \sigma L_\theta + \sum_\sigma \varepsilon_\theta \|L_\theta\|$$

and the proof is then completed. □

The characterization (iii) in Theorem I.2.2. treats operating characteristics, which will be defined now. When $\mathcal{E} = (\mathcal{X},\mathcal{A},\mu_\theta : \theta\in\Theta)$ is a pseudo experiment, $(T,\mathcal{S})$ a decision space ($(T,\mathcal{S})$ a measurable space) and ρ a decision rule in $\mathcal{E}$, we denote $\mu_\theta\rho$ the operating characteristics in $\mathcal{E}$, where

$$(\mu_\theta\rho)(S) = \int \rho(S|x)\mu_\theta(dx) \; ; \quad S\in\mathcal{S}$$

When $\mathcal{E}$ is an experiment $(\mu_\theta \rho)(S)$ expresses the probability of making a decision in S when θ is the underlying parameter.

Therefore (iii) of Theorem I.2.2. says how the operating characteristics in $\mathcal{F}$ can be approximated by the operating characteristics in $\mathcal{E}$.

In connection with inequalities in Chapter IV, a special case of the next proposition will be needed, but this proposition is also useful in other situations.

PROPOSITION I.2.4.

Let $\mathcal{E}$ and $\mathcal{F}$ be two pseudo experiments with the same parameter space Θ, where Θ is finite. Let furthermore $\{\varepsilon^{(n)}\}_{n=1}^{\infty}$ be a sequence of non-negative, real-valued functions on Θ such that

$$\varepsilon_\theta^{(n)} \to \varepsilon_\theta , \quad \forall\theta \in \Theta$$

Assume that $\mathcal{E}$ is $\varepsilon^{(n)}$-deficient with respect to $\mathcal{F}$ (for k-decision problems) for $n = 1,2,\ldots$.

Then $\mathcal{E}$ will be ε-deficient with respect to $\mathcal{F}$ (for k-decision problems).

The famous Markow-kernel theorem for ε-deficiency, which for instance can be found in Corollary B.3.5. in reference [4], will also be formulted here, since it will be of great use later on.

PROPOSITION I.2.5.

Let $\mathcal{E} = (\mathcal{X},\mathcal{A},\mu_\theta : \theta\in\Theta)$ and $\mathcal{F} = (\mathcal{Y},\mathcal{B},\nu_\theta : \theta\in\Theta)$ be two pseudo experiments, where Θ is finite and $\mathcal{Y}$ is a Borel-set in a complete, separable metric space and where $\mathcal{B}$ consists of Borel-subsets. Let ε be a non-negative function on Θ. Then the following equivalence will hold:

$$\mathcal{E} \text{ is } \varepsilon\text{-deficient with reespect to } \mathcal{F}$$

$$\Updownarrow$$

there is a Markow-kernel M from $(\mathcal{X},\mathcal{A})$ to $(\mathcal{Y},\mathcal{B})$ such that

$$\|\mu_\theta M-\nu_\sigma\| \leq \varepsilon_\theta \; ; \qquad \forall\theta\in\Theta$$

If we represent Markow-kernels by Markow-matrices in the case of finite pseudo experiments, we get the following corollary of Proposition I.2.5.

COROLLARY I.2.6.

Let $\mathcal{E} = (\mathcal{X},\mathcal{A},\mu_\theta:\theta\in\Theta)$ and $\mathcal{F} = (\mathcal{Y},\mathcal{B},\mu_\theta:\theta\in\Theta)$ be two pseudo experiments, where

$$\mathcal{X} = \{1,\ldots,r\}$$
$$\mathcal{Y} = \{1,\ldots,k\}$$
$$\Theta = \{1,\ldots,s\}$$

and $\mathcal{A} = \mathcal{P}(\mathcal{X})$, $\mathcal{B} = \mathcal{P}(\mathcal{Y})$.

Then the following will hold:

$$\mathcal{E} > \mathcal{F}$$

$$\Updownarrow$$

$$\exists M\in\mathcal{M}_{r,k}: P_{\mathcal{E}}M = P_{\mathcal{F}}$$

We will also present a generalization of Neyman-Pearson's lemma for later use.

PROPOSITION I.2.7.

Let $(\mathcal{X},\mathcal{A},\mu)$ be a measure-space, and let f_1 and f_2 be measurable, μ-integrable functions defined on $\mathcal{X}$.

Assume that there to a constant α is a randomization δ

satisfying

(I.2.2) $$\int \delta f_1 d\mu = \alpha$$

Let $\mathcal{C}$ be the class of all ramdomizations for which (I.2.2.) holds. Then it follows that

i) Among all elements in $\mathcal{C}$ there is one that maximises $\int \delta f_2 d\mu$.

ii) A necessary and sufficient condition for an element δ in $\mathcal{C}$ to maximize $\int \delta f_2 d\mu$ is the existence of a constant c such that

$$\delta(x) = \begin{cases} 1 & \text{when } f_2(x) < c\, f_1(x) \\ 0 & \text{when } f_2(x) < c\, f_1(x) \end{cases}$$

PROOF: See reference [2] page 83. □

I.3. PSEUDO DICHOTOMIES

In this section we'll give a few characterizations of ε-deficiency for pseudo dichotomies. These results are generalizations of the theory on pseudo derivatives, which forms the basis of local comparison of experiments.

Let henceforth (in I.3.) $\mathcal{E} = (\mathcal{X}, \mathcal{A}, \mu_1, \mu_2)$ and $\mathcal{F} = (\mathcal{Y}, \mathcal{B}, \nu_1, \nu_2)$ be pseudo dichotomies which have the following two properties:

(I.3.1) μ_1 and ν_1 are probability measurees

(I.3.2.) $\Delta_1(\mathcal{E}, \mathcal{F}) = 0$

Note that (I.3.2.) is equivalent to $\mu_2(\mathcal{X}) = \nu_2(\mathcal{Y})$ because of (I.3.1.). This means that μ_2 and ν_2 are arbitrary finite measures with the same total mass.

PROPOSITION I.3.1.

$$\mathcal{E} \text{ is } (\varepsilon_1,\varepsilon_2)\text{-deficient with respect to } \mathcal{F}$$

$$\Updownarrow$$

$$\text{(I.3.3.)} \qquad \|a_1\mu_1+a_2\mu_2\| > \|a_1\nu_1+a_1\nu_2\|-\varepsilon_1|a_1|-\varepsilon_2|a_2|, \quad \forall a_1,a_2\in R$$

PROOF: From the assumptions (I.3.1) and (I.3.2) it follows by using Proposition I.2.1. that $\mathcal{E}$ is $(\varepsilon_1,\varepsilon_2)$-deficient with respect to $\mathcal{F}$ if and only if $\mathcal{E}$ is $(\varepsilon_1,\varepsilon_2)$-deficient with respect to $\mathcal{F}$ for testing problems. The proposition is then a consequence of Corollary B.2.3. in Reference [4]. ⊔

PROPOSITION I.3.2.

$$\mathcal{E} \text{ is } (\varepsilon_1,\varepsilon_2)\text{-deficient with respect to } \mathcal{F}$$

$$\Updownarrow$$

$$\text{(I.3.4.)} \qquad \|\xi\mu_1-\mu_2\| > \|\xi\nu_1-\nu_2\|-\varepsilon|\xi|-\varepsilon_2 , \qquad \forall \xi\in R$$

PROOF: This result follows quite easily from Proposition I.3.1. It is enough to show that (I.3.3.) and (I.3.4.) are equivalent.

It is trivial that (I.3.3.) implies (I.3.4.) (simply choose $a_1 = \xi$ and $a_2 = -1$).

Assume therefore now that (I.3.4.) holds and let $a_1,a_2\in R$. If $a_2 = 0$, then (I.3.3.) will hold because $\|\mu_1\| = \|\nu_1\| = 1$. If $a_2 \neq 0$, we choose $\xi = -\frac{a_1}{a_2}$ and from (I.3.4.) we then have

$$\|-\frac{a_1}{a_2}\mu_1-\mu_2\| > \|-\frac{a_1}{a_2}\nu_1-\nu_2\|-\varepsilon_1|-\frac{a_1}{a_2}|-\varepsilon_2$$

so multiplication by $|a_2|$ gives us (I.3.3.). ⊔

The concepts introduced in the next definition will be important both in this and subsequent chapters.

DEFINITION I.3.3.

Let $\mathcal{E} = (\mathcal{X}, \mathcal{A}, \mu_1, \mu_2)$. We then define

$$U_{\mathcal{E}}(\xi) = \|\xi\mu_1 - \mu_2\| \; ; \qquad \xi \in R$$

and denote $U_{\mathcal{E}}$ the U-function of $\mathcal{E}$.

Let furthermore

$$V_{\mathcal{E}} = \{(\int\delta d\mu_1, \int\delta d\mu_2) \mid \delta: \mathcal{X} \to [0,1] \text{ is } \mathcal{A}\text{-measurable}\}$$

and denote $V_{\mathcal{E}}$ the V-set of $\mathcal{E}$.

Finally we define

$$\beta_{\mathcal{E}}(\alpha) = \sup\{y \mid (\alpha, y) \in V_{\mathcal{E}}\} \; ; \qquad \alpha \in [0,1]$$

and denote $\beta_{\mathcal{E}}$ the β-function of $\mathcal{E}$.

One of the reeasons for us to introduce the U-function, the V-set and the β-finction of a pseudo dichotomy, is that each of them characterize pseudo dichotomies up to an equivalence. This will be shown later.

Proposition I.3.2. can now be reformualted with the aid of the U-function.

COROLLARY I.3.4.

$$\mathcal{E} \text{ is } (\varepsilon_1, \varepsilon_2)\text{-deficient with respect to } \mathcal{F}$$

$$\Updownarrow$$

(I.3.5.) $$U_{\mathcal{E}}(\xi) \geq U_{\mathcal{F}}(\xi) - \varepsilon_1 |\xi| - \varepsilon_2 \,, \qquad \forall \xi \in R$$

PROOF: This is seen directly from Proposition I.3.2. and Definition I.3.3. □

COROLLARY I.3.5.

$$\mathcal{E} > \mathcal{F} \iff U_{\mathcal{E}} > U_{\mathcal{F}}$$
$$\mathcal{E} \sim \mathcal{F} \iff U_{\mathcal{E}} = U_{\mathcal{F}}$$

PROOF: This follows from Corollary I.3.4. by considering (0,0)-deficiency. □

Corollary I.3.5. shows that the U-function is well suited for describing "more informative" and "equivalence" between pseudo dichotomies which satisfy (I.3.1.) and (I.3.2.). We can now see that the U-function characterizes the pseudo dichotomy up to an equivalence. Furthermore dot-deficiencies between pseudo dichotomies can easily be expressed by the U-function, as the next proposition says.

PROPOSITION I.3.6.

$$\dot{\delta}(\mathcal{E},\mathcal{F}) = \tfrac{1}{2}\sup_{\xi}[U_{\mathcal{F}}(\xi)-U_{\mathcal{E}}(\xi)]^{+}$$

$$\dot{\Delta}(\mathcal{E},\mathcal{F}) = \tfrac{1}{2}\sup_{\xi}|U_{\mathcal{F}}(\xi)-U_{\mathcal{E}}(\xi)|$$

PROOF: By applying Corollary I.3.4. we get

$$\begin{aligned}\dot{\delta}(\mathcal{E},\mathcal{F}) &= \tfrac{1}{2}\inf_{\xi}\{\varepsilon>0 \mid \mathcal{E} \text{ is } (0,\varepsilon)\text{-deficient with respect to } \mathcal{F}\} \\ &= \tfrac{1}{2}\inf\{\varepsilon>0 \mid U_{\mathcal{E}}(\xi) \geq U_{\mathcal{F}}(\xi)-\varepsilon,\ \forall\xi\in R\} \\ &= \tfrac{1}{2}\inf\{\varepsilon>0 \mid U_{\mathcal{F}}(\xi)-U_{\mathcal{E}}(\xi) \leq \varepsilon,\ \forall\xi\in R\} \\ &= \tfrac{1}{2}\sup[U_{\mathcal{F}}(\xi)-U_{\mathcal{E}}(\xi)]^{+}\end{aligned}$$

The expression for $\dot{\Delta}(\mathcal{E},\mathcal{F})$ follows from this because $\dot{\Delta}(\mathcal{E},\mathcal{F}) = \dot{\delta}(\mathcal{E},\mathcal{F}) \vee \dot{\delta}(\mathcal{F},\mathcal{E})$. □

We shall now consider V-sets, and we start by showing that every V-set is compact and convex. Since this result is based on the finiteness of Θ only, and not necessarily that $\Theta = 2$ we present this result in its general version.

PROPOSITION I.3.7.

Let $\mathcal{E} = (\mathcal{X},\mathcal{A},\mu_\theta:\theta\in\Theta)$, where $\Theta = \{1,\dots,s\}$ be a pseudo experiment, and let

$$\mathcal{M} = \{\delta \mid \delta:\mathcal{X} \to [0,1] \text{ is } \mathcal{A}\text{-measurable}\}$$

Define now

$$V = \{(\int\delta d\mu_1,\dots,\int\delta d\mu_s) \mid \delta\in\mathcal{M}\}$$

Then V is a compact and convex subset of R^s.

PROOF: First we'll show that V is convex.

Let $v_1,v_2\in V$ and let $t\in[0,1]$. Then there are $\delta_1,\delta_2\in\mathcal{M}$ such that $v_i = (\int\delta_i d\mu,\dots,\int\delta_i d\mu_s)$, $i = 1,2$. Consequently

$$tv_1+(1-t)v_2 = (\int[t\delta_1+(1-t)\delta_2]d\mu_1,\dots,\int[t\delta_1+(1-t)\delta_2]d\mu_s)$$

and since $t\delta_1+(1-t)\delta_2\in\mathcal{M}$ (because $\delta_1(x),\delta_2(x),t\in[0,1]$, $\forall x\in X$), this implies that $tv_1+(1-t)v_2\in V$, which means that V is convex

We now show that V is compact.

It is sufficient to show that V is closed and bounded. Since μ_θ, $\theta\in\Theta$ is a finite measure, $M = \bigvee_{\theta\in\Theta} |\mu_\theta|(\mathfrak{X})$ will be a real number, because Θ is finite. We therefore see that

$$|\int\delta d\mu_\theta| \leq \int|\delta|d|\mu_\sigma| \leq \int d|\mu_\sigma| = |\mu_\sigma|(\mathfrak{X}) \leq M$$

so V is bounded in each component, and because V has a finite number of components (namely s), V itself will be bounded.

In order to show that V is closed, it is enough to show that every sequence in V has a convergent subsequence. Let $\{v_n\}_{n=1}^{\infty}$ be a sequence in V. Then there is a sequence $\{\delta_n\}_{n=1}^{\infty}$ in $\mathcal{M}$ such that

$$v_n = (\int\delta_n d\mu_1,\ldots,\int\delta_n d\mu_s).$$

Define now

$$\mu(A) = \frac{\sum_{\theta=1}^{s} |\mu_\theta|(A)}{\sum_{\theta=1}^{s} |\mu_\theta|(\mathfrak{X})}; \quad A\in\mathcal{A}.$$

It is then easy to see that μ is a probability measure on $(\mathfrak{X},\mathcal{A})$. Since $\{\delta_n\}_{n=1}^{\infty}$ is uniformly integrable. (Because the sequence is uniformly bounded), the weak compactness theorem tells us that there is a subsequence $\{\delta_{n'}\}$ of $\{\delta_n\}$ and a $\mathcal{A}$-measurable $\delta:\mathfrak{X}\to R$ such that $\delta_{n'}\to\delta$ weakly :

$$\int\delta_{n'} h d\mu \to \int\delta h d\mu$$

for every bounded, measurable $h:\mathfrak{X}\to R$.

We realize that $\delta\in\mathcal{M}$ because

$$\int_A\delta_{n'}\delta\mu \to \int_A\delta d\mu$$

and $\int_A d_{n'} d\mu\in[0,1]$ for every $A\in\mathcal{A}$, so $\int_A\delta d\mu\in[0,1]$ for every $A\in\mathcal{A}$, and consequently $0<\delta<1$ a.e. $[\mu]$. Then δ can be modified on subset of μ-measure 0 such that $0<\delta<1$ without changing the value of $\int\delta h d\mu$.

Finally, for $\theta\in\{1,\ldots,s\}$, we have

$$\int\delta_{n'} d\mu_\theta = \int\delta_{n'}\frac{d\mu_\theta}{d\mu}d\mu \to \int\delta\frac{d\mu_\theta}{d\mu}d\mu = \int\delta d\mu_\theta$$

since $h = \frac{d\mu_\theta}{d\mu}$ is bounded (while μ_θ is finite) and measurable.

This shows, because Θ is finite, that

$$(\int\delta_{n'}d\mu_1,\ldots,\int\delta_{n'}d\mu_s) \to (\int\delta d\mu_1,\ldots,\int\delta d\mu_s)\in V$$

so $\{v_n\}$ has a convergent subsequence in V. □

The next proposition gives us some important properties of the V-set of a pseudo dichotomy.

PROPOSITION I.3.8.

Let $\mathcal{E} = (\mathfrak{X},\mathcal{A},\mu_1,\mu_2)$ be a pseudo dichotomy. The V-set, $V_{\mathcal{E}}$, of $\mathcal{E}$ will then have the following properties:

i) $V_{\mathcal{E}}$ is compact and convex

ii) $(0,0)$, $(1,\mu_2(\mathfrak{X}))\in V_{\mathcal{E}}$

iii) $V_{\mathcal{E}}$ is symmetrical about the point $(\frac{1}{2},\frac{1}{2}\mu_2(\mathfrak{X}))$.

PROOF: i) Follows from Proposition I.3.7 with $s = 2$.

ii) Can be seen by choosing respectively $\delta\equiv 0$ and $\delta\equiv 1$.

iii) If $\delta:\mathfrak{X}\to[0,1]$ is $\mathcal{A}$-measurable, then $\delta' = 1-\delta$ will have the properties: $\delta':\mathfrak{X}\to[0,1]$ and δ' is $\mathcal{A}$-measurable.

Furthermore

$$(\int\delta' d\mu_1,\int\delta' d\mu_2) = (\int d\mu_1-\int d\mu_1,\int d\mu_2-\int\delta d\mu_2)$$

$$= (1-\int\delta d\mu_1,\mu_2(\mathfrak{X})-\int\delta d\mu_2) = (1,\mu_2(\mathfrak{X}))-(\int\delta d\mu_1,\int\delta d\mu_2)$$

so we see that $V_{\mathcal{E}}$ is symmetrical about the point $(\frac{1}{2},\frac{1}{2}\mu_2(\mathfrak{X}))$. □

Since $V_{\mathcal{E}}$ is compact and convex, it is possible to consider the support function $H_{\mathcal{E}}$ of $V_{\mathcal{E}}$, which is defined by

$$H_{\mathcal{E}}(a) = \sup_{v\in V_{\mathcal{E}}} \langle a,v\rangle$$

where $a \in R^2$ and $\langle \cdot , \cdot \rangle$ denotes the usual Euclidian scalar product on R^2.

Let now $H_{\varepsilon_1,\varepsilon_2}$ be the support function of the set $V_{\varepsilon_1,\varepsilon_2} = [-\frac{\varepsilon_1}{2},\frac{\varepsilon_1}{2}]\times[-\frac{\varepsilon_2}{2},\frac{\varepsilon_2}{2}]$. It is then easy to show that

$$H_{\varepsilon_1,\varepsilon_2}(a_1,a_2) = \tfrac{1}{2}(\varepsilon_1|a_1|+\varepsilon_2|a_2|).$$

With this we have come to another characterization of $(\varepsilon_1,\varepsilon_2)$-deficiency between pseudo dichotomies satisfying (I.3.1) and (I.3.2).

PROPOSITION I.3.9.

$\mathcal{E}$ is $(\varepsilon_1,\varepsilon_2)$-deficient with respect to $\mathcal{F}$

$$\Updownarrow$$

$$H_{\mathcal{E}} + H_{\varepsilon_1,\varepsilon_2} \geq H_{\mathcal{F}}.$$

PROOF: In order to show this equivalence we show a useful equality, which holds for any measure μ on $(\mathcal{X},\mathcal{A})$:

$$\|\mu\| = 2 \sup_{0\leq\delta\leq 1} \int \delta d\mu - \mu(\mathcal{X}).$$

We see this from the following

$$\|\mu\| = \sup_{\|\delta\|\leq 1} \int \delta d\mu = 2(\sup_{\|\delta\|\leq 1} \int \frac{\delta+1}{2}\, d\mu - \tfrac{1}{2}\mu(\mathcal{X})) = 2 \sup_{0\leq\delta\leq 1} \int \delta d\mu - \mu(\mathcal{X}).$$

According to Proposition I.3.1 we know that $\mathcal{E}$ is $(\varepsilon_1,\varepsilon_2)$-deficient with respect to $\mathcal{F}$ if and only if

(I.3.3) $\|a_1\mu_1+a_2\mu_2\| \geq \|a_1\nu_1+a_2\nu_2\| - \varepsilon_1|a_1| - \varepsilon_2|a_2|, \quad \forall a_1,a_2 \in R$

where we as usual let $\mathcal{E} = (\mathcal{X},\mathcal{A},\mu_1,\mu_2)$ and $\mathcal{F} = (\mathcal{Y},\mathcal{B},\nu_1,\nu_2)$. But now we have

$$H_{\mathcal{E}}(a_1,a_2) = \sup_{(x_1,x_2)\in V_{\mathcal{E}}} (a_1x_1+a_2x_2) = \sup_{0\leq\delta\leq 1} (a_1\int\delta d\mu_1+a_2\int\delta d\mu_2)$$

$$= \sup_{0\leq\delta\leq 1} \int\delta d(a_1\mu_1+a_2\mu_2) = \tfrac{1}{2}(\|a_1\mu_1+a_2\mu_2\|+a_1+a_2\mu_2(\mathcal{X}))$$

because of the equation above. Therefore (I.3.3) is equivalent to

$$2H_{\mathcal{E}}(a_1,a_2)-a_1-a_2\mu_2(\mathcal{X}) \geq 2H_{\mathcal{F}}(a_1,a_2)-a_1-a_2\nu_2(\mathcal{Y})$$
$$- 2H_{\varepsilon_1,\varepsilon_2}(a_1,a_2), \ \forall a_1,a_2\in R$$

which in turn, since $\mu_2(\mathcal{X}) = \nu_2(\mathcal{Y})$, is equivalent to

$$H_{\mathcal{E}}(a_1,a_2) + H_{\varepsilon_1,\varepsilon_2}(a_1,a_2) \geq H_{\mathcal{F}}(a_1,a_2), \ \forall a_1,a_2\in R$$

and the proof is completed. □

It is now possible to describe $(\varepsilon_1,\varepsilon_2)$-deficiency by means of V-sets.

PROPOSITION I.3.10.

$\mathcal{E}$ is $(\varepsilon_1,\varepsilon_2)$-deficient with respect to $\mathcal{F}$

$$\Updownarrow$$

$$V_{\mathcal{E}} + V_{\varepsilon_1,\varepsilon_2} \supset V_{\mathcal{F}}$$

PROOF: This is simply a reformulation of the previous proposition since we have the following two properties of the support function ϕ_K of a compact, convex set K:

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

and

$$K_1 \subset K_2 \iff \phi_{K_2} \leq \phi_{K_2}.$$ □

COROLLARY I.3.11.

$$\mathcal{E} > \mathcal{F} \iff V_{\mathcal{E}} \supset V_{\mathcal{F}}$$
$$\mathcal{E} \sim \mathcal{F} \iff V_{\mathcal{E}} = V_{\mathcal{F}}.$$

PROOF: This follows from Proposition I.3.10 by considering (0,0)-deficiency. □

Corollary I.3.11 shows that the V-sets are well suited for describing "more informative" and "equivalence" between pseudo dishotomies satisfying (I.3.1) and (I.3.2). As we have pointed out before, the V-set characterizes the pseudo dichotomy up to an equivalence. Later we'll discuss the geometrical aspects of this corollary.

We now proceed with a study of the relationship between ε-deficiency and the β-function. Let $\mathcal{E} = (\mathcal{X},\mathcal{A},\mu_1,\mu_2)$ be a pseudo dichotomy. Then $\beta_{\mathcal{E}}$ is defined on $[0,1]$ by

$$\beta_{\mathcal{E}}(\alpha) = \sup\{y \mid (\alpha,y)\in V_{\mathcal{E}}\}.$$

We extend the domain of $\beta_{\mathcal{E}}$ to R by defining

$$\beta_{\mathcal{E}}(\alpha) = \begin{cases} \beta\ (0), & \alpha<0 \\ \beta\ (1), & \alpha>1 \end{cases} .$$

Our intention with this is to be able to present the next result in a simpler form.

PROPOSITION I.3.12.

$\mathcal{E}$ is $(\varepsilon_1,\varepsilon_2)$-deficient with respect to $\mathcal{F}$

$$\Updownarrow$$

(I.3.6) $$\sup\{\beta_{\mathcal{E}}(x) \mid x\in[\alpha-\tfrac{\varepsilon_1}{2}, \alpha+\tfrac{\varepsilon_1}{2}]\} \geq \beta_{\mathcal{F}}(\alpha)-\tfrac{\varepsilon_2}{2},\quad \forall\alpha\in[0,1].$$

PROOF: let $\mathcal{E} = (\mathcal{X},\mathcal{A},\mu_1,\mu_2)$ and $\mathcal{F} = (\mathcal{Y},\mathcal{B},\nu_1,\nu_2)$. Let furthermore

$$H = \{(x,y)\,|\,y > \mu_2(\mathcal{X})x,\ x\in R\}$$

where $\mu_2(\mathcal{X}) = \nu_2(\mathcal{Y})$ is the total mass of μ_2 and ν_2. We now put $\hat{K} = K\cap H$ for any $K \subset R^2$, and the following equivalences will then hold:

(I.3.7) $\mathcal{E}$ is $(\varepsilon_1,\varepsilon)$-deficient with respect to $\mathcal{F}$

$\Updownarrow$

(I.3.8) $V_{\mathcal{E}} + V_{\varepsilon_1,\varepsilon_2} \supset V_{\mathcal{F}}$

$\Updownarrow$

(I.3.9) $\hat{V}_{\mathcal{E}} + \hat{V}_{\varepsilon_1,\varepsilon_2} \supset \hat{V}_{\mathcal{F}}$

The equivalence between (I.3.7) and (I.3.8) is due to Proposition I.3.10. The equivalence between (I.3.8) and (I.3.9) comes from the fact that $V_{\mathcal{E}}+V_{\varepsilon_1,\varepsilon_2}$, like $V_{\mathcal{F}}$, is symmetrical about the point $(\frac{1}{2},\frac{1}{2}\mu_2(\mathcal{X}))$, and because we can apply the following lemma:

LEMMA: Let $A,B \subset R^2$ and assume that A and B are symmetrical about $a\in R^2$. Then we have:

$$A \subset B \iff \hat{A} \subset \hat{B}.$$

PROOF OF THE LEMMA: Assume first that $A \subset B$. Then $\hat{A} = A\cap H \subset B\cap H = \hat{B}$, so $\hat{A} \subset \hat{B}$.

Assume then that $\hat{A} \subset \hat{B}$, and let $g:R^2\to R^2$ be defined by

$$g(x_1,x_2) = 2a-(x_1,x_2)$$

which means that x and $g(x)$ lie symmetrical about a. Then $A = \hat{A}\cup g(\hat{A})$ and $B = \hat{B}\cup g(\hat{B})$ since both A and B are symmetrical about a. Consequently $\hat{A} \subset \hat{B} \Rightarrow g(\hat{A}) \subset g(\hat{B}) \Rightarrow$ $\hat{A} \cup g(\hat{A}) \subset \hat{B} \cup g(\hat{B}) \Rightarrow A \subset B$, which completes the proof of the lemma.

We have therefore proved the equivalence between (I.3.7) and (I.3.9) (because the fact that $V_{\mathcal{E}} + V_{\varepsilon_1,\varepsilon_2}$ is symmetrical about $(\frac{1}{2},\frac{1}{2}\mu_2(\mathcal{X}))$ follows easily from the symmetry of $V_{\mathcal{E}}$ about $(\frac{1}{2},\frac{1}{2}\mu_2(\mathcal{X}))$ and the symmetry of $V_{\varepsilon_1,\varepsilon_2}$ about $(0,0)$).

Further it is clear that (I.3.9) is equivalent to

(I.3.10) $\quad \beta_{\mathcal{F}}(\alpha) \leq \sup\{y \mid (\alpha,y) \in V_{\mathcal{E}} + V_{\varepsilon_1,\varepsilon_2}\}, \quad \forall \alpha \in [0,1].$

We now finish this proof by applying the next lemma.

LEMMA: $\sup\{y \mid (\alpha,y) \in V_{\mathcal{E}} + V_{\varepsilon_1,\varepsilon_2}\} = \sup\{\beta_{\mathcal{E}}(x) \mid x \in [\alpha - \frac{\varepsilon_1}{2}, \alpha + \frac{\varepsilon_1}{2}]\} + \frac{\varepsilon_2}{2}$, $\forall \alpha \in [0,1]$.

PROOF OF LEMMA: We have $\sup\{y \mid (\alpha,y) \in V_{\mathcal{E}} + V_{\varepsilon_1,\varepsilon_2}\} = \sup\{y \mid (v_1,v_2) \in V_{\mathcal{E}}$ and $|\alpha - v_1| \leq \frac{\varepsilon_1}{2}, |y - v_2| \leq \frac{\varepsilon_2}{2}\} = \sup\{y \mid |\alpha - v_1| \leq \frac{\varepsilon_1}{2}$ and $(v_1,y) \in V_{\mathcal{E}}\} + \frac{\varepsilon_2}{2} = \sup\{\beta_{\mathcal{E}}(x) \mid x \in [\alpha - \frac{\varepsilon_1}{2}, \alpha + \frac{\varepsilon_1}{2}]\} + \frac{\varepsilon_2}{2}$, which completes the proof of the lemma. □

COROLLARY I.3.13.

$$\mathcal{E} > \mathcal{F} \iff \beta_{\mathcal{E}} > \beta_{\mathcal{F}}$$
$$\mathcal{E} \sim \mathcal{F} \iff \beta_{\mathcal{E}} = \beta_{\mathcal{F}}.$$

PROOF: This follows directly from Proposition I.3.12 by considering (0,0)-deficiency. □

This corollary shows that the concepts "more informative" and "equivalence" between pseudo dichotomies can be expressed quite easily trough the β-functions. In particular we see that the β-function characterizes the pseudo dichotomy up to an equivalence. Furthermore we can express the dot deficiency between pseudo dichotomies with the aid of β-functions, such as the next proposition says.

PROPOSITION I:3.14.

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

$$\dot{\Delta}(\mathcal{E},\mathcal{F}) = \sup_\alpha|\beta_{\mathcal{F}}(\alpha)-\beta_{\mathcal{E}}(\alpha)|.$$

PROOF: $\dot{\delta}(\mathcal{E},\mathcal{F}) = \frac{1}{2}\inf\{\varepsilon>0 \mid \mathcal{E}$ is $(0,\varepsilon)$-deficient with respect to $\mathcal{F}\}$
$= \frac{1}{2}\inf\{\varepsilon>0 \mid \beta_{\mathcal{E}}(\alpha)>\beta_{\mathcal{F}}(\alpha)-\frac{\varepsilon}{2},\ \forall\alpha\in[0,1]\}$
$= \frac{1}{2}\inf\{\varepsilon>0 \mid \beta_{\mathcal{F}}(\alpha)-\beta_{\mathcal{E}}(\alpha)<\frac{\varepsilon}{2},\ \forall\alpha\in[0,1]\} = \sup_\alpha(\beta_{\mathcal{F}}(\alpha)-\beta_{\mathcal{E}}(\alpha))^+.$

The expression for $\dot{\Delta}(\mathcal{E},\mathcal{F})$ comes from the fact that $\dot{\Delta}(\mathcal{E},\mathcal{F}) = \dot{\delta}(\mathcal{E},\mathcal{F})\vee\dot{\delta}(\mathcal{F},\mathcal{E})$. □

We shall also give a characterization of "more informative", which holds under certain additional assumptions on $\mathcal{E}$ and $\mathcal{F}$.

PROPOSITION I.3.15.

Let $\mathcal{E} = (\mathcal{X},\mathcal{A},\mu_1,\mu_2)$ and $\mathcal{F} = (\mathcal{Y},\mathcal{B},\nu_1,\nu_2)$ be two pseudo dichotomies, where $\mu_1,\nu_1>0$, $\mu_2<<\mu_1$, $\nu_2<<\nu_1$ and $\Delta_1(\mathcal{E},\mathcal{F}) = 0$.

We define

$$s_{\mathcal{E}} = \frac{d\mu_2}{d\mu_1},\ s_{\mathcal{F}} = \frac{d\nu_2}{d\nu_1},\ F_{\mathcal{E}} = \mu_1 s_{\mathcal{E}}^{-1} \text{ and } F_{\mathcal{F}} = \nu_1 s_{\mathcal{F}}^{-1}.$$

Then the following equivalence holds:

$$\mathcal{E} > \mathcal{F}$$

$$\Updownarrow$$

(I.3.11) $\int\phi dF_{\mathcal{E}} > \int\phi dF_{\mathcal{F}}$ for every convex function $\phi:R\to R$.

PROOF: According to Theorem I.2.2 $\mathcal{E} > \mathcal{F}$ will be equivalent to

(I.3.12) $$\psi(\mathcal{E}) \geq \psi(\mathcal{F}), \quad \forall\psi\in\Psi^{(2)}$$

Assume first that $\mathcal{E} > \mathcal{F}$ and let $\psi\in\Psi^{(2)}$. Then, since $\mu_1>0$ and $\mu_2<<\mu_1$, we have

$$\psi(\mathcal{E}) = \int\psi(\frac{d\mu_1}{d\mu_1},\frac{d\mu_2}{d\mu_1})d\mu_1 = \int\psi(1,s_{\mathcal{E}})d\mu_1$$
$$= \int\psi(1,x)(\mu_1 s_{\mathcal{E}}^{-1})(dx) = \int\psi(1,x)F_{\mathcal{E}}(dx)$$

by applying the change of variable formula. But since every convex function $\phi:R\to R$ can be written as $\lim_{n\to\infty}\psi_n(1;)$ for a suitable pointwise increasing sequence $\{\psi_n\}_{n=1}^{\infty}$ in $\Psi^{(2)}$, we see (from monotone convergence theorem) that

$$\int\phi dF_{\mathcal{E}} \geq \int\phi dF_{\mathcal{F}}$$

holds for every convex $\phi:R\to R$ and (I.3.11) holds.

Assume now that (I.3.11) holds, and let $\psi\in\Psi^{(2)}$. Because $\phi(x) = \psi(1,x)$ is convex, we know that

$$\int\psi(1,x)dF_{\mathcal{E}} \geq \int\psi(1,x)dF_{\mathcal{F}}.$$

Consequently, due to the equalities $\psi(\mathcal{E}) = \int\psi(1,x)dF_{\mathcal{E}}$ and $\psi(\mathcal{F}) = \int\psi(1,x)dF_{\mathcal{F}}$, (I.3.12) will hold, which implies that $\mathcal{E} > \mathcal{F}$, and the proof is completed. □

I.4. An important example.

In this section we shall calculate $V_{\mathcal{E}}$, $\beta_{\mathcal{E}}$ and $U_{\mathcal{E}}$ of a certain kind of pseudo dichotomy $\mathcal{E}$, which will be of importance in the following chapters.

Let

$$\mathcal{E} = (\{1,\ldots,n\}, \mathcal{P}(\{1,\ldots,n\}), \mu_1, \mu_2)$$

where

$$\mu_1(\{j\}) = \frac{1}{n} \quad \text{and} \quad \mu_2(\{j\}) = x_j; \quad j = 1,\ldots,n.$$

Here $x_1,\ldots,x_n$ are arbitrary real numbers.

First we'll determine $V_{\mathcal{E}}$. Because $\mathcal{X}$ is finite, one can show that $V_{\mathcal{E}} = \langle V_{\mathcal{E}}' \rangle$, where

$$V_{\mathcal{E}}' = \{(\int \delta d\mu_1, \int \delta d\mu_2) \mid \delta : \mathcal{X} \to [0,1] \text{ is non-randomized}\}$$

by applying separating hyperplane theorem. A non-randomized decision rule δ is such that $\delta(j) = \delta_j \in \{0,1\}$; $j = 1,\ldots,n$. This result can be shown analogously to the fact that "a risk set is the convex hull of the non randomized risk set" (see reference [3]).

This simplifies the work in connection with determining $V_{\mathcal{E}}$ considerably, because $V_{\mathcal{E}}'$ is a finite set and quite easy to determine.

Let $\delta : \mathcal{X} \to [0,1]$ be non-randomized and put $\delta(j) = \delta_j$; $j = 1,\ldots,n$. Then

$$(\int \delta d\mu_1, \int \delta d\mu_2) = (\sum_{j:\delta_j=1} \frac{1}{n}, \sum_{j:\delta_j=1} x_j)$$

and consequently

$$V_{\mathcal{E}} = \langle \{(\tfrac{k}{n}, \sum_{i=1}^{k} x_{j_i}) \mid k \in \{0,\ldots,n\} \text{ and } \{j_1,\ldots,j_k\} \subset \{1,\ldots,n\}$$

$$\text{where } j_{i_1} \neq j_{i_2} \text{ when } i_1 \neq i_2 \rangle$$

$$= \langle \{(\tfrac{k}{n}, \sum_{j=1}^{k} x_{[j]}) \mid k = 0,1,\ldots,n\} \cup \{(\tfrac{k}{n}, \sum_{j=1}^{k} x_{(j)}) \mid k = 0,1,\ldots,n\} \rangle$$

since all the points $(\frac{k}{n}, \sum_{j=1}^{k} x_{j_i})$, by varying $j_1,\ldots,j_k$, lie on the line segment between $(\frac{k}{n}, \sum_{j=1}^{k} x_{[j]})$ and $(\frac{k}{n}, \sum_{j=1}^{k} x_{(j)})$.

This means that V_ℓ is given by

$$V_\ell = \langle \{(\frac{k}{n}, \sum_{j=1}^{k} x_{[j]}) \mid k = 0,1,\ldots,n\} \cup \{(\frac{k}{n}, \sum_{j=1}^{k} x_{(j)}) \mid k = 0,1,\ldots,n\} \rangle .$$

From this it is easy to find β_ℓ, which is defined by $\beta_\ell(\alpha) = \sup\{y \mid (\alpha,y) \in V_\ell\}$. We see that

$$\beta_\ell(\alpha) = \sum_{j=1}^{k} x_{[j]} \quad \text{when} \quad \alpha = \frac{k}{n},\ k = 0,1,\ldots,n$$

and that β_ℓ is piecewise linear and continous on $[0,1]$.

We let μ denote the counting measure on $\{1,\ldots,n\}$, and it is then possible to calculate U_ℓ:

$$U_\ell(\xi) = \| \xi\mu_1 - \mu_2 \| = \int \left| \frac{d(\xi\mu_1 - \mu_2)}{d\mu} \right| d\mu = \sum_{j=1}^{n} \left| \frac{\xi}{n} - x_j \right| .$$

Consequently the expression for U_ℓ is

$$U_\ell(\xi) = \sum_{j=1}^{n} \left| \frac{\varepsilon}{n} - x_j \right|, \quad \xi \in R.$$

In our example we have started off by determining V_ℓ, and then we have found β_ℓ. We shall now give some comments on an alternative manner of proceeding. It is namely possible to attack the problem differently, by first calculating β_ℓ and thereafter use the wellknown geometrical properties of V-sets in order to determine V_ℓ. This method is based on a generalized version of Neyman-Pearson's lemma.

Let $\alpha \in [0,1]$. We wish to calculate

$$\beta_\ell(\alpha) = \sup\{y \mid (\alpha,y) \in V_\ell\}.$$

Since

$$V_\ell = \{(\int \delta d\mu_1, \int \delta d\mu_2) \mid \delta \text{ is a function from } \{1,\ldots,n\} \text{ to } [0,1]\}.$$

This is the same as maximizing $\int \delta d\mu_2$ under the constraint $\int \delta d\mu_1 = \alpha$, where δ is a function from $\{1,\ldots,n\}$ to $[0,1]$.

By introducing $\mu = \mu_1 + |\mu_2|$, we se that $\mu_1, \mu_2 << \mu$ and

$$\int \delta d\mu_i = \int \delta f_i d\mu, \; i = 1,2,$$

where $f_i = \frac{d\mu_i}{d\mu}$. We are then in the siutation described in Proposition I.2.7. (The generalized version of Neyman-Pearson's lemma.) This proposition assumes the existence of a maximizing δ and it says that this δ must satisfy

$$\delta(x) = \begin{cases} 1 & \text{when} \quad f_2(x) > cf_1(x) \\ \gamma & \text{when} \quad f_2(x) = cf_1(x) \\ 0 & \text{when} \quad f_2(x) < cf_1(x), \end{cases}$$

where c and γ are constants $(0<\gamma<1)$ such that

$$\int \delta f_1 d\mu = \alpha.$$

After some elementary calculations, we now get

$$\beta_\ell(\alpha) = \sum_{j=1}^{k} x_{[j]} + (n\alpha - k) x_{[k+1]}, \quad \text{when} \quad \frac{k}{n} < \alpha < \frac{k+1}{n}; \; k = 0,1,\ldots,n-1$$

and

$$\beta_\ell(1) = \sum_{j=1}^{k} x_j$$

which is the same result as the one we got earlier.

According to Proposition I.3.8 we know that V_ℓ is compact and convex and that V_ℓ is symmetrical about the point $(\frac{1}{2}, \frac{1}{2}\sum x_j)$. This implies that "the lower boundary" of V_ℓ is determined by the graph of the following function:

$$\underline{\beta}_\ell : [0,1] \to R$$

$$\underline{\beta}_\ell(\alpha) = \sum_j x_j - \beta\,(1-\alpha); \quad \alpha \in [0,1]$$

because β_ℓ and $\underline{\beta}_\ell$ are symmetrical about $(\frac{1}{2},\frac{1}{2}\sum_j x_j)$.

Consequently

$$V_\ell = \{(\alpha,y) \mid \underline{\beta}_\ell(\alpha) < y < \beta_\ell(\alpha),\ \alpha\in[0,1]\}$$

and it is easy to realize that this is the same set as the one we found originally.

If we didn't know the symmetry-property of the V-sets, we could have found $\underline{\beta}_\ell$ alternativly by using the generalized version of Neyman-Pearson's lemma in order to minimize

$$\int \delta f_2 d\mu$$

among all decision rules δ satisfying $\int \delta f_1 d\mu = \alpha$.

This shows that the generalized version of Neyman-Pearson's lemma plays a fundamental role in the example of this section. Since these pseudo dichotomies will be of great importance in chapter II on majorization, this generalization (Theorem 5 in referance [2]) is quite essential as regards characterizations of majorization.

We shall end this chapter by giving a concrete example in order to illustrate β_ℓ, $\underline{\beta}_\ell$, V_ℓ geometrically.

Let $n = 4$ and $x = (6,4,1,-1)$.

The following table gives us a few values of β_ℓ and $\underline{\beta}_\ell$.

α	$\beta_\ell(\alpha)$	$\underline{\beta}_\ell(\alpha)$
0	0	0
$\frac{1}{4}$	6	-1
$\frac{1}{2}$	10	0
$\frac{3}{4}$	11	4
1	10	10

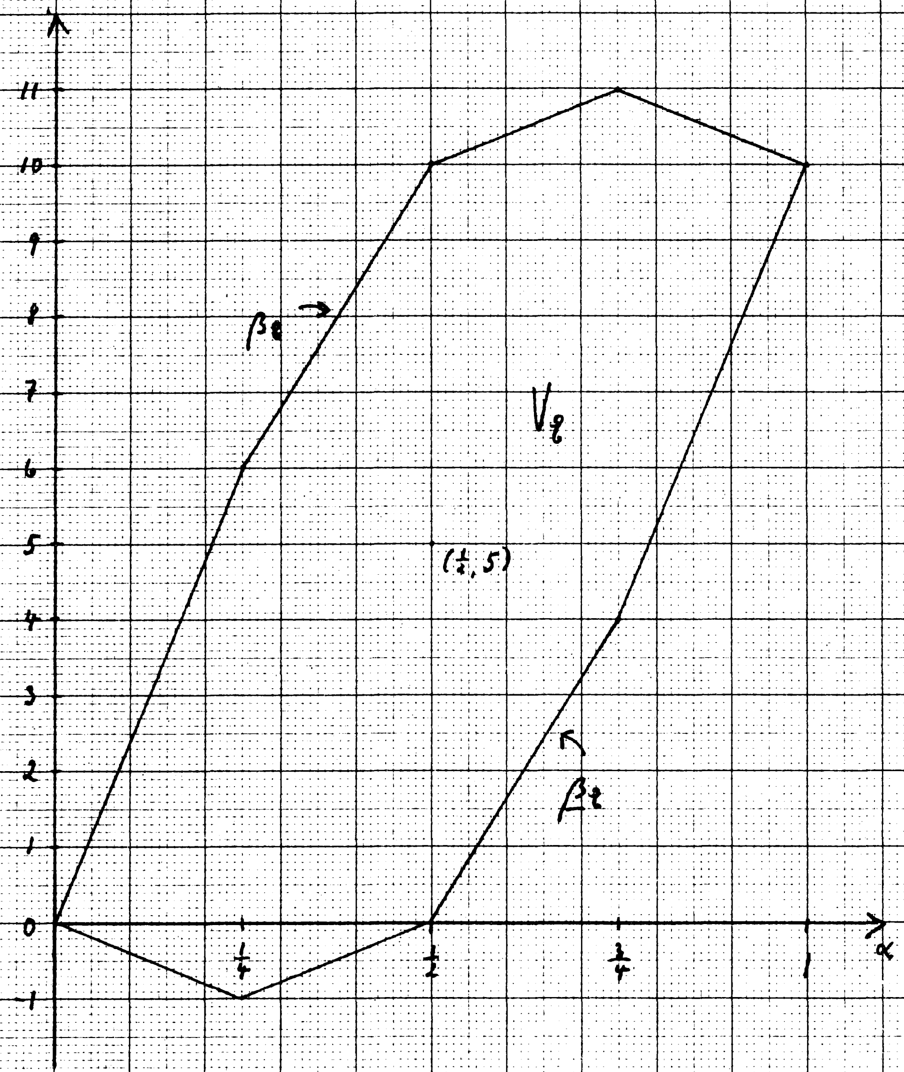
On fig. I.4.1 β_ℓ, $\underline{\beta}_\ell$ and V_ℓ are all drawn, and we see how the graphs of β_ℓ and $\underline{\beta}_\ell$ constitutes respectively the "upper" and "lower" boundary of V_ℓ. We also see that V_ℓ is symmetric about the point $(\frac{1}{2},5)$. Note that β_ℓ is concave and $\underline{\beta}_\ell$ convex; this holds in the general case, too.

When calculating V_ℓ, one has to treat five different intervals separately, and the result is:

$$V_\ell(\xi) = \begin{cases} -\xi + 10 & \text{when } \xi \leq -4 \\ -\frac{\xi}{2} + 12 & \text{when } -4 < \xi \leq 4 \\ 10 & \text{when } 4 < \xi \leq 16 \\ \frac{\xi}{2} + 2 & \text{when } 16 < \xi \leq 24 \\ \xi - 10 & \text{when } 24 < \xi \end{cases}$$

By drawing the graph of U_ℓ, we see that this function is convex. This also holds in the general case, which easily can be shown analytically.

Figure I.4.1



CHAPTER II: MAJORIZATION

II.1. Definition and characterizations.

The mathematical concept "majorization" is used in different context in litterature. Most common is majorization between vectors, which we shall study in this chapter.

DEFINITION II.1.1.

Let $x = (x_1, \ldots, x_n)$ and $y = (y_1, \ldots, y_n)$ be two vectors in R^n. If

$$\text{(II.1.1)} \qquad \sum_{j=1}^{k} x_{[j]} \leq \sum_{j=1}^{k} y_{[j]}; \quad k = 1, \ldots, n-1$$

and

$$\text{(II.1.2)} \qquad \sum_{j=1}^{n} x_j = \sum_{j=1}^{n} y_j$$

hold, we say that x is majorized by y, and in that case we write $x \prec y$.

That x is majorized by y expresses that the components of x "are less spread out" than the components of y.

EXAMPLE II.1.2.

The concept of majorization as defined above can be used to describe whether a certain income-distribution over a population is "more equal" than another such income-distribution of the same amount of money. If $x = (x_1, \ldots, x_n)$ and $y = (y_1, \ldots, y_n)$ denote the different individual incomes in a population of n individuals according to two ways of distributing the total income, we can say that the income-distribution x is "more equal" than the income-distribution y when $x \prec y$. This means that the sum of the k greatest incomes in the distribution y is at least as great as the sum of the k greatest incomes in the distribution x, where k runs through $\{1,\dots,n-1\}$.

We shall now list several characterizations of this majorization concept. The next theorem therefore gives different conditions on the vectors x and y, each of these being equivalent to $x \prec y$. These equivalences are well-known, and they can be found in reference [3].

We remind ourselves that K_y, whenever $y \in R^n$, denotes the convex hull of the set of all possible permutations of y, and that $\mathcal{M}^D_{n,n}$ is the set of all doubly-stochastic $n \times n$ matrices.

THEOREM II.1.3.

Let $x = (x_1,\dots,x_n)$ and $y = (y_1,\dots,y_n)$ be two arbitrary vectors in R^n. Then (II.1.3)-(II.1.9) are all equivalent:

(II.1.3) $x \prec y$

(II.1.4) $\sum_{j=1}^{k} x_{(j)} \leq \sum_{j=1}^{k} y_{(j)}$; $k = 1,\dots,n-1$, and $\sum_{j=1}^{n} x_j = \sum_{j=1}^{n} y_j$

(II.1.5) $\sum_j |x_j - a| \leq \sum_j |y_j - a|$; $\forall a \in R$, and $\sum_j x_j = \sum_j y_j$

(II.1.6) $\sum_j (x_j - a)^+ \leq \sum_j (y_j - a)^+$; $\forall a \in R$, and $\sum_j x_j = \sum_j y_j$

(II.1.7) $\sum_j \phi(x_j) \leq \sum_j \phi(y_j)$ for every convex $\phi: R \to R$, and $\sum_j x_j = \sum_j y_j$

(II.1.8) $x \in K_y$

(II.1.9) $\exists M \in \mathcal{M}^D_{n,n}: x = yM$

II.2. Majorization as a statistical concept.

We shall in this section show that majorization can be considered as a statistical concept.

Let $x = (x_1,\ldots,x_n) \in R^n$. We then define $\mathcal{E}_{\frac{1}{n}e,x}$ as the pseudo dichotomy

$$(\{1,\ldots,n\},\ \mathcal{P}(\{1,\ldots,n\}),\mu_1,\mu_2),$$

where $\mu_1(\{j\}) = \frac{1}{n}$ and $\mu_2(\{j\} = x_j;\ j = 1,\ldots,n.$

This implies that $\mathcal{E}_{\frac{1}{n}e,x}$ has a pseudo experiment matrix $P_{\mathcal{E}_{\frac{1}{n}e,x}}$ defined by

$$P_{\mathcal{E}_{\frac{1}{n}e,x}} = \begin{pmatrix} \frac{1}{n},\ldots,\frac{1}{n} \\ x,\ldots,x_n \end{pmatrix}$$

In this situation $\mathcal{E}_{\frac{1}{n}e,x}$ is denoted a majorization pseudo dichotomy, and if $x \in K_n$ (if x is a probability vector) $\mathcal{E}_{\frac{1}{n}e,x}$ is denoted a majorization dichotomy.

Majorization between vectors x and y now turns out to be equivalent to the relation "more informative than" between the corresponding majorization pseudo dichotomies. This is most easily seen by applying the Markov-kernel criterion for "more informative" in this situation.

PROPOSITION II.2.1.

Let $x,y \in R^n$. Then the following holds:

$$x \prec y$$

$$\Updownarrow$$

$$\mathcal{E}_{\frac{1}{n}e,x} \leq \mathcal{E}_{\frac{1}{n}e,y}\ .$$

PROOF: By applying Corollary I.2.6 we get:

$$\mathcal{E}_{\frac{1}{n}e,x} \leq \mathcal{E}_{\frac{1}{n}e,y}$$

$$\Updownarrow$$

$$\exists M \in \mathcal{M}_{n,n} : P_{\mathcal{E}_{\frac{1}{n}e,x}} = P_{\mathcal{E}_{\frac{1}{n}e,y}} M$$

$$\Updownarrow$$

$$\exists M = (m_{ij})_{i,j=1,1}^{n,n} \in \mathcal{M}_{n,n} : \begin{pmatrix} \frac{1}{n}, \dots, \frac{1}{n} \\ x_1, \dots, x_n \end{pmatrix} = \begin{pmatrix} \frac{1}{n}, \dots, \frac{1}{n} \\ y_1, \dots, y_n \end{pmatrix} \begin{pmatrix} m_{11}, \dots, m_{1n} \\ \vdots \qquad \vdots \\ m_{n1}, \dots, m_{nn} \end{pmatrix}$$

$$\Updownarrow$$

$$\exists M \in \mathcal{M}_{n,n} : \sum_{i=1}^{n} m_{ij} = 1; \; j = 1, \dots, n \quad \text{and} \quad x = yM$$

$$\Updownarrow$$

$$\exists M \in \mathcal{M}_{n,n}^{D} : x = yM$$

$$\Updownarrow$$

$$x \prec y$$

where the last equivalence is due to Theorem II.1.3. □

Proposition II.2.1 gives the connection between majorization and the theory on pseudo dichotomies. Since we have several characterizations of the "more informative"-concept, it is natural to pose the following two questions:

- which characterizations of majorization in Theorem II.1.3 are consequences of the theory of pseudo dichotomies?
- can the theory of pseudo dichotomies also give new characterizations and interpretations?

The rest of this chapter is devoted these two questions.

PROPOSITION II.2.2.

Let $x,y \in R^n$ and assume that $\sum_j x_j = \sum_j y_j$. Then the statements (II.2.1)-(II.2.4) are all equivalent.

(II.2.1) $\mathcal{E}_{\frac{1}{n}e,x} \leq \mathcal{E}_{\frac{1}{n}e,y}$

(II.2.2) $\beta_{\mathcal{E}_{\frac{1}{n}e,x}} \leq \beta_{\mathcal{E}_{\frac{1}{n}e,y}}$

(II.2.3) $U_{\mathcal{E}_{\frac{1}{n}e,x}} \leq U_{\mathcal{E}_{\frac{1}{n}e,y}}$

(II.2.4 $\exists M \in \mathcal{M}_{n,n} : P_{\mathcal{E}_{\frac{1}{n}e,x}} \leq P_{\mathcal{E}_{\frac{1}{n}e,y}} M$

PROOF: The equivalence between (II.2.1) and (II.2.2), (II.2.3), (II.2.4) follow respectively from Corollary I.3.13, Corollary I.3.5 and Corollary I.2.7. □

By using the expressions developed in section I.4 for the β-function and the U-function of a majorization pseudo dichotomy, we see that

(II.2.2) <=> $\sum_{j=1}^{k} x_{[j]} \leq \sum_{j=1}^{k} y_{[j]}$; $k = 1,\ldots,n-1$

and that

(II.2.3) <=> $\sum_j |x_j - a| \leq \sum_j |y_j - a|$, $\forall a \in R$.

Besides it has just been shown in Proposition II.2.1 that

(II.2.4) $\exists M \in \mathcal{M}^D_{n,n}: x = yM$.

This means that we, by using the theory of comparison of pseudo experiments, have proved the equivalences between (II.1.3), (II.1.5) and (II.1.9) in Theorem II.1.3 in a new way. We shall also comment the other characterizations in this theorem.

The equivalence between (II.1.5) and (II.1.6) is immediate and can be seen by using the fact that $|a| = 2a^+ - a$ and that $\sum_j x_j = \sum_j y_j$. As regards (II.1.8), it can be shown to be equivalent to (II.1.9) by applying Birkhoff's theorem (which says that the set of all doubly-stochastic $n\times n$ matrices is the convex hull of the set of all $n\times n$ permutation matrices). Furthermore one easily realizes that (II.1.3) and (II.1.4) are equivalent since $\sum_j x_j = \sum_j y_j$, but this can also be seen after a geometrical discussion. We know that

$$x \prec y \iff \ell_{\frac{1}{n}e,x} < \ell_{\frac{1}{n}e,y} \iff V_{\ell_{\frac{1}{n}e,x}} \subset V_{\ell_{\frac{1}{n}e,y}}$$

Because of the symmetry-property of the V-sets (see Proposition I.3.8) this is equivalent to the following: "the lower boundary" of $V_{\ell_{\frac{1}{n}e,x}}$ lies above "the lower boundary" of $V_{\ell_{\frac{1}{n}e,y}}$. Since the breakpoints on the lower boundary of $V_{\ell_{\frac{1}{n}e,x}}$ are among the points $(\frac{k}{n}, \sum_{j=1}^{k} x_{[j]})$; $k = 1,\dots,n-1$, one realizes (from this informal argument) that (II.1.3) and (II.1.4) are equivalent.

We can also obtain the equivalence between (II.1.3) and (II.1.7) (the characterization of majorization by inequalities for convex functions) as a result of the theory in chapter I.

PROPOSITION II.2.3.

Let $x,y\in R^n$ and assume that $\sum_j x_j = \sum_j y_j$. Then the following equivalence holds:

$$x \prec y$$

$$\Updownarrow$$

$$\sum_j \phi(x_j) \le \sum_j \phi(y_j) \quad \text{for every convex function } \phi: R\to R.$$

PROOF: We see that $\mathcal{E}_{\frac{1}{n}e,x}$ and $\mathcal{E}_{\frac{1}{n}e,y}$ satisfy the assumptions in Proposition I.3.15, and from this we know that

$$\mathcal{E}_{\frac{1}{n}e,x} < \mathcal{E}_{\frac{1}{n}e,y}$$

$$\Updownarrow$$

(I.2.5) $\int \phi dF_{\mathcal{E}_{\frac{1}{n}e,x}} < \int \phi dF_{\mathcal{E}_{\frac{1}{n}e,y}}$ for every convex function $\phi: R \to R$.

With the same notation as in Proposition I.3.15, we have

$$\int \phi dF_{\mathcal{E}_{\frac{1}{n}e,x}} = \int \phi d\mu_1 s^{-1}_{\mathcal{E}_{\frac{1}{n}e,x}} = \int \phi o s_{\mathcal{E}_{\frac{1}{n}e,x}} d\mu_1 = \sum_j \phi(nx_j)\frac{1}{n}$$

from the change of variable formula. Thus (II.2.5) is equivalent to

(II.2.6) $\sum_j \phi(nx_j) < \sum_j \phi(ny_j)$ for every convex $\phi: R \to R$.

But since $x \to \phi(nx)$ is convex if and only if $x \to \phi(x)$ is convex, we have completed the proof by applying Proposition II.2.1. □

We now turn to the second question that was posed earlier in this chapter: Can the theory of pseudo experiments give us new characterizations of majorization?

The first result in this direction is available when we return to the definition of "more informative".

PROPOSITION II.2.4.

Let $x,y \in R^n$. Then we have

$$x \prec y$$

$$\Updownarrow$$

to every decision space $T = \{1,\dots,k\}$. where $k = 1,2,\dots,$ and to every bounded loss function $L_\theta(t)$; $\theta = 1,2$, $t \in T$, and to every decision rule ρ in $\mathcal{E}_{\frac{1}{n}e,x}$, there corresponds a decision rule δ in $\mathcal{E}_{\frac{1}{n}e,y}$ such that

$$r_{\mathcal{E}_{\frac{1}{n}e,y}}(\theta,\delta) \le r_{\mathcal{E}_{\frac{1}{n}e,x}}(\theta,\rho); \quad \theta = 1,2,$$

where $r_{\mathcal{E}_{\frac{1}{n}e,x}}(\theta,\rho)$ denotes the risk in $\mathcal{E}_{\frac{1}{n}e,x}$ by using the decision rule ρ when θ is the underlying value of the parameter.

PROOF: This is simply the definition of $\mathcal{E}_{\frac{1}{n}e,x} \le \mathcal{E}_{\frac{1}{n}e,y}$ (see Definition I.1.1) combined with Proposition II.2.1). □

When $x,y \in K_n$, $r_{\mathcal{E}_{\frac{1}{n}e,x}}(\theta,\rho)$ and $r_{\mathcal{E}_{\frac{1}{n}e,y}}(\theta,\delta)$ will in fact be the risk functions in the original sense, because $\mathcal{E}_{\frac{1}{n}e,x}$ and $\mathcal{E}_{\frac{1}{n}e,y}$ then are experiments. When $x \notin K_n$ or $y \notin K_n$, Proposition II.2.4 still holds, but $\mathcal{E}_{\frac{1}{n}e,x}$ or $\mathcal{E}_{\frac{1}{n}e,y}$ are in that case no longer experiments and the probabilistic interpretation disappears.

Loosely speaking we can say that $x \prec y$ if and only if every finite decision problem can be solved better, or just as good, in

$\mathcal{E}_{\frac{1}{n}e,y}$ as in $\mathcal{E}_{\frac{1}{n}e,x}$. Actually it is possible to consider a larger class of decision problems and still conserve the validity of this proposition, but this will not be proved here (see Theorem 7.5 in referance [5]).

The next proposition vill also give a statistical description of majorization, and now by means of operating characteristics.

PROPOSITION II.2.5.

Let $x,y \in R^n$ and assume that $\sum_j x_j = \sum_j y_j$. Then the statements (II.2.7)-(II.2.9) are all equivalent

(II.2.7) $\quad x \prec y$

(II.2.8) $\quad \forall k \in N,\ \forall \rho \in \mathcal{M}_{n,k},\ \exists \delta \in \mathcal{M}_{n,k} : e\delta = e\rho \ \text{ and } \ y\delta = x\rho$

(II.2.9) $\quad \forall \rho \in [0,1]^n,\ \exists \delta \in [0,1]^n\colon \sum_j \delta_j = \sum_j \rho_j \ \text{ and } \ \sum_j y_j \delta_j = \sum_j x_j \rho_j.$

PROOF: The equivalence between (II.2.7) and (II.2.8) follows from Theorem I.2.3 (iii) by introducing matrix notation for decision rules.

The equivalence between (II.2.8) and (II.2.9) follows from Proposition I.2.1. This will be shown in detail later in a more general version (see Propostion III.2.8). □

(II.2.8) expresses that there to every finite decision space and every decision rule (represented by a Markow matrix ρ) in $\mathcal{E}_{\frac{1}{n}e,x}$, corresponds a decision rule (represented by the Markow matrix δ) in $\mathcal{E}_{\frac{1}{n}e,y}$ parrying the first one in the sense that the operating characteristics are equal.

(II.2.9) expresses the same idea, but for testing problems (2-decision problems) only.

It's interesting to note that (II.2.8) is "quite close to" the following well-known characterization of $x \prec y$

(II.2.10) $$\exists M \in \mathcal{M}^{D}_{n,n} : x = yM.$$

The equivalence between (II.2.8) and (II.2.10) can in fact be seen <u>directly</u> in an easy way:

The implication from (II.2.8) to (II.2.10) follows by choosing $k = n$ and $\rho = I$ (I: the $n \times n$ identity matrix). Then there is a $\delta \in \mathcal{M}_{n,n}$ such that

$$e\delta = e\rho = eI = e \quad \text{and} \quad y\delta = x\rho = xI = x,$$

which means that $\delta \in \mathcal{M}^{D}_{n,n}$ and $x = y\delta$, and (II.2.10) holds.

Conversely we prove the implication from (II.3.10) to (II.3.8) by, for given $k \in N$ and $\rho \in \mathcal{M}_{n,k}$, putting $\delta = M\rho$. Then it's easy to see that $\delta \in \mathcal{M}_{n,k}$ and that $e\delta = e\rho$. Furthermore

$$y\delta = y(M\rho) = (yM)\rho = x\rho$$

so (II.2.8) holds.

On the other hand it is harder to realize the implication from (II.2.9) to (II.2.8) directly (or alternativly the implication from (II.2.9) to (II.2.10)). This suggests that the reduction from ε-deficiency to ε-deficiency for 2-decision problems (which the implication from (II..8) to (II.2.9) represents) is not trivial.

II.3. An example showing the statistical content in the concept of majorization.

This section gives an example of a practical problem in which majorization occurs, and where the statistical interpretation of the concept is illustrated.

EXAMPLE II.3.1.

A statistician is confronted with the following problem: Two boxes are given; box 1 and box 2, each containing two dice, one red and one blue. We denote the red die in box 1 by R_1, and the blue die by B_1. Analogously R_2 and B_2 are the red and the blue die respectivly in box 2.

We have certain informations on the dice. All the dice have sides showing the numbers $1,2,\dots,6$, and in each box there is exactly <u>one</u> die, which is just. When we denote a die just, we mean that the probability of each of the six possible outcomes is $1/6$. Furthermore we know that those two dice that are just, are of the same colour. This implies that either R_1 and R_2 are just (while B_1 and B_2 are not) or B_1 and B_2 are just (while R_1 and R_2 are not). Besides we have some knowledge of those dice that are not just. The table below shows the probability of the different sides coming up in a throw with the non-just die from each box.

Side number	Box 1	Box 2
1	0.10	0.05
2	0.05	0.30
3	0.15	0.30
4	0.30	0.15
5	0.25	0.05
6	0.15	0.15

The statistician's task is to choose one of the boxes, and from certain experiments he is allowed to perform with the dice in this box, he should tell whether the red dice are just or not. Thus he faces the following problem: Which box should be chosen in order to have as much information as possible before answering the "colour-problem".

We have by this presented two different problems:

Problem 1 is the decision problem the statistician faces after he has chosen a box, nemaly to answer the question: are the red dice just?"

Problem 2 is whether we should choose box 1 or box 2 in order to solve problem 1 in the best possible way. It is this problem we are interested in her.

We now define problem 1 precisely, by giving the following information: After having chosen which box he will use, the statistician shall pick one of the dice in this box and throw this die 25 times. On the based of the 25 observed results he shall then answer this question: are the red dice just? He must give one of the answers "yes", "no" and "I don't know", and he then looses or wins a certain amount of money depending on the relation between his answer and the correct answer according to the next table:

Correct answer \ Given answer	"yes"	"no"	"I don't know"
"yes"	50	-20	-16
"no"	-20	50	-23

Positive numbers indicates profit and negative numbers indicates loss to the statistician. By this problem 1 is well defined, and we see that this is a decision problem.

As we have mentioned before problem 2 is our man interest, and we shall now show how this can be solved.

We introduce the following two majorization dichotomies:

$$\mathcal{E}_{\frac{1}{6}e, x_1} \quad \text{and} \quad \mathcal{E}_{\frac{1}{6}e, x_2},$$

where

$$x_1 = (0.10,\ 0.05,\ 0.15,\ 0.30,\ 0.25,\ 0.15)$$

and

$$x_2 = (0.05,\ 0.30,\ 0.30,\ 0.15,\ 0.05,\ 0.15)$$

The product experiment $(\mathcal{E}_{\frac{1}{6}e, x_i})^{25}$, $i = 1,2$, will then consist in throwing one die from box i 25 times and observe the result.

We now wish to find out which box to choose in order to have as much information as possible when we shall decide the colour of the just dice. Thus it is needed to compare the selection of box 1 to the selection of box 2 with respect to the solvation of problem

1. Therefore we compare the two experiments $(\mathcal{E}_{\frac{1}{6}e,x_1})^{25}$ and $(\mathcal{E}_{\frac{1}{6}e,x_2})^{25}$.

It turns out to be sufficient to compare $\mathcal{E}_{\frac{1}{6}e,x_1}$ and $\mathcal{E}_{\frac{1}{6}e,x_2}$, and this is quite simple. We see (for instance from Definition II.1.1) that

$$x_1 \prec x_2$$

according to Proposition II.2.1. This means that

$$\mathcal{E}_{\frac{1}{6}e,x_1} < \mathcal{E}_{\frac{1}{6}e,x_2}.$$

From the general theory of product experiments it follows that

$$(\mathcal{E}_{\frac{1}{6}e,x_1})^{25} < (\mathcal{E}_{\frac{1}{6}e,x_2})^{25}.$$

This implies that there to every decision problem (as long as the decision space is Borel-isomorph; see Theorem 7.5 in referance [5]), to every bounded loss function and to every decision rule in $(\mathcal{E}_{\frac{1}{6}e,x_2})^{25}$ having a risk function which is uniformly less than or equal to the risk function of the decision rule in $(\mathcal{E}_{\frac{1}{6}e,x_2})^{25}$. As a special case, this will hold for the decision problem that problem 1 represents.

In our example one should choose box 2. It is important to note that we arrive at the same conclusion whatever decision space and loss function we might consider. Besides one ought to choose box 2 whatever number of throws we are allowed to make with the die.

CHAPTER III. ε-MAJORIZATION

III.1. Definition.

We have earlier seen the following fundamental result:

If $x,y \in R^n$ and $\sum_j x_j = \sum_j y_j$, then

$$x \prec y \iff \mathcal{E}_{\frac{1}{n}e,x} < \mathcal{E}_{\frac{1}{n}e,y}.$$

Therefore x is majorized by y if and only if the majorization pseudo dichotomy determined by y is more informative than the majorization pseudo dichotomy determined by x. Since "more informative" is the same as "(0,0)-deficiency", it is natural to ask which relations between x and y that correspond to $\mathcal{E}_{\frac{1}{n}e,y}$ being ε-deficient with respect to $\mathcal{E}_{\frac{1}{n}e,x}$. In this chapter we shall consider this question in the case of (0,ε)-deficiency.

DEFINITION III.1.1.

Let $x,y \in R^n$ be such that $\sum_j x_j = \sum_j y_j$ and let $\varepsilon > 0$.

We then say that x is ε-majorized by y, and in that case we write $x \prec_\varepsilon y$, if $\mathcal{E}_{\frac{1}{n}e,y}$ is (0,ε)-deficient with respect to $\mathcal{E}_{\frac{1}{n}e,x}$.

We demand that $\sum_j x_j = \sum_j y_j$ in this definition because we wish $\Delta_1(\mathcal{E}_{\frac{1}{n}e,x}, \mathcal{E}_{\frac{1}{n}e,y}) = 0$ to hold, since this is needed to assure that ε-deficiency is equivalent to ε-deficiency for 2-decision problems. This gives source to several interesting characterizations of ε-majorization, and besides $\sum_j x_j = \sum_j y_j$ is a necessary condition of usual majorization.

We see that ε-majorization generalizes majorization, like the next proposition says.

PROPOSITION III.1.2.

Let $x,y \in R^n$ be such that $\sum_j x_j = \sum_j y_j$. Then

$$x \prec y \iff x \underset{0}{\prec} y.$$

PROOF: This is seen directly from Definition III.1.1 with $\varepsilon = 0$ because "(0,0)-deficiency is the same as "more informative". □

This implies that all the results we will get on ε-majorization for $\varepsilon > 0$, will give us results on majorization by simply letting $\varepsilon = 0$.

II. Characterizations.

The results in this section are all different characterizations of ε-majorization, and they are consequences of the general theory in section I.3 on pseudo dichotomies.

The first characterization we will present of ε-majorization connects the concept to inequalities between the partial sums that we know from the definition of majorization.

PROPOSITION III.2.1.

Let $x,y \in R^n$ and assume that $\sum_j x_j = \sum_j y_j$. Let $\varepsilon > 0$. Then the following holds:

$$x \underset{\varepsilon}{\prec} y$$

$$\Updownarrow$$

(III.2.1) $$\sum_{j=1}^{k} x_{[j]} < \sum_{j=1}^{k} y_{[j]} + \frac{\varepsilon}{2};\ k = 1,\ldots,n-1.$$

PROOF: According to Proposition I.3.12 and Definition III.1.1 we have

$$x \underset{\varepsilon}{\prec} y$$

$$\Updownarrow$$

$$\mathcal{E}_{\frac{1}{n}e,y} \text{ is } (0,\varepsilon)\text{-deficient with respect to } \mathcal{E}_{\frac{1}{n}e,x}$$

$$\Updownarrow$$

$$\beta_{\mathcal{E}_{\frac{1}{n}e,y}}(\alpha) > \beta_{\mathcal{E}_{\frac{1}{n}e,x}}(\alpha) - \frac{\varepsilon}{2},\ \forall\alpha\in[0,1].$$

From I.4 we know that

$$\beta_{\mathcal{E}_{\frac{1}{n}e,x}}\left(\frac{k}{n}\right) = \sum_{j=1}^{k} x_{[j]};\ k = 1,2,\ldots,n.$$

Furthermore $\beta_{\mathcal{E}_{\frac{1}{n}e,x}}$ is continous, and linear on the intervals $[\frac{k-1}{n},\frac{k}{n}]$, $k = 1,\ldots,n$. $\beta_{\mathcal{E}_{\frac{1}{n}e,y}}$ has got the same properties. The linearity and the continuity implies that (III.2.2) is equivalent to the same statement when we let α run through the set $\{\frac{k}{n}|k = 0,1,\ldots,n\}$, and this means that (III.2.1) and (III.2.2) are equivalent. □

The next result is a consequence of "the U-criterion for ε-deficiency".

PROPOSITION III.2.2.

Let $x,y \in R^n$ and $\varepsilon > 0$. Assume that $\sum_j x_j = \sum_j y_j$. Then we have

$$x \underset{\varepsilon}{\prec} y$$
$$\Updownarrow$$
$$\sum_j |x_j - \xi| \le \sum_j |y_j - \xi| + \varepsilon, \quad \forall \xi \in R.$$

PROOF: By applying Corollary I.3.4 and the expression for the U-function in section I.4, we get this equivalence immediately. □

COROLLARY III.2.3.

Let $x,y \in R^n$ and $\varepsilon > 0$. Assume that $\sum_j x_j = \sum_j y_j$. Then

$$x \underset{\varepsilon}{\prec} y$$
$$\Updownarrow$$
$$\sum_j (x_j - a)^+ \le \sum_j (y_j - a)^+ + \frac{\varepsilon}{2}, \quad \forall a \in R.$$

PROOF: We realize this by using that $|a| = 2^+ - a$ and $\sum_j x_j = \sum_j y_j$ in Proposition III.2.2. □

One of the most interesting results on majorization is:
If $x,y \in R^n$ and $\sum_j x_j = \sum_j y_j$, we have

$$x \prec y \iff x \in K_y.$$

Here K_y (see section I.1) denotes the convex hull of the set of all permutations of y. This characterization of majorization is closely connected to this result:

$$x \prec y \iff \text{there exists a doubly-stochastic matrix } M \text{ such that } x = yM.$$

This "nice" geometrical description is also available within ε-majorization. This can be shown by using "the Markov-kernel criterion" for ε-deficiency.

PROPOSITION III.2.4.

Let $x,y \in R^n$ and $\varepsilon > 0$. Assume that $\sum_j x_j = \sum_j y_j$. Then the following equivalence holds:

$$x \underset{\varepsilon}{\prec} y$$
$$\Updownarrow$$

(III.2.3) $$\exists M \in \mathcal{M}^D_{n,n} : \| x - yM \|_0 < \varepsilon .$$

PROOF: Let $\mathcal{E}_{\frac{1}{n}e,x} = (\{1,\ldots,n\}, \mathcal{P}(\{1,\ldots,n\}), \nu_1, \nu_2)$ and $\mathcal{E}_{\frac{1}{n}e,y} = (\{1,\ldots,n\}, \mathcal{P}(\{1,\ldots,n\}), \mu_1, \mu_2)$, where ν_i and μ_1 defined in the usual sense. The Markow-kernel criterion now gives us

$$x \underset{\varepsilon}{\prec} y$$
$$\Updownarrow$$

(III.2.4) there exists a Markow-kernel M such that

$$\| \mu_\theta M - \nu_\theta \| < \varepsilon_\theta, \; \theta = 1,2,$$

$\varepsilon_1 = 0$ and $\varepsilon_2 = \varepsilon$.

We reformulate (III.2.4) by considering the two inequalities for $\theta = 1,2$ separately.

When $\theta = 1$ (III.2.4) gives

$$\|\mu_1 M-\nu_1\| < 0$$
$$\Updownarrow$$
$$\mu_1 M = \nu_1$$
$$\Updownarrow$$
$$\sum_{i=1}^{n} M(\{j\}|i)\frac{1}{n} = \frac{1}{n};\ j = 1,\ldots,n$$
$$\Updownarrow$$
$$\sum_{i=1}^{n} M(\{j\}|i) = 1;\ j = 1,\ldots,n.$$

When we let the Markow-kernel M be represented by the Markow matrix $(m_{ij})_{i,j=1,1}^{n,n}$ defined by

$$m_{ij} = M(\{j\}|i);\ i,j\in\{1,\ldots,n\}.$$

Then (III.24) for $\theta = 1$ is equivalent to $(m_{ij})_{i,j=1,1}^{n,n}$ is doubly-stochastic.

Furthermore

$$\|\mu_2 M-\nu_2\| = \sum_{j=1}^{n} |(\mu_2 M)(\{j\})-\nu_2(\{j\})| = \sum_{j=1}^{n} |\sum_{i=1}^{n} m_{ij}y_i-x_j|.$$

Without danger of confusion, we now define $M = (m_{ij})_{i,j=1,1}^{n,n}$ and thus

$$\sum_{i=1}^{n} m_{ij}y_i = (yM)_j$$

so

$$\|\mu_2 M-\nu_2\| = \sum_{j=1}^{n} |(yM)_j-x_j| = \|x-yM\|_0.$$

By this we see that (III.2.3) and (III.2.4) are equivalent and the proof is completed. □

Proposition III.2.4 says that $x \prec_\varepsilon y$ if and only if x can be approximated within the $\|\cdot\|_0$ norm by the image of y under a doubly-stochastic transformation.

COROLLARY III.2.5.

Let $x,y \in R^n$ and $\varepsilon > 0$. Assume that $\sum_j x_j = \sum_j y_j$. Then we have

$$x \underset{\varepsilon}{\prec} y$$
$$\Updownarrow$$
$$d_0(x, K_y) \leq \varepsilon$$

PROOF: Assume first that $x \underset{\varepsilon}{\prec} y$. According to Proposition III.2.4 there must then exist a $M \in \mathcal{M}^D_{n,n}$ such that $\|x-yM\|_0 \leq \varepsilon$, so

$$d_0(x,K_y) = \inf_{z \in K_y} \|x-z\|_0 \leq \|x-yM\|_0 \leq \varepsilon$$

since $yM \in K_y$ (from Birkhoff's theorem we know that

(III.2.6) $K_y = \langle \{y\Pi \mid \Pi$ is a permutation-matrix on $\{1,\dots,n\}\} \rangle$
$= \{yM \mid M \in \mathcal{M}^D_{n,n}\}$.)

Conversely, assume that $d_0(x,K_y) \leq \varepsilon$. Since K_y is compact, there is a $z \in K_y$ such that

$$d_0(x,K_y) = \|x-z\|_0$$

and because $z \in K_y$, we must have that $z = yM$ for a suitable $M \in \mathcal{M}^D_{n,n}$ (see (III.2.6)). This shows that (III.2.5) holds. □

Another major result from the theory of majorization is:

Let $x,y \in R^n$ and assume that $\sum_j x_j = \sum_j y_j$. Then

$$x \prec y$$
$$\Updownarrow$$

(III.2.7) $\sum_j \phi(x_j) \leq \sum_j \phi(y_j)$ for every convex function $\phi: R \to R$.

This can also be generalized to include ε-majorization.

PROPOSITION III.2.6.

Let $x,y \in R^n$ and $\varepsilon > 0$. Assume that $\sum_j x_j = \sum_j y_j$. Then

$$x \underset{\varepsilon}{\prec} y$$
$$\Updownarrow$$
$$\sum_j \phi(x_j) \leq \sum_j \phi(y_j) + \frac{\varepsilon}{2}(\phi^-(\bar{y})-\phi^+(q))$$

for every convex function $\phi : R \to R$,

where $q = x_1 \wedge y_{(1)}$, $\bar{y} = x_{[1]} \vee y_{[1]}$ and where ϕ^+ and ϕ^- denotes rightsided and leftsided derivative respectivly.

PROOF: Assume that $x \underset{\varepsilon}{\prec} y$. According to Corollary III.2.3 the following will hold:

(III.2.9) $$\sum_j (y_j-a)^+ \geq \sum_j (x_j-a)^+ - \frac{\varepsilon}{2}, \quad \forall a \in R.$$

We shall now show start by showing that this implies that (III.2.8) holds for all convex functions that are a maximum of a finite number of linear functionals. Let

$$\phi(x) = \bigvee_{i=1}^{N} (a_i x+b_i); \quad x \in R.$$

It is then easy to show that

$$\phi(x) = a_1 x+b_1 + \sum_{i=1}^{N-1} (a_{i+1}x+b_{i+1}-a_i x-b_i)^+.$$

We may here assume that $a_1 < a_2 < \ldots < a_N$ (because the convexity implies that $a_1 \leq a_2 \leq \ldots \leq a_N$ and if $a_i = a_{i+1}$ we might as well eliminate the functional corresponding to $i+1$ in the maximum above.

Thus

$$\phi(x) = a_1 x+b_1 + \sum_{i=1}^{N-1} (a_{i+1}-a_i)(x + \frac{b_{i+1}-b_i}{a_{i+1}-a_i})^+$$

so

$$\sum_j \phi(y_j) = \sum_j [a_1 y_j + b_1 + \sum_{i=1}^{N-1} (a_{i+1}-a_i)(y_j + \frac{b_{i+1}-b_i}{a_{i+1}-a_i})^+$$

$$= a_1 \sum_j y_j + nb_1 + \sum_{i=1}^{N-1} (a_{i+1}-a_i) \sum_{j=1}^{n} (y_j + \frac{b_{i+1}-b_i}{a_{i+1}-a_i})^+$$

$$> a_1 \sum_j x_j + nb_1 + \sum_{i=1}^{N-1} (a_{i+1}-a_i) [\sum (x_j + \frac{b_{i+1}-b_i}{a_{i+1}-a_i})^+ - \frac{\varepsilon}{2}]$$

$$= \sum_j \phi(x_j) - \frac{\varepsilon}{2} \sum_{i=1}^{N-1} (a_{i+1}-a_i) = \sum_j \phi(x_j) - \frac{\varepsilon}{2}(a_N - a_1).$$

Furthermore it's clear that it is only the behaviour of ϕ on $[q,\bar{y}]$ that matters as regards our inequalities since $x_i, y_i \in [q,\bar{y}]$, $i = 1,\dots,n$. Consequently we can assume that the piecewise linear convex function above is such that

$$\phi(q) = a_1 q + b_1 \quad \text{and} \quad \phi(\bar{y}) = a_N \bar{y} + b_N$$

(because otherwise we elementate "the first and last" linear functionals so that this will hold!)

Let furthermore this choice (and this can be done generally) be such that there exist $\delta_1, \delta_2 > 0$ such that

$$x \in [q, q+\delta_1] \Rightarrow \phi(x) = a_1 x + b_1$$

and

$$x \in [\bar{y}-\delta_2, \bar{y}] \Rightarrow \phi(x) = a_N x + b_N.$$

This implies that

$$\phi^-(\bar{y}) = a_N \quad \text{and} \quad \phi^+(q) = a_1$$

and the inequality

$$\sum_j \phi(y_j) > \sum_j \phi(x_j) - \frac{\varepsilon}{2}(\phi^-(\bar{y}) - \phi^+(q))$$

therefore holds for all piecewise linear, convex functions ϕ.

We can now perform the final step, by approximating an arbitrary convex function ϕ with piecewise linear, convex functions. Let $\phi:R\to R$ be a convex function. We now define, for $m = 1,2,\ldots$, a function ϕ_m by

$$\phi_m(q+\frac{i}{2^m}(\bar{y}-q)) = \phi(q+\frac{i}{2^m}(\bar{y}-q)$$

for $i = 0,1,\ldots,2^m$, and where ϕ_m is linear on the intervals $[q+\frac{i-1}{2^m}(\bar{y}-q), q+\frac{i}{2^m}(\bar{y}-q)]$, $i = 0,1,\ldots,2^m$.

Then the following is clear:

i) Since ϕ_m is a maximum of a finite number of linear functionals, we have

$$\sum_j \phi_m(y_j) > \sum_j \phi_m(x_j) - \frac{\varepsilon}{2}(\phi_m^-(\bar{y})-\phi_m^+(q))$$

for $m = 1,2,\ldots$.

ii) $\phi_m(x)\downarrow\phi(x)$, $\forall x\in[q,\bar{y}]$ because ϕ is convex and ϕ_m is equal to ϕ at all the partition-points.

iii) $\phi_m^-(\bar{y})\uparrow\phi^-(\bar{y})$ and $\phi_m^+(y)\downarrow\phi^+(q)$.

By letting $m\to\infty$ in the inequality (III.2.10), we get the desired result (III.2.8).

Conversely, assume that (III.2.8) holds. It is now enough to show that the inequalities (III.2.9) hold, because according to Corollary III.2.3 this means that $x \underset{\varepsilon}{\prec} y$.

If $q = \bar{y}$, we must have $x_1 =\ldots= x_n = y_1 = \ldots = y_n$ and (III.2.9) holds trivially.

Assume therefore that $q < \bar{y}$, and we shall then show that (III.2.9) holds for every $a\in R$. We treat three different cases separately:

i) Let $a\in\langle q,\bar{y}\rangle$ and define $\phi(x) = (x-a)^+$. Then ϕ is convex and $\phi^+(q) = 0$, $\phi^-(\bar{y}) = 1$. From (III.2.8) we now get

$$\sum_j (y_j-a)^+ > \sum_j (x_j-a)^+ - \frac{\varepsilon}{2}$$

and (III.2.9) holds.

ii) Let $a \in \langle -\infty, q]$. Then $\sum_j (x_j-a)^+ = \sum_j (x_j-a)$ and $\sum_j (y_j-a)^+ = \sum_j (y_j-a)$, so because $\sum_j x_j = \sum_j y_j$, (III.2.9) will hold.

iii) Let $a \in [\bar{y}, \infty\rangle$. Then $\sum_j (x_j-a)^+ = \sum_j (y_j-a)^+ = 0$, and again (iii.2.9) holds.

The proof is then completed. □

We shall give some other characterizations of ε-majorization, and they all have in common that they describe the statistical content of the concept.

PROPOSITION III.2.7.

Let $x, y \in R^n$ and $\varepsilon > 0$. Assume that $\sum_j x_j = \sum_j y_j$. Then the following will hold:

$$x \underset{\varepsilon}{\prec} y$$

$$\Updownarrow$$

(III.2.11) $\forall k \in N, \forall \rho \in \mathcal{M}_{n,k}, \exists \delta \in \mathcal{M}_{n,k}$: $e\delta = e\rho$ and $\| y\delta - x\rho \|_0 < \varepsilon$.

PROOF: This follows directly from Theorem I.2.2 (iii) by introducing matrix notation for decision rules, like we did when proving Proposition III.2.4. □

This proposition can be given a statistical interpretation, at least when $\mathcal{E}_{\frac{1}{n}e,x}$ and $\mathcal{E}_{\frac{1}{n}e,y}$ are dichotomies (): when $x, y \in K_n$). In fact $x \underset{\varepsilon}{\prec} y$ if and only if operating characteristics in $\mathcal{E}_{\frac{1}{n}e,x}$ relative to a finite decision space can be approximated by operating characteristics in $\mathcal{E}_{\frac{1}{n}e,y}$ in the sense that (III.2.11) says. If we let $\mathcal{E}_{\frac{1}{n}e,y} = (\mathcal{Y},\mathcal{B},P_1,P_2)$ and $\mathcal{E}_{\frac{1}{n}e,x} = (\mathcal{X},\mathcal{A},Q_1,Q_2)$ be two dichotomies, we see that

$$e\delta = e\rho \iff (P_1\delta)(\{j\}) = (Q_1\rho)(\{j\});\ j = 1,\ldots,k$$

which means that the operating characteristics in $\mathcal{E}_{\frac{1}{n}e,y}$ and $\mathcal{E}_{\frac{1}{n}e,x}$ are equal when $\theta = 1$. Furthermore

$$\|y\delta - x\rho\| \leq \varepsilon \iff \|P_2\delta - Q_2\rho\| \leq \varepsilon$$

which means that the statistical distance between the operating characteristics when $\theta = 2$ is at most ε.

In the next proposition we have a similar statement, except that it says that is is enough to consider testing problems in order to conclude ε-majorization. This is due to the fundamental reduction result for ε-deficiency, Proposition I.2.1.

PROPOSITION III.2.8.

Let $x,y \in R^n$ and let $\varepsilon > 0$. Assume that $\sum_j x_j = \sum_j y_j$. Then

$$x \underset{\varepsilon}{\prec} y$$
$$\Updownarrow$$

(III.2.12) $\forall \rho \in [0,1]^n \exists \delta \in [0,1]^n: \sum_j \delta_j = \sum_j \rho_j$ and $|\sum_j y_j\delta_j - \sum_j x_j\rho_j| \leq \frac{\varepsilon}{2}$.

PROOF: This follows from Proposition I.2.1 and Proposition III.2.7 that $x \underset{\varepsilon}{\prec} y$ is equivalent to

(III.2.13) $\forall \rho \in \mathcal{M}_{n,2}, \exists \delta \in \mathcal{M}_{n,2}: e\delta = e\rho$ and $\|y\delta - x\rho\|_0 \leq \varepsilon$.

But in this case we have

$$\delta = \begin{pmatrix} \delta_1 & 1-\delta_1 \\ \vdots & \vdots \\ \delta_n & 1-\delta_n \end{pmatrix} \quad \text{and} \quad \rho = \begin{pmatrix} \rho_1 & 1-\rho_1 \\ \vdots & \vdots \\ \rho_n & 1-\rho_n \end{pmatrix}$$

so

$$e\delta = e\rho$$
$$\Updownarrow$$
$$(\sum_i \delta_i, n-\sum_i \delta_i) = (\sum_i \rho_i, n-\sum_i \rho_i)$$
$$\Updownarrow$$
$$\sum_i \delta_i = \sum_i \rho_i$$

Furthermore $y\delta = (\sum_i y_i\delta_i, \sum_i y_i - \sum_i y_i\delta_i)$ and $x\rho = (\sum_i x_i\rho_i, \sum_i x_i - \sum_i x_i\rho_i)$, and because $\sum_j x_j = \sum_j y_j$, we see

$$\|y\delta - x\rho\|_0 < \varepsilon$$
$$\Updownarrow$$
$$|\sum_i y_i\delta_i - \sum_i x_i\rho_i| + |\sum_i y_i\delta_i - \sum_i x_i\rho_i| < \varepsilon$$
$$\Updownarrow$$
$$|\sum_i y_i\delta_i - \sum_i x_i\rho_i| < \frac{\varepsilon}{2}.$$

Consequently (III.2.12) and (III.2.13) are equivalent and the proof is then completed. □

The next charaterization of ε-majorization if of special interest from a decision theoretical viewpoint. It gives a connection to risk sets in the different pseudo dichotomies.

PROPOSITION III.2.9.

Let $x,y \in R^n$ and $\varepsilon > 0$. Assume that $\sum_j x_j = \sum_j y_j$. Then the following equivalence holds:

$$x \underset{\varepsilon}{\prec} y$$
$$\Updownarrow$$
$$V_{\mathcal{E}_{\frac{1}{n}e,x}} \subset V_{\mathcal{E}_{\frac{1}{n}e,y}} + \{0\} \times [-\tfrac{\varepsilon}{2}, \tfrac{\varepsilon}{2}]$$

PROOF: This follows directly from Proposition I.3.10 by considering $(0,\varepsilon)$-deficiency. □

In section I.4 we have found the extreme points of $V_{\mathcal{E}_{\frac{1}{n}e,x}}$ and $V_{\mathcal{E}_{\frac{1}{n}e,y}}$, and the proposition therefore gives us a new geometrical idea of ε-majorization. By drawing the V-sets in the plane R^2, one sees immediatly that the V-criterion and the β-criterion are equivalent. This is caused by the graph of the β-function being "the upper boundary" of V, and that V is symmetrical about $(\frac{1}{2}, \frac{1}{2}\sum_j x_j)$.

We shall now show that Proposition III.2.9 also is interesting from a decision theoretical viewpoint. Assume that $\mathcal{E}_{\frac{1}{n}e,x} = (\mathcal{X}, \mathcal{A}, \mu_1, \mu_2)$ and $\mathcal{E}_{\frac{1}{n}e,y} = (\mathcal{Y}, \mathcal{B}, \nu_1, \nu_2)$ both are pseudo dichotomies satisfying (I.3.1) and (I.3.2). We consider the decision problem D that consists in estimating θ with "0-1 loss" : we let

$$\Theta = T = \{1,2\} \quad \text{and} \quad L_\theta(t) = \begin{cases} 1 & \text{when } \theta \neq t \\ 0 & \text{when } \theta = t. \end{cases}$$

We let $R^D_{\mathcal{E}_{\frac{1}{n}e,x}}$ and $R^D_{\mathcal{E}_{\frac{1}{n}e,y}}$ denote the risk set in respectively $\mathcal{E}_{\frac{1}{n}e,x}$ and $\mathcal{E}_{\frac{1}{n}e,y}$ relative to the decision problem D. Let now δ be a decision rule in $\mathcal{E}_{\frac{1}{n}e,y}$ and define $\delta(y) = \delta(\{2\}|y)$.

Then

$$r^D_{\mathcal{E}_{\frac{1}{n}e,y}}(1,\delta) = \int[\int L_1(t)\delta(dt|y)]\nu_1(dy) = \int\delta(y)\nu_1(dy) = \int\delta d\nu_1$$

and

$$r^D_{\mathcal{E}_{\frac{1}{n}e,y}}(2,\delta) = \int[\int L_2(t)\delta(dt|y)]\nu_2(dy) = \int(1-\delta(y))\nu_2(dy) = \sum_j y_j - \int\delta d\nu_2$$

Consequently

$$R^D_{\mathcal{E}_{\frac{1}{n}e,y}} = \{(\int\delta d\nu_1, \sum_j y_j - \int\delta d\nu_2)\,|\,\delta:\{1,\dots,n\} \to [0,1]\}$$

and analogously

$$R^D_{\mathcal{E}_{\frac{1}{n}e,x}} = \{(\int\rho d\mu_1, \sum_j x_j - \int\rho d\mu_2)\,|\,\rho:\{1,\dots,n\} \to [0,1]\}.$$

We now let $\alpha = \sum_j x_j = \sum_j y_j$ and define the transformation $g:R^2 \to R^2$ by

$$g(v_1,v_2) = (v_1,\alpha-v_2);\ \ (v_1,v_2)\in R^2.$$

It is then easy to show

(III.2.15) $R^D_{\mathcal{E}_{\frac{1}{n}e,x}} = g(V_{\mathcal{E}_{\frac{1}{n}e,x}})$ and $R^D_{\mathcal{E}_{\frac{1}{n}e,y}} + V_{\varepsilon_1,\varepsilon_2} = g(V_{\mathcal{E}_{\frac{1}{n}e,y}} + V_{\varepsilon_1,\varepsilon_2})$.

This leads to the following result:

PROPOSITION III.2.10.

Let $x,y \in R^n$ and $\varepsilon > 0$. Assume that $\sum_j x_j = \sum_j y_j$. Then

$$x \underset{\varepsilon}{\prec} y$$
$$\Updownarrow$$

(III.2.16) $$R^D_{\mathcal{E}_{\frac{1}{n}e,x}} \subset R^D_{\mathcal{E}_{\frac{1}{n}e,y}} + \{0\} \times [-\tfrac{\varepsilon}{2}, \tfrac{\varepsilon}{2}].$$

PROOF: It is enough to show that (III.2.14) and (III.2.16) are equivalent.

Assume that (III.2.14) holds. Then

$$g(V_{\mathcal{E}_{\frac{1}{n}e,x}}) \subset (V_{\mathcal{E}_{\frac{1}{n}e,y}} + V_{0,\varepsilon})$$

so $R^D_{\mathcal{E}_{\frac{1}{n}e,x}} \subset R^D_{\mathcal{E}_{\frac{1}{n}e,y}} + \{0\} \times [-\frac{\varepsilon}{2}, \frac{\varepsilon}{2}]$ because of (III.2.15). Hence (III.2.16) holds.

Conversely assume that (III.2.16) holds. Due to (III.2.15) we then have

$$g(V_{\mathcal{E}_{\frac{1}{n}e,x}}) \subset g(V_{\mathcal{E}_{\frac{1}{n}e,y}} + V_{0,\varepsilon}).$$

Since g is injective, this implies that

$$V_{\mathcal{E}_{\frac{1}{n}e,x}} = g^{-1}(g(V_{\mathcal{E}_{\frac{1}{n}e,x}})) \subset g^{-1}(g(V_{\mathcal{E}_{\frac{1}{n}e,y}} + V_{0,\varepsilon})) = V_{\mathcal{E}_{\frac{1}{n}e,y}} + V_{0,\varepsilon}$$

and (III.2.14) holds. □

Only when $\mathcal{E}_{\frac{1}{n}e,x}$ and $\mathcal{E}_{\frac{1}{n}e,y}$ are dichotomies the statistical content of Proposition III.2.10 is clear, and in that case (III.2.15) expresses a relation between "the usual risk sets" in

$\ell_{\frac{1}{n}e,x}$ and $\ell_{\frac{1}{n}e,y}$ relative to D. The proposition is of course valied for more general mass distributions.

The next characterization of ε-majorization we present is particularly interesting because it shows a connection to usual majorization.

When $y = (y_1,\ldots,y_n) \in R^n$ and $\varepsilon > 0$, we define

$$y_\varepsilon = (y_{[1]} + \tfrac{\varepsilon}{2}, y_{[2]}, \ldots, y_{[n-1]}, y_{[n]} - \tfrac{\varepsilon}{2}).$$

PROPOSITION III.2.11.

Let $x,y \in R^n$ and $\varepsilon > 0$. Assume that $\sum_j x_j = \sum_j y_j$. Then the following holds:

$$x \prec_\varepsilon y$$
$$\Updownarrow$$
$$x \prec y_\varepsilon .$$

PROOF: Since $\sum_j x_j = \sum_j y_j = \sum_j (y_\varepsilon)$; the following holds

$$x \prec y_\varepsilon$$
$$\Updownarrow$$
$$\sum_{j=1}^{k} x_{[j]} < \sum_{j=1}^{k} (y_\varepsilon)_{[j]}, \quad k = 1,\ldots,n-1$$
$$\Updownarrow$$
$$\sum_{j=1}^{k} x_{[j]} < \sum_{j=1}^{k} y_{[j]} + \frac{\varepsilon}{2}, \quad k = 1,\ldots,n-1$$
$$\Updownarrow$$
$$x \prec_\varepsilon y$$

where the last equivalence follows from Proposition III.2.1. □

This proposition gives a useful description of ε-majorization, because we now can use the known results from the theory of usual majorization to obtain results on ε-majorization.

By combining Proposition III.2.11 and Corollary III.2.5, we get an interesting geometrical property of the d_0-metric. First we present a useful lemma.

LEMMA III.2.12.

Let $y \in R^n$ and define, for $\varepsilon > 0$,

$$q_\varepsilon = (\tfrac{\varepsilon}{2}, 0, \ldots, 0, -\tfrac{\varepsilon}{2}) \in R^n .$$

Then

$$K_{q_\varepsilon} = \{ v \in R^n \mid \sum_j v_j = 0 \quad \text{and} \quad \| v \|_0 \leq \varepsilon \} .$$

PROOF: According to Theorem II.1.3 and Proposition III.2.11 we have

$$\begin{gathered} v \in K_{q_\varepsilon} \\ \Updownarrow \\ v \prec q_\varepsilon \\ \Updownarrow \\ v \underset{\varepsilon}{\prec} (0, \ldots, 0) \\ \Updownarrow \end{gathered}$$

(III.2.17) $$\sum_j v_j = 0 \quad \text{and} \quad \sum_j |v_j - a| \leq n|a| + \varepsilon, \ \forall a \in R$$

where the last equivalende is due to Proposition III.2.2.

But now (III.2.17) is equivalent to

(III.2.18) $$\sum_j v_j = 0 \quad \text{and} \quad \sum_j |v_j| \leq \varepsilon$$

which we see by applying the triangle inequality. Since $\| v \|_0 = \sum_j |v_j|$ the proof is completed. □

We are now able to prove the followin "nice" geometrical result:

PROPOSITION III.2.13.

Let $y \in R^n$ and $\varepsilon > 0$. We define $y_\varepsilon \in R^n$ by $y_\varepsilon = (\frac{\varepsilon}{2}, 0, \ldots, 0, -\frac{\varepsilon}{2})$. Then we have:

$$K_{y_\varepsilon} = K_y + K_{q_\varepsilon}.$$

PROOF: According to Proposition III.2.11 and Corollary III.2.5 we have

$$K_{y_\varepsilon} = \{x \mid x \prec y_\varepsilon\} = \{x \mid x \underset{\varepsilon}{\prec} y\} = \{x \mid \sum_j x_j = \sum_j y_j \text{ and } d_0(x, K_y) \leq \varepsilon\} = K_y + K_{q_\varepsilon}$$

where the last equality is shown in the following way:

Assume that $x \in R^n$ is such that $\sum_j x_j = \sum_j y_j$ and $d_0(x, K_y) \leq \varepsilon$. Then there is a $y' \in K_y$ such that $d_0(x, y') \leq \varepsilon$ (since $d_0(x, K_y) = \inf\{d_0(x, y') \mid y' \in K_y\}$ is obtained because K_y is compact and $y' \to d_0(x, y')$ is continous). Put $v = x - y'$. Then

$$\sum_j v_j = \sum_j x_j - \sum_j y'_j = \sum_j x_j - \sum_j y_j = 0$$

and $\|v\|_0 = \|x - y'\|_0 = d_0(x, y') \leq \varepsilon$, and by applying Lemma III.2.12 we know that $v \in K_{q_\varepsilon}$. Thus $x = y' + v \in K_y + K_{q_\varepsilon}$. This shows that

$$\{x \mid \sum_j x_j = \sum_j y_j \text{ and } d_0(x, K_y) \leq \varepsilon\} \subset K_y + K_{q_\varepsilon}.$$

Assume then that $x \in K_y + K_{q_\varepsilon}$. Then there exists a $y' \in K_y$ and $v \in K_{q_\varepsilon}$ such that $x = y' + v$. According to Lemma III.2.1 we then have

$$\sum_j x_j = \sum_j y'_j + \sum_j v_j = \sum_j y_j + 0 = \sum_j y_j.$$

Furthermore

$$d_0(x,K_y) \leq d_0(x,y') = \|x-y'\|_0 = \|y'+v-y'\|_0 = \|v\|_0 < \varepsilon$$

since $v \in K_{q_\varepsilon}$. This shows that

$$\{x \mid \sum_j x_j = \sum_j y_j \text{ and } d_0(x,K_y) \leq \varepsilon\} \supset K_y + K_{q_\varepsilon}$$

and the proof is completed. □

We shall complete this section by giving a simple example, which is intended to demonstrate some of the concepts of geometrical nature in this chapter.

EXAMPLE III.2.14.

Let $n = 5$ and $x = (\frac{2}{5},\frac{2}{5},\frac{1}{5},0,0)$, $y = (\frac{3}{5},\frac{1}{5},\frac{1}{10},\frac{1}{20},\frac{1}{20})$. Figure III.2.1 shows $\beta_{\ell_{\frac{1}{5}e,x}}$, $\beta_{\ell_{\frac{1}{5}e,y}}$, $V_{\ell_{\frac{1}{5}e,x}}$ and $V_{\ell_{\frac{1}{5}e,y}}$.

The first thing we notice is that neither x is majorized by y nor y is majorized by x. For instance we realize this fact by seeing that neighter $\beta_{\ell_{\frac{1}{5}e,x}} \leq \beta_{\ell_{\frac{1}{5}e,y}}$ nor $\beta_{\ell_{\frac{1}{5}e,y}} \leq \beta_{\ell_{\frac{1}{5}e,x}}$ holds. On the other hand $\beta_{\ell_{\frac{1}{5}e,y}}+0.1 \geq \beta_{\ell_{\frac{1}{5}e,x}}$ and this is also seen in Figure III.2.1. Thus, according to Proposition III.2.1, $x \underset{\varepsilon}{\prec} y$ for $\varepsilon = 0.2$. The figure also shows $V_{\ell_{\frac{1}{5}e,y}}+\{0\}\times[-\frac{\varepsilon}{2},\frac{\varepsilon}{2}]$, and we see that this set contains $V_{\ell_{\frac{1}{5}e,x}}$. This illustrates the close connection between the β-criterion (Proposition III.2.1) and the V-criterion (Proposition III.2.9) for ε-deficiency.

From a statistical viewpoint it is interesting that the power of the Neyman-Pearson test with size α (the power of the most powerful test for the hypothesis $\theta = 1$ against the alternative $\theta = 2$) is best in $\mathcal{E}_{\frac{1}{5}e,y}$ when $\alpha \in \langle 0, 0.4 \rangle$ and best in $\mathcal{E}_{\frac{1}{5}e,x}$ when $\alpha \in \langle 0.4, 1 \rangle$. If we for instance have a testing problem and want size 5%, it will be preferable to choose $\mathcal{E}_{\frac{1}{5}e,y}$. Then the strongest test will have power 0.15, while the strongest test in $\mathcal{E}_{\frac{1}{5}e,x}$, of the same size, has power 0.10. On the other hand, if the size is 40% (which is very uncommon!) $\mathcal{E}_{\frac{1}{5}e,x}$ is to be preferred.

Figure III.2.2 illustrates $R^D_{\mathcal{E}_{\frac{1}{5}e,x}}$ and $R^D_{\mathcal{E}_{\frac{1}{5}e,y}}$. These sets lie symmetrical to $V_{\mathcal{E}_{\frac{1}{5}e,x}}$ and $V_{\mathcal{E}_{\frac{1}{5}e,y}}$ respectivly with respect to the line $y = \frac{1}{2}$, and they are interesting from a statistical viewpoint. In fact it is easy to compare $\mathcal{E}_{\frac{1}{5}e,x}$ and $\mathcal{E}_{\frac{1}{5}e,y}$ as regards minimax- and Bayes-solutions in the decision problem D that consists in estimating θ with "0-1 loss" (see the description of D before Proposition III.2.10).

In I.7 in fererance [1] it is explained how to represent decision rules by their risk points and how to find minimax- and Bayes-solutions geometrically on the basis of $R^D_{\mathcal{E}_{\frac{1}{5}e,x}}$ and $R^D_{\mathcal{E}_{\frac{1}{5}e,y}}$. In our example we see that the minimax-risk in $\mathcal{E}_{\frac{1}{5}e,x}$ is approximately 0.34, while the minimax-risk in $\mathcal{E}_{\frac{1}{5}e,y}$ is 0.3. If one uses the minimax principle in this decision problem, $\mathcal{E}_{\frac{1}{5}e,y}$ is to be preferred. On the other hand it is important to realize that

Figure III.2.1

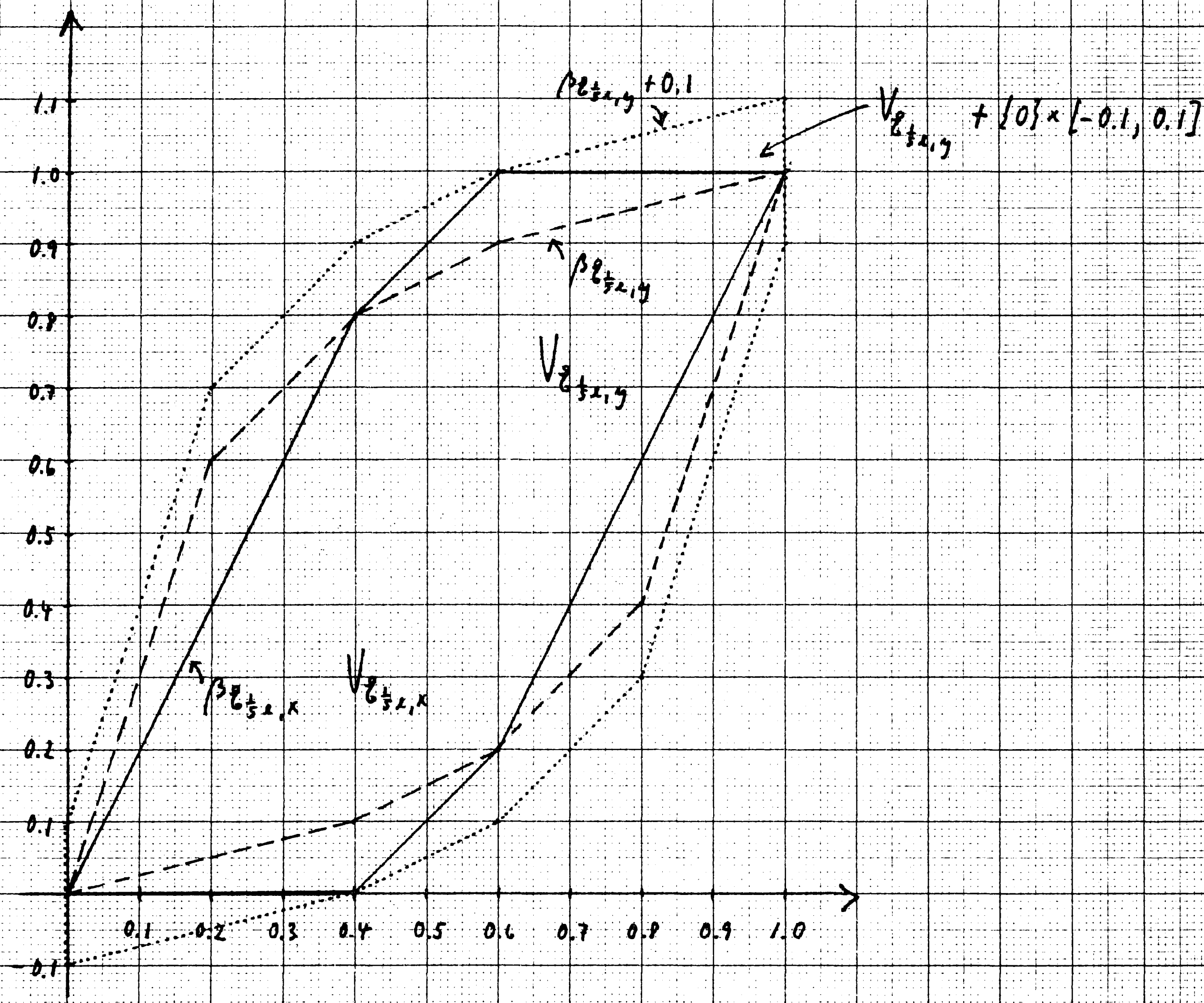
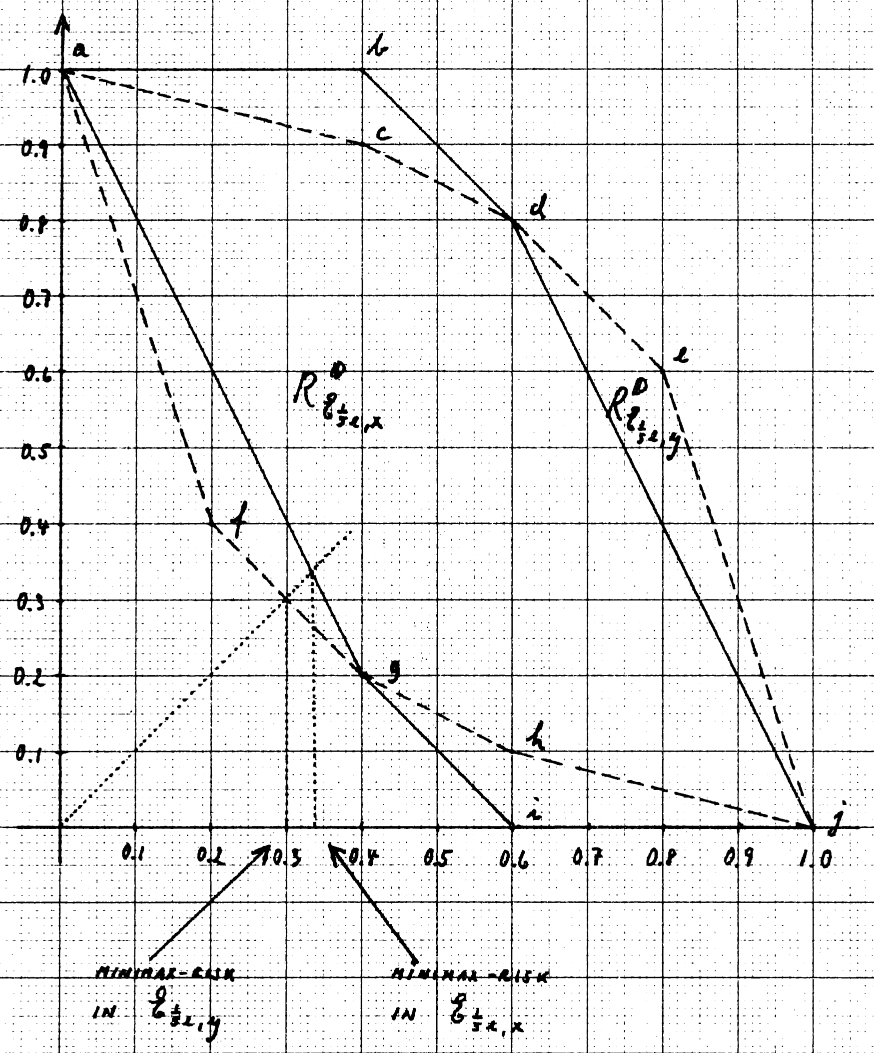


Figure III.2.2



there are other decision problems where $\ell_{\frac{1}{5}e,x}$ gives the smallest minimax-risk. This is the case because $\ell_{\frac{1}{5}e,x} < \ell_{\frac{1}{5}e,y}$ does not hold.

We can also decide which experiment we should prefer when we use the Bayes principle, but then the answer will depend on the a priori distributing on Θ. Once again we consider the decision problem D described above, and we also have an a priori distribution that gives masses $1-\lambda$ and λ to $\theta = 1$ and $\theta = 2$ respectively, where $\lambda \in [0,1]$. By representing such a distribution by the vector $(1-\lambda,\lambda)$, the points in a risk set having the same Bayes risk will lie on a straight line that is perpendicular to $(1-\lambda,\lambda)$. We find the minimum Bayes risk geometrically by considering the set of all such lines that have a non-empty intersection with the risk set, and then find the smallest 1. coordinate of points that lie on these lines and on the line $y = x$. (This is explained in detail in I.7 in reference [1].)

The nest table shows "the minimum Bayes point" in $R^D_{\ell_{\frac{1}{5}e,x}}$ and $R^D_{\ell_{\frac{1}{5}e,y}}$ respectivly as a function of λ.

λ	$\ell_{\frac{1}{5}e,x}$	$\ell_{\frac{1}{5}e,y}$
$\lambda \in [0,\frac{1}{4}\rangle$	a	a
$\lambda = \frac{1}{4}$	a	$\langle a,f \rangle$
$\lambda \in \langle \frac{1}{4},\frac{1}{3} \rangle$	a	f
$\lambda = \frac{1}{3}$	$\langle a,f \rangle$	f
$\in \langle \frac{1}{3},\frac{1}{2} \rangle$	g	f
$\lambda = \frac{1}{2}$	$\langle g,i \rangle$	$\langle f,g \rangle$
$\in \langle \frac{1}{2},\frac{2}{3} \rangle$	i	g
$\lambda = \frac{2}{3}$	i	$\langle g,h \rangle$
$\lambda \in \langle \frac{2}{3},\frac{4}{5} \rangle$	i	h
$\lambda = \frac{4}{5}$	i	$\langle h,j \rangle$
$\lambda \in \langle \frac{4}{5},1]$	i	j

From this table we see that in certain cases there is not just one "minimum Bayes point", but that a while line segment can have this property. For instance: when $\lambda = \frac{1}{4}$ all the points on the line segment $\langle a,f \rangle$ will be "minimum Bayes points" in $R^D_{\ell_{\frac{1}{5}e,y}}$.

On the basis of this table one can calculate minimum Bayes risk as a function of λ. In $\ell_{\frac{1}{5}e,x}$ and $\ell_{\frac{1}{5}e,y}$ we denote this variable by $B(\lambda|\ell_{\frac{1}{5}e,x})$ and $B(\lambda|\ell_{\frac{1}{5}e,y})$ respectivly. Then

$$B(|\ell_{\frac{1}{5}e,x}) = (1-\lambda)r^D_{\ell_{\frac{1}{5}e,x}}(1,\delta^*)+\lambda r^D_{\ell_{\frac{1}{5}e,x}}(2,\delta^*),$$

where δ^* is "the" Bayes-rule in $\ell_{\frac{1}{5}e,x}$ with respect to λ.

Since $(r^D_{\ell_{\frac{1}{5}e,x}}(1,\delta^*),r^D_{\ell_{\frac{1}{5}e,x}}(2,\delta^*))$ is the "minmum Bayes point" in

$\ell_{\frac{1}{5}e,x}$ relative to λ, it is easy to calculate $B(\lambda|\ell_{\frac{1}{5}e,x})$ on the basis of the table above. Analogously $B(\lambda|\ell_{\frac{1}{5}e,y})$ is calculated. We get

$$B(\lambda|\ell_{\frac{1}{5}e,x}) = \begin{cases} \lambda & \text{when } \lambda\in[0,\frac{1}{3}] \\ 0.4-0.2\lambda & \text{when } \lambda\in<\frac{1}{3},\frac{1}{2}] \\ 0.6-0.6\lambda & \text{when } \lambda\in<\frac{1}{2},1] \end{cases}$$

and

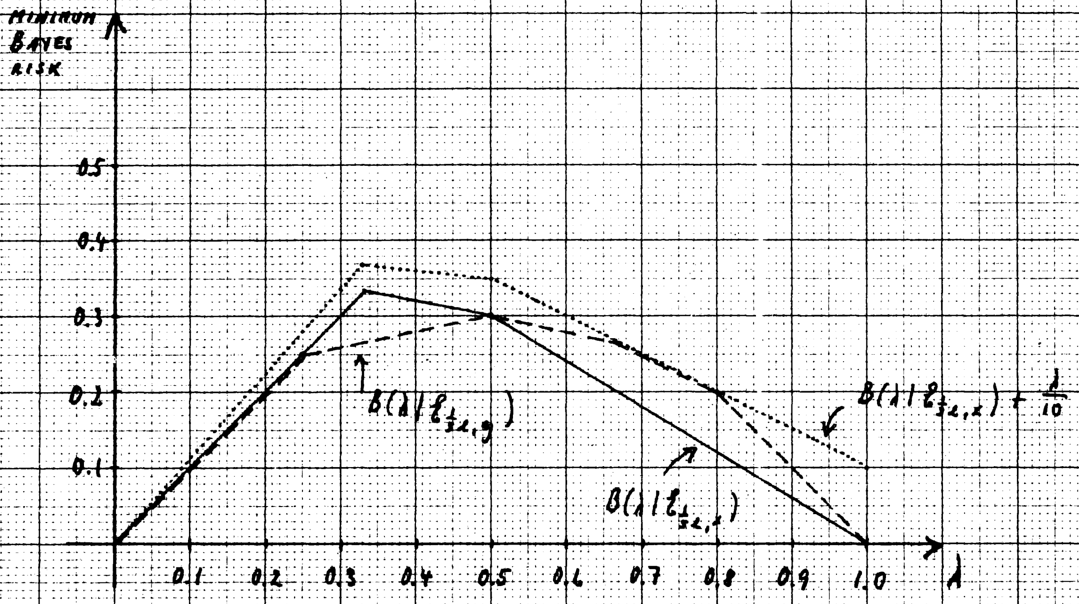
$$B(\lambda|\ell_{\frac{1}{5}e,y}) = \begin{cases} \lambda & \text{when } \lambda\in[0,\frac{1}{4}] \\ 0.2+0.2\lambda & \text{when } \lambda\in<\frac{1}{4},\frac{1}{2}] \\ 0.4-0.2\lambda & \text{when } \lambda\in<\frac{1}{2},\frac{2}{3}] \\ 0.6-0.5\lambda & \text{when } \lambda\in<\frac{2}{3},\frac{4}{5}] \\ 1-\lambda & \text{when } \lambda\in<\frac{4}{5},1] \end{cases}$$

In figure III.2.3 the graphs of $B(\lambda|\ell_{\frac{1}{5}e,x})$ and $B(\lambda|\ell_{\frac{1}{5}e,y})$ are drawn. We see that the minimum Bayes risk with respect to D is smallest in $\ell_{\frac{1}{5}e,y}$ when $\lambda\in<\frac{1}{4},\frac{1}{2}>$ and smallest in $\ell_{\frac{1}{5}e,x}$ when $\lambda\in<\frac{1}{2},1>$, and that they otherwise are equal. Thus: If one is interested in solving D by using the Bayes principle for a certain a priori distribution, the election between $\ell_{\frac{1}{5}e,x}$ and $\ell_{\frac{1}{5}e,y}$ can be done on the basis of these conclutions.

We can also illustrate a consequence of the fact that $x \underset{\varepsilon}{\prec} y$ for $\varepsilon = 0.2$. According to Proposition I.2.3 we know that

$$B(\lambda|\ell_{\frac{1}{5}e,y}) \leq B(\lambda|\ell_{\frac{1}{5}e,x}) + \sum_{\theta} \varepsilon_\theta \lambda_\theta \|L_\theta\| .$$

Figure III.2.3



It is possible to sharpen this inequality a little in our situation, because the loss function is non-negative. This is done in Theorem 6 in reference [6]. We then get

$$B(\lambda|\ell_{\frac{1}{5}e,y}) \leq B(\lambda|\ell_{\frac{1}{5}e,x}) + \tfrac{1}{2}\sum_{\theta} \varepsilon_\theta \lambda_\theta \|L_\theta\|$$

and thus

$$B(\lambda|\ell_{\frac{1}{5}e,y}) \leq B(\lambda|\ell_{\frac{1}{5}e,x}) + \frac{\lambda}{10}. \tag{III.2.19}$$

In figure III.2.3 the graph of $B(\lambda|\ell_{\frac{1}{5}e,x})+\frac{\lambda}{10}$ is also drawn, and we see that this figure confirms the inequality (III.2.19). On the basis of $x \underset{\varepsilon}{\prec} y$ we have therefore found an inequality that gives an upper bound for $B(\lambda|\ell_{\frac{1}{5}e,y})-B(\lambda|\ell_{\frac{1}{5}e,x})$, namely $\frac{\lambda}{10}$. We also see from Figure III.2.3 that this upper bound i attained when $\lambda\in[\frac{2}{3},\frac{4}{5}]$, while it otherwise is "too high" (except when $\lambda = 0$).

III.3. Product majorization.

In this short section we shall define a certain product between vectors, and show how ε-majorization is preserved under such products.

Let $x^{(1)}\in R^n$ and $x^{(2)}\in R^m$. We then define

$$x^{(1)}\otimes x^{(2)} = (x_1^{(1)}x_1^{(2)},\ldots,x_1^{(1)}x_m^{(2)},x_2^{(1)}x_1^{(2)},\ldots$$
$$\ldots,x_2^{(1)}x_m^{(2)},\ldots,x_n^{(1)}x_1^{(2)},\ldots,x_n^{(1)}x_m^{(2)}).$$

PROPOSITION III.3.1.

Let $x^{(1)},y^{(1)}\in K_n$ and $x^{(2)},y^{(2)}\in K_m$, and let $\varepsilon_1,\varepsilon_2>0$. Then the following holds:

$$x^{(i)} \prec_{\varepsilon_i} y^{(i)}, \ i = 1,2 \Rightarrow x^{(1)} \otimes x^{(2)} \prec_{\varepsilon_1+\varepsilon_2} y^{(1)} \otimes y^{(2)}.$$

PROOF: Since $x^{(1)} \prec_{\varepsilon_1} y^{(1)}$, $\mathcal{E}_{\frac{1}{n}e, y^{(1)}}$ is $(0,\varepsilon)$-deficient with respect to $\mathcal{E}_{\frac{1}{5}e, x^{(1)}}$, and because $x^{(2)} \prec_{\varepsilon_2} y^{(2)}$, $\mathcal{E}_{\frac{1}{n}e, y^{(2)}}$ is $(0,\varepsilon_2)$-deficient with respect to $\mathcal{E}_{\frac{1}{n}e, x^{(2)}}$. According to Proposition 5.18 in reference [5], $\mathcal{E}_{\frac{1}{n}e, y^{(1)}} \times \mathcal{E}_{\frac{1}{n}e, y^{(2)}}$ will then be $(0,\varepsilon_1+\varepsilon_2)$-deficient with respect to $\mathcal{E}_{\frac{1}{n}e, x^{(1)}} \times \mathcal{E}_{\frac{1}{n}e, x^{(2)}}$.

It is easy to realize from the definition of $\otimes$ above that

$$\mathcal{E}_{\frac{1}{n}e, y^{(1)}} \times \mathcal{E}_{\frac{1}{n}e, y^{(2)}} \sim \mathcal{E}_{\frac{1}{nm}e, y^{(1)} \otimes y^{(2)}}$$

and that

$$\mathcal{E}_{\frac{1}{n}e, x^{(1)}} \times \mathcal{E}_{\frac{1}{m}e, x^{(2)}} \sim \mathcal{E}_{\frac{1}{nm}e, x^{(1)} \otimes x^{(2)}}.$$

Thus $\mathcal{E}_{\frac{1}{nm}e, y^{(1)} \otimes y^{(2)}}$ will be $(0,\varepsilon_1+\varepsilon_2)$-deficient with respect to $\frac{1}{nm}e, x^{(1)} \otimes x^{(2)}$, and this shows that

$$x^{(1)} \otimes x^{(2)} \prec_{\varepsilon_1+\varepsilon_2} y^{(1)} \otimes y^{(2)}. \qquad \square$$

Besides we remark that this result can be generalized to an arbitrary, finite number of factors.

CHAPTER IV. DOT-DEFICIENCIES AS A MEASURE OF DISTANCE

IV.1. Definition and calculation of dot-deficiency.

We know that majorization is a pre-ordering on R^n. This means that $\prec$ has the following properties:

(IV.1.1) $\forall x \in R^n : x \prec x$

(IV.1.2) $\forall x,y,z \in R^n : x \prec y$ and $y \prec z \Rightarrow x \prec z$.

If we consider the restriction to $\mathcal{D}^n$, $\prec$ will be a partial ordering, so $\prec$ will in addition to (IV.1.1) and (IV.1.2) satisfy

(IV.1.3) $\forall x,y \in R^n : x \prec y$ and $y \prec x \Rightarrow x = y$.

On the other hand $\prec$ won't be a total ordering, even though we restrict ourselves to consider vectors in $\mathcal{D}^n$ lying in the same hyperplane $H_\alpha = \{x \in R^n \mid \sum_j x_j = \alpha\}$. In fact there exists $x,y \in H_\alpha \cap \mathcal{D}^n$ such that x is not majorized by y and y is not majorized by x. In that case we say that x and y are not comparable.

When we consider ε-majorization on the contrary, it is possible to compare arbitrary vectors in the same hyperplane H_α by using a suitable $\varepsilon > 0$. The following statements all hold for $\prec_\varepsilon$:

PROPOSITION IV.1.1.

Let α be an arbitrary real number. Then the following statements hold:

(IV.1.4) $\forall \varepsilon_1, \varepsilon_2 > 0, \forall x,y \in H_\alpha : \varepsilon_1 < \varepsilon_2$ and $x \prec_{\varepsilon_1} y \Rightarrow x \prec_{\varepsilon_2} y$

(IV.1.5) $\forall\varepsilon > 0, \forall x \in R^n : x \underset{\varepsilon}{\prec} x$

(IV.1.6) $\forall\varepsilon_1, \varepsilon_2 > 0, \forall x,y,z \in H_\alpha : x \underset{\varepsilon_1}{\prec} y$ and $y \underset{\varepsilon_2}{\prec} z \Rightarrow x \underset{\varepsilon_1+\varepsilon_2}{\prec} z$

(IV.1.7) $\forall x,y \in H_\alpha, \exists\varepsilon > 0 : x \underset{\varepsilon}{\prec} y$.

PROOF: (IV.1.4) follows from the following well-known result: $\mathcal{E}$ is ε-deficient with respect to $\mathcal{F}$ and $\eta > \varepsilon \Rightarrow$, $\mathcal{E}$ is η-deficient with respect to $\mathcal{F}$.

(IV.1.5) follows from (IV.1.4) and the fact that $\mathcal{E} > \mathcal{E}$.

(IV.1.6) follows from the following result: $\mathcal{E}$ is ε-deficient with respect to $\mathcal{F}$ and $\mathcal{F}$ is η-deficient with respect to $\mathcal{G}$ $\Rightarrow$ $\mathcal{E}$ is $(\varepsilon+\eta)$-deficient with respect to $\mathcal{G}$.

(IV.1.7) is seen from the β-criterion (Proposition III.2.1) by, for given $x,y \in H_\alpha$, choosing $\varepsilon = \sum_j |x_j|$. ␣

In this chapter property (IV.1.7) will be studied closer. This statement tells us that two arbitrary vectors x and y in the same hyperplane H_α can be campared by simply choosing ε big enough. It is therefore natural to wonder how big it is necessary to choose ε to make $x \underset{\varepsilon}{\prec} y$ hold; or equivalently: what is the smallest $\varepsilon > 0$ such that $x \underset{\varepsilon}{\prec} y$?

We find the answer to this question by considering the dot-deficiency $\dot{\delta}(\mathcal{E}_{\frac{1}{n}e,y}, \mathcal{E}_{\frac{1}{n}e,x})$ between $\mathcal{E}_{\frac{1}{n}e,y}$ and $\mathcal{E}_{\frac{1}{n}e,x}$. This quantity is defined as

$$\dot{\delta}(\mathcal{E}_{\frac{1}{n}e,y}, \mathcal{E}_{\frac{1}{n}e,x}) = \tfrac{1}{2}\inf\{\varepsilon > 0 \mid \mathcal{E}_{\frac{1}{n}e,y} \text{ is } (0,\varepsilon)\text{-deficient with respect to } \mathcal{E}_{\frac{1}{n}e,x}\}$$

We now introduce the following definition:

DEFINITION IV.1.2.

Let $x,y \in R^n$ and assume that $\sum_j x_j = \sum_j y_j$. We then define

$$\dot{\delta}(y,x) = \dot{\delta}(\mathcal{E}_{\frac{1}{n}e,y}, \mathcal{E}_{\frac{1}{n}e,x})$$

and this quantity is denoted by the dot-deficiency between y and x.

Furthermore we define

$$\dot{\Delta}(y,x) = \dot{\Delta}(\mathcal{E}_{\frac{1}{n}e,y}, \mathcal{E}_{\frac{1}{n}e,x}).$$

The existence of the dot-deficiency between y and x is assured by the fact that $x \underset{\varepsilon}{\prec} y$ for $\varepsilon = \sum_j |x_j|$, which implies that the infimum is taken over a non-empty set that has 0 as a lower bound.

Besides we note the following:

Let $x,y \in R^n$ and assume that $\sum_j x_j = \sum_j y_j$. Then

$$x \prec y \iff \dot{\delta}(y,x) = 0.$$

We shall now show a method of calculating $\dot{\delta}(x,y)$. Let $x = (x_1,\ldots,x_n) \in R^n$. We then define

$$\tilde{x} = (x_{[1]}, x_{[1]}+x_{[2]}, \ldots, x_{[1]}+\ldots+x_{[n]})$$

$$\text{i.e.:} \quad (\tilde{x})_k = \sum_{j=1}^{k} x_{[j]}; \; k = 1,\ldots,n.$$

This notation is used in our next proposition which gives a simple formula for the dot-deficiency between y and x.

PROPOSITION IV.1.3.

Let $x, y \in R^n$ and assume that $\sum_j x_j = \sum_j y_j$. Then the following equation holds:

$$\overset{\bullet}{\delta}(y,x) = (\hat{x}-\hat{y})_{[1]}.$$

PROOF: $\overset{\bullet}{\delta}(y,x) = \tfrac{1}{2}\inf\{\varepsilon>0 \mid x \underset{\varepsilon}{\prec} y\} =$

$$= \tfrac{1}{2}\inf\{\varepsilon>0 \mid \sum_{j=1}^{k} x_{[j]} < \sum_{j=1}^{k} y_{[j]} + \tfrac{\varepsilon}{2},\ k = 1,\dots,n-1\}$$

$$= \tfrac{1}{2}\inf\{\varepsilon>0 \mid 2(\sum_{j=1}^{k} x_{[j]} - \sum_{j=1}^{k} y_{[j]}) < \varepsilon,\ k = 1,\dots,n-1\}$$

$$= \tfrac{1}{2}(\sup_{k=1,\dots,n-1} 2(x_{[j]}-y_{[j]}) \vee 0) = (\hat{x}-\hat{y})_{[1]}$$

where the last equality is due to the fact that $(\hat{x}-\hat{y})_{[n]} = 0$, so that

$$\sup_{k=1,\dots,n-1} \sum_{j=1}^{k} (x_{[j]}-y_{[j]}) > 0.$$ □

By introducing the notation $|x| = (|x_1|,\dots,|x_n|)$ when $x = (x_1,\dots,x_n) \in R^n$ we also have the following result:

COROLLARY IV.1.4.

Let $x, y \in R^n$ and assume that $\sum_j x_j = \sum_j y_j$. Then we have

$$\overset{\bullet}{\Delta}(y,x) = (|\hat{x}-\hat{y}|)_{[1]}.$$

PROOF: This is seen from Proposition IV.1.3 because

$$\overset{\bullet}{\Delta}(y,x) = \overset{\bullet}{\Delta}(\ell_{\frac{1}{n}e,y}, \ell_{\frac{1}{n}e,x}) = \overset{\bullet}{\delta}(\ell_{\frac{1}{n}e,y}, \ell_{\frac{1}{n}e,x}) \vee$$

$$\overset{\bullet}{\delta}(\ell_{\frac{1}{n}e,x}, \ell_{\frac{1}{n}e,y}) = \overset{\bullet}{\delta}(y,x) \vee \overset{\bullet}{\delta}(x,y).$$ □

EXAMPLE IV.1.5.

This is just a simple example showing how to find $\dot{\delta}$ and $\dot{\Delta}$. Let $n = 4$ and

$$x = (2,6,2,9) \quad \text{and} \quad y = (4,8,7,0).$$

Then we have

$$\sum_j x_j = \sum_j y_j = 19$$

and

$$\hat{x} = (9,15,17,19) \quad \text{and} \quad \hat{y} = (8,15,19,19)$$

so we find

$$\hat{x}-\hat{y} = (1,0,-2,0)$$
$$\hat{y}-\hat{x} = (-1,0,2,0)$$
$$|\hat{x}-\hat{y}| = (1,0,2,0).$$

According to Proposition IV.1.3 and Corollary IV.1.4, we then get

$$\dot{\delta}(y,x) = (\hat{x}-\hat{y})_{[1]} = 1$$
$$\dot{\delta}(x,y) = (\hat{y}-\hat{x})_{[1]} = 2$$
$$\dot{\Delta}(y,x) = (|\hat{y}-\hat{x}|)_{[1]} = 2.$$

The dot-deficiency between y and x can also be given a geometrical interpretation, like the next proposition says.

PROPOSITION IV.1.6.

Let $x,y \in R^n$ be such that $\sum_j x_j = \sum_j y_j$. We then have

$$\dot{\delta}(y,x) = \tfrac{1}{2} d_0(x,K_y).$$

PROOF: $\dot{\Delta}(y,x) = \frac{1}{2}\inf\{\varepsilon>0 \,|\, \mathcal{E}_{\frac{1}{n}e,y}$ is $(0,\varepsilon)$-deficient with respect to $\mathcal{E}_{\frac{1}{n}e,x}\} = \frac{1}{2}\inf\{\varepsilon>0 \,|\, x \underset{\varepsilon}{\prec} y\} = \frac{1}{2}\inf\{\varepsilon>0 \,|\, d_0(x,K_y) \le \varepsilon\} = \frac{1}{2}d_0(x,K_y)$ according to Corollary III.2.5. □

IV.2. Dot-deficiency and inequalities.

We are sometimes interested in making inequalities of the type (III.2.8) for convex functions as sharp as possible. In such situations it can be useful to calculate the dot-deficiency first, and then apply the following result:

PROPOSITION IV.2.1.

Let $x,y \in R^n$ and assume that $\sum_j x_j = \sum_j y_j$. If $\phi : R \to R$ is a convex function, the following will hold

$$(IV.2.1) \qquad \sum_j \phi(x_j) \le \sum_j \phi(y_j) + \dot{\delta}(y,x)(\phi^-(\bar{y}) - \phi^+(q)),$$

where $\bar{y} = x_{[1]} \vee y_{[1]}$ and $q = x_{(1)} \wedge y_{(1)}$.

PROOF: We have

$$2\dot{\delta}(y,x) = \inf\{\varepsilon>0 \,|\, x \underset{\varepsilon}{\prec} y\}.$$

Put now $\varepsilon_0 = \inf\{\varepsilon>0 \,|\, x \underset{\varepsilon}{\prec} y\}$. First we shall show that $x \underset{\varepsilon_0}{\prec} y$. From the definition of ε_0 we see that there is a sequence $\{\varepsilon_n\}_{n=1}^{\infty}$ of positive, real numbers such that

$$\varepsilon_n \downarrow \varepsilon_0 \quad \text{and} \quad x \underset{\varepsilon_n}{\prec} y.$$

This means that $\ell_{\frac{1}{n}e,y}$ is $(0,\varepsilon_n)$-deficient with respect to $\ell_{\frac{1}{n}e,x}$ for $n = 1,2,\ldots$. According to Proposition I.2.4 we know that $\ell_{\frac{1}{n}e,y}$ is $(0,\varepsilon_0)$-deficient with respect to $\ell_{\frac{1}{n}e,x}$. Hence $x \prec_{\varepsilon_0} y$, and (IV.2.1) now follows from Proposition III.2.6. □

COROLLARY IV.2.2.

Let $x,y \in R^n$ and assume that $\sum_j x_j = \sum_j y_j$. If $\phi:[q,\bar{y}]$ is convex the following will hold:

(IV.2.2)
$$\sum_j \phi(x_j) - \mathring{\delta}(y,x)(\phi^-(\bar{y}) - \phi^+(q)) < \sum_j \phi(y_j) < \sum_j \phi(x_j) + \mathring{\delta}(x,y)(\phi^-(\bar{y}) - \phi^+(q)).$$

PROOF: This follows directly from Proposition IV.2.1 by applying this result twice and the put the inqualities together. □

EXAMPLE IV.2.3.

The entropy of a discrete probability distribution on at set with n elements and probabilities $p_1,\ldots,p_n$ respectivly, is defined as

$$H(p) = H(p_1,\ldots,p_n) = -\sum_{j=1}^{n} p_j \ell n p_j,$$

where $p = (p_1,\ldots,p_n) \in K_n$ and where define $p_j \ell n r_j = 0$ when $p_j = 0$.

We now define $\phi:[0,1] \to R$ by

$$\phi(x) = \begin{cases} x \ell n x, & \text{when } x \in \langle 0,1] \\ 0, & \text{when } x = 0 \end{cases}$$

This implies that

(IV.2.3)
$$H(p) = -\sum_j \phi(p_j)$$

and by two times derivation, we see that ϕ is convex (ϕ is also continous).

Let now p and q represent two probability distributions on a set with n elements): we let $p,q \in K_n$. We define

$$\bar{y} = p_{[1]} \vee q_{[1]} \quad \text{and} \quad q = p_{(1)} \wedge q_{(1)}$$

and with the conventions $\ell n \infty = \infty$ and $\ell n\, 0 = -\infty$, we get from (IV.2.2) that

$$\text{(IV.2.4)} \qquad \sum_j \phi(p_j) - \mathring{\delta}(q,p)\,\ell n \frac{\bar{y}}{q} \leq \sum_j \phi(q_j) \leq \sum_j \phi(p_j) + \mathring{\delta}(p,q)\,\ell n \frac{\bar{y}}{q}.$$

By multiplying (IV.2.3) by -1 and using (IV.2.3), we get

$$\text{(IV.2.5)} \qquad H(p) - \mathring{\delta}(p,q)\,\ell n \frac{\bar{y}}{q} \leq H(q) \leq H(p) + \mathring{\delta}(q,p)\,\ell n \frac{\bar{y}}{q}.$$

These inequalities give us an upper and a lower bound of the entropy in q, and these bounds are expressed by the entropy in p.

We also have that

$$|H(p)-H(q)| \leq \mathring{\delta}(q,p)\,\ell n \frac{\bar{y}}{q} \vee \mathring{\delta}(p,q)\,\ell n \frac{\bar{y}}{q}$$

$$= (\mathring{\delta}(q,p) \vee \mathring{\delta}(p,q)\,\ell n \frac{\bar{y}}{q} = \mathring{\Delta}(p,q)\,\ell n \frac{\bar{y}}{q}.$$

Thus we have shown that

$$\text{(IV.2.6)} \qquad |H(p)-H(q)| \leq \mathring{\Delta}(p,q)\,\ell n \frac{\bar{y}}{q}.$$

The inequality (IV.2.6) will also hold in a more general situation, like the next corollary says.

COROLLARY IV.2.4.

Let $x,y \in R^n$ and assume that $\sum_j x_j = \sum_j y_j$. Let further

$\phi:[q,\bar{y}]\to R$ be a convex function (where q and $\bar{y}$ are defined as before). Then the following inequality will hold:

$$\text{(IV.2.7)}\qquad \left|\sum_j \phi(y_j) - \sum_j \phi(x_j)\right| \leq \dot{\Delta}(x,y)(\phi^-(\bar{y})-\phi^+(q)).$$

PROOF: This is an immediate consequence of Corollary IV.2.2 by using the same approach as in Example IV.2.3. □

CHAPTER V. MULTI-DIMENSIONAL MAJORIZATION

V.1. Multi-dimensional maorization.

The concept of majorization can be extended to majorization between matrices. In this chapter we shall present such a concept, and also point out how it can be studied within the theory of comparison of pseudo experiments.

Let $M_{m,n}$ denote the set of all real $m \times m$ matrices. We then define a majorization-concept on $M_{m,n}$ in the following way (see page 430 in referance [3]):

DEFINITION V.1.1.

Let $X, Y \in M_{m,n}$. We then say that X is majorized by Y, and in that case we write $X \prec Y$, if there exists a dubly-stochastic $n \times n$ matrix M such that

$$X = YM.$$

We realize that this is a generalization of majorization between vectors, by simply choosing $m = 1$.

DEFINITION V.1.2.

Let $X \in M_{m,n}$. We then define the finite pseudo experiment $\mathcal{E}_X$ by

$$\mathcal{E}_X = (\mathcal{X}, \mathcal{A}, \mu_\theta : \theta \in \Theta),$$

where $\mathcal{X} = \{1,\ldots,n\}$, $\mathcal{A} = \mathcal{P}(\mathcal{X})$, $\Theta = \{1,\ldots,m+1\}$ and where μ_θ, $\theta \in \Theta$ are decided by the pseudo experiment matrix $P_{\mathcal{E}_X}$ defined by

$$P_{\ell_X} = \begin{pmatrix} \frac{1}{n}, \ldots, \frac{1}{n} \\ x_{11}, \ldots, x_{1n} \\ \vdots \qquad \vdots \\ x_{m1}, \ldots, x_{mn} \end{pmatrix},$$

where $X = (x_{ij})_{ij,=1,1}^{m,n}$.

With the aid of this definition it is now possible to find the connection between multi-dimensional majorization and the concept of "more informative".

PROPOSITION V.1.3.

Let $X,Y \in M_{m,n}$. Then the following equivalence holds:

$$X \prec Y \iff \ell_X < \ell_Y.$$

PROOF: According to Corollary I.2.6 we will have:

$$\ell_X < \ell_Y \iff \exists M \in \mathcal{M}_{n,n}:\ P_{\ell_X} = P_{\ell_Y} M.$$

But this again will be equivalent to the existance of $M \in \mathcal{M}_{n,n}^{D}$ such that $X = YM$. This is seen by writing out all the equations contained in the matrix equation $P_{\ell_X} = P_{\ell_Y} M$, and by noting that

$$(\tfrac{1}{n}, \ldots, \tfrac{1}{n}) = (\tfrac{1}{n}, \ldots, \tfrac{1}{n})M$$

if and only if M is doubly-stochastic. By using Definition V.1.1 the proof is then completed. □

Since multi-dimensional majorization now has been reduced to "more informative" between pseudo experiments, we can use this theory to give a couple of characterizations of $X \prec Y$.

PROPOSITION V.1.4.

Let $X,Y \in M_{m,n}$. Then the following equivalence hold:

$$X \prec Y$$
$$\Updownarrow$$
$$\sum_{j=1}^{n} \phi(\tfrac{1}{n}, x^{(j)}) < \sum_{j=1}^{n} \phi(\tfrac{1}{n}, y^{(j)}), \quad \forall \phi \in \Psi^{(n+1)},$$

where $x^{(j)}$ and $y^{(j)}$ denote the j-th coloum vector in X and Y respectivly.

PROOF: According to Theorem I.2.2 we have

$$\ell_X < \ell_Y \iff \forall \phi \in \Psi^{(m+1)}: \phi(\ell_X) < \phi(\ell_Y).$$

Therefore it is needed to calculate $\phi(\ell_X)$. Let μ be the counting measure on $\{1,\dots,n\}$, and put $f_i = d\mu_i | d\mu$, where

$$\mu_i(\{j\}) = \begin{cases} \frac{1}{n} & \text{when } i = 0 \\ x_{ij} & \text{when } i > 0, \end{cases}$$

$j = 1,\dots,n$. Then

$$f_i(j) = \begin{cases} \frac{1}{n} & \text{when } i = 0 \\ x_{ij} & \text{when } i > 0, \end{cases}$$

$j = 1,\dots,n$, and we have

$$\phi(\ell_X) = \int \phi(f_i: i \in \{0,1,\dots,m\}) d\mu = \sum_{j=1}^{n} \phi(f_i(j): i \in \{0,\dots,n\})$$

$$= \sum_{j=1}^{n} \phi(\tfrac{1}{n}, x_{1j},\dots,x_{mj}) = \sum_{j=1}^{n} \phi(\tfrac{1}{n}, x^{(j)}).$$

The proposition follows from this equality. □

COROLLARY V.1.5.

Let $X,Y \in M_{m,n}$, and let $\|\cdot\|$ denote an arbitrary norm on R^{m+1}. Then we have

$$X \prec Y \Rightarrow \sum_{j=1}^{n} \|(\tfrac{1}{n}, x_{1j},\dots,x_{mj})\| < \sum_{j=1}^{n} \|(\tfrac{1}{n}, y_{1j},\dots,y_{mj})\|.$$

PROOF: This follows from Proposition V.1.4 because every norm on R^{m+1} is a sublinear functional on R^{m+1}. □

Within the theory of comparison of pseudo experiments one speaks of "more informative for k-decision problems" (see I.1), and it is therefore also possible to introduce a corresponding concept within multi-dimensional majorization.

DEFINITION V.1.6.

Let $X,Y \in M_{m,n}$ and $k \in \{1,2,\dots\}$. We say that X is majorized by Y for k-decision problems, and in that case we write $X \underset{k}{\prec} Y$, if $\mathcal{E}_X \underset{k}{\leq} \mathcal{E}_Y$.

We now know from the general theory that the following will hold:

(V.1.1) $\quad X \underset{k+1}{\prec} Y \Rightarrow X \underset{k}{\prec} Y$

(V.1.2) $\quad X \prec Y \Rightarrow \forall k \in \{1,2,\dots\}: X \underset{k}{\prec} Y.$

When $k = 2$ the characterization in Proposition V.1.4 turns out to be of a more simple kind.

PROPOSITION V.1.7.

Let $X,Y \in M_{m,n}$ and assume that $\sum_j x_{ij} = \sum_j y_{ij}$, $i = 1,\dots,m$. Then the following equivalences will hold

(I.1.3) $$X \underset{2}{\leq} Y$$

$$\Updownarrow$$

(I.1.4) $$\sum_{j=1}^{n} |a_0 + \sum_{i=1}^{m} a_i x_{ij}| \leq \sum_{j=1}^{n} |a_0 + \sum_{i=1}^{m} a_i y_{ij}|, \quad \forall (a_0,\dots,a_m) \in R^{m+1}$$

$$\Updownarrow$$

(V.1.5) $$\sum_{j=1}^{n} (a_0 + \sum_{i=1}^{m} a_i x_{ij})^+ \leq \sum_{j=1}^{n} (a_0 + \sum_{i=1}^{m} a_i y_{ij})^+, \quad \forall (a_0,\dots,a_m) \in R^{m+1}.$$

PROOF: This follows from Proposition V.1.4 by reducing a maxmum of two linear functionals to a simple type and then use that $\sum_j x_{ij} = \sum_j y_{ij}$. This principle is the basis of Corollary B.2.3 in referance [4], which says:

Assume $\Delta_1(\ell_X, \ell_Y) = 0$. Then

$$\ell_X < \ell_Y \iff \|\sum_i a_i \mu_i\| < \|\sum_i a_i \nu_i\|, \quad \forall a \in R^{m+1},$$

where $\ell_X = (\mu_i : i \in \{0,\ldots,m\})$ and $\ell_Y = (\nu_i : i \in \{0,\ldots,m\})$.

But now

$$\Delta_1(\ell_X, \ell_Y) = 0$$
$$\Updownarrow$$
$$\mu_i(\{1,\ldots,n\}) = \nu_i(\{1,\ldots,n\}), \quad i = 0,\ldots,m$$
$$\Updownarrow$$
$$\sum_j x_{ij} = \sum_j y_{ij}, \quad i = 1,\ldots,m$$

and furthermore

$$\|\sum_i a_i \mu_i\| = \sum_j |\sum_i a_i \mu_i(\{j\})| = \sum_j |\frac{a_0}{n} + \sum_i a_i x_{ij}|$$

The equivalence between (V.1.3) and (V.1.4) then follows y replacing a_0 b $a_0 n$. The equivalence between (V.1.4) and (V.1.5) is simple and follows from $\sum_j x_{ij} = \sum_j y_{ij}$, $i = 1,\ldots,m$, by using the equation $|b| = 2b^+ - b$. □

COROLLARY V.1.8.

Let $X, Y \in M_{m,n}$, and assume that $\sum_j x_{ij} = \sum_j y_{ij}$, $\forall i$. Let further $k \in \{2,3,\ldots\}$. Then we have

(V.1.6) $$X \prec Y$$

$$\Downarrow$$

(V.1.7) $$X \underset{k}{\prec} Y$$

$$\Downarrow$$

(V.1.8) $$\sum_j (a_0+\sum_i a_i x_{ij})^+ \leq \sum_j (a_0+\sum_i a_i y_{ij})^+, \ \forall (a_0,\ldots,a_m)\in R^{m+1}.$$

PROOF: The first implication is seen from (V.1.2), and the second from Proposition V.1.7. □

PROPOSITION V.1.9.

Let $X,Y\in M_{m,n}$. Then we have:

$$X \underset{k}{\prec} Y$$

$$\Updownarrow$$

$$\forall \phi\in\Psi_k^{(m+1)}: \sum_j \phi(\tfrac{1}{n}, x^{(j)}) \leq \sum_j \phi(\tfrac{1}{n}, y^{(j)}).$$

PROOF: This follows from Theorem I.2.2. □

The next proposition characterizes $\prec$ and $\underset{k}{\prec}$ by means of relations between the operating characteristics.

PROPOSITION V.1.10.

Let $X,Y\in M_{m,n}$. Then we have

(V.1.9) $$X \prec Y$$

$$\Updownarrow$$

(V.1.10) $$\forall k\in\{1,2,\ldots\}, \forall \rho\in\mathcal{M}_{n,k}, \exists \delta\in\mathcal{M}_{n,k}: \sum_j \rho_{jt} = \sum_j \delta_{jt}, \forall t$$

and

$$X\rho = Y\delta.$$

In addition the following equivalence holds:

(V.1.11) $$X \underset{k}{<} Y$$

$$\Updownarrow$$

(V.1.12) $$\forall \rho \in \mathcal{M}_{n,k}, \exists \delta \in \mathcal{M}_{n,k}: \sum_j \rho_{jt} = \sum_j \delta_{jt}, \ \forall t$$

and

$$X\rho = Y\delta.$$

PROOF: Let $\mathcal{E}_X = (\mu_i : i \in \{0,\ldots,m\})$ and $\mathcal{E}_Y = (\nu_i : i \in \{0,\ldots,m\})$. According to Theorem I.2.2 $\mathcal{E}_X < \mathcal{E}_Y$ if and only if the following holds:

To every $k \in \{1,2,\ldots\}$, and to every randomization ρ from $\{1,\ldots,n\}$ to $\{1,\ldots,k\}$ there exists a randomization δ from $\{1,\ldots,n\}$ to $\{1,\ldots,k\}$ such that

$$\mu_i \rho = \nu_i \delta, \ \forall i.$$

But a randomization ρ from $\{1,\ldots,n\}$ to $\{1,\ldots,k\}$ can be represented by a Markow-matrix ρ, where $\rho \in \mathcal{M}_{n,k}$. Furthermore we have

$$(P_{\mathcal{E}_X}\rho)_{it} = \langle (P_{\mathcal{E}_X})_{(i)}, \rho^{(t)} \rangle = \sum_j \mu_i(\{j\})\rho_{jt}$$

$$= \int \rho(\{t\}|x)\mu_i(dx) = (\mu_i\rho)(\{t\}); \ i \in \{0,\ldots,m\}, \ t \in \{1,\ldots,k\}.$$

Thus we see that

$$\mu_i \rho = \nu_i \delta, \ \forall i$$

$$\Updownarrow$$

$$P_{\mathcal{E}_X}\rho = P_{\mathcal{E}_Y}\delta$$

$$\Updownarrow$$

$$\sum_j \frac{1}{n}\rho_{jt} = \sum_j \frac{1}{n}\delta_{jt}, \ \forall t \quad \text{and} \quad X\rho = Y\delta$$

and the equivalence between (V.1.9) and (V.1.10) has been shown.

The equivalence between (V.1.11) and (V.1.12) can be shown analogously. □

With the aid of Proposition V.1.10, we can say even more about the relation between $\prec$ and $\underset{k}{\prec}$ than (V.1.2) tells us.

PROPOSITION V.1.11.

Let $X,Y \in M_{m,n}$. Then we have

$X \prec Y \iff X \underset{n}{\prec} Y.$

PROOF: The implication $X \prec Y \Rightarrow X \underset{n}{\prec} Y$ is trivial (see. V.1.2).

We shall now show the converse implication, and let us therefore assume that $X \underset{n}{\prec} Y$. From Proposition V.1.10 we then know:

$$\forall \rho \in \mathcal{M}_{n,n}, \exists \delta \in \mathcal{M}_{n,n}: \sum_j \rho_{jt} = \sum_j {}_{jt}, \ \forall t$$

and

$$X\rho = Y\delta.$$

Let now I_n denote the $n \times n$ identity matrix and choose $\rho = I_n$. Then there exists a $\delta \in \mathcal{M}_{n,n}$ such that

$$\sum_j \delta_{jt} = \sum_j \rho_{jt} = 1, \ t = 1,\dots,n$$

and

$$Y\delta = X\rho = XI_n = X.$$

This means that there is a doubly-stochastic $n \times n$ matrix δ such that $X = Y\delta$, and according to Definition V.1.1 $X \prec Y$ must hold. □

COROLLARY V.1.12.

Let $X,Y \in M_{n,2}$ and assume that

$$\sum_{j=1}^{2} x_{ij} = \sum_{j=1}^{2} y_{ij}, \ i = 1,..,m.$$

Then the following equivalence will hold:

$$X \prec Y$$

$$\Updownarrow$$

$$\sum_{j=1}^{2}(a_0+\sum_{i=1}^{m}a_i x_{ij})^+ \leq \sum_{j=1}^{2}(a_0+\sum_{i=1}^{m}a_i y_{ij})^+, \ \forall(a_0,\ldots,a_m)\in R^{m+1}.$$

PROOF: This follows easily by combining Proposition V.1.11 and Proposition V.1.7. □

REFERENCES

[1] Ferguson, T.S. (1967): "Mathematical Statistics: A decision theoretic approach". Academic Press, New York.

[2] Lehmann, E.L. (1959): "Testing statistical hypothesis". Wiley, New York.

[3] Marshall, A. W. and Olkin, I. (1979): "Inequalities: Theory of majorization and its applications". Academic Press, New York.

[4] Torgersen, E. N. (1972): "Local comparison of experiments when the parameter set is one dimensional". Stat. Research Report, University of Oslo.

[5] Torgersen, E. N. and Lindquist, B. (1975): "Notes on comparison of experiments". Stat. Memoirs, University of Oslo.

[6] Torgersen, E.N. (1976): "Comparison of statistical experiments". Scand. Journ. of Statistics, 3.