Metrika, Volume 30, 1983, page 55–61.

Solution of a Statistical Optimization Problem by Rearrangement Methods

By *L. Rüschendorf*, Freiburg[1])

Summary: Inequalities for the rearrangement of functions are applied to obtain a solution of a statistical optimization problem. This optimization problem arises in situations where one wants to describe the influence of stochastic dependence on a statistical problem.

1. Introduction

Let $P_1, \ldots, P_n$ be n elements of $\mathcal{M}^1(R^1, \mathcal{B}^1)$ – the set of all probability measures on $(R^1, \mathcal{B}^1)$ – and define $\mathcal{M}(P_1, \ldots, P_n)$ to be the set of all probability distributions on $(R^n, \mathcal{B}^n)$ with P_i as i-th marginal distribution, $1 \leqslant i \leqslant n$. For measurable functions $\varphi: R^n \to R^1$ define

$$m := \inf \{\int \varphi \, dP: P \in \mathcal{M}(P_1, \ldots, P_n)\} \tag{1}$$

assuming the integrals in (1) to exist.

$\mathcal{M}(P_1, \ldots, P_n)$ is convex, tight and closed and, therefore, by Prohorov's theorem compact w.r.t. weak-topology. To prove tightness, let $K_i \in R^1$, $1 \leqslant i \leqslant n$, be compact sets with $P_i(K_i) \geqslant 1 - \epsilon/n$, $1 \leqslant i \leqslant n$, then $K = K_1 \times \ldots \times K_n$ is compact and using Fréchet's lower bounds $P(K) \geqslant \sum_{i=1}^{n} P_i(K_i) - (n-1) \geqslant 1 - \epsilon$ for all $P \in \mathcal{M}(P_1, \ldots, P_n)$. Therefore, the set on the right hand side of (1) is an interval and there exists a solution P^* of (1) if φ is bounded and continuous. (For bounded measurable φ there exists in general only a solution in the set of normed additive set functions with marginals P_i, $1 \leqslant i \leqslant n$.)

Problem (1) arises in situations where one wants to describe the influence of dependence on a statistical problem. Some typical examples are the following ones:

[1]) Dr. *L. Rüschendorf*, Institut für Mathematische Stochastik der Albert-Ludwigs-Universität, Hebelstr. 27, D–7800 Freiburg im Breisgau.

0026-1335/83/010055-61$2.50

1) Let f: $[0,1] \to R^1$ be measurable and assume that $a := \int f d\lambda^1$ exists. A rough estimator for a is $T_n := (1/n) \sum_{i=1}^{n} f(U_i)$, where $(U_i)_{1 \leqslant i \leqslant n}$ are i.i.d. random variables with $P^{U_1} = R(0,1)$ – the uniform distribution on $[0,1]$. The problem of determining a best unbiased estimator (minimum variance) for a of the type $T'_n = (1/n) \sum_{i=1}^{n} f(V_i)$, where $P^{V_i} = R(0,1)$, $1 \leqslant i \leqslant n$, is a problem of type (1) with $\varphi(x_1, \ldots, x_n) = (\sum_{i=1}^{n} f(x_i))^2$ and $P_i = R(0,1)$, $1 \leqslant i \leqslant n$. (In simulation studies the above estimator is combined with approximation techniques and applied usually in higher dimensions).
2) Let φ be a test for the simple hypotheses $H = \{P_0^{(n)}\}$ and $K = \{P_1^{(n)}\}$ with $P_0, P_1 \in M^1(R^1, B^1)$. Then

$$\sup \{\int \varphi \, dP; P \in M(P_0, \ldots, P_0)\}$$

is the worst level attained by this test if one is not sure that the underlying observations are independent.
3) Let $X = (X_1, \ldots, X_n)$ be a n-dimensional random variable with $X_i \geqslant 0$ and $P^{X_i} = P_i$, $1 \leqslant i \leqslant n$. Then with $\varphi(x_1, \ldots, x_n) = \prod_{i=1}^{n} x_i$, $\sup \{\int \varphi \, dP; P \in M(P_1, \ldots, P_n)\}$ gives an upper bound for $E \prod_{i=1}^{n} X_i$ which is better than the bounds $\prod_{i=1}^{n} \| X_i \|_{d_i}$, $d_i > 0$, $1 \leqslant i \leqslant n$, $\sum_{i=1}^{n} (1/d_i) = 1$ obtained by Hölders-inequality.
Similarly, for φ: $R^n \to R^1$ convex, $\inf \{\int \varphi \, dP; P \in M(P_1, \ldots, P_n)\}$ gives a better lower bound to $E\varphi(X)$ than $\varphi(EX)$ – the bound given by Jensen's inequality. Indeed,

$$\varphi(EX) = \inf \{\int \varphi \, dP; EX_i \text{ is the first moment of the } i\text{-th marginal distribution of } P\}.$$

A solution of (1) by means of duality-theory was given by *Gaffke/Rüschendorf* [1980], *Rüschendorf* [1980]. The present paper relates (1) to the problem of rearrangement of functions.

2. Rearrangement of Functions

Let f, g: $[0,1] \to \bar{R}^1$ be measurable functions. Then g is called a rearrangement of f if $\lambda \{g \geqslant c\} = \lambda \{f \geqslant c\}$, $\forall \, c \in R^1$; in other words, g and f have the same distribution function under λ – the restriction of Lebesgue-measure to $[0,1]$.

Rearrangements of functions were introduced by *Hardy/Littlewood/Polya* [1952]. They have important applications in many parts of analysis and were studied intensively by *Luxemburg* [1967], *Chong/Rice* [1971] and *Day* [1972]. To f there exist especially nondecreasing and nonincreasing rearrangements f^*, f_* and in generalization to the discrete case for $f, g \in L^1(\lambda)$

$$\int f^* g_* \, d\lambda \leqslant \int fg d\lambda \leqslant \int f^* g^* d\lambda. \tag{2}$$

Furthermore, $\int f^* g_* \, d\lambda = \int f_* g^* d\lambda$ and $\int f^* g^* \, d\lambda = \int f_* g_* \, d\lambda$.

(1) is related to rearrangements by means of the following lemma. Let F_i be the distribution function of P_i, $1 \leqslant i \leqslant n$, and let $F_i^{-1}(x) = \sup \{y \in R^1; F_i(y) \leqslant x\}$, $x \in [0,1]$, be the generalized inverse of F_i, $1 \leqslant i \leqslant n$.

Lemma 1: Let U be a random variable on $(\mathcal{M}, \mathcal{A}, P)$ with $P^U = R(0,1)$. Then

$$\mathcal{M}(P_1, \ldots, P_n) = \{P^{(f_1(U), \ldots, f_n(U))};\ f_i \text{ is a rearrangement of } F_i^{-1},\ 1 \leqslant i \leqslant n\}. \tag{3}$$

Proof: If f_i is a rearrangement of F_i^{-1}, then

$$P^{f_i(U)} = \lambda^{f_i} = \lambda^{F_i^{-1}} = P_i,\ 1 \leqslant i \leqslant n.$$

So the right hand side of (3) is contained in the left hand side.

A theorem of *Rohlin* [1952] [cf. also *Parthasaraty; Whitt*, Lemma 2.7] on the isomorphism of measure spaces implies that each $Q \in \mathcal{M}^1(R^n, \mathcal{B}^n)$ has a representation $Q = \lambda^{(f_1, \ldots, f_n)}$, where f_i: $[0,1] \to R^1$ are measurable. For $Q \in \mathcal{M}(P_1, \ldots, P_n)$ $\lambda^{f_i} = P_i = \lambda^{F_i^{-1}}$, $1 \leqslant i \leqslant n$, which implies that f_i is a rearrangement of F_i^{-1}, $1 \leqslant i \leqslant n$. □

As consequence we obtain:

Theorem 2.

$$m = \inf \{\int \varphi(f_1(t), \ldots, f_n(t))\, d\lambda(t); f_i \text{ is a rearrangement of } F_i^{-1},\ 1 \leqslant i \leqslant n\}. \tag{4}$$

□

Rearrangement-inequalities are closely connected with a generalization of Schur-order in the discrete case to measurable functions f, g: $[0,1] \to R^1$. For $f, g \in L^1(\lambda)$ one defines Schur-order by

$$\begin{aligned} & f \leqslant g \text{ if } \int_0^x f^*(t)\, d\lambda(t) \leqslant \int_0^x g^*(t)\, d\lambda(t), \qquad \forall x \in (0,1) \\ & f < g \text{ if } f \leqslant g \text{ and } \int_0^1 f^*(t)\, d\lambda(t) = \int_0^1 g^*(t)\, d\lambda(t). \end{aligned} \tag{5}$$

Characterizations and properties of the 'continuous' Schur-order were intensively discussed by *Ryff* [1965], *Luxemburg* [1967] and *Chong/Rice* [1971]. A famous theorem of *Hardy/Littlewood/Polya* [1952], *Chong* [1974, Theorems 2.3, 2.5], states that for $f, g \in L^1(\lambda)$ $f \preccurlyeq g$ is equivalent to $\int \varphi \circ f d\lambda \leqslant \int \varphi \circ g d\lambda$ for all convex, nondecreasing $\varphi: R^1 \to R^1$, while

$$f < g \text{ is equivalent to } \int \varphi \circ f d\lambda \leqslant \int \varphi \circ g d\lambda \quad \text{for all convex } \varphi. \tag{6}$$

Some simple relations fulfilled by $<, \preccurlyeq$ are the following [cf. *Day*, 6.1, 6.2]

$$\begin{aligned} &f^* + g_* < f + g < f^* + g^* \\ &f^* - g^* < f - g < f^* - g_* \\ &f^* g_* \preccurlyeq fg \preccurlyeq f^* g^*, \text{ if } fg \in L^1(\lambda). \end{aligned} \tag{7}$$

Consider the following conditions on $\varphi: (R^n, \mathcal{B}^n) \to (R^1, \mathcal{B}^1)$.

We shall omit those arguments of φ which are the same in a certain formula.

A) $\varphi(u_i + h, u_j + h) - \varphi(u_i + h, u_j) - \varphi(u_i, u_j + h) + \varphi(u_i, u_j)$
$\geqslant 0, \forall u_i, u_j \in R^1, h \geqslant 0, i \neq j.$

B) $\varphi(u_i + h) - 2\varphi(u_i) + \varphi(u_i - h) \geqslant 0, \quad \forall u_i \in R^1, h \geqslant 0.$

Functions satisfying condition A) are called L-superadditive. For a discussion of L-superadditive functions cf. *Marshall/Olkin* [1979, Chapter 6, C, D].

If φ has continuous second partial derivatives, then A), B) are equivalent to $(\partial^2 \varphi)/(\partial u_i \, \partial u_j) \geqslant 0$, $1 \leqslant i, j \leqslant n$.

Let $P^* \in M(P_1, \ldots, P_n)$ be defined by $P^* = P^{(F_1^{-1}(U), \ldots, F_n^{-1}(U))}$, where $P^U = R(0,1)$.

Corollary 3.

a) If $P_1, \ldots, P_n$ have first moments and φ satisfies condition A),then

$$\sup\{\int \varphi dP; P \in M(P_1, \ldots, P_n)\} = \int \varphi dP^*.$$

b) If φ satisfies conditions A), B) and $P_i' \in M^1(R^1, \mathcal{B}^1)$ have distribution functions G_i, $1 \leqslant i \leqslant n$, with $G_i^{-1} < F_i^{-1}$, $1 \leqslant i \leqslant n$, then

$$\int \varphi(G_1^{-1}(t), \ldots, G_n^{-1}(t)) \, d\lambda(t) \leqslant \int \varphi(F_1^{-1}(t), \ldots, F_n^{-1}(t)) \, d\lambda(t).$$

Proof.

a) By Theorem 2

$$\sup \{\int \varphi \, dP; P \in M(P_1, \ldots, P_n)\}$$
$$= \sup \{\int \varphi(f_1(t), \ldots, f_n(t)) \, d\lambda(t); f_i \text{ is a rearrangement of } F_i^{-1}, 1 \leqslant i \leqslant n\}.$$

In the case that F_i^{-1} are nonnegative and bounded, a) follows from a theorem of *Lorentz* [1953] on the rearrangement of functions. The condition of nonnegativity is not necessary for Lorentz's result while the general case (of integrable F_i^{-1}, $1 \leqslant i \leqslant n$) can be obtained similarly to the extension of a theorem of Hardy, Littlewood, Polya proved by *Chong* [1974, Theorem 2.5].

b) is implied similarly by Theorem 1 of *Ky Fan/Lorentz* [1954]. □

Examples:

a) If P_i have support in R_+, $1 \leqslant i \leqslant n$, then $\varphi(x_1, \ldots, x_n) = \prod_{i=1}^{n} x_i$, $x_i \geqslant 0$, $1 \leqslant i \leqslant n$, satisfies condition A), so that $\int \varphi dP^* = E \prod_{i=1}^{n} F_i^{-1}(U)$, $P^U = R(0,1)$ is the best upper bound for $E \prod_{i=1}^{n} X_i$ obtainable by fixing marginal distributions. For the case of continuous distributions with compact support cf. *Gaffke/Rüschendorf* [1980].

b) Let $\varphi(x_1, \ldots, x_n) = \max_{1 \leqslant i \leqslant n} x_i - \min_{1 \leqslant i \leqslant n} x_i$, $x_j \in R^1$, $1 \leqslant j \leqslant n$.
Then φ is convex, $-\varphi$ satisfies condition A). Therefore, $\int \varphi dP^* =$ $= \inf \{\int \varphi dP; P \in M(P_1, \ldots, P_n)\}$ and $\int \varphi dP^*$ is better than the lower bound obtained from Jensen's inequality. This result was obtained by *Schaefer* [1976] using Fréchet-bounds.

c) (6) and Lemma 1 (with $n = 1$) imply that for $P, Q \in M^1(R^1, B^1)$ with existing first moments:

$$\int \varphi dP \leqslant \int \varphi dQ, \ \forall \text{ convex } \varphi: R^1 \to R^1 \text{ is equivalent to}$$

$$F^{-1} < G^{-1}; F, G \text{ are the distribution functions of } P, Q, \tag{8}$$

which is equivalent to

$$\int_x^\infty (t - x) \, dF(t) \leqslant \int_x^\infty (t - x) \, dG(t), \quad \forall x \in R^1.$$

This characterization of convex ordering of distributions on (R^1, B^1) was proved in a different way by *Stoyan* [1972]. For applications of this ordering cf. *Stoyan* [1977].

d) For $\varphi: R^1 \to R^1$ convex and integrable, real random variables X, Y with distribution functions F, G:

$$E\varphi(F^{-1}(U)+G^{-1}(1-U)) \leqslant E\varphi(X+Y) \leqslant E\varphi(F^{-1}(U)+G^{-1}(U))$$
$$E\varphi(F^{-1}(U)-G^{-1}(U)) \leqslant E\varphi(X-Y) \leqslant E\varphi(F^{-1}(U)-G^{-1}(1-U)). \tag{9}$$

If $X \cdot Y \in L^1(P)$, then for all nondecreasing, convex φ

$$E\varphi(F^{-1}(U)\,G^{-1}(1-U)) \leqslant E\varphi(XY) \leqslant E\varphi(F^{-1}(U)\,G^{-1}(U)). \tag{10}$$

If $X_1, \ldots, X_n$ are integrable with $P^{X_i} = P_i$, $1 \leqslant i \leqslant n$, then

$$E\varphi(X_1 + \ldots + X_n) \leqslant E\varphi(\sum_{i=1}^{n} F_i^{-1}(U)) \tag{11}$$

where F_i are the distribution functions of P_i, $1 \leqslant i \leqslant n$ and φ is assumed to be convex.

(9), (10) follow from (2), (6) and (7) observing that $F^{-1}(1-t)$ is the nonincreasing rearrangement of $F^{-1}(t)$. (11) follows from Corollary 3 since $f(x_1, \ldots, x_n) = \varphi(x_1 + \ldots + x_n)$, φ convex, satisfies condition A). Otherwise, it is also implied by the obvious generalization of (7)

$$f_1 + \ldots + f_n < f_1^* + \ldots + f_n^*.$$

(10) generalizes a wellknown result of Hoeffding on the extreme correlation of two random variables [cf. *Whitt*, Lemma 2.3] while (9), (11) generalize results on the distance between distributions [cf. *Dall'Aglio*] and solve e.g. the problem of construction of random variables with maximum variance of the sum.

e) Part b) of Corollary 3 implies that

$$\sup\{\textstyle\int \varphi dP; P \in M(P_1, \ldots, P_n)\} \leqslant \sup\{\textstyle\int \varphi dP; P \in M(Q_1, \ldots, Q_n)\}$$

for φ satisfying A), B) and $Q_i \in M^1(R^1, B^1)$, $1 \leqslant i \leqslant n$, if $F_i^{-1} < G_i^{-1}$, $1 \leqslant i \leqslant n$ where F_i, G_i are the distribution functions of P_i, Q_i, $1 \leqslant i \leqslant n$.

Remark: The results of this paper can be generalized partially to more general spaces using results of *Luxemburg* [1967] on 'adequate' measure spaces.

References

Chong, K.-M.: Some extensions of a theorem of Hardy, Littlewood and Polya and their applications. Canad. J. Math. **26**, 1974, 1321–1340.

Chong, K.-M., and *N.M. Rice*: Equimeasurable rearrangement of functions. Queen's Papers in Pure and Applied Mathematics **28**, 1971.

Dall'Aglio, G.: Fréchet classes and compatibility of distribution functions. Symposia Mathematica **9**, 1972, 131–150.

Day, P.W.: Rearrangement inequalities. Canad. J. Math. **24**, 1972, 930–943.

Gaffke, N., and *L. Rüschendorf*: On a class of extremal problems in statistics. Mathem. Operationsforschung und Statistik. Series Optimization **12**, 1981, 123–136.

Hardy, G.H., J.E. Littlewood and *G. Polya*: Inequalities. 2nd ed. Cambridge 1952.

Ky Fan, and *G.G. Lorentz*: An integral inequality. Amer. Math. Monthly **61**, 1954, 626–631.

Lorentz, G.G.: An inequality for rearrangements. Amer. Math. Monthly **60**, 1953, 176–179.

Luxemburg, W.A.J.: Rearrangement invariant Banach function spaces. Queen's Papers in Pure and Applied Mathematics **10**, 1967, 83–144.

Marshall, A.W., and *I. Olkin*: Inequalities: Theory of Majorization and Its Applications. New York 1979.

Parthasaraty, K.R.: Probability Measures on Metric Spaces. New York 1967.

Rohlin, V.A.: On the Fundamental Ideas of Measure Theory. Amer. Math. Soc. Transl. **71**, 1952.

Rüschendorf, L.: Sharpness of Fréchet-bounds. Z. Wahrscheinlichkeitstheorie verw. Gebiete **57**, 1981, 293–302.

Ryff, J.V.: Orbits of L^1 functions under doubly stochastic transformations. Trans. Amer. Math. Soc. **117**, 1965, 92–100.

Schaefer, M.: Notes on the k-dimensional Jensen inequality. Ann. Prob. **4**, 1976, 502–504.

Stoyan, D.: Halbordnungsrelationen für Verteilungsgesetze. Math. Nachr. **52**, 1972, 315–331.

–: Qualitative Eigenschaften und Abschätzungen stochastischer Modelle. München 1977.

Whitt, M.: Bivariate distributions with given marginals. Ann. Statist. **4**, 1976, 1280–1289.

Received June 20, 1980
(revised version October 22, 1980)