Stochastic Inequalities
IMS Lecture Notes – Monograph Series
Volume 22 (1993)

EXTREMAL PROBLEMS FOR PROBABILITY DISTRIBUTIONS: A GENERAL METHOD AND SOME EXAMPLES

By L. MATTNER

Universität Hamburg

A general method for treating extremal problems for probability distributions is presented. It is based on a Lagrange multiplier rule for constrained extremal problems in cones of Banach spaces. Some concrete problems are discussed.

1. Introduction

The purpose of this article is to report on a general method for solving extremal problems for probability distributions, as well as to present some examples developed in detail in the author's thesis Mattner (1990a) of which Mattner (1990b) is the relevant part in this context.

Additionally, a new and, hopefully, illuminating example (number 2 below) is treated.

The idea underlying the method to be presented is quite simple, namely: Just apply the existing Lagrange multiplier theory for extremal problems in Banach Spaces and modify it slightly, in such a way that the essential side condition of positivity is taken care of. This will lead to a necessary condition to be satisfied by any solution of a given extremal problem, provided that the functional to be extremized as well as functionals representing side conditions are sufficiently well-behaved, e.g. continuously Fréchet-differentiable.

Before stating a general theorem, let us look at a specific example which in fact motivated my study.

EXAMPLE 1 Let X and Y denote independent and identically distributed real random variables with

$$(1) \qquad E[X] = 0, \quad Var(X) = 1.$$

AMS 1991 *subject classifications.* Primary 60E15; Secondary 49K27, 62H05.

Key words and phrases. Characterizations of probability distributions, expectation inequalities, extremal problems, Lagrange multiplier rule.

The problem is to maximize the expected distance of X and Y, under the above constraints:

$$E[|X-Y|] \quad \overset{!}{=} \quad \max. \tag{2}$$

This problem has received its solution a long time ago: the maximum is attained if and only if X and Y are uniformly distributed over the interval $[-\sqrt{3},\sqrt{3}]$. In fact, Plackett (1947) considered a more general problem. He gave an argument which, for the present case, essentially reduces to writing the expected distance as

$$E[|X-Y|] = -2\int_{-\infty}^{\infty} F(1-F)dx, \tag{3}$$

where F denotes the distribution function of X, and performing a variation with respect to F. But he did this only formally, neither worrying about the existence of a solution nor making sure in an adequate way that his variations where still distribution functions. Nevertheless, he somehow arrived at the solution stated above. A proof of its correctness was first indicated by Moriguti (1951, footnote 5, p. 534) and given explicitly by Hartley and David (1954, p. 88), roughly speaking by applying the Cauchy-Schwarz inequality to the right-hand side of (3) after first manipulating that integral in such a way as to make sure that equality will hold for the presumed solution of Plackett.

2. A General Method

Thus it appears that the method used to solve Plackett's problem is unsystematic and also intrinsically univariate, the latter since it relies on manipulations involving the distribution function. These remarks apply as well to a more recent proof based on Terrell (1983) and given in Baringhaus and Henze (1990). Hence before trying to solve similar and perhaps more complicated problems, one should look for a general method yielding Plackett's result. To this end we observe that the problem may be viewed as an extremal problem in a Banach space.

Namely, let

$$M \;:=\; \{\mu : \mu \quad \text{signed Borel measure with} \quad ||\mu|| < \infty\},$$

where

$$||\mu|| \;:=\; \int (1+x^2)\;|d\mu(x)| \tag{4}$$

and $|d\mu(x)|$ denotes integration with respect to the total variation of μ. Plackett's problem, which I prefer to write as a minimization problem, may

then be written as follows:
"Minimize

$$\varphi_0(\mu) := -\int\int |x-y|\, d\mu(x)\, d\mu(y)$$

subject to the constraints

$$\begin{array}{rcll} \psi_1(\mu) & := & \int x\, d\mu(x) & = 0, \\ \psi_2(\mu) & := & \int x^2 d\mu(x) - 1 & = 0, \\ \psi_3(\mu) & := & \mu(\mathbb{R}) - 1 & = 0, \\ \mu & \geq & 0." & \end{array}$$

If the last condition were absent, we would just have an extremal problem in a Banach space with finitely many one-dimensional side conditions. What makes things slightly complicated is that $\mu \geq 0$ is an inequality constraint of infinite dimensional character. However, it may be written as $\mu \in C$, where C is the cone of the positive measures in M. And in fact, there is a Lagrange multiplier rule applicable in such cases:

THEOREM 1 *Let Z be a Banach space,*

$$\begin{array}{rcccl} \varphi_0 & : & Z & \to & \mathbb{R}, \\ \varphi_i & : & Z & \to & \mathbb{R}, \quad i = 1, \dots, m, \\ \psi_j & : & Z & \to & \mathbb{R}, \quad j = 1, \dots, n, \end{array}$$

continuously Fréchet-differentiable, and

(5) $$C \text{ a convex cone in } Z.$$

Define

$$\mathcal{L}(z) := \lambda_0 \varphi_0(z) + \sum_{i=1}^{m} \lambda_i \varphi_i(z) + \sum_{j=1}^{n} \alpha_j \psi_j(z).$$

If $z \in Z$ minimizes φ_0 subject to

$$\begin{array}{rcll} \varphi_i(z) & \leq & 0, & i = 1, \dots, m, \\ \psi_j(z) & = & 0, & j = 1, \dots, n, \\ z & \in & C, & \end{array}$$

then there exist $\lambda_0, \lambda_1, \dots, \lambda_m, \alpha_1, \dots, \alpha_n \in \mathbb{R}$ with

$$\begin{array}{ll} (i) & \text{not all } \lambda_i \text{ and } \alpha_j \text{ vanish}, \\ (ii) & \lambda_i \geq 0, \quad i = 0, \dots, m, \\ (iii) & \langle \mathcal{L}'(z), w \rangle \ \geq 0, \quad w \in C, \\ (iv) & \langle \mathcal{L}'(z), z \rangle \ = 0. \end{array}$$

This is proved in Mattner (1990a,b) by an application of an abstract multiplier rule given in Tikhomirov (1986). Here we try to explain the theorem by looking at two special cases, which should make it easy to memorize it.

In case one we assume that C is the whole space. Then condition (iii) may be applied to any w as well als to $-w$. So it just says that $\mathcal{L}'(z)$ is the zero functional. This contains condition (iv), and the theorem reduces to the ordinary Lagrange multiplier rule for Banach spaces.

In case two we assume that the ordinary constraints given by the φ_i and ψ_j are absent. Consider a typical point $z \in C$. It will usually lie on the boundary. In fact, it is easy to see, that the cone of positive measures in Plackett's example has empty interior. If now z minimizes φ_0 in C, then those one-sided directional derivatives $\langle \varphi_0'(z),\ w\rangle$ of φ_0 for which $w \in C$ have to be nonnegative. This yields condition (iii) with $\mathcal{L} = \varphi_0$. In the direction of z we may even perform a two-sided derivative, which accordingly has to vanish. This is condition (iv). So the theorem is seen to be true in both cases, and it is at least plausible that it is true in general.

3. Examples

We will now look at several examples, beginning with Plackett's problem.

EXAMPLE 1 (CONT.) A standard argument involving tightness, Fatou's lemma and integration to the limit of uniformly integrable sequences (see Mattner (1990a), pp. 16-17, for details) shows that a solution of the extremal problem exists. The functionals involved here are either quadratic or linear, and continuous by choice of the norm (4). Hence the derivative of the Lagrange functional $\mathcal{L}$ is given by

$$\begin{aligned}\langle \mathcal{L}'(\mu), \nu\rangle &= -2\lambda_0 \textstyle\int\int |x-y|\, d\mu(y)\, d\nu(x) + \alpha_1 \int x d\nu(x) \\ &\quad + \alpha_2 \textstyle\int x^2 d\nu(x) + \alpha_3 \nu(\mathbb{R}) \\ &= \textstyle\int l(x)\, d\nu(x),\end{aligned}$$

where

$$l(x) = -2\lambda_0 \int |x-y| d\mu(y) + \alpha_1 x + \alpha_2 x^2 + \alpha_3.$$

Assuming that μ is any solution, we may apply (iii) to any Dirac measure $\nu = \delta_x$, in order to get

$$l(x) \begin{cases} \geq 0, & x \in \mathbb{R} \\ = 0, & x \in \operatorname{supp} \mu, \end{cases}$$

where the equality in supp μ, the support of μ, follows from the nonnegativity of l and (iv).

This is the typical preliminary result obtained when working with the above Lagrange multiplier rule: We have an integral relation for the extremal measure μ consisting of an integral inequality in the whole space and an integral equality in the support of μ. A difficulty is that we don't know the support since we don't know μ. From this point on, the arguments to follow will have to make use of the specific properties of the extremal problem at hand.

Assume for a moment that λ_0 vanishes. Then l is a nonnegative polynomial of degree at most 2, not vanishing identically and hence having at most one zero. This implies that the support of μ is a singleton, which is impossible since the variance of μ equals one. This contradiction shows that we may assume without loss of generality that $2\,\lambda_0 = 1$.

It is now convenient to rewrite l as

$$l(x) = \int_x^{\infty} (1 - F(x))\,dy + \int_{-\infty}^{x} F(y)\,dy + \alpha_1 x + \alpha_2 x^2 + \alpha_3,$$

where F denotes the distribution function of μ and the familiar area formula for the expectation of a random variable has been applied. Clearly, l possesses at least one-sided derivatives given by

$$l'(x\pm) = 2F(x\pm) - 1 + \alpha_1 + 2\alpha_2 x.$$

We have

$$0 \leq 2\,(F(x+) - F(x-)) = l'(x-) - l'(x+) \leq 0, \quad x \in \operatorname{supp}\mu,$$

the latter inequality holding because l assumes its minimum value at every $x \in \operatorname{supp}\mu$. This shows that F is continuous and that l' exists and vanishes in the support of μ. Hence

$$F(x) = a + bx, \quad x \in \operatorname{supp}\mu.$$

Using the continuity of F, it readily follows that μ is uniform over some interval, which is determined by the mean and variance.

The purpose of the above example was to illustrate the multiplier method in one of the simplest nontrivial cases, rather than to match other proofs of Plackett's result in brevity. The remaining examples were not previously treated by simpler methods and illustrate various aspects of the present method.

EXAMPLE 2 This example will show that the problem of the support of the extremal measure is not a trivial matter. Bentkus (1991) raised the question of what happens if the constraint

$$E[|X|^3] \leq \beta$$

is added in Example 1.

It makes things slightly easier to omit the condition on the mean, which turns out to be satisfied anyway if we replace (1) by

$$E[X^2] = 1, \; E[|X|^3] \le \beta. \tag{6}$$

The argument given in Example 1 applies virtually without change, leading to the existence of a solution μ for every $\beta > 1$ and to the condition

$$F(x) = \frac{1}{2} + bx + cx|x| =: \varphi(x), \quad x \in \operatorname{supp} \mu, \tag{7}$$

for the corresponding continuous distribution function F, where c is known to be nonnegative.

In case $b \ge 0, \varphi$ is strictly increasing, which forces F to agree with φ on an interval symmetrical with respect to the origin, i.e. F has a density f given by

$$f(x) = b + 2c|x|, \quad |x| \le A. \tag{8}$$

However, in case $b < 0, \varphi$ increases in $(-\infty, \frac{b}{c}]$, decreases in $[\frac{b}{c}, -\frac{b}{c}]$, and increases again in $[-\frac{b}{c}, \infty)$, strictly in each case. Now (7) and continuity of F allow for several possibilities. In each case F has to coincide with φ in one or two compact intervals and be constant in the complementary intervals. Instead of calculating $E[|X - Y|]$ for each of these possible F, it is more convenient to observe that the extremal distribution has to be symmetrical with respect to the origin, leaving a density given by

$$f(x) = b + 2c|x|, \quad -\frac{b}{c} \le |x| \le A \tag{9}$$

as the only possibility. In fact, if the distribution of X is extremal satisfying (6), so is that of $-X$. The representation (3) shows that the functional to be minimized is strictly convex on the set of the probability measures satisfying (6). Hence, the solution is unique, i.e. X is distributed as $-X$.

Taking $\int f dx = \int x^2 f(x) dx = 1$ into account, we get from (8) and (9) by trite calculations

$$\begin{aligned} f(x) &= (1-\alpha)\tfrac{1}{2}\sqrt{\tfrac{2+\alpha}{6}} + \alpha\tfrac{2+\alpha}{6}|x|, \quad |x| \le \sqrt{\tfrac{6}{2+\alpha}}, \\ \beta(f) &:= \int |x|^3 f(x) dx = \tfrac{3\sqrt{6}}{10}\,\tfrac{5+3\alpha}{(2+\alpha)^{\frac{3}{2}}}, \\ E_f[|X-Y|] &= \sqrt{\tfrac{6}{2+\alpha}}\left(\tfrac{2}{3} + \tfrac{\alpha}{6} - \tfrac{\alpha^2}{30}\right) \end{aligned}$$

for $b \geq 0$ and

$$\begin{aligned}
f(x) &= \left(\tfrac{1}{\alpha}\sqrt{1-\alpha+\tfrac{2}{3}\alpha^2-\tfrac{\alpha^3}{6}}|x| - \tfrac{1-\alpha}{\alpha}\right)\sqrt{1-\alpha+\tfrac{2}{3}\alpha^2-\tfrac{\alpha^3}{6}}, \\
&\quad (1-\alpha)\left(1-\alpha+\tfrac{2}{3}\alpha^2-\tfrac{\alpha^3}{6}\right)^{-\frac{1}{2}} \leq |x| \\
&\leq \left(1-\alpha+\tfrac{2}{3}\alpha^2-\tfrac{\alpha^3}{6}\right)^{-\frac{1}{2}}, \\
\beta(f) &:= \int |x|^3 f(x)dx = \frac{1-\frac{3}{2}\alpha+\frac{3}{2}\alpha^2-\frac{3}{4}\alpha^3+\frac{3}{20}\alpha^4}{\left(1-\alpha+\frac{2}{3}\alpha^2-\frac{\alpha^3}{6}\right)^{\frac{3}{2}}}, \\
E_f[|X-Y|] &= \frac{1-\frac{\alpha}{3}+\frac{2}{15}\alpha^2}{\sqrt{1-\alpha+\frac{2}{3}\alpha^2-\frac{\alpha^3}{6}}}
\end{aligned}$$

for $b \leq 0$, where in each case α is allowed to vary between zero and one. We observe that, for the above densities f, $E_f[|X-Y|]$ is strictly increasing in $\beta(f)$, by considering both as functions of α. The largest value of $\beta(f)$ is $\frac{3\sqrt{3}}{4}$, corresponding to the uniform density. Hence we may conclude that the f with $\beta(f) = \min\left(\beta, \frac{3\sqrt{3}}{4}\right)$ is the solution of our problem. The support of these solutions is disconnected for $1 < \beta < \frac{4\sqrt{2}}{5}$ and connected for $\frac{4\sqrt{2}}{5} \leq \beta < \infty$, which was hardly obvious at the outset.

Incidentally, it follows from the above that we get an extremal problem without solution if we replace (6) by

$$E[X^2] = 1, \quad E[|X|^3] = \beta$$

for some $\beta > \frac{4\sqrt{2}}{5}$.

The next two examples concern multivariate extremal problems.

EXAMPLE 3 The expected distance makes sense also for multivariate random variables. How large can it be given the second moment of the euclidean norm? The answer is given by the following theorem.

THEOREM 2 *If X and Y are independent and identically distributed random vectors in d-space, then*

$$E[|X-Y|] \leq \sqrt{2}\sqrt{E[|X|^2]} \cdot \begin{cases} \sqrt{\frac{2}{3}}, & d=1, \\ \frac{\pi}{\sqrt{6}}, & d=2, \\ \frac{\Gamma^2(\frac{d}{2})}{\Gamma(\frac{d}{2}-\frac{1}{4})\Gamma(\frac{d}{2}+\frac{1}{4})}, & d \geq 3, \end{cases}$$

where equality and $E|X|^2 = 1$ occurs if and only if

$$\begin{array}{ll}
X \sim U([-\sqrt{3}, \sqrt{3}]), & d=1 \\
f_X(x) = \frac{1}{2\pi}\frac{\frac{2}{3}}{\sqrt{1-\frac{2}{3}|x|^2}}\, 1\left(|x| \leq \sqrt{\frac{3}{2}}\right), & d=2 \\
X \sim U(\{x \in \mathbb{R}^d : |x| = 1\}), & d \geq 3.
\end{array}$$

Here $|\cdot|$ *denotes the euclidean norm and* U *stands for uniform distribution.*

This is proved in Mattner (1990a,b) and, independently and in a different way, in Buja, Logan, Reeds and Shepp (1990). Of course, for $d = 1$ we just get a reformulation of Plackett's result.

The case $d = 2$ is particularly interesting. A heuristic for it runs as follows. The multiplier rule leads as before to an integral relation for a measure μ maximizing the expected distance given $E[|X|^2] = 1$:

$$(10) \qquad - \int\limits_{\mathbb{R}^2} |x - y| \, d\mu(y) + \alpha_1 + \alpha_2 \, |x|^2 \begin{cases} \geq 0, & x \in \mathbb{R}^2 \\ = 0, & x \in \text{ supp } \mu. \end{cases}$$

Let us formally apply the Laplacian to the above equality. Because of $\Delta|\cdot| = \frac{1}{|\cdot|}$, valid in $\mathbb{R}^2$, this should lead to something like

$$\int_{\mathbb{R}^2} \frac{d\mu(y)}{|x - y|} = 4 \, \alpha_2 \quad (x \in \text{ supp } \mu \subset \mathbb{R}^2).$$

Now read this relation "three-dimensionally": under the plausible assumption that the support of μ is a circular disk, it says that the spatial potential of μ is constant on it and hence μ has to be the electrostatic equilibrium distribution of unit charge on that disk. The latter is known and has, taking the condition $E[|x|^2] = 1$ into account, the density given in the theorem.

For a rigorous proof note that the euclidean norm is a so-called negative-definite function, which almost by definition implies that the functional

$$\mu \mapsto - \int \int |x - y| \, d\mu(x) \, d\mu(y)$$

is convex on the set of those probability measures with finite second moments. The convexity is in fact strict (This is true for arbitrary dimensions. We encountered the one-dimensional and elementary case in the previous example.). An application of the Kuhn-Tucker theorem shows that (10) is also sufficient for μ to be extremal. A not completely trivial calculation shows that the density of the theorem fulfills (10). See Mattner (1990a,b) for details.

So far we have solved given extremal problems. Now we are reversing the question and ask for an extremal problem having a given solution.

EXAMPLE 4 Can we characterize the uniform distribution over a ball in d-space as the solution of an extremal problem similar to Plackett's? Replacing $|x - y|$ by some unspecified function $K_d(x, y)$ and proceeding formally as above, we get

$$\int K_d(x, y) \, d\mu(y) = \alpha_1 + \alpha_2 \, |x|^2 \quad (x \in \text{ supp } \mu)$$

as a necessary condition for any extremal μ. Applying again the Laplacian to this equation, we get a constant on the right-hand side and want to get $f(x)$, a density of μ, on the left-hand side. This will be the case if $\Delta_x K_d(x,y)$ (Laplacian with respect to x) is the Dirac measure located at y. This suggests to take $K_d(x,y) = u(x-y)$ with u a constant multiple of a fundamental solution of the Laplacian. Thus we are led to guess the following theorem, proved in Mattner (1990a,b).

THEOREM 3 *If*

$$K_d(x,y) = \begin{cases} -|x-y|, & d=1, \\ \log \frac{1}{|x-y|}, & d=2, \\ \frac{1}{|x-y|^{d-2}}, & d \geq 3, \end{cases}$$

and if X and Y are independent and identically distributed random vectors in d-space with $E[|x|^2] = 1$, then $E[K_d(X,Y)]$ is minimal if and only if $X \sim U(\{x \in \mathbb{R}^d : |x| \leq r_d\})$ for some suitable r_d.

EXAMPLE 5 Can we characterize any given probability distribution as in Example 4? It is proved in Mattner (1990a,b) that the answer is "yes" under some regularity conditions, as well as that it is often "no" if we try $K(x,y) = u(x-y)$ for some function u:

THEOREM 4 *If $|u(x)| \leq A(1+x^2)$ for some finite A and if $E[u(X-Y)]$ is extremal for $X \sim N(0,1)$ under the constraints (1), then u is a polynomial of degree at most 2 and there are other extremal distributions.*

Acknowledgements

It is a pleasure to thank Professors Moshe Shaked and Y. L. Tong for organizing this conference and inviting me to present this paper. Financial support by the National Science Foundation and the Deutsche Forschungsgemeinschaft is gratefully acknowledged. Also I wish to thank a referee for his careful reading of this paper.

REFERENCES

BARINGHAUS, L., AND HENZE, N. (1990). A consistent test for uniformity with unknown limits based on D'Agostinos D. *Statist. Prob. Lett.* **9** 299-304.

BENTKUS, V. (1991). Personal communication.

BUJA, A., LOGAN, B. F., REEDS, J. R. AND SHEPP, L. A. (1990). Inequalities and positive-definite functions arising from a problem in multidimensional scaling. Preprint.

HARTLEY, H. O. AND DAVID, H. A. (1954). Universal bounds for mean range and extreme observation. *Ann. Math. Statist.* **25** 85-99.

MATTNER, L. (1990a). Behandlung einiger Extremalprobleme für Wahrscheinlichkeitsverteilungen. Dissertation, Universität Hannover.

MATTNER, L. (1990b). Extremal problems for probability distributions. Preprint.

MORIGUTI, S. (1951). Extremal properties of extreme value distributions. *Ann. Math. Statist.* **22** 523-536.

PLACKETT, R. L. (1947). Limits of the ratio of mean range to standard deviation. *Biometrika* **34** 120-122.

TERRELL, G. R. (1983). A characterization of rectangular distributions. *Ann. Probab.* **11** 823-826.

TIKHOMIROV, V. M. (1986). *Theory of Extremal Problems.* Wiley, New York.

INSTITUT FÜR MATHEMATISCHE STOCHASTIK
UNIVERSITÄT HAMBURG
BUNDESSTR. 55
D-2000 HAMBURG 13
GERMANY