Alexander Shapiro

Worst-case distribution analysis of stochastic programs

Abstract. We show that for even quasi-concave objective functions the worst-case distribution, with respect to a family of unimodal distributions, of a stochastic programming problem is a uniform distribution. This extends the so-called "Uniformity Principle" of Barmish and Lagoa (1997) where the objective function is the indicator function of a convex symmetric set.

Key words. Min-max Stochastic Programming – Ambiguous Chance Constraints – Robust Optimization – Uniformity Principle

1. Introduction

Let $g : \mathbb{R}^n \to \mathbb{R}$ be a measurable function and μ be a (Borel) probability measure (distribution) on $\mathbb{R}^n$. By $\mathbb{E}_\mu[g(W)]$ we denote the expectation of the random variable $g(W)$, where W is a random vector with probability distribution μ. In this paper we discuss the following optimization problem

$$\underset{\mu \in \mathcal{P}}{\text{Min}} \left\{ \phi(\mu) := \mathbb{E}_\mu[g(W)] \right\}, \tag{1}$$

where $\mathcal{P}$ is a family of probability distributions on $\mathbb{R}^n$. In case the objective function $g(w, x)$ also depends on a decision vector $x \in X$, one can try to solve the corresponding min-max stochastic programming problem of maximization of the optimal value $\vartheta(x)$ of the associated problem (1) with respect to $x \in X$. In this respect an optimal solution of (1), for a fixed x, can be viewed as the worst-case distribution from the specified family $\mathcal{P}$ of distributions.

The min-max approach to stochastic programming has been discussed in a number of publications (e.g., [2, 3, 5, 6, 12, 13]). If the family $\mathcal{P}$ is defined by constraints of the form $\mathbb{E}_\mu[h_i(W)] = b_i, i = 1, \ldots, r$, then (1) becomes the classical problem of moments (see, e.g., [10] and references therein). The case of moment constraints was studied extensively in the min-max approach to stochastic programming. On the other hand, quite often, it is natural to assume that the random data can vary in a specified region $\Xi \subset \mathbb{R}^n$ around some reference value $\bar{w} \in \mathbb{R}^n$. For example, one can use box-type constraints of the form $\Xi := \{w : \bar{w}_i - a_i \leq w_i \leq \bar{w}_i + b_i, \ i = 1, \ldots, n\}$ restricting variability of the data. In that case we may consider families of distributions supported on the specified region Ξ. In particular, if we take the family of Dirac measures $\mathcal{P} := \{\delta_w : w \in \Xi\}$, then problem (1) becomes the problem of minimization of $g(w)$ subject to $w \in \Xi$ (recall that

A. Shapiro: School of Industrial and Systems Engineering, Georgia Institute of Technology, Atlanta, Georgia 30332-0205, USA. e-mail: `ashapiro@isye.gatech.edu`

Dirac measure δ_w is the measure of mass one at w). If, moreover, the function $g(w, x)$ depends on x and the corresponding optimal value $\vartheta(x)$ is used as a constraint of the form $\vartheta(x) \geq 0$, then such constraints correspond to the robust approach to optimization problems. Note also that if $g(\cdot) := \mathbb{1}_S(\cdot)$ is the indicator function of a set $S \subset \mathbb{R}^n$, then $\phi(\mu) = \mu(S)$. The associated worst-case chance constrained problems were called ambiguous chance constrained problems in [4].

We assume throughout the paper that all measures $\mu \in \mathcal{P}$ are supported on a compact set $\Xi \subset \mathbb{R}^n$ and consider families of unimodal distributions. Recall that a univariate distribution is said to be *unimodal*, with mode m, if its cumulative distribution function is convex on the interval $(-\infty, m)$ and concave on the interval $(m, +\infty)$. Equivalently, the distribution is unimodal if it is a mixture of the Dirac distribution δ_m and a distribution with density function that is nondecreasing on $(-\infty, m]$ and nonincreasing on $[m, +\infty)$, i.e., this density function is quasi-concave. By a result due to Khintchine we have that a distribution is unimodal with mode $m = 0$ iff it is the distribution of the product $W = UZ$, where U and Z are independent random variables, U is uniformly distributed on the interval $[0, 1]$ and the distribution of Z is arbitrary. Moreover, the unimodal distribution is supported on the interval $[a, b]$, with $a \leq 0 \leq b$, iff the random variable Z is supported on $[a, b]$. We refer to UZ as the Khintchine representation of the considered unimodal distribution. Note that the probability of the event "$W = 0$" is equal to the probability of "$Z = 0$" and can be positive.

In this paper we consider families of distributions $\mathcal{P}$ formed by distributions of random vectors $W = (W_1, \ldots, W_n)$ which can be represented in the product form $W = UZ$ (the product UZ is taken componentwise, i.e., $W_i = U_i Z_i$, $i = 1, \ldots, n$), with random vectors $U = (U_1, \ldots, U_n)$ and $Z = (Z_1, \ldots, Z_n)$ satisfying the following conditions: (i) U has uniform distribution on a convex compact set $K \subset \mathbb{R}^n$, (ii) Z has an arbitrary distribution on a convex compact set $C \subset \mathbb{R}^n$. (iii) U and Z are independent. We refer to $\mathcal{P}$ as *Khintchine's family* of distributions associated with sets K and C.

In particular, if $K := [0, 1]^n$, then U_i are independent and if, moreover,

$$C := \times_{i=1}^n [a_i, b_i] = \{z \in \mathbb{R}^n : a_i \leq z_i \leq b_i, \ i = 1, \ldots, n\},$$

where $a_i < 0 < b_i$, $i = 1, \ldots, n$, then the marginal distributions of W_i's are unimodal (with mode 0) on the respective intervals $[a_i, b_i]$. If $K := [-1, 1]^n$ and $C := \times_{i=1}^n [0, b_i]$, then the marginal distributions of W_i's are unimodal and even (symmetric) on $[-b_i, b_i]$. In both cases random variables W_i, $i = 1, \ldots, n$, are independent, if Z_i are independent.

2. Derivations

Let $\mathcal{P}$ be Khintchine's family of distributions associated with sets K and C. Consider the corresponding random vector $W = UZ$ having a distribution μ from this family. We can write

$$\mathbb{E}_\mu[g(W)] = \mathbb{E}_Z \{E_{W|Z}[g(W)]\} = \mathbb{E}[h(Z)], \tag{2}$$

where $h(z) := \mathbb{E}[g(W) \mid Z = z]$ is the conditional expectation of $g(W)$ given $Z = z$. Let $\mathcal{M}$ be the family of (Borel) probability measures on the set C, and consider the optimization problem

$$\underset{\nu \in \mathcal{M}}{\text{Min}} \, \mathbb{E}_\nu[h(Z)]. \tag{3}$$

It is well known in the theory of the problem of moments (and is not difficult to show directly) that it suffices to solve problem (3) with respect to (Dirac) measures $\nu = \delta_z$, $z \in C$. Of course, any Dirac measure δ_z is the product of its marginal measures δ_{z_i}, $i = 1, \dots, n$. Also if Z has distribution δ_z, i.e., $Z \equiv z$, then by independence of U and Z we have that $W = UZ$ has uniform distribution on the set

$$zK = \{w \in \mathbb{R}^n : w = (z_1 u_1, \dots, z_n u_n), \ u \in K\}.$$

Consider the family of distributions $\mathcal{U} \subset \mathcal{P}$ formed by distributions of random vectors Uz, where U is uniformly distributed on the set K and $z \in C$ is an arbitrary (deterministic) vector. By the above discussion we obtain that problem (1) is equivalent to the problem

$$\underset{\mu \in \mathcal{U}}{\text{Min}} \, \mathbb{E}_\mu[g(W)]. \tag{4}$$

Proposition 1. *Let $\mathcal{P}$ be Khintchine's family of distributions. Then optimization problems (1) and (4) are equivalent in the sense that they have the same optimal value and any optimal solution of (4) is an optimal solution of (1).*

In particular, consider the family of distributions, denoted $\mathcal{F}_U$, formed by distributions of random vectors W with mutually independent components W_i, $i = 1, \dots, n$, such that each W_i has a unimodal distribution, with mode 0, on the interval $[a_i, b_i]$, $a_i < 0 < b_i$. We refer to $\mathcal{F}_U$ as the family of *unimodal distributions* on the (rectangular) set $\Xi = \times_{i=1}^n [a_i, b_i]$. If, moreover, $a_i = -b_i$ and the distribution of each W_i is even (symmetric), then we refer to the corresponding family $\mathcal{F}_E \subset \mathcal{F}_U$ as the family of *even unimodal distributions* on $\times_{i=1}^n [-b_i, b_i]$. By Khintchine's theorem we have that $\mathcal{F}_U$ and $\mathcal{F}_E$ are subfamilies of the corresponding Khintchine's families associated with sets $K := [0, 1]^n$, $C := \times_{i=1}^n [a_i, b_i]$, and $K := [-1, 1]^n$, $C := \times_{i=1}^n [0, b_i]$, respectively. Note that the inclusions $\mathcal{F}_U \subset \mathcal{P}$ and $\mathcal{F}_E \subset \mathcal{P}$ are strict since it is not assumed in the definition of the corresponding Khintchine's family $\mathcal{P}$ that the components of Z are mutually independent. Here the family $\mathcal{U}$, corresponding to $\mathcal{F}_U$, is formed by uniform distributions on rectangular sets $\times_{i=1}^n I_i$, where each I_i is a subinterval either of $[0, b_i]$ or $[a_i, 0]$ with one of its edges being zero. Similarly, the family $\mathcal{U}$ corresponding to $\mathcal{F}_E$ is formed by uniform distributions on rectangular sets $\times_{i=1}^n I_i$, where each I_i is a symmetric subinterval of $[-b_i, b_i]$.

Corollary 1. *Let $g(w)$ be a measurable function and $\mathcal{F}_U$ ($\mathcal{F}_E$) be the family of unimodal (even unimodal) distributions on $\times_{i=1}^n [a_i, b_i]$ (on $\times_{i=1}^n [-b_i, b_i]$). Then the problem of minimization of $\phi(\mu) := \mathbb{E}_\mu[g(W)]$ subject to $\mu \in \mathcal{F}_U$ (subject to $\mu \in \mathcal{F}_E$), is equivalent to the reduced problem of minimization of $\phi(\mu)$ subject to $\mu \in \mathcal{U}$, where $\mathcal{U}$ is the family of uniform distributions corresponding to $\mathcal{F}_U$ (to $\mathcal{F}_E$).*

By taking $g(\cdot) := \mathbb{1}_S(\cdot)$, where S is a nonempty subset of $\mathbb{R}^n$, we obtain that the problem of minimization of $\phi(\mu) := \mu(S)$ subject to $\mu \in \mathcal{F}_E$, is equivalent to the problem of minimization of $\phi(\mu)$ subject to $\mu \in \mathcal{U}$. For a convex set S this result was derived in [1], and in a more complete form in [7].

Recall that a function $g : \mathbb{R}^n \to \mathbb{R}$ is said to be quasi-concave if for any $w_1, w_2 \in \mathbb{R}^n$ and $\gamma \in [0, 1]$ the inequality

$$g(\gamma w_1 + (1 - \gamma)w_2) \geq \min\{g(w_1), g(w_2)\}$$

holds. It is said that $g(\cdot)$ is even if $g(w) = g(-w)$ for all $w \in \mathbb{R}^n$.

Theorem 1. *Let $\mathcal{P}$ be Khintchine's family of distributions associated with sets $K :=$ $[-1, 1]^n$ and $C := \times_{i=1}^n [0, b_i]$, where $b_i > 0$, $i = 1, \ldots, n$, $\Xi := \times_{i=1}^n [-b_i, b_i]$, and $g : \Xi \to \mathbb{R}$ be a bounded, quasi-concave, even function. Then problem (1) attains its optimal value at the uniform on the set Ξ distribution.*

In order to prove the above theorem we use the following lemmas. Recall that the Brunn-Minkowski inequality asserts that if A, $B \subset \mathbb{R}^d$ are nonempty compact sets, then

$$\mathrm{vol}(A + B)^{1/d} \geq \mathrm{vol}(A)^{1/d} + \mathrm{vol}(B)^{1/d}, \tag{5}$$

where $\mathrm{vol}(A)$ denotes volume of set A (e.g., [11]).

Lemma 1. *Let S be a convex compact subset of $\mathbb{R}^n$ and $\lambda(t)$ be the $(n-1)$-dimensional volume of the set $L_t := \{z \in S : z_1 = t\}$. Then the function $\lambda(t)$ is quasi-concave.*

Proof. By convexity of S we have that for any $t_1, t_2 \in \mathbb{R}$ and $\gamma \in [0, 1]$,

$$\gamma L_{t_1} + (1 - \gamma)L_{t_2} \subset L_{\gamma t_1 + (1-\gamma)t_2}.$$

Together with the Brunn-Minkowski inequality this implies

$$\begin{aligned}
\lambda(\gamma t_1 + (1 - \gamma)t_2) &\geq \mathrm{vol}(\gamma L_{t_1} + (1 - \gamma)L_{t_2}) \\
&\geq \left[\gamma \lambda(t_1)^{1/(n-1)} + (1 - \gamma)\lambda(t_2)^{1/(n-1)}\right]^{n-1} \\
&\geq \min\{\lambda(t_1), \lambda(t_2)\},
\end{aligned}$$

which completes the proof. $\square$

Lemma 2. *Let S be a convex symmetric, with respect to the origin, subset of $\mathbb{R}^n$ and $\mathbb{P}_z$ be the probability measure uniformly distributed on the rectangular set $R_z := \times_{i=1}^n [-z_i, z_i]$, where $z_i > 0$, $i = 1, \ldots, n$. Then the function $\rho(z) := \mathbb{P}_z(S \cap R_z)$ is monotonically nonincreasing in each $z_i \in (0, +\infty)$.*

Proof. Let us fix, say, $z_i = 1/2$, $i = 2, \ldots, n$, and consider $\rho(z)$ as a function of the first component $z_1 = z$ (with some abuse of notation we view now $\rho(z)$ as a function of $z \in \mathbb{R}$). Observe that $\rho(z) = v(z)/(2z)$, where $v(z) := \mathrm{vol}(S \cap R_z)$. Consider the set $L_t := \{w \in S \cap R_z : w_1 = t\}$, and let $\lambda(t)$ be the $(n-1)$-dimensional volume of L_t. We have then that $v(z) = \int_{-z}^z \lambda(t)dt$. Also by Lemma 1 the function $\lambda(t)$ is quasi-concave and is even since S is symmetric. Consequently, $\lambda(t)$ is monotonically nonincreasing on $(0, +\infty)$. It follows that the function $v(\cdot)$ is Lipschitz continuous on $(0, +\infty)$, and

hence $\rho(\cdot)$ is Lipschitz continuous on any interval $[\alpha, \beta] \subset (0, +\infty)$. Consequently, $\rho(\cdot)$ is almost everywhere differentiable on $[\alpha, \beta]$ and

$$\rho(\beta) - \rho(\alpha) = \int_\alpha^\beta \rho'(t)dt = \tfrac{1}{2}\int_\alpha^\beta \frac{v'(t)t - v(t)}{t^2}\,dt.$$

Therefore, it suffices to show that $\rho'(z)$ is less than or equal to zero for every $z > 0$ where $\rho'(z)$ exists, or, equivalently, that

$$v'(z) \leq v(z)/z \tag{6}$$

for every $z > 0$ where $v'(z)$ exists.

We have that $v(\cdot)$ is differentiable at z iff $\lambda(\cdot)$ is continuous at z, in which case $v'(z) = 2\lambda(z)$. Moreover, since $\lambda(\cdot)$ is monotonically nonincreasing on $(0, +\infty)$, it follows that for $z > 0$,

$$v(z) = \int_{-z}^z \lambda(t)dt = 2\int_0^z \lambda(t)dt \geq 2z\lambda(z).$$

Consequently, (6) holds, and hence the proof is complete. $\square$

Proof of Theorem 1. By Proposition 1 it suffices to consider optimization problem (1) with respect to even (symmetric) uniform distributions only. Together with Lemma 2 this implies that the assertion of Theorem 1 holds in case $g(\cdot) := \mathbb{1}_S(\cdot)$, where S is a convex symmetric set.

Now, since $g(\cdot)$ is bounded, we can assume without loss of generality that $0 \leq g(w) \leq c$ for some $c > 0$ and every $w \in \Xi$. We have that if $R := \times_{i=1}^n [-b_i, b_i]$ is a symmetric rectangular subset of $\mathbb{R}^n$ and μ is the uniform distribution on R, then

$$\mathbb{E}_\mu[g(W)] = \kappa \int_R g(w)dw = \kappa \int_0^c \vartheta(t)dt$$

where $\kappa := \left(\prod_{i=1}^n (2b_i)\right)^{-1}$ and $\vartheta(t)$ is volume of the set $S_t := \{w \in R : g(w) \geq t\}$. Since $g(\cdot)$ is quasi-convex and even, we have that for every t the set S_t is convex and symmetric. It remains to note that $\kappa\vartheta(t) = \mu(S_t)$ where μ is the uniform measure on the set R, and to apply the asserted result for the indicator function $\mathbb{1}_{S_t}(\cdot)$. $\square$

Note that Khintchine's family of distributions considered in Theorem 1 includes the corresponding family of even unimodal distributions. For the family $\mathcal{P} = \mathcal{F}_E$ of even unimodal distributions and $g(\cdot) := \mathbb{1}_S(\cdot)$, with S being a convex symmetric set, the result of Theorem 1 was first proved by Barmish and Lagoa [1], where it was called the "Uniformity Principle". The original proof in [1] is quite complicated. Simplified derivations of the Uniformity Principle are discussed in [7–9].

References

1. Barmish, B.R., Lagoa, C.M.: The uniform distribution: a rigorous justification for the use in robustness analysis. Math. Control, Signals, Systems **10**, 203–222 (1997)
2. Dupačová, J.: Minimax approach to stochastic linear programming and the moment problem. Optimierung, Sstochastik und mathematische Methoden der Wirtschaftswissenschaften **58**, 466–467 (1978)
3. Dupačová, J.: The minimax approach to stochastic programming and an illustrative application. Stochastics **20**, 73–88, (1987)
4. Erdoğan, E., Iyengar, G.: Ambiguous chance constrained problems and robust optimization, preprint. IOER Department, Columbia University, 2004
5. Gaivoronski, A.A.: A numerical method for solving stochastic programming problems with moment constraints on a distribution function. Ann. Oper. Res. **31**, 347–369 (1991)
6. Jagannathan, R.: Minimax procedure for a class of linear programs under uncertainty. Oper. Res. **25**, 173–177 (1997)
7. Kan, Yu.S.: On justification of the uniformity principle in the optimization problem for the probabilistic performance index. Avtomatika i Telemekhanika **1**, 54–70 (2000) [in Russian]
8. Kan, Yu.S., Kibzun, A.I.l: Sensitivity analysis of worst-case distribution for probability optimization problems. In: S.P. Uryasev (ed), Probabilistic Constrained Optimization: Theory and Applications. Kluwer, 2000, pp. 31–46
9. Kibzun, A.I.: On the worst-case distribution in stochastic optimization problems with probability function. Automation and Remote Control **59**, 1587–1597 (1998)
10. H.J. Landau (ed): Moments in mathematics. Proc. Sympos. Appl. Math., 37, Providence, RI: Amer. Math. Soc., 1987
11. Schneider, R.: Convex Bodies: The Brunn-Minkowski Inequality. Cambridge: Cambridge University Press, 1993
12. Shapiro, A., Kleywegt, A.: Minimax analysis of stochastic programs. Optimization Meth. Software **17**, 523–542 (2002)
13. Žáčková, J.; On minimax solutions of stochastic linear programming problems. Čas. Pěst. Mat. **91**, 423–430 (1966)