STOCHASTIC LINEAR PROGRAMS WITH RECOURSE AND ARBITRARY MULTIVARIATE DISTRIBUTIONS, SLP

Stochastic linear programs (SLPs) can be seen as a generalization of linear programming problems where at least some coefficients in the objective function and/or the constraints are random. The motivation for such a formulation is that in many practical applications, the problem data are not known with certainty, for instance because they represent information about the future.

As an example, consider a problem statement from the area of production planning. Today (at a first stage), a decision maker has to decide upon an input plan x which yields goods Tx by means of some technological process in order to meet the uncertain demand h in the future (a second stage). Since it is likely that the number of produced goods fails to meet the demand, a recourse action y is required that allows to compensate the discrepancy " $h - Tx$ " as soon as the demand is known. Such a correction induces additional cost $q'y$ in the second stage. The objective is to find a decision x that minimizes the direct cost $c'x$ of the first stage plus the expected cost $Q(x)$ induced by x for compensations. Formally, this can be written as

$$\begin{aligned} \min \quad & c'x + Q(x) \\ \text{s. t.} \quad & Ax = b \\ & x \geq 0 \end{aligned} \quad (1)$$

where $Q(x) = \int_{\Xi} Q(x, \xi) dF(\xi)$ is the expectation of the *recourse function*

$$Q(x, \xi) = \min_y \{q'y : Wy = h(\xi) - T(\xi)x\}. \quad (2)$$

The second-stage cost (2) depends on x and the realization of some K -dimensional random vector ξ with values in the compact convex set $\Xi \subset \mathbf{R}^K$ and distribution function F . Note that

recourse function

nonanticipativity

fixed recourse

simple recourse

stochastic linear program with recourse

deterministic equivalent problem

decomposition

approximation techniques

$Q(x, \xi) : \Xi \rightarrow \mathbf{R}$ is convex for all x satisfying the constraints in (1). While the recourse action y may be different for each ξ , the first-stage decision x is independent of which event actually occurs. This property is known as *nonanticipativity*. The meaning is that the current decision is only based on what is known today.

Throughout this article, it is assumed that the cost coefficients q as well as the recourse matrix W are fixed since the penalization for compensating the discrepancy $h(\xi) - T(\xi)x$ is likely to be of a deterministic nature. The case of a non-random W is known as *fixed recourse*. In particular, $W = (-I, I)$ with I being the identical matrix is called *simple recourse*, meaning that each deviation of $h(\xi) - T(\xi)x$ from zero is penalized by its absolute value. However, in a more general formulation both q and W may also depend on the realization of ξ (see [1], [12] for a more thorough discussion). Furthermore, it is assumed that the demand h and the technology matrix T are linear affine in ξ , i.e.

$$\begin{aligned} h(\xi) &= h_0 + h_1\xi_1 + \dots + h_K\xi_K \\ T(\xi) &= T_0 + T_1\xi_1 + \dots + T_K\xi_K. \end{aligned}$$

If the distribution of ξ is discrete, the *stochastic linear program with recourse* given by (1) and (2) can be written as a large deterministic problem where the expectations are written as a finite sum, and all constraints are duplicated for each realization of ξ . The resulting *deterministic equivalent problem* may be solved by straightforward application of standard linear programming methods (provided that the discrete set of possible outcomes for ξ is of relatively low cardinality). However, it exhibits a typical block structure that can be exploited by special *decomposition* algorithms (see also [1], [12] or [13]).

Otherwise, if the distribution of ξ is continuous, one can use *approximation techniques* ([2], [10], [11]) where the original random vector ξ is replaced by another one $\hat{\xi}$. Typically, $\hat{\xi}$

is discrete, and the problem reduces to a discretely distributed stochastic program. These techniques take advantage of the convexity of the recourse function, yielding upper and lower bounds for the expected recourse cost. This allows to quantify the accuracy of the approximation and, if not sufficient, to improve it by constructing a better approximation to $Q(x)$.

The Jensen lower bound. In order to outline the basic concepts (see e.g. [12]), assume that ξ is one-dimensional and denote the expectation by $\bar{\xi} := E\xi = \int_{\Xi} \xi dF(\xi)$. Recall that $\varphi(\xi) := Q(\hat{x}, \xi)$ is convex in ξ for any fixed $\hat{x}$. Therefore, it can be bounded from below by a linear function $\psi(\xi)$ that supports φ at some point $\hat{\xi}$, i.e. (assuming that φ is (sub)differentiable in $\hat{\xi}$)

$$\psi(\xi) = \varphi(\hat{\xi}) + \varphi'(\hat{\xi})(\xi - \hat{\xi}).$$

Due to linearity, the expected value of this lower bound is given by

$$E\psi(\xi) = \varphi(\hat{\xi}) + \varphi'(\hat{\xi})(E\xi - \hat{\xi}) = \psi(\bar{\xi}).$$

Obviously, the best lower bound is given by $\hat{\xi} = \bar{\xi}$ since no linear function supporting φ has a value larger than $\varphi(\bar{\xi})$ in $\bar{\xi}$. This is stated by *Jensen's inequality* $\varphi(E\xi) \leq E\varphi(\xi)$ for convex functions of random variables. Applied to the situation where $\varphi(\xi)$ is convex over the domain of ξ , one obtains a lower bound to the recourse function from

$$\varphi(\bar{\xi}) \leq \int_{\Xi} \varphi(\xi) dF(\xi). \quad (3)$$

Note that (3) also holds for multidimensional $\xi \in \Xi \subset R^K$, $K \geq 1$, regardless of any correlation between the components of ξ .

Jensen's inequality
Edmundson-Madansky upper bound

For one-dimensional random variables, the lower bounding function $\psi(\xi)$ is illustrated in the picture above, together with an upper bound which can be derived as follows.

The Edmundson-Madansky upper bound (independent case). For simplicity, the one-dimensional case is considered first. Let the support of ξ be given by the interval $\Xi = [a_0, a_1] \subset \mathbf{R}$. The idea of the *Edmundson-Madansky upper bound* is to introduce a discrete random variable $\hat{\xi}$ with the same expectation, i.e. $E\hat{\xi} = \bar{\xi}$, attaining the values a_0 and a_1 with probabilities

$$p_{a_0} := P(\hat{\xi} = a_0) = \frac{a_1 - \bar{\xi}}{a_1 - a_0}$$

$$p_{a_1} := P(\hat{\xi} = a_1) = \frac{\bar{\xi} - a_0}{a_1 - a_0} = (-1) \frac{a_0 - \bar{\xi}}{a_1 - a_0}$$

Obviously, the linear function $\Psi(\xi)$ through the points $(a_0, \varphi(a_0))$ and $(a_1, \varphi(a_1))$ is above $\varphi(\xi)$ for all $\xi \in \Xi$ due to convexity. This implies

$$\varphi(\xi) \leq \frac{a_1 - \xi}{a_1 - a_0} \cdot \varphi(a_0) + \frac{\xi - a_0}{a_1 - a_0} \cdot \varphi(a_1)$$

for all $\xi \in \Xi$, and integrating this inequality yields

$$\begin{aligned} & \int_{\Xi} \varphi(\xi) dF(\xi) \\ & \leq \frac{a_1 - \bar{\xi}}{a_1 - a_0} \cdot \varphi(a_0) + \frac{\bar{\xi} - a_0}{a_1 - a_0} \cdot \varphi(a_1) \\ & = p_{a_0} \cdot \varphi(a_0) + p_{a_1} \cdot \varphi(a_1) = E\varphi(\hat{\xi}) \end{aligned}$$

as an upper bound for the expectation $E\varphi(\xi)$. The EM-bound on intervals can be extended easily to multivariate distributions, where $\Xi = [a_{10}, a_{11}] \times \dots \times [a_{K0}, a_{K1}] \subset \mathbf{R}^K$ is the rectangular support of ξ , if either $Q(x, \cdot)$ is separable in the components of ξ or the elements of ξ are stochastically independent. In the former case, the bound may be applied to each component separately. Here, the more general latter case is of interest.

Denote the vertices of Ξ by a^ν , $\nu = (\nu_1, \dots, \nu_K)$, $\nu_i \in \{0, 1\}$, such that $a_i^\nu = a_{i\nu_i}$. Analogously to the above, $\hat{\xi}$ is a discrete random vector with independent components and

$E\hat{\xi}_i = \bar{\xi}_i$, attaining values a^ν with probability

$$p(a^\nu) := P(\hat{\xi} = a^\nu) = \frac{\prod_{i=1}^K (-1)^{\nu_i} (a_{i\bar{\nu}_i} - \bar{\xi}_i)}{\prod_{i=1}^K (a_{i1} - a_{i0})}, \quad (4)$$

where $\bar{\nu}_i = 1 - \nu_i$, and the EM-inequality contains the product of all combinations of each interval bound, i.e.

$$\int_{\Xi} \varphi(\xi) dF(\xi) \leq \sum_{\nu} p(a^\nu) \cdot \varphi(a^\nu) = E\varphi(\hat{\xi}).$$

The Edmundson-Madansky upper bound (dependent case). If the components of ξ are dependent, the EM-bound is more difficult to evaluate (unfortunately, the notation is also somewhat cumbersome). For $\mathcal{B} := \{\Lambda : \Lambda \subset \{1, \dots, K\}\}$ and $\delta_\Lambda(\nu) := \prod_{i \in \Lambda} (-1)^{\bar{\nu}_i}$, $\Lambda \in \mathcal{B}$,

$$m_\Lambda := \int_{\Xi} \left(\prod_{i \in \Lambda} \xi_i \right) dF(\xi)$$

are the crossmoments of ξ for any $\Lambda \in \mathcal{B}$, and $\rho_\Lambda := m_\Lambda - \prod_{i \in \Lambda} \bar{\xi}_i$. Using these definitions and $\bar{\Lambda} = \{1, \dots, K\} \setminus \Lambda$, it has been shown in [6], that the probabilities of the discrete outcomes a^ν are determined by

$$p(a^\nu) = \prod_{i=1}^K (a_{i1} - a_{i0})^{-1} \left[\prod_{i=1}^K (-1)^{\nu_i} (a_{i\bar{\nu}_i} - \bar{\xi}_i) + \sum_{\Lambda \in \mathcal{B}} (\delta_{\bar{\Lambda}}(\bar{\nu}) \prod_{i \in \bar{\Lambda}} a_{i\bar{\nu}_i}) \delta_\Lambda(\nu) \rho_\Lambda \right]$$

Obviously, $\rho_\Lambda = 0$ if the components of ξ are independent which is equivalent to the bound in (4). The calculation of the EM-bound requires to evaluate 2^K points and, in the dependent case, the same number of crossmoments. Therefore, for higher dimensions not only the computational effort but also the number of discrete outcomes increases exponentially, yielding a deterministic equivalent problem that might be too large to be solved.

Bounds on simplices. Instead of rectangles, one can use a regular simplex $\Delta \supset \Xi$ containing the support of ξ . In this case, Jensen's inequality where Ξ is replaced by Δ in (3) can be applied immediately to obtain a lower bound.

To derive an upper bound, the affine independent vertices $v_0, v_1, \dots, v_K$ of Δ are considered as discrete outcomes. Since this are only $K+1$ points, the complexity is no longer exponential in the dimension of ξ . Note that for any $\xi \in \Delta$, the system of linear equations

$$\begin{aligned} p_0(\xi) + p_1(\xi) + \dots + p_K(\xi) &= 1 \\ v_0 p_0(\xi) + v_1 p_1(\xi) + \dots + v_K p_K(\xi) &= \xi \end{aligned}$$

or briefly

$$Vp(\xi) = \begin{pmatrix} 1 \\ \xi \end{pmatrix} \text{ with } V = \begin{pmatrix} 1 & 1 & \dots & 1 \\ v_0 & v_1 & \dots & v_K \end{pmatrix}$$

has the unique solution

$$p(\xi) = V^{-1} \begin{pmatrix} 1 \\ \xi \end{pmatrix}.$$

Analogously to the above, a discrete random variable $\hat{\xi}$ is constructed attaining values $v_0, v_1, \dots, v_K$ with probabilities $p_0, p_1, \dots, p_K$ so that the expectation of the discrete distribution is equivalent to those of the original one, i.e. $E\hat{\xi} = p_0 v_0 + p_1 v_1 + \dots + p_K v_K = \bar{\xi}$. This yields the following version of the Edmundson-Madansky inequality:

$$\int_{\Delta} \varphi(\xi) dF(\xi) \leq \sum_{i=0}^K p_i \varphi(v_i),$$

where the vector of probabilities is given by

$$p = \int_{\Delta} V^{-1} \begin{pmatrix} 1 \\ \xi \end{pmatrix} dF(\xi) = V^{-1} \begin{pmatrix} 1 \\ \bar{\xi} \end{pmatrix}$$

Note that this equation holds for both the independent and the dependent case.

Improving bounds. The advantageous feature of approximation techniques is that the accuracy can be quantified by the difference between the Jensen and the Edmundson-Madansky bound. If not sufficient, the approximation can be improved by dividing the rectangular support Ξ (or the simplex Δ , respectively) into convex disjunct subsets like in the following picture.

A finite collection of such subsets is called a *partition*. Dividing an element of an existing partition yields a “refined” partition, and the associated bound is at least as good as the former one (monotonicity of bounds). If the subsets become arbitrary small, the approximated recourse function converges to the original one. However, for computational reasons the partition cannot become arbitrarily fine. Also, dividing Ξ (or Δ , respectively) without strategy may increase the computational effort dramatically. Hence, sophisticated refinement strategies are required that analyze the accuracy of the bounds for each subset and determine those that should be further divided. This yields a sequential approximation to the original problem.

Generalizations and alternative methods. When deriving the upper and lower bounds, only randomness in the right-hand side was taken into account so far. If the latter are deterministic (i.e. $h(\xi) := h_0$) but the coefficients $q(\xi)$ in the objective are uncertain, one can apply the same procedure to the dual problem. Since this is a maximization problem with concave recourse function, the Jensen inequality provides an upper while the Edmundson-Madansky inequality yields a lower bound. For uncertainty in both the objective and the right-hand side of constraints, extensions of the approximation scheme described above are required (see e.g. [7]). It has to be mentioned that there are other concepts apart from the bounds derived here which are also applicable for non-compact support of the random data [4] or derive sharper lower bounds by exploiting second moment information [3].

Alternatively, one may approximate the original problem by sampling from continuous distributions to obtain a deterministic equivalent, or one may use *Bender's decomposition* together with variance reduction techniques to handle a large number of scenarios [9]. Other sophisticated approaches like stochastic quasi-gradient methods [5] or stochastic decomposition [8] combine sampling with techniques known from convex optimization, for example subgradient or cutting plane procedures.

References

- [1] BIRGE, J.R., AND LOUVEAUX, F.: *Introduction to stochastic programming*, Springer, 1997.
- [2] BIRGE, J.R., AND WETS, R.J.-B.: ‘Designing approximation schemes for stochastic optimization problems, in particular for stochastic programs with recourse’, *Mathematical Programming Study* **27** (1986), 54–102.
- [3] EDIRISINGHE, N.C.P.: ‘New second-order bounds on the expectation of saddle functions with applications to stochastic linear programming’, *Operations Research* **44** (1996), 909–922.
- [4] EDIRISINGHE, N.C.P., AND ZIEMBA, W.T.: ‘Bounding the expectation of a saddle function with application to stochastic programming’, *Mathematics of Operations Research* **19** (1994), 314–340.
- [5] ERMOLIEV, Y.: ‘Stochastic quasigradient methods’, in Y. ERMOLIEV AND R.J.-B. WETS (eds.): *Numerical techniques for stochastic optimization*, Springer, 1988, pp. 141–185.
- [6] FRAUENDORFER, K.: ‘Solving SLP recourse problems with arbitrary multivariate distributions – the dependent case’, *Mathematics of Operations Research* **13** (1988), 377–394.
- [7] FRAUENDORFER, K.: *Stochastic two-stage programming*, Springer, 1992.
- [8] HIGLE, J.L., AND SEN, S.: *Stochastic decomposition – a statistical method for large scale stochastic linear programming*, Kluwer, 1996.
- [9] INFANGER, G.: *Planning under uncertainty*, Boyd & Fraser, 1994.
- [10] KALL, P.: ‘Bounds for and approximations to stochastic linear programs with recourse’, in K. MARTI AND P. KALL (eds.): *Stochastic programming methods and technical applications*, Springer, 1998, pp. 1–21.
- [11] KALL, P., RUSZCZYŃSKI, A., AND FRAUENDORFER, K.: ‘Approximation techniques in stochastic programming’, in Y. ERMOLIEV AND R.J.-B. WETS (eds.): *Numerical techniques for stochastic optimization*, Springer, 1988, pp. 33–64.

partition

Bender's decomposition

- [12] KALL, P., AND WALLACE, S.W.: *Stochastic programming*, Wiley, 1994.
- [13] RUSZCZYŃSKI, A.: ‘Decomposition methods in stochastic programming’, *Mathematical Programming* **79** (1997), 333–353.

Karl Frauendorfer

Institute of Operations Research
University of St. Gallen
St. Gallen, Switzerland

E-mail address: karl.frauendorfer@unisg.ch

Michael Schürle

Institute of Operations Research
University of St. Gallen
St. Gallen, Switzerland

E-mail address: michael.schuerle@unisg.ch

AMS1991 Subject Classification: 90C15.

Key words and phrases: stochastic programming, approximation, bounding the expectation.