Inequalities in Statistics and Probability
IMS Lecture Notes–Monograph Series Vol. 5 (1984), 178–186

APPROXIMATIONS AND ERROR BOUNDS IN STOCHASTIC PROGRAMMING

BY JOHN R. BIRGE and ROGER J.-B. WETS[1]
University of Michigan and University of Kentucky

We review and complete the approximation results for stochastic programs with recourse. Since this note is to serve as a preamble to the development of software for stochastic programming problelms, we also address the question of how to easily find a (starting) solution.

We consider the *stochastic program with (fixed) recourse* (Wets, 1983))

(0.1) find $x \in \mathcal{R}_+^{n_1}$ such that $Ax = b$ and $z = cx + \mathcal{Q}(x)$ is minimized

where A is $m_1 \times n_1$, $b \in \mathcal{R}^{m_1}$, and

(0.2) $$\mathcal{Q}(x) = E\{Q(x,\xi)\} = \int Q(x,\xi)\, P(d\xi)$$

with P a probability measure defined on $\Xi \subset \mathcal{R}^{n_2}$, and

(0.3) $$Q(x,\xi) = \inf_{y\in\mathcal{R}_+^{n_2}}\{qy \mid Wy = \xi - Tx\},$$

W is $m_2 \times n_2$, T is $m_2 \times n_1$, $q \in \mathcal{R}^{n_2}$ and $\xi \in \mathcal{R}^{n_2}$. We think of Ξ as the set of possible values of a random vector. Technically this means that Ξ is the support of the probability measure P. We shall assume that $\bar{\xi} = E\{\xi\}$ exists.

Many properties are known about problems of this type (Wets (1983)). For our purposes, the most important ones are

(0.4) $\xi \mapsto Q(x,\xi)$ is a convex piecewise linear function for all feasible x, i.e. $x \in K = K_1 \cap K_2$

where

$$K_1 = \{x \mid Ax = b,\ x \geq 0\}$$

$$K_2 = \{x \mid \text{for every } \xi \in \Xi, \text{ there exists a } y \geq 0 \text{ such that } Wy = \xi - Tx\},$$

and

(0.5) $x \mapsto Q(x,\xi)$ is a convex piecewise linear function which imples that

(0.6) $x \mapsto \mathcal{Q}(x)$ is a convex function, finite on K_2 (as follows from the integrability condition on Ξ).

It is also useful to consider an equivalent formulation of (0.1) that stresses the fact that choosing x corresponds to generating a *tender* $\chi = Tx$ to be bid by the decision maker against the outcomes ξ of the random events, viz.

(0.7) find $x \in \mathcal{R}_+^{n_1}$, $\chi \in \mathcal{R}^{m_2}$ such that $Ax = b$, $Tx = \chi$, and $z = cx + \psi(\chi)$ is minimized,

where

(0.8) $\Psi(\chi) = \mathrm{E}\{\psi(\chi,\xi)\} = \int \psi(\chi,\xi)\, \mathrm{P}(d\xi)$,

and

(0.9) $\psi(\chi,\xi) = \inf_{y\in\mathcal{R}_+^{n_2}}\{qy \mid Wy = \xi - \chi\}$.

---

1 Supported in part by a grant of the National Science Foundation.
AMS 1980 subject classifications. 90C15.
Key words and phrases: Stochastic programming, error bounds, recourse problem.

The functions ψ and Ψ have basically the same properties as Q and $\mathcal{Q}$, replacing naturally K_2 by the set

$$L_2 = \{\chi \mid \text{for every } \xi \in \Xi, \text{ there exists a } y \geq 0 \text{ such that } Wy = \xi - \chi\}.$$

Let z^* denote the optimal value of (0.1) or equivalently (0.7). We are interested in finding bounds on z^* by approximating $\mathcal{Q}$ or Ψ.

1. Lower Bounds. A lower bound for z^* can be obtained by solving the linear program

(1.1) find $x \geq 0, y \geq 0$ such that $Ax = b$, $Tx + Wy = \bar{\xi}$, and $cx + qy = z$ is minimized.

To see this note that (1.1) can also be expressed as

(1.2) find $x \in \mathcal{R}_+^{n_1}$ such that $Ax = b$ and $z = cx + Q(x,\bar{\xi})$ is minimized,

and with $\bar{z}$ denoting the optimal value of (1.2). We certainly have that $\bar{z} \leq z^*$ if we show that

$$Q(\cdot,\bar{\xi}) \leq \mathcal{Q}(\cdot). \tag{1.3}$$

But this follows from (0.4) and Jensens' inequality:

$$Q(x, \mathrm{E}\xi) \leq \mathrm{E}\{Q(x,\xi)\} \tag{1.4}$$

for every $x \in K_2$. There is another way to obtain this inequality, relying on the dual solution to (1.1):

(1.5) find $\sigma \in \mathcal{R}^{m_1}$, $\pi \in \mathcal{R}^{m_2}$ such that $\sigma A + \pi T \leq c$, $\pi W \leq q$, $\sigma b + \pi\bar{\xi} = w$ is maximized.

Let $(\bar{\sigma}, \bar{\pi})$ be an optimal solution to this linear program. Since $\bar{\pi} W \leq q$, it follows again from the duality theory of linear programming that

$$Q(x,\xi) = \sup_{\pi \in \mathcal{R}^{m_2}} \{\pi(\xi - Tx) \mid \pi W \leq q\} \geq \bar{\pi}(\xi - Tx)$$

and also that, for $x \in K$,

$$cx + \mathcal{Q}(x) \geq cx + \bar{\pi}\cdot E\xi - \bar{\pi}Tx = \bar{\pi}\bar{\xi} + (c - \bar{\pi}T)x$$
$$= \bar{\pi}\bar{\xi} + \bar{\sigma}Ax = \bar{\pi}\bar{\xi} + \bar{\sigma}b = w_{\text{opt}} = \bar{z}.$$

Hence

$$\bar{z} \leq \inf_{x \in K} cx + \mathcal{Q}(x) = z^*. \tag{1.6}$$

Madansky (1960) was the first to point out that this type of reasoning provided error bounds for stochastic programs. We can refine this lower bound in a number of ways.

The first one is to use a sharper version of Jensens' inequality. Let $\mathcal{S}^\nu = \{\Xi_l, l = 1, \dots, \nu\}$ be a partition of Ξ and let us denote by $\bar{\xi}^l$ the conditional expectation of ξ given that its values are in Ξ_l, i.e., $\bar{\xi}^l = E\{\xi | \xi \in \Xi_l\}$. Also, let $f_l = P(\Xi_l)$, i.e., f_l is the probability that $\xi \in \Xi_l$. The convexity of $Q(x, \cdot)$ yields

$$Q(x, \bar{\xi}) \leq \Sigma_{l=1}^{\nu} f_l Q(x, \bar{\xi}^l) \leq E\{Q(x, \xi)\} = \mathcal{Q}(x) \tag{1.7}$$

as follows from a generalization of Jensens' inequality (Perlman (1974)). Denote by $\bar{z}^\nu$ the optimal value of the linear program:

(1.8) find $x \geq 0$ such that $Ax = b$, $Tx + Wy^l = \bar{\xi}^l$, $l = 1, \dots, \nu$, and $cx + \Sigma_{l=1}^{\nu} f_l q y^l = z$ is minimized,

which can also be written in the form

(1.9) find $x \in \mathcal{R}_+^{n_1}$ such that $Ax = b$ and $z = cx + \Sigma_{l=1}^{\nu} f_l Q(x, \bar{\xi}^l)$ is minimized.

In view of (1.7), it follows that

$$\bar{z} \leq \bar{z}^\nu \leq z^*. \tag{1.10}$$

The same reasoning shows that if $\mathcal{S}^{\nu'} = \{\Xi_k, k = 1, \ldots, \nu'\}$ is a finer partition of Ξ, i.e., for all $k = 1, \ldots, \nu'$, $\Xi_k \subset \Xi_l$, for some $\Xi_l \in \mathcal{S}^\nu$, and if $\bar{z}^{\nu'}$ is the optimal value of the linear program of type (1.8) that corresponds to this partition, then

(1.11) $$\bar{z} \leqslant \bar{z}^\nu \leqslant \bar{z}^{\nu'} \leqslant z^*.$$

In fact the $\bar{z}^\nu$ converge to z^* provided that the partitions $\mathcal{S}^\nu$ are such that the probability measures they generate, viz.

$$P^\nu(A) = \Sigma_{\{l|\bar{\xi}^l \in A\}} P(\Xi_l),$$

converge in distribution to P, as follows from Theorem (3.9) of Wets (1983). The suggestion to rely on conditional expectations to refine (1.6) is due to Kall (1974) and to Huang, Ziemba and Ben-Tal (1977) who give a detailed analysis of these bounds when Ψ is separable.

Another method is to proceed as follows: For every $\xi \in \Xi$, and some $\xi \in$ co Ξ (the convex hull of Ξ), we define

(1.12) $$\phi(\hat{\xi},\xi) = \inf cx + \hat{p}q\hat{y} + (1-\hat{p})qy_\xi \text{ such that } Ax = b, Tx + W\hat{y} = \hat{\xi},$$
$$Tx + Wy_\xi = \xi, x \geqslant 0, \hat{y} \geqslant 0, y_\xi \geqslant 0$$

with $\hat{p} \in [0,1]$. If (0.1) is solvable, so is (1.12) for all $\xi \in \Xi$ as follows directly from Walkup and Wets (1967). Let x° solve (0.1) and for all ξ

$$y^\circ(\xi) \in \operatorname{argmin}_{y \in \mathcal{R}_+^{m_2}} \{qy | Wy = \xi - Tx^\circ\}.$$

It is well known that the $y^\circ(\xi)$ can be chosen so that as a function of ξ, $y^\circ(\cdot)$ is measurable, cf. Walkup and Wets (1967). Now let $\hat{\xi} = \bar{\xi}$ and $\bar{y}^\circ = E\{y^\circ(\xi)\}$. The triple $(x^\circ, \bar{y}^\circ, y^\circ(\xi))$ is a feasible solution of the linear program (1.12) when $\hat{\xi} = \bar{\xi}$. However in general it is not an optimal solution.

Whence

(1.13) $$\phi(\bar{\xi},\xi) \leqslant cx^\circ + \hat{p}q\bar{y}^\circ + (1-\hat{p})qy^\circ(\xi)$$

and integrating this on both sides with respect to P we obtain

(1.14) $$E\{\phi(\bar{\xi},\xi)\} \leqslant cx^\circ + \mathcal{Q}(x^\circ) = z^*$$

which gives us a new lower bound for z^*. This bound can be refined in many ways: first instead of using just one point $\bar{\xi}$ we can use a collection of points obtained as conditional expectations of a partition of Ξ. Second we can increase the number of points that are taken to build (1.12) as an approximation to (0.1). A detailed discussion appears in Birge (1982).

A lower bound of a somewhat different nature still using the convexity of Q, but not based on Jensens' inequality per se, can be obtained as follows. Let $\{\xi^l, l = 1, \ldots, \nu\}$ be a collection of points in Ξ and let

$$\pi^l \in \operatorname{argmax} [\pi(\xi^l - Tx) \mid \pi W \leqslant q].$$

Then $\pi^l \in \partial_\xi Q(x, \xi^l)$, i.e. the subgradient of Q with respect to ξ at ξ^l (for given x). We have that $Q(x,\xi^l) = \pi^l(\xi^l - Tx)$ and

(1.15) $$Q(x,\xi) \geqslant \pi^l(\xi - Tx) \text{ for all } \xi \in \Xi.$$

The last inequality follows from the simple observation that

$$Q(x,\xi) = \sup [\pi(\xi - Tx) \mid \pi W \leqslant q]$$

and that π^l is a feasible, but not necessarily optimal, solution for the sup-problem defining Q. Since (1.15) holds for every l, we have

$$Q(x,\xi) \geqslant \max_{1 \leqslant l \leqslant \nu} \pi^l(\xi - Tx).$$

Integrating on both sides yields

(1.16) $$\mathcal{Q}(x) \geq E\{\max_{1 \leq l \leq \nu} \pi^l(\xi - Tx)\}.$$

In general finding the maximum for each ξ may be difficult. But we may assign each π^l to a subregion of Ξ; this bound is not as tight as (1.16) but we can refine it by taking successively finer and finer partitions. However one should not forget that (1.16) involves a rather simple integral and the expression to the right could be evaluated numerically (to an acceptable degree of accuracy) without major difficultiies. Note that the calculation of this lower bound does not require the ξ^l to be conditional expectations or chosen in any specific manner, however it should be obvious that a well chosen spread of the $\{\xi^l, l = 1, \ldots, \nu\}$ will give us sharper bounds. Also, the use of larger samples, i.e. by increasing ν, will also yield a better lower bound.

2. Upper Bounds. If $\mathcal{Q}(x)$ is easily computable, a simple upper bound is given by $z^* \leq c\hat{x} + \mathcal{Q}(\hat{x})$ for any feasible $\hat{x}$ in K. In particular, if $\bar{x}$ solves (1.1) and it turns out that $\bar{x} \in K$, then we have that

(2.1) $$\bar{z} = c\bar{x} + Q(\bar{x},\bar{\xi}) \leq z^* \leq c\bar{x} + \mathcal{Q}(\bar{x}).$$

In general we cannot infer that $\bar{x} \in K$ simply from knowing that $\bar{x}$ solves (1.1), unless we know that we are dealing with a stochastic program with complete recourse, or more generally with relatively complete recourse, Wets (1983), i.e., when $K = \{x | Ax = b, x \geq 0\}$. Refinements of this bound, relying on different values of x may be found in Kall (1979) and Birge (1980) but they always involve the evaluation of $\mathcal{Q}(x)$.

Without evaluating $\mathcal{Q}$, we may find upper bounds for $\mathcal{Q}$ by considering the extreme points of coΞ. Let us assume in what follows that Ξ is compact, then so is its convex hull and $\Xi = \text{co}(\text{ext } \Xi)$ where ext Ξ are the extreme points of Ξ. Since $Q(x,\xi)$ is convex in ξ, we have that for all $\xi \in \Xi$

$$\begin{aligned} Q(x,\xi) &\leq \sup_{\xi \in \Xi} Q(x,\xi), \\ &= Q(x,e^{(x)}), \quad \text{for some } e^{(x)} \in \text{ext } \Xi. \\ &= \max_{e \in \text{ext } \Xi} Q(x,e). \end{aligned}$$

Now $e^{(x)}$ may depend on x, but we always have that

(2.2) $$\mathcal{Q}(x) \leq \max_{e \in \text{ext } \Xi} Q(x,e) = Q(x,e^{(x)}),$$

and hence

(2.3) $$z^* \leq \inf_{x \in K}[cx + (\max_{e \in \text{ext } \Xi} Q(x,e))].$$

If there are only a finite number of extreme points of Ξ, as is usually the case in practice, the function appearing on the right hand side of the inequality can be minimized without major difficulties. Let $\{e^j, j = 1, \ldots, J\} = \text{ext } \Xi$ be this finite collection of extreme points. We have to solve the mathematical program

(2.4) find $x \in \mathcal{R}_+^{n_1}$ and $\theta \in \mathcal{R}$ such that $Ax = b$, $Q(x,e^j) \leq \theta$ for $j = 1, \ldots, J$ and $cx + \theta$ is minimized.

The last condition can also be expressed as

$$\theta \geq qy^j, \; Wy^j = e^j - Tx, \; y^j \geq 0 \quad \text{for } j = 1, \ldots, J.$$

Thus (2.4) becomes equivalent to the linear program

(2.5) find $x \in \mathcal{R}_+^{n_1}$, $\theta \in R$ and ($y^j \in \mathcal{R}_+^{n_2}$, $j = 1, \ldots, J$) such that $Ax = b$, $Tx + Wy^j = e^j$, $\theta \geq qy^j$ for $j = 1, \ldots, J$ and $cx + \theta$ is minimized.

The optimal value yields the upper bound for z^*.

This is a very crude bound. We can improve on this, as follows: every $\xi \in \Xi$ also belongs to co(ext Ξ). We can thus find $\{\lambda_j(\xi), j = 1, \ldots, J\}$ such that $\lambda_j(\xi) \geq 0$, $\Sigma_{j=1}^J \lambda_j(\xi) = 1$, and $\Sigma_{j=1}^J \lambda_j(\xi)\, e^j = \xi$. We write $\lambda_j(\xi)$ to indicate the dependence of the λ_j on ξ. By convexity of $Q(x,\cdot)$,

$$Q(x,\xi) \leq \Sigma_{j=1}^J \lambda_j(\xi)\, Q(x,e^j).$$

Taking the expectation on both sides yields

$$\begin{aligned} \mathcal{Q}(x) &\leq \int_\Xi \Sigma_{j=1}^J \lambda_j(\xi)\, Q(x,e^j)\, P(d\xi) \\ &= \int_\Lambda \Sigma_{j=1}^J \lambda_j Q(x,e^j)\, G(d\lambda) \end{aligned} \tag{2.6}$$

where G is the distribution function induced by P on $\Lambda = \{\lambda \in \mathcal{R}^J | \Sigma_{j=1}^J \lambda_j = 1, \lambda_j \geq 0\}$.

If coΞ is a simplex, then each $\xi \in \Xi$ is obtained by a unique convex combination of the extreme points. It is not difficult to actually derive G, calculate the last integral and then minimize the resulting function to obtain an upper bound for z^*. In general Ξ is not a simplex. We shall see later what to do in the general case, but there is an important class of problems that reduces to the case where Ξ is a simplex.

Suppose the random variables (of the m_2 vector) are independent. Then the distribution function (or the probability measure) is separable and (2.6) can be written as

$$\begin{aligned} \mathcal{Q}(x) &= \int_{\alpha_{m_2}}^{\beta_{m_2}} P_{m_2}(d\xi_{m_2}) \ldots \int_{\alpha_2}^{\beta_2} P_2(d\xi_2) \int_{\alpha_1}^{\beta_1} P_1(d\xi_1)\, Q(x, (\xi_1,\xi_2, \ldots, \xi_{m_2})) \\ &\leq \int_{\alpha_{m_2}}^{\beta_{m_2}} P_{m_2}(d\xi_{m_2}) \ldots \int_{\alpha_2}^{\beta_2} P_2(d\xi_2) \int_0^1 G_1(d\lambda_1)\, Q^1(x, (\lambda_1,\xi_2, \ldots, \xi_{m_2})) \end{aligned} \tag{2.7}$$

where

$$Q^1(x, (\lambda_1,\xi_2, \ldots, \xi_{m_2})) = (1-\lambda_1)Q(x, (\alpha_1,\xi_2, \ldots, \xi_{m_2})) + \lambda_1 Q(x, (\beta_1,\xi_2, \ldots, \xi_{m_2}))$$

and for each i, $\Xi_i = [\alpha_i,\beta_i]$ and $\Xi = \times_{i=1}^{m_2} \Xi_i$. Since $\xi_1 = (1 - \lambda_1)\, \alpha_1 + \lambda\beta_1$ we get the following expression for $\lambda_1(\xi_1)$:

$$\lambda_1 = (\xi_1 - \alpha_1)/(\beta_1 - \alpha_1), \quad \text{and} \quad 1 - \lambda_1 = (\beta_1 - \xi_1)/(\beta_1 - \alpha_1).$$

Hence, with $\mu_1 = E\{\xi_1\}$,

$$\begin{aligned} &\int_0^1 Q^1(x, (\lambda_1, \xi_2, \ldots, \xi_m))\, G_1(d\lambda_1) = \\ &((\beta_1-\mu_1)/(\beta_1-\alpha_1))\, Q(x, (\alpha_1, \xi_2, \ldots, \xi_{m_2})) + ((\mu_1-\alpha_1)/(\beta_1-\alpha_1))\, Q(x, (\beta_1,\xi_2, \ldots, \xi_{m_2})) \end{aligned} \tag{2.8}$$

which we can substitute in (2.7) for the integral with respect to λ_1. We can repeat this process for each ξ_i to obtain a bound on $\mathcal{Q}$ involving only the evaluation of the function $Q(x,\cdot)$ at the vertices of the retangular region Ξ.

The whole argument really boils down to the use of the simple inequality for real-valued convex functions ϕ of a random variable ξ, with distribution P on $[\alpha,\beta]$ and expectation μ.

$$\int_\alpha^\beta \phi(\xi)P(d\xi) \leq ((\beta-\mu)/(\beta-\alpha))\phi(\alpha) + ((\mu-\alpha)/(\beta-\alpha))\phi(\beta) \tag{2.9}$$

This inequality is due to Edmundson. Madansky (1960) used it in the context of stochastic programs (with simple recourse) to obtain a simple version of (2.7). A much refined version of this upper bound can be obtained by partitioning the interval $[\alpha,\beta]$ and using (2.9) for each interval in the partition, substituting the end points of the subinterval for α and β, and the conditional expectation (with respect to this subinterval) for μ. In the case of stochastic programs with simple recourse this was carried out by Huang, Ziemba and Ben-Tal (1977) and by Kall and Stoyan (1982) who also consider stochastic problems of a more general nature.

Also, when P is not separable we can improve somewhat on (2.3) by observing that we can use (2.9) with respect to one random variable, say ξ_1. We have

$$\int Q(x,\xi)\,P(d\xi_1,\xi_2,\ldots,\xi_{m_2}) \leq \sup_{\{(\xi_2,\ldots,\xi_{m_2})\mid \xi\in\Xi\}} \int Q(x,\xi)\,P(d\xi_1,\xi_2,\ldots,\xi_{m_2})$$
$$= \sup_{\{(\bar e^1,\ldots,\bar e^{\xi})\mid e^j=(e_1^j,\bar e^j)\in \operatorname{ext}\Xi\}}[((\beta_1-\mu_1(e^j))/(\beta_1-\alpha_1))\,Q(x,(\alpha_1,\bar e^j))$$
$$+((\mu_1(e^j)-\alpha_1)/(\beta_1-\alpha_1))\,Q(x,(\beta_1,\bar e j)]$$

where $\mu_1(e^j)$ is the conditional expectation of ξ_1 given $\bar e^j$ (the last (m_2-1) coordinates of e^j). From this it follows that

$$\text{(2.10)}\qquad \mathcal{Q}(x) \leq \min_{1\leq i\leq m_2} \sup_{\{\bar e^j \mid e^j\in \operatorname{ext}\Xi\}}[((\beta_1-\mu_1(e^j))/(\beta_1-\alpha_1))\,Q(x,(\alpha_i,e^{-j}))$$
$$+((\mu_i(e^j)-\alpha_i)/(\beta_i-\alpha_i))\,Q(x,(\beta_i,\bar e^j))],$$

where it must be understood that $\bar e^j$ consists of the (m_2-1) components of e^j that are not indexed by i. Further refinements through the partitioning of Ξ and the use of the corresponding conditional means, tighten up this inequality.

Another refinement of (2.3), in the case of nonseparable measure P, can be obtained by considering simplicial decompositions of Ξ, assuming naturally that Ξ admits such a decomposition (which means that Ξ should be polyhedral). Let $\mathcal{S} = \{\mathcal{S}^l, l = 1, \ldots, L\}$ be such a decomposition (technically $\mathcal{S}$ is a complex whose cells $\mathcal{S}^l$ are simplices). Let $\{e_0^l, \ldots, e_{m_2}^l\}$ be the vertices of the simplex $\mathcal{S}^l$, assuming that $\dim \Xi = m_2$. Then each $\xi \in \mathcal{S}^l$ determines a unique vector of barycentric coordinates $(\lambda_0^l, \ldots, \lambda_{m_2}^l)$ such that

$$\xi = \Sigma_{j=0}^{m_2}\lambda_j^l e_j^l,\ \lambda_j^l \geq 0,\ \Sigma_0^{m_2}\lambda_j^l = 1.$$

On $\mathcal{S}^l$, we are thus given a simple formula for the relationship between the distribution of ξ and the induced distribution for the λ_j^ls. We have

$$\int_{\mathcal{S}^l} Q(x,\xi)\,P(d\xi) \leq \int_\Lambda \Sigma_{j=0}^{m_2}\lambda_j Q(x,e_j^l)\,G_l(d\lambda) = \tilde Q_l(x,\mathcal{S}^l)$$

where $\Lambda = \{\lambda \in \mathcal{R}^{m_2+1} \mid \Sigma_{j=0}^{m_2}\lambda_j^l = 1,\ \lambda_j^l \geq 0\}$ and G_l is the measure induced by the preceding transformation. If we assume that the measure P is absolutely continuous (with respect to the Lebesgue measure on $\mathcal{R}^{m_2}$), then P assigns zero measure to every face (of dimension less than m_2) of the simplices $\mathcal{S}^l$ and hence

$$\text{(2.11)}\qquad \mathcal{Q}(x) = \Sigma_l \int_{\mathcal{S}^l} Q(x,\xi)\,P(d\xi) \leq \Sigma_l \tilde Q_l(x,\mathcal{S}^l).$$

This new upper bound can again be refined in two ways, first by considering finer simplicial decompositions, and second by considering for every ξ the smallest upper bound given by a number of possible simplicial representations. We sketch this out.

Suppose Ξ is a convex polytope (of dimension m_2) and $\{v^1, \ldots, v^J\}$ is a finite collection of points in Ξ that includes the extreme points of Ξ. Let $\mathcal{P}$ be the set of all (m_2+1) subsets of $\{v^1, \ldots, v^J\}$ such that $\operatorname{co}(v^{j_0}, \ldots, v^{j_{m_2}})$ is a m_2-simplex. The convexity of $Q(x,\cdot)$ yields

$$Q(x,\xi) \leq \Sigma_{i=0}^{m_2}\lambda_{j_i} Q(x,v^{j_i})$$

where

$$\Sigma_{i=0}^{m_2}\lambda_{j_i}v^{j_i} = \xi,\ \Sigma_{i=0}^{m_2}\lambda_{j_i} = 1,\ \lambda_{j_i} \geq 0,$$

i.e. $\xi \in \operatorname{co}(v^{j_0}, \ldots, v^{j_{m_2}})$. With $\mathcal{P}(\xi)$ denoting the elements of $\mathcal{P}$ that have ξ in their convex hull,

$$Q(x,\xi) \leq \inf_{\{(v^{j_0},\ldots,v^{j_{m_2}})\in\mathcal{P}(\xi)\mid \Sigma_{i=0}^{m_2}\lambda_{j_i}v^{j_i}=\xi\}} \Sigma_{i=0}^{m_2}\lambda_{j_i} Q(x,v^{j_i}).$$

Each element of $\mathcal{P}(\xi)$ induces a measure on Λ, we can integrate on both sides to obtain an upper bound on $\mathcal{Q}$ and thus also on z^*.

3. Getting a Starting Solution. The inequalities, and thus the resulting error bounds, presented above depend upon the chosen sample points of Ξ or the partitioning scheme used. Choices for initial samples can be based on the solutions of simplified problems in which the constraints have been relaxed. It is convenient to use here the version (0.7)–(0.8)–(0.9) of the original problem. We shall assume that we are dealing with stochastic programs with relatively complete recourse ($K = K_1$). In terms of (0.7) this means that if $x \in K_1$ and $\chi = Tx$, then $\chi = L_2$, cf. the expression for L_2 following (0.9).

Suppose χ^0 is a guess at the optimal tender, i.e. as part of a pair (x^0, χ^0) solving (0.7). Cost considerations might lead us to such a choice, but there is no guarantee that χ^0 is actually part of a feasible pair for problem (0.7), that we repeat here for convenience sake:

(0.7) find $x \in \mathcal{R}_+^{n_2}$, $\chi \in \mathcal{R}^{m_2}$ such that $Ax = b$, $Tx = \chi$, and $z = cx + \Psi(\chi)$ is minimized.

To obtain a feasible solution we might solve the linear program (with $h^+ \geq 0, h^- \geq 0$)

(3.1) find $x \in \mathcal{R}_+^{n_1}$, $u^+ \in \mathcal{R}_+^{m_2}$, $u^- \in \mathcal{R}_-^{m_2}$ such that $Ax = b$, $Tx + u^+ + u^- = \chi^0$, and $z = cx + h^+u^+ - h^-u^-$ is minimized.

We can use the resulting solution to start the optimization algorithm. In the case of simple recourse, a suitable choice of h^+ and h^- may be the vectors q^+ and q^- that determine the recourse costs. Recall that for stochastic programs with simple recourse, the function ψ as defined by (0.9) is given by $\psi(\chi, \xi) = \sum_{i=1}^{m_2} \psi_i(\chi_i, \xi_i)$ and

$$\psi_i(\chi_i, \xi_i) = \inf\{q_i^+ y_i^+ + q_i^- y_i^- \mid y_i^+ - y_i^- = \xi_i - \chi_i, y_i^+ \geq 0, y_i^- \geq 0\},$$
$$= \begin{cases} q_i^+(\xi_i - \chi_i) & \text{if } \chi_i \leq \xi_i, \\ q_i^-(\chi_i - \xi_i) & \text{if } \chi_i \geq \xi_i. \end{cases}$$

In this situation, we could proceed as follows: for every $i = 1, \ldots, m_2$, solve the single constraint stochastic program

(3.2) find $x \in \mathcal{R}_+^{n_1}$, $\chi_i \in \mathcal{R}$ such that $T_i x = \chi_i$, and $z_i = cx + \Psi_i(\chi_i)$ is minimized.

Here T_i is the i-th row of T and

$$\Psi_i(\chi_i) = E\{\Psi_i(\chi_i, \xi_i)\}.$$

This problem is equivalent to

(3.3) find $x \in \mathcal{R}_+^{n_1}$, $\chi_i \in \mathcal{R}$ such that $\chi_i = T_i x$, and

$$z_i = cx + \int_{\xi_i \geq \chi_i} q_i^+(\xi_i - \chi_i) F_i(d\xi_i) + \int_{\xi_i \leq \chi_i} q_i^-(\chi_i - \xi_i) F_i(d\xi_i)$$

with F_i denoting the marginal distribution function of ξ_i. The optimal solution of (3.2) is the pair (x^0, χ_i^0) such that

$$x_j^0 \geq 0 \text{ for } j = 1, \ldots, n,$$
$$\sum_{j=1}^{n_1} t_{ij} x_j^0 = \chi_i^0,$$
$$\theta \in -\partial\psi_i(\chi_i^0) = [q_i^+ - q_i F_i^+(\chi_i^0), q_i^+ - q_i F_i \chi_i^0)],$$
$$c_j - \theta t_{ij} \geq 0 \quad \text{for } j = 1, \ldots, n,$$
$$(c_j - \theta t_{ij}) x_j = 0 \quad \text{for } j = 1, \ldots, n,$$

where $q_i = q_i^+ + q_i^-$, $F_i(z) = P[\xi_i < z]$, and $F_i^+(z) = P[\xi_i \leq z]$.

In order to simplify the presentation, we make the following assumptions:

(i) F_i is strictly continuously increasing on its support,

(ii) $T_i \geq 0$,

(iii) $\inf_j c_j t_{ij}^{-1} \in [-q_i^-, q_i^+]$.

The last assumption is only introduced to render the problem nontrivial. Without such a condition the problem is either unbounded or of a type that has no practical interest. With this, we have

$$\theta = \inf_j(c_j/t_{ij}) = c_s/t_{is},$$
$$\chi_i = F_i^{-1}((q_i^+ - c_s/t_{is})/q_i).$$

This method gives us a starting vector χ^0, which we can then use to generate a feasible pair $(\hat{x}, \hat{\chi})$, as indicated at the beginning of this section. Some justification for this choice comes from the fact that we are solving for each i the problem "optimally". This boils down to finding the solution to a newsboy problem (having more than one supply source). For a detailed study of this class of problems, when viewed as simple stochastic programs, consult Wets (1974).

If we are not dealing with simple recourse we may still proceed in a very similar manner. For each i, the problem to be solved is

(3.4) find $x \in \mathcal{R}_+^{n_1}$, $\chi_i \in \mathcal{R}$ such that $T_i x = \chi_i$ and $cx + \int_{\Xi_i} \inf\{qy | W_i y = \xi_i - \chi_i\} dP_i(\xi_i)$ is minimized.

Here again P_i is the marginal distribution of ξ_i and $\Xi_i \subset \mathcal{R}$ its support. We note that the integrand above is

$$\int_{\xi_i < \chi_i} (q_j/w_{ij})_{\min}(\xi_i - \chi_i) dP_i(\xi_i) + \int_{\xi_i \geq \chi_i} (q_j/w_{ij})_{\max}(\xi_i - \chi_i) dP_i(\xi_i),$$

assuming here that

$$(q_j/w_{ij})_{\min} = \inf_{1 \leq j \leq n_2}(q_j/w_{ij}),$$
$$(q_j/w_{ij})_{\max} = \sup_{1 \leq j \leq n_2}(q_j/w_{ij}),$$

and the coefficients w_{ij} appearing in $(q_j/w_{ij})_{\min}$ and $(q_j/w_{ij})_{\max}$ are negative and positive respectively. The infimum in (3.4) then occurs at a point such that

$$0 = (\partial(cx)/\partial\chi_i) - ((q_j/w_{ij})_{\max} + ((q_j/w_{ij})_{\min})F(\chi_i) + (q_j/w_{ij})_{\max}$$

If we restrict χ_i to $\chi_i = t_{ij}x_j$ for fixed j,

$$\chi_i^0 \in \operatorname{argmin}_j[(c_j/t_{ij})\chi_{ij} + \int_{\Xi_i} \inf\{qy | W_i y = \xi_i - \chi_{ij}\} dP_i(\xi_i)]$$

where

$$\chi_{ij} = F^{-1}(((q_j/w_{ij})_{\max} - (c_j/t_{ij}))/((q_j/w_{ij})_{\max} + (q_j/w_{ij})_{\min})).$$

Again this leads us to a vector χ^0. The intuitive justification for the use of this vector is the same as in the case of stochastic programs with simple recourse.

After the intial choice of χ^0, other values of χ may be chosen by minimizing the expected error in approximating the function $\Psi(\chi)$, by using an a priori distribution on χ. As new χ values are found in an optimization procedure, this distribution may be changed using Bayesian updates; in the case of simple recourse the expected error is easily measurable since $\psi(\chi)$ can be evaluated precisely on each subregion.

REFERENCES

BIRGE, J. (1980). Solution methods for stochastic dynamic linear programs, *Technical Report No. 80–29*, Systems Optimization Laboratory, Stanford University, Stanford, CA.

BIRGE, J. (1982). The value of the stochastic solution in stochastic linear programs with fixed recourse. *Mathematical Programming 24* 314–325.

HUANG, C., ZIEMBA, W., and BEN-TAL, A. (1977). Bounds on the expectation of a convex function of a random variable: with applications to stochastic programming. *Operations Research 25* 315–325.

KALL, P. (1974). Approximations to stochastic programming with complete fixed recourse. *Num. Math 22* 333–339.

KALL, P. (1979). Computational methods for solving two-stage stochastic linear programming problems. *ZAMP 30* 261–271.

KALL, P. and STOYAN, D. (1982). Solving stochastic programming problems with recourse, including error bounds. *Math. Operationsforschung Statist. Ser. Optimization 13* 431–447.

MANDANSKY, A. (1960). Inequalities for stochastic linear programming problems. *Management Science 6* 197–204.

PERLMAN, M. (1974). Jensen's inequality for a convex vector-valued function on an infinite dimensional space. *J. Multivariate Analysis 4* 52–65.

WALKUP, D. and WETS, R. (1967). Stochastic programs with recourse. *SIAM J. Applied Math. 15* 1299–1314.

WETS, R. (1974). *Stochastic programming,* Lecture Notes.

WETS, R. (1983). Stochastic programming: approximation schemes and solution techniques, in *Mathematical Programming 1982: The State-of-the-Art*. Springer-Verlag, Berlin and New York.