GENERATING UPPER BOUNDS ON THE EXPECTED VALUE OF A CONVEX FUNCTION WITH APPLICATIONS TO STOCHASTIC PROGRAMMING

John R. Birge
Industrial and Operations Engineering
The University of Michigan
Ann Arbor, MI 48109

Roger J-B Wets
Mathematics
University of California
Davis, CA 95616

Technical Report 85-14

GENERATING UPPER BOUNDS ON THE EXPECTED VALUE OF A CONVEX FUNCTION WITH APPLICATIONS TO STOCHASTIC PROGRAMMING

John R. Birge*
Industrial and Operations Engineering
The University of Michigan
Ann Arbor, MI 48109

Roger J-B. Wets*
Mathematics
University of California
Davis, CA 95616

ABSTRACT

Bounds on the expected value of a convex function are obtained by means of an approximating generalized moment problem. Numerical implementation is discussed in the context of stochastic programming problems.

KEYWORDS: Generalized moment problems, stochastic programming, inequalities in probability.

*This research has been partially supported by the National Science Foundation.

1. INTRODUCTION: THE GENERALIZED MOMENT PROBLEM

The (generalized) moment problem:

find P: $\mathcal{A}$ ---> [0,1] such that (1.1)

$\int v_i(\xi)\ P(d\xi) \leq \beta_i,\quad i = 1,\ldots,s$

$\int v_i(\xi)\ P(d\xi) = \beta_i,\quad i = s+1,\ldots,m,$

and $z = \int v_0(\xi)\ P(d\xi)$ is maximized

with $\mathcal{A}$ the sigma-field of events defined on Ξ --here a subset of R^N-- is of general interest in statistics, in stochastic optimization, etc. The underlying premise is that some information is available about certain moments, or generalized moments, of an unknown probability distribution. This determines a class $\mathcal{P}$, i.e., the probability measures that satisfy the constraints of (1.1). This limited information is to be used in order to obtain an upper (or/and lower) bound on some other moment of the distributions in this class. This problem has been studied in detail in the classical framework of statistical theory, see for example [9] where the accent is placed on those problems of type (1.1) that can be solved analytically (in the framework provided by Chebyshev systems). The connection between the generalized moment problem and optimization theory has been clarified by Kemperman [8], but so far the computational tools provided by linear or nonlinear programming have only been used sparingly in the development of general solution procedures for generalized moment problems. This paper is one contribution in that direction.

Dupačová [2.3] was the first to rely on some (classical) results for the moment problem to obtain bounds for the optimal value of certain stochastic programming problems; the class $\mathcal{P}$ is usually determined by first--or possibly second order--moments and v_0 is a 1-dimensional convex or concave nondifferentiable function. In such cases it is possible to obtain an explicit characterization of the extremal measure that solves the corresponding moment problem. This is not the case in general. Although it is known that if the support of the probability measures Ξ is convex compact, then the linear programming problem (1.1) admits as optimal solution, an extremal measure whose support is concentrated on at most m+1 points of Ξ (see [1, Theorem 6.9] for a constructive proof of this result). There is usually no closed form expression that allows us to easily identify this extremal measure. There are no conceptual difficulties in designing solution procedures for solving (1.1), see [1, Section 6], however some of the operations that must be carried out may be in practice extremely onerous. Much depends on the properties of the functions v_0 and v_i, $i = 1,\dots,m$. In this paper we shall be mostly concerned with the case when v_0 is convex, and the functions v_i, $i = 1,\dots,m$ are linear or piecewise linear. We extend and sharpen the results of Gassmann and Ziemba [4] who were first in developing an implementable procedure for the restricted, but important, case when v_0 is convex and there is just one constraint on the expectation (with respect to P).

2. LINEAR CONSTRAINTS

Most of the literature devoted to generalized moment problems works with the assumption that Ξ is compact [7], [1, Section 6]. Here we drop this assumption. It is possible to do so by relying on an appropriate compactification of R^n, viz. by adding to R^n the space of directions of recession.

Let us suppose that $\Xi \subset R^N$ is a nonempty convex set, if not we would work with co Ξ the convex hull of Ξ. Let

$\text{rc } \Xi :=$ cone of the directions of recession of Ξ.

This corresponds to the largest (closed) convex cone such that

$\xi + (\text{rc } \Xi) \subset \Xi$ for all ξ in Ξ;

it is called the recession cone of Ξ. It is nonempty, in fact $0 \in \text{rc } \Xi$, since by assumption Ξ is nonempty. For more about recession cones, consult [10, Section 8]. Let

$\text{ext } \Xi :=$ extreme points of Ξ.

If Ξ has no extreme points, set $\text{ext } \Xi = \{0\}$, and let

$\text{ext-rc}\,\Xi :=$ extreme directions of $\text{rc-}\Xi$

by which we mean a collection of points in R^N such that

$\text{rc}\,\Xi = \text{pos } (\text{ext-rc}\,\Xi)$,

(where pos T denotes the positive hull of the points in $T \subset R^N$, i.e.

$$\text{pos } T := \{ \sum_{k=1}^{q} \mu_k t^k \mid \mu_k \geq 0,\ t^k \in T,\ \ q \text{ finite}\}),$$

and no element of ext-rc Ξ can be obtained as a positive linear combination of other elements of Ξ. The elements of ext-rc Ξ are positively linearly independent. Since Ξ is convex, we have

$$\Xi = \text{co}(\text{ext}\,\Xi) + \text{pos } (\text{ext-rc}\,\Xi).$$

Every point ξ in Ξ has at least one representation of the form:

$$\xi = \int_{\text{ext } \Xi} e\lambda(\xi, de) + \int_{\text{ext-rc } \Xi} r\mu(\xi, dr) \qquad (2.1)$$

with $\lambda(\xi,\cdot)$ a probability measure on $(\text{ext } \Xi, \mathcal{E})$ and $\mu(\xi, \cdot)$ a nonnegative measure on $(\text{ext-rc } \Xi, \mathcal{R})$; $\mathcal{E}$ and $\mathcal{R}$ are the Borel fields. In fact, the theorems of Carathéodory and Steinitz guarantee that for each ξ there exists one representation involving no more than N+1 points of ext Ξ and 2N points of ext-rc Ξ, i.e. the measures $\lambda(\xi, \cdot)$ and $\mu(\xi, \cdot)$ have then finite support.

Now, suppose v_0: Ξ ---> $\bar{R}$ is a proper convex function, and consider ξ as as given by (2.1) with

$$\hat{\xi} := \int_{\text{ext-rc}\Xi} r\mu(\xi, dr),$$

then

$$\xi = \int_{\text{ext }\Xi} (e + \hat{\xi})\, \lambda\, (\xi, de), \tag{2.2}$$

and

$$v_0(\xi) \leq \int_{\text{ext}\Xi} v_0(e + \hat{\xi})\, \lambda\, (\xi, de). \tag{2.3}$$

By

$$\text{rc } v_0(\zeta) := \lim_{t \to \infty} \frac{v_0(\xi+t\zeta) - v_0(\xi)}{t} = \sup_{t>0} \frac{v_0(\xi+t\zeta) - v_0(\xi)}{t} \tag{2.4}$$

we denote the <u>recession function</u> of v_0 in the direction ζ [10, Section 8]. This is a sublinear function (positively homogeneous and convex). Since v_0 is convex, the ratio

$$\lambda^{-1}(v_0(\xi + \zeta) - v_0(\xi))$$

is a monotone nondecreasing function of λ (when $\lambda>0$) and thus the supremum is "attained" at $\lambda = \infty$. This justifies the equivalence of the two formulas in (2.4). It also means that for all $e \in \text{ext } \Xi$,

$$v_0(e + \hat{\xi}) \leq v_0(e) + \text{rc } v_0(\hat{\xi})$$

and hence (2.3) yields

$$v_0(\xi) \leq \int_{\text{ext}\Xi} v_0(e)\, \lambda\, (\xi, de) \quad + \text{ rc } v_0(\hat{\xi}). \tag{2.5}$$

Since rc v_0 is a sublinear function, this implies that

$$v_0(\xi) \leq \int_{\text{ext}\Xi} v_0(e)\, \lambda(\xi, de) + \int_{\text{rc-ext}\Xi} \text{rc } v_0(r)\, \mu(\xi, dr). \tag{2.6}$$

Integrating on both sides with respect to P, we obtain

$$\int v_0(\xi)\, P(d\xi) \leq \int_{\text{ext}\Xi} v_0(e)\, \lambda(de) + \int_{\text{rc-ext}\Xi} \text{rc } v_0(r)\, \mu(dr) \tag{2.7}$$

where

$$\lambda(\cdot) = \int_{\Xi} \lambda(\xi, \cdot)\ P(d\xi) \tag{2.8}$$

and

$$\mu(\cdot) = \int_{\Xi} \mu(\xi, \cdot)\ P(d\xi), \tag{2.9}$$

(assuming that λ and μ have been chosen measurable with respect to ξ).

Moreover

$$\bar{\xi} := \int \xi\ P(d\xi) = \int_{\text{ext}\,\Xi} e\lambda(de) + \int_{\text{ext-rc}\,\Xi} r\mu(dr), \tag{2.10}$$

$$1 = \int \lambda(de) = \int_{\Xi} P(d\xi) \int_{\text{ext}\Xi} \lambda(\xi, de) = \int 1 P(d\xi) \tag{2.11}$$

and both λ(defined on E) and μ(defined on R) are nonnegative.

The inequality (2.7) holds for any nonnegative measures λ and μ that satisfy (2.8)-(2.11) and (2.1). Given the bound provided by (2.7) one may be tempted to ignore (2.8) and (2.9)--i.e., that it must be possible to "disintegrate" λ and μ in measures $\lambda(\xi, \cdot)$ and $\mu(\xi, \cdot)$ that satisfy (2.1) -- but this would render (2.7) invalid as an easy example will show readily. However, there always exists one pair (λ,μ) for which it is not necessary to verify if λ and μ can be "disintegrated" so as to satisfy (2.1): namely, if (λ,μ) maximizes the right-hand side of (2.7)! We have thus shown

2.1 THEOREM. Suppose $\Xi \subset R^N$ is a nonempty convex set, P is a probability measure on (Ξ, A), and v_0 is a proper (nowhere $-\infty$ and finite-valued somewhere) extended-real valued convex function defined on R^N. Then

$$\int v_0(\xi)\ P(d\xi) \le \sup_{(\lambda,\mu)} \int_{\text{ext}\Xi} v_0(e)\ \lambda(de) + \int_{\text{ext-rc}\Xi} \text{rc}\ v_0(r)\ \mu(dr) \tag{2.12}$$

where λ is a probability measure on (ext Ξ, E), and μ is a nonnegative measure on (rc ext Ξ, R) such that

$$\int_{\text{ext}\,\Xi} e\,\lambda(de) + \int_{\text{ext-rc}\,\Xi} r\mu(dr) = \xi\cdot P(d\xi) = \bar{\xi}. \tag{2.13}$$

This yields an upper bound on the optimal value of (1.1) when the constraints are determined by the linear system $\int \xi\ P(d\xi) = \bar{\xi}$. It should be emphasized that (2.12) yields in some sense the worst possible bound for $\int v_0(\xi)\ P(d\xi)$ among all those generated by (2.7)-(2.11), but it may be the only one that is sufficiently easy to compute and to integrate into an approximation scheme for solving stochastic optimization problems. The next result goes even further in that direction. It sharpens and extends [4, Theorem 4] or [1, Section 5(v)] for the case of bounded convex polyhedral support.

2.2 COROLLARY. Suppose $\Xi \subset C = co(e^1,\dots,e^p) + pos(r^1,\dots,r^q)$ and h: $R^N \longrightarrow \bar{R}$ is a convex function such that $h \geq v_0$ on Ξ

Then

$$\int v_0(\xi)\ P(d\xi) \leq \sup_{(\lambda,\mu)}\Bigg[\sum_{j=1}^{p} \lambda_j h(e^j) + \sum_{t=1}^{q} \mu_i (rc\ h)(r^i) \,\Big|\, \lambda_j \geq 0,\ \mu_i \geq 0,$$

$$\sum_{j=1}^{p} \lambda_j e^j + \sum_{i=1}^{q} \mu_i r^i = \xi,\ \sum_{j=1}^{p} \lambda_j e^j = 1 \Bigg]. \tag{2.14}$$

PROOF. Note that

$$\int v_0(\xi)\ P(d\xi) \leq \int_\Xi h(\xi)\ P(d\xi) = \int_C h(\xi)\ P(d\xi)$$

where P has been (trivially) extended to C. It now suffices to apply Theorem 2.1. ■

All the results and remarks of this section apply equally well to the case when P is simply a bounded measure, in particular to the case when P is the restriction of some probability to a subset of the space of events, making of course the obvious adjustments. This simple observation yields directly the versions of Theorem 2.1 and Corollary 2.2 with conditional expectations.

3. PIECEWISE LINEAR CONSTRAINTS.

The extension to the case when the constraints are piecewise linear, or more precisely piecewise affine, is straightforward. Details are worked out in this section. The importance of this case, rests on the potential use of piecewise linear approximations for handling the (general) nonlinear case, see Section 4 for an elementary example.

Suppose Ξ is partitioned in L subregions Ξ_ℓ, $\ell = 1,\dots,L$, and in (1.1) the functions v_i, that define the constraints, are piecewise linear: for $i=1,\dots,m$,

$$v_i(\xi) = a_{i\ell} \cdot \xi - \alpha_{i\ell} \qquad \text{when} \qquad \xi \in \Xi_\ell. \tag{3.1}$$

We then have

$$\int v_i(\xi)\, P(d\xi) = \sum_{\ell=1}^{L} \int_{\Xi} (a_{i\ell} \cdot \xi - \alpha_{i\ell})\, P(d\xi) \tag{3.2}$$

$$= \sum_{\ell=1}^{L} (a_{i\ell} \cdot \bar{\xi}^{\ell} - \alpha_{i\ell})\, p_\ell \tag{3.3}$$

where

$$\bar{\xi}^{\ell} = E\{\xi \mid \xi \in \Xi_\ell\},$$

$$p_\ell = P(\Xi_\ell).$$

Thus, the (generalized) moment problem becomes:

find Q: $A \dashrightarrow [0,1]$, a probability measure, such that (3.4)

$$\bar{\xi}^{\ell} = E_Q\{\xi \mid \xi \in \Xi_\ell\},\quad p_\ell = Q(\Xi_\ell)$$

$$\sum_{\ell=1}^{L} (p_\ell\, a_{i\ell}) \cdot \bar{\xi}^{\ell} - \sum_{\ell=1}^{L} p_\ell \alpha_{i\ell} \leq \beta_i, \quad i=1,\dots,s,$$

$$\sum_{\ell=1}^{L} (p_\ell\, a_{i\ell}) \cdot \bar{\xi}^{\ell} - \sum_{\ell=1}^{L} p_\ell \alpha_{i\ell} = \beta_i, \quad i=s+1,\dots,m.$$

and $z = \int v_0(\xi)\, Q(d\xi)$ is maximized.

Our objective is to obtain an upper bound on the optimal value of this problem. Let P be any probability measure that satisfies the constraints of

(3.4); P could be the measure that we are trying to approximate.

Let C_ℓ, $\ell=1,\ldots,L$, be a collection of (nonempty) convex sets such that for all ℓ, $\Xi_\ell \subset C_\ell$. One possibility is to choose for all $\ell = 1,\ldots,L$,

C_ℓ = co Ξ: = the convex hull of Ξ .

Let $\lambda_\ell'(\xi,\cdot)$ be probability measures defined on (ext C_ℓ, E_ℓ) and $\mu_\ell'(\xi, \cdot)$ nonnegative measures on (ext-rc C_ℓ, R_ℓ) such that for all $\xi \in \Xi_\ell$

$$\xi = \int_{\text{ext } C_\ell} e\, \lambda_\ell' (\xi, de) + \int_{\text{ext-rc } C_\ell} r\mu_\ell' (\xi, dr). \tag{3.5}$$

Here E_ℓ (R_ℓ resp.) is the Borel field on ext C_ℓ(ext-rc C_ℓ resp.) and we assume that λ_ℓ' and μ_ℓ' are A-measurable with respect to ξ. For any $A \in E_\ell$, set

$$\lambda_\ell(A) := \int_{\Xi_\ell} \lambda_\ell'(\xi, A)\ P(d\xi). \tag{3.6}$$

and for any $B \in R_\ell$, set

$$\mu_\ell(B) := \int_{\Xi_\ell} \mu_\ell'(\xi, B)\ P(d\xi). \tag{3.7}$$

We have from (3.6):

$$\int_{\text{ext } C_\ell} \lambda_\ell(de) = \lambda_\ell(\text{ext } C_\ell) = P(\Xi_\ell) = p_\ell. \tag{3.8}$$

From (3.2), after replacing ξ by its representation (3.5), and interchanging the summands using the definitions of λ_ℓ and μ_ℓ, we obtain:

$$\int v_i(\xi)\ P(d\xi) = \Big(\sum_{\ell=1}^{L} \int_{\text{ext } C_\ell} a_{i\ell}\cdot e\lambda_\ell de + \int_{\text{ext-rc } C_\ell} a_{i\ell}\cdot r\, \mu_\ell(dr) - \alpha_\ell p_\ell \Big). \tag{3.9}$$

From the above, by the same arguments as in Section 2, in particular, by the convexity of v_0 on C_ℓ, we prove the following generalization of Theorem 2.1.

3.1 THEOREM. _Suppose_ $\Xi \subset R^N$ _is nonempty_, $\{\Xi_\ell, \ell=1,\ldots,L\}$ _a partition of_ Ξ, $\{C_\ell, \ell=1,\ldots,L\}$ _nonempty convex sets such that_ $\bar{\Xi}_\ell \subset C_\ell$ _for all_ ℓ, P _a probability measure_, v_0 _an extended real-valued proper convex function on_ R^N, and for $i = 1,\ldots,m$, v_i is a piecewise affine function on R^N with

$$v_i(\xi) = a_{i\ell}\cdot\xi - \alpha_{i\ell} \qquad \text{when } \xi \in \Xi_\ell.$$

Then

$$\int v_0(\xi)\, P(d\xi) \le \sup_{(\lambda_\ell,\mu_\ell)} \left[\sum_{\ell=1}^{L} \int_{\text{ext } C_\ell} v_0(e)\, \lambda_\ell(de) + \int_{\text{ext-rc } C_\ell} (\text{rc } v_0)(r)\mu_\ell(dr) \right] \tag{3.10}$$

where for $\ell=1,\ldots,L$, λ_ℓ and μ_ℓ are nonnegative measures on $(\text{ext } C_\ell, \mathcal{E}_\ell)$ and $(\text{ext-rc } C_\ell, R_\ell)$ respectively, such that $\lambda_\ell(\text{ext } C_\ell) = P(\Xi_\ell) =: p_\ell$ (3.11)

and for $i=1,\ldots,m$,

$$\sum_{\ell=1}^{L} \left(\int_{\text{ext } C_\ell} a_{i\ell}\cdot e\,\lambda_\ell(de) + \int_{\text{ext-rc } C_\ell} a_{i\ell}\cdot r\mu_\ell(dr) \right) = \sum_{\ell=1}^{L} (p_\ell a_{i\ell})\cdot\bar{\xi}_\ell. \tag{3.12}$$

Observe that no convexity conditions are necessary on the functions v_i, $i=1,\ldots,m$. Suppose, for example with $N=1$, that

$$a_{ik} = 1, \qquad a_{i\ell} = 0 \quad \text{if } \ell \neq k,$$

the i-th condition of (3.12) would require that the measures be chosen so that

$$\int_{\text{ext } C_k} e\lambda_k\,(de) + \int_{\text{ext-rc } C_k} r\,\mu_k(dr) = p_k\bar{\xi}_k.$$

This means that the measures on the extremal structure of C_k must satisfy this conditional expectation condition. Approximation schemes can be built by requiring that the chosen measures satisfy conditional expectation conditions that involve finer and finer partitions of Ξ. Kall and Stoyan [6] and Huang, Vertinsky, and Ziemba [5] have studied constraints of that type. The constraints (3.7), however, are much more general, in that they allow for tighter restrictions so that sharper bounds may be obtained. The approximations in [5], [6] rely on the extreme points of the Ξ_ℓ for all $\ell=1,\ldots,L$, and therefore require the evaluation of the v_i at each of these points. Here only the extreme points of some convex set containing Ξ are needed, indeed simply choose $\Xi_\ell = \text{co}\ \Xi$ for all $\ell=1,\ldots,L$. Restriction of C_ℓ to subsets of Ξ, however, still imposes additional restrictions on the set of feasible probability measures and can, therefore, improve the bounds.

There is also in the piecewise affine case a generalization of Corollary 2.2. We do not need v_0 convex, only that it be dominated by a convex function.

3.2 COROLLARY. *Suppose* $\Xi_\ell \subset C_\ell := \text{co}(e^1,\ldots,e^{r_\ell}) + \text{pos}(r^1,\ldots,r^{q_\ell})$, $\{\Xi_\ell,\ \ell=1,\ldots,L\}$ *is a partition of* Ξ, *and* h: $R^N \longrightarrow \bar{R}$ *is a proper convex function such that* $h \geq v_0$ *on* Ξ. *Then for any probability measure* $P: A \longrightarrow [0,1]$, *we have*

$$\int v_0(\xi)P(d\xi) \leq \sup_{(\lambda,\mu)} \left[\sum_{\ell=1}^{L} \left(\sum_{k=1}^{r_\ell} h(e^k)\lambda_{\ell k} + \sum_{k=1}^{q_\ell} (\text{rc } h)(r^k)\mu_{\ell k} \right) \right] \tag{3.13}$$

$$\text{such that} \quad \sum_{\ell=1}^{L} \left(\sum_{k=1}^{r_\ell} a_{i\ell} e^k \lambda_{\ell k} + \sum_{k=1}^{q_\ell} a_{i\ell} r^k \mu_{\ell k} \right) = \sum_{\ell=1}^{L} p_\ell a_{i\ell} \cdot \bar{\xi}^\ell \quad \text{for } i = 1,\ldots,m,$$

$$\sum_{k=1}^{r_\ell} \lambda_{\ell k} = p_\ell, \quad \ell = 1,\ldots,L$$

$$\lambda_{\ell k} \geq 0,\ \mu_{\ell k} \geq 0$$

where for $\ell=1,\ldots,L$, $p_\ell := P(\Xi_\ell)$ *and* $\bar{\xi}^\ell$ *is the conditional expectation (with respect to* P) *given that* $\xi \in \Xi_\ell$.

Of course to obtain this upper bound on $\int v_0(\xi)P(d\xi)$ it is not necessary to know P. It suffices to know the values of p_ℓ and $\bar{\xi}^\ell$. In fact we do not even need the individual values of the $\bar{\xi}^\ell$. It is only necessary to know for each $i = 1,\ldots,m$,

$$\alpha_i := \sum_{\ell=1}^{L} p_i\, a_{i\ell} \cdot \bar{\xi}^\ell.$$

More generally, suppose it is known that

$$\alpha_i^- \le \int v_i(\xi)\ P(d\xi) \le \alpha_i^+$$

but the precise value of this integral (with respect to P) is not known. In this case the constraint (3.12) could be replaced by

$$\alpha_i^- \le \sum_{\ell=1}^{L}\left(\int_{\text{ext } C_\ell} a_{i\ell}\cdot e\ \lambda_\ell(de) + \int_{\text{ext-rc } C_\ell} a_{i\ell}\cdot r\ \mu_\ell(dr)\right) \le \alpha_i^+ . \tag{3.14}$$

This is the case when the generalized moment problem involves inequalities and the v_i are piecewise linear. This yields immediately an extension of Theorem 3.1 (to the inequality case) that turns out to be quite useful when dealing with higher order moments approximations. If it is known for example, that

$$\int \xi^2\ P(d\xi) \le \alpha_i, \tag{3.15}$$

then by defining a piecewise affine (lower) approximation v_i such that $v_i(\xi) \le \xi^2$, the optimization problem (3.10) with the constraint (3.14) -- defining $\alpha_i^-: = 0$ and $\alpha_i^+ = \alpha_i$ -- yields an upper bound on all probability measures satisfying the second moment condition (3.15). The bound can be improved by refining the approximation v_i.

For other than piecewise linear functions, limited results are still available. Suppose for example v_i is concave and

$$\int v_i(\xi)\ P(d\xi) \le \alpha_i . \tag{3.16}$$

We can substitute for its representation (3.5) to obtain

$$\sum_{\ell=1}^{L}\left(\int_{\Xi_\ell} v_i\left(\int_{\text{ext } C_\ell} e\ \lambda'(\xi,de) + \int_{\text{ext-rc } C_\ell} r\ \mu'(\xi,\ dr)\right)\right)P(d\xi) \le \alpha_i .$$

Now if we use the concavity of v_i and the definitions (3.6) of λ_ℓ and (3.7) of μ_ℓ, we have

$$\sum_{\ell=1}^{L}\left(\int_{\text{ext } C_\ell} v_i(e)\lambda_\ell(de) + \int_{\text{ext-rc } C_\ell} rc\ v_i(r)\mu_\ell(dr)\right) \le \alpha_i . \tag{3.17}$$

Therefore we can replace the constraint (3.16) by (3.17) and again obtain an upper bound on the expectation of v_0 that satisfies (3.16). Obviously the same technique can also be used if v_i is convex and we know a lower bound on the expectation of v_i.

4. EXAMPLES.

The results of Section 3 can be used to obtain bounds for a variety of lower dimensional cases. In this section we consider two piecewise affine approximations of the second moment function, $v_i(\xi) = \xi^2$. We take ξ to be a 1-dimensional random variable distributed on the interval $[\beta_1, \beta_2]$ with $\beta_1 < 0 < \beta_2$. Let $P([\beta_1, 0]) = p_1$, $P((0, \beta_2]) = p_2$ and suppose $P(\{0\}) = 0$.

As upper approximate we define

$$v_i(\xi) = \begin{cases} -\gamma_1 \xi, & \xi \leq 0 \\ \gamma_2 \xi, & \xi \geq 0 \end{cases} \tag{4.1}$$

with $\gamma_1, \gamma_2 \geq 0$. If $\gamma_1 = \gamma_2$, then $v_i(\xi) = |\xi|$. In general we define γ_1 and γ_2 so as to best fit the case at hand, see Figure 1.

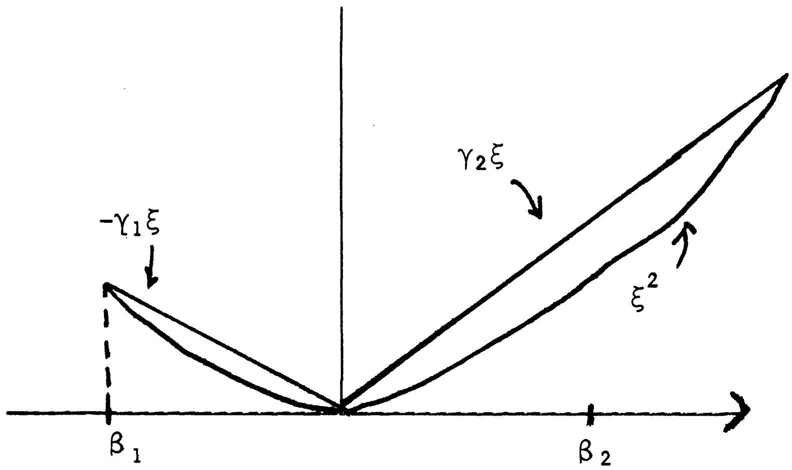

Figure 1. Upper "absolute value" approximate.

With $C = co(\beta_1 = e^1, \beta_2 = e^2)$, the linear program (3.13) becomes:

find $\lambda_{11} \geq 0$, $\lambda_{12} \geq 0$, $\lambda_{21} \geq 0$, $\lambda_{22} \geq 0$ such that (4.2)

$$\begin{aligned} -\gamma_1\beta_1\lambda_{11} - \gamma_1\beta_2\lambda_{12} + \gamma_2\beta_1\lambda_{21} + \gamma_2\beta_2\lambda_{22} &= \alpha, \\ \lambda_{11} + \lambda_{12} &= p_1, \\ \lambda_{21} + \lambda_{22} &= p_2, \end{aligned}$$

and $w = \sum_{\ell=1}^{2} \sum_{k=1}^{2} h(\beta_k)\lambda_{\ell k}$ is maximized.

We set

$$\alpha := \int_{\beta_1}^{\beta_2} v_i(\xi)\, P(d\xi) = -\gamma_1 \int_{\beta_1}^{0} \xi P(d\xi) + \gamma_2 \int_0^{\beta_2} \xi\, P(d\xi) = -\gamma_1 p_1 \bar{\xi}_1 + \gamma_2 p_2 \bar{\xi}_2$$

where $\bar{\xi}_1$ and $\bar{\xi}_2$ denote conditional expectation with respect to $[\beta_1, 0]$ and $(0, \beta_2]$ respectively. Thus we can view (4.2) as finding a probability that assigns weights to the extreme points of $[\beta_1, \beta_2]$, viz. $(\lambda_{11} + \lambda_{21})$ to β_1 and $(\lambda_{12} + \lambda_{22})$ to β_2 in such a way that a certain generalized conditional expectation condition is satisfied.

There are four possible bases -- each one corresponding to having 1 variable $\lambda_{\ell k}$ nonbasic -- depending on the values of the coefficients. The following table gives the solution values and optimality conditions when $\gamma_1 = \gamma_2 > 0$, and for 2 of the 4 possible bases (the other two have λ_{21} and λ_{22} nonbasic resp.):

Nonbasic Variable	λ_{11}	λ_{12}
Dual feasibility	$h(\beta_2) \geq h(\xi_1)$	$h(\beta_2) \geq h(\xi_1)$
Primal feasibility	$p_1 \leq p_2 \left(\frac{\beta_2 - \bar{\xi}_2}{\beta_2 - \bar{\xi}_1}\right)$	$p_1 \leq p_2 \left(\frac{\beta_2 - \bar{\xi}_2}{\beta_1 - \bar{\xi}_1}\right)$
Solution:	$\lambda_{12} = p_1$	$\lambda_{22} = p_2$
	$\lambda_{21} = p_2 - \lambda_{22}$	$\lambda_{12} = p_1 - \lambda_{11}$
	$\lambda_{22} = \frac{1}{\beta_2 - \beta_1}\left[[p_1(\beta_2 - \bar{\xi}_1) + p_2(\beta_2 - \bar{\xi}_1)]\right]$	$\lambda_{11} = \ldots$

Table 1: Optimality Conditions: $\gamma_1 = \gamma_2 > 0$

The dual feasibility conditions are the same for all γ_1 and γ_2 positive. The primal feasibility conditions depend on the relative sizes of the slopes.

We can also work with a lower approximate for $\xi \mapsto \xi^2$, for example with the cup-shaped function:

$$v_i(\xi) = \begin{cases} \gamma_1(\beta_1 - \xi) & \text{if} \quad \beta_1 \le \xi \le \beta_{11} \\ 0 & \text{if} \quad \beta_{11} \le \xi \le \beta_{21} \\ \gamma_2(\xi - \beta_{21}) & \text{if} \quad \beta_{21} \le \xi \le \beta_2. \end{cases}$$

where $\beta_1 < \beta_{11} < 0 < \beta_{21} < \beta_2$. See Figure 2.

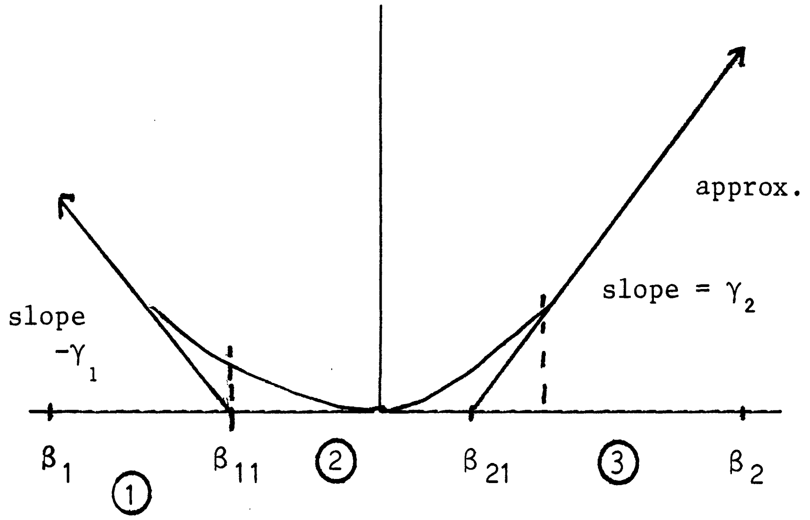

Figure 2. Cup-shaped approximation for ξ^2 .

If v_i is this cup-shaped function, the associated linear program (that yields an upper bound) reads:

$$\text{find } \lambda_{11} \ge 0, \quad \lambda_{12} \ge 0, \ \ldots, \ \lambda_{32} \ge 0 \text{ such that} \tag{4.3}$$

$$-\gamma_1\beta_1\lambda_{11} \ -\gamma_1\beta_2\lambda_{12} + \gamma_3\beta_1\lambda_{31} + \gamma_3\beta_2\lambda_{32} = \alpha \ .$$

$$\lambda_{\ell 1} + \lambda_{\ell 2} = p_\ell, \qquad \ell = 1,2,3$$

$$\text{and } w = \sum_{\ell=1}^{3} \sum_{k=1}^{2} h(\beta_k)\lambda_{\ell k} \text{ is maximized,}$$

where

$$\alpha := -p_1\gamma_1\bar{\xi}_1 + p_3\gamma_3\bar{\xi}_3.$$

There are 8 feasible bases, The solutions are of the same type as in Table 1.

The solution of (4.6) is useful in conjunction with the solution of (4.2). As we noted earlier, if $v_i(\xi) \leq \xi^2$ then the solution of (3.13) yields an upper bound on $\int v_0(\xi)P(d\xi)$. If, however, $v_i(\xi) \geq \xi^2$, then the solution of (3.13) yields a lower bound on the supremum over all distributions satisfying the second moment condition. The two bounds can be used to determine how well the second moment condition has been approximated.

5. APPLICATIONS IN STOCHASTIC PROGRAMMING

Bounds on the expectation of a convex function can be particularly useful in the solution of stochastic optimization problems. In these problems, it is often necessary to perform an optimization to evaluate a function at a single point. Taking the expectation of such functions presents a formidable task. By limiting the support of the distribution to a finite number of points, as in the development above, the set of problems to optimize is limited, and one can efficiently obtain bounds on the expectation.

A typical stochastic optimization problem has the following form:

$$\inf_{x \in K} \quad c(x) + \int_{\Xi} Q(x,\xi)\ P(d\xi), \tag{5.1}$$

where $Q(x,\xi)$ is itself the optimal value of a mathematical program whose coefficients depend on x and ξ. To solve (5.1), we need to evaluate $\int Q(x,\xi)$ $P(d\xi)$ for many values of x chosen in an optimization procedure. An upper bound on the integral that is easily calculable may lead to useful bounds on the optimal value of (5.1).

As an example, consider

$$Q(x,\xi) = \min \quad 5y_1 + 10y_2 + 10y_3 \tag{5.2}$$

$$\text{s.t.} \quad y_1 + y_2 = \xi_1 - x_1$$

$$y_1 + y_3 = \xi_2 - x_2,$$

$$y_1, y_2, y_3 \geqq 0.$$

Note that this problem has the values

$$Q(x,\xi) = \begin{cases} 10\xi_1 + 5x_2 - 5\xi_2 - 10x_1 & \xi_1 - x_1 \geq \xi_2 - x_2 \geq 0 \\ 10\xi_2 + 5x_1 - 5\xi_1 - 10x_2 & \xi_2 - x_2 \geq \xi_1 - x_1 \geq 0 \\ +\infty & \xi_1 - x_1 < 0 \text{ or } \xi_2 - x_2 < 0. \end{cases}$$

We wish to consider the case where x = (0,0) and $P[\xi_1 \leq \hat{\xi}_1, \xi_2 \leq \hat{\xi}_2] = \int_0^{\hat{\xi}_1} \int_0^{\hat{\xi}_2} 4e^{-2\xi_1} e^{-2\xi_2} d\xi_1 \, d\xi_2$, where $\hat{\xi}_1$ and $\hat{\xi}_2$ are any given nonnegative values of the random variables ξ_1 and ξ_2. In this way, ξ_1 and ξ_2 are independent exponential random variables with means 1/2.

We wish to investigate how to find an upper bound on

$$\int Q((0,0),\xi) \; P(d\xi) \tag{5.3}$$

without computing the integral exactly. We first note that the exact value is 6.25. The other methods we will try will be Gassmann and Ziemba's approach, a solution using Corollary 3.2 and not exploiting the independence of the random variables, and a solution using Corollary 3.2 and the independence of the random variables.

A lower bound on (5.3) is found using Jensen's Inequality as $Q((0,0), \bar{\xi}) = 2.5$. An upper bound is available using the Gassmann - Ziemba approach which reduces to (3.13) with $a_{i\ell} = e_\ell$ and L=n. Using the extreme directions of $\Xi = [0,\infty) \times [0,\infty)$ as (1,0) and (0,1), an extreme point of (0,0), and noting that rc Q((0,0), (1,0)) = rc Q((0,0), (0,1)) = 10, the following version of

(3.13) is obtained,

$$\int Q((0,0),\xi)\ P(d\xi) \leq \sup_{(\lambda,\mu)} \quad 10\,\mu_1 + 10\,\mu_2 \tag{5.4}$$

$$\begin{aligned} \text{s.t.} \quad & 0\lambda_1 + \mu_1 && = 1/2 \\ & 0\lambda_1 \qquad + \mu_2 && = 1/2 \\ & \lambda_1 && = 1 \\ & \lambda_1,\ \mu_1,\ \mu_2 \geq 0 . \end{aligned}$$

The only feasible solution to (5.4) is $\mu_1 = \mu_2 = 1/2$, $\lambda_1 = 1$, and the upper bound so obtained is 10. By using the independence of the random variables, we can see that, for any value $\hat{\xi}_2$ of ξ_2, $Q((0,0),(\xi_1, \hat{\xi}_2))$ defined on $[0,+\infty)$ has a value of $10\xi_2$ at (0,0) and a recession function value of 10 in the direction of increasing ξ_1. A Gassmann-Ziemba bound on the integral over ξ_1, is therefore, $10\hat{\xi}_2 + 5$ for any $\hat{\xi}_2$. Using this value, we can then solve the problem in ξ_2 and again obtain a value of 5 at (0,0) and a recession function of 10. The resulting upper bound is again 10, so the use of independence does not improve the bound.

To provide a more precise bound on (5.3), we include more constraints in (3.13). Note that $E(\xi_1^2) = E({\xi_2}^2) = 1/2$. A function $v_i\ (\xi_1, \xi_2)$ such that $v_i \leq {\xi_i}^2$ is

$$v_i(\xi) = \begin{cases} 0 & \text{if } \xi_i \leq 1/2 \\ 2\xi_i - 1 & \text{if } \xi_i \geq 1/2, \end{cases} \tag{5.5}$$

a cup-shaped function as in Figure 2. This can be used with bounds of 1/2 for i=1,2 to obtain additional constraints to mean value constraints. The region of Ξ must then be partitioned into Ξ^1, Ξ^2, Ξ^3, and Ξ^4 as in Figure 3. In (3.13),

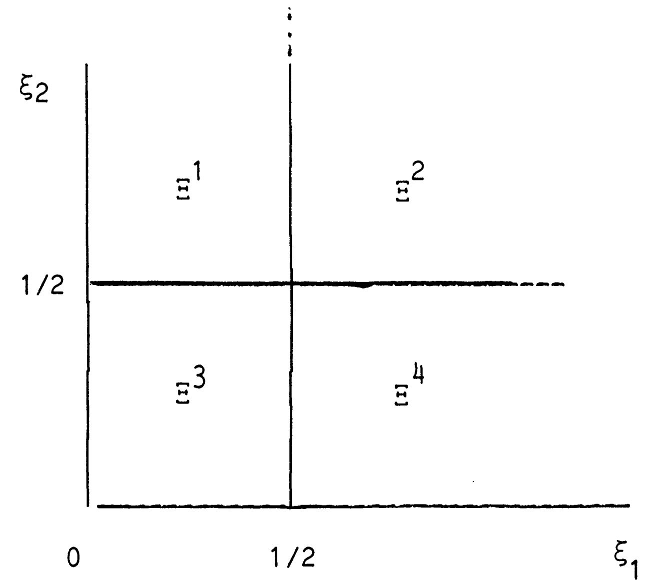

Figure 3. Partitions of Ξ.

we use $h = Q((0,0), \xi)$ and $C^{\ell} = \Xi^{\ell}$ with constraints on $v_{\ell} = e_{\ell-2}$, $\ell = 3,4$, and v_i as in (5.5). The resulting linear program has an optimal value of 8.98.

We can improve on this solution value by using the independence of ξ_1 and ξ_2. We consider first $\int Q((0,0), (\xi_1, 0))\ P^1(d\xi_1)$ where P^1 is the marginal probability distribution of ξ_1. Clearly, this has a value of $10\ \xi_1 = 5$ since $Q((0,0), (\xi_1, 0))$ is linear. Next, consider $Q((0,0), (\xi_1, 1/2))P^1(d\xi_1)$. For this problem, we use $v_1(\xi)$ as in (5.5) and an upper bound of 1/2 and solve the resulting program (3.13) in ξ_1 only. The result is an upper bound of 6.25. We note here that using the partition $\Xi_1^1 = [0, 1/2]$, $\Xi_2^2 = (1/2, +\infty)$, without the constraint on $v_1(\xi)$ leads to an upper bound of 7.24. The restriction on the second moment therefore allows for a 14% improvement in the bound.

The bound on $\int Q((0,0), (\xi_1, 1/2))P^1(d\xi_1)$ can then be used to bound (5.3). We use that 5 is a bound at $\hat{\xi}_2 = 0$, that 6.25 is a bound at $\hat{\xi}_2 = 1/2$ for $\int Q(0,0), (\xi_1, \hat{\xi}_2)P^1(d\xi_1)$ and that the recession function value of $\int Q((0,0), (\xi_1,\xi_2))P^1(d\xi_1)$ is 10. This yields another problem (3.13) that can be solved with a constraint on $v_2(\xi)$ to yield an overall upper bound of 8.12.

This represents a 10% improvement over the bound without using independence, and a 19% improvement over the bound which only considers mean values. The results are summarized in Table 2.

Constraints	Independent Computation	Upper Bound
Means	No	10.00
Means	Yes	10.00
Means & Second Moments	No	8.98
Means & Second Moments	Yes	8.12

Table 2. Bounds on $\int Q((0,0), \xi) P(d\xi)$.

This example provides one indication of the usefulness of including constraints on second moments in the solution of stochastic optimization problems. As a second example, we consider a region that is not bounded in any direction and a nonlinear function v_0. Let

$$Q(x, \xi) = \begin{cases} (\xi - x)^2 & \text{if } |\xi - x| \leq 5 \\ 10x - 25 & \text{if } \xi - x \geq 5 \\ -10x - 25 & \text{if } \xi - x \leq 5, \end{cases} \tag{5.6}$$

where ξ is normally distributed. We wish to bound $\int Q(0, \xi) P(d\xi)$ where $\bar{\xi} = 0$ and $E(\xi^2) = 1/4$. We use $v_1 = \xi$ for the mean value constraint and

$$v_2(\xi) = \begin{cases} 2\xi - 1 & \text{if } \xi \geq 1/2, \\ 0 & \text{if } -1/2 \leq \xi \leq 1/2, \\ -2\xi - 1 & \text{if } \xi \leq -1/2, \end{cases} \tag{5.7}$$

where $v_2(\xi) \leq \xi^2$. The resulting linear program (3.13), with $\Xi^1 = (-\infty, -1/2]$, $\Xi^2 = (-1/2, 1/2]$, $\Xi^3 = (1/2, +\infty)$, and a bound of 1/4 on $\int v_2(\xi) P(d\xi)$ has an optimal value of 1.5. Without the bound on $\int v_2(\xi) P(d\xi)$, however, the linear program is unbounded because of the distribution's symmetry. In this case, it is essential to include additional piecewise linear constraints to the mean value constraint.

REFERENCES

[1] J. Birge and R. Wets, Designing approximation schemes for stochastic optimization problems, in particular for stochastic programs with recourse, Mathematical Programming Study, forthcoming.

[2] J. Dupačová, Minimaxová úloha stochastického lineárního programování a momentovy problem, Ekonomicko - matematicky Obzor, 13(1977), 279-306.

[3] J. Dupačová, Minimax approach to stochastic linear programming and the moment problem. Recent results, ZAMM 58(1978), T466-467.

[4] H. Gassmann and W. Ziemba, A tight upper bound for the expectation of a convex function of a multi-variate random variable, Mathematical Programming Study, forthcoming.

[5] C. Huang, I. Vertinsky and W. Ziemba, Sharp bounds on the value of perfect information, Operations Research 26(1977), 315-325.

[6] P. Kall and D. Stoyan, Solving stochastic programming problems with recourse including error bounds, Mathematische Operationsforschung und Statistik, Serie Optimization 13(1982), 431-447.

[7] A. Karr, Extreme points of certain sets of probability measures, with applications, Mathematics of Operations Research, 8(1983), 74-85.

[8] J. Kemperman, The general moment problem. A geometric approach, Annals of Mathematical Statistics 39(1968), 93-112.

[9] M. Krein and A. Nudelman, The Markov Moment Problem and Extermal Problems, Translation Mathematical Monographs 50, American Mathematical Society, Providence, 1977.

[10] R.T. Rockafellar, Convex Analysis, Princeton University Press, Princeton, 1970.