

**Bounding the Expectation of Convex Functions
with Limited Distribution Information**

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

José H. Dulá
34 Voie du Roman Pays
C.O.R.E.
1348 Louvain-la-Neuve, Belgium

Technical Report 87-15

**Bounding the Expectation of Convex Functions
with Limited Distribution Information***

John R. Birge

Industrial and Operations Engineering

University of Michigan

Ann Arbor, MI, USA 48109

José H. Dulá

34 Voie du Roman Pays

C.O.R.E.

1348 Louvain-la-Neuve, Belgium

Keywords: integration, stochastic programming, moment problem, duality, approximation.

Abbreviated title: Bounding convex functions.

Mathematics Subject Classification: 90C15.

Abstract: This paper considers bounds on the expectation of a convex function of a random variable when only limited information is available about the underlying distribution. The problem is presented as a generalized moment problem. A special class of functions is shown to have an easily computable solution to this problem with first and second moment constraints. Extensions are given for general convex functions on finite intervals or with finitely valued recession functions.

* This research has been partially supported by the National Science Foundation under Grant ECS-8304065, by the Office of Naval Research Grant N00014-86-K-0628 and by the National Research Council under a Research Associateship at the Naval Postgraduate School, Monterey, California.

1. Introduction. Many problems in applied mathematics involve an integral or an expectation of a convex function. For example, in mechanics, the average load on a given structure may be expressed as the expectation of convex functions (or linear combinations of convex functions) of the stresses. In finance, the present value of an option on a stock may be the expected value of a convex function of the future value of the stock price. In computer systems, the performance (throughput) may be a concave (i.e., the negative of a convex) function of the system load. The average performance is then the expectation of this performance function. Our interest has been in functions arising in *stochastic programs*, mathematical optimization models that involve random variables.

The basic problem is to find:

$$E_f(x) = E\{f(x(\omega))\} = \int f(x(\omega))P(d\omega), \quad (1.1)$$

where x is a random vector mapping the probability space, (Ω, \mathcal{A}, P) , onto $(\mathfrak{R}^N, \mathcal{B}^N, F)$, F is the distribution function of x , and $x \in X \subset \mathfrak{R}^N$. The *expectation functional*, $E_f(x)$, can also be written as a Lebesgue–Stieltjes integral with respect to F :

$$E_f(x) = \int_{\mathfrak{R}^N} f(x)dF(x). \quad (1.2)$$

Difficulties arise in evaluating $E_f(x)$ when either the function f is difficult to evaluate (for example, requiring a complex simulation for each function evaluation as in the computer system example) or the distribution function F is not known exactly (for example, when the demands on the computer system are not known beforehand). Many approximation formulas for integrals (1.2) have been given (see Davis and Rabinowitz [8]), but they either do not provide efficient error bounds or impose strict differentiability requirements on the integrand. In this paper, we review methods for evaluating upper bounds on $E_f(x)$ with limited distribution information and convex f . In addition, we provide a method for evaluating upper bounds on a large class of convex functions on \mathfrak{R}^1 given first and second moment information about the distribution of x . We also show how this result can be extended to a general bound for convex functions on \mathfrak{R}^N . The bounds extend from basic properties of convex functions and from a generalized moment problem interpretation of the problem.

Section 2 provides background on previous approaches to bounding expectations. The generalized moment problem interpretation is given in Section 3. Section 4 presents basic results and some examples in \mathfrak{R}^1 . Section 5 provides the multi-dimensional extension.

2. Background and previous integral approximations. For $f(x) = x$, the expectation functional is the *first moment* or *mean* of the random vector, x , with respect to the distribution function, F . For $f(x) = x^i$, the expectational functional is the *i th moment* of x with respect to F . A function $f : X \rightarrow (-\infty, +\infty]$ for X a convex set is *convex on X* if and only if $f((1 - \lambda)x + \lambda y) \leq (1 - \lambda)f(x) + \lambda f(y)$ for all $\lambda \in (0, 1)$ and every x and y in X . In the following, we assume that X is convex and that every convex function f on X is *proper*, i.e., there exists $x \in X$ such that $f(x) < +\infty$ and $f(x) > -\infty$ for all $x \in X$. A function f is *concave on X* if $-f$ is convex on X . A function f is *affine on X* if f is finite-valued, convex and concave on X . If $f(\lambda x) = \lambda f(x)$ for all $\lambda \in (0, \infty)$ and $x \in \mathfrak{R}^N$, then f is *positively homogeneous (of degree one)*. A positively homogeneous function $f : \mathfrak{R}^N \rightarrow (-\infty, +\infty]$ that is also convex on \mathfrak{R}^N is *sublinear*. A sublinear function associated with any convex function is the *recession function*, f_0^+ , defined as (see [31])

$$(f_0^+)(y) = \sup\{f(x + y) - f(x) \mid f(x) < \infty\}. \quad (2.1)$$

In the sequel, we also consider optimization problems of the form,

$$\sup_{x \in C} f(x), \quad (2.2)$$

where C is a subset of a linear space. A point x is *feasible* in (2.2) if $x \in C$. A feasible point x is an *extreme point* of C if there does not exist y and z in C such that $x = (1 - \lambda)y + \lambda z$, $0 < \lambda < 1$ other than $x = y = z$. A point x^* is *optimal* in (2.2) if it is feasible and $f(x^*) = \sup_{x \in C} f(x)$. In this case, x^* *attains* the supremum of f over C . When the supremum of f is assumed to be attained for some $x^* \in C$, we replace “supremum” by “maximum.”

For general functions f , the basic procedures to approximate $E_f(x)$ use some form of a discrete approximation for the distribution of x . For $X = [a, b] \subset \mathfrak{R}^1$, the most basic procedures are the *midpoint* and the *trapezoidal* approximations. If one interval is used,

these approximations apply to a uniform distribution on $[a,b]$. They are $M_1(f) = f((a + b)/2)$ for the midpoint and $T_1(f) = (f(a) + f(b))/2$ for the trapezoidal approximation. The approximations are improved by dividing $[a,b]$ into subintervals, appropriately weighting the subintervals and applying M_1 and T_1 on each subinterval.

A more sophisticated procedure is *gaussian quadrature* to find an integral formula that fits all polynomials up to some degree. As noted by Miller and Rice [29], this can be used to find a discretization with N values that matches the first $N + 1$ moments of the distribution of x . To match the first three moments of the uniform distribution on $[a,b]$, for example, gaussian quadrature selects two points, $(a + b)/2 \pm (\sqrt{3}/6)(b - a)$, with equal probability, $1/2$.

A difficulty with using the gaussian quadrature formulas is that they do not generally provide bounds on the expectation. Restrictions on higher-order derivatives and Peano's theorem [30] may be used to provide bounds but that requires, at least, differentiability of f and a density function that may not be available. Generalizations of the mid-point and trapezoidal approximations do, however, obtain bounds on the expectation of a convex function. Jensen's inequality [21] can be interpreted as a generalization of the mid-point approximation that provides a lower bound on the expected value of convex f through the following:

$$\int_{\mathbb{R}^N} f(x)dF(x) \geq f\left(\int_{\mathbb{R}^N} x dF(x)\right), \quad (2.3)$$

where $\bar{x} = \int_{\mathbb{R}^N} x dF(x)$ is assumed finite.

Madansky, following Edmundson, ([27], [12]) provided a generalization of the trapezoidal approximation, called the *Edmundson–Madansky inequality*, that gives an upper bound on the expectation of a convex function. For $N = 1$, the basic inequality is:

$$\int_{\mathbb{R}^N} f(x)dF(x) \leq \frac{((b - \bar{x})f(a) + (\bar{x} - a)f(b))}{(b - a)}, \quad (2.4)$$

where $[a, b]$ is a finite interval. The Edmundson–Madansky inequality (2.4) can also be extended to multiple dimensions and infinite intervals (see, for example, [1], [14], and [16]).

Refinements of the Jensen and Edmundson–Madansky inequalities are possible by subdividing the interval (or, more generally, the region) into smaller pieces on which the

bounds can be reapplied as in the traditional mid-point and trapezoidal approximations (see [3], [15], [19], and [22]). These refinements require additional functional evaluations and conditional expectations on the subregions. As has been observed, the Jensen lower bound is generally reasonably accurate relative to the Edmundson–Madansky upper bound (e.g., [17]), which requires more function evaluations. The primary concern is then in obtaining more accurate upper bounds without additional computational effort. This paper shows that this is possible for a class of convex functions, if second moment information is available. The next section provides the generalized moment problem formulation that leads to this new bound.

3. Generalized moment problem. To obtain bounds that hold for all distributions with certain properties, we can find

$Q \in \mathcal{P}$ a set of probability measures on (X, \mathcal{B}^N) subject to

$$\begin{aligned} \int_X v_i(x)Q(dx) &\leq \alpha_i, i = 1, \dots, s, \\ \int_X v_i(x)Q(dx) &= \alpha_i, i = s + 1, \dots, M, \end{aligned} \tag{3.1}$$

$$\text{to maximize } \int_X f(x)Q(dx),$$

where M is finite and the v_i are bounded, continuous functions. A solution of (3.1) obtains an upper bound on the expectation of f with respect to any probability measure satisfying the conditions above. Problem 3.1 is a *generalized moment problem* ([25]). When the v_i are powers of x , the constraints restrict the moments of x with respect to Q . In this context, (3.1) determines an upper bound when only limited moment information on a distribution is available.

Problem 3.1 can also be interpreted as an *abstract linear program* since the objective and constraints are linear functions of the probability measure. The solution is then an extreme point (see [33] for a discussion of properties) in the infinite dimensional space of probability measures. The following theorem, proven in [23, Theorem 2.1], gives the explicit solution properties.

Theorem 1. *Suppose X is compact. Then the set of feasible measures in (3.1), \mathcal{Q} , is convex and compact (with respect to the weak* topology), and \mathcal{Q} is the closure of the convex hull of the extreme points of \mathcal{Q} . If f is continuous relative to X , then an optimum (maximum or minimum) of $\int_X f(x)Q(dx)$ is attained at an extreme point of \mathcal{Q} . The extremal measures of \mathcal{Q} are those measures that have finite support, $\{x_1, \dots, x_L\}$, with $L \leq M + 1$, such that the vectors,*

$$\begin{pmatrix} v_1(x_1) \\ v_2(x_1) \\ \vdots \\ v_M(x_1) \end{pmatrix}, \dots, \begin{pmatrix} v_1(x_L) \\ v_2(x_L) \\ \vdots \\ v_M(x_L) \end{pmatrix}, \quad (3.2)$$

are linearly independent.

Kemperman [24] showed that the supremum is attained under more general continuity assumptions and provides conditions (which we assume to hold) for \mathcal{Q} to be nonempty. Dupačová's (formerly Žáčková) [9, 10, 36] work on a minimax approach to stochastic programming led to the use of the moment problem as a bounding procedure for stochastic programs. She showed that (3.1) attains the Edmundson–Madansky bound (and the Jensen bound if the objective is minimized) when the only constraint in (3.1) is $v_1 = x$, i.e., the constraints fix the first moment of the probability measure. She also provided some properties of the solution with an additional second moment constraint ($v_2(x) = x^2$) for a specific objective function f . The general problem can be solved using the *generalized linear programming* procedure in [7, Chapter 24]. This procedure is given below.

Generalized Linear Programming Procedure for the Generalized Moment Problem (GLP)

Step 0. Initialization. Identify a set of $L \leq M + 1$ linearly independent vectors as in (3.2) that satisfy the constraints in (3.1). (Note that a *phase-one objective* ([7]) may be used if such a starting solution is not immediately available. For $N = 1$, the gaussian quadrature points may be used as mentioned above.) Let $\nu = L$, $k = 1$, go to 1.

Step 1. Master problem solution. Find $p_1 \geq 0, \dots, p_\nu \geq 0$ such that

$$\begin{aligned} \sum_{l=1}^{\nu} p_l &= 1, \\ \sum_{l=1}^{\nu} v_l(x_l) p_l &\leq \alpha_i, i = 1, \dots, s, \\ \sum_{l=1}^{\nu} v_l(x_l) p_l &= \alpha_i, i = s + 1, \dots, M, \text{ and} \\ z &= \sum_{l=1}^{\nu} f(x_l) p_l \text{ is maximized.} \end{aligned} \tag{3.3}$$

Let $\{p_1^k, \dots, p_\nu^k\}$ attain the optimum in (3.3), and let $\{\theta^k, \pi_1^k, \dots, \pi_M^k\}$ be the associated *dual multipliers* such that

$$\begin{aligned} \theta^k + \sum_{i=1}^M \pi_i^k v_i(x_l) &= f(x_l), \text{ if } p^l > 0, l = 1, \dots, \nu, \\ \theta^k + \sum_{i=1}^M \pi_i^k v_i(x_l) &\geq f(x_l), \text{ if } p^l = 0, l = 1, \dots, \nu, \\ \pi_i^k &\geq 0, i = 1, \dots, s. \end{aligned} \tag{3.4}$$

Step 2. Subproblem solution. Find $x^{\nu+1}$ that maximizes

$$\rho(x, \theta^k, \pi^k) = f(x) - \theta^k - \sum_{i=1}^M \pi_i^k v_i(x). \tag{3.5}$$

If $\rho(x^{\nu+1}, \theta^k, \pi^k) > 0$, let $\nu = \nu + 1$, $k = k + 1$ and go to 1.

Otherwise, stop, $\{p_1^k, \dots, p_\nu^k\}$ are the optimal probabilities associated with $\{x_1, \dots, x_\nu\}$ in a solution to [5].

The proof of the convergence of GLP is given in [7, Chapter 24]. This result is used in [13] to solve a class of problems (3.1). The difficulty in GLP is in the solution of

the subproblem (3.5), which generally involves a nonconvex function. Birge and Wets [3] describe how to solve (3.5) with constrained first and second moments, if convexity properties of ρ can be identified. Cipra [6] describes other methods for this problem based on discretizations and random selections of candidate points, x_i .

This paper considers specific conditions so that (3.1) can be solved without requiring the repeated nonconvex optimization in (3.5). The goal is to show that a large class of problems for $N = 1$ require only $L = 2$ points of support that can be identified in one line search, that the points of support are analytically calculable for certain functions, and that the solution procedure can be extended to a general class of functions in multiple dimensions. The next section describes the basic results.

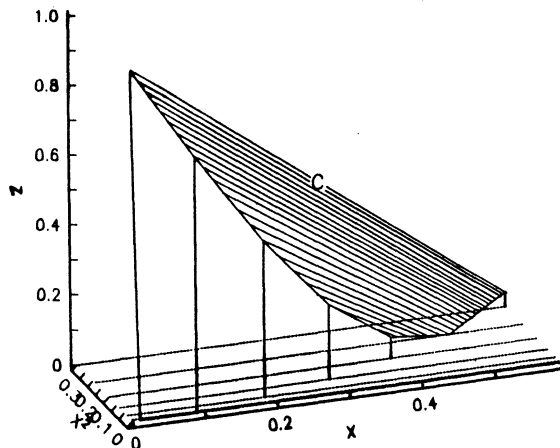


Figure 1. The generalized moment problem in \mathfrak{R}^1 .

4. Two-point support functions. In this section, we restrict (3.1) to the case of $N = 1$, $s = 0$, and $M = 2$, where again the constraints correspond to first and second moment constraints. To distinguish this case of (3.1), we refer to it as Problem (3.1-(1,0,2)). The problem is illustrated geometrically in Figure 1. Here, $X = [0, .6]$ and $C =$ the convex hull of $(x, x^2, f(x))$ for $x \in X$. The first moment, \bar{x} , and the second moment, $x^{(2)}$, are shown. The objective in (3.1-(1,0,2)) is to find $y^* = (\bar{x}, x^{(2)}, z^*) \in C$ that maximizes z . A generalization of Carathéodory's theorem ([35]) for the convex hull of a connected set tells us that y^* can be expressed as a convex combination of at most three extreme points of C , giving us a special case of Theorem 1. Therefore, an optimal solution to (3.1-(1,0,2)) can be written, $\{x^*, p^*\}$, where the points of support, $x^* = \{x_1^*, x_2^*, x_3^*\}$ have probabilities,

$p^* = \{p_1^*, p_2^*, p_3^*\}$. We will show that, for a broad class of *two-point support* functions, this can indeed be further restricted to two points, $\{x_1^*, x_2^*\}$, with positive probability.

A useful result of the linear programming interpretation of (3.1) is the presence of a dual problem to (3.1). The dual to (3.1) is known as a *semi-infinite* program ([18]) that appears, for example, in Chebyshev approximation ([32]). For the one-dimensional, two-moment constraint problem considered here, this dual is to find θ, π_1, π_2 such that

$$\theta + \pi_1 x + \pi_2 x^2 \geq f(x), \forall x \in X, \quad (4.1)$$

and $\theta + \pi_1 \bar{x} + \pi_2 \bar{x}^2$ is minimized.

Note that (4.1) involves three variables and an infinite number of constraints in contrast to the infinite dimensional, finitely constrained *primal problem* (3.1-(1,0,2)). Note also that an optimal solution to (4.1) is a quadratic function that dominates f and that has minimum expectation with respect to any probability measure in \mathcal{Q} .

The optimality conditions on a feasible solution to (3.1), $x^* = \{x_1^*, x_2^*, x_3^*\}$, with associated probabilities, $p^* = \{p_1^*, p_2^*, p_3^*\}$, are that there exist dual variables, $\theta^*, \pi_1^*, \pi_2^*$, such that

$$\begin{aligned} \theta^* + \pi_1^* x_i^* + \pi_2^* (x_i^*)^2 &= f(x_i^*) \text{ if } p_i^* > 0, \\ \theta^* + \pi_1^* x + \pi_2^* x^2 &\geq f(x), \forall x \in X, \end{aligned} \quad (4.2)$$

where the first condition is known as *complementary slackness* condition and the second condition is *dual feasibility*. A useful interpretation of these conditions in terms of the function ρ defined in (3.5) is that x_i^* has a positive probability, p_i^* , only if $\rho(x_i^*, \theta^*, \pi^*) = 0$ and x_i^* maximizes ρ over X for fixed (θ^*, π^*) . It is convenient to let $\rho(x, \theta, \pi) = f(x) - q(x)$, where $q(x, \pi, \theta) = \theta + \pi_1 x + \pi_2 x^2$. The following lemmas give additional properties for an optimal solution $\{x^*, p^*\}$ to (3.1-(1,0,2)). We assume that f is always convex below. The first lemma refers to feasible solutions of (3.1-(1,0,2)).

Lemma 1. *If a feasible solution to (3.1-(1,0,2)) is obtained with positive probability support points, $x_1 < x_2 < x_3$, then another feasible solution exists with support points, x_1 and x_4 , where $x_2 < x_4 < x_3$.*

Proof: This result is straightforward. Feasibility of $\{x_1, x_2, x_3\}$ implies there exists $\{p_1, p_2, p_3\} > 0$ such that $\sum_{i=1}^3 p_i(x_i, x_i^2) = (\bar{x}, x^{(2)})$ or $(x_1, x_1^2) + p_2(x_2 - x_1, x_2^2 - x_1^2) + p_3(x_3 - x_1, x_3^2 - x_1^2) = (\bar{x}, x^{(2)})$. This implies that the line $(x_1 + t(\bar{x} - x_1), x_1^2 + t(x^{(2)} - x_1^2))$ is a convex combination of the lines, $(x_1 + t(x_i - x_1), x_1^2 + t(x_i^2 - x_1^2))$, $i = 2, 3$. All three lines intersect (x, x^2) at (x_1, x_1^2) . The convexity property implies $(x_1 + t(\bar{x} - x_1), x_1^2 + t(x^{(2)} - x_1^2))$ intersects (x, x^2) at x_4 such that $x_2 < x_4 < x_3$. ■

The next lemma considers f with a derivative f' that has local convexity or concavity properties. These properties form the basis for the bounding approach given in this paper.

Lemma 2. *If $\rho(\hat{x}, \hat{\theta}, \hat{\pi}) = 0$ for some $(\hat{\theta}, \hat{\pi})$ feasible in (4.1), \hat{x} is in the interior of X , and f' is convex or concave on $A = (\hat{x} - \eta, \hat{x})$ or $B = (\hat{x}, \hat{x} + \eta)$ for some $\eta > 0$, then, there exists $\delta > 0$, such that, for any $\epsilon \geq 0, \epsilon < \delta$, $f'(\hat{x} - \epsilon) \geq q'(\hat{x} - \epsilon, \hat{\theta}, \hat{\pi})$ for f' convex or concave on A and $f'(\hat{x} + \epsilon) \leq q'(\hat{x} + \epsilon, \hat{\theta}, \hat{\pi})$ for f' convex or concave on B .*

Proof: Let $\hat{\rho}(\cdot) = \rho(\cdot, \hat{\theta}, \hat{\pi})$ and let $\hat{q}(\cdot) = q(\cdot, \hat{\theta}, \hat{\pi})$. Note that \hat{q} is differentiable on X and that, for f convex, f is continuously differentiable on all but a countable number of points in X . Since $\hat{\rho}$ is then densely differentiable, there exist right and left open neighborhoods of \hat{x} on which $\hat{\rho}$ is differentiable. The following limits are well-defined

$$\hat{\rho}'^-(\hat{x}) = \lim_{t \downarrow 0} \hat{\rho}'(\hat{x} - t); \hat{\rho}'^+(\hat{x}) = \lim_{t \downarrow 0} \hat{\rho}'(\hat{x} + t). \quad (4.3)$$

If $\hat{\rho}(\hat{x}) = 0$ for feasible $(\hat{\theta}, \hat{\pi})$ in (4.1), then \hat{x} maximizes $\hat{\rho}$. Therefore, $0 \in [\hat{\rho}'^-(\hat{x}), \hat{\rho}'^+(\hat{x})]$. By the local convexity (or concavity) assumption for f' around \hat{x} , there exist intervals $(\hat{x} - t, \hat{x})$ or $(\hat{x}, \hat{x} + t)$ ($t > 0$) over which $\hat{\rho}'$ has constant sign. Suppose $f'(\hat{x} - \epsilon) \leq q'(\hat{x} - \epsilon, \hat{\theta}, \hat{\pi})$ for all $0 < \epsilon < t$ and $f'(\hat{x} - s) < q'(\hat{x} - s, \hat{\theta}, \hat{\pi})$ for some $0 < s < t$, then $\hat{\rho}(\hat{x}) < \hat{\rho}(\hat{x} - t)$, contradicting the maximality of \hat{x} . A similar argument holds for $(x, x + t)$, proving the result. ■

The previous lemma considers local convexity properties of f' when it exists. The following results refer to functions with derivatives that are convex and then concave.

Lemma 3. *Let $g(x) = h(x) - c(x)$ be a function such that $g(x)$ is increasing on \mathfrak{R} , h is convex on $(-\infty, y)$ and concave on (y, ∞) , and $c(x)$ is an affine function on \mathfrak{R} . Then there exists a partition of $(-\infty, \infty)$ into subintervals, $I_1 = (-\infty, a_1)$, $I_2 = [a_1, a_2)$, $I_3 = [a_2, a_3)$, $I_4 = [a_3, +\infty)$, $-\infty \leq a_1 \leq a_2 \leq a_3 \leq \infty$, such that $g(x) \geq 0$ for all $x \in I_1 \cup I_3$ and $g(x) \leq 0$ for all $x \in I_2 \cup I_4$.*

Proof: First note that g is continuous on $(-\infty, c)$ and (c, ∞) . Also, g cannot change sign from negative to positive twice on $(-\infty, c)$ by the convexity of h in that region. If g has two sign changes on $(-\infty, y)$, then there exists intervals, I_1, I_2, I_3 such that g is positive on $I_1 \cup I_3$ and negative on I_2 . For g increasing, $g(y) \geq 0$. By h concave on (y, ∞) , g has at most one sign change on (y, ∞) , giving the result when g has two sign changes on $(-\infty, y)$. A similar argument applies if g has one or no sign changes on $(-\infty, c)$. ■

The next lemma considers the case where $g = \rho'$ is constant on an interval. In the following, we use the notation $X = [a, b]$ for convenience. It is assumed that this also includes the cases $X = (-\infty, b]$, $X = [a, +\infty)$, and $X = (-\infty, +\infty)$ unless explicitly stated otherwise.

Lemma 4. *If f is convex on $X = [a, b]$ with derivative f' defined as a convex function on (a, c) and as a concave function on (c, b) for $a \leq c \leq b$ and if $\rho(x, \hat{\theta}, \hat{\pi}) = 0$ for some $(\hat{\theta}, \hat{\pi})$ feasible in (4.1) and for all $x \in (\hat{x} - \epsilon, \hat{x} + \epsilon)$ for some $\hat{x} \in X$ and $\epsilon > 0$, then there exists an interval $D \supset (\hat{x} - \epsilon, \hat{x} + \epsilon)$ such that $\rho(x, \hat{\theta}, \hat{\pi}) = 0$ for all $x \in D$ and $\rho(x, \hat{\theta}, \hat{\pi}) > 0$ for all $x \notin$ the closure of D .*

Proof: Let $D = (d, e)$ be the largest interval including $(\hat{x} - \epsilon, \hat{x} + \epsilon)$ such that $\hat{\rho}(x) = 0$ for all $x \in D$. By Lemma 2 and the assumption, if $d > a$, there exists an interval, $(d - \eta, d)$, and, if $e < b$, there exists an interval $(e, e + \eta)$, for $\eta > 0$, such that $\hat{\rho}' > 0$ on $(d - \eta, d)$ and $\hat{\rho}' < 0$ on $(e, e + \eta)$. By the convex-concave assumption, f' is convex on $[a, d)$ and concave on $(e, b]$ since \hat{q}' is affine. Hence, $\hat{\rho}'$ would be strictly positive on $[a, d)$ and strictly negative on $(e, b]$. This would imply that $\hat{\rho}$ is negative on $(e, b]$, violating the feasibility of $(\hat{\theta}, \hat{\pi})$ in

(4.1). Hence, $e = b$, giving the result.■

The convex–concave property is now used to derive our main result about two–point support functions.

Theorem 2. *If f is convex with derivative f' defined as a convex function on (a, c) and as a concave function on (c, b) for $X = [a, b]$ and $a \leq c \leq b$, then there exists an optimal solution to (3.1–(1,0,2)) with at most two support points, $\{x_1, x_2\}$, with positive probabilities, $\{p_1, p_2\}$.*

Proof: Let $\{\hat{\theta}, \hat{\pi}\}$ be an optimal solution to (4.1). First, assume that there does not exist $\epsilon > 0$, $\hat{x} \in (a, b)$ such that $\rho(x, \hat{\theta}, \hat{\pi}) = 0$ for all $x \in (\hat{x} - \epsilon, \hat{x} + \epsilon)$. By Lemmas 1 and 3, the only isolated points where $\hat{\rho}$ could be 0 and maximized are a_1 and a_3 if $[a, b] \supset I_2 \cup I_3$. If $[a, b] \not\supset I_2$, then a can replace a_1 and if $[a, b] \not\supset I_3$, then b can replace a_3 , but, in either case, at most two points meet the conditions for optimality.

If there exists $\epsilon > 0$, $\hat{x} \in (a, b)$ such that $\rho(x, \hat{\theta}, \hat{\pi}) = 0$ for all $x \in (\hat{x} - \epsilon, \hat{x} + \epsilon)$, then Lemma 4 implies that any optimal solution $\{x_1, x_2, x_3\}$ must be in the closure of D and that the $\hat{\rho}(x) = 0$ for all $x \in D$. By Lemma 2, we can select x_4 in (x_2, x_3) such that there exists $\{p_1, p_4\}$ so that $\{x_1, x_4, p_1, p_4\}$ is feasible in (3.1–(1,0,2)). The optimality conditions still hold for $\hat{\rho}(x_4) = 0$. Hence, $\{x_1, x_4, p_1, p_4\}$ is optimal in (3.1–(1,0,2)).■

A corollary of Theorem 2 is that any function f that has a convex or concave derivative has the two–point support property. The class of functions that meets the criteria of Theorem 2 contains many useful examples. Some of these functions are given below:

1. Polynomials defined over ranges with at most one third derivative sign change.
2. Exponential functions of the form, $c_0 e^{c_1 x}$, $c_0 \geq 0$.
3. Logarithmic functions of the form, $\log_k(cx)$, for any $k \geq 0$.
4. Certain hyperbolic functions such as $\sinh(cx)$, $c, x \geq 0$, $\cosh(cx)$.
5. Certain trigonometric functions such as $\tan^{-1}(cx)$, $c, x \geq 0$.

In fact, Theorem 2 can be applied to provide an upper bound on the expectation of any convex function with known third derivative when the distribution function has a known third moment, $x^{(3)}$. Suppose $a > 0$ (if not, then this argument can be applied on $[a, 0]$ and $[0, b]$), then let $g(x) = \alpha x^3 + f(x)$. The function g is still convex on $[0, b]$ for $\alpha \geq 0$. By defining $\alpha \geq (-1/6) \min(0, \inf_{x \in [a, b]} f'''(x))$, g' is convex on $[a, b]$, and an upper bound, $UB(g)$, on $E_g(x)$ has a two-point support. The expectation of f is then bounded by

$$E_f(x) \leq UB(g) - \alpha x^{(3)}. \quad (4.4)$$

The conditions in Theorem 2 are only sufficient for a two-point support function. They are not necessary. The following function, for example, has an optimal two-point support at $x^* = \{1/3, 1\}$ for any corresponding feasible p_1 and p_2 when $X = [0, 1]$.

$$f(x) = \begin{cases} 6/5 - 4x + 5x^2 & \text{if } 0 \leq x < .2, \\ -2x + 1 & \text{if } .2 \leq x < .4, \\ 1 - (2/5)x - 8x^2 + 10x^3 & \text{if } .4 \leq x < .6, \\ 1 - 4x + 4x^2 & \text{if } .6 \leq x \leq 1. \end{cases} \quad (4.5)$$

The function defined in (4.5) does not, however, meet the conditions of Theorem 2.

Note also that not all functions are two-point support functions (although bounds such as (4.4) are generally available). A function requiring three support points, for example, is $f(x) = (1/2) - \sqrt{(1/4) - (x - (1/2))^2}$. This function and its optimal dominating quadratic function are illustrated in Figure 2.

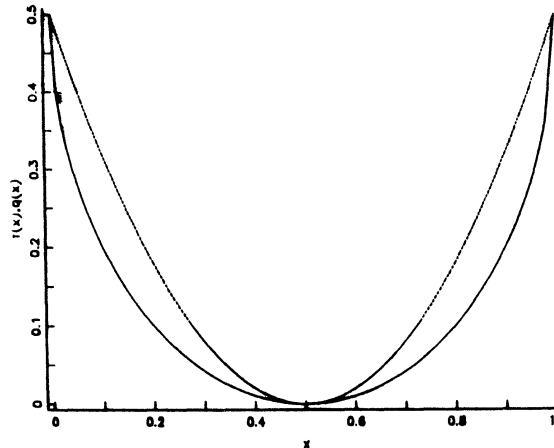


Figure 2. A function requiring three support points.

Given that a function is a two-point support function, the points $\{x_1, x_2\}$ can be found using a line search to find a maximum. For example, if some candidate $x_1 < \bar{x}$ is given, then a feasible corresponding x_2 is

$$x_2 = \frac{x^{(2)} - \bar{x}x_1}{\bar{x} - x_1}, \quad (4.6)$$

where $p_1 = (x_2 - \bar{x})/(x_2 - x_1)$ and $p_2 = 1 - p_1$. Note that the problem is obviously not feasible if $(x_1, x_2) \not\geq \bar{x}$. The solution of (3.1-(1,0,2)) then reduces to maximizing :

$$\gamma(x_1) = p_1(x_1)f(x_1) + p_2(x_1)f(x_2(x_1)) \text{ subject to } x_1 \in [a, \bar{x}]. \quad (4.7)$$

A line search to find the maximum in (4.7) can be performed efficiently using, for example, Lemaréchal and Mifflin's procedure in [26] if $\gamma \in C^2$ or Mifflin and Strodios's [28] method without derivatives. Table 1 gives the values (under "2-M") that were obtained by this procedure for three two-point support functions with distributions on $[0, 1]$. Figures 3-5 illustrate that the optimal points, $\{x_1^*, x_2^*\}$, may be at either endpoint or interior to $[0, 1]$. The table gives the expectation for a random variable with beta distribution (under "Beta") with the given first and second moments for comparison. The Edmundson-Madansky upper bound ("E-M") is also provided. The "S-L" value (for "semi-linear bound") given in Table 1 is discussed below.

Table 1. Bounding values

Function	\bar{x}	$x^{(2)}$	Beta	2-M	S-L	E-M
e^{-x}	0.500	0.333	0.622	0.624	0.651	0.684
x^3	0.833	0.714	0.625	0.629	0.675	0.833
$\sin(\pi(x+1))$ +1	0.500	0.333	0.363	0.384	0.577	1.000

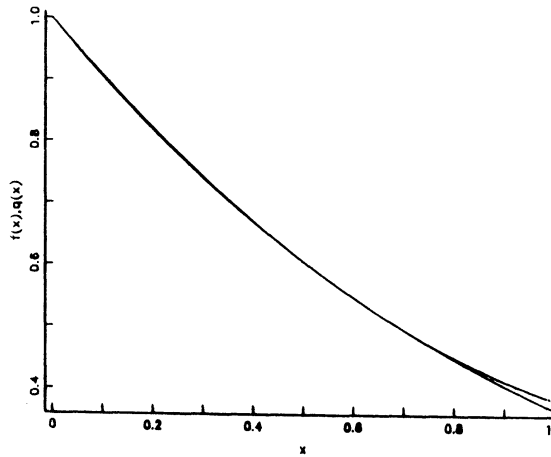


Figure 3. Optimal bounding function for e^{-x^2} .

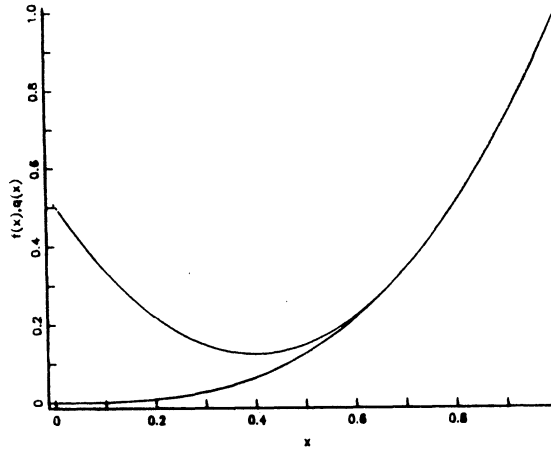


Figure 4. Optimal bounding function for x^3 .

A line search is not necessary for solving (3.1-(1,0,2)) in certain cases. The support points can be calculated analytically for *semi-linear, convex functions* that are defined by

$$f(x) = \begin{cases} q^-(c-x) & \text{if } x \leq c, \\ q^+(x-c) & \text{if } x > c, \end{cases}$$

where $q^+ + q^- \geq 0$. These functions clearly meet the conditions of Theorem 2 and, hence, require at most two points of support. The analytical results depend on the interval, $[a, b]$. If $[a, b] = [0, 1]$, then consider the nonintersecting intervals, $A = (0, x^{(2)}/(2\bar{x}))$, $B = [x^{(2)}/(2\bar{x}), (1 - x^{(2)})/(2(1 - \bar{x}))]$, and $C = ((1 - x^{(2)})/(2(1 - \bar{x})), 1)$. The points of support for a semi-linear, convex function defined on $[0, 1]$ are

$$\{x_1^*, x_2^*\} = \begin{cases} \{0, x^{(2)}/\bar{x}\} & \text{if } c \in A, \\ \{c-d, c+d\} & \text{if } c \in B, \\ \{(\bar{x} - x^{(2)})/(1 - \bar{x}), 1\} & \text{if } c \in C, \end{cases} \quad (4.8)$$

where $d = \sqrt{c^2 - 2c\bar{x} + x^{(2)}}$. This result can be obviously extended to all finite intervals. It results from **analyzing the zeroes of γ** .

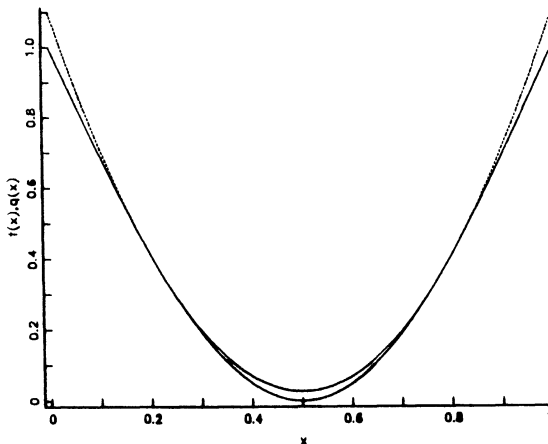


Figure 5. Optimal bounding function for $\sin(\pi(x+1)) + 1$.

Infinite intervals can also be solved analytically for semi-linear, convex functions. For $X = [0, \infty)$, the results are as in (4.8) with $B = [x^{(2)}/(2\bar{x}), \infty)$ and $C = \emptyset$. For the interval $(-\infty, \infty)$, the points of support are those for interval B in (4.8). We note that special cases for these supports of semi-linear, convex functions were considered in [9], [20], and [34].

Semi-linear, convex functions are common in decision problems to represent penalties for being above or below a preferred value, c . They can also be used, however, to provide bounds for other convex functions when only the first and second moments of the distribution function are known. For example, the following function dominates any convex function on $[0, 1]$,

$$\nu(x, c) = \begin{cases} (c-x)((f(0) - f(c))/c) + f(c) & \text{if } x \leq c, \\ (x-c)((f(1) - f(c))/(1-c)) + f(c) & \text{if } x > c, \end{cases} \quad (4.9)$$

where $c \in (0, 1)$. The function ν is a semi-linear, convex function for any convex function, f . The values of ν at the points in (4.8) can, therefore, be used to bound $\int \nu(x)Q(dx) \geq \int f(x)Q(dx)$ for any Q that is feasible in (3.1-(1,0,2)). The results of using ν to bound $f(x) = e^{-x}$, x^3 and $\sin(\pi(x+1)) + 1$ are given in Table 1 as mentioned above. More extensive results in [11] indicate that this approximation is accurate for many functions and moment values.

Semi-linear, convex functions can also be defined to dominate functions defined on unbounded ranges if those functions have finitely valued recession functions in each direc-

tion. If $f0^+(y) < \infty$ for all $y \in (-\infty, \infty)$, then there exists some c such that $f(c) < \infty$ and $\nu(x, c) = (f0^+)(x - c) + f(c) \geq f(x)$ for all x . The corresponding formulas for semi-linear, convex functions on unbounded ranges can then again be used to bound $E_f(x)$ for any feasible distribution that meets the conditions in (3.1-(1,0,2)). Bounds on the expectation of sublinear functions are especially computable because $(f0^+)(y) = f(y)$ for sublinear f by definition. The next section discusses how to use these bounds in multiple dimensions.

5. Bounds in multiple dimensions. The use of the generalized programming formulation is limited in multiple dimensions because of the difficulty in solving the sub-problem (3.5). Another difficulty is that, even with a bound only on the first moment of the distribution function and constrained cross-moments, $E[x_i x_j], i \neq j$, the moment problem solution ([14]) involves positive weights on 2^N extreme points of X when X is a compact set in \mathfrak{R}^N (and a corresponding number of recession function evaluations when X is not compact ([5])). These computational disadvantages for large values of N suggest that a looser but more computationally efficient upper bound on the value of (3.1) may be more useful than solving (3.1) exactly for large N .

If a *separable function*, $\nu(x, c) = \sum_{i=1}^N \nu_i(x(i), c(i))$, is available, it offers an obvious advantage by only requiring single integrals. In this case, we would like to find $\nu(x, c) = \sum_{i=1}^N \nu_i(x(i), c(i)) \geq f(x)$ where each $\nu_i(x(i), c(i))$ is a semi-linear, convex function. Methods for constructing these functions to bound the optimal value of a linear program with random right-hand side are discussed in [2] and [4]. In these *stochastic linear programs*,

$$f(x) = \min_{y \in \mathfrak{R}^{n+}} \{qy \mid Ay = x\}, \quad (5.1)$$

where $q \in \mathfrak{R}^n$ and $A \in \mathfrak{R}^{n \times N}$ are known parameters of the problem. Note that f defined in (5.1) is sublinear.

The functions ν_i are found by solving for

$$q_i^\pm = \min_{y \in \mathfrak{R}^{n+}} \{qy \mid Ay = \pm e_i\}, \quad (5.2)$$

where e_i is the i th unit vector in \mathfrak{R}^N . We then have

$$\nu_i(x_i, 0) = \begin{cases} q_i^+ x_i & \text{if } x_i \geq 0, \\ -q_i^- x_i & \text{if } x_i < 0, \end{cases} \quad (5.3)$$

whose expectation can now be bounded using the support points found in Section 4. Note that this bound requires $2N$ function evaluations instead of the exponential number of function evaluations required in the Edmundson–Madansky bound.

Different values of c_i from 0 in (5.3) are used when a deterministic vector also appears in the right-hand side of the linear program in (5.1) (i.e., the constraints are $Ay = x - t$ for some deterministic vector t). More precise bounding functions for f in (5.1) are possible for distributions on compact regions ([2]). Linear transformations can be used to obtain other separable, semi-linear convex upper bounding functions ([4]).

6. Conclusions. This paper describes how to bound the expectation of convex functions when only limited distributional information is available. Given the difficulties in estimating random phenomenon, limited information in terms of bounds on the mean and second moments of distributions is a general practical situation. The bounds provided in this paper allow for efficient computation. We note that these procedures can be extended to lower bounds on the expectation of concave functions obviously. Since the Jensen inequality is the solution of the generalized moment problem to minimize the expectation of a convex function subject to a first moment equality constraint and an upper bound on the second moment constraint, upper and lower bounds are computable on the expectation of general functions that can be expressed as linear combinations of convex and concave functions. Given this extension and the use of upper bounding separable, semi-linear convex functions as in Section 5, the two-point support bounds apply to a wide range of problems.

REFERENCES

1. A. Ben-Tal and E. Hochman, More Bounds on the Expectation of a Convex Function of a Random Variable, *J. Appl. Prob.* 9 (1972) 803–812.
2. J. R. Birge and S.W. Wallace, A Sepable Piecewise Linear Upper Bound for Stochastic Linear Programs, *SIAM J. Control and Optimization*, to appear.

3. J.R. Birge and R. J-B. Wets, Designing Approximation Schemes for Stochastic Optimization Problems, in particular, for Stochastic Programs with Recourse, *Mathematical Programming Study* 27 (1986) 54–102.
4. J.R. Birge and R. J-B. Wets, Sublinear Upper Bounds for Stochastic Linear Programs with Recourse, Technical Report No. 86–26, Department of Industrial and Operations Engineering, The University of Michigan (Ann Arbor, MI, 1986).
5. J.R. Birge and R. J-B. Wets, Computing Bounds for Stochastic Programming Problems by Means of a Generalized Moment Problem, *Mathematics of Operations Research* 12 (1987) 149–162.
6. T. Cipro, Moment Problem with Given Covariance Structure in Stochastic Programming, *Ekonom.-Mat. Obzor* 21 (1985) 66–77.
7. G.B. Dantzig, *Linear Programming and Extensions*, Princeton University Press, Princeton, NJ, 1963.
8. P. J. Davis and P. Rabinowitz, *Methods of Numerical Integration*, Academic Press, Orlando, Florida, 1984.
9. J. Dupačová, Minimax Stochastic Programs with Nonconvex Nonseparable Penalty Functions, in: *Progress in Operations Research*, A. Prékopa, ed., North-Holland, Amsterdam, 1976, pp. 303–316.
10. J. Dupačová, Minimax Approach to Stochastic Linear Programming and the Moment Problem. Recent Results, *ZAMM* 58 (1977) T466–467.
11. J. Dulá, Bounds on the Expectation of Convex Functions, Ph. D. dissertation, The University of Michigan, Ann Arbor, Michigan, 1986.
12. H. P. Edmundson, Bounds on the Expectation of a Convex Function of a Random Variable, The Rand Corporation Paper 982, Santa Monica, CA, 1956.
13. Y. Ermoliev, A. Gaivoronski, and C. Nedeva, Stochastic Optimization Problems with Partially Known Distribution Functions, *SIAM J. Control and Optimization*, to appear.
14. K. Frauendorfer, Solving S.L.P. Recourse Problems with Arbitrary Multivariate Dis-

- tributions – the Dependent Case, *Mathematics of Operations Research*, to appear.
15. K. Frauendorfer and P. Kall, A Solution Method for SLP Recourse Problems with Arbitrary Multivariate Distributions – the Independent Case, Institut für Operations Research der Universität Zürich, Technical Report, 1986.
 16. H. Gassmann and W. T. Ziemba, A Tight Upper Bound for the Expectation of a Convex Function of a Multivariate Random Variable, *Mathematical Programming Study* 27 (1986) 39–53.
 17. D. Hausch and W.T. Ziemba, Bounds on the Value of Information in Uncertain Decision Problems II, *Stochastics* 10 (1983) 181–217.
 18. R. Hettich, An Implementation of a Discretization Method for Semi-Infinite Programming, *Mathematical Programming* 34 (1986) 354–361.
 19. C.C. Huang, W. Ziemba, and A. Ben-Tal, Bounds on the Expectation of a Convex Function of a Random Variable: with Applications to Stochastic Programming, *Operations Research* 25 (1979) 315–325.
 20. R. Jagannathan, A Minimax Procedure for a Class of Linear Programs under Uncertainty, *Operations Research* 25 (1977) 173–177.
 21. J.L. Jensen, Sur les fonction convexes et les inégalités entre les valeurs moyennes, *Acta Mathematica* 30 (1906) 175–193.
 22. P. Kall and D. Stoyan, Solving Stochastic Programming Problems with Recourse including Error Bounds, *Math. Operationsforsch. Statist. Ser. Optim.* 13 (1982) 431–447.
 23. A. Karr, Extreme Points of Certain Sets of Probability Measures, with Applications, *Mathematics of Operations Research* 8 (1983) 74–85.
 24. J.H.B. Kemperman, The General Moment Problem, a Geometric Approach, *The Annals of Mathematical Statistics* 39 (1968) 93–122.
 25. M. Krein and A. Nudelman, The Markov Moment Problem and Extremal Problems, Translation of Mathematical Monographs, 50, The American Mathematical Society, Providence, RI, 1977.

26. C. Lemaréchal and R. Mifflin, Global and Superlinear Convergence of an Algorithm for One-Dimensional Minimization of Convex Functions, *Mathematical Programming* 24 (1982) 241–256.
27. A. Madansky, Bounds on the Expectation of a Convex Function of a Multivariate Random Variable, *Ann. Math. Statist.* 30 (1959) 743–746.
28. R. Mifflin and J. Strodriot, Reliable and Rapid Univariate Minimization without Derivatives, presented at TIMS/ORSA Joint National Meeting, Los Angeles, CA, April, 1986.
29. A. C. Miller, III, and T. R. Rice, Discrete Approximations of Probability Distributions, *Management Science* 29 (1983) 352–362.
30. G. Peano, Residuo in formulas de quadratura, *Mathesis* 39 (1914) 5–10.
31. R.T. Rockafellar, *Convex Analysis*, Princeton University Press, Princeton, New Jersey, 1970.
32. T. J. Rivlin, *An Introduction to the Approximation of Functions*, Blaisdell, Waltham, MA, 1969.
33. N. M. Roy, Extreme Points of Convex Sets in Infinite Dimensional Spaces, *Am. Mathematical Monthly* 94 (1987) 409–422.
34. H. Scarf, A Minimax Solution of an Inventory Problem, in: *Studies in the Mathematical Theory of Inventory and Production*, K. J. Arrow, S. Karlin, and H. Scarf, eds., Stanford University Press, Stanford, CA, 1958.
35. F. A. Valentine, *Convex Sets*, McGraw-Hill, New York, 1964.
36. J. Žáčková, On Minimax Solutions of Stochastic Linear Programming Problems, *Časopis pro Pěstování Matematiky* 91 (1966) 423–430.