To the Graduate Council:

I am submitting herewith a thesis written by Phillip D. Yates entitled "Bounds on the Expectation of a Convex Function of a Multivariate Random Variable." I have examined the final copy of this thesis for form and content and recommend that it be accepted in partial fulfillment of the requirements for the degree of Master of Science, with a major in Mathematics.

Yueh-er Kuo

Yueh-er Kuo, Major Professor

We have read this thesis

and recommend its acceptance:

Charles B. Collins

Julius Smith

Accepted for the Council:

Associate Vice Chancellor and

Dean of The Graduate School

BOUNDS ON THE EXPECTATION OF A CONVEX FUNCTION OF A MULTIVARIATE RANDOM VARIABLE

A Thesis

Presented for the

Master of Science

Degree

The University of Tennessee, Knoxville

Phillip D. Yates

August 1996

ACKNOWLEDGMENTS

The author wishes to express his gratitude to Dr. Yueh-er Kuo for her assistance in the completion of this thesis. Her combination of open-mindedness and gentle insistence allowed me to accomplish what at times appeared a daunting task. Appreciation is also extended to Dr. J. Smith and Dr. C. Collins for both their time in reading this thesis and their advice.

ABSTRACT

The idea of bounding E[g(X)], where g is a convex function of a multivariate random variable X, has received considerable attention in the literature. The need for such a bound arises from the complexities associated with computing E[g(X)], especially in a stochastic programming setting, when the dimension of X is large. The first contribution in this area was Jensen's inequality (1906). The purpose here is to discuss the 1st order bounds derived by Madansky (1959), Ben-Tal and Hochman's (1972) refinement of the Madansky bound using additional information, and Gassman and Ziemba's (1986) extension of the Madansky bound. An example of bounding E[f(X)], where f is not necessarily convex, will also be discussed.

TABLE OF CONTENTS

CHAPTER		PAGE
I.	INTRODUCTION	1
II.	BOUNDS IN A LINEAR PROGRAMMING CONTEXT	3
III.	PRELIMINARY BOUNDS	9
IV.	REFINED BOUNDS	17
V.	A FURTHER REFINEMENT WITH EXTENSION	26
BIBLIOGRAPHY		38
VITA		41

CHAPTER I

INTRODUCTION

In this chapter the motivating concepts that led to an interest in determining the bounds on the expectation of a convex function of a multivariate random variable will be presented. We will also outline the various contributions to the literature that have been presented in partial solution to the above problem and the purpose of this thesis.

The general problem is simply stated as follows:

Problem: Determine an upper and lower bound for $\int g(x_1,...,x_n)dF(X_1,...,X_n)$ where g is a convex function and F is a cumulative distribution function on $\Re^n$.

The evaluation of such an expectation can be a difficult task should

1. F not be complete,
2. F not exist in closed form,
3. g be difficult or expensive to calculate,
4. the evaluation of a higher dimensional integral.

J. Dula (1992) claims that it becomes virtually impossible to evaluate the integral directly when the dimension of our sample space increases. Prekopa and Wets (1986) state that numerical integration routines exist for n=3 in only a few special cases. Since typical applications of stochastic programs require the dimension of our sample space to be over 5 to 100 dimensions, there is obviously a need to tackle the problem a different way. Thus a clearly defined need exists to establish bounds on the expectation of a convex function of a multivariate random variable. Since the problem as stated is unsolved, we will discuss methods to solve the problem when various restrictions are imposed.

In chapter II the General Moment Problem (GMP) will be presented and a solution to it that makes use of mathematical programming. This chapter will contain a slightly more general version of the problem to be discussed by this thesis; but, some of the methodologies employed appear in later contributions that will be addressed in subsequent

chapters.

With regards to establishing bounds on the expectation of convex functions of a multivariate random variable, Jensen's inequality (1906) was the first contribution and will be presented in chapter III. Additionally, A. Madansky's (1959) upper bound will be presented in chapter III. A. Madansky's upper bound can be viewed as a linear or first-order approximation. By requiring that the random variable have compact range, obtainable 1st moment information, and both with and without an independence assumption, Madansky was able to propose upper bounds in the multivariate case.

Ben-Tal and Hochman (1972) refined first-order bounds by requiring additional information on conditional means. Their contributions are to be presented in chapter IV. The last contribution to be discussed is the work of Gassman and Ziemba (1986) in chapter V. They were able to refine the Madansky upper bound while still only using 1st moment information. Additionally, they present results that allow for the relaxation of the independence assumption, refine bounds by obtaining bounds on subsets containing the sample space, and obtain bounds for convex domains which may be unbounded. Lastly, references for related research areas and recent advances will be provided should the reader desire further study.

CHAPTER II

BOUNDS IN A LINEAR PROGRAMMING CONTEXT

Before discussing the multivariate case, we shall look at an interesting formulation of the problem that makes use of mathematical programming. All of the techniques outlined in this thesis can be thought of as a solution to some form of the General Moment Problem (GMP). J. Dula (1992) poses this problem as follows:

General Moment Problem (GMP): Determine, should it exist, $\sup_{\mu \in M} \int_T h(t)\mu(dt)$, where M is a set of probability measures defined on the space T. The set M is determined by the conditions

$$\int_T g_j(t)\mu(dt) = y_j, \quad \text{j=1,...k,}$$

where $h, g_1, ..., g_k$ are real valued Borel measurable functions on a fixed measurable space T.

Sposito (1975) outlines a method for determining an upper bound to the above problem when X is a univariate bounded random variable and f(X) is a function of X, not necessarily convex, that we wish to determine the bound for the expectation of. A lower bound is not presented since Jensen's inequality, to be discussed in chapter III, generally provides a good lower bound.

So given a bounded random variable $X \in [a,b]$ where a,b are finite and the distribution is unknown, we wish to find an upper bound U for the expected value of f(X), i.e. $E[f(X_0)] \le U$ for some $X_0 \in [a,b]$. The condition imposed to solve this problem is that we know the expectation for n+1 functions of X, say $E[g_i(X)] = c_i$ (i=0,1,...,n). The relation between this problem and the GMP is obvious.

We first "linearize" the problem by assuming that there are y_i (i=0,...,n) such that

$$f(X) \le \sum_{i=0}^{n} y_i g_i(X) \qquad X \in [a,b].$$

So that when we take expectations we have

$$E[f(X)] \le \sum_{i=0}^{n} y_i E[g_i(X)] = y'c,$$

where c is a vector whose i-th element c_i is the expectation of $g_i(X)$. So to find our bound U, we wish to solve the following linear programming LP_1 minimization problem

find a vector $y' = (y_0, ..., y_n)$ to minimize $y'c$

subject to $f(x) \le \sum_{i=0}^{n} y_i g_i(x) \quad \forall x \in [a,b].$ (2.1)

In going from the random variable X to x we have insured that our "linearization" approach does indeed yield a sharp upper bound by considering our entire sample space. Additionally, we will impose that condition (2.1) be satisfied on some set of finite points $a = x_1 < x_2 < ... < x_{m-1} < x_m = b$ to make the problem computationally workable. Our new LP modified minimization problem is now

$$\min\ y'c$$

$$\text{s.t. } Ay \ge b$$

where $A = [a_{ji}] = [g_i(x_j)]$ and $b' = (f(x_1), f(x_2), ..., f(x_{m-1}), f(x_m))$ with the associated linear dual LD problem

$$\max\ w'b$$

$$\text{s.t. } A'w = c$$

$$w \ge 0$$

or

$$\max \sum_{j=1}^{m} w_j f(x_j)$$

$$\text{s.t. } \sum_{j=1}^{m} w_j g_0(x_j) = c_0$$

$$\vdots$$

$$\sum_{j=1}^{m} w_j g_n(x_j) = c_n$$

$$w_j \geq 0 \quad \text{for j=1,2,...,m.}$$

At this point, one should think of w as some discrete probability measure on [a,b]. Since the choice of our x_j (j=2,...,m-1) was arbitrary, we can think of our original LP_1 problem as the limiting form of our modified LP problem where the x_j can become large but still finite in number. Choosing the weights w_j, as in determining the optimal vector y, is a nontrivial matter however. Correspondingly for the associated dual problem as well. Thus to determine the limiting form of the LD problem we replace summations with integrals and denote the modified LD problem as LD_1

$$\max \int_a^b f(x)dw(x) \qquad \forall w(x)$$

$$\text{s.t.} \int_a^b g_0(x)dw(x) = c_0$$

$$\vdots$$

$$\int_a^b g_n(x)dw(x) = c_n$$

$$w(x) \geq 0 \qquad \forall x \in [a,b].$$

Now if y^0 and $w_0(x)$, an optimal vector with an attendant probability measure, are optimal finite solutions of LP_1 and its dual LD_1 respectively we have

$$y^0 c = \int_a^b f(x)dw_0(x) \leq \int_a^b [\sum_{i=0}^{n} y_i^0 g_i(x)]dw_0(x) = y^0 c$$

since f (X) was assumed to be less than or equal to the linear combination of y and g, so that

$$\int_a^b [f(x) - \sum_{i=0}^{n} y_i^0 g_i(x)]dw_0(x) = 0$$

or

$$\int_a^b h(x)dw_0(x) = 0$$

where

$$h(x) = f(x) - \sum_{i=0}^{n} y_i^0 g_i(x).$$

Since $h(x) \le 0 \ \ \forall x \in [a,b]$ we must have $w_0(x) = 0$ whenever $f(x) \ne \sum_{i=0}^{n} y_i^0 g_i(x)$.

Sposito interprets this to mean that "w_0 concentrates all its positive mass on the zeros of the function h(x)." We then determine y^0 by determining all the ways in which h(x)=0 and assign the weight points for w according to the constraints in the limited form of the LD problem. In practice this can be quite difficult for all but simple examples. We then choose the scheme with corresponding measure $w_0(x)$, not necessarily unique, which gives a maximum for $\int_a^b f(x)dw(x) = I$. So that if the roots for h, $s_0 = \{x \mid h(x) = 0\}$, which maximizes I with corresponding probability measure w_0 we have

$$y^0 c = \int_a^b f(x)dw_0(x) = \sum_{x_j \in s_0} (w_0)_j f(x_j) = U$$

where U is the promised upper bound.

In order to establish that LD_1 is the dual of LP_1 and LP_1 does indeed have a finite optimal solution, Sposito introduces the following theorem (which is a modified form of Theorems 1 and 2 of Kingman (1963) and stated without proof) when the following hold:

1. Let $\wp(c)$ is the set of random variables X in $\Omega \subset E^1$ such that $E[g_i(X)] = c_i$,i=0,...,n, with $g_0(X) = c_0 = 1 \ \left(\int_a^b g_0 dw(x) = 1\right)$.
2. Let $I = \{z \in E^{n+1} \mid z_1 = c_1, ..., z_n = c_n\}$.
3. Let $H = \{z \in E^{n+1} \mid z_0 = E[f(X)], z_1 = E[g_0(X)], ..., z_n = E[g_n(X)]\}$.

Theorem 2.1. If $I \cap H$ contains at least one point and $\wp(c) \ne \varnothing$, then either

(i) there exists n+1 numbers $y_0^0, y_1^0, ..., y_n^0$ such that

$$f(t) \le \sum_{i=0}^{n} y_i^0 g_i(t) \qquad \forall t \qquad (2.2)$$

and sup $E[f(x)] = y^0 c$ for $X \in \wp(c)$, or

(ii) no vector y satisfies (2.2) and sup $E[f(X)] \not< \infty$ for $X \in \wp(c)$.

Lemma 2.2. (Weak Duality) If there exists $y^0 \in E^{n+1}$ such that $f(t) \le \sum_{i=0}^{n} y_i^0 g_i(t)$ for all t, then sup $E[f(X)] \le \inf y'c$ where $X \in \wp(c)$ and y is restricted by (2.2).

Proof: Consider y such that $0 \le \sum_{i=0}^{n} y_i g_i(t) - f(t) \quad \forall t$.

Then $0 \le y'c - E[f(X)]$ or $E[f(X)] \le y'c$.

Therefore sup $E[f(X)] \le y'c$ for $X \in \wp(c)$

and sup $E[f(X)] \le \inf y'c$ for $X \in \wp(c)$, y restricted by (2.2).■

Condition (i) of Theorem 2.1 thus guarantees a solution $y^0 \in E^{n+1}$ in light of Lemma 2.2, i.e. $\exists y^0 \in E^{n+1}$ such that sup $E[f(X)] = \inf y'c = y^0 c$, y as restricted by (2.2). So that by letting $\Omega = [a, b]$, we know that if y^0 exists and solves the primal problem LP_1, its associated dual LD_1 has a finite optimal solution as well and

$$y^0 c = \int_a^b f(x) dw_0(x).$$

The results presented can be difficult to implement in practice. As an example, let us determine a bound for the second moment of a univariate random variable.

Example 2.1. Suppose we have a univariate bounded random variable X, a≤X≤b, where E[X]=μ and we wish to determine a bound for $E[X^2]$. Since $h(x) = x^2 - (y_0 + y_1 x) \le 0$ on [a,b] we know that h(x) can have at most two zeros on the interval [a,b]. Furthermore, since h is quadratic and nonpositive function for all x in [a,b], we observe that the roots for h(x) are x=a and x=b. The dual problem is

maximize $\int_a^b x^2 dw(x)$ over all probability measures w(x)

st $\int_a^b x\ dw(x) = \mu$

$\int_a^b dw(x) = 1.$

To find the measure $w_0(x)$ we must solve

$$w_0(a) + w_0(b) = 1$$

$$aw_0(a) + bw_0(b) = \mu$$

or

$$aw_0(a) + b[1 - w_0(a)] = \mu.$$

Therefore

$$w_0(a) = \frac{(b-\mu)}{(b-a)}$$

and

$$w_0(b) = 1 - w_0(a) = \frac{(\mu - a)}{(b-a)}.$$

So we have

$$\sup_w \int_a^b x^2 dw(x) = \int_a^b x^2 dw_0(x) = a^2 \frac{(b-\mu)}{(b-a)} + b^2 \frac{(\mu - a)}{(b-a)} = \mu(a+b) - ab.$$

CHAPTER III

PRELIMINARY BOUNDS

In this chapter our first results concerning the upper and lower bounds on the expectation of a convex function of a multivariate random variable will be presented. The main body of this chapter is to be found in Madansky (1959). Madansky was able to formulate his results by considering the moment spaces of multivariate distributions, an extension of the work by Dresher (1953). But first some definitions are needed.

Definition 3.1. A real function g defined on a segment (a,b), where $-\infty \le a < b \le \infty$, is called convex if the inequality holds

$$g((1-\lambda)x+\lambda y) \le (1-\lambda)g(x)+\lambda g(y)$$

whenever $a<x<b$, $a<y<b$, and $0 \le \lambda \le 1$. Equivalently, whenever $a<s<t<u<b$

$$\frac{g(t)-g(s)}{t-s} \le \frac{g(u)-g(t)}{u-t}.$$

Definition 3.2. A set E is said to be convex if whenever $x \in E$ and $y \in E$, and $0<\lambda<1$, the element $z_\lambda = (1-\lambda)x + \lambda y$ is in E as well.

Example 3.1. The constant function and the identity map are convex functions.

Should a function be twice differentiable, there is a convenient result that tests for convexity.

Proposition 3.1. Let f be a real-valued twice differentiable function defined on an open convex set $E \subset \Re^n$. Then f is convex if, and only if,

$$y' H_x y \ge 0 \qquad x \in E, y \in \Re^n,$$

where the matrix H is the Hessian matrix whose (i,j)th component is $\frac{\partial^2 f(x)}{\partial x_i \partial x_j}$.

Provided a function's Hessian matrix is positive semidefinite one can avoid the laborious

task of applying Definition 3.1. The next 2 examples will appear in later examples.

Example 3.2. Consider $f(x)=x^2$ where $f:\Re\rightarrow\Re$. By Proposition 3.1 f is convex.

Example 3.3. Consider $f(X)=\exp(X)$ where $f:\Re^n \rightarrow\Re$ where $X=X_1+\ldots X_n$. By Proposition 3.1 f is convex.

Definition 3.3. Given a set E, the convex hull H of E is the smallest convex set containing E.

To restate the problem of interest once again, we wish to determine a bound for the expectation $\int g(x_1,\ldots,x_n)dF(X_1,\ldots,X_n)$ where g is a convex function and F is a cumulative distribution function on $\Re^n$, i.e. determine finite U and L such that $L \le \int gdF \le U$.

Dresher (1953) presented a way, in the univariate case, in which an integral inequality like the one just mentioned can be geometrically interpreted. The bounds are obtained by dictating that a specified point lie in the convex hull of a given curve. The encountered difficulty was being able to specify the boundary of the appropriate convex hull. He was able to derive several inequalities, notably Jensen's inequality and the Holder and Minkowski inequalities as well, by exploiting the geometric properties that these inequalities share. They share the same set of boundaries, the so-called lower boundaries of a convex hull. The convex hull is determined by examining a corresponding moment space which will be defined shortly. Madansky (1959) then extended these results to include multivariate distributions. Jensen's inequality is the only lower bound presented in this chapter.

We first need some relevant background material concerning moment spaces of multivariate distributions. Let $F(X) = F(X_1,\ldots,X_r)$ be an r-variate cumulative distribution function on a bounded r-dimensional rectangle I, and $\{f_i(x_1,\ldots,x_r) = f_i(x), i = 1,\ldots,n\}$ be a set of n continuous functions on I.

Definition 3.4. The ith moment of F(X) with respect to $\{f_i(x)\}$ is:

$$\mu_i(F) = \int_I f_i(x)dF = E[f_i(x)].$$

Similarly we have the following definition of a moment space:

Definition 3.5. The nth moment space M_n with respect to $\{f_i(x)\}$ is the set of all points $\mu = (\mu_1,..., \mu_n)$ in n-dimensional Euclidean space, E^n, whose coordinates are the moments $\mu_1(F),...,\mu_n(F)$ with respect to $\{f_i(x)\}$ for some distribution function F(X), i.e.

$$\mu = \{\mu_i \mid \mu_i = E[f_i(x)], i = 1,...,n\}.$$

Now let C_n be the surface generated by $\{z_i = f_i(x), i = 1,..., n, x \in I\}$ and H_n be the convex hull of C_n. Using the results of Dresher, Karlin, and Shapley (1950), Madansky establishes that H_n is identical to M_n and M_n is closed, bounded, and convex. This link is crucial in that it establishes the convex hull based, for which we want to determine the boundary of, on the moment space. Therefore, since we wish to determine the bounds for the moment space, we will restrict the admissible functions to be continuous convex functions, i.e. so that the bounds for the expectation exist and are finite.

Let $g(X_1,..., X_r) = g(X)$ be a continuous convex function of an r-dimensional random variable X on a r-dimensional rectangle I and C_{r+1} the surface traced out in (r+1)Euclidean space by $z_1 = x_1, z_2 = x_2,....,z_r = x_r, z_{r+1} = g(x_1,....,x_r)$. As in Dresher (1953), Madansky then determines the boundary of the convex hull of C_{r+1}, H_{r+1}, in order to obtain the inequalities expressing E[g(X)] in terms of g(E(X)). We can see the direction that we would like to take by stipulating that g be continuous and convex on a bounded rectangle and now we want finite 1st moments as well.

Now a point x_0 is on the boundary of H_{r+1} if, and only if, $x_0 \in H_{r+1}$ and there exist β_i, i=0,....,r+1 such that

$$\sum_{i=1}^{r+1} \beta_i x_{0i} + \beta_0 = 0$$

and

$$\sum_{i=1}^{r+1} \beta_i x_i + \beta_0 \geq 0 \qquad \forall x \in I.$$

This can be geometrically interpreted as $x_0 \in H_{r+1}$ if, and only if, there is a supporting hyperplane to H_{r+1} at that point. The point x_0 is on the lower (upper) boundary of H_{r+1} provided H_{r+1} lies in the positive (negative) halfspace in relation to the supporting hyperplane to H_{r+1} at x_0. Once again let I be a bounded r-dimensional rectangle defined by the 2^r vertices of the form $(a^1_{\phi 1}, a^2_{\phi 2}, ..., a^r_{\phi r})$ where ϕi takes on either 1 or 2 for i=1,....,r. Furthermore, to preserve an ordering scheme, let $a^i_1 < a^i_2 \ \forall i$ and let $x_{r+1} = g(x_1, ..., x_r)$ be defined as before for all $(x_1, ..., x_r) \in I$. We claim that points of the form $(x_1, x_2, ..., x_r, g(x_1, ..., x_r))$ form the lower boundary of H_{r+1}. Clearly the formulated points are in the hull, i.e. $(x_1, x_2, ..., x_r, g(x_1, ..., x_r)) \in H_{r+1}$. Additionally, since g is continuous and convex, given any point $(x^0_1, x^0_2, ..., x^0_r, g(x^0_1, ..., x^0_r))$, there is a supporting hyperplane to H_{r+1}, not necessarily unique, at that point, specifically the plane tangent to the surface $x_{r+1} = g(x_1, ..., x_r))$ at that point. We have the 2r hyperplanes $x_{11} = a^1_1, x_{21} = a^1_2, x_{12} = a^2_1, x_{22} = a^2_2, ..., x_{1r} = a^r_1, x_{2r} = a^r_2$ to serve as bounds for H_{r+1} on its sides.

We are now able to characterize the upper boundary for H_{r+1} (Theorem 1 in Madansky (1959)) based upon the previous discussion.

Proposition 3.2. The upper boundary for H_{r+1} is identical with the upper boundary of the convex hull of the 2^r points $(a^1_{\phi 1}, ..., a^r_{\phi r}, g(a^1_{\phi 1}, ..., a^r_{\phi r}))$.

And if we let $H^*(x^0_1, ..., x^0_r)$ be the point at which the ray

$$(x_1 = x^0_1, x_2 = x^0_2, ..., x_r = x^0_r, x_{r+1} = \theta \mid -\infty \leq \theta \leq \infty)$$

intersects the upper boundary of H_{r+1} we are in the position of being able to state our first

main result.

Theorem 3.1. If $X = (X_1, ..., X_r)$ is a random vector such that Pr{X∈I}=1, and g(X) is a continuous convex function of X over the bounded r-dimensional rectangle I, then

$$g(E[X_1], ..., E[X_r]) \le E[g(X_1, ..., X_r)] \le H^*(E[X_1], ..., E[X_r]).$$

Proof: By the above discussion and Proposition 3.2, given an arbitrary point $(x_1, ..., x_{r+1}) \in H_{r+1}$, since $(x_1, ..., x_r, g(x_1, ..., x_r))$ is on the lower boundary and $H^*(x_1, ..., x_r)$ is on the upper boundary we have

$$(x_1, ..., x_r, g(x_1, ..., x_r)) \le (x_1, ..., x_{r+1}) \le (x_1, ..., x_r, H^*(x_1, ..., x_r)).$$

Since $H_{r+1} = M_{r+1}$ we can stipulate $(x_1, ..., x_{r+1})$ to be $(\mu_1, ..., \mu_{r+1})$ where

$$\mu_i = \int_I x_i dF = E[X_i]$$

and

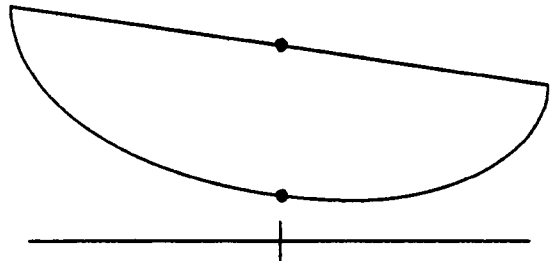
$$\mu_{r+1} = \int_I g(x_1, ..., x_r) dF = E[g(X_1, ..., X_r)]$$

for some distribution function F. ■

As stated in the introduction, the above result relies heavily on the fact that our random variables have compact range and obtainable and finite first moment information. One should also immediately recognize the left-handed inequality stated in the theorem as Jensen's inequality. In chapter V, we will be able to loosen the compact range restriction but at a cost.

Now let us consider the univariate case, i.e. r=1. In this instance, our moment space M_2 is the curve defined by $z_1 = x_1, z_2 = g(x_1)$ where $a_1^1 \le x_1 \le a_2^1$ and as previously indicated $a_1^1 < a_2^1$. Since we only have the 2 points $(a_1^1, g(a_1^1)), (a_2^1, g(a_2^1))$ to consider, the convex hull in this case is simply the line segment joining them. Edmundson (1956) was able to determine the upper bound in this case to be

$$E[g(X_1)] \le \frac{g(a_2^1) - g(a_1^1)}{a_2^1 - a_1^1}(E[X_1] - a_1^1) + g(a_1^1). \qquad (3.1)$$



In higher dimensions a precise description of the upper boundary for H_r is more difficult to visualize. One evaluates $H^*(x_1^0,...,x_r^0)$ by finding equations for the $\binom{2r}{r+1}$ hyperplanes by joining r+1 points of the form $(a_{\phi 1}^1,...,a_{\phi r}^r, g(a_{\phi 1}^1,...,a_{\phi r}^r))$. One then determines the maximum value of the (r+1)st coordinate on these hyperplanes where the first r coordinates are $(x_1^0,...,x_r^0)$. The maximum value obtained is $H^*(x_1^0,...,x_r^0)$.

In the bivariate case, r=2, we are able to give an explicit description of the upper boundary for the appropriate convex hull H_3.

Let $D = g(a_1^1, a_1^2) + g(a_2^1, a_2^2) - g(a_2^1, a_1^2) - g(a_1^1, a_2^2)$.

1. If D>0 then the upper boundary for H_3 is determined by

   $(a_1^1, a_1^2, g(a_1^1, a_1^2)), (a_2^1, a_2^2, g(a_2^1, a_2^2)), (a_2^1, a_1^2, g(a_2^1, a_1^2))$ when

   $x_2 \le a_1^2 + \left[\frac{a_2^2 - a_1^2}{a_2^1 - a_1^1}\right](x_1 - a_1^1)$ and by $(a_1^1, a_1^2, g(a_1^1, a_1^2)), (a_2^1, a_2^2, g(a_2^1, a_2^2)), (a_1^1, a_2^2, g(a_1^1, a_2^2))$

   for $x_2 \ge a_1^2 + \left[\frac{a_2^2 - a_1^2}{a_2^1 - a_1^1}\right](x_1 - a_1^1)$.

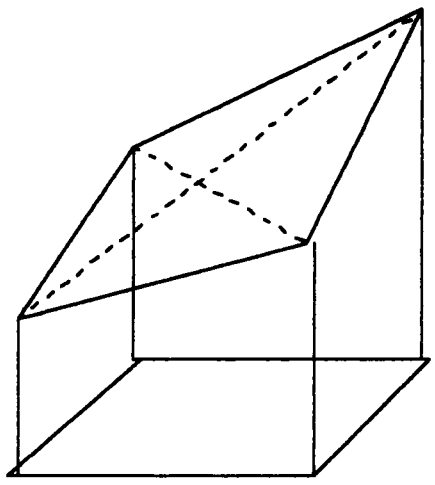
2. If D<0 then the upper boundary is determined by

$(a_1^1, a_1^2, g(a_1^1, a_1^2)), (a_2^1, a_1^2, g(a_2^1, a_1^2)), (a_1^1, a_2^1, g(a_1^1, a_2^1))$ for $x_2 \le a_2^2 - \left[\frac{a_2^2 - a_1^2}{a_2^1 - a_1^1}\right](x_1 - a_1^1)$

and by $a_2^1, a_2^2, g(a_2^1, a_2^2)), (a_2^1, a_1^2, g(a_2^1, a_1^2)), (a_1^1, a_2^2, g(a_1^1, a_2^2))$ for

$x_2 \ge a_2^2 - \left[\frac{a_2^2 - a_1^2}{a_2^1 - a_1^1}\right](x_1 - a_1^1)$.

3. If D=0 then the four points are coplanar.



Example 3.5. Let r=1, $g(X_1) = X_1^2$ and a=0, b=1. Using the inequality (3.1) we see that $E[g(X_1)] \le E[X_1]$. This is precisely the bound as computed in Example 2.1. But, the machinery needed to develop it is less formidable and more easily extended.

Example 3.6. Let r=2, $g(X_1, X_2) = e^{X_1 + X_2}$ with $a_1^1 = a_1^2 = 0, a_2^1 = a_2^2 = 1$. Then

$$E[e^{X_1+X_2}] \le \begin{cases} (e-1)[E(X_1) + E(X_2)] + (e-1)^2 E(X_2) + 1 & \text{if } E(X_1) \ge E(X_2) \\ (e-1)[E(X_1) + E(X_2)] + (e-1)^2 E(X_1) + 1 & \text{if } E(X_2) \ge E(X_1) \end{cases}.$$

As we can see from the bivariate case, r=2, we have

$$g(X_1) \le \frac{X_1 - a_1^1}{a_2^1 - a_1^1} g(a_2^1) + \frac{a_2^1 - X_1}{a_2^1 - a_1^1} g(a_1^1).$$

So that for convex $g(X_1, ..., X_r)$ one obtains

$$g(X_1,....,X_r) \le \frac{X_r - a_1^r}{a_2^r - a_1^r} g(X_1,....,X_{r-1},a_2^r) + \frac{a_2^r - X_r}{a_2^r - a_1^r} g(X_1,....,X_{r-1},a_1^r)$$

for fixed $X_1,....,X_{r-1}$. And for fixed $X_1,....,X_{j-1}$, and g convex in X_j for

$$g(X_1,....,X_j,a_{\phi j+1}^{j+1},.....,a_{\phi r}^r)$$

one uses the above bound successively to obtain

$$g(X_1,....,X_r) \le \sum_{\phi} \prod_{j=1}^{r} (-1)^{\bar{\phi}} \frac{(a_{\phi}^j - X_j)}{a_2^j - a_1^j} g(a_{\phi 1}^1,...,a_{\bar{\phi} r}^r)$$

for $\bar{\phi} i = 3 - \phi i$. So that if the components of our random vector are independent, by the properties of expectation we have

$$E[g(X_1,....,X_r)] \le \sum_{\phi} \prod_{j=1}^{r} (-1)^{\phi j} \frac{(a_{\phi j}^j - E[X_j])}{a_2^j - a_1^j} g(a_{\phi 1}^1,...,a_{\bar{\phi} r}^r) = H(E[X]).$$

Applying this to the bivariate Example 3.6 we obtain a sharper bound of

$$E[e^{X_1+X_2}] \le (e-1)[E(X_1) + E(X_2)] + (e-1)^2 E(X_1)E(X_2) + 1.$$

Notice that no distributional assumption is made.

Madansky claims this is true in general by noting that

$$\sum_{\phi} \prod_{j=1}^{r} (-1)^{\phi j} \frac{(a_{\phi j}^j - E[X_j])}{a_2^j - a_1^j} = 1$$

and

$$\prod_{j=1}^{r} (-1)^{\phi j} \frac{(a_{\phi j}^j - E[X_j])}{a_2^j - a_1^j} \ge 0.$$

Therefore the point $(E[X_1],...,E[X_r], H(E[X]))$ is in the convex hull of the 2^r points $(a_{\phi 1}^1,....., a_{\phi r}^r, g(a_{\phi 1}^1,....., a_{\phi r}^r))$ and so $H(E[X]) \le H^*(E[X])$.

CHAPTER IV

REFINED BOUNDS

In chapter III, an upper bound due to Madansky was presented under the assumption that our multivariate random variable have a compact range with finite and obtainable 1st moment information. The bound was then sharpened should the components be independent. Ben-Tal and Hochman (1972) were able to extend Madansky's bound by requiring that additional information be known, namely information about the absolute deviation from the mean. Using this added information, Madansky's upper bound, and Jensen's inequality they were able to provide both a better upper and lower bound for the expectation of a convex function of a multivariate random variable. An example due to Mangasarian (1964) will be discussed that makes use of the improved bounds.

Once again let $T = (T_1,...,T_r)$ be a multivariate random variable with finite first moment and F_T its distribution function. Now for $-\infty < a < b < \infty$, $a < \mu < b$ one defines the following sets:

$$\bar{\theta}_1 = \{F_T \mid \int_a^b dF_T = 1\},$$

$$\bar{\theta}_2 = \{F_T \mid \int_a^b t dF_T = \mu\} \cap \bar{\theta}_1,$$

$$\bar{\theta}_3 = \{F_T \mid \int_a^b |t - \mu| \, dF_T = d\} \cap \bar{\theta}_2,$$

and

$$\bar{\theta}_4 = \{F_T \mid \int_\mu^b dF_T = \beta\} \cap \bar{\theta}_3.$$

One can immediately see that the $\bar{\theta}_1$, $\bar{\theta}_2$ correspond to the finite expectation property as imposed by Madansky. We will also require that $0 < d \leq d_{\max}$ and $\underline{\beta} \leq \beta \leq \bar{\beta}$ where these bounds are to be discussed later. The inclusion of the sets $\bar{\theta}_3$ and $\bar{\theta}_4$ provides us with the additional information needed to refine Madansky's bound.

Similarly we assume that $g(T_1,...,T_r) = g(T)$ is a convex function in [a,b]. We will then address the issue of bounding E[g(T)] by determining the extrema on each of four previously mention sets, i.e. determine

$$M_i = Ext_{\theta_i}\{E_{F_T} g(T)\},\ i=1,2,3,4$$

into the corresponding

$$\underline{M}_i = \min_{\theta_i}\{E_{F_T} g(T)\},\ i=1,2,3,4$$

$$\overline{M}_i = \max_{\theta_i}\{E_{F_T} g(T)\},\ i=1,2,3,4.$$

The bounds for the extrema of M_1 and M_2 are easily obtained using the fact that g is convex on [a,b] and using the results from chapter III. The extrema for M_1 are therefore determined to be

$$\underline{M}_1 = \min_{\theta_1} E[g(T)] = \min_{\theta_1} \int_a^b g dF_T = \min_{a \le t \le b} g(t)$$

and similarly

$$\overline{M}_1 = \max_{\theta_1} E[g(T)] = \max_{a \le t \le b} g(t).$$

For determining the bounds for M_2 we will appeal to the results from chapter III and make use of the bounds developed there. We have

$$\underline{M}_2 = \min_{\theta_2} E[g(T)] \ge \min_{\theta_2} g(E[T]) = g(\mu) \text{ by Jensen's inequality}$$

and

$$\overline{M}_2 = \frac{b-\mu}{b-a} g(a) + \frac{\mu - a}{b-a} g(b)$$

which is precisely Madansky's upper bound.

In order to determine bounds for M_3 and M_4, Ben-Tal and Hochman require that information about d and β be known. If the bounds arise in connection with a stochastic programming problem, it is up to the user to specify these constants based on prior distributional knowledge or compute them directly. In principle this may be an onerous

assumption since $|t-\mu|$ is a convex function of t. Nonetheless, since

$$\bar{\theta}_4 \subset \bar{\theta}_3 \subset \bar{\theta}_2 \subset \bar{\theta}_1$$

we have

$$\underline{M}_1 \le \underline{M}_2 \le \underline{M}_3 \le \underline{M}_4 \le E[g(T)] \le \overline{M}_4 \le \overline{M}_3 \le \overline{M}_2 \le \overline{M}_1$$

and hence $\underline{M}_i$ and $\overline{M}_i$ are improved bounds for i=3,4. In order to make our computations more manageable let $X = T - \mu$ so that our mean for M_i is zero. Thus

$$Ext_{\bar{\theta}_i} E_{F_T} g(T) = Ext_{\bar{\theta}_i} E_{F_X} g(\mu + X) \quad i=3,4$$

where

$$\theta_3 = \{F_X \mid \int_{a-\mu}^{b-\mu} dF_X = 1, \int_{a-\mu}^{b-\mu} x dF_X = 0, \int_{a-\mu}^{b-\mu} |x| dF_X = d\}$$

and

$$\theta_4 = \{F_X \mid \int_0^{b-\mu} dF_X = \beta\} \cap \theta_3 .$$

We are then in a position to determine our first improved lower bound.

Theorem 4.1. $\underline{M}_4 = \beta g(\mu + \frac{d}{2\beta}) + (1-\beta) g(\mu - \frac{d}{2(1-\beta)})$.

Proof: Let $F \in \theta_4$ and consider the probability function $\hat{F}$ which assigns probability

$$P_1 = F(b-\mu) - F(0) = \beta$$

to the point

$$x_1 = \frac{1}{\beta} \int_0^{b-\mu} x dF_X \qquad 0 \le x_1 \le b-\mu$$

and probability

$$P_2 = F(0) - F(\mu - a) = 1 - \beta$$

to the point

$$x_2 = \frac{1}{1-\beta} \int_{\mu-a}^{0} x dF_X \qquad \mu - a \le x_2 < 0 .$$

Then the expected value of $\hat{F}$ is $E_{\hat{F}}(X) = x_1P_1 + x_2P_2 = \int_{a-\mu}^{b-\mu} x dF_X = E_F(X) = 0$ and

$E_{\hat{F}}(|X|) = |x_1|P_1 + |x_2|P_2 = \int_{a-\mu}^{b-\mu} |x| dF_X = E_F(|X|) = d$. Therefore $\hat{F} \in \theta_4$. So that when we take the expectation of g(T) in terms of X we find

$$\begin{aligned} E_F g(X+\mu) &= \int_{a-\mu}^{b-\mu} g(x+\mu) dF_X \\ &= \int_{a-\mu}^{0} g(x+\mu) dF_X + \int_{0}^{b-\mu} g(x+\mu) dF_X \\ &= (1-\beta)\int_{a-\mu}^{0} g(x+\mu) d\left(\frac{F_X}{1-\beta}\right) + \beta \int_{0}^{b-\mu} g(x+\mu) d\left(\frac{F_X}{\beta}\right) \end{aligned}$$

and applying Jensen's inequality to each of the above integrals

$$\begin{aligned} &\geq (1-\beta) g\left[\mu + \int_{a-\mu}^{0} x\, d\left(\frac{F_X}{1-\beta}\right)\right] + \beta g\left[\mu + \int_{0}^{b-\mu} x d\left(\frac{F_X}{\beta}\right)\right] \\ &= P_2 g(\mu + x_2) + P_1 g(\mu + x_1) \qquad (4.1) \\ &= E_{\hat{F}} g(X+\mu) = E_F(T). \end{aligned}$$

So that for any $F \in \theta_4, \hat{F} \in \theta_4$ yields a lower bound to E[g(T)]. The result is proved by noting that

$$-\int_{a-\mu}^{0} x dF_X = \int_{0}^{b-\mu} x dF_X = \frac{1}{2} d$$

and substituting this into equation (4.1). ∎

The lower bound for M_3 will be determined using Theorem 4.1. But first we will need two lemmas.

Lemma 4.1. $d \leq d_{\max} = \dfrac{2(b-\mu)(\mu-a)}{b-a}$.

Proof: $d = E|T-\mu| \leq \dfrac{b-\mu}{b-a}|a-\mu| + \dfrac{\mu-a}{b-a}|b-\mu|$

$$= \frac{(b-\mu)(\mu-a)+(\mu-a)(b-\mu)}{b-a}$$

$$= \frac{2(b-\mu)(\mu-a)}{b-a} = d_{\max}.$$

The above inequality follows from Madansky's bound since $g(T) = |T - \mu|$ is obviously convex. ■

And now for our second lemma:

Lemma 4.2. $\underline{\beta} = \frac{d}{2(b-\mu)} \le \beta \le 1 - \frac{d}{2(\mu-a)} = \overline{\beta}.$

Proof: Once again noting

$$-\int_{a-\mu}^{0} x dF_X = \int_{0}^{b-\mu} x dF_X = \frac{1}{2} d,$$

we have

$$\beta = \int_{0}^{b-\mu} dF_X = \frac{1}{b-\mu}\int_{0}^{b-\mu} (b-\mu) dF_X$$

$$\ge \frac{1}{b-\mu}\int_{0}^{b-\mu} x dF_X = \frac{d}{2(b-\mu)} = \underline{\beta}$$

and

$$1-\beta = \int_{a-\mu}^{0} dF_X = \frac{1}{\mu-a}\int_{a-\mu}^{0} (\mu-a) dF_X$$

$$\ge \frac{1}{\mu-a}\int_{a-\mu}^{0} (-x) dF_X = \frac{d}{2(\mu-a)}$$

hence $\beta \le 1 - \frac{d}{2(\mu-a)} = \overline{\beta}.$

Using Lemma 4.1 we see that

$$\overline{\beta} - \underline{\beta} = 1 - \frac{d}{2}\left(\frac{1}{\mu-a} + \frac{1}{b-\mu}\right) = 1 - \frac{d}{2}\frac{b-a}{(\mu-a)(b-\mu)}$$

$$\ge 1 - \frac{d_{\max}(b-a)}{2(\mu-a)(b-\mu)} = 1 - 1 = 0.$$

Since $\underline{\beta} > 0$ and $\overline{\beta} < 1$ the lemma is proven. ■

Using Lemma 4.2 and Theorem 4.1 we see that the bound for $\underline{M}_3$ is determined to be

$$\underline{M}_3 = \min_{\underline{\beta} \le \beta \le \overline{\beta}} \underline{M}_4(\beta) = \min_{\underline{\beta} \le \beta \le \overline{\beta}} \left(\beta g(\mu + \frac{d}{2\beta}) + (1-\beta) g(\mu - \frac{d}{2(1-\beta)}) \right).$$

Ben-Tal and Hochman were also able to show that the above lower bound has a unique solution.

Theorem 4.2. $\overline{M}_3 = g(\mu) + \frac{d}{2}\left[\frac{g(b) - g(\mu)}{b - \mu} + \frac{g(a) - g(\mu)}{\mu - a} \right]$.

Proof: Let $F \in \theta_3$ and consider the probability function F^* which assigns probability

$$P_1 = \frac{1}{\mu - a} \int_{a-\mu}^{0} -x dF_X \qquad \text{to } x_1 = (a - \mu)$$

$$P_2 = \frac{1}{b - \mu} \int_{0}^{b-\mu} x dF_X \qquad \text{to } x_2 = (b - \mu)$$

$$P_3 = 1 - P_1 - P_2 \qquad \text{to } 0.$$

Noting that $P_1 > 0$, $P_2 > 0$, $P_1 + P_2 < 1$ we have that F^* is well defined. Since $E_{F^*}(x) = x_1 P_1 + x_2 P_2 + 0 P_3 = 0$ and $E_{F^*}(|x|) = d$ we have $F^* \in \theta_3$.

Applying Madansky's upper bound we see that

$$E_F g(X + \mu) = \int_{a-\mu}^{b-\mu} g(x + \mu) dF$$

$$= (1 - \beta) \int_{a-\mu}^{0} g(x + \mu) d\left(\frac{F_X}{1 - \beta} \right) + \beta \int_{0}^{b-\mu} g(x + \mu) d\left(\frac{F_X}{\beta} \right)$$

$$\le \frac{(1-\beta)}{\mu - a} \left[\left(0 - \int_{a-\mu}^{0} x d\left(\frac{F_X}{1 - \beta} \right) \right) g(a) + \left(\int_{a-\mu}^{0} x d\left(\frac{F_X}{1 - \beta} \right) - (a - \mu) \right) g(\mu) \right]$$

$$+ \frac{\beta}{b - \mu} \left[\left((b - \mu) - \int_{0}^{b-\mu} x d\left(\frac{F_X}{\beta} \right) \right) g(\mu) + \left(\int_{0}^{b-\mu} x d\left(\frac{F_X}{\beta} \right) - 0 \right) g(b) \right]$$

$$= g(x_1 + \mu)P_1 + g(x_2 + \mu)P_2 + g(0 + \mu)P_3 = E_{F^*} g(x + \mu).$$

So for any $F \in \theta_3$, $F^* \in \theta_3$ yields a lower bound for E[g(T)]. The theorem is obtained by substituting

$$x_1 = a - \mu, \qquad x_2 = b - \mu, \qquad P_1 = \frac{d}{2(\mu - a)}, \qquad P_2 = \frac{d}{2(b - \mu)}$$

in $E_{F^*} g(x + \mu)$. ■

We can similarly obtain a bound for $\overline{M}_4$. If we attached the probability associated with $0\ (1 - P_1 - P_2)$ to the points $\pm \varepsilon$ such that the probability that $x \geq 0$ were equal to β, i.e. $\Pr(x \geq 0) = \beta$ we have that $F^* \in \theta_4$. Letting $\varepsilon \to 0$ we are able to conclude that

$$\overline{M}_4 = \sup_{\theta_4} E[g(T)] = \max_{\theta_3} E[g(T)] = \overline{M}_3.$$

Ben-Tal and Hochman also extended Madansky's results in the independent random vector component case. Let $T_1, T_2, ..., T_n$ be independent random variables of finite mean, with the associated parameters of $[a_i, b_i]$, μ_i, d_i, and β_i for each T_i, and $g(T_1, T_2, ..., T_n)$ be a convex function of the vector $(T_1, T_2, ..., T_n)$. As performed by Madansky, one can obtain bounds for $E[g(T_1, T_2, ..., T_n)]$ by bounding $E[g(T_1, t_2, ..., t_n)]$, i.e. fix $t_2, ..., t_n$, and repeat this procedure successively n times. Operationalizing this procedure one can obtain the following bounds which are stated without proof.

$$\underline{M}_4 = \sum_{\Phi_2} \left(\prod_{k=1}^{n} \beta_{\phi k}^{k} \right) g(a_{\phi 1}^{1}, a_{\phi 2}^{2}, ..., a_{\phi n}^{n})$$

where ϕk is 1 or 2,

Φ_2 is the set of 2^n n-dimensional vectors whose components are all 1's and/or 2's.

$$\beta_1^k = \beta_k \qquad \beta_2^k = 1 - \beta_k$$

$$a_1^k = \mu_k + \frac{d_k}{2\beta_k} \qquad a_2^k = \mu_k + \frac{d_k}{2(1 - \beta_k)};$$

$\underline{M}_3 = \min_{\underline{\beta}_k \le \beta \le \overline{\beta}_k} \underline{M}_4(\beta_1,...,\beta_n)$ for k=1,2,...,n and

$$\overline{M}_4 = \overline{M}_3 = \sum_{\Phi_3} \left(\prod_{k=1}^{n} P_{\phi k}^{k} \right) g(a_{\phi 1}^{1}, a_{\phi 2}^{2}, ..., a_{\phi n}^{n})$$

where ϕk is 1,2, or 3,

Φ_3 is the set of 3^n n-dimensional vectors whose components are all 1's, 2's and/or 3's,

$$P_1^k = \frac{d_k}{2(\mu_k - a_k)}, \qquad a_1^k = a_k,$$

$$P_2^k = \frac{d_k}{2(b_k - \mu_k)}, \qquad a_2^k = b_k,$$

$$P_3^k = 1 - P_1^k - P_2^k, \qquad a_3^k = \mu_k.$$

As an example of the above results let us consider an example that is due to Mangasarian (1964). Consider the following problem:

$$E\max_x \{g(c,x) \mid h(x) \ge 0\}$$

where g is a scalar function of a n-dimensional random vector c of unknown distribution and a m-dimensional vector x and h is a k-dimensional vector whose components are scalar functions of x. The proposed question of interest is finding the expected value of perfect information which is given as $E\max_x \{g(c,x) \mid h(x) = 0\} - E\max_x \{g(c,x) \mid h(x) \ge 0\}$. Should c have a continuous distribution, it may prove difficult to evaluate the perfect information explicitly since it may require the solution of an infinite number of programming problems.

Mangasarian proved that should g(c,x) be convex in c for each feasible x, and if c is defined over a bounded n-dimensional rectangle whose components are independent random variables, then denoting $\alpha(c) = E\max_x \{g(c,x) \mid h(x) \ge 0\}$ we have

$$\overline{M}_2 \ge E[\alpha(c)] \ge \max_x \{E[g(c,x)] \mid h(x) \ge 0\} \equiv E[g(c\bar{x}(E(c)))] \ge \underline{M}_2$$

where $\bar{x}(E(c))$ is the solution of $\max_x\{E[g(c,x)] \mid h(x) \geq 0\}$ and $\overline{M}_2$, $\underline{M}_2$ are the Madansky and Jensen's bound respectively for $\alpha(c)$. Since $\alpha(c)$ is convex we can apply the previously developed bounds to obtain sharper upper bounds via $\overline{M}_3$ and $\underline{M}_4$.

Example 4.1. Consider $E[\max_{x_1,x_2}\{c_1x_1^2 + c_2x_2^2\}]$

such that $0 \leq x_1 \leq 10, 0 \leq x_2 \leq 5$ where both c_i are uniformly distributed as follows, $c_1 \sim U[-1/2, 1/2], c_2 \sim U[0,1]$. Mangasarian determined the numerical bounds and the exact solution to be as follows

$$37.5 \geq 20.833 = E[\max_x\{g(c,x) \mid h(x) \geq 0\}] \geq 8.33 \geq 6.25 = 6.25.$$

So the true value of 20.833 lies in the interval [8.33,37.5] which is rather broad. Applying Ben-Tal and Hochman's bound we obtain $\overline{M}_3$ and $\underline{M}_4$ of 21.875 and 20.312 respectively. The new interval of [20.312,21.875] is a substantial reduction.

CHAPTER V

A FURTHER REFINEMENT WITH EXTENSION

The purpose of this chapter is to discuss selected contributions of Gassman and Ziemba with regards to bound the convex expectation of a multivariate random variable. In Gassman and Ziemba (1986), results were obtained that provided a further refinement to Madansky's bounds. Improved bounds were obtained by partitioning the sample space and computing bounds for the respective subsets, Madansky's independence assumption is relaxed, and bounds were obtained that are applicable to random variables with convex domains which may be unbounded. Unlike the Ben-Tal and Hochman (1972) bounds discussed in the preceding chapter, these bounds will be derived requiring only obtainable first moment information.

Once again let $X = (X_1,...,X_n)$ be a multivariate random variable with distribution function F with finite mean μ. Let U be the support of X, i.e. the smallest closed subset such that $\int_U dF = 1$. Supposing that U is convex, let g be a convex function where $E[g(X)] = \bar{g}$ exists and is finite.

Madansky (1959) treated the case where U was compact and where the components of X were both independent and not necessarily independent. Briefly restating those results, with n=1 yields

$$E[g(X)] = \bar{g} \le \frac{b-\mu}{b-a}g(a) + \frac{\mu-a}{b-a}g(b)$$

where U=[a,b]. For finite n, where $X = (X_1,...,X_n)$ and whose components are independent, and for the n-dimensional rectangle $U = \prod_{i=1}^{n}[a_i,b_i]$ we have

$$\bar{g} \le \sum_{\phi 1=0}^{1}\cdots\sum_{\phi n=0}^{1}\left(\prod_{k=1}^{n}\delta_{\phi k}^{k}\right)g(d_{\phi 1}^{1},...,d_{\phi n}^{n})$$

where $d_0^i = a_i$, $d_1^i = b_i$, $\delta_0^k = \frac{b_k - \mu}{b_k - a_k}$ and $\delta_1^k = \frac{\mu - a_k}{b_k - a_k}$.

Let Gr(g) be the graph of g, i.e. Gr(g)={(x,y)|x∈U,y =g(x)} and let H be its convex hull. Since integration with respect to a probability measure can be viewed as taking a convex combination, $(\mu,\bar{g}) \in H$ provided μ and $\bar{g}$ exist. One can therefore see that $\bar{g} \le \sup\{y|\ (\mu,y)\in H\}$. Madansky assumed that U was convex and compact to guarantee that the sup{y| (x,y)∈H}=H*(X) would exist and be finite. Our first result relaxes the restrictions on U but requires that it be contained in a bounded convex polyhedron.

Theorem 5.1. Let V be a bounded convex polyhedron containing U and $\{v_1,...,v_m\}$, where $\{v_1,...,v_m\}$ are the extreme points of V, i.e. a point whose exclusion doesn't affect the set's convexity, of V. Assume g can be extended to a convex function from V into $\Re$ which shall be denoted by g as well. Then

$$\bar{g} \le z = \max\{\sum_{i=1}^{m} g(v_i)\lambda_i \mid \lambda_i \ge 0, \sum \lambda_i = 1, \sum \lambda_i v_i = \mu\}.$$

Proof: Given (x,y)∈H, there exist k, $\alpha_i \ge 0$ and $u_i \in U$, i=1,...,k such that $\sum_{i=1}^{k} \alpha_i = 1, \sum_{i=1}^{k} \alpha_i u_i = x, \sum_{i=1}^{k} \alpha_i g(u_i) = y$. Now for every i, there exist $u_{ij} > 0$, j=1,...,m such that we have $\sum_{j}^{m} u_{ij} = 1, \sum_{j}^{m} u_{ij} v_j = u_i$, i.e. $x = \sum_i \sum_j \alpha_i u_{ij} v_j$. Since g is convex

$$y = \sum_{i=1}^{k} g(u_i)\alpha_i = \sum_i \alpha_i g\left(\sum_j u_{ij} v_j\right) \le \sum_i \sum_j \alpha_i u_{ij} g(v_j).$$

Setting $\hat{\lambda}_j = \sum_i u_{ij}\alpha_i$ gives $x = \sum_j \hat{\lambda}_j v_j, \sum_j \hat{\lambda}_j = 1, \hat{\lambda}_j \ge 0$, and

$$y \le \sum_j \hat{\lambda}_j g(v_j) \le \max\left\{\sum_{i=1} g(v_j)\lambda_j \mid \lambda_j \ge 0, \sum \lambda_j = 1, \sum \lambda_j v_j = x\right\}.$$

One should also note that this holds for $(x,y)=(\mu,H^*(\mu))\in H$! ■

It is worth noting that the above result shows Madansky's bound to be a special case. Should U be polyhedral, say a n-dimensional rectangle, then $H^*(\mu)=z$. Should $n=1$ so that $U=[a,b]$, $z=\max\{\lambda_1 g(a) + \lambda_2 g(b) \mid \lambda_1 + \lambda_2 = 1, \lambda_1 a + \lambda_2 b = \mu, \lambda_1, \lambda_2 \ge 0\}$. Thus $\lambda_1 = \frac{b-\mu}{b-a}, \lambda_2 = 1 - \lambda_1 = \frac{\mu - a}{b-a}$, to derive the familiar Madansky bound of

$$z = \frac{b-\mu}{b-a} g(a) + \frac{\mu - a}{b-a} g(b).$$

Should V be a simplex, a unique convex combination of the extreme points $\{v_i\}$ yields μ so that the maximization process of Theorem 5.1 is trivial. Lastly, Birge and Wets (1986) provided a related result on a family $\{v_x(\bullet), x \in U\}$ of probability measures on the set of extreme points of U, ext U, such that $\int_{ext U} s v_x(ds) = x \quad \forall x \in U$. This last observation serves as a computational starting point in determining bounds. One can compute various bounds dependent upon the shape of V and its associated extreme points. Similar to the problem faced by Sposito (1975) in chapter II, determining the optimal probability measure is a nontrivial matter which has yet to be resolved. The weights λ_i in Theorem 5.1 can be thought of as the probability measure on ext $V=\{v_1,\ldots,v_m\}$ where $\int_{ext U} s v_x(ds) = x$ holds only at $x=\mu$.

Example 5.1. Let $g(x,y) = e^x$ and density f be defined as

$$f(x,y) = \begin{cases} 3\sqrt{1-x^2-y^2}/2\pi & x^2 + y^2 \le 1 \\ 0 & ow \end{cases}.$$

Then $U = \{(x,y) \mid x^2 + y^2 \le 1\}$ and we are able to explicitly compute the E[g(x,y)] to be

$$\bar{g} = \frac{3}{2\pi}\int_{-1}^{1}\int_{-\sqrt{1-x^2}}^{\sqrt{1-x^2}} e^x \sqrt{1-x^2-y^2}\,dydx = \frac{3}{e} \approx 1.1036.$$

Applying Jensen's inequality we see that $\bar{g} \ge g(\mu) = g(0,0) = 1$. To apply Theorem 5.1 we need to specify a convex polyhedra which contains U. Consider the following convex sets defined as

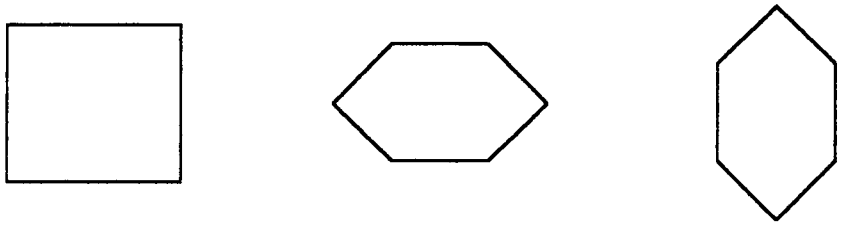
$V_1 = Co\{(-1,1),(1,-1),(-1,1),(1,1)\}$,

$$V_2 = Co\left\{\left(\frac{-\sqrt{3}}{3},-1\right),\left(\frac{\sqrt{3}}{3},-1\right),\left(\frac{2\sqrt{3}}{3},0\right),\left(\frac{\sqrt{3}}{3},1\right),\left(\frac{-\sqrt{3}}{3},1\right),\left(\frac{-2\sqrt{3}}{3},0\right)\right\},$$

and

$$V_3 = Co\left\{\left(-1,\frac{-\sqrt{3}}{3}\right),(-1,1),\left(0,\frac{2\sqrt{3}}{3}\right),\left(1,\frac{\sqrt{3}}{3}\right),\left(1,\frac{-\sqrt{3}}{3}\right),\left(0,\frac{-2\sqrt{3}}{3}\right)\right\}.$$

Geometrically we have



V_1 V_2 V_3

where V_3 is obtained by rotating V_2 90 degrees. The upper bound on V_1 was determined to be $z_1 = \frac{e^2+1}{2e} \approx 1.5431$ using Theorem 5.1. Bounds for V_2 and V_3 were found to be $z_2 = \frac{1}{2}\left(e^{\frac{2}{3}\sqrt{3}} + e^{\frac{-2}{3}\sqrt{3}}\right) \approx 1.7441$ and $z_3 = \frac{e^2+1}{2e} \approx 1.5431$ respectively. One immediately notices that the bounds produced are dependent upon the choice of polyhedra, not invariant under rotation, and a smaller area does not necessarily yield a tighter bound (the area for $V_2 = 2\sqrt{3}$ which is less than the area for $V_1 = 4$ but $\bar{g} \approx 1.1036 < z_1 < z_2$).

Example 5.2. Let $g(x,y) = (x+y)^2 + x + 2y$ with a distribution function F given by

$$F(x,y) = \begin{cases} 1 & 0 \le x,y \le 1 \\ 0 & ow \end{cases}.$$

Since X and Y are independent we can compute the bound using Theorem 5.1 and compare it with the Madansky upper bound developed in presented in chapter III. First, Jensen's inequality yields a lower bound of $\bar{g} \ge g\left(\frac{1}{2},\frac{1}{2}\right) = 2.5$. Applying Madansky's upper bound in the independence case we see that

$\bar{g} \le \frac{1}{4}[g(0,0) + g(0,1) + g(1,0) + g(1,1)] = 3$ whereas if we apply Theorem 5.1

$$\bar{g} \le \max\{3\lambda_2 + 2\lambda_3 + 7\lambda_4 \mid \lambda_1 + \lambda_2 + \lambda_3 + \lambda_4 = 1, \lambda_3 + \lambda_4 = \frac{1}{2}, \lambda_2 + \lambda_4 = \frac{1}{2}\}$$

noting $\lambda_4 = \frac{1}{2} - \lambda_2, \lambda_3 = \lambda_2$ and substituting into $3\lambda_2 + 2\lambda_3 + 7\lambda_4$ yields

$$= \max\{\frac{7}{2} - 2\lambda_2 \mid 0 \le \lambda_2 \le \frac{1}{2}\} = 3.5.$$

Now

$$\bar{g} = \int_{I\times I} (x+y)^2 + x + 2y\,dx\,dy = \frac{8}{3} \approx 2.6667.$$

As we can see, Madansky's bound is better in this instance. A lack of consistency is regrettable.

In light of Example 5.1, the shape/orientation of the convex polyhedra is crucial to determining the quality of an upper bound. The bound of Theorem 5.1 can be sharpened by partitioning V into convex polyhedra V^j, j=1,...,k such that $V = \bigcup_{j=1}^{k} V^j$, $V^i \cap V^j = \varnothing$ for $i \ne j$.

If we let $\eta_j = \int_{V^j} dF$ with corresponding $\mu_j = \frac{1}{\eta_j}\int_{V^j} x dF$ for j=1,...,k and z_j as the upper bound for $E[g \mid V^j]$ as obtained from Theorem 5.1 then we have the following proposition

Proposition 5.1. $\sum_{j=1}^{k} \eta_j g(\mu_j) \le \bar{g} \le \sum_{j=1}^{k} \eta_j z_j$.

The lower bounds are obtained by applying Jensen's inequality on the subsets which comprise the partition and the upper bound is a 'weighted' sum of subset upper bounds generated by Theorem 5.1. If $\eta_j = 0$ we set $\mu_j = \mu$ to remove any ambiguity.

Example 5.3. Let $g(x,y) = (x+y)^2 + x + 2y$ with an associated density of

$$f(x,y) = \begin{cases} \frac{1}{\pi} & x^2 + y^2 \le 1 \\ 0 & ow \end{cases}.$$

Let V_1 be as before and $V_4 = Co\{\left(\frac{-1}{2},-1\right),(1,-1),\left(1,\frac{1}{2}\right),\left(\frac{1}{2},1\right),(-1,1),\left(-1,\frac{-1}{2}\right)\}$ which is geometrically



V_4.

Now

$$\bar{g} = \frac{1}{\pi}\int_{-1}^{1}\int_{-\sqrt{1-x^2}}^{\sqrt{1-x^2}}((x+y)^2 + x + 2y)dxdy = \frac{4\sqrt{2}}{3\pi} \approx 0.6002.$$

Applying Jensen's inequality and Theorem 5.1 we see that $\bar{g} \ge g(0,0) = 0$ and $z_1 = 4$ with $z_4 = 2.25$. Neither upper bound appears to be that sharp. Now consider the following partitions of V_1 and V_4. Let

$$V_1^1 = Co\{(1,1),(1,-1),(-1,1)\},$$

$$V_1^2 = Co\{(1,-1),(-1,1),(-1,-1)\},$$

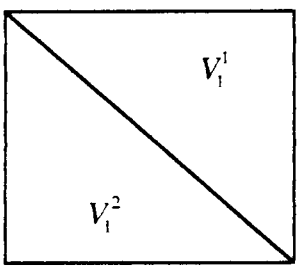
$$\overline{V_1^1} = Co\{(-1,-1),(1,-1),(1,1)\},$$

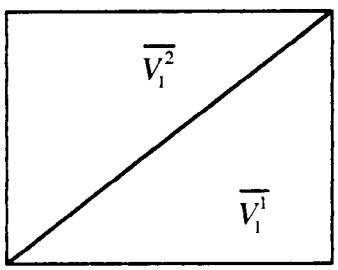
$$\overline{V_1^2} = Co\{(-1,-1),(-1,1),(1,1)\},$$

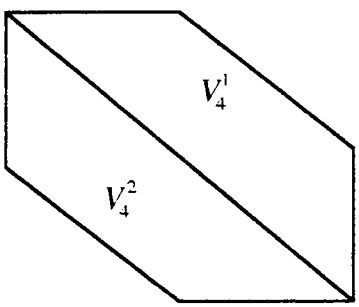
$$V_4^1 = Co\{(-1,1),\left(\frac{1}{2},1\right),\left(1,\frac{1}{2}\right),(1,-1)\},$$

and

$$V_4^2 = Co\{(-1,1),\left(-1,-\frac{1}{2}\right),\left(-\frac{1}{2},-1\right),(1,-1)\}.$$







Applying Proposition 5.1 to V_1^1 and V_1^2 we find that $\eta_1 = \eta_2 = \frac{1}{2}$, $\mu_1 = \frac{2\sqrt{2}}{3\pi}(1,1)$ and $\mu_2 = \frac{-2\sqrt{2}}{3\pi}(1,1)$. So $z_1 = \frac{7\sqrt{2}}{3\pi}, z_2 = \frac{\sqrt{2}}{3\pi}$ and hence $\bar{g} \le z_1 + z_2 = \frac{8\sqrt{2}}{3\pi} \approx 1.200$. Using $\overline{V_1^1}$ and $\overline{V_1^2}$ we find that $\bar{g} \le 4 - \frac{8\sqrt{2}}{3\pi} \approx 2.800$. For V_4^1 and V_4^2 we determine

$$\bar{g} \le z_1 - z_2 = \frac{6\sqrt{2}}{3\pi} \approx 0.9003.$$

The restriction that our random variable have compact range is not a very realistic

assumption. We shall now briefly discuss a generalization which extends to random variables of unbounded range. The basic idea is that we will not allow our convex function g to 'grow' too rapidly while restricting the range of our random variable to lie in a pointed cone.

We still require that U be a convex range but possibly unbounded and that $\int_U gdF < \infty$. Using properties of improper integrals we know that for every $\varepsilon > 0$ there is a compact and convex $\hat{U} \subset U$ such that $\left|\int_{U \setminus \hat{U}} gdF\right| < \varepsilon, \int_{U \setminus \hat{U}} dF < \varepsilon$. One can then apply the previous results on $\hat{U}$ instead of U. One then obtains a lower bound of

$$\bar{g} \geq (1-\varepsilon)E[g \mid \hat{U}] + \int_{U \setminus \hat{U}} dF \cdot E[g \mid U \setminus \hat{U}] \geq (1-\varepsilon)E[g \mid \hat{U}] - \varepsilon$$

with an upper bound of

$$\bar{g} = E_U[g] = \int_{\hat{U}} dF \cdot E[g \mid \hat{U}] + \int_{U \setminus \hat{U}} dF \cdot E[g \mid U \setminus \hat{U}] \leq E[g \mid \hat{U}] + \varepsilon.$$

Unfortunately, being able to determine the 'best' $\hat{U}$ given a particular ε may be a difficult task. Gassman and Ziemba were able to obtain results under the restriction that (1) U be contained in (the translation of) a pointed cone in $\Re^n$, where a pointed cone C is a set C such that if x and -x are elements of C then x=0, and (2) g satisfy the linear growth condition: $\lim_{t \to \infty} g(x+td)/\|x+td\| \leq \delta < \infty$ for every x in U and for all directions d of U, where a direction d of U, $d \in \Delta$, provided x+td is in U for every x in U, $t \geq 0$ and $d \neq 0$. Assuming that the growth of g is 'well-behaved' in the appropriate regions the following result holds true.

Theorem 5.2. Assume there exists a polyhedron V with extreme points $\{v_1, \ldots, v_k\}$ and extreme directions $\{d_1, \ldots, d_l\}$ with $U \subset V$. Provided g can be extended to a convex function on V and the previously stated assumptions (1) and (2) hold, then

$$\bar{g} \leq \max\left\{\sum_{i=1}^{k} \lambda_i g(v_i) + \sum_{j=1}^{l} \delta\mu_j \|d_j\| \mid \lambda_i \geq 0, \mu_j \geq 0, \sum \lambda_i = 1, \sum_i (\lambda_i v_i + \mu_j d_j) = \mu\right\}.$$

Proof: If we can show that for all $(x,y)\in H$ the following holds

$$y \le \max\left\{\sum_i \lambda_i g(v_i) + \sum_j \delta\mu_j \|d_j\| \mid \lambda_i \ge 0, \mu_j \ge 0, \sum \lambda_i = 1, \sum_i \lambda_i v_i + \sum_j \mu_j d_j = x\right\}$$

then the result will be proved since $(\mu,\bar{g}) \in H$. Without loss of generality let the norm of $d_j = 1$ for j=1,....,l. Now for $(x,y)\in H$ there exist $\gamma_i \ge 0, u_i \in U, i = 1,...,I$ such that

$$\sum \gamma_i = 1, \sum \gamma_i u_i = x, \sum \gamma_i g(u_i) = y.$$

Now for each i we have $\alpha_{ij} \ge 0, \beta_{ij} \ge 0$ such that

$$\sum_j^k \alpha_{ij} = 1 \mu_i = \sum_{j=1}^k \alpha_{ij} v_i + \sum_{j=1}^l \beta_{ij} d_j$$

so that

$$x = \sum_i \sum_j \gamma_i \alpha_{ij} v_j + \sum_i \sum_j \gamma_i \beta_{ij} d_j.$$

Noting that d is a direction of U for i=1,....,I, we can write the linear combination for y as a function of the above expression for x. Making use of the fact that g is convex and the triangle inequality we have

$$y = \sum \gamma_i g(u_i) = \sum_i \gamma_i g\left(\sum_j \alpha_{ij} v_j + \sum_j \beta_{ij} d_j\right)$$

$$\le \sum_i \left\{\gamma_i g\left(\sum_j \alpha_{ij} v_j\right) + \delta \left\|\sum_j \beta_{ij} d_j\right\|\right\} \le \sum_i \sum_j \gamma_i \alpha_{ij} g(v_j) + \sum_i \sum_j \gamma_i \beta_{ij} d_j.$$

Then set

$$\hat{\lambda}_j = \sum_i \gamma_i \alpha_{ij}, j = 1,...,k, \hat{\mu}_j = \sum_i \gamma_i \beta_{ij}, j = 1,...,l,$$

and observe that

$$\hat{\lambda}_j \ge 0, \sum_j \hat{\lambda}_j = 1, \hat{\mu}_j \ge 0, \sum_{j=1}^k \hat{\lambda}_j v_j + \sum_{j=1}^l \hat{\mu}_j d_j = x.$$

Now for λ_j, μ_j satisfying the above expression we have in particular

$$y \le \max \sum_{j=1}^{k} \lambda_j g(v_j) + \sum_{j=1}^{l} \mu_j \delta. \ \blacksquare$$

A few comments are in order.

1. If d and -d are both extreme directions for V then the linear program as stated in Theorem 5.2 is unbounded. This situation was excluded by requiring that U lie in a pointed cone.
2. For a bounded V, the set of extreme directions is empty, i.e. l=0, so that Theorem 5.2 reduces to Theorem 5.1.
3. For one dimension, our set V is either $[a,\infty),(-\infty,b]$. So that for d=1, Gassman and Ziemba (1986) were able to explicitly derive the upper bound as $g(a)+\delta(\mu - a)$.

Since it may be difficult to check condition (2) for every possible x and d, the above inequality in Theorem 5.2 may be replaced by equality should (1a) V be contained in the translate of an orthant in $\Re^n$, (2a) V satisfies the growth condition

$$\max_{i=1,\ldots,k} \lim_{t\to\infty} \frac{g(v_i + td_j)}{\|v_i + td_j\|} = \delta_j < \infty, j = 1,\ldots,l.$$

One can also obtain bounds via Theorem 5.2 on subsets of V although computationally this is more difficult to implement. The following example is from Gassman and Ziemba (1986).

Example 5.4. Let $f(x,y) = \begin{cases} e^{-x} \cdot 1/2\sqrt{x} & \text{if } |y| \le \sqrt{x}, x > 0 \\ 0 & \text{ow} \end{cases}$

with $g(x,y) = \frac{x}{2} + \frac{1}{x+2} + \sqrt{x} - \sqrt{2x - y^2}$ for $|y| \le \sqrt{x}$.

Here U is the area inside the parabola $y^2 = x$ and g is extended to a convex function on the right half-plane by linear continuation, i.e. setting

$$g(x,y) = \frac{x}{2} + \frac{1}{x+2} - \sqrt{x} + |y| \qquad |y| > \sqrt{x}.$$

The expected value of (x,y) is (1,0) and the lower bound for g was found to be $\bar{g} \geq g(1,0) = \frac{11}{6} - \sqrt{2} \approx 0.4191$. The exact value for g was computed to be $\bar{g} \approx 0.6062$.

Computing the upper bounds via Theorem 5.2 and on the polyhedra presented earlier are summarized in the following table.

Polyhedron	Extreme Points	Extreme rays	δ_1	δ_2	Upper bound on E[g]
V_1	$\left(0,\frac{1}{4}\right),\left(0,-\frac{1}{4}\right)$	$y = \pm\left(x + \frac{1}{4}\right)$	$\frac{3}{4}\sqrt{2}$	$\frac{3}{4}\sqrt{2}$	2.25
V_2	(0,1),(0,-1)	$y = \pm\left(\frac{1}{4}x + 1\right)$	$3/\sqrt{17}$	$3/\sqrt{17}$	2.25
V_3	$\left(0,\frac{1}{2}\right),\left(0,-\frac{1}{2}\right)$	$y = \pm\left(\frac{1}{2}x + \frac{1}{2}\right)$	$2/\sqrt{5}$	$2/\sqrt{5}$	2.00

Much sharper results can be obtained by partitioning the above subsets. Since the user would need to specify the extreme points, the extreme directions, the desired measure, and the 'growth' bounds an example is not presented here. This is obviously not an automative task.

In conclusion, the results listed in the previous four chapters involves some severe constraints and/or operational limitations. With regards to the GMP, as presented in chapter II, choosing the appropriate probability measure can be a formidable task. Birge and Wets (1987) show that if h is convex, the g_j are linear or piecewise affine, and the random variable have convex range, then the probability measure that solves the moment problem is defined over the set of extreme points and extreme rays of the range of the compact variable. Should the number of extreme points and extreme rays be finite then this reduces to an LP problem. Dupacova (1966) obtained improvements over the Madansky bound by incorporating 2nd moment information as well. Perhaps most importantly, requiring that the random variable have compact range excludes members of the exponential family of distributions. Recent research has addressed this limitation, especially for the normal distribution, but is still computationally difficult to implement. Also, linearizing the convex

function g and partitioning the sample space in an appropriate manner loses both information and allows for wide ranging results. Birge and Wets (1987) allow for an unbounded random variable range provided that these can be enclosed in the union of compact convex sets and recession cones. Dula (1986,1992) was able to produce bounds for the expectation of sublinear (positively homogeneous of degree one) functions via an inner linearization scheme by Birge and Wets (1989). He is thus able to compute tight bounds over the entire space $\Re^n$ without the independence assumption as well. As usual, the area is teeming with ideas and a need for improvement.

BIBLIOGRAPHY

BIBLIOGRAPHY

1. Ben-Tal, A. and Hochman, E. (1972). "More bounds on the expectation of a convex function of a random variable," *Journal of Applied Probability* 9, 803-812.

2. Birge, J. and Wets, R.J.-B. (1987). "Computing bounds for stochastic programming problems by means of a generalized moment problem," *Mathematics of Operations Research* 12, 149-162.

3. Birge, J. and Wets, R.J.-B. (1989). "Sublinear upper bounds for stochastic programs with recourse," *Mathematical Programming* 43, 131-149.

4. Dresher, M. (1953). "Moment spaces and inequalities," *Duke Math. Jour.* 20, 261-271.

5. Dresher, M., Karlin, S. and Shapley, L. (1950). "Polynomial games," *Ann. Math. Studies* 20, 161-180.

6. Dula, J. (1986). "Bounds on the expectation of convex functions." Ph.D. Dissertation, The University of Michigan (Ann Arbor, MI).

7. Dula, J. (1992). "An upper bound on the expectation of simplicial functions of multivariate random variables," *Mathematical Programming* 55, 69-80.

8. Dupacova, J. (as Zackova) (1966). "On minimax solutions of stochastic linear programming problems," *Casopis Pro Pestovani Matematiky* 91, 423-430.

9. Edmundson, H. (1956). "Bounds on the expectation of a convex function of a random variable," *The RAND Corporation Paper* 982 (November 26, 1956).

10. Gassman, H. and Ziemba, W. (1986). "A tight upper bound for the expectation of a convex function of a multivariate random variable," *Mathematical Programming Study* 27, 39-53.

11. Heyman, D. and Sobel, M. (1982). Stochastic Models in Operations Research Vol.1. McGraw-Hill, New York.

12. Jensen, J. (1906). "Sur les fonctions convexes et les inegalites entre les valuers moyennes," *Acta Mathematica 30*, 175-193.

13. Kingman, J. (1963). "On inequalities of the tchebychev type," *Proc. Cambridge Phil. Soc.* 59, 135-46.

14. Madansky, A. (1959). "Bounds on the expectation of a convex function of a multivariate random variable," *Annals of Mathematical Statistics* 30, 743-746.

15. Mangasarian, O. (1964). "Nonlinear programming problems with stochastic objective functions," *Management Science* 12, 353-359.

16. Prekopa, A. and Wets, R.J.-W. (1986). "Stochastic Programming 84, Foreward," *Mathematical Programming Study* 27.

17. Rudin, W. (1987). Real and Complex Analysis 3rd edition. McGraw-Hill, New York.

18. Sposito, V. (1975). Linear and Nonlinear Programming. Iowa State University Press, Ames.

VITA

The author received a Bachelor of Science degree in Mathematics from North Carolina State University in 1993. That same year he enrolled at the University of Tennessee to begin study towards a Master of Science degree in Mathematics. In 1995 he started to study for a Master of Science degree in Statistics at the University of Tennessee.