

Higher-Order Upper Bounds on the Expectation of a Convex Function

Steftcho P. Dokov and David P. Morton
Graduate Program in Operations Research and Industrial Engineering
The University of Texas at Austin
Austin, Texas 78712

January 2002

Abstract We develop a decreasing sequence of upper bounds on the expectation of a convex function. The n -th term in the sequence uses moments and cross-moments of up to degree n from the underlying random vector. Our work has application to a class of two-stage stochastic programs with recourse. The objective function of such a model can defy computation when: (i) the underlying distribution is assumed to be known only through a limited number of moments or (ii) the function is computationally intractable, even though the distribution is known. A tractable approximating model arises by replacing the objective function by one of our bounding elements. We justify this approach by showing that as n grows, solutions of the order- n approximation solve the true stochastic program.

1 Introduction

In this paper we derive a new class of upper bounds on the expected value of a convex function of a random vector. Computing one of our bounds requires knowledge of the moments and cross-moments of the underlying random vector up to degree n . The bounds can be applied to approximately solve an important class of two-stage stochastic programs. The need for such approximations arises for two reasons. First, the “true” distribution of the stochastic program’s random parameters may be unknown, but we may be willing to specify, or estimate, some of the distribution’s moments. Second, the distribution may be (assumed) known, but because the stochastic program’s objective function is a difficult multivariate expectation, we cannot compute it exactly.

The development of first-order upper bounds on the expectation of a convex function starts with Edmundson [19] in the univariate case. Madansky [31, 32] and Frauendorfer [20] generalize this bound to the multivariate setting in the respective cases when the components of the random vector are independent and dependent. Both [20] and [31, 32] assume the support of the underlying distribution is contained in a multidimensional rectangle. These are called “first-order” bounds because they only use the mean, and in the dependent case

degree-one cross-moments, of the underlying random vector. Generalizations of the so-called Edmundson-Madansky (*EM*) bound can be made to non-rectangular sets including simplices and general polyhedral domains (see, e.g., [13, 21, 23]).

For first-order bounds, the advantage of using simplicial sets (relative to rectangular sets) is that the required computational effort drops from exponential to linear in the dimension of the random vector. So, we first develop our higher-order bounds on multidimensional rectangles but then do so on simplices, too. We note that when the support of the random vector is contained in a general polyhedron, a linear program must be solved in order to compute the bound. In contrast, when using simplicial or rectangular sets the bound is available in analytical form.

Dupačová [12] initiated the investigation of bounds derived when the underlying distribution is known only in limited manner (see also [13, 14]). In stochastic programming, these bounds may be viewed as arising through “games against nature” in which we first select a decision vector and then nature chooses the worst (or, more generally, an extreme) distribution from a prespecified class, e.g., distributions with given first-order information.

When first-order bounds are too weak and the distribution is known, the bounds can be sequentially improved by applying them in a conditional fashion on an iteratively refined partition of the random vector’s support [4, 6, 21, 22, 25]. An alternative (or complement) is to tighten the bounds by using higher-order moment information. Upper bounds using second-order information are developed in [3, 10, 11, 12, 27]. Bounds based on convexity can be extended to handle convex-concave saddle functions. For first-order bounds of this type see [17, 18] and [21]. Edirisinghe [15] develops second-order bounds for expectations of such saddle functions and employs these bounds in a sequential refinement scheme in [16].

Typically, bounds can be categorized as being based on distributional approximations or functional approximations. Sometimes, both views are applicable. Distributional approximations replace the original distribution with another (usually discrete) distribution that eases computing expectations. Functional approximations replace the original function with a simpler (e.g., separable) function to facilitate computation of the bound. For bounds in the literature motivated by functional approximations see, e.g., [5, 6, 7, 33, 34, 36].

The bounds we develop may be viewed from both the distributional- and functional-approximation perspectives. Specifically, one can interpret our bounds as discrete distributional approximations derived using convexity. The weights in the resulting convex combination are probability weights because they are nonnegative and sum to one. On the other hand, once we derive the bounding expressions, we recognize that they are actually the well-known Bernstein polynomials from approximation theory [2]. Hence, our bounds can be seen as expectations of these polynomial approximating functions.

The focus in this paper is on upper bounds, but for completeness we note that there has been much less work on finding lower bounds on the expectation of a convex function. Apparently, convexity more easily allows a variety of majorizing schemes that yield upper bounds rather than minorizing schemes that can be used to generate lower bounds. The classic inequality of Jensen [26] provides a first-order lower bound on the expectation of a convex function and a second-order lower bound is developed in [15]. In the case of hyper-rectangular support and independent components, the bound of [15] is tightened in [9].

We call the higher-order upper bounds we derive Edmundson-Madansky-type upper bounds because in the univariate case our first-order bound is simply the *EM* bound. In the multivariate case, our first-order bound on a hyper-rectangular domain is the upper bound of [20]. Our first-order bound on a simplicial domain is that of [13] and [23], when their domain sets are restricted to be simplices.

This paper is organized as follows. In Section 2 we develop a class of higher-order upper bounds on the expectation of a convex function in the univariate case. Section 3 interprets the univariate bounds as both distributional- and functional-approximations, and examines the geometry of the second-order bound. Section 4 generalizes the bounds to the multivariate case for hyper-rectangular and simplicial support of the underlying distribution. Section 5 provides a convergence result for the bounds when applied to stochastic programming problems and the paper is summarized in Section 6.

2 The Univariate Case

The Edmundson-Madansky inequality [19] for a convex function $f : [a, b] \rightarrow \mathbf{R}$ of a univariate random variable ξ with support contained in a bounded interval $[a, b]$, $a < b$, can be derived by expressing ξ as the following convex combination

$$\xi = \frac{b - \xi}{b - a}a + \frac{\xi - a}{b - a}b.$$

Hence, by convexity of f we have

$$f(\xi) \leq \frac{b - \xi}{b - a}f(a) + \frac{\xi - a}{b - a}f(b).$$

Taking the expectation of both sides and letting $m_1 = \mathbf{E}\xi$ yields

$$\mathbf{E}f(\xi) \leq \frac{b - m_1}{b - a}f(a) + \frac{m_1 - a}{b - a}f(b) \equiv EM,$$

where we refer to *EM* as the Edmundson-Madansky bound. This three-step procedure can be generalized to produce a tighter upper bound that uses higher-order moments for the random

variable ξ . The binomial identity

$$\sum_{i=0}^n \binom{n}{i} \frac{(\xi - a)^i (b - \xi)^{n-i}}{(b - a)^n} = 1$$

can be rewritten

$$\xi = \sum_{i=0}^n \binom{n}{i} \frac{(\xi - a)^i (b - \xi)^{n-i}}{(b - a)^n} \left(a + \frac{i}{n}(b - a) \right). \quad (1)$$

This representation of ξ is a convex combination on the uniform grid of points $a + (i/n)(b - a)$, $i = 0, 1, \dots, n$, with corresponding weights $\binom{n}{i}(\xi - a)^i(b - \xi)^{n-i}/(b - a)^n$, $i = 0, 1, \dots, n$. Convexity of f implies

$$f(\xi) \leq \sum_{i=0}^n \binom{n}{i} \frac{(\xi - a)^i (b - \xi)^{n-i}}{(b - a)^n} f \left(a + \frac{i}{n}(b - a) \right) \equiv B_n(f; \xi), \quad (2)$$

where $B_n(f; \xi)$ is the n -th order Bernstein [2] polynomial. Taking the expectation of both sides of (2) yields

$$\mathbf{E}f(\xi) \leq \sum_{i=0}^n \binom{n}{i} \frac{\mathbf{E}[(\xi - a)^i (b - \xi)^{n-i}]}{(b - a)^n} f \left(a + \frac{i}{n}(b - a) \right) \equiv EM_n, \quad (3)$$

where computing $\mathbf{E}[(\xi - a)^i (b - \xi)^{n-i}]$ requires moments $\mathbf{E}\xi^i$, $i = 1, \dots, n$. We refer to EM_n defined in (3) as the n -th order Edmundson-Madansky bound because it involves the first n moments of the random variable ξ and $EM_1 = EM$. In order to show that $\{EM_n\}_{n=1}^\infty$ is a decreasing sequence it clearly suffices to prove $B_n(f; \xi) \leq B_{n-1}(f; \xi)$, $n = 2, 3, \dots$, and the following lemma establishes this property provided f is convex.

Lemma 1 *For a convex function $f : [a, b] \rightarrow \mathbf{R}$ the Bernstein polynomials defined in (2) form a decreasing sequence bounded below by $f(\xi)$, i.e.,*

$$f(\xi) \leq B_n(f; \xi) \leq B_{n-1}(f; \xi). \quad (4)$$

Proof. We established $f(\xi) \leq B_n(f; \xi)$, in (2). To verify the other inequality in (4) we express $a + i(b - a)/n$ using the following convex combination

$$\frac{(n - i)a + ib}{n} = \frac{i}{n} \left(\frac{(n - i)a + (i - 1)b}{n - 1} \right) + \frac{n - i}{n} \left(\frac{(n - 1 - i)a + ib}{n - 1} \right).$$

Hence, convexity of f implies

$$f \left(a + \frac{i}{n}(b - a) \right) \leq \frac{i}{n} f \left(\frac{(n - i)a + (i - 1)b}{n - 1} \right) + \frac{n - i}{n} f \left(\frac{(n - 1 - i)a + ib}{n - 1} \right). \quad (5)$$

Applying inequality (5) to the defining expression (2) of $B_n(f; \xi)$ we obtain

$$\begin{aligned} B_n(f; \xi) &\leq \sum_{i=1}^n \binom{n}{i} \frac{(\xi - a)^i (b - \xi)^{n-i}}{(b - a)^n} \binom{i}{n} f\left(a + \frac{(i-1)(b-a)}{n-1}\right) \\ &\quad + \sum_{i=0}^{n-1} \binom{n}{i} \frac{(\xi - a)^i (b - \xi)^{n-i}}{(b - a)^n} \binom{n-i}{n} f\left(a + \frac{i(b-a)}{n-1}\right). \end{aligned}$$

Changing the range of the first summation to $i = 0, \dots, n-1$ and simplifying yields

$$\begin{aligned} B_n(f; \xi) &\leq \sum_{i=0}^{n-1} \binom{n-1}{i} \frac{(\xi - a)^{i+1} (b - \xi)^{n-i-1}}{(b - a)^n} f\left(a + \frac{i(b-a)}{n-1}\right) \\ &\quad + \sum_{i=0}^{n-1} \binom{n-1}{i} \frac{(\xi - a)^i (b - \xi)^{n-i}}{(b - a)^n} f\left(a + \frac{i(b-a)}{n-1}\right) \\ &= \left(\frac{\xi - a}{b - a} + \frac{b - \xi}{b - a}\right) \sum_{i=0}^{n-1} \binom{n-1}{i} \frac{(\xi - a)^i (b - \xi)^{n-1-i}}{(b - a)^{n-1}} f\left(a + \frac{i}{n-1}(b-a)\right) \\ &= B_{n-1}(f; \xi), \end{aligned}$$

which completes the proof. ■

Taking the expectation in (4) yields monotonicity of the upper bounds EM_n , and we summarize this in the following theorem.

Theorem 2 *Let $f : [a, b] \rightarrow \mathbf{R}$ be a convex function, and let ξ be a random variable with moments $\mathbf{E}\xi^n$, $n = 1, 2, \dots$, and support contained in $[a, b]$. Then*

$$\mathbf{E}f(\xi) \leq EM_n \leq EM_{n-1},$$

where EM_n is defined in (3).

From Theorem 2 we can bound the expectation, $\mathbf{E}f(\xi)$, of a convex function of a univariate random variable by a decreasing sequence of upper bounds $\{EM_n\}_{n=1}^\infty$. The bound EM_n requires knowing the moments $\mathbf{E}\xi^i$, $i = 1, 2, \dots, n$, and requires $n + 1$ evaluations of f on a uniform grid of points over $[a, b]$.

3 Geometry and Interpretation of the Bounds

In this section we turn to a geometric view of EM_n , interpret it as a functional- and a distributional-approximation, and finally consider the sensitivity of EM_2 with respect to its defining parameters. We start with the well-known interpretation of the classical Edmundson-Madansky bound EM_1 (see, e.g., [28, Section 3.4.2]). In order to approximate $\mathbf{E}f(\xi)$ we

replace $f(\xi)$ with $B_1(f; \xi)$, the affine function defined by the pair of points $(a, f(a))$ and $(b, f(b))$ as illustrated in Figure 1. The expectation of this approximating function is simply $EM_1 = \mathbf{E}B_1(f; \xi)$ and is given by the point $(m_1, B_1(f; m_1))$ in Figure 1. On the other hand, from the definition of $EM_1 = \frac{b-m_1}{b-a}f(a) + \frac{m_1-a}{b-a}f(b)$ we can view the original, and possibly continuous, distribution of ξ as being replaced with an approximating distribution that takes values a and b with respective probabilities $\frac{b-m_1}{b-a}$ and $\frac{m_1-a}{b-a}$.

More generally, that EM_n is a distributional approximation is clear from (3). The discrete approximating distribution takes values on the uniform grid of points $a + (i/n)(b-a)$, $i = 0, \dots, n$, with respective probabilities $\binom{n}{i} \mathbf{E}[(\xi-a)^i(b-\xi)^{n-i}]/(b-a)^n$, $i = 0, \dots, n$. The functional approximation view of EM_n follows from the fact that $EM_n = \mathbf{E}B_n(f; \xi)$ and $B_n(f; \xi)$ is a polynomial approximation of degree n majorizing $f(\xi)$. In the remainder of this section we focus on the geometric interpretation of EM_2 . We have the following two expressions for the second-order upper bound EM_2 and the second-order Bernstein polynomial $B_2(f; \xi)$.

$$B_2(f; \xi) = \frac{(\xi - \frac{a+b}{2})(\xi - b)}{(a - \frac{a+b}{2})(a - b)}f(a) + \frac{(\xi - a)(\xi - \frac{a+b}{2})}{(b - a)(b - \frac{a+b}{2})}f(b) \\ + \frac{(\xi - a)(\xi - b)}{(\frac{a+b}{2} - a)(\frac{a+b}{2} - b)} \frac{1}{2} \left(\frac{f(a) + f(b)}{2} + f\left(\frac{a+b}{2}\right) \right), \quad (6)$$

and

$$EM_2 = \frac{B - m_1}{B - A} \left[\frac{\frac{a+b}{2} - A}{\frac{a+b}{2} - a} f(a) + \frac{A - a}{\frac{a+b}{2} - a} \left(\frac{f(\frac{a+b}{2}) + \frac{f(a)+f(b)}{2}}{2} \right) \right] \\ + \frac{m_1 - A}{B - A} \left[\frac{b - B}{b - \frac{a+b}{2}} \left(\frac{f(\frac{a+b}{2}) + \frac{f(a)+f(b)}{2}}{2} \right) + \frac{B - \frac{a+b}{2}}{b - \frac{a+b}{2}} f(b) \right], \quad (7)$$

where $A = m_1 - \sigma^2/(b - m_1)$, $B = m_1 + \sigma^2/(m_1 - a)$, and $\sigma^2 = m_2 - m_1^2$. In (6) we express $B_2(f; \xi)$ in the form of Lagrange's interpolation formula (see, e.g., [8]). This shows $B_2(f; \xi)$ is a quadratic that passes through the points:

$$\begin{array}{c|c|c|c} \xi : & a & \frac{a+b}{2} & b \\ \hline f(\xi) : & f(a) & \frac{1}{2} \left(\frac{f(a)+f(b)}{2} + f\left(\frac{a+b}{2}\right) \right) & f(b) \end{array},$$

and, of course, from (2) we know $B_2(f; \xi)$ majorizes f . From (7) we see that EM_2 is a convex combination with weights $\frac{B-m_1}{B-A} = \frac{b-m_1}{b-a}$ and $\frac{m_1-A}{B-A} = \frac{m_1-a}{b-a}$ of two expressions which are themselves convex combinations of the points involved in the construction of $B_2(f; \xi)$, provided $A \leq \frac{a+b}{2} \leq B$.

Figure 1 illustrates $f(\xi)$, $B_1(f; \xi)$, $B_2(f; \xi)$, the first- and second-order upper bounds EM_1 , and EM_2 and finally, for reference, the first- and second-order lower bounds of Jensen [26] and

Edirisinghe [15]. Figure 1 is valid for EM_2 provided $A \leq \frac{a+b}{2} \leq B$. Figures 2 and 3 illustrate EM_2 in the other two cases: $(a+b)/2 \leq A$ and $B \leq (a+b)/2$, respectively.

As we discussed in the introduction, one of the motivations for studying bounds on $\mathbf{E}f(\xi)$ using only limited moment information on ξ is that the “true” distribution may be unknown. Indeed, we may not even know these moments with certainty and so we close this section by briefly discussing the sensitivity of EM_2 with respect to m_1 and σ^2 . To do so, it is convenient to re-express EM_2 as

$$EM_2 = \frac{1}{(b-a)^2} \left[f(a) + f(b) - 2f\left(\frac{a+b}{2}\right) \right] \sigma^2 + \left[\left(\frac{b-m_1}{b-a}\right)^2 f(a) + \frac{2(m_1-a)(b-m_1)}{(b-a)^2} f\left(\frac{a+b}{2}\right) + \left(\frac{m_1-a}{b-a}\right)^2 f(b) \right]. \quad (8)$$

The coefficient of σ^2 is nonnegative because f is convex and hence EM_2 is increasing in σ^2 assuming m_1 is held constant. Holding σ^2 constant and computing the second derivative of EM_2 in (8) with respect to m_1 shows that EM_2 is convex in m_1 . So, if m_1 is fixed but σ^2 is only known to be in the interval $[\underline{\sigma}^2, \bar{\sigma}^2]$ then we can conclude $EM_2(\bar{\sigma}^2)$ is an upper bound on $\mathbf{E}f(\xi)$. Similarly, if σ^2 is fixed but $m_1 \in [\underline{m}_1, \bar{m}_1]$ then $\max\{EM_2(\underline{m}_1), EM_2(\bar{m}_1)\}$ is an upper bound on $\mathbf{E}f(\xi)$.

4 The Multivariate Case

In this section we extend EM_n to handle multivariate distributions with support contained either in a hyper-rectangle or in a simplex. In both cases the components of $\xi = (\xi_1, \xi_2, \dots, \xi_d)^\top$ can have a general form of dependency.

4.1 Hyper-rectangular domain

We begin by assuming the support of ξ is contained in the hyper-rectangle $\times_{j=1}^d [a_j, b_j]$, $a_j < b_j$, $j = 1, \dots, d$, and that $f : \times_{j=1}^d [a_j, b_j] \rightarrow \mathbf{R}$ is convex. Now we generalize the three-step procedure ((1) \rightarrow (2) \rightarrow (3)) to this multivariate case. Each point ξ in the domain, $\times_{j=1}^d [a_j, b_j]$, can be represented as a convex combination formed from the uniform grid of $(n+1)^d$ points given by $\times_{j=1}^d \{a_j + i_j(b_j - a_j)/n : i_j = 0, \dots, n\}$ where n is a prespecified integer, and so

$$\xi = \sum_{i_1=0}^n \cdots \sum_{i_d=0}^n \prod_{j=1}^d \binom{n}{i_j} \left(\frac{\xi_j - a_j}{b_j - a_j} \right)^{i_j} \left(\frac{b_j - \xi_j}{b_j - a_j} \right)^{n-i_j} \begin{pmatrix} a_1 + \frac{i_1(b_1 - a_1)}{n} \\ \vdots \\ a_d + \frac{i_d(b_d - a_d)}{n} \end{pmatrix}.$$

Because this representation of ξ is a convex combination we have

$$\begin{aligned} f(\xi) &\leq \sum_{i_1=0}^n \cdots \sum_{i_d=0}^n \prod_{j=1}^d \binom{n}{i_j} \left(\frac{\xi_j - a_j}{b_j - a_j} \right)^{i_j} \left(\frac{b_j - \xi_j}{b_j - a_j} \right)^{n-i_j} f \left(\begin{matrix} a_1 + \frac{i_1(b_1-a_1)}{n} \\ \vdots \\ a_d + \frac{i_d(b_d-a_d)}{n} \end{matrix} \right) \\ &\equiv B_{n,d}^{rec}(f; \xi). \end{aligned} \quad (9)$$

Here, $B_{n,d}^{rec}(f; \xi)$ is the n -th order Bernstein polynomial for a function f defined on a hyper-rectangle $\times_{j=1}^d [a_j, b_j]$ (see, e.g., [30]). By taking expectations in (9) we obtain

$$\begin{aligned} \mathbf{E}f(\xi) &\leq \sum_{i_1=0}^n \cdots \sum_{i_d=0}^n \mathbf{E} \left[\prod_{j=1}^d \binom{n}{i_j} \frac{(\xi_j - a_j)^{i_j} (b_j - \xi_j)^{n-i_j}}{(b_j - a_j)^n} \right] f \left(\begin{matrix} a_1 + \frac{i_1(b_1-a_1)}{n} \\ \vdots \\ a_d + \frac{i_d(b_d-a_d)}{n} \end{matrix} \right) \\ &\equiv EM_{n,d}^{rec}, \end{aligned} \quad (10)$$

where $EM_{n,d}^{rec}$ denotes the n -th order Edmundson-Madansky upper bound when the random vector ξ has support contained in a d -dimensional rectangle. If the components of the random vector ξ are dependent then $EM_{n,d}^{rec}$ can be computed provided we know all cross-moments $\mathbf{E}[\xi_1^{i_1} \xi_2^{i_2} \cdots \xi_d^{i_d}]$, $i_j = 0, \dots, n$, $j = 1, \dots, d$. When ξ has independent components $EM_{n,d}^{rec}$ simplifies to

$$EM_{n,d}^{rec} = \sum_{i_1=0}^n \cdots \sum_{i_d=0}^n \prod_{j=1}^d \binom{n}{i_j} \mathbf{E} \left[\frac{(\xi_j - a_j)^{i_j} (b_j - \xi_j)^{n-i_j}}{(b_j - a_j)^n} \right] f \left(\begin{matrix} a_1 + \frac{i_1(b_1-a_1)}{n} \\ \vdots \\ a_d + \frac{i_d(b_d-a_d)}{n} \end{matrix} \right).$$

To compute the bound in this case, we need to know the moments of each component, i.e., $\mathbf{E}\xi_j^1, \dots, \mathbf{E}\xi_j^n$, $j = 1, \dots, d$. Monotonicity of $EM_{n,d}^{rec}$ in n is achieved in a similar way to the univariate case treated in Lemma 1. In particular we apply (5) componentwise to get $B_{n,d}^{rec}(f; \xi) \leq B_{n-1,d}^{rec}(f; \xi)$. Then, by taking the expectation, we conclude that $EM_{n,d}^{rec} = \mathbf{E}B_{n,d}^{rec}(f; \xi) \leq \mathbf{E}B_{n-1,d}^{rec}(f; \xi) = EM_{n-1,d}^{rec}$. We summarize this discussion in the following theorem.

Theorem 3 *Let $f : \times_{j=1}^d [a_j, b_j] \rightarrow \mathbf{R}$ be a convex function, and let ξ be a random vector with support contained in $\times_{j=1}^d [a_j, b_j]$ and with cross-moments $\mathbf{E}[\xi_1^{i_1} \xi_2^{i_2} \cdots \xi_d^{i_d}]$, $i_j = 0, \dots, n$, $j = 1, \dots, d$. Then*

$$\mathbf{E}f(\xi) \leq EM_{n,d}^{rec} \leq EM_{n-1,d}^{rec},$$

where $EM_{n,d}^{rec}$ is defined in (10).

From the defining expression of $EM_{n,d}^{rec}$ in (10) we see that computing $EM_{n,d}^{rec}$ requires $(n+1)^d$ function evaluations of f . Frauendorfer [20] generalized the first-order Edmundson-Madansky inequality to the multivariate setting; when $n = 1$ we recover Frauendorfer's bound as $EM_{1,d}^{rec}$. The number of function evaluations required to compute $EM_{n,d}^{rec}$ increases at an exponential rate in the dimension d and at a polynomial rate in the "order" parameter n . Therefore, the upper bounds $EM_{n,d}^{rec}$ can only be applied for modest-sized values of these parameters. As we will see in the next section, if the distribution's support is instead contained in a multidimensional simplex then the number of function evaluations for the corresponding upper bound increases at polynomial rates in both n and d , so that it may be more attractive for practical application.

4.2 Simplicial domain

We now consider the case when the support of the random d -vector ξ is contained in $co\{u^1, u^2, \dots, u^{d+1}\}$, the convex hull of $d+1$ points in general position. Each point in this simplex can be expressed as a convex combination of the extreme points, u^1, \dots, u^{d+1} . The weights of such a convex combination, say p_1, \dots, p_{d+1} , are uniquely specified by the following system of equations

$$\begin{aligned} p_1 u^1 + \dots + p_{d+1} u^{d+1} &= \xi \\ p_1 + p_2 + \dots + p_{d+1} &= 1. \end{aligned} \tag{11}$$

Now we generalize the three-step procedure ((1) \rightarrow (2) \rightarrow (3)) to this simplicial case. Each point ξ in the domain, $co\{u^1, u^2, \dots, u^{d+1}\}$, can be represented as the following convex combination

$$\xi = \sum_{i \in I_n} \binom{n}{i_1, \dots, i_{d+1}} \left(\prod_{j=1}^{d+1} p_j^{i_j} \right) \frac{i_1 u^1 + \dots + i_{d+1} u^{d+1}}{n} \tag{12}$$

where

$$I_n = \{i = (i_1, \dots, i_{d+1}) : i_1 + \dots + i_{d+1} = n, i_1, \dots, i_{d+1} \in \mathbf{Z}_+\},$$

and the points

$$\frac{i_1 u^1 + \dots + i_{d+1} u^{d+1}}{n}, \quad i \in I_n, \tag{13}$$

form a uniform grid over the simplex $co\{u^1, u^2, \dots, u^{d+1}\}$. Here, \mathbf{Z}_+ denotes the set of nonnegative integers. To verify the identity (12) let $\eta = (\eta_1, \dots, \eta_{d+1})$ have a multinomial distribution with parameters n, p_1, \dots, p_{d+1} . Then each η_k has a binomial marginal distribution with parameters n, p_k . Hence, by changing the order of summation, the right-hand side of (12) may

be written

$$\begin{aligned} \frac{1}{n} \sum_{k=1}^{d+1} u^k \left[\sum_{i \in I_n} \binom{n}{i_1, \dots, i_{d+1}} \left(\prod_{j=1}^{d+1} p_j^{i_j} \right) i_k \right] &= \frac{1}{n} \sum_{k=1}^{d+1} u^k \mathbf{E} \eta_k \\ &= \frac{1}{n} \sum_{k=1}^{d+1} u^k n p_k = \sum_{k=1}^{d+1} p_k u^k = \xi, \end{aligned} \quad (14)$$

where the last equality follows from (11). From the convexity of f and (12)

$$\begin{aligned} f(\xi) &\leq \sum_{i \in I_n} \binom{n}{i_1, \dots, i_{d+1}} \left(\prod_{j=1}^{d+1} p_j^{i_j} \right) f \left(\frac{i_1 u^1 + \dots + i_{d+1} u^{d+1}}{n} \right) \\ &\equiv B_{n,d}^{sim}(f; \xi). \end{aligned} \quad (15)$$

Here, $B_{n,d}^{sim}(f; \xi)$ is the n -th order Bernstein polynomial for a function f defined on the simplex $co\{u^1, u^2, \dots, u^{d+1}\}$ (see, e.g., [30]). By taking expectations in (15) we obtain

$$\begin{aligned} \mathbf{E} f(\xi) &\leq \sum_{i \in I_n} \binom{n}{i_1, \dots, i_{d+1}} \mathbf{E} \left[\prod_{j=1}^{d+1} p_j^{i_j} \right] f \left(\frac{i_1 u^1 + \dots + i_{d+1} u^{d+1}}{n} \right) \\ &\equiv EM_{n,d}^{sim}, \end{aligned} \quad (16)$$

where $EM_{n,d}^{sim}$ denotes the n -th order Edmundson-Madansky upper bound when the random vector ξ has support contained in a d -dimensional simplex. If the components of the random vector ξ are dependent then $EM_{n,d}^{sim}$ can be computed provided we know all cross-moments $\mathbf{E}[\xi_1^{i_1} \xi_2^{i_2} \dots \xi_d^{i_d}]$, $i = (i_1, \dots, i_{d+1}) \in I_n$. Note that it is straight-forward to compute $\mathbf{E} \left[\prod_{j=1}^{d+1} p_j^{i_j} \right]$, $i \in I_n$, by first expressing the p_j 's as linear combinations of the ξ_i 's—using the inverse of the matrix on the left-hand side of (11)—and then forming expectations of the required expressions. When ξ has independent components, $EM_{n,d}^{sim}$ simplifies to

$$\mathbf{E} f(\xi) \leq \sum_{i \in I_n} \binom{n}{i_1, \dots, i_{d+1}} \left(\prod_{j=1}^{d+1} \mathbf{E} [p_j^{i_j}] \right) f \left(\frac{i_1 u^1 + \dots + i_{d+1} u^{d+1}}{n} \right).$$

In this case we need to know the marginal moments, $\mathbf{E} \xi_j^{i_j}$, $i_j = 1, \dots, n$, $j = 1, \dots, d+1$. The next theorem states that the upper bounds $EM_{n,d}^{sim}$ defined in (16) decrease monotonically in n .

Theorem 4 *Let $f : co(u^1, u^2, \dots, u^{d+1}) \rightarrow \mathbf{R}$ be a convex function, and let ξ be a random vector with support contained in $co(u^1, u^2, \dots, u^{d+1})$ where the points u^1, u^2, \dots, u^{d+1} are in general position in \mathbf{R}^d . Then*

$$\mathbf{E} f(\xi) \leq EM_{n,d}^{sim} \leq EM_{n-1,d}^{sim},$$

where $EM_{n,d}^{sim}$ is defined in (16).

Proof. First we express every point from the uniform grid, (13), defined by a prespecified integer n , as a convex combination of the points forming the coarser uniform grid defined by the integer $n - 1$, i.e., for $i \in I_n$

$$\frac{i_1 u^1 + \dots + i_{d+1} u^{d+1}}{n} = \sum_{k=1}^{d+1} \frac{i_k}{n} \left(\frac{i_1 u^1 + \dots + (i_k - 1) u^k + \dots + i_{d+1} u^{d+1}}{n - 1} \right). \quad (17)$$

To verify (17), note that the coefficient of u^j on the right-hand side of (17),

$$\frac{i_j}{n} \frac{i_j - 1}{n - 1} + \left(\sum_{k=1, k \neq j}^{d+1} \frac{i_k}{n} \frac{i_j}{n - 1} \right),$$

reduces to i_j/n , the left-hand side coefficient of u^j , by applying $\sum_{k=1, k \neq j}^{d+1} i_k = n - i_j$. Because f is convex, we obtain

$$f \left(\frac{i_1 u^1 + \dots + i_{d+1} u^{d+1}}{n} \right) \leq \sum_{k=1}^{d+1} \frac{i_k}{n} f \left(\frac{i_1 u^1 + \dots + (i_k - 1) u^k + \dots + i_{d+1} u^{d+1}}{n - 1} \right).$$

Combining this inequality with (15) yields

$$B_{n,d}^{sim}(f; \xi) \leq \sum_{i \in I_n} \binom{n}{i_1, \dots, i_{d+1}} \left(\prod_{j=1}^{d+1} p_j^{i_j} \right) \sum_{k=1}^{d+1} \frac{i_k}{n} \cdot f \left(\frac{i_1 u^1 + \dots + (i_k - 1) u^k + \dots + i_{d+1} u^{d+1}}{n - 1} \right).$$

We now change the order of summation to obtain

$$\begin{aligned} B_{n,d}^{sim}(f; \xi) &\leq \sum_{k=1}^{d+1} p_k \sum_{\substack{i_1 + \dots + (i_k - 1) + \dots + i_{d+1} = n - 1 \\ i_j \geq 0 \ \forall j, \ i_k - 1 \geq 0}} \binom{n - 1}{i_1, \dots, i_k - 1, \dots, i_{d+1}} \\ &\quad \cdot p_1^{i_1} \dots p_k^{i_k - 1} \dots p_{d+1}^{i_{d+1}} f \left(\frac{i_1 u^1 + \dots + (i_k - 1) u^k + \dots + i_{d+1} u^{d+1}}{n - 1} \right) \\ &= \sum_{k=1}^{d+1} p_k B_{n-1,d}^{sim}(f; \xi) = B_{n-1,d}^{sim}(f; \xi). \end{aligned}$$

Taking expectations yields the monotonicity result

$$EM_{n,d}^{sim} = \mathbf{E} B_{n,d}^{sim}(f; \xi) \leq \mathbf{E} B_{n-1,d}^{sim}(f; \xi) = EM_{n-1,d}^{sim}. \quad \blacksquare$$

First-order bounds whose computation requires optimization over a polyhedral set containing ξ 's support are developed in [13, 23]. Our first-order bound, $EM_{1,d}^{sim}$, is equivalent to their bounds when their polyhedral set is restricted to be a simplex. The defining expression of

$EM_{n,d}^{sim}$, (16), involves the sum across $i \in I_n$ and function evaluations of f at the uniform simplicial grid of points, (13). The number of evaluations of f required to compute $EM_{n,d}^{sim}$ is therefore $|I_n| = \binom{n+d}{n}$.

In Section 4.1 we developed $EM_{n,d}^{rec}$ which requires $(n+1)^d$ function evaluations over a d -dimensional hyper-rectangle, and in this section we developed $EM_{n,d}^{sim}$ which requires $\binom{n+d}{n} \leq (n+1)^d$ function evaluations over a d -dimensional simplex. The computational effort for both $EM_{n,d}^{rec}$ and $EM_{n,d}^{sim}$ grows polynomially in the order n of the approximation. In contrast the computational effort for $EM_{n,d}^{rec}$ grows exponentially in the dimension d while that for $EM_{n,d}^{sim}$ is polynomial in d . For modest-sized values of n and d the effort required to compute $EM_{n,d}^{sim}$ can be dramatically less than that of $EM_{n,d}^{rec}$. For example, $n = 5$ and $d = 5$ yields $\binom{10}{5} = 252$ versus $6^5 = 7776$ function evaluations.

5 A Convergence Result

In this section we apply the higher-order Edmundson-Madansky upper bounds $EM_{n,d}$, derived in the previous section, to stochastic programming. In particular, we show that in the limit, as the order n grows, optimizing an n -th order approximation yields an optimal solution to the original problem.

We consider a stochastic programming problem formulated as

$$(P) \quad z^* = \min_{x \in X} \mathbf{E}f(x, \xi),$$

where X is a deterministic compact set, $f(x, \cdot)$ is convex $\forall x \in X$, and we assume that (P) has a finite optimal solution. The support of ξ is contained in a hyper-rectangle or simplex which we denote Ξ . An example of (P) is a two-stage stochastic program with recourse in which the recourse function $f(x, \cdot)$ is defined as the optimal value of a linear program given x and $\xi = (T, h)$, i.e.,

$$\begin{aligned} f(x, \xi) = cx + \min_{y \geq 0} \quad & qy \\ \text{s.t.} \quad & Wy = Tx + h. \end{aligned} \tag{18}$$

Note that if the linear program in (18) has a finite optimal solution for all $x \in X$ and $\xi \in \Xi$ then $f(x, \cdot)$ is convex over domain Ξ , $\forall x \in X$. This holds because the randomness, $\xi = (T, h)$, only appears on the right-hand side of the linear program.

Often we cannot solve (P) directly because computing the expectation $\mathbf{E}f(x, \xi)$ is too expensive or impossible. So, instead we will apply the upper bounds from Section 4.1 or 4.2 and solve

$$(P_n) \quad z_n^* = \min_{x \in X} \mathbf{E}B_{n,d}(f; x, \xi).$$

Here, $B_{n,d}(f; x, \xi)$ can be either the Bernstein polynomial $B_{n,d}^{rec}$ defined in (9) for a hyper-rectangular domain, or the Bernstein polynomial $B_{n,d}^{sim}$ defined in (15) for a simplicial domain, where we have extended the notation to include the decision vector x . Similarly, we extend the notation using $EM_{n,d}(f; x) \equiv \mathbf{E}B_{n,d}(f; x, \xi)$. When solving an approximating problem of type (P_n) , in place of (P) , we want its solution x_n^* and solution value z_n^* to have desirable properties relative to (P) as n grows large. Epi-convergence plays a key role in verifying asymptotic optimality of such sequences (see, e.g., [35]). (A sequence of functions $\{g_n\}$ is said to epi-converge to g , written $g_n \xrightarrow{\text{epi}} g$, if the epi-graphs of g_n , $\{(x, \alpha) : \alpha \geq g_n(x)\}$, converge to that of g .) So, in the following proposition we obtain an epi-convergence property for approximations based on Bernstein polynomials.

Proposition 5 *Let $f : X \times \Xi \rightarrow \mathbf{R}$, where Ξ is a hyper-rectangle or simplex. Then, the Bernstein polynomials $B_{n,d}(f; x, \xi)$ converge pointwise to $f(x, \xi)$, $\forall x \in X, \xi \in \Xi$ as $n \rightarrow \infty$. If, in addition, f is continuous on $X \times \Xi$, and $f(x, \cdot)$ is convex on Ξ for all $x \in X$ then*

$$B_{n,d}(f; x, \xi) \xrightarrow{\text{epi}} f(x, \xi) \quad \forall \xi \in \Xi \quad (19)$$

and

$$\mathbf{E}B_{n,d}(f; x, \xi) \xrightarrow{\text{epi}} \mathbf{E}f(x, \xi). \quad (20)$$

Proof. Pointwise convergence of the Bernstein polynomials, $B_{n,d}(f; x, \xi) \rightarrow f(x, \xi)$ as $n \rightarrow \infty$, is provided by [24] for rectangular domain and [29] for simplicial domain. Next recall the monotonicity properties, $f(x, \xi) \leq B_{n+1,d}(f; x, \xi) \leq B_{n,d}(f; x, \xi)$, established enroute to Theorems 3 and 4. From [1, Proposition 3.12] pointwise convergence and monotonicity, coupled with (lower semi-) continuity of the limiting function, implies epi-convergence of the Bernstein polynomials, i.e., (19). Note that the epi-convergence here is with respect to x and holds separately for all $\xi \in \Xi$. Finally, (20) follows from (19) (see [6, Theorem 2.7]), provided $f(x, \xi)$ is measurable, finite with probability one and $f(x, \xi)$ and $B_{1,d}(f; x, \xi)$ have finite expectation. These results are immediate from the continuity of $f(x, \cdot)$ and the fact that the support of ξ is contained in the compact set Ξ . ■

The following theorem infers asymptotic optimality for solutions of (P_n) from the epi-convergence results established in Proposition 5.

Theorem 6 *Let the hypotheses of Proposition 5 on f and Ξ hold, and also assume that X is compact. Let x_n^* denote an optimal solution to (P_n) . Then every accumulation point of $\{x_n^*\}_{n=1}^\infty$ solves (P) and $z_n^* \rightarrow z_n$.*

Proof. The theorem's conclusion follows from [6, Corollary 2.5], provided we verify: (i) $\mathbf{E}B_{n,d}(f; x, \xi) \xrightarrow{\text{epi}} \mathbf{E}f(x, \xi)$ and (ii) the sequence of minimizing sets $\text{argmin}_{x \in X} EM_{n,d}(f; x)$ is contained in a bounded set. We established condition (i) in Proposition 5 and condition (ii) is immediate from the hypothesis that X is compact. ■

The convergence results in Theorem 6 justify solving (P_n) in place of (P) when n is sufficiently large. As we have indicated above, solving (P_n) instead of (P) , i.e., minimizing the upper bound $EM_{n,d}(f; x) \equiv \mathbf{E}B_{n,d}(f; x, \xi)$ instead of minimizing $\mathbf{E}f(x, \xi)$ might be helpful when the latter function cannot be computed exactly because the underlying random vector ξ has a continuous distribution, a discrete distribution with a large number of realizations, or has an unknown distribution for which only the first few moments are estimated.

6 Summary

In this paper we developed two decreasing sequences of higher-order upper bounds on the expectation of a convex function of a random vector whose support is contained in either a hyper-rectangle ($EM_{n,d}^{rec}$) or a simplex ($EM_{n,d}^{sim}$). The first-order terms of these two sequences reduce to bounds already available in the stochastic programming literature [13, 20, 23], and the general order- n bounds have strong connections to the Bernstein [2] polynomials from approximation theory. The $EM_{n,d}$ bounds allow the random d -dimensional vector ξ to have dependent components and can incorporate n -th order moment information for general values of n . The effort to compute $EM_{n,d}^{rec}$ grows exponentially as the dimension d grows. In contrast, the effort to compute $EM_{n,d}^{sim}$ grows polynomially in d . In applying the $EM_{n,d}$ bounds to a class of stochastic programs, we established a convergence result that allows solving the approximating stochastic program (P_n) in place of the original program (P) for sufficiently large values of n .

References

- [1] Attouch, H., Wets, R.J.-B. (1981): Approximation and convergence in nonlinear optimization. In: Mangasarian, O., Meyer, R., Robinson, S., eds., *Nonlinear Programming 4*, Academic Press, New York, pp. 367–394
- [2] Bernstein, S. (1912-13): Démonstration du théorème de Weierstrass, fondée sur le calcul des probabilités. *Commun. Soc. Math. Kharkov* (2) **13**, 1–2
- [3] Birge, J.R., Dulá, J.H. (1991): Bounding separable recourse functions with limited distribution information. *Ann. Oper. Res.* **30**, 277–298
- [4] Birge, J.R., Wallace, S.W. (1986): Refining bounds for stochastic linear programs with linearly transformed independent random variables. *Oper. Res. Lett.* **5**, 73–77
- [5] Birge, J.R., Wallace, S.W. (1988): A separable piecewise linear upper bound for stochastic linear programs. *SIAM J. Control Optimization* **26**, 725–739
- [6] Birge, J.R., Wets, R.J.-B. (1986): Designing approximation schemes for stochastic optimization problems, in particular, for stochastic programs with recourse. *Math. Program. Study* **27**, 54–102
- [7] Birge, J.R., Wets, R.J.-B. (1989): Sublinear upper bounds for stochastic programs with recourse. *Math. Program.* **43**, 131–149
- [8] Davis, P.J. (1975): *Interpolation and Approximation*. Dover Publications, New York
- [9] Dokov, S.P., Morton, D.P. (2001): Second order lower bounds on the expectation of a convex function. *Stochastic Programming E-Print Series*, <http://dochost.rz.huberlin.de/speps/>.
- [10] Dulá, J.H. (1992): An upper bound on the expectation of simplicial functions of multivariate random variables. *Math. Program.* **55**, 69–80
- [11] Dulá, J.H., Murthy, R.V. (1992): A Tchebysheff-type bound on the expectation of sublinear polyhedral functions. *Oper. Res.* **40**, 914–922
- [12] Dupačová, J. (as Žáčková) (1966): On minimax solutions of stochastic linear programming problems. *Časopis pro Pěstování Matematiky* **91**, 423–429
- [13] Dupačová, J. (1976): Minimax stochastic programs with nonconvex nonseparable penalty functions. In: A.Prékopa, ed., *Progress in Operations Research*, Mathematica Societatis János Bolyai, Eger, Hungary, pp. 303–316

- [14] Dupačová, J. (1987): The minimax approach to stochastic programming and an illustrative application. *Stochastics* **20**, 73–88
- [15] Edirisinghe, N.C.P. (1996): New second-order bounds on the expectation of saddle functions with applications to stochastic linear programming. *Oper. Res.* **44**, 909–922
- [16] Edirisinghe, N.C.P., You, G-M. (1996): Second-order scenario approximation and refinement in optimization under uncertainty. *Ann. Oper. Res.* **64**, 143–178
- [17] Edirisinghe, N.C.P., Ziemba, W.T. (1994): Bounds for two-stage stochastic programs with fixed recourse. *Math. Oper. Res.* **19**, 292–313
- [18] Edirisinghe, N.C.P., Ziemba, W.T. (1994): Bounding the expectation of a saddle function with application to stochastic programming. *Math. Oper. Res.* **19**, 314–340
- [19] Edmundson, H.P. (1956): Bounds on the expectation of a convex function of a random variable. Technical report, The Rand Corporation Paper 982, Santa Monica, California
- [20] Frauendorfer, K. (1988): Solving SLP recourse problems with arbitrary multivariate distributions – the dependent case. *Math. Oper. Res.* **13**, 377–394
- [21] Frauendorfer, K. (1992): Stochastic Two-Stage Programming. *Lecture Notes in Economics and Mathematical Systems* 392, Springer-Verlag, Berlin
- [22] Frauendorfer, K., Kall, P. (1988): A Solution Method for SLP Recourse Problems with Arbitrary Multivariate Distributions – the independent case. *Problems in Control and Information Theory* **17**, 177–205
- [23] Gassmann, H., Ziemba, W.T. (1986): A tight upper bound for the expectation of a convex function of a multivariate random variable. *Math. Program. Study* **27**, 39–53
- [24] Hildebrandt, T.H., Schoenberg, I.J. (1933): On linear functional operations and the moment problems. *Ann. of Math. (2)* **34**, 317–328
- [25] Huang, C.C., Ziemba, W.T., Ben-Tal, A. (1977): Bounds on the expectation of a convex function of a random variable: with applications to stochastic programming. *Oper. Res.* **25**, 315–325
- [26] Jensen, J.L. (1906): Sur les fonctions convexes et les inégalités entre les valeurs moyennes. *Acta Math.* **30**, 175–193

- [27] Kall, P. (1991): An upper bound for slp using first and total second moments. *Ann. Oper. Res.* **30**, 267–276
- [28] Kall, P., Wallace, S.W. (1994): *Stochastic Programming*, John Wiley & Sons, Chichester
- [29] Lorentz, G.G. (1953): *Bernstein Polynomials*, Mathematical Expositions. No. 8, University of Toronto Press, Toronto
- [30] Lorentz, G.G. (1986): *Bernstein Polynomials*. 2nd Ed. Chelsea Publishing Company, New York
- [31] Madansky, A. (1959): Bounds on the expectation of a convex function of a multivariate random variable. *Ann. Math. Stat.* **30**, 743–746
- [32] Madansky, A. (1960): Inequalities for stochastic linear programming problems. *Manage. Sci.* **6**, 197–204
- [33] Morton, D.P., Wood, R.K. (1999): Restricted-recourse bounds for stochastic linear programming. *Oper. Res.* **47**, 943–956
- [34] Powell, W.B., Frantzeskakis, L.F. (1994): Restricted recourse strategies for dynamic networks with random arc capacities. *Transportation Science* **28**, 3–23
- [35] Rockafellar, R.T., Wets, R.J.-B. (1998): *Variational Analysis*. Springer-Verlag, Berlin
- [36] Wallace, S.W. (1987): A piecewise linear upper bound on the network recourse function. *Math. Program.* **38**, 133–146
- [37] Wets, R.J.-B. (1974): Stochastic programs with fixed recourse: the equivalent deterministic program. *SIAM Rev.* **16**, 309–339

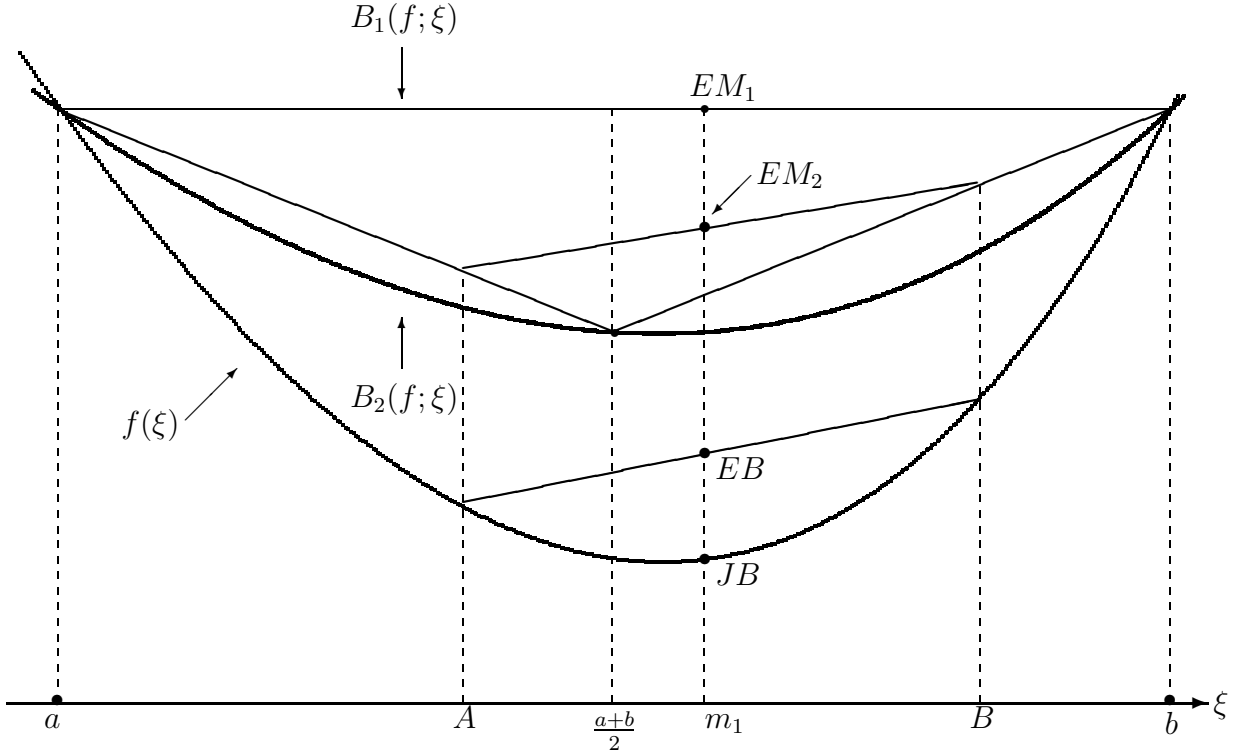


Figure 1: This figure illustrates the classical upper bound $EM \equiv EM_1$ [19] and the second-order upper bound EM_2 which we derive in Section 2 as the second term of the decreasing sequence of upper bounds $\{EM_n\}_{n=1}^{\infty}$. EM_1 involves only the mean m_1 while EM_2 involves both the mean and variance, m_1 and σ^2 , of the underlying random variable ξ . In this figure we show EM_2 in the case when ξ has mean and variance such that $A \leq (a+b)/2 \leq B$, where $A = m_1 - \sigma^2/(b - m_1)$ and $B = m_1 + \sigma^2/(m_1 - a)$. The linear $B_1(f; \xi)$ and quadratic $B_2(f; \xi)$ Bernstein [2] polynomials that majorize $f(\xi)$ are shown. Finally, the lower bounds EB of Edirisinghe [15] and JB of Jensen [26] are also indicated in the graph.

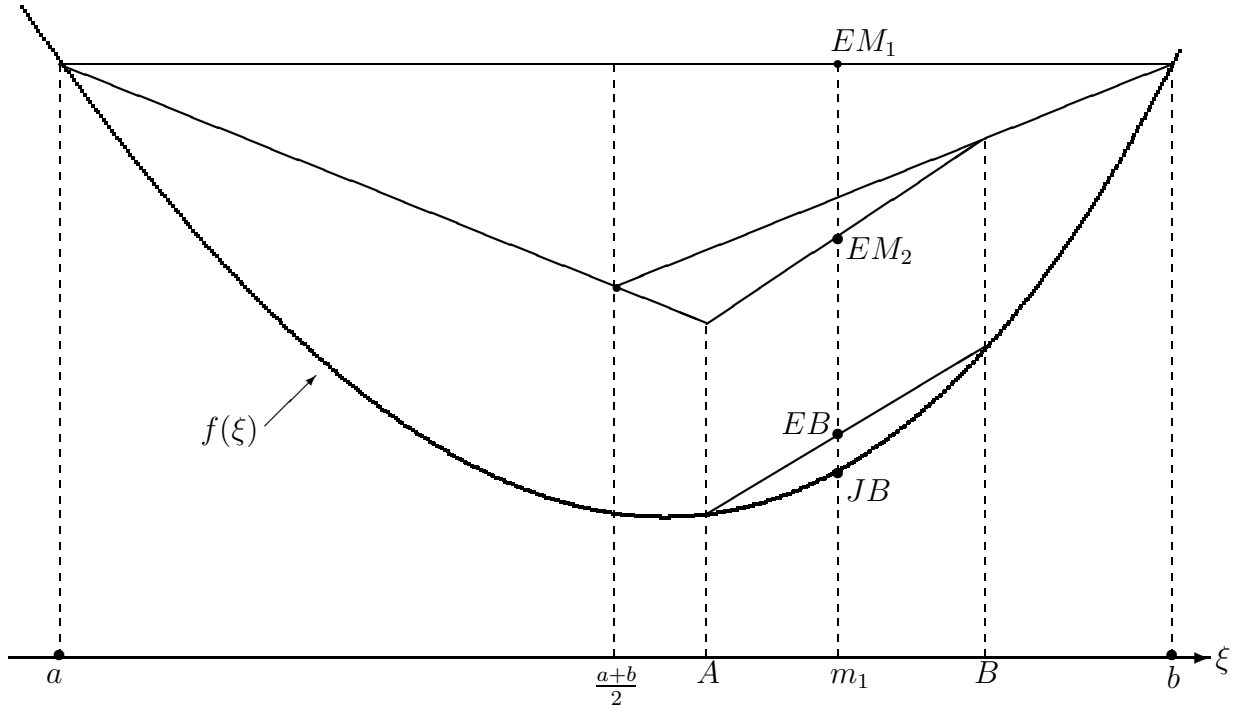


Figure 2: The figure illustrates upper bounds EM_1 and EM_2 and lower bounds JB and EB in the case when $(a + b)/2 \leq A \equiv m_1 - \sigma^2/(b - m_1)$.

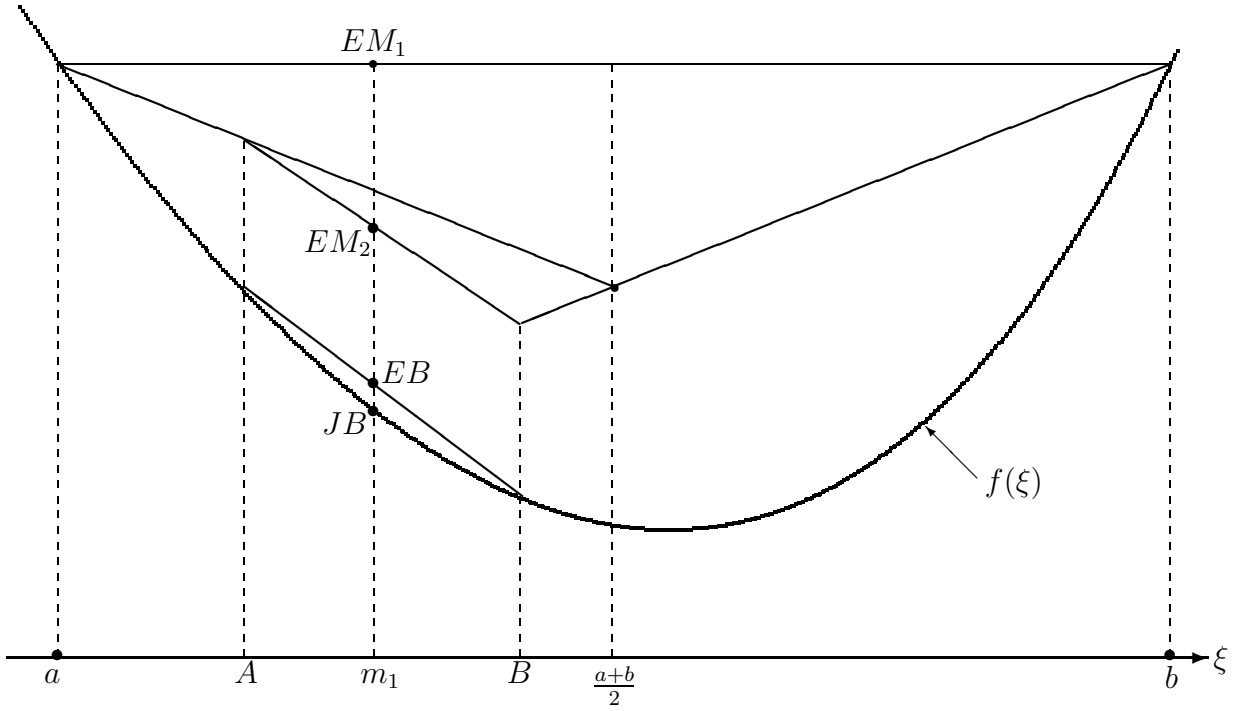


Figure 3: Similar to the previous two figures, here we illustrate upper bounds EM_1 and EM_2 and lower bounds JB and EB in the case when $m_1 + \sigma^2/(m_1 - a) \equiv B \leq (a + b)/2$.