

GENERALIZED CHEBYCHEV INEQUALITIES: THEORY AND APPLICATIONS IN DECISION ANALYSIS

JAMES E. SMITH

Duke University, Durham, North Carolina

(Received March 1993; revisions received February, June 1994; accepted August 1994)

In many decision analysis problems, we have only limited information about the relevant probability distributions. In problems like these, it is natural to ask what conclusions can be drawn on the basis of this limited information. For example, in the early stages of analysis of a complex problem, we may have only limited fractile information for the distributions in the problem; what can we say about the optimal strategy or certainty equivalents given these few fractiles? This paper describes a very general framework for analyzing these kinds of problems where, given certain "moments" of a distribution, we can compute bounds on the expected value of an arbitrary "objective" function. By suitable choice of moment and objective functions we can formulate and solve many practical decision analysis problems. We describe the general framework and theoretical results, discuss computational strategies, and provide specific results for examples in dynamic programming, decision analysis with incomplete information, Bayesian statistics, and option pricing.

In many decision analysis problems, because of computational or cognitive limitations, we have only limited information about the probability distributions involved in the problem. In these cases it is natural to ask what conclusions can be drawn on the basis of this limited information. For example, in the early stages of a decision analysis, we often have only rough estimates of, say, the 10th, 50th, and 90th percentiles of the distributions in the problem; what can we say about the optimal policy and certainty equivalent on the basis of these few assessments? In a Bayesian statistics problem, we might have extensive sample data, but only limited information about the prior; what can we say about the posterior distribution given this limited prior information? This paper describes how we can compute bounds on the quantities of interest given limited information about the underlying distribution.

The classic problem of this kind is where we are given power moments (mean, variance, etc.) for a random variable and are interested in its cumulative distribution. In this case, we can use "Chebychev's inequalities" to compute bounds on the distribution given an arbitrary number of power moments. Here we consider generalizations of these inequalities that let the "moments" be expectations of arbitrary "moment functions" defined on very general spaces; provide bounds on the expected value of an arbitrary "objective function"; and allow constraints on the set of underlying distributions. Our goal in this paper is to describe the general structure of this class of problems, review and synthesize the relevant theory, and demonstrate how, by suitable choice of moment and objective functions, we can formulate and solve many practical decision analysis problems using these techniques.

We illustrate the general framework and results by considering four specific examples. In the first example, we apply the results of the classical moment problem in dynamic programming: We develop a recursive procedure for calculating the moments of the uncertain return (or total reward) and use these moments to calculate bounds on the return distribution and its certainty equivalent. In the second example, we consider Howard's (1971) "Entrepreneur's Problem" and calculate bounds on certainty equivalents and optimal policies given limited fractile information for the state variables in the problem. In the third example, we consider a Bayesian forecasting problem and compute bounds on the posterior distribution given a few fractiles from the prior. In the final example, we compute bounds on the value of a call option given market prices for the underlying stock and related options.

The foundation for the results presented here can be traced back to the classical "problem of moments," the problem of constructing a probability distribution which matches a prescribed sequence of moments. The original "Chebychev's inequalities" were stated without proof by Chebychev in 1873 and were proven by Markov (a student of Chebychev) in his Ph.D. dissertation and independently by Stieltjes in 1884. Since then there have been many advances in the theory of the moment problem and from it flow developments in functional analysis, probability and statistics, and approximation theory. For a survey of recent results, see Landau (1987) and the other papers in that volume, especially Kemperman (1987) and Diaconis (1987). For a survey of the early history, see Shohat and Tamarkin (1943).

Despite these theoretical advances, there is still a perception that "the theory is not up to the demands of applications" (Diaconis, p. 129). One often-cited reason

Subject classifications: Decision analysis, inference: bounds given by incomplete information. Probability: generalized Chebychev inequalities. Programming, linear: applications in probability and decision analysis.

Area of review: DECISION ANALYSIS, BARGAINING AND NEGOTIATION.

for this is that the bounds given by Chebychev's inequalities—particularly the simple two-moment version given in most introductory probability and statistics texts—are quite loose. The more general versions are rarely used because of the lack of simple closed-form expressions for the bounds and the lack of “reasonably good numerical procedures” for handling the variety of cases which arise in practice (Kemperman, p. 20, Diaconis, p. 129). We pay particular attention to these criticisms by discussing general computational procedures and illustrating them with specific and practical examples.

This paper is organized as follows. In Section 1, we introduce the general framework and the four illustrative examples. In Section 2, we consider the problem of computing bounds given by moments when we place no restrictions on the allowed set of probability distributions. We view the problem as a linear programming problem, give versions of the fundamental theorem of linear programming and the duality theorem, and present results for the examples. We describe a general computational strategy for solving these problems in an appendix. In Section 3, we obtain tighter bounds by placing restrictions on the allowed set of distributions. We give some general duality results (in subsection 3.1) and consider specific examples where the distributions are constrained to be unimodal with a fixed mode (subsection 3.2), where the distributions are constrained to have a density function that lies between specified lower and upper bounds (subsection 3.3), and where the distributions are constrained to have entropy greater than a specified lower bound (subsection 3.4). In Section 4, we offer a few concluding remarks. All proofs are given in an appendix.

Our main contributions are in the review and synthesis of the theory of generalized Chebychev inequalities, the description of computational strategies, and the demonstration of applications in decision analysis. The main theoretical results—the fundamental theorem and duality results—were developed though not in their present form in Isii (1963), but seem little known in the decision analysis and operations research communities. The decision analysis applications, the general computational strategy (in Appendix A), and entropy constraints (in subsection 3.4) all appear to be new.

1. GENERAL FRAMEWORK AND ILLUSTRATIVE EXAMPLES

Our general framework follows that of Isii. Let X be an abstract space (with elements x) and \mathcal{B} a σ -algebra of measurable subsets of X . We assume that we are given $n + 1$ real-valued *moment functions* $f_i(x)$, $i = 0, 1, \dots, n$, defined on X and measurable with respect to \mathcal{B} . The expectations of these moment functions, referred to as *moments* μ_i , are assumed to be known and finite; i.e., we know that

$$\mu_i = E_P[f_i] \equiv \int_X f_i(x) dP(x) \quad \text{for } i = 0, 1, \dots, n,$$

though we may not know the underlying distribution (or nonnegative measure) P . For convenience, we take $f_0(x) \equiv 1$ so that $\mu_0 = E_P[f_0] = 1$ for all probability distributions P and write the vectors of moment functions and moments as $\mathbf{f} = (f_0, f_1, \dots, f_n)$ and $\boldsymbol{\mu} = (\mu_0, \mu_1, \dots, \mu_n)$. We assume that the moment functions are linearly independent on X in that there is no nonzero vector $\boldsymbol{\lambda}$ such that $\boldsymbol{\lambda}^T \mathbf{f}(x) = 0$ for all $x \in X$.

Given a real-valued *objective function* ϕ that is defined on X and measurable with respect to \mathcal{B} , our goal is to compute

$$\inf_{P \in \mathcal{A}(\boldsymbol{\mu})} E_P[\phi] \quad \text{and} \quad \sup_{P \in \mathcal{A}(\boldsymbol{\mu})} E_P[\phi], \quad (1)$$

the lower and upper bounds on the expectation of ϕ over the set of *allowed distributions* whose moments match $\boldsymbol{\mu}$. We let \mathcal{A} denote the set of allowed distributions and let $\mathcal{A}(\boldsymbol{\mu})$ denote the subset of \mathcal{A} matching the specified moments; i.e., those distributions P in \mathcal{A} such that $\boldsymbol{\mu} = E_P[\mathbf{f}]$. It is convenient to let \mathcal{A} include distributions with total mass not necessarily equal to 1 and then enforce the scaling requirement through the moment constraint $\mu_0 = E_P[f_0] = 1$. The distributions in \mathcal{A} are defined on (X, \mathcal{B}) and are assumed to be such that the moment and objective functions are integrable with respect to each distribution P in \mathcal{A} . In Section 2 we take \mathcal{A} to be the set \mathcal{D} of all such distributions, and in Section 3 we restrict the set of allowed distributions by taking \mathcal{A} to be a convex subset of \mathcal{D} .

This framework is quite general. The space X may be discrete or continuous, multidimensional, even infinite dimensional. The moment and objective functions may be any real-valued function, subject only to the measurability requirements. To illustrate this framework and our interest in the lower and upper bounds (1), we consider four specific examples that will be used throughout the paper.

Example 1: Dynamic Programming and the Classical Moment Problem

In many decision tree or dynamic programming problems, it is easy to compute the moments of the return distribution (sometimes called a “value lottery” or “risk profile”) but very difficult to compute the exact distribution or its certainty equivalent. For example, let us consider a discrete-time, discrete-space dynamic programming problem. Let the states be indexed by integers j and k , and let \mathcal{A} be the set of actions available. Let $p_{jk}(a)$ denote the probability of making the transition from state j to state k if action $a \in \mathcal{A}$ is chosen, let $r_{jk}(a)$ denote the reward earned in this case and let $v_n(j)$ denote the uncertain return (or total reward) for an n -stage problem starting in state j , assuming that the optimal strategy is followed. For $n = 0$, we have $v_0(j) = 0$ for all j and thus $E[v_0(j)] = 0$ for all j . For $n > 0$, we can write a recursive equation for the maximum expected return:

$$E[v_n(j)] = \max_{a \in \mathcal{A}} \sum_k p_{jk}(a)(r_{jk}(a) + E[v_{n-1}(k)]).$$

A risk-averse decision maker may be interested in the distribution of returns and its certainty equivalent as well as the expected return. While the distribution and certainty equivalent are, in general, difficult to compute, we can develop a recursive equation for the higher-order moments of $v_n(j)$ analogous to the one given for the expected return. If we assume the optimal policy a^* is followed and use the binomial expansion, the i th moment of $v_n(j)$ is given by the recursive equation

$$\begin{aligned} E[v_n^i(j)] &= \sum_k p_{jk}(a^*)(r_{jk}(a^*) + v_n(k))^i \\ &= \sum_k \left(p_{jk}(a^*) \sum_{l=0}^i \binom{i}{l} r_{jk}(a^*)^{i-l} E[v_{n-1}^l(k)] \right). \end{aligned}$$

Using this recursive formula, we can efficiently compute the leading moments of the return distribution. We can then use these moments to construct an approximate return distribution, which matches these moments and computes its certainty equivalent.¹ To determine the possible ranges of these distributions and certainty equivalents, we can compute Chebychev-type bounds.

To place this problem in our general framework, we let x represent the uncertain return of the process ($v_N(j)$ for the current stage N and state j) and take X to be the real-line \mathcal{R}^1 . The moment functions are the power functions, $f_i(x) = x^i$, and the moments are the ordinary moments about the origin, $E[x^i]$. To compute the bounds on the cumulative return distribution at a point d , we take $\phi(x) = 1$ if $x \leq d$, and $\phi(x) = 0$ if $x > d$, because the expected value of this step function is equal to the value of the cumulative distribution at the point d , $F(d)$. To compute the bounds on the entire cumulative distribution, we vary this point d . To compute the bounds on the certainty equivalent, we take $\phi(x)$ to be the decision maker's utility function and use the bounds on expected utility given by (1) to generate bounds on the certainty equivalent.

To make this example concrete, we will assume that we have computed the first four moments of the return distribution: $\mu_1 = E[x] = 5$, $\mu_2 = E[x^2] = 26$, $\mu_3 = E[x^3] = 140$, and $\mu_4 = E[x^4] = 778$. These are the moments of the normal distribution with mean 5 and standard deviation 1. We assume that the decision maker's risk preferences are captured by an exponential utility function of the form $u(x) = -\exp(-x/2)$; if the distribution is normal, then the exact certainty equivalent is $-2 \operatorname{Ln}(-E[u(x)]) = \4.75 .

Example 2: Decision Analysis With Incomplete Probability Assessments

In recent years, a number of decision analysis researchers have developed methods for reducing the probability and utility assessments required to complete an analysis. One approach is to examine all solutions that are consis-

tent with a given, limited set of assessments and see if these limited assessments are sufficient to recommend a particular action (see, e.g., Fishburn 1965, Hazen 1986, Rios Insua and French 1992, and Moskowitz, Preckel and Yang 1993). Much of this work is based on linear programming and assumes a finite number of states of nature. We will study an example with continuous states and will focus on incompleteness in the probability assessments; the payoffs and utilities are assumed to be known.

Our specific example is based on Howard's "Entrepreneur's Problem" (Howard 1971) and concerns an entrepreneur who is trying to determine a price for a new product. He faces an uncertain demand curve and has an unknown cost of production. If he charges $\$p$ for the product, the quantity sold q will be determined by $q(p) = 80[\ln 50 - \ln p] + \epsilon_q$, where ϵ_q is a random error in the estimate of the demand curve. The total cost c (measured in $\$$) to produce q units of the product is given by $c(q) = 700 + 4q + 400(1 - \exp(-q/50)) + \epsilon_c$, where ϵ_c is a random error in the cost curve. The entrepreneur's profit π is given by $\pi(p, q, c) = pq - c$ and his utility function is assumed to be $u(\pi) = -\exp(-\pi/250)$.

The entrepreneur's probability assessments are incomplete. As is common in the early stages of analysis, we will assume that the entrepreneur has specified only the 10th, 50th, and 90th fractiles for each uncertainty ($-12.82, 0$, and 12.82 for ϵ_q and $-128.2, 0$, and 128.2 for ϵ_c) and has not yet made any assertions about their joint distribution. Suppose that the entrepreneur decides to charge $\$25$ for the product; what bounds can we place on his certainty equivalent? What can we say about the set of potentially optimal prices? As a point of reference, if we assume that the errors ϵ_q and ϵ_c are independent and normally distributed with mean 0 and standard deviations of 10 and 100, respectively (these assumptions are consistent with the specified fractiles), the optimal price is $\$22.72$ and the corresponding certainty equivalent is $\$123.70$.

To place this problem in our framework, let x be a random vector of errors (ϵ_q, ϵ_c) and take $X = \mathcal{R}^2$. The moment functions f_1, f_2, \dots, f_6 are indicator functions on the events whose probabilities are specified and the moments are the specified probabilities. For example, for the first fractile specified, $f_1(\epsilon_q, \epsilon_c) = 1$ if $\epsilon_q \leq -12.82$ and 0 otherwise, and $\mu_1 = 0.10$. To compute the bounds on the certainty equivalent for a fixed price p , we take the objective function ϕ to be the function $u(p, \epsilon_q, \epsilon_c)$ that describes the entrepreneur's utilities as a function of the price p and errors ϵ_q and ϵ_c . To check whether a price p is potentially optimal, we apply a result of Hazen that says that p is potentially optimal if and only if there does not exist a convex combination of pricing strategies that dominates p , in that $E_P[u(p, \epsilon_q, \epsilon_c)] < E_P[u^*(\epsilon_q, \epsilon_c)]$ for all $P \in \mathcal{A}(\mu)$, where $u^*(\epsilon_q, \epsilon_c)$ denotes a convex combination of utility functions for different prices. To

check for dominance, we take the objective function ϕ to be $u(p, \epsilon_q, \epsilon_c) - u^*(\epsilon_q, \epsilon_c)$ and compute the lower bound in (1). If this lower bound is less than 0, p is not dominated by this combination of prices. If the lower bound is greater than 0, p is dominated and is not potentially optimal.²

Example 3: Bayesian Analysis With an Incompletely Specified Prior

In Bayesian statistics problems, we typically start with a model of the sampling process and a prior distribution for the unknown model parameters. As we observe sample data, we update our prior to obtain a posterior distribution for the model parameters. In practice, it is often difficult to obtain complete, precise assessments of the prior distribution and, consequently, a number of researchers have studied the sensitivity of the posterior distribution to the assumed prior (see Berger 1990 for a comprehensive review). Here we consider a specific example where we have an incompletely specified prior.

Our example is based on a problem given in Clemen (1991, pp. 309–310) and concerns a salesman's estimates of future sales. The salesman's errors, $\epsilon = (\text{actual sales} - \text{forecast sales})$, are assumed to be normally distributed with an unknown mean m and a standard deviation of 3,000 units. Because of the salesman's quota-based incentive system, the decision maker thinks the salesman tends to underestimate sales and believes that there is a 5% chance that m is less than 1,000 units, a 50% chance that m is less than 1,700 units, and a 95% chance that m is less than 2,400 units. Given 14 observations of actual and forecast sales (see Clemen for the sample data) and this limited information about the prior, the problem is to calculate bounds on the posterior distribution for m . As a point of reference, the specified prior fractiles are consistent with a normal distribution with a mean of 1,700 units and a standard deviation of 426 units. The sample data has a mean of 2,418 and, if the prior is normal, the posterior is normal with a mean of 1,858 and a standard deviation of 376.

To place this problem in our framework, we let x be the unknown mean m of the error distribution and take $X = \mathcal{R}^1$. As in the previous example, the moment functions f_1 , f_2 , and f_3 are step functions on the events whose probabilities are specified and the moments μ_i are the specified probabilities. Given a prior P and observations $\epsilon = (\epsilon_1, \epsilon_2, \dots, \epsilon_{14})$, the posterior cumulative distribution at d is given by Bayes' Rule as

$$E_P[\phi|\epsilon] \equiv \frac{E_P[\phi(x)L(\epsilon|m=x)]}{E_P[L(\epsilon|m=x)]},$$

where ϕ is the step function with a step at d and $L(\epsilon|m=x)$ denotes the likelihood of the observations ϵ given a mean error m equal to x . We will assume that the errors are independent (given m), so $L(\epsilon|m=x)$ is a product of normal densities with a mean m equal to x and a standard deviation of 426. Unfortunately, $E_P[\phi|\epsilon]$ is not a

simple expectation of the kind required in (1). However, we can use a "linearization technique," due to Lavine (1991), to convert this problem into a series of problems of the form of (1). For any real k , define $c(x, k) = (\phi(x) - k)L(\epsilon|m=x)$ and note that $E_P[\phi|\epsilon] > k$ if and only if $E_P[c(x, k)] > 0$. Using this, we can compute bounds on $E_P[\phi|\epsilon]$ by varying k until the corresponding upper or lower bound on $E_P[c(x, k)]$ is equal to 0.

We can calculate bounds on other quantities of interest by choosing different forms for ϕ . For example, to compute bounds on the predictive distribution for the error on a particular sales forecast, we would take $\phi(x)$ to be the cumulative probability for a normal distribution with mean x and standard deviation 3,000. The prior predictive distribution is then given by $E_P[\phi]$ and bounds are given by (1). Upon incorporation of the sample data, the posterior predictive distribution is given by $E_P[\phi|\epsilon]$ and bounds can be calculated using the linearization technique. Similarly, bounds on the mean of the prior and posterior distributions can be calculated by taking $\phi(x) = x$.

Example 4: Option Pricing

In a market context, we can use the results of this paper in a "risk-neutral" pricing framework to determine bounds on the value of one security given market prices for other related securities. To illustrate, suppose that we are given current (time 0) market prices for a stock and for a series of (European) call options on this stock that expire at time t . Let $X = [0, \infty)$ represent the possible stock prices at time t . At expiration, the holder of a call option may either buy the stock for the "strike price" K_i or let the option expire worthless. Assuming optimal exercise, the value of the option at expiration is then given by $(x - K_i)^+ \equiv \max[0, x - K_i]$.

Provided the markets do not allow risk-free arbitrage opportunities, there exists a "risk-neutral" probability distribution P , such that the price of every security is equal to its (risk-neutral) expected future value discounted at the risk-free rate for borrowing and lending. For example, if P is the risk-neutral distribution and r the risk-free rate, the current stock price (X_0) is given by $E_P[X/(1+r)^t]$ and the current price of the a option is given by $E_P[(X - K_i)^+/(1+r)^t]$. In general, the risk-neutral distribution will be unique if and only if the set of securities is sufficient to give "complete" markets. See Harrison and Kreps (1979) for a detailed discussion of this theory; see Nau and McCardle (1991) and Smith and Nau (1995) for discussions relating this theory to decision analysis.

In our framework, the moments are the observed market prices of the securities and the moment functions describe the discounted future values of the securities as a function of the stock price at expiration (x): for the stock $f_i(x) = x/(1+r)^t$ and for the call options $f_i(x) = (x - K_i)^+/(1+r)^t$, where K_i denotes the call's strike price. To compute bounds on a call option with a \$30

strike price, we take $\phi(x) = (x - 30)^+ / (1 + r)^t$. We can compute bounds on the underlying cumulative “risk-neutral” distribution by taking ϕ to be a step function as in the first example.

To make the example concrete, suppose that the current stock price is \$40 and call options that expire in 4 months with strike prices \$35, \$40, and \$45 have current prices of \$6.26, \$3.08, and \$1.26. These prices are consistent with the Black–Scholes model with a risk-free discount rate of 5% per year and an annual volatility (σ) of 30%; in this case, the exact price for a call option with a \$30 strike price is \$10.59. In the Black–Scholes model, the risk-neutral distribution is log-normal: $\ln(X/X_0)$ is normally distributed with mean $(r - \sigma^2/2)t$ and variance $\sigma^2 t$.

2. BOUNDS WITHOUT DISTRIBUTION CONSTRAINTS

We first consider the problem of computing bounds where we place no constraints on the allowed set of distributions. Focusing on the upper bound, the problem (1) can be rewritten as

$$\sup_{P \in \mathcal{D}(\boldsymbol{\mu})} E_P[\phi], \quad (2)$$

where $\mathcal{D}(\boldsymbol{\mu})$ represents the class of all distributions matching the given set of moments $\boldsymbol{\mu}$, i.e., the set of *feasible distributions*. This problem can be viewed as a linear programming problem in standard form: The decision variables are the amount of mass assigned to each point x in X and are required to be nonnegative; the objective, $E_P[\phi]$, and constraints, $\boldsymbol{\mu} = E_P[\mathbf{f}]$, are linear functions of these decision variables. If the space X is finite, (2) is a conventional linear program. If X is infinite (as in our examples), (2) is a semi-infinite linear program with an infinite number of decision variables and a finite number of constraints. Viewing (2) as a linear program, we can identify basic solutions and state versions of the fundamental and duality theorems of linear programming. Appendix A describes computational strategies for solving (2) in the specific case where $X = \mathcal{R}^k$.

2.1. Fundamental Theorem

Analogous to basic solutions in conventional linear programming, we define a *basic distribution* as a discrete probability distribution with mass points x_1, x_2, \dots, x_k such that the vectors $\mathbf{f}(x_1), \mathbf{f}(x_2), \dots, \mathbf{f}(x_k)$ are linearly independent. As there are $n + 1$ moment functions, these basic distributions have at most $n + 1$ points. We let $\Delta(\boldsymbol{\mu})$ denote the set of basic distributions that match moments $\boldsymbol{\mu}$, i.e., the set of *basic feasible distributions*.

With this definition of a basic distribution, we can state a version of the fundamental theorem of linear programming as follows. Mulholland and Rogers (1955) proved this result for the specific case in which X is the real line \mathcal{R}^1 . Isii worked in the more general framework used here but focused on discrete distributions with no more than

$n + 1$ points of support without considering the linear independence condition. The proof given in the appendix is an extension of Isii’s.

Fundamental Theorem. *Given the problem (2),*

- if there is a feasible distribution, there is a feasible basic distribution;*
- if there is an optimal distribution, there is an optimal basic distribution;*
- $\sup_{P \in \mathcal{D}(\boldsymbol{\mu})} E_P[\phi] = \sup_{P \in \Delta(\boldsymbol{\mu})} E_P[\phi].$

Parts a and b of the theorem paraphrase the conventional fundamental theorem of linear programming as stated in, for example, Luenberger (1984). Part c addresses a concern in semi-infinite linear programs that is not a concern in conventional linear programs: Even though the bound may be finite, there may be no distribution in $\mathcal{D}(\boldsymbol{\mu})$ achieving this bound. The message of the theorem is that, even if the bound is not achieved, the problem (2) requiring a search of the set of all feasible distributions $\mathcal{D}(\boldsymbol{\mu})$ can be reduced to a search of the set of basic feasible distributions $\Delta(\boldsymbol{\mu})$.³

To illustrate the use of this theorem, we consider its application in the dynamic programming example; the other examples are discussed in subsection 2.3. Here, we are given four moments of the return distribution and seek bounds on the cumulative distribution and certainty equivalent. The results are summarized in Figure 1.

To compute bounds on the cumulative at a point d , $F(d)$, we take ϕ to be a step function with a step at d . For each d , the bounds on $F(d)$ are achieved (or approached) by a 3-point discrete distribution; each discrete distribution is unique in that it is the only 3-point distribution that has mass at the point d and matches the specified moments. Moreover, these 3-point moment-matching distributions provide both upper and lower bounds on $F(x)$ at each mass point. By varying d , we generate the upper and lower bounding envelopes shown in Figure 1. Here, we see that the bounds, though the best possible based only on the given moments, are not very tight.⁴

Provided we restrict the underlying random variable, the bounds on the certainty equivalent are much tighter. The bounds on the certainty equivalent are generated by taking ϕ to be the utility function given in Section 1 and, if we place no restriction on X , are equal to $-\infty$ and 4.760. These bounds are not achieved by any distribution in $\Delta(\boldsymbol{\mu})$: They are given by constructing the 3-point moment-matching distribution with mass at d and taking the limit as d approaches $\pm\infty$. If we restrict the range of the underlying random variable to some finite interval $[a, b]$, the bounds are achieved by the 3-point moment-matching distributions with mass at a (for the lower bound) and b (for the upper bound). For example, if we take the interval $[a, b]$ to be $[0, 10]$, the bounds are given by 4.743 and 4.753.

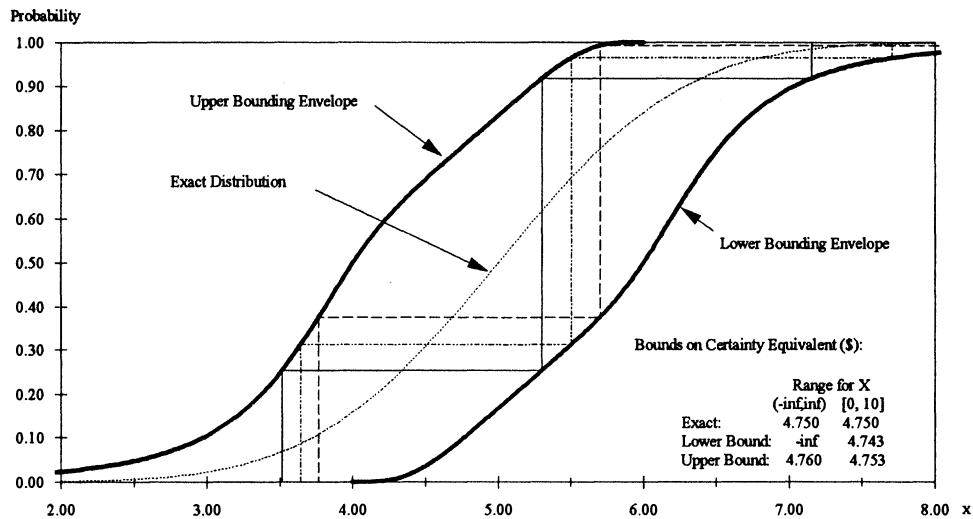


Figure 1. Bounds on return distribution for the dynamic programming example.

2.2. Duality Theorem

Associated with each primal problem of the form (2) is a corresponding dual problem. Instead of seeking a feasible distribution that maximizes $E_P[\phi]$, we seek a vector $\lambda = (\lambda_0, \lambda_1, \dots, \lambda_n)$ that solves

$$\inf_{\lambda} \{ \lambda^T \mu : \lambda^T f(x) \geq \phi(x) \text{ for all } x \in X \}. \quad (3)$$

It is not hard to see that the solution to the dual problem (3) provides an upper bound on the value of the primal problem: Taking expectations on both sides of the inequality $\lambda^T f(x) \geq \phi(x)$ yields $\lambda^T \mu \geq E_P[\phi]$ for any distribution P in $\mathcal{D}(\mu)$. These observations imply the following results.

- Weak Duality Lemma.** a. if P and λ are feasible for (2) and (3), respectively, then $E_P[\phi] \leq \lambda^T \mu$.
 b. if P and λ are feasible for (2) and (3), respectively and $E_P[\phi] = \lambda^T \mu$, then P and λ are optimal for their respective problems.

The duality theorem of conventional linear programming establishes the equality of the solutions to the primal and dual problems: if either the primal or dual problem has a finite optimal solution, so does the other and the corresponding values of the objective function are equal (see Luenberger 1984, p. 89). In our context, we can state the following slightly weaker version of the duality theorem. This result was first stated in its general form in Isii and can be proven as a special case of the more general duality theorem of the next section.

Duality Theorem. If μ is an interior point of $\{E_P[f(x)] : P \in \mathcal{D}\}$, then

$$\begin{aligned} & \sup_{P \in \mathcal{D}(\mu)} E_P[\phi] \\ &= \inf_{\lambda} \{ \lambda^T \mu : \lambda^T f(x) \geq \phi(x) \text{ for all } x \in X \}. \end{aligned} \quad (4)$$

Furthermore, if the primal problem is bounded, the dual has a finite optimal solution.

This duality result is weaker than the conventional result in that we require μ to be an interior point of the set $\{E_P[f(x)] : P \in \mathcal{D}\}$. The interior point restriction requires that μ be the moments of some distribution but rules out, for example, moments that uniquely determine the underlying distribution.⁵

To illustrate the use of the duality theorem, let us consider the dynamic programming example where we are given four power moments and compute bounds on $F(d)$ by taking the objective function ϕ to be a step function with a step at d . In this case, the functions $\lambda^T f$ are polynomials of degree 4. Figure 2 shows several polynomials $\lambda^T f$ that dominate ϕ (i.e., $\lambda^T f(x) \geq \phi(x)$ for all x), and hence are feasible for the dual problem. Geometric arguments suggest that if the polynomial is to have the least expectation, it must be equal to ϕ at three points including the point d (like the polynomial with the bold line in Figure 2). These three points are precisely the points of the discrete distribution that solves the corresponding primal problem. This result holds in general and can be stated as follows.

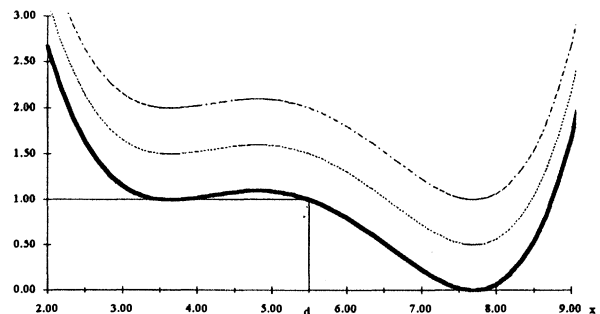


Figure 2. The dual minimization problem for the dynamic programming example.

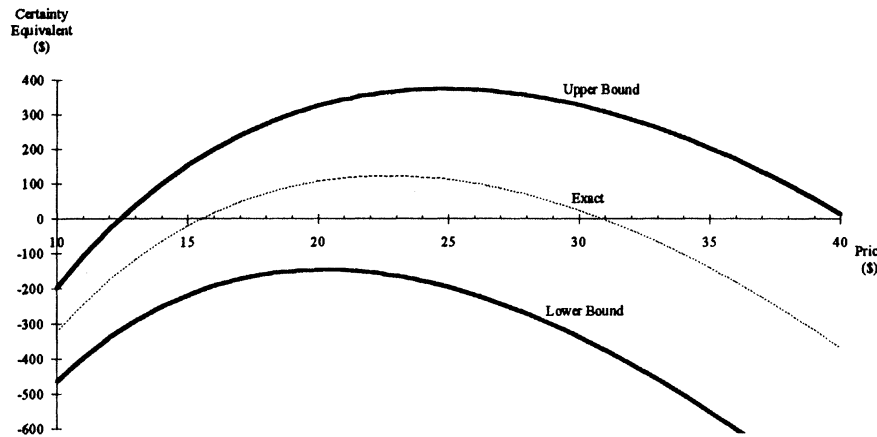


Figure 3. Bounds on certainty equivalents for incomplete assessment example.

Complementary Slackness Condition. If P and λ are optimal solutions to the primal and dual problems respectively, then P has mass only at those points x such that $\lambda^T f(x) = \phi(x)$.

The complementary slackness condition provides a valuable method for checking solutions for optimality. Given a feasible probability distribution, we can construct a polynomial $\lambda^T f$ satisfying the conditions of the complementary slackness condition. If this polynomial dominates ϕ , then the distribution is an optimal solution to (2). Conversely, given a feasible polynomial $\lambda^T f$, we can check its optimality by seeing if it is possible to construct a feasible distribution with mass restricted to the points x where $\lambda^T f(x) = \phi(x)$.

The dual perspective also offers some insight as to which objective functions ϕ will yield tight bounds: the tightness of the bounds depends on how well the function ϕ can be approximated by dominating and dominated polynomials of the form $\lambda^T f$. At one extreme, if ϕ is a linear combination of moment functions, then $E[\phi]$ is precisely determined. If ϕ is poorly approximated by the polynomials, as is the case with step function and power moments (see Figure 2), the bounds on $E[\phi]$ will not be very tight. If ϕ is well approximated by the polynomials, as is the case with power moments and exponential utility (when X is bounded below), the bounds on $E[\phi]$ will be much tighter.

2.3. Examples

The results for the other examples are similar to the dynamic programming example and reinforce the same general conclusions.

Example 2: Decision Analysis With Incomplete Probability Assessments

In this example, we are given limited fractile information for two uncertainties, demand (ϵ_q) and costs (ϵ_c) and are asked to compute bounds on the certainty equivalent for various prices p . To ensure finite bounds, we restrict ϵ_q to $[-30, 30]$ and ϵ_c to $[-300, 300]$. The results are

summarized in Figure 3. Both the upper and lower bounds are given by 4-point distributions. For every price p , the upper bound is given by a distribution with mass .1 at $(\epsilon_q, \epsilon_c) = (-12.82, -300)$, 0.4 at $(0, -128.2)$, 0.4 at $(12.82, 0)$, and 0.1 at $(30, 128.2)$; the lower bound is given by a distribution with mass 0.1 at $(\epsilon_q, \epsilon_c) = (-30, 300)$, 0.4 at $(-12.82, 128.2)$, 0.4 at $(0, 0)$, and 0.1 at $(12.82, -128.2)$. In both cases, the two uncertainties are dependent.

For any price p , the bounds on the certainty equivalent are quite loose. What can we say about the optimal p ? As discussed in Section 1, we can check for dominance and potential optimality by computing bounds on utility differences. An immediate consequence of the fundamental theorem in this context is that we can check dominance and potential optimality by considering the set of basic distributions $\Delta(\mu)$ rather than the set of all feasible distributions $\mathcal{D}(\mu)$. In the example, we find that the set of potentially optimal prices consists of $[19.6, 25.4]$. (If we assume the uncertainties are normal and independent, the exact optimum is \$22.72.) Thus we find that, even though we cannot obtain tight bounds on the certainty equivalents, we can establish a reasonably narrow range of potentially optimal prices. The dual problem offers some insight as to why this is the case: while the utility function $u(p, \epsilon_q, \epsilon_c)$ is not well approximated by the step functions whose expectations are assessed, the utility differences $u(p_1, \epsilon_q, \epsilon_c) - u(p_2, \epsilon_q, \epsilon_c)$ are more easily approximated as ϵ_q and ϵ_c have similar impacts on utility regardless of price.

Example 3: Bayesian Analysis with an Incompletely Specified Prior

In this example, we are given sample data and three fractiles of the prior distribution for the mean of the sampling distribution and are asked to compute bounds on the posterior distribution. The results are summarized in Figure 4. Here we see that, except for the tails of the distribution, the bounds are quite loose. As described in Section 1, the bounds on the posterior distribution are

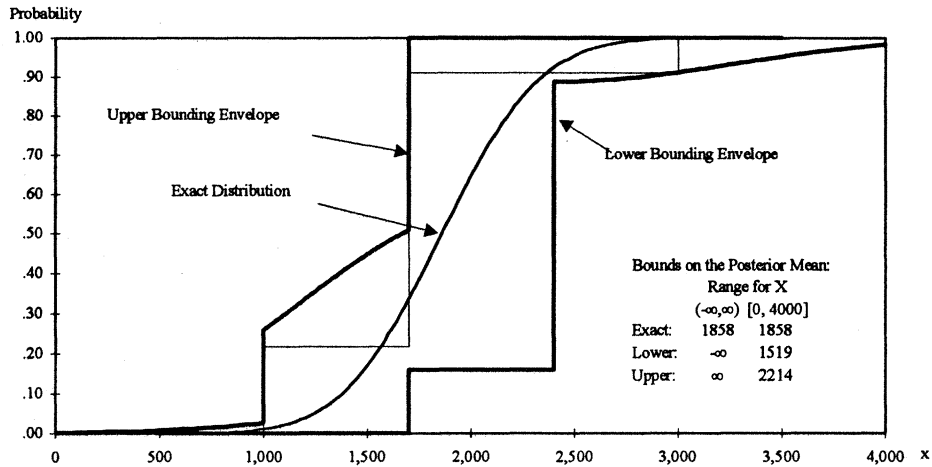


Figure 4. Bounds on posterior distribution in the Bayesian example.

calculated by computing bounds on $E_P[c(x, k)]$, where $c(x, k) = (\phi(x) - k)L(\epsilon|m = x)$, ϕ is the step function and L is the likelihood function, and varying k until the upper (or lower) bound is equal to zero; here P ranges over the set of allowed priors. As required by the Fundamental Theorem, the bounds on the posterior are achieved (or approached) by a discrete prior distribution with no more than 4 points. For example, the lower bound at the point $x = 3,000$ is achieved by a discrete prior distribution with masses 0.05, 0.45, 0.45, 0.05 at $-\infty, 1,000 + \epsilon, 1,700 + \epsilon,$ and $3,000 + \epsilon$, where ϵ is an infinitesimal positive quantity. The observed sample data (14 observations with a mean of 2,418) is inconsistent with a mean of $-\infty$, and the corresponding posterior distribution (shown as the faint line in Figure 4) assigns masses 0, 0.217, 0.694, and 0.088 to these same points.⁶

Again the dual perspective offers some insight as to why these bounds are so loose. Except for the extreme tails where $L(\epsilon|m = x) \approx 0$, the objective function $c(x, k) = (\phi(x) - k)L(\epsilon|m = x)$ is poorly approximated by linear combinations of the step functions whose expectations were specified. For example, the objective function corresponding to the bound at $d = 3,000$ is shown in Figure 5. Provided we restrict the range of the underlying variable, the bounds on the posterior mean are somewhat tighter as the objective function as with $\phi(x) = x$ the

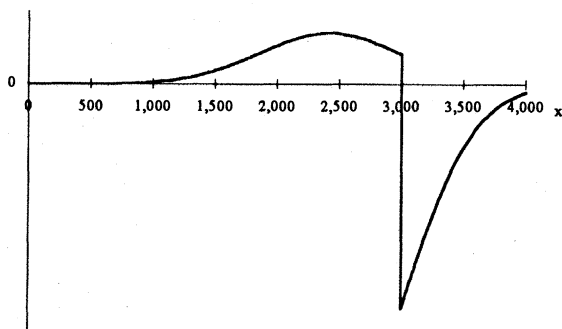


Figure 5. Objective function for the Bayesian example.

objective function $c(x, k)$ is better approximated by the step functions whose expectations are given.

Example 4: Option Pricing

The results for the option pricing example are summarized in Figure 6. Here we are given prices for a stock and call options on that stock with strike prices of \$35, \$40, and \$45. Like the earlier examples, we find that the bounds on the underlying risk-neutral distribution are quite loose. This is a reflection of the fact that the step functions used in computing these bounds are poorly approximated by the given moment functions: a “step” security that pays \$1 if and only if the stock price is less than, say, \$30 cannot be approximated very well by a portfolio consisting of the stock and given call options. The bounds on the value of a call option with a strike price of \$30 are somewhat tighter as we can better approximate its payoffs using the given securities.

3. BOUNDS WITH DISTRIBUTION CONSTRAINTS

In the previous section, we studied the problem of computing bounds where we placed no constraints on the underlying distribution. In these cases, the bounds were always achieved (or approached) by a discrete distribution and the resulting bounds were often quite loose. To achieve tighter bounds, it is natural to try to constrain the underlying distributions to rule out these discrete distributions. For example, the structure of some problems might suggest that the underlying distribution is unimodal or continuous. In this section, we study the problem of computing bounds where we restrict the underlying distributions to a convex set \mathcal{A} of allowed distributions⁷ and focus on the problem of computing

$$\sup_{P \in \mathcal{A}(\mu)} E_P[\phi], \tag{5}$$

where $\mathcal{A}(\mu)$ denotes the subset of \mathcal{A} matching the moments μ .

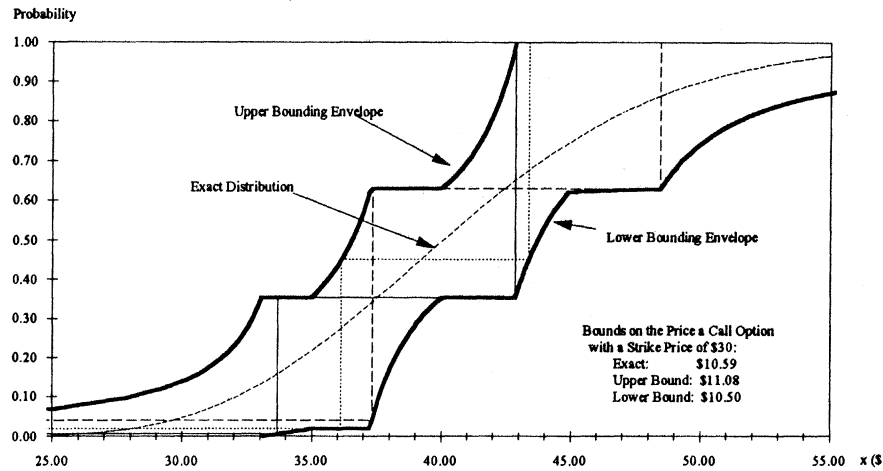


Figure 6. Bounds on the risk neutral distribution in the option pricing example.

This restriction to convex sets \mathcal{A} allows us to preserve some of the geometry underlying the unconstrained problem even though (5) is not a linear program. Some examples of useful convex constraint sets \mathcal{A} include: a) the set of all distributions \mathcal{D} ; b) the set of all distributions symmetric about a specific point; c) the set of distributions that are unimodal with a specific mode; d) the set of distributions with bounded density functions; and e) the set of all distributions with entropy greater than a specific value. Moreover, since the intersection of two convex sets is again a convex set, these examples may be combined to yield further examples of convex constraint sets. We give general results applicable to all convex constraint sets and then discuss examples c, d, and e to illustrate these general results.

3.1. General Results

Though (5) is no longer a linear program, the problem still possesses many of the same geometric features: we are maximizing a linear functional over a convex set of feasible distributions $\mathcal{A}(\mu)$. In the unconstrained case, we can identify the basic solutions as the extreme points of $\mathcal{A}(\mu)$ (see Winkler 1988 for a precise statement of and conditions for this result), and the fundamental theorem allows us to reduce our search for optimal solutions to this set of extreme points. Though this intuition still applies in this more general setting, we have not yet made enough assumptions about $\mathcal{A}(\mu)$ to be sure that it possesses extreme points, let alone be sure that the solution to (5) is achieved by one. Rather than make the assumptions required to make the fundamental theorem hold in general, we will instead identify the equivalent of “basic solutions” for specific examples of \mathcal{A} .

In contrast to the fundamental theorem, the duality results carry over directly to this more general setting. The dual problem corresponding to (5) can be written as

$$\inf_{\lambda} \{ \lambda^T \mu + \sup_{P \in \mathcal{A}} E_P[\phi - \lambda^T f] \}. \tag{6}$$

To relate this problem to the problem considered in the previous section, let \mathcal{A} be the set of all distributions \mathcal{D} . Since \mathcal{D} includes distributions that place arbitrarily large amounts of mass at any single point (or measurable set), for any fixed λ , we have

$$\begin{aligned} \sup_{P \in \mathcal{A}} E_P[\phi - \lambda^T f] \\ = \begin{cases} 0 & \text{if } \lambda^T f(x) \geq \phi(x) \text{ for all } x \in X \\ \infty & \text{otherwise.} \end{cases} \end{aligned}$$

Thus (6) reduces to

$$\inf_{\lambda} \{ \lambda^T \mu : \lambda^T f(x) \geq \phi(x) \text{ for all } x \in X \},$$

which is exactly the dual problem (3) considered in the previous section.

As in the unconstrained case, it is not hard to see that solutions to the dual problem (6) provide an upper bound on the solutions to the primal problem (5). To see this note that for any fixed λ , we have

$$\begin{aligned} \Phi(\lambda) \equiv \lambda^T \mu + \sup_{P \in \mathcal{A}} E_P[\phi - \lambda^T f] &\geq \lambda^T \mu \\ &+ \sup_{P \in \mathcal{A}(\mu)} E_P[\phi - \lambda^T f] = \sup_{P \in \mathcal{A}(\mu)} E_P[\phi]. \end{aligned}$$

Thus, we have established a generalized version of the weak duality lemma.

Weak Duality Lemma

- a. if P and λ are feasible for (5) and (6), respectively, then $E_P[\phi] \leq \Phi(\lambda)$.
- b. if P and λ are feasible for (5) and (6), respectively and $E_P[\phi] = \Phi(\lambda)$, then P and λ are optimal for their respective problems.

The duality theorem and complementary slackness conditions of the unconstrained case generalize perfectly to the constrained case. These results were first stated in their general form by Isii (1963) but can also be viewed as

an application of the Lagrange duality theorems developed in Hurwicz (1958) and discussed in Luenberger (1969). A proof based on Isii's is given in the Appendix.

Duality Theorem. *If μ is an interior point of $\{E_P[f]: P \in \mathcal{A}\}$, then*

$$\sup_{P \in \mathcal{A}(\mu)} E_P[\phi] = \inf_{\lambda} \{\lambda^\top \mu + \sup_{P \in \mathcal{A}} E_P[\phi - \lambda^\top f]\}.$$

Furthermore, if the primal problem is bounded, the dual problem has a finite optimal solution.

Complementary Slackness Condition: *If P and λ are optimal solutions to the primal and dual problems respectively, then P attains the supremum in (6) for the specified λ .*

We illustrate the use of these results by considering three examples of constraint sets. These examples are chosen both for their practical interest and to illustrate a variety of different solution strategies. In subsection 3.2, we consider the case of unimodal distributions and solve the problem by transforming the constrained problem to an equivalent unconstrained one. In subsection 3.3, we consider the case of distributions with bounded density functions and use the duality results to identify the form of the basic solutions for this problem. In subsection 3.4, we consider distributions that satisfy an entropy constraint and solve the problem using the duality theorem directly. In the case of the classical moment problem, the unimodal restriction was first treated by Johnson and Rogers (1951) and the bounded density function restriction dates back to Markov (see Shohat and Tamarkin 1943, pp. 82–87). The entropy restriction is apparently new.

3.2. Unimodal Distributions

Let $X = \mathcal{R}^1$. A distribution P on X is *unimodal with mode m* if its cumulative probability distribution $F(x)$ is convex on $(-\infty, m]$ and concave on $[m, \infty)$. The mode may be a point of discontinuity, but apart from this, unimodality requires the existence of a density function that is nondecreasing on $(-\infty, m]$ and nonincreasing on $[m, \infty)$. Without loss of generality, we can assume that the mode m is equal to 0 because we can always translate a problem with mode m to one with mode 0.

We can compute bounds over the unimodal distributions by establishing a one-to-one correspondence between the set of unimodal distributions with mode 0 and the set of all distributions \mathcal{D} that allows us to transform the unimodal problem back to the unconstrained problem treated in Section 2. The key to this transformation is the following result due to Khintchine (see Feller 1971, p. 158 for a proof).

Khintchine's Proposition. *A distribution is unimodal with mode 0 if and only if it is the distribution of a product YZ where Y and Z are independent random variables and Z is uniformly distributed on $[0, 1]$.*

To see how this result allows us to transform the unimodal problem back to the unconstrained problem, note that if X has a unimodal distribution, then for any function $f(x)$ we may write $E[f(x)]$ as

$$E[f(x)] = \int_{y=-\infty}^{\infty} \int_{z=0}^1 f(yz) dz dF_y(y) = E[g(y)],$$

where $g(y) = \int_0^1 f(yz) dz$. Since the distribution for Y is not constrained in any way, we can transform the moment and objective functions f_i and ϕ for the unimodal problem to new moment and objective functions for an equivalent unconstrained problem on Y . The optimal values for the two problems are equal and, given an optimal distribution for the unconstrained problem, we can use Khintchine's proposition to transform it to a solution to the unimodal problem: if the solution to the unconstrained problem is a discrete distribution with masses p_i at points x_i , the solution to the unimodal problem is a mixture of uniform distributions on $[0, x_i]$ (or $[x_i, 0]$ if $x_i < 0$) with mixing weights p_i . Thus, the basic solutions in the unimodal case are mixtures of $n + 1$ or fewer of these uniform distributions.

To illustrate the results in this case, we consider the dynamic programming example and suppose that the underlying distribution is known to be unimodal with mode 5. The results are summarized in Figure 7. Comparing these results with those in Figure 1, we see that we have obtained much tighter bounds on the cumulative distribution, particularly in the tails of the distribution. As in the unconstrained case, there is no lower bound on the certainty equivalent unless we place some restriction on the range of the underlying random variable. If we do restrict the range, we see that the bounds on the certainty equivalent are slightly tighter in the unimodal case than in the unconstrained case.

3.3. Distributions With Bounded Density Functions

Suppose that in addition to knowing the moments of the distribution, we know that the underlying distribution is continuous with a bounded density function. For example, if $X = \mathcal{R}^k$ we can take the allowed set of distributions to be the set of all measures P such that $dP(x) = p(x) dx$ and are bounded in that $p(x)$ lies between some specified lower and upper bounds l and u : $l(x) \leq p(x) \leq u(x)$ for all x . More generally, we consider measures that are absolutely continuous with respect to some given measure P_1 , so P can be written $dP(x) = p(x)dP_1(x)$ where p is the Radon–Nikodym derivative of P with respect to P_1 and p lies between some specified lower and upper bounds l and u .⁸

In this case, the dual problem (5) can be written in the form:

$$\inf_{\lambda} \{\lambda^\top \mu + \sup_{l(x) \leq p(x) \leq u(x)} \int_X [\phi(x) - \lambda^\top f(x)] p(x) dP_1(x)\}.$$

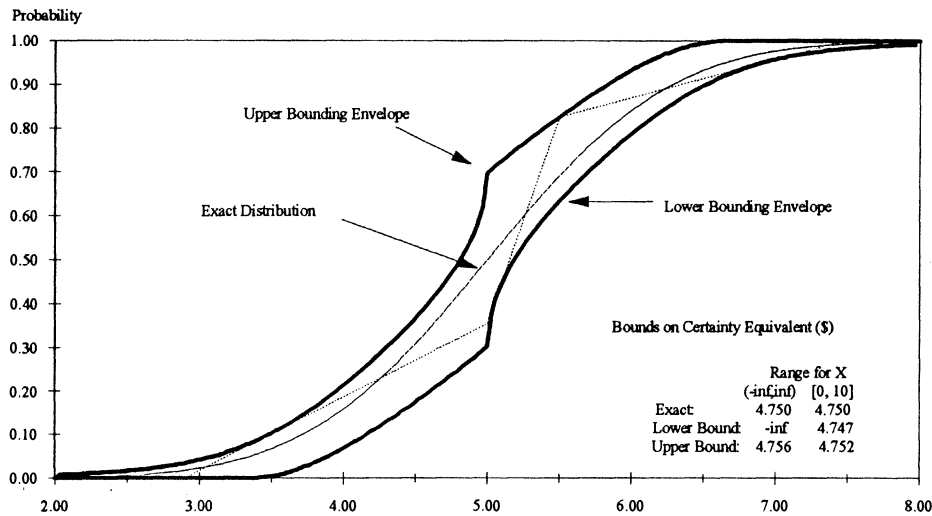


Figure 7. Results for the dynamic programming example with unimodal distributions.

For any fixed λ , the supremum is obtained by concentrating as much mass as possible in regions where $\phi(x) > \lambda^T f(x)$ and as little as possible where $\phi(x) < \lambda^T f(x)$. Thus the optimal Radon–Nikodym derivative p^* must satisfy

$$p^*(x) = \begin{cases} u(x) & \text{if } \phi(x) > \lambda^T f(x) \\ l(x) & \text{if } \phi(x) < \lambda^T f(x) \end{cases}$$

and we can restrict our search for optimal solutions to the set of basic distributions of this form.

To illustrate the results in this case, we again consider the dynamic programming example. We take P_1 to be Lebesgue measure on $(-\infty, \infty)$, so the Radon–Nikodym derivative p is a density function in the usual sense and assume uniform upper and lower bounds on the density function: $l(x) \equiv 0$ and $u(x) \equiv 0.5$. The upper bound of 0.5 corresponds to a maximum value of 0.39 for the

“true” normal distribution. The results are summarized in Figure 8. Comparing these results with the unconstrained case, we find substantially tighter bounds on the cumulative distribution and slightly tighter bounds on the certainty equivalent. Comparing these results with the unimodal case, we find tighter bounds on the cumulative near the mode of the distribution, but looser bounds in the tails; the bounds on the certainty equivalents are very similar.

3.4. Entropy Constraints

Another useful constraint set requires the allowed distributions to be “not too unusual” by requiring them to have entropy greater than some specified amount. To formalize this notion, given distributions P_1 and P_2 such that P_1 is absolutely continuous with respect to P_2 , the entropy of P_1 relative to P_2 , is

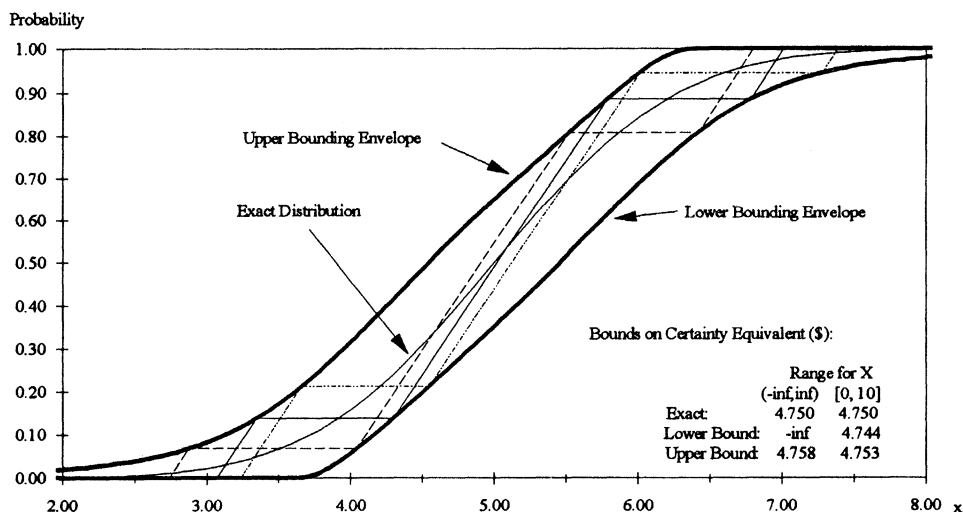


Figure 8. Results for the dynamic programming example with bounded density.

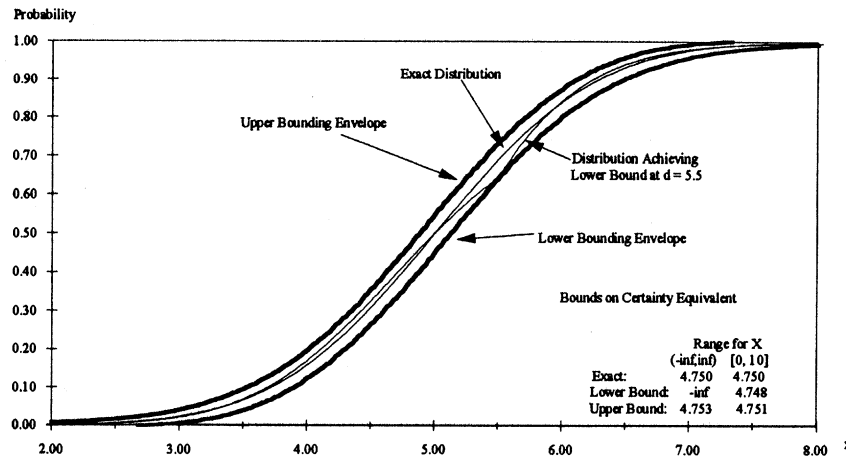


Figure 9. Results for the dynamic programming example with entropy restrictions.

$$H(P_1, P_2) \equiv - \int_X p(x) \ln p(x) dP_2(x),$$

where $p(x)$ is the Radon–Nikodym derivative of P_1 with respect to P_2 and $0 \ln 0$ is taken to be equal to 0. For distributions P_1 that are not absolutely continuous with respect to P_2 , $H(P_1, P_2)$ is defined to be $-\infty$. For our purposes, it suffices to interpret $H(P_1, P_2)$ as a measure of how much P_1 differs from the “prior” distribution P_2 : if P_1 and P_2 are identical, then $p(x) = 1$ and $H(P_1, P_2) = 0$; otherwise $H(P_1, P_2) < 0$. (See Kullback 1959 and Jaynes 1983 for complete discussions of the entropy measure and its interpretations.)

In our case, we assume that we are given the prior P_2 as well as a lower bound, H_0 , on the entropy of the allowed distributions. The problem (5) can then be written:

$$\begin{aligned} &\sup_P E_P[\phi] \\ &\text{subject to } E_P[f] = \mu \text{ and } H(P, P_2) \geq H_0. \end{aligned}$$

Since the constraint set is convex and the objective function is linear, the optimal solution will lie at a boundary of the constraint set and the entropy constraint will hold with equality. Introducing a Lagrange multiplier γ for the entropy constraint ($\gamma < 0$), we can write the dual problem (6) as

$$\begin{aligned} &\inf_{\lambda, \gamma} \sup_P \{E_P[\phi] + \lambda^T(\mu - E_P[f]) \\ &\quad + \gamma(H_0 - H(P, P_2))\}. \end{aligned} \tag{7}$$

Since $H(P, P_2) = -\infty$ for distributions P that are not absolutely continuous with respect to P_2 , when considering optimal solutions to (7), we can assume that P is absolutely continuous with respect to P_2 and focus on computing the optimal Radon–Nikodym derivative p^* . Differentiating (7) with respect to p and setting the result equal to zero gives the form of p^* as a function of λ and γ .

$$p^*(x; \lambda, \gamma) = \exp\left(-\frac{1}{\gamma} (\phi(x) - \lambda^T f(x) + \gamma)\right).$$

Substituting p^* back into (7), reduces (7) to an equivalent unconstrained optimization problem

$$\inf_{\lambda, \gamma} \{-E_{P^*}[\phi] + \lambda^T \mu + \gamma H_0\}, \tag{8}$$

where P^* denotes the distribution corresponding to the Radon–Nikodym derivative $p^*(x; \lambda, \gamma)$. The gradient for (8) is given by $[(\mu - E_{P^*}[f]), (H_0 - H(P^*, P_2))]$, so the first-order conditions for (8) requiring the gradient to be equal to zero correspond to feasibility in the original problem (6). The Hessian is also readily computed and can be shown to be positive definite (provided the moment functions f_i are linearly independent on the support of P_2). Thus, the minimization problem (8) can be solved using variations of Newton’s method (keeping $\gamma < 0$) and numerical approximations of the integrals involved.

To illustrate the results for the case of entropy constraints, we present results for each of the four examples and compare them with the results obtained when we place no constraints on the underlying distribution. In each example, we see that the bounds with entropy constraints are substantially tighter than those found earlier.

Example 1: Dynamic Programming

In this example, we are given power moments of the return (or total reward) distribution and asked to compute bounds on the distribution and its certainty equivalent. Here we take the prior P_2 to be Lebesgue measure so the Radon–Nikodym derivative p is a density function in the usual sense. We take H_0 to be 1.400 as compared to a maximum entropy for these moment constraints of 1.419.⁹ The results are shown in Figure 9. Here we see the bounds on the cumulative distribution are substantially tighter than in the other cases. The bounds on the certainty equivalent are somewhat tighter as well.

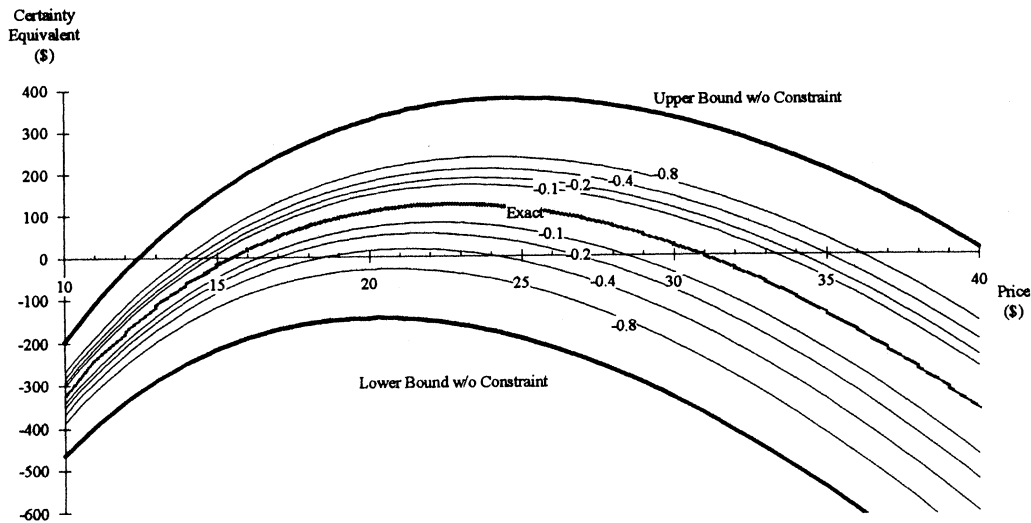


Figure 10. Results for incomplete assessment example with entropy constraints (the curves are labeled by entropy cutoff).

Example 2: Decision Analysis With Incomplete Probability Assessment

To illustrate the sensitivity of the bounds to the entropy cutoff, we consider the “Entrepreneur’s Problem” and solve for bounds given a variety of different entropy cutoffs. Here, we are given fractile information for sales and costs and are asked to compute bounds on certainty equivalents for various prices and to determine the range of potentially optimal prices. The results are summarized in Figure 10 and Table I. Here we have taken the prior P_2 to be the multivariate normal distribution that is consistent with the specified fractiles and truncated the range of the variables as in subsection 2.4. We give results for a variety of cutoffs, ranging from -0.1 to -0.8 , as compared to a maximum of 0.0 . As expected, the bounds on both certainty equivalents and optimal prices become tighter and tighter as the entropy cutoff approaches the maximum possible entropy. As the cutoff reaches the maximum, the set of allowed distributions collapses to the “exact” multivariate normal distribution that we have taken as our prior. All of these bounds are significantly tighter than those given by placing no constraint on the underlying distribution (or, equivalently, taking the entropy cutoff to be $-\infty$).

Table I
Potentially Optimal Prices for Incomplete Assessment Example With Entropy Constraints

Entropy Cutoff	Lower Bound	Upper Bound
0 (Exact)	22.7	22.7
-0.1	21.9	23.4
-0.2	21.5	23.6
-0.4	21.0	23.8
-0.8	20.5	24.1
$-\infty$ (No Constraint)	19.6	25.4

While it is difficult to suggest a precise entropy cutoff that is appropriate in this problem, on the basis of these results, it seems safe to conclude that this project is a “go” (its certainty equivalent is greater than 0) and that the product should be priced at about \$22.70. While it would be nice to have more complete assessments, in this case it seems unlikely that the conclusions would change substantially.

Example 3: Bayesian Analysis With an Incompletely Specified Prior

In this example, we are given fractiles of the prior distribution and sample data and compute bounds on the posterior distribution and its mean; the results are summarized in Figure 11. Here we have taken P_2 to be the normal distribution with parameters consistent with the specified fractiles of the prior distribution. In this case, the maximum possible entropy is 0.0 and we have taken our entropy cutoff to be -0.02 . Comparing Figures 5 and 11, we see that the bounds on the posterior distribution here are much tighter than in the unrestricted case and are particularly tight at the points corresponding to the specified prior fractiles. Given restrictions on the range of X , we find that the bounds on the posterior mean are much tighter as well: $[1,519, 2,214]$ versus $[1,821, 1,895]$.

Example 4: Option Pricing

In the option pricing example, we are given prices for a stock and three call options on the stock and seek bounds on the underlying risk-neutral distribution and the price of call option with a \$30 strike price. The results for this example are summarized in Figure 12. Here we have taken the prior P_2 to be a lognormal distribution (as assumed in the standard Black–Sholes model) with parameters consistent with the given stock and call

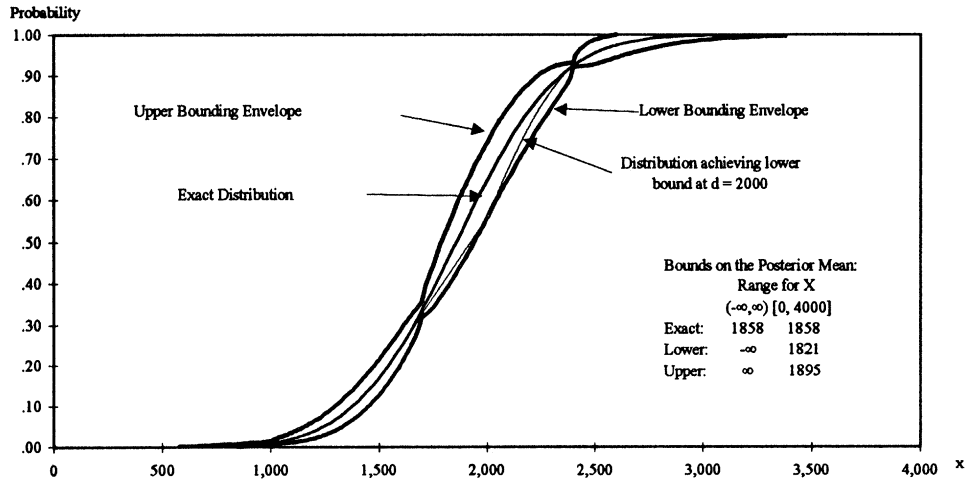


Figure 11. Results for the Bayesian example with entropy restrictions.

prices. In this case, the maximum possible entropy is 0 and we take our entropy cutoff to be -0.02 . Here, we see that the bounds on the risk-neutral distribution are much tighter than they were in the unrestricted case (compare Figures 12 and 6), and the bounds on the price of a call option with a \$30 strike price are tighter as well: [10.53, 10.66] versus [10.50, 11.08].

4. SUMMARY AND CONCLUSIONS

We have reviewed the theory of generalized Chebychev inequalities, discussed computational strategies, and illustrated their use in a variety of applications in decision analysis. We conclude with a few remarks about the usefulness of these bounds and remaining questions of theory and practice.

The usefulness of these bounds depends both on how tight they are and how easy they are to compute. The tightness of the bounds depends on how well the objective function is approximated by linear combinations of the moment functions and the set of allowed

distributions. If the objective function is well-approximated by the moment functions, as was the case with the exponential utility function and power moments (when X was bounded below), tight bounds may be obtained without placing any further restrictions on the allowed distributions. If, however, the objective function is poorly approximated by the moment functions, as was the case of the step functions and power moments, we must constrain the set of allowed distributions to obtain tight bounds.

The choice of an appropriate constraint set raises several practical concerns. While in some applications, the appropriate constraint set may be obvious (symmetry might suggest a symmetric unimodal distribution with a mode at 0), in other cases the appropriate constraint set may not be at all clear. For example, our intuition about a problem may suggest that the underlying distribution is unimodal without specifying a particular mode. Similarly our intuition may suggest that the underlying distribution has near-maximal entropy without specifying a particular

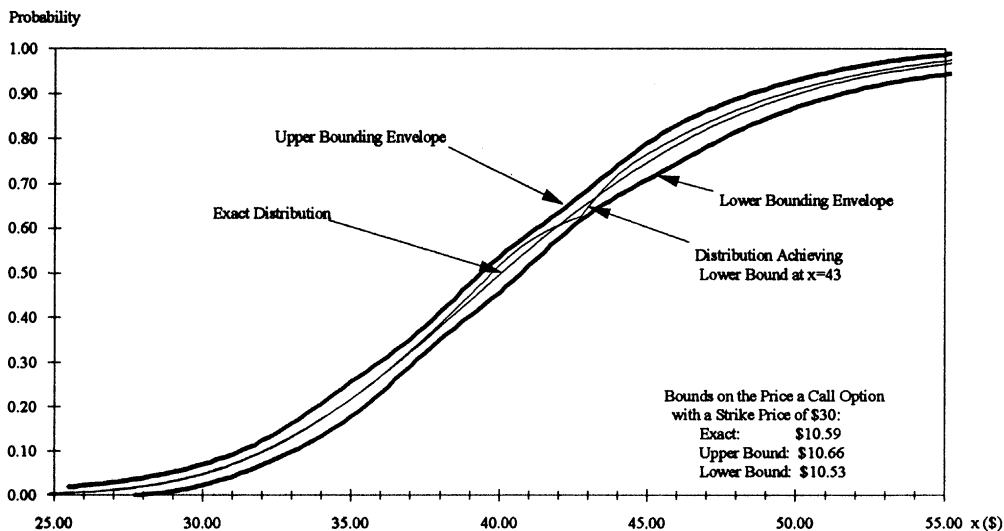


Figure 12. Results for the option pricing example with entropy constraints.

entropy cutoff. In these cases, it may be helpful to examine bounds for varying modes or entropy cutoffs even if it is impossible to determine precise constraints and, hence, precise bounds.

The constraint sets also pose computational and theoretical questions. While it is possible to give a general computational method for the unconstrained case, the algorithms for the constrained case are ad hoc in that they exploit the particular properties of the constraint set; are general algorithms possible? To address this we need to find a general characterization of the set of the basic feasible solutions in the constrained case. In the unconstrained case, the basic solutions are combinations of $n + 1$ distributions that are, in a sense, “extreme points” of \mathcal{D} ; does a similar result hold in general?

In terms of decision analysis applications, we see at least two directions for further work. Here we have focused on problems where we have incomplete but precise information about the underlying distribution. What if the information is imprecise as well as incomplete? One way to relax the precision assumption is to work with ranges of moment values. For example, rather than assessing a precise probability or fractile, one might specify bounds on the probabilities or fractiles. This would lead us to replace the equality constraints in our basic problem with inequality constraints and the results of this paper would generalize in much the same way as the conventional linear programming results generalize from equality to inequality constraints.

More generally, what if we have incomplete and imprecise information about the decision-maker’s preferences as well as his beliefs? One way to proceed is to start with statements about preferences among particular gambles (modeled as moment inequalities) and then seek to determine preferences for other gambles that are consistent with the specified preferences. Here we might assume some functional form for the decision-maker’s utility function, say an exponential utility with unknown risk-tolerance parameter, and compute bounds over the allowed set of utility functions as well as the allowed set of probability distributions. Unfortunately, the problem of computing bounds in this context is not only nonlinear but may, in fact, be nonconvex. (See Moskowitz, Preckel and Yang.)

Alternatively, given incomplete information about the decision-maker’s beliefs and preferences, one might proceed as in the option pricing example: one might assess “prices” for particular gambles and then seek to determine prices for other gambles that are consistent with the assessed prices. Provided the decision-maker wishes to be coherent and not allow arbitrage opportunities, there will be a “risk-neutral” distribution such that the price of any gamble is equal to its (risk-neutral) expected value. Then, rather than using the decision-maker’s probabilities and utilities, we would use these “risk-neutral” probabilities in the framework developed here to determine bounds on the prices for other gambles. This

approach is discussed in Nau and McCardle (1991) and Nau (1994).

In summary, we see generalized Chebychev inequalities as providing a very general framework for studying decision analysis problems where we have incomplete information about the underlying distribution. Having described the common mathematical structure of this class of problems and analyzed several examples in detail, the hope is that this discussion has made the reader aware of the broad range of applications in decision analysis and given some insights as to what kinds of results, both quantitative and qualitative, should be expected in these applications.¹⁰

APPENDIX A: A GENERAL COMPUTATIONAL STRATEGY

In this appendix, we give a general computational strategy for computing bounds in the unconstrained case. Specifically, we focus on the case where $X = \mathcal{R}^k$ or a subset of \mathcal{R}^k and solve (2):

$$\sup_{P \in \mathcal{D}(\mu)} \mathbf{E}_P[\phi].$$

Given the relationship between this problem and the conventional linear programming problem, it is natural to apply standard linear programming algorithms here. We propose a 3-step procedure which we will briefly describe and illustrate. We will not attempt to completely specify an algorithm or establish any convergence results.¹¹

STEP 1. Approximate the (potentially) infinite space X by a finite grid and then solve the resulting finite linear program using the simplex method. When choosing a grid to approximate X , the grid must be fine enough and large enough to ensure that the approximation of (2) has a feasible solution. Because the solutions to (2) tend to “pick out” those points where the moment and objective functions (or their derivatives) are discontinuous, it is a good idea to explicitly include these points in the approximating grid. Because the solution to the approximate problem is feasible for the original problem, the solution to the approximate problem provides a lower bound on the optimal value of the original problem.

To illustrate this procedure, we consider the dynamic programming example and compute an upper bound on the cumulative return distribution at the point $d = 5.5$. The objective function ϕ in this case is the unit step function with the step at 5.5. We approximate X with a grid with 11 evenly-spaced points spanning the interval $[0, 10]$ and add to the grid the point 5.5, where ϕ has a discontinuity. The optimal solution to the approximate linear program involves the five points: 3.0, 4.0, 5.5, 7.0, and 8.0, and has an objective function value of 0.934.

STEP 2. Using the solution in Step 1 as a guide, construct a feasible and approximately optimal solution to

the dual problem. The key to this step is to approximately locate the points x_0, x_1, \dots, x_q ($q \leq n$) involved in the optimal distribution. If the points x_0, x_1, \dots, x_q are the points of the exact solution, there would be a polynomial $\lambda^T f$ such that $\lambda^T f(x_i) = \phi(x_i)$ for $i = 0, 1, \dots, q$ and $\lambda^T f(x) \geq \phi(x)$ for all x . If this is the case, $\lambda^T f$ and ϕ will also be tangent at those points where f and ϕ are both continuous. Given a set of points x_0, x_1, \dots, x_q that are not necessarily optimal, we can select $n + 1$ of these conditions, q equations of the form $\lambda^T f(x_i) = \phi(x_i)$ and $n - q + 1$ equations of the form $\lambda^T f'(x_i) = \phi'(x_i)$, and solve for λ . If λ is feasible for the dual problem (i.e., $\lambda^T f(x) \geq \phi(x)$ for all x), then the objective function value $\lambda^T \mu$ provides an upper bound on the exact optimal solution.

To ensure that the approximate solution λ is feasible for the dual problem, it is often necessary to first group adjacent points of the approximate solution. In the example, the optimal solution in Step 1 assigns mass at the five points 3.0, 4.0, 5.5, 7.0, and 8.0. If we construct a polynomial $\lambda^T f$ equal to ϕ at these points, we find that $\lambda^T f(x) < \phi(x)$ between for $3 < x < 4$ and $7 < x < 8$; thus λ is not feasible for the dual problem. This suggests that ϕ and the optimal $\lambda^T f$ are tangent at a point between 3 and 4 and at a point between 7 and 8. Replacing these adjacent points by their midpoint gives the points $(x_0, x_1, x_2) = (3.5, 5.5, 7.5)$ and a polynomial $\lambda^T f$ similar to the one shown with the solid line in Figure 2. This λ is feasible for the dual problem and has an objective function value ($\lambda^T \mu$) of 0.969.

STEP 3. If greater accuracy is desired, "polish" the approximate solution to the dual problem by solving a nonlinear programming problem. Viewing the approximate solution to the dual problem λ constructed in Step 2 as a function of the contact points $x = (x_0, x_1, \dots, x_q)$, we attempt to improve $\lambda(x)$ by varying x to minimize $\lambda(x)^T \mu$. In performing this minimization, we fix the points x_i corresponding to discontinuities in the moment and objective functions (assuming that we have correctly identified these discontinuities in Step 1) and seek a local minimum in the vicinity of the original feasible solution to the dual problem (other minima may not correspond to a feasible solution). Provided that the initial points are sufficiently close to the optimal points, the minimizing polynomial will be the optimal solution for the dual problem and the minimizing points $x_0^*, x_1^*, \dots, x_q^*$ will be the points of the optimal distribution.

In the example, this polishing procedure gives the three points $(x_0^*, x_1^*, x_2^*) = (3.639, 5.5, 7.694)$. The corresponding λ is feasible and has objective function value ($\lambda^T \mu$) of 0.964. We find the masses p_j assigned to the points x_j^* by selecting q moments, say $\mu_0, \mu_1, \dots, \mu_q$, and solving $\sum_{j=0}^q f_i(x_j^*) p_j = \mu_i$ for $i = 0, 1, \dots, q$, we can verify the feasibility of the distribution by checking that the distribution matches the other moments as well. In the example, the resulting distribution has

masses $(p_0, p_1, p_2) = (0.311, 0.653, 0.036)$, is feasible for the primal problem (2), and has an objective function value ($E_P[\phi]$) equal to the dual objective function value (0.964). Since both primal and dual solutions are feasible and have equal objective function values, by condition b of the weak duality lemma, both are optimal for their respective problems.

APPENDIX B: PROOFS

We give proofs for fundamental theorem in the unconstrained case and the duality theorem and complementary slackness condition in the constrained case. The duality theorem and complementary slackness condition in the unconstrained case are a special case of the constrained results.

B.1. Proof of the Fundamental Theorem (Unconstrained Case)

To prove the fundamental theorem, we will instead prove the following main proposition which simultaneously establishes all three parts of the fundamental theorem.

Proposition. *For every $P_1 \in \mathcal{D}(\mu)$, there exists a $P_2 \in \Delta(\mu)$ such that $E_{P_2}[\phi] \geq E_{P_1}[\phi]$.*

The proof of the proposition exploits the convexity of the moment space $\mathcal{M} \equiv \{(E_P[f], E_P[\phi]): P \in \mathcal{D} \text{ and } E_P[1] = 1\}$ and its relation to the moment curve $\mathcal{S} \equiv \{(f(x), \phi(x)): x \in X\}$ and is established with the aid of two lemmas. The first lemma concerns the representation of points in the convex hull of \mathcal{S} , denoted $\text{conv}(\mathcal{S})$, and the second establishes the equality of \mathcal{M} and $\text{conv}(\mathcal{S})$.

Lemma 1. *For any point $s \in \text{conv}(\mathcal{S})$, there exists q ($q \leq n + 2$) points s_1, s_2, \dots, s_q in \mathcal{S} and positive weights w_1, w_2, \dots, w_q such that $s = \sum_{i=1}^q w_i s_i$. Furthermore, the points s_1, s_2, \dots, s_q may be selected so that they are linearly independent.*

Since $f_0 \equiv 1$, the weights w_i must sum to one, and $\sum_{i=1}^q w_i s_i$ is a convex combination of points in \mathcal{S} .

Proof. The fact that s can be represented as a positive convex combination of a finite number of points in \mathcal{S} follows from standard results of convex analysis. We now show that, if the points representing s are not linearly independent, s may be represented by a smaller set of points that are linearly independent. Since $\mathcal{S} \subseteq \mathcal{R}^{n+2}$, no more than $n + 2$ points can be linearly independent.

Suppose that s may be represented as a positive convex combination $\sum_{i=1}^q w_i s_i$ of points s_1, s_2, \dots, s_q that are linearly dependent. Then there exists numbers $\alpha_1, \alpha_2, \dots, \alpha_{n+2}$, not all zero, such that $\sum_{i=1}^{n+2} \alpha_i s_i = \mathbf{0}$. For each r such that $\alpha_r \neq 0$, we may write s_r and s as

$$s_r = -\sum_{i \neq r} (\alpha_i / \alpha_r) s_i \quad \text{and} \quad s = \sum_{i=1}^q (w_i - (\alpha_i / \alpha_r) w_r) s_i.$$

If we take r to be such that $(\omega_r/\alpha_r) = \min\{\omega_i/\alpha_i: \alpha_i > 0\}$ (if no $\alpha_i > 0$, multiply all of the α_i by -1), all of the $(w_i - (\alpha_i/\alpha_r)w_r)$ are nonnegative. Thus, we have a new convex combination with weights $(w_i - (\alpha_i/\alpha_r)w_r)$ and points s_1, s_2, \dots, s_k . The point s_r now has a zero weight and it, and any other point with zero weight, can be dropped to yield a positive convex combination with fewer points. We repeat this reduction process until we arrive at a positive convex combination with linearly independent points.

Lemma 2. $\mathcal{M} = \text{conv}(\mathcal{S})$.

Proof. We first show that $\text{conv}(\mathcal{S}) \subseteq \mathcal{M}$. For every point $s \in \mathcal{S}$, there exists a distribution in \mathcal{D} —the Dirac measure δ_x —that places its entire mass on a point x such that $(f(x), \phi(x)) = s$. Thus $\mathcal{S} \subseteq \mathcal{M}$. Since \mathcal{M} is convex, $\text{conv}(\mathcal{S}) \subseteq \mathcal{M}$.

We show that $\mathcal{M} \subseteq \text{conv}(\mathcal{S})$ by induction on the number of moments n . In the case $n = 0$, the result $\mathcal{M} \subseteq \text{conv}(\mathcal{S})$ reduces to $\{E_P[\phi]: P \in \mathcal{D} \text{ and}$

$$E_P[1] = 1\} \\ \subseteq \{\alpha\phi(x_1) + (1 - \alpha)\phi(x_2): x_1, x_2 \in X, \\ 0 \leq \alpha \leq 1\},$$

which is true for any set X . As our induction hypothesis, we assume that $\mathcal{M} \subseteq \text{conv}(\mathcal{S})$ holds for any $n - 1$ moment functions and for any set X .

Suppose that, for some X , $\mathcal{M} \subseteq \text{conv}(\mathcal{S})$ does not hold for n moments. Then there exists a point $\mu_0 \in \mathcal{M}$ such that $\mu_0 \notin \text{conv}(\mathcal{S})$. Since $\text{conv}(\mathcal{S})$ is convex, there exists a hyperplane that separates μ_0 from $\text{conv}(\mathcal{S})$; i.e., there exists a nonzero vector $\lambda \in \mathbb{R}^{n+2}$ such that $\lambda^T \mu_0 \leq 0$ and $\lambda^T \mu \geq 0$ for all $\mu \in \text{conv}(\mathcal{S})$. Since $\mu_0 \in \mathcal{M}$ there exists a distribution P_0 such that $\mu_0 = E_{P_0}[(f, \phi)]$, this implies $E_{P_0}[\lambda^T(f, \phi)] \leq 0$. Considering the points $\mu \in \mathcal{S}$, we have $\lambda^T(f(x), \phi(x)) \geq 0$ for all $x \in X$. Thus, we must have $E_{P_0}[\lambda^T(f, \phi)] = 0$ which implies that P_0 is concentrated on the set of points $X' = \{x: \lambda^T(f(x), \phi(x)) = 0\}$ (i.e., $P_0(X') = 1$). Note that this implies that the functions (f, ϕ) are linearly dependent on X' .

Since the functions (f, ϕ) are linearly dependent on X' , we can drop one of the moment functions, say f_n , and apply the induction hypothesis with X replaced by X' . Let $f' = (f_0, f_1, \dots, f_{n-1})$ denote the reduced set of moment functions, \mathcal{M}' the reduced moment space $\{(E_P[f'], E_P[\phi]): P \in \mathcal{D} \text{ and } E_P[1] = 1\}$, and \mathcal{S}' the reduced moment curve $\{(f'(x), \phi(x)): x \in X'\}$. Since $\mu'_0 \equiv E_{P_0}[(f', \phi)] \in \mathcal{M}'$, the induction hypothesis implies $\mu'_0 \in \text{conv}(\mathcal{S}')$.

We now show that this implies $\mu_0 \in \text{conv}(\mathcal{S})$. Since $\mu'_0 \in \text{conv}(\mathcal{S}')$, by Lemma 1, there exists $q (q \leq n)$ points $x_1, x_2, \dots, x_q \in X'$, and positive weights w_1, w_2, \dots, w_q such that $\mu'_0 = \sum_{i=1}^q w_i(f'(x_i), \phi(x_i))$. Because of the linear dependence of the functions (f, ϕ)

on X' , there exists a vector $a \in \mathbb{R}^{n+1}$ such that $f_n(x) = a^T(f'(x), \phi(x))$ for all $x \in X'$ and $\mu_{n0} = a^T \mu'_0$. Since

$$\mu_{n0} = a^T \mu'_0 = \sum_{i=1}^q w_i a^T(f'(x_i), \phi(x_i)) = \sum_{i=1}^q w_i f_n(x_i),$$

we have $\mu_0 = \sum_{i=1}^q w_i(f(x_i), \phi(x_i))$ and $\mu_0 \in \text{conv}(\mathcal{S})$, contradicting our assumption that $\mu_0 \notin \text{conv}(\mathcal{S})$. Thus, $\mathcal{M} \subseteq \text{conv}(\mathcal{S})$, which coupled with $\text{conv}(\mathcal{S}) \subseteq \mathcal{M}$, implies that $\text{conv}(\mathcal{S}) = \mathcal{M}$.

Proof of Main Proposition. Consider any point $(\mu, \gamma) = E_{P_1}[(f, \phi)] \in \mathcal{M}$. (Note that $\gamma < \infty$ by our assumption that $P_1 \in \mathcal{D}$.) By Lemmas 1 and 2, there exists $q (q \leq n + 2)$ points $x_1, x_2, \dots, x_q \in X$, and positive weights w_1, w_2, \dots, w_q , such that the q vectors $(f(x_i), \phi(x_i))$ are linearly independent and $(\mu, \gamma) = \sum_{i=1}^q w_i(f(x_i), \phi(x_i))$. The points $(f(x_i), \phi(x_i))$ are the vertices of a simplex Ψ that contains the point (μ, γ) . The half-line $\zeta \equiv \{(\mu, \tau): \tau \geq \gamma\}$ must intersect a proper face of Ψ at some maximal point $(\mu, \bar{\tau})$, where $\bar{\tau} = \max\{\tau: (\mu, \tau) \in \Psi\}$. Since this point lies on a proper face of Ψ , it can be represented as a positive convex combination of a subset of the q points $(f(x_i), \phi(x_i))$ containing no more than $n + 1$ points. Thus we have established the existence of a discrete distribution P_2 with mass at $n + 1$ or fewer points such that $E_{P_2}[f] = \mu$ and $E_{P_2}[\phi] = \bar{\tau} \geq E_{P_1}[\phi] = \tau$.

We still need to show that the points x_1, x_2, \dots, x_k involved in P_2 are such that the vectors $f(x_1), f(x_2), \dots, f(x_k)$ are linearly independent. Suppose they are not linearly independent. Then there exists a nonzero vector $\lambda = (\lambda_1, \lambda_2, \dots, \lambda_k)$ such that $\sum_{i=1}^k \lambda_i f(x_i) = 0$. Since the vectors $(f(x_i), \phi(x_i))$ are linearly independent, we must have $\sum_{i=1}^k \lambda_i \phi(x_i) \neq 0$. Without loss of generality, we may assume $\sum_{i=1}^k \lambda_i \phi(x_i) > 0$ (otherwise replace λ by $-\lambda$). Let $p_1, p_2, \dots, p_k > 0$ denote the masses assigned to the points x_1, x_2, \dots, x_k in P_2 . Since $E_{P_2}[f] = \mu$, we have $\sum_{i=1}^k p_i f(x_i) = \mu$. Since the masses p_i are positive, there exists some $\epsilon > 0$ such that $p_i + \epsilon \lambda_i > 0$ for all i . Then we have

$$\sum_{i=1}^q (p_i + \epsilon \lambda_i) f(x_i) = \mu \quad \text{and} \\ \tau' \equiv \sum_{i=1}^q (p_i + \epsilon \lambda_i) \phi(x_i) \\ > \sum_{i=1}^q p_i \phi(x_i) = \bar{\tau}.$$

Thus (μ, τ') is in Ψ and lies on the half-line ζ above the point $(\mu, \bar{\tau})$, contradicting our definition of $(\mu, \bar{\tau})$ as a maximal intersection of ζ and Ψ . Thus the vectors $f(x_1), f(x_2), \dots, f(x_k)$ must be linearly independent and $P_2 \in \Delta(\mu)$.

B.2. Proof of Duality Results (Constrained Case)

The proof of the duality theorem is a standard Lagrange duality type proof (see Luenberger 1969, pp. 213–238) and runs as follows.

Proof. Let $\bar{\gamma}$ and $\underline{\gamma}$ be defined as:

$$\bar{\gamma} \equiv \sup_{P \in \mathcal{A}(\mu)} E_P[\phi] \quad \text{and}$$

$$\underline{\gamma} \equiv \inf_{\lambda} \{ \lambda^T \mu + \sup_{P \in \mathcal{A}} E_P[\phi - \lambda^T f] \}.$$

We seek to show that $\bar{\gamma} = \underline{\gamma}$. Since $\bar{\gamma} \leq \underline{\gamma}$ was established in the weak duality lemma, we focus on showing that $\bar{\gamma} \geq \underline{\gamma}$. The result is trivially true if $\bar{\gamma} = \infty$, so we assume that $\bar{\gamma} \leq \infty$. The set $\mathcal{M} \equiv \{(E_P[f], E_P[\phi]): P \in \mathcal{A}\}$ is a convex subset of \mathcal{R}^{n+2} and the point $(\mu, \bar{\gamma})$ is, by definition of $\bar{\gamma}$, a boundary point of \mathcal{M} . Thus there exists a hyperplane that supports \mathcal{M} at the point $(\mu, \bar{\gamma})$; i.e., there exists a nonzero vector $(w, \alpha) \in \mathcal{R}^{n+2}$ such that

$$w^T \mu + \alpha \bar{\gamma} \geq w^T E_P[f] + \alpha E_P[\phi], \quad \text{for all } P \in \mathcal{A}. \quad (\text{B.1})$$

We now show that $\alpha > 0$. To see that $\alpha \geq 0$, consider the point $(\mu, \bar{\gamma} + \epsilon)$ for $\epsilon > 0$. Because of the definition of $\bar{\gamma}$, this point lies in the half-plane opposite of \mathcal{M} , so $w^T \mu + \alpha \bar{\gamma} + \alpha \epsilon \geq w^T E_P[f] + \alpha E_P[\phi]$ for all $\epsilon > 0$; this implies $\alpha \geq 0$. We have $\alpha \neq 0$, since otherwise μ would be a boundary point of $\{E_P[f]: P \in \mathcal{A}\}$, which contradicts our assumption that μ is an interior point. Thus $\alpha > 0$.

Defining $\lambda = -1/\alpha w$, (B.1) can be rewritten as $-\lambda^T \mu + \bar{\gamma} \geq -\lambda^T E_P[f] + E_P[\phi]$ for all $P \in \mathcal{A}$, which implies that

$$\bar{\gamma} \geq \lambda^T \mu + \sup_{P \in \mathcal{A}} E_P[\phi - \lambda^T f].$$

Thus, we have established $\bar{\gamma} \geq \underline{\gamma}$ and the existence of a λ achieving $\underline{\gamma}$ in the case where $\bar{\gamma}$ is finite.

Complementary Slackness Condition. Since $P \in \mathcal{A}(\mu)$, $E_P[\phi] = \lambda^T \mu + E_P[\phi - \lambda^T f]$. By the duality theorem, $E_P[\phi] = \lambda^T \mu + \sup_{P \in \mathcal{A}} E_P[\phi - \lambda^T f]$. Thus, $E_P[\phi] - \lambda^T E_P[f] = \sup_{P \in \mathcal{A}} E_P[\phi - \lambda^T f]$.

NOTES

1. Methods for doing this are discussed in Smith (1990, 1993). This recursive formula is easily generalized to problems with discounting as well as infinite horizon problems. This “moment approach” has been applied in a recent extension of the Utility Fuel Inventory Model (Morris et al. 1987) to compute distributions on costs associated with supply and demand disruptions.
2. In this problem, we need only check for dominance by individual prices. Because $u(p, \epsilon_q, \epsilon_c)$ is concave in p for all ϵ_q and ϵ_c , every convex combination of prices is dominated by an individual price, namely

the convex combination of prices (by Jensen’s inequality). Thus if there is a u^* that dominates p , there is a single price p^* that dominates p .

3. Assuming the bound is finite, there are various conditions that ensure the existence of a distribution achieving the bound including the requirement that the set $\{(E_P[f], E_P[\phi]): P \in \mathcal{D}\}$ be closed (see Glashoff and Gustafson 1983, p. 79). Part c of the theorem, unlike parts a and b, does not hold in semi-infinite linear programs in general; here it follows from our assumption that $f_0(x) = 1$ and $\mu_0 = 1$.
4. Notice that in this example the bounds are attained (or approached) by 3-point discrete distributions while the fundamental theorem suggests that we need to consider distributions with as many as $n + 1 = 5$ points. This kind of “degeneracy” is common in problems of the form of (2). In fact, we can often use the properties of the moment and objective functions to a priori restrict our search to distributions with fewer than $n + 1$ points. For example, with the power moments, we can be sure that the bounds on $F(d)$ will be achieved (or approached) by a discrete distribution with no more than $(n + 2)/2$ points including the point d . Many of the properties of power moments generalize to moment functions $\{f_0, f_1, \dots, f_n\}$ that form a “Chebychev system”; see Karlin and Studden (1966) for a detailed discussion of these results.
5. If μ is in $\{E_P[f(x)]: P \in \mathcal{D}\}$ but is not an interior point, then the distributions in \mathcal{D} are concentrated on a subset X_0 of X (i.e., $P(X_0) = 1$ for all $P \in \mathcal{D}$); if this is the case, (4) holds provided we replace X by the subset X_0 (see Kingman 1963). Another condition ensuring that (4) holds is that the set $\{(E_P[f], E_P[\phi]): P \in \mathcal{D}\}$ be closed; in this case, if the primal problem is bounded, the primal must have a solution though the dual may not (see Glashoff and Gustafson 1983, p. 79).
6. Note that, except for the tails of the distribution, the bounds on the posterior are looser than the bounds on the prior! For example, at the point $x = 3,000$, the lower bound on the posterior is 0.883 while in the prior, since that the 95th percentile is 2,400, the lower bound is 0.95.
7. \mathcal{A} is convex if for every $P_1, P_2 \in \mathcal{A}$, $\alpha P_1 + (1 - \alpha)P_2 \in \mathcal{A}$ for any α such that $0 \leq \alpha \leq 1$.
8. A measure P is “absolutely continuous” with respect to a measure P_1 , if $P(E) = 0$ for all events E such that $P_1(E) = 0$. If P is absolutely continuous with respect to P_1 , there exists a unique Radon–Nikodym derivative p such that $dP(x) = p(x)dP_1(x)$ (see Royden, 1968, p. 238). Radon–Nikodym derivatives are sometimes referred to as “generalized probability density functions.”
9. One way to interpret this entropy cutoff is to appeal to the “entropy concentration theorem” which can be summarized as follows. Suppose we draw N samples from P_2 (or a probability distribution proportional to P_2) and estimate P_1 using some k parameter

family of distributions that satisfy the given moment constraints (the family of distributions and estimates must satisfy certain regularity conditions). Let $H(*, P_2)$ denote the maximum entropy value consistent with the given moments (1.419 in the example) and $H(P_1, P_2)$ the entropy of the estimated distribution. Then asymptotically as $N \rightarrow \infty$, $2N[H(*, P_2) - H(P_1, P_2)]$ follows a χ^2 distribution with $k - n - 1$ degrees of freedom. If we take $N = 1,000$ and $k = 20$ in the example, we find that 99.9% ($= P[\chi_{15}^2 \geq 2 \cdot 1,000 \cdot (1.419 - 1.400)]$) of the possible distributions P_1 will have entropy greater than 1.40. See Jaynes (1983, pp. 315–336) and Kullback (1959, pp. 97–106) for precise statements of this “entropy concentration theorem.” In general, the maximum entropy distribution will have a Radon–Nikodym derivative of the form $p^*(x; \lambda) = \exp(-1 - \lambda^T f(x))$.

10. The author is grateful for the helpful comments provided by Kevin McCardle, Bob Nau, Bob Winkler, two anonymous referees, and the area and associate editors.
11. This strategy is similar to the “Three-Phase Algorithm” described by Glashoff and Gustafson though the two proposals differ in their second and third steps. Their second step does not require the approximate solution to the dual problem to be feasible. Their third step solves a set of nonlinear equations where we solve a nonlinear programming problem in fewer dimensions.

REFERENCES

- BERGER, J. O. 1990. Robust Bayesian Analysis: Sensitivity to the Prior. *J. Statist. Plan. and Infer.* **25**, 303–328.
- CLEMEN, R. T. 1991. *Making Hard Decisions: An Introduction to Decision Analysis*. PWS-Kent, Boston.
- DIACONIS, P. 1987. Application of the Method of Moments in Probability and Statistics. In *Moments in Mathematics*, H. J. Landau (ed.). Proceedings of Symposia in Applied Mathematics, Series III, 37, American Mathematical Society, 125–142.
- FELLER, W. 1971. *An Introduction to Probability Theory and Its Applications, Volume II*. John Wiley, New York.
- FISHBURN, P. C. 1965. Analysis of Decisions with Incomplete Knowledge of Probabilities. *Opns. Res.* **13**, 217–237.
- GLASHOFF, K., AND S.-A. GUSTAFSON. 1983. *Linear Optimization and Approximation*. Springer-Verlag, New York.
- HARRISON, J. M., AND D. M. KREPS. 1979. Martingales and Arbitrage in Multiperiod Securities Markets. *J. Econ. Theory* **20**, 381–408.
- HAZEN, G. B. 1986. Partial Information, Dominance, and Potential Optimality in Multiattribute Decision Theory. *Opns. Res.* **34**, 296–310.
- HOWARD, R. A. 1971. Proximal Decision Analysis. *Mgmt. Sci.* **17**, 507–541.
- HURWICZ, L. 1958. Programming in Linear Spaces. In *Studies in Linear and Non-Linear Programming*. K. J. Arrow, L. Hurwicz, and H. Uzawa. Stanford University Press, Stanford, Calif.
- ISII, K. 1963. On Sharpness of Tchebycheff-Type Inequalities. *Ann. Inst. Stat. Math.* **14**, 185–197.
- JAYNES, E. T. 1983. *Papers on Probability, Statistics, and Statistical Physics*, R. D. Rosenkrantz (ed.). Reidel, Dordrecht, The Netherlands.
- JOHNSON, N. L., AND C. A. ROGERS. 1951. The Moment Problem for Unimodal Distributions. *Ann. Math. Stat.* **22**, 433–439.
- KARLIN, S., AND W. J. STUDDEN. 1966. *Tchebycheff Systems: With Applications in Analysis and Statistics*. Interscience Publishers, New York.
- KEMPERMAN, J. H. B. 1987. Geometry of the Moment Problem. In *Moments in Mathematics*. H. J. Landau (ed.). Proceedings of Symposia in Applied Mathematics, Series III, vol. 37, American Mathematical Society, 16–53.
- KINGMAN, J. F. C. 1963. On Inequalities of the Tchebychev Type. *Proc. Camb. Phil. Soc.* **59**, 135–147.
- KULLBACK, S. 1959. *Information Theory and Statistics*. John Wiley, New York.
- LANDAU, H. J. 1987. Classical Background on the Moment Problem. In *Moments in Mathematics*, H. J. Landau (ed.). Proceedings of Symposia in Applied Mathematics, Series III, vol. 37, American Mathematical Society, 1–15.
- LAVINE, M. 1991. Sensitivity in Bayesian Statistics: The Prior and Likelihood. *J. Am. Stat. Assoc.* **86**, 396–399.
- LUENBERGER, D. G. 1969. *Optimization by Vector Space Methods*. John Wiley, New York.
- LUENBERGER, D. G. 1984. *Linear and Nonlinear Programming*. Addison-Wesley, Reading, Mass.
- MORRIS, P. A., M. J. SANDLING, R. B. FANCHER, M. A. KOHN, H. CHAO AND S. W. CHAPEL. 1987. A Utility Fuel Inventory Model. *Opns. Res.* **35**, 169–184.
- MOSKOWITZ, H., P. V. PRECKEL AND A. YANG. 1993. Decision Analysis With Incomplete Probability and Utility Information. *Opns. Res.* **41**, 864–879.
- MULHOLLAND, H. P., AND C. A. ROGERS. 1958. Representation Theorems for Distribution Functions. *Proc. London Math. Soc. (3)* **8**, 177–223.
- NAU, R. F., AND K. F. MCCARDLE. 1991. Arbitrage, Rationality, and Equilibrium. *Theory and Decision* **31**, 199–240.
- NAU, R. F. 1994. Coherent Decision Analysis With Inseparable Probabilities and Utilities. *J. Risk and Uncert.* **10**, 71–91.
- RIOS INSUA, D., AND S. FRENCH. 1991. A Framework for Sensitivity Analysis in Discrete Multi-Objective Decision Making. *Eur. J. Opnl. Res.* **54**, 176–190.
- ROYDEN, H. L. 1968. *Real Analysis, Second Edition*. Macmillan, New York.
- SHOHAT, J. A., AND J. D. TAMARKIN. 1943. *The Problem of Moments*. Mathematical Surveys, No. 1, American Mathematical Society, New York.
- SMITH, J. E. 1990. Moment Methods in Decision Analysis. Unpublished Ph.D. Dissertation, Department of Engineering–Economic Systems, Stanford University, Stanford, Calif.
- SMITH, J. E. 1993. Moment Methods in Decision Analysis. *Mgmt. Sci.* **39**, 340–358.
- SMITH, J. E., AND R. F. NAU. 1995. Valuing Risky Projects: Option Pricing Theory and Decision Analysis. *Mgmt. Sci.* (to appear).
- WINKLER, G. 1988. Extreme Points of Moment Sets. *Math. O. R.* **13**, 581–587.