Entropy Methods For Univariate Distributions in Decision Analysis
Ali E. Abbas

Citation: AIP Conference Proceedings 659, 339 (2003); doi: 10.1063/1.1570551
View online: http://dx.doi.org/10.1063/1.1570551
View Table of Contents: http://scitation.aip.org/content/aip/proceeding/aipcp/659?ver=pdfcov
Published by the AIP Publishing

Articles you may be interested in
On Bayesian Inference, Maximum Entropy and Support Vector Machines Methods
AIP Conf. Proc. 872, 43 (2006); 10.1063/1.2423259

On The Relationship between Bayesian and Maximum Entropy Inference
AIP Conf. Proc. 735, 445 (2004); 10.1063/1.1835243

Maximum Probability and Maximum Entropy methods: Bayesian interpretation

An Entropy Approach for Utility Assignment in Decision Analysis
AIP Conf. Proc. 659, 328 (2003); 10.1063/1.1570550

Using Bayesian Analysis and Maximum Entropy To Develop Non-parametric Probability Distributions for the Mean and Variance
Entropy Methods For Univariate Distributions in Decision Analysis

Ali E. Abbas

Department of Management Science and Engineering, Stanford University, Stanford, Ca, 94305

Abstract. One of the most important steps in decision analysis practice is the elicitation of the decision-maker's belief about an uncertainty of interest in the form of a representative probability distribution. However, the probability elicitation process is a task that involves many cognitive and motivational biases. Alternatively, the decision-maker may provide other information about the distribution of interest, such as its moments, and the maximum entropy method can be used to obtain a full distribution subject to the given moment constraints. In practice however, decision makers cannot readily provide moments for the distribution, and are much more comfortable providing information about the fractiles of the distribution of interest or bounds on its cumulative probabilities. In this paper we present a graphical method to determine the maximum entropy distribution between upper and lower probability bounds and provide an interpretation for the shape of the maximum entropy distribution subject to fractile constraints, (FMED). We also discuss the problems with the FMED in that it is discontinuous and flat over each fractile interval. We present a heuristic approximation to a distribution if in addition to its fractiles, we also know it is continuous and work through full examples to illustrate the approach.

INTRODUCTION

In many Decision Analysis applications, a decision-maker is interested in an uncertain quantity \( \theta \). When data sets are available she may provide moments of the distribution but in many situations that arise in practice, moments are not available and the decision maker provides percentiles, (fractiles), or other information representing her belief about the shape of the distribution. The maximum entropy formulation for moments and (or) fractile constraints is given as:

\[
\max_{f(x)} \int_a^g f(x) \ln(f(x)) \, dx - \int_a^g f(x) \, dx \\
\text{subject to } \int_a^g h_i(x) f(x) \, dx = \mu_i, \quad i = 0,1,...,n;
\]

Where, \([a,g]\) is the support of the maximum entropy distribution, \( h_i(x) \) is either an indicator function over an interval for fractile constraints or \( x \) raised to a certain

CP659, Bayesian Inference and Maximum Entropy Methods in Science and Engineering: 22nd International Workshop, edited by C. J. Williams
© 2003 American Institute of Physics 0-7354-0119-5/03/$20.00

339
power, for moment constraints, and \( \mu \)'s are a given sequence of fractiles or moments.

The solution to this problem has the form:

\[
f(x) = \exp\left(-\sum_{i=0}^{k} \alpha_i h_i(x)\right)
\]

Where \( \alpha_i \) is the Lagrange multiplier for each fractile or moment constraint.

The maximum entropy solution subject to fractile constraints produces a discontinuous stair case density function satisfying the constraints in each interval. We will refer to this distribution as the Fractile Maximum Entropy Distribution (FMED). If the density function has an unbound support in any direction, the unbound FMED has an exponential tail in that direction.

An example of a FMED distribution given five common fractiles used in practice, (1%, 25%, 50%, 75%, and 99%), and a bounded support is shown in Figure 1.

![FIGURE 1](image-url)

In the next section we will present a solution to a more general formulation, where bounds on the cumulative distribution are available rather than precise fractiles. We will refer to this distribution as the Taut String distribution. Note that the FMED is a special case of the taut string distribution when upper and lower bounds coincide. We will also present a heuristic approximation to a distribution if in addition to its fractiles, we also know it is continuous. We will refer to this distribution as the Mid Point Maximum Entropy Distribution, MMED, and compare it to other approaches.

**THE MAXIMUM ENTROPY DISTRIBUTION BETWEEN UPPER AND LOWER PROBABILITY BOUNDS**

In many decision analysis applications, a decision-maker is interested in a probability distribution for some random quantity, \( \theta \). The decision analyst may help
elicit her probability distribution, however, the elicitation of a full probability
distribution is a task that involves many cognitive and motivational biases. [1], [2].
Alternatively, the decision-maker may provide other information about the fractiles
of the distribution or bounds on these fractiles. The decision maker may also consult an
expert who provides another cumulative distribution for $\theta$. Faced with two probability
distributions, or upper and lower bounds on the cumulative distribution, we would like
to find an unbiased distribution that lies between these bounds. A similar problem that
arises is when the analyst has knowledge of some utility values or bounds on some
utility values, obtained during a utility assessment, and would like to find an unbiased
utility curve between the upper and lower utility bounds. [3].

We will present an exact graphical method to solve these problems based on the
maximum entropy principle. [4]. We will refer to the solution as the “Taut String”
distribution. Without significant loss of generality, we will focus on the maximum
entropy distribution between upper and lower bounds of two probability distributions.

Now let us move to the formulation of our problem: given two cumulative
probability distributions, $R$ and $Q$ for the same quantity $\theta$, what is the maximum
entropy distribution, $P$, which lies between their upper and lower bounds? The
distributions, $P$ and $Q$, may be continuous or discrete.

We will start the analysis with the case where there is stochastic dominance between
the two distributions, this implies the two distributions do not cross, and then
generalize to the case where stochastic dominance does not exist.

The problem can be formulated in discrete form as follows:

$$\text{maximize } - \sum_{i=1}^{n} p_i \log(p_i)$$

such that

$$\sum_{i=1}^{n} p_i = 1, \quad \text{Constraint (I)}$$

$$\sum_{j=1}^{r_i} r_j \leq \sum_{j=1}^{p_i} p_j \leq \sum_{j=1}^{q_j} q_j, \quad \forall i = 1, \ldots, n \quad \text{Constraints (II)}$$

$$p_j \geq 0, \quad \forall i = 1, \ldots, n \quad \text{Constraints (III)}$$

Where, $r_j, q_j$ and $p_j$ are the discrete probabilities of outcome $\theta_j$ for distributions $R$, $P$, and $Q$ respectively. For shorthand we will use $R_i = \sum_{j=1}^{r_i} r_j$, and $Q_i = \sum_{j=1}^{q_j} q_j$.

We will present the solution to this problem in discrete form but the solution to the
continuous form follows by analogy. We start the analysis by solving the formulation
using the first constraint alone and any strictly additive concave function, $\sum_{i=1}^{n} h(p_i)$, where $h(p_i)$ is any strictly concave term of $p_j$(not necessarily the entropy expression).  

341
maximize $\sum_{i=1}^{n} h(p_i)$

subject to $\sum_{i=1}^{n} p_i = 1$  

Using the method of Lagrange multipliers,

$$L = \sum_{i=1}^{n} h(p_i) - \lambda \left( \sum_{i=1}^{n} p_i - 1 \right)$$  

$$\frac{\partial L}{\partial p_i} = h'(p_i) - \lambda = 0 \Rightarrow h'(p_i) = \lambda \ \forall i = 1, \ldots, n$$  

From the strict concavity of $h(p)$, equality of the derivatives implies that

$$p_i = p_j \ \forall i, j \in 1, \ldots, n.$$  

Now if we take the derivative with respect to the Lagrange multiplier and use the results of Equation 7, we get,

$$\frac{\partial L}{\partial \lambda} = \sum_{i=1}^{n} p_i - 1 = 0 \Rightarrow n p_i = 1$$

$$\Rightarrow p_i = \frac{1}{n}, \ \forall i = 1, \ldots, n$$  

This is the familiar uniform solution obtained when maximizing the entropy of a distribution but note that this solution is invariant for any strictly additive concave function, $\sum_{i=1}^{n} h(p_i)$, and is not unique to the entropy expression. In other words we can maximize any concave term, or minimize any convex term, and obtain the same result.

Now we consider the second constraint. The solution to the formulation with the second constraint alone appears in many network flow and supply chain problems where the objective function to be minimized is a convex cost function. The solution, if one exists, is unique and can be determined using the Karush-Kuhn-Tucker optimality conditions. The solution is also invariant to the maximization of any additive concave term or the minimization of any additive convex term. [5].

Now we move on to the third constraint: The non-negativity of the probability values implies we select a non-decreasing probability distribution out of all the possible solutions satisfying constraints (I) and (II).

Using the previous results, the solution to our problem is thus equivalent to the maximization of any additive concave function, or the minimization of any additive
convex function subject to the given constraints. With no loss of generality, let us choose to minimize an additive convex function, $\sum_{i=1}^{n} \sqrt{p_i^2 + \Delta^2}$, where $\Delta$ is the discretizing interval for the values of the variable $\theta$. This new objective function is actually the distance (path) in the plane of the cumulative probability distribution starting from point A to point B, as shown in Figure 2. So in fact our problem is reduced to the problem of finding the shortest path in the plane, (new objective function), that lies between the two distributions, (condition 2), starts at a probability zero and ends with a probability 1, (required by condition 1 and automatically satisfied by condition 2), and is non-decreasing, (required by condition 3 and is satisfied by being the shortest path as will follow).

The problem of choosing a distribution that maximizes the entropy subject to lower and upper constraints now has a geometric interpretation. It is the shortest path from the first to the last point among the paths that lie between the lower and upper bounds. To find this path, imagine pins in the plane at the points $(\theta_i, R_i)$ and $(\theta_i, Q_i)$ for $0 \leq i \leq n$, where $\theta_i$ is the value of the variable at which the assessment took place, $R_i = r_1 + r_2 + \ldots + r_i$, and $Q_i = q_1 + \ldots + q_i$. Now thread a string between the pins at $(\theta_i, R_i)$ and $(\theta_i, Q_i)$ for $1 \leq i \leq n$, and pull the string taut. The taut string traces out the shortest path and is also the maximum entropy solution. The taut string distribution is non-decreasing, satisfying condition 3, because both upper and lower bounds are non-decreasing. This is shown in Figure 2.

Note that the taut string does not have to be linear but takes the shape of the shortest path, which could be one of the bounds themselves depending on the geometry of the upper and lower bounds.

![Taut String solution between two distributions with Stochastic Dominance](image)

**FIGURE 2.** The taut string distribution between two distributions with stochastic dominance.
Now we move to the case where there is no stochastic dominance between the two distributions. In this case we are interested in the maximum entropy distribution between their upper bound, $\max(R_i, Q_i)$, and lower bound, $\min(R_i, Q_i)$. Once again, we can solve for this distribution using a taut string between the upper and lower bounds. This is justified by the following argument: If there is no stochastic dominance, then the distributions will cross at least once. This crossing point comprises both upper and lower bounds at that value of $\theta$ so the taut string must pass through it. Now if we start at the first crossing, the non-decreasing value of the cumulative probability distributions guarantees that further crossings will have a higher cumulative probability. A taut string through them thus satisfies constraints (III). The fact that the distributions end at unity satisfies constraint (I) and since the taut string lies between the upper and lower bounds, constraints (II) are satisfied. This is shown in Figure 3.

![Figure 3](image.png)

**Figure 3.** The taut string distribution between the upper and lower bounds of two distributions where there is no stochastic dominance.

We summarize the previous results in the following theorem:

**Theorem 1: The Taut String Distribution**

"The distribution that maximizes the entropy and lies between the upper and lower bounds of two cumulative distributions can be determined graphically by a taut-string distribution."

344
A special case of the taut string distribution is the maximum entropy distribution given fractile constraints. We can think of fractiles constraints as upper and lower bounds that coincide and so the taut string must pass through them. This explains the piece wise linear shape of the FMED in Figure 1 (a). Note however, that the probability density function corresponding to this solution is discontinuous. Figure 1 (b). In the next section we will explore some methods that build on the taut string solution given fractile constraints and incorporate knowledge of continuity of the corresponding probability density function.

**MID-POINT MAXIMUM ENTROPY DISTRIBUTION (MMED)**

In this section we show a heuristic approximation to a distribution if in addition to its fractiles and support, we also know it is continuous. We will make use of a simple geometric property of the FMED to obtain this approximation:

**Property 1: Equal Area Constraint**

If the probability density function, from which the fractiles were assessed, is continuous and monotonically increasing or decreasing on any finite interval, then this probability density function crosses the FMED at least once in each fractile interval.

This property is a result of the equal area constraint for both density functions on each fractile interval. Further, if the distribution is unimodal, then it crosses the FMED exactly once every fractile interval except for the interval containing the mode, which can cross twice. Figure 4 shows a comparison of the Beta(4,6) distribution and the FMED constructed from its fractiles.

![Beta(4,6) vs. FMED](image)

**FIGURE 4.** Beta (4,6) distribution and FMED. An intersection of the two distributions occurs at each fractile interval.
Constructing The Mid-point Maximum Entropy Distribution

Property (1) says there will be at least one crossing, however, we are uncertain about its precise location. For simplicity, let us assume it occurs at the mid-interval for each fractile of the FMED. The mid-point of each interval is where the actual intersection is probabilistically equally likely to be above or below. Now, we connect these mid points and normalize the resulting distribution. We will call this distribution the Mid-Point Maximum Entropy distribution.

Figure 5 (a) shows the construction of the MMED from the unbound FMED given fractiles of a Beta (4,6). The mid-intervals for each inner fractile are connected. The resulting MMED is shown in Figure 5 (b) and compared to the original Beta (4,6). Using the unbound FMED is a good exploitation to knowledge of zero values of the density function at the end points of the support. In Figure 6 (a) we show the bounded FMED vs. a truncated exponential, \( f(x) = 0.019e^{-x/0.01} \), and in Figure 6 (b), we compare it to the corresponding MMED.

![Image 5](image5.png)

FIGURE 5. (a). Construction of the MMED by connecting the midpoints of inner fractile intervals of the unbound FMED. (b) Comparison of MMED and Beta (4,6).

![Image 6](image6.png)

FIGURE 6. (a). Truncated exponential vs. bounded FMED from its fractiles. (b) Truncated exponential vs. MMED constructed by connecting mid-interval points of the FMED.
Comparison With Other Approaches

Several other approaches for estimating density functions using incomplete information have been proposed in literature. One method uses the theory of splines to connect the fractiles on the cumulative distribution curve. While this method yields good results, it is more complex and there is some arbitrariness in the order of the splicing polynomial. In addition, constraints may be needed to prevent the polynomial from going negative.

Other methods suggest using a skew logistic distribution whose parameters are calculated from three fractiles. [6]. Figure 7 shows the cumulative distribution for a Beta(4,6), the MMED, and a skew logistic distribution constructed from fractiles of the Beta (4,6). The skew logistic distribution method yields very good results and only requires three fractiles to determine the parameters of the equivalent distribution. The problem, (as mentioned in Lindley’s paper), is that there is a limit on the amount of skewness that can be modeled by a skew logistic distribution and it can only be applied to unimodal distributions.

In Figure 8, we show a comparative example for the “closeness” of a Beta (4,6) to its estimated distributions using different methods. We use the total variation as a measure of “closeness”. The total variation between two distributions is the sum of absolute difference in probabilities.

\[
Total \ Vari\ation = \sum_{i=1}^{n} |p_i - q_i| \quad (9)
\]

FIGURE 7. Skew Logistic Distribution vs. MMED for a Beta (4,6)

FIGURE 8. Comparative example for the “closeness” of a Beta (4,6) to its estimated distributions using different methods.
The total variation also has an interesting interpretation. It can be shown that if the total variation between two distributions, \( P \) and \( Q \), is \( \alpha \), then the maximum difference between any two discrete probabilities, \( p_i \) and \( q_i \), cannot exceed \( \frac{\alpha}{2} \).

From the results of Figure 8, we can see that the MMED outweighs the other approaches on the total variation measure. Its main drawback, however, is that its density function, despite continuous, is piecewise linear. On the other hand, it is very simple to construct and gives good approximations for inferences about a continuous, or discrete, distribution with fractile constraints.

![Total Variation vs. Estimation Method](image.png)

**FIGURE 8.** Comparison of total variation for different estimation methods

### CONCLUSIONS

We have shown an application of the maximum entropy principle to the assignment of univariate distributions and presented a graphical method to determine the maximum entropy distribution between upper and lower probability bounds. We also presented a heuristic approximation to a continuous distribution given fractile constraints. We compared this approximation to several other approaches to illustrate its simplicity and closeness to the original distribution.

### REFERENCES


