Soft Computing (2004)
DOI 10.1007/s00500-004-0367-6

ORIGINAL PAPER

James J. Buckley

Maximum entropy principle with imprecise side-conditions

Published online: 27 April 2004

Abstract In this paper we consider the maximum entropy principle with imprecise side-conditions, where the imprecise side-conditions are modeled as fuzzy sets. Our solution produces fuzzy discrete probability distributions and fuzzy probability density functions.

Keywords Maximum entropy · Fuzzy constraints · Fuzzy probability

1 Introduction

We first discuss the maximum entropy principle, subject to crisp (non-fuzzy) constraints, in the next section. This presentation is based on [1]. Then we show how this principle may be extended to handle fuzzy constraints (fuzzy numbers model the imprecision) in Sect. 3. In Sect. 3 we obtain solutions like a fuzzy discrete probability distribution, the fuzzy normal probability distribution, the fuzzy negative exponential distribution, etc. which are all contained in [2]–[7].

Let us now introduce the notation we will use in the paper. We place a "bar" over a symbol to denote a fuzzy set. So, $\bar{A}$, $\bar{B}$, $\bar{x}$, ... all represent fuzzy sets. If $\bar{A}$ is a fuzzy set, then $\bar{A}(x) \in [0,1]$ is the membership function for $\bar{A}$ evaluated a real number x. An α-cut of $\bar{A}$, written $\bar{A}[\alpha]$, is defined as $\{x|\bar{A}(x) \geq \alpha\}$, for $0 < \alpha \leq 1$. $\bar{A}[0]$ is separately defined as the closure of the union of all the $\bar{A}[\alpha]$, $0 < \alpha \leq 1$. A fuzzy number $\bar{N}$ is a fuzzy subset of the real numbers satisfying: (1) $\bar{N}(x) = 1$ for some x (normalized); and (2) $\bar{N}[\alpha]$ is a closed, bounded, interval for $0 \leq \alpha \leq 1$. A triangular fuzzy number $\bar{T}$ is defined by three numbers $a_1 < a_2 < a_3$ where the graph of $y = \bar{T}(x)$ is a triangle with base on the interval $[a_1, a_3]$ and vertex at $x = a_2$ ($\bar{T}(a_2) = 1$). We write $\bar{T} = (a_1/a_2/a_3)$ for triangular fuzzy numbers. A triangular shaped fuzzy number has curves, not straight line segments, for the sides of the triangle. For any fuzzy number $\bar{N}$ we have $\bar{N}[\alpha] = [n_1(\alpha), n_2(\alpha)]$ for all α, which describes the closed, bounded, intervals as functions of α.

J. J. Buckley
University of Alabama at Birmingham,
Department of Mathematics, Birmingham,
Alabama, 35294, USA
E-mail: buckley@math.uab.edu
Tel.: 205-934-2154
Fax: 205-934-9025

2 Maximum entropy principle

We first consider discrete probability distributions and then continuous probability distributions. The entropy principle have not gone uncriticized, and this literature, together with that justifying the principles, has been surveyed in [1].

2.1 Discrete probability distributions

We start with a discrete, and finite, probability distribution. Let $X = \{x_1, \ldots, x_n\}$ and $p_i = P(x_i)$, $1 \leq i \leq n$, where we use P for probability. We do not know all the p_i values exactly but we do have some prior information, possibly through expert opinion, about the distribution. This information could be in the form of: (1) its mean; (2) its variance; or (3) interval estimates for the p_i. The decision problem is to find the "best" $p = (p_1, \ldots, p_n)$ subject to the constraints given in the information we have about the distribution. A measure of uncertainty in our decision problem is computed by $H(p) = H(p_1, \ldots, p_n)$ where

$$H(p) = -\sum_{i=1}^{n} p_i \ln(p_i) \ , \tag{1}$$

for $p_1 + \cdots + p_n = 1$ and $p_i \geq 0$, $1 \leq i \leq n$. Define $0\ln(0) = 0$. $H(p)$ is called the entropy (uncertainty) in the decision problem.

Let $\mathcal{F}$ denote the set of feasible probability vectors p. $\mathcal{F}$ will contain all the p satisfying the constraints dictated by the prior information about the distribution. The maximum entropy principle states that the "best" p, say p^*, has the maximum entropy subject to $p \in \mathcal{F}$. Therefore p^* solves

$$\max[-\sum_{i=1}^{n} p_i \ln(p_i)] \ , \tag{2}$$

subject to $p \in \mathcal{F}$. With only the constraint that $p_1 + \cdots + p_n = 1$ and $p_i \geq 0$ all i the solution is the uniform distribution $p_i = 1/n$ all i.

It is easy to extend this decision problem to the infinite case of $X = \{x_1, \ldots, x_n, \ldots\}$.

Example 2.1.1

Suppose we have prior information, possibly through expert opinions, about the mean m of the discrete probability distribution. Our decision problem is

$$\max\left[-\sum_{i=1}^{n} p_i \ln(p_i)\right] \ , \tag{3}$$

subject to

$$p_1 + \cdots + p_n = 1, \quad p_i \geq 0, \ 1 \leq i \leq n \ , \tag{4}$$

$$\sum_{i=1}^{n} x_i p_i = m \ . \tag{5}$$

The solution is [1]

$$p_i^* = \exp[\lambda - 1] \exp[\mu x_i] \ , \tag{6}$$

for $1 \leq i \leq n$ and λ and μ are Lagrange multipliers whose values are obtained from the constraints

$$\exp[\lambda - 1] \sum_{i=1}^{n} \exp[\mu x_i] = 1 \ , \tag{7}$$

$$\exp[\lambda - 1] \sum_{i=1}^{n} x_i \exp[\mu x_i] = m \ . \tag{8}$$

An example where the constraints are $p_1 + \cdots + p_n = 1$, $p_i \geq 0$ all i and $a_i \leq p_i \leq b_i$ all i with $a_1 + \cdots + a_n \leq 1 \leq b_1 + \cdots + b_n$ is in [1].

Example 2.1.2

Now assume that $X = \{0, 1, 2, 3, \ldots\}$ so that we have a discrete, but infinite, probability distribution. If we have prior information about the expected outcome m, then the decision problem is

$$\max\left[-\sum_{i=0}^{\infty} p_i \ln(p_i)\right] \ , \tag{9}$$

subject to

$$\sum_{i=0}^{\infty} p_i = 1, \quad p_i \geq 0, \ all \ i \ , \tag{10}$$

$$\sum_{i=0}^{\infty} i p_i = m \ . \tag{11}$$

The solution, using Lagrange multipliers, is [1]

$$p_i^* = \left(\frac{1}{m+1}\right)\left(\frac{m}{m+1}\right)^i, \quad i = 0, 1, 2, 3, \ldots \ . \tag{12}$$

which is the geometric probability distribution.

2.2 Continuous probability distributions

Let E be $(a, b), -\infty < a < b < \infty$, or $(0, \infty)$, or $(-\infty, \infty)$. The probability density function over E will be written as $f(x)$. That is, $f(x) \geq 0$ for $x \in E$ and $f(x)$ is zero outside E. We do not know the probability density function exactly but we do have some prior information, possibly through expert opinion, about the distribution. This information could be in the form of: (1) its mean; or (2) its variance. The decision problem is to find the "best" $f(x)$ subject to the constraints given in the information we have about the distribution. A measure of uncertainty (entropy) in our decision problem is $H(f(x))$ computed by

$$H(f(x)) = -\int_E f(x) \ln[f(x)] \mathrm{d}x \ , \tag{13}$$

for $f(x) \geq 0$ on E and the integral of $f(x)$ over E equals one. Define $0 \ln(0) = 0$. $H(f(x))$ is called the entropy (uncertainty) in the decision problem.

Let $\mathcal{F}$ denote the set of feasible probability density functions. $\mathcal{F}$ will contain all the $f(x)$ satisfying the constraints dictated by the prior information about the distribution. The maximum entropy principle states that the "best" $f(x)$, say $f^*(x)$, has the maximum entropy subject to $f(x) \in \mathcal{F}$. Therefore $f^*(x)$ solves

$$\max[-\int_E f(x) \ln[f(x)] \mathrm{d}x] \ , \tag{14}$$

subject to $f(x) \in \mathcal{F}$. With only the constraint that $\int_E f(x) \mathrm{d}x = 1$, $f(x) \geq 0$ on E and $E = (a, b)$ the solution is the uniform distribution on E.

Example 2.2.1

Suppose we have prior information, possibly through expert opinions, about the mean m and variance σ^2 of the probability density. Our decision problem is

$$\max\left\{-\int_E f(x) \ln[f(x)] \mathrm{d}x\right\} \ , \tag{15}$$

subject to

$$\int_E f(x) \mathrm{d}x = 1, \quad f(x) \geq 0 \ on \ E \ , \tag{16}$$

$$\int_E x f(x) \mathrm{d}x = m \ , \tag{17}$$

$$\int_E (x - m)^2 f(x) \mathrm{d}x = \sigma^2 \ . \tag{18}$$

The solution, using the calculus of variations, is [1]

$$f^*(x) = \exp[\lambda - 1]\exp[\mu x]\exp[\gamma(x-m)^2] \ , \quad (19)$$

where the constants λ, μ, γ are determined from the constraints given in equations (16) through (18).

Example 2.2.2

Let $E = (0, \infty)$ and omit the constraint that the variance must equal the positive number σ^2. That is, in Example 2.2.1 drop the constraint in equation (18). Then the solution is [1]

$$f^*(x) = (1/m)\exp\left[-\frac{x}{m}\right], \quad x \ge 0 \ , \quad (20)$$

the negative exponential.

Example 2.2.3

Now assume that $E = (-\infty, \infty)$ together with all the constraints of Example 2.2.1. The solution is [1] the normal probability density with mean m and variance σ^2.

3 Maximum entropy principle with imprecise side-conditions

We first consider discrete probability distributions and then continuous probability distributions. We will only consider imprecise side-conditions relating to the mean and variance of the unknown probability distribution. These imprecise conditions will be stated as the mean is "approximately" m and the variance is "approximately" σ^2. We will model this imprecision using triangular fuzzy numbers.

How will we obtain these fuzzy numbers? Let us first present a simple method based on expert opinion which is adopted from estimating job times in project scheduling ([8], Chapter 13). Suppose we have a group of N experts all to estimate the mean of some probability distribution and we solicit the following numbers from the ith member: (1) a_i = the "pessimistic" value of m, or the smallest possible value; (2) b_i = the most likely value of m; and (3) c_i = the "optimistic" value of m, or the highest possible value. We average these numbers over all the experts producing m_1, m_2, m_3 (m_1 = average of the a_i, etc.) and then we use the triangular fuzzy number $\bar{m} = (m_1/m_2/m_3)$ for "approximately" m. Similarly we get $\bar{\sigma}^2 = (\sigma_1^2/\sigma_2^2/\sigma_3^2)$ with $\sigma_1^2 > 0$.

A second method of obtaining fuzzy sets for the mean and variance is through getting a random sample $y_1, \ldots, y_m$ and computing its mean $\bar{y}$ (crisp number here, not a fuzzy set) and variance s^2. Let us consider how we now map this data into a triangular shaped fuzzy number $\bar{m}$ for the mean. Further details may be found in [2]–[7]. We propose to find the $(1-\beta)100\%$ confidence interval for m, for all $0.01 \le \beta < 1$. Starting at 0.01 is arbitrary and you could begin at 0.001, or 0.005, etc. Denote these confidence intervals as

$$[m_1(\beta), m_2(\beta)] \ , \quad (21)$$

for $0.01 \le \beta < 1$. Add to this the interval $[\bar{y}, \bar{y}]$ for the 0% confidence interval for m. Then we have $(1-\beta)100\%$ confidence interval for m for $0.01 \le \beta \le 1$.

Now place these confidence intervals, one on top of the other, to produce a triangular shaped fuzzy number $\bar{m}$ whose α-cuts are the confidence intervals. We have

$$\bar{m}[\alpha] = [m_1(\alpha), m_2(\alpha)] \ , \quad (22)$$

for $0.01 \le \alpha \le 1$. All that is needed is to finish the "bottom" of $\bar{m}$ to make it a complete fuzzy number. We will simply drop the graph of $\bar{m}$ straight down to complete its α-cuts so

$$\bar{m}[\alpha] = [m_1(0.01), m_2(0.01)] \ , \quad (23)$$

for $0 \le \alpha < 0.01$. In this way we are using more information in $\bar{m}$ than just a point estimate, or just a single interval estimate. Notice that $\bar{m}[0]$ is the 99% confidence interval for m. In a similar manner we may obtain a triangular shaped fuzzy number for the variance.

We now show how to solve the maximum entropy principle with imprecise side-conditions through a series of examples patterned after the examples in the previous section.

3.1 Discrete probability distributions

Example 3.1.1

This is the same as Example 2.1.1 except Eq. (5) becomes

$$\sum_{i=1}^{n} x_i p_i = \bar{m} \ . \quad (24)$$

We solve by taking α-cuts. So the above equation becomes

$$\sum_{i=1}^{n} x_i p_i = \bar{m}[\alpha] \ , \quad (25)$$

for $\alpha \in [0,1]$. Now we solve the decision problem, Eqs. (3), (4) and (25), for each $m \in \bar{m}[\alpha]$ giving

$$\bar{\Omega}[\alpha] = \{p^* \mid m \in \bar{m}[\alpha]\} \ , \quad (26)$$

for each α. We put these α-cuts together to obtain the fuzzy set $\bar{\Omega}$, a fuzzy subset of $\mathbf{R}^n$.

We can not project the joint fuzzy probability distribution $\bar{\Omega}$ onto the coordinate axes to get the marginal fuzzy probabilities because the α-cuts of $\bar{\Omega}$ are not "rectangles" in $\mathbf{R}^n$. In fact, $\bar{\Omega}$ is a fuzzy subset of the hyperplane $\{p = (p_1, \ldots, p_n) | p_1 + \cdots + p_n = 1\}$.

How can we compute fuzzy probabilities using $\bar{\Omega}$? The basic method is contained in [2]–[7]. Let A be a subset on X. Say $A = \{x_1, x_2, \ldots, x_6\}$. We want $\bar{P}(A)$ the fuzzy probability of A. It is to be determined by its α-cuts

$$\bar{P}(A)[\alpha] = \{p_1 + \cdots + p_6 \mid p \in \bar{\Omega}[\alpha]\} \ , \tag{27}$$

for all α. Now this α-cut will be an interval so let $\bar{P}(A)[\alpha] = [\tau_1(\alpha), \tau_2(\alpha)]$. Then the optimization problems give the end points of this interval

$$\tau_1(\alpha) = \min\{p_1 + \cdots + p_6 \mid p \in \bar{\Omega}[\alpha]\} \ , \tag{28}$$

$$\tau_2(\alpha) = \max\{p_1 + \cdots + p_6 \mid p \in \bar{\Omega}[\alpha]\} \ , \tag{29}$$

all α.

Next we might ask is the mean of $\bar{\Omega}$ equal to $\bar{m}$. We now see if this is true. The fuzzy mean is computed by α-cuts. Let this unknown fuzzy mean be $\bar{M}$. Then

$$\bar{M}[\alpha] = \{ \sum_{i=1}^{n} x_i p_i \mid p \in \bar{\Omega}[\alpha] \} \ , \tag{30}$$

all α. But each $p \in \bar{\Omega}[\alpha]$ corresponds to a $m \in \bar{m}[\alpha]$ so the sum in Eq. (30) equals the m that produced the p we choose in $\bar{\Omega}[\alpha]$. Hence, $\bar{M}[\alpha] = \bar{m}[\alpha]$ for all α and $\bar{M} = \bar{m}$.

Example 3.1.2

This is the same as Example 2.1.2 except Eq. (11) is

$$\sum_{i=0}^{\infty} i p_i = \bar{m} \ . \tag{31}$$

As in the previous example we solve by α-cuts producing $\bar{\Omega}[\alpha]$ and $\bar{\Omega}$.

It is easier to see what we get in this case because the $p \in \bar{\Omega}[\alpha]$ are given by Eq. (12) for all $m \in \bar{m}[\alpha]$. We again may find that the fuzzy mean of $\bar{\Omega}$ is $\bar{m}$.

3.2 Continuous probability distributions

Example 3.2.1

This example continues Example 2.2.1 but now we have fuzzy mean $\bar{m}$ and fuzzy variance $\bar{\sigma}^2$. We solve by α-cuts. That is, we solve the optimization problem in Example 2.2.1 for all m in $\bar{m}[\alpha]$ and for all σ^2 in $\bar{\sigma}^2[\alpha]$. This produces $\bar{\Omega}[\alpha]$ and $\bar{\Omega}$. That is

$$\bar{\Omega}[\alpha] = \{f^*(x) \mid m \in \bar{m}[\alpha], \ \sigma^2 \in \bar{\sigma}^2[\alpha] \} \ . \tag{32}$$

How do we compute fuzzy probabilities with this joint fuzzy distribution? Let G be a subset of E. Then an α-cut of $\bar{P}(G)$ is

$$\bar{P}(G)[\alpha] = \left\{ \int_G f^*(x)\mathrm{d}x \mid m \in \bar{m}[\alpha], \ \sigma^2 \in \bar{\sigma}^2[\alpha] \right\} \ , \tag{33}$$

for all α. $\bar{P}(G)$ is a fuzzy subset of $\mathbf{R}$ and its interval α-cuts are given in the above equation.

We may also find the fuzzy mean and fuzzy variance of $\bar{\Omega}$ and compute $\bar{m}$ and $\bar{\sigma}^2$, respectively. For example, if we denote the fuzzy mean of $\bar{\Omega}$ as $\bar{M}$ its alpha-cuts are

$$\bar{M}[\alpha] = \left\{ \int_E x f^*(x)\mathrm{d}x \mid m \in \bar{m}[\alpha], \ \sigma^2 \in \bar{\sigma}^2[\alpha] \right\} \ , \tag{34}$$

for all α. Now the integral in the above equation equals m for each m in the alpha-cut of $\bar{m}$ and σ^2 in the alpha-cut of $\bar{\sigma}^2$. So $\bar{M}[\alpha] = \bar{m}[\alpha]$ for all α and $\bar{M} = \bar{m}$.

Example 3.2.2

This is the same as Example 2.2.2 but it has a fuzzy mean $\bar{m}$. Solving by α-cuts we obtain the fuzzy negative exponential [2,4,5].

Example 3.2.3

The same as Example 2.2.3 having a fuzzy mean and a fuzzy variance. Solving by α-cuts we get the fuzzy normal [2,4,5] with mean $\bar{m}$ and variance $\bar{\sigma}^2$.

Let $N(c, d)$ denote the normal probability density with mean c and variance d. Then

$$\bar{\Omega}[\alpha] = \{N(m, \sigma^2) \mid m \in \bar{m}[\alpha], \ \sigma^2 \in \bar{\sigma}^2[\alpha] \} \ , \tag{35}$$

for $\alpha \in [0, 1]$. We compute with the fuzzy normal as follows

$$\bar{P}(G)[\alpha] = \left\{ \int_G N(m, \sigma^2)\mathrm{d}x \mid m \in \bar{m}[\alpha], \ \sigma^2 \in \bar{\sigma}^2[\alpha] \right\} \ , \tag{36}$$

for all α giving fuzzy probability $\bar{P}(G)$. We may also find that the fuzzy mean of $\bar{\Omega}$ is $\bar{m}$ and the fuzzy variance of $\bar{\Omega}$ is $\bar{\sigma}^2$.

4 Summary and conclusions

We solved the maximum entropy principle with imprecise side-conditions, which were modeled as fuzzy sets, producing fuzzy probability distributions [2]–[7]. It seems very natural if you start with a fuzzy mean, variance, etc, you need to end up with a fuzzy probability distribution. Fuzzy probability distributions produce fuzzy means, variances, etc.

Acknowledgements The author wished to thank professor L.A. Zadel for suggesting this problem via the "BISC-Group" posting service.

References

1. Buckley JJ (1985) Risk analysis 5:303–313
2. Buckley JJ (2003) Fuzzy probabilities: new approach and applications, Physica-Verlag, Heidelberg
3. Buckley JJ, Eslami E (2003) Uncertain probabilities I: the discrete case. Soft Comput 7:500–505
4. Buckley JJ, Eslami E. Uncertain probabilities II: the continuous case. Soft Computing 8 (2004) 193–199

5. Buckley JJ, Uncertain probabilities III: the continuous case. Soft Computing 8 (2004) 200–206
6. Buckley JJ, Reilly K, Zheng X. Fuzzy probabilities for web planning. Soft Computing (to appear)
7. Buckley JJ, Fuzzy probabilities and fuzzy sets for web planning. Springer, Berlin Heidelberg New York 2004
8. Taha HA (1992) Operations research, 5th edn. Macmillan, New York