Mathematical Forum, Vol. IX, No. 1, March 1992, pp. 23–28

MAXIMUM-ENTROPY PROBABILITY DISTRIBUTIONS WHEN THERE ARE INEQUALITY CONSTRAINTS ON PROBABILITIES : AN ALTERNATIVE APPROACH

J. N. KAPUR

Jawahar Lal Nehru University
and
Mathematical Sciences Trust Society
C-766 New Friends Colony
New Delhi 110065

ABSTRACT

A new approach is given for finding the Maximum-Entropy Probability Distributions when there are inequality constraints on probabilities in addition to the usual moment constraints.

INDEX TERMS

Maximum-Entropy Principle/Inequality Constraints

1. *Introduction* :

The following constraints are prescribed on the probabilities $p_1, p_2, \cdots, p_n$ of a probability distribution $P=(p_1, p_2, \cdots, p_n)$

$$a_i \leqslant p_i \leqslant b_i, \quad i=1, 2, \cdots, n \tag{1}$$

$$\sum_{i=1}^{n} p_i = 1 \tag{2}$$

$$\sum_{i=1}^{n} p_i g_{ri} = a_r, \quad r=1, 2, \cdots, m \tag{3}$$

There may be an infinity of probability distributions satisfying (1), (2) and (3) ; our object is to find that distribution out of these, which has the maximum entropy.

Freund and Saxena [1] solved this problem by maximizing Shannon's [3, 5] entropy measure

$$\sum_{i=1}^{n} p_i \ln p_i \tag{4}$$

subject to constraints (1) and (2) only. Even in this simple case, the algorithm they got was not very simple. The problem of maximizing (4) subject to (1), (2) and (3) has not been solved and its solution is not going to be easy.

We can get suboptimal solutions by maximizing (4) subject to (2) and (3) and then adjusting the probabilities to satisfy (1) or maximizing (4) subject to (1) and (2) and then adjusting the probabilities to satisfy (3). Both the methods may give us solutions, which are far from optimal.

Recently, Kapur [2] used the inverse maximum-entropy principle of Kapur and Kesavan [4] to get the measure

$$-\sum_{i=1}^{n} (p_i-a_i) \ln (p_i-a_i)-\sum_{i=1}^{n} (b_i-p_i) \ln (b_i-p_i) \tag{5}$$

Its maximization subject to (2) and (3) leads to a solution which automatically satisfies (1). Computationally the method is extremely simple and elegant.

However, while maximization of (4) gives us a solution which makes probabilities as equal as possible subject to the constraints, maximization of (5) gives us probabilities which make $P_1, P_2, \cdots, P_n$ as equal as possible, subject to the given constraints where

$$P_i=\frac{p_i-a_i}{b_i-p_i}, \qquad i=1, 2, \cdots, n \tag{6}$$

The two objectives are the same only if

$$a_1=a_2 \cdots =a_n=a, \qquad b_1=b_2 \cdots =b_n=b \tag{7}$$

In other words maximization of (5) will give us probabilities which will divide the intervals (a_i, b_i) in as equal ratios as possible, subject to the given constraints.

In many practical problems, this is a more desirable objective than making all the probabilities equal subject to the given constraints.

However, in these problems where the objective is to make probabilities as equal as possible, subject to the given constraints, in spite of the non-symmetric inequality constraints, we have to maximize (4) subject to (1), (2) and (3).

In the present paper, we suggest an alternative approach to solve the problem by modifying the problem in a different manner.

2. *The New Method* :

We assume that p_i is a random variable which lies in the interval (a_i, b_i) with density function $f_i(x)$ so that

$$\int_{a_i}^{b_i} f_i(x)\, dx=1, \qquad i=1, 2, \cdots, n \tag{8}$$

We define $\hat{p}_i$ as the expected value of this random variate so that

$$\hat{p}_i=\int_{a_i}^{b_i} x\, f_i(x)\, dx \tag{9}$$

Now the constraints (2) and (3) are replaced by

$$\sum_{i=1}^{n} \int_{a_i}^{b_i} x\, f_i(x)\, dx=1 \tag{10}$$

$$\sum_{i=1}^{n} g_{ri} \int_{a_i}^{b_i} x\, f_i(x)\, dx=\alpha_i, \quad r=1, 2, \cdots, m \tag{11}$$

We now seek to maximize

$$-\sum_{i=1}^{n} \int_{a_i}^{b_i} f_i(x) \ln f_i(x)\, dx \tag{12}$$

subject to (10) and (11) to get $f_1(x), f_2(x), \cdots, f_n(x)$ and then use (9) to get $\hat{p}_1, \hat{p}_2, \cdots, \hat{p}_n$

In view of (8) and (9), we shall get automatically

$$a_i \leqslant \hat{p}_i \leqslant b_i \tag{13}$$

Even in this method, we are strictly not attaching the original problem. We are using expected values of probabilities instead of probabilities themselves, and instead of taking entropy of expected values, viz.

$$-\sum_{i=1}^{n} \hat{p}_i \ln \hat{p}_i \tag{14}$$

we are using the entropy measure (12).

The method aims to make the density functions as uniform as possible, subject to the given constraints.

3. *The Solution* :

Maximizing (12) subject to (8), (10) and (11) by using Lagrange's method we get

$$-1-\ln f_i(x)-(u-1)-\lambda_i x-\sum_{r=1}^{m} \lambda_m g_{ri}\, x=0$$

or

$$f_i(x)=C_i \exp\left[-x(\lambda_0+\lambda_1 g_{1i}+\lambda_2 g_{2i}+\cdots+\lambda_m g_{mi}\right] \qquad i=1, 2, \cdots, n \tag{15}$$

where the $n+m+1$ Lagrange multipliers are determined by using $n+m+1$ equations (8), (10) and (11) ; so that

$$C_i \int_{a_i}^{b_i} \exp(-A_i x)\, dx=1, \qquad i=1, 2, \cdots, n \tag{16}$$

$$\sum_{r=1}^{m} C_i \int_{a_i}^{b_i} \exp(-A_i x)\, dx=1 \tag{17}$$

$$\sum_{r=1}^{m} C_i g_{ri} \int_{a_i}^{b_i} x \exp(-A_i x)\, dx=a_r, \quad r=1, 2, \cdots, m \tag{18}$$

where $$A_i=\lambda_0+\lambda_1 g_{1i}+\lambda_2 g_{2i}+\cdots+\lambda_m g_{mi} \tag{19}$$

so that $$C_i\left[\frac{e^{-A_i a_i}-e^{-A_i b_i}}{A_i}\right]=1, \quad i=1, 2, \cdots, n \tag{20}$$

$$\sum_{i=1}^{n} C_i\left[\frac{a_i\, e^{-A_i a_i}-b_i\, e^{-A_i b_i}}{A_i}-\frac{e^{-A_i b_i}-e^{-A_i a_i}}{A_i^2}\right]=1 \tag{21}$$

$$\sum_{i=1}^{n} C_i g_{ri} \left[\frac{a_i e^{-A_i a_i} - b_i e^{-A_i b_i}}{A_i} - \frac{e^{-A_i b_i} - e^{-A_i a_i}}{A_i^2}\right] = a_i$$

$$i=1, 2, \cdots, m \qquad (22)$$

Substituting for C_i in (21); (22), we get

$$\sum_{i=1}^{n} \left[\frac{a_i e^{-A_i a_i} - b_i e^{-A_i b_i}}{e^{-A_i a_i} - e^{-A_i b_i}} + \frac{1}{A_i}\right] = 1 \qquad (23)$$

$$\sum_{i=1}^{n} g_{ri} \left[\frac{a_i e^{-A_i a_i} - b_i e^{-A_i b_i}}{e^{-A_i a_i} - e^{-A_i b_i}} + \frac{1}{A_i}\right] = a_r, \; r=1, 2, \cdots, m \qquad (24)$$

From the $(m+1)$ equations (23) and (24), we can find $\lambda_0, \lambda_1, \cdots, \lambda_m$ and then from (20) we can find $C_1, C_2, \cdots, C_n$ and then we can find $f_1(x), f_2(x), \cdots, f_n(x)$ and $\hat{p}_1, \hat{p}_2, \cdots, \hat{p}_n$

4. *A Special Case* :

We consider the special case considered by Freund and Saxena [1] i.e., we consider only the natural constraints. In this case $A_i = \lambda_0$ and this is determined by using (23)

$$\left[\frac{a_i e^{-\lambda_0 a_i} - b_i e^{-\lambda_0 b_i}}{e^{-\lambda_0 a_i} - e^{-\lambda_0 b_i}} + \frac{1}{\lambda_0}\right] = 1 \qquad (25)$$

and then C_i's are determined by using (20), so that

$$C_i = \frac{\lambda_0}{e^{-\lambda_0 a_i} - e^{-\lambda_0 b_i}}, \quad i=1, 2, \ldots, n \qquad (26)$$

and

$$\hat{p}_i = \frac{a_i e^{-\lambda_0 a_i} - b_i e^{-\lambda_0 b_i}}{e^{-\lambda_0 a_i} - e^{-\lambda_0 b_i}} + \frac{1}{\lambda_0}, \quad i=1, 2, \ldots, n \qquad (27)$$

If

$$a_1 = a_2 \cdots = a_n = a, \; b_1 = b_2 \cdots = b_n = b \qquad (28)$$

then $\hat{p}_i$ is independent of i and since

$$\sum_{i=1}^{n} \hat{p}_i = 1 \text{ each } \hat{p}_i = \frac{1}{n}$$

so that the distribution in the uniform distribution if $a < \frac{1}{n} < b$. This is expected.

A Necessary Condition for a Solution :

Since

$$a_i \leqslant p_i \leqslant b_i \tag{29}$$

$$\sum_{i=1}^{n} a_i \leqslant \sum_{i=1}^{n} p_i \leqslant \sum_{i=1}^{n} b_i \tag{30}$$

so that $\sum_{i=1}^{n} a_i$ should be less than or equal to unity and $\sum_{i=1}^{n} b_i$ should be greater than or equal to unity.

References

1. D. Freund and Saxena (1984), 'An algorithm for a class of entropy problem', Ops. Res. *32*, 210–215.

2. J. N. Kapur (1990), 'Maximum entropy probability distributions when there are inequality constraints on moments and equality constraints on means', Aligarh Journal of Statistics, vol. 9, 28–35.

3. J. N. Kapur (1990), *Maximum-Entropy Models in Science and Engineering*, John Wiley, New York.

4. J. N. Kapur and H. K. Kesavan (1987), *Generalised Maximum-Entropy Principle* (with Applications), Sandford Educational Press, Waterloo.

5. C. E. Shannon (1940), 'A mathematical theory of communication', Bell Sys. Tech. Journal, *27*, 379–423, 623–659.