On the calculations of maximum entropy distributions having prescribed the moments

1985 J. Phys. A: Math. Gen. 18 L337

(http://iopscience.iop.org/0305-4470/18/7/001)

J. Phys. A: Math. Gen. 18 (1985) L337–L339. Printed in Great Britain

LETTER TO THE EDITOR

On the calculations of maximum entropy distributions having prescribed the moments

A Kociszewski

Institute of Mathematics and Physics, Academy of Technology and Agriculture, Bydgoszcz, Poland

Received 12 February 1985

Abstract. It is shown that maximum entropy distributions having prescribed the finite number of moments can be determined from a suitable system of differential equations.

In many branches of physics there arises the problem of determining the probability distribution from the limited number of known statistical moments. According to the maximum entropy (information) principle (see, for example, Jaynes 1957, Ingarden and Urbanik 1962), the probability distribution, having prescribed the n moments (of one variable in the range (a, b))

$$M_k = \int_a^b \mathrm{d}x\, x^k p(x), \qquad k = 1, \ldots, n, \tag{1}$$

may be found by maximisation of entropy

$$S = -\int_a^b \mathrm{d}x\, p(x) \ln p(x) \tag{2}$$

under constraint (1). A formal derivation of entropy-maximising distributions may be carried out by means of the Lagrange multipliers method, which yields

$$p(x) = \exp\left(-\sum_{k=0}^{n} \lambda_k x^k\right). \tag{3}$$

where the Lagrange multipliers $\lambda_0, \lambda_1, \ldots, \lambda_n$ are chosen to satisfy tthe moments condition

$$M_k = \int_a^b \mathrm{d}x\, x^k \exp\left(-\sum_{l=0}^{n} \lambda_l x^l\right) \qquad k = 0, 1, \ldots, n. \tag{4}$$

The normalisation of the probability distribution implies $M_0 = 1$ which gives

$$\lambda_0 = \ln \int_a^b \mathrm{d}x\, \exp\left(-\sum_{l=1}^{n} \lambda_l x^l\right). \tag{5}$$

The above results and the difficulties in calculating the Lagrange multipliers $\boldsymbol{\lambda} = (\lambda_0, \lambda_1, \ldots, \lambda_n)$ from the system of equations (4) (when the values of moments $\boldsymbol{M} = (M_0, M_1, \ldots, M_n)$ are given) have been known for a long time. The analytical results are known only in a few cases (see, for example Collins and Wragg 1977). In the

0305-4470/85/070337+03$02.25

following we will show that it is possible to determine the Lagrange multipliers from a certain system of differential equations.

Suppose that for the values $\boldsymbol{\lambda}^0 = (\lambda_1^0, \ldots, \lambda_n^0)$ the integrations in (4) and (5) are carried out, which give the moments $\boldsymbol{M}^0 = (1, M_1^0, \ldots, M_n^0)$. Suppose also that we want to determine the values $\boldsymbol{\lambda}^1 = (\lambda_0^1, \lambda_1^1, \ldots, \lambda_n^1)$ being the solutions of equations (4) for some given moments $\boldsymbol{M}^1 = (1, M_1^1, \ldots, M_n^1)$. Using the real parameter α we may consider a path $\boldsymbol{M}(\alpha) = (1, M_1(\alpha), \ldots, M_n(\alpha))$ in the space of moments connecting $\boldsymbol{M}^0$ and $\boldsymbol{M}^1$, such that $\boldsymbol{M}(\alpha = 0) = \boldsymbol{M}^0$ and $\boldsymbol{M}(\alpha = 1) = \boldsymbol{M}^1$. For example we may put $\boldsymbol{M}(\alpha) = \alpha \boldsymbol{M}^0 + (1-\alpha)\boldsymbol{M}^1$ but the other parametrisations are also possible. In the next step we differentiate (4) with respect to α which gives

$$\frac{\mathrm{d}M_k}{\mathrm{d}\alpha} = \sum_{l=0}^{n} \frac{\partial M_k}{\partial \lambda_l} \frac{\mathrm{d}\lambda_l}{\mathrm{d}\alpha} = -\sum_{l=0}^{n} M_{k+l} \frac{\mathrm{d}\lambda_l}{\mathrm{d}\alpha}, \tag{6}$$

for $k = 0, 1, \ldots, n$. Equation (6) may be rewritten in the matrix form

$$\mathbf{V}(\alpha) = \mathbf{D} \frac{\mathrm{d}\boldsymbol{\lambda}}{\mathrm{d}\alpha}, \tag{7}$$

where $\mathbf{V}(\alpha) = (0, \mathrm{d}M_1(\alpha)/\mathrm{d}\alpha, \ldots, \mathrm{d}M_n(\alpha)/\mathrm{d}\alpha)$ is the function of parameter α and the matrix $\mathbf{D}$ has the following elements

$$D_{k,l} = -M_{k+l} = -\int_a^b \mathrm{d}x\, x^{k+l} \exp\left(-\sum_{t=0}^{n} \lambda_t x^t\right) \tag{8}$$

for k, $l = 0, 1, \ldots, n$. Therefore the matrix $\mathbf{D}$ contains the moments $\boldsymbol{M}(\alpha)$ and the higher moments $M_{n+1}, \ldots, M_{2n}$. Integrating (4) by parts gives

$$M_k = (k+1)^{-1}\left(\sum_{l=1}^{n} l\lambda_l M_{l+k} + W_k(b, \boldsymbol{\lambda}) - W_k(a, \boldsymbol{\lambda})\right), \tag{9}$$

for $k = 0, 1, \ldots, n$, where

$$W_k(x, \boldsymbol{\lambda}) = x^{k+1} \exp\left(-\sum_{l=0}^{n} \lambda_l x^l\right). \tag{10}$$

From (9) we obtain the following recurrence relations for higher moments

$$M_{n+k} = (n\lambda_n)^{-1}\left((k+1)M_k - \sum_{l=1}^{n-1} l\lambda_l M_{l+k} + W_k(a, \boldsymbol{\lambda}) - W_k(b, \boldsymbol{\lambda})\right). \tag{11}$$

As a result, the matrix $\mathbf{D}$ is a simple function of Lagrange multipliers $\boldsymbol{\lambda}$ and the parameter α (by virtue of dependence $M(\alpha)$), which we write $\mathbf{D} = \mathbf{D}(\boldsymbol{\lambda}, \alpha)$. In the final step we consider the expression

$$\sum_{k,l=0}^{n} \Delta_k \Delta_l D_{k,l} = -\int_a^b \mathrm{d}x \left(\sum_{k=0}^{n} \Delta_k x^k\right)^2 \exp\left(-\sum_{l=0}^{n} \lambda_l x^l\right) \leqslant 0, \tag{12}$$

where the equality sign is valid only if $\Delta_k = 0$ for all $k = 0, 1, \ldots, n$. From (12) it follows that the matrix $\mathbf{D}$ is always strictly negative definite and therefore $\mathbf{D}$ has the inverse matrix $\mathbf{D}^{-1} = \mathbf{D}^{-1}(\boldsymbol{\lambda}, \alpha)$. As the result, (7) takes the form

$$\frac{\mathrm{d}\boldsymbol{\lambda}}{\mathrm{d}\alpha} = \mathbf{D}^{-1}(\boldsymbol{\lambda}, \alpha)\mathbf{V}(\alpha). \tag{13}$$

Equations (13) are the system of non-autonomous ordinary differential equations but generally can be solved only numerically. It is very important in practical applications that matrix $\mathbf{D}^{-1}(\boldsymbol{\lambda}, \alpha)$ does not contain any integrands. The point $\boldsymbol{\lambda}(\alpha = 1)$ of the trajectory $\boldsymbol{\lambda}(\alpha)$ being the solution of (13) (with the initial condition $\boldsymbol{\lambda}(\alpha = 0) = \boldsymbol{\lambda}^0$) is the solution of (4) for $\boldsymbol{M} = \boldsymbol{M}^1$, i.e., $\boldsymbol{\lambda}(\alpha = 1) = \boldsymbol{\lambda}^1$. Using (13) and the standard procedures for solving the system of ordinary differential equations, it is therefore possible to determine the maximum entropy distributions having prescribed the finite number of moments.

It is possible to generalise this approach to probability distributions with more than one variable. More detailed considerations and some applications will be given in a forthcoming paper.

References

Collins R and Wragg A 1977 *J. Phys. A: Math. Gen.* **10** 1441-63
Ingarden R S and Urbanik K 1962 *Acta Phys. Pol.* **21** 281
Jaynes E T 1957 *Phys. Rev.* **106** 620-30