The existence conditions for maximum entropy distributions, having prescribed the first three moments

1986 J. Phys. A: Math. Gen. 19 L823

(http://iopscience.iop.org/0305-4470/19/14/001)

J. Phys. A: Math. Gen. 19 (1986) L823-L827. Printed in Great Britain

LETTER TO THE EDITOR

The existence conditions for maximum entropy distributions, having prescribed the first three moments

Andrzej Kociszewski

Institute of Mathematics, Higher Pedagogical School, Chodkiewicza 30, 85-064 Bydgoszcz, Poland

Received 22 May 1986

Abstract. The criteria for existence of maximum entropy distributions, having prescribed the first three moments (in a semi-infinite range), are given. Some relationships between moments and Lagrange multipliers are also presented.

The maximum entropy formalism (Jaynes 1957, Ingarden and Urbanik 1962) is applied in physics for estimating an unknown probability distribution given only partial data (for a review see Jaynes 1978). Maximum entropy formalism is also applied in line-shape problems (e.g. Powles and Carraza 1970, Meinander and Tabisz 1984) where usually some moments are partial data. Maximising entropy, given the moments $\boldsymbol{M} = \{M_1, \ldots, M_n\}$ as the constraints, one obtains a probability distribution uniquely determined by M. For one variable x in a semi-infinite range, the maximising entropy probability density function (PDF) is in the form

$$P(x, \boldsymbol{\lambda}) = z_0^{-1}(\boldsymbol{\lambda}) \exp\left(-\sum_{i=1}^{n} \lambda_i x^i\right). \tag{1}$$

Obviously, the PDF (1) exists only if the system of the following non-linear equations

$$M_k = m_k(\boldsymbol{\lambda}) = z_k(\boldsymbol{\lambda})/z_0(\boldsymbol{\lambda}) \qquad k = 1, \ldots, n \tag{2}$$

where

$$z_k(\boldsymbol{\lambda}) = \int_0^\infty \mathrm{d}x\, x^k \exp\left(-\sum_{i=1}^{n} \lambda_i x^i\right) \tag{3}$$

has a resolution with respect to the Lagrange multipliers $\boldsymbol{\lambda} = \{\lambda_1, \ldots, \lambda_n\}$. Equations (2) can be solved numerically (except for the case $n = 1$) but it requires a lot of numerical calculations which strongly increase with n. It is therefore very important to know beforehand whether, given M, equations (2) have a resolution at all with respect to $\boldsymbol{\lambda}$. The well known Liapounov inequality for absolute moments requires that the following system of inequalities be satisfied:

$$M_1 > 0 \qquad \text{and} \qquad U_k^{k-1} > U_{k-1}^k \qquad \text{for } k = 2, \ldots, n \tag{4}$$

where U_k are the so-called relative moments defined as $U_k = M_k/M_1^k$. From (4) for $n = 1$ we have $M_1 > 0$ only, which is the necessary and sufficient condition for the existence of (1) (being the exponential PDF) in this case. For $n = 2$ we obtain $M_1 > 0$ and $U_2 > 1$ from (4), but it was found (see, e.g., Dowson and Wragg 1973) that if

$U_2 > 2$, equations (2) have no solutions with respect to $\boldsymbol{\lambda}$ and the corresponding PDF (1) do not exist. For $n \geq 3$ the conditions of solvability of (2) are so far unknown except for the trivial necessary conditions (4), which must be satisfied for any probability distribution in a semi-infinite range. The necessary and sufficient conditions of solvability of (2) for $n = 3$ are the main results of this letter and may be described as follows. The maximum entropy distributions, having prescribed the first three moments M_1, M_2, M_3, exist if and only if $M_1 > 0$, $U_2 > 1$, $U_3 > U_2^2$ and for $1 < U_2 < 2$ also $U_3 < f(U_2)$ are satisfied, where the function $f(U_2)$ can be determined numerically ($f(U_2)$ is presented for some values of U_2 in table 1). We have an important conclusion: the conditions of existence of PDF (1) for $n > 2$ require, in part, some numerical calculation. The conditions of solvability of (2) described above for $n = 3$ are represented by the region (denoted by I) between the two shaded areas in figure 1. In turn, the shaded areas in figure 1 represent those relative moments U_2 and U_3 which fulfil the Liapounov inequalities (4) and do not fulfil the conditions of solvability of (2) in the case under consideration. Now we describe the method of obtaining the results presented above. The functions $m_k(\boldsymbol{\lambda})$ have the following property:

$$m_k(a\lambda_1, a^2\lambda_2, a^3\lambda_3) = a^k m_k(\lambda_1, \lambda_2, \lambda_3) \qquad k = 1, 2, 3. \tag{5}$$

From (2) and (5) for the functions $u_k(\boldsymbol{\lambda}) = m_k(\boldsymbol{\lambda})/m_1^k(\boldsymbol{\lambda})$ we have

$$u_k(\lambda_1, \lambda_2, \lambda_3) = u_k(a\lambda_1, a^2\lambda_2, a^3\lambda_3) \qquad k = 2, 3. \tag{6}$$

Table 1. Dependence of $f(U_2)$ on U_2 corresponding to curve $C(\infty)$ in figure 1.

U_2	1.1	1.2	1.3	1.4	1.5	1.6	1.7	1.8	1.9
$f(U_2)$	1.30	1.62	1.97	2.36	2.80	3.29	3.84	4.45	5.17

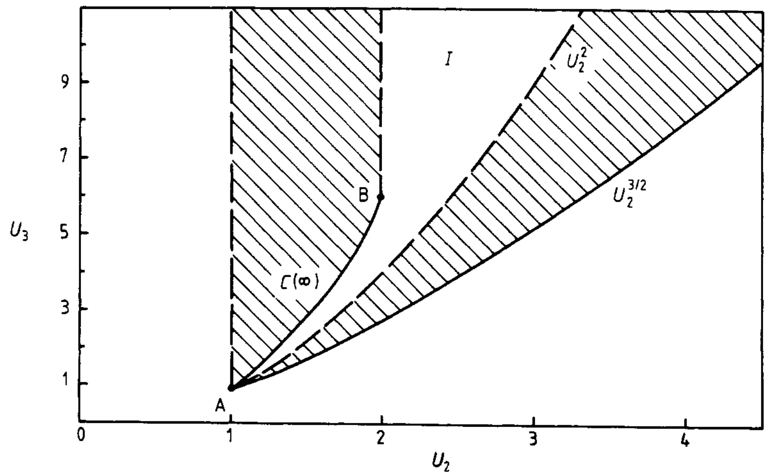

Figure 1. The existence conditions for maximum entropy distributions, having prescribed the first three moments, are represented in terms of relative moments U_2 and U_3 by the region I. Region I together with the two shaded areas represent the Liapounov inequalities for relative moments.

According to (6) for $a=\lambda_3^{-1/3}$ we obtain

$$U_k=u_k(\alpha,\beta,1) \qquad k=2,3 \tag{7}$$

where $\alpha=\lambda_1/\lambda_3^{1/3}$ and $\beta=\lambda_2/\lambda_3^{2/3}$. It means that the relative moments U_2, U_3 are functions of two parameters α and β. For $k=1$ in (5) we obtain $M_1=\lambda_3^{-1/3}m_1(\alpha,\beta,1)$, which means that by changing λ_3, the M_1 may vary from zero to infinity independently of the values of α and β. Denoting by I the image of the function $\boldsymbol{u}(\alpha,\beta,1)=\{u_2(\alpha,\beta,1),u_3(\alpha,\beta,1)\}$ defined by (2) and (7), the necessary and sufficient conditions of existence of PDF (1) for $n=3$ may be described as $M_1>0$ and $\{U_2,U_3\}\in I$. Now we determine the image I of $\boldsymbol{u}$. The function $\boldsymbol{u}(\alpha,\beta,1)$ has the following properties:

$$\frac{\partial(u_2,u_3)}{\partial(\alpha,\beta)}=-3[m_1(\alpha,\beta,1)]^{-6}\frac{\partial(m_1,m_2,m_3)}{\partial(\lambda_1,\lambda_2,\lambda_3)}>0 \tag{8}$$

and

$$\lim_{\alpha\to+\infty}u_k(\alpha,\beta,1)=k! \qquad k=2,3 \tag{9}$$

$$\lim_{\alpha\to-\infty}u_k(\alpha,\beta,1)=1 \qquad k=2,3 \tag{10}$$

where in (9) and (10) β is kept constant. The property (8) may be obtained from the following chain rule for the Jacobians:

$$\frac{\partial(m_1,m_2,m_3)}{\partial(\lambda_1,\lambda_2,\lambda_3)}=\frac{\partial(m_1,m_2,m_3)}{\partial(m_1,u_2,u_3)}\frac{\partial(m_1,u_2,u_3)}{\partial(\alpha,\beta,\lambda_3)}\frac{\partial(\alpha,\beta,\lambda_3)}{\partial(\lambda_1,\lambda_2,\lambda_3)} \tag{11}$$

taking into account that the matrix $(\partial m_k/\partial\lambda_l)$ is strictly negative definite (see, e.g., Kociszewski 1985). Let us denote by $C(\beta)$ the curve on the U_2U_3 plane, which is obtained for constant β and varying α in $\boldsymbol{u}(\alpha,\beta,1)$ from $-\infty$ to $+\infty$. According to (9) and (10), for any β the curve $C(\beta)$ connects two points (1, 1) and (2, 6) (denoted by A and B) on the U_2U_3 plane. In turn, from (8) it follows that the curves $C(\beta_1)$ and $C(\beta_2)$ never intersect for $\beta_1\neq\beta_2$. As a result, any two curves $C(\beta_1)$ and $C(\beta_2)$ enclose some bounded region $I(\beta_1,\beta_2)$ on the U_2U_3 plane and if $\beta_1<\beta<\beta_2$ the curve $C(\beta)$ belongs to the interior of $I(\beta_1,\beta_2)$. This means that image I may be determined by finding $I(\beta_1,\beta_2)$ for $\beta_1\to+\infty$ and $\beta_2\to-\infty$. The calculation of $I(\beta_1,\beta_2)$ for β_1 is relatively simple, because there is the bounded and continuous curve $C(\infty)=\lim_{\beta\to\infty}C(\beta)$, which may be determined as follows. Putting $a=\lambda_2^{1/2}/\lambda_3^{2/3}$ into (6) one obtains

$$U_k=u_k(\alpha,\beta,1)=u_k(\alpha/\beta^{1/2},1,\beta^{-3/2}) \qquad k=2,3. \tag{12}$$

Carrying out $\beta\to\infty$ and keeping $\alpha/\beta^{1/2}=\text{constant}$, we obtain that $C(\infty)$ is given parametrically by $\boldsymbol{u}(\gamma,1,0)$ where γ varies from $-\infty$ to $+\infty$. The curve $C(\infty)$ may be easily determined numerically using the method presented in Kociszewski (1985) and, corresponding to $C(\infty)$, the function $U_3=f(U_2)$ where $1\leqslant U_2\leqslant 2$ is tabulated in table 1. Let us note that $C(\infty)$ is determined by two moments of the PDF (1), i.e. by a cut-off Gaussian PDF. $I(\infty,\beta)$ is the region between $C(\infty)$ and $C(\beta)$ and now we will determine $I(\infty,\beta)$ for $\beta\to-\infty$. To this end, it is convenient to put $a=-\lambda_2/\lambda_3$ in (6) $(\lambda_2<0)$ which gives $\boldsymbol{u}(\alpha,\beta,1)=\boldsymbol{u}[\delta(\xi+\frac{1}{4}),-\delta,\delta]$ where $\delta=(-\beta)^3$ and $\xi=\alpha/\beta^2-\frac{1}{4}$. If $\xi>\frac{1}{12}$, the PDF has one maximum at zero. For $\xi<-\frac{1}{4}$ the PDF also has one maximum but at the point $x^{\max}=\frac{1}{3}[1+\frac{1}{2}(1-12\xi)^{1/2}]$. If $-\frac{1}{4}<\xi<\frac{1}{12}$ the PDF has two maxima, one at zero and the other at $x^{\max}$, which are separated by a minimum at the point $x^{\min}=\frac{1}{3}[1-\frac{1}{2}(1-12\xi)^{1/2}]$. For $\xi=0$ the heights of maxima are equal. For $\delta\gg 1$ the maxima of the PDF are very sharp and are separated by a deep minimum. For δ tending to infinity the maxima may coexist only if $\delta\xi$ is kept constant. This gives the asymptotic expression for $z_k(\alpha, \beta, 1)$ in the form

$$\int_0^\infty \mathrm{d}x\, x^k \exp[-\delta(\xi+\tfrac{1}{4}-x+x^2)x] = \frac{k!}{(\delta/4)^{k+1}} + \exp(-\tfrac{1}{2}\delta\xi)(2)^{-k}\left(\frac{2\pi}{\delta}\right)^{1/2}. \tag{13}$$

From (13) for very large δ we obtain the following expression:

$$U_k = u_k(\alpha, \beta, 1) = h_k(\omega, \eta) = \frac{(k!\,\omega+\eta^k)(\omega+1)^{k-1}}{(\omega+\eta)^k} \qquad k=2,3 \tag{14}$$

where $\eta = (-\beta)^3/8$ and $\omega = \exp(4\eta\xi)$. The curve $C(\beta)$ for very large but negative β is therefore determined by (14) where η = constant and ω (being the monotonically increasing function of α) varies from zero to infinity. Resolving, by determination of (14), the function $U_2 = h_2(\omega, \eta)$ with respect to ω one obtains for $1 < U_2 < 2$ one positive solution ω_1 in the form

$$\omega_1(\eta, U_2) = \frac{-b-[b^2-4(U_2-1)(U_2-2)\eta^2]^{1/2}}{2(U_2-2)} \tag{15}$$

where $b = 2\eta U_2 - \eta^2 - 2$. Resolving $U_2 = h_2(\omega, \eta)$ with respect to ω for $U_2 > 2$ we obtain two positive solutions, one ω_1 given by (15) and the other ω_2 in the form

$$\omega_2(\eta, U_2) = \frac{-b+[b^2-4(U_2-1)(U_2-2)\eta^2]^{1/2}}{2(U_2-2)}. \tag{16}$$

The $I(\infty, \beta)$ for negative and large β may be described as follows. Given $U_2 > 2$, the following inequalities must be satisfied:

$$h_3(\omega_1(\eta, U_2), \eta) < U_3 < h_3(\omega_2(\eta, U_2), \eta) \tag{17}$$

and given $1 < U_2 < 2$, the following inequalities must be satisfied:

$$h_3(\omega_1(\eta, U_2), \eta) < U_3 < f(U_2) \tag{18}$$

where $f(U_2)$ corresponds to the curve $C(\infty)$. It is easy to calculate that

$$\lim_{\eta\to\infty} h_3(\omega_1(\eta, U_2), \eta) = U_2^2 \qquad \text{for } U_2 > 1 \tag{19}$$

and

$$\lim_{\eta\to\infty} h_3(\omega_2(\eta, U_2), \eta) = \infty \qquad \text{for } U_2 > 2. \tag{20}$$

As a result, from (17)-(20) we obtain that the image I of u is defined by the inequalities $1 < U_2^2 < U_3$ together with $U_3 < f(U_2)$ if $1 < U_2 < 2$. In this way the results presented in figure 1 are proven. In passing we have presented some asymptotic formulae for moments. The method presented in this letter may be applied to other maximum entropy distributions, having prescribed the moments, and the calculations are in progress.

References

Dowson D C and Wragg A 1973 *IEEE Trans. Inf. Theor.* **IT-16** 226–30
Ingarden R S and Urbanik K 1962 *Acta. Phys. Pol.* **21** 281
Jaynes E T 1957 *Phys. Rev.* **106** 620–30
—— 1978 *The Maximum Entropy Formalism* ed R D Levine and M Tribus (Cambridge, MA: MIT Press) pp 15–118
Kociszewski A 1985 *J. Phys. A: Math. Gen.* **18** L337–9
Meinander N and Tabisz G C 1984 *Chem. Phys. Lett.* **110** 388–93
Powles J G and Carraza B 1970 *Magnetic Resonance* ed C K Cogan (New York: Plenum) pp 133–63