Maximum entropy histograms

1977 J. Phys. A: Math. Gen. 10 1441

(http://iopscience.iop.org/0305-4470/10/9/007)

J. Phys. A: Math. Gen., Vol. 10, No. 9, 1977. Printed in Great Britain. © 1977

Maximum entropy histograms

R Collins and A Wragg
Department of Mathematics, University of Salford, Salford M5 4WT, UK

Received 9 August 1976, in final form 28 April 1977

Abstract. The problem is considered of assigning a probability distribution on the basis of a limited number of moments. The Jaynes–Shannon maximum-entropy method is adapted to numerical computation by the construction of a suitable type of histogram. A brief account is given of the numerical methods involved and some numerical examples and results are presented, including an application to ion implantation.

1. Introduction

The estimation of a probability density function $w(x)$ from a limited number of known parameters is a well established problem. In its usual form the known (or assumed) parameters consist of some or all of the moments

$$\mu_m = \langle x^m \rangle = \int_R \mathrm{d}x\, x^m w(x) \qquad m = 1, 2, 3, \ldots \tag{1.1}$$

where it is convenient to use $\langle f(x) \rangle$ to denote the expectation value of $f(x)$, and R denotes the range of variation of x. Even in the theoretical case where moments of all orders exist and are known, the complete set $\{\mu_m\}$ does not necessarily uniquely determine $w(x)$ (see for example Shohat and Tamarkind 1950 for a full discussion). The usual practical situation is one where only a few moments are available, either from experimental measurements or from calculations. The problem is then to assign a $w(x)$ which is consistent with these and is most 'reasonable' in some sense. This situation has occurred in several branches of physics in recent years (cf Porter 1965, Collins 1965, 1967, Powles and Carazza 1970a, b).

One approach to this problem which has become quite widely adopted is to select a $w(x)$ which maximises the entropy H (subject to the given μ_m) where H is defined by

$$H = -\int \mathrm{d}x\, w(x) \ln w(x). \tag{1.2}$$

The entropy concept in this sense was developed by a number of workers, notably Wiener, Shannon and their collaborators (for a full discussion see for instance Shannon and Weaver 1949). The inference argument used here was developed by Jaynes (1957) with particular reference to physics. A comprehensive account of its use in statistical mechanics is given by Katz (1967). An early account of the more general physical significance of these ideas is given by Brillouin (1956).

Despite much theoretical discussion, until recently very little in the way of actual computation was done using these ideas and the only explicit results available were those for μ_1 only, for a one-sided $w(x)$ defined over $0 \leq x < \infty$, and for μ_2 only (where a normal distribution is obtained). These two cases were put together on a systematic basis by Wragg and Dowson (Wragg and Dowson 1970, Dowson and Wragg 1973) who showed that even in this relatively simple case, the solutions of Jaynes–Shannon (JS) type only exist over $0 \leq x < \infty$ if the (hitherto unsuspected) relation

$$\mu_2 \leq 2\mu_1^2 \tag{1.3}$$

is satisfied.

Although a solution of JS type over $0 \leq x < \infty$ does not exist in the strict mathematical sense when $\mu_2 > 2\mu_1^2$, the entropy is bounded and the upper bound of H can be approached in practice by some distributions. These distributions characteristically contain a component which is completely determined for finite x, together with a small weight at infinity accounting for the second moment μ_2. If μ_2 is increased the small weight at infinity changes but the first component remains unchanged, and in this sense it can be considered as an admissible solution of JS type.

Severe computational difficulties were encountered in extending the JS formalism to include higher-order moments. In an attempt to overcome these, we decided to try direct numerical computation of a maximum-entropy distribution in the form of a histogram with a fixed number of blocks. To carry out this procedure, an algorithm was developed, and embodied in a working computer program. The results obtained so far seem to be sufficiently encouraging to present them here and to describe the method used in more detail.

The method described here does not provide a different answer to the moments problem from that implicit in the JS formalism. However, experience so far indicates that it is considerably easier to handle numerically, and also is easier to use in the anomalous case, that is when $\mu_2 > 2\mu_1^2$ over $0 \leq x < \infty$, and a truncated range must be used. Moreover, in this latter case, it is not necessary with the method described here, to specify in advance the truncation range, which is determined automatically during the computation. This represents a major advantage from the practical point of view.

To illustrate the application of the method to a specific problem in physics, we consider the depth distribution of atoms implanted in a solid by ionic bombardment. This is a case where physical theory yields only a limited number of moments of the distribution, and the overall shape has to be inferred in some manner from these. Previous methods of carrying out the fitting of a distribution to the moments have depended to a greater or lesser extent on prior assumptions as to the general form of the distribution. In § 4.6 a particular case is treated by out present methods, and the results are compared with those of previous statistical treatments.

2. Summary of earlier results

2.1. General background

Suppose M moments $\mu_1, \mu_2, \ldots, \mu_M$ are given, as defined by (1.1). Then, direct maximisation of H (as given by (1.2)) subject to these M constraints and the normalisation condition

$$\int \mathrm{d}x\, w(x) = 1 \tag{2.1}$$

can be carried out by the usual method of Lagrange undetermined multipliers (see, e.g. Jaynes 1957). The formal result is

$$w(x)=c\,\exp(-\beta_1 x-\beta_2 x^2-\ldots-\beta_M x^M) \tag{2.2}$$

where

$$c=\left(\int \mathrm{d}x\,\exp(-\beta_1 x-\ldots-\beta_M x^M)\right)^{-1} \tag{2.3}$$

and the β_m are to be chosen to satisfy equations (1.1) for the given μ_m. The precise conditions under which such a set $\{\beta_m\}$ *do* exist satisfying these conditions in the general case do not seem to be known. Only in the simple cases $M=1$ and $M=2$ are the conditions known in any completeness.

2.2. *First moment only given*

If the range R of x is finite ($a\leqslant x\leqslant b$) or semi-infinite ($a\leqslant x<\infty$ or $-\infty<x\leqslant b$) then a solution of the form

$$w(x)=c\,\mathrm{e}^{-\beta x} \tag{2.4}$$

can be found, for which β is positive or negative depending on R, provided only that $\mu_1=\mu$ lies within R, which is an obvious condition for consistency.

For the semi-infinite ranges we get

$$\beta=\begin{cases}1/(\mu-a) & a\leqslant x<\infty\\ 1/(\mu-b) & -\infty<x\leqslant b\end{cases} \tag{2.5}$$

and for the finite range

$$\mu=\frac{(1+\beta b)\,\mathrm{e}^{-\beta b}-(1+\beta a)\,\mathrm{e}^{-\beta a}}{\beta(\mathrm{e}^{-\beta b}-\mathrm{e}^{-\beta a})} \tag{2.6}$$

which must be solved for β numerically.

If we consider the infinite range $-\infty<x<\infty$ then the entropy H is not bounded for a fixed value of μ, and (2.4) no longer holds. A search for larger and larger values of H would lead to distributions having a larger and larger spread.

2.3. *First two moments only*

In this case the maximum-entropy probability density function given by the variational calculation has the form (putting $M=2$):

$$w(x)=c\,\exp(-\beta_1 x-\beta_2 x^2) \tag{2.7}$$

where in this case for $-\infty\leqslant x\leqslant\infty$

$$c=\left(\frac{\beta_2}{\pi}\right)^{1/2}\exp\left(-\frac{\beta_1^2}{4\beta_2}\right). \tag{2.8}$$

The existence or otherwise of real β_1, β_2 satisfying the moment equations depends on the range R involved, but in any case, necessary conditions are

$$a < \mu_1 < b \tag{2.9}$$

and

$$\mu_1^2 < \mu_2 \tag{2.10}$$

where a may be $-\infty$, and b may be $+\infty$.

2.3.1. Unrestricted range $-\infty < x < \infty$. In this case, (2.9) and (2.10) are also sufficient conditions and β_1, β_2 are given explicitly by:

$$\beta_1 = -\mu_1/(\mu_2 - \mu_1^2), \tag{2.11}$$

$$\beta_2 = 1/2(\mu_2 - \mu_1^2). \tag{2.12}$$

2.3.2. Finite range $a \leqslant x \leqslant b$. In this case also, (2.9) and (2.10) are sufficient conditions (Dowson and Wragg 1973) but the calculation of β_1 and β_2 from μ_1 and μ_2 has to be done numerically.

2.3.3. Semi-infinite range $a \leqslant x < \infty$. This case also was investigated by Dowson and Wragg for the particular case $a = 0$. Transformed to the case of non-zero a, the result they proved was that conditions (2.9) and (2.10) are no longer sufficient for the existence of real values of β_1, β_2 but must be supplemented by the further condition:

$$\mu_2 \leqslant 2\mu_1(\mu_1 - a) + a^2. \tag{2.13}$$

If this inequality is reversed, then no absolutely continuous probability distribution exists on $a \leqslant x < \infty$ which has the given values for μ_1 and μ_2 and which maximises the entropy.

For simplicity we restrict attention to the case $a = 0$ and if (2.13) is not satisfied (i.e. if $2\mu_1^2 < \mu_2$) we term this an 'anomalous case'. In fact although no one distribution in $0 \leqslant x < \infty$ maximises the entropy H, the upper bound $1 + \ln \mu_1$ of H for a given μ_1 can be approached arbitrarily closely. To see this consider a distribution of the form

$$w = \begin{cases} c\,\mathrm{e}^{-\beta x} & 0 \leqslant x < X \\ \epsilon & X \leqslant x \leqslant X + \epsilon \\ 0 & X + \epsilon < x \end{cases} \tag{2.14}$$

where X is large and ϵ is small†. This is shown diagrammatically in figure 1. For given μ_1 and μ_2 it can be shown that

$$\mu_1\beta = 1 + \frac{\lambda}{\mu_1}\epsilon + \left(1 + \frac{\lambda^2}{\mu_1^2}\right)\epsilon^2 + \mathrm{O}(\epsilon^3, \mathrm{e}^{-\lambda/\epsilon}) \tag{2.15}$$

$$\mu_1 c = 1 + \frac{\lambda}{\mu_1}\epsilon + \frac{\lambda^2}{\mu_1^2}\epsilon^2 + \mathrm{O}(\epsilon^3, \mathrm{e}^{-\lambda/\epsilon}) \tag{2.16}$$

$$\frac{\epsilon X}{\lambda} = 1 + \frac{2\mu_1}{\lambda}\epsilon - \frac{1}{4\lambda}\left(1 + \frac{2}{\lambda} - \frac{2\mu_1^2}{\lambda^2}\right)\epsilon^2 + \mathrm{O}(\epsilon^3, \mathrm{e}^{-\lambda/\epsilon}) \tag{2.17}$$

$$H = 1 + \ln \mu_1 - \frac{\lambda}{\mu_1}\epsilon + \left[\ln\left(\frac{\mu_1}{\epsilon}\right) + 1 + \frac{\lambda^2}{2\mu_1^2}\right]\epsilon^2 + \mathrm{O}(\epsilon^3, \mathrm{e}^{-\lambda/\epsilon}) \tag{2.18}$$

† The authors are indebted to one of the referees for suggesting this approach.

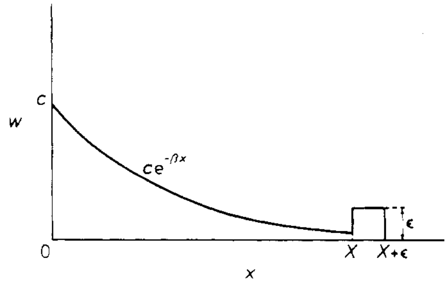

Figure 1. High-entropy distribution for the 'anomalous case' $2\mu_1^2 < \mu_2$, with two moments given.

where

$$\lambda = (\mu_2 - 2\mu_1^2)^{1/2}. \tag{2.19}$$

Clearly, as $\epsilon \to 0$, $X \to \infty$ and the entropy H approaches arbitrarily closely to its upper bound. The maximum-entropy continuous distribution over $0 \leq x \leq R$ (discussed in § 4.5) closely parallels this as $R \to \infty$.

Clearly for $2\mu_1^2 < \mu_2$ the upper bound of H, for any specified value of μ_1, is independent of μ_2. Following thermodynamic practice, we denote by $S(\mu_1, \mu_2)$ the upper bound of H for given μ_1 and μ_2. The graph of S against μ_2 for given μ_1 is shown in figure 2 where we take $\mu_1 = 1$ for simplicity. As $\mu_2 \to \mu_1^2+$, of course, $S \to -\infty$ since $w(x) \to \delta(x - \mu_1)$. The full curve (calculated using the results of Wragg and Dowson 1970) shows the smooth variation of the (attainable) maximum value S of H as a function of μ_2, reaching its largest value of $1 + \ln \mu_1$ as $\mu_2 \to 2\mu_1^2$.

The broken line shows the persistence of this value of S into the anomalous region as an upper bound to H, although unattainable exactly by any given distribution with the given two moments. The curves for $R = 3$ and $R = 4$ are given to show the corresponding variation of S with μ_2 for a finite range R.

The decrease of S as $\mu_2 \to \infty$ for fixed R is quite a striking feature, and corresponds to the increasing concentration of the distribution near $x = 0$ and $x = R$. For $R = \infty$, and

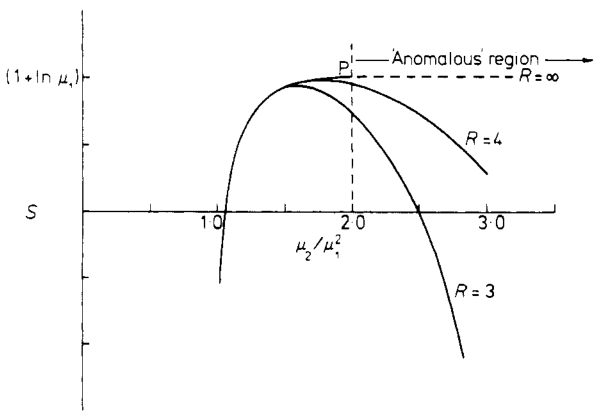

Figure 2. Dependence on μ_2 of the entropy upper bound S for the two-moment case with range R.

fixed μ_1, S is clearly not an analytic function of μ_2, since there is a transition at the point P. If we put

$$\mu_2 = 2\mu_1^2(1-h) \qquad 0<h\ll 1 \tag{2.20}$$

then a straightforward (although algebraically tedious) calculation gives the result

$$S = 1+\ln \mu_1 - \tfrac{1}{2}h^2 + \mathrm{O}(h^3) \tag{2.21}$$

so that S and $\partial S/\partial \mu_2$ are continuous across the transition point P, while $\partial^2 S/\partial \mu_2^2$ has a discontinuity of magnitude $1/(4\mu_1^4)$.

The fact that S becomes a non-analytic function of μ_2 as $R\to\infty$, suggests a comparison with a phase transition of a physical system, where the thermodynamic entropy also takes a maximum, subject to constraints on constants of motion such as energy and particle number, and where the entropy also becomes non-analytic as the volume $V\to\infty$. In such a comparison, the two 'phases' of the system would correspond roughly to the continuous part of $w(x)$ and the small, discrete, 'weight at infinity' as in figure 1.

Consideration of the following two points, however, indicates that the analogy should not be taken too far. Firstly the underlying reason for the discontinuity at P seems to be quite different from that in a physical phase change. For a physical phase transition the discontinuity is produced by taking a large number N of strongly interacting systems with distributions in volume V, and letting $N\to\infty$ as well as $V\to\infty$. No phase discontinuity could result by allowing V to tend to infinity for a fixed number of particles. The discontinuity observed at the point P, however, arises from a single distribution; there is no suggestion of coupling between two or more distributions. The second point arises if, for example, we regard μ_2 as analogous to the internal energy U of a physical system which would imply that $\partial S/\partial \mu_2$ is analogous to $1/T$. In this case we would have $1/T\to 0$ $(T\to\infty)$ as the point P is approached from the left on the $R=\infty$ curve of figure 2. Since S and $\partial S/\partial T$ would also be continuous, no latent heat or change of specific heat is involved, hence the only possible real physical analogy would be a third-order phase change at infinite temperature.

2.4. More than two moments specified

There seems to be no *general* theory covering these cases. Individual cases have been calculated numerically. For example, Wilson and Wragg (1973) give maximum-entropy distributions calculated using $M=2$, 4 and 6 with moments calculated from $w(x)=4x\,\mathrm{e}^{-2x}$ $(0\leq x<\infty)$. Powles and Carazza (1970a) treat the particular case of a symmetric distribution ($\mu_m=0$ for all odd values of m) with $M=2$ and 4. They derive the result that the necessary and sufficient condition for the existence of a maximum-entropy distribution given by

$$w(x) = c\exp(-\beta_2 x^2 - \beta_4 x^4) \qquad -\infty<x<\infty \tag{2.22}$$

is

$$\mu_2^2 \leq \mu_4 \leq 3\mu_2^2. \tag{2.23}$$

The left-hand inequality in (2.23) is an obvious necessary condition. The right-hand inequality expresses the fact that for given μ_2 only, the maximum-entropy distribution is a Gaussian

$$w = \frac{1}{(2\pi\mu_2)^{1/2}} e^{-x^2/2\mu_2} \tag{2.24}$$

which gives $\mu_4 = 3\mu_2^2$. If μ_4 is specified to be greater than $3\mu_2^2$, we have another 'anomalous case', similar to that discussed in § 2.3. In this case the entropy H of a distribution with given μ_2 and μ_4 has an upper bound S given by

$$S = \tfrac{1}{2}\ln(2\pi\mu_2\,\mathrm{e}) \tag{2.25}$$

which is the entropy of the distribution given by (2.24). This upper bound can be approached arbitrarily closely, for example by a distribution approximating to (2.24), together with two small localised perturbations at $\pm X$, where X is large. More precisely, put

$$w = (1-\epsilon^2)\frac{1}{a(2\pi)^{1/2}} e^{-x^2/2a^2} + w' \tag{2.26}$$

where

$$w' = \begin{cases} \epsilon/2X & X < |x| < X(1+\epsilon) \\ 0 & \text{other values of } x. \end{cases} \tag{2.27}$$

Then, for specified moments μ_2 and μ_4 satisfying $3\mu_2^2 < \mu_4$, it can be shown that

$$X = \left(\frac{k}{\epsilon}\right)^{1/2}\left[1 + \left(\frac{3\mu_2}{k} - 1\right)\frac{\epsilon}{2} + \mathrm{O}(\epsilon^2)\right], \tag{2.28}$$

$$a^2 = \mu_2 - k\epsilon + \mathrm{O}(\epsilon^2), \tag{2.29}$$

$$H = S - \frac{k}{2\mu_2}\epsilon + \mathrm{O}(\epsilon^2 \ln \epsilon), \tag{2.30}$$

where

$$k = (\mu_4 - 3\mu_2^2)^{1/2}, \tag{2.31}$$

so that, as $\epsilon \to 0$, the height $\epsilon/2X$ and width ϵX of the perturbations both approach zero, while their displacement X from the origin increases indefinitely. The entropy H meanwhile becomes arbitrarily close to S.

When higher additional moments are specified (Powles and Carazza consider specifically the case of μ_2, μ_4, and μ_6), then the straightforward classification of anomalous cases by simple moment inequalities seems not to hold, and the general theory of these more complex cases remains to be worked out. We consider later some specific numerical cases.

3. The histogram approach

3.1. High-entropy approximations in general

The rigorous results derived so far and expressed in relations (2.13), (2.23) show that the problem of finding a probability density function which gives a *strict* maximum to the entropy for given moments may only be solved under quite severe restrictions for the cases $M=2$ and $M=4$. Further similar restrictions may well come to light for higher values of M. From the point of view of an experimentalist or theoretician who simply wants to assign a distribution to a set of moments which is as unbiased as possible, a distribution which has the largest entropy of a particular class of functions may well have an entropy close to the upper bound, even it if is not a strict maximum for *all* absolutely continuous distributions. In particular, suppose we can find a class of functions $\{w(x)\}_N$ (where N is an integer parameter) such that for large N the $\{w(x)\}_N$ can in some sense approximate to any continuous distribution. One method might be to construct the $\{w(x)\}_N$ from the first N members of a complete set of orthonormal functions over the given interval. (The drawback to this particular procedure is the difficulty of ensuring that the resulting probability density functions are everywhere positive. (See, for example, Wilson and Wragg (1973, p 172) where the $\{w_N(x)\}$ based on Laguerre polynomials are noted to be negative near the origin.)) For a given N, denote the maximum entropy H_N for a fixed set of moments by S_N (by analogy with the thermodynamic case) and denote by S the corresponding constrained upper bound (if it exists) for the entropy H over *all* continuous distributions, again for the set of moments given. Then as $N\to\infty$ we would expect $S_N\to S$ and the corresponding $w_N(x)\to w(x)$. Experience so far indicates that, near the optimum distribution, H is rather insensitive to the precise approximation to $w(x)$ which is used. Hence we may expect to obtain a reasonable approach to S even with a finite value of N. For most of the 'coordinate' functions in common use a large value of N is required to represent a $w(x)$ with anything like a local peak unless the effective range of $w(x)$ is know in advance (which is rarely the case).

This difficulty can be overcome to quite a large extent by choosing a $w_N(x)$ to be a histogram of a particular type described in the next section.

3.2. Common-area histograms in the semi-infinite range

Suppose for simplicity that we are dealing with the semi-infinite range $0\leqslant x<\infty$, and that $w_N(x)$ is a histogram approximation to $w(x)$ with N blocks. If the histogram is of the usual type in which each block is of the same width, then we need to know the effective range R_{eff} of x before we start. R_{eff} could be defined, for example, by

$$\int_0^{R_{\text{eff}}} \mathrm{d}x\, w(x)=1-\epsilon \tag{3.1}$$

where ϵ is some arbitrary small number. Since it is rarely possible to estimate R_{eff} beforehand, this would enormously complicate the calculation.

Instead we define $w_N(x)$ to be histogram with N blocks, with an area $1/N$ common to all blocks. A typical common-area histogram for a one-sided distribution with $N=5$ is shown in figure 3. If a_n denotes the width of the nth block, then

$$w_n=1/Na_n \tag{3.2}$$

and

$$x_n=a_1+a_2+\ldots+a_n. \tag{3.3}$$

The main advantage of histograms of this type is that they are characterised by the set of widths $(a_1, a_2, \ldots, a_N)$ of the blocks, which can be regarded as coordinates, and that the upper limit

$$R=a_1+a_2+\ldots+a_N=x_N \tag{3.4}$$

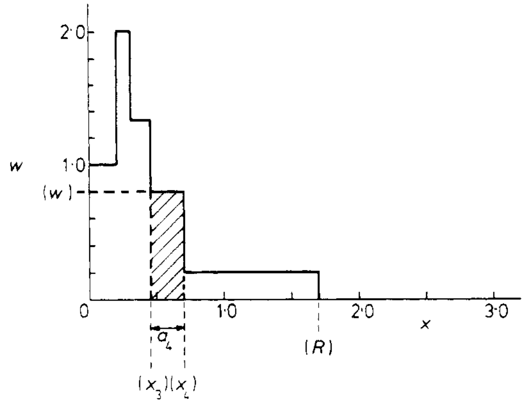

Figure 3. Typical five-block common-area histogram.

to the range need not be specified in advance, which is a great advantage from the algorithmic point of view. The normalisation is automatically built-in to the calculation. A secondary advantage (but useful from the representational point of view) is that the ordinates are more closely bunched where w is largest (where most detail of the slope of a distribution is usually needed). For conciseness we refer to a common-area histogram of this type as a 'c-histogram'.

3.3. *Entropy maximisation of a c-histogram*

The information entropy H_N of an N-block c-histogram over $0 \leq x \leq R$ is given by

$$H_N = -\int_0^R \mathrm{d}x\, w \ln w = -\sum_{n=1}^{N} \int_{x_{n-1}}^{x_n} \mathrm{d}x\, w \ln w. \tag{3.5}$$

Writing w_n for the constant value of w over $x_{n-1} < x < x_n$ we have

$$w = w_n = 1/Na_n \qquad x_{n-1} < x < x_n \tag{3.6}$$

where

$$a_n = x_n - x_{n-1}. \tag{3.7}$$

Substitution from (3.6) and (3.7) in (3.5) leads (after some simplification) to

$$H_N = \ln N + \frac{1}{N} \sum_{n=1}^{N} \ln a_n. \tag{3.8}$$

The problem is then to find the set $\{\boldsymbol{a}\} \equiv (a_1, a_2, \ldots, a_N)$ which give H_N its greatest value S_N subject to the fixed values of the moments μ_m ($m = 1, 2, \ldots, M$) about the origin. We have, using (3.6)

$$\mu_k = \int_0^R \mathrm{d}x\, x^k w(x) = \frac{1}{k+1}\left(w_1 x_1^{k+1} + \sum_{n=2}^{N} w_n (x_n^{k+1} - x_{n-1}^{k+1})\right) \tag{3.9}$$

and using (3.6) and (3.7) this can be written

$$N(k+1)\mu_k = x_1^k + \sum_{n=2}^{N} (x_{n-1}^k + x_{n-1}^{k-1}x_n + \ldots + x_{n-1}x_n^{k-1} + x_n^k) \tag{3.10}$$

$$= \sum_{n=1}^{N} \sigma_{kn} \tag{3.11}$$

where

$$\sigma_{kn} = x_{n-1}^k + x_{n-1}^{k-1}x_n + \ldots + x_{n-1}x_n^{k-1} + x_n^k \tag{3.12}$$

and $x_0 = 0$.

The constrained maximum of H_N is found by introducing Lagrange multipliers β_k, and locating the stationary values of

$$H_N^* = H_N + \sum_{k=1}^{M} \beta_k\left(\mu_k - \frac{1}{N(k+1)} \sum_{n=1}^{N} \sigma_{kn}\right) \tag{3.13}$$

which are given by

$$0 = \partial H_N^*/\partial a_l. \tag{3.14}$$

Using (3.8) and (3.13), this eventually reduces to

$$0 = \frac{1}{a_l} - \sum_{k=1}^{M} \left(\beta_k - \frac{\sigma_{kl}}{(k+1)a_l} + \frac{x_l^k}{a_l} + \sum_{n=l+1}^{N} \sigma_{k-1,n}\right). \tag{3.15}$$

The numerical determination of a set $(x_1, \ldots, x_N; \beta_1, \ldots, \beta_M)$ satisfying equations (3.11) and (3.15) is discussed in § 4.1.

In general we denote by $S_N(\mu_1, \mu_2, \ldots, \mu_M)$ the maximum entropy of an N-block histogram with given moments $\mu_1, \ldots, \mu_M$.

3.4. First moment only

Relations (3.11), (3.12) reduce to (putting $M = 1$)

$$2N\mu_1 = (2N-1)a_1 + (2N-3)a_2 + \ldots + 3a_{N-1} + a_N. \tag{3.16}$$

We have also

$$\sigma_{0l} = 1 \tag{3.17}$$

$$\sigma_{1l} = x_l + x_{l-1} \tag{3.18}$$

so that (3.15) reduces to

$$a_l = \frac{2}{\beta_1(2N-2l+1)}. \tag{3.19}$$

Substitution from (3.19) into (3.16) gives immediately

$$\beta_1 = 1/\mu_1 \tag{3.20}$$

so that

$$a_l = 2\mu_1/(2N-2l+1), \qquad l = 1, 2, \ldots, N. \tag{3.21}$$

This c-histogram for the case $\mu_1 = 1$, $N = 20$ is shown in figure 4, with the curve $w = e^{-x}$ (given by the maximum-entropy Jaynes–Shannon formalism) shown for comparison.

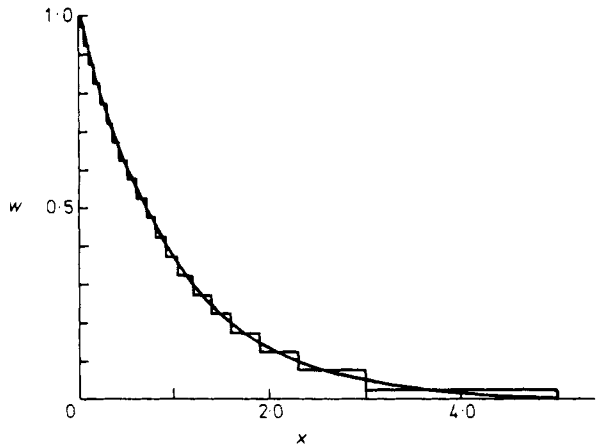

Figure 4. The JS distribution $w = e^{-x}$, and twenty-block c-histogram with $\mu_1 = 1{\cdot}0$.

As will be seen, the agreement is about as good as possible with a histogram of this size. From (3.21) we have

$$w_l = (2N - 2l + 1)/2N\mu_1. \tag{3.22}$$

This is the only case in which the w_l and a_l can be obtained explicitly in a simple form.

The formal proof that the histogram converges to the analytic function is as follows. Let T_r denote the finite harmonic sum

$$T_r = 1 + \frac{1}{2} + \frac{1}{3} + \ldots + \frac{1}{r}. \tag{3.23}$$

Then we have the asymptotic expansion (see for example Stewart 1946, p 485):

$$T_n \sim \gamma + \ln n + \frac{1}{2n} - \frac{B_1}{2n^2} + \frac{B_2}{4n^4} - \frac{B_3}{6n^6} + \ldots \tag{3.24}$$

where γ is Euler's constant and $B_1, B_2, \ldots$ are the Bernoulli numbers. Using (3.19) we have

$$\begin{aligned} x_l &= 2\mu_1\left(\frac{1}{2N-1} + \frac{1}{2N-3} + \ldots + \frac{1}{2N-2l+1}\right) \\ &= 2\mu_1(T_{2N} - \tfrac{1}{2}T_N - T_{2N-2l} + \tfrac{1}{2}T_{N-l}). \end{aligned} \tag{3.25}$$

Substituting from (3.24) for the harmonic sums in (3.25), we get after re-arrangement

$$x_l = \mu_1\left[\ln\left(\frac{N}{N-l}\right) + \mathrm{O}\left(\frac{1}{N^2}, \frac{1}{l^2}\right)\right]. \tag{3.26}$$

Using (3.22) this becomes

$$x_l = -\mu_1\left[\ln\left(\mu_1 w_l - \frac{1}{2N}\right) + \mathrm{O}\left(\frac{1}{N^2}, \frac{1}{l^2}\right)\right]. \tag{3.27}$$

Now keep μ_1 and w fixed, let $l = N(1-\mu w)$ and let $N \to \infty$. Then we can omit the suffix l on x and from (3.27) we get

$$\lim_{N\to\infty} x = -\mu_1 \ln(\mu_1 w) \tag{3.28}$$

in agreement with the result of § 2.2 which in this case ($a = 0$) is

$$w = \frac{1}{\mu_1} e^{-x/\mu_1}. \tag{3.29}$$

4. Numerical results

4.1. Numerical solution of the maximisation equations

It was shown in § 3.3 that the determination of a maximum entropy histogram was equivalent to finding the stationary values of

$$H_N^* = \ln N + \frac{1}{N}[\ln x_1 + \ln(x_2 - x_1) + \ldots + \ln(x_N - x_{N-1})] + \sum_{k=1}^{M} \beta_k\left(\mu_k - \frac{1}{N(k+1)}\sum_{n=1}^{N} \sigma_{kn}\right) \tag{4.1}$$

where

$$\sigma_{kn} = x_{n-1}^k + x_{n-1}^{k-1} x_n + \ldots + x_{n-1} x_n^{k-1} + x_n^k \tag{4.2}$$

and $x_0 = 0$.

We seek therefore a set

$$\boldsymbol{X} = (x_1, x_2, \ldots, x_N; \beta_1, \beta_2, \ldots, \beta_M) \tag{4.3}$$

which satisfies the $N+M$ non-linear algebraic equations

$$F_i(x_1, x_2, \ldots, x_N; \beta_1, \beta_2, \ldots, \beta_M) = 0 \tag{4.4}$$

for $i = 1, 2, \ldots, N+M$, where

$$F_n \equiv \partial H_N^*/\partial x_n \tag{4.5}$$

for $n = 1, 2, \ldots, N$, and

$$F_{N+k} \equiv \mu_k - \frac{1}{N(k+1)} \sum_{n=1}^{N} \sigma_{kn} \tag{4.6}$$

for $k = 1, 2, \ldots, M$.

An iterative scheme for solving (4.4) involves starting from a point

$$\boldsymbol{X}_0 = (x_1^{(0)}, x_2^{(0)}, \ldots, x_N^{(0)}; \beta_1^{(0)}, \beta_2^{(0)}, \ldots, \beta_M^{(0)}) \tag{4.7}$$

and solving the $N+M$ simultaneous linear equations

$$\mathbf{A}(\boldsymbol{X}_s)\boldsymbol{E}_{s+1} = \boldsymbol{R}(\boldsymbol{X}_s) \tag{4.8}$$

for $s = 0, 1, 2, \ldots$, and taking

$$\boldsymbol{X}_{s+1} = \boldsymbol{X}_s + \boldsymbol{E}_s. \tag{4.9}$$

The matrix of coefficients, $\mathbf{A}(\boldsymbol{X}_s)$, and the right-hand side, $\boldsymbol{R}(\boldsymbol{X}_s)$, in (4.8) have the forms

$$\mathbf{A}(\boldsymbol{X}_s)=\left[\frac{\partial F_i}{\partial x_1},\frac{\partial F_i}{\partial x_2},\ldots,\frac{\partial F_i}{\partial x_N};\frac{\partial F_i}{\partial \beta_1},\frac{\partial F_i}{\partial \beta_2},\ldots,\frac{\partial F_i}{\partial \beta_M}\right] \tag{4.10}$$

and

$$\boldsymbol{R}(\boldsymbol{X}_s)=[-F_i] \tag{4.11}$$

for $i=1,2,\ldots,N+M$, with the individual elements evaluated at the point $\boldsymbol{X}_s$.

In view of the underlying wish to use the maximum-entropy approach to fit a sequence of histograms that progressively match an increasing number of specified distribution moments, a numerical algorithm was sought which would cater, for fixed N, with arbitrary values of M. In order to produce such an algorithm it is necessary to express $\mathbf{A}(\boldsymbol{X}_s)$ and $\boldsymbol{R}(\boldsymbol{X}_s)$ in recursive form with respect to their dependence on M. To illustrate this dependence more clearly we introduce the notation $\mathbf{A}^{(M)}$ for $\mathbf{A}(\boldsymbol{X}_s)$ and $\boldsymbol{R}^{(M)}$ for $\boldsymbol{R}(\boldsymbol{X}_s)$, the evaluation of the elements at the point $\boldsymbol{X}_s$ being understood.

We note firstly that $\mathbf{A}^{(2)}$ and $\boldsymbol{R}^{(2)}$ have the forms

$$\mathbf{A}^{(2)}=\left[\begin{array}{c|cc} a^{(2)}_{p,q} & a^{(2)}_{p,N+1} & a^{(2)}_{p,N+2} \\ \hline a^{(2)}_{N+1,q} & 0 & 0 \\ a^{(2)}_{N+2,q} & 0 & 0 \end{array}\right] \tag{4.12}$$

and

$$\boldsymbol{R}^{(2)}=\begin{bmatrix} r^{(2)}_p \\ r^{(2)}_{N+1} \\ r^{(2)}_{N+2} \end{bmatrix}. \tag{4.13}$$

In (4.12), the entries $a^{(2)}_{p,q}$ are the elements of an $N\times N$ symmetric tridiagonal matrix with main diagonal:

$$-\frac{1}{N}\left(\frac{1}{x_1^2}+\frac{1}{(x_2-x_1)^2}+\frac{4\beta_2}{3}\right),$$
$$-\frac{1}{N}\left(\frac{1}{(x_2-x_1)^2}+\frac{1}{(x_3-x_2)^2}+\frac{4\beta_2}{3}\right),\ldots,\ -\frac{1}{N}\left(\frac{1}{(x_N-x_{N-1})^2}+\frac{2\beta_2}{3}\right) \tag{4.14}$$

and super-diagonal

$$\frac{1}{N}\left(\frac{1}{(x_2-x_1)^2}-\frac{\beta_2}{3}\right),\frac{1}{N}\left(\frac{1}{(x_3-x_2)^2}-\frac{\beta_2}{3}\right),\ldots,\frac{1}{N}\left(\frac{1}{(x_N-x_{N-1})^2}-\frac{\beta_2}{3}\right). \tag{4.15}$$

The remaining elements are defined by:

$$a^{(2)}_{N+1,q}=a^{(2)}_{p,N+1}=\begin{cases} -\dfrac{1}{N} & p\mp q=1,2,\ldots,N-1 \\[2ex] -\dfrac{1}{2N} & p\mp q=N \end{cases} \tag{4.16}$$

and

$$a^{(2)}_{N+2,q}=a^{(2)}_{p,N+2}=\begin{cases}-\dfrac{1}{3N}(4x_1+x_2) & p\mp q=1\\ -\dfrac{1}{3N}(x_{p-1}+4x_p+x_{p+1}) & p\mp q=2,3,\ldots,N-1\\ -\dfrac{1}{3N}(x_{N-1}+2x_N) & p\mp q=N.\end{cases} \tag{4.17}$$

In (4.13) the elements of $\boldsymbol{R}^{(2)}$ are given by:

$$r^{(2)}_p=\begin{cases}-\dfrac{1}{N}\left(\dfrac{1}{x_1}-\dfrac{1}{x_2-x_1}\right)-\beta_1 a^{(2)}_{1,N+1}-\beta_2 a^{(2)}_{1,N+2} & p=1\\ -\dfrac{1}{N}\left(\dfrac{1}{x_p-x_{p-1}}-\dfrac{1}{x_{p+1}-x_p}\right)-\beta_1 a^{(2)}_{p,N+1}-\beta_2 a^{(2)}_{p,N+2} & p=2,3,\ldots,N-1\\ -\dfrac{1}{N}\left(\dfrac{1}{x_N-x_{N-1}}\right)-\beta_1 a^{(2)}_{N,N+1}-\beta_2 a^{(2)}_{N,N+2} & p=N\end{cases} \tag{4.18}$$

with

$$r^{(2)}_{N+1}=-F_{N+1} \tag{4.19}$$

and

$$r^{(2)}_{N+2}=-F_{N+2}. \tag{4.20}$$

The matrix $\mathbf{A}^{(m)}$ can now be defined recursively in terms of $\mathbf{A}^{(m-1)}$ by noting that

$$\mathbf{A}^{(m)}=\left[\begin{array}{c|cccc} a^{(m-1)}_{p,q}+b^{(m)}_{p,q} & a^{(m-1)}_{p,N+1} & \ldots & a_{p,N+m-1} & b^{(m)}_{p,N+m}\\ \hline a^{(m-1)}_{N+1,q} & & & & \\ \vdots & & & 0 & \\ a^{(m-1)}_{N+m-1,q} & & & & \\ b^{(m)}_{N+m,q} & & & & \end{array}\right] \tag{4.21}$$

The elements $b^{(m)}_{p,q}$ in (4.21) constitute an $N\times N$ symmetric tridiagonal matrix with main diagonal

$$-\frac{\beta_m}{(m+1)N}[2m(m-1)x_1^{m-1}+(m-1)(m-2)x_1^{m-3}x_2+\ldots+2x_2^{m-2}],$$
$$-\frac{\beta_m}{(m+1)N}\{[2x_1^{m-2}+\ldots+m(m-1)x_2^{m-2}]$$
$$+[m(m-1)x_2^{m-2}+(m-1)(m-2)x_2^{m-3}x_3+\ldots+2x_3^{m-2}]\},\ldots,$$
$$-\frac{\beta_m}{(m+1)N}(2x_{N-1}^{m-2}+\ldots+m(m-1)x_N^{m-2}) \tag{4.22}$$

and super-diagonal

$$-\frac{\beta_m}{(m+1)N}[(m-1)x_1^{m-2}+2(m-2)x_1^{m-3}x_2+\ldots+(m-1)x_2^{m-2}],\ldots,$$
$$-\frac{\beta_m}{(m+1)N}[(m-1)x_{N-1}^{m-2}+2(m-2)x_{N-1}^{m-3}x_N+\ldots+(m-1)x_N^{m-2}]. \tag{4.23}$$

The non-zero elements of the $(N+m)$th row and the $(N+m)$th column of $\mathbf{A}^{(m)}$ are defined by

$$b^{(m)}_{N+m,q}=b^{(m)}_{p,N+m}=\begin{cases} -\dfrac{1}{(m+1)N}\{mx_1^{m-1}+[mx_1^{m-1}+(m-1)x_1^{m-2}x_2+\ldots+x_2^{m-1}]\} \\ \qquad p=q=1 \\ -\dfrac{1}{(m+1)N}\{(x_{p-1}^{m-1}+2x_{p-1}^{m-2}x_p+\ldots+mx_p^{m-1}) \\ +[mx_p^{m-1}+(m-1)x_p^{m-2}x_{p+1}+\ldots+x_{p+1}^{m-1})\} \\ \qquad p=q=2,\ldots,N-1 \\ -\dfrac{1}{(m+1)N}(x_{N-1}^{m-1}+2x_{N-1}^{m-2}x_N+\ldots+mx_N^{m-1}) \\ \qquad p=q=N. \end{cases} \tag{4.24}$$

Similarly, $\boldsymbol{R}^{(m)}$ can be defined recursively in terms of $\boldsymbol{R}^{(m-1)}$. We note that

$$\boldsymbol{R}^{(m)}=\begin{bmatrix} r_p^{(m-1)}+U_p^{(m)} \\ r_{N+1}^{(m-1)} \\ \vdots \\ r_{N+m-1}^{(m-1)} \\ U_{N+m}^{(m)} \end{bmatrix} \tag{4.25}$$

where

$$U_p^{(m)}=-\beta_m b^{(m)}_{p,N+m} \qquad p=1,2,\ldots,N \tag{4.26}$$

and

$$U^{(m)}_{N+m}=-F_{N+m}. \tag{4.27}$$

An ALGOL program incorporating recursive procedures has been written to implement the iterative solution of the maximisation equations. The program can be used in several modes depending on the amount of information available for the problem under study.

4.2. *A particular continuous distribution*

To demonstrate consistency with the results from the analytic JS formulation we consider a case previously discussed (Wragg and Dowson 1970, figure 1) with density function

$$w(x)=4x\,\mathrm{e}^{-2x} \qquad 0\leq x<\infty \tag{4.28}$$

and moments

$$\mu_k=(k+1)!/2^k. \tag{4.29}$$

Here $\mu_2/\mu_1^2=3/2<2$ so the Wragg–Dowson criterion is satisfied. Graphs of the corresponding histograms with $N=20$ and $M=2$ and 3 are given in figures 5 and 6 with

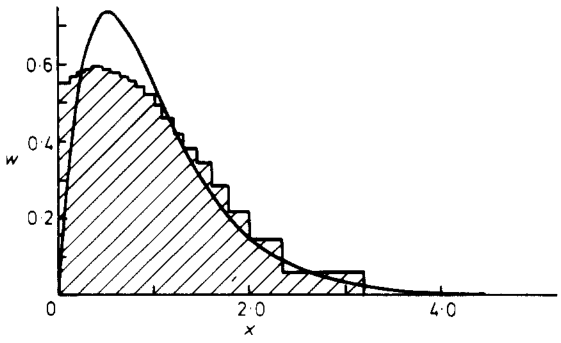

Figure 5. Twenty-block c-histogram for $\mu_1 = 1$, $\mu_2 = 1{\cdot}5$, with the continuous distribution $w = 4x\,e^{-2x}$.

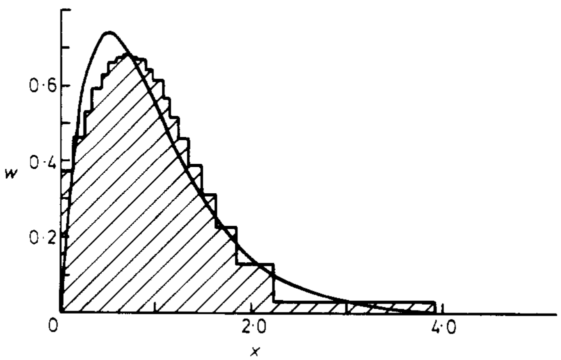

Figure 6. Twenty-block c-histogram for $\mu_1 = 1$, $\mu_2 = 1{\cdot}5$, $\mu_3 = 3{\cdot}0$, with the continuous distribution $w = 4x\,e^{-2x}$.

the graph of equation (4.28) for comparison. The successively closer approximation to the full curve is quite marked, and closely parallels that given by the earlier approach.

4.3. A 'two-peak' distribution

To test the computational method on a distribution with a more detailed structure, a distribution consisting of two isosceles triangles was selected, given by

$$w(x) = \begin{cases} 2x & 0 \leqslant x \leqslant 1/2 \\ 2(1-x) & 1/2 \leqslant x \leqslant 1 \\ 2(x-1) & 1 \leqslant x \leqslant 3/2 \\ 4-2x & 3/2 \leqslant x \leqslant 2 \\ 0 & \text{other values of } x. \end{cases} \tag{4.30}$$

Elementary algebra gives the formula for the moments as:

$$\mu_k = \frac{2}{(k+1)(k+2)}\left(2(1+2^{k+1}) - \frac{1}{2^{k+1}}(1+3^{k+2})\right). \tag{4.31}$$

Obviously to obtain any reasonable histogram which would approximate to (4.30) one would expect more moments to be required than the previous cases. In fact five moments and 20 blocks in the histogram give the fit shown in figure 7, where the histogram reproduces to a remarkable degree the main features of the original distribution, which is also shown for comparison. This implies that, for the moments specified, there is little 'room for manoeuvre' in assigning a probability distribution, since the double-triangle distribution giving rise to these moments is in no particular sense a maximum-entropy distribution. Incorporation of higher moments still would be expected to reduce the 'room for manoeuvre' still further, and the corresponding c-histogram to approach the double-triangle distribution still more closely as regards general features, although formal convergence of the histogram (as both $N \to \infty$, and $M \to \infty$) to the double triangle has not been established.

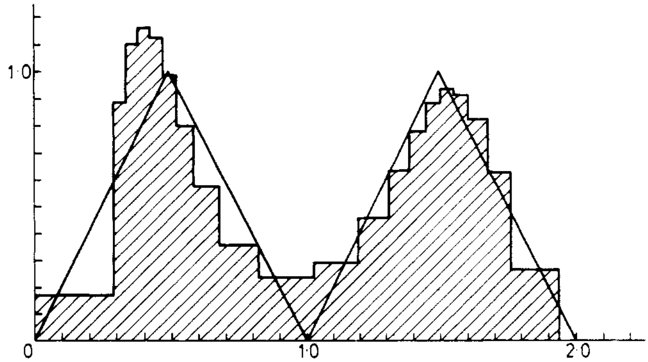

Figure 7. Twenty-block c-histogram fitted, with five moments, to a double-triangle distribution.

4.4. *An 'anomalous' distribution*

It has already been noted (Dowson and Wragg 1973) that the use of the maximum-entropy criterion for fitting distributions over $0 \leq x < \infty$ has one somewhat unsatisfactory aspect. This occurs when the first two moments μ_1, μ_2 are specified and $\mu_2 > 2\mu_1^2$. For this case no maximum-entropy distribution having the specified moments exists. For any finite range R the Jaynes–Shannon formalism *does* give admissible distributions but as $R \to \infty$ these always converge, for given x, to the negative exponential distribution $(1/\mu_1)\exp(-x/\mu_1)$ irrespective of the given value of μ_2.

In order to see the extent to which this behaviour is reproduced by the present approach using histograms we examine the χ^2 distribution with one degree of freedom cited by Wragg and Dowson (1970, 1973). In our present notation this distribution has density function

$$w(x) = (2\pi)^{-1/2} x^{-1/2} \, e^{-x/2} \tag{4.32}$$

with moments $\mu_1 = 1$, $\mu_2 = 3$. For this distribution $\mu_2/\mu_1^2 = 3 > 2$, and hence no maximum-entropy distribution exists over $0 \leq x < \infty$ satisfying these moments. For conciseness we term this an 'anomalous' case.

The current method when presented with these moments gives, for finite N, a distribution over a finite range $R = x_N$ even though this is not specified in advance. In view of the discussion above it would be expected that the c-histogram for this value of N would correspond closely with the JS distribution over $0 \leqslant x \leqslant x_N$. This is illustrated for the case $N = 10$ in figure 8.

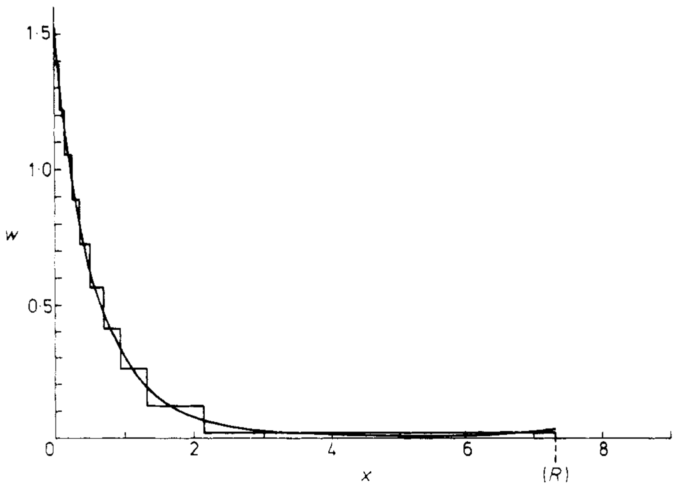

Figure 8. 'Anomalous case' c-histogram for $\mu_1 = 1$, $\mu_2 = 3$, and the corresponding JS distribution over the same range.

The histogram was obtained first and then the JS continuous distribution obtained for the same value of $R = x_N$. It can be seen that the main features of the continuous distribution are well represented by the histogram, with the exception of the slight increase in the $w_{\mathrm{JS}}(x)$ as $x \to R$, which is beyond the resolving power of the histogram to reproduce.

The correspondence between c-histograms with finite N and JS distributions over finite ranges means, of course, that taking increasingly large values of N will produce histograms which approximate more and more closely the limiting distribution $(1/\mu_1)\exp(-x/\mu_1)$. It is interesting, however, to compare the ten-block histogram fitted to the χ_1^2 distribution (shown in figure 9).

4.5. Convergence of the histogram

Two convergence problems arise in connection with the histograms produced by the method described here. To discuss them, let $w_N(x|M, \boldsymbol{\mu})$ denote the c-histogram of maximum entropy with N blocks, corresponding to M given moments $(\mu_1, \mu_2, \ldots, \mu_M)$. The first convergence question concerns the existence of a limiting distribution $w(x|M, \boldsymbol{\mu})$ given by

$$\lim_{N\to\infty} w_N(x|M, \boldsymbol{\mu}) = w(x|M, \boldsymbol{\mu}). \tag{4.33}$$

Assuming the existence of the limit implied by (4.33), the second question arises when the $(\mu_1, \ldots, \mu_M)$ are a subject of an indefinitely large number of moments, any of which

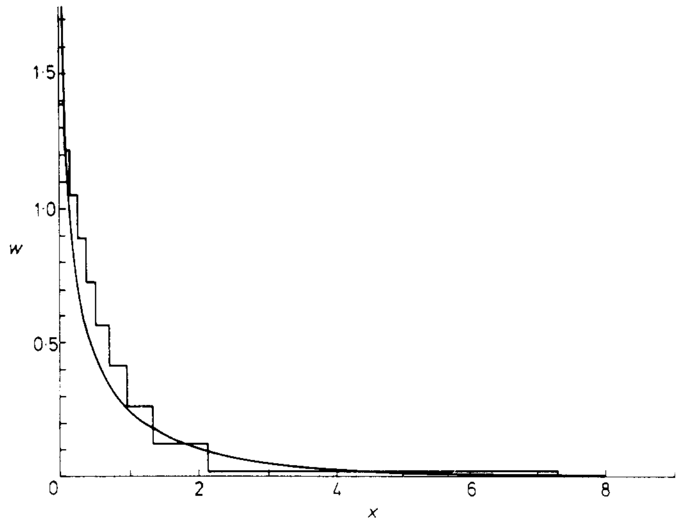

Figure 9. 'Anomalous case' ten-block c-histogram for $\mu_1=1$, $\mu_2=3$, and the χ^2 distribution $w=(2\pi)^{-1/2}x^{-1/2}\,e^{-\frac{1}{2}x}$.

are available numerically. The question then arises of the existence of a limiting distribution

$$w(x|\boldsymbol{\mu})=\lim_{M\to\infty} w(x|M,\boldsymbol{\mu}). \tag{4.34}$$

So far the only case in which the limit given by (4.33) has been *proved* to exist is in the simplest case $M=1$ discussed in § 3.4. All the other results obtained numerically are at least consistent with the following conjecture: 'For a set of moments $(\mu_1,\ldots,\mu_M)$ where there is a continuous maximum-entropy JS probability density function $w_{\mathrm{JS}}(x)$ with the moments specified, then this is the limiting probability density function $w(x|M,\boldsymbol{\mu})$'.

For the anomalous cases mentioned earlier the situation is more complicated. Some insight into the problem may be obtained by considering the $M=2$ case discussed earlier with $\mu_1=1$, $\mu_2=3$. It has been shown (Wragg and Dowson 1970, Dowson and Wragg 1973) that for a finite range $0\leqslant x\leqslant R$, there exists a maximum-entropy JS function:

$$w_R(x)=c\,\exp(-\beta x+\gamma x^2) \tag{4.35}$$

where c, β and γ are positive constants, and

$$1=\int_0^R \mathrm{d}x\, w_R, \tag{4.36}$$

$$1=\int_0^R \mathrm{d}x\, x w_R, \tag{4.37}$$

$$3=\int_0^R \mathrm{d}x\, x^2 w_R. \tag{4.38}$$

A particular case of (4.35) is shown in figure 8. The range can be split up into $0 \leqslant x \leqslant 5$, over which the Jaynes–Shannon $w_R(x)$ is approximately equal to $c \exp(-\beta x)$, and the region $5 < x \leqslant R$ over which the term in γ in (4.35) is significant. As R increases indefinitely, $c \to 1$, and $\beta \to 1$, and the γ-significant region moves steadily to the right. The value of γ is such as to satisfy (4.38) which for large R (and c, β approximately unity) becomes approximately

$$3 = \int_0^R \mathrm{d}x\, x^2 \exp(\gamma x^2 - x). \tag{4.39}$$

For given x then we have

$$\lim_{R \to \infty} w_R(x) = \mathrm{e}^{-x} \tag{4.40}$$

which of course does not satisfy the equation (4.38), and so the limiting distribution does not give the required second moment, even though each member of the sequence of distributions *does* give the required value. Present indications are that the behaviour of the histograms parallels this, suggesting that the maximum-entropy criterion is ill adapted to practical calculations for anomalous cases, which, however, are somewhat exceptional. For given N the histogram corresponds closely to the continuous JS distribution over the same range, and as $N \to \infty$ then for fixed x, the histogram approaches the curve e^{-x}, the region of appreciable discrepancy moving off to the right with increasing H (increasing range). This is in substantial agreement with the results for the rather artificial distribution shown in figure 1.

4.6. Depth distributions from ion implantation

When a beam of energetic ions impinges on a solid, the incident ions penetrate the solid, are gradually slowed by collisions with the atoms of the solid, and after neutralisation of their original charge, come to rest in the solid as implanted atoms. The theory of the slowing down process has been discussed in a number of papers, notably by Winterbon and his colleagues. (See for example Winterbon *et al* (1970, to be referred to as WSS) for a full discussion, where numerous further references are given.) It is a problem of some importance to be able to predict the eventual distributions both of the implanted atoms and the energy they release. Using an integro-differential equation to describe the slowing down process, a set of equations is derived yielding expressions for the moments of the depth distributions of energy and implanted atoms (WSS, p 42). The corresponding complete distributions must then be inferred from these moments. In the WSS paper this is done by several different statistical methods. The distributions obtained agree fairly well as regards their main features but differ considerably in detail (WSS, p 52).

To illustrate the use of the present method we select the particular case where the ion beam is incident normally on the solid surface, and the implanted atoms are of mass equal to that of those of the solid target. In this case the expression for the nth moment μ_n (in our notation) of the distribution of released energy (or 'damage' in WSS terminology) given on page 42 of the WSS paper simplifies to

$$\mu_n = \sum_l (2l+1) A_l^n. \tag{4.41}$$

Here the A_l^n are numbers resulting from the slowing-down theory which can be expressed in terms of integrals involving Legendre polynomials. They depend on an integral scattering parameter m and are tabulated, up to $l = 5$, on page 37 of the WSS paper.

Following WSS we select the case $m = 1/3$, giving the following moments:

$$\begin{aligned} \mu_0 &= 1{\cdot}000000 \qquad \text{(simple normalisation)} \\ \mu_1 &= 0{\cdot}505500 \\ \mu_2 &= 0{\cdot}359350 \\ \mu_3 &= 0{\cdot}315100 \\ \mu_4 &= 0{\cdot}322818 \\ \mu_5 &= 0{\cdot}374316. \end{aligned} \tag{4.42}$$

Although the values given in (4.42) are obviously specified to much higher accuracy than experimental measurements would justify, it was felt appropriate not to round-off to a smaller number of significant figures until the sensitivity of the method to small changes in the moments had been tested. Corresponding c-histograms for both four and five moments have been computed and are shown in figure 10. For comparison the

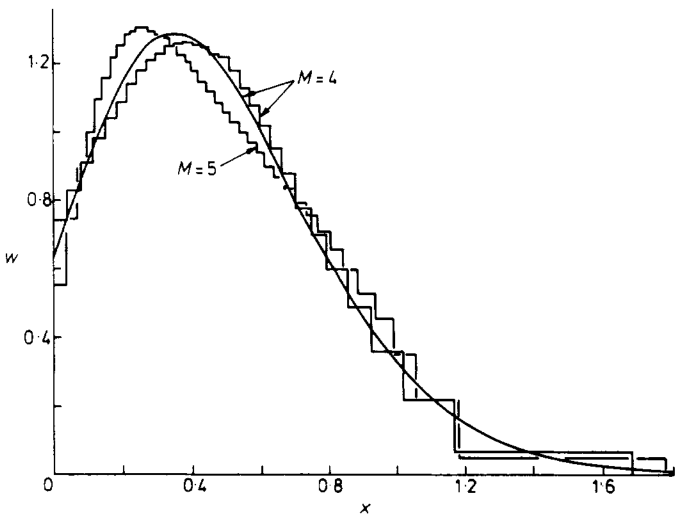

Figure 10. Continuous and histogram maximum-entropy distributions for the WSS moments.

continuous JS distribution is shown for the first four moments specified. The determination of this JS curve gave rise to numerical difficulties, and so far it has not proved possible to obtain the five-moment JS curve at all. The determination of the c-histograms, on the other hand, did not present any special technical problems. This seems to be typical, in that even in cases where the continuous JS distribution for a given set of moments does exist over $0 \leqslant x < \infty$, it is more difficult to obtain numerically.

From figure 10 it can be seen that the specification of the fifth moment has the effect of moving the maximum of the damage distribution function from $x = 0{\cdot}4$ to $x = 0{\cdot}27$, i.e. nearer to $x=0$ (the solid surface, in physical terms). This is in broad qualitative agreement with the wss curves for the Edgeworth series where the introduction of higher terms in the series moves the maximum back from 0·5 to 0·3. A significant difference between figure 10 and all the wss curves lies in the ratio w_m/w_0, where w_m denotes the maximum value of the distribution function, and w_0 denotes the corresponding value at $x=0$. For all the distributions of figure 10 this has a value of about 2·2, whereas for the wss curves, it ranges from 2·6 to 4·1, that for the Edgeworth series being 3·4. The reason is probably that the wss distribution are all fitted to the given moments over $-\infty < x < \infty$, and contain significant contributions from the physically unrealistic region $x<0$. It would appear that this leads to an underestimate of the damage immediately below the surface (as compared with that further in) by something approaching 50%. In this respect, the present method seems much more reliable in that the distribution is only defined (and the moments evaluated) over the region inside the solid where it actually exists. This is quite apart from the fact that it makes no prior assumptions about the shape of the distribution, whereas all the statistical methods used in the wss paper depend on the assumption of a particular functional form (usually related to a Gaussian) which is assigned to an *ad hoc* basis.

To assess the effects on the histograms of small variations in the specified moments, it was necessary to resort to empirical methods, since so far it has not been found possible to express the histogram parameters as analytic functions of the moments. A 20-block histogram was calculated specifying the first three moments given by relations (4.42). Some or all of these were then varied by ±1% and the corresponding 20-block histograms computed. Some idea of the variations produced is given in table 1 which shows the resulting values of the moments μ_4 and μ_5 (which may be compared with the values in (4.42)) and the corresponding entropy $S_{20}(\mu_1, \mu_2, \mu_3)$. It can easily be seen that while the changes in μ_4 and μ_5 are small and apparently nearly linear functions of the $\Delta\mu_1$, $\Delta\mu_2$, $\Delta\mu_3$ the changes in S are quite large and very non-linear. For example $(\Delta\mu_1, \Delta\mu_2, \Delta\mu_3)=(0, +1\%, 0)$ gives $\Delta S = 0{\cdot}0176$ while $(\Delta\mu_1, \Delta\mu_2, \Delta\mu_3)=(0, -1\%, 0)$ gives $\Delta S=-0{\cdot}0300$, a change nearly twice as great, and amounting to 18% of the original entropy.

Table 1. Variations in estimated values of μ_4, μ_5 and entropy resulting from small changes in specified values of μ_1, μ_2 and μ_3.

$\Delta\mu_1$	$\Delta\mu_2$	$\Delta\mu_3$	μ_4	μ_5	S_{20}
0	0	0	0·3209	0·3642	0·1662
+1%	0	0	0·3267	0·3832	0·1390
−1%	0	0	0·3159	0·3482	0·1827
0	+1%	0	0·3104	0·3345	0·1838
0	−1%	0	0·3328	0·3999	0·1362
0	0	+1%	0·3310	0·3870	0·1599
0	0	−1%	0·3111	0·3429	0·1710
+1%	−1%	0	0·3394	0·4228	0·0984
+1%	−1%	+1%	0·3505	0·4499	0·0857

Examples of the actual histograms produced are shown in figure 11. Clearly the shape obtained for the distribution is quite sensitive to small variations in the specified moments, although the variations are no greater than those produced by different statistical methods of calculation for the *same* moments as shown in the wss paper.

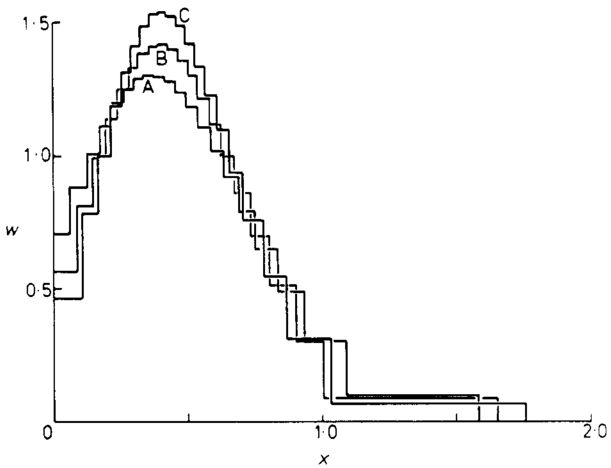

Figure 11. Histogram sensitivity to small variations in the first three moments specified. A, μ_1, μ_2, μ_3 unchanged; B, $\Delta\mu_1 = +1\%$; C, $\Delta\mu_1 = +1\%$, $\Delta\mu_2 = -1\%$.

5. Summary and conclusions

Results so far suggest that the approach described here is a useful way of assigning a probability distribution to a variate when only a limited number of moments are available. Moreover the method offers significant practical advantages over the continuous Jaynes–Shannon formalism although the difficulty associated with distributions described as anomalous is reproduced. Numerical investigations indicate that the entropy of a distribution determined by the method is very sensitive to variations in the prescribed moments. A particularly attractive feature of the method, unlike most numerical estimates of semi-infinite distributions, is that the upper limit of the histogram does not have to be specified in advance, but is assigned automatically by the program. Results so far have been confined to the semi-infinite range; however, this limitation is not inherent, and it is hoped to present results for distributions over $-\infty < x < \infty$ before long.

Acknowledgments

The authors would like to express their appreciation of some very stimulating and constructive comments by the referees, which have directed attention to several important points and implications for further study.

References

Brillouin L 1956 *Science and Information Theory* (New York: Academic)
Collins R 1965 *Proc. Phys. Soc.* **86** 199–211
—— 1967 *Phase Stability of Metals and Alloys* eds P S Rudman, J Stringer and R Jaffee (New York: McGraw-Hill) pp 499–522
Dowson D C and Wragg A 1973 *IEEE Trans. Inf. Theor.* **IT-19** 689–93

Jaynes E T 1957 *Phys. Rev.* **106** 620–30
Katz A 1967 *Principles of Statistical Mechanics* (San Francisco: Freeman)
Porter C E (ed.) 1965 *Statistical Theories of Spectra: Fluctuations* (New York: Academic)
Powles J G and Carazza B 1970a *Magnetic Resonance* ed. C K Googan (New York: Plenum) pp 133–61
—— 1970b *J. Phys. A: Gen. Phys.* **3** 335–41
Shannon C E and Weaver W 1949 *Mathematical Theory of Communication* (Urbana: University of Illinois Press)
Shohat J A and Tamarkin J D 1950 *The Problem of Moments* (Providence, RI: American Mathematical Society)
Stewart C A 1946 *Advanced Calculus* (London: Methuen)
Wilson G A and Wragg A 1973 *J. Inst. Math. Applic.* **12** 165–73
Winterbon K B, Sigmund P and Sanders J B 1970 *K. Danske Vidensk. Selsk., Mat.-fys. Meddr.* **37** No. 14
Wragg A and Dowson D C 1970 *IEEE Trans. Inf. Theor. (Corresp.)* **IT-16** 226–30