MAXIMUM-ENTROPY PROBABILITY DISTRIBUTIONS FOR CONTINUOUS RANDOM VARIATES

BY

J.N. KAPUR

I I T, Kanpur

(Received : August, 1981)

SUMMARY

Maximum-Entropy Probability Distributions (MEPD'S) are sought to be obtained when the continuous random variate varies over the range $(-\infty, \infty)$ and the moments prescribed are (*i*) mean alone (*ii*) variance alone (*iii*) coefficient of variation alone (*iv*) mean and mean deviation about some fixed point (*v*) $E \ln (1+x^2)$. The MEPD'S are also obtained for the range $[0, \infty)$ when the moments prescribed are : (*i*) mean m and $E \mid x-m \mid$ (*ii*) mean m and $E \mid x-m_0 \mid$ (*iii*) $E (\ln x)$ and $E \ln (1+x))$ (*iv*) $E (\ln (1+x)^2))$ (*v*) $E (\ln x)$ and $(E (\ln x)^2$ (*vi*) $E (x^n)$ and $E (\ln x)$ (*vii*) variance alone. Finally the MEPD's are obtained for the range [0, 1] when the median or quartiles or quantiles are prescribed.

1. INTRODUCTION

Let x be a continuous random variate varying over the interval $[a, b]$ and let $f(x)$ be its probability density function. Let the values of k of its moments be prescribed so that

$$\int_{o}^{b} f(x)\, d x=1, \int_{a}^{b} f(x)\, g_r(x)\, dx=a_r,\ r=1, 2, \ldots, k \qquad \ldots(1)$$

In general, there will be an infinity of probability density functions satisfying (1). If no other information is available about $f(x)$, according to Jayne's [5] Maximum-Entropy Principle, we should choose that density function, out of all those satisfying (1), which is 'least committed' to the unknown information, which is 'most-unbiased', which is 'most random', which is 'most uniform' and which is 'most uncertain'. More precisely, we have to choose the density function $f(x)$ which maximizes the informational entropy

$$S=-\int_a^b f(x)\ \ln f(x)\ dx, \qquad \ldots(2)$$

subject to (1) being satisfied. It is easily shown by using Euler-Lagrange equation of calculus of variations that the Maximum-Entropy Probability Density Function (MEPDF) is given by

$$f(x)=exp\ [-\lambda_0-\lambda_1\ g_1\ (x)\ldots-\lambda_k\ g_k\ (x)\ ], \qquad \ldots(3)$$

where the Lagrange multipliers $\lambda_0,\lambda_1,\lambda_2,\ldots,\lambda_k$ are determined by using (1).

It is obvious that the MEPDF depends on (1) the range of the variate and (*ii*) the moments prescribed.

The following results are already available in the literature

TABLE I

Range	*Moments Prescribed*	*MEPD*	*References*
(∞, ∞)	$E(x^2)$	Normal	[2, 3, 4, 6 9, 10, 11, 2]
,,	$E(x), E(x^2)$	,,	
,,	$B(x)=m. E(x-m)^2$	,,	
,,	$E(\|x\|)$	Laplace	[2, 6]
,,	$E(\|x-m\|)$	,,	
,,	$E(\ln(1+x^2))$	Cauchy	[2]
$[0, \infty]$	$E(x)$	Exponential	[3, 4, 6, 11, 12]
,,	$E(x), E(\ln x)$	Gamma	[2, 6, 12]
,,	$E(x)=m, E(x-m)^2=\sigma^2$	Truncated Normal if $\sigma^2<m^2$ Exponential if $\sigma^2=m^2$ Does not exist if $\sigma^2>m^2$	[1, 12, 14]
[0, 1]	No Moment	Uniform	[2, 3, 4, 6, 7, 11, 12]
,,	Mean m	Uniform if $m=\frac{1}{2}$, Truncated Normal otherwise.	[7]
,,	$E[\ln x], E[\ln(1-x)]$	Beta	[7, 9, 12]
,,	$E(x), E(x^2)$	Truncated Normal or Truncated U or uniform	[7]

The object of the present paper is to fill in the gaps in our existing knowledge, to examine the existence and uniqueness of MEPD's and to discuss the case when instead of moments, quantiles are prescribed.

2. MEPD's When x Varies Form $-\infty$ to ∞

When mean m and variance σ^2 (or first and second order moments about the origin) are both prescribed, the distribution with the maximum entropy is the normal distribution $N(m\ \sigma^2)$ and its entropy is $\ln [\sqrt{2\pi e}\ \sigma]$. We now discuss some related cases

(*i*) *If mean alone is prescribed*, then from (1) and (2)

$$f(x)=exp\ (-\lambda_0-\lambda_1\ x), \qquad ...(4)$$

where

$$\exp(-\lambda_0)\int_{-\infty}^{\infty}\exp(-\lambda_1\ x)\ dx=1 \qquad ...(5)$$

$$\exp(-\lambda_0)\int_{-\infty}^{\infty}x\exp(-\lambda_1\ x)\ dx=m$$

Equations (5) do not determine λ_0 and λ_1. As such, when mean value is prescribed, the MEPD does not exist. To see the reason for this, we note that the entropy for the normal distribution is indepedent of m and it can be made as large as we like by making σ sufficiently large. Thus even if we confine ourselves to normal distributions with the given mean, we find that the entropy is unbounded and there is no distribution with a finite maximum entropy. There is however a distribution with minimum entropy and this is the degenerate normal distribution with given mean m and variance zero. This is the Dirac delta function distribution which arises when the value m is taken with probability unity and all other variate values have probability zero.

The result can be generalised to show that if moments of first r orders are prescribed, no MEPD exists if r is odd.

(*ii*) *If variance alone is prescribed*, then we have to maximize

$$-\int_{-\infty}^{\infty}f(x)\ \ln f(x)\ dx \qquad ...(6)$$

subject to

$$\int_{-\infty}^{\infty} f(x)\, dx=1, \quad \int_{-\infty}^{\infty} x^2 f(x)\, dx-\left[\int_{-\infty}^{\infty} x f(x)\, dx\right]^2=0 \quad ..(7)$$

Using Euler-Lagrange equation, we get

$$1+\ln f(x)+(\lambda_0-1)+\lambda_1 \left[x^2-2x \int_{-\infty}^{\infty} x f(x)\, dx\right]=0 \quad ...(8)$$

Substituting
$$\int_{-\infty}^{\infty} x f(x)\, dx=m, \quad ...(9)$$

(8) gives
$$f(x)=\exp\left[=\lambda_0-\lambda_1 (x^2-2mx)\right] \quad ...(10)$$

Using (7), (9) and (10), we get

$$f(x)=\frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{1}{2}\frac{(x-m)^2}{\sigma^2}\right], \quad ...(11)$$

so that the MEPD is $N(m, \sigma^2)$ where m is arbitrary. Thus when variance alone is prescribed, the MEPD is *any* normal distribution with given variance σ^2 but with any *arbitrary* mean m. Thus in this case, *the MEPD is not unique*. The reason for this non-uniqueness is obvious from the fact that the expression for the entropy of a normal distribution is independent of m.

Goldman [3] sought to find the MEPD when the variance alone is prescribed. However he found the MEPD when $E(x^2)$ alone is prescribed.

(*iii*) *If coefficient of variation alone is prescribed as a,* then the constraint is

$$\int_{-\infty}^{\infty} x^2 f(x)\, dx-(1+a^2)\left[\int_{-\infty}^{\infty} x f(x)\, dx\right]^2=0 \quad ...(12)$$

Proceeding as before, we get

$$f(x)=\frac{1}{\sqrt{2\pi}\, am} \exp\left[-\frac{1}{2}\frac{(x-m)^2}{a^2 m^2}\right], \quad ...(13)$$

where m is arbitrary. The entropy of this can be made as large as we please and as such no MEPD exists in this case. A minimum entropy probability distribution exists and is the Dirac delta function distribution with zero mean and zero variance.

(*iv*) *If mean and mean deviation about a fixed point are prescribed,* the MEPD is

$$f(x)=A_e^{-vx}\, e^{-\mu|x-m|} \qquad \ldots(14)$$

If mean deviation about a fixed point alone is prescribed, the MEPD is Laplace distribution. If both mean and mean deviation about the mean are prescribed, the MEPDE is still the Laplace distribution. If mean (m) and mean deviation about the origin (σ) are prescribed, we get

$$\left.\begin{aligned} f(x)=A_e^{-vx}\, e^{-\mu|x|}\ ;\ A\int_0^{\infty} e^{-vx-\mu x}\, dx + A\int_{-\infty}^{0} e^{-vx+\mu x}\, dx=1 \\ A\int_0^{\infty} xe^{-vx-\mu x}\, dx + A\int_{-\infty}^{0} xe^{-vx+\mu x}\, dx=m \\ A\int_0^{\infty} xe^{-vx-\mu x}\, dx - A\int_{-\infty}^{0} xe^{-vx+\mu x}\, dx=\sigma, \end{aligned}\right] \qquad \ldots(15)$$

from which we can attempt to determinee A, v and μ. The MEPD is not the Laplace distribution

(*v*) If $E[\ln(1+x^2)]$ is prescribed as c, we get

$$f(x)=\frac{A}{(1+x^2)^b},\ -\infty<x<\infty \qquad \ldots(16)$$

where

$$\int_{-\infty}^{\infty}\frac{A}{(1+x^2)^b}\, dx=1,\ A\int_{-\infty}^{\infty}\frac{\ln(1+x^2)}{(1+x^2)^b}dx=c \qquad \ldots(17)$$

or

$$2A\int_0^{\pi/2}\cos^{2b-1}\theta\, d\theta=1, -2A\int_0^{\pi/2}\cos^{2b-2}\theta\ \ln\cos^2\theta\, d\theta=c \quad \ldots(18)$$

Let

$$\int_0^{\pi/2}(\cos^2\theta)^m d\theta=f(m), \text{ then } \int_0^{\pi/2}(\cos^2\theta)^m \ln(\cos^2\theta)d\theta=f'(m), \ldots(19)$$

so that

$$c=-\frac{f'(b-1)}{f(b-1)}=-\frac{f'(m)}{f(m)}\ ;\ m=b-1 \quad \ldots(20)$$

but

$$f(m)=\frac{\pi(1/2)\Gamma(m+1/2)}{2\ \Gamma(m+1)}, \quad \ldots(21)$$

so that

$$c=-\frac{f'(m)}{f(m)}=\frac{\Gamma'(m+1)}{\Gamma(m+1)}-\frac{\Gamma'(m+1/2)}{\Gamma(m+1/2)}$$

$$=\frac{\Gamma'(b)}{\Gamma(b)}-\frac{\Gamma'(b-1/2)}{\Gamma(b-1/2)} \quad \ldots(22)$$

or $$c=\psi(b)-\psi(b-\tfrac{1}{2})\ ;\ \psi(b)=\frac{\Gamma'(b)}{\Gamma(b)}, \quad \ldots(23)$$

which $\psi(b)$ is digamma function which is the derivative of the logarithm of the gamma function. If $b=1$, $c=2\ \ln 2=1.386264$. As c increases, b decreases. As $c\to o$, $b\to\infty$, as $c\to\infty$, $b\to\frac{1}{2}$.

Thus Cauchy distribution is the MEPD if $E(\ln(1+x^2))$ is prescribed as $2\ \ln\ 2$. If $[\ln(1+x^2)]$ is prescribed as some other value between 0 and ∞, the MEPD is the generalised Cauchy distribution (16) where the exponent b is determined by the prescribed value of a.

Figure 1 gives the relation between b and c. If $c<2\ \ln 2$, $b>1$. The generalised Cauchy distribution can be more useful than the Cauchy distribution, since here moments of higher order can exist if b is sufficiently large.

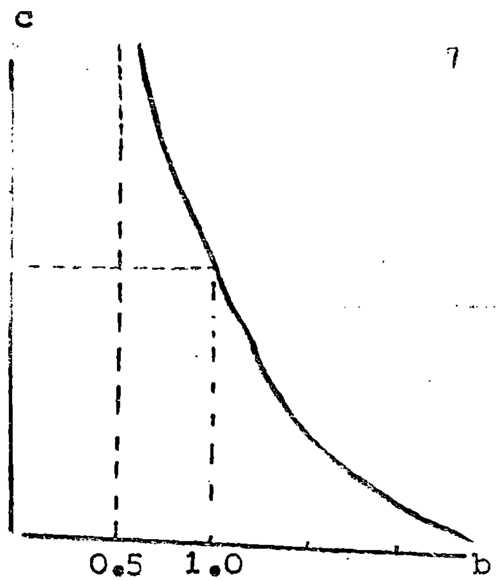

Fig. 1

3. MEPD'S WHEN THE RANDOM VARIATE VARIES FROM O TO ∞

(i) If both arithmetic mean m and geometric mean g are prescribed, it is well-known that the MEPDF is given by the gamma density function

$$f(x)=\frac{a^{\gamma}}{\tau(\gamma)}e^{-ax}\, x^{\gamma-1},\ 0\leqslant x<\infty \qquad ...(24)$$

where

$$m=\frac{\gamma}{a},\ \ln g=\frac{\tau'(\gamma)}{\tau(\gamma)}-\ln a, \qquad ...(25)$$

so that $\ln m-\ln g=\ln \gamma-\Psi(\gamma)$ ...(26)

Knowing m and g, we can find γ from (26) if $m>g$, and then a can be determined from (25).

(ii) If the mean deviation about a fixed point m_o *is prescribed as* σ_o, *then* the MEPDF is given by

$$f(x)=A\, e^{-b|x-m_o|},\ x>0, \qquad ...(27)$$

where

$$A=\frac{1}{2m_o}[1+bm_o-b\sigma_o],\ e^{-bm_o}=\frac{2(1-b\sigma_o)}{1+bm_o-b\sigma_o} \qquad ...(28)$$

Thus in this case, the MEPD is truncated Laplace distribution (29) where A and b are given by (29), b lies between $(m_o+\sigma_o)^{-1}$ and $(\sigma_o)^{-1}$ and $A \gtrless \frac{1}{2m_o}$ according as $m_o \gtrless \sigma_o$.
We get the following tables for A and b

TABLE II

A

σ_o \ m_o	.1	.25	.5	1.0	2.0
1	.500	.454	.466	.481	.489
2	.302	.250	.250	.278	.229
3	.224	.185	.167	.158	.153
4	.180	.150	.154	.125	.119
5	·150	.128	.114	.106	.100

TABLE III

b

σ_o \ m_o	1	2	3	4	5
1	.758	.825	.908	.958	.982
2	.398	.384	.393	.413	.434
3	.277	.260	.256	.259	.266
4	.214	.200	.194	.192	.193
5	.175	.163	.157	.154	.154

The mean of the MEPD is given by

$$m=A\int_0^\infty xe^{-b|x-m_o|}\,dx=m_o+\frac{(1-b\sigma_o)(1+bm_o)}{b^2m_o}, \qquad \ldots(31)$$

so that the mean of the MEPD is always greater than m_o. This is unlike the case of the MEPD over the range $(-\infty,\infty)$ where the mean was same as m_o.

(*iii*) *If both the mean and the mean deviation about the mean are prescribed,* MEPDF is

$$f(x)=Ae^{-ax-b|x-m|} \qquad \ldots(32)$$

and here $a\neq o$, unlike the situation over the range $(-\infty,\infty)$. Similarly if *the mean deviation about the mean is prescribed*, the discussion is much more complicated, illustrating the general principle that in general, the discussion of MEPD over the range $[0,\infty]$ is more difficult than over the range $(-\infty,\infty)$. This is similar to the experience of Dowson and Wragg [1] and Wragg and Dowson [14].

(*iv*) *If* $E(ln\ (1+x^2))$ *is prescribed*, the MEPD is unilateral generalised Cauchy distribution.

(*v*) *If* $E(ln\ x)$ *and* $E(ln\ (1+x))$ *are prescribed* as a and b respectively, then the MEPD is the beta distribution of the second kind with

$$f(x)=\frac{1}{B(m,n)}\frac{x^{m-1}}{(1+x)^{m+n}} \qquad \ldots(33)$$

$$a=\frac{1}{B(m,n)}\left[\frac{\partial}{\partial m}B(m,n)-\frac{\partial}{\partial n}B(m,n)\right]$$

$$=\Psi(m)-\Psi(n) \qquad \ldots(34)$$

$$b=-\frac{1}{B(m,n)}\frac{\partial}{\partial m}B(m,n)=\Psi(m+n)-\Psi(n) \qquad \ldots(35)$$

$$b-a=\Psi(m+n)-\Psi(n) \qquad \ldots(36)$$

Since $\Psi(x)$ is a monotone increasing function of x and $a>o$, $b>a$, we find $m>n$ and $b>a$. If m and n are integers.

$$a=\frac{1}{n}+\frac{1}{n+1}+\ldots+\frac{1}{m-1},\quad b=\frac{1}{n}+\frac{1}{n+1}+\ldots+\frac{1}{m+n-1} \qquad \ldots(37)$$

Since $\Psi(x)$ goes from $-\infty$ to ∞, it should be possible to solve for m and n from (34)–(36) when $b>a$.

(*vi*) *If* $E(x^n)$ *and* $E(ln\ x)$ *are prescribed*, we get the MEPDF

$$f(x)=\frac{n}{l^{c+1}\,\Gamma\left(\frac{c+1}{n}\right)}\exp\left[-\left(\frac{x}{b}\right)^n\right]x^c \qquad \ldots(38)$$

If we put $n=1$, we get gamma distribution and if we put $c=n-1$, we get Weibull's distribution which is of great importance in reliability theory. Tribus [12] gave it as an example of a distribution which cannot be deduced from the maximum-entropy principle. The other example he gave was of Cauchy distribution. We have seen that not only these can be derived, but these can be generalised. Thus (38) is a generalised Weibull's distribution. If

$$E(x^n)=m^n,\ E(\ln x)=\ln g, \qquad \ldots(39)$$

we can find b and c by using

$$m^n=b^n\frac{c+1}{n},\ \ln g=\ln b+\frac{1}{n}\Psi\left(\frac{c+1}{n}\right) \qquad .\ (40)$$

For Weibull's distribution, $c=n-1$, so that

$$m^n=b^n,\ \ln g=\ln b+\frac{1}{n}\Psi(1)=\ln b+\frac{1}{n}\Gamma'(1),$$

so that $m=b$, $\ln\frac{g}{b}=-\frac{1}{n}(0.547216), \ln\frac{g}{m}=-\frac{0.577216}{n}$ $\qquad \ldots(41)$

Thus if $E(x^n)$ is prescribed as m^n and $E(\ln x)$ is prescribed as $\ln g$, where g and m are related by (41), then the MEPD is the Weibull distribution.

(*vii*) The discussion when mean and variance are prescribed has been given by Wragg and Dowson [14]. From their discussion the following additional results can be deduced.

(*a*) *If the variance alone is prescribed*, the entropy is unbounded and the MEPD does not exist. This result is quite different from the corresponding result for the range $(-\infty, \infty)$.

(*b*) *If the coefficient of variation alone is prescribed* and is less than or equal to unity, no MEPD exists.

(*c*) *If the second moment about a fixed point m_o is prescribed* as σ_o, the MEPD exists and is a truncated normal distribution. For this distribution, the mean will always be greater than m_o and the variance will always be less than σ_o^2. In the particular case when $m_o=0$, mean comes out to be $\sigma_o\sqrt{2\pi}$ and variance comes out to be $\sigma_o^2(1-2/\pi)$, (*viii*) *If both $E(\ln x)$ and $E(\ln x)^2$ are prescribed*, the MEPD is lognormal distribution given by

$$f(x)=\frac{1}{x}\frac{1}{\sqrt{2\pi}\sigma}\exp\left[-\frac{(\ln x-m)^2}{\sigma^2}\right],\ 0<x<\infty \qquad ...(42)$$

(*ix*) If *geometric mean alone is prescribed as g*, then the MEPD is the Pareto distribution given by

$$f(x)=(\mu-1)\,c^{\mu-1}x^{-\mu},\ \mu=1+[\ln g/c]^{-1},\ x\geqslant c \qquad ...(43)$$

4. MEPD'S OVER A FINITE INTERVAL [0,1] WHEN QUANTILES ARE PRESCRIBED

(*i*) *If the mediam is prescribed as* a, then MEPDF is discrete and is given by

$$f(x)=\frac{1}{2a},\ 0\leqslant x<a \qquad ...(44)$$

$$=\frac{1}{2(1-a)},\ a<x<1$$

This would be continuous if and only if $a=1/2$ and then the MEPD is the uniform distribution. If $a\neq 1/2$, we can approximate the MEPD by a continuous distribution having entropy k times the maximum entropy where k can be made as near to unity as we like.

(*ii*) *If the three quartiles are prescribed* as Q_1, Q_2, Q_3, the MEPD is again discrete and is given by

$$f(x)=\frac{1}{4Q_1},\ O\leqslant x<Q_1$$

$$=\frac{1}{4(Q_2-Q_1)},\ Q_1<x<Q_2 \qquad ...(45)$$

$$=\frac{1}{4(Q_3-Q_2)},\ Q_2<x<Q_3$$

$$=\frac{1}{4(1-Q_3)},\ Q_3<x<1$$

so that the MEPDF is uniform in each of the four intervals (O,Q_1), (Q_1, Q_2), (Q_2, Q_3), $(Q_3, 1)$. This can again be approximated by a continuous density function having entropy as close to maximum entropy as we like.

(*iii*) Similarly *if* $(n-1)$ *quantiles are prescribed* as $Q_1, Q_2, \ldots,$ Q_{n-1}, the MEPDF is uniform over the n intervals (O, Q_1) $(Q_1, Q_2), \ldots$ $(Q_{n-1}, 1)$ and we can approximate it by a continuous distribution having entropy as near to the maximum entropy as we like.

5. Table Of New Results Obtained in This Paper

TABLE IV

Range $(-\infty, \infty)$		Range$(0, \infty)$	
Moment Prescribed	*MEPD*	*Moments Prescribed*	*MEPD*
Mean	Does not exist	$E(\mid x-m_0 \mid)$	Truncated Laplace with mean m_0
Variance	Normal with arbitrary mean	$E(x)=m$, $E(\mid x-m$	Modified turncated Laplace
Coefficient of variation	Does not exist	$E(\ln(1+x^2))$	Unilateral Generalied Chauchy
$E(x)$ and $E(\mid x-m \mid)$	Modified Laplace	$E(\ln x)$, $E(\ln(1+x))$ $E(x^n)$, $E(\ln x)$	Beta distribution of second kind Generalised Weibull
$E(\ln(1+x^2))$	Generalised Cauchy	$E(\ln x)$, $E(\ln x)^2$	Log Normal
Range (c, ∞)		Range $(0,1)$	
$E(\ln x)$	Pareto	Median, quartiles	Union of discrete uniform density functions

References

[1] Dowson D.C. and Wragg A. (1961) : Maximum-Entropy Distribution having prescribed first and second order moments, IEEE Trans. IT-19, 689-693.

[2] Gokhale D.V. (1978) : Maximum Entropy characterisation of some distributions. In *Statistical Distributions in Scientific work* Vol. III, edited by Patel, Golz and Ord 299-304, MA Redial, Boston.

[3] Goldman S. (1955) : *Information Theory*, Prentice Hall, New York.

[4] Guiasu S. (1977) : *Information With Applications*, McGraw Hill New York.

[5] Jaynes E.T. (1957) : Information Theory and Statistical Mechanics, *Physical Review*, **106**, 620-630, **108**, 171-197.

[6] Kagan A.M., Lenmik J.V. Rao C.R. (1973) : *Characterisation Problems in Mathematical Statistics*, Chapter XIII, Wiley, New York.

[7] Kapur J.N. (1982) : Maximum-Entropy probability distribution for a continuous random variate over a finite interval, *Jour. Math. Phys. Sci.* **16**, 97-109.

[8] Kapur J.N. (1984) : *Maximum-Entropy Models in Science and Engineering*, South Asia Publishers. New Delhi (to appear).

[9] Linsman J.H. and Van Zaylen M.C. (1972), : Note on the generations of the most probable frequency distributions Statistica, *Netherlandica*, **26**, 19-23.

[10] Rao C.R. (1974) : *Linear Statistical Inference and Its Applications*, Wiley Eastern. New Delhi.

[11] Reza F.M. (1961) : *Introduction To Information Theory*, McGraw Hill, New York.

[12] Tribus M. (1966) : *Rational Descriptions, Decisions and Designs*, Pergamon Press, Oxford.

[13] Verdugo Lazo A.C.G. and Rathie P.N. (1975) : On the Entropy of continuous probability distribution, *IEEE Trans. Inf. Theory IT* **24**, 120-123.

[14] Wragg A. Dowson D.C. (1970) : Fitting continuous Probability Density Functions over (0, ∞) using Information Theory, *IEEE Trans. IT*-**16**, 226-230.