Ann. Inst. Statist. Math.
Vol. 42, No. 2, 269–279 (1990)

A CLASSIFICATION OF THE MAIN PROBABILITY DISTRIBUTIONS BY MINIMIZING THE WEIGHTED LOGARITHMIC MEASURE OF DEVIATION*

SILVIU GUIASU

Department of Mathematics, York University, 4700 Keele Street, North York, Ontario, Canada M3J 1P3

(Received June 13, 1988; revised May 25, 1989)

Abstract. The paper reanalyzes the following nonlinear program: Find the most similar probability distribution to a given reference measure subject to constraints expressed by mean values by minimizing the weighted logarithmic deviation. The main probability distributions are examined from this point of view and the results are summarized in a table.

Key words and phrases: Probability distribution similar to a reference measure, constraints expressed by mean values, logarithmic measure of deviation, Shannon entropy, Kullback-Leibler number.

1. Introduction

As it is well-known, *some* main probability distributions have been reobtained by maximizing the Shannon entropy or by minimizing the Kullback-Leibler number subject to constraints expressed by mean values of some random variables (Kullback and Leibler (1951), Jaynes (1957*a*, 1957*b*), Kullback (1959), Ingarden (1963), Ingarden and Kossakowski (1971), Guiasu (1977, 1986), Preda (1982*a*, 1982*b*)). To give only one example, if the mean μ and the variance σ^2 are given, then on the real line, in the class of all probability distributions compatible with μ and σ^2, the normal distribution $N(\mu, \sigma^2)$ maximizes the Shannon entropy. Thus, as the Shannon entropy is a generally accepted measure of the amount of uncertainty contained by a probability distribution, we see that the normal distribution $N(\mu, \sigma^2)$ is the most random, or unbiased, probability distribution subject to the given constraints μ and σ^2.

Generally, we are dealing with two types of constraints: a) *Constraints*

*The author acknowledges the NSERC Canada Research Grant A-5712.

of type 1: mean values, or moments, as global indicators imposed on a probability distribution; b) *Constraints of type* 2: the rough shape of the probability distribution as sketched by a reference measure.

The object of this paper is to summarize the constraints under which the main probability distributions may be obtained by minimizing the weighted logarithmic deviation from a reference measure.

2. The weighted logarithmic deviation

Let $(\Omega, \mathscr{B})$ be a measurable space, w and Q two σ-finite measures on $\mathscr{B}$, and P a probability measure on $\mathscr{B}$, such that $P \simeq Q \simeq w$, where $\simeq$ means "equivalent to". Throughout the paper, w is the *weighting measure* and Q the *reference measure.* Let

$$p = dP/dw \quad \text{and} \quad q = dQ/dw$$

be the corresponding Radon-Nikodym derivatives. Suppose that $0 < p < +\infty$ and $0 < q < +\infty$ almost everywhere. We write $w = 1$ when Ω is countable and we have $w(\{\omega\}) = 1$ for every $\omega \in \Omega$. On the real line, m_L denotes the Lebesgue measure.

The *logarithmic deviation of P from Q weighted by w* is

$$D(P: Q; w) = \int p \ln (p/q) dw \,. \tag{2.1}$$

If Q is a probability measure, then $D(P: Q; m_L)$ on the real line, or $D(P: Q; 1)$ in the discrete case, is the Kullback-Leibler number. Also, $-D(P: m_L; m_L)$ on the real line, or $-D(P: 1; 1)$ in the discrete case, is the Shannon entropy.

PROPOSITION 2.1. *If Q is a finite measure, then*

$$D(P: Q; w) \geq \ln Q(\Omega) \tag{2.2}$$

with equality if and only if P is similar to Q, namely $P = Q/Q(\Omega)$.

PROOF. Inequality (2.2) is an immediate consequence of the inequality

$$t \ln t \geq t - 1, \quad (t > 0)\,, \tag{2.3}$$

with equality if and only if $t = 1$.

Proposition 2.1 tells us that the logarithmic deviation of the probability measure P from the finite reference measure Q cannot be smaller than the non-negative number $\ln Q(\Omega)$. The lower bound is obtained when P is in fact the probability measure $Q/Q(\Omega)$ induced by the finite measure Q. If Q is a probability measure itself, then $D(P: Q; w) \geq 0$, with equality if and only if $P = Q$.

Let $\boldsymbol{\theta} = (1, \theta_1, \ldots, \theta_n)$ be a numerical vector from $\mathbb{R}^{n+1}$, and let us denote by $\boldsymbol{h} = (1, h_1, \ldots, h_n)$ a vector whose components are measurable real functions defined on $(\Omega, \mathscr{B})$. The inner product in $\mathbb{R}^{n+1}$ is denoted by $(\cdot, \cdot)$. Suppose that the functions h_j $(j = 1, \ldots, n)$ are P-integrable. We are interested here in the following non-linear program:

Program A: $\min\limits_{P} D(P: Q; w)$ subject to:

$$\boldsymbol{\theta} = \int \boldsymbol{h} dP . \tag{2.4}$$

PROPOSITION 2.2. *The solution of Program* A *is*

$$p = qe^{-(\boldsymbol{\tau}, \boldsymbol{h})} \tag{2.5}$$

where the numerical vector $\boldsymbol{\tau} = (\tau_0, \tau_1, \ldots, \tau_n) \in \mathbb{R}^{n+1}$ *of Lagrange's multipliers satisfies the vector equality*

$$\boldsymbol{\theta} = \int \boldsymbol{h} q e^{-(\boldsymbol{\tau}, \boldsymbol{h})} dw . \tag{2.6}$$

For this solution we get

$$D(P: Q; w) = -(\boldsymbol{\tau}, \boldsymbol{\theta}) . \tag{2.7}$$

PROOF. For any $\boldsymbol{\tau}$, using either (2.3) or the simple inequality $\ln x \leq x - 1$ (with equality if and only if $x = 1$), we get

$$D(P: Q; w) \geq -(\boldsymbol{\tau}, \boldsymbol{\theta}) + 1 - \int q e^{-(\boldsymbol{\tau}, \boldsymbol{h})} dw , \tag{2.8}$$

with equality if and only if (2.5) is true. The right-hand side of the inequality (2.8), as a function of $\boldsymbol{\tau}$, attains its minimum if (2.6) holds; but (2.6) is immediately obtained by introducing (2.5) into (2.4). Finally, (2.5), (2.6) and (2.8) imply (2.7).

Remark. The number n is called the *order of classification.*

3. The classification of the main probability distributions

The method of minimizing the weighted logarithmic deviation from a reference measure subject to some given moments may be used in order to classify known probability distributions. The attached tables summarize the constraints under which the main probability distributions may be obtained by this optimization method. Each probability distribution mentioned in the tables is viewed as the solution of the nonlinear Program A subject to the constraints of type 1 (moments) and of type 2 (reference measures). For each probability distribution, the tables give the domain of definition Ω, the order of classification n, the measurable vector $\boldsymbol{h}$ whose mean vector $\boldsymbol{\theta}$ represents the constraints of type 1, the density q of the reference measure Q, and the weighted measure w. The row *RBP* contains the relationships between the parameters involved (the mean values θ_i, the Lagrange multipliers τ_i, and the standard parameters of the probability distribution). The last row gives the expression of the minimum logarithmic deviation $D(P{:}\, Q; w)$ in terms of the standard parameters of the respective probability distribution P of density p.

The evaluation of the integrals and series involved in the applications of equalities (2.4)–(2.7) to the different probability distributions examined in the tables follows the standard techniques from calculus (see Gradshteyn and Ryzhik (1980) for useful standard formulas). As far as the special functions are concerned, according to the usual notations, Γ is the gamma function, ψ is the psi (digamma) function, i.e. $\psi = \Gamma'/\Gamma$, B is the beta function, and ζ is the Euler-Riemann function. When Ω is the real line or an interval on the real line, the corresponding measurability is taken in Borel's sense.

For each probability distribution, the tables give the classification corresponding to the maximum order n and, in some cases of interest, even for a smaller value of n. The order n of classification cannot exceed the number of parameters of the corresponding probability distribution. Any probability distribution with an analytical expression for its density has a classification of order 0. The order of classification of a probability distribution essentially depends on both types of constraints (the mean values and the reference measure). Program A has been solved for some constraints as simple as possible, but they are not the only choice. A simple classification uses a reference measure and elementary measurable functions $\{h_i\}$ that are independent of any external parameter, in which case the mean values $\{\theta_i\}$ completely characterize the standard parameters of the respective probability distribution.

In order to be more explicit, let us show, briefly, how the results mentioned in the attached tables are obtained. Let us take the second order classification for the gamma distribution (the first column of Table 2). In this case

Table 1.

	Probability distributions		
	Normal	Normal	Exponential
$p(x)$	$(2\pi\sigma^2)^{-1/2}e^{-(x-\mu)^2/(2\sigma^2)}$ $(\sigma > 0)$	$(2\pi\sigma^2)^{-1/2}e^{-(x-\mu)^2/(2\sigma^2)}$ $(\sigma > 0)$	$\alpha e^{-\alpha x}$ $(\alpha > 0)$
Ω	$(-\infty, +\infty)$	$(-\infty, +\infty)$	$(0, +\infty)$
n	2	1	1
$\boldsymbol{h}(x)$	$h_1(x) = x$; $h_2(x) = x^2$	$h_1(x) = x$	$h_1(x) = x$
$q(x)$	1	$e^{-x^2/(2\sigma^2)}$	1
w	m_L	m_L	m_L
RBP	$\tau_1 = -\theta_1/(\theta_2 - \theta_1^2)$; $\tau_2 = 1/[2(\theta_2 - \theta_1^2)]$; $\mu = -\tau_1/(2\tau_2)$; $\sigma^2 = 1/(2\tau_2)$; $\tau_0 = \ln[\sigma(2\pi)^{1/2}] + \mu^2/(2\sigma^2)$	$\tau_1 = -\theta_1/\sigma^2$; $\mu = -\sigma^2\tau_1$; $\tau_0 = \ln[\sigma(2\pi)^{1/2}] + \mu^2/(2\sigma^2)$	$\tau_1 = 1/\theta_1$; $\alpha = \tau_1$; $\tau_0 = -\ln\alpha$
Min dev	$-\ln[\sigma(2\pi)^{1/2}] - (1/2)$	$-\ln[\sigma(2\pi)^{1/2}] + \mu^2/(2\sigma^2)$	$\ln\alpha - 1$

Table 2.

	Probability distributions		
	Gamma	Gamma	Maxwell-Boltzmann
$p(x)$	$[\beta^\alpha\Gamma(\alpha)]^{-1}x^{\alpha-1}e^{-x/\beta}$ $(\alpha > 0, \beta > 0)$	$[\beta^\alpha\Gamma(\alpha)]^{-1}x^{\alpha-1}e^{-x/\beta}$ $(\alpha > 0, \beta > 0)$	$(4/\pi^{1/2})\beta^{3/2}x^2e^{-\beta x^2}$ $(\beta > 0)$
Ω	$(0, +\infty)$	$(0, +\infty)$	$(0, +\infty)$
n	2	1	1
$\boldsymbol{h}(x)$	$h_1(x) = -\ln x$; $h_2(x) = x$	$h_1(x) = x$	$h_1(x) = x^2$
$q(x)$	x^{-1}	$x^{\alpha-1}$	x^2
w	m_L	m_L	m_L
RBP	$\tau_1/\tau_2 = \theta_2$; $\ln\tau_2 - \psi(\tau_1) = \theta_1$; $\alpha = \tau_1$; $\beta = 1/\tau_2$; $\tau_0 = \alpha\ln\beta + \ln\Gamma(\alpha)$	$\tau_1 = \alpha/\theta_1$; $\beta = 1/\tau_1$; $\tau_0 = \alpha\ln\beta + \ln\Gamma(\alpha)$	$\tau_1 = 3/(2\theta_1)$; $\beta = \tau_1$; $\tau_0 = \ln(\pi^{1/2}/4) - (3/2)\ln\beta$
Min dev	$\alpha\psi(\alpha) - \alpha - \ln\Gamma(\alpha)$	$-\alpha\ln\beta - \ln\Gamma(\alpha) - \alpha$	$\ln(4/\pi^{1/2}) + (3/2)\ln\beta - 3/2$

$$\Omega = (0, +\infty), \quad h_1(x) = -\ln x, \quad h_2(x) = x,$$
$$q(x) = x^{-1}, \quad w = m_L.$$

The solution of Program A, given by (2.5), becomes

$$p(x) = x^{\tau_1 - 1} e^{-\tau_0 - \tau_2 x}. \tag{3.1}$$

Denote $\alpha = \tau_1$, $\beta = 1/\tau_2$. Using (3.1), the constraints (2.4) become

$$1 = \int_0^{+\infty} p(x)dx = e^{-\tau_0}\beta^\alpha \Gamma(\alpha),$$

$$\theta_1 = -\int_0^{+\infty} \ln x\, p(x)dx = \psi(\alpha) - \ln \beta, \quad \theta_2 = \int_0^{+\infty} xp(x)dx = \alpha\beta.$$

There are numerical tables containing the values of the function $\ln x - \psi(x)$ (see Abramowitz and Stegun (1970), for instance). The mean values θ_1 and θ_2 uniquely determine the parameters α and β. We get

$$\tau_0 = \alpha \ln \beta + \ln \Gamma(\alpha),$$

and (3.1) becomes the gamma distribution

$$p(x) = [\beta^\alpha \Gamma(\alpha)]^{-1} x^{\alpha - 1} e^{-x/\beta}.$$

The minimum deviation (2.7) is

$$D(P: Q; w) = -\tau_0 - \tau_1\theta_1 - \tau_2\theta_2 = \alpha\psi(\alpha) - \alpha - \ln \Gamma(\alpha).$$

Looking at Tables 1–7, we can see that the constraints of type 2 are as important as the constraints of type 1. The Cauchy distribution, for instance, has no finite moments; therefore we cannot consider the mean value of any polynomial as a constraint of type 1 and the only natural classification for it seems to have the order zero.

As the optimization criterion for obtaining a probability distribution is the same, Tables 1–7 allows us to see similarities and differences between the main probability distributions by simply examining what kind of constraints have been imposed in each case. Thus, the analogies between the Rayleigh, Weibull and Maxwell-Boltzmann distributions are evident. Also, we can easily notice the analogy between the negative binomial distribution and gamma distribution (in the first order classification).

The minimization of the weighted logarithmic measure of deviation does not only allow us to classify the main known probability distributions but also opens the possibility of obtaining new probability distributions

Table 3.

	Probability distributions		
	Chi-square	Beta	Beta
$p(x)$	$[2^{\nu/2}\Gamma(\nu/2)]^{-1}x^{\nu/2-1}e^{-x/2}$ ($\nu > 0$, integer)	$x^{\alpha-1}(1-x)^{\beta-1}B(\alpha,\beta)$ ($\alpha > 0, \beta > 0$)	$x^{\alpha-1}(1-x)^{\beta-1}B(\alpha,\beta)$ ($\alpha > 0, \beta > 0$)
Ω	$(0, +\infty)$	$(0, 1)$	$(0, 1)$
n	1	2	1
$\boldsymbol{h}(x)$	$h_1(x) = -\ln x$	$h_1(x) = -\ln x$; $h_2(x) = -\ln(1-x)$	$h_1(x) = -\ln x$
$q(x)$	$x^{-1}e^{-x/2}$	$[x(1-x)]^{-1}$	$x^{-1}(1-x)^{\beta-1}$
w	m_L	m_L	m_L
RBP	$\psi(\tau_1) = -\theta_1 - \ln 2$; $\nu = 2\tau_1$; $\tau_0 = (\nu/2)\ln 2 + \ln\Gamma(\nu/2)$	$\psi(\tau_1+\tau_2) - \psi(\tau_1) = \theta_1$; $\psi(\tau_1+\tau_2) - \psi(\tau_2) = \theta_2$; $\alpha = \tau_1$; $\beta = \tau_2$; $\tau_0 = \ln B(\alpha,\beta)$	$\psi(\tau_1+\beta) - \psi(\tau_1) = \theta_1$; $\alpha = \tau_1$; $\tau_0 = \ln B(\alpha,\beta)$
Min dev	$(\nu/2)\psi(\nu/2) - \ln\Gamma(\nu/2)$	$-\ln B(\alpha,\beta) + \alpha\psi(\alpha) - (\alpha+\beta)\psi(\alpha+\beta) + \beta\psi(\beta)$	$-\ln B(\alpha,\beta) - \alpha\psi(\alpha+\beta) + \alpha\psi(\alpha)$

Table 4.

	Probability distributions		
	F-distribution	Weibull	Rayleigh
$p(x)$	$\left[B\left(\frac{\nu_1}{2},\frac{\nu_2}{2}\right)\right]^{-1}\left(\frac{\nu_1}{\nu_2}\right)^{\nu_1/2}x^{\nu_1/2-1}\left(1+\frac{\nu_1}{\nu_2}x\right)^{-(\nu_1+\nu_2)/2}$ ($\nu_1 > 0, \nu_2 > 0, \nu_2$ integer)	$\alpha\beta x^{\alpha-1}e^{-\beta x^{\alpha}}$ ($\alpha > 0, \beta > 0$)	$2\beta x e^{-\beta x^2}$ ($\beta > 0$)
Ω	$(0, +\infty)$	$(0, +\infty)$	$(0, +\infty)$
n	2	1	1
$\boldsymbol{h}(x)$	$h_1(x) = -\ln x$; $h_2(x) = \ln\left(1+\frac{\nu_1}{\nu_2}x\right)$	$h_1(x) = x^{\alpha}$	$h_1(x) = x^2$
$q(x)$	1	$x^{\alpha-1}$	x
w	m_L	m_L	m_L
RBP	$\psi(\tau_2-\tau_1-1) - \psi(\tau_1+1) - \ln(\nu_2/\nu_1) = \theta_1$; $\psi(\tau_2) - \psi(\tau_2-\tau_1-1) = \theta_2$; $\nu_1 = 2(\tau_1+1)$; $\nu_2 = 2(\tau_2-\tau_1-1)$; $\tau_0 = \ln B(\nu_1/2, \nu_2/2) + (\nu_1/2)\ln(\nu_2/\nu_1)$	$\tau_1 = 1/\theta_1$; $\beta = \tau_1$; $\tau_0 = -\ln(\alpha\beta)$	$\tau_1 = 1/\theta_1$; $\beta = \tau_1$; $\tau_0 = -\ln(2\beta)$
Min dev	$-\ln B(\nu_1/2, \nu_2/2) - (1-\nu_1/2)\psi(\nu_1/2) - [(\nu_1+\nu_2)/2]\psi\{(\nu_1+\nu_2)/2\} + (1+\nu_2/2)\psi(\nu_2/2) - \ln(\nu_2/\nu_1)$	$\ln(\alpha\beta) - 1$	$\ln(2\beta) - 1$

Table 5.

	Probability distributions		
	Student-T	Cauchy	Pareto
$p(x)$	$(1 + x^2/\nu)^{-(\nu+1)/2}/[\nu^{1/2}B(\nu/2, 1/2)]$ ($\nu > 0$, integer)	$\lambda/[\pi(\lambda^2 + x^2)]$ ($\lambda > 0$)	$\alpha k^{\alpha} x^{-\alpha-1}$ ($\alpha > 0$)
Ω	$(-\infty, +\infty)$	$(-\infty, +\infty)$	$[k, +\infty)$
n	1	0	0
$\boldsymbol{h}(x)$	$h_1(x) = (1 + x^2/\nu)$		
$q(x)$	1	$(\lambda^2 + x^2)^{-1}$	$x^{-\alpha-1}$
w	m_L	m_L	m_L
RPB	$\psi(\tau_1) - \psi(\tau_1 - 1/2) = \theta_1$; $\nu = 2\tau_1 - 1$; $\tau_0 = (1/2) \ln \nu + \ln B(\nu/2, 1/2)$	$\tau_0 = \ln (\pi/\lambda)$	$\tau_0 = -\ln \alpha - \alpha \ln k$
Min dev	$-\ln B(\nu/2, 1/2) - (1/2) \ln \nu - [(\nu + 1)/2][\psi\{(\nu + 1)/2\} - \psi(\nu/2)]$	$\ln (\lambda/\pi)$	$\ln \alpha + \alpha \ln k$

corresponding to different constraints. It is striking in some sense that very few constraints suffice to determine the main probability distributions. Obviously, the class of such constraints, of both type 1 and type 2, may be diversified. Let us notice, for instance, that with one exception (the beta distribution), all the reference measures used in Tables 1–7 are unimodal. But the present approach, based on the nonlinear Program A, enables us to construct multimodal probability distributions as well. The relative modes, if known, cannot be included among the constraints of type 1, but they can be incorporated in the shape of the reference measure. To give only a simple example, according to (2.5), the closest probability distribution (in the sense of the logarithmic deviation D) to the bimodal reference measure $q(x) = 4x^2(1 - x^2) + 1$, subject to the mean value θ_1 of $h_1(x) = x$, has the form

$$p(x) = [4x^2(1 - x^2) + 1]e^{-\tau_0 - \tau_1 x}$$

where τ_0 and τ_1 satisfy (2.6).

4. Conclusion

By minimizing the weighted logarithmic deviation we can construct the probability distribution which is the most similar to a given reference measure subject to given mean values of some random variables. The main probability distributions may be classified from this point of view. While almost all these main probability distributions are unimodal, the procedure

Table 6.

	Probability distributions				
	Uniform	Uniform	Poisson	Hypergeometric	Zeta
$p(x)$	$(b-a)^{-1}$ $(a<b)$	$1/N$	$e^{-\lambda}\lambda^x/x!$ $(\lambda>0)$	$\binom{r}{x}\binom{M-r}{N-x}\Big/\binom{M}{N}$ $(r<M, N<M)$ (r, M, N integers)	$x^{-s}/\zeta(s)$ $(s>1)$
Ω	$[a, b]$	$\{1,\ldots,N\}$	$\{0, 1,\ldots\}$	$\{0, 1,\ldots,N\}$	$\{1, 2,\ldots\}$
n	0	0	1	0	1
$\boldsymbol{h}(x)$			$h_1(x)=x$		$h_1(x)=\ln x$
$q(x)$	1	1	$(x!)^{-1}$	$\binom{r}{x}\binom{M-r}{N-x}$	1
w	m_L	1	1	1	1
RBP	$\tau_0=\ln(b-a)$	$\tau_0=\ln N$	$\tau_1=-\ln\theta_1$; $\tau_0=\lambda$	$\tau_0=\ln\binom{M}{N}$	$\zeta'(\tau_1)/\zeta(\tau_1)=-\theta_1$; $\tau_1=s$; $\tau_0=\ln\zeta(s)$
Min dev	$-\ln(b-a)$	$-\ln N$	$-\lambda+\lambda\ln\lambda$	$-\ln\binom{M}{N}$	$-\ln\zeta(s)+s\zeta'(s)/\zeta(s)$

Table 7.

	Probability distributions		
	Binomial	Negative binomial	Geometric
$p(x)$	$\binom{N}{x}\alpha^x(1-\alpha)^{N-x}$ $(0<\alpha<1)$	$\binom{x-1}{r-1}\lambda^r(1-\lambda)^{x-r}$ $(0<\lambda<0)$ $(r>0, \text{integer})$	$\lambda(1-\lambda)^{x-1}$ $(0<\lambda<1)$
Ω	$\{0, 1, \ldots, N\}$	$\{r, r+1, \ldots\}$	$\{1, 2, \ldots\}$
n	1	1	1
$\boldsymbol{h}(x)$	$h_1(x) = x$	$h_1(x) = x$	$h_1(x) = x$
$q(x)$	$\binom{N}{x}$	$\binom{x-1}{r-1}$	1
w	1	1	1
RBP	$\tau_1 = \ln(N-\theta_1) - \ln\theta_1$; $\alpha = \theta_1/N = e^{-\tau_1}/(1+e^{-\tau_1})$; $\tau_0 = -N\ln(1-\alpha)$	$\tau_1 = \ln[\theta_1/(\theta_1 - r)]$; $\lambda = 1 - e^{-\tau_1}$; $\tau_0 = r\ln(1/\lambda - 1)$	$\tau_1 = \ln[\theta_1/(\theta_1 - 1)]$; $\lambda = 1 - e^{-\tau_1}$; $\tau_0 = \ln(1/\lambda - 1)$
Min dev	$N[\alpha\ln\alpha + (1-\alpha)\ln(1-\alpha)]$	$\frac{r}{\lambda}[\lambda\ln\lambda + (1-\lambda)\ln(1-\lambda)]$	$[\lambda\ln\lambda + (1-\lambda)\ln(1-\lambda)]/\lambda$

enables us to construct multimodal probability distributions as well, subject to given mean values.

Acknowledgements

The author is indebted to the referees of this paper for their valuable and constructive comments.

References

Abramowitz, M. and Stegun, I. A. (eds.) (1970). *Handbook of Mathematical Functions*, Dover Publications, New York.

Gradshteyn, I. S. and Ryzhik, I. M. (1980). *Table of Integrals, Series and Products*, Academic Press, New York.

Guiasu, S. (1977). *Information Theory with Applications*, McGraw-Hill, New York.

Guiasu, S. (1986). An optimization problem related to the zeta-function, *Canad. Math. Bull.*, **29**, 70–73.

Ingarden, R. S. (1963). Information theory and variational principles in statistical theories, *Bull. Acad. Polon. Sci. Ser. Math. Astronom. Phys.*, **11**, 541–547.

Ingarden, R. S. and Kossakowski, A. (1971). Poisson probability distribution and information thermodynamics, *Bull. Acad. Polon. Sci. Ser. Math.*, **19**, 83–86.

Jaynes, E. T. (1957*a*). Information theory and statistical mechanics, *Phys. Rev.*, **106**, 620–630.

Jaynes, E. T. (1957*b*). Information theory and statistical mechanics, *Phys. Rev.*, **108**, 171–182.

Kullback, S. (1959). *Information Theory and Statistics*, Wiley, New York; Chapman and Hall, London.

Kullback, S. and Leibler, R. A. (1951). On information and sufficiency, *Ann. Math. Statist.*, **22**, 79–86.

Preda, V. (1982*a*). Binomial distributions and information thermodynamics, *Bull. Acad. Polon. Sci. Ser. Math.*, **30**, 569–573.

Preda, V. (1982*b*). The Student distribution and the principle of maximum entropy, *Ann. Inst. Statist. Math.*, **34**, 335–338.