academicJournals

Vol. 6(4), pp. 51-57, April 2013
DOI: 10.5897/AJMCSR 2013.0463
ISSN 2006-9731 ©2013 Academic Journals
http://www.academicjournals.org/AJMCSR

African Journal of Mathematics and Computer Science Research

Full Length Research Paper

Information theoretic methods in parameter estimation

D. S. Hooda[1]*, Ketki Kulkarni[1] and Parmil Kumar[2]

[1]Jaypee University of Engineering and Technology, A. B. Road, Raghogarh, Distt. Guna, M. P. India.
[2]Department of Statistics, University of Jammu, Jammu, India

Accepted 20 March, 2013

In the present communication entropy optimization principles namely maximum entropy principle and minimum cross entropy principle are defined and a critical approach of parameter estimation methods using entropy optimization methods is described in brief. Maximum entropy principle and its applications in deriving other known methods in parameter estimation are discussed. The relation between maximum likelihood estimation and maximum entropy principle has been derived. The relation between minimum divergence information principle and other classical method minimum Chi-square is studied. A comparative study of Fisher's measure of information and minimum divergence measure is made. Equivalence of classical parameter estimation methods and information theoretic methods is studied. An application for estimation of parameter estimation when interval proportions are given is discussed with a numerical example.

Key words: Parameter estimation, maximum entropy principle, minimum divergence measure, maximum likelihood estimation, minimum interdependence principle.

INTRODUCTION

Shannon (1948) introduced the term 'entropy' which measures the uncertainty contained in a message. Thus the information contained in a message can be measured by the uncertainty removed by it. Shannon defined a function $H(p_1, p_2,....p_n)$ satisfying the properties intuitively expected of a measure of uncertainty like continuity, symmetry, additivity, being maximum when all outcomes are equally alike and being minimum when one of the outcomes is bound to occur. He proved that the only function which satisfies all these requirements was

$$s = H(p_1, p_2,....p_n) = -\sum_{i=1}^{n} p_i \log p_i \quad (1)$$

Equation (1) was called Shannon's entropy and is non-negative. Suppose that we know nothing about pi's except that:

$$\sum_{i=1}^{n} p_i = 1, \quad \text{and } p_i \geq 0 \quad \text{for each i.} \quad (2)$$

Then according to Laplace's principle of insufficient reasons which states that in the absence of any constraint except the natural constraints

$$\sum_{i=1}^{n} p_i = 1 or \int_a^b f(x)dx = 1$$

we should choose the uniform distribution that is,

$$p_1 = p_2 = \ldots\ldots = p_n = 1/n \quad (3)$$

There are infinite numbers of probability distributions which are consistent with Equation (2). Out of these Equation (3), viz., uniform distribution has maximum entropy. Let us suppose that we are given additional piece of information in the form of constraints:

$$\sum_{i=1}^{n} p_i g_r(X_i) = a_r, r = 1,2,....,m \quad (4)$$

*Corresponding author. E-mail:ds_hooda@rediffmail.com. Tel: 07544-267160. Fax: 07544-267011.

The probability distributions consistent with Equation (2) may not be consistent with Equation (4) and consequently, the maximum value of entropy of the distributions consistent with Equation (2) and (4) will be less than or equal to log n and the minimum value of this will be non-negative. Janyes (1957) generalized Laplace's principle to a more general situation and suggested that we should choose p_i 's so as to maximize S subject to the constraints (2) and (4). He called it the maximum entropy principle (MEP). According to this principle, we should use all the information we have and scrupulously avoid any information we do not have, Kapur (1989) has studied the application of maximum entropy principle in science and engineering.

The concept of the directed divergence of a probability distribution P from another probability distribution Q was given by Kullback and Leibler (1951) who introduced the following measure:

$$D(P:Q) = \sum_{i=1}^{n} p_i \log(p_i / q_i) \quad (5)$$

This is also called as measure of "distance", "discrepancy" or "discrimination" of one distribution from the other and D(P:Q) is always non-negative. In view of the fact that there is zero discrepancy if two distributions are identical, we require: D (P:Q) = 0 if P = Q.

If P = (p_1, p_2 , …..,p_n) is a given distribution and U = $(1/n, 1/n, ..., 1/n)$ is the uniform distribution , then the directed divergence of P from U, is given by

$$D(P:U) = \log n - S(P) \quad (6)$$

Thus, maximizing S (P) under certain constraints is equivalent to minimizing D (P:Q) under the same constraints. Hence, the principle of maximum entropy requires us to choose a distribution which is as close to the uniform distribution as possible subject to the given constraints. This is a new insight into the concept of maximum entropy which is provided by Kullback-Leibler's measure of directed divergence. Let Q = (q_1, q_2,.....,q_n) be a priori distribution of a distribution P = (p_1,p_2,.....,p_n) obtained on the basis of previous experience, intuition or some theory. In the presence of constraints, however, we would like to choose P which satisfies all the given constraints and should be as close to Q as possible. Thus, we choose P which minimize D (P:Q) subject to the given constraints. Since D (P: Q) is a measure of directed divergence of P from Q or of discrimination of information (MDIP). When Q = U, this includes the principle of maximum entropy as a special case. To minimize D (P:Q) subject to some linear constraints, we also require that D (P:Q) is a convex function of P and Q so that its local minimum should also be its global minimum. For the distribution of continuous random variable, MDI principle given by Kullback and Leibler (1951) requires us to minimize:

$$\int f(x) \log(f(x) / g(x)) dx \quad (7)$$

Subject to

$$\int_a^b f(x) = 1 \ \ and \ \ \int_a^b g_r(x) f(x) dx = a_r, r = 1,2,.....,m.$$

Let f(x, θ) be the probability density function (pdf) of a random variable X, where functional form of pdf is known except for the parameter θ. This parameter θ can be a scalar or a vector. One of the most important tasks in statistical inference is of estimating θ on basis of a random sample (x_1, x_2,.....,x_n) drawn from the population. The classical statistical methods of parameter estimation are: methods of moments, least squares, minimum chi-square, maximum likelihood, minimum distance and a recent one called method of probability weighted moment. Amongst all methods, Fisher's (1921) method of maximum likelihood is widely accepted and is considered as one of the best method for parameter estimation.

Akaike's (1971) work paved the way for the information theoretic approach in parameter estimation. Lind and Solana (1988) method is based on the principle of least information. Kapur (1989) compared the Gauss' method of estimation with a method based on the principle of maximum entropy.

In the present paper, we present a critical appraisal of parameter estimation methods using entropy optimization principles and compare these with classical methods such as method of moments and method of maximum likelihood. The basic principle is that, subject to the information available we should choose θ in such a way that the entropy is as large as possible or the distribution as nearly uniform as possible.

MAXIMUM ENTROPY PRINCIPLE IN PARAMETER ESTIMATION AND ITS APPLICATION

Let us consider f(x,θ) as the given functional form of pdf and we have to estimate the parameter θ for a given random sample x_1, x_2,..... x_n from the population. Fisher (1921) suggested the method of maximum likelihood that is, θ should be chosen such that it maximizes the likelihood function:

$$L(x,\theta) = \prod_{i=1}^{n} f(x_i), \quad (8)$$

or $$\log L(x, \theta) = \sum_{i=1}^{n} \log f(x_i, \theta) \quad (9)$$

Now a probability distribution can be formed such that

$$p_i = \frac{f(x_i,\theta)}{\sum_{i=1}^{n} f(x_i,\theta)}, \qquad i = 1,2,.....n, \tag{10}$$

where f (x_i, θ) is the value of pdf at X = x_i. For making p_i's as equal as possible, we choose parameter θ such that it maximizes Burg's (1972) measure of entropy for this distribution. However, it may be noted that we can use any other entropy to measure the uncertainty contained in the probability distribution of a random variable in an experiment. Burg's entropy measure for probability distribution

(p_1, p_2....., p_n; $p_i > 0$; $\sum_{i=1}^{n} p_i = 1$) is given by

$$H(P) = \sum_{i=1}^{n} \log p_i \tag{11}$$

Substituting (10) in (11), we have

$$H(P) = \sum_{i=1}^{n} \log \frac{f(x_i,\theta)}{\sum_{i=1}^{n} f(x_i,\theta)} \tag{12}$$

For maximizing Equation (12) with respect to θ, we put the first derivative of Equation (12) with respect to θ equal to zero and thus we get

$$\frac{\partial H(p)}{\partial \theta} = \sum_{i=1}^{n} \left(\frac{1}{f(x_i,\theta)} \cdot \frac{\partial f(x_i,\theta)}{\partial \theta} \right) - \frac{\sum_{i=1}^{n} \frac{\partial f(x_i,\theta)}{\partial \theta}}{\sum_{i=1}^{n} \partial f(x_i,\theta)} = 0 \tag{13}$$

But Fisher's method of maximum likelihood requires to solve

$$\sum_{i=1}^{n} \left(\frac{1}{f(x_i,\theta)} \cdot \frac{\partial f(x_i,\theta)}{\partial \theta} \right) = 0 \tag{14}$$

Since $\sum_{i=1}^{n}$ f (x_i, θ) is not independent of θ, therefore Equation (13) and (14) will give different estimates of θ. Hence Equation (13) is different from Fisher's method of maximum likelihood estimation (MLE).

Remarks

It may be noted f(x_1, θ), f(x_2, θ),....., f(x_n, θ) are not probabilities. Actually, these are the values of pdf at x_1,x_2,....., x_n. Their sum is not necessarily unity or independent of θ as x_1, x_2,....., x_n represents only a random sample and not all the values which the variate X can take. We have formed probability distribution of Equation (10) from these values on driving by the sum of the values of PDF.

Relation between maximum likelihood estimation (MLE) and maximum entropy principle (MEP)

Let x_1, x_2,....., x_n be a random sample from a population with pdf f(x, θ). We choose or estimate parameter θ in terms of the sample values such that it maximizes likelihood function. But according to MEP, we choose the value of θ such that the uncertainty that remains after the sample values are known as large as possible or, we can say that the entropy of the sample itself has to be a minimum. Thus, the Shannon's (1948) entropy is given by

$$H_S = -\sum_{I=1}^{n} f(x_i,\theta) \log f(x_i,\theta) \tag{15}$$

Or

$$H_S = -\frac{1}{n}\left[\log f(x_1,\theta) + \log f(x_2,\theta) + ... + \log f(x_n,\theta)\right]$$

$$= -\frac{1}{n}\left[\sum_{i=1}^{n} \log f(x_i,\theta)\right]$$

$$= -\frac{1}{n}\left[\log\left[\prod_{i=1}^{n} f(x_i,\theta)\right]\right]$$

$$= -\frac{1}{n}\left[\log L(x,\theta)\right] \tag{16}$$

Where $L(x,\theta)$ is the maximum likelihood function given by Equation (8). Thus, we choose θ such that it minimizes the entropy of the sample or maximizes the likelihood function. It implies that maximum entropy principle leads to the principle of maximum likelihood.

Now let us consider $\phi(x,\theta)$ as the cumulative distribution function of the second distribution in case of minimum cross entropy principle. We shall choose θ such that for the chosen value of θ it minimizes the entropy of the sample .The distribution function f(x, θ) is as close as possible to the distribution function determined by the random sample x_1, x_2,....., x_n.
Thus, minimum discrimination information statistic (refer to Kumar (2001) is given by:

$$D(\phi' : f) = \sum_{i=1}^{n} \phi'(x_i,\theta) \log \frac{\phi'(x_i,\theta)}{f(x_i,\theta)} =$$

$$\sum_{i=1}^{n} \phi'(x_i,\theta) \log \phi'(x_i,\theta) - \sum_{i=1}^{n} \log f(x_i,\theta) d\phi(x_i,\theta) \cdot \tag{17}$$

Equation (8) attains minimum when its second part is maximum. Consequently, we choose θ which can maximize:

$$\sum_{i=1}^{n} \log f(x_i,\theta) d\phi(x_i,\theta) =$$

$$\frac{1}{n}\left[\log f(x_1,\theta) + \log f(x_2,\theta) + ... + \log f(x_n,\theta)\right]$$

$$= \frac{1}{n}\left[\log\left[\prod_{i=1}^{n} f(x_i,\theta)\right]\right]$$

$$= \frac{1}{n}\left[\log L(x,\theta)\right]. \quad (18)$$

Thus, both maximum entropy and minimum cross entropy principles lead to maximum likelihood principle given by Fisher (1921).

MINIMUM DIVERGENCE INFORMATION PRINCIPLE IN PARAMETER ESTIMATION AND APPLICATIONS

It is very interesting and useful to study the relation between traditional methods and minimum divergence information principle in parameter estimation

Relation between measures of minimum divergence information and minimum chi-square

Let us consider that there are n classes and $Np_1, Np_2, ..., Np_n$ be the expected frequencies on the basis of parameter θ in these classes. Further, we consider that $Nq_1, Nq_2, ..., Nq_n$ are the observed frequencies in these n classes. Then we choose θ so as to minimize divergence measure D (P:Q) or D (Q:P).

Let qi = pi + ε_i ,

where ε_i is very small

Then $\sum_{i=1}^{n} \varepsilon_i = 0$, since $\sum_{i=1}^{n} p_i = \sum_{i=1}^{n} q_i = 1$

We have, D (P: Q) = $\sum_{i=1}^{n} p_i \log \frac{p_i}{q_i}$ (19)

$$= \sum_{i=1}^{n} p_i \log \frac{p_i}{p_i(1+\varepsilon_i/p_i)}$$

$$= -\sum_{i=1}^{n} p_i \log(1+\varepsilon_i/p) \quad (20)$$

$$\cong \frac{1}{2}\sum_{i=1}^{n} \frac{\varepsilon_i^2}{p_i} = \frac{1}{2}\sum_{i=1}^{n} \frac{(q_i-p_i)^2}{p_i}$$

Next, similarly we have

$$D(Q:P) \cong \frac{1}{2}\sum_{i=1}^{n} \frac{(q_i-p_i)^2}{q_i} \quad (21)$$

It may be pointed here that Equation (20) corresponds to modified chi-square while Equation (21) is chi-square statistic. Thus, from Equations (20) and (21) we can infer that θ is chosen to minimize

either $\frac{1}{2}\sum_{i=1}^{n} \frac{(O_i-E_i)^2}{E_i}$ or $\frac{1}{2}\sum_{i=1}^{n} \frac{(O_i-E_i)^2}{O_i}$,

where O_i and E_i are observed and expected frequencies in the ith class.

Fisher's measure of information (FMI) and minimum divergence measure

Let f(x, θ) = f and f (x, $\theta+\Delta\theta$) = g, be the two density functions, then divergence measure of f from g is given by

$$D(f:g) = \int_X f \log \frac{f}{g} dx$$

$$= \int_X f(x,\theta) \log \frac{f(x,\theta)}{f(x,\theta+\Delta\theta)} dx$$

$$= -\int_X f(x,\theta) \log \frac{f(x,\theta+\Delta\theta)+f(x,\theta)-f(x,\theta)}{f(x,\theta)} dx$$

$$= -\int_X f(x,\theta) \log(1+\frac{f(x,\theta+\Delta\theta)-f(x,\theta)}{f(x,\theta)}) dx$$

When $\Delta\theta \to 0$, we have

$$D(f:g) = -\int_X f(x,\theta) \log(1+\frac{\partial f(x,\theta)}{\partial\theta}\frac{\Delta\theta}{f(x,\theta)}) dx$$

$$= -\int_X f(x,\theta)\left[\frac{\partial f(x,\theta}{\partial\theta}\frac{\Delta\theta}{f(x,\theta)} - \frac{1}{2}(\frac{\partial f(x,\theta)}{\partial\theta}\frac{\Delta\theta}{f(x,\theta)})^2 + ...\right] dx \quad (22)$$

Since $\int_X f(x, \theta)\, dx = 1$, therefore

$$\int_X \frac{\partial f}{\partial \theta} dx = 0 \text{ and } \int_X \frac{\partial^2 f}{\partial \theta^2} dx = 0 \quad (23)$$

Equations (22) and (23) together gives

$$D(f:g) = \frac{1}{2}(\Delta\theta)^2 \int_X \frac{1}{f(x,\theta)} \left(\frac{\partial f(x,\theta)}{\partial \theta} \right)^2 dx + \ldots\ldots\ldots \quad (24)$$

$$\int_X \frac{1}{f(x,\theta)} \left(\frac{\partial f(x,\theta)}{\partial \theta} \right)^2$$

In Equation (24) is called Fisher's information measure. It can be noted Fisher's measure of information measures the power of discrimination or divergence between two density functions $f(x,\theta)$ and $f(x,\theta+\Delta\theta)$. Thus, greater the value of FMI, greater is the power of discrimination or it can be said that it gives us more information about θ.

Fisher's measure of information is different in many aspects from Shannon's measure of information and Kullback-Leibler's measure of divergence. Shannon's measure of information gives us information about the probability density functions while FMI gives information about the estimators of population parameters. When interval is finite FMI measures the directed divergence of $f(x,\theta)$ from $f(x,\theta+\Delta\theta)$, while Shannon's measure gives the directed divergence of f (x, θ) from uniform density function.

Fisher's measure of information gives directed divergence of f (x, θ) from density function depending on both f and θ, while Shannon's measure gives the directed divergence of f(x, θ) from a density function which is independent of both f and θ. The Kullback-Leibler measure of directed divergence can discriminate between any two density functions f(x, θ) and g(x, θ) while FMI discriminate between f(x, θ) and $f(x,\theta+\Delta\theta)$ only. Thus, these measures have different purposes, to decide the relative merits of information measures difficulty arises when the problems of discriminate are viewed in isolation.

In generalized model, these measures are considered in relation with the probability distributions and their moments.

EQUIVALENCE OF CLASSICAL AND INFORMATION THEORETIC METHODS OF PARAMETER ESTIMATION

Some classical statistical methods and information theoretic methods in parameter estimation have equivalence which we discuss thus.

Entropy optimization Principle and Laplace's principle of insufficient Reasoning

If the constraints are absent in Jaynes (1957), MEP, then maximization of uncertainty gives the uniform distribution. Thus, the Laplace principle is a special case of MEP. However, Hadgiwas (1981) has shown that the MEP and the MDI principles can be deduced from the principle of insufficient reasoning and thus, MEP and MDI can be regarded as the special case of Laplace's principle, while Laplace's principle can be regarded as a particular case of MDI principle when there are no constraints and the prior distribution is uniform.

Minimum discrimination information and maximum likelihood principle

A correspondence between the MDI and Fisher's maximum likelihood principle has been established. Suppose we are given g(x) then we find f(x) which minimizes

$$D(f:g) = \int_X f(x) \log \frac{f(x)}{g(x)} dx = \int_X f(x) \log f(x)\, dx - \int_X f(x) \log g(x)\, dx \quad (25)$$

and satisfies the given constraints or we may be given f(x) and have to find g(x) so that we have to maximize

$$\int_X (\log g(x))\, f(x)\, dx = \int_X \log g(x)\, df(x), \quad (26)$$

where F(x) is the cumulative distribution function of X. We have shown that maximization of Equation (26) corresponds to maximization of the likelihood function. Thus, maximum likelihood principle can be regarded as a special case of MDI principle.

Entropy optimization principle and Gauss's principle of minimum interdependence (PMI)

If the probability distributions of the individual random variables are included in the set of constraints, as the marginal probability distributions of the joint probability distribution, the PMI is equivalent to the maximum entropy principle (MEP) which is also a particular case of Kullback and Leibler's (1951), MDI principle if a priori joint probability density function is the independent product density of n individual variables.

According to Guass's principle of PMI if we know the density functions of n random variables and some mixed

moments of these variables, we should choose that the joint density function for these variables which is as close to independence as possible subject to the given constraints that is, the joint density function should be as close as possible to the product of the density functions of the independent variables subject to joint moments having the prescribed values. Thus, Kullback and Leibler's MDI and Gauss's principle of PMI are equivalent.

ESTIMATION OF PARAMETER WHEN INTERVAL PROPORTIONS ARE GIVEN

Let us consider a random variable X over the interval [a,b] and let the random sample be arranged in order as

$$a = x_0 < x_1 < x_2 < \ldots\ldots\ldots < x_i < x_{i+1} < \ldots\ldots\ldots < x_n < x_{n+1} = b \quad (27)$$

So that the interval [a, b] is divided into (n +1) subintervals and Q_0, Q_1... Q_n are the given proportions of the population in these (n + 1) subintervals. Let us define a probability function over subinterval (x_i, x_{i+1}) as

$$Pi = \int_{x_i}^{x_{i+1}} f(x, \theta)\, dx, \qquad i = 0, 1, 2, \ldots\ldots n, \quad (28)$$

where θ is the population parameter. Thus, (P_0, P_1, ..., P_n) gives us a probability distribution depending on θ. Now, we have to choose parameter θ such that P_0, P_1,..... P_n are as close as possible to given Q_0, Q_1,....., Q_n. This can be achieved by minimizing the measure of cross entropy or directed divergence. We can make use of any measure of cross entropy that gives rise to a convex function of θ. But here, we minimize the Kullback Leibler measure of cross entropy,

$$D(Q: P) = \sum_{i=0}^{n} Q_i \log \frac{Q_i}{P_i} =$$

$$\sum_{i=0}^{n} Q_i \log Q_i - \sum_{i=0}^{n} Q_i \log P_i \cdot \quad (29)$$

Minimization of Equation (29) is same as maximization of

$\sum_{i=0}^{n} Q_i \log P_i$. So, we have to maximize

$$\sum_{i=0}^{n} Q_i \log P_i = \int_{x_i}^{x_{i+1}} Q_i \log \int_{x_i}^{x_{i+1}} f(x_i, \theta)\, dx \quad (30)$$

This principle have wide applications in estimating parameters when interval proportions are given to us, e.g. proportions of students in different intervals of marks obtained, proportion of failed equipments in different intervals of time etc. Let us consider the case when $f(x_i, \theta)$, functional form of distribution is exponentially distributed with unknown parameter θ. Then, Equation (30) reduces to maximize:

$$\phi = \sum_{i=0}^{n} Q_i \log \int_{x_i}^{x_{i+1}} \theta e^{-x\theta} dx$$

$$= \sum_{i=0}^{n} Q_i \log(e^{-x_{i+1}\theta} + e^{-x_i\theta}) \quad (31)$$

The above principle is illustrated in the following example having randomly generated population data.

Example

Let us consider a randomly generated population of size 50 (from exponential distributed with mean = 20) with interval proportions as:

Intervals:	0-10	10-20	20-30	30-40	40-50	>75
Frequency:	19	13	4	4	7	3
Q_i=Proportion:	0.38	0.26	0.08	0.08	0.14	0.06

Here $x_0=0$, $x_1=10$, $x_2=20$, $x_3=30$, $x_4=40$, $x_5=60$, $x_6=\infty$.

We choose θ which maximizes Equation (31), that is,

$$\phi(\theta) = \sum_{i=0}^{n} Q_i \log(e^{-x_{i+1}\theta} + e^{-x_i\theta}) = 0.38 \log (1-e^{-10\theta}) + 0.26 \log (e^{-10\theta} - e^{-20\theta}) + 0.08 (e^{-20\theta} - e^{-30\theta}) + 0.08 (e^{-30\theta} - e^{-40\theta}) + 0.14(e^{-40\theta} - e^{-50\theta}) + 0.06\, e^{-50\theta}$$
$$= -0.26 \times 10\theta - 0.08 \times 20\theta - 0.08 \times 30\theta - 0.14 \times 40\theta - 0.06 \times 50\theta + (0.38 + 0.26 + 0.08 + 0.08 + 0.14) \log (1-e^{-10\theta}) = -15.2\theta + 0.94 \log (1-e^{-10\theta}) \quad (32)$$

To maximize Equation (32), differentiate it with respect to θ and put the resultant form equal to zero, we get

$$\phi'(\theta) = -15.2 + \frac{0.94 \times 10 e^{-10\theta}}{1-e^{-10\theta}} = 0$$

$$\Rightarrow \quad 9.4\, e^{-10\theta} = 15.2 - 15.2 e^{-10\theta}$$

$$\Rightarrow \quad 24.6\, e^{-10\theta} = 15.2.$$

Taking log both sides, we get:

$$\hat{\theta} = \frac{1}{10}\log\frac{24.6}{15.2}$$

$$mean = \frac{1}{\hat{\theta}} \cong 20.77$$

The estimated value of the parameter is quite close to the population parameter value that is, we have small bias. Further, we can study the asymptotic behaviour of the estimator.

Conclusion

Entropy optimization principles and their applications in statistics is mainly the study of two well known entropy optimization principles namely maximum entropy principle and Kullback-Leibler minimum information principle and their generalizations by considering some generalized measures of entropy and cross entropy. These principles have found wide applications in various branches of science and engineering. We have presented a critical appraisal of classical statistical methods vis-à-vis entropy optimization principles in parameter estimation. Further, parameter estimation methods using entropy optimization principles have been discussed with illustrations and examples.

REFERENCES

Akaike H (1971). Information-theoretical considerations on estimation problems. Inf. Cont. 19(3):181-194.

Burg JP (1972). The relationship between maximum entropy spectra and maximum likelihood spectra. In D. G. Childers, Editor, Modern Spectral Analysis. pp. 130-131.

Fisher RA (1921). On the mathematical foundations of theoretical Statistics. Phil. Trans. Roy. Soc. 222(A):309-368.

Hadgiwas N (1981). The maximum entropy principle as a consequence of the principle of Laplace. J. Stat. Phy. 26:807-815.

Jaynes ET (1957). Information theory and statistical mechanics." Phy. Rev. 106:620-630.

Kapur JN (1989). Maximum Entropy Models in Science in Engineering. Wiley Eastern, New Delhi. P. 244.

Kullback S, Leibler RA (1951). On Information and Sufficiency. Ann. Math. Stat. 22:79-86.

Kumar P (2001). Entropy Optimization Principles and Their Applications, Thesis submitted to HAU, Hisar, India. P. 95.

Lind NC, Solana V (1988). Cross Entropy Estimation of Random Variables with Fractile constraints. Paper no. 11, Institute for Risk Research, University of Waterloo, Canada. pp. 34-45.

Shannon CE (1948). A Mathematical Theory of Communication. Bell System Tech. J. 27:379-423.