International Journal of Theoretical and Applied Sciences **4**(1): 33-35(2012)

ISSN No. (Print) : 0975-1718
ISSN No. (Online) : 2249-3247

Entropy Optimization and Its Application to Parametric Estimation

A. Ray* and S.K. Majumder**

**Department of Mathematics, Rajyadharpur Deshbandhu Vidyapith, Serampore, Hooghly, (WB), India*

***Department of Mathematics, Bengal Engineering and Science University (B.E.S.U), Shibpur, Howrah, (WB), India*

(Received 12 April 2012 Accepted 28 April 2012)

ABSTRACT : In the present paper we will establish a fundamental result which has been used in the principle of minimum cross entropy application to the estimation of parameter. Our another result is by using Havrda-Charvat measure of entropy which has been described by Kullback's minimum cross entropy. At the end of this paper we have represented some useful inequalities by using different measures of entropy.

Keywords: Shannon's measure of entropy, Havrda-Charvat α-entropy, Lind and Solanas principle, Berg measure, Csizer's measure, Euler-Lagranges equation of calculus of variation.

1. INTRODUCTION

Lind and Solona [2] introduced the principle of minimum cross-entropy in two stages. Vinocha and Singla [3] in their paper used an important result which also has been used by [2]. Our discussion is as follows:

1.1. Discussion: As discussed in [1], let the random sample be $x_0, x_1, x_2, ..., x_n, x_{n+1}$ without loss of generality we assume that

$$x_0 < x_1 < \cdots < x_n < x_{n+1} \qquad ...(1)$$

In the first stage we assume θ to be known. Since the area under the density function $g(x, \theta)$ in the $(n + 1)$ subintervals are not equal if possible we choose another density function $h(x, \theta)$ that has an area $1/(n + 1)$ in each subinterval and that is close to $g(x, \theta)$ as possible i.e. we choose $g(x, \theta)$ to minimize,

$$\int_{x_0}^{x_{n+1}} h(x,\theta)\ln\frac{h(x,\theta)}{g(x,\theta)}dx \qquad ...(2)$$

Subject to the constraints

$$\int_{x_i}^{x_{i+1}} h(x,\theta)dx = \frac{1}{n+1},\ i = 0, 1, 2, ..., n \qquad ...(3)$$

This gives $$\frac{h(x,\theta)}{g(x,\theta)} = \frac{1}{k_i},\quad x_i < x < x_{i+1} \qquad ...(4)$$

Where, $$\int_{x_i}^{x_{i+1}} g(x,\theta)dx = \frac{k_i}{n+1} = P_i \qquad ...(5)$$

In the second stage we choose θ in such a way as to make the two distributions with density functions. $g(x, \theta)$ and $h(x, \theta)$ as close as possible. For this we minimize either.

$$\sum_{i=0}^{n}\int_{x_i}^{x_{i+1}} g(x,\theta)\ln\frac{g(x,\theta)}{h(x,\theta)}dx \text{ or, } \sum_{i=0}^{n}\int_{x_i}^{x_{i+1}} h(x,\theta)\ln\frac{h(x,\theta)}{g(x,\theta)}dx \qquad ...(6)$$

subject to (3) gives us to minimize

$$\sum_{i=0}^{n} k_i \ln k_i \quad \text{or,} -\sum_{i=0}^{n}\ln k_i \qquad ...(7)$$

which are basically Shannon's and Berg measure of entropy respectively.

II. ANALYSIS

Here we will show below three important result using Lagrange's equation of calculus of variation.

2.1. Result: We will show minimizing either of the measures

$$\int_{x_0}^{x_{n+1}} g(x,\theta)\phi\left[\frac{h(x,\theta)}{g(x,\theta}\right]dx \text{ or } \int_{x_0}^{x_{n+1}} h(x,\theta)\phi\left[\frac{g(x,\theta)}{h(x,\theta}\right]dx \qquad ...(8)$$

where $\phi(.)$ as a twice differentiable convex function with $\phi(1) = 0$ subject to (3) gives us (4) as it was described earlier.

2.1.1. Proof: 1st Part

Let the Lagrangian function be

$$L \equiv g(x,\theta)\phi\left[\frac{h(x,\theta)}{g(x,\theta)}\right] - \lambda\left\{h(x,\theta)\} - \frac{1}{n+1}\right\}$$

$$\therefore \frac{\partial L}{\partial h} = g(x,\theta)\phi'\left[\frac{h(x,\theta)}{g(x,\theta)}\right].\frac{1}{g(x,\theta)} - \lambda$$

Euler-Lagranges equation gives $$\lambda = \phi'\left[\frac{h(x,\theta)}{g(x,\theta)}\right] \qquad ...(9)$$

$\because$ Since $\phi(.)$ is a convex function so $\phi(.)$ is a monotonic increasing function and so $\phi'^{-1}(.)$ exists

$\therefore$ from (9) $\dfrac{h(x,\theta)}{g(x,\theta)} = \phi'^{-1}(\lambda) = \dfrac{1}{k_i}$ (say) ...(10)

Which is same as (4)

2.1.2. Proof : 2nd Part

Let $L \equiv h(x,\theta)\phi\left[\dfrac{g(x,\theta)}{h(x,\theta)}\right] - \lambda\left\{h(x,\theta) - \dfrac{1}{n+1}\right\}$

$\therefore$ Euler-Lagrange's equation gives,

$$\lambda = \phi\left[\frac{g(x,\theta)}{h(x,\theta)}\right] - \frac{g(x,\theta)}{h(x,\theta)}\phi'\left[\frac{g(x,\theta)}{h(x,\theta)}\right] \quad ...(11)$$

$$\therefore \lambda = \phi(1) + \left(\frac{g(x,\theta)}{h(x,\theta)} - 1\right)\phi'\left[\frac{g(x,\theta)}{h(x,\theta)}\right] + \frac{1}{2}\left(\frac{g(x,\theta)}{h(x,\theta)} - 1\right)^2 \phi''\left[\frac{g(x,\theta)}{h(x,\theta)}\right] - \frac{g(x,\theta)}{h(x,\theta)}\phi'\left[\frac{g(x,\theta)}{h(x,\theta)}\right]$$

$$\therefore \quad \lambda = -\phi'\left[\frac{g(x,\theta)}{h(x,\theta)}\right] + \frac{1}{2}\left(\frac{g(x,\theta)}{h(x,\theta)} - 1\right)^2 \phi''\left[\frac{g(x,\theta)}{h(x,\theta)}\right]$$

$\therefore \quad \lambda + \phi'\left[\dfrac{g(x,\theta)}{h(x,\theta)}\right] > 0$ [$\because \phi''(.) > 0$ as $\phi(x)$ is convex]

Let, $\phi'\left[\dfrac{g(x,\theta)}{h(x,\theta)}\right] + \lambda = \lambda_1$ (say)

$\therefore \quad \phi'\left[\dfrac{g(x,\theta)}{h(x,\theta)}\right] = \lambda_1 - \lambda$ or, $\dfrac{g(x,\theta)}{h(x,\theta)} = \phi^{-1}(\lambda_1 - \lambda) = \dfrac{1}{k_i}$ (say) ...(12)

which is the desired result as in equation (4)

If we consider $\phi(x) = \ln\left(\dfrac{1}{x}\right)$ then $\phi'(x) = -\dfrac{1}{x}$ and $\phi''(x) = +\left(\dfrac{1}{x^2}\right) > 0$

$\therefore \phi\left(\dfrac{g(x,\theta)}{h(x,\theta)}\right) = \ln\left(\dfrac{h(x,\theta)}{g(x,\theta)}\right)$ which when put into (8) gives us (2).

2.2 Result: Here we will minimize the cross entropy of $g(x, \theta)$ relative to $h(x, \theta)$ and of $h(x, \theta)$ relative to $g(x, \theta)$ by using the Havrda-Charvat measure of entropy of the second stage of the method discussed in the introduction and we will also find the result when $\alpha \to 1$.

Let us minimize the Havrda-Charvat measure of cross entropy

$$\frac{1}{\alpha-1}\left[\int_{x_0}^{x_{n+1}}\left\{g(x,\theta)^{\alpha} h(x,\theta)^{1-\alpha} dx\right\} - 1\right]$$

$$\text{or, } \frac{1}{\alpha-1}\left[\int_{x_0}^{x_{n+1}}\left\{h(x,\theta)^{\alpha} g(x,\theta)^{1-\alpha} dx\right\} - 1\right] \quad ...(13)$$

subject to (3)

Let the Lagrangian be

$$L \equiv \frac{1}{\alpha-1}\left(g(x,\theta)^{\alpha} h(x,\theta)^{1-\alpha} - 1\right) + \lambda\left(\left(h(x,\theta) - \frac{1}{n+1}\right)\right)$$

$$\frac{\partial L}{\partial h(x,\theta)} = 0 \text{ gives } \lambda = \left(\frac{g(x,\theta)}{h(x,\theta)}\right)^{\alpha} \therefore \frac{g(x,\theta)}{h(x,\theta)} = \lambda^{\frac{1}{\alpha}} = \frac{1}{k_i} \text{ (say)} \quad ...(14)$$

$\therefore h(x,\theta) = \dfrac{g(x,\theta)}{k_i}$ $\therefore$ $\displaystyle\int_{x_i}^{x_{i+1}} h(x,\theta)dx = \frac{1}{n+1}$ gives

$\displaystyle\int_{x_i}^{x_{i+1}} g(x,\theta)dx = \frac{k_i}{n+1}$ which is basically equation ...(5)

Again when $L \equiv \dfrac{1}{\alpha-1}\left(h(x,\theta)^{\alpha} g(x,\theta)^{1-\alpha} - 1\right) + \lambda\left(h(x,\theta) - \dfrac{1}{n+1}\right)$

$$\therefore \frac{\partial L}{\partial h(x,\theta)} = 0 \Rightarrow \lambda = \frac{\alpha}{1-\alpha}\left(\frac{g(x,\theta)}{h(x,\theta)}\right)^{1-\alpha}$$

$$\therefore \frac{g(x,\theta)}{h(x,\theta)} = \left(\frac{1-\alpha}{\alpha}\right)^{\frac{1}{1-\alpha}} \lambda^{\frac{1}{1-\alpha}} = \frac{1}{\mathrm{k_i}} \text{ (say)} \quad ...(15)$$

2.2.1. Now from (13) we have to minimize

$$\frac{\left\{\sum_{i=0}^{n}\int_{x_i}^{x_{i+1}} h(x,\theta)\left(\frac{g(x,\theta)}{h(x,\theta)}\right)^{\alpha} dx\right\} - 1}{\alpha-1} \quad ...(16)$$

$$\text{Or, min } \frac{\left\{\sum_{i=0}^{n}\int_{x_i}^{x_{i+1}} h(x,\theta)k_i^{\alpha} dx\right\} - 1}{\alpha-1} \quad ...(17)$$

$$\textit{i.e.} \text{ min } \frac{\left\{\sum_{i=0}^{n} k_i^{\alpha}.\frac{1}{n+1}\right\} - 1}{\alpha-1} \quad \left[\because \int_{x_i}^{x_{i+1}} h(x,\theta)dx = \frac{1}{n+1}\right] \quad ...(18)$$

$$\text{i.e. min } \frac{\frac{1}{n+1}\left\{\sum_{i=0}^{n} k_i^{\alpha}\right\} - 1}{\alpha-1} \quad ...(19)$$

Now takin $\alpha \to 1$ we get

Min $\sum k_i \ln k_i$

$$\text{i.e. maximize } -\sum_{i=0}^{n} k_i \ln k_i \quad ...(20)$$

which is Shannon's measure of entropy,

2.2.2. Again we will minimize

$$\frac{\left\{\sum_{i=0}^{n}\int_{x_i}^{x_{i+1}} g(x,\theta)\left(\frac{h(x,\theta)}{g(x,\theta)}\right)^{\alpha} dx\right\}-1}{\alpha-1} \quad ...(21)$$

$$\text{i.e. min } \frac{\left\{\sum_{i=0}^{n}\int_{x_i}^{x_{i+1}} g(x,\theta)\frac{1}{k_i^{\alpha}} dx\right\}-1}{\alpha-1} \quad ...(22)$$

$$\text{i.e. min } \frac{\left\{\sum_{0}^{n}\frac{1}{k_i^{\alpha}}.\frac{k_i}{n+1}\right\}-1}{\alpha-1}$$

$$\left[\because \int_{x_i}^{x_{i+1}} g(x,\theta)dx = \frac{k_i}{n+1}, x_i < x < x_{i+1}\right] \quad ...(23)$$

$$\text{i.e. min } \frac{\frac{1}{n+1}\left\{\sum_{0}^{n} k_i^{1-\alpha}\right\}-1}{\alpha-1} \quad ...(24)$$

Taking $\alpha \to 1$, we get

$$\text{min } -\sum_{i=0}^{n}\ln k_i \quad ...(25)$$

$$\text{i.e. maximize } \sum_{i=0}^{n}\ln k_i \quad ...(26)$$

which is the Bergs measure of entropy.

2.3. Result : Now we will show some inequality which are important result too. At first let us consider the beta distribution, now from [4] we know it is maximum when m = 1, n = 1 and S_{max} is a concave function of moment values we will prove.

2.3.1. ln $B(m, n) + m + n - 2 > 0$ ...(27)

We know that Shannon and Berg both the measure of entropy are concave, permutationally symmetric and maximum for the uniform distribution, Shannon's measure of entropy is defined for zero probabilities also and always positive (≥ 0), but Berg's measure is not finite for degenerate probability distribution and is always negative (≤ 0).

Now let us consider $f(x,m,n) = \frac{1}{B(m,n)} x^{m-1}(1-x)^{n-1}$ ...(28)

∴ Bergs measure of entropy gives us

$$\int_0^1 \ln f(x)dx < 0 \quad ...(29)$$

$$i.e. \ -\ln B(m,n) + (m-1)\int_0^1 \ln x dx + (n-1)\int_0^1 \ln(1-x)dx < 0 \quad ...(30)$$

$$i.e. \ -\ln B(m,n) + \{(m-1)+(n-1)\}\int_0^1 \ln x dx < 0 \quad ...(31)$$

$$i.e. \ -\ln B(m,n) + (m+n-2)(-1) < 0 \quad ...(32)$$

$$i.e. \ \ln B(m,n) + m + n - 2 > 0 \quad ...(33)$$

When $m = 1 = n$ then In $B(m,n) + m + n - 2 = 0$ ...(34)

$\therefore \ln B(m,n) + m + n - 2 \geq 0$ [Proved].

2.3.2. Now we will finish with the following result

$$\frac{1}{1-\alpha}\ln\frac{[B(m,n)]^{\alpha}}{B(m\alpha-\alpha+1, n\alpha-\alpha+1)} \geq 0 \quad ...(35)$$

Let us consider equation (28)

$$\therefore f^{\alpha}(x) = \frac{x^{m\alpha-\alpha}(1-x)^{n\alpha-\alpha}}{[B(m,n)]^{\alpha}} \quad ...(36)$$

Now Havrda-Charvat α entropy gives us

$$\frac{1}{\alpha-1}\left[\ln\int_0^1 f^{\alpha}(x)dx\right] = \frac{1}{\alpha-1}\left[\ln\int_0^1 \frac{x^{m\alpha-\alpha}(1-x)^{n\alpha-\alpha}dx}{[B(m,n)]^{\alpha}}\right] \quad ...(37)$$

$$= \frac{1}{\alpha-1}\left[\ln\frac{\int_0^1 x^{(m\alpha-\alpha+1)-1}(1-x)^{(n\alpha-\alpha+1)-1}dx}{[B(m,n)]^{\alpha}}\right]$$

$$= \frac{1}{\alpha-1}\left[\ln\frac{B(m\alpha-\alpha+1, n\alpha-\alpha+1)}{[B(m,n)]^{\alpha}}\right]$$

$$= \frac{1}{1-\alpha}\left[\ln\frac{[B(m,n)]^{\alpha}}{B(m\alpha-\alpha+1, n\alpha-\alpha+1)}\right] \quad ...(38)$$

$$\geq 0\left[\because \frac{1}{\alpha-1}\int_0^1 f^{\alpha}(x)dx \geq 0\right] \quad \text{[Proved]}$$

III. CONCLUSION

The results in this paper show that we get the same result when we use $\phi(.)$ a convex function twice differentiable which is basically Csiszer's function. As it has been already showed in [1] that Kull-back Lieber's measure is a special case of Csiszer's measure and which is also a limiting case of the Havrda-Charvat measure of entropy.

REFERENCES

[1] Kapur, J.N. and Kesavan, H.K. Entropy Optimization principles with applications, (Academic Press, London, 1992).

[2] Lind, N.C. and Solana, V. Cross entropy estimation of random variable with fractile constraints, Paper 11, Institute of Risk Research, University of Waterloo (1988).

[3] Vinocha, O.P. and Singla Prem, On lines principle of minimum information, *International Journal of Theoritical & Applied Sciences*, 22-26 (2010).

[4] Kapur, J.N., Maximum-entropy models in science and engineering, (Wiley Eastern Limited, New Delhi, 1989).