Aligarh Journal of Statistics
Vol. 32 (2012), 115-122

MOMENTS OF THE MAXIMUM LIKELIHOOD ESTIMATES OF SOME INFORMATION MEASURES

R. Ahmad

ABSTRACT

The paper contains exact expressions for the first two moments of the maximum likelihood estimates of the combined information measure introduced by Kaur (1983). Further, If utility Schemes are also associated with the probability schemes, then one gets the generalized combined 'useful' information measure. Exact expressions for the first two moments of the maximum likelihood estimates of this measure are also derived.

1. INTRODUCTION

Shannon's (1948) entropy measure

$$I(P) = -\sum_{i=1}^{k} p_i \log p_i \tag{1.1}$$

of the probability distribution $P = (p_1, p_2, ...p_k), p_i \geq 0, \sum_{i=1}^{k} p_i = 1,$ was generalized by Belis and Guiasu (1968) defined 'useful' information as:

$$I(U;P) = -\sum_{i=1}^{k} u_i p_i \log p_i \tag{1.2}$$

by attaching a utility $u_i (> 0)$ to the event occurring with probability p_i. Following the same idea Emptoz (1976) and Sharma et al. (1978) generalized the Havrda- Charvat (1967) entropy measure

$$I^{\alpha}(P) = (2^{1-\alpha} - 1)^{-1} \sum_{i=1}^{k} (p_i^{\alpha} - p_i), \alpha > 0 , \tag{1.3}$$

through the introduction of utilities u_i to

$$I^{\alpha}(U,P) = (2^{1-\alpha} - 1)^{-1} \sum_{i=1}^{k} u_i (p_i^{\alpha} - p_i), \alpha > 0 , \tag{1.4}$$

to which function they also gave the name of 'useful' information.

Kaur (1983) defined a combined information measure of two probability distributions

$$P = (p_1, p_2, \ldots\ldots, p_k), p_i > 0, \sum_{i=1}^{k} p_i = 1$$

and $Q = (q_1, q_2, \ldots\ldots, q_k), q_i > 0, \sum_{i=1}^{k} q_i = 1,$

of a set of k mutually exclusive and exhaustive events as:

$$I^{\alpha,\beta,\gamma}(P;Q) = (2^{\alpha-\beta+\gamma} - 1)^{-1} \sum_{i=1}^{k} (p_i^{\alpha} q_i^{\beta-\alpha} - p_i), \beta \neq \alpha + \gamma \tag{1.5}$$

If we attach utilities $U = (u_1, u_2, \ldots\ldots, u_k), u_i > 0$ and $V = (v_1, v_2, \ldots\ldots, v_k), v_i > 0,$ to the probability distributions P and Q, we get the measure

$$I^{\alpha,\beta,\gamma}(W;P,Q) = (2^{\alpha-\beta+\gamma} - 1)^{-1} \sum_{i=1}^{k} w_i (p_i^{\alpha} q_i^{\beta-\alpha} - p_i), \beta \neq \alpha + \gamma \tag{1.6}$$

where $w_i = u_i v_i$, which may call the combined 'useful' information measure in analogy with Belis and Guiasu (1968).

Putting $\beta = 1, \gamma = 0$ and then limiting α tends to 1, Kaur's measure yields

$$I(P;Q) = \sum_{i=1}^{k} p_i \log_2 \left(\frac{p_i}{q_i} \right) \tag{1.7}$$

which is Kullback's (1959) measure of relative information that the distribution $P = (p_1, p_{2,} \ldots\ldots, p_k)$ provides about the distribution $Q = (q_1, q_2, \ldots\ldots, q_k)$.

We may, therefore, regard Kaur's measure as a generalized measure of the relative information that the distribution P provides about the distribution Q. If P is the distribution determined on the basis of an experiment, then this measure may be considered as measure of the information on Q furnished by the experiment.

Moments of the statistical estimates of Shannon entropy and $I(U;P)$ have been studied by several authors. The work in this direction started with the papers of Miller (1955) and Basharin (1959), who derived asymptotic mean and variance of an of an M.L.E. of Shannon entropy. Exact expressions for these moments have been obtained by Rogers and Green (1955) and Hutcheson and Shenton (1974). Further in a work, Sharma et al. (1977) derived asymptotic mean and variance of M.L.E. of $I(U;P)$. Exact expressions for these moments have been obtained by Sharma and Mohan (1978, 79).

Applications of $I^{\alpha,\beta,\gamma}(P;Q)$ and $I^{\alpha,\beta,\gamma}(W;P,Q)$ have been made recently in the theory of questionnaire (refer Picard 1972) and in the analysis of business and Accounting data (refer Sharma et al. (1976,1978). Thus there arose a need of further studying statistical estimators of $I^{\alpha,\beta,\gamma}(P;Q)$ and $I^{\alpha,\beta,\gamma}(W;P,Q)$.

We have obtained in section 2 the first two moments of the maximum likelihood estimates (M.L.E.) of $I^{\alpha,\beta,\gamma}(W;P,Q)$. By putting $w_i = 1$, we also get first two moments of the *m.l.e.* of $I^{\alpha,\beta,\gamma}(P;Q)$ which in section 3.

2. MOMENTS OF THE M.L.E. OF $I^{\alpha,\beta,\gamma}(W;P,Q)$

Let $n_i, i = 1,2,.......,k$ be the frequency of the occurrence of the event in a random sample of size N, such that $\sum_{i=1}^{k} n_i = N$. Then the *m.l.e.* of $I^{\alpha,\beta,\gamma}(W;P,Q)$ is given by

$$\hat{I}^{\alpha,\beta,\gamma}(W;P,Q) = (2^{\alpha-\beta+\gamma}-1)^{-1}\sum_{i=1}^{k} w_i(\hat{p}_i^{\alpha} q_i^{\beta-\alpha} - \hat{p}_i), \beta \neq \alpha+\gamma \tag{2.1}$$

where

$\hat{p} = \frac{n_i}{N}$ is the *m.l.e.* of $p_i, i = 1,2,.......,k$

and $n_i's$ follow the multinomial distribution. Then (2.1) can be written as:

$$\hat{I}^{\alpha,\beta,\gamma}(W;P,Q) = 2^{\alpha-\beta+\gamma} - 1)^{-1}\left[N^{-\alpha}\left(\sum_{i=1}^{k} w_i n_i^{\alpha} q_i^{\beta-\alpha} - \frac{1}{N}\sum_{i=1}^{k} w_i n_i\right)\right]. \tag{2.2}$$

We consider

$$M(t) = M(t_1, t_2, \ldots\ldots t_k; n_1, n_2, \ldots\ldots, n_k) = \left(p_1 e^{t_1} + p_2 e^{t_2} + \ldots\ldots + p_k e^{t_k}\right)^N \tag{2.3}$$

where $\sum_{i=1}^{k} p_i = 1$ and $\sum_{i=1}^{k} n_i = N$.

Let $x_i = \sum_{r=1, r\neq 1}^{k} p_r e^{t_r}$ in (2.3), we get

$$M(t) = (x_i + p_i e^{t_i})^N$$

$$= \sum_{a=0}^{N} \binom{N}{a} \left(p_i e^{t_i}\right)^a x_i^{N-a}. \tag{2.4}$$

Let us consider the parameter is a natural number and then differentiating (2.4) α times w.r.t. t_i we get

$$\frac{\partial^{\alpha} M(t)}{\partial t_i^{\alpha}} = \sum_{a=0}^{N} \binom{N}{a} p_i^a x_i^{N-a} a^{\alpha} e^{t_i a} \tag{2.5}$$

Setting each $t_i = 0$ in (2.5) gives

$$E(n_i^{\alpha}) = \sum_{a=0}^{N} \binom{N}{a} p_i^a (1-p_i)^{N-a} a^{\alpha} \tag{2.6}$$

$$= (1-p_i)^N \sum_{a=0}^{N} \binom{N}{a} p_i^a (1-p_i)^{-a} E^a 0^{\alpha}$$

(E is the shift operator defined by $E^r u_x = u_{x+r}$)

$$= (1 - p_i)^N (1 + \frac{p_i}{1 - p_i} E)^N 0^\alpha$$

$$= (1 + p_i \Delta)^N 0^\alpha \qquad (2.7)$$

where Δ is the usual different operator; and $\Delta^r 0^\alpha$ is the $r-th$ difference operator of 0^α.

Now, differentiating (2.5) α times w.r.t. $t_j (i \neq j)$ we have

$$\frac{\partial^{2\alpha} M(t)}{\partial^\alpha_{t_i} \partial^\alpha_{t_j}} = \sum_{a=0}^{N} \binom{N}{a} p_i^a a^\alpha e^{t_i a} \frac{\partial^\alpha (x_i^{N-a})}{\partial^\alpha_{t_j}}$$

$$= \sum_{a=0}^{N} \binom{N}{a} p_i^a a^\alpha e^{t_i a} \sum_{b=0}^{N-a} \binom{N-a}{b} x_j^{N-a-b} p_j^b \qquad (2.8)$$

where $x_j = \sum_{r=1, r \neq i, j}^{k} p_r e^{t_r}$

setting each $t_i = 0$ in (2.8), we get

$$E(n_i^\alpha n_j^\alpha)$$

$$= \sum_{a=0}^{N} \binom{N}{a} p_i^a a^\alpha e^{t_i a} \sum_{b=0}^{N-a} \binom{N-a}{b} x_{j*}^{N-a-b} p_j^b (ab)^\alpha, \; (x_{j*} = 1 - p_i - p_j)$$

$$= \sum_{a=0}^{N} \binom{N}{a} p_i^a (x_{j*} + p_j E^a)^{N-a} 0^\alpha$$

$$= \sum_{a=0}^{N} \binom{N}{a} p_i^a (1 - p_i - p_j + p_j E^a)^{N-a} 0^\alpha \qquad (2.9)$$

Now taking expectations on both sides of (2.2), we get

$$E\left[\hat{I}^{\alpha,\beta,\gamma}(W; P, Q)\right] = (2^{\alpha-\beta+\gamma} - 1)^{-1} \left[N^{-\alpha} \sum_{i=1}^{k} w_i E(n_i^\alpha) q_i^{\beta-\alpha} - \overline{W} \right] \quad (2.10)$$

where $\overline{W} = \sum_{i=1}^{k} w_i p_i$.

Invoking (2.7), this proves

$$E\left[\hat{I}^{\alpha,\beta,\gamma}(W;P,Q)\right]$$
$$= (2^{\alpha-\beta+\gamma}-1)^{-1}\left[N^{-\alpha}\sum_{i=1}^{k} w_i (1+p_i\Delta)^N 0^\alpha q_i^{\beta-\alpha} - \overline{W}\right] \qquad (2.11)$$

Further, we have

$$\left[\hat{I}^{\alpha,\beta,\gamma}(W;P,Q)\right]^2$$
$$= N^{-2\alpha}(2^{\alpha-\beta+\gamma}-1)^{-2}\sum_{i=1}^{k}\left\{\begin{array}{l} w_i^2 n_i^{2\alpha} q_i^{2\beta-2\alpha} + \sum\limits_{\substack{i=1\\ j\neq i}}^{k} w_i w_j n_i^\alpha n_j^\alpha q_i^{\beta-\alpha} \\ + \dfrac{N^{2\alpha}}{k}\overline{w}^2 - 2N^\alpha \overline{w} w_i n_i^\alpha q_i^{\beta-\alpha} \end{array}\right\}$$
$$(2.12)$$

Taking expectation both the sides and using (2.7) and (2.9), (2.12) becomes

$$E\left[\hat{I}^{\alpha,\beta,\gamma}(W;P,Q)\right]^2.$$
$$= N^{-2\alpha}(2^{\alpha-\beta+\gamma}-1)^{-2}\sum_{i=1}^{k}\left\{ w_i^2 \ (1+p_i\Delta)^N 0^{2\alpha} q_i^{2\beta-2\alpha}.\right.$$
$$+\frac{N^{2\alpha}}{k}\overline{w}^2 - 2N^\alpha \overline{w} w_i (1+p_i\Delta)^N 0^\alpha q_i^{\beta-\alpha}$$
$$\left. + \sum_{a=0}^{N}\binom{N}{a}\sum_{\substack{j=1\\ j\neq i}}^{k} w_i w_j p_i^a (1-p_i-p_j+p_j E^a)^{N-a} 0^\alpha \right\}$$
$$(2.13)$$

3. FIRST TWO MOMENTS OF THE M.L.E. OF $I^{\alpha,\beta,\gamma}(W;P,Q)$

For $w_i = 1, (i = 1,2,......,k), I^{\alpha,\beta,\gamma}(W;P,Q)$ reduces to $I^{\alpha,\beta,\gamma}(P,Q)$. Hence, the first two moments of the $m.l.e.$ M.L.E. of to $I^{\alpha,\beta,\gamma}(P,Q)$ are obtained by putting $w_i = 1$ in (2.11) and (2.13) respectively, and are given by:

$$E\left[\hat{I}^{\alpha,\beta,\gamma}(P;Q)\right] = (2^{\alpha-\beta+\gamma}-1)^{-1}\left[N^{-\alpha}\sum_{i=1}^{k}(1+p_i\Delta)^N 0^{\alpha} q_i^{\beta-\alpha} - 1\right] \quad (3.1)$$

and

$$E\left[\hat{I}^{\alpha,\beta,\gamma}(P;Q)\right]^2 = N^{-2\alpha}(2^{\alpha-\beta+\gamma}-1)^{-2}\sum_{i=1}^{k}\left\{(1+p_i\Delta)^N 0^{2\alpha} q_i^{2\beta-2\alpha}\right. .$$

$$+\frac{N^{2\alpha}}{k} - 2N^{\alpha}(1+p_i\Delta)^N 0^{\alpha} q_i^{\beta-\alpha}$$

$$\left. + \sum_{j=1}^{k}\sum_{a=0}^{N}\binom{N}{a} p_i^a (1-p_i-p_j+p_jE^a)^{N-a} 0^{\alpha}\right\} \quad (3.2)$$

Acknowledgement

The author is thankful to Professor D. D. Joshi, Ex, Pro-Vice -chancellor I.G.N.O.U. , New, Delhi, for his offering constructive suggestion during the preparation of this paper. He is also grateful to the Editor and the Referee for his valuable comments and suggestions that helped to revise the manuscript in its present form.

REFERENCES

Basharin, G.P (1959): On a statistical estimate for the entropy of a sequence of independent random variables. Theory Prob. App., **4**, 333-336.

Bilis, M. and Guiasu, S. (1968): A quantitative – qualitative measure of information in cybernetic systems. *IEEE Trans. Inform. Theory.* **4**, 593-594.

Emptoz , H. (1976): Information de-type β integrant un concept d' utilities. *CR Acad. Sci. Paris **28**, 2A,* 911-14.

Guiasu, S., and Picard, C.F. (1971): Borne inferieure de la longueur utile de certains codes, C. R. Acad. Sci. Paris, Ser. A **273**, 248-251.

Havrda, H. and Charvat, F. (1967): Quantification method of classification process: Concept of structural β- entropy. *Kybernatica,* 3, 30-35.

Hutcheson, k. and Shanton, L.R. (1974): Some moments of an estimate of Shannon's measure of information. *Comm. Statist.*, **3**, 89-94.

Kaur, P. (1983): A combined information measure. *Pure App. Math. Sci. vol. XVIII. No.* **1-2** , 59-65.

Kullback , S. (1958): *Information Theory and Statistics, Wiley New York.*

Miller, G.A. (1955): Note on the bias of information estimate. Information Theory in Psychology, H. Quastler ed. , The free Press, Glencoe, 95-100.

Picard, C.F. (1972): *Graphs et Questionnaires.* Gautheir-Villars, Paris.

Rogers, M. S. and Green, B. F. (1955): The moments of sample information when the alternatives are equally likely, Information theory in Psychology, H. Quastler ed., The Free Press, Glencoe, 101-107.

Shannon, C. E. (1948): The mathematical theory of communication. *Bell System Tech. Journal, 27,* 379-423.

Sharma, B.D., Mitter, J., Mohan, Man (1978): On measure of useful information. *Information and Control,* **39**, 323-336

Sharma, B.D., Mohan Man, and Dev, J. (1976): Financial statements: An information approach. Indian J. Accounting, **6**, 29-48

Sharma, B.D., Mohan, Man, and Mitter, J. (1977): On the mean and variance of an estimate of 'useful' information. Statistica, **37**, 427-438

Sharma, B.D., and Man Mohan (1978): An exact expression for the mean of an m.l.e. of "useful" Information. *Pure Appl. Math. Sci.*, **7**, 29-33.

Sharma, B.D. and Man, and Mohan (1979): Exact variance of an M.L.E. of 'useful' information. Statistica, **39**, 633-643.

Received : 13-04-2012
Revised : 15.10.2012

Rafiq Ahmad
Department of Statistics and Operations Research
Women's College
Aligarh Muslim University, Aligarh (India)
Email: dr.ahmadstats@gmail.com