SENRA Academic Publishers, British Columbia
Vol. 8, No. 1, pp. 2795-2799, February 2014
Online ISSN: 1920-3853; Print ISSN: 1715-9997

Canadian Journal of pure&applied sciences

Short Communication

APPLICATIONS OF INFORMATION MEASURES FOR THE STUDY OF GAUSSIAN DISTRIBUTIONS

*Om Parkash[1] and Mukesh[2]
Department of Mathematics, Guru Nanak Dev University, Amritsar 143005, India

ABSTRACT

Finding the relationships between information measures and statistical constants leads to the applicability of information theory to the field of statistics. In the existing literature of information theory, there are many well known information theoretic measures, each with its own merits, limitations and areas of applications. In the present communication, we have developed new generalized parametric divergence measure and provided its applications along with the other parametric information theoretic measure to the field of Statistics by establishing the relationships of information contents with some statistical constants of Gaussian distribution.

Keywords: Gaussian distribution, entropy, divergence, chi-square statistics, covariance matrix.

INTRODUCTION

In the literature, there exist certain analytical expressions for the entropy of univariate continuous distributions studied by Lazo and Rathie (1978) and Cover and Thomas (1991) whereas for multivariate distributions very few results have been provided by Ahmed and Gokhale (1989). Of course, Darbellay and Vajda (2000) developed a series of analytical expressions for the entropy and the mutual information of continuous multivariate distributions.

Another study Ginebra (2007) provided the applications of information measures to the field of statistics by stating a minimal set of requirements that must be satisfied by all such measures. By doing so, the author provided the links of information and uncertainty in a probability distribution. Kitsos and Toulias (2010) introduced a three-parameter generalized normal distribution to study the generalized measures of entropy and for this generalized normal measure, the Kullback-Leibler's (1951) information measure was evaluated to extend the well known result for the normal distribution. Parkash and Thukral (2010) emphasized that statistics is extensively used for the measurement of statistical constants whereas measures of information are used to study diversity and equitability. These two fields have been used independent of each other for data analysis. The authors developed the interrelations between the two and proved that statistical measures can be used as information measures.

Many researchers have provided axiomatic derivations of Shannon's (1948) measure of entropy, given by

$$H(P) = -\sum_{i=1}^{n} p_i \ln p_i \ , \qquad (1.1)$$

of a discrete-variate probability distribution $P = (p_1, p_2, \ldots, p_n)$ from different sets of plausible axioms where a great deal of mathematical sophistication and rigour has been exercised in the process. However, when it comes to the derivation of the corresponding measure

$$H[f(x)] = -\int_a^b f(x) \ln f(x) dx \ , \qquad (1.2)$$

which is called differential entropy, for the entropy of continuous-variate probability distribution with density function $f(x)$, all pretensions of rigour are given up and heuristic arguments are freely employed. The major argument in favour of (1.2) is not its rigorous derivation from clearly stated axioms, but the fact that it gives useful results. Similar arguments are used to deduce various measures of entropy for the continuous variate distributions. It is well known that the Gaussian distribution density has the greatest Shannon's (1948) entropy of all distribution densities. All the other distribution densities, having the same second order moments as the Gaussian distribution density, have smaller Shannon entropy than the Gaussian density.

In statistics, the entropy corresponds to the maximum likelihood method, in which Kullback-Leibler (1951) divergence measure given by

$$D(P;Q) = \sum_{i=1}^{n} p_i \ln \frac{p_i}{q_i} \ , \qquad (1.3)$$

*Corresponding author email: omparkash777@yahoo.co.in

connects Boltzmann-Shannon entropy and the expected log-likelihood function.

Recently, Parkash and Mukesh (2013) have developed the relations between parametric information measures and chi-square statistic by using the optimization principles. Further, Parkash and Mukesh (2012) have provided the applications of divergence measure by developing an optimizational principle for minimizing risk in portfolio analysis. Parkash and Mukesh (personal communication) have also investigated and introduced a new generalized measure of entropy and provided its applications to the field of queueing theory. This new entropy measure with real parameters α and β is given by

$$H_{\alpha,\beta}(P)=\frac{1}{\beta-\alpha}\left(\sum_{i=1}^{n}p_i^{\alpha-\beta+1}-1\right), \tag{1.4}$$

$\alpha\neq\beta, \beta<\alpha+1, -\infty<\alpha<\infty$.

Some other characterizations and generalization of the measures of entropy and directed divergence along with their detailed properties have been provided by Havrada and Charvat (1967) and Parkash and Mukesh (2012, 2011).

In the present communication, we have developed new generalized information measure and established the relationships between information measures and the statistical constants for Gaussian distribution densities.

2 A New generalized parametric measure of divergence

In this section, we propose a new generalized parametric measure of divergence for the probability distributions $P=\left\{(p_1,p_2,\ldots,p_n), p_i\geq 0, \sum_{i=1}^{n}p_i=1\right\}$ and $Q=\left\{(q_1,q_2,\ldots,q_n), q_i\geq 0, \sum_{i=1}^{n}q_i=1\right\}$ and study its properties. This new entropy measure with real parameters α and β is given by

$$D_{\alpha,\beta}(P;Q)=\frac{1}{\alpha-\beta}\left(\sum_{i=1}^{n}p_i^{\alpha-\beta+1}q_i^{\beta-\alpha}-1\right), \tag{2.1}$$

$\alpha\neq\beta, \beta<\alpha+1, -\infty<\alpha<\infty$.

We have $\lim_{\alpha\to\beta}D_{\alpha,\beta}(P;Q)=\sum_{i=1}^{n}p_i\ln\frac{p_i}{q_i}$, which shows that $D_{\alpha,\beta}(P;Q)$ is a generalization of Kullback-Leibler's (1951) measure of divergence.

For $\beta=1$, we get $D_{\alpha,1}(P;Q)=\frac{1}{\alpha-1}\left(\sum_{i=1}^{n}p_i^{\alpha}q_i^{1-\alpha}-1\right)$.

Again $D_{\alpha,\beta}(P;Q)$ is the generalization of Havrada-Charvat's (1967) divergence measure.

Next, to prove that the measure (2.1) is a valid measure of directed divergence, we study its essential properties as follows:

1) $D_{\alpha,\beta}(P;Q)$ is a continuous function of $p_1,p_2,\ldots,p_n$ and $q_1,q_2,\ldots,q_n$.
2) $D_{\alpha,\beta}(P;Q)\geq 0$ and vanishes if and only if $P=Q$.
3) We can deduce from condition (2) that the minimum value of $D_{\alpha,\beta}(P;Q)$ is zero.
4) We shall now prove that $D_{\alpha,\beta}(P;Q)$ is a convex function of both P and Q. This result is important in establishing the property of global minimum.

Let

$$D_{\alpha,\beta}(P;Q)=f(p_1,p_2,\ldots,p_n;q_1,q_2,\ldots,q_n)$$
$$=\frac{1}{\alpha-\beta}\left(\sum_{i=1}^{n}p_i^{\alpha-\beta+1}q_i^{\beta-\alpha}-1\right).$$

Thus

$$\frac{\partial f}{\partial p_i}=\frac{1}{\alpha-\beta}\left[(\alpha-\beta+1)p_i^{\alpha-\beta}q_i^{\beta-\alpha}\right],$$

$$\frac{\partial^2 f}{\partial p_i^2}=(\alpha-\beta+1)p_i^{\alpha-\beta-1}q_i^{\beta-\alpha}, \ \forall i=1,2,\ldots,n,$$

and $\frac{\partial^2 f}{\partial p_i\partial p_j}=0$, $\forall i,j=1,2,\ldots,n; i\neq j$.

Similarly

$$\frac{\partial f}{\partial q_i}=-p_i^{\alpha-\beta+1}q_i^{\beta-\alpha-1},$$

$$\frac{\partial^2 f}{\partial q_i^2}=(\alpha-\beta+1)p_i^{\alpha-\beta+1}q_i^{\beta-\alpha-2}, \ \forall i=1,2,\ldots,n,$$

and $\frac{\partial^2 f}{\partial q_i\partial q_j}=0$, $\forall i,j=1,2,\ldots,n; i\neq j$.

Hence, the Hessian matrix of the second order partial derivatives of f with respect to $p_1,p_2,\ldots,p_n$ is given by

$H=a_{ij}$,

where

$$a_{ij}=\begin{cases}(\alpha-\beta+1)p_i^{\alpha-\beta-1}q_i^{\beta-\alpha}, & i=j\\ 0, & i\neq j\end{cases}.$$

This Hessian matrix is positive definite. Similarly one can prove that the Hessian matrix of second order partial derivatives of f with respect to $q_1,q_2,\ldots,q_n$ is positive definite. Thus, we conclude that $D_{\alpha,\beta}(P;Q)$ is a convex function of both $p_1, p_2, \ldots, p_n$ and $q_1, q_2, \ldots, q_n$. Moreover, with the help of numerical data shown in the table 1, we have presented $D_{\alpha,\beta}(P;Q)$ as shown in the figure 1.

Table 1. $D_{\alpha,\beta}(P;Q)$ against p for $n = 2, \alpha = 3$ and $\beta = 1.1$.
Table. 1.

P	Q	$D_{\alpha,\beta}(P;Q)$
0.0	0.5	1.4380
0.1	0.5	0.9233
0.2	0.5	0.5205
0.3	0.5	0.2317
0.4	0.5	0.0580
0.5	0.5	0.0000
0.6	0.5	0.0580
0.7	0.5	0.2317
0.8	0.5	0.5205
0.9	0.5	0.9233
1.0	0.5	1.4380

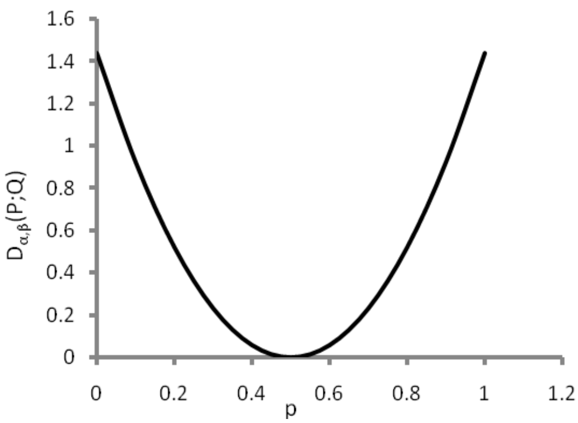

Fig. 1. Convexity of $D_{\alpha,\beta}(P;Q)$ with respect to P.

Under the above conditions, the function $D_{\alpha,\beta}(P;Q)$ is a valid parametric measure of directed divergence.

3 Relationships between information measures and Gaussian distribution

In this section, we make use of generalized information measures of entropy and divergence, discussed in the above sections, to establish their relationships with certain statistical constants for Gaussian distributed random variables.

3.1 Generalized measure of entropy of Gaussian distributed random variable

We now consider the continuous version of the generalized two parametric measure of entropy (1.4) with density function $f_x(\xi)$ given by

$$H_{\alpha,\beta} = \frac{1}{\beta-\alpha}\left[\int_{-\infty}^{\infty} f_x^{\alpha-\beta+1}(\xi)d\xi - 1\right], \quad (3.1)$$

$\alpha \neq \beta, \beta < \alpha + 1, -\infty < \alpha < \infty$.

The Gaussian distribution density $N(\mu, \sigma^2)$ with the expectation value μ and variance σ^2 is given by

$$f_x(\xi) = \frac{1}{\sqrt{2\pi}\sigma}\exp\left[-\frac{1}{2}\frac{(\xi-\mu)^2}{\sigma^2}\right]. \quad (3.2)$$

An extension to a multi-dimensional Gaussian random variable of dimension n provides the distribution density

$$f_{\underline{x}}(\underline{\xi}) = \frac{1}{(2\pi)^{\frac{n}{2}}|\Sigma|^{\frac{1}{2}}}\exp\left[-\frac{1}{2}(\underline{\xi}-\underline{\mu})^T.\Sigma^{-1}.(\underline{\xi}-\underline{\mu})\right], \quad (3.3)$$

with the expectation vector

$\underline{\mu} = E\{\underline{x}\}$,

and the covariance matrix

$\Sigma = E\{(\underline{x}-\underline{\mu})(\underline{x}-\underline{\mu})^T\}$.

If we look at $H_{\alpha,\beta}$, we realize that this information is equal to the expected value of a distribution density

$$H_{\alpha,\beta} = \frac{1}{\beta-\alpha}\left[E\left\{f_x^{\alpha-\beta+1}(\xi)\right\} - 1\right], \quad (3.4)$$

which is a typical form of an information function. Such functions often contain an expected value of a quantity, defined as information, which consist of a distribution density. Shannon's (1948) information has, for instance, such a form as $H = -E\{\ln[p(x)]\}$.

Now we insert the multi-dimensional Gaussian distribution density into the expectation function (3.4) and we get

$$H_{\alpha,\beta} = \frac{1}{\beta-\alpha}\left\{\left[\frac{1}{(2\pi)^{\frac{n}{2}}|P|^{\frac{1}{2}}}\right]^{\alpha-\beta}\int_{\underline{\Xi}}\frac{1}{(2\pi)^{\frac{n}{2}}|\Sigma|^{\frac{1}{2}}}\exp\left[-\frac{1}{2}(\alpha-\beta+1).(\underline{\xi}-\underline{\mu})^T.\Sigma^{-1}.(\underline{\xi}-\underline{\mu})\right]d\underline{\xi} - 1\right\}, \quad (3.5)$$

and we use a new covariance matrix $\Sigma_{new} = \left[(\alpha-\beta+1)\Sigma^{-1}\right]^{-1}$ in the exponent, so that

$$\int_{\underline{\Xi}}\frac{1}{(2\pi)^{\frac{n}{2}}|\Sigma_{new}|^{\frac{1}{2}}}\exp\left[-\frac{1}{2}(\underline{\xi}-\underline{\mu})^T.\Sigma_{new}^{-1}.(\underline{\xi}-\underline{\mu})\right]d\underline{\xi} = 1 \cdot \quad (3.6)$$

To obtain the determinant $|\Sigma_{new}|$, we have to calculate the matrix Σ_{new} as given below

$$\Sigma_{new} = \left[(\alpha-\beta+1)\Sigma^{-1}\right]^{-1} = \frac{1}{(\alpha-\beta+1)}\Sigma,$$

and thus we are able to determine the determinant of the $[n \times n]$-dimensional matrix

$$|\Sigma_{new}| = \left|\frac{1}{(\alpha-\beta+1)}\Sigma\right| = \frac{1}{(\alpha-\beta+1)^n}|\Sigma|.$$

The integral of the new distribution density upon multiplying both sides by $(\alpha-\beta+1)^{-\frac{n}{2}}$ gives

$$\int_{\Xi} \frac{1}{(2\pi)^{\frac{n}{2}}|\Sigma|^{\frac{1}{2}}} \exp\left[\begin{array}{l} -\frac{1}{2}(\underline{\xi}-\underline{\mu})^{T} \\ .(\alpha-\beta+1)\Sigma^{-1}.(\underline{\xi}-\underline{\mu}) \end{array}\right] d\underline{\xi} = (\alpha-\beta+1)^{-\frac{n}{2}}, \quad (3.7)$$

and thus the integral in equation (3.5) is solved. We therefore write

$$H_{\alpha,\beta} = \frac{1}{\beta-\alpha}\left\{\left[\frac{1}{(2\pi)^{\frac{n}{2}}|\Sigma|^{\frac{1}{2}}}\right]^{\alpha-\beta} (\alpha-\beta+1)^{-\frac{n}{2}} - 1\right\}, \alpha \neq \beta . \quad (3.8)$$

Thus, the given information measure depends upon the determinant of the covariance matrix which describes the uncertainty of the expectation value of the Gaussian distribution density.

3.2 Generalized measure of divergence of Gaussian distributed random variable

We now consider the continuous version of the generalized measure of divergence (2.1) for the continuous-variate probability distributions with density functions $f_p(\xi)$ and $f_q(\xi)$. This measure is given by

$$D_{\alpha,\beta} = \frac{1}{\alpha-\beta}\left[\int_{-\infty}^{\infty} f_p^{\alpha-\beta+1}(\xi) f_q^{\beta-\alpha}(\xi) d\xi - 1\right], \quad (3.9)$$

$\alpha \neq \beta, \beta < \alpha + 1, -\infty < \alpha < \infty$.

In the computation of the divergence, we have to assume two Gaussian distribution densities, both with scalar random variables, to reduce the computational effort, given by the equations (3.2) and

$$f_{x_s}(\xi) = \frac{1}{\sqrt{2\pi}\sigma_s} \exp\left[-\frac{1}{2}\frac{(\xi-\mu_s)^2}{\sigma_s^2}\right]. \quad (3.10)$$

The dummy variable ξ is the same for both densities. Otherwise one of the two distribution densities would be independent of the integration variable and could be extracted out of the integral.

Thus, corresponding to distribution densities (3.2) and (3.10), the equation (3.9) attains the following form:

$$D_{\alpha,\beta} = \frac{1}{\alpha-\beta}\left\{\int_{\Xi} \frac{(\sigma_s)^{\alpha-\beta}}{\sqrt{2\pi}(\sigma)^{\alpha-\beta+1}} \exp\left[\begin{array}{l} -\frac{1}{2}(\alpha-\beta+1)\frac{(\xi-\mu)^2}{\sigma^2} \\ +\frac{1}{2}(\alpha-\beta)\frac{(\xi-\mu_s)^2}{\sigma_s^2} \end{array}\right] d\xi - 1\right\} \quad (3.11)$$

The exponent in the equation (3.11) is given by

$$E = -\frac{1}{2}(\alpha-\beta+1)\frac{(\xi-\mu)^2}{\sigma^2} + \frac{1}{2}(\alpha-\beta)\frac{(\xi-\mu_s)^2}{\sigma_s^2}$$

$$= -\frac{1}{2}\left[\begin{array}{l} \xi^2\left(\frac{\alpha-\beta+1}{\sigma^2}+\frac{\beta-\alpha}{\sigma_s^2}\right) - 2\xi\left(\mu\frac{\alpha-\beta+1}{\sigma^2}+\mu_s\frac{\beta-\alpha}{\sigma_s^2}\right) \\ \qquad + \left(\mu^2\frac{\alpha-\beta+1}{\sigma^2}+\mu_s^2\frac{\beta-\alpha}{\sigma_s^2}\right) \end{array}\right]$$

$$= -\frac{1}{2}\left[a\xi^2 - 2b\xi + c\right]$$

$$= -\frac{a}{2}\left[\left(\xi-\frac{b}{a}\right)^2 + \frac{c}{a} - \left(\frac{b}{a}\right)^2\right], \quad (3.12)$$

where

$$a = \frac{\alpha-\beta+1}{\sigma^2} + \frac{\beta-\alpha}{\sigma_s^2}, \quad (3.13)$$

$$b = \mu\frac{\alpha-\beta+1}{\sigma^2} + \mu_s\frac{\beta-\alpha}{\sigma_s^2}, \quad (3.14)$$

and

$$c = \mu^2\frac{\alpha-\beta+1}{\sigma^2} + \mu_s^2\frac{\beta-\alpha}{\sigma_s^2}. \quad (3.15)$$

Now using equation (3.12) into equation (3.11), we get

$$D_{\alpha,\beta} = \frac{1}{(\alpha-\beta)}\left(\begin{array}{c} \frac{(\sigma_s)^{\alpha-\beta}}{(\sigma)^{\alpha-\beta+1}} \exp\left\{-\frac{a}{2}\left[\frac{c}{a}-\left(\frac{b}{a}\right)^2\right]\right\} \\ \int_{\Xi} \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{a}{2}\left(\xi-\frac{b}{a}\right)^2\right\} d\xi - 1 \end{array}\right). \quad (3.16)$$

Using the identity

$$\int_{\Xi} \frac{1}{\sqrt{2\pi}\frac{1}{\sqrt{a}}} \exp\left\{-\frac{a}{2}\left(\xi-\frac{b}{a}\right)^2\right\} d\xi = 1, \quad (3.17)$$

in the equation (3.16), we get

$$D_{\alpha,\beta} = \frac{1}{(\alpha-\beta)}\left(\frac{(\sigma_s)^{\alpha-\beta}}{(\sigma)^{\alpha-\beta+1}} \exp\left\{-\frac{a}{2}\left[\frac{c}{a}-\left(\frac{b}{a}\right)^2\right]\right\}\frac{1}{\sqrt{a}} - 1\right). \quad (3.18)$$

Inserting the terms for a, b and c from equations (3.13), (3.14) and (3.15) in equation (3.18), we get

$$D_{\alpha,\beta} = \frac{1}{(\alpha-\beta)}\left(\begin{array}{l} \frac{(\sigma_s)^{\alpha-\beta}}{(\sigma)^{\alpha-\beta+1}\sqrt{\frac{\alpha-\beta+1}{\sigma^2}+\frac{\beta-\alpha}{\sigma_s^2}}} \\ \exp\left\{-\frac{\sqrt{\frac{\alpha-\beta+1}{\sigma^2}+\frac{\beta-\alpha}{\sigma_s^2}}}{2}\left[\frac{\mu^2\frac{\alpha-\beta+1}{\sigma^2}+\mu_s^2\frac{\beta-\alpha}{\sigma_s^2}}{\sqrt{\frac{\alpha-\beta+1}{\sigma^2}+\frac{\beta-\alpha}{\sigma_s^2}}} - \left(\frac{\mu\frac{\alpha-\beta+1}{\sigma^2}+\mu_s\frac{\beta-\alpha}{\sigma_s^2}}{\sqrt{\frac{\alpha-\beta+1}{\sigma^2}+\frac{\beta-\alpha}{\sigma_s^2}}}\right)^2\right]\right\} - 1 \end{array}\right)$$

(3.19)

Thus given parametric measure of cross entropy is expressible in terms of standard deviations of the Gaussian distribution densities.

If the expected values for both the random variables are equal, we get

$$D_{\alpha,\beta} = \frac{1}{(\alpha-\beta)}\left[\frac{(\sigma_s)^{\alpha-\beta}}{(\sigma)^{\alpha-\beta+1}\sqrt{\frac{\alpha-\beta+1}{\sigma^2}+\frac{\beta-\alpha}{\sigma_s^2}}} - 1\right].$$

(3.20)

CONCLUDING REMARKS

We have derived the relationships between the information measures and the statistical constants of Gaussian distributions to develop the link between information theory and statistics. With similar arguments, the relations between information measures and other standard distributions can be studied.

ACKNOWLEDGEMENTS

The authors are thankful to the University Grants Commission, New Delhi, for providing financial assistance for the preparation of the manuscript.

REFERENCES

Ahmed, NA. and Gokhale, DV. 1989. Entropy expressions and their estimators for multivariate distributions. IEEE Transactions on Information Theory. 35:688-692.

Cover, TM. and Thomas, JA. 1991. Elements of information theory. Wiley, New York, USA.

Darbellay, GA. and Vajda, I. 2000. Entropy expressions for multivariate continuous distributions. IEEE Transactions on Information Theory. 46:709-712.

Ginebra, J. 2007. On the measure of the information in a statistical experiment. Bayesian Analysis. 2:167-212.

Havrada, JH. and Charvat, F. 1967. Quantification methods of classification process: Concept of structural α -entropy. Kybernetika. 3:30-35.

Kitsos, CP. and Toulias, TL. 2010. New information measures for the generalized normal distribution. Information. 1:13-27.

Kullback, S. and Leibler, RA. 1951. On information and sufficiency. Annals of Mathematical Statistics. 22:79-86.

Lazo, ACGV. and Rathie, PN. 1978. On the entropy of continuous probability distributions. IEEE Transactions on Information Theory. IT-24:102-122.

Parkash, O. and Mukesh. 2013. A study of maximum entropy principle for queueing theory. Canadian Journal of Mathematics. (Communicated).

Parkash, O. and Mukesh. 2012. Development of optimizational principle for minimizing risk in portfolio analysis. Global and Stochastic Analysis: An International Journal. 2:21-26.

Parkash, O. and Mukesh. 2012. New generalized parametric measures of entropy and cross entropy. American Journal of Mathematics and Sciences. 1:91-96.

Parkash, O. and Mukesh. 2013. Relation between information measures and chi-square distribution. International Journal of Pure and Applied Mathematics. 84:517-524.

Parkash, O. and Mukesh. 2011. Two new symmetric divergence measures and information inequalities. International Journal of Mathematics and Applications. 4:165-179.

Parkash, O. and Thukral, AK. 2010. Statistical measures as measures of diversity. International Journal of Biomathematics. 3:173-185.

Shannon, CE. 1948. A mathematical theory of communication. Bell System Technical Journal. 27:379-423.

Received: Sept 27, 2013; Accepted: Nov 5, 2013