APPROXIMATE CONVEXITY OF USEFUL INFORMATION OF J-DIVERGENCE OF TYPE ALFA IN CONNECTION WITH J-S MIXTURE DISTANCE MODELS

Rohit Kumar Verma[1], M. BhagyaLaxmi[2]

[1] *Associate Professor, Department of Mathematics, BhartiVishwavidyalaya, Durg, C.G., India*
[2] *Research Scholar Ph.D. Department of Mathematics, BhartiVishwavidyalaya, Durg, C.G., India*

---***---

Abstract

Inthis work we review Kullback-Leibler divergence and Jeffery's distance divergence measures for the flexible family of multivariate R-norm.we use jefferys divergence measure to compare the multivariate R-norm. A J-divergence Measuredbased on Renyi's-Tsallis Entropy much like kullback-Leibler divergence is related to Shannon's entropy .in this paper ,we have characterized the sum of two general measures associated with two distribution with discrete random variables.one of these measures is logarithmic ,while other contains the power of variable named as J-divergencebased on Renyi's –Tsallis entropy measures. Some illustrative examples are given to support the finding and further exhibit and adequacy of measure.

Keywords– Shannon's Entropy, Kullbuck–Leiblerdivergence,J-divergence, Information Measure, J- Shannon.

1.INTRODUTION

1.1 KULLBACK-LEIBLER DIVERGENCE (KL- DIVERGENCE)

The relative entropyχ from Q to P for discrete probability distributions P and Q specified on the same sample space is defined as in [12,13,15]

$D_{KL}(P \parallel Q)= \sum_{x\epsilon\ \chi} P(x) \log \frac{P(x)}{Q(x)} \equiv\ D_{KL}(P \parallel Q) = -\sum_{x\epsilon\ \chi} P(x) \log \frac{Q(x)}{P(x)}$

This study has developed several new generalized measures of relevant relative information and examined their specific cases. These metrics have also yielded novel and useful information measures, as well as their relationship with various entropy measurements.

Relative entropy Entropy is only defined in this following way

(1) If, for all x, Q(x) = 0 $\Longrightarrow P(x) = 0$ (absolutely continuity)
(2) If Q(x)$\neq$ 0 $\Longrightarrow P(x) = +\infty$

For distribution P and Q of continuous random variable ,relative entropy is defined to be the integral

$$D_{KL}(P \parallel Q) = \int_{-\infty}^{+\infty} p(x) \log \frac{p(x)}{q(x)}\,\mathrm{dx}$$

P, q are probability densities of P and Q.

It has the following properties:

1) If P and Q probability measures on a measurable space χ , and P is absolutely continuous with respect to Q, the relative entropy from Q to P is defined as
2) $D_{KL}(P \parallel Q) = \int \log\left(\frac{P(dx)}{Q(dx)}\right)$ P(dx),where$\frac{P(dx)}{Q(dx)}$ is the Random-Nikodymderivative of Q with respective to Q.

3) By the chain rule this can be written as

$$D_{KL}(P \parallel Q) = \int \frac{P(dx)}{Q(dx)} \log \frac{p(x)}{q(x)} \mathrm{Q(dx)}$$

Which is the $x\epsilon\chi$,entropy of P relative to Q.

4) If μ is any measure on $\chi$for which

$$\mathrm{P}(dx) = P(x), \mu(x) \text{ and}$$
$$\mathrm{Q(dx)} = q(x), \mu(x)$$

Meaning that P and Q both absolutely continuous with respect to μ .Relative entropy from Q to P is given by

$$D_{KL}(P \parallel Q) = \int p(x) \log\left(\frac{p(x)}{q(x)}\right) \mu(dx)$$

1.2 Entropy type measure and KL -divergence

1.3 Definition

Let P={$(p_1, p_2, \dots \dots \dots \dots, p_n)$},$\forall\ 0 \leq p_i \leq 1$ be a discrete probability distribution of aSet of events E = $\{E_1, E_2, \dots \dots \dots \dots, E_n\}$ on the basis of an experiment whose predicted probability distribution Q = $\{q_1, q_2, \dots \dots \dots \dots, q_n\}$, $0 \leq q_i \leq 1$, in information theory ,the following measures are well known:

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

1.3 Definition

In order to understand how the KL divergence works ,remember the formula for the expected value of a function .given a function f with x being a discrete variable, the expected value of $f(x)$ is defined as

$$X[f(x)] = \sum_x f(x)\, p(x)$$

Where $p(x)$ is the probability density function of the variable x. for the continuous case we have

$$X[f(x)] = \int_{-\infty}^{\infty} f(x)\, p(x) \mathrm{dx}$$

1.4 Definition

Ratio of $\frac{f(x)}{g(x)}$

It is evident from a review of the definitions of the anticipated value and the KL divergence that they are fairly comparable. while deciding$Z(x) = \left(\log\frac{f(x)}{g(x)}\right)$

We can see that:

$$X[Z(x)] = X_{x\sim f(x)}\left(\log\frac{f(x)}{g(x)}\right)$$
$$= \int_{-\infty}^{\infty} f(x) \log\left(\frac{f(x)}{g(x)}\right) dx$$

$=D_{KL}$ (F ‖ G)

Let us examine the quantity $\frac{f(x)}{g(x)}$ first .We can compare two probability density functions, let's say f and g, by calculating their ratio.

Ratio = $\frac{f(x)}{g(x)}$

1.5 Definition

1.6 Ratio for entire dataset

Using the product of the individual ratio and the full dataset $X = x_1, x_2, \dots \dots .., x_n$, we can calculate the ratio of the entire set. Be aware that this is only true if examples x_i unrelated to one another

Ratio = $\prod_{i=1}^{n} \frac{f(x_i)}{g(x_i)}$

1.7 Definition

1.8 RatioVS. KL-divergence

The log-ratio proved to be a useful tool for comparing two probability densities, f and g. The predicted value of the log-ratio is what the KL-divergence is Setting f(x)= $\log\left(\frac{f(x)}{g(x)}\right)$ results in

$$X[Z(x)] = X_{x\sim f(x)}\left[\left(\log\frac{f(x)}{g(x)}\right)\right]$$

$$= \int_{-\infty}^{\infty} f(x)\log\left(\frac{f(x)}{g(x)}\right)dx$$

$$= D_{KL}(F \parallel G)$$

1.9 THEOREM

What makes the KL divergence consistently positive?

Animportant property of the KL-divergence is that its always non-negative,that is $D_{KL}(F, \parallel, G) \geq 0$ for any valid FG .we can prove this using Jensen's inequality.

Jensen's inequality states that, if a function $f(x)$is convex ,then

$$X[Z(x)] \geq f(X[x])$$

To show that $D_{KL}(F, \parallel, G) \geq 0$ we first make use of the expected value:

$$D_{KL}(F \parallel G) = \int_{-\infty}^{\infty} f(x)\log\left(\left(\frac{f(x)}{g(x)}\right)\right)$$

$$=X_{x\sim f(x)}\left[\left(\log\frac{f(x)}{g(x)}\right)\right]$$

$$= -X_{x\sim f(x)}\left[\left(\log\frac{g(x)}{f(x)}\right)\right]$$

Because $-\log x$ is a convex function we can apply jensen's inequality:

$$-X_{x\sim f(x)}\left[\left(\log\frac{g(x)}{f(x)}\right)\right] \geq -\log\left(X_{x\sim f(x)}\left[\frac{g(x)}{f(x)}\right]\right)$$

$$= -\log\int_{-\infty}^{\infty} f(x)\frac{g(x)}{f(x)}dx$$

$$=-\log\int_{-\infty}^{\infty} g(x)\,dx$$

$$= -\log 1$$

$$= 0$$

1.10 R-Norm Measures

Some new generalized R-norm measures of useful relative information have been defined and their particular cases have been studied.from these measure new useful R-norm information measure have been derived. we have obtained j- divergence corresponding to each measure of useful relative R-norm information

We consider the function

$$D_R(P:Q;U) = \frac{R}{1-R}\left[\emptyset(1) - \emptyset\left(\frac{\sum_{i=1}^{n} u_{i}\, p_i^R q_i^{1-R}}{\sum_{i=1}^{n} u_i p_i}\right)^{\frac{1}{R}}\right]$$

1.11 J-divergence Measure

Let π_k= $\{\varepsilon = (v_1, v_2, \dots\dots\dots v_k): v_p \geq 0\,, p\ = 1,2, \dots\dots, k\ ;\ \sum_{p=1}^{k} v_p\ = 1\}$,k≥ 2 be set of k-complete probability distribution for any probability distribution $\varepsilon = (v_1, v_2, \dots\dots\dots v_k)\ \in\ \pi_k$

Shannon [2] defined an entropy as

$$\mathrm{H}(\varepsilon)\ =\ -\sum_{p=1}^{k}(v_p)\log(v_p)$$

For any $\varepsilon, \delta\ \in\ \pi_k$,Kullback and Leibler[6,12] defined a divergence measure as

$D^{K-L}(\varepsilon, \delta) = \sum_{p=1}^{k}(v_p)\log\left(\frac{v_p}{w_p}\right)$………………(1)

It is well known that $D^{K-L}(\varepsilon, \delta)$ is nonnegative, additive but not symmetric.To obtain symmetric measure, one can define

$\mathrm{J}(\varepsilon, \delta) =\ D^{K-L}(\varepsilon, \delta) +\ D^{K-L}(\delta, \varepsilon) = \sum_{p=1}^{k}(v_p - w_p)\log\left(\frac{v_p}{w_p}\right)$………………………..(2)

This is referred to as the J-divergence. It is evident that D^{K-L} and J-divergence share the majority of their features.It is evident that if w=0 and v=0,D^{K-L} is undefined.this suggests that in order to construct D^{K-L}, distribution E must be completely continuous with respect to distribution f.

Litegebe and satish [25]define a new information measure as

$H^{a}_{T-H}(\epsilon) =$
$\left\{\frac{1}{a-1}\left(1-\sum_{p\in k} v_p^a\right) - \sum_{p\in k}(v_p)log(w_p) + \frac{2^{a-1}}{2^{a-1}-1}\left(1-\ \sum_{p=1}^{k} v_p^a\right) - \right.$
$\left.\sum_{p=1}^{k} v_p\ \log v_p\right\}$……………………..(1)

A combination formulation of Havrda-charvat and Tsallis entropy of order " a" was introduced in amount (1).

A generalized usable relative information measure of order " a" that Bhaker and Hooda examined given below

$$D_a(P:Q;U) = \frac{1}{1-a}\log\left(\frac{\sum_{i=1}^{n} u_i\ p_i^a q_i^{1-a}}{\sum_{i=1}^{n} u_i p_i}\right)$$

1.12 Useful measure of j-divergence of type "a"

Kullback and Leilbler[20] and Jeffrey's[6] introduced a symmetric divergence called J-divergence of type "a" is given by

$$J_a(P:Q;U) = D_a(P:Q;U) + D_a(Q:P;U)$$

$= \frac{2}{1-a}\log\left(\frac{\sum_{i=1}^{n} u_i\, p_i^a q_i^{1-a}}{\sum_{i=1}^{n} u_i p_i}\right), a > 0$

In case utilities are ignored i.e. $u_i = 1$ for each i

Equation reduced to

$$J_a(P:Q) = \frac{2}{1-a}\log\left(\frac{\sum_{i=1}^{n} p_i^a q_i^{1-a}}{\sum_{i=1}^{n} p_i}\right)$$

2. Our Claims:

Claim I

$\frac{1}{1-a}\log\left(\frac{\sum_{i=1}^{n} u_i\, p_i^a q_i^{1-a}}{\sum_{i=1}^{n} u_i p_i}\right)$is a convex function of Q.

The steps that follow demonstrate this.

Step1: **For**$\frac{\mathbf{1}}{\boldsymbol{a-1}} > 0, a > 1$ **is a convex function of Q**

Let $\phi = \log\left(\frac{\sum_{i=1}^{n} u_i\, p_i^a q_i^{1-a}}{\sum_{i=1}^{n} u_i p_i}\right)$ if we differentiate ϕ partially w.r.t. q_i taking all p_i and u_i fixed then $\sum_{i=1}^{n} u_i q_i$

Thus, $\sum_{i=1}^{n} p_i u_i \geq \sum_{i=1}^{n} q_i u_i$ is constant

Hence$\phi = \psi \log(\sum_{i=1}^{n} u_i\ p_i^a q_i^{1-a})$

Where $\frac{1}{\psi} = ,\quad \sum_{i=1}^{n} p_i u_i \geq \sum_{i=1}^{n} q_i u_i > 0$

It implies

$$\frac{\partial\phi}{\partial p_i} = \psi\ \frac{\partial}{\partial p_i}[\log(u_i p_i^a q_i^{1-a})]$$

$$= \psi\ \frac{\partial}{\partial p_i}[\log u_i + \log {p_i}^a + \log {q_i}^{1-a}]$$

$$= \psi\left[0 + \frac{a\, {p_i}^{1-a}}{{p_i}^a} + 0\right]$$

$$= \psi\ \frac{a\, {p_i}^{1-a}}{{p_i}^a}$$

and$\frac{\partial^2\phi}{\partial {p_i}^2} = \psi a \frac{\partial}{\partial p_i}\left(\frac{\partial\phi}{\partial p_i}\right)$

$$= \psi a \frac{\partial}{\partial p_i}({p_i}^{1-a}{p_i}^{-a})$$

$$= \psi a[(1-a){p_i}^{1-a-1}{p_i}^{-a} + {p_i}^{1-a}(-a){p_i}^{-a-1}]$$

$= \psi\ a\ [(1-a){p_i}^{1-a-1-a} - a\ {p_i}^{1-a-a-1}]$

$$= \psi\ a[(1-a){p_i}^{-2a} - a{p_i}^{-2a}]$$

$$= \psi a[{p_i}^{-2a} - a{p_i}^{-2a} - a{p_i}^{-2a}]$$

$$= \psi a({p_i}^{-2a} - 2a{p_i}^{-2a})$$

$$= \psi a(1-2a){p_i}^{-2a}$$

$$= \text{A positive value}$$

Hence, $\log\left(\frac{\sum_{i=1}^n u_i p_i^a q_i^{1-a}}{\sum_{i=1}^n u_i p_i}\right)$ is a convex function of Q.

Step2: For $\boldsymbol{a} < 1, \frac{\mathbf{1}}{\boldsymbol{a-1}} < 0$ is a Monotonic increasing convex function of Q

Since $\frac{\mathbf{1}}{\boldsymbol{a-1}} > 0, for\ a\ > 1$ is a convex function of Q.

Therefore $\frac{1}{1-a}\log\left(\frac{\sum_{i=1}^n u_i p_i^a q_i^{1-a}}{\sum_{i=1}^n u_i p_i}\right)$ is a convex function of Q for all $a < 0$,

Provided, $\sum_{i=1}^n {p_i}^a u_i \ \geq\ \sum_{i=1}^n q_i u_i {p_i}^{a-1}$,$a \geq 1$. Since $\phi(x)$is monotonic increasing function of x. Then $a\ > 1$ gives

$$\log\left(\frac{\sum_{i=1}^n u_i\ p_i^a q_i^{1-a}}{\sum_{i=1}^n u_i p_i}\right) \geq 1$$

$$\Rightarrow \quad \phi\left[\log\left(\frac{\sum_{i=1}^n u_i\ p_i^a q_i^{1-a}}{\sum_{i=1}^n u_i p_i}\right)\right] \geq\ \phi(1)$$

Thus $D_a(P : Q; U) \geq 0$. Since an increasing convex function of a convex function is a convex function and $\phi(x)$ is a is monotonic increasing convex function.

Therefore $D_a(P : Q; U)$ is a convex function of Q.

Claim II

2.1 Relation between J-divergence and J-shannon

J-divergence [6,7,8]

J(p;q) = $\frac{1}{2}$ D$(p \parallel q) + \frac{1}{2}\ D(p \parallel q)$

= $\frac{1}{2}\sum_i p_i\ \log\frac{p_i}{q_i} + \frac{1}{2}\sum_i q_i\ \log\frac{p_i}{q_i}$

and the Jensen-Shannon divergence [9,8,10]

JS-divergence (p; q) = $\frac{1}{2}$ D$\left(p \parallel \frac{1}{2}(p+q)\right) + \frac{1}{2}\ D\left(q \parallel \frac{1}{2}(p+q)\right)$

$JS(p;q) = \frac{1}{2}\sum_i p_i \ln\frac{p_i}{\frac{1}{2}(p_i+q_i)} + \frac{1}{2}\sum_i q_i \ln\frac{q_i}{\frac{1}{2}(p_i+q_i)}$ ……………..(1)

Are related by the inequality

$$JS(p;q) \le min\left\{\frac{1}{4} J(p;q), \ln\frac{2}{1+e^{-J(p;q)}}\right\} \qquad \text{..................(2)}$$

The inequalityis described by Lin [9]

$JS(p;q) \le \frac{1}{2} J(p;q)$

The first part of equation [1] is described by Taneja [10]

$JS(p;q) \le \frac{1}{4} J(p;q)$

We note that many interesting measures between probability distribution can be written as an f-divergence [23,24,25]

$$\chi_f(p;q) = \sum_i p_i\, f\left(\frac{q_i}{p_i}\right)$$

$$= \langle\, f\left(\frac{q_i}{p_i}\right)\rangle$$

$$\ge 0$$

The relation $\chi_f(p;q) \ge 0$ follows from an application of Jensen's inequality [24] for convex function

$$\langle f(x)\rangle \ge f(x)$$

Now suppose that we can write

$f_w(x) = f_v(x) - k f_u(x)$,where$f_u, f_v, f_w$are all convex and normalized and k is constant.

Then,

$$f_w = f_v - w f_u \ge 0$$

Equivalently$f_v \ge k\, f_u$

The desired inequality given

$$f_{jeffery}(x) = \frac{1}{2}(x)\ln(x) - \frac{1}{2}\ln(x)$$

$f_{JS} = \frac{1}{2}\ln\frac{2}{1+x} + \frac{1}{2}x\ln\frac{2}{1+x^{-1}}$

$$f_w(x) = f_{jeffery}(x) - 4 f_{JS}(x)$$

This inequality has the same form as the asymptotic scaling between Jensen-shannon and Symeterized KL-divergencefor infinitely different distributions.

Jensen's inequality $\langle f(x)\rangle \ge f(x)$ implies that

$$\langle\ln\frac{2}{1+e^x}\rangle \le \ln\frac{2}{1+e^x}$$

Therefore $JS(p;q) = \frac{1}{2}\sum_i p_i \ln\frac{p_i}{\frac{1}{2}(p_i+q_i)} + \frac{1}{2}\sum_i q_i \ln\frac{q_i}{\frac{1}{2}(p_i+q_i)}$ (from 1)

$$= \frac{1}{2}\sum_i p_i \ln \frac{2}{1+e^{\frac{q_i}{p_i}}} + \frac{1}{2}\sum_i q_i \ln \frac{2}{1+e^{\frac{p_i}{q_i}}}$$

$$\leq \frac{1}{2}\ln \frac{2}{1+exp\{-D(p \parallel q)\}} + \frac{1}{2}\ln \frac{2}{1+exp\{-D(q \parallel p)\}}$$

$$\leq \ln \frac{2}{1+exp\{jeffery(p;q)\}}$$

i.e. JS$(p;q) \leq \ln \frac{2}{1+exp\{jeffery(p;q)\}}$

Therefore JS$(p;q)$ = Jeffery(p; q) = 0 (if $p_i = q_i$)

andJS$(p;q)$ = Jeffery(p; q) = $+\infty(if\ p_i.q_i = 0)$

since the J-divergence range between zero and positive infinity, whereas the Jensen-shannon divergence ranges between zero and ln 2 ,this inequality has the correct limit and identical and orthogonal.

Conclusions

The KL-divergence is a widely used metric to assess how well the R-norm fits the data. It was demonstrated that generalized relative entropies, whether Renyi or Tsallies, in the discrete situation can be readily extended to the measure-theoretic context, much as in the case of kullback-Leibler relative entropy. The definition of new generalized R-norm measures of meaningful relative information has been completed, and their specific cases have been examined. For any measure of valuable R-norm information, we have a corresponding J-divergence. The metrics described in this study can be applied to further information theory results and we also find theJ-divergence and the Jensen-Shannon divergence are shown to be related by an inequality that involves a transcendental function of the J-divergence.

REFERENCES:

[1] Arizono,I.,&Ohta,H.(1989):A test for normality based on Kullback-Leibler information.

[2] Shannon C. E.(1948): A mathematical theory of communication", Bell System Technical Journal,27, 379-423, 623-659

[3] Hooda, D. S. and Bhaker U.S.(1997): A generalized useful information measure and coding theorems, SoochowJournal of Mathematics, Vol 23(1), 53-62.

[4] Havrda, J. F. and Charvat, F. (1967): Quantification method of classification Processes, the concept of structural α –entropy, Kybernetika, Vol. 3, no. 1, pp. 30-35.

[5] Hooda, D. S. (1986):On generalized measure of relative useful information, Soochow Journal of Mathematics, Vol. 12, 23-32.

[6] Jeffreys. H.(1948):Theory of probability" Clarendon Press,Oxford, 2nd edition, 1.

[7] Kullback, Sand Leibler, R. A. (1951):"On information and sufàciency". Ann. Math. Statist., 22:79–86.

[8] Crooks.G.E.(2016):On measures of entropy and information, 2016. Tech.

[9] Lin, J.(1991): Divergence measures based on the Shannon entropy. IEEE Trans. Inf. Theory, 37:145–151.

[10] Topsøe, F (2000): Some inequalities for information diver-gence and related measures of discrimination". IEEE Trans. Inf. Theory, 46(4):1602–1609.

[11] Fuglede, B. and Topsøe, F.(2003): Jensen-Shannon di-vergence and Hilbert space embedding". In IEEE

[12] Taneja, I. J. (2005):Refinement inequalities among symmetric divergence measure. Aust. J. Math. Anal.Appl.

[13]Bouhlel, N and Dziri, A. (2019):Kullback–Leibler divergence between multivariate generalized gaussian distributions, IEEE Signal Processing Letters, 26(7):1021–1025.

[14] Tim, van Erven and Peter,Harremos(2014):Rényi divergence and Kullback-Leiblerdivergence. IEEE.Transactions on Information Theory, 60(7):3797–3820.

[15] Rui, F. Vigelis, Luiza, H. F. Andrade,De and Charles, C. Cavalcante(2020): Properties of a generalized divergence related to Tsallis generalized divergence". IEEE Transactions on Information Theory,66(5):2891–2897.

[16] Pengfei, Yang and Biao, Chen (2019):RobustKullback-Leibler divergence and universal hypothesistesting for continuous distributions". IEEE Transactions on Information Theory, 65(4):23602373.

[17] Yufeng, Zhang;Jialu, Pan;Wanwei, Liu;Zhenbang, Chen;Kenli, Li; Wang,Ji;Zhiming, Liu and Hongmei, Wei (2023):Kullback-leibler divergence-based out-of-distribution detection with flow-based generative models, IEEE Transactions on Knowledge and Data Engineering, pages 1–14.

[18] ZiadRached, FadyAlajaji, and Campbell, L Lorne(2004): The Kullback-Leibler divergence rate between markov sources, IEEE Transactions on Information Theory, 50(5):917–921.

[19] Nielsen, Frank(2021): On the Kullback-Leibler divergence between discrete normal distributions, Journal of the Indian Institute of Science,arXiv preprint arXiv:2109.14920.

[20] Kullback, Solomon(1997):Information theory and statistics, Courier Corporation.

[21] Hershey,John R and Olsen,Peder A.(2007): Approximating Kullback-Leibler divergence between Gaussian mixture models, IEEE International Conference on Acoustics, Speech and Signal Processing-ICASSP'07, volume 4, pages IV–317.

[22] Thomas M Cover and Joy A Thomas(2012): Elements of information theory. John Wiley & Sons.

[23] Karim T Abou-Moustafa and Frank P Ferrie(2012):A note on metric properties for some divergence measures" .The gaussian case. In Asian Conference on Machine Learning, pages 1–15. PMLR.

[24] Ali, S. M. and Silvey, S. D. (1966): "A general class of coefàcients of divergence of one distribution from another".J. Roy. Statist. Soc. B, 28(1):131–142.

[25] Csiszar, I (1967): Information-type measures of difference of probability distributions and indirect observation."Studia Sci. Math. Hungar., 2:299–318.

[26] J. L. W. V. Jensen. "Sur les functions convex set les ineqalities entre les valeursmoyennes. Act a Mathematical, 30(1):175–193.

[27] Litegebe and Satish(2016):Some inequalities in information theory using Tsallis entropy, International Journal of Mathematics and Mathematical sciences, Hindawai, pages 1-4, April.