Journal of
Mathematical
Inequalities
Volume 17, Number 4 (2023), 1639–1646 doi:10.7153/jmi-2023-17-107

AN APPLICATION OF KY FAN INEQUALITY: ON KULLBACK–LEIBLER DIVERGENCE BETWEEN A PROBABILITY DISTRIBUTION AND ITS NEGATION

RAMAZAN TINAZTEPE

(*Communicated by N. Elezović*)

Abstract. The negation of a probability distribution has been defined in the context of belief theory by [13]. Since then, its generalizations and applications have been studied. Although a probability distribution and its negation have been studied separately in terms of entropy in the literature, the relative entropy (Kullback-Leibler divergence) between them has not been investigated. In this article, we investigate the difference between a probability distribution and its negation in terms of the Kullback-Leibler divergence. In particular, we establish some bounds for the Kullback-Leibler divergence in terms of logical entropy of a probability distribution. To do this, we not only use the generalized Ky Fan inequality but also provide an extension of the generalized Ky Fan inequality. As a result, this article provides a nice application of the generalized Ky Fan inequality in a special topic in information science.

1. Introduction

Discrete probability distributions have been one of the most useful tools in probability, statistics, information theory and so many other areas of science. Depending on the needs arising in different areas of science, new concepts related to probability distributions are defined and developed.

In information science, one of the important problems is to be able to extract information from a source that has uncertainities, e.g. fuzzy sets or belief functions. Related to this problem, the question of how to define the negation of a probability distribution has been raised. This question is important because the negation of an information might help to determine the uncertainty level.

Although there have been some attempts to define the negation of a probability distribution in different contexts, the definition proposed in [13] in the context of belief functions in evidence theory (also called the Demspter-Shafer Theory in the literature. See [14] for further reading) has been the most influential in the last decade. It has been defined in the following way:

Mathematics subject classification (2020): 94A17, 26D07.
Keywords and phrases: Jensen inequality, generalized Ky Fan inequality, negation of a probability distribution, Kullback-Leibler divergence.

Let X be discrete random variable taking the values $x_1,\ldots,x_n$ with probability distribution function P and $\overline{P}(x_i)$ denote $\overline{P}(X=x_i)$. The negation of P (denoted by $\overline{P}$) is defined as

$$\overline{P}(x_i) = \frac{1-P(x_i)}{n-1}.$$

In [13] it has been shown that this definition of negation is unique in that it is based on maximum entropy according to the Dempster-Shafer theory. In [13], the logical entropy is used. The logical entropy for a discrete probability is given by

$$H(P) = 1 - \sum_i^n P(x_i)^2.$$

Note that the logical entropy $H(P)$ and $\sum_i^n P(x_i)^2$ can be considered as heterogeneity distribution and homogeneity of the distribution [3] respectively. Note that the maximum value of $\sum_i^n P(x_i)^2$ is 1 and it approaches to 1 as one of the values of the random variable approaches to 1 and the minimum value of $\sum_i^n P(x_i)^2$ is $1/n$ and this happens if the distribution becomes a uniform distribution (homogeneous in terms of probabilities).

There have been studies following the definition of negation in [13]. Based on [13], negation of bivariate probability distributions have been defined in [9] and its properties have been listed. Also, [15] extended the negation of a probability distribution using the idea of a nonextensive statistic based on the Tsallis entropy. [4] gave the application of negation of a distribution in target recognition based on sensor fusion. [10] proposed a multi-criteria decision making based on the negation distribution.

Although the definition of negation distribution given by [13] has some merit in belief theory, it is imperfect in literal meaning of negation. First, following the meaning of negation in classical sense, one expects that the negation of negation of a probability distribution is itself. However, this is not the case. Note that

$$\overline{\overline{P}} \neq P$$

unless $P = \overline{P}$, which occurs if and only if P is uniform discrete probability distribution, i.e., $P(x_i) = \frac{1}{n}$. Moreover, consecutive negations of a probability distribution approaches to the uniform distribution, which can be stated as

$$\lim_{k\to\infty} P_k = U,$$

where $P_1 = \overline{P}$ and $P_{i+1} = \overline{P_i}$ and U denotes the uniform distribution. Thus, it can be seen as a transformation to approximate the uniform distribution rather than negation.

Secondly, while $n=2$ corresponds to the usual definition of negation, the classical meaning of negation weakens as n gets larger. We observe this as

$$P(x_i) < \overline{P}(x_i) \qquad \text{if} \qquad P(x_i) < \frac{1}{n}$$

$$P(x_i) > \overline{P}(x_i) \qquad \text{if} \qquad P(x_i) > \frac{1}{n}.$$

The flaws mentioned above makes it important to evaluate the difference between a probability distribution and the distribution of its negation in terms of statistical distance. For this purpose, we use the Kullback-Leibler divergence in this article.

In mathematical statistics, the Kullback-Leibler divergence, D_{KL} (also called relative entropy) is given as a measure of how a probability distribution P differs from a second, reference probability distribution Q. For discrete probability distributions P and Q defined on the same probability space, $\mathscr{X}$, the Kullback-Leibler divergence from Q to P is defined to be

$$D_{KL}(P \parallel Q) = \sum_{x_i \in \mathscr{X}} P(x_i) \ln\left(\frac{P(x_i)}{Q(x_i)}\right).$$

In the studies related to negation of probability distribution in the literature, the Kullback-Leibler divergence between the probability distribution and its negation has not been investigated to the best of our knowledge. In this article, we relate the Kullback-Leibler divergence between a probability distribution and its negation with logical entropy. In particular, we provide some bounds for the the Kullback-Leibler divergence between a probability distribution and its negation in terms of its logical entropy. These bounds follow as a direct application of the generalized Ky Fan Inequality. However, we not only use the Ky Fan Inequality but also extend it in a direction for our special purpose. This extension is also a novel finding of this article.

In the next section we establish the generalized Ky Fan inequality and its extension. For the sake of self-containment, we also provide Jensen's inequality and explain how the generalized Ky Fan Inquality is obtained from Jensen's Inequality. Then we provide some bounds for Kullback-Leibler divergence between the probability distribution and its negation.

2. Ky Fan inequality

The Ky Fan inequality is a mathematical inequality that is derived from Jensen's inequality. For the sake of self-containment, we start from Jensen inequality and derive the others subsequently. Jensen's inequality is one of the well-known inequalities of mathematical analysis, which characterizes the convex functions.

THEOREM 1. (Jensen's Inequality) *Let* $f:\mathbb{R}\to\mathbb{R}$ *be a convex function and* λ_i *be positive real numbers for* $i=1,\ldots,n$ *that satisfy* $\sum_{i=1}^{n}\lambda_i = 1$. *In this case,*

$$f\left(\sum_{i=1}^{n}\lambda_i x_i\right) \leqslant \sum_{i=1}^{n}\lambda_i f(x_i)$$

holds for any real numbers $x_i \in \mathbb{R}$ *for* $i=1,\ldots,n$.

Note that for a random variable X with a discrete probability distribution, the Jensen inequality can be stated as

$$f(\mathbb{E}(X)) \leqslant \mathbb{E}(f(X))$$

for any convex function f.

The Jensen inequality yields different inequalities for different choices of convex functions. Many generalizations and refinements of this inequality have been obtained, for example, see [5, 12, 6] and the references therein. Also, related to classical entropy, see [2] and [8].

The following inequality is known as the generalized Ky Fan inequality in the literature. It follows from Jensen's inequality for the choice of $f(x) = \ln(x) - \ln(1-x)$ for $x \in (0, \frac{1}{2}]$.

THEOREM 2. (Generalized Ky Fan inequality) *If λ_i is a positive real number for $i = 1, \ldots, n$ satisfying $\sum_{i=1}^{n} \lambda_i = 1$, then*

$$\frac{\prod_{i=1}^{n} a_i^{\lambda_i}}{\prod_{i=1}^{n}(1-a_i)^{\lambda_i}} \leqslant \frac{\sum_{i=1}^{n} \lambda_i a_i}{\sum_{i=1}^{n} \lambda_i (1-a_i)}$$

holds for real numbers $a_i \in (0, \frac{1}{2}]$ for $i = 1, \ldots, n$ and the equality holds if $x_1 = \cdots = x_n$.

In the literature, there are refinements or generalizations of the inequality above and other inequalities associated with Ky Fan, for example, see [7, 6, 1, 11] and the references therein. We present a new one in Theorem 3.

Note that Theorem 2 requires the condition $a_i \leqslant \frac{1}{2}$. The following theorem flexes this condition. Our motivation to do so is that if we assume that a_i's represent the probabilities of a discrete probability distribution, then one of the a_i's may be bigger than $\frac{1}{2}$.

THEOREM 3. *Assume that $a_i \in (0, \frac{1}{2}]$ for $i \in \{1, \ldots n\} \setminus \{m\}$ and $\frac{1}{2} \leqslant a_m \leqslant 1$. If γ_i is a positive real number for $i = 1, \ldots, n$ such that $\sum_{i=1}^{n} \gamma_i = 1$, then*

$$\frac{\prod_{i=1}^{n} a_i^{\gamma_i}}{\prod_{i=1}^{n}(1-a_i)^{\gamma_i}} \leqslant \frac{a_m^{2\gamma_m}}{(1-a_m)^{2\gamma_m}} \left(\frac{\gamma_m(1-a_m) + \sum_{i\neq m}^{n} \gamma_i a_i}{\gamma_m(a_m) + \sum_{i\neq m}^{n} \gamma_i (1-a_i)} \right).$$

Proof. We start with rewriting the ratio of products in the following way:

$$\frac{\prod_{i=1}^{n} a_i^{\gamma_i}}{\prod_{i=1}^{n}(1-a_i)^{\gamma_i}} = \frac{a_m^{2\gamma_m}}{(1-a_m)^{2\gamma_m}} \left(\frac{\prod_{i\neq m}^{n} a_i^{\gamma_i}}{\prod_{i\neq m}^{n}(1-a_i)^{\gamma_i}} \right) \frac{(1-a_m)^{\gamma_m}}{a_m^{\gamma_m}} \tag{1}$$

Since $1 - a_m \leqslant \frac{1}{2}$, we can use Theorem 2 in the following way:

$$\begin{aligned} &\frac{a_m^{2\gamma_m}}{(1-a_m)^{2\gamma_m}} \left(\frac{\prod_{i\neq m}^{n} a_i^{\gamma_i}}{\prod_{i\neq m}^{n}(1-a_i)^{\gamma_i}} \right) \frac{(1-a_m)^{\gamma_m}}{a_m^{\gamma_m}} \\ \leqslant\ &\frac{a_m^{2\gamma_m}}{(1-a_m)^{2\gamma_m}} \left(\frac{\gamma_m(1-a_m) + \sum_{i\neq m}^{n} \gamma_i a_i}{\gamma_m(a_m) + \sum_{i\neq m}^{n} \gamma_i (1-a_i)} \right). \quad \square \end{aligned}$$

Note that this theorem coincides with Theorem 2 when $x_m = \frac{1}{2}$.

3. Kullback-Leibler divergence between a probability distribution and its negation

In the following theorem, we provide an upper bound for the Kullback-Leibler divergence between probability distribution and its negation.

THEOREM 4. *Let P and $\overline{P}$ be discrete probability distribution and its negation respectively. In this case,*

$$D_{KL}(P \,\|\, \overline{P}) \leqslant \ln(n-1) + \ln\left(\frac{1-H(P)}{H(P)}\right)$$

provided that $P(x_i) \leqslant 1/2$. The equality holds if $P(x_i) = \dfrac{1}{n}$ for $i = 1, \ldots, n$.

Proof. $D_{KL}(P \,\|\, \overline{P}) = \sum_{i=1}^{n} P(x_i) \ln\left(\dfrac{(n-1)P(x_i)}{1-P(x_i)}\right)$.
Let $t_i = P(x_i)$. So

$$D_{KL}(P \,\|\, \overline{P}) = \sum_{i=1}^{n} t_i \ln\left(\frac{(n-1)t_i}{1-t_i}\right)$$
$$= \ln(n-1) + \ln\left(\prod_{i=1}^{n} \frac{t_i^{t_i}}{(1-t_i)^{t_i}}\right).$$

Now letting $a_i = \lambda_i = t_i$ in Theorem 2 yields

$$\frac{\prod_{i=1}^{n} t_i^{t_i}}{\prod_{i=1}^{n} (1-t_i)^{t_i}} \leqslant \frac{\sum_{i=1}^{n} t_i^2}{\sum_{i=1}^{n} t_i(1-t_i)}. \tag{3}$$

So we get

$$D_{KL}(P \,\|\, \overline{P}) = \ln(n-1) + \ln\left(\prod_{i=1}^{n} \frac{t_i^{t_i}}{(1-t_i)^{t_i}}\right)$$
$$\leqslant \ln(n-1) + \ln\left(\frac{\sum_{i=1}^{n} t_i^2}{1-\sum_{i=1}^{n} t_i^2}\right).$$

Note that if $P(x_i) = \frac{1}{n}$ then $\overline{P}(x_i) = \frac{1}{n}$, i.e. P is uniform distribution, then $D_{KL}(P \,\|\, \overline{P}) = 0$. □

Using the generalized Ky Fan inequality, we can find a lower bound for $D_{KL}(\overline{P} \,\|\, P)$ in terms of logical entropy of the probability distribution P.

THEOREM 5. *If P and $\overline{P}$ be discrete probability distribution and its negation respectively, then*

$$D_{KL}(\overline{P} \,\|\, P) \geqslant \ln\left(\frac{n-1-H(P)}{H(P)}\right) - \ln(n-1)$$

provided that $P(x_i) \leqslant 1/2$. The equality holds if $P(x_i) = \frac{1}{n}$ for $i = 1, \ldots, n$.

Proof. Note that

$$D_{KL}(\overline{P} \parallel P) = \sum_{i=1}^{n} \frac{(1-P(x_i))}{n-1} \ln\left(\frac{1-P(x_i)}{(n-1)P(x_i)}\right). \tag{5}$$

Letting $P(x_i) = t_i$ yields

$$\begin{aligned}\sum_{i=1}^{n} \frac{1-t_i}{n-1} \ln\left(\frac{1-t_i}{(n-1)t_i}\right) &= \sum_{i=1}^{n} \ln\left(\frac{1-t_i}{(n-1)t_i}\right)^{\frac{1-t_i}{n-1}} \\ &= \ln\left(\prod \frac{(1-t_i)^{\dfrac{1-t_i}{n-1}}}{t_i^{\dfrac{1-t_i}{n-1}}}\right) - \ln(n-1).\end{aligned}$$

Taking the reciprocals of both sides in the inequality in Theorem 2 yields a new inequality:

$$\frac{\prod_{i=1}^{n}(1-a_i)^{\gamma_i}}{\prod_{i=1}^{n} a_i^{\gamma_i}} \geqslant \frac{\sum_{i=1}^{n} \gamma_i(1-a_i)}{\sum_{i=1}^{n} \gamma_i a_i}.$$

The new inequality is used below. Letting $\gamma_i = \frac{1-t_i}{n-1}$ and $a_i = t_i$ yields

$$\begin{aligned}\ln\left(\prod \frac{(1-t_i)^{\frac{1-t_i}{n-1}}}{t_i^{\frac{1-t_i}{n-1}}}\right) &\geqslant \ln\left(\frac{\sum_{i=1}^{n} \frac{1-t_i}{n-1}(1-t_i)}{\sum_{i=1}^{n} \frac{1-t_i}{n-1} t_i}\right) \\ &= \ln\left(\frac{\sum_{i=1}^{n}(1-t_i)^2}{1-\sum_{i=1}^{n} t_i{}^2}\right) \\ &= \ln\left(\frac{n-2+\sum_{i=1}^{n} t_i^2}{1-\sum_{i=1}^{n} t_i{}^2}\right).\end{aligned}$$

Subtracting $\ln(n-1)$ from both sides yields

$$D_{KL}(\overline{P} \parallel P) \geqslant \ln\left(\frac{n-2+\sum_{i=1}^{n} t_i^2}{1-\sum_{i=1}^{n} t_i{}^2}\right) - \ln(n-1).$$

It is clear that $D_{KL}(\overline{P} \parallel P) = 0$ if $P(x_i) = \frac{1}{n}$. $\square$

Now using Theorem 3, we extend Theorem 4 to all probability distributions.

THEOREM 6. *Let P and $\overline{P}$ be discrete probability distribution and its negation respectively. If $P(x_M) \geqslant \frac{1}{2}$ for some $1 \leqslant M \leqslant n$, then*

$$D_{KL}(P \parallel \overline{P}) \leqslant 2x_M \ln\left(\frac{x_M}{1-x_M}\right) + \ln(n-1) + \ln\left(\frac{1-H(P)-\theta}{H(P)+\theta}\right)$$

where $\theta = 2x_M^2 - x_M$.

Proof. We use the same idea in the proof of Theorem 4. Now taking $\gamma_i = x_i$ in Theorem 3 yields

$$\frac{\prod_{i=1}^{n} x_i^{x_i}}{\prod_{i=1}^{n}(1-x_i)^{x_i}} \leqslant \frac{x_M^{2x_M}}{(1-x_M)^{2x_M}} \frac{\sum_{i=1}^{n} x_i{}^2 - 2x_M^2 + x_M}{1-(\sum_{i=1}^{n} x_i^2 - 2x_M^2 + x_M)}.$$

The rest of the proof is just to apply natural logarithm to both sides. Note that the equality occurs only if $n = 2$ and $x_1 = x_2$. □

Note that taking $x_M = \frac{1}{2}$ in Theorem 6 coincide with Theorem 4 for $x_M = \frac{1}{2}$.

Conclusion

In the light of Theorem 6, we can conclude that the Kullback-Leibler divergence between a distribution and its negation becomes higher when there is an x_M such that $P(x_M)$ is very close to 1. Thus, this definition of negation makes more sense for such probability distributions. However, a definition of negation addressing the imperfections listed in the introduction and maximizing the relative entropy remains to be an open problem.

REFERENCES

[1] E. BECKENBACH AND R. BELLMAN, *Inequalities*, Springer Science & Business Media, 2012.

[2] S. S. DRAGOMIR AND C. J. GOH, *Some bounds on entropy measures in Information Theory*, Applied Mathematics Letters, **10** (3), (1997), 23–28.

[3] D. ELLERMAN, *An introduction to logical entropy and its relation to Shannon Entropy*, International Journal of Semantic Computing, **7**, 2 (2013), 121–145.

[4] X. GAO AND Y. DENG, *The generalization negation of probability distribution and its application in target recognition based on sensor fusion*, International Journal Of Distributed Sensor Networks, **15** (2019).

[5] J. LIAO AND A. BERG, *Sharpening Jensen's inequality*, The American Statistician, **73**, 3 (2019), 278–281.

[6] D. S. MITRINOVIC, J. PECARIC AND A. M. FINK, *Classical and new inequalities in analysis*, vol. 61, Springer Science & Business Media, (2013).

[7] M. MOSLEHIAN, *Ky fan inequalities*, Linear And Multilinear Algebra, **60**, (2012), 1313–1325.

[8] S. SIMIC, *Jensen's inequality and new entropy bounds*, Applied Mathematics Letters, **22** (8), (2009), 1262–1265.

[9] A. SRIVASTAVA AND S. MAHESHWARI, *Some new properties of negation of a probability distribution*, International Journal Of Intelligent Systems, **33**, (2018), 1133–1145.

[10] C. SUN, S. LI AND Y. DENG, *Determining weights in multi-criteria decision making based on negation of probability distribution under uncertain environment*, Mathematics, **8**, (2020), 1–15.

[11] R. TINAZTEPE, G. TINAZTEPE AND Y. USER, *Refinements of the Ky Fan Inequality and its additive analogue via abstract convexity*, Journal of Nonlinear and Convex Analysis, **24**, 11 (2023), 2403–2413.

[12] S. WALKER, *On a lower bound for the Jensen inequality*, SIAM Journal On Mathematical Analysis, **46**, (2014), 3151–3157.

[13] R. YAGER, *On the maximum entropy negation of a probability distribution*, IEEE Transactions On Fuzzy Systems, **23**, (2014), 1899–1902.

[14] R. YAGER AND L. LIU, *Classic works of the Dempster-Shafer theory of belief functions*, Springer, 2008.

[15] J. ZHANG, R. LIU, J. ZHANG AND B. KANG, *Extension of Yager's negation of a probability distribution based on Tsallis entropy*, International Journal Of Intelligent Systems, **35**, (2020), 72–84.

(Received March 19, 2023)

Ramazan Tinaztepe
Imam Abdulrahman Bin Faisal University
Deanship of Preparatory Year and Supporting Studies
Dammam, 31441, P.O. Box 1982, KSA
e-mail: ttinaztepe@iau.edu.sa