Conference Proceedings Paper – Entropy

A Note on Bound for Jensen-Shannon Divergence by Jeffreys

Takuya Yamano [1,*]

1   Department of Mathematics and Physics, Faculty of Science, Kanagawa University,
    2946, 6-233 Tsuchiya, Hiratsuka, Kanagawa 259-1293, Japan;
    E-Mails: yamano@amy.hi-ho.ne.jp

*   Author to whom correspondence should be addressed; E-Mail: yamano@amy.hi-ho.ne.jp;
    Tel.: +81-45-472-8796; Fax: +81-45-472-8796.

Abstract: We present a lower bound on the Jensen-Shannon divergence by the Jeffrers'
divergence when $p_i \geq q_i$ is satisfied. In the original Lin's paper 〔IEEE Trans. Info.
Theory, 37, 145 (1991)〕, where the divergence was introduced, the upper bound in terms
of the Jeffreys was the quarter of it. In view of a recent shaper one reported by Crooks, we
present a discussion on upper bounds by transcendental functions of Jeffreys by comparing
those values for a binary distribution.

1. Introduction

The Jensen-Shannon divergence $JS(p; q)$ is a similarity measure between two probability
distributions p and q. It is presently used in various disciplines ranging from information theory to
electron orbital patterns. It was introduced in the descrete case as [1, 2]

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

$$(1)$$

In terms of the Kullback-Leibler divergence $D(p; q) = \sum_i p_i ln \frac{p_i}{q_i}$, it can also be expressed as

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

(2)

In addition, the square root of the JS divergence becomes a metric in that it satisfies the triangle inequality [3]. The JS divergence can be bounded by other divergence. For example, the variational distance $V(p; q) = \sum_i |p_i - q_i|$ can be used both for upper and lower bounds:

$$JS(p; q) \leq \frac{1}{2} V(p; q), \quad ([1]), \quad JS(p; q) \geq \frac{1}{8} V^2(p; q) + \frac{1}{1152} V^4(p; q) + \cdots$$

(3)

The second inequality (lower bound) is explained in Appendix. The Jeffreys divergence $J(p; q)$, on the other hand, is an old measure [4] and is defined as $J(p; q) \triangleq D(p; q) + D(q; p)$. However bounds of JS divergence in terms of the Jeffreys divergence is not so much examined since the introduction by Lin [1]. It was shown that the JS divergence is upper-bounded by the quarter of the Jeffreys, i.e., $J(p; q)/4$ without mentioning whether there exists the best possible (We present an alternative proof of the Lin's upper bound in Appendix). Later, Crooks [5] presented a sharper one than the Lin's bound. In addition, it is stated that there is no corresponding *lower* bound in terms of $J(p; q)$. In this article, however, we show that there exists in a limited case.

2. Results and Discussion

2.1. A lower bound

We here are interested in obtaining a lower bound in terms of the Jeffreys divergence. It should practically be better to have both upper and lower bounds in terms of the same divergence. However, it was reported that there is no lower bound by Jeffreys [5]. We disprove this statement below. By a heuristic method (though it is easily verified), we find the following inequality:

$$\ln \frac{2x}{1 + x} \geq \frac{1}{4} \ln(1 + 2\ln x), \quad (x \geq 1)$$

(4)

The equality holds when $x = 1$. Substituting $x = p_i/q_i$ and taking an average with the probability p_i, we have

$$\frac{1}{2} \sum_i p_i \ln \frac{2p_i}{p_i + q_i} \geq \frac{1}{2} \frac{1}{4} \sum_i p_i \ln(1 + \ln \left(\frac{p_i}{q_i}\right)^2)$$

(5)

Similarly for the probability q_i, an inequality

$$\frac{1}{2} \sum_i q_i \ln \frac{2p_i}{p_i + q_i} \geq \frac{1}{2} \frac{1}{4} \sum_i q_i \ln(1 + \ln \left(\frac{q_i}{p_i}\right)^2)$$

(6)

holds. Since the function $\ln(1 + 2x)$ is convex, the Jensen's inequality tells $\langle \ln(1 + 2x) \rangle \leq \ln(1 + 2\langle x \rangle)$. Therefore, the right-hand sides of the above two inequalities are bounded from below respectively by

$$\frac{1}{4} \frac{1}{2} \ln\left(1 + 2D(p; q)\right) \text{ and } \frac{1}{4} \frac{1}{2} \ln(1 + 2D(q; p))$$

Summing the both sides and applying a Jensen's inequality with the equal weight, we reach a lower bound:

$$JS(p; q) \geq \frac{1}{4} \ln(1 + J(p; q))$$

$$(8)$$

2.2. Comparison of upper bounds

The Lin's upper bound

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

$$(9)$$

is based on the inequality of the arithmetic and geometric means, i.e., $(p_i + q_i)/2 \geq \sqrt{p_i q_i}$. On the other hand, a sharper one derived by Crooks uses the Jensen's inequality and it is expressed as [5],

$$JS(p; q) \leq \ln \left[\frac{2}{1 + e^{-\frac{1}{2} J(p;q)}} \right]$$

$$(10)$$

It is expressed by a transcendental function of the Jeffreys divergence. Here, we pursuit yet another simpler one in the following.

First, we recognize that the Kullback-Leibler divergence satisfies

$$D(p; q) \leq \sum_i \frac{p_i^2}{q_i} - 1$$

$$(11)$$

This inequality is easily confirmed by a function $x \ln x \geq 1, (0 \leq x \leq \infty)$, and thus the inequality $\sum_i p_i \left[\frac{p_i}{q_i} - \ln \left(\frac{p_i}{q_i} \right) \right] \geq 1$ holds. Therefore, as in the same line in [5], we can evaluate the following way:

$$D \left(p; \frac{p+q}{2} \right) \leq \sum_i \frac{p_i^2}{\frac{p_i + q_i}{2}} - 1 = \sum_i p_i \frac{2}{1 + e^{\ln \frac{q_i}{p_i}}} - 1 \leq \frac{2}{1 + e^{-D(p;q)}} - 1$$

$$(12)$$

where we have used the Jensen's inequality for a concave function at the last inequality, i.e.

$$\langle \frac{1}{1 + e^{-x}} \rangle \leq \frac{1}{1 + e^{-\langle x \rangle}}$$

In a similar way, we have

$$D \left(q; \frac{p+q}{2} \right) \leq \frac{2}{1 + e^{-D(q;p)}} - 1$$

$$(13)$$

Putting together the inequalities (12) and (13), we have

$$JS(p; q) \leq \frac{1}{2} \left(\frac{2}{1 + e^{-D(p;q)}} - 1 \right) + \frac{1}{2} \left(\frac{2}{1 + e^{-D(q;p)}} - 1 \right)$$

$$(14)$$

Lastly, using the Jensen's inequality once more, we have an upper bound

$$JS(p; q) \leq \frac{2}{1 + e^{-\frac{1}{2} J(p;q)}} - 1$$

$$(15)$$

This bound is similar to (10) and is expressed by a transcendental function of the Jeffreys, however it is simpler. For practical purposes, it is important to examine how the upper bound (15) relates to the previous ones. In Figure 1, we show three curves of the upper bounds in the JS-Jeffreys plane. These

curves correspond to inequalities (9), (10) and (15). We find that the bound (15) is sharper than Lin's, however not than Crook's.

In addition, the comparison of these bounds for a binary distribution is shown in Figure 2, in which $p = (t, 1-t)$ and $q = (1-t, t)$, where $0 \leq t \leq 1$. We observe that the bounds give nearly same value around t=1/2, while the differences become prominent as t departs towards either 0 or 1. Among these bounds we find that the Crooks' bound is sharper, however it indicates that there still seems to exist a sharper one for entire range.

Figure 1. Comparison of the bound functions in the JS-Jeffreys plane: Lin (red), Crooks (green) and the present (blue).
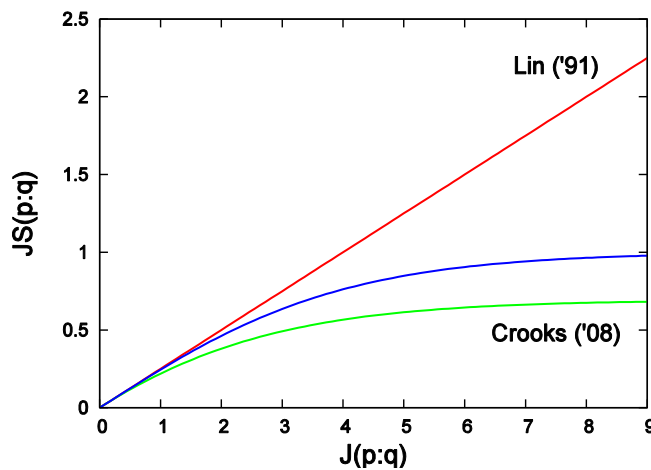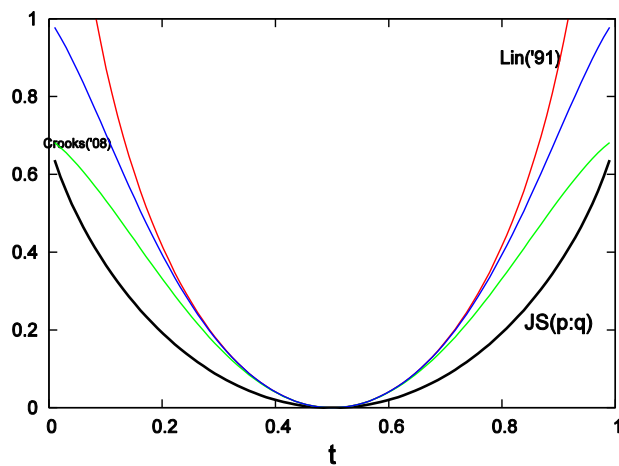

Figure 2. Comparison of the bounds for a binary distribution: Lin (red), Crooks (green) and the present (blue).

4. Conclusions

We have derived a lower bound of the Jensen-Shannon divergence in terms of the Jeffrey's divergence under a condition of $p_i \geq q_i$: $JS(p; q) \geq \frac{1}{4}\ln(1 + J(p; q))$. Regarding the upper bound, we found a simpler one that is another transcendental function of the Jeffrey's divergence along the line of the

previous study. It is sharper than Lin's, but not than Crooks'. Comparisons of alternative upper bounds may practically be useful and it is worth seeking for a sharper one.

Appendix

On lower bound by variational distance

The Kullback-Leibler divergence is lower bounded by the variational distance V, which is known as the Pinsker inequality [6, 7]:

$$D(p; q) \geq \frac{1}{2}V^2 + \frac{1}{36}V^4 + \cdots$$

(Various bounds on a *generalized* Kullback-Leibler divergence was given in [8]). Therefore, an equality

$$D\left(p; \frac{p+q}{2}\right) = D\left(q; \frac{p+q}{2}\right) \geq \frac{1}{8}V^2 + \frac{1}{1152}V^4 + \cdots$$

holds, where $V\left(p; \frac{p+q}{2}\right) = V\left(q; \frac{p+q}{2}\right) = V(p; q)/2$ is used. Thus, for the JS divergence, the second inequality of Eq. (3) follows.

An alternative proof of the Lin's bound [1]:

The Lin's bound $JS(p; q) \leq \frac{1}{4}J(p; q)$ can be derived as follows. Let us put the ratio of the two probability functions as $\theta_i = \frac{q_i}{p_i}$. Then, we have

$$\frac{1}{2}\sum_i p_i \ln \frac{p_i}{\frac{p_i + q_i}{2}} = \frac{1}{2}\{\ln 2 + \sum_i p_i (1 + \theta_i)^{-1}\}$$

Noting $(1 + \theta_i)^{-1} = \int_0^\infty e^{-(1+\theta_i)x} dx$ and applying the Schwartz inequality to it,

$$\int_0^\infty e^{-(1+\theta_i)x} dx \leq \sqrt{\int_0^\infty e^{-2x} dx} \sqrt{\int_0^\infty e^{-2\theta_i x} dx} = \frac{1}{\sqrt{2}} \frac{1}{\sqrt{2\theta_i}}$$

we have

$$\langle \ln(1 + \theta_i)^{-1} \rangle \leq -\ln 2 - \frac{1}{2}\langle \ln \theta_i \rangle = -\ln 2 + \frac{1}{2}D(p; q)$$

Thus, we can evaluate as

$$\frac{1}{2}\sum_i p_i \ln \frac{2p_i}{p_i + q_i} \leq \frac{1}{2} \cdot \frac{1}{2}D(p; q)$$

Similarly, we have

$$\frac{1}{2}\sum_i q_i \ln \frac{2q_i}{p_i + q_i} \leq \frac{1}{2} \cdot \frac{1}{2}D(q; p)$$

Therefore, putting together the above two inequalities, we have the Lin's bound:

$$JS(p; q) \leq \frac{1}{4}(D(p; q) + D(q; p)) = \frac{1}{4}J(p; q)$$

Acknowledgments

Author Contributions

The author conducted this research and wrote the whole part of the paper.

Conflicts of Interest

The author declares no conflict of interest.

References and Notes

1. Lin, J. Divergence measures based on the Shannon entropy. IEEE Trans. Info. Theory 1991, 37, 145-151.
2. Rao, C. R. Differential Geometry in Statistical Inference. IMS-Lecture Notes 1987, 10, 217–225.
3. Endres D. and Schindelin, J. A new metric for probability distributions. IEEE Trans. Inf. Theory, 2003 49, 1858-18.
4. Jeffreys, H. An invariant form for the prior probability in estimation problems. Proc. Roy. Soc. Lon., Ser. A 1946, 186, 453-461.
5. Crooks, G.E. Inequalities between the Jensen–Shannon and Jeffreys divergences. Tech. Note 004 (2008) http://threeplusone.com/Crooks-inequality.pdf
6. Kraft, O. A note on exponential bounds for binomial probabilities. Ann. Inst. Statist. Math. 1969, 21, 219-220.
7. Fedotov, A., Harremoes, P., Topsoe, F. Refinements of Pinsker inequality. IEEE Trans. Inf. Theory, 2003, 49 1491-1498.
8. Yamano, T. Bounds on divergence in nonextensive statistical mechanics. J. Indones. Math. Soc. 2013, 19, 89-97.