Ultimate Sharpening of Jensen's Inequality

Sang Kyu Lee, Jaeho Chang and Hyoung-Moon Kim
Department of Applied Statistics, Konkuk University, Seoul, Korea

1. Introduction

- Jensen's inequality is one of the fundamental mathematical inequalities and it has been used in many statistical and mathematical proofs. For example, the nonnegativity of Kullback-Leibler divergence in information theory(Cover and Thomas, 2012).
- The Jensen inequality is defined in Jensen (1906) as $E\big(\phi(X)\big) - \phi\big(E(X)\big) \geq 0$, and the Jensen gap is defined as $E\big(\phi(X)\big) - \phi\big(E(X)\big)$, where X is a random variable and $\phi(X)$ is the convex function of X. Sometimes in machine learning researches, $E\big(\phi(X)\big)$ is approximated by using $\phi\big(E(X)\big)$ and the errormatic value of the Jensen gap.
- The improvements over the Jensen gap have been discussed on other researches, but, we propose new bounds(or approximations) for the Jensen gap. We will show that these bounds are the ultimate bounds when some conditions are satisfied. Also, these bounds are applicable to various examples; moment generating function, arithmetic-geometric mean, power mean, partitioned case and information entropy. But due to the space limitation, we only show 2 examples here. Also we only include important references.

2. Main Idea and Results

Theorem 1. Let Xbe a random variable with mean μ, and $P\big(X \in (a,b)\big) = 1$, where $-\infty \leq a < b \leq \infty$. Assume $E|X-\mu|^k < \infty$ for $k = 2m,\ m = 1,2,3,\ldots$. Let $\phi(x)$ be a $(k+1)$ times differentiable function on $\forall x \in (a,b)$, and define the function

$$r^{(k)}(x;\mu) = \frac{1}{k!}\phi^{(k)}\{g(x)\} = \frac{\phi(x)-\phi(\mu)}{(x-\mu)^k} - \sum_{i=1}^{k-1}\frac{\phi^{(i)}(\mu)}{i!\,(x-\mu)^{k-i}}$$

where $\phi^{(k)}(x) = \frac{d^k}{dx^k}\phi(x)$ and $g(x) \in (\mu, x)$. In addition, let

$$m_k = \inf_{x\in(a,b)} r^{(k)}(x;\mu), \qquad M_k = \sup_{x\in(a,b)} r^{(k)}(x;\mu).$$

Then lower bound and upper bound of the $E[\phi(X)] - \phi(E[X])$ become

$$\sum_{i=1}^{k-1}\frac{\mu_i}{i!}\phi^{(i)}(\mu) + \mu_k m_k \quad \text{and} \quad \sum_{i=1}^{k-1}\frac{\mu_i}{i!}\phi^{(i)}(\mu) + \mu_k M_k,$$

where $\mu_k = E[X-\mu]^k$.

- The main result in Liao and Berg (2018) is the special case of ours when $k=2$, because $r^{(2)}(x;\mu) = \frac{\phi(x)}{(x-\mu)^2} - \frac{\phi^{(1)}(\mu)}{x-\mu}$ is equivalent to the function given in Liao and Berg (2018). Next, three lemmas are given in order to obtain m_k and M_k.

Lemma 1. Let $\phi(x)$ be $(k+1)$-times differentiable function on $\forall x \in (a,b)$. Then, if $\phi^{(k-1)}(x)$ is strictly convex, there is always an unique $g(x) \in (\mu, x)$, with $\mu \in (a,b)$ such that $\phi^{(k)}\{g(x)\} = \frac{\phi^{(k-1)}(x)-\phi^{(k-1)}(\mu)}{x-\mu}$ for $\forall x \in (a,b)$, $k = 1,2,3,\ldots$.

Lemma 2. If $\phi^{(k-1)}(x)$ is strictly convex on $\forall x \in (a,b)$, $g(x)$ is also strictly increasing on $\forall x \in (a,b)$ where $k = 1,2,3,\ldots$.

Lemma 3. Let $\phi(x)$ be $(k+1)$-times differentiable function on $\forall x \in (a,b)$, $\phi^{(k-1)}(x)$ is strictly convex and $g(x)$ be differentiable function in $\forall x \in (a,b)$, then $\phi^{(k+1)}\{g(x)\}g'(x) > 0,\ \mu < g(x) < x$, where $k = 1,2,3,\ldots$.

- We obtain a similar result if $\phi^{(k-1)}$ is a strictly concave function. Thus, it is obvious from Lemma 3 that if $\phi^{(k-1)}$ is strictly convex (concave), $\phi^{(k)}\{g(x)\}$ is strictly increasing (decreasing) in x, respectively. If $\phi^{(k-1)}(x)$ is strictly convex,

$$\inf \phi^{(k)}\{g(x)\} = \lim_{x\to a}\phi^{(k)}\{g(x)\} \text{ and } \sup \phi^{(k)}\{g(x)\} = \lim_{x\to b}\phi^{(k)}\{g(x)\}$$

if $\phi^{(k-1)}(x)$ is strictly concave,

$$\inf \phi^{(k)}\{g(x)\} = \lim_{x\to b}\phi^{(k)}\{g(x)\} \text{ and } \sup \phi^{(k)}\{g(x)\} = \lim_{x\to a}\phi^{(k)}\{g(x)\}.$$

Note that the limits of $\phi^{(k)}\{g(x)\}$ can be finite or infinite.

- Next, we explore the conditions under which the bound given in Theorem 1 is an ultimate bound. After, $f \in C^k(I)$ means that f is k- times continuously differentiable on I, $f \in C^\infty(I)$ means that f is infinitely differentiable on I and $f \in C^\omega(I)$ means f is analytic on I.

Lemma 4. A function $\phi(x) \in C^\infty(a,b)$ is analytic on (a,b) if and only if for given $\mu \in (a,b)$ there is an interval (c,d) containing μ such that the remainder term $R_k^{\phi,\mu}(x)$ converges to 0 for all $x \in (c,d)$ where $R_k^{\phi,\mu}(x) = \phi(x) - \sum_{i=1}^{k-1}\frac{\phi^{(i)}(\mu)}{i!}(x-\mu)^i$.

Theorem 2. If $\phi(x) \in C^\infty(a,b)$ is an analytic function for given $E[X] = \mu \in (a,b)$, let $E|X-\mu|^k < \infty$ for all $k = 2m, m = 1,2,3,\ldots$, then

$$E[\phi(X)] - \phi(E[X]) = \lim_{k\to\infty}\sum_{i=1}^{k-1}\frac{\mu_i}{i!}\phi^{(i)}(\mu) \qquad (1)$$

and (1) is the ultimate lower bound of the Jensen gap for $\phi(x)$, if it satisfies conditions from Theorem 1 and Lemma 3.

- Hence, by Theorem 2 and Lemma 4, the bound given in Theorem 1 is an ultimate bound if the function $\phi(x)$ is analytic and it satisfies the conditions in Theorem 1 and Lemma 3. Next, we explore the case where the composite of real analytic functions may have an ultimate bound.

Lemma 5. Denote an analytic function p in some I as $p \in C^\omega(I)$. Let $p \in C^\omega(I)$ where $I \subset R$, and p takes real values in $J \subset R$ and $\phi \in C^\omega(J)$. Then $\phi \circ p \in C^\omega(I)$.

Theorem 3. Let $h = \phi \circ p$, where $\phi(x) = -\log x\,\mathbb{1}(x>0)$ and p is the density of X which is analytic in $I = \{x\,|\,p(x) > 0\}$ then

$$E[h(X)] - h(\mu) = \lim_{k\to\infty}\sum_{i=1}^{k-1}\frac{h^{(i)}(\mu)\mu_i}{i!},$$

where $\mathbb{1}$ is the indicating function.

- Hence, by Lemma 5 and Theorem 3, we know that the composite of real analytic functions can have an ultimate bound similar to that in Theorem 2. Information entropy is a specific example of such cases.

3. Examples

Example 1. (Moment Generating Function) Let X a random variable with finite moments on support (a,b). Then, we can get the lower bound of the moment generating function $E[e^{tX}]$ using Theorem 1 with $\phi(x) = e^{tx}$. Note that the domain can be $(a,b) = (0,\infty)$. Let $X \sim Exponential(1)$ and $t = \frac{1}{2}$, then the actual value of the Jensen gap is $E[e^{tX}] - e^{tE[X]} = 2 - \sqrt{e} \approx 0.351$. The results, attained by Theorem 1 are in Figure 1 and Table 1. Note that the results of ours are better approximated than the results of Liao and Berg (2018) and Walker (2014).

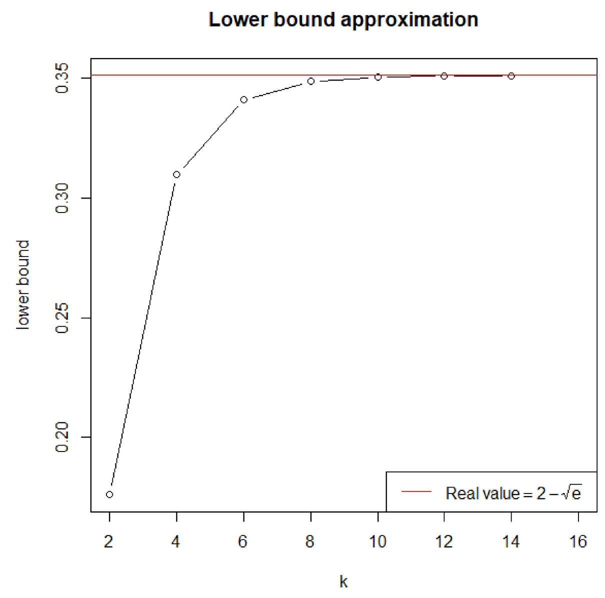

Figure 1. Lower bounds of an exponential distribution MGF with kth central moments

Actual value of Jensen Gap	$2-\sqrt{e} \approx 0.351$
From Liao and Berg (2018)	0.176
From Walker (2014)	0.271

k	2	4	6	8	10
Approximated Lower bound	0.176	0.310	0.341	0.349	0.351

Table 1. Lower bounds of an exponential distribution MGF with kth central moments

Example 2. (Information Entropy) Let X be a positive random variable with a support (a,b) and let $E[X] = \mu_X$. The information entropy is defined as

$$H(X) = -\int_x \log f(x)\,dF(x)$$

where $f(x)$ is the probability density function of X, log is the natural log function and $F(x)$ is the cumulative distribution function of X. By Lemma 5, we know that $\phi(x) = -\log f(x)$ is an analytic and also a strictly convex or concave function of x. Thus, the Jensen gap is, $H(X) + \log f(E[X])$. For the numerical illustration, let a random variable X has the distribution of $Truncated\,gamma(\alpha,\beta)$ with a support *(1,2)*. Then,

$$f(x) = \frac{g(x)}{G(2)-G(1)} = \frac{1}{G(2)-G(1)}\frac{1}{\Gamma(\alpha)\beta^\alpha}x^{\alpha-1}e^{-\frac{x}{\beta}}$$

where g is the density function and G is the distribution function of $Gamma(\alpha,\beta)$. Hence,

$$\phi(x) = -\log f(x) = \log\big(G(2)-G(1)\big) + \log\Gamma(\alpha) + \alpha\log\beta + (1-\alpha)\log x + \frac{x}{\beta}$$

and

$$\phi^{(k)}(x) = (1-\alpha)\frac{1}{x} + \frac{1}{\beta} \text{ for } k=1, \text{ and } (-1)^{k-1}\Gamma(k)(1-\alpha)\frac{1}{x^k} \text{ for } k \geq 2.$$

Note that we assume α is not equal to 1 because the function $\phi^{(k)}(x)$ becomes a constant if $\alpha = 1$. The results are in Figure 2 and Table 2. They are well approximated.

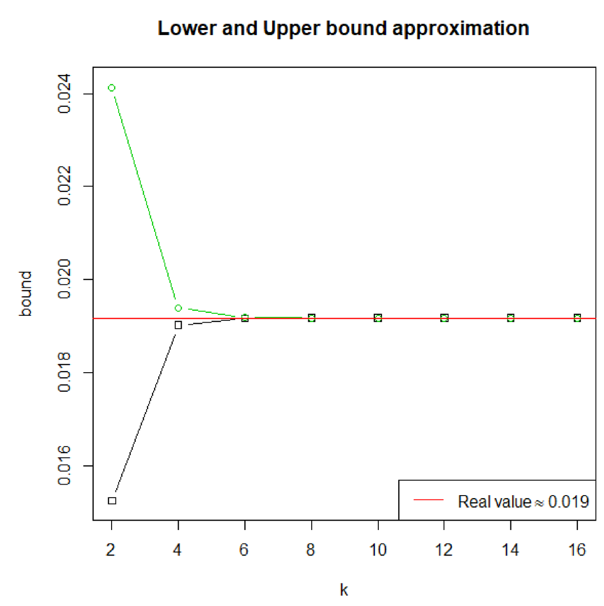

Figure 2. Lower bounds and upper bounds for entropy of Truncated Gamma(2,1) on $x \in (1,2]$ with kth central moments

Actual value of Jensen gap for TruncGamma(2,1) entropy
$-E(\log f(x)) + \log f(\mu) \approx 0.019$

k	2	4
Lower bound	0.015	0.019
Upper bound	0.024	0.019

Table 2. Lower bounds and upper bounds for entropy of Truncated Gamma(2,1) on $x \in (1,2]$ with kth central moments

4. Concluding Remarks

- This article proposes a new bound for the Jensen gap as an extension of the Liao and Berg (2018).
- It showed that if the function is analytic and satisfies certain conditions related to convexity, the bounds converge to the Jensen gap monotonically as the dimension of the Taylor series increases. Hence, we obtain the ultimate bound for the Jensen gap. To the best of our knowledge, this is the first bound to be developed. Furthermore, we showed that the composite of real analytic functions can also have an ultimate bound.

References

Cover, T. M. and Thomas, J. A. (2012), *Elements of information theory*, John Wiley & Sons.

Jensen, J. L. W. V. (1906), "Sur les fonctions convexes et les inegalites entre les valeurs moyennes," *Acta mathematica*, 30, 1, 175-193.

Liao, J. G. and Berg, A. (2018), "Sharpening Jensen's Inequality," *The American Statistician*, in process.

Walker, S. G. (2014), "On a lower bound for the Jensen inequality," *SIAM Journal on Mathematical Analysis*, 46, 5, 3151-3157.