BOUNDS ON THE JENSEN GAP, AND IMPLICATIONS FOR MEAN-CONCENTRATED DISTRIBUTIONS

XIANG GAO, MEERA SITHARAM, ADRIAN E. ROITBERG

DEPARTMENT OF CHEMISTRY, AND DEPARTMENT OF COMPUTER & INFORMATION SCIENCE & ENGINEERING, UNIVERSITY OF FLORIDA, GAINESVILLE, FL 32611, USA.
qasdfgtyuiop@gmail.com
URL: https://scholar.google.com/citations?user=t2nOdxQAAAAJ

ABSTRACT. This paper gives upper and lower bounds on the gap in Jensen's inequality, i.e., the difference between the expected value of a function of a random variable and the value of the function at the expected value of the random variable. The bounds depend only on growth properties of the function and specific moments of the random variable. The bounds are particularly useful for distributions that are concentrated around the mean, a commonly occurring scenario such as the average of i.i.d. samples and in statistical mechanics.

1. INTRODUCTION

The widely used Jensen's inequality for convex functions, attributed to Danish mathematician Johan Jensen, dates back to 1906[12]. The literature contains numerous bounds on the Jensen gap, defined as $\mathcal{J}(f, X \sim \mathcal{P}) = \mathbb{E}[f(X)] - f(\mathbb{E}[X])$, where X is a random variable with distribution $\mathcal{P}$, and the function f might be convex or nonconvex [20].

Example consequences and applications of the known bounds are: a number of famous classical inequalities such as the generalized mean inequality, and a special case the inequality of arithmetic and geometric means, the Hölder's inequality, etc. [17]; commonly used results in information theory, e.g. non-negativity of Kullback-Leibler divergence [7]; variational bounds for negative log likelihood used in statistics and machine learning methods such as the expectation maximization algorithm[8], and variational inference[18].

Computing a hard-to-compute $\mathbb{E}[f(X)]$ appears in theoretical estimates in a variety of scenarios from statistical mechanics to machine learning theory. A common approach to tackle this problem is to make the approximation $\mathbb{E}[f(X)] \approx f(\mathbb{E}[X])$ (for example $\langle \frac{1}{X} \rangle \approx \frac{1}{\langle X \rangle}$), and then show that the error, i.e., the Jensen gap, would be small enough for the application. Since the error itself is as hard to compute as $\mathbb{E}[f(X)]$, inequalities on the Jensen gap would help by giving easy-to-compute bounds. Moments are commonly used to characterize distributions of random variables because of their relative ease to compute for many distributions. By establishing the connection between the Jensen gap and moments, we create a powerful tool for error estimation based on moment estimates.

As a concrete scenario, the Jensen gap has many useful interpretations in statistical mechanics such as the difference of average non-equilibrium work and change of free energy, an important quantity to characterize the deviation of a thermodynamical process from a quasi-static process, as in Jarzynski equality[11], and the fluctuation of thermodynamical quantities around their ensemble average, which is of common interest in physics. In machine learning theory, stochastic gradient descent is employed to minimize the so-called loss function, when learning the parameters of a function from a parametrized family; in this case the training inputs are sampled from a distribution and the loss function is an expectation, to be minimized with respect to the parameters.

In such scenarios, a common type of random variable has a distribution concentrated around its mean as described below.

(1) In estimating $\xi = f(\mathbb{E}[X])$, an empirical average from samples is often used as an estimation of expectation, i.e. $\mathbb{E}[X] \approx \bar{X} = \frac{1}{|\mathcal{M}|} \sum_{X_i \in \mathcal{M}} X_i$ and $\xi \approx \hat{\xi} = f(\bar{X})$. The bias of $\hat{\xi}$, i.e. $\mathbb{E}_{\mathcal{M}}[\hat{\xi}] - \xi$ is given by the Jensen gap where the random variable $\bar{X}$ has a distribution concentrated around its mean. The asymptotic growth behavior of the Jensen gap therefore gives an idea how fast we can push the bias to zero by increasing N.

(2) Random variables with a distribution concentrated around the mean are very common in statistical mechanics. Since the number of particles in the system is usually the order of Avogadro constant $N_A \sim 10^{23}$, the distribution is so sharp that all the Jensen gaps become negligible. However, this is not the case in computer simulation or microscopic experiments, which usually have a much smaller system size. The asymptotic growth behavior of thermodynamic fluctuation (defined as the Jensen gap of function $\sqrt{\cdot}$ of random variable E^2) with the system size guides the simulation/experiment setup.

Moments play an important role in studying random variables with a distribution concentrated on the mean, especially when the random variable is an empirical average of i.i.d random

variables[14]. Our results will use moments to express the asymptotic growth behavior of the Jensen gap.

Next we give an elementary example to illustrate the inspiration behind our results, which establish the connection between the Jensen gap and the (absolute centered) moment $\sigma_p = \sqrt[p]{\mathbb{E}\left[|X - \mu|^p\right]}$, where $\mu = \mathbb{E}[X]$ is the expectation of random variable X. Assume that for $\alpha > 0$, $f(x)$ is α-Hölder continuous over $\mathbb{R}$, i.e. there exists a positive number M such that for any $x \in \mathbb{R}$, $|f(x) - f(\mu)| \leq M|x - \mu|^\alpha$. Then we have

$$(1.1) \quad |\mathbb{E}[f(X)] - f(\mathbb{E}[X])| \leq \int |f(X) - f(\mu)|\, d\mathcal{P}(X)$$

$$\leq M \int |x - \mu|^\alpha\, d\mathcal{P}(X) \leq M\sigma_\alpha^\alpha$$

Similarly, if $f(x) - f(\mu) \geq M|x - \mu|^\alpha$ or $f(\mu) - f(x) \geq M|x - \mu|^\alpha$, we can obtain an elementary lower bound on the Jensen gap

$$(1.2) \quad |\mathbb{E}[f(X)] - f(\mathbb{E}[X])| = \int |f(X) - f(\mu)|\, d\mathcal{P}(X)$$

$$\geq M \int |x - \mu|^\alpha\, d\mathcal{P}(X) \geq M\sigma_\alpha^\alpha$$

Our main results generalize these two elementary bounds as described in the next section.

1.1. **Contribution and Comparison.**

We prove an upper and lower bound on the Jensen gap, summarized below, and demonstrate their tightness. In the following, "upper bound of A" means $|\mathcal{J}(f, X \sim \mathcal{P})| \leq A$ which means $-A \leq \mathcal{J}(f, X \sim \mathcal{P}) \leq A$, while "lower bound of A" means either $\mathcal{J}(f, X \sim \mathcal{P}) \geq A$ or $-\mathcal{J}(f, X \sim \mathcal{P}) \geq A$.

- For functions that approach $f(\mu)$ at $x \to \mu$ no slower than $|x - \mu|^\alpha$, and grow as $x \to \pm\infty$ no faster than $\pm|x|^n$ for $n \geq \alpha$,

  $$|\mathbb{E}[f(X)] - f(\mathbb{E}[X])| \leq M\left(\sigma_\alpha^\alpha + \sigma_n^n\right) \leq M\left(1 + \sigma_n^{n-\alpha}\right)\sigma_n^\alpha$$

  where $M = \sup_{x \neq \mu} \frac{|f(x) - f(\mu)|}{|x - \mu|^\alpha + |x - \mu|^n}$. This implies that $\mathbb{E}[f(X)] - f(\mathbb{E}[X])$ approaches 0 no slower than σ_n^α as $\sigma_n \to 0$.

- For functions that either decrease or increase (but do not decrease on one side and increase on the other) to $f(\mu)$ as $x \to \mu$ no faster than $|x - \mu|^\alpha$, and grow to infinity at $x \to \infty$ no slower than $|x|^\beta$ for $0 \leq \beta \leq \alpha$,

  $$|\mathbb{E}[f(X)] - f(\mathbb{E}[X])| \geq M\frac{\sigma_{\alpha/2}^\alpha}{1 + \sigma_{\alpha-\beta}^{\alpha-\beta}}$$

  where $M = \inf_{x \neq \mu}\left\{[f(x) - f(\mu)] \cdot \left(\frac{1}{|X - \mu|^\beta} + \frac{1}{|X - \mu|^\alpha}\right)\right\}$. This implies that $\mathbb{E}[f(X)] - f(\mathbb{E}[X])$ decreases to 0 no faster than $\sigma_{\alpha/2}^\alpha$ as $\sigma_{\alpha/2} \to 0$ as long as $\sigma_{\alpha-\beta}$ does not grow to infinity at the same time.

Although neither our upper bounds nor our lower bounds require the function to be convex or concave, the condition in our lower bound is naturally satisfied by convex or concave functions as we will show in Section 3.2.

In order to illustrate the flavor of the main results, we give simple examples that are direct consequences. We also compare the consequences of our main results with known lower bounds[1, 15, 19, 2, 20, 13] and upper bounds[6, 16]. A major advantage of our result over these known results is their relative generality: our conditions on the function, its domain, and

the distribution are weak (for example we do not require the function to be convex, and we do not require the distribution to be discrete). Among the above-mentioned bounds, [15] and [16] are only for discrete distributions, and are omitted from the comparisons below. Besides the bounds as listed above, the Jensen's gap can also be estimated by Jensen-Ostrowski type inequalities[4, 9, 5, 10, 3].

Example 1.1. *Consider the Jensen gap of* $\sin(x)$ *and random variables with mean at* 0. *Observe that* $\sin(x)$ *has a power series* $\sin(x) = x - \frac{x^3}{6} + \frac{x^5}{120} + \cdots$, *and by choosing* $\alpha = n = 1$, *we get* $|\mathcal{J}(\sin, X \sim \mathcal{P})| \leq \sigma_1^1$. *Also, since* $\sin'(0) = 1 \neq 0$, *we can obtain a different result by studying* $g(x) = \sin(x) - x$ *(which has the same gap behavior, discussed in Section 2.2 and Section 3.2) instead. This time, by choosing* $\alpha = n = 3$, *we can see that* $|\mathcal{J}(\sin, X \sim \mathcal{P})| \leq \frac{\sigma_3^3}{6}$. *If we are interested in the asymptotic behavior of the Jensen gap when the distribution is concentrated around the mean, we can conclude immediately that* $|\mathcal{J}(\sin, X \sim \mathcal{P})|$ *decreases to* 0 *no slower than* $\sim \sigma_3^3$ *and* $\sim \sigma_1^1$. *It is also possible to choose* $\alpha = n = 2$ *and obtain* $|\mathcal{J}(\sin, X \sim \mathcal{P})| \leq \frac{\sigma_2^2}{\pi}$. *Although this result is not as good as the* σ_3^3 *version in terms of asymptotic behavior for non-heavy-tailed distributions such as Gaussian distribution and Laplace distribution, the second moment is usually more available than the third moment. Our lower bound result is not useful in this example. In fact, since* $\sin(x)$ *is odd, any even distribution* $\mathcal{P}$ *will result in a zero Jensen gap regardless of its moments. That is, it is impossible to obtain a non-trivial bound that is a function of only moments.*

Results in [6, 19] *require the function to be convex, and are therefore not useful for this example. Since the domain is not* $[0, A)$ *or* $(0, A]$ *as required by* [2], *or* $[0, +\infty)$ *as required by* [1], *these results do not apply either. The result in* [13] *gives the same bound as our* σ_2^2 *version, which is better than the* $\mathcal{J}(\sin, X \sim \mathcal{P}) \geq -\frac{\sigma_2^2}{2}$ *given by* [20]. *The fact that* [13] *gives the same bound as ours is not just a coincidence, but can be attributed to the connections between our results and* [13] *as described in Section 2.2.*

Example 1.2. *Consider* $\cos(x)$ *and random variables with mean at* 0. *Observe that* $\cos(x)$ *has a power series* $\cos(x) = 1 - \frac{x^2}{2} + \cdots$, *we can choose* $\alpha = n = 2$ *and see that* $|\mathcal{J}(\cos, X \sim \mathcal{P})| \leq \frac{\sigma_2^2}{2}$. *If we are interested in the asymptotic behavior, we can conclude that* $|\mathcal{J}(\cos, X \sim \mathcal{P})|$ *will decrease no slower than* $\sim \sigma_2^2$, *i.e. the variance of the distribution. Again, our lower bound result is not useful in this example. In fact, although non-trivial, it is possible to construct a probability distribution* $\mathcal{P}$ *that makes the Jensen gap equal* 0 *and has arbitrary moments,*[1] *that is, it is impossible to obtain a non-trivial bound that is a function of only moments.*

Results in [6, 19] *require the function to be convex, therefore not useful for this example. Since the domain is not* $[0, A)$ *or* $(0, A]$ *as required by* [2], *or* $[0, +\infty)$ *as required by* [1], *these results do not apply. Both our result and* [13, 20] *are able to get* $\mathcal{J}(\cos, X \sim \mathcal{P}) \geq -\frac{\sigma_2^2}{2}$. *Since* $\cos(x) - 1 \leq 0$, *it is not hard to see that* $\mathcal{J}(\cos, X \sim \mathcal{P}) \leq 0$, *which is also given by the result of* [13].

Example 1.3. *Consider* $\log(x)$ *and random variable* $X \in [a, +\infty)$ *with* $a > 0$ *that has* $\mathbb{E}[X] = 1$. *Since* $\log'(x) = 1 \neq 0$, *we study* $g(x) = \log(x) - (x - 1)$ *instead, which preserves gap behavior as in the* $\sin$ *example above. Note that* $\log(x) = (x - 1) - \frac{1}{2}(x - 1)^2 + \cdots$. *By choosing* $\alpha = n = 2$, *we have*

$$|\mathcal{J}(\log, X \sim \mathcal{P})| \leq \frac{a - 1 - \log(a)}{(1 - a)^2}\sigma_2^2$$

---

[1]To construct such a probability distribution, we can choose a discrete $\mathcal{P}$ whose support is a subset of $\{\pm 2\pi k | k \in \mathbb{N}\}$. By appropriate choice of probability values at discrete points, it is possible to make $\mathcal{P}$ have any desired set of moments.

i.e. $|\mathcal{J}(\log, X \sim \mathcal{P})|$ will decrease no slower than $\sim \sigma_2^2$, i.e. the variance of the distribution, as $\sigma_2 \to 0$. Since the $\log$ function is concave, our lower bound is useful (see Section 3.2). Choosing $\alpha = 2$ and $\beta = 1$, we get

$$-\mathcal{J}(\log, X \sim \mathcal{P}) \geq \frac{1}{2} \cdot \frac{\sigma_1^2}{1 + \sigma_1}$$

whereby the Jensen gap approaches 0 no faster than σ_1^2 as $\sigma_1 \to 0$.

The estimate given by [6] is

$$\mathcal{J}(\log, X \sim \mathcal{P}) \leq \frac{1}{2a^2} \min_{c \geq a} \left\{ \mathbb{E}\left[(X - c)^2\right] + (1 - c)^2 \right\} = \frac{\sigma_2^2}{2a^2}$$

for $a < 1$

$$\frac{a - 1 - \log(a)}{(1 - a)^2} < \frac{1}{2a^2}$$

which means that we get a better result than [6]. Since the domain is not $[0, A)$ or $(0, A]$ as required by [2], or $(0, +\infty)$ as required by [19], or $[0, +\infty)$ as required by [1], these results do not apply. The result in [20] in this case falls back to Jensen's inequality $\mathcal{J}(\log, X \sim \mathcal{P}) \geq 0$. The result in [13] gives the same upper bound as ours and fall back to Jensen's inequality for the lower bound.

Example 1.4. *Consider $f(x) = \sqrt{x}$ and random variables on $[0, +\infty)$ that has mean at 1. Since $f'(1) = \frac{1}{2} \neq 0$, we study $g(x) = \sqrt{x} - \frac{x-1}{2}$ instead. Note that $\sqrt{x} = 1 + \frac{x-1}{2} - \frac{1}{8}(x - 1)^2 + \cdots$. By choosing $n = \alpha = 2$, we see that $\left|\mathcal{J}(\sqrt{\cdot}, X \sim \mathcal{P})\right| \leq \frac{\sigma_2^2}{2}$, i.e. $\left|\mathcal{J}(\sqrt{\cdot}, X \sim \mathcal{P})\right|$ will decrease no slower than $\sim \sigma_2^2$, i.e. the variance of the distribution. Also, since $\sqrt{x}$ is concave, our lower bound is useful. By setting $\alpha = 2$ and $\beta = 1$, we get*

$$-\mathcal{J}(\sqrt{\cdot}, X \sim \mathcal{P}) \geq \frac{1}{8} \cdot \frac{\sigma_1^2}{1 + \sigma_1}$$

whereby the Jensen gap will approach 0 no faster than σ_1^2.

Since the second order derivative is not bounded, [6] does not apply to this example. Since $\frac{\sqrt{x} - \sqrt{0}}{x}$ is not defined on 0 and does not have a power series on 0, results in [2] do not apply. Since the domain is not $(0, +\infty)$ as required by [19], that result does not apply. Since $-\sqrt{\cdot}$ is superquadratic, [1] applies and has a result $-\mathcal{J}(\sqrt{\cdot}, X \sim \mathcal{P}) \geq -\mathbb{E}\left[\sqrt{|X - 1|}\right] = -\sigma_{1/2}^{1/2}$, which is not even an improvement of Jensen's inequality $-\mathcal{J}(\sqrt{\cdot}, X \sim \mathcal{P}) \geq 0$. Again, the result in [20] falls back to Jensen's inequality $-\mathcal{J}(\sqrt{\cdot}, X \sim \mathcal{P}) \geq 0$. The result in [13] gives the same upper bound as ours and fall back to Jensen's inequality for the lower bound.

Example 1.5. *Consider $f(x) = x^4$ and random variables that have mean at 1. Since $f'(1) = 4 \neq 0$, we study $g(x) = x^4 - 4(x - 1)$ instead. By choosing $\alpha = 2$, $n = 4$, we see that $|\mathcal{J}(f, X \sim \mathcal{P})| \leq \frac{7 + \sqrt{41}}{2}(1 + \sigma_4^2)\sigma_4^2$, i.e. $|\mathcal{J}(f, X \sim \mathcal{P})|$ will decrease no slower than $\sim \sigma_4^2$. Also, since $f(x) = x^4$ is convex, our lower bound is useful. By choosing $\alpha = \beta = 2$, we get $\mathcal{J}(f, X \sim \mathcal{P}) \geq 2\sigma_1^2$ whereby the Jensen gap will decrease to 0 no faster than σ_1^2.*

Since the second order derivative is not bounded, results in [6] do not apply to this example. Since the domain is not $[0, A)$ or $(0, A]$ as required by [2], or $(0, +\infty)$ as required by [19], or $[0, +\infty)$ as required by [1], these results do not apply. Again [20] falls back to Jensen's inequality $\mathcal{J}(f, X \sim \mathcal{P}) \geq 0$. The Jensen gap of x^4 on $\mathbb{R}$ with $\mu = 1$. The result in [13] gives a trivial upper bound $\mathcal{J}(f, X \sim \mathcal{P}) \leq +\infty$ and a lower bound $\mathcal{J}(f, X \sim \mathcal{P}) \geq 2\sigma_2^2$. This lower bound gives better numerical values and is usually easier to compute compared with ours.

2. FIRST MAIN RESULT: UPPER BOUND

We first prove an upper bound on the Jensen gap and discuss the tightness of this bound in Section 2.1. Note that our upper bound is useful even when the function f is not convex. Next, we show how to use shifts to expand the scope of our upper bound in Section 2.2.

The upper bound in the following theorem holds for any probability distribution as long as the relevant moments are well defined.

Theorem 2.1. *If $f : I \to \mathbb{R}$, where I is a closed subset of $\mathbb{R}$ and $\mu \in I$, satisfies the following conditions:*

(1) *f is bounded on any compact subset of I.*
(2) *$|f(x) - f(\mu)| = O(|x - \mu|^\alpha)$ at $x \to \mu$ for $\alpha > 0$.*
(3) *$|f(x)| = O(|x|^n)$ as $x \to \infty$ for $n \geq \alpha$*

then for a random variable X with probability distribution $\mathcal{P}$ and expectation μ, the following inequality holds:

$$(2.1) \qquad |\mathbb{E}[f(X)] - f(\mu)| \leq M(\sigma_\alpha^\alpha + \sigma_n^n) \leq M\left(1 + \sigma_n^{n-\alpha}\right)\sigma_n^\alpha$$

where $M = \sup_{x \in I \backslash \{\mu\}} \frac{|f(x) - f(\mu)|}{|x - \mu|^\alpha + |x - \mu|^n}$ does not depend on the probability distribution $\mathcal{P}$.

Proof. We begin by showing that $g(x) = \frac{|f(x) - f(\mu)|}{|x - \mu|^\alpha + |x - \mu|^n}$ is bounded on $I \backslash \{\mu\}$:

Since $|f(x)| = O(|x|^n)$ and $|x - \mu|^\alpha + |x - \mu|^n = \Theta(|x|^n)$ at $x \to \infty$, there exists d_1 that $g(x)$ is bounded on $|x - \mu| \geq d_1$. Also, at $x \to \mu$, since $|f(x) - f(\mu)| = O(|x - \mu|^\alpha)$ and $|x - \mu|^\alpha + |x - \mu|^n = \Theta(|x - \mu|^\alpha)$, there exists $d_2 < d_1$ such that $g(x)$ is bounded on $|x - \mu| \leq d_2$. Since the set $d_1 \leq |x - \mu| \leq d_2$ is compact, the numerator is bounded on this set, and the denominator is bounded from below by $d_2^\alpha + d_2^n$, $g(x)$ is therefore bounded on $d_1 \leq |x - \mu| \leq d_2$. In summary, $g(x)$ is bounded on $\mathbb{R} \backslash \{0\}$.

Let $M = \sup_{x \in I \backslash \{\mu\}} \frac{|f(x) - f(\mu)|}{|x - \mu|^\alpha + |x - \mu|^n}$, we then have:

$$|f(x) - f(\mu)| = (|x - \mu|^\alpha + |x - \mu|^n) \cdot g(x) \leq M(|x - \mu|^\alpha + |x - \mu|^n)$$

So the Jensen gap is

$$|\mathbb{E}[f(X)] - f(\mathbb{E}[X])| = \int_{\mathbb{R}} |f(X) - f(\mu)| \, d\mathcal{P}(X)$$

$$\leq M \int_{\mathbb{R}} |X - \mu|^\alpha + |X - \mu|^n \, d\mathcal{P}(X) \leq M(\sigma_\alpha^\alpha + \sigma_n^n)$$

Also note that $\sigma_\alpha \leq \sigma_n$ for $\alpha \leq n$, we then have

$$M(\sigma_\alpha^\alpha + \sigma_n^n) \leq M\left(1 + \sigma_n^{n-\alpha}\right)\sigma_n^\alpha$$

∎

If we are only interested in distributions concentrated around μ, we can further simplify the inequality to the corollary below:

Corollary 1. *For functions that satisfy the condition in Theorem 2.1, there exists a positive number M' independent of the distribution such that*

$$(2.2) \qquad |\mathbb{E}[f(X)] - f(\mu)| \leq M'\sigma_n^\alpha$$

for sufficiently small σ_n,

2.1. **Tightness of upper bound.** We show that modulo the preceding constant M', the inequality 2.2 is sharp.

Proposition 1. *Let $f(x)$ be a function that satisfies the condition in Theorem 2.1 with $I = \mathbb{R}$ and has $|f(x) - f(\mu)| \geq M |x - \mu|^{\alpha}$ on $x \in \mathbb{R}$ for some $M > 0$. Then for any $\sigma_n > 0$ there exists probability distribution $\mathcal{P}$ that makes*

$$|\mathbb{E}[f(X)] - f(\mathbb{E}[X])| \geq M\sigma_n^{\alpha}$$

Proof. Let $\mathcal{P}$ be discrete with

$$\mathcal{P}(\{\mu + \sigma_n\}) = \mathcal{P}(\{\mu - \sigma_n\}) = \frac{1}{2}$$

$$\mathcal{P}(\mathbb{R}\backslash\{\mu + \sigma_n, \mu - \sigma_n\}) = 0$$

The Jensen gap can then be written as

$$|\mathbb{E}[f(X)] - f(\mathbb{E}[X])| \geq M \int |X - \mu|^{\alpha} \, d\mathcal{P}(X) = M\sigma_n^{\alpha}$$

∎

The following proposition shows that the σ_n in inequality (2.2) cannot be replaced by σ_{β} for any $\beta < n$:

Proposition 2. *There exists a function f that satisfies the condition in Theorem 2.1 such that for any $0 < \beta < n$ and $\sigma_n > 0$, there exists a probability distribution $\mathcal{P}$ that makes $\frac{|\mathcal{J}(f, X \sim \mathcal{P})|}{\sigma_{\beta}^{\alpha}}$ arbitrarily large.*

Proof. Let $\mathcal{P}$ be discrete with

$$\mathcal{P}(\{\mu\}) = 1 - p$$

$$\mathcal{P}(\{\mu + a\}) = \mathcal{P}(\{\mu - a\}) = p/2$$

$$\mathcal{P}(\mathbb{R}\backslash\{\mu, \mu + a, \mu - a\}) = 0$$

Then σ_{β} can be written as

$$\sigma_{\beta} = \sqrt[\beta]{p} \cdot a$$

Let $f(x) = |x - \mu|^{\alpha} + |x - \mu|^{n}$. The absolute value of the Jensen gap can be written as

$$|\mathcal{J}(f, X \sim \mathcal{P})| = p \cdot (a^{\alpha} + a^{n}).$$

Then the ratio

$$\frac{|\mathcal{J}(f, X \sim \mathcal{P})|}{\sigma_{\beta}^{\alpha}} = p^{1 - \frac{\alpha}{\beta}} \cdot \left(1 + a^{n - \alpha}\right).$$

Note that $\sigma_n = \sqrt[n]{p} \cdot a$. We then have $a = \frac{\sigma_n}{\sqrt[n]{p}}$. Then we can write the ratio as

$$\frac{|\mathcal{J}(f, X \sim \mathcal{P})|}{\sigma_{\beta}^{\alpha}} = p^{1 - \frac{\alpha}{\beta}} \cdot \left(1 + \frac{\sigma_n^{n - \alpha}}{p^{1 - \frac{\alpha}{n}}}\right) = p^{1 - \frac{\alpha}{\beta}} + p^{\alpha\left(\frac{1}{n} - \frac{1}{\beta}\right)} \cdot \sigma_n^{n - \alpha}.$$

Since $\frac{1}{n} - \frac{1}{\beta} < 0$ and p can take any value in $(0, 1)$, it is always possible to make the ratio arbitrarily large. ∎

2.2. **Expanding the scope of the upper bound by linear shifts .** When referring to random variables with distribution peaked around its mean, i.e. random variables with small σ_n, the larger the α in inequality (2.2), the tighter the upper bound. However, for many f, it is impossible to find an $\alpha > 1$. For example, for functions that are differentiable at μ and have a $f'(\mu) \neq 0$, the largest α we can obtain is $\alpha = 1$. Also, for the case of convex functions that are strictly increasing at $x = \mu$, it is impossible to find an $\alpha > 1$:

Proposition 3. *Let $f(x)$ be a convex function that is strictly increasing near μ. Then for any $\alpha > 1$, we have*

$$\lim_{x \to \mu} \frac{|x - \mu|^{\alpha}}{f(x) - f(\mu)} = 0$$

Proof. Since f is convex and strictly increasing, we have $f'_+(\mu) > 0$ and $f'_-(\mu) > 0$. So

$$\lim_{x \to \mu^+} \frac{|x - \mu|^{\alpha}}{f(x) - f(\mu)} = \lim_{x \to \mu^+} (x - \mu)^{\alpha-1} \cdot \frac{x - \mu}{f(x) - f(\mu)} = 0 \cdot \frac{1}{f'_+(\mu)} = 0$$

Same argument holds for $x \to \mu^-$. ∎

Although the inability to get an $\alpha > 1$ seems to be a major limitation, fortunately for most cases we can eliminate this limitation by shifting the function by a linear function, because this does not change its convexity or the Jensen gap. For functions that are differentiable at $x = \mu$, from Taylor's theorem with Peano's form of remainder, we know that

$$f(x) = f(\mu) + f'(\mu)(x - \mu) + o(x - \mu)$$

We can therefore study $g(x) = f(x) - f'(\mu)(x - \mu)$ instead of $f(x)$. We will then have $g(x) - g(\mu) = o(x - \mu)$, which has an α value at least as large as $f(x)$. If further $f(x)$ has well defined second derivative, we then have

$$f(x) = f(\mu) + f'(\mu)(x - \mu) + \frac{f''(\xi_L)}{2}(x - \mu)^2$$

that is

$$(2.3) \qquad\qquad g(x) - g(\mu) = \frac{f''(\xi_L)}{2}(x - \mu)^2$$

which implies $\alpha = 2$. If $f''(\mu) = 0$, we can apply similar arguments to higher order derivatives to find the best α.

Note that if we define $h(x; \mu) \equiv \frac{f''(\xi_L)}{2} \equiv \frac{f(x)-f(\mu)-f'(\mu)(x-\mu)}{(x-\mu)^2}$, then (2.3) can be written as $g(x) - g(\mu) = h(x; \mu)(x - \mu)^2$, which further gives

$$(2.4) \qquad\qquad \inf_x h(x; \mu) \cdot \mathrm{Var}[X] \leq \mathcal{J}(f, X \sim \mathcal{P}) \leq \sup_x h(x; \mu) \cdot \mathrm{Var}[X]$$

as shown in[13]. If $|f(x)| \neq O(x^2)$ at $x \to \infty$, then $\sup_x h(x; \mu) = +\infty$ or $\inf_x h(x; \mu) = -\infty$ or both, which means at least half of (2.4) will become a trivial inequality $-\infty \leq \mathcal{J}(f, X \sim \mathcal{P})$ or $\mathcal{J}(f, X \sim \mathcal{P}) \leq +\infty$. On the other hand, if $|f(x)| = O(x^2)$ at $x \to \infty$, we then have $n = 2$. If this is the case, the preceding constant M in Theorem 2.1 can then be written as $M = \frac{1}{2}\sup_x |h(x; \mu)|$ and Equation (2.1) therefore becomes $-\sup_x |h(x; \mu)| \cdot \sigma_2^2 \leq \mathcal{J} \leq \sup_x |h(x; \mu)| \cdot \sigma_2^2$, which is equivalent to (2.4) in half or in full[2]. Due to these connections,

---

[2]If $|\sup_x h(x; \mu)| > |\inf_x h(x; \mu)|$, then we must have $\sup_x |h(x; \mu)| = \sup_x h(x; \mu)$, which means the $\mathcal{J}(f, X \sim \mathcal{P}) \leq$ part of Theorem 2.1 and of (2.4) are equivalent. If $|\sup_x h(x; \mu)| < |\inf_x h(x; \mu)|$, then we must have $-\sup_x |h(x; \mu)| = \inf_x h(x; \mu)$, which means the $\leq \mathcal{J}(f, X \sim \mathcal{P})$ part of Theorem 2.1 and of (2.4) are equivalent. If $|\sup_x h(x; \mu)| = |\inf_x h(x; \mu)|$ and $h(x; \mu)$ is not constant, then we must have $\sup_x |h(x; \mu)| = -\inf_x h(x; \mu) = \sup_x h(x; \mu)$, hence in this case, Theorem 2.1 is fully equivalent to (2.4).

Lemma 1 in [13] gives a convenient way to compute the M in equation (2.1) when the $f'(x)$ is convex or concave.

3. SECOND MAIN RESULT: LOWER BOUND

We first prove our lower bound for conditions similar to the upper bound case. The tightness of this bound will be discussed in Section 3.1 followed, in Section 3.2, by strong implications for convex functions, and expanding the scope via linear function shifts.

The lower bound given in the following theorem holds for any probability distribution as long as the relevant moments are well-defined.

Theorem 3.1. *If function $f : I \to \mathbb{R}$, where I is a closed subset of $\mathbb{R}$ and $\mu \in I$, satisfies the following conditions:*

(1) $f(x) - f(\mu) > 0$ *at* $x \neq \mu$
(2) $f(x) - f(\mu) = \Omega\left(|x - \mu|^{\alpha}\right)$ *at* $x \to \mu$ *for* $\alpha > 0$
(3) $f(x) - f(\mu) = \Omega\left(|x - \mu|^{\beta}\right)$ *at* $x \to \infty$ *for* $0 \leq \beta \leq \alpha$

then for random variable X with probability distribution $\mathcal{P}$ that has expectation μ, the following inequality holds:

$$(3.1) \qquad \mathcal{J}(f, X \sim \mathcal{P}) \geq M \frac{\sigma_{\alpha/2}^{\alpha}}{1 + \sigma_{\alpha-\beta}^{\alpha-\beta}}$$

where $M = \inf_{x \in I \setminus \{\mu\}} \left\{ [f(x) - f(\mu)] \cdot \left(\frac{1}{|X - \mu|^{\beta}} + \frac{1}{|X - \mu|^{\alpha}} \right) \right\} > 0$ *does not depend on the probability distribution $\mathcal{P}$.*

Proof. Let

$$g(x) = \left(\frac{1}{|x - \mu|^{\beta}} + \frac{1}{|x - \mu|^{\alpha}} \right)^{-1}$$

from the definition of M, we know that $f(x) - f(\mu) \geq M \cdot g(x)$.

We first prove that $M > 0$. It is easy to see that $g(x)$ is positive at $x \neq \mu$, $g(x) = \Theta\left(|x - \mu|^{\alpha}\right)$ at $x \to \mu$, and $g(x) = \Theta\left(|x - \mu|^{\beta}\right)$ at $x \to \infty$. Therefore, there exists positive M_1, M_2 and $d_1 \leq d_2$ such that $f(x) - f(\mu) \geq M_1 \cdot g(x)$ at $|x - \mu| \leq d_1$ and $f(x) - f(\mu) \geq M_2 \cdot g(x)$ at $|x - \mu| \geq d_2$. Since $d_1 \leq |x - \mu| \leq d_2$ is compact and both $f(x) - f(\mu)$ and $g(x)$ are positive in this interval, there exists $M_3 > 0$ such that $f(x) - f(\mu) \geq M_3 \cdot g(x)$. Taking $M' = \min\{M_1, M_2, M_3\} > 0$, we have $f(x) - f(\mu) \geq M'g(x)$. That is, $\frac{f(x) - f(\mu)}{g(x)}$ is bounded from below by some positive number. Therefore

$$M = \inf_{x \in I \setminus \{\mu\}} \frac{f(x) - f(\mu)}{g(x)} > 0$$

Since $f(x) - f(\mu) \geq M \cdot g(x)$, we have

$$(3.2) \quad \mathcal{J}(f, X \sim \mathcal{P}) \geq M \int \left(\frac{1}{|X - \mu|^{\beta}} + \frac{1}{|X - \mu|^{\alpha}} \right)^{-1} d\mathcal{P}(X)$$

$$= M \int \frac{|X - \mu|^{\alpha}}{|X - \mu|^{\alpha-\beta} + 1} d\mathcal{P}(X)$$

The Cauchy–Schwarz inequality can be used to simplify the above inequality: Let

$$g_1(X) = \sqrt{\frac{|X - \mu|^\alpha}{|X - \mu|^{\alpha - \beta} + 1}}$$

and

$$g_2(X) = \sqrt{|X - \mu|^{\alpha - \beta} + 1}$$

Cauchy–Schwarz inequality

$$\mathbb{E}\left[g_1^2(X)\right] \mathbb{E}\left[g_2^2(X)\right] \geq \mathbb{E}\left[g_1(X)g_2(X)\right]^2$$

can be rewritten as

$$\mathbb{E}\left[g_1^2(X)\right] \geq \frac{\mathbb{E}\left[g_1(X)g_2(X)\right]^2}{\mathbb{E}\left[g_2^2(X)\right]}$$

Note that

$$\mathbb{E}\left[g_1^2(X)\right] = \int \frac{|X - \mu|^\alpha}{|X - \mu|^{\alpha - \beta} + 1} d\mathcal{P}(X)$$

$$\mathbb{E}\left[g_2^2(X)\right] = \int \left(|X - \mu|^{\alpha - \beta} + 1\right) d\mathcal{P}(X) = 1 + \sigma_{\alpha - \beta}^{\alpha - \beta}$$

$$\mathbb{E}\left[g_1(X)g_2(X)\right]^2 = \left(\int |X - \mu|^{\alpha/2} d\mathcal{P}(X)\right)^2 = \sigma_{\alpha/2}^\alpha$$

We therefore have

$$\int \frac{|X - \mu|^\alpha}{|X - \mu|^{\alpha - \beta} + 1} d\mathcal{P}(X) \geq \frac{\sigma_{\alpha/2}^\alpha}{1 + \sigma_{\alpha - \beta}^{\alpha - \beta}}$$

Plugging into (3.2), we have

$$\mathcal{J}(f, X \sim \mathcal{P}) \geq M \frac{\sigma_{\alpha/2}^\alpha}{1 + \sigma_{\alpha - \beta}^{\alpha - \beta}}$$

∎

Note that if we replace all $f(x) - f(\mu)$ with $f(\mu) - f(x)$, and at the same time replace $\mathcal{J}(f, X \sim \mathcal{P})$ with $-\mathcal{J}(f, X \sim \mathcal{P})$, Theorem 3.1 still holds. Also note that by replacing the Cauchy–Schwarz inequality with Hölder's inequality in the proof of the above theorem, we can get a more general but less pleasing result:

Theorem 3.2. *The inequality* (3.1) *in Theorem 3.1 can be replaced by the following inequality:*

$$(3.3) \qquad \mathcal{J}(f, X \sim \mathcal{P}) \geq M \frac{\left[\sum_{l=0}^{(k+1)/q-1} \binom{(k+1)/q - 1}{l} \sigma_{\alpha/p + l(\alpha - \beta)}^{\alpha/p + l(\alpha - \beta)}\right]^p}{\left[\sum_{l=0}^{k} \binom{k}{l} \sigma_{l(\alpha - \beta)}^{l(\alpha - \beta)}\right]^{p/q}}$$

where $k \geq 1$ is an integer, q can be any positive factor of $(k+1)$ except 1, and $p = \frac{q}{q-1}$.

Proof. Following the same steps as in the Cauchy–Schwarz version, but this time introducing a new integral parameter $k \geq 1$, and setting

$$g_1(X) = \sqrt[p]{\frac{|X - \mu|^\alpha}{|X - \mu|^{\alpha - \beta} + 1}}$$

and

$$g_2(X) = \sqrt[q]{\left(|X - \mu|^{\alpha - \beta} + 1\right)^k}$$

we have

$$\mathbb{E}\left[g_1^p(X)\right] = \int \frac{|X - \mu|^\alpha}{|X - \mu|^{\alpha-\beta} + 1} d\mathcal{P}(X)$$

$$\mathbb{E}\left[g_2^q(X)\right] = \int \left(|X - \mu|^{\alpha-\beta} + 1\right)^k d\mathcal{P}(X) = \sum_{l=0}^{k} \binom{k}{l} \sigma_{l(\alpha-\beta)}^{l(\alpha-\beta)}$$

$$\mathbb{E}\left[g_1(X)\,g_2(X)\right] = \int |X - \mu|^{\alpha/p}\left(|X - \mu|^{\alpha-\beta} + 1\right)^{k/q - 1/p} d\mathcal{P}(X)$$

$$= \int |X - \mu|^{\alpha/p}\left(|X - \mu|^{\alpha-\beta} + 1\right)^{(k+1)/q - 1} d\mathcal{P}(X)$$

$$= \sum_{l=0}^{(k+1)/q - 1} \binom{(k+1)/q - 1}{l} \sigma_{\alpha/p + l(\alpha-\beta)}^{\alpha/p + l(\alpha-\beta)}.$$

From Hölder's inequality, we know that

$$\mathbb{E}\left[g_1^p(X)\right] \geq \frac{\mathbb{E}\left[g_1(X)\,g_2(X)\right]^p}{\mathbb{E}\left[g_2^q(X)\right]^{\frac{p}{q}}} = \frac{\left[\sum_{l=0}^{(k+1)/q - 1} \binom{(k+1)/q - 1}{l} \sigma_{\alpha/p + l(\alpha-\beta)}^{\alpha/p + l(\alpha-\beta)}\right]^p}{\left[\sum_{l=0}^{k} \binom{k}{l} \sigma_{l(\alpha-\beta)}^{l(\alpha-\beta)}\right]^{p/q}}$$

which immediately yields inequality (3.3). ∎

Although general, inequality (3.3) is too cumbersome to be useful. To simplify it, we can take $q = k + 1$ and therefore $p = 1 + \frac{1}{k}$. We then have

$$(3.4) \qquad \mathcal{J}(f, X \sim \mathcal{P}) \geq M \frac{\sigma_{\alpha/\left(1+\frac{1}{k}\right)}^{\alpha}}{\left[\sum_{l=0}^{k} \binom{k}{l} \sigma_{l(\alpha-\beta)}^{l(\alpha-\beta)}\right]^{1/k}}$$

Note that applying inequality (3.4) to $f(x) = |x|^\alpha$, we obtain

$$(3.5) \qquad \sigma_\alpha \geq \sigma_{\alpha/\left(1+\frac{1}{k}\right)}$$

which is a special case of the inequality

$$\mathbb{E}\left[|X|^r\right] \leq \mathbb{E}\left[|X|^s\right]^{\frac{r}{s}}$$

for $0 < r < s$.

3.1. **Tightness of the lower bound.** Since inequality (3.5) is a special case of (3.4) and since (3.5) is sharp, it follows that (3.4) is sharp.

In inequality (3.4), as the centered absolute moments decrease to 0, since the denominator decreases to 1, it is the numerator that characterizes how fast the Jensen gap decreases to 0. Since $\sigma_r \leq \sigma_s$ for $r \leq s$, having a larger subscript in the numerator means a tighter result. In (3.4), the subscript of the numerator can be increased to a value arbitrarily close to α by choosing larger k, but as a side effect, this also brings higher orders of moments into the denominator.

Therefore, a natural question is whether we can increase the subscript of the numerator of (3.4) without bringing in higher orders of moments? The following proposition shows that the answer is no, by showing that if we increase the subscript higher than that proposed in (3.4), it is possible to construct a sequence of probability distributions that make the moments in the denominator decrease to 0 while at the same time making the ratio between the Jensen gap and

the numerator go to zero (therefore making it impossible to find a M to make the $\geq$ in (3.4) hold):

Proposition 4. *Let $f(x) = \Theta(|x|^{\beta})$ at $x \to \infty$ be a function that satisfies the condition in Theorem 3.1. Then for any $q > \alpha / \left(1 + \frac{1}{k}\right)$, there exists a sequence of probability distributions $\mathcal{P}^{(1)}, \mathcal{P}^{(2)}, \ldots$ such that $\sigma_r^{(j)}$ is non-increasing with respect to j for all $r \leq k(\alpha - \beta)$ and*

$$\lim_{j \to +\infty} \frac{\mathcal{J}\left(f, X \sim \mathcal{P}^{(j)}\right)}{\left[\sigma_q^{(j)}\right]^{\alpha}} = 0$$

Proof. Let $m = k(\alpha - \beta)$. Without loss of generality, assume $\mu = 0$, $f(x)$ is even, and $f(0) = 0$. Let $\mathcal{P}$ be a discrete probability distribution that has

$$\mathcal{P}\left(\{j\}\right) = \mathcal{P}\left(\{-j\}\right) = \frac{1}{2j^m}$$

$$\mathcal{P}\left(\{0\}\right) = 1 - \frac{1}{j^m}$$

$$\mathcal{P}\left(\mathbb{R} \backslash \{0, \pm j\}\right) = 0$$

It is easy to see that

$$\sigma_r^{(j)} = j^{1 - \frac{m}{r}}$$

does not increase as j increases for $r \leq m$, and

$$\mathcal{J}\left(f, X \sim \mathcal{P}^{(j)}\right) = \frac{f(j)}{j^m} = \Theta\left(j^{\beta - m}\right)$$

at $j \to +\infty$. We then have

$$\frac{\mathcal{J}\left(f, X \sim \mathcal{P}^{(j)}\right)}{\left[\sigma_q^{(j)}\right]^{\alpha}} = \Theta\left[j^{\beta - m - \alpha\left(1 - \frac{m}{q}\right)}\right]$$

In the case that $q > \alpha / \left(1 + \frac{1}{k}\right)$, we have $\beta - m - \alpha\left(1 - \frac{m}{q}\right) < 0$. Therefore

$$\Theta\left[j^{\beta - m - \alpha\left(1 - \frac{m}{q}\right)}\right] \to 0$$

as $j \to +\infty$. ∎

3.2. **The lower bound for convex functions.**

The conditions for Theorem 3.1 are hard for a general function to satisfy. In fact, Jensen's inequality only holds for convex functions, so it is not surprising that we are unable to obtain a lower bound of the Jensen gap. In this section, we show that convexity implies the conditions in Theorem 3.1. The argument in this section also applies to concave functions.

In order for the condition in Theorem 3.1 to be satisfied, a convex function $f(x)$ needs to be non-increasing at $(-\infty, \mu]$ and non-decreasing at $[\mu, +\infty)$. The following proposition shows that it is always possible to shift a convex function by a linear function to make it so.

Proposition 5. *For any convex function $f(x)$, and any real number a satisfying $f'_{-}(\mu) \leq a \leq f'_{+}(\mu)$, the linear shift $g(x) = f(x) - a(x - \mu)$ is non-increasing at $(-\infty, \mu]$ and non-decreasing at $[\mu, +\infty)$. Specially, if $f(x)$ is differentiable at μ, a is unique and given by $a = f'(x)$.*

Proof. For $\mu \le x < x'$, we have

$$\frac{g(x') - g(x)}{x' - x} = \frac{f(x') - f(x)}{x' - x} - a \ge f'_+(\mu) - a \ge 0$$

That is, $g(x') \ge g(x)$. Similar argument applies to the $x \le \mu$ half of $g(x)$. ∎

The convexity also implies that α can only take values from $[1, +\infty)$, as shown in the following proposition:

Proposition 6. *There does not exist any convex function that has $f(x) - f(\mu) = \Omega(|x - \mu|^\alpha)$ at $x \to \mu$ with $\alpha < 1$.*

Proof. Since $f(x) - f(\mu) = \Omega(|x - \mu|^\alpha)$ as $x \to \mu$, there exists positive number d and M such that $f(x) - f(\mu) \ge M|x - \mu|^\alpha$ at $|x - \mu| \le d$. Then for any x that has $\mu < x < \mu + d$, we have

$$\frac{f(x) - f(\mu)}{x - \mu} \ge M(x - \mu)^{\alpha - 1}$$

Since $\frac{f(x) - f(\mu)}{x - \mu}$ is non-decreasing with respect to x, if $\alpha < 1$, $(x - \mu)^{\alpha - 1}$ will becomes arbitrarily high as $x \to \mu^+$, making it impossible to for the above inequality to be true as x decreases to μ. ∎

The following proposition shows that for convex functions, it is possible to find a β at least as large as 1:

Proposition 7. *If $f(x)$ is convex, then at $x \to \infty$, $|f(x)|$ is either constant or $\Omega(|x|)$.*

Proof. If $f(x)$ is constant, then this proposition automatically holds true. Otherwise, let $x_0 < x_1$ be two real numbers such that $f(x_0) \ne f(x_1)$. Without loss of generality, let us assume $f(x_0) < f(x_1)$. Since $f(x)$ is convex, then for' any $x > x_1$

$$\frac{f(x) - f(x_0)}{x - x_0} \ge \frac{f(x_1) - f(x_0)}{x_1 - x_0} > 0$$

That is

$$f(x) - f(x_0) \ge \frac{f(x_1) - f(x_0)}{x_1 - x_0} \cdot (x - x_0)$$

i.e., $f(x) = \Omega(x)$ at $x \to +\infty$. Considering all the cases, i.e. when $x \to -\infty$, and when $f(x_0) > f(x_1)$, we get $|f(x)| = \Omega(|x|)$. ∎

Although for the lower bound case our result has no similar equivalence relation with [13] as the one discussed in Section 2.2, the preceding constant M in our Theorem 3.1, after the linear shift, can be written as $M = 2 \cdot \sup_{x \ne \mu} h(x; \mu)$ when $\alpha = \beta = 2$. For this special case, when $f'(x)$ is convex or concave, the Lemma 1 in [13] is still helpful.

4. FURTHER DISCUSSION, CONCLUSION AND OPEN PROBLEMS

The procedure in the proofs of Theorems 2.1 and 3.1 can be thought of as a general scheme, which is also followed by [13], of obtaining bounds on the Jensen gap. This procedure first writes $f(x) - f(\mu)$ as a product of two functions, say $s(x)t(x)$, where the sup and inf of s are easy to compute and the integral $\int t(X) d\mathcal{P}(X)$ can be easily computed or further bounded. We then have

$$\inf s(x) \cdot \int t(X) d\mathcal{P}(X) \le \mathcal{J}(f, X \sim \mathcal{P}) \le \sup s(x) \cdot \int t(X) d\mathcal{P}(X)$$

The above formula gives a very general way to bound the Jensen gap. For example, instead of using $t(x) = |x - \mu|^\alpha + |x - \mu|^n$ as in Theorem 2.1, the reader can choose a more general form

$t\left(x\right) = \sum_{\alpha \leq \eta \leq n} a_\eta \left|x - \mu\right|^\eta$ where values of η and a_η are chosen based on the application to better approximate $f\left(x\right)$, and obtain an improved upper bound

$$\mathcal{J} \leq \sup \frac{f\left(x\right)}{t\left(x\right)} \cdot \left(\sum_{\alpha \leq \eta \leq n} a_\eta \sigma_\eta^\eta\right)$$

Similarly, instead of using $t\left(x\right) = \left(\frac{1}{\left|x-\mu\right|^\alpha} + \frac{1}{\left|x-\mu\right|^\beta}\right)^{-1}$ as in Theorem 3.1, the reader can choose $t\left(x\right) = \left(\sum_{\beta \leq \eta \leq \alpha} \frac{a_\eta}{\left|x-\mu\right|^\eta}\right)^{-1}$ where values of η and a_η depend on the application, and obtain an improved lower bound

$$\mathcal{J} \geq \inf \frac{f\left(x\right)}{t\left(x\right)} \cdot \frac{\sigma_{\alpha/2}^\alpha}{\sum_{\beta \leq \eta \leq \alpha} a_\eta \sigma_{\alpha-\eta}^{\alpha-\eta}}$$

or

$$\mathcal{J} \geq \inf \frac{f\left(x\right)}{t\left(x\right)} \cdot \frac{\sigma_{\alpha/\left(1+\frac{1}{k}\right)}^\alpha}{\left(\sum_{\beta \leq \eta_1, \cdots, \eta_k \leq \alpha} a_{\eta_1} \cdots a_{\eta_k} \sigma_{k\alpha-\eta_1-\cdots-\eta_k}^{k\alpha-\eta_1-\cdots-\eta_k}\right)^{1/k}}$$

We have obtained general upper and lower bounds on Jensen's gap that depend on the asymptotic growth of the function and related moments of the random variable's distribution and compared the new bounds with existing upper and lower bounds. Although fairly general, some conditions in our theorems are still too strong for some situations. For example, in our upper bound, we require the function to grow no faster than polynomial at $x \to \infty$, which excludes some useful functions, such as exponential functions. Also, we require the function to be bounded on any compact subset of $\mathbb{R}$ in our upper bound, which exclude the study of the Jensen gap for functions like $\log(x)$, $\frac{1}{x}$ with random variable X on $(0, +\infty)$. Future work is proposed to extend our results to include such cases.

5. DECLARATIONS

5.1. **Availability of data and material.** Not applicable

5.2. **Competing interests.** The authors declares no competing interests.

5.4. **Authors' contributions.** XG proves most theorems and is responsible for most part of paper writing. MS has major contribution on the organization of the paper and paper writing of the whole paper, and provides critical insights on Section 3.2 and Proposition 3. AER has contribution on technical discussion and supervision of the overall project.

5.6. **Authors information.** Author's email:
   Xiang Gao: qasdfgtyuiop@gmail.com
   Meera Sitharam: sitharam@cise.ufl.edu
   Adrian E. Roitberg: roitberg@ufl.edu

REFERENCES

[1] SHOSHANA ABRAMOVICH, GRAHAM JAMESON, AND GORD SINNAMON, Refining Jensen's inequality, *Bulletin mathématique de la Société des Sciences Mathématiques de Roumanie*, p.p. 3–14, (2004).

[2] SHOSHANA ABRAMOVICH and LARS-ERIK PERSSON, Some new estimates of the 'Jensen gap', *Journal of Inequalities and Applications*, 2016(1):39, (2016).

[3] P. CERONE, S. S. DRAGOMIR, and E. KIKIANTY, Jensen–ostrowski inequalities and integration schemes via the darboux expansion, *Ukrains'kyi Matematychnyi Zhurnal*, 69(08), pp. 1123–1140, (2017).

[4] PIETRO CERONE, SEVER S. DRAGOMIR, and EDER KIKIANTY, Jensen–ostrowski type inequalities and applications for f-divergence measures, *Applied Mathematics and Computation*, 266, pp. 304–315, (2015).

[5] PIETRO CERONE, SEVER S. DRAGOMIR, and EDER KIKIANTY, On inequalities of Jensen-Ostrowski type, *Journal of Inequalities and Applications*, 2015(1):328, (2015).

[6] DANILO COSTARELLI and RENATO SPIGLER, How sharp is the Jensen inequality?, *Journal of Inequalities and Applications*, 2015(1):69, (2015).

[7] THOMAS M. COVER and JOY A. THOMAS, *Elements of information theory*, John Wiley & Sons, 2012.

[8] ARTHUR P. DEMPSTER, NAN M. LAIRD, and DONALD B. RUBIN, Maximum likelihood from incomplete data via the em algorithm, *Journal of the royal statistical society, Series B (methodological)*, pp. 1–38, (1977).

[9] SEVER DRAGOMIR, Jensen and ostrowski type inequalities for general lebesgue integral with applications, *Annales Universitatis Mariae Curie-Sklodowska, section A–Mathematica*, 70(2):29, (2016).

[10] SILVESTRU DRAGOMIR, P. CERONE, and EDER KIKIANTY, Ostrowski and Jensen type inequalities for higher derivatives with applications, *Journal of Inequalities and Special Functions*, 7(1), pp. 61–77, 2016.

[11] C. JARZYNSKI, Nonequilibrium equality for free energy differences, *Phys. Rev. Lett.*, 78, pp. 2690–2693, (1997).

[12] JOHAN LUDWIG WILLIAM VALDEMAR JENSEN, Sur les fonctions convexes et les inégalités entre les valeurs moyennes, *Acta Mathematica*, 30(1), pp. 175–193, (1906).

[13] J. G. LIAO and ARTHUR BERG, Sharpening Jensen's inequality, *The American Statistician*, 73:3, 278-281, (2019) DOI: 10.1080/00031305.2017.1419145.

[14] DANIEL M. PACKWOOD, Moments of sums of independent and identically distributed random variables, *arXiv preprint arXiv:1105.6283*, (2011).

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

[16] SLAVKO SIMIC, On an upper bound for Jensen's inequality, *JIPAM. Journal of Inequalities in Pure & Applied Mathematics [electronic only]*, 10(2): Paper No. 60, 5 p., (2009).

[17] J. MICHAEL STEELE, *The Cauchy-Schwarz Master Class: an Introduction to the Art of Mathematical Inequalities*, Cambridge University Press, 2004.

[18] MARTIN J. WAINWRIGHT, MICHAEL I. JORDAN, et al, Graphical models, exponential families, and variational inference, *Foundations and Trends® in Machine Learning*, 1(1–2), pp. 1–305, (2008).

[19] STEPHEN G. WALKER, *On a lower bound for the Jensen inequality*, *SIAM Journal on Mathematical Analysis*, 46(5), pp. 3151–3157, 2014.

[20] SANJO ZLOBEC, *Jensen's inequality for nonconvex functions*, *Mathematical Communications*, 9(2), pp. 119–124, (2004).