Sharp Comparisons of Jensen Gaps under Atom-Mass Constraints

Denis Sheremet

Abstract

For a convex function f, write

$$J_f(X) = \mathbb{E}f(X) - f(\mathbb{E}X)$$

for the Jensen gap. Classical Jensen–variance refinements compare $J_f(X)$ with

$$\mathrm{Var}(X) = J_{x^2}(X).$$

This paper studies the more general problem of comparing two Jensen gaps,

$$J_f(X) \quad \text{and} \quad J_g(X),$$

for finite atomic random variables whose atom masses are bounded below by a fixed number α. Under a coordinatewise monotonicity assumption on the relative second divided difference

$$R_{f,g}(x, y, z) = \frac{f[x, y, z]}{g[x, y, z]},$$

we prove sharp two-sided inequalities

$$C_\alpha^-(f, g; \mu)J_g(X) \le J_f(X) \le C_\alpha^+(f, g; \mu)J_g(X), \qquad \mathbb{E}X = \mu.$$

The constants are given by one-dimensional two-point formulas, and the extremal laws have atom masses α and $1 - \alpha$. The Jensen–variance case is recovered by taking $g(x) = x^2$. Applications are given to power gaps, exponential gaps, power-divergence comparisons, Hellinger-type divergences, shifted and reciprocal powers, and comparisons between Kullback–Leibler divergence and Pearson chi-square divergence. We also prove a positive-cone closure theorem, which generates broad classes of admissible numerators from elementary admissible families.

1 Introduction

Jensen's inequality states that, for a convex function f,

$$f(\mathbb{E}X) \le \mathbb{E}f(X).$$

The difference

$$J_f(X) = \mathbb{E}f(X) - f(\mathbb{E}X)$$

is the Jensen gap. A classical way to sharpen Jensen's inequality is to compare this gap with the variance of X:

$$J_f(X) \asymp \mathrm{Var}(X).$$

Sharp mean–variance Jensen-type bounds are classical; Pittenger [6] gave a short two-point proof for such inequalities under fixed mean and variance, and noted their relation to the broader theory of Tchebycheff systems developed by Karlin and Studden [1].

The present paper considers a different but related constrained problem. Instead of fixing the variance, we impose a lower atom-mass constraint

$$p_i \geq \alpha.$$

This condition occurs naturally in equal-weight finite inequalities, where $\alpha = 1/N$. It prevents the extremal distribution from using arbitrarily small atoms.

The starting observation is that variance itself is a Jensen gap:

$$\mathrm{Var}(X) = J_{x^2}(X).$$

Thus Jensen–variance bounds are a special case of the more general problem of comparing

$$J_f(X) \quad \text{with} \quad J_g(X).$$

The correct object for this comparison is the relative second divided difference

$$R_{f,g}(x,y,z) = \frac{f[x,y,z]}{g[x,y,z]}.$$

When $g(x) = x^2$, one has $g[x,y,z] = 1$, and hence

$$R_{f,x^2}(x,y,z) = f[x,y,z].$$

The present theorem therefore extends Jensen–variance bounds from the denominator J_{x^2} to a general denominator Jensen gap J_g.

The ratio of divided differences is a standard object in the theory of Cauchy or difference means. The Cauchy mean value theorem for divided differences asserts that, under suitable regularity and nonvanishing assumptions,

$$\frac{[x_1,\ldots,x_n]f}{[x_1,\ldots,x_n]g} = \frac{f^{(n-1)}(t)}{g^{(n-1)}(t)}$$

for some t between the nodes. This is the basis of the Cauchy mean values studied by Leach and Sholander [2] and Losonczi [3, 4]. The present paper uses this divided-difference quotient not as a mean itself, but as the sharp local comparison coefficient for two Jensen gaps.

The main theorem says that if $R_{f,g}$ is coordinatewise monotone, then the sharp constants are obtained from two-point laws with masses

$$\alpha, \qquad 1-\alpha.$$

In the nondecreasing case,

$$C_\alpha^-(f,g;\mu) = \inf_{u<\mu} R_{f,g}\left(u, \mu, \frac{\mu - \alpha u}{1-\alpha}\right),$$

and

$$C_\alpha^+(f,g;\mu) = \sup_{v>\mu} R_{f,g}\left(\frac{\mu - \alpha v}{1-\alpha}, \mu, v\right),$$

with the natural restriction that all points lie in the domain. A reflected theorem holds when $R_{f,g}$ is nonincreasing.

The proof is elementary. It uses three ingredients: the two-point divided-difference identity, a decomposition of the Jensen gap after separating one extreme atom, and the coordinatewise monotonicity of $R_{f,g}$ to compare the remaining tail. The proof is close in spirit to the two-point extremal mechanism appearing in mean–variance Jensen bounds, but here the fixed parameter is the minimal atom mass α, not the variance.

2 Standing assumptions and divided differences

Let $I \subset \mathbb{R}$ be an interval. For distinct points $x, y, z \in I$, define

$$f[x,y] = \frac{f(x) - f(y)}{x - y},$$

and

$$f[x,y,z] = \frac{f[x,y] - f[y,z]}{x - z}.$$

Repeated nodes are interpreted in the usual limiting sense. In particular, if f is differentiable at μ, then

$$f[x,\mu,\mu] = \frac{f(x) - f(\mu) - f'(\mu)(x-\mu)}{(x-\mu)^2}, \qquad x \neq \mu.$$

If $f \in C^2(I)$, then

$$f[x,x,x] = \frac{1}{2} f''(x).$$

We shall use standard properties of divided differences: symmetry in the nodes, linearity in the function, continuity in the nodes, and the integral representation over a simplex. These are recalled, for instance, in Leach and Sholander [2] and Losonczi [3, 4].

All random variables in this paper are finite atomic. Equal support values are merged into a single atom, and the lower atom-mass condition is imposed after merging atoms.

For $0 < \alpha \leq 1/2$ and $\mu \in I$, let $\mathcal{A}_\alpha(\mu)$ denote the class of all nonconstant finite atomic random variables X supported in I such that

$$\mathbb{E}X = \mu,$$

and every distinct atom of X has probability at least α.

We assume throughout the main theorem that

$$g[x,y,z] > 0$$

for every nondegenerate triple in I, and also for repeated nodes by continuity. In particular,

$$J_g(X) > 0$$

for every nonconstant admissible X.

Define the relative second divided difference

$$R_{f,g}(x,y,z) = \frac{f[x,y,z]}{g[x,y,z]}.$$

Since second divided differences are symmetric in the nodes, $R_{f,g}$ is symmetric as well.

Remark 2.1. *The direct hypothesis of the main theorem is coordinatewise monotonicity of $R_{f,g}$. In the nondecreasing case, increasing one of the three arguments while keeping the other two fixed does not decrease $R_{f,g}$; in the nonincreasing case, it does not increase $R_{f,g}$. We do not use as a general principle the implication*

$$\frac{f''}{g''} \text{ monotone} \quad \Longrightarrow \quad R_{f,g} \text{ coordinatewise monotone.}$$

For each application, the required coordinatewise monotonicity is verified separately or derived from a known mean-value theorem.

3 Basic identities

Lemma 3.1 (Bregman divided-difference identity)**.** *Let X be finite atomic with $\mathbb{E}X = \mu$. Then*

$$J_f(X) = \mathbb{E}\Big[(X-\mu)^2 f[X,\mu,\mu]\Big].$$

Proof. Since $\mathbb{E}(X-\mu) = 0$,

$$J_f(X) = \mathbb{E}f(X) - f(\mu) = \mathbb{E}\left(f(X) - f(\mu) - f'(\mu)(X-\mu)\right).$$

The identity

$$f(x) - f(\mu) - f'(\mu)(x-\mu) = (x-\mu)^2 f[x,\mu,\mu]$$

gives the claim. □

Lemma 3.2 (Two-point identity)**.** *Let $a < \mu < b$, and let*

$$Z = \begin{cases} a, & p, \\ b, & q, \end{cases} \qquad p + q = 1, \qquad pa + qb = \mu.$$

Then

$$J_f(Z) = \operatorname{Var}(Z) f[a,\mu,b].$$

Consequently,

$$\frac{J_f(Z)}{J_g(Z)} = R_{f,g}(a,\mu,b).$$

Proof. From $pa + qb = \mu$,

$$p = \frac{b-\mu}{b-a}, \qquad q = \frac{\mu-a}{b-a}.$$

The interpolation identity for second divided differences gives

$$pf(a) + qf(b) - f(\mu) = \Big(p(a-\mu)^2 + q(b-\mu)^2\Big)\, f[a,\mu,b].$$

The factor in parentheses is $\operatorname{Var}(Z)$. Applying the same identity to g gives the ratio formula. □

Lemma 3.3 (Decomposition)**.** *Let X be finite atomic with $\mathbb{E}X = \mu$. Suppose a is an atom of X with mass p, and let $q = 1 - p$. Let*

$$Y \sim X \mid X \neq a, \qquad y = \mathbb{E}Y.$$

Let

$$Z = \begin{cases} a, & p, \\ y, & q. \end{cases}$$

Then $\mathbb{E}Z = \mu$, and

$$J_f(X) = J_f(Z) + qJ_f(Y).$$

The same identity holds for g.

Proof. Since $\mu = pa + qy$,

$$J_f(X) = pf(a) + q\mathbb{E}f(Y) - f(\mu).$$

Adding and subtracting $qf(y)$ yields

$$J_f(X) = (pf(a) + qf(y) - f(\mu)) + q\left(\mathbb{E}f(Y) - f(y)\right),$$

which is exactly $J_f(X) = J_f(Z) + qJ_f(Y)$. □

4 The increasing case

Theorem 4.1 (Sharp comparison for nondecreasing relative divided differences)**.** *Let $I \subset \mathbb{R}$, let $0 < \alpha \leq 1/2$, and let $\mu \in I$. Let f, g be such that all second divided differences below are well-defined and*

$$g[x, y, z] > 0$$

for every nondegenerate triple. Suppose that

$$R_{f,g}(x, y, z) = \frac{f[x, y, z]}{g[x, y, z]}$$

is nondecreasing in each argument.

Define

$$C_\alpha^-(f, g; \mu) = \inf_{\substack{u<\mu \\ u\in I \\ (\mu-\alpha u)/(1-\alpha)\in I}} R_{f,g}\left(u, \mu, \frac{\mu - \alpha u}{1 - \alpha}\right),$$

and

$$C_\alpha^+(f, g; \mu) = \sup_{\substack{v>\mu \\ v\in I \\ (\mu-\alpha v)/(1-\alpha)\in I}} R_{f,g}\left(\frac{\mu - \alpha v}{1 - \alpha}, \mu, v\right).$$

Then, for every $X \in \mathcal{A}_\alpha(\mu)$,

$$C_\alpha^-(f, g; \mu)J_g(X) \leq J_f(X) \leq C_\alpha^+(f, g; \mu)J_g(X).$$

Both constants are sharp.

Proof. We prove the lower inequality first. Let $X \in \mathcal{A}_\alpha(\mu)$. Since X is nonconstant and has mean μ, its support has points on both sides of μ. Let a be the leftmost atom of X. Then $a < \mu$. Write

$$p = \mathbb{P}(X = a), \qquad q = 1 - p.$$

Since $X \in \mathcal{A}_\alpha(\mu)$, $p \geq \alpha$. Let

$$Y \sim X \mid X \neq a, \qquad y = \mathbb{E}Y.$$

Then

$$\mu = pa + qy, \qquad a < \mu < y.$$

Let

$$Z = \begin{cases} a, & p, \\ y, & q. \end{cases}$$

By the decomposition lemma,

$$J_f(X) = J_f(Z) + qJ_f(Y), \qquad J_g(X) = J_g(Z) + qJ_g(Y).$$

For the two-point part,

$$J_f(Z) = R_{f,g}(a, \mu, y)J_g(Z).$$

For the tail Y, the Bregman divided-difference identity gives

$$J_f(Y) = \mathbb{E}\Big[(Y - y)^2 f[Y, y, y]\Big], \qquad J_g(Y) = \mathbb{E}\Big[(Y - y)^2 g[Y, y, y]\Big].$$

Every atom x of Y satisfies $x \geq a$, and $y > \mu$. Therefore, by coordinatewise monotonicity,

$$R_{f,g}(x, y, y) \geq R_{f,g}(a, \mu, y).$$

Thus

$$f[x, y, y] \geq R_{f,g}(a, \mu, y)g[x, y, y].$$

Multiplying by $(x - y)^2$ and averaging gives

$$J_f(Y) \geq R_{f,g}(a, \mu, y)J_g(Y).$$

Combining this with the two-point part,

$$J_f(X) \geq R_{f,g}(a, \mu, y)J_g(X).$$

Now

$$y = \frac{\mu - pa}{1 - p}.$$

For fixed $a < \mu$, this is increasing in p, because

$$\frac{d}{dp}\frac{\mu - pa}{1 - p} = \frac{\mu - a}{(1 - p)^2} > 0.$$

Since $p \geq \alpha$,

$$y \geq \frac{\mu - \alpha a}{1 - \alpha}.$$

Therefore

$$R_{f,g}(a, \mu, y) \geq R_{f,g}\left(a, \mu, \frac{\mu - \alpha a}{1 - \alpha}\right).$$

Taking the infimum over all admissible $a < \mu$ gives

$$J_f(X) \geq C_\alpha^-(f, g; \mu) J_g(X).$$

We now prove the upper inequality. Let b be the rightmost atom of X. Then $b > \mu$. Write

$$p = \mathbb{P}(X = b), \qquad q = 1 - p,$$

so that $p \geq \alpha$. Let

$$Y \sim X \mid X \neq b, \qquad y = \mathbb{E}Y.$$

Then

$$\mu = qy + pb, \qquad y < \mu < b.$$

Let

$$Z = \begin{cases} y, & q, \\ b, & p. \end{cases}$$

Again,

$$J_f(X) = J_f(Z) + qJ_f(Y), \qquad J_g(X) = J_g(Z) + qJ_g(Y).$$

The two-point part gives

$$J_f(Z) = R_{f,g}(y, \mu, b) J_g(Z).$$

For the tail Y, every atom x satisfies $x \leq b$. If $x \leq y$, then

$$(x, y, y) \leq (y, \mu, b)$$

coordinatewise. If $x \geq y$, then by symmetry of second divided differences,

$$R_{f,g}(x, y, y) = R_{f,g}(y, y, x),$$

and

$$(y, y, x) \leq (y, \mu, b)$$

coordinatewise. Thus

$$R_{f,g}(x, y, y) \leq R_{f,g}(y, \mu, b).$$

It follows that

$$J_f(Y) \leq R_{f,g}(y, \mu, b) J_g(Y).$$

Combining the two-point part and the tail part gives

$$J_f(X) \leq R_{f,g}(y, \mu, b) J_g(X).$$

Now

$$y = \frac{\mu - pb}{1 - p}.$$

For fixed $b > \mu$, this is decreasing in p, because

$$\frac{d}{dp} \frac{\mu - pb}{1 - p} = \frac{\mu - b}{(1 - p)^2} < 0.$$

Since $p \geq \alpha$,

$$y \leq \frac{\mu - \alpha b}{1 - \alpha}.$$

Thus

$$R_{f,g}(y, \mu, b) \le R_{f,g}\left(\frac{\mu - \alpha b}{1-\alpha}, \mu, b\right).$$

Taking the supremum over all admissible $b > \mu$ gives

$$J_f(X) \le C_\alpha^+(f, g; \mu) J_g(X).$$

Sharpness follows from the two-point laws

$$X_u^\alpha = \begin{cases} u, & \alpha, \\ \dfrac{\mu - \alpha u}{1-\alpha}, & 1-\alpha, \end{cases}$$

for the lower constant and

$$Y_v^\alpha = \begin{cases} \dfrac{\mu - \alpha v}{1-\alpha}, & 1-\alpha, \\ v, & \alpha, \end{cases}$$

for the upper constant. These laws belong to $\mathcal{A}_\alpha(\mu)$ and attain the corresponding two-point quotient. Minimizing and maximizing sequences prove sharpness if the infimum or supremum is not attained. □

5 The reflected theorem

Theorem 5.1 (Sharp comparison for nonincreasing relative divided differences)**.** *Under the assumptions of Theorem 4.1, suppose instead that $R_{f,g}$ is nonincreasing in each argument. Define*

$$\widetilde{C}_\alpha^-(f, g; \mu) = \inf_{\substack{v>\mu \\ v \in I \\ (\mu - \alpha v)/(1-\alpha) \in I}} R_{f,g}\left(\frac{\mu - \alpha v}{1-\alpha}, \mu, v\right),$$

and

$$\widetilde{C}_\alpha^+(f, g; \mu) = \sup_{\substack{u<\mu \\ u \in I \\ (\mu - \alpha u)/(1-\alpha) \in I}} R_{f,g}\left(u, \mu, \frac{\mu - \alpha u}{1-\alpha}\right).$$

Then, for every $X \in \mathcal{A}_\alpha(\mu)$,

$$\widetilde{C}_\alpha^-(f, g; \mu) J_g(X) \le J_f(X) \le \widetilde{C}_\alpha^+(f, g; \mu) J_g(X).$$

Both constants are sharp.

Proof. The proof is the reflected version of Theorem 4.1. For the lower inequality, one separates the rightmost atom; for the upper inequality, one separates the leftmost atom. All inequalities in the tail comparison are reversed because $R_{f,g}$ is nonincreasing. The same two-point laws prove sharpness. □

6 The Jensen–variance case

Taking $g(x) = x^2$ gives

$$J_g(X) = \operatorname{Var}(X), \qquad x^2[a,b,c] = 1.$$

Thus

$$R_{f,x^2}(a,b,c) = f[a,b,c].$$

If $f[x,y,z]$ is nondecreasing in each argument, then Theorem 4.1 gives

$$C_\alpha^-(f,x^2;\mu)\operatorname{Var}(X) \le J_f(X) \le C_\alpha^+(f,x^2;\mu)\operatorname{Var}(X),$$

where

$$C_\alpha^-(f,x^2;\mu) = \inf_{\substack{u<\mu \\ u\in I \\ (\mu-\alpha u)/(1-\alpha)\in I}} f\left[u,\mu,\frac{\mu-\alpha u}{1-\alpha}\right],$$

and

$$C_\alpha^+(f,x^2;\mu) = \sup_{\substack{v>\mu \\ v\in I \\ (\mu-\alpha v)/(1-\alpha)\in I}} f\left[\frac{\mu-\alpha v}{1-\alpha},\mu,v\right].$$

If $f \in C^2(I)$ and f'' is nondecreasing, then $f[x,y,z]$ is nondecreasing in each argument. The reflected statement applies when $f[x,y,z]$ is nonincreasing, for instance when f'' is nonincreasing.

7 Equal-weight finite form

Let $x_1,\dots,x_N \in I$ be not all equal, and set

$$\mu = \frac{1}{N}\sum_{i=1}^N x_i.$$

Let X be uniformly distributed on the multiset $\{x_1,\dots,x_N\}$. After merging equal values, every atom has mass at least $1/N$. Hence $\alpha = 1/N$.

In the nondecreasing case, define

$$C_N^-(f,g;\mu) = \inf_{\substack{u<\mu \\ u\in I \\ (N\mu-u)/(N-1)\in I}} R_{f,g}\left(u,\mu,\frac{N\mu-u}{N-1}\right),$$

and

$$C_N^+(f,g;\mu) = \sup_{\substack{v>\mu \\ v\in I \\ (N\mu-v)/(N-1)\in I}} R_{f,g}\left(\frac{N\mu-v}{N-1},\mu,v\right).$$

Then

$$C_N^-(f,g;\mu)\left(\sum_{i=1}^N g(x_i) - Ng(\mu)\right) \le \sum_{i=1}^N f(x_i) - Nf(\mu),$$

and

$$\sum_{i=1}^N f(x_i) - Nf(\mu) \le C_N^+(f,g;\mu)\left(\sum_{i=1}^N g(x_i) - Ng(\mu)\right).$$

The lower extremal vectors have the form

$$\left(u, \frac{N\mu - u}{N-1}, \dots, \frac{N\mu - u}{N-1}\right),$$

and the upper extremal vectors have the form

$$\left(\frac{N\mu - v}{N-1}, \dots, \frac{N\mu - v}{N-1}, v\right).$$

The next result is a useful closure principle. It gives a broad constructive class of pairs to which the main theorem applies.

Theorem 7.1 (Finite positive cones of admissible numerators)**.** *Let $I \subset \mathbb{R}$ be an interval, and let g satisfy*

$$g[x, y, z] > 0$$

for every nondegenerate triple in I. Let $f_1, \dots, f_m$ be functions on I such that, for every $j = 1, \dots, m$,

$$R_{f_j,g}(x, y, z) = \frac{f_j[x, y, z]}{g[x, y, z]}$$

is coordinatewise nondecreasing. Let

$$f = \ell + \sum_{j=1}^{m} A_j f_j, \qquad A_j \geq 0,$$

where ℓ is affine. Then

$$R_{f,g}(x, y, z) = \frac{f[x, y, z]}{g[x, y, z]}$$

is coordinatewise nondecreasing. Hence Theorem 4.1 applies to the pair (f, g).

The reflected statement also holds: if every $R_{f_j,g}$ is coordinatewise nonincreasing, then $R_{f,g}$ is coordinatewise nonincreasing, and Theorem 5.1 applies.

Proof. Since ℓ is affine,

$$\ell[x, y, z] = 0.$$

By linearity of divided differences,

$$f[x, y, z] = \sum_{j=1}^{m} A_j f_j[x, y, z].$$

Therefore

$$R_{f,g}(x, y, z) = \frac{f[x, y, z]}{g[x, y, z]} = \sum_{j=1}^{m} A_j \frac{f_j[x, y, z]}{g[x, y, z]} = \sum_{j=1}^{m} A_j R_{f_j,g}(x, y, z).$$

A finite nonnegative linear combination of coordinatewise nondecreasing functions is coordinatewise nondecreasing. The nonincreasing case is identical. □

Remark 7.2 (Integral positive cones)**.** *The same conclusion holds for*

$$f = \ell + \int_{\Theta} f_\theta \, d\nu(\theta)$$

with a positive measure ν, provided the integral is well defined and second divided differences may be passed under the integral sign. The finite version above is the form used in the applications below.

Remark 7.3 (Natural scope)**.** *The full theorem-level class in this paper is the class of pairs (f, g) for which $R_{f,g}$ is coordinatewise monotone. The positive-cone theorem is a broad constructive subclass. It includes many examples, but it does not imply that arbitrary mixtures in both numerator and denominator are admissible. The denominator must remain fixed in the closure argument.*

8 Power-gap comparisons

Let $I = (0, \infty)$ and

$$f(x) = x^p, \qquad g(x) = x^q, \qquad p > q > 1.$$

For a random variable $X > 0$ with $\mathbb{E}X = \mu$, write

$$J_p(X) = \mathbb{E}X^p - \mu^p.$$

Proposition 8.1 (Monotonicity for power divided differences)**.** *For $p > q > 1$, the function*

$$R_{p,q}(a, b, c) = \frac{x^p[a, b, c]}{x^q[a, b, c]}$$

is nondecreasing in each of a, b, c on $(0, \infty)$.

Proof. We use the multivariable extended means of Leach and Sholander [2]. For a sequence $x = (x_0, \dots, x_n)$, they define functions f_r by $f_r^{(n)}(t) = t^{r-n}$ and the mean

$$E(r, s; x) = \left(\frac{\delta_n f_r(x)}{\delta_n f_s(x)}\right)^{1/(r-s)} \qquad (r \neq s).$$

They prove that $E(r, s; x)$ increases with each component x_k.

Take $n = 2$ and $x = (a, b, c)$. Then $f_r''(t) = t^{r-2}$. Since $(x^r)'' = r(r-1)t^{r-2}$, the second divided difference of x^r satisfies

$$x^r[a, b, c] = r(r-1) f_r[a, b, c].$$

Therefore

$$R_{p,q}(a, b, c) = \frac{p(p-1)}{q(q-1)} \frac{f_p[a, b, c]}{f_q[a, b, c]} = \frac{p(p-1)}{q(q-1)} E(p, q; a, b, c)^{p-q}.$$

Since $p - q > 0$ and $E(p, q; \cdot)$ is coordinatewise increasing, $R_{p,q}$ is coordinatewise nondecreasing. $\square$

Corollary 8.2 (Sharp power-gap comparison)**.** *Let $p > q > 1$, let $X > 0$ be finite atomic with $\mathbb{E}X = \mu > 0$, and suppose that every atom has mass at least α. Then*

$$C^-_{\alpha,p,q}(\mu) J_q(X) \le J_p(X) \le C^+_{\alpha,p,q}(\mu) J_q(X),$$

where

$$C^-_{\alpha,p,q}(\mu) = \inf_{0<u<\mu} R_{p,q}\left(u, \mu, \frac{\mu - \alpha u}{1 - \alpha}\right),$$

and

$$C^+_{\alpha,p,q}(\mu) = \sup_{\mu<v<\mu/\alpha} R_{p,q}\left(\frac{\mu - \alpha v}{1 - \alpha}, \mu, v\right).$$

Both constants are sharp. By homogeneity,

$$C^\pm_{\alpha,p,q}(\mu) = \mu^{p-q} C^\pm_{\alpha,p,q}(1).$$

Example 8.3 (J_3 versus variance)**.** *For $p = 3$ and $q = 2$,*

$$x^3[a, b, c] = a + b + c, \qquad x^2[a, b, c] = 1.$$

Thus

$$\mu \frac{2 - \alpha}{1 - \alpha} \operatorname{Var}(X) \le \mathbb{E}X^3 - \mu^3 \le \mu \frac{1 + \alpha}{\alpha} \operatorname{Var}(X).$$

Both constants are sharp.

9 Exponential-gap comparisons

Let

$$f(x) = e^{\lambda x}, \qquad g(x) = e^{\eta x}, \qquad 0 < \eta < \lambda.$$

For a random variable X with $\mathbb{E}X = \mu$, write

$$J_\lambda(X) = \mathbb{E}e^{\lambda X} - e^{\lambda \mu}.$$

Proposition 9.1 (Monotonicity for exponential divided differences)**.** *For $0 < \eta < \lambda$, the function*

$$R_{\lambda,\eta}(a,b,c) = \frac{e^{\lambda x}[a,b,c]}{e^{\eta x}[a,b,c]}$$

is nondecreasing in each of a, b, c.

Proof. Let G_r be any function satisfying

$$G_r''(t) = e^{rt}.$$

Leach and Sholander define an exponential analogue $F(r,s;x)$ of their multivariable extended mean by

$$F(r,s;x) = \frac{1}{s-r} \log \frac{\delta_n G_s(x)}{\delta_n G_r(x)} \qquad (r \neq s),$$

with the usual limiting extension for $r = s$. They state that this exponential mean has the analogue of the variable-monotonicity property of their Theorem 4: it is increasing in each component of x.

Take $n = 2$ and $x = (a,b,c)$. Since $G_r'' = e^{rt}$, the second divided differences satisfy

$$e^{rx}[a,b,c] = r^2 G_r[a,b,c]$$

for $r \neq 0$; the case $r = 0$ is obtained by continuity, and here we only need $0 < \eta < \lambda$. For $r = \eta$ and $s = \lambda$,

$$F(\eta,\lambda;a,b,c) = \frac{1}{\lambda - \eta} \log \frac{G_\lambda[a,b,c]}{G_\eta[a,b,c]}.$$

Hence

$$\frac{G_\lambda[a,b,c]}{G_\eta[a,b,c]} = \exp\Big((\lambda-\eta)F(\eta,\lambda;a,b,c)\Big).$$

Since $F(\eta,\lambda;\cdot)$ is coordinatewise increasing and $\lambda - \eta > 0$, this ratio is coordinatewise nondecreasing. Finally,

$$R_{\lambda,\eta}(a,b,c) = \frac{e^{\lambda x}[a,b,c]}{e^{\eta x}[a,b,c]} = \frac{\lambda^2}{\eta^2} \frac{G_\lambda[a,b,c]}{G_\eta[a,b,c]},$$

and multiplication by the positive constant λ^2/η^2 preserves coordinatewise monotonicity. $\square$

Corollary 9.2 (Sharp exponential-gap comparison)**.** *Let $0 < \eta < \lambda$, and let X be finite atomic, supported in an interval $I \subset \mathbb{R}$, with $\mathbb{E}X = \mu$ and atom masses at least α. Then*

$$C^-_{\alpha,\lambda,\eta}(\mu) J_\eta(X) \le J_\lambda(X) \le C^+_{\alpha,\lambda,\eta}(\mu) J_\eta(X),$$

where

$$C^{-}_{\alpha,\lambda,\eta}(\mu) = \inf_{\substack{u<\mu \\ u,(\mu-\alpha u)/(1-\alpha)\in I}} R_{\lambda,\eta}\left(u, \mu, \frac{\mu-\alpha u}{1-\alpha}\right),$$

and

$$C^{+}_{\alpha,\lambda,\eta}(\mu) = \sup_{\substack{v>\mu \\ v,(\mu-\alpha v)/(1-\alpha)\in I}} R_{\lambda,\eta}\left(\frac{\mu-\alpha v}{1-\alpha}, \mu, v\right).$$

Both constants are sharp.

10 Further admissible classes

The preceding examples, together with the positive-cone theorem, generate several useful families.

10.1 Power-divergence generators

For $r \in \mathbb{R}$, define the Cressie–Read power-divergence generator

$$\phi_r(x) = \frac{x^r - r(x-1) - 1}{r(r-1)}, \qquad r \neq 0, 1,$$

with the limiting cases

$$\phi_1(x) = x \log x - x + 1, \qquad \phi_0(x) = -\log x + x - 1.$$

Affine terms have zero second divided difference, and

$$\phi_r''(x) = x^{r-2} > 0$$

for all real r, with the limiting interpretations at $r = 0, 1$.

Proposition 10.1 (Power-divergence monotonicity)**.** *If $r > s$, then*

$$\frac{\phi_r[x, y, z]}{\phi_s[x, y, z]}$$

is coordinatewise nondecreasing on $(0, \infty)$*.*

Proof. Let h_r be the function used in the Leach–Sholander extended mean construction, normalized by

$$h_r''(x) = x^{r-2}.$$

Then h_r and ϕ_r have the same second derivative; hence they differ by an affine function and have the same second divided differences:

$$h_r[x, y, z] = \phi_r[x, y, z].$$

For $n = 2$, the Leach–Sholander mean satisfies

$$E(r, s; x, y, z) = \left(\frac{h_r[x, y, z]}{h_s[x, y, z]} \right)^{1/(r-s)} \qquad (r \neq s),$$

and is increasing in each variable. Therefore, for $r > s$,

$$\frac{\phi_r[x, y, z]}{\phi_s[x, y, z]} = E(r, s; x, y, z)^{r-s}$$

is coordinatewise nondecreasing. The limiting cases $r = 0, 1$ or $s = 0, 1$ follow by continuity of divided differences in the parameter. □

Hence, if $X > 0$, $\mathbb{E}X = \mu$, and all atom masses are at least α, then

$$C^-_{\alpha,r,s}(\mu)J_{\phi_s}(X) \le J_{\phi_r}(X) \le C^+_{\alpha,r,s}(\mu)J_{\phi_s}(X),$$

with the sharp two-point constants supplied by Theorem 4.1.

For probability distributions $P = (p_i)$ and $Q = (q_i)$, put $X_i = p_i/q_i$ under Q. Then $\mathbb{E}_Q X = 1$ and

$$D_{\phi_r}(P\|Q) = \mathbb{E}_Q \phi_r(X) = J_{\phi_r}(X).$$

Thus, if $q_i \ge \alpha$ for every i, then

$$C^-_{\alpha,r,s}D_{\phi_s}(P\|Q) \le D_{\phi_r}(P\|Q) \le C^+_{\alpha,r,s}D_{\phi_s}(P\|Q),$$

with sharp two-point constants.

10.2 Hellinger-type divergences

The squared Hellinger generator without the conventional factor 1/2 is

$$h(x) = (\sqrt{x} - 1)^2 = x + 1 - 2\sqrt{x}.$$

It is a constant multiple of the power-divergence generator $\phi_{1/2}$:

$$\phi_{1/2}(x) = 2(\sqrt{x} - 1)^2.$$

Therefore the previous power-divergence comparison gives sharp comparisons between Hellinger-type divergence, KL divergence, and Pearson chi-square divergence under the lower reference-mass condition $q_i \ge \alpha$.

In particular, with

$$H_0^2(P, Q) = \sum_i (\sqrt{p_i} - \sqrt{q_i})^2,$$

one obtains sharp constants a_α, b_α such that

$$a_\alpha \chi^2(P\|Q) \le H_0^2(P, Q) \le b_\alpha \chi^2(P\|Q)$$

whenever $q_i \ge \alpha$. The constants are again the corresponding two-point constants.

10.3 Shifted and reciprocal powers

Let $s \in \mathbb{R}$ and work on an interval where $s + x > 0$. If $p > q > 1$, then

$$f(x) = (s + x)^p, \qquad g(x) = (s + x)^q$$

has coordinatewise nondecreasing relative divided difference, because it is obtained from the power case by the translation $y = s + x$.

If $p > q > 0$, then

$$f(x) = (s + x)^{-p}, \qquad g(x) = (s + x)^{-q}$$

has coordinatewise nonincreasing relative divided difference, and the reflected theorem applies. The constants may be infinite if the admissible support can approach the pole $x = -s$.

Similarly,

$$f_p(x) = (1 - x)^{-p}, \qquad g_q(x) = (1 - x)^{-q}, \qquad p > q > 0,$$

on $x < 1$ has coordinatewise nondecreasing relative divided difference, because the substitution $y = 1 - x$ reverses the monotonicity of the negative-power quotient. However, if the support lies in $[0, 1)$ and $\mu \ge \alpha$, then the upper constant is infinite: an atom of mass α may approach the pole $x = 1$.

10.4 Positive mixtures

The positive-cone theorem gives immediate finite-mixture extensions. The integral versions below are valid under the standard interchange assumption: the integral is finite on the relevant interval and second divided differences may be passed under the integral sign.

For example, if $q > 1$ and

$$f(x) = \ell(x) + \int_{(q,\infty)} x^p \, d\nu(p), \qquad \nu \geq 0,$$

then R_{f,x^q} is coordinatewise nondecreasing, provided the preceding interchange is justified. Similarly, if $\eta > 0$ and

$$f(x) = \ell(x) + \int_{(\eta,\infty)} e^{\lambda x} \, d\nu(\lambda), \qquad \nu \geq 0,$$

then $R_{f,e^{\eta x}}$ is coordinatewise nondecreasing, again provided divided differences may be passed under the integral sign. Reflected versions hold for positive mixtures of negative powers.

11 KL divergence versus chi-square divergence

Let

$$f(x) = x \log x, \qquad f(0) = 0, \qquad g(x) = (x-1)^2.$$

On $I = [0, \infty)$, one has $g[a, b, c] = 1$. For $x > 0$,

$$f''(x) = \frac{1}{x}.$$

By the Hermite–Genocchi formula,

$$f[a, b, c] = \frac{1}{2} \int_{\Delta_2} \frac{1}{\lambda_1 a + \lambda_2 b + \lambda_3 c} \, d\lambda.$$

Hence $f[a, b, c]$ is nonincreasing in each argument, and Theorem 5.1 applies.

Let $\mu = 1$. For $1 < v < 1/\alpha$, set

$$u_\alpha(v) = \frac{1 - \alpha v}{1 - \alpha}.$$

Define

$$K_\alpha^-(v) = \frac{\alpha v \log v + (1-\alpha) u_\alpha(v) \log u_\alpha(v)}{\alpha(v-1)^2 + (1-\alpha)(u_\alpha(v) - 1)^2}.$$

For $0 < u < 1$, set

$$v_\alpha(u) = \frac{1 - \alpha u}{1 - \alpha},$$

and define

$$K_\alpha^+(u) = \frac{\alpha u \log u + (1-\alpha) v_\alpha(u) \log v_\alpha(u)}{\alpha(u-1)^2 + (1-\alpha)(v_\alpha(u) - 1)^2}.$$

Let

$$c_\alpha^- = \inf_{1 < v < 1/\alpha} K_\alpha^-(v), \qquad c_\alpha^+ = \sup_{0 < u < 1} K_\alpha^+(u).$$

Then for every finite atomic $X \geq 0$ with $\mathbb{E}X = 1$ and atom masses at least α,

$$c_\alpha^- \mathbb{E}(X-1)^2 \leq \mathbb{E}X \log X \leq c_\alpha^+ \mathbb{E}(X-1)^2.$$

Both constants are sharp.

Let $P = (p_i)$ and $Q = (q_i)$ be probability distributions on a finite alphabet, with $q_i > 0$, and set

$$X_i = \frac{p_i}{q_i}$$

under Q. Then

$$\mathbb{E}_Q X = 1, \qquad \mathbb{E}_Q X \log X = D_{\mathrm{KL}}(P\|Q),$$

and

$$\mathbb{E}_Q(X-1)^2 = \chi^2(P\|Q).$$

If $q_i \geq \alpha$ for every i, then after merging equal likelihood ratios the atom-mass condition is satisfied. Hence

$$c_\alpha^- \chi^2(P\|Q) \leq D_{\mathrm{KL}}(P\|Q) \leq c_\alpha^+ \chi^2(P\|Q).$$

12 Limitations and counterexamples

The coordinatewise monotonicity assumption on $R_{f,g}$ is structural. It cannot be replaced by convexity alone.

Example 12.1 (A convex polynomial need not be admissible)**.** *Let*

$$g(x) = x^2, \qquad f(x) = x^4 - x^3 + 10x^2, \qquad x \geq 0.$$

Then

$$f''(x) = 12x^2 - 6x + 20 > 0,$$

so f is convex on $[0,\infty)$. However,

$$R_{f,g}(a,b,c) = f[a,b,c]$$

and

$$f[a,b,c] = a^2 + b^2 + c^2 + ab + ac + bc - (a+b+c) + 10.$$

Taking $b = c = 0.1$ gives

$$f[a, 0.1, 0.1] = a^2 - 0.8a + 9.83,$$

which decreases for small positive a. Thus convexity of f does not imply coordinatewise monotonicity of R_{f,x^2}.

Example 12.2 (Positive mixtures in both numerator and denominator)**.** *Let*

$$f(x) = x^4, \qquad g(x) = x^2 + x^5.$$

Both are convex on $[0,\infty)$. However, with $b = 1$ and $c = 2$,

$$\frac{f[0,1,2]}{g[0,1,2]} = \frac{7}{16}, \qquad \frac{f[1,1,2]}{g[1,1,2]} = \frac{11}{27} < \frac{7}{16}.$$

Thus allowing positive mixtures in both numerator and denominator does not preserve coordinatewise monotonicity. The positive-cone theorem keeps the denominator fixed.

Example 12.3 (No positive lower variance constant for a convex function)**.** *Let*

$$f(x) = \log(1 + e^x), \qquad x \in \mathbb{R}.$$

Then f is convex, but for

$$X_M = \begin{cases} -M, & 1/2, \\ M, & 1/2, \end{cases}$$

one has $\mathbb{E}X_M = 0$ and $\mathrm{Var}(X_M) = M^2$, whereas

$$J_f(X_M) = \frac{1}{2}f(M) + \frac{1}{2}f(-M) - f(0) \sim \frac{M}{2}.$$

Hence

$$\frac{J_f(X_M)}{\mathrm{Var}(X_M)} \to 0.$$

Thus convexity alone does not imply a positive Jensen–variance lower bound on unbounded intervals.

13 Boundary singularities

The lower atom-mass condition controls the probabilities of atoms, but it does not necessarily control their locations. Therefore functions with singularities at the boundary may have infinite comparison constants.

For example, let

$$f(x) = -\log x, \qquad g(x) = x^2, \qquad I = (0, \infty).$$

Fix $\mu > 0$ and consider

$$X_\varepsilon = \begin{cases} \varepsilon, & \alpha, \\ \dfrac{\mu - \alpha\varepsilon}{1 - \alpha}, & 1 - \alpha. \end{cases}$$

Then $\mathbb{E}X_\varepsilon = \mu$, but

$$J_{-\log}(X_\varepsilon) \to +\infty$$

as $\varepsilon \downarrow 0$, while $\mathrm{Var}(X_\varepsilon)$ remains finite. Hence the upper Jensen–variance constant is infinite unless an additional lower support bound is imposed.

Similarly, for

$$f_n(x) = \frac{1}{(1 - x)^n}$$

on $[0, 1)$, the upper Jensen–variance constant is infinite whenever $\mu \geq \alpha$, because an atom of mass α may approach the pole $x = 1$.

14 Relation to earlier mean–variance bounds

In the classical mean–variance setting one fixes $\mu = \mathbb{E}X$ and $\sigma^2 = \mathrm{Var}(X)$ and seeks sharp bounds for $\mathbb{E}f(X)$ or $J_f(X)$. In Pittenger's formulation [6], for a distinguished point x_0, the extremal law has one atom at x_0 and another atom at

$$a_0 = \mu + \frac{\sigma^2}{\mu - x_0},$$

with mass

$$p_0 = \frac{(\mu - x_0)^2}{\sigma^2 + (\mu - x_0)^2}.$$

Thus the atom mass is determined by the prescribed variance.

In the present paper, the variance is not fixed. Instead, a lower atom-mass constraint $p_i \geq \alpha$ is imposed, and two Jensen gaps are compared. The variance case is the special denominator $g(x) = x^2$. The sharp constants are obtained by optimizing over the two-point family with fixed atom masses α and $1 - \alpha$.

15 Conclusion

We have proved sharp comparisons of Jensen gaps under lower atom-mass constraints. The main theorem extends Jensen–variance bounds from the special denominator

$$J_{x^2}(X) = \mathrm{Var}(X)$$

to a general denominator Jensen gap $J_g(X)$. The sharp constants are obtained from two-point laws with masses α and $1-\alpha$, and the proof is based on coordinatewise monotonicity of the relative second divided difference

$$R_{f,g}(x, y, z) = \frac{f[x, y, z]}{g[x, y, z]}.$$

The equal-weight case $\alpha = 1/N$ gives sharp finite-sum inequalities. The examples show that the method applies naturally to Jensen–variance bounds, power gaps, exponential gaps, and finite-alphabet divergence comparisons.

References

[1] S. Karlin and W. J. Studden, *Tchebycheff Systems: With Applications in Analysis and Statistics*, Wiley, New York, 1966.

[2] E. B. Leach and M. C. Sholander, *Multi-variable extended mean values*, Journal of Mathematical Analysis and Applications 104 (1984), 390–407.

[3] L. Losonczi, *Equality of Cauchy mean values*, Publicationes Mathematicae Debrecen 57 (2000), 217–230.

[4] L. Losonczi, *Inequalities for Cauchy mean values*, Mathematical Inequalities & Applications 5 (2002), 349–359.

[5] L. Losonczi, *Ratio of Stolarsky means: monotonicity and comparison*, Publicationes Mathematicae Debrecen 75 (2009), 221–238.

[6] A. O. Pittenger, *Sharp mean-variance bounds for Jensen-type inequalities*, Statistics & Probability Letters 10 (1990), 91–94.