TOPIC. Inequalities; measures of spread. This lecture explores the implications of Jensen's inequality for g-means in general, and for harmonic, geometric, arithmetic, and related means in particular. Some corollaries are the Hölder and Cauchy-Schwarz inequalities. We close with a discussion of various measures of spread: the standard deviation, the mean absolute deviation, and Gini's mean difference.

***g*-means, revisited.** Jensen's inequality has implications for g-means. To consider a simple case, suppose that g is a convex function from a closed bounded interval J to $\mathbb{R}$ and that X is a (necessarily) integrable random variable taking values in J. Then Jensen's inequality says that $g(X)$ has an expectation and

$$g\bigl(E(X)\bigr) \le E\bigl(g(X)\bigr).$$

If g is also continuous and strictly increasing on J, we may apply the strictly increasing inverse function g^{-1} to this inequality to get

$$E(X) \le g^{-1}\bigl(E(g(X))\bigr) = E_g(X); \tag{1}$$

that is, the g-mean of X exists and is at least as large as the ordinary mean. Moreover, strict inequality holds in (1) if g is strictly convex and X is nondegenerate. The following theorem asserts that these conclusions hold even if J and/or $g(J)$ is unbounded. There are some minor complications in this general setting since the endpoints of J and/or $g(J)$ may be infinite, whereas the definition of a convex function requires that both its domain and range be subsets of $\mathbb{R}$.

Theorem 1 (The *g*-means theorem). *Let J be a closed subinterval of $[-\infty,\infty]$ and let $g\colon J \to [-\infty,\infty]$ be continuous. Put*

$$I = \{\, x \in J : |x| < \infty \text{ and } |g(x)| < \infty \,\}. \tag{2}$$

Suppose that one of $\uparrow$ and $\downarrow$ below holds, and also one of $\vee$ and $\wedge$:

$\uparrow$: *g is strictly increasing on J;* $\quad\vee$: *g is convex on I;*

$\downarrow$: *g is strictly decreasing on J;* $\quad\wedge$: *g is concave on I.*

$I = \{\, x \in J : |x| < \infty \text{ and } |g(x)| < \infty \,\}.$

Let X be an integrable random variable taking values in J. Then the g-mean $E_g(X)$ of X exists and satisfies:

$$E(X) \le E_g(X) \quad \text{if } \uparrow \text{ and } \vee \text{ hold, or if } \downarrow \text{ and } \wedge \text{ hold;} \tag{3}$$

$$E(X) \ge E_g(X) \quad \text{if } \uparrow \text{ and } \wedge \text{ hold, or if } \downarrow \text{ and } \vee \text{ hold.} \tag{4}$$

When X is nondegenerate, strict inequality holds in (3) (respectively, (4)) when g is strictly convex (respectively, concave).

Proof I will treat the case where g is strictly increasing and convex; the other cases follow from this one by changing the sign of g and/or x. The argument leading up to the theorem establishes (3) when X takes values in I (with probability one), so it will suffice to reduce the general case to that situation. Since X is integrable we have

$$P[\,|X| < \infty\,] = 1; \tag{5}$$

indeed if (5) fails, then $E(|X|) = \infty$. Since g is convex on I, there exists a point $x_0 \in I$ and a finite number β such that

$$g(x) \ge g(x_0) + \beta(x - x_0)$$

for all $x \in I$. Since g is continuous on J and I contains all the points of J except possibly the endpoints, this inequality holds for all $x \in J$ too. Thus $g(X)$ is bounded below by the integrable random variable $Y := g(x_0) + \beta(X - x_0)$; this implies that $g(X)$ has an expectation and thus that the g-mean of X exists. If $E\bigl(g(X)\bigr)$ is infinite, then $E_g(X)$ is the right endpoint of J and (3) holds trivially. Otherwise $g(X)$ is integrable, so

$$P[\,|g(X)| < \infty\,] = 1. \tag{6}$$

Together (5) and (6) imply that X takes values in I with probability one; that completes the reduction. ∎

Harmonic, geometric, and other means, revisited. Let X be a random variable taking values in $[-\infty, \infty]$. For $-\infty < p < \infty$, the ***p*-norm** of X is defined to be

$$\|X\|_p := \begin{cases} \big(E(|X|^p)\big)^{1/p}, & \text{if } p \neq 0, \\ \exp\big(E(\log(|X|)\big), & \text{if } p = 0. \end{cases} \tag{7}$$

$\|X\|_p$ exists for all $p \neq 0$; $\|X\|_0$ exists if and only if $\log(|X|)$ has an expectation. Figure 8.2 graphs the p-norm of X versus p for a couple of random variables X; that figure motivates the following theorem.

Theorem 2 (The *p*-norm theorem). *Let X take values in $[-\infty, \infty]$ and define $\|X\|_p$ by (7). Then*

(M1) *$\|X\|_p$ is non-decreasing in p: if $p < q$ and $\|X\|_p$ and $\|X\|_q$ exist, then $\|X\|_p \leq \|X\|_q$.*

(M2) *$\|X\|_0$ exists if $\|X\|_p > 0$ for some $p < 0$, or if $\|X\|_q < \infty$ for some $q > 0$.*

(M3) *$\|X\|_p$ is strictly increasing on $\{\, p : 0 < \|X\|_p < \infty \,\}$ provided X is nondegenerate.*

Proof We may assume $X \geq 0$. Everything follows from the g-means theorem. I will do just part of it here, and leave the rest to Exercise 3. Suppose $q > 0$ and $\|X\|_q < \infty$. I claim $\|X\|_0$ exists and satisfies $\|X\|_0 \leq \|X\|_q$, or, equivalently, that $\log(X)$ has an expectation and

$$e^{E(\log(X))} \leq \big(E(X^q)\big)^{1/q} \iff e^{qE(\log(X))} \leq E(X^q)$$
$$\iff e^{E(\log(X^q))} \leq E(X^q) \iff E_g(Y) \leq E(Y)$$

for $Y = X^q$ and $g(y) = \log(y)$. The g-means theorem implies that the final inequality is valid (in particular, that $E_g(Y)$ exists) because:

- Y is integrable
- Y takes values in the closed interval $J = [0, \infty]$, and
- g is continuous and strictly increasing on J, and (strictly) concave on $I = \{\, y \in J : |y| < \infty \text{ and } |g(y)| < \infty \,\} = (0, \infty)$. ∎

Corollary 1. *Suppose $x_1, \ldots, x_k$ are nonnegative finite numbers and $p_1, \ldots, p_k$ are strictly positive numbers summing to 1. Then*

$$\prod_{i=1}^k x_i^{p_i} \leq \sum_{i=1}^k p_i x_i; \tag{8}$$

moreover, strict inequality holds in (8) unless all the x_i's are equal.

Proof Let X be a random variable taking the value x_i with probability p_i, for $i = 1, \ldots, k$. Then the RHS of (8) is $E(X) = \|X\|_1$, while the LHS is

$$\exp\Big(\sum_{i=1}^k p_i \log(x_i)\Big) = \exp\big(E(\log(X))\big) = \|X\|_0.$$

By assumption, $\|X\|_1 < \infty$; the p-norm theorem implies that $\|X\|_0$ exists and that $\|X\|_0 \leq \|X\|_1$. Strict inequality holds here if X is nondegenerate, i.e., the x_i's are not all the same. ∎

Theorem 3 (Hölder's inequality). *Let X and Y be random variables taking values in $[-\infty, \infty]$. Let p and q be positive, finite numbers such that $1/p + 1/q = 1$. Then*

$$E(|X||Y|) \leq \|X\|_p \, \|Y\|_q \tag{9}$$

The products in (9) are evaluated using the convention that $c \times \infty = \infty \times c = \infty$ if $0 < c \leq \infty$, but $= 0$ if $c = 0$.

Proof Without loss of generality, suppose X and Y are nonnegative.

• *Case 1:* $\|X\|_p = 1 = \|Y\|_q$. Since $\|X\|_p = 1$, we have $E(X^p) = 1$ and $X^p < \infty$ with probability one. Similarly $E(Y^q) = 1$ and Y^q is finite with probability one. Applying (8) with $k = 2$, $x_1 = X^p$, $x_2 = Y^q$, $p_1 = 1/p$ and $p_2 = 1/q$ gives

$$XY = (X^p)^{1/p}(Y^q)^{1/q} \leq \frac{1}{p}X^p + \frac{1}{q}Y^q.$$

Taking expectations here gives

$$E(XY) \leq \frac{1}{p}E(X^p) + \frac{1}{q}E(Y^q) = \frac{1}{p} + \frac{1}{q} = 1 = \|X\|_p\|Y\|_q.$$

(9): $E(XY) \le (E(X))^{1/p}(E(Y))^{1/q} = \|X\|_p\|Y\|_q$.

Case 2: $0 < \|X\|_p < \infty$ *and* $0 < \|Y\|_q < \infty$. Put

$$X^* = \frac{X}{\|X\|_p} \quad \text{and} \quad Y^* = \frac{Y}{\|Y\|_q}.$$

Then $\|X^*\|_p = 1 = \|Y^*\|_q$ (check this!), so Case 1 gives

$$E(X^*Y^*) \le 1 \Longrightarrow E\Big(\frac{XY}{\|X\|_p\|Y\|_q}\Big) \le 1$$
$$\Longrightarrow E(XY) \le \|X\|_p\|Y\|_q.$$

Case 3: $\|X\|_p = 0$ *or* $\|Y\|_q = 0$. Suppose $\|X\|_p = 0$. Then

$$E(X^p) = 0 \Longrightarrow P[\,X^p = 0\,] = 1 \quad (\text{since } X^p \ge 0)$$
$$\Longrightarrow P[\,X = 0\,] = 1 \Longrightarrow P[\,XY = 0\,] = 1$$
$$\Longrightarrow E(XY) = 0 = \|X\|_p\|Y\|_q.$$

Case 4: $\|X\|_p > 0$ *and* $\|Y\|_q > 0$ *and at least one if infinite*. Here $\|X\|_p\|Y\|_q$ is infinite, so (9) holds trivially. ∎

There is an addendum to Hölder's inequality which can be established by pushing the arguments in the proof further. One says two random variables U and V are ***linearly dependent*** if there exist finite numbers a and b, not both 0, such that $aU + bV = 0$ with probability one. The proof of the following theorem is left to Exercise 5.

Theorem 3, continued. *Suppose* $\|X\|_p < \infty$ *and* $\|Y\|_q < \infty$. *Then equality holds in* (9) *if and only if*

$$|X|^p \text{ and } |Y|^q \text{ are linearly dependent.} \tag{10}$$

The Cauchy-Schwarz inequality. For any random variable X, ***Root Means Square (RMS)*** of X is defined to be

$$\|X\|_2 = \sqrt{E(X^2)}. \tag{11}$$

X is said to be ***square-integrable*** if and only if $E(X^2) < \infty$, or, equivalently, $\|X\|_2 < \infty$.

Theorem 4 (The Cauchy-Schwarz inequality). *Let X and Y be random variables taking values in* $[-\infty, \infty]$. *Then*

$$E(|XY|) \le \|X\|_2\,\|Y\|_2. \tag{12}$$

If X and Y are both square-integrable, then XY is integrable and

$$|E(XY)| \le \|X\|_2\,\|Y\|_2; \tag{13}$$

moreover, equality holds in (12) *if and only if $|X|$ and $|Y|$ are linearly dependent, while equality holds in* (13) *if and only if X and Y are linearly dependent.*

Proof Taking $p = 2 = q$ in Hölder's inequality (which is legitimate, since $1/p + 1/q = 1/2 + 1/2 = 1$) gives

$$E(|X||Y|) \le \|X\|_2\,\|Y\|_2.$$

Now suppose $\|X\|_2$ and $\|Y\|_2$ are both finite. According to the addendum to Hölder's inequality, equality holds in (12) if and only if $|X|^2$ and $|Y|^2$ are linearly dependent; this is clearly equivalent to $|X|$ and $|Y|$ being linearly dependent. Moreover, XY is integrable because $|XY|$ has a finite expectation, and (13) holds since

$$|E(XY)| \le E(|XY|),$$

with equality if and only if $P[\,XY \ge 0\,] = 1$ or $P[\,XY \le 0\,] = 1$. The final claim in the theorem follows. ∎

Example 1. Suppose X and Y are square-integrable random variables. They are then also integrable (because, e.g., $E(|X|) = \|X\|_1 \le \|X\|_2 = \sqrt{E(X^2)} < \infty$). Put

$$X^* = X - \mu_X \quad \text{and} \quad Y^* = Y - \mu_Y$$

where $\mu_X = E(X)$ and $\mu_Y = E(Y)$. According to the Cauchy-Schwarz inequality X^*Y^* is integrable and

$$|E(X^*Y^*)| \le \|X^*\|_2 \, \|Y^*\|_2,$$

with equality if and only X^* and Y^* are linearly dependent. Now

$$E(X^*Y^*) = E\big((X-\mu_X)(Y-\mu_Y)\big) := \mathrm{Cov}(X,Y) \tag{14}$$

and, e.g.,

$$\|X^*\|_2 = \sqrt{E(X-\mu_X)^2} := \sigma_X \tag{15}$$

so we have shown that the absolute value of the correlation coefficient

$$\rho(X,Y) := \frac{\mathrm{Cov}(X,Y)}{\sigma_X \sigma_Y} \tag{16}$$

is always less than or equal in one, and equals one if and only if $X-\mu_X$ and $Y-\mu_Y$ are linearly dependent. •

Some measures of spread. Suppose X is an integrable random variable with mean $\mu = E(X)$. The ***variance*** of X is

$$\mathrm{Var}(X) := E\big((X-\mu)^2\big) = E(X)^2 - \mu^2; \tag{17}$$

this may be infinite. The square root of the variance is the ***standard deviation***, or ***root mean square deviation***:

$$\sigma_X := \sqrt{\mathrm{Var}(X)} = \|X-\mu\|_2. \tag{18}$$

This measure of spread is especially important because of the CLT.

Example 2. Suppose $U \sim \mathrm{Uniform}(0,1)$. Then U is integrable with mean $1/2$ and expected square

$$E(U^2) = \int_0^1 u^2\,du = \frac{u^3}{3}\Big|_0^1 = \frac{1}{3}.$$

Thus

$$\mathrm{Var}(U) = \frac{1}{3} - \frac{1}{4} = \frac{1}{12} \quad \text{and} \quad \sigma_U = \frac{1}{\sqrt{12}} = \frac{1}{2\sqrt{3}}. \tag{19}$$ •

Again suppose X is integrable with mean μ. The quantity

$$\delta_X := E(|X-\mu|) = \|X-\mu\|_1 \tag{20}$$

is called the ***mean absolute deviation*** of X; sometimes this term is shortened to the ***mean deviation***. The mean deviation can never exceed the standard deviation, since

$\|X-\mu\|_1 \le \|X-\mu\|_2 \, \|1\|_2$

$$\delta_X = \|X-\mu\|_1 \le \|X-\mu\|_2 = \sigma_X, \tag{21}$$

iff $|X-\mu| = c|1|$ $c \in \mathbb{R}$

(1) $c=0$

with strict inequality unless (why?) X is degenerate or there exist distinct numbers x_1 and x_2 such that $P[\,X = x_1\,] = 1/2 = P[\,X = x_2\,]$.

(2) $c \ne 0$ $X = \mu \pm c$ $\mathbb{E}X = (\mu+c)p_1 + (\mu-c)p_2 = \mu + c(p_1 - p_2) = \mu \Rightarrow p_1 = p_2 = \frac{1}{2}$

Example 2, continued. For $U \sim \mathrm{Uniform}(0,1)$ we have

$$\delta_U = E(|U-1/2|) = 2\int_0^{1/2} v\,dv = v^2\Big|_0^{1/2} = \frac{1}{4}. \tag{22}$$

Note that $\delta_U = 1/4 < 1/(2\sqrt{3}) = \sigma_U$, in agreement with (21). •

The mean and standard deviations are sometimes criticized for comparing X to a particular measure of location, namely $E(X)$. To get around this, one can use ***Gini's mean difference***

$$\Delta_X := E\big(|X^* - X|\big), \tag{23}$$

where X^* is distributed like X but is independent of it.

Example 2, continued. For $U \sim \text{Uniform}(0,1)$ we have

$$\Delta = E\big(|U^* - U|\big) = \iint_{0 \le u,v \le 1} |v-u|\,du\,dv = 2\int_{v=0}^{1}\Big[\int_{u=0}^{v}(v-u)\,du\Big]dv = \frac{1}{3}. \quad (24)$$

Notice that

$$\delta_U = \frac{1}{4} < \frac{1}{3} = \Delta_U < \frac{1}{2} = 2\delta_U$$

and

$$\Delta_U = \frac{1}{3} = \frac{2}{\sqrt{3}}\,\frac{1}{2\sqrt{3}} = \frac{2}{\sqrt{3}}\,\sigma_U < \sqrt{2}\,\sigma_U. \qquad \bullet$$

How does Δ_X compare to δ_X and to σ_X for the general integrable X? Clearly

$$\begin{aligned}\Delta_X &= E(|X^* - X|) = \|X^* - X\|_1 \le \|X^* - X\|_2 \\ &= \sqrt{E(X^* - X)^2} = \sqrt{\text{Var}(X^* - X)} \\ &= \sqrt{\text{Var}(X^*) + \text{Var}(X)} = \sqrt{2}\,\sigma_X.\end{aligned}$$

However, this is not the best bound: one can show that in general

$$\Delta_X \le \frac{2}{\sqrt{3}}\,\sigma_X; \quad (25)$$

equality holds here when $X \sim U$. Moreover one can show that in general

$$\delta_X \le \Delta_X \le 2\delta_X. \quad (26)$$

If X is nondegenerate, there is equality on the left in (26) if and only if X takes only two values (with probability one). There is no nondegenerate X for which equality holds on the right; however given any $\epsilon > 0$ there is a nondegenerate X (depending on ϵ) such that $\Delta_X \ge (2-\epsilon)\delta_X$. These assertions are explored in the exercises.

Exercise 1. Let Y be a standard Cauchy random variable and put $X = e^Y$. Let $\|X\|_p$ be defined by (7). Show that: (i) $0 < |X| < \infty$; (ii) $\|X\|_p = \infty$ for all $p > 0$; (iii) $\|X\|_p = 0$ for all $p < 0$; and (iv) $\|X\|_0$ does not exist. $\diamond$

Exercise 2. Let X be a random variable taking values in $[-\infty, \infty]$. Show that

$$\|1/X\|_p = 1/\|X\|_{-p} \quad (27)$$

for each nonzero real number p, and that

$$\begin{aligned}&1/X \text{ has a geometric mean} \iff X \text{ has a geometric mean} \\ &\quad \Longrightarrow \|1/X\|_0 = 1/\|X\|_0.\end{aligned} \quad (28)\ \diamond$$

Exercise 3. Complete the proof of the p-norm theorem (Theorem 2). First argue that $0 < q$ and $\|X\|_q < \infty$ imply $\|X\|_0 < \|X\|_q$ if X is nondegenerate. Then argue that $0 < p < q$ and $\|X\|_q < \infty$ imply $\|X\|_p \le \|X\|_q$, with strict inequality if X is nondegenerate. Finally use the result of the preceding exercise. $\diamond$

Exercise 4. Suppose X and Y are independent nonnegative random variables. How does the p-norm of the product XY relate to the p-norms of X and Y? Are there any problem cases? $\diamond$

Exercise 5. Prove the addendum to Hölder's inequality.

Exercise 6. Suppose $p \in [1, \infty)$ and X and Y are two random variables such that $\|X\|_p < \infty$ and $\|Y\|_p < \infty$. Show that

$$\|X + Y\|_p \le \|X\|_p + \|Y\|_p; \quad (29)$$

this is called ***Minkowski's inequality***. When does equality hold in (29)? [Hint: for $p > 1$ write $|X+Y|^p \le |X+Y|^{p-1}|X| + |X+Y|^{p-1}|Y|$ and apply Hölder's inequality.] $\diamond$

Exercise 7. Find the variance of all the random variables in Example 7.4 and Exercise 7.6. •

Exercise 8. Suppose T is a random variable having a (normalized) t-distribution with n degrees of freedom. (a) Show that the variance of T is infinite if $n = 2$, and equals $n/(n-2)$ if $n > 2$. (b) The standardized variable $T^* := T/\mathrm{SD}(T) = \sqrt{(n-2)/n}\,T$ has mean 0 and variance 1; moreover the distribution of T^* is approximately normal, at least for large n. Using SPLUS or the equivalent, produce a table that shows that the 0.975 quantile of T^* is approximately the 0.975 quantile of $Z \sim N(0,1)$, namely 1.96, even for small values of n. (c) Does a similar relationship hold for other the quantiles? ⋄

Exercise 9. Let the covariance between two square integrable random variables be defined by (14). Throughout this exercise, let X, Y, Z, X_1, ..., X_n be square integrable random variables. Show that:

$$\mathrm{Cov}(X,Y) = E(XY) - E(X)E(Y); \tag{30}$$

$$\mathrm{Cov}(X,Y) = \mathrm{Cov}(Y,X); \tag{31}$$

$$\mathrm{Cov}(X+Y,Z) = \mathrm{Cov}(X,Z) + \mathrm{Cov}(Y,Z); \tag{32}$$

$$\mathrm{Var}\Big(\sum_{i=1}^n X_i\Big) = \sum_{i=1}^n \mathrm{Var}(X_i) + 2\sum_{i=1}^{n-1}\sum_{j=i+1}^n \mathrm{Cov}(X_i,X_j); \tag{33}$$

$$X \text{ and } Y \text{ independent} \Longrightarrow \mathrm{Cov}(X,Y) = 0. \tag{34}$$ ⋄

Exercise 10. Let δ_X and Δ_X be respectively the mean absolute deviation (20) and Gini's mean difference (23) of the random variable X. (a) Show that $\delta_X = \Delta_X$ if X is degenerate, or if $X \sim \mathrm{Binomial}(1,p)$ for some $0 < p < 1$. (b) Find a sequence X_1, X_2, ... of random variables, each taking on just the values -1, 0, and 1, such that $\Delta_{X_n}/\delta_{X_n} \to 2$ as $n \to \infty$. ⋄

Exercise 11. Let X and Y be (not necessarily integrable) iid random variables with distribution function F and representing function R. Let $\Delta := E(|X-Y|)$ be Gini's mean difference (23). Show that

$$\Delta = 2\int_{-\infty}^{\infty} F(x)\big(1-F(x)\big)\,dx = 2\int_0^1 (2u-1)R(u)\,du. \tag{35}$$

[Hint: for the first equality, show that

$$\Delta = 2E\Big(\int_{-\infty}^{\infty} I_{\{X \le t < Y\}}\,dt\Big)$$

and interchange the expectation and integration; justify the interchange. For the second equality, give separate arguments depending on whether X is integrable or not. When X is integrable (so $E(|X|) = \int_0^1 |R(u)|\,du$ is finite) argue that

$$\Delta = E|Y-X| = \cdots = 2\iint_{(u,v):\,u<v} \big(R(v)-R(u)\big)\,du dv = \cdots ;$$

fill in the dots and justify the steps. For the justifications, you need the following fact, which is developed in the next lecture — a double integral can be done as an iterated integral (in either order) provided the integrand is nonnegative, or the double integral is absolutely convergent.] ⋄

Exercise 12. Let X be an integrable random variable with mean μ, mean deviation δ, mean difference Δ, and standard deviation σ. Use the results of the preceding exercise and Exercise 7.3 to show that:

$$\Delta \le 2\delta; \tag{36}$$

$$\delta \le \Delta; \tag{37}$$

$$\Delta \le 2\sigma/\sqrt{3}; \tag{38}$$

$$\begin{bmatrix}\text{if } 0 < \sigma < \infty, \text{ equality holds in (38) iff } X \text{ has} \\ \text{a uniform distribution on some finite interval}\end{bmatrix}. \tag{39}$$

[Hint: For (36), use the first expression for Δ in formula (35). For (38) and (39), use the second expression.] ⋄