MATH 556: MATHEMATICAL STATISTICS I

SOME INEQUALITIES

1. **Jensen's Inequality:** A function $g(x)$ is **convex** if, for $0 < \lambda < 1$,

$$g(\lambda x + (1-\lambda)y) \le \lambda g(x) + (1-\lambda) g(y)$$

for all x and y. More generally, we have that $g(x)$ is convex if, for $n \ge 2$ and constants $\lambda_i, i = 1, \ldots, n$, with $0 < \lambda_i < 1$, and $\lambda_1 + \cdots + \lambda_n = 1$

$$g\left(\sum_{i=1}^{n} \lambda_i x_i\right) \le \sum_{i=1}^{n} \lambda_i g\left(x_i\right)$$

for all vectors $(x_1, \ldots, x_n)$. We may regard this definition as stating

$$g\left(\mathbb{E}_{F_n}[X]\right) \le \mathbb{E}_{F_n}[g(X)] \tag{1}$$

where

$$\mathbb{E}_{F_n}[X] = \int x \, dF_n(x) \qquad \mathbb{E}_{F_n}[g(X)] = \int g(x) \, dF_n(x)$$

where

$$F_n(x) = \sum_{i=1}^{n} \lambda_i \mathbb{1}_{[x_i, \infty)}(x). \tag{2}$$

is the cdf of the discrete distribution on $\{x_1, \ldots, x_n\}$ with probabilities $\{\lambda_1, \ldots, \lambda_n\}$. Now, for any F_X, we can find infinite sequences $\{(x_i, \lambda_i), i = 1, 2, \ldots\}$ such that for all x

$$\lim_{n \longrightarrow \infty} F_n(x) = F_X(x).$$

For example, for $n \ge 2$, using the quantile function $Q_X(p)$ corresponding to F_X we may take

$$\lambda_i = \frac{i}{n+1} \qquad x_i = Q_X(\lambda_i) \qquad \text{for } i = 1, 2, \ldots, n.$$

As g is convex, it is also continuous, so we may pass limits through the integrals and note that

$$\lim_{n \longrightarrow \infty} \mathbb{E}_{F_n}[X] = \mathbb{E}_X[X] \qquad \lim_{n \longrightarrow \infty} \mathbb{E}_{F_n}[g(X)] = \mathbb{E}_X[g(X)]$$

which yields Jensen's inequality by substitution into (1).

Note: If relevant derivatives are well-defined, another way to view this result uses the tangent to g; for $x_0, x \in \mathbb{R}$, by convexity

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

which we evaluate for $x_0 = \mathbb{E}_X[X] = \mu$

$$g(x) \ge g(\mu) + g'(\mu)(x - \mu)$$

so that, replacing x by X and taking expectations we have

$$\mathbb{E}_X[g(X)] \ge g(\mu) + g'(\mu)(\mathbb{E}_X[X] - \mu) = g(\mu).$$

Equality holds if and only if g is linear.

Function $g(x)$ is **convex** if for all x, $g''(x) \geq 0$. Then

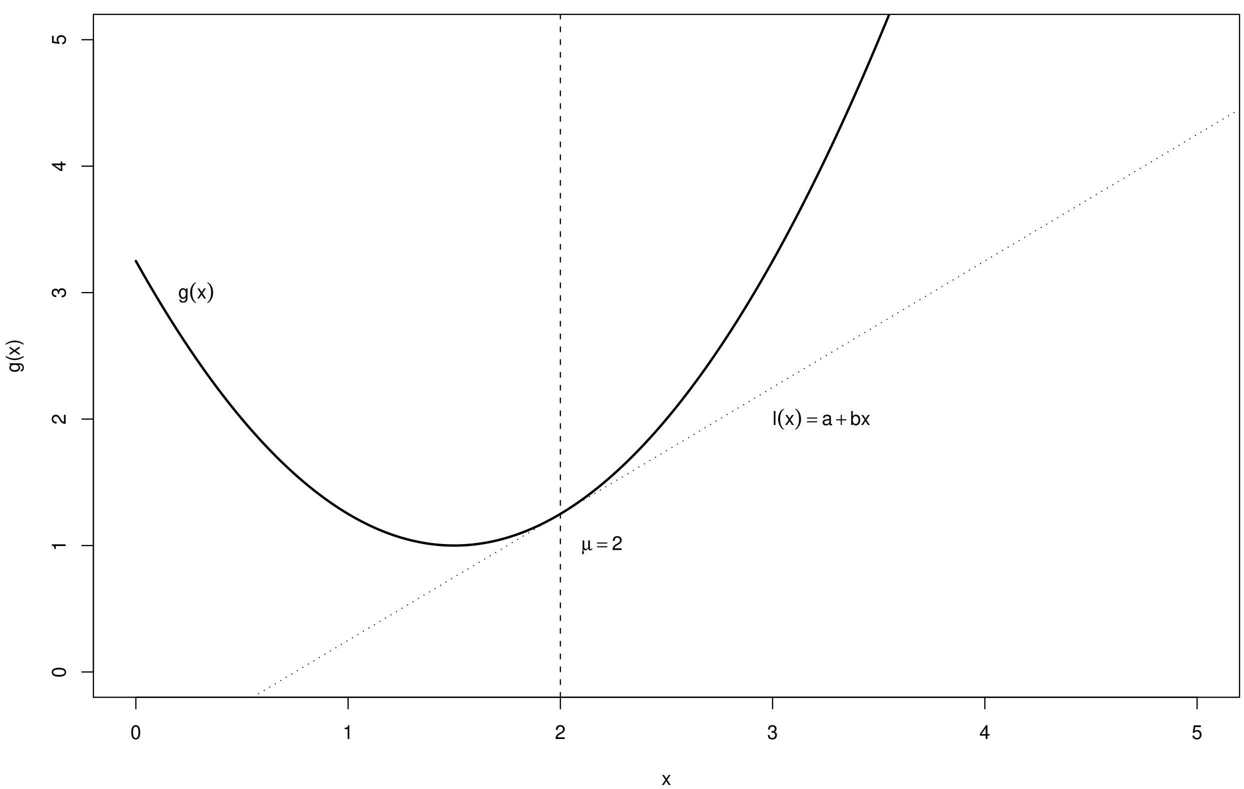
$$\mathbb{E}_X[g(X)] \geq g(\mathbb{E}_X[X])$$

with equality if and only if $g(x)$ is **linear**, that is for every line $a + bx$ that is a tangent to g at μ

$$P_X[g(X) = a + bX] = 1.$$

To see this, let $l(x) = a+bx$ be the equation of the tangent at $x = \mu$. Then, for each x, $g(x) \geq a+bx$ as in the figure, and

$$\mathbb{E}_X[g(X)] \geq \mathbb{E}_X[a + bX] = a + b\mathbb{E}_X[X] = l(\mu) = g(\mu) = g(\mathbb{E}_X[X]).$$



If $g(x)$ is linear, then equality follows by properties of expectations. Conversely, suppose that

$$\mathbb{E}_X[g(X)] = g(\mathbb{E}_X[X]) = g(\mu)$$

but $g(x)$ is convex, but **not** linear. Let $l(x) = a + bx$ be the tangent to g at μ. Then by convexity we have that $g(x) - l(x) > 0$, so

$$\int (g(x) - l(x))\, dF_X(x) = \int g(x)\, dF_X(x) - \int l(x)\, dF_X(x) > 0$$

and hence $\mathbb{E}_X[g(X)] > \mathbb{E}_X[l(X)]$; but $l(x)$ is linear, so $\mathbb{E}_X[l(X)] = a + b\mathbb{E}_X[X] = g(\mu)$, yielding a contradiction

$$\mathbb{E}_X[g(X)] > g(\mathbb{E}_X[X]).$$

- If $g(x)$ is **concave** then $-g(x)$ is convex, and $\mathbb{E}_X[g(X)] \leq g(\mathbb{E}_X[X])$
- $g(x) = x^2$ is **convex**, thus $\mathbb{E}_X\left[X^2\right] \geq \{\mathbb{E}_X[X]\}^2$
- $g(x) = \log x$ is **concave**, thus $\mathbb{E}_X[\log X] \leq \log\{\mathbb{E}_X[X]\}$

2. **Chebychev's Lemma:** If X is a random variable, then for non-negative function h, and $c > 0$,

$$P_X\left[h(X) \geq c\right] \leq \frac{\mathbb{E}_X\left[h(X)\right]}{c}$$

Proof Suppose that X has mass or density function f_X with support $\mathbb{X}$. Let $\mathcal{A} = \{x \in \mathbb{X} : h(x) \geq c\}$. Then, as $h(x) \geq c$ on $\mathcal{A}$,

$$\begin{aligned}
\mathbb{E}_X\left[h(X)\right] = \int h(x)\, dF_X(x) &= \int_{\mathcal{A}} h(x)\, dF_X(x) + \int_{\mathcal{A}'} h(x)\, dF_X(x) \\
&\geq \int_{\mathcal{A}} h(x)\, dF_X(x) \\
&\geq \int_{\mathcal{A}} c\, dF_X(x) = c\, P_X\left[X \in \mathcal{A}\right] = c\, P_X\left[h(X) \geq c\right]
\end{aligned}$$

and the result follows.

- **Special Case I: The Markov Inequality** If $h(x) = |x|^r$ for $r > 0$, so

  $$P_X\left[|X|^r \geq c\right] \leq \frac{\mathbb{E}_X\left[|X|^r\right]}{c}.$$

  Alternately: if $P_Y[Y \geq 0] = 1$ and $P_Y[Y = 0] < 1$, then for any $r > 0$

  $$P_Y[Y \geq r] \leq \frac{\mathbb{E}_Y\left[Y\right]}{r}$$

  with equality if and only if

  $$P_Y[Y = r] = p = 1 - P_Y[Y = 0]$$

  for some $0 < p \leq 1$.

- **Special Case II: The Chebychev Inequality** Suppose that X is a random variable with expectation μ and variance σ^2. Then $h(x) = (x - \mu)^2$ and $c = k^2\sigma^2$, for $k > 0$,

  $$P_X\left[(X - \mu)^2 \geq k^2\sigma^2\right] \leq 1/k^2$$

  or equivalently

  $$P_X\left[|X - \mu| \geq k\sigma\right] \leq 1/k^2.$$

  Setting $\epsilon = k\sigma$ gives

  $$P_X\left[|X - \mu| \geq \epsilon\right] \leq \sigma^2/\epsilon^2$$

  or equivalently

  $$P_X\left[|X - \mu| < \epsilon\right] \geq 1 - \sigma^2/\epsilon^2.$$

3. **Cauchy-Schwarz Inequality:** For random variable X and functions $g_1()$ and $g_2()$, we have that

$$\{\mathbb{E}_X[g_1(X)g_2(X)]\}^2 \leq \mathbb{E}_X[\{g_1(X)\}^2]\mathbb{E}_X[\{g_2(X)\}^2] \tag{3}$$

with equality if and only if either $\mathbb{E}_X[\{g_1(X)\}^2] = 0$ or $\mathbb{E}_X[\{g_2(X)\}^2] = 0$, or

$$P_X[g_1(X) = cg_2(X)] = 1$$

for some $c \neq 0$.

Proof Let $X_1 = g_1(X)$ and $X_2 = g_2(X)$, and let

$$Y_1 = aX_1 + bX_2 \qquad Y_2 = aX_1 - bX_2$$

and as $\mathbb{E}_{Y_1}[Y_1^2], \mathbb{E}_{Y_2}[Y_2^2] \geq 0$, we have that

$$a^2\mathbb{E}_X[X_1^2] + b^2\mathbb{E}_X[X_2^2] + 2ab\mathbb{E}_X[X_1X_2] \geq 0$$

$$a^2\mathbb{E}_X[X_1^2] + b^2\mathbb{E}_X[X_2^2] - 2ab\mathbb{E}_X[X_1X_2] \geq 0$$

Set $a^2 = \mathbb{E}_X[X_2^2]$ and $b^2 = \mathbb{E}_X[X_1^2]$. If either a or b is zero, the inequality clearly holds. We may thus consider $\mathbb{E}_X[X_1^2], \mathbb{E}_X[X_2^2] > 0$: we have

$$2\mathbb{E}_X[X_1^2]\mathbb{E}_X[X_2^2] + 2\{\mathbb{E}_X[X_1^2]\mathbb{E}_X[X_2^2]\}^{1/2}\mathbb{E}_X[X_1X_2] \geq 0$$

$$2\mathbb{E}_X[X_1^2]\mathbb{E}_X[X_2^2] - 2\{\mathbb{E}_X[X_1^2]\mathbb{E}_X[X_2^2]\}^{1/2}\mathbb{E}_X[X_1X_2] \geq 0$$

Rearranging, we obtain that

$$-\{\mathbb{E}_X[X_1^2]\mathbb{E}_X[X_2^2]\}^{1/2} \leq \mathbb{E}_X[X_1X_2] \leq \{\mathbb{E}_X[X_1^2]\mathbb{E}_X[X_2^2]\}^{1/2}$$

that is $\{\mathbb{E}_X[X_1X_2]\}^2 \leq \mathbb{E}_X[X_1^2]\mathbb{E}_X[X_2^2]$ or, in the original form

$$\{\mathbb{E}_X[g_1(X)g_2(X)]\}^2 \leq \mathbb{E}_X[\{g_1(X)\}^2]\mathbb{E}_X[\{g_2(X)\}^2].$$

Now, for equality:

$$\{\mathbb{E}_X[g_1(X)g_2(X)]\}^2 = \mathbb{E}_X[\{g_1(X)\}^2]\mathbb{E}_X[\{g_2(X)\}^2] \tag{4}$$

- If $\mathbb{E}_X[\{g_j(X)\}^2] = 0$ for $j = 1$ or 2, then $P_X[g_j(X) = 0] = 1$. The left-hand side of (3) is certainly non-negative, so must be zero.
- If $\mathbb{E}_X[\{g_j(X)\}^2] > 0$ for $j = 1, 2$, but $g_1(X) = cg_2(X)$ with probability one for some $c \neq 0$. In this case we replace $g_1(X)$ in the left- and right- hand sides of (3) to conclude that

$$\{\mathbb{E}_X[cg_2(X)^2]\}^2 = \mathbb{E}_X[\{cg_2(X)\}^2]\mathbb{E}_X[\{g_2(X)\}^2] = c^2\mathbb{E}_X[\{g_2(X)\}^2]$$

  and equality follows. Conversely, assume that (4) holds. If both sides equate to zero, then we must have at least one term on the right-hand side equal to zero, so $\mathbb{E}_X[\{g_j(X)\}^2] = 0$ for $j = 1$ or 2. If both sides equate to a positive constant then both $\mathbb{E}_X[\{g_j(X)\}^2] > 0$ so

$$\mathbb{E}_X[\{g_1(X)\}^2] = \frac{\{\mathbb{E}_X[g_1(X)g_2(X)]\}^2}{\mathbb{E}_X[\{g_2(X)\}^2]}.$$

Let $Z = g_1(X) - cg_2(X)$. Assume that Z is not zero with probability 1: we then have

$$0 < \mathbb{E}_Z[Z^2] = \mathbb{E}_X[\{g_1(X)\}^2] + c^2\mathbb{E}_X[\{g_2(X)\}^2] - 2c\mathbb{E}_X[g_1(X)g_2(X)]$$

However the right-hand side can be written,

$$\mathbb{E}_X[\{g_1(X)\}^2] + \left(c\{\mathbb{E}_X[\{g_2(X)\}^2]\}^{1/2} - \frac{\mathbb{E}_X[g_1(X)g_2(X)]}{\{\mathbb{E}_X[\{g_2(X)\}^2]\}^{1/2}}\right)^2 - \left(\frac{\mathbb{E}_X[g_1(X)g_2(X)]}{\{\mathbb{E}_X[\{g_2(X)\}^2]\}^{1/2}}\right)^2$$

Now if we set

$$c = \frac{\mathbb{E}_X[g_1(X)g_2(X)]}{\mathbb{E}_X[\{g_2(X)\}^2]}$$

the second term is zero, so we must then have

$$\mathbb{E}[\{g_1(X)\}^2] - \frac{\{\mathbb{E}[g_1(X)g_2(X)]\}^2}{\mathbb{E}[\{g_2(X)\}^2]} > 0$$

but this contradicts assumption (4). Hence Z must be zero with probability 1, that is $g_1(X) = cg_2(X)$ with probability 1.

4. **Hölder's Inequality:** Suppose $p, q > 1$ satisfy $p^{-1} + q^{-1} = 1$. Then

$$|\mathbb{E}_{X,Y}[XY]| \leq \mathbb{E}_{X,Y}[|XY|] \leq \{\mathbb{E}_X[|X|^p]\}^{1/p}\{\mathbb{E}_Y[|Y|^q]\}^{1/q}$$

for random variables X and Y

Lemma Let $a, b > 0$ and $p, q > 1$ satisfy $p^{-1} + q^{-1} = 1$. Then

$$p^{-1}\,a^p + q^{-1}\,b^q \geq ab$$

with equality if and only if $a^p = b^q$. To see this, fix $b > 0$. Let

$$g(a; b) = p^{-1}\,a^p + q^{-1}\,b^q - ab.$$

We require that $g(a; b) \geq 0$ for all a. Differentiating wrt a for fixed b yields $g^{(1)}(a; b) = a^{p-1} - b$, so that $g(a; b)$ is minimized (the second derivative is strictly positive at all a) when $a^{p-1} = b$, and at this value of a, the function takes the value

$$p^{-1}\,a^p + q^{-1}\,(a^{p-1})^q - a(a^{p-1}) = p^{-1}\,a^p + q^{-1}\,a^p - a^p = 0$$

as, $1/p + 1/q = 1 \implies (p-1)q = p$. As the second derivative is strictly positive at all a, the minimum is attained at the **unique** value of a where $a^{p-1} = b$, where, raising both sides to power q yields $a^p = b^q$.

Proof (of Hölder's Inequality, given in the continuous case) For the first inequality,

$$\mathbb{E}_{X,Y}[|XY|] = \iint |xy| f_{X,Y}(x,y)\,dx\,dy \geq \iint xy f_{X,Y}(x,y)\,dx\,dy = \mathbb{E}_{X,Y}[XY]$$

and

$$\mathbb{E}_{X,Y}[XY] = \iint xy f_{X,Y}(x,y)\,dx\,dy \geq \iint -|xy| f_{X,Y}(x,y)\,dx\,dy = -\mathbb{E}_{X,Y}[|XY|]$$

so

$$-\mathbb{E}_{X,Y}[|XY|] \leq \mathbb{E}_{X,Y}[XY] \leq \mathbb{E}_{X,Y}[|XY|] \qquad \therefore \qquad |\mathbb{E}_{X,Y}[XY]| \leq \mathbb{E}_{X,Y}[|XY|].$$

For the second inequality, using

$$a = \frac{|X|}{\{\mathbb{E}_X[|X|^p]\}^{1/p}} \qquad b = \frac{|Y|}{\{\mathbb{E}_Y[|Y|^q]\}^{1/q}}.$$

in the lemma, we have that

$$p^{-1}\frac{|X|^p}{\mathbb{E}_X[|X|^p]} + q^{-1}\frac{|Y|^q}{\mathbb{E}_Y[|Y|^q]} \geq \frac{|XY|}{\{\mathbb{E}_X[|X|^p]\}^{1/p}\{\mathbb{E}_Y[|Y|^q]\}^{1/q}}$$

and taking expectations yields, on the left hand side,

$$p^{-1}\frac{\mathbb{E}_X[|X|^p]}{\mathbb{E}_X[|X|^p]} + q^{-1}\frac{\mathbb{E}_Y[|Y|^q]}{\mathbb{E}_Y[|Y|^q]} = p^{-1} + q^{-1} = 1$$

and on the right hand side

$$\frac{\mathbb{E}_{X,Y}[|XY|]}{\{\mathbb{E}_X[|X|^p]\}^{1/p}\{\mathbb{E}_Y[|Y|^q]\}^{1/q}}$$

and the result follows.

Note: here we have equality if and only if

$$P_{X,Y}[|X|^p = c|Y|^q] = 1$$

for some non zero constant c.

Corollaries:

(a) Setting $p = q = 2$ in the Hölder Inequality, we have

$$|\mathbb{E}_{X,Y}[XY]| \leq \mathbb{E}_{X,Y}[|XY|] \leq \left\{\mathbb{E}_X[|X|^2]\right\}^{1/2}\left\{\mathbb{E}_Y[|Y|^2]\right\}^{1/2}$$

as in the Cauchy-Schwarz inequality.

(b) Let μ_X and μ_Y denote the expectations of X and Y respectively. Then

$$|\mathbb{E}_{X,Y}[(X-\mu_X)(Y-\mu_Y)]| \leq \left\{\mathbb{E}_X[(X-\mu_X)^2]\right\}^{1/2}\left\{\mathbb{E}_Y[(Y-\mu_Y)^2]\right\}^{1/2}$$

so that

CORRECTED $\{\mathbb{E}_{X,Y}[(X-\mu_X)(Y-\mu_Y)]\}^2 \leq \mathbb{E}_X[(X-\mu_X)^2]\mathbb{E}_Y[(Y-\mu_Y)^2]$

and defining the left-hand side as the square of the **covariance** between X and Y, $\text{Cov}_{X,Y}[X,Y]$, we have

$$\{\text{Cov}_{X,Y}[X,Y]\}^2 \leq \text{Var}_X[X]\ \text{Var}_Y[Y].$$

(c) **Lyapunov's Inequality**: Suppose $P_Y[Y=1] = 1$. Then, for $1 < p < \infty$

$$\mathbb{E}_X[|X|] \leq \{\mathbb{E}_X[|X|^p]\}^{1/p}.$$

Let $1 < r < p$. Then

$$\mathbb{E}_X[|X|^r] \leq \{\mathbb{E}_X[|X|^{pr}]\}^{1/p}$$

and letting $s = pr > r$ yields

$$\mathbb{E}_X[|X|^r] \leq \{\mathbb{E}_X[|X|^s]\}^{r/s}$$

so that

$$\{\mathbb{E}_X[|X|^r]\}^{1/r} \leq \{\mathbb{E}_X[|X|^s]\}^{1/s}$$

for $1 < r < s < \infty$.

(d) **Minkowski's Inequality:** Suppose that $1 \leq p < \infty$. Then

$$\{\mathbb{E}_{X,Y}[|X+Y|^p]\}^{1/p} \leq \{\mathbb{E}_X[|X|^p]\}^{1/p} + \{\mathbb{E}_Y[|Y|^p]\}^{1/p}$$

for random variables X and Y.

Proof Write

$$\begin{aligned} \mathbb{E}_{X,Y}[|X+Y|^p] &= \mathbb{E}_{X,Y}[|X+Y||X+Y|^{p-1}] \\ &\leq \mathbb{E}_{X,Y}[|X||X+Y|^{p-1}] + \mathbb{E}_{X,Y}[|Y||X+Y|^{p-1}] \end{aligned}$$

by the triangle inequality $|x+y| \leq |x| + |y|$. Using Hölder's Inequality on the terms on the right hand side, for q selected to satisfy $1/p + 1/q = 1$,

$$\begin{aligned} \mathbb{E}_{X,Y}[|X+Y|^p] \leq \{\mathbb{E}_X[|X|^p]\}^{1/p} \left\{\mathbb{E}_{X,Y}[|X+Y|^{q(p-1)}]\right\}^{1/q} \\ + \{\mathbb{E}_Y[|Y|^p]\}^{1/p} \left\{\mathbb{E}_{X,Y}[|X+Y|^{q(p-1)}]\right\}^{1/q} \end{aligned}$$

and dividing through by $\left\{\mathbb{E}_{X,Y}[|X+Y|^{q(p-1)}]\right\}^{1/q}$ yields

$$\frac{\mathbb{E}_{X,Y}[|X+Y|^p]}{\left\{\mathbb{E}_{X,Y}[|X+Y|^{q(p-1)}]\right\}^{1/q}} \leq \{\mathbb{E}_X[|X|^p]\}^{1/p} + \{\mathbb{E}_Y[|Y|^p]\}^{1/p}$$

and the result follows as $q(p-1) = p$, and $1 - 1/q = 1/p$.