Econ 240A: Some Useful Inequalities: Theory, Examples and Exercises[*]

Chuan Goh

8 February 2001

Lawrence Klein: "If the Devil promised you a theorem in return for your immortal soul, would you accept the bargain?"
Harold Freeman: "No. But I would for an inequality."
(Reported in Klein, 1991)

This handout expands on the text by presenting a number of inequalities that are frequently used in proving theorems in statistics.

Theorem 1 (Markov's Inequality). *Let Y be a random variable and $g(\cdot)$ be a function such that $g(y) \geq 0$ for all $y \in \Re$. Assume that $E[g(Y)]$ exists. Then for any $\epsilon > 0$*

$$P[g(Y) \geq \epsilon] \leq \frac{1}{\epsilon} E[g(Y)].$$

Proof. We give a proof for Y continuous. The proof for Y discrete is similar[1]. Let $B_1 = \{y : g(y) \geq \epsilon\}$ and $B_2 = \{y : g(y) < \epsilon\}$. Then

$$\begin{aligned} E[g(Y)] &= \int_{B_1} g(y)f(y)dy + \int_{B_2} g(y)f(y)dy \\ &\geq \int_{B_1} g(y)f(y)dy \\ &\geq \epsilon \int_{B_1} f(y)dy \\ &= \epsilon P[g(Y)] \geq \epsilon. \end{aligned}$$

□

The next theorem will be used in the proof of the law of large numbers:

[*] The material in this handout is a slightly modified version of Section 2.8 in Poirier (1995)

[1] Of course Y may be neither discrete nor continuous, and so this proof would not be acceptable in a probability theory course like Stat 205A.

Corollary 1.1 (Chebyshev's Inequality). *Let Y be a random variable with finite mean μ and variance σ^2. Then for all $\delta > 0$*

$$P[\,|Y - \mu| \geq \delta\sigma] \leq \frac{1}{\delta^2}.$$

Proof. Set $g(y) = (y - \mu)^2$ and $\epsilon = \delta^2\sigma^2$ and apply Markov's Inequality. □

Theorem 2 (Cauchy-Schwarz Inequality). *Let random variables $Y_n, n = 1, 2$ have finite second moments. Then*

$$[E(Y_1Y_2)]^2 \leq E(Y_1^2)E(Y_2^2),$$

with equality if $P(Y_2 = kY_1) = 1$ for some constant k.

Proof. See the solution to Problem 22 from Chapter 3 in McFadden (2001) (Problem Set 3). □

Note that if $\sigma_n^2 \equiv Var[Y_n] > 0$, $n = 1, 2$ a direct consequence of the Cauchy-Schwarz Inequality is that

$$\rho_{12} \equiv \frac{\sigma_{12}}{\sqrt{\sigma_1\sigma_2}} \in [-1, 1],$$

where $\sigma_{12} \equiv Cov[Y_1, Y_2]$. Also it follows from the same inequality that the covariance matrix

$$\Sigma_{12} \equiv \begin{bmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{bmatrix}$$

is positive definite if $|\rho_{12}| < 1$ and positive semidefinite if $|\rho_{12}| = 1$.

The following inequality is a generalization of the Cauchy-Schwarz Inequality:

Theorem 3 (Hölder's Inequality). *Let c, d be real numbers such that $c > 1, d > 1, c^{-1} + d^{-1} = 1$. Let Y_1, Y_2 be random variables such that $E[|Y_1|^c] < \infty$ and $E[|Y_2|^d < \infty$. Then*

$$E[|Y_1Y_2|] \leq [E|Y_1|^c]^{\frac{1}{c}}[E|Y_2|^d]^{\frac{1}{d}}.$$

Proof. See Billingsley (1995, p.80). □

Definition 1. *A continuous function $g(\cdot)$ is called convex iff $\forall x, y \in \Re$ and $\lambda \in [0, 1]$, $g[\lambda x + (1 - \lambda)y] \leq \lambda g(x) + (1 - \lambda)g(y)$. $g(\cdot)$ is strictly convex if the inequality is strict. $g(\cdot)$ is concave iff $-g(x)$ is convex and strictly concave iff $-g(x)$ is strictly convex.*

Theorem 4 (Jensen's Inequality). *Let Y be a random variable and $g(\cdot)$ a convex function. If $E[Y]$ and $E[g(Y)]$ exist, then*

$$E[g(Y)] \geq g[E(Y)].$$

If $g(\cdot)$ is strictly convex, the inequality is strict unless Y is a constant with probability one.

Proof. See Lehmann & Casella (1998, p. 46). □

Example 1. *$g(y) = y^2$ is convex, so Jensen's inequality implies that $E[Y^2] \geq [E(Y)]^2$. It follows immediately that $Var(Y) = E[Y^2] - [E(Y)]^2 \geq 0$.*

Example 2. *Suppose Y is a nondegenerate positive random variable with finite mean. Then Jensen's Inequality implies*

(a) $E[Y^{-1}] > [E(Y)]^{-1}$,

(b) $E[\log Y] < \log[E(Y)]$.

Theorem 5. *Let Y be a positive random variable such that the required moments exist:*

(a) $E(Y^\alpha) \leq [E(Y)]^\alpha$, $\alpha \in [0, 1]$.

(b) $E(Y^\alpha) \geq [E(Y)]^\alpha$, $\alpha \leq 0$ *or* $\alpha \geq 1$.

(c) $h(\alpha) \equiv \frac{E(Y^{\alpha-1})}{E(Y^\alpha)}$ *is a nonincreasing function of* $\alpha > 0$.

Proof. Exercise. Use Jensen's inequality. □

The following theorems are stated without proof:

Theorem 6. *Let Y and Z be independent random variables such that the required moments exist:*

(a) $E[(\frac{Y}{Z})^\alpha] \geq \frac{E(Y^\alpha)}{E(Z^\alpha)}$ *for all* α.

(b) If $E(Y) = E(Z) = 0$ *and* $E(Z^{-1})$ *exists, then* $Var(\frac{Y}{Z}) \geq \frac{Var(Y)}{Var(Z)}$.

Theorem 7. *Suppose Y is a random variable with $E(|Y|^\alpha) < \infty$ for some $\alpha > 0$. Then $E[|Y|^\beta] < \infty$ for $0 \leq \beta \leq \alpha$.*

Theorem 8 (Minkowski's Inequality). *Let Y, Z be random variables such that $E(|Y|^\alpha) < \infty$ and $E(|Z|^\alpha) < \infty$ for some $1 \leq \alpha < \infty$. Then*

$$[E(|Y + Z|^\alpha)]^{\frac{1}{\alpha}} \leq [E(|Y|^\alpha)]^{\frac{1}{\alpha}} + [E(|Z|^\alpha)]^{\frac{1}{\alpha}}.$$

Theorem 9. *Let Y be a random variable such that $E[|Y|^\alpha] < \infty$. Then*

$$\lim_{y \to \infty} y^\alpha P(|Y| \geq y) = 0.$$

Theorem 10 (Information Inequality). *Let $\sum_i a_i$ and $\sum_i b_i$ be convergent series of positive numbers such that $\sum_i a_i \geq \sum_i b_i$. Then*

$$\sum_{i=1}^{\infty} a_i \log(\frac{b_i}{a_i}) \leq 0,$$

where the inequality is strict unless $a_i = b_i$, $i = 1, 2, 3, \ldots$.

Exercises

1.

Let A and B be arbitrary events. Prove *Boole's inequality*: $P(A\cap B) \geq P(A) - P(B^c) = 1 - P(A^c) - P(B^c)$.

2.

Let A, B, C be events such that $A\cap B \subset C$. Show that $P(C^c) \leq P(A^c)+P(B^c)$.

3.

Let Y be a discrete random variable with $P(Y=-1)=\frac{1}{8}, P(Y=0)=\frac{3}{4}$ and $P(Y=1)=\frac{1}{8}$. Show that Chebyshev's inequality holds exactly for $\delta = 2$.

4.

Show that if $g(\cdot)$ is concave, $E[g(Y)] \leq g[E(Y)]$.

5.

Assume $\mu \equiv E(Y) < \infty$ and $g(\cdot)$ is a nondecreasing function. Use Jensen's inequality to show that $E[g(Y)(Y-\mu)] \geq 0$.

6.

Suppose $y_n, n = 1, \ldots, N$ are positive numbers. Define the following:

- Arithmetic mean: $\bar{y} = \frac{1}{N}\sum_{n=1}^{N} y_n$,
- Geometric mean: $\bar{y}_G = [\prod_{n=1}^{N} y_n]^{\frac{1}{N}}$,
- Harmonic mean: $\bar{y}_H = [\frac{1}{N}\sum_{n=1}^{N}\frac{1}{y_n}]^{-1}$.

Use Jensen's inequality to show $\bar{y}_H \leq \bar{y}_G \leq \bar{y}$.

7.

Let $Y_n, n = 1, \ldots, N$ be jointly distributed with $E(Y_n) = \mu_n$ and $Var(Y_n) = \sigma_n^2, n = 1, \ldots, N$. Let $A_n \equiv \{Y_n : |Y_n - \mu_n| \leq \sqrt{N}\delta\rho_n\}$, where $\delta > 0$. Prove the *multivariate Chebyshev's inequality*: $P(A_1 \cap A_2 \cap \cdots \cap A_N) \geq 1 - \delta^{-2}$.

8.

Use Jensen's inequality to prove *Liapounov's inequality*: If $E[|Y|^r]$ exists and $0 < s < r < \infty$, then $[E(|Y|^s)]^{\frac{1}{s}} \leq [E(|Y|^r)]^{\frac{1}{r}}$.

9.

Let $M_Y(t)$ be the mgf of a random variable Y. Show that $P(Y \geq c) \leq \exp(-ct)M_Y(t)$ for $0 < t < h < \infty$.

References

[1] Billingsley, P. (1995): *Probability and Measure*, Third Edition. New York: J. Wiley & Sons.

[2] Klein, L. (1991): "The Statistics Seminar, MIT, 1942–1943." *Statistical Science*, 6, 320–338.

[3] Lehmann, E. L. and G. Casella (1998): *Thoery of Point Estimation*, Second Edition. New York: Springer.

[4] McFadden, D. L. (2001): *Statistical Tools for Economists.* Unpublished volume, University of California, Berkeley.

[5] Poirier, D. J. (1995): *Intermediate Statistics and Econometrics: A Comparative Approach.* Cambridge, Mass.: MIT Press.