STA 623 – Fall 2009 – Dr. Charnigo

Section 4.7: Inequalities

Hölder's Inequality and Cauchy-Schwarz Inequality. Let p and q be positive numbers with $1/p + 1/q = 1$. Then, for any random variables X and Y, we have

$$|E[XY]| \leq E[|XY|] \leq (E[|X|^p])^{1/p}(E[|Y|^q])^{1/q}$$

whenever all of these expectations exist as finite numbers. Your textbook authors provide the proof.

The special case when $p = q = 2$ is called the Cauchy-Schwarz Inequality,

$$|E[XY]| \leq E[|XY|] \leq (E[X^2])^{1/2}(E[Y^2])^{1/2}.$$

Applications of Hölder's Inequality and Cauchy-Schwarz Inequality.

1. Suppose that X and Y are random variables with finite second moments. Can we derive the relationship $|\rho_{XY}| \leq 1$ from the Cauchy-Schwarz Inequality? We need to exercise some care here since $E[XY]$ is not the same as $Cov[X, Y]$ unless $E[X]$ or $E[Y]$ happens to equal 0.

2. By letting Y equal 1 with probability 1, we see that Hölder's Inequality yields

$$E[|X|] \leq (E[|X|^p])^{1/p}$$

at any $p \in (1, \infty)$ for which both expectations exist as finite numbers. What is the interpretation of this result when $p = 2$?

3. Let $a_1, \ldots, a_n$ and $b_1, \ldots, b_n$ be arbitrary real constants. I claim that

$$\sum_{i=1}^{n} |a_i b_i| \leq \left(\sum_{i=1}^{n} |a_i|^p\right)^{1/p} \left(\sum_{i=1}^{n} |b_i|^q\right)^{1/q}.$$

Can you explain how this follows from Hölder's Inequality?

4. A special case of item 3 arises when $b_1 = \cdots = b_n = 1$ and $p = q = 2$,

$$\sum_{i=1}^{n} |a_i| \leq n^{1/2} \left(\sum_{i=1}^{n} a_i^2\right)^{1/2}.$$

Thus, if you are told that $\sum_{i=1}^{n} a_{i,n}^2 \to 0$ as $n \to \infty$, you may conclude that $n^{-1/2} \sum_{i=1}^{n} |a_{i,n}| \to 0$. (Here I write $a_{i,n}$ rather than a_i since the summands may change with n.) Is this conclusion valid without the absolute value, $n^{-1/2} \sum_{i=1}^{n} a_{i,n} \to 0$? Is this conclusion valid without the $n^{-1/2}$, $\sum_{i=1}^{n} |a_{i,n}| \to 0$?

Jensen's Inequality. Let X be a random variable whose support set is contained in an interval $I \subset \mathbb{R}$, and suppose that g is a convex function from I to $\mathbb{R}$. That is, suppose that $g(tu + (1-t)v) \leq tg(u) + (1-t)g(v)$ for $t \in (0, 1)$, $u \in I$, and $v \in I$. (If g is twice differentiable, then nonnegativity of the second derivative implies convexity of g.) Provided that both expectations exist as finite numbers, we have

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

Your textbook authors provide the proof.

A function g is said to be concave if $-g$ is convex. (If g is twice differentiable, then nonpositivity of the second derivative implies concavity of g.) Jensen's

Inequality is reversed for concave g,

$$E[g(X)] \leq g(E[X]).$$

Remarks on Jensen's Inequality.

1. To remember the direction of the inequality, think of the convex function $g(x) := x^2$. You know that $E[X^2]$ must be greater than or equal to $(E[X])^2$ because their difference is $Var[X]$, which must be nonnegative.

2. In some instances, we may wish to obtain the stronger conclusion that

$$E[g(X)] > g(E[X]).$$

However, the stronger conclusion is not universally valid. For instance, if X is degenerate or g is linear, then we have equality. A sufficient condition for the stronger conclusion is that X not be degenerate and g have strictly positive second derivative.

An application of Jensen's Inequality. Let X have the gamma distribution with known shape parameter $n \in \{2, 3, \ldots\}$ and unknown rate parameter $\lambda \in (0, \infty)$. Suppose that we wish to "guess" λ after observing a realization of X. Noting that the expected value of X is n/λ, we may think that the realization of X is a good guess for n/λ and hence the realization of n/X is a good guess for λ. While this strategy does not seem unreasonable, on average we will [choose one and then justify: underestimate / overestimate] λ.

Note that we did not need to compute $E[n/X]$ to conclude that it differed from λ. But, just for fun, let us compute $E[n/X]$ anyway: