Lecture 10: Probability inequalities

Ciaran Evans

Last time

- Wald tests for single coefficients:

$$Z = \frac{\widehat{\beta}_j - 0}{\widehat{SE}(\widehat{\beta}_j)} \qquad \text{under } H_0, \; Z \approx N(0,1)$$

- Tests for nested models:

$$G = 2(\log L_{\text{full}} - \log L_{\text{reduced}}) \qquad \text{under } H_0, \; G \approx \chi^2_q$$

What we need

We need to show that

$$\widehat{\beta} \approx N(\beta, \mathcal{I}^{-1}(\beta))$$

This requires:

- a notion of convergence of random variables
- asymptotic results about MLEs
- hypothesis testing fundamentals

Roadmap:

1. Preliminary machinery – probability inequalities, types of convergence, theorems about convergence
2. Properties of MLEs – consistency and asymptotic normality
3. Hypothesis testing theory – types of hypotheses, types of error, and types of hypothesis test (Neyman-Pearson, Wald, Likelihood ratio)

Markov's inequality

Theorem: Let Y be a non-negative random variable, and suppose that $\mathbb{E}[Y]$ exists. Then for any $t > 0$,

$$P(Y \geq t) \leq \frac{\mathbb{E}[Y]}{t} \quad \Longleftrightarrow \quad tP(Y \geq t) \leq \mathbb{E}[Y]$$

Pf: (continuous case, discrete case is similar)

$$tP(Y \geq t) = t\int_t^\infty f(y)\,dy = \int_t^\infty t\,f(y)\,dy$$

$$\leq \int_t^\infty y\,f(y)\,dy$$

$$\leq \int_0^\infty y\,f(y)\,dy = \mathbb{E}[Y]$$

$$\Rightarrow \quad P(Y \geq t) \leq \frac{\mathbb{E}[Y]}{t}$$

//

Chebyshev's inequality

Recall: $Var(\bar{X}) = \frac{\sigma^2}{n} \to 0$ as $n \to \infty$

Theorem: Let Y be a random variable, and let $\mu = \mathbb{E}[Y]$ and $\sigma^2 = Var(Y)$. Then

$$P(|Y - \mu| \geq t) \leq \frac{\sigma^2}{t^2}$$

With your neighbor, apply Markov's inequality to prove Chebyshev's inequality.

$\sigma^2 = Var(Y) = \mathbb{E}[(Y-\mu)^2]$

Pf: $P(|Y-\mu| \geq t) = P((Y-\mu)^2 \geq t^2)$

$\leq \frac{\mathbb{E}[(Y-\mu)^2]}{t^2} = \frac{\sigma^2}{t^2}$ //

Cauchy-Schwarz inequality

Theorem: For any two random variables X and Y,

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

Casella & Berger, Theorem 4.7.3

Example: The *correlation* between X and Y is defined by

$$\rho(X, Y) = \frac{Cov(X, Y)}{\sqrt{Var(X)}\sqrt{Var(Y)}}$$

Using the Cauchy-Schwarz inequality, show that $-1 \le \rho(X, Y) \le 1$.

$$Cov(X,Y) = \mathbb{E}[XY] - \mathbb{E}[X]\,\mathbb{E}[Y]$$
$$= \mathbb{E}[(X-\mu_X)(Y-\mu_Y)]$$

$$|\mathbb{E}[(X-\mu_X)(Y-\mu_Y)]| \le \mathbb{E}[|(X-\mu_X)(Y-\mu_Y)|]$$
$$\le \left(\mathbb{E}[(X-\mu_X)^2]\right)^{\frac{1}{2}} \left(\mathbb{E}[(Y-\mu_Y)^2]\right)^{\frac{1}{2}}$$
$$= \sqrt{Var(X)}\ \sqrt{Var(Y)}$$

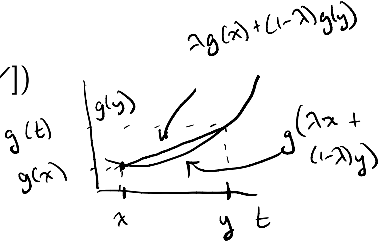
$$\Rightarrow |cov(X,Y)| \le \sqrt{Var(X)}\ \sqrt{Var(Y)} \quad \Rightarrow \frac{|Cov(X,Y)|}{\sqrt{Var(X)}\sqrt{Var(Y)}} \le 1 \quad //$$

Jensen's inequality

Theorem: For any random variable Y, if g is a convex function, then

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

Recall: convex function.

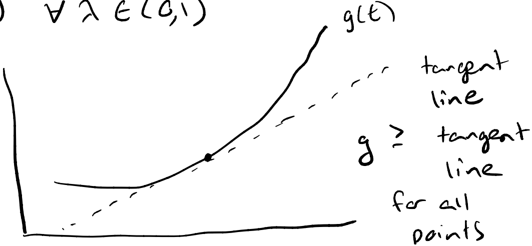


Def: g is convex if

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

$\forall x, y$ and $\forall \lambda \in (0,1)$

Equivalently:



Example:

$$g(t) = t^2 \quad \text{is convex}$$

Apply Jensen's inequality:

$$\mathbb{E}[X^2] \geq (\mathbb{E}[X])^2$$

$$\iff \underbrace{\mathbb{E}[X^2] - (\mathbb{E}[X])^2}_{Var(X)} \geq 0$$

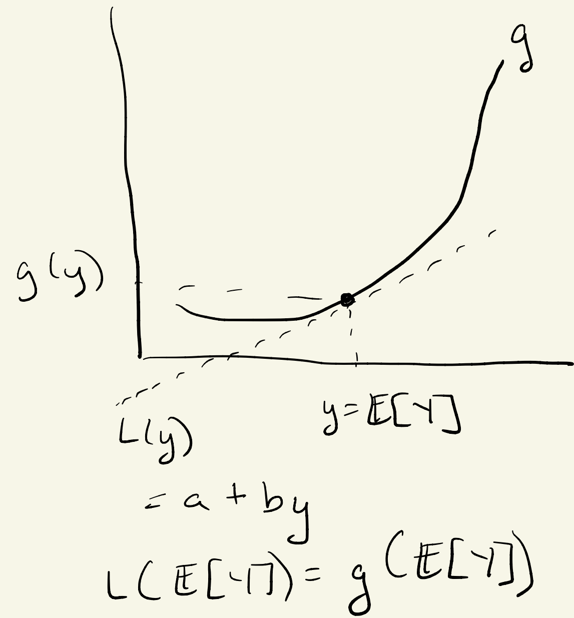
Thm: if g is convex, then

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

Pf: Let $L(y)$ be the tangent line to $g(y)$ at the point $y = \mathbb{E}[Y]$

$L(y) = a + by$ for some a, b

By convexity, $g(y) \geq a + by \quad \forall y$



$$\begin{aligned}
\Rightarrow \mathbb{E}[g(Y)] &\geq \mathbb{E}[L(Y)] \\
&= \mathbb{E}[a + bY] \\
&= a + b\,\mathbb{E}[Y] \\
&= L(\mathbb{E}[Y]) \\
&= g(\mathbb{E}[Y])
\end{aligned}$$

$$\Rightarrow \mathbb{E}[g(Y)] \geq g(\mathbb{E}[Y]) \qquad //$$