Beginning asymptotics

Ciaran Evans

Course plan

So far: maximum likelihood estimation

- Univariate and multivariate estimation
- Applications and connections to regression models
- Cases where support depends on the parameter (e.g. $Uniform(0, \theta)$)
- Invariance of MLE
- Situations without a closed form solution (e.g. Newton's method for GLMs)

Still to come:

- Asymptotic properties of the MLE
- Hypothesis tests and confidence intervals for parameters of interest
- Other approaches to estimation

Motivation: the Titanic data

Data on 891 passengers on the *Titanic*. Variables include:

- `Survived`
- `Pclass`
- `Sex`
- `Age`

$$Survived_i | \mathbf{x}_i \sim Bernoulli(p_i)$$

$$\log\left(\frac{p_i}{1-p_i}\right) = \beta_0 + \beta_1 Male_i + \beta_2 Age_i + \beta_3 Class2_i + \beta_4 Class3_i$$

Fitting the model in R

$\hat{\beta}$ $SE(\hat{\beta})$ test statistic p-values

	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	3.777	0.401	9.416	4.682e-21
Sexmale	-2.523	0.207	-12.164	4.811e-34
Age	-0.037	0.008	-4.831	1.359e-06
Pclass2	-1.310	0.278	-4.710	2.472e-06
Pclass3	-2.581	0.281	-9.169	4.761e-20

Suppose I want to know whether there is a relation between age and the probability of survival, after accounting for passenger class and sex. What hypotheses would I test?

$H_0: \beta_2 = 0$ (no relationship)

$H_A: \beta_2 \neq 0$ (some relationship

Need:

- test statistic
- null distribution (distribution of test statistic under H_0)

z-tests for single coefficients

	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	3.777	0.401	9.416	4.682e-21
Sexmale	-2.523	0.207	-12.164	4.811e-34
Age	-0.037	0.008	-4.831	1.359e-06
Pclass2	-1.310	0.278	-4.710	2.472e-06
Pclass3	-2.581	0.281	-9.169	4.761e-20

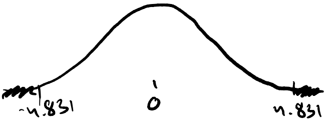
$H_0: \beta_2 = 0 \qquad H_A: \beta_2 \neq 0$

Test statistic:

$$Z = \frac{\hat{\beta}_2 - 0}{\hat{SE}(\hat{\beta}_2)} = \frac{-0.037 - 0}{0.008} \approx -4.831$$

Null distribution: Under H_0, $Z \approx N(0,1)$

$N(0,1)$



What we need

We need to show that

- $\widehat{\beta} \approx$ Normal
- We can find $\mathbb{E}[\hat{\beta}]$ and $Var(\hat{\beta})$

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 $t > 0$,

$f =$ pdf of Y

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

Pf: (continuous case, discrete case is similar)

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

$(t > 0,\ y \geq 0)$

$$\geq \int_t^\infty t f(y) dy$$

($y \geq t$ on (t, ∞), range of integration)

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

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

//

Chebyshev's inequality

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

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

We can apply Markov's inequality to prove Chebyshev's inequality.

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

<u>Pf</u>:

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

(markov's inequality)
$(Y-\mu)^2 \geq 0$

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

$$= \frac{\sigma^2}{t^2}$$

//

Cauchy-Schwarz inequality

Theorem: For any two random variables X and Y,

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

(see 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, we can show that $-1 \leq \rho(X, Y) \leq 1$. we will prove $|\rho(X,Y)| \leq 1$

Let $\mu_X = \mathbb{E}[X]$, $\mu_Y = \mathbb{E}[Y]$

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

C-S: $$|Cov(X,Y)| = |\mathbb{E}[(X-\mu_X)(Y-\mu_Y)]| \leq (\mathbb{E}[(X-\mu_X)^2])^{\frac{1}{2}}(\mathbb{E}[(Y-\mu_Y)^2])^{\frac{1}{2}} = \sqrt{Var(X)}\sqrt{Var(Y)}$$

$$\Rightarrow \frac{|Cov(X,Y)|}{\sqrt{Var(X)}\sqrt{Var(Y)}} \leq 1 \quad \Rightarrow |\rho(X,Y)| \leq 1$$

//

Jensen's inequality

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

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

Example: $g(t) = t^2$ is convex

Jensen's inequality: $\mathbb{E}[Y^2] \geq (\mathbb{E}[Y])^2$

$\Rightarrow \mathbb{E}[Y^2] - (\mathbb{E}[Y])^2 \geq 0$

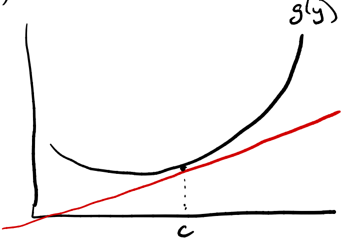
$\Rightarrow \text{var}(Y) \geq 0$

(variance is non-negative)

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

Let L_c be the tangent line to g at some point c

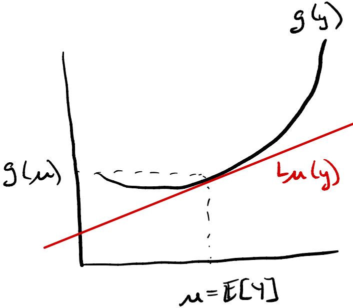
If g is differentiable, then g is convex if $g(y) \geq L_c(y)$ for all points y, for all points c

(g lies above all tangent lines)

Jensen's inequality

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

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



pf: (case where g is differentiable)

Let $L_\mu(y)$ be tangent line to $g(y)$ at the point $\mu = \mathbb{E}[Y]$. $L_\mu(y) = a + by$ for some a, b

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

$\Rightarrow$ For r.v. Y, $\quad g(Y) \geq a + bY$

$\Rightarrow \mathbb{E}[g(Y)] \geq \mathbb{E}[a+bY] = a + b\,\mathbb{E}[Y] = a + b\mu$

$= L_\mu(\mu) = g(\mu) = g(\mathbb{E}[Y])$

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