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

	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?

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

What we need

We need to show that

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

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 $t > 0$,

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

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.

Cauchy-Schwarz inequality

Theorem: For any two random variables X and Y,

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

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 $-1 \leq \rho(X, Y) \leq 1$.

Jensen's inequality

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

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