Lecture 15: Asymptotic normality and beginning testing

Ciaran Evans

Convergence of the MLE

Suppose that we observe $Y_1, Y_2, Y_3, \ldots$ iid from a distribution with probability function $f(y|\theta)$, where $\theta \in \mathbb{R}^d$ is the parameter(s) we are trying to estimate. Let

$$\ell_n(\theta) = \sum_{i=1}^{n} \log f(Y_i|\theta)$$

log likelihood

$$\widehat{\theta}_n = \text{argmax}_\theta \; \ell_n(\theta)$$

MLE

$$\mathcal{I}_1(\theta) = -\mathbb{E}\left[\frac{\partial^2}{\partial\theta^2} \log f(Y_i|\theta)\right]$$

Fisher info for one observation

Theorem: Under certain regularity conditions,

(a) $\widehat{\theta}_n \xrightarrow{p} \theta$ (consistency)

(b) $\sqrt{n}(\widehat{\theta}_n - \theta) \xrightarrow{d} N(0, \mathcal{I}_1^{-1}(\theta))$ (asymptotic normality)

Proof of asymptotic normality

① $$\sqrt{n}(\hat{\theta} - \theta) \approx \frac{\frac{1}{\sqrt{n}} \ell_n'(\theta)}{-\frac{1}{n} \ell_n''(\theta)}$$ (Taylor expansion)

② $$\frac{1}{\sqrt{n}} \ell_n'(\theta) \xrightarrow{d} N(0, \mathcal{I}_1(\theta))$$ (CLT)

③ $$-\frac{1}{n} \ell_n''(\theta) \xrightarrow{P} \mathcal{I}_1(\theta)$$ (WLLN)

④ $$\sqrt{n}(\hat{\theta} - \theta) \xrightarrow{d} \frac{1}{\mathcal{I}_1(\theta)} \cdot N(0, \mathcal{I}_1(\theta))$$

$$= N(0, \mathcal{I}_1^{-1}(\theta))$$

Slutsky's theorem

Let $\{X_n\}, \{Y_n\}$ be sequences of random variables, and suppose that $X_n \xrightarrow{d} X$ and $Y_n \xrightarrow{p} c$ for some constant c. Then:

- $X_n + Y_n \xrightarrow{d} X + c$
- $X_n Y_n \xrightarrow{d} cX$
- $\dfrac{X_n}{Y_n} \xrightarrow{d} \dfrac{X}{c}$ (if $c \neq 0$)

Some sufficient regularity conditions

- The dimension of θ does not change with n
- $f(y \mid \theta)$ is a sufficiently smooth function of θ
- We can swap integration & differentiation (see C&B section 2.4)
- θ is <u>identifiable</u> (basically, $f(y \mid \theta_1) \neq f(y \mid \theta_2)$ if $\theta_1 \neq \theta_2$)
- θ is not on the boundary of the parameter space

  e.g., Bernoulli(θ): θ cannot be 0 or 1

  Normal(μ, σ^2): σ^2 cannot be 0

A counterexample

Suppose $Y_1, Y_2, \ldots \overset{iid}{\sim} Uniform(0, \theta)$.

$\hat{\theta} = Y_{(n)}$ Previously: $\hat{\theta} \xrightarrow{P} \theta$

$n(\hat{\theta} - \theta) \xrightarrow{d} -1 \cdot \text{Exponential}\left(\frac{1}{\theta}\right)$

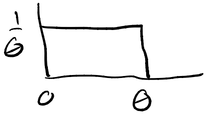
$$\sqrt{n}(\hat{\theta} - \theta) = \underbrace{\frac{1}{\sqrt{n}}}_{\to 0} \cdot \underbrace{n(\hat{\theta} - \theta)}_{\xrightarrow{d} -\text{Exp}\left(\frac{1}{\theta}\right)} \xrightarrow{P} 0$$

(Slutsky's)

(Not Normal)

Regularity conditions violated

- $f(y|\theta)$ is not smooth at θ



$\Rightarrow$ derivatives are not defined at θ

$\Rightarrow$ $\mathcal{I}_1(\theta)$ is not defined

- can't exchange integration & differentiation

A counterexample

Suppose $Y_1, Y_2, \ldots \overset{iid}{\sim} Bernoulli(p)$.

- $\widehat{p} = \bar{Y} = \frac{1}{n} \sum_i Y_i$
- If $p = 0$ or $p = 1$, what is $\sqrt{n}(\widehat{p} - p)$?

If $p = 0$: all $Y_i = 0 \Rightarrow \hat{p} = 0$

If $p = 1$: all $Y_i = 1 \Rightarrow \hat{p} = 1$

$\sqrt{n}(\hat{p} - p) = 0 \overset{d}{\not\to}$ Normal

In general, if $Var(\hat{\theta}) = 0$, won't get the asymptotic normality

Where we are going

So far:

- How can we estimate parameters/ fit a model?
- Asymptotic properties of MLEs

Next:

- How can we use our estimates for inference?

Future:

- General hypothesis testing framework
- Other methods for testing hypotheses
- Confidence intervals

Hypothesis tests for a population mean

Let $X_1, X_2, \ldots$ be an iid sample from a population with mean μ and variance σ^2. We want to test

value hypothesized under H_0

$$H_0 : \mu = \mu_0 \qquad H_A : \mu \neq \mu_0$$

Test statistic:

$$Z_n = \frac{\bar{X}_n - \mu_0}{\sigma / \sqrt{n}}$$

CLT: $\sqrt{n}(\bar{X}_n - \mu) \xrightarrow{d} N(0, \sigma^2)$

<u>if</u> H_0 is true, $\sqrt{n}(\bar{X}_n - \mu_0) \xrightarrow{d} N(0, \sigma^2)$

$$\Rightarrow \frac{\sqrt{n}(\bar{X}_n - \mu_0)}{\sigma} \xrightarrow{d} N(0,1)$$

$$\Rightarrow \quad Z_n \approx N(0,1) \quad \text{under } H_0$$

(null distribution)

Hypothesis tests for a population mean

Let $X_1, X_2, \ldots$ be an iid sample from a population with mean μ and variance σ^2. We want to test

$$H_0 : \mu = \mu_0 \qquad H_A : \mu \neq \mu_0$$

Test statistic: $Z_n = \dfrac{\overline{X}_n - \mu_0}{\sigma/\sqrt{n}}$

What if σ is unknown?

$$S = \sqrt{\frac{1}{n-1} \sum_i (X_i - \bar{X}_n)^2} \qquad S \xrightarrow{P} \sigma$$

$$\frac{\bar{X}_n - \mu_0}{S/\sqrt{n}} = \underbrace{\frac{\bar{X}_n - \mu_0}{\sigma/\sqrt{n}}}_{\xrightarrow{d} N(0,1)} \cdot \underbrace{\frac{\sigma}{S}}_{\xrightarrow{P} 1} \xrightarrow{d} N(0,1)$$

(Slutsky's)

Alternative: $\dfrac{\bar{X}_n - \mu_0}{S/\sqrt{n}} \approx t_{n-1}$ as $\nu \to \infty$, $t_\nu \to N(0,1)$

Hypothesis tests for a population mean

Let $X_1, X_2, \dots$ be an iid sample from a population with mean μ and variance σ^2. We want to test

$$H_0 : \mu = \mu_0 \qquad H_A : \mu \neq \mu_0$$

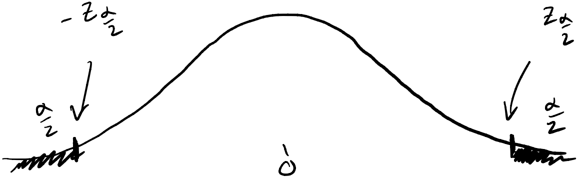
Test statistic: $Z_n = \dfrac{\overline{X}_n - \mu_0}{\sigma/\sqrt{n}}$

Rejecting: Should we reject H_0 when Z_n is close to 0, or when Z_n is far away from 0?

Reject when $|Z_n| >$ cutoff

Distribution of Z_n under H_0: $N(0,1)$

reject when $|Z_n| > z_{\frac{\alpha}{2}}$



Hypothesis tests for a population mean

Let $X_1, X_2, \ldots$ be an iid sample from a population with mean μ and variance σ^2. We want to test

$$H_0 : \mu = \mu_0 \qquad H_A : \mu \neq \mu_0$$

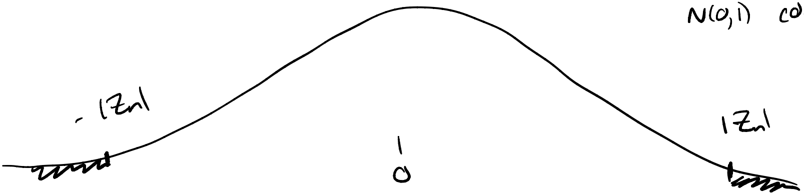
Test statistic: $Z_n = \dfrac{\overline{X}_n - \mu_0}{\sigma/\sqrt{n}}$

Rejecting: Reject H_0 when $|Z_n| > z_{\alpha/2}$

p-value: How do we calculate a p-value?

p-value $= 2\,\Phi(-|Z_n|)$

Φ: $N(0,1)$ cdf



Wald test for one parameter