Lecture 27: Interval estimation

Ciaran Evans

Motivation

Suppose we have data $(X_1, Y_1), ..., (X_n, Y_n)$ with

$$Y_i \sim Bernoulli(p_i)$$

$$\log\left(\frac{p_i}{1-p_i}\right) = \beta^T X_i$$

So far, we have discussed:

- Finding point estimates $\widehat{\beta}$
- Testing hypotheses about the true (but unknown) parameters β

Question: What are the limitations of point estimates and hypothesis tests for inference about β?

Confidence interval

```
##                 Estimate    Std. Error   z value       Pr(
## (Intercept)  2.641506279 0.1213233066  21.77246 4.233346
## WBC         -0.289290446 0.0134349261 -21.53272 7.689284
## PLT         -0.006561464 0.0005932064 -11.06101  1.93894
```

Question: How would I calculate a 95% confidence interval for β_1 (the change in the log odds of dengue for a one-unit increase in WBC, holding PLT fixed)?

$$\underbrace{\hat{\beta}_1 \pm z_{\frac{\alpha}{2}} SE(\hat{\beta}_1)}_{1-\alpha \text{ Wald CI}}$$

$z_{\frac{\alpha}{2}}$ = upper $\frac{\alpha}{2}$ quantile of $N(0,1)$

95% CI : $-0.289 \pm 1.96 (0.0134)$

$= (-0.315, -0.262)$

Confidence interval

```
##                  Estimate   Std. Error   z value      Pr(
## (Intercept)  2.641506279 0.1213233066  21.77246 4.233346
## WBC         -0.289290446 0.0134349261 -21.53272 7.689284
## PLT         -0.006561464 0.0005932064 -11.06101  1.93894
```

95% confidence interval for β_1: (-0.315, -0.262)

β_1 is either in this interval or not (β_1 is not a random variable)

Question: How do I interpret this confidence interval?

95% confident: if we take many samples from the population, and we calculate many intervals, 95% should contain the true (unknown) parameter

Deriving the coverage probability

Suppose $\hat{\theta} \approx N(\theta, Var(\hat{\theta})) \Rightarrow \dfrac{\hat{\theta} - \theta}{SE(\hat{\theta})} \approx N(0,1)$

$$(1-\alpha) \text{ Wald interval : } \widehat{\theta} \pm z_{\alpha/2} SE(\widehat{\theta})$$

$$P\left(\hat{\theta} - z_{\frac{\alpha}{2}} SE(\hat{\theta}) \leq \theta \leq \hat{\theta} + z_{\frac{\alpha}{2}} SE(\hat{\theta})\right) \approx 1-\alpha$$

<u>endpoints</u> are random (function of data)

$$= P\left(-z_{\frac{\alpha}{2}} SE(\hat{\theta}) \leq \hat{\theta} - \theta \leq z_{\frac{\alpha}{2}} SE(\hat{\theta})\right)$$

$$= P\left(-z_{\frac{\alpha}{2}} \leq \underbrace{\frac{\hat{\theta} - \theta}{SE(\hat{\theta})}}_{\approx N(0,1)} \leq z_{\frac{\alpha}{2}}\right) \approx 1-\alpha$$

Note: $\theta_0 \in \left[\hat{\theta} - z_{\frac{\alpha}{2}} SE(\hat{\theta}), \hat{\theta} + z_{\frac{\alpha}{2}} SE(\hat{\theta})\right]$

$\Leftrightarrow \left|\dfrac{\hat{\theta} - \theta_0}{SE(\hat{\theta})}\right| \leq z_{\frac{\alpha}{2}}$

i.e. the α-level Wald test of $H_0: \theta = \theta_0$ vs $H_A: \theta \neq \theta_0$ fails to reject

Formal definition

Let $\theta \in \Theta$ be a parameter of interest, and $X_1, \dots, X_n$. Let $C(X_1, \dots, X_n) \subseteq \Theta$ be a set constructed from $X_1, \dots, X_n$ ($\Rightarrow C(X_1, \dots, X_n)$ is a random set).

Then $C(X_1, \dots, X_n)$ is a $1-\alpha$ confidence set for θ if

$$\inf_{\theta \in \Theta} P_\theta \left(\theta \in C(X_1, \dots, X_n) \right) = 1-\alpha$$

(i.e., $\forall \theta \in \Theta$, $P_\theta \left(\theta \in C(X_1, \dots, X_n) \right) \geq 1-\alpha$)

Inverting a test

level α test of : $P_{\theta_0}((X_1,\dots,X_n) \in \mathcal{R}(\theta_0)) \le \alpha$

$H_0: \theta = \theta_0$

$H_A: \theta \neq \theta_0$

Theorem: Let $\theta \in \Theta$ be a parameter of interest. For each value of $\theta_0 \in \Theta$, consider testing $H_0 : \theta = \theta_0$ vs. $H_A : \theta \neq \theta_0$, and let $\mathcal{R}(\theta_0)$ be the rejection region for a level α test. Let $C(X_1, \dots, X_n) = \{\theta_0 \in \Theta : (X_1, \dots, X_n) \notin \mathcal{R}(\theta_0)\}$. Then $C(X_1, \dots, X_n)$ is a $1 - \alpha$ confidence set for θ.

Pf:
$$\theta_0 \in C(X_1,\dots,X_n) \iff (X_1,\dots,X_n) \notin \mathcal{R}(\theta_0)$$

$$\begin{aligned} P_{\theta_0}(\theta_0 \in C(X_1,\dots,X_n)) &= P_{\theta_0}((X_1,\dots,X_n) \notin \mathcal{R}(\theta_0)) \\ &= 1 - P_{\theta_0}((X_1,\dots,X_n) \in \mathcal{R}(\theta_0)) \\ &\geq 1 - \alpha \qquad \text{(level } \alpha \text{ test)} \end{aligned}$$

//

Example

Suppose $X_1, ..., X_n \overset{iid}{\sim} Uniform[0, \theta]$. Let's invert the LRT to find a confidence interval for θ.

$H_0: \theta = \theta_0$ vs. $H_A: \theta \neq \theta_0$

LRT: reject H_0 when $\dfrac{\sup_\theta L(\theta|X)}{L(\theta_0|X)} = \dfrac{L(\hat{\theta}|X)}{L(\theta_0|X)} > k$

$L(\theta|X) = \left(\frac{1}{\theta}\right)^n \mathbb{1}\{X_{(n)} \leq \theta\}$ $\qquad \hat{\theta}_{MLE} = X_{(n)}$

$\Rightarrow$ reject H_0 when $\dfrac{\theta_0^n}{X_{(n)}^n \mathbb{1}\{\theta_0 \geq X_{(n)}\}} > k$

reject H_0 when $\theta_0 < X_{(n)}$ or when $\left(\dfrac{\theta_0}{X_{(n)}}\right)^n > k \Rightarrow \dfrac{\theta_0}{X_{(n)}} > k^{\frac{1}{n}}$

Example

Suppose $X_1, \ldots, X_n \overset{iid}{\sim} \mathit{Uniform}[0, \theta]$. Let's invert the LRT to find a confidence interval for θ.

reject H_0 when $\theta_0 < X_{(n)}$ or $\frac{\theta_0}{X_{(n)}} > k^{\frac{1}{n}}$

$\Rightarrow$ fail to reject H_0 when

$$X_{(n)} \leq \theta_0 \leq X_{(n)} k^{\frac{1}{n}}$$

$\Rightarrow$ confidence set for θ

$$[X_{(n)}, X_{(n)} k^{\frac{1}{n}}] = [X_{(n)}, X_{(n)} k']$$

So we need k' s t $P_\theta(\theta \in [X_{(n)}, X_{(n)} k']) \geq 1-\alpha$

$$P_\theta(\theta \in [X_{(n)}, X_{(n)} k']) = P_\theta(X_{(n)} k' \geq \theta)$$

$$= P_\theta\left(X_{(n)} \geq \frac{\theta}{k'}\right)$$

$$= 1 - P_\theta\left(X_{(n)} < \frac{\theta}{k'}\right)$$

$$= 1 - \left(P\left(X_i \leq \frac{\theta}{k'}\right)\right)^n$$

Example

Suppose $X_1, ..., X_n \overset{iid}{\sim} Uniform[0, \theta]$. Let's invert the LRT to find a confidence interval for θ.

$$P_\theta(\theta \in [X_{(n)}, X_{(n)}k']) = 1 - (P(X \le \frac{\theta}{k'}))^n$$

$$P(X \le \frac{\theta}{k'}) = \frac{\theta/k'}{\theta} = \frac{1}{k'}$$

$$= 1 - \left(\frac{1}{k'}\right)^n$$

want : $1 - \left(\frac{1}{k'}\right)^n = 1 - \alpha$

$$\Rightarrow k' = \frac{1}{\alpha^{\frac{1}{n}}}$$

$\Rightarrow$ $1-\alpha$ CI for θ : $\left[X_{(n)}, \frac{X_{(n)}}{\alpha^{1/n}}\right]$