Wald tests

Proof: asymptotic distribution of the MLE

Last time: WTS $\sqrt{n}(\hat{\theta} - \theta) \xrightarrow{d} N(0, \mathcal{I}_1^{-1}(\theta))$

So far:

1) $\sqrt{n}(\hat{\theta} - \theta) \approx \dfrac{-\frac{1}{\sqrt{n}} \ell'(\theta)}{\frac{1}{n} \ell''(\theta)}$

2) $\frac{1}{n} \ell''(\theta) \xrightarrow{P} -\mathcal{I}_1(\theta)$

3) $\mathbb{E}\left[-\frac{1}{\sqrt{n}} \ell'(\theta)\right] = -\sqrt{n}\, \mathbb{E}\left[\frac{\partial}{\partial \theta} \log f(Y_i; \theta)\right]$
$= 0$

4) $\text{var}\left(-\frac{1}{\sqrt{n}} \ell'(\theta)\right) = \text{var}\left(\frac{\partial}{\partial \theta} \log f(Y_i; \theta)\right)$

Claim: $\mathrm{Var}\left(\frac{\partial}{\partial \theta} \log f(Y_i;\theta)\right) = \mathcal{I}_1(\theta)$

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

$$= -\mathbb{E}\left[\frac{\partial}{\partial \theta}\left(\frac{1}{f(Y_i;\theta)} \cdot \frac{\partial}{\partial \theta} f(Y_i;\theta)\right)\right]$$

$$= -\mathbb{E}\left[\frac{1}{f(Y_i;\theta)} \cdot \frac{\partial^2}{\partial \theta^2} f(Y_i;\theta) - \frac{\partial}{\partial \theta} f(Y_i;\theta) \cdot \frac{\frac{\partial}{\partial \theta} f(Y_i;\theta)}{f(Y_i;\theta)^2}\right]$$

$$= -\mathbb{E}\left[\frac{\frac{\partial^2}{\partial \theta^2} f(Y_i;\theta)}{f(Y_i;\theta)} - \left(\frac{\frac{\partial}{\partial \theta} f(Y_i;\theta)}{f(Y_i;\theta)}\right)^2\right]$$

$$-\mathbb{E}\left[\frac{\frac{\partial^2}{\partial \theta^2} f(Y_i;\theta)}{f(Y_i;\theta)}\right] = -\int \frac{\partial^2}{\partial \theta^2} f(y;\theta) \cdot \frac{1}{\cancel{f(y;\theta)}} \cdot \cancel{f(y;\theta)}\, d\theta$$

$$= -\frac{\partial^2}{\partial \theta^2} \int f(y;\theta)\, d\theta \qquad = 0$$

$$\mathbb{E}\left[\left(\frac{\frac{\partial}{\partial\theta} f(Y_i;\theta)}{f(Y_i;\theta)}\right)^2\right] = \mathbb{E}\left[\left(\frac{\partial}{\partial\theta}\log f(Y_i;\theta)\right)^2\right]$$

$$\mathrm{Var}(X) = \mathbb{E}(X^2) - (\mathbb{E}(X))^2$$

$$\Rightarrow \mathbb{E}\left[\left(\frac{\partial}{\partial\theta}\log f(Y_i;\theta)\right)^2\right] = \mathrm{Var}\left(\frac{\partial}{\partial\theta}\log f(Y_i;\theta)\right) + \left(\mathbb{E}\left[\frac{\partial}{\partial\theta}\log f(Y_i;\theta)\right]\right)^2$$

$$= \mathrm{Var}\left(\frac{\partial}{\partial\theta}\log f(Y_i;\theta)\right) \quad //$$

- $\sqrt{n}(\hat{\theta} - \theta) \approx \dfrac{-\frac{1}{\sqrt{n}} \ell'(\theta)}{\frac{1}{n} \ell''(\theta)}$

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

- <u>CLT</u>: $\frac{1}{\sqrt{n}} \ell'(\theta) = \sqrt{n} \left(\frac{1}{n} \sum_{i=1}^{n} \frac{\partial}{\partial \theta} \log f(Y_i; \theta) \right)$

$$\xrightarrow{d} N\left(\mathbb{E}\left(\frac{\partial}{\partial\theta} \log f(Y_i;\theta) \right), \mathrm{Var}\left(\frac{\partial}{\partial\theta} \log f(Y_i;\theta) \right) \right)$$

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

$$\frac{\sqrt{n}\left(\frac{1}{n} \sum_{i=1}^{n} X_i - \mu \right)}{\sigma} \xrightarrow{d} N(0,1)$$

$\mathbb{E}[X_i]$

$$\Rightarrow \sqrt{n}\left(\frac{1}{n} \sum X_i \right) \xrightarrow{d} N(\mu, \sigma^2)$$

$\mathrm{Var}(X_i)$

$$\Rightarrow \frac{1}{n} \sum_i X_i \;\hat{\sim}\; N\left(\mu, \frac{\sigma^2}{n}\right)$$

<u>Slutsky's</u>: $\dfrac{\frac{1}{\sqrt{n}} l'(\theta)}{\frac{1}{n} l''(\theta)} \xrightarrow{d} \dfrac{1}{\mathcal{I}_1(\theta)} \cdot N(0, \mathcal{I}_1(\theta))$

$$= N\left(0, \frac{\mathcal{I}_1(\theta)}{(\mathcal{I}_1(\theta))^2}\right)$$

$$= N\left(0, \frac{1}{\mathcal{I}_1(\theta)}\right)$$

$$\sqrt{n}(\hat{\theta} - \theta) \approx N(0, \mathcal{I}_1^{-1}(\theta))$$

$$\Rightarrow \hat{\theta} \approx N\left(\theta, \frac{1}{n\mathcal{I}_1(\theta)}\right) = N\left(\theta, \frac{1}{\mathcal{I}(\theta)}\right)$$

$$\mathcal{I}(\theta) = -\mathbb{E}\left[\frac{\partial^2}{\partial\theta^2} l(\theta)\right] = -\mathbb{E}\left[\frac{\partial^2}{\partial\theta^2} \sum_{i=1}^{n} \log f(Y_i;\theta)\right]$$

$$= \sum_{i=1}^{n} \underbrace{-\mathbb{E}\left[\frac{\partial^2}{\partial\theta^2} \log f(Y_i;\theta)\right]}_{\mathcal{I}_1(\theta)}$$

$$= n\mathcal{I}_1(\theta)$$

Logistic regression:

$\hat{\beta}$ is the MLE of β

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

$$\mathcal{I}(\beta) = X^T W X$$

design matrix

weight matrix

$$\begin{bmatrix} p_1(1-p_1) & & & 0 \\ & p_2(1-p_2) & & \\ & & \ddots & \\ 0 & & & p_n(1-p_n) \end{bmatrix}$$

Wald tests for single parameters

Logistic regression model for the dengue data:

$$Y_i \sim Bernoulli(p_i)$$

$$\log\left(\frac{p_i}{1 - p_i}\right) = \beta_0 + \beta_1 WBC_i + \beta_2 PLT_i$$

> Researchers want to know if there is a relationship between white blood cell count and the probability a patient has dengue, after accounting for platelet count. What hypotheses should the researchers test?

$H_0: \beta_1 = 0$

$H_A: \beta_1 \neq 0$

Wald tests for single parameters

$H_0: \beta_1 = 0$

$H_A: \beta_1 \neq 0$

```
m1 <- glm(Dengue ~ WBC + PLT, data = dengue,
          family = binomial)
summary(m1)
```

```
...
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)  2.6415063  0.1213233   21.77   <2e-16 ***
## WBC         -0.2892904  0.0134349  -21.53   <2e-16 ***
## PLT         -0.0065615  0.0005932  -11.06   <2e-16 ***
## ---
```

Wald test statistic

$\hookrightarrow Z = \dfrac{\hat{\beta}_1 - \beta_1^0}{SE_{\hat{\beta}_1}} \approx N(0,1)$

m1$weights

$\mathcal{I}^{-1}(\hat{\beta}) = (X^T W X)^{-1} = \begin{bmatrix} 0.1213^2 & & \text{(other stuff)} \\ & 0.0134^2 & \\ \text{(other stuff)} & & 0.00059^2 \end{bmatrix}$

$Z = -21.53 = \dfrac{-0.289}{0.0134}$

Wald tests for multiple parameters

Logistic regression model for the dengue data:

$$Y_i \sim Bernoulli(p_i)$$

$$\log\left(\frac{p_i}{1-p_i}\right) = \beta_0 + \beta_1 WBC_i + \beta_2 PLT_i$$

> Researchers want to know if there is any relationship between white blood cell count or platelet count, and the probability a patient has dengue. What hypotheses should they test?

H_0: $\beta_1 = \beta_2 = 0$

H_A: at least one of $\beta_1, \beta_2 \neq 0$

$H_0: \beta_1 = \beta_2 = 0$

H_A: at least one of $\beta_1, \beta_2 \neq 0$

Wald tests for multiple parameters

```
m1 <- glm(Dengue ~ WBC + PLT, data = dengue,
          family = binomial)
summary(m1)
```

```
...
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)  2.6415063  0.1213233   21.77   <2e-16 ***
## WBC         -0.2892904  0.0134349  -21.53   <2e-16 ***
## PLT         -0.0065615  0.0005932  -11.06   <2e-16 ***
## ---
...
```

> Can the researchers test their hypotheses using this output?

Nope!

Wald tests for multiple parameters

$$\beta = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{bmatrix} = \begin{bmatrix} \beta_{(1)} \\ \beta_{(2)} \end{bmatrix} \qquad \beta_{(1)} = \beta_0 \qquad \beta_{(2)} = \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix} \qquad \hat{\beta} = \begin{bmatrix} \hat{\beta}_{(1)} \\ \hat{\beta}_{(2)} \end{bmatrix}$$

$H_0: \beta_{(2)} = \beta^0_{(2)}$ e.g. $\beta^0_{(2)} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$

H_A: not H_0

$$\hat{\beta} \approx N\left(\begin{bmatrix} \beta_{(1)} \\ \beta_{(2)} \end{bmatrix}, \mathcal{I}^{-1}(\beta) \right) \qquad \mathcal{I}^{-1}(\beta) = \begin{bmatrix} \mathcal{I}^{11} & \mathcal{I}^{12} \\ \mathcal{I}^{21} & \mathcal{I}^{22} \end{bmatrix}$$

$$\hat{\beta}_{(2)} \approx N\left(\beta_{(2)}, \mathcal{I}^{22}\right)$$

e.g. $\mathcal{I}^{11} \in \mathbb{R}^{1\times 1}$

$\mathcal{I}^{12} \in \mathbb{R}^{1\times 2}$

$\mathcal{I}^{21} \in \mathbb{R}^{2\times 1}$

$\mathcal{I}^{22} \in \mathbb{R}^{2\times 2}$

Class activity

https://sta712-f22.github.io/class_activities/ca_lecture_10.html

- Wald tests for the dengue data