Convergence of the MLE, Wald tests

Recap: convergence of the MLE

Under regularity conditions,

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

written another way:

$$\sqrt{n}\, \mathcal{I}_1^{\frac{1}{2}}(\theta)(\hat{\theta}_n - \theta) \xrightarrow{d} N(0, I)$$

identity matrix

$$\mathcal{I}_1 = \mathcal{I}_1^{\frac{1}{2}} \mathcal{I}_1^{\frac{1}{2}}$$

Now

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

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

$$= -\mathbb{E}\left[\frac{\partial^2}{\partial \theta^2} \log f(Y_1, \ldots, Y_n|\theta)\right]$$

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

$$\Rightarrow \mathcal{I}^{\frac{1}{2}}(\theta)(\hat{\theta}_n - \theta) \xrightarrow{d} N(0, I)$$

Regularity conditions

Some sufficient regularity conditions:

- The dimension of θ does not change with n
- $f(y|\theta)$ is a sufficiently smooth function of θ
- we can swap integration and differentiation
- θ is identifiable (basically, $f(y|\theta_1) \neq f(y|\theta_2)$ if $\theta_1 \neq \theta_2$)
- θ is not on the boundary of the parameter space

Counterexample: $X_1, X_2, X_3, \ldots \overset{iid}{\sim} \text{Uniform}(0, \theta)$

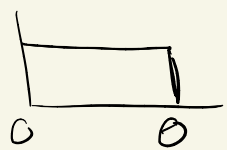
$$\hat{\theta}_n = X_{(n)}$$

Similar to previous lecture, $n(\hat{\theta}_n - \theta) \xrightarrow{d} -\text{Exponential}\left(\frac{1}{\theta}\right)$

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

<u>Regularity conditions violated</u>:

- $f(x|\theta)$ is not smooth at θ, so $\mathcal{I}_1(\theta)$ is not even defined



- can't exchange derivatives & integrals (domain of integration depends on θ)

Counterexample:

Suppose we measure some outcome for n individuals, and we take two measurements for each individual:

$$Y_{11}, Y_{12} \sim N(\mu_1, \sigma^2)$$

$$Y_{21}, Y_{22} \sim N(\mu_2, \sigma^2)$$

$$\vdots$$

$$Y_{n1}, Y_{n2} \sim N(\mu_n, \sigma^2)$$

We want to estimate σ^2. Some math gives

$$\hat{\sigma}^2 = \frac{1}{4n} \sum_{i=1}^{n} (Y_{i1} - Y_{i2})^2 \xrightarrow{P} \frac{\sigma^2}{2}, \quad \text{\underline{not} to } \sigma^2$$

<u>Regularity condition violated</u>: # parameters increases with n

$$\mathcal{I}^{\frac{1}{2}}(\theta)(\hat{\theta}_n - \theta) \xrightarrow{d} N(0, I)$$

Application to logistic regression

$Y_i \sim \text{Bernoulli}(p_i)$

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

$\mathcal{I}(\beta) = X^T W X$ $W = \text{diag}(p_i(1-p_i))$

$$(X^T W X)^{\frac{1}{2}}(\hat{\beta} - \beta) \xrightarrow{d} N(0, I)$$

we don't actually know β, so we don't know true W

$\Rightarrow$ use $\hat{W} = \text{diag}(\hat{p}_i(1-\hat{p}_i))$

$$(X^T \hat{W} X)^{\frac{1}{2}}(\hat{\beta} - \beta) \xrightarrow{d} N(0, I)$$ (basically, Slutsky's)

$\Rightarrow$ $\hat{\beta} \approx N(\beta, (X^T \hat{W} X)^{-1})$

$\Rightarrow$ $\hat{\beta}_j \approx N(\beta_j, \text{Var}(\hat{\beta}_j))$

$$(X^T \hat{W} X)^{-1} = \begin{bmatrix} \text{var}(\hat{\beta}_0) & & \text{(other stuff)} \\ & \text{var}(\hat{\beta}_1) & \\ \text{(other stuff)} & & \ddots \end{bmatrix}$$

Wald tests for single parameters

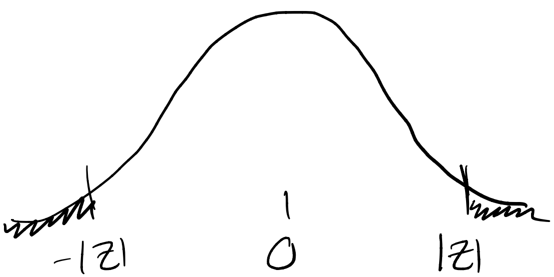
Logistic regression model for the dengue data:

$$Y_i \sim Bernoulli(p_i)$$

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

$H_0 : \beta_1 = 0 \qquad H_A : \beta_1 \neq 0$

$$Z = \frac{\hat{\beta}_1 - 0}{SE(\hat{\beta}_1)} \approx N(0,1)$$



p-value = "probability of our data or more extreme"

$= P(|Z| > |z|) \qquad Z \sim N(0,1)$

Class activity

https://sta711-s23.github.io/class_activities/ca_lecture_16.html

Class activity

```
m1 <- glm(Dengue ~ WBC + PLT, data = dengue, family = binomial)
X <- model.matrix(m1)
solve(t(X) %*% diag(m1$weights) %*% X)
```

$\leftarrow \widehat{Var}(\hat{\beta}) \approx (X^T W X)^{-1}$

```
##               (Intercept)           WBC           PLT
## (Intercept)  1.471934e-02 -4.937020e-04 -5.125888e-05
## WBC         -4.937020e-04  1.804972e-04 -3.221337e-06
## PLT         -5.125888e-05 -3.221337e-06  3.518938e-07
```

same result!

```
vcov(m1)
```

```
##               (Intercept)           WBC           PLT
## (Intercept)  1.471934e-02 -4.937020e-04 -5.125888e-05
## WBC         -4.937020e-04  1.804972e-04 -3.221337e-06
## PLT         -5.125888e-05 -3.221337e-06  3.518938e-07
```

```
summary(m1)$coefficients[,2]^2
```

```
##  (Intercept)          WBC          PLT
## 1.471934e-02 1.804972e-04 3.518938e-07
```

Class activity