Lecture 14

Diagnostics and checking assumptions

1. Empirical log means / empirical logit / scatterplots. Apply any transformations needed
2. Fit the model
3. (Quantile) residual plots. Apply any transformations needed
4. VIFs and Cook's distance. Consider removing/modifying variables or observations
5. GOF tests (if applicable)

What if assumptions are violated?

- Fix any shape violations first
- Then address distributional violations
  - Robust variance estimates
  - Use a different distribution/model for the response

NB regression

Quasi-Poisson models

hurdle models + zero-inflated models

time permitting: methods for correlated data

Recap: maximum likelihood asymptotics

Suppose $Y_1, \ldots, Y_n \sim F_\theta$ w/ probability function $f(y;\theta)$

log likelihood: $\ell(\theta) = \sum_{i=1}^{n} \log f(Y_i;\theta)$

MLE: $\hat{\theta}$ maximizes $\ell(\theta)$

Let score $U(\theta) = \frac{\partial \ell}{\partial \theta} = \ell'(\theta)$

$\hat{\theta}$ solves $U(\theta) = 0$

Properties: <u>If</u> model is correct: (w/ regularity conditions)

$\hat{\theta} \xrightarrow{P} \theta$ (true parameter)

$E_\theta[U(\theta)] = 0$ $\quad$ $\mathrm{Var}_\theta(U(\theta)) = \mathcal{I}(\theta) = -E_\theta[\ell''(\theta)]$

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

$\mathcal{I}_1(\theta) =$ Fisher info. for $n=1$

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

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

Maximum likelihood with mis-specified models

$Y_1, \dots, Y_n \sim G$ ← no θ, b/c we have the wrong model

w/ probability function g

Assume (incorrectly!) that $Y_i \sim F_\theta$, and we estimate θ

Still write down $\ell(\theta) = \sum_{i=1}^{n} \log f(Y_i; \theta)$

Estimate: $\hat{\theta}$ solves $U(\theta) = 0$
↖ $\ell'(\theta)$

<u>Q</u>: what is $\hat{\theta}$ actually estimating??

Let θ^* be value of θ that solves $\mathbb{E}_g[U(\theta)] = 0$ ← expectation wrt <u>true</u> distribution

<u>Now</u>: θ^* is <u>not</u> true parameter; rather, θ^* is the parameter which <u>best</u> approximates true model (g) in space of models considered $(f(\cdot;\theta))$

As $n \to \infty$ $\hat{\theta} \xrightarrow{P} \theta^*$

$\hat{\theta} \xrightarrow{p} \theta^*$

$u(\theta)$

Let $\quad V_n(\theta) = \mathrm{Var}_g(\ell'(\theta))$ no longer equal (unless model is correctly specified)

$\quad J_n(\theta) = -\mathbb{E}_g[\ell''(\theta)]$

Asymptotics: $\quad -\frac{1}{n}\ell''(\theta) \xrightarrow{p} J_1(\theta^*)$

$$\frac{1}{\sqrt{n}}\ell'(\theta) \xrightarrow{d} N(0, V_1(\theta^*))$$

$$\sqrt{n}(\hat{\theta} - \theta^*) \xrightarrow{d} N(0, \underbrace{J_1(\theta^*)^{-1} V_1(\theta^*) J_1(\theta^*)^{-1}}_{\text{sandwich variance}})$$

$$\hat{\theta} \approx N(\theta^*, \hat{J}_n(\hat{\theta})^{-1} \hat{V}_n(\hat{\theta}) \hat{J}_n(\hat{\theta})^{-1})$$

Poisson regression: $\quad u(\beta) = X^T(Y-\mu) = \sum_{i=1}^{n}(Y_i - \mu_i)X_i$

μ_i: $\exp\{\beta^T X_i\}$

$$\hat{J}_n(\hat{\beta}) = \sum_{i=1}^{n} \hat{\mu}_i X_i X_i^T = X^T \mathrm{diag}(\hat{\mu}_i) X$$

$$\hat{V}_n(\hat{\beta}) = \sum_{i=1}^{n} (Y_i - \hat{\mu}_i)^2 X_i X_i^T = X^T \mathrm{diag}((Y_i - \hat{\mu}_i)^2) X$$

Class activity

https://sta712-f23.github.io/class_activities/ca_lecture_14.html