

Robust Sandwich Variance Estimation

Define the fisher information \mathcal{A} as the per-observation information

$$\mathcal{A} = \mathcal{I}_1(\theta) = -\mathbb{E} \left[\frac{\partial^2}{\partial \theta \partial \theta^\top} \ell_i(\theta) \right]$$

and variance of score function, \mathcal{B} , as

$$\mathcal{B} = \text{Var}(S_i(\theta)) = \mathbb{E}[S_i(\theta)S_i(\theta)^\top]$$

By asymptotic distribution, we have

$$\sqrt{n}(\hat{\theta} - \theta_0) \xrightarrow{d} \mathcal{N}(0, \mathcal{A}^{-1}\mathcal{B}\mathcal{A}^{-1})$$

Equivalently

$$\Rightarrow \text{Var}(\hat{\theta}) = \frac{1}{n} \mathcal{A}^{-1} \mathcal{B} \mathcal{A}^{-1}$$

If $\mathcal{A}^{-1} = \mathcal{B}$, then

$$\begin{aligned} \Rightarrow \text{Var}(\hat{\theta}) &= \frac{1}{n} \mathcal{A}^{-1} \mathcal{B} \mathcal{A}^{-1} \\ &= \frac{1}{n} \mathcal{I}_1^{-1}(\theta) \\ &= (n \mathcal{I}_1(\theta))^{-1} \\ &= \mathcal{I}^{-1}(\theta) \end{aligned}$$

Empirically

$$\hat{\mathcal{A}} = \frac{1}{n} \sum_{i=1}^n \frac{\partial}{\partial \theta} \ell_i(\theta), \quad \hat{\mathcal{A}} \xrightarrow{P} \mathcal{A}$$

$$\hat{\mathcal{B}} = \frac{1}{n} \sum_{i=1}^n S_i(\theta) S_i(\theta)^\top, \quad \hat{\mathcal{B}} \xrightarrow{P} \mathcal{B}$$

$$\Rightarrow \widehat{\text{Var}}(\hat{\theta}) = \frac{1}{n} \hat{\mathcal{A}}^{-1} \hat{\mathcal{B}} \hat{\mathcal{A}}^{-1}$$

MSL0M

$$\log \mathbf{O}_i = \beta_0 + \log z_i$$

Log-likelihood for individual i :

$$\ell_i(\beta) = w_i \{y_i \log p_i + (1 - y_i) \log(1 - p_i)\}$$

The Score function can be written as

$$S_i(\beta) = w_i r_i g_i = w_i r_i \begin{pmatrix} 1 \\ G_i/z_i \\ E_i/z_i \\ G_i E_i/z_i \end{pmatrix}$$

with expected information:

$$I_1(\beta) = w_i p_i (1 - p_i) g_i g_i^\top$$

$$\begin{aligned}
\mathcal{A} &= \mathbb{E}_{\mathcal{Y}}[I_1(\beta)] \\
&= \mathbb{E}[w_i p_i (1 - p_i) g_i g_i^\top] \\
&= \mathcal{I}_1(\theta)
\end{aligned}$$

$$\hat{\mathcal{A}} = \frac{1}{n} \sum w_i \hat{p}_i (1 - \hat{p}_i) g_i g_i^\top$$

The variance of score function is

$$\begin{aligned}
\mathcal{B} &= \mathbb{E}_{\mathcal{Y}}[S(\beta_0) S(\beta_0)^\top] \\
&= \mathbb{E}[w_i^2 (y_i - p_i)^2 g_i g_i^\top] \\
&= \mathbb{E}[\mathbb{E}[w_i^2 (y_i - p_i)^2 g_i g_i^\top \mid \sigma(G_i, E_i, C_i)]] \\
&= \mathbb{E}[w_i^2 g_i g_i^\top p_i (1 - p_i)]
\end{aligned}$$

with

$$\hat{\mathcal{B}} = \frac{1}{n} \sum w_i^2 r_i^2 g_i g_i^\top$$

Therefore the robust sandwich variance estimator is

$$\Rightarrow \widehat{\text{Var}}(\hat{\theta}) = \frac{1}{n} \left(\frac{1}{n} \sum w_i \hat{p}_i (1 - \hat{p}_i) g_i g_i^\top \right)^{-1} \left(\frac{1}{n} \sum w_i^2 r_i^2 g_i g_i^\top \right) \left(\frac{1}{n} \sum w_i \hat{p}_i (1 - \hat{p}_i) g_i g_i^\top \right)^{-1}$$