Model Selection & Hypothesis Testing

Recall we want to test $H_0: \theta \in \Theta_0$ $H_1: \theta \in \Theta_1$

and want to infer $I(\theta \in \Theta_0)$

We can define a decision $\phi(y) = \underset{\phi \in [0,1]}{\text{argmin}}\ \mathbb{E}_{\theta|y}[L(\theta, \phi)]$

We considered the loss type I: γ type II: 1 and found

$$\phi(y) = I\left[\frac{\mathbb{P}(\theta \in \Theta_1 | Y)}{\mathbb{P}(\theta \in \Theta_0 | Y)} \leq \gamma\right] = I\left[\mathbb{P}(\theta \in \Theta_0 | Y) \geq \frac{1}{1+\gamma}\right]$$

If we have a point hypothesis s.t. $\mathbb{P}(\theta \in \Theta_0) = 0$ $\mathbb{P}(\theta \in \Theta_1) = 0$ and θ is cont. then

this setup can breakdown.

One way to accomondate simple nulls then use a hierachial model to test

$\mathbb{P}(H_0) = \mathbb{P}(\theta \in \Theta_0)$ and $\mathbb{P}(H_1) = 1 - \mathbb{P}(H_0)$ and set up the prior

$$\mathbb{P}(\theta) = \mathbb{P}(H_0)\mathbb{P}(\theta | \theta \in \Theta_0) + \mathbb{P}(H_1)\mathbb{P}(\theta | \theta \in H_1)$$

This way, $\dfrac{\mathbb{P}(\theta \in \Theta_1 | Y)}{\mathbb{P}(\theta \in \Theta_0 | Y)} = \dfrac{\mathbb{P}(\theta \in \Theta_1, Y)}{\mathbb{P}(\theta \in \Theta_0, Y)} = \dfrac{\int \mathbb{P}(Y|\theta)\mathbb{P}(\theta) I(\theta \in \Theta_1)\, d\theta}{\int \mathbb{P}(Y|\theta)\mathbb{P}(\theta) I(\theta \in \Theta_0)\, d\theta}$

$$= \frac{\int \mathbb{P}(Y|\theta)\mathbb{P}(\theta | \theta \in \Theta_1) I(\theta \in \Theta_1)\, d\theta \ \ \mathbb{P}(H_1)}{\int \mathbb{P}(Y|\theta)\mathbb{P}(\theta | \theta \in \Theta_0) I(\theta \in \Theta_0)\, d\theta \ \ \mathbb{P}(H_0)} = \underbrace{\frac{\mathbb{P}(Y | \theta \in \Theta_1)}{\mathbb{P}(Y | \theta \in \Theta_0)}}_{\substack{\text{Bayes Factor} \\ BF_{10}}} \times \frac{\mathbb{P}(H_1)}{\mathbb{P}(H_0)}$$

one to zero

$$BF_{10}(y) = \frac{\dfrac{\mathbb{P}(\theta \in \Theta_1 \mid Y)}{\mathbb{P}(\theta \in \Theta_0 \mid Y)}}{\dfrac{\mathbb{P}(\theta \in \Theta_1)}{\mathbb{P}(\theta \in \Theta_0)}}$$

Post.–Prior Odds Ratio.

A useful table by Kass/Raftery

BF	Evidence
1–3	Barely mention
3–20	positive
20–150	strong
>150	v. strong

Variable Selection

Suppose we regress $Y \sim X$, where $X \in \mathbb{R}^{n \times p}$. Want to compare all possible models from including or excluding predictors. 2^p models

We indicate which model we are considering with $\theta \in \{0,1\}^p$

Want to make inferences on $\hat{\theta} = \underset{\phi \in \{0,1\}^p}{\arg\min}\ \mathbb{E}_{\theta \mid Y}[L(\theta, \phi)]$

two common choices:

(i) $L_m(\theta, \phi) = I(\theta = \phi) = 1 - \prod_{j=1}^{p} I(\theta_j = \phi_j)$

$$\phi_m(y) = \underset{\phi \in \{0,1\}^p}{\arg\min}\ \mathbb{E}_{\theta \mid y}[I(\theta \neq \phi)] = \underset{\phi}{\arg\min}\left[1 - \mathbb{P}(\theta = \phi \mid Y)\right]$$

$$= \underset{\phi}{\arg\max}\ \mathbb{P}(\theta = \phi \mid Y) = \hat{\theta}_{map}$$

(ii) $L_H(\theta, \phi) = \sum_{j=1}^{p} \underbrace{L(\theta_j, \phi_j)}_{} = \sum_{j=1}^{p}\left[\gamma(1-\phi_j)\theta_j + \phi_j(1-\theta_j)\right]$

L:

0	γ
1	0

$$\phi_H(y) = \underset{\phi}{\arg\min}\ \mathbb{E}_{\theta|Y}\left[\sum_{j=1}^{p} L(\theta_j,\phi_j)\right] = \underset{\phi}{\arg\min} \sum_{j=1}^{p} \mathbb{E}_{\theta_j|Y}\left[L(\theta_j,\phi_j)\right]$$

$$\Rightarrow \left(\phi_H(y)\right)_{j=1}^{p} = \underset{\phi_j}{\arg\min}\ \mathbb{E}_{\theta_j|Y}\left[L(\theta_j,\phi_j)\right] = I\Big[\underbrace{\mathbb{E}[\theta_j|Y]}_{\mathbb{P}(\theta_j=1|Y)} \geq \frac{1}{1+\gamma}\Big]$$

post. prediction.

"posterior prob. of inclusion"

<u>Rmk:</u> Hard to compute.

<u>Ex</u>: For simplicity assume $Y|\beta,\sigma^2,\theta \sim N(\underbrace{X_\theta}_{X[,\theta=1]} \underbrace{\beta_\theta}_{\beta[\theta=1]}, \sigma^2 I_n)$

$$\beta_\theta|\sigma^2 \sim N(\beta_{0,\theta}, \sigma^2\Sigma_\theta) \qquad \beta_{0,\theta}=0, \quad \Sigma_\theta = g(X_\theta^T X_\theta)^{-1}$$

— typically $g=n$.

$$\sigma^2 \sim \text{Inv}\,\chi^2(\nu,\tau^2)$$

For prior on θ by $\theta_j \overset{iid}{\sim} \text{Bern}(\alpha)$

From the model

$$Y|\sigma^2,\theta \sim N\left(X_\theta\beta_{0,\theta}, \sigma^2(I_n + X_\theta\Sigma_\theta X_\theta^T)\right)$$

$$Y|\theta \sim t_\nu\left(X\beta_{0,\theta}, \tau^2(I_n + X_\theta\Sigma_\theta X_\theta^T)\right)$$

$$\mathbb{P}(\theta|Y) = \frac{\mathbb{P}(Y|\theta)\mathbb{P}(\theta)}{\sum_{\vec{\theta}\in\{0,1\}^p} \mathbb{P}(Y|\vec{\theta})\mathbb{P}(\vec{\theta})}$$