ST 790: Bayesian Biostatistics
Project Description

Spring, 2010

Project report Due: March 10, 2010

40% of your course grade will depend upon successful, on time completion of the midterm project. Project reports are to result in thorough but concise, professional quality technical reports of not more than 10 double spaced pages (not including the raw data, computer outputs and `R` or `WinBUGS` codes). Projects are to be turned in by the deadline stated above.

Sign the **Honor Pledge**: *I have neither given nor received unauthorized aid on this assignment* on your cover page, before you submit the project.

Suppose we observe data $X = (x_1, x_2, \ldots, x_n) \sim f(X|\theta)$ where $f(X|\theta)$ denotes the conditional density of data X given the parameter $\theta \in \Theta$. Let $\theta \sim \pi(\cdot)$, where $\pi(\cdot)$ is a (prior) probability density defined on the parameter space Θ.

Consider the problem of comparing two competing hypotheses:

$$H_0 : \theta \in \Theta_0 \;\; vs. \;\; H_1 : \theta \in \Theta_1,$$

where $\Theta_0 \cap \Theta_1 = \emptyset$ and $\Theta_0 \cup \Theta_1 \subseteq \Theta$. Assume that $\Pr[\theta \in \Theta_j] = \int_{\Theta_j} \pi(\theta) d\theta > 0$ for $j = 0, 1$. Define the Bayes factor (BF) as:

$$BF(X) = \frac{\Pr[\theta \in \Theta_1|X]}{\Pr[\theta \in \Theta_0|X]} \cdot \frac{\Pr[\theta \in \Theta_0]}{\Pr[\theta \in \Theta_1]}.$$

Let the "test statistic" be $T(X) = \log(BF(X))$ and we would decide to reject the null hypothesis H_0 (in favor of the alternative hypothesis H_1) if $T(X) > t_0$ for some $t_0 \geq 0$. To determine the cutoff value t_0 we use the following criteria:

Let the "*average (Bayes) type I error rate*" be:

$$AE_1(t_0) = \Pr[T(X) > t_0|\theta \in \Theta_0] = \frac{\Pr[T(X) > t_0, \theta \in \Theta_0]}{\Pr[\theta \in \Theta_0]},$$

where the numerator is computed using the joint distribution of (X, θ) while the denominator is computed using the prior distribution. Similarly let the "*average (Bayes) type II error rate*" be:

$$AE_2(t_0) = \Pr[T(X) \leq t_0 | \theta \in \Theta_1] = \frac{\Pr[T(X) \leq t_0, \theta \in \Theta_1]}{\Pr[\theta \in \Theta_1]}.$$

Next, we define the "total weighted error rate" as:

$$TWE(t_0) = wAE_1(t_0) + (1 - w)AE_2(t_0) \quad \text{for a given } \; w \in [0, 1].$$

Finally, we obtain the cutoff value as:

$$\hat{t}_0 = \arg\min\{TWE(t_0) : t_0 \in [0, 100]\}.$$

Hence the above (Bayes) test will reject H_0 if $T(X) > \hat{t}_0$. Notice that $\hat{t}_0$ as defined above depends only on the sample n and the criteria weight w given the model with sampling density $f(X|\theta)$ and prior density $\pi(\theta)$.

Sample size determination rule: Given the weights $w \in [0, 1]$ and $\alpha \in (0, 1)$, find the (minimum) sample size $n > 1$, such that $TE(\hat{t}_0(n)) \leq \alpha$, where $TE(t_0) = AE_1(t_0) + AE_2(t_0)$ denotes the "*total (Bayes) error rate*" of the hypotheses problem. In other words, the optimal sample size is given by

$$\hat{n} = \arg\min\{TE(\hat{t}_0(n) : n = 2, 3, \ldots\}$$

---

1. Let $X|\theta \sim Bin(\theta, n)$ where $\theta \in \Theta = [0, 1]$. Let $\Theta_0 = \{\theta_0\}$ and $\Theta_1 = \{\theta \in [0, 1] : \theta \neq \theta_0\}$. Suppose $\pi(\theta) = 0.5I(\theta = \theta_0) + 0.5\pi_1(\theta)I(\theta \neq \theta_0)$, where $\pi_1(\theta)$ represents a $Beta(a, b)$ density. Obtain $\hat{n}$ as defined above for the following cases:

   (a) $\theta_0 = 0.25, a = b = 1, w = 0.90$ and $\alpha = 0.25$

   (b) $\theta_0 = 0.75, a = b = 1, w = 0.95$ and $\alpha = 0.25$

   (c) $\theta_0 = 0.5, a = b = 1, w = 0.5$ and $\alpha = 0.25$

   (d) $\theta_0 = 0.5, a = b = 1, w = 0.05$ and $\alpha = 0.25$

   (e) $\theta_0 = 0.5, a = b = 1, w = 0.10$ and $\alpha = 0.25$

   Given $\theta_0 = 0.5, a = b = 1, w = 0.90$, for a selected grid of values of $\alpha \in (0.15, 0.35)$, obtain $\hat{n}(\alpha)$ and plot $\log(\hat{n}(\alpha))$ vs. α.

2. Let $x_j|\theta_j \overset{ind.}{\sim} Bin(\theta_j, n)$ for $j = 1, 2$ where $\theta = (\theta_1, \theta_2) \in [0, 1]^2$. Let $\Theta_0 = \{\theta \in [0, 1]^2 : \theta_1 = \theta_2\}$ and $\Theta_1 = \{\theta \in [0, 1]^2 : \theta_1 \neq \theta_2\}$. Suppose $\pi(\theta) = 0.5\pi_0(\eta)I(\theta_1 = \theta_2 = \eta) + 0.5\pi_1(\theta_1)\pi_1(\theta_2)I(\theta_1 \neq \theta_2)$, where each $\pi_j(\cdot)$ represents a $Beta(a, b)$ density for $j = 0, 1$. Obtain $\hat{n}$ as defined above for the following cases:

   (a) $a = b = 1, w = 0.90$ and $\alpha = 0.25$

   (b) $a = b = 1, w = 0.95$ and $\alpha = 0.25$

   (c) $a = b = 1, w = 0.5$ and $\alpha = 0.25$

   (d) $a = b = 1, w = 0.10$ and $\alpha = 0.25$

   Given $a = b = 1, w = 0.90$, for a selected grid of values of $\alpha \in (0.15, 0.35)$, obtain $\hat{n}(\alpha)$ and plot $\log(\hat{n}(\alpha))$ vs. α.

   Given $a = b = 1, w = 0.10$, for a selected grid of values of $\alpha \in (0.15, 0.35)$, obtain $\hat{n}(\alpha)$ and plot $\log(\hat{n}(\alpha))$ vs. α.

3. Let $x_j|\theta_j \overset{ind.}{\sim} Bin(\theta_j, n)$ for $j = 1, 2$ where $\theta = (\theta_1, \theta_2) \in [0, 1]^2$. Let $\Theta_0 = \{\theta \in [0, 1]^2 : \theta_1 \leq \theta_2\}$ and $\Theta_1 = \{\theta \in [0, 1]^2 : \theta_1 > \theta_2\}$. Suppose $\pi(\theta) = \pi_0(\theta_1)\pi_0(\theta_2)$, where $\pi_0(\cdot)$ represents a $Beta(a, b)$ density. Obtain $\hat{n}$ as defined above for the following cases:

   (a) $a = b = 1, w = 0.9$ and $\alpha = 0.30$

   (b) $a = b = 1, w = 0.95$ and $\alpha = 0.30$

   (c) $a = b = 1, w = 0.10$ and $\alpha = 0.30$

   (d) $a = b = 1, w = 0.05$ and $\alpha = 0.30$

   (e) $a = b = 1, w = 0.50$ and $\alpha = 0.30$

   Given $a = b = 1, w = 0.90$, for a selected grid of values of $\alpha \in (0.15, 0.35)$, obtain $\hat{n}(\alpha)$ and plot $\log(\hat{n}(\alpha))$ vs. α.

   Given $a = b = 1, w = 0.5$, for a selected grid of values of $\alpha \in (0.15, 0.35)$, obtain $\hat{n}(\alpha)$ and plot $\log(\hat{n}(\alpha))$ vs. α.