

STA 360: Homework 5

Samuel Eure

10/9/2018

Problem 5.2

Find and plot $P(\theta_A < \theta_B \mid \mathbf{y}_A, \mathbf{y}_B)$

This can be found Monte Carlo by obtaining a lot of iid posterior samples of $(\theta_A \mid \mathbf{y}_A)$ and $(\theta_B \mid \mathbf{y}_B)$ and taking the average of $(\theta_A \mid \mathbf{y}_A) < (\theta_B \mid \mathbf{y}_B)$

In this situation, the variance of θ_a or θ_B , (and the exam scores) is not known, so we must account for this in our approximation by assigning a prior and posterior belief to σ^2 (the variance of exam scores).

If $\nu_0 = k_0$ is the weight we wish to apply to our prior belief (i.e. the hypothetical previous number of sample), and our prior belief of what σ^2 should be is equal to σ_0^2 , then

$$\sigma^2 \sim \text{Inv-Gamma}\left(\frac{\nu_0}{2}, \frac{\nu_0 \sigma_0^2}{2}\right)$$

Since θ_a and θ_B are being modeled as the center of the distribution of exam scores, let the prior for θ_i be

$$\theta_i \mid \sigma^2 \sim \text{Normal}\left(\mu_0, \frac{\sigma^2}{k_0}\right)$$

Thus, the posterior distributions are

$$(\sigma^2 \mid y_1, \dots, y_n) \sim \text{Inv-Gamma}\left(\frac{\nu_0 + n}{2}, \frac{(\nu_0 + n)\sigma_n^2}{2}\right)$$

With

$$\sigma_n^2 = \frac{1}{\nu_0 + n} \left[\nu_0 \sigma_0^2 + (n-1) \sum_{i=1}^n \frac{(y_i - \bar{y})^2}{n-1} + \frac{k_0 n}{k_0 + n} (\bar{y} - \mu_0)^2 \right]$$

and a posterior of θ_i of

$$(\theta_i \mid \sigma^2, y_1, \dots, y_n) \sim \text{Normal}\left(\frac{k_0 \mu_0 + \bar{y} n}{k_0 + n}, \frac{\sigma^2}{k_0 + n}\right)$$

Thus

```

nA = 16; nB = 16;
muOA = 75; sigmaOA = 10;
muOB = 75; sigmaOB = 10;
yA = 75.2; sA = 7.3;
yB = 77.5; sB = 8.1;
priorWeights = 2^c(0:14)

samples = 100000;

results = c();

for (p in 1:length(priorWeights)){
  k = priorWeights[p];

  postGuessVarA = (1/(k+nA))*(k*sigmaOA^2 + (nA-1)*sA^2 + ((k*nA)/(k+nA))*(yA - muOA )^2);
  postVarianceA = 1/rgamma(samples, (k+nA)/2,((k+nA)*postGuessVarA)/2);
  postThetaA = rnorm(length(postVarianceA), (k*muOA+nA*yA)/(nA+k), sqrt(postVarianceA/(k+nA)));

  postGuessVarB = (1/(k+nB))*(k*sigmaOB^2 + (nB-1)*sB^2 + ((k*nB)/(k+nB))*(yB - muOB )^2);
  postVarianceB = 1/rgamma(samples, (k+nB)/2,((k+nB)*postGuessVarB)/2);
  postThetaB = rnorm(length(postVarianceB), (k*muOB+nB*yB)/(nB+k), sqrt(postVarianceB/(k+nB)));
  results = c(results, mean(postThetaA < postThetaB));
}

print("probabilities given the prior weight is 1, 2, 4, ..., 1024")

## [1] "probabilities given the prior weight is 1, 2, 4, ..., 1024"

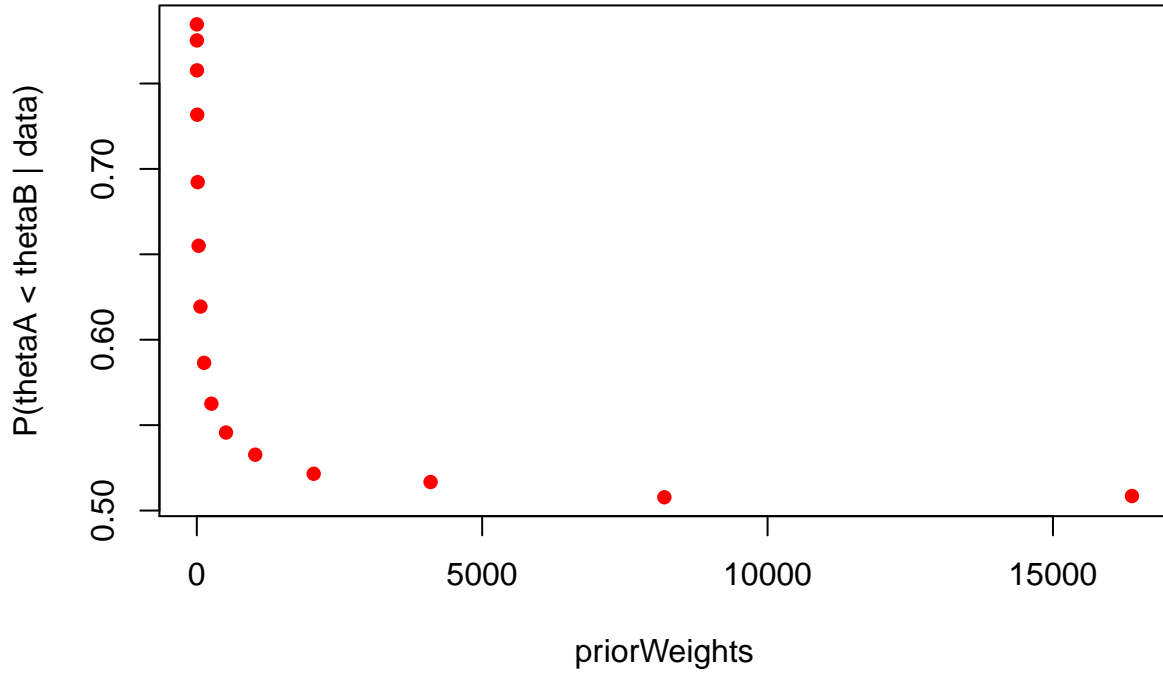
print(results)

## [1] 0.78463 0.77522 0.75771 0.73175 0.69225 0.65501 0.61940 0.58647
## [9] 0.56257 0.54568 0.53270 0.52153 0.51670 0.50779 0.50850

plot(priorWeights, results, col = 'red', pch = 16,
     main = "Prior Weight vs P( thetaA < thetaB | data )", ylab = "P(thetaA < thetaB | data)")

```

Prior Weight vs P(thetaA < thetaB | data)



As shown in the graph above, statisticians who held a weak or moderate prior belief that $\theta_A = \theta_B$ should be more confident that $\theta_A < \theta_B \mid data$ than those with a strong prior belief that $\theta_A = \theta_B$ since $P(\theta_A < \theta_B \mid data)$ is around 0.6 to 0.8, as shown in the plot above. However, for someone who was fairly confident that $\theta_A = \theta_B$, their posterior belief of $P(\theta_A < \theta_B \mid data)$ should be around 0.5 since the prior weight should flood out the sampling data.

Midterm Practice Problems

Practice Problem 1

Let $(Y_1, \dots, Y_n \mid \theta) \sim \text{Poisson}(\theta)$.

(1) Find the conjugate family of priors.

$$p(\theta \mid Y_1, \dots, Y_n) \propto p(\theta) \prod_{i=1}^n \frac{\theta^{y_i} e^{-\theta}}{y_i!} \propto p(\theta) \theta^{n\bar{y}} e^{-n\theta} \Rightarrow p(\theta) \propto \theta^{c_1} e^{-c_2\theta}$$

for constants c_1, c_2 , which makes $p(\theta)$ have a Gamma distribution, so the conjugate family of priors is Gamma.

(2) Find the posterior distribution

Let $\theta \sim \text{Gamma}(a, b)$

$$p(\theta \mid Y_1, \dots, Y_n) \propto (\theta^{a-1} e^{-\theta b}) \theta^{n\bar{y}} e^{-n\theta} = \theta^{a+n\bar{y}-1} e^{-\theta(n+b)}$$

$$\Rightarrow (\theta \mid y_1, \dots, y_n) \sim \text{Gamma}(a + n\bar{y}, n + b)$$

(3) Write down the posterior mean as a weighted average

$$\mathbb{E}(\theta \mid y_1, \dots, y_n) = \frac{a + n\bar{y}}{n + b} = \left(\frac{a}{b}\right)\left(\frac{b}{n + b}\right) + \left(\bar{y}\right)\left(\frac{n}{n + b}\right)$$

Practice Problem 2

Let $(Y_1, \dots, Y_n \mid \alpha, \beta) \sim^{iid} \text{Gamma}(\alpha, \beta)$, with α known.

(1) Find the conjugate family of priors for β .

$$p(\beta \mid y_1, \dots, y_n, \alpha) \propto p(\beta) \prod_{i=1}^n \beta^\alpha e^{-\beta y_i} \propto p(\beta) \beta^{n\alpha} e^{-\beta n\bar{y}} \propto \beta^{c_1} e^{-\beta c_2}$$

for constants c_1, c_2 . Thus, the posterior distribution of β has the shape of a Gamma distribution, so the family of conjugate priors is Gamma.

(2) Find the posterior

Let $\beta \sim \text{Gamma}(a, b)$. Then,

$$p(\beta \mid y_1, \dots, y_n, \alpha, a, b) \propto \left(\beta^{a-1} e^{-\beta b}\right) \beta^{n\alpha} e^{-\beta n\bar{y}} = \beta^{a+n\alpha-1} e^{-\beta(b+n\bar{y})} \Rightarrow (\beta \mid \alpha, y_1, \dots, y_n, a, b) \sim \text{Gamma}(a+n\alpha, b+n\bar{y})$$

(3) Give interpretations of the prior hyperparameters

Since $n\bar{y}$ is equal to the sum of the values of the observed data, $b + n\bar{y}$ can be interpreted as the sum of all data observed, for example, it could represent the total counts based on prior count data and current count data, or it could represent the total waiting time at this up to this point. Relatedly, a can be thought of as proportional to the prior sample size, allowing $\frac{a}{b}$, the prior mean, to represent a prior belief of β , which is the rate parameter, allowing the inverse of the prior mean to represent the prior mean of the value of Y .

Practice Problem 3

Let $(Y_1, \dots, Y_n \mid \theta) \sim^{iid} \text{Bernoulli}(\theta)$

(1) Find the conjugate family of priors.

$$p(\theta \mid y_1, \dots, y_n) \propto p(\theta) p(y_1, \dots, y_n \mid \theta) \propto p(\theta) \prod_{i=1}^n \theta^{y_i} (1 - \theta)^{1-y_i} = p(\theta) \theta^{n\bar{y}} (1 - \theta)^{n(1-\bar{y})}$$

Which implies the prior should include terms such as $\theta^{c_1}(1-\theta)^{c_2}$, which looks like a Beta distribution. Thus, the conjugate prior family for θ should be Beta.

(2) Find the posterior distribution.

Let $\theta \sim \text{Beta}(a, b)$ with $a, b \in \mathbb{Z}$ and positive.

$$\begin{aligned} p(\theta \mid y_1, \dots, y_n) &= p(\theta) \left(\theta^{n\bar{y}} (1-\theta)^{n(1-\bar{y})} \right) \propto \theta^{a-1} (1-\theta)^{b-1} \theta^{n\bar{y}} (1-\theta)^{n(1-\bar{y})} = \theta^{a+n\bar{y}-1} (1-\theta)^{b+n-n\bar{y}-1} \\ \implies p(\theta \mid y_1, \dots, y_n) &\propto \theta^{a+n\bar{y}-1} (1-\theta)^{b+n-n\bar{y}-1} \propto \text{dbeta}(a+n\bar{y}, b+n-n\bar{y}) \end{aligned}$$

So the posterior distribution is $\text{Beta}(a+n\bar{y}, b+n-n\bar{y})$.

(3) Give interpretations of the prior parameters.

The prior parameter a represents the prior number of successes from the bernoulli distribution, and the prior parameter b represents the prior number of failures from the bernoulli distribution.

(4) Find the posterior predictive distribution.

Let \tilde{Y} be a future pull from the Bernoulli distribution. Sample space of θ is $[0, 1] \subset \mathbb{R}$. Thus,

$$\begin{aligned} p(\tilde{Y} = 1 \mid y_1, \dots, y_n) &= \int_0^1 p(\tilde{Y} \mid \theta, y_1, \dots, y_n) p(\theta \mid y_1, \dots, y_n) d\theta = \\ &= \int_0^1 (\theta) \frac{\Gamma(a+b+n)}{\Gamma(a+n\bar{y})\Gamma(b+n-n\bar{y})} \left(\theta^{a+n\bar{y}-1} (1-\theta)^{b+n-n\bar{y}-1} \right) d\theta = \\ &= \frac{\Gamma(a+b+n)}{\Gamma(a+n\bar{y})\Gamma(b+n-n\bar{y})} \int_0^1 \left(\theta^{a+n\bar{y}+1-1} (1-\theta)^{b+n-n\bar{y}-1} \right) d\theta \end{aligned}$$

Since the integrand has the shape of a $\text{Beta}(a+n\bar{y}+1, b+n-n\bar{y})$, the integral must integrate to

$$\frac{\Gamma(a+n\bar{y}+1)\Gamma(b+n(1-\bar{y}))}{\Gamma(a+b+n+1)}$$

Thus,

$$p(\tilde{Y} = 1 \mid y_1, \dots, y_n) = \frac{\Gamma(a+b+n)}{\Gamma(a+n\bar{y})\Gamma(b+n-n\bar{y})} \frac{\Gamma(a+n\bar{y}+1)\Gamma(b+n(1-\bar{y}))}{\Gamma(a+b+n+1)}$$

Since $a, b, n, n\bar{y}$ are all non-negative integers, $\Gamma(\nu) = (\nu-1)!$, for $\nu \in (a, b, n, n\bar{y})$. Thus,

$$p(\tilde{Y} = 1 \mid y_1, \dots, y_n) = \frac{a+n\bar{y}}{a+b+n}$$

due to cancelations of factorials, and thus

$$p(\tilde{Y} = 0 \mid y_1, \dots, y_n) = \frac{b + n(1 - \bar{y})}{a + b + n}$$

Thus,

$$p(\tilde{Y} = 1 \mid y_1, \dots, y_n) = \left(\frac{a + n\bar{y}}{a + b + n} \right)^{\tilde{Y}} \left(\frac{b + n(1 - \bar{y})}{a + b + n} \right)^{1 - \tilde{Y}}$$

Which is a Bernoulli($\frac{a+n\bar{y}}{a+b+n}$) posterior predictive distribution.

Practice Problem 4

Let $(Y_1, \dots, Y_n \mid \theta, \sigma^2) \sim^{iid} \text{Normal}(\theta, \sigma^2)$, with known σ^2 .

(1) Find the conjugate family of priors for θ .

$$p(\theta \mid y_1, \dots, y_n) \propto p(\theta)p(y_1, \dots, y_n \mid \theta) \propto p(\theta) \left(\prod_{i=1}^n e^{-\frac{(y_i - \theta)^2}{2\sigma^2}} \right) = p(\theta) \left(\exp \left[-\frac{n\theta^2 - 2n\bar{y}\theta + \sum_{i=1}^n y_i^2}{2\sigma^2} \right] \right) \propto p(\theta) \exp \left[-c_1(\theta - c_2) \right]$$

For some constants c_1, c_2 , which gives implies the prior has the shape of a Normal distribution. Thus, the conjugate family for the prior is a Normal distribution.

(2) Find the posterior distribution

Let $\theta \sim \text{Normal}(\mu_0, \tau_0^2)$ for known prior hyperparameters μ_0, τ_0^2 with $\tau_0^2 < \infty$. Then,

$$p(\theta \mid y_1, \dots, y_n) \propto \exp \left[\frac{(\theta - \mu_0)^2}{2\tau_0^2} \right] \left(\exp \left[-\frac{n\theta^2 - 2n\bar{y}\theta + \sum_{i=1}^n y_i^2}{2\sigma^2} \right] \right) = \exp \left[-\frac{1}{2} \left(\frac{\theta^2 - 2\theta\mu_0 + \mu_0^2}{\tau_0^2} + \frac{n\theta^2 - 2n\bar{y}\theta + \sum_{i=1}^n y_i^2}{\sigma^2} \right) \right]$$

Considering only what is in the brackets of the exponential and ignoring the one-half for a moment, the portion within the round brackets simplifies to

$$= \theta^2 \left(\frac{1}{\tau_0^2} + \frac{n}{\sigma^2} \right) - 2\theta \left(\mu_0 \frac{1}{\tau_0^2} + \bar{y} \frac{n}{\sigma^2} \right) + \left(\frac{\mu_0^2}{\tau_0^2} + \frac{\sum_{i=1}^n y_i^2}{\sigma^2} \right)$$

Letting

$$\tau_n^2 = \frac{1}{\frac{1}{\tau_0^2} + \frac{n}{\sigma^2}}$$

and

$$\mu_n = \frac{\mu_0 \frac{1}{\tau_0^2} + \bar{y} \frac{n}{\sigma^2}}{1/\tau_n^2}$$

the posterior distribution simplifies to

$$p(\theta \mid y_1, \dots, y_n) \propto \exp \left[\frac{(\theta - \mu_n)^2}{2\tau_n^2} \right]$$

Which implies that the posterior distribution for $\theta \mid y_1, \dots, y_n$ is Normal (μ_n, τ_n^2)

(3) Give an interpretation of the prior parameters.

μ_0 can be thought of as our prior belief of what the average value of θ should be (i.e. what value we expect θ to be) and τ_0^2 is our prior belief of the variance of θ .

(4) Find the posterior predictive distribution.

If we knew what θ, σ^2 were, then $\tilde{Y} \sim \text{Normal}(\theta, \sigma^2)$ regardless of any data sampled. Since we do not know what θ is, to find $p(\tilde{Y} \mid y_1, \dots, y_n)$, let $\tilde{Y} = \theta + \tilde{\nu}$ where $\tilde{\nu} \mid \theta, \sigma^2 \sim \text{normal}(0, \sigma^2)$. Thus, the posterior predictive distribution for Y is equal to the posterior predictive distribution of $\theta + \tilde{\nu}$. Thus, \tilde{Y} is normal with

$$\mathbb{E}[\theta + \tilde{\nu} \mid \sigma^2, y_1, \dots, y_n] = \mathbb{E}[\theta \mid \sigma^2, y_1, \dots, y_n] + \mathbb{E}[\tilde{\nu} \mid \sigma^2, y_1, \dots, y_n] = \mu_n + 0$$

and

$$\text{Var}[\theta + \tilde{\nu} \mid \sigma^2, y_1, \dots, y_n] = \tau_n^2 + \sigma^2$$

Thus, the posterior predictive distribution of \tilde{Y} is Normal $(\mu_n, \tau_n^2 + \sigma^2)$.