

STAT 513 fa 2020 hw 8

1. Suppose you are to observe the random variables X_1, \dots, X_n coming from the hierarchical model

$$\begin{aligned} X_1, \dots, X_n | \lambda &\stackrel{\text{ind}}{\sim} \text{Exponential}(\lambda) \\ \lambda &\sim \text{Inverse-Gamma}(\alpha, \beta), \end{aligned}$$

where the pdf of the $\text{Inverse-Gamma}(\alpha, \beta)$ distribution is given by

$$f(\lambda; \alpha, \beta) = \frac{1}{\Gamma(\alpha)\beta^\alpha} \left(\frac{1}{\lambda}\right)^{\alpha+1} \exp\left(-\frac{1}{\lambda\beta}\right) \mathbf{1}(\lambda > 0).$$

(a) Find an expression for the prior (unconditional) mean of λ .

We have

$$\begin{aligned} \mathbb{E}\lambda &= \int_0^\infty \lambda \frac{1}{\Gamma(\alpha)\beta^\alpha} \left(\frac{1}{\lambda}\right)^{\alpha+1} \exp\left(-\frac{1}{\lambda\beta}\right) d\lambda \\ &= \frac{\Gamma(\alpha-1)\beta^{\alpha-1}}{\Gamma(\alpha)\beta^\alpha} \underbrace{\int_0^\infty \frac{1}{\Gamma(\alpha-1)\beta^{\alpha-1}} \left(\frac{1}{\lambda}\right)^{(\alpha-1)+1} \exp\left(-\frac{1}{\lambda\beta}\right) d\lambda}_{=1} \\ &= \frac{1}{(\alpha-1)\beta}, \end{aligned}$$

using the fact that $\Gamma(\alpha) = (\alpha-1)\Gamma(\alpha-1)$.

(b) Write down the joint pdf of the data $X_1, \dots, X_n | \lambda$ conditional on λ .

The pdf of the data X_1, \dots, X_n conditional on λ is given by

$$\begin{aligned} f(X_1, \dots, X_n | \lambda) &= \prod_{i=1}^n \left(\frac{1}{\lambda}\right) \exp\left(-\frac{X_i}{\lambda}\right) \\ &= \left(\frac{1}{\lambda}\right)^n \exp\left(-\frac{n\bar{X}_n}{\lambda}\right). \end{aligned}$$

(c) Find the posterior distribution of $\lambda | X_1, \dots, X_n$.

We have

$$\begin{aligned}\pi(\lambda|X_1, \dots, X_n) &\propto \left(\frac{1}{\lambda}\right)^n \exp\left(-\frac{n\bar{X}_n}{\lambda}\right) \frac{1}{\Gamma(\alpha)\beta^\alpha} \left(\frac{1}{\lambda}\right)^{\alpha+1} \exp\left(-\frac{1}{\lambda\beta}\right) \\ &\propto \left(\frac{1}{\lambda}\right)^{n+\alpha+1} \exp\left[-\frac{1}{\lambda}\left(n\bar{X}_n + \frac{1}{\beta}\right)\right] \\ &\propto \text{pdf of the Inverse-Gamma}(n + \alpha, 1/(n\bar{X}_n + \beta^{-1})) \text{ distribution.}\end{aligned}$$

So

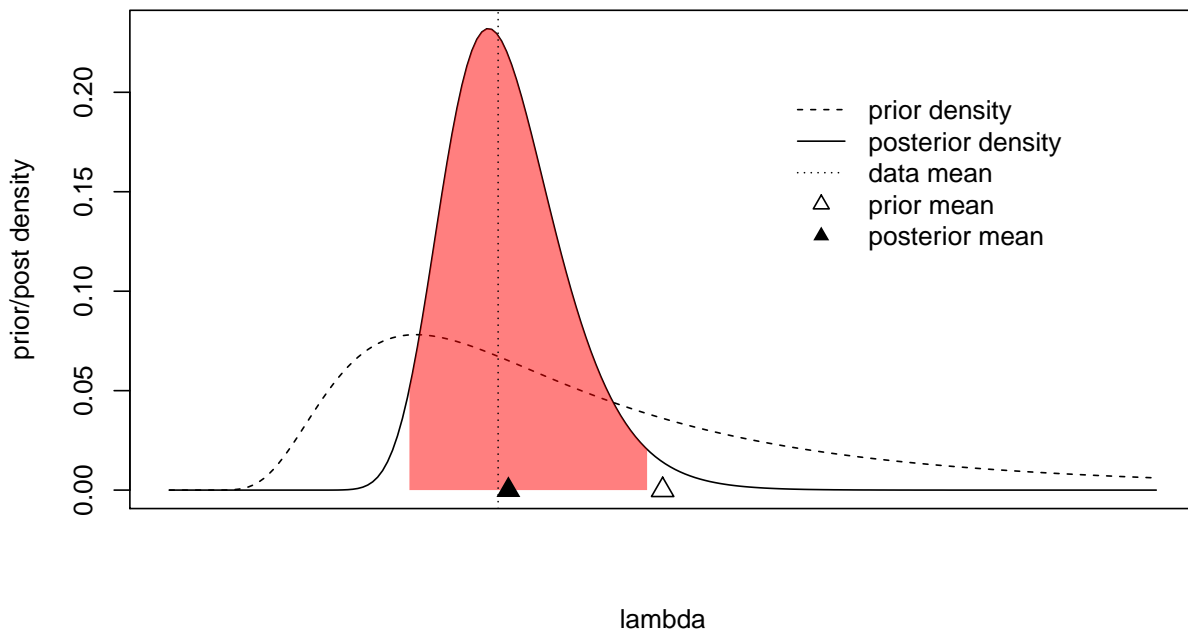
$$\lambda|X_1, \dots, X_n \sim \text{Inverse-Gamma}(n + \alpha, 1/(n\bar{X}_n + \beta^{-1})).$$

- (d) Give an expression for the posterior mean $\mathbb{E}[\lambda|X_1, \dots, X_n]$ in terms of α , β , n , and the data mean \bar{X}_n .

Based on the solution to part (a), the posterior mean of $\lambda|X_1, \dots, X_n$ is given by

$$\mathbb{E}\lambda|X_1, \dots, X_n = \frac{1}{(n + \alpha - 1)/(n\bar{X}_n + \beta^{-1})} = \frac{n\bar{X}_n + \beta^{-1}}{n + \alpha - 1}.$$

- (e) Refer to the plot below, which shows the prior and posterior densities of λ under $\beta = 1/30$ and $\alpha = 3$ for an observed value of $\bar{X}_n = 10$ based on a sample of size $n = 30$. The shaded region represents the 95% equal-tails Bayesian credible interval for $\lambda|X_1, \dots, X_n$.



- i. Give the prior mean of λ .

The prior mean is

$$\mathbb{E}\lambda = \frac{1}{(3-1)(1/30)} = 15.$$

- ii. Give the posterior mean of $\lambda|X_1, \dots, X_n$.

The posterior mean is

$$\mathbb{E}[\lambda|X_1, \dots, X_n] = \frac{1}{(30+3-1)/(30(10)+30)} = 10.3125.$$

- iii. Find the endpoints of the 95% equal-tails credible interval for λ , which is depicted in the plot above. *Hint: Use the fact*

$$\lambda \sim \text{Inverse-Gamma}(\alpha, \beta) \iff 1/\lambda \sim \text{Gamma}(\alpha, \beta)$$

to find a credible interval for $1/\lambda$, and then transform the interval so that it contains λ with probability 0.95, conditional on the data.

First find the equal-tails credible interval for $1/\lambda|X_1, \dots, X_n$: We have

$$\begin{aligned} P(1/\lambda < \Gamma_{n+\alpha, (n\bar{X}_n+\beta^{-1})^{-1}, 1-\alpha/2} | X_1, \dots, X_n) \\ = P(1/\lambda > \Gamma_{n+\alpha, (n\bar{X}_n+\beta^{-1})^{-1}, \alpha/2} | X_1, \dots, X_n) = \alpha/2, \end{aligned}$$

where for $\beta = 1/30$, $\alpha = 3$, $\bar{X}_n = 10$, and $n = 30$

$$\begin{aligned} \Gamma_{n+\alpha, (n\bar{X}_n+\beta^{-1})^{-1}, 1-\alpha/2} &= 0.0688354 \\ \Gamma_{n+\alpha, (n\bar{X}_n+\beta^{-1})^{-1}, \alpha/2} &= 0.1368923. \end{aligned}$$

So we have

$$P(1/\lambda < 0.0688354 | X_1, \dots, X_n) = P(1/\lambda > 0.1368923 | X_1, \dots, X_n) = \alpha/2,$$

or

$$P(\lambda > 1/0.0688354 | X_1, \dots, X_n) = P(\lambda < 1/0.1368923 | X_1, \dots, X_n) = \alpha/2,$$

giving the interval

$$(1/0.1368923, 1/0.0688354) = (7.305013, 14.52741).$$

2. Suppose we are interested in estimating a success probability p based on the number of successes in n Bernoulli trials, and suppose that we use the following Bayesian setup:

$$\begin{aligned} Y|p &\sim \text{Binomial}(n, p) \\ p &\sim \text{Uniform}(0, 1). \end{aligned}$$

- (a) What beliefs about p does the prior distribution reflect?

The pdf of the prior distribution is flat, spreading the probability mass evenly across the parameter space $(0, 1)$ such that all values of p are equally likely. A flat prior expresses the absence of any prior beliefs about p .

- (b) Find the posterior distribution of $p|Y$.

The prior pdf is given by $\pi(p) = 1$ for $p \in (0, 1)$, so the posterior pdf of $p|Y$ is given by

$$\pi(p|Y) \propto \binom{n}{Y} p^Y (1-p)^{n-Y} \cdot 1 \propto p^{(Y+1)-1} (1-p)^{(n-Y+1)-1},$$

which is proportional to the pdf of the Beta($Y + 1, n - Y + 1$) distribution. So we have

$$p|Y \sim \text{Beta}(Y + 1, n - Y + 1).$$

(c) Give an expression for the posterior mean of $p|Y$.

If $X \sim \text{Beta}(\alpha/\beta)$, then $\mathbb{E}X = \alpha/(\alpha + \beta)$, so we have

$$\mathbb{E}[p|Y] = \frac{Y + 1}{n + 1 + 1}.$$

(d) Suppose $n = 10$ and $Y = 3$ is observed. Give the equal-tails 95% credible interval for $p|Y$.

For $n = 10$ and $Y = 3$, we have $p|Y \sim \text{Beta}(4, 8)$, so the 95% equal-tails credible interval for $p|Y$ is the interval between the lower and upper 0.025 quantiles of the $\text{Beta}(4, 8)$ distribution. These quantiles are given by

$$\text{qbeta}(.025, 4, 8) = 0.1092634 \quad \text{and} \quad \text{qbeta}(.975, 4, 8) = 0.6097426,$$

so the 95% credible interval is $(0.1092634, 0.6097426)$.

(e) i. Run a simulation in R to compare the coverage of the 95% frequentist Wald-type confidence interval for p given by

$$\hat{p} \pm z_{0.025} \sqrt{\hat{p}(1 - \hat{p})/n},$$

where $\hat{p} = Y/n$ and the coverage of the 95% Bayesian credible interval: Generate 1000 values of p from the $\text{Uniform}(0, 1)$ distribution, and for each of these, generate a value of Y from the $\text{Binomial}(n, p)$ distribution. Then compute the upper and lower bounds of the frequentist confidence interval and of the Bayesian credible interval and record whether the true value of p was captured in the interval or not. Then compute the proportion of times each interval captured the true value of p . Use $n = 10$. Show your R code and give the coverage results.

```
n <- 10
S <- 1000

covered.Bayes <- covered.Wald <- numeric()

for( s in 1:S)
{
  p <- runif(1)
  Y <- rbinom(1,n,p)

  p.hat <- Y/n

  upper.Wald <- p.hat + qnorm(.975)*sqrt(p.hat*(1-p.hat)/n)
```

```

lower.Wald <- p.hat - qnorm(.975)*sqrt(p.hat*(1-p.hat)/n)

covered.Wald[s] <- ( p > lower.Wald) & (p < upper.Wald)

upper.Bayes <- qbeta(.975,Y+1,n-Y+1)
lower.Bayes <- qbeta(.025,Y+1,n-Y+1)

covered.Bayes[s] <- ( p > lower.Bayes) & (p < upper.Bayes)
}
mean(covered.Bayes)
mean(covered.Wald)

```

I got 0.947 coverage for the 95% Bayes credible interval and 0.793 coverage for the 95% Wald-type interval.

- ii. Interpret your simulation; which interval performed better? What happens when you use $n = 100$?

The Bayesian credible interval performed much better than the frequentist Wald-type interval under $n = 10$. Note: This doesn't mean Bayesian methods are better than frequentist methods. There are frequentist confidence intervals which perform much better than the Wald interval! For $n = 100$, I get 0.953 coverage for the 95% Bayes credible interval and 0.935 coverage for the 95% Wald-type interval. The Wald-type interval is, after all, based on the asymptotic convergence of the quantity $(\hat{p} - p)/\sqrt{\hat{p}(1 - \hat{p})/n}$ to the $\text{Normal}(0, 1)$ distribution, so its coverage approaches the nominal, that is the specified coverage as $n \rightarrow \infty$.

- (f) Suppose you wish to test the hypotheses

$$H_0: p \leq 1/2 \text{ versus } H_1: p > 1/2.$$

Let $n = 10$ and suppose you observe $Y = 6$.

- i. Compute the prior odds in favor of H_0 over H_1 and interpret the number.

The prior odds in favor of H_0 over H_1 are

$$\frac{\pi_0}{\pi_1} = \frac{P(p \leq 1/2)}{P(p > 1/2)} = \frac{1/2}{1/2} = 1.$$

Since the prior odds are equal to 1, the null hypothesis is regarded as no more or less likely to be true than the alternate hypothesis.

- ii. Compute the posterior odds in favor of H_0 over H_1 and interpret the number.

The posterior odds in favor of H_0 over H_1 are

$$\frac{\pi_0}{\pi_1} = \frac{P(p \leq 1/2|Y)}{P(p > 1/2|Y)} = \frac{\text{pbeta}(1/2, 6+1, 10-6+1)}{1 - \text{pbeta}(1/2, 6+1, 10-6+1)} = \frac{0.2744141}{1 - 0.2744141} = 0.3781965.$$

Given that $Y = 6$ was observed, H_0 is about a third as likely as H_1 to be true.

iii. Compute the Bayes factor associated with observing $Y = 6$.

The Bayes factor is the ratio of the posterior odds over the prior odds. Since the prior odds were equal to 1, the Bayes factor is in this case equal to the posterior odds: 0.3781965. This is the factor by which the odds in favor of H_0 over H_1 were changed by the observation of the data $Y = 6$.

iv. In your opinion, is the evidence against H_0 strong enough to reject it?

There is some evidence in the data against H_0 , but it does not seem very strong.

3. Let \mathbf{Y} be an $n \times 1$ vector of random variables and \mathbf{X} be an $n \times p$ matrix of real numbers such that

$$\begin{aligned}\mathbf{Y}|\boldsymbol{\beta} &\sim \text{Normal}(\mathbf{X}\boldsymbol{\beta}, \sigma^2\mathbf{I}_n) \\ \boldsymbol{\beta} &\sim \text{Normal}(\mathbf{0}, \sigma^2\kappa_0^{-1}\mathbf{I}_p),\end{aligned}$$

where $\boldsymbol{\beta}$ is a $p \times 1$ vector of random variables and σ^2 and κ_0 are known, fixed constants. This is the multiple linear regression model with Normally distributed errors and with the regression coefficients regarded as random variables. It can be shown that the posterior distribution of $\boldsymbol{\beta}|\mathbf{Y}$ is given by

$$\boldsymbol{\beta}|\mathbf{Y} \sim \text{Normal}((\mathbf{X}^T\mathbf{X} + \kappa_0\mathbf{I}_p)^{-1}\mathbf{X}^T\mathbf{Y}, \sigma^2(\mathbf{X}^T\mathbf{X} + \kappa_0\mathbf{I}_p)^{-1}).$$

- (a) Suggest a Bayesian estimator for $\boldsymbol{\beta}$ based on the posterior distribution of $\boldsymbol{\beta}|\mathbf{Y}$.
- (b) Install the R package `hdi` and read in the data set `riboflavin` which is included in that package. Enter `?riboflavin` into the console to learn more about the data; the data contain for $n = 71$ samples the values of a response variable measuring riboflavin production and the values of 4088 predictor variables, which are gene expression levels. Store the responses in a vector \mathbf{Y} and store in a matrix \mathbf{X} the values of the first 100 predictors, centering both \mathbf{Y} and the columns of \mathbf{X} to have mean 0 and scaling them to have standard deviation 1. All of this is achieved with the following code:

```
library(hdi)
data(riboflavin)
Y <- scale(riboflavin[,1])
X.allcovariates <- scale(riboflavin[,-1])
X <- X.allcovariates[,1:100]
```

Try to compute the least-squares estimator of $\boldsymbol{\beta}$ based on the data in the matrix \mathbf{X} and the vector \mathbf{Y} . Explain why you get an error.

The matrix $\mathbf{X}^T\mathbf{X}$ is not invertible, because \mathbf{X} does not have full-column rank; it cannot have full-column rank because the number of columns, $p = 100$, exceeds the number of rows, $n = 71$. Therefore we cannot compute the least-squares estimator.

- (c) Set $\kappa_0 = 1$ and compute the Bayesian estimator you suggested in part (a). Give the first 5 entries of the estimated value of β . *Hint: You can make the $p \times p$ identity matrix \mathbf{I}_p in R with the command `diag(p)`.*

The matrix $\mathbf{X}^T\mathbf{X} + \kappa_0\mathbf{I}_p$ is invertible, so we can compute the Bayesian estimator of β . With the code

```
p <- ncol(X)
k0 <- 1
beta.hat.Bayes <- solve( t(X) %*% X + k0 * diag(p)) %*% t(X) %*% Y
round(beta.hat.Bayes[1:5], 3)
```

we get that the first five entries of the Bayes estimator are 0.107, -0.102 , 0.136, 0.007, and 0.100

- (d) Set $\kappa_0 = 10$ and do the same thing.

The first five entries are now 0.084, -0.055 , 0.034, 0.074, -0.018 .

- (e) Describe how the estimates changed.

The larger value of κ_0 resulted in estimates closer to zero for the regression coefficients.

- (f) For what value of κ_0 would your Bayesian estimator be the same as the least-squares estimator?

As $\kappa_0 \rightarrow 0$, the Bayesian estimator approaches the least-squares estimator.

- (g) What happens if $\kappa_0 \rightarrow \infty$? *Hint: You can try very large values of κ_0 and see what happens.*

As $\kappa_0 \rightarrow \infty$, all the estimated regression coefficients approach 0.

- (h) What beliefs about β are reflected by choosing κ_0 to be very large?

Larger choices of κ_0 reflect prior beliefs that the regression coefficients are close to 0, that is, for each gene, we begin with the belief that it is probably not important; the data must then convince us otherwise.