STATS 551 Homework 1 Solution

Introduction & Single Parameter Models

Responsible GSI: Wayne Wang

Practice of Bayes Formula (3×10 points). Suppose that if $\theta = 1$, then y has a normal distribution with mean 1 and standard deviation σ, and if $\theta = 2$, then y has a normal distribution with mean 2 and standard deviation σ. Also, suppose $\Pr(\theta = 1) = 0.5$ and $\Pr(\theta = 2) = 0.5$.

1. For $\sigma = 2$, derive the formula for the marginal probability density for y, p(y), and sketch/visualize it in R.
2. What is $Pr(\theta = 1|y = 1)$, again supposing $\sigma = 2$?
3. Describe how the posterior density of θ changes in shape as σ is increased and as it is decreased.

Solution:

1.

$$\begin{aligned}
p(y) &= \int p(y|\theta)p(\theta)d\theta \\
&= p(y|\theta = 1)(0.5) + p(y|\theta = 2)(0.5) \\
&= 0.5(N(y|1, 2^2) + N(y|2, 2^2)).
\end{aligned}$$

2.

$$p(\theta = 1|y = 1) = \frac{p(y = 1|\theta = 1)p(\theta = 1)}{p(y = 1)} \approx 0.5312.$$

3.

$$p(\theta|y) = \frac{\exp\{-\frac{1}{2\sigma^2}(y-\theta)^2\}}{\exp\{-\frac{1}{2\sigma^2}(y-1)^2\} + \exp\{-\frac{1}{2\sigma^2}(y-2)^2\}}, \quad \theta \in \{1, 2\},$$

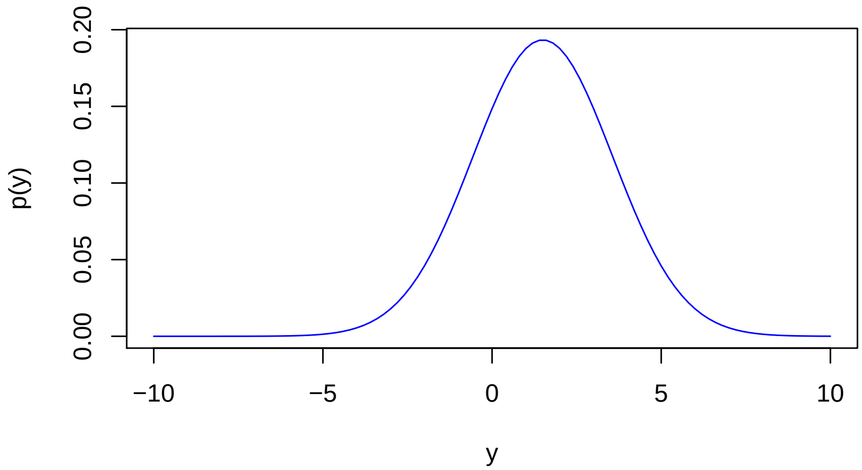


Figure 1: Marginal density of y.

and

$$p(1|y) = \frac{1}{1 + \exp(\frac{2y-3}{2\sigma^2})},$$
$$p(2|y) = \frac{1}{1 + \exp(-\frac{2y-3}{2\sigma^2})}.$$

So, when $\sigma \to \infty$, $p(\theta|y) \to p(\theta)$. When $\sigma \to 0$, consider cases if $y < 3/2$, then $p(\theta = 1|y) \to 1$; if $y > 3/2$, then $p(\theta = 2|y) \to 1$.

Normal distribution with unknown mean (3×10 points). A random sample of n students is drawn from a large population, and their weights are measured. The average weight of the n sampled students is $\bar{y} = 150$ pounds. Assume the weights in the population are normally distributed with unknown mean θ and known standard deviation 20 pounds. Suppose your prior distribution for θ is normal with mean 180 and standard deviation 40.

1. Give your posterior distribution for θ. (Your answer will be a function of n.)
2. A new student is sampled at random from the same population and has a weight of $\tilde{y}$ pounds. Give a posterior predictive distribution for $\tilde{y}$. (Your answer will still be a function of n.)

3. For $n = 10$, give a 95% posterior interval for θ and a 95% posterior predictive interval for $\tilde{y}$. Do the same for $n = 100$.

Solution:

1.

$$
\begin{aligned}
p(\theta|y) &\propto p(y|\theta)p(\theta) \\
&= \exp\{-\frac{(n40^2 + 20^2)\theta^2 - 2\theta(40^2 n\bar{y} + 20^2(180)) + (40^2 n\bar{y^2} + 20^2(180))}{2(20^2)(40^2)}\} \\
&\sim N(y|\mu_n, \tau_n^2),
\end{aligned}
$$

where the last equality can be seen by completing the square inside the exponential function by adding/subtracting appropriate constants, and

$$
\begin{aligned}
\mu_n &= \frac{\frac{n}{20^2}\bar{y} + \frac{1}{40^2}(180)}{\frac{n}{20^2} + \frac{1}{40^2}}, \\
\tau_n^2 &= \frac{1}{\frac{n}{20^2} + \frac{1}{40^2}}.
\end{aligned}
$$

2.

$$
\begin{aligned}
p(\tilde{y}|y) &= \int p(\tilde{y}|\theta)p(\theta|y)d\theta \\
&\propto \int \exp\{-\frac{1}{2(20^2)}(\tilde{y} - \theta)^2\} \exp\{-\frac{1}{2\tau_n^2}(\theta - \mu_n)^2\}d\theta \\
&= \exp\{-\frac{\tau_n^2\tilde{y}_n^2 + 20^2\mu_n^2}{2(20^2)(\tau_n^2)}\} \int \exp\{-\frac{(\tau_n^2 + 20^2)\theta^2 - 2\theta(\tilde{y}\tau_n^2 + 20^2\tau_n^2)}{2(20^2)(\tau_n^2)}\}d\theta \\
&\propto \exp\{-\frac{\tilde{y}^2 + \mu_n^2 - 2\tilde{y}\mu_n}{2(20^2 + \tau_n^2)}\} \\
&\sim N(\tilde{y}|\mu_n, 20^2 + \tau_n^2).
\end{aligned}
$$

Again, by completing the squares, you'll find that the expression inside the integral results in the density function for

$$
N(\frac{\tau_n^2\tilde{y} + 20^2\mu_n}{20^2 + \tau_n^2}, \frac{1}{\frac{1}{20^2} + \frac{1}{\tau_n^2}}),
$$

so as a result the integration is 1.

3. From previous questions we have for $n = 10$,

$$p(\theta|y) = N(\theta|150.73, 6.25^2)$$
$$p(\tilde{y}|y) = N(\tilde{y}|150.73, 20.95^2).$$

So the corresponding 95% posterior interval for θ and posterior predictive interval for $\tilde{y}$ are $[138.48, 162.98]$ and $[109.67, 191.79]$, respectively. For $n = 100$,

$$p(\theta|y) = N(\theta|150.07, 1.997^2)$$
$$p(\tilde{y}|y) = N(\tilde{y}|150.07, 20.10^2).$$

So the corresponding 95% posterior interval for θ and posterior predictive interval for $\tilde{y}$ are $[146.16, 153.98]$ and $[110.67, 189.47]$, respectively.

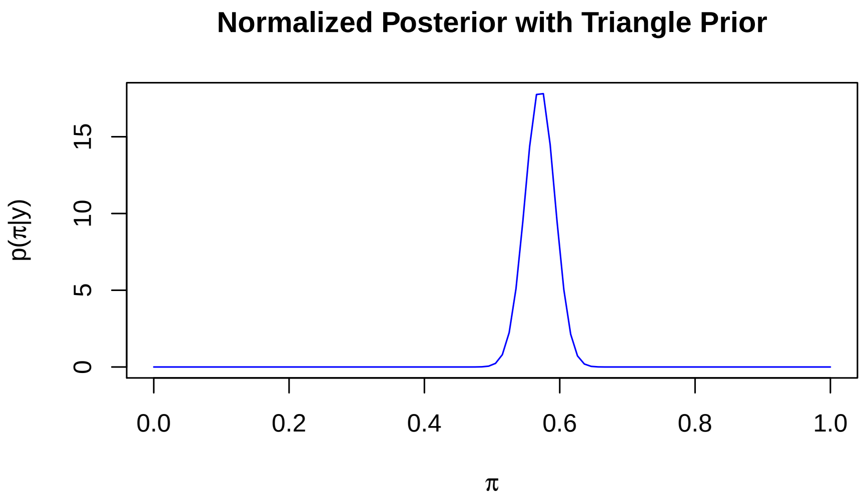
Nonconjugate single parameter model (2×20 points). Suppose you observe $y = 285$ from the model Binomial$(500, \theta)$, where θ is an unknown parameter. Assume the prior on θ has the following form

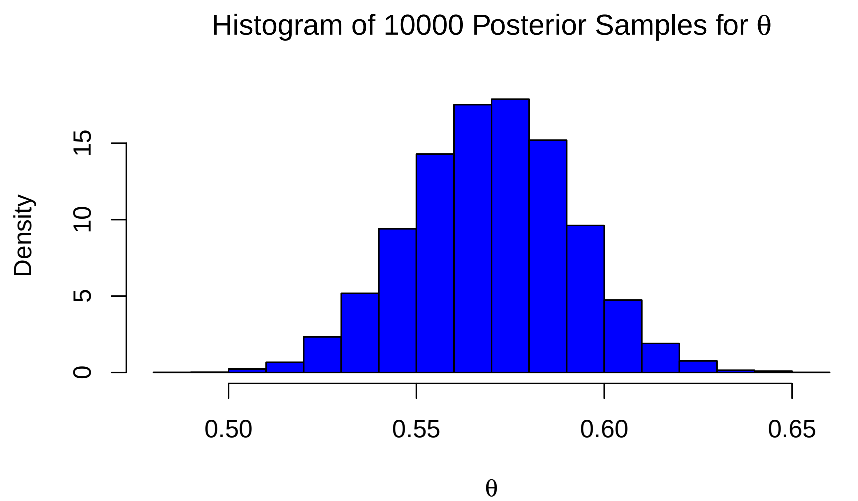
$$p(\theta) = \begin{cases} 8\theta, & 0 \le \theta < 0.25 \\ \frac{8}{3} - \frac{8}{3}\theta, & 0.25 \le \theta \le 1 \\ 0, & \text{otherwise.} \end{cases}$$

1. Compute the unnormalized posterior density function on a grid of m points for some large integer m. Using the grid approximation, compute and plot the normalized posterior density function $p(\theta|y)$, as a function of θ.

2. Sample 10000 draws of θ from the posterior density and plot a histogram of the draws.

Solution: See the attached `R` code and plots.


Normalized Posterior with Triangle Prior
p(π|y)
0
5
10
15
0.0
0.2
0.4
0.6
0.8
1.0
π


Histogram of 10000 Posterior Samples for θ
Density
0
5
10
15
0.50
0.55
0.60
0.65
θ

```
# Define the prior density on theta
triangle.prior <- function(x) {
  if (x >= 0 && x < 0.25)
    8 * x
  else if (x >= 0.25 && x <= 1)
    8/3 - 8 * x/3
  else 0
}

# Define the unnormalized posterior density
posterior.function <- function(theta, n, y) {
  (theta^y) * (1 - theta)^(n - y) * triangle.prior(theta)
}

# Obtain a grid of m points from 0 to 1
m <- 100
grid.points <- seq(from = 0, to = 1, length.out = m)

# Compute (at y=285) the unnormalized poterior density
unnormal.post.ord <-
  posterior.function(theta = grid.points, n = 500, y = 285)

# Compute the normalizing constant using grid approximation
k <- 1/m # step size
normal.constant <- sum(k * unnormal.post.ord)
post.ord <-
  unnormal.post.ord/normal.constant # normalized posterior density

# Plot the normalized posterior density
plot(x = grid.points, y = post.ord, type = "l", col = "blue",
     xlab = expression(pi),
     ylab = expression(paste("p(", pi, "|y)")),
     main = "Normalized␣Posterior␣with␣Triangle␣Prior")

# Draw from the posterior distribution
set.seed(123)
posterior.triangle.1 <- sample(grid.points, size = 10000,
                               replace = T, prob = post.ord)
hist(posterior.triangle.1, col = "blue",
     main = expression(
       paste("Histogram␣of␣10000␣Posterior␣Samples␣for␣",theta)),
```

```
xlab = expression(theta),
probability = T)
```