HW3

Alexander Grunenwald, Logan Morgoff, Grant Gilly, Refat Zeyed

2024-10-06

Question 1

Part a.)

$$(Y \mid \frac{1}{\lambda}) \propto \lambda^{\frac{n}{2}+\alpha-1} \exp -\frac{\lambda n v}{2} + \beta - 1 \mid v = \frac{1}{n}\sum(y_i - \mu)^2$$
$$P(Y \mid \lambda^{-1}) \sim Gamma(\frac{n}{2} + \alpha, \frac{nv}{2} + \beta)$$

Part b.)

$$\text{Given } y_i \in [8.4, 10.1, 9.4], \mu = 8, \alpha = 3, \beta = 2, v = \frac{1}{3}||y - \mu|| = 2.1766$$
$$P(\lambda \mid y, \mu, \alpha, \beta) \sim Gamma(4.5, 5.265)$$
$$\mathbb{E}[\lambda \mid y] = \frac{4.5}{5.265} = 0.8547$$

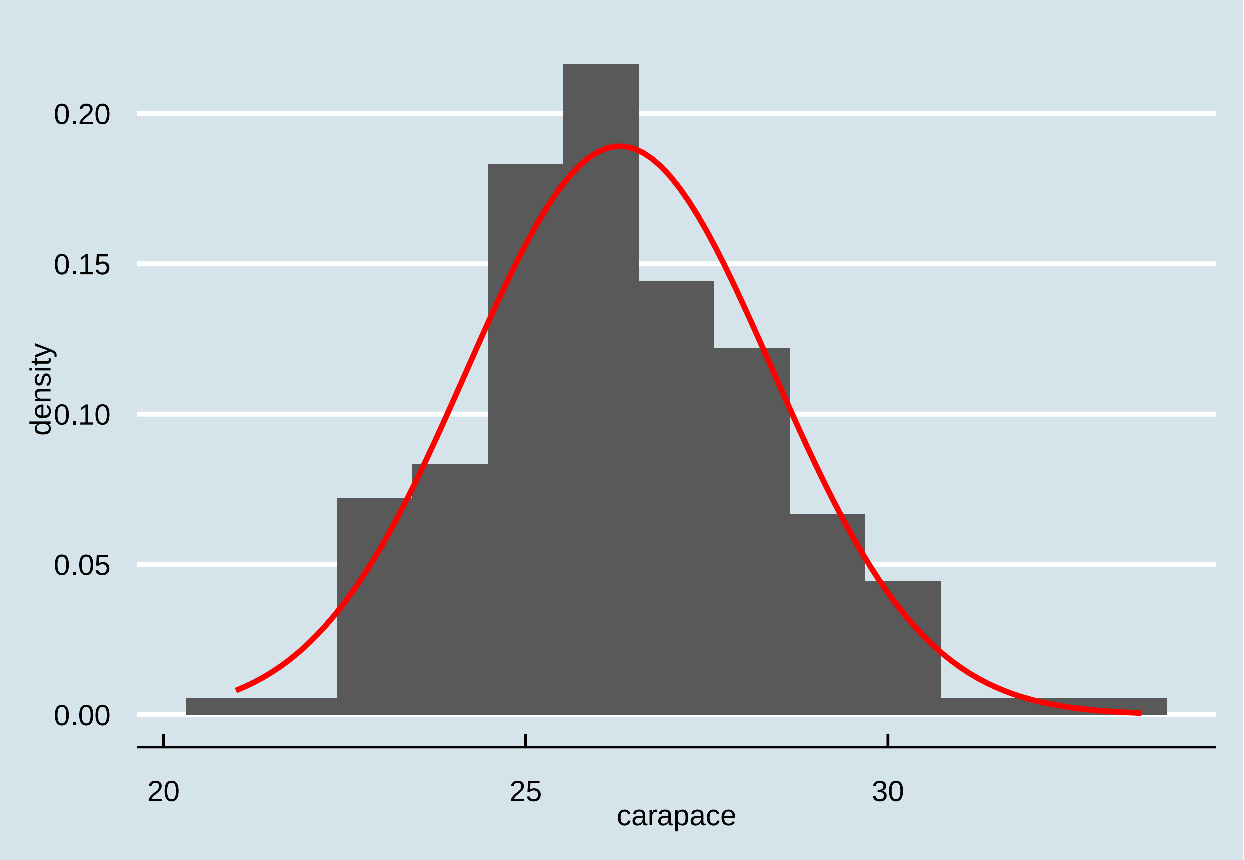
Question 2

```
crab <- read.table("data/Crabs.txt", header=T)
head(crab)
```

```
##   crab y weight width color spine
## 1    1 8   3.05  28.3     2     3
## 2    2 0   1.55  22.5     3     3
## 3    3 9   2.30  26.0     1     1
## 4    4 0   2.10  24.8     3     3
## 5    5 4   2.60  26.0     3     3
## 6    6 0   2.10  23.8     2     3
```

```
carapace <- crab$width
sample_mean <- mean(carapace)
sample_sd <- sqrt(var(carapace))
num_bins <- as.integer(sqrt(length(carapace)))

ggplot(data=crab, mapping = aes(x=carapace)) + geom_histogram(aes(y=after_stat(density)), bins = num_bi
                args = list(mean = sample_mean, sd = sample_sd),
                color = "red", linewidth = 1) + theme_economist()
```



The Normal distribution with the parameters of the smaple mean and variance seems to be a good fit to this distribution.

Given that the prior and sampling distribution are Normal

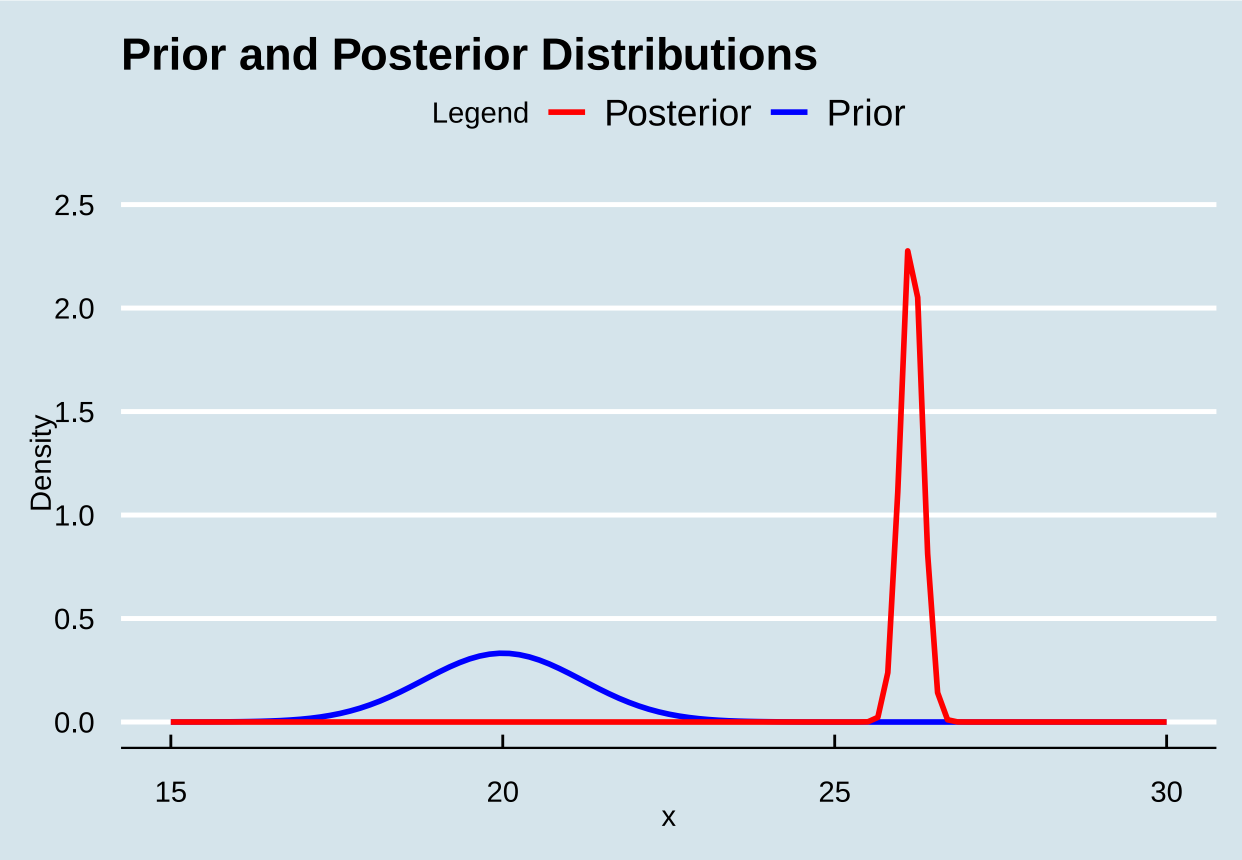
$$(y \mid \mu\sigma^2) \sim N(\mu \mid y, \sigma^2 = 4.8)$$
$$(\mu \mid \mu_o = 20, \tau_0^2 = 1.2) \sim N(20, 1.2)$$

Then the posterior distribution will be normal with the following parameters

$$P(\mu \mid \sigma^2 = 4.8, y) \propto \exp[-\frac{1}{2}(\frac{\sum(\mu - \mu_0)}{\tau_0^2} + \frac{\sum(y_i - \mu)^2}{4.8})]$$
$$P(\mu \mid \bar{y}) = N(\mu \mid \mu_n, \tau_n^2) = N(26.1256, 0.02739267)$$

```
ggplot() +
  stat_function(
    fun = dnorm,
    args = list(mean = 20, sd = 1.2),
    aes(color = "Prior"),
    linewidth = 1
  ) +
  stat_function(
    fun = dnorm,
    args = list(mean = 26.156, sd = 0.1655073),
    aes(color = "Posterior"),
    linewidth = 1
```

```r
  ) +
  xlim(15, 30) +
  ylim(0, 2.5) +
  labs(
    title = "Prior and Posterior Distributions",
    x = "x",
    y = "Density",
    color = "Legend"
  ) +
  scale_color_manual(values = c("Prior" = "blue", "Posterior" = "red")) +
  theme_economist()
```




```r
26.156 + 0.1655073*(2)
```

```
## [1] 26.48701
```

```r
26.156 - 0.1655073*(2)
```

```
## [1] 25.82499
```

I observe that the centrality of the posterior has shifted after the data update as well as the Margin of Error has significantly decreased in relation to the prior.

Question 3

For School 1

```r
s1_school <- read.table("data/school1.txt")

# Param Decleration
mu_0 <- 5
kappa_0 <- 1
a <- 1
b <- 4

y1_bar <- mean(s1_school$V1)
n1 <- length(s1_school$V1)
s1_var <- var(s1_school$V1)

# Posterior Param Calculation
kappa_n <- kappa_0 + n1
mu_n <- (kappa_0 * mu_0 + n1 * y1_bar) / kappa_n
a_n <- a + n1 / 2
b_n <- b + 0.5 * sum((s1_school$V1 - y1_bar)^2)

# Sample from inverse Gamma
sigma2_s1_posterior <- 1 / rgamma(10000, shape = a_n, rate = b_n)

# Calculate Posterior Distribution for Means
sigma_s1_posterior <- sqrt(sigma2_s1_posterior)
mu_s1_posterior <- rnorm(10000, mean = mu_n, sd = sigma_s1_posterior / sqrt(kappa_n))

posterior_mean_s1 <- mean(mu_s1_posterior)
credible_interval_s1 <- quantile(mu_s1_posterior, probs = c(0.025, 0.975))

credible_interval_s1
```

```
##      2.5%     97.5%
##  7.784174 10.779696
```

```r
posterior_mean_s1
```

```
## [1] 9.292482
```

For School 2

```r
s2_school <- read.table("data/school2.txt")
# Param Decleration
mu_0 <- 5
kappa_0 <- 1
a <- 1
b <- 4
```

```r
y2_bar <- mean(s2_school$V1)
n2 <- length(s2_school$V1)
s2_var <- var(s2_school$V1)

# Posterior Param Calculation
kappa_n <- kappa_0 + n2
mu_n <- (kappa_0 * mu_0 + n2 * y2_bar) / kappa_n
a_n <- a + n2 / 2
b_n <- b + 0.5 * sum((s2_school$V1 - y2_bar)^2)

# Sample from inverse Gamma
sigma2_s2_posterior <- 1 / rgamma(10000, shape = a_n, rate = b_n)

# Calculate Posterior Distribution for Means
sigma_s2_posterior <- sqrt(sigma2_s2_posterior)
mu_s2_posterior <- rnorm(10000, mean = mu_n, sd = sigma_s2_posterior / sqrt(kappa_n))

posterior_mean_s2 <- mean(mu_s2_posterior)
credible_interval_s2 <- quantile(mu_s2_posterior, probs = c(0.025, 0.975))

credible_interval_s2
```

```
##     2.5%    97.5%
## 5.188244 8.720031
```

```r
posterior_mean_s2
```

```
## [1] 6.951046
```

Sigma for School 1

```r
quantile(sigma_s1_posterior, probs = c(0.025, 0.975))
```

```
##     2.5%    97.5%
## 2.917877 5.053373
```

```r
mean(sigma_s1_posterior)
```

```
## [1] 3.81152
```

Sigma for School 2

```r
quantile(sigma_s2_posterior, probs = c(0.025, 0.975))
```

```
##     2.5%    97.5%
## 3.317150 5.852693
```

```r
mean(sigma_s2_posterior)
```

```
## [1] 4.367907
```

Probability that mu1 is greater than mu2

```r
mean(mu_s1_posterior > mu_s2_posterior)
```

```
## [1] 0.9789
```