STATS 551 HW2

Minxuan Chen

2023/10/2

We contribute to each of the problem equally by finishing them individually and having a discussion to agree on the answers. And we improved our solution to Problem 3 together during the discussion.

Problem 1

Problem 2

(a)

Prior

$$\theta_A \sim \text{gamma}(120, 10), \theta_B \sim \text{gamma}(12, 1), p\left(\theta_A, \theta_B\right) = p\left(\theta_A\right) \times p\left(\theta_B\right)$$

Likelihood

$$\begin{aligned} y_A|\theta_A &\sim \mathcal{P}(\theta_A), y_A \text{ i.i.d.} \\ y_B|\theta_B &\sim \mathcal{P}(\theta_B), y_B \text{ i.i.d.} \\ y_A &\perp\!\!\!\perp y_B \end{aligned}$$

Posterior

$$\begin{aligned} \theta_A|\mathbf{y}_A, \mathbf{y}_B &\sim \text{gamma}(120 + 10\bar{y}_A, 10 + 10) = \text{gamma}(237, 20) \\ \theta_B|\mathbf{y}_A, \mathbf{y}_B &\sim \text{gamma}(12 + 13\bar{y}_B, 1 + 13) = \text{gamma}(125, 14) \end{aligned}$$

```r
ya <- c(12,9,12,14,13,13,15,8,15,6)
yb <- c(11,11,10,9,9,8,7,10,6,8,8,9,7)
sum(ya)
```

```
## [1] 117
```

```r
sum(yb)
```

```
## [1] 113
```

Posterior mean

$$E(\theta_A|\mathbf{y}_A, \mathbf{y}_B) = \frac{237}{20} = 11.85, \quad E(\theta_B|\mathbf{y}_A, \mathbf{y}_B) = \frac{125}{14} = 8.929$$

Posterior variance

$$Var(\theta_A|\mathbf{y}_A, \mathbf{y}_B) = \frac{237}{400}, \quad Var(\theta_B|\mathbf{y}_A, \mathbf{y}_B) = \frac{125}{196}$$

Posterior confidence interval

```r
#theta_A
cat("theta_A: ", qgamma(c(.025, .975), 237, 20), "\n")
```

```
## theta_A:  10.38924 13.40545
```

```r
#theta_B
cat("theta_B: ", qgamma(c(.025, .975), 125, 14))
```

```
## theta_B:  7.432064 10.56031
```

(b)

Posterior

$$\theta_B|\mathbf{y}_A, \mathbf{y}_B \sim \text{gamma}(12n_0 + 13\bar{y}_B, n_0 + 13) = \text{gamma}(12n_0 + 113, 13 + n_0)$$

```r
#posterior expectation is (12n_0+185)/(13+n_0)
(12*1:50+113)/(13+1:50)
```

```
##  [1]  8.928571  9.133333  9.312500  9.470588  9.611111  9.736842  9.850000
##  [8]  9.952381 10.045455 10.130435 10.208333 10.280000 10.346154 10.407407
## [15] 10.464286 10.517241 10.566667 10.612903 10.656250 10.696970 10.735294
## [22] 10.771429 10.805556 10.837838 10.868421 10.897436 10.925000 10.951220
## [29] 10.976190 11.000000 11.022727 11.044444 11.065217 11.085106 11.104167
## [36] 11.122449 11.140000 11.156863 11.173077 11.188679 11.203704 11.218182
## [43] 11.232143 11.245614 11.258621 11.271186 11.283333 11.295082 11.306452
## [50] 11.317460
```

For arbitrary gamma prior, $\theta_B \sim \text{gamma}(\alpha, \beta)$, to make the posterior expectation of θ_B to be close to that of θ_A, just make them equal and solve

$$\frac{\alpha + 113}{\beta + 13} = \frac{237}{20}$$

we get

$$\alpha = 41.05 + 11.83\alpha, \quad \alpha, \beta > 0$$

Therefore, if the parameters of θ_B's prior locate at the line $\alpha = 41.05 + 11.83\beta$, we can get exactly the same posterior expectation. To get a close result, you can add small perturbation to this line.

(c)

No. Since the prior $p(\theta_A, \theta_B) = p(\theta_A) \times p(\theta_B)$ implies the independence between type A and type B. In other words

$$p(\theta_B|\mathbf{y}_A, \mathbf{y}_B) = p(\theta_B|\mathbf{y}_B)$$

We can prove it.
Proof

$$p(\theta_A, \theta_B|\mathbf{y}_A, \mathbf{y}_B) \propto p(\mathbf{y}_A, \mathbf{y}_B|\theta_A, \theta_B)p(\theta_A, \theta_B)$$

Note that type A and type B are independent given θ_A, θ_B (due to different type, i.e. two populations), so

$$\begin{aligned}
p(\theta_A, \theta_B|\mathbf{y}_A, \mathbf{y}_B) &\propto p(\mathbf{y}_A|\theta_A, \theta_B)p(\mathbf{y}_B|\theta_A, \theta_B)p(\theta_A)p(\theta_B) \\
&\propto p(\mathbf{y}_A|\theta_A)p(\theta_A) \cdot p(\mathbf{y}_B|\theta_B)p(\theta_B) \\
&\propto p(\theta_A|\mathbf{y}_A)p(\theta_B|\mathbf{y}_B)
\end{aligned}$$

The second line is because in the sampling model we assume type A or B is only influenced by its own parameter.
We integrate over θ_A,

$$p(\theta_B|\mathbf{y}_A, \mathbf{y}_B) = C \cdot p(\theta_B|\mathbf{y}_B)$$

Note that both sides are pdf of θ_B, so if we integrate over θ_B again, we must have $C = 1$, i.e.

$$p(\theta_B|\mathbf{y}_A, \mathbf{y}_B) = p(\theta_B|\mathbf{y}_B)$$

It doesn't make sense to have $p(\theta_A, \theta_B) = p(\theta_A)p(\theta_B)$ because it's known that type B mice are related to type A mice. The previous prior implies there's no connection between type A and B.

(d)

The test quantity is

$$T(y, \theta) = \frac{\bar{\mathbf{y}}_A}{sd(\bar{\mathbf{y}}_A)}$$

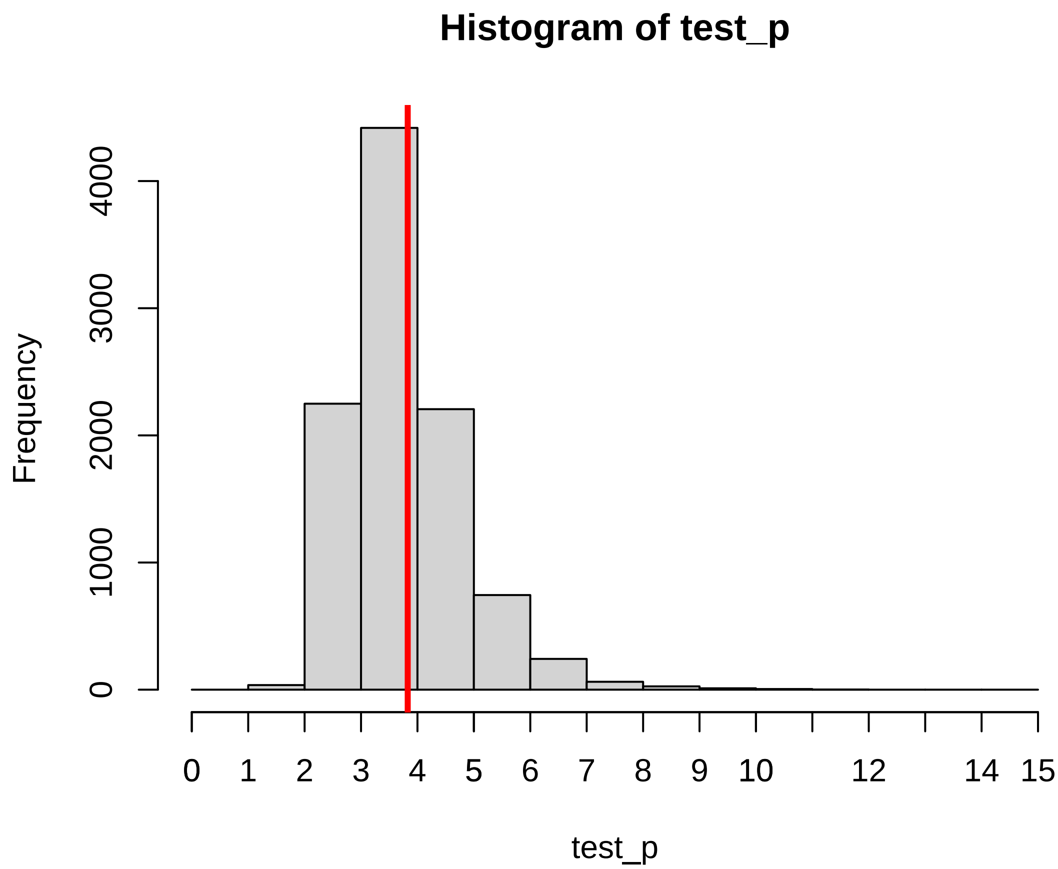
We need to calculate this value in the replicated data.

```r
#set.seed(2304933)
S <- 10000

post_test <- function(a,b,n){
  #sample posterior theta
  theta_p <- rgamma(S,a,b)
```

```r
  #sample posterior y, one row is one replicated data
  y_p <- matrix(rpois(n*S, rep(theta_p, n)), ncol=n)

  #calculate posterior test quantity
  test_p <- rowMeans(y_p)/apply(y_p, 1, sd)

  return (test_p)
}

test_p <- post_test(237, 20, 10)

#observed test quantity
test_obs <- mean(ya)/sd(ya)
#calculate bayesian p-value
hat_p_value <- mean(test_p>=test_obs)
print(hat_p_value)
```

```
## [1] 0.3915
```

```r
hist(test_p, breaks=seq(0, 15), cex=1, xlim=c(0,15))
axis(side = 1, at = seq(0,15))
abline(v=test_obs, col='red', lwd=3)
```



From the Bayesian p-value (0.4045, not extreme) and the distribution of test quantity, we conclude that using Poisson sampling model is appropriate and adequate.

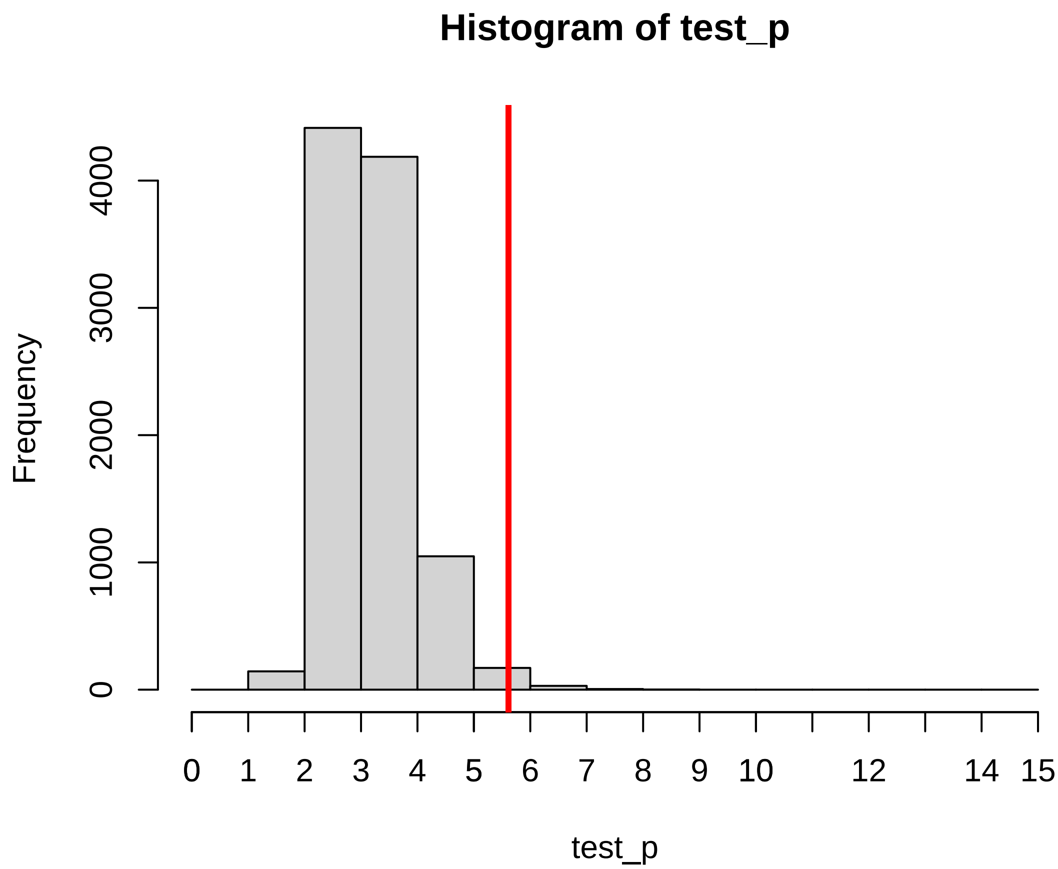
(e)

Similarly

```
test_p <- post_test(125, 14, 13)

#observed test quantity
test_obs <- mean(yb)/sd(yb)
#calculate bayesian p-value
hat_p_value <- mean(test_p>=test_obs)
print(hat_p_value)
```

```
## [1] 0.008
```

```
hist(test_p, breaks=seq(0, 15), cex=1, xlim=c(0,15))
axis(side = 1, at = seq(0,15))
abline(v=test_obs, col='red', lwd=3)
```



Note that the Bayesian p-value is extreme (0.0074), which shows that the Poisson sampling model and the prior of type B is not adequate.

Problem 3

(a)

Prior

$$\theta \sim \Gamma(\alpha, \beta) \propto \theta^{\alpha-1} e^{-\beta\theta}$$

Likelihood

$$y_i|\theta \sim \mathcal{E}(\theta) \propto \theta e^{-\theta y_i}, y_i \text{ i.i.d.}$$
$$\to p(\vec{y}|\theta) \propto \prod_i \theta e^{-\theta y_i} = \theta^n e^{-\theta \sum_i y_i} = \theta^n e^{-\theta \cdot n\bar{y}}$$

Posterior

$$p(\theta|\vec{y}) \propto p(\vec{y}|\theta)p(\theta) \propto \theta^n e^{-\theta \cdot n\bar{y}} \theta^{\alpha-1} e^{-\beta\theta} \propto \theta^{\alpha+n-1} e^{-(\beta+n\bar{y})\theta}$$
$$\sim \Gamma(\alpha + n, \beta + n\bar{y})$$

Since the posterior distribution of θ is still gamma distribution, which is in the same probability distribution family as the prior $p(\theta)$, by definition, the gamma prior distribution is conjugate for exponential distribution likelihood.

(b)

For $\Gamma(\alpha, \beta)$, the coefficient of variation is

$$\frac{\sqrt{\alpha/\beta^2}}{\alpha/\beta} = \frac{1}{\sqrt{\alpha}}$$

It's known that this value is 0.5, so for the prior distribution

$$\frac{1}{\sqrt{\alpha}} = 0.5 \rightarrow \alpha = 4$$

In the posterior distribution of θ, the coefficient of variation is

$$\frac{\sqrt{\alpha + n/(\beta + +n\bar{y})^2}}{\alpha + n/(\beta + n\bar{y})} = \frac{1}{\sqrt{\alpha + n}} = 0.1$$

Therefore

$$n = 96$$

So, 96 light bulbs need to be tested.

(c)

We need to consider the prior and posterior distribution of $\phi = 1/\theta$. Since the prior of θ is $\Gamma(\alpha, \beta)$, the prior of ϕ should be

$$\phi \sim \text{Inv-}\Gamma(\alpha, \beta)$$

the coefficient of variation is

$$\frac{\sqrt{\beta^2/((\alpha - 1)^2(\alpha - 2))}}{\beta/(\alpha - 1)} = 0.5 \rightarrow \alpha = 6$$

Similarly, the posterior of ϕ should be

$$\phi|\vec{y} \sim \text{Inv-}\Gamma(\alpha + n, \beta + n\bar{y})$$

so

$$\frac{\sqrt{(\beta + n\bar{y})^2/((\alpha + n - 1)^2(\alpha + n - 2))}}{(\beta + n\bar{y})/(\alpha + n - 1)} = 0.1 \rightarrow n = 96$$

The answer doesn't change.

(d)

Prior

$$\theta \sim \Gamma(\alpha, \beta) \propto \theta^{\alpha - 1} e^{-\beta\theta}$$

Data

$$y|\theta \sim \mathcal{E}(\theta), y \geq 100, \text{ and no exact value}$$

Since we don't know the exact value of y, so now the condition of posterior becomes $y \geq 100$,i.e.

$$\begin{aligned}
p(\theta|y \geq 100) \propto p(y \geq 100|\theta)p(\theta) &\propto \left(\int_{100}^{+\infty} \theta e^{-\theta y} dy\right) \cdot \theta^{\alpha-1} e^{-\beta\theta} \\
&\propto e^{-100\theta}\theta^{\alpha-1}e^{-\beta\theta} \\
&\propto \theta^{\alpha-1}e^{-(\beta+100)\theta} \\
&\sim \Gamma(\alpha, \beta+100)
\end{aligned}$$

Therefore, posterior mean and variance is

$$E(\theta|y \geq 100) = \frac{\alpha}{\beta+100}, \quad Var(\theta|y \geq 100) = \frac{\alpha}{(\beta+100)^2}$$

(e)

By part (a), when $y = 100$, posterior distribution of θ is

$$\theta|y = 100 \sim \Gamma(\alpha+1, \beta+100)$$

Posterior mean and variance is

$$E(\theta|y = 100) = \frac{\alpha+1}{\beta+100}, \quad Var(\theta|y = 100) = \frac{\alpha+1}{(\beta+100)^2}$$

We find that this posterior variance is higher than that in part (d).

As for the reason, by the law of total variance

$$Var(\theta) = Var(E(\theta|Y)) + E(Var(\theta|Y))$$

thus

$$Var(\theta) \geq Var(E(\theta|Y)), Var(\theta) \geq E(Var(\theta|Y))$$

Note that we can apply these inequalities to conditional variance, i.e.

$$Var(\theta|y \geq 100) \geq E(Var(\theta|y)|y \geq 100)$$

This inequality implies that, **on average**, the variance of θ decreases as we acquire more information.

However, it's important to note that simply taking $y = 100$ on the right-hand side is different from averaging over the distribution $y|y \geq 100$.

In this particular case, the increase in variance may be attributed to a numerical coincidence rather than a natural phenomenon. For instance, in (e), if we are told that $y = 100000$, it's highly probable that the variance would decrease, as the term $(\beta + 100000)^2$ significantly influences the comparison between $\alpha/(\beta + 100)^2$ and $(\alpha + 1)/(\beta + 100000)^2$.