Discussion 3 Solution

Huaiyu

September 25, 2019

```
library(ggplot2)
```

Exercise

Tumor counts: A cancer laboratory is estimating the rate of tumorigenesis in two strains of mice, A and B. They have tumor count data for 10 mice in strain A and 13 mice in strain B. Type A mice have been well studied and information from other laboratories suggests that type A mice have tumor counts that are approximately Poisson-distributed with a mean of 12. Tumor count rates for type B mice are unknown, but type B mice are related to type A mice. The observed tumor counts for the two populations are $y_A = (12, 9, 12, 14, 13, 13, 15, 8, 15, 6)$ and $y_B = (11, 11, 10, 9, 9, 8, 7, 10, 6, 8, 8, 9, 7)$.

(1) Find the posterior distributions, means, variances and 95% quantile based confidence intervals for θ_A and θ_B. Assuming a Poisson sampling distribution for each group and the following prior distribution:

$\theta_A \sim gamma(120, 10), \theta_B \sim gamma(12, 1), p(\theta_A, \theta_B) = p(\theta_A) \times p(\theta_B)$

```
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)
sya <- sum(ya)
syb <- sum(yb)
na <- length(ya)
nb <- length(yb)
c(sya, syb, na, nb)
```

```
## [1] 117 113  10  13
```

$$\theta_A \mid \mathbf{y}_a \sim Gamma(120 + 117, 10 + 10) = Gamma(237, 20)$$

$$\theta_B \mid \mathbf{y}_b \sim Gamma(12 + 113, 1 + 13) = Gamma(125, 14)$$

$$\mathbb{E}(\theta_A) = 237/20 = 11.85, \mathbb{E}(\theta_B) = 125/14 = 8.92$$

$$\text{Var}(\theta_A) = 237/400 = 0.593, \text{Var}(\theta_B) = 125/196 = 0.638$$

95% quantile-based confidence intervals can be solved by setting the CDF of the Gammas to p, and solving for θ. Alternatively,

```
qgamma(c(0.025, 0.975), 237, 20)
```

```
## [1] 10.38924 13.40545
```

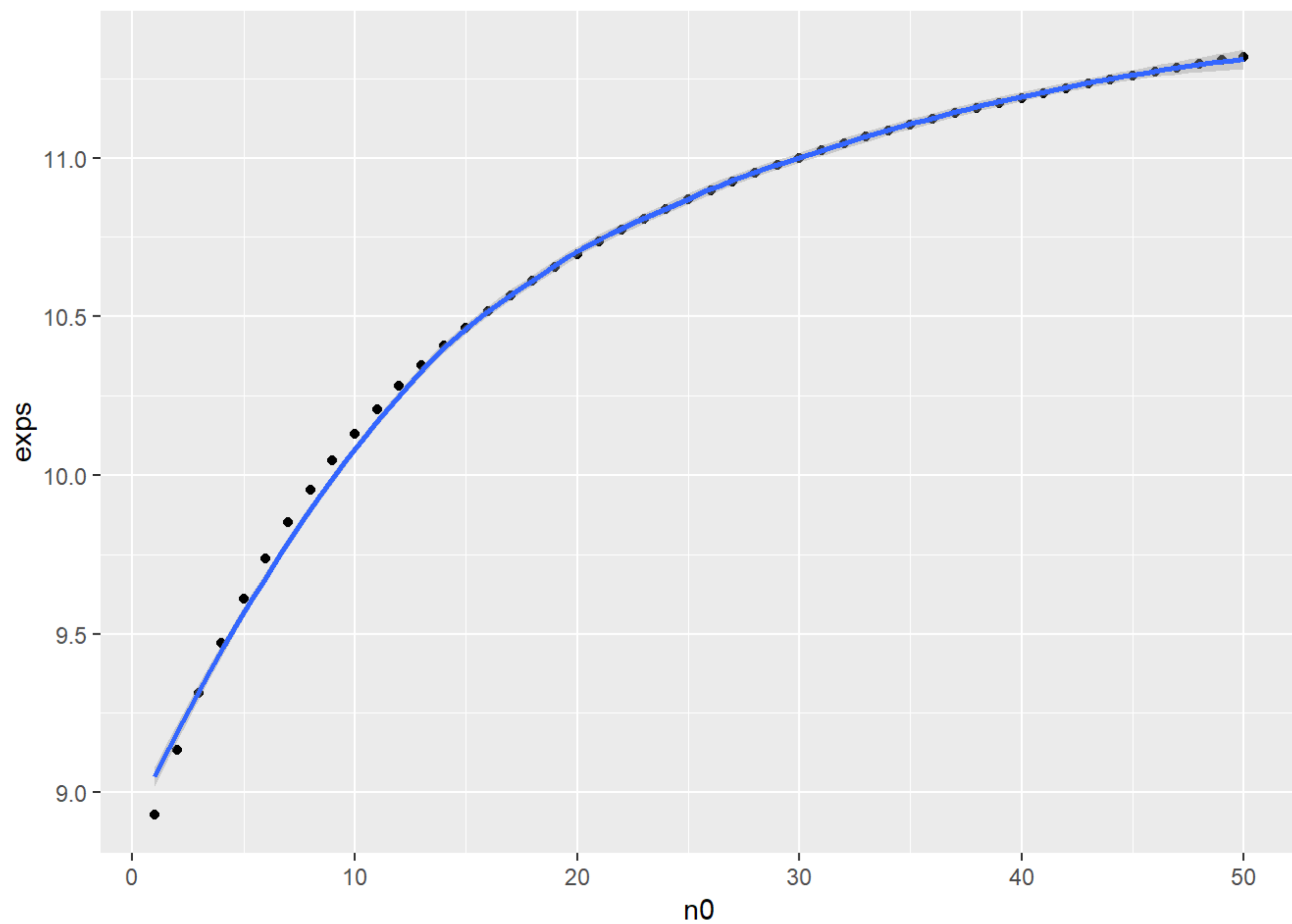
```
qgamma(c(0.025, 0.975), 125, 14)
```

```
## [1]  7.432064 10.560308
```

(2). Compute and plot the posterior expectation of θ_B under the prior distribution $\theta_B \sim gamma(12 \times n_0, n_0)$ for each value of $n_0 \in 1, 2, \ldots, 50$. Describe what sort of prior beliefs about θ_B would be necessary in order for the posterior expectation of θ_B to be closed to that of θ_A.

```
n0 <- 1:50
exps <- (12 * n0 + sum(yb)) / (n0 + length(yb))
qplot(n0, exps, geom = c('point', 'smooth'))
```

```
## `geom_smooth()` using method = 'loess' and formula 'y ~ x'
```



According to the graph, n_0 values should be closed to 50.

(3). For the prior distribution given in part (1), obtain $Pr(\theta_B < \theta_A | y_A, y_B)$ via Monte Carlo sampling.

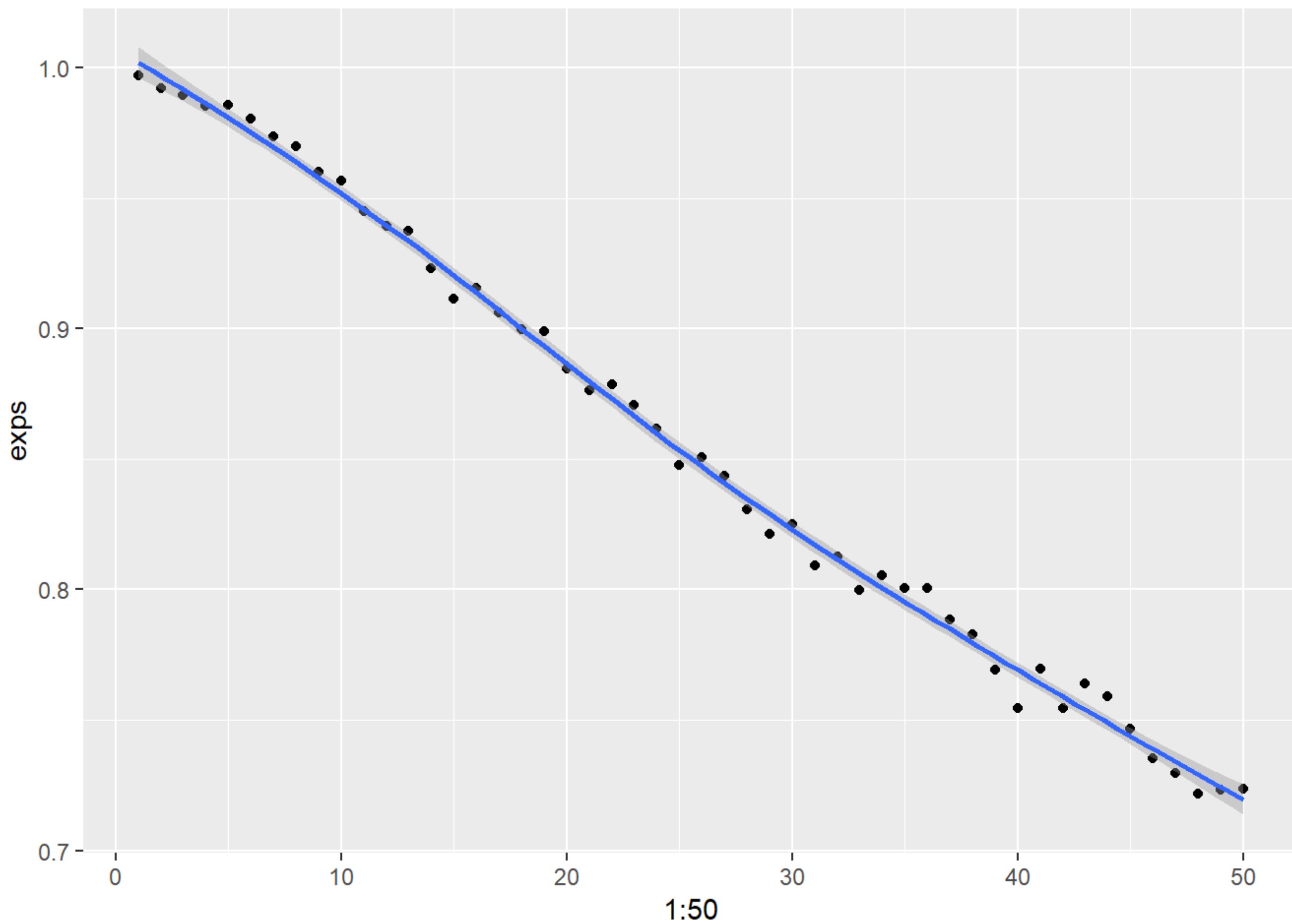
```
theta.a = rgamma(5000, 237, 20)
theta.b = rgamma(5000, 125, 14)
mean(theta.b < theta.a)
```

```
## [1] 0.9948
```

(4). For a range of values of n_0, obtain $Pr(\theta_B < \theta_A | y_A, y_B)$ for $\theta_A \sim gamma(120, 10)$ and $\theta_B \sim gamma(12 \times n_0, n_0)$. Describe how sensitive the conclusions about the event $\{\theta_B < \theta_A\}$ are to the prior distribution on θ_B.

```
exps <- sapply(1:50, function(n0) {
  mean(rgamma(5000, (12 * n0) + 113, n0 + 13) < rgamma(5000, 237, 20))
})
qplot(1:50, exps, geom = c('point', 'smooth'))
```

```
## `geom_smooth()` using method = 'loess' and formula 'y ~ x'
```



Over the values of $n_0 \in 1, 2, \ldots, 50$, the probability decreses with n_0 and the lowest probability is 0.7268 when $n_0 = 50$.

(5). Find the $Pr(\tilde{y}_B < \tilde{y}_A)$ given prior in part (1).

```
N <- 10000
thetaa.mc <- rgamma(N, 237, 20)
thetab.mc <-rgamma(N, 125, 14)
ya.mc <- rpois(N, thetaa.mc)
yb.mc <- rpois(N, thetab.mc)
mean(ya.mc > yb.mc)
```

```
## [1] 0.6984
```

Reference. A First Course in Bayesian Statistical Methods. Peter D. Hoff. Chaper 3-4. Springer.