Assignment 7

Sam Olson

The Assignment

Your assignment is to conduct a Bayesian analyis of relative risk for these data. You will do so using both the Multinomial probability model and the Poisson probability model.

1. Derive Posterior Distributions

Using prior distributions and prior parameter values as given in Section 4 (Bayesian Analysis), derive the joint posterior distributions for $\boldsymbol{\theta} = (\theta_{11}, \theta_{10}, \theta_{01}, \theta_{00})$ in the multinomial model, and $\boldsymbol{\lambda} = (\lambda_{11}, \lambda_{10}, \lambda_{01}, \lambda_{00})$ in the Poisson model.

Answer

Multinomial–Dirichlet model

The likelihood is multinomial with parameters
$\mathbf{y} = (y_{11}, y_{10}, y_{01}, y_{00})$ and probabilities
$\boldsymbol{\theta} = (\theta_{11}, \theta_{10}, \theta_{01}, \theta_{00})$:

$$f(\mathbf{y} \mid \boldsymbol{\theta}) \propto \theta_{11}^{y_{11}} \theta_{10}^{y_{10}} \theta_{01}^{y_{01}} \theta_{00}^{y_{00}}, \quad \sum_{i,j} \theta_{ij} = 1.$$

Given a Dirichlet prior
$\boldsymbol{\theta} \sim \text{Dirichlet}(\gamma_{11}, \gamma_{10}, \gamma_{01}, \gamma_{00})$ with
$\gamma_{11} = 0.10, \ \gamma_{10} = 0.75, \ \gamma_{01} = 0.10, \ \gamma_{00} = 0.75$, conjugacy gives the posterior:

$$\boldsymbol{\theta} \mid \mathbf{y} \ \sim \ \text{Dirichlet}(\gamma_{11} + y_{11}, \gamma_{10} + y_{10}, \gamma_{01} + y_{01}, \gamma_{00} + y_{00}).$$

Poisson–Gamma model

Assume independent $Y_{ij} \sim \text{Poisson}(\lambda_{ij})$ with priors
$\lambda_{ij} \sim \text{Gamma}(\alpha_{ij}, \beta_{ij})$ (shape–rate form):

$$(\alpha_{ij}, \beta_{ij}) = \begin{cases} (0.5, 5), & ij = 11 \\ (3.0, 4), & ij = 10 \\ (0.5, 5), & ij = 01 \\ (3.0, 4), & ij = 00 \end{cases}$$

The posterior distributions are

$$\lambda_{ij} \mid y_{ij} \sim \text{Gamma}(\alpha_{ij} + y_{ij}, \beta_{ij} + 1).$$

2. Posterior Expected Values

Using the data of Table 2 give posterior expected values for θ and λ, and the cell probabilities that would correspond to these means.

Answer

Multinomial–Dirichlet posterior means

Let $G = \sum \gamma_{ij} = 1.70$ and $N + G = 153.7$.
Posterior means:

$$E[\theta_{11} \mid y] = \frac{15.10}{153.7} = 0.0982,$$

$$E[\theta_{10} \mid y] = \frac{50.75}{153.7} = 0.3302,$$

$$E[\theta_{01} \mid y] = \frac{3.10}{153.7} = 0.0202,$$

$$E[\theta_{00} \mid y] = \frac{84.75}{153.7} = 0.5514.$$

Poisson–Gamma posterior means

$$E[\lambda_{ij} \mid y] = \frac{\alpha_{ij} + y_{ij}}{\beta_{ij} + 1}.$$

$$E[\lambda_{11}] = 15.5/6 = 2.5833,$$
$$E[\lambda_{10}] = 53/5 = 10.6,$$
$$E[\lambda_{01}] = 3.5/6 = 0.5833,$$
$$E[\lambda_{00}] = 87/5 = 17.4.$$

Normalize to probabilities:

$$\tilde{\theta}_{ij} = \frac{E[\lambda_{ij} \mid y]}{\sum E[\lambda_{ij} \mid y]}.$$

$$\tilde{\theta}_{11} = 0.0829, \quad \tilde{\theta}_{10} = 0.3401, \quad \tilde{\theta}_{01} = 0.0187, \quad \tilde{\theta}_{00} = 0.5583.$$

3. Evaluate Relative Risk

Evaluate relative risk RR at the expected values of the posterior distributions you derived in exercise 1 and compare to the observed or sample-based RR based only on the observed cell counts. Briefly explain why this is not the posterior expected value of RR (although we hope it is not far off).

Answer

The relative risk is

$$RR = \frac{\Pr(D \mid R)}{\Pr(D \mid R^c)} = \frac{\theta_{11}(\theta_{01} + \theta_{00})}{\theta_{01}(\theta_{11} + \theta_{10})}.$$

Sample estimate (Eq. 2)

$$\widehat{RR} = \frac{15(3 + 84)}{3(15 + 50)} = \frac{1305}{195} = 6.69.$$

Posterior plug-in values

$$RR(E[\theta]) = 6.50, \qquad RR(\tilde{\theta}) = 6.04.$$

Interpretation

Because RR is nonlinear in $\boldsymbol{\theta}$,
$$E[RR \mid y] \neq RR(E[\boldsymbol{\theta} \mid y]),$$
so these plug-in values are not the true posterior means of RR, though they are typically close.

Under the Dirichlet posterior, the theta values are dependent (they must sum to 1), and RR involves both sums and ratios of these components. Expectations do not commute through such ratios/sums. So, predominantly due to Jensen's, they will not be equal unless they have zero variance.

4. Posterior Distributions of RR

Find the posterior distributions of RR under both the multinomial and Poisson models. Give summary and 95% credible intervals. Produce some type of a graphical display of the posterior distributions. How do the results compare for the multinomial and Poisson models? How do the results compare to the initial sample version of RR in expression (2)? What would you conclude about the relation between time since vaccination and the chances of getting chickenpox in this population of school children? For example, what is your posterior probability that the risk factor of having more than 5 years since vaccination is positively related to the chance of contracting chickenpox in these students?

Answer

```
set.seed(1)
library(gtools)

# Data
```

```r
y <- c(15, 50, 3, 84)
g <- c(0.10, 0.75, 0.10, 0.75)
a <- c(0.5, 3.0, 0.5, 3.0)
b <- c(5, 4, 5, 4)
S <- 3e5

# Multinomial-Dirichlet draws
theta <- rdirichlet(S, g + y)
RR_dir <- theta[,1]*(theta[,3]+theta[,4]) / (theta[,3]*(theta[,1]+theta[,2]))

# Poisson-Gamma draws
lam <- cbind(
  rgamma(S, a[1]+y[1], b[1]+1),
  rgamma(S, a[2]+y[2], b[2]+1),
  rgamma(S, a[3]+y[3], b[3]+1),
  rgamma(S, a[4]+y[4], b[4]+1)
)
p <- lam / rowSums(lam)
RR_pois <- p[,1]*(p[,3]+p[,4]) / (p[,3]*(p[,1]+p[,2]))
```

```r
library(knitr)
library(kableExtra)

summarize_rr <- function(x) {
  c(
    Mean      = mean(x),
    Median    = median(x),
    SD        = sd(x),
    `2.5%`    = unname(quantile(x, 0.025)),
    `97.5%`   = unname(quantile(x, 0.975)),
    `Pr(RR>1)` = mean(x > 1),
    `Pr(RR>5)` = mean(x > 5)
  )
}

tab <- rbind(
  `Multinomial-Dirichlet` = summarize_rr(RR_dir),
  `Poisson-Gamma`         = summarize_rr(RR_pois)
)

# Turn into a data.frame and move model names into a column
tab_df <- as.data.frame(tab)
tab_df$Model <- rownames(tab_df)
tab_df <- tab_df[, c("Model", setdiff(colnames(tab_df), "Model"))]

# --- IMPORTANT: drop rownames so kable doesn't add an extra column ---
rownames(tab_df) <- NULL

# Build table
kable(
  tab_df,
  format  = if (knitr::is_latex_output()) "latex" else "html",
  booktabs = TRUE,
```

```
  row.names = FALSE,
  digits    = c(0, rep(2, ncol(tab_df) - 1)),
  caption   = "Posterior summaries for RR under two Bayesian models",
  align     = c("l", rep("r", ncol(tab_df) - 1)),
  format.args = list(big.mark = ",")
) |>
  kable_styling(full_width = FALSE, position = "center",
                latex_options = c("hold_position","striped")) |>
  column_spec(1, bold = TRUE) |>
  add_header_above(c(" " = 1, "Posterior Summary" = ncol(tab_df) - 1))
```

Table 1: Posterior summaries for RR under two Bayesian models

| Model	Posterior Summary						
	Mean	Median	SD	2.5%	97.5%	Pr(RR>1)	Pr(RR>5)
Multinomial–Dirichlet	9.49	7.08	9.51	2.4	31.06	1	0.72
Poisson–Gamma	8.39	6.50	7.38	2.3	25.83	1	0.68

```
# Sample RR line
RR_hat <- 15*(3+84) / (3*(15+50))

par(mfrow = c(1,2))
# Define the range you want to cover
xmax <- ceiling(max(RR_dir, RR_pois))  # round up to next integer

# Dirichlet-Multinomial RR histogram with bin width 1
hist(RR_dir,
     breaks = seq(0, xmax, by = 1),   # ← each bin is 1 wide
     freq = FALSE,
     main = "RR | Multinomial-Dirichlet",
     xlab = "RR",
     xlim = c(0, 80))
abline(v = RR_hat, lty = 2)

# Poisson-Gamma RR histogram with bin width 1
hist(RR_pois,
     breaks = seq(0, xmax, by = 1),   # ← each bin is 1 wide
     freq = FALSE,
     main = "RR | Poisson-Gamma",
     xlab = "RR",
     xlim = c(0, 80))
abline(v = RR_hat, lty = 2)
```

RR | Multinomial–Dirichlet

RR | Poisson–Gamma
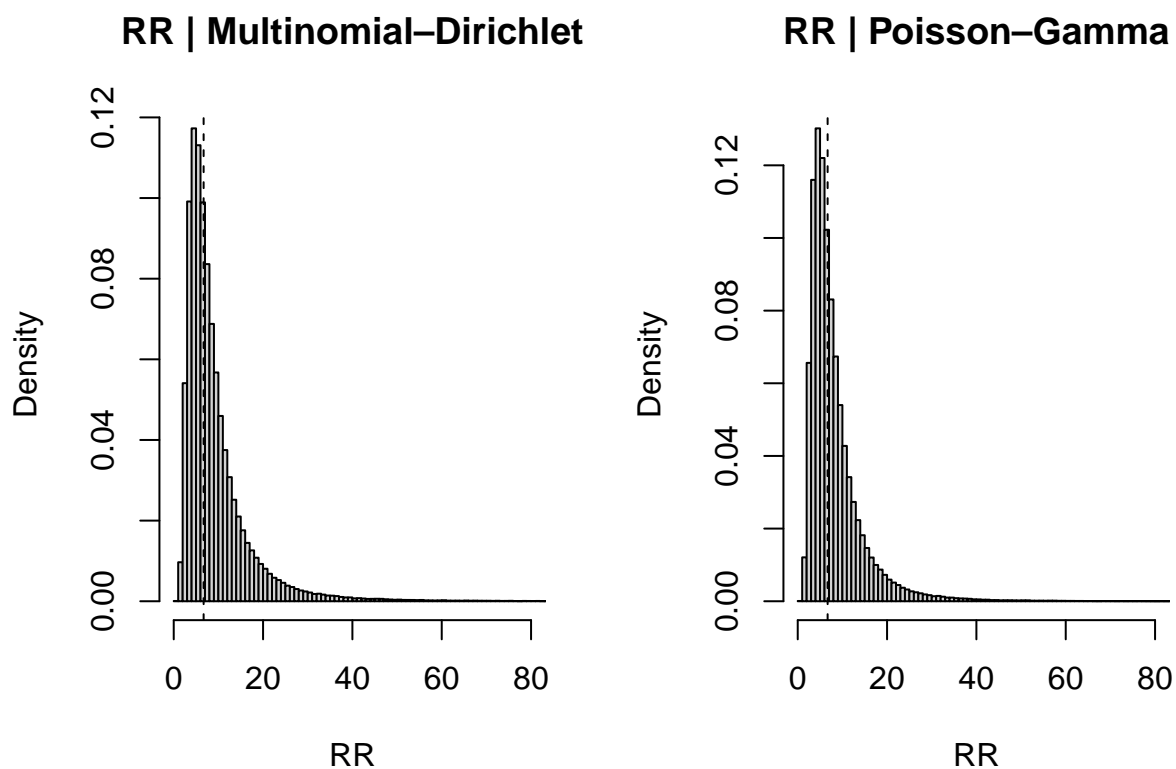

```
breaks <- seq(0, xmax, by = 1)

# Compute density estimates
d_dir  <- density(RR_dir, from = 0, to = 40)
d_pois <- density(RR_pois, from = 0, to = 40)

# Plot both densities overlaid
plot(d_dir, type = "l", lwd = 2, col = "steelblue",
     xlim = c(0, 30), ylim = c(0, max(d_dir$y, d_pois$y)),
     main = "Posterior Distributions of RR",
     xlab = "Relative Risk (RR)")
lines(d_pois, lwd = 2, col = "firebrick")
abline(v = RR_hat, lty = 2, col = "black")

legend("topright",
       legend = c("Multinomial-Dirichlet", "Poisson-Gamma", "Sample RR"),
       col = c("steelblue", "firebrick", "black"),
       lwd = c(2, 2, 1), lty = c(1, 1, 2), bty = "n")
```

Posterior Distributions of RR

Legend:
- Multinomial–Dirichlet
- Poisson–Gamma
- Sample RR