

Homework 5

STT 465, Bayesian Statistical Methods

Lowell Monis

October 29, 2025

Question 1

The files `menchild30bach.txt` and `menchild30nobach.txt` contain data on the number of children of men in their 30s with and without bachelor's degrees, respectively. Download the files onto your computer, import the data into R using `yA <- scan("menchild30bach.txt"); yB <- scan("menchild30nobach.txt")`.

```
yA <- scan("data/menchild30bach.txt"); yB <- scan("data/menchild30nobach.txt")
```

Denote the two groups of data by $y^A = (y_1^A, \dots, y_{n_A}^A)$ and $y^B = (y_1^B, \dots, y_{n_B}^B)$, respectively. We assume Poisson sampling models: $Y_i^A \sim \text{Poisson}(\theta_A)$, $Y_i^B \sim \text{Poisson}(\theta_B)$.

But now we parameterize θ_A and θ_B as $\theta_A = \theta$, $\theta_B = \theta \times \gamma$. In this parameterization, γ represents the relative rate $\frac{\theta_B}{\theta_A}$. We consider the independent prior for θ and γ : $\theta \sim \text{Gamma}(a_\theta, b_\theta)$ and $\gamma \sim \text{Gamma}(a_\gamma, b_\gamma)$. We now use Gibbs sampler to approximate the posterior distribution $p(\theta, \gamma \mid y^A, y^B)$.

(a) 16 points

Show that the full conditional distribution of θ given y^A, y^B, γ takes the form:

$$p(\theta \mid y^A, y^B, \gamma) = \text{Gamma}(\tilde{a}_\theta, \tilde{b}_\theta)$$

where

$$\begin{aligned}\tilde{a}_\theta &= a_\theta + \sum_{i=1}^{n_A} y_i^A + \sum_{i=1}^{n_B} y_i^B \\ \tilde{b}_\theta &= b_\theta + n_A + n_B \cdot \gamma\end{aligned}$$

Answer

We need to find the full condition distribution $p(\theta \mid y^A, y^B, \gamma)$. Using Bayes' theorem, we know that:

$$p(\theta \mid y^A, y^B, \gamma) \propto p(y^A \mid \theta) \cdot p(y^B \mid \theta, \gamma) \cdot p(\theta)$$

$p(\gamma)$ has been absorbed into the proportionality since it does not depend on θ . The likelihood of y^A is given by:

$$y_i^A \sim \text{Poisson}(\theta_A) = \text{Poisson}(\theta)$$

We can reduce $\theta_A = \theta$ from the given conditions. Further, from the Poisson distribution,

$$p(y^A | \theta) = \prod_{i=1}^{n_A} \frac{e^{-\theta} \theta^{y_i^A}}{y_i^A!} \propto e^{-n_A \theta} \theta^{\sum_{i=1}^{n_A} y_i^A}$$

Since we are looking to derive the condition distribution for θ , terms that do not depend on θ are treated as constants and absorbed into the proportionality.

Similarly, the likelihood of θ^B is given by $y_i^B \sim \text{Poisson}(\theta \gamma) = \text{Poisson}(\theta \gamma)$, and

$$p(y^B | \theta, \gamma) = \prod_{i=1}^{n_B} \frac{e^{-\theta \gamma} (\theta \gamma)^{y_i^B}}{y_i^B!} \propto e^{-n_B \theta \gamma} (\theta \gamma)^{\sum_{i=1}^{n_B} y_i^B} \propto e^{-n_B \theta \gamma} \theta^{\sum_{i=1}^{n_B} y_i^B} \gamma^{\sum_{i=1}^{n_B} y_i^B} \propto e^{-n_B \theta \gamma} \theta^{\sum_{i=1}^{n_B} y_i^B}$$

The prior for θ is given by $\theta \sim \text{Gamma}(a_\theta, b_\theta)$. From this, we get:

$$p(\theta) = \frac{b_\theta^{a_\theta}}{\Gamma(a_\theta)} \theta^{a_\theta-1} e^{-b_\theta \theta} \propto \theta^{a_\theta-1} e^{-b_\theta \theta}$$

Now, we can combine the proportionality components into the formula received via the Bayes' theorem.

$$\begin{aligned} p(\theta | y^A, y^B, \gamma) &\propto p(y^A | \theta) \cdot p(y^B | \theta, \gamma) \cdot p(\theta) \\ p(\theta | y^A, y^B, \gamma) &\propto e^{-n_A \theta} \theta^{\sum_{i=1}^{n_A} y_i^A} \cdot e^{-n_B \theta \gamma} \theta^{\sum_{i=1}^{n_B} y_i^B} \cdot \theta^{a_\theta-1} e^{-b_\theta \theta} \end{aligned}$$

$$p(\theta | y^A, y^B, \gamma) \propto e^{-\theta(n_A + n_B \gamma + b_\theta)} \cdot \theta^{a_\theta-1 + \sum_{i=1}^{n_A} y_i^A + \sum_{i=1}^{n_B} y_i^B}$$

The combined proportionality is of the form $e^{-\theta \cdot \text{rate}} \theta^{\text{shape}-1}$, which are the parameters of a Gamma distribution.

Thus, the full conditional distribution is:

$$\theta | y^A, y^B, \gamma \sim \text{Gamma}(\tilde{a}_\theta, \tilde{b}_\theta)$$

From the proportionality we received by combining, the updated shape is

$$\tilde{a}_\theta = a_\theta + \sum_{i=1}^{n_A} y_i^A + \sum_{i=1}^{n_B} y_i^B$$

and the updated rate is

$$\tilde{b}_\theta = b_\theta + n_A + n_B \gamma$$

Hence, we have verified the claim. ■

(b) 16 points

Show that the full conditional distribution of γ given y^A, y^B, θ takes the form:

$$p(\gamma | y^A, y^B, \theta) = \text{Gamma}(\tilde{a}_\gamma, \tilde{b}_\gamma)$$

where

$$\begin{aligned} \tilde{a}_\gamma &= a_\gamma + \sum_{i=1}^{n_B} y_i^B \\ \tilde{b}_\gamma &= b_\gamma + n_B \cdot \theta \end{aligned}$$

Answer

Now, we essentially repeat the process in (a), but for γ . By Bayes' theorem, the proportionality we can use is:

$$p(\gamma \mid y^A, y^B, \theta) \propto p(y^B \mid \theta, \gamma) \cdot p(\gamma)$$

Similarly, $p(\theta)$ and $p(y^A \mid \theta)$ have been absorbed into the proportionality since they do not depend on γ and are treated like constants. The likelihood of y^B is given by $y_i^B \sim \text{Poisson}(\theta_B) = \text{Poisson}(\theta\gamma)$. I will be simplifying my working here as compared to (a), since this follows the same absorption of constants not dependent on γ into the proportionality:

$$p(y^B \mid \theta, \gamma) = \prod_{i=1}^{n_B} \frac{e^{-\theta\gamma} (\theta\gamma)^{y_i^B}}{y_i^B!} \propto e^{-n_B\theta\gamma} \gamma^{\sum_{i=1}^{n_B} y_i^B}$$

The prior for γ is given by $\gamma \sim \text{Gamma}(a_\gamma, b_\gamma)$, which gives us:

$$p(\gamma) = \frac{b_\gamma^{a_\gamma}}{\Gamma(a_\gamma)} \gamma^{a_\gamma-1} e^{-b_\gamma\gamma} \propto \gamma^{a_\gamma-1} e^{-b_\gamma\gamma}$$

We now proceed with combining the components of the original proportionality given by Bayes' theorem:

$$\begin{aligned} p(\gamma \mid y^A, y^B, \theta) &\propto p(y^B \mid \theta, \gamma) \cdot p(\gamma) \\ p(\gamma \mid y^A, y^B, \theta) &\propto e^{-n_B\theta\gamma} \gamma^{\sum_{i=1}^{n_B} y_i^B} \cdot \gamma^{a_\gamma-1} e^{-b_\gamma\gamma} \\ p(\gamma \mid y^A, y^B, \theta) &\propto e^{-\gamma(n_B\theta+b_\gamma)} \cdot \gamma^{a_\gamma-1+\sum_{i=1}^{n_B} y_i^B} \end{aligned}$$

Similarly as (a), this is a Gamma distribution:

$$\gamma \mid y^A, y^B, \theta \sim \text{Gamma}(\tilde{a}_\gamma, \tilde{b}_\gamma)$$

Here, the updated shape is

$$\tilde{a}_\gamma = a_\gamma + \sum_{i=1}^{n_B} y_i^B$$

and the updated rate is

$$\tilde{b}_\gamma = n_B\theta + b_\gamma$$

Hence, we have verified the claim. ■

(c) 10 points

Use the number of children data, and set the parameter values in the prior: $a_\theta = 2$, $b_\theta = 1$, $a_\gamma = b_\gamma = 8$. Run a Gibbs sampler of 5000 iterations.

Make the traceplots of θ and γ . What do you observe?

Further plot the autocorrelation functions (from lag-1 to lag-40) for θ and γ (using the R function `acf` to compute the autocorrelations). What do you observe?

Based on the traceplots and autocorrelation function plots, describe how you use the 5000 samples from Gibbs sampler to approximate $E(\theta_B - \theta_A \mid y^A, y^B)$.

Answer

We first create variables to store the given information from the question.

```
# required descriptive information
nA <- length(yA)
nB <- length(yB)
sA <- sum(yA)
sB <- sum(yB)

# prior parameters
a_theta <- 2
b_theta <- 1
a_gamma <- 8 -> b_gamma
```

We can now set up the Markov Chain Monte Carlo experiment with 5,000 iterations.

```
S <- 5000 # iterations

# storage
THETA <- numeric(S) -> GAMMA

# prior mean as initial value
THETA[1] <- a_theta/b_theta
GAMMA[1] <- a_gamma/b_gamma
```

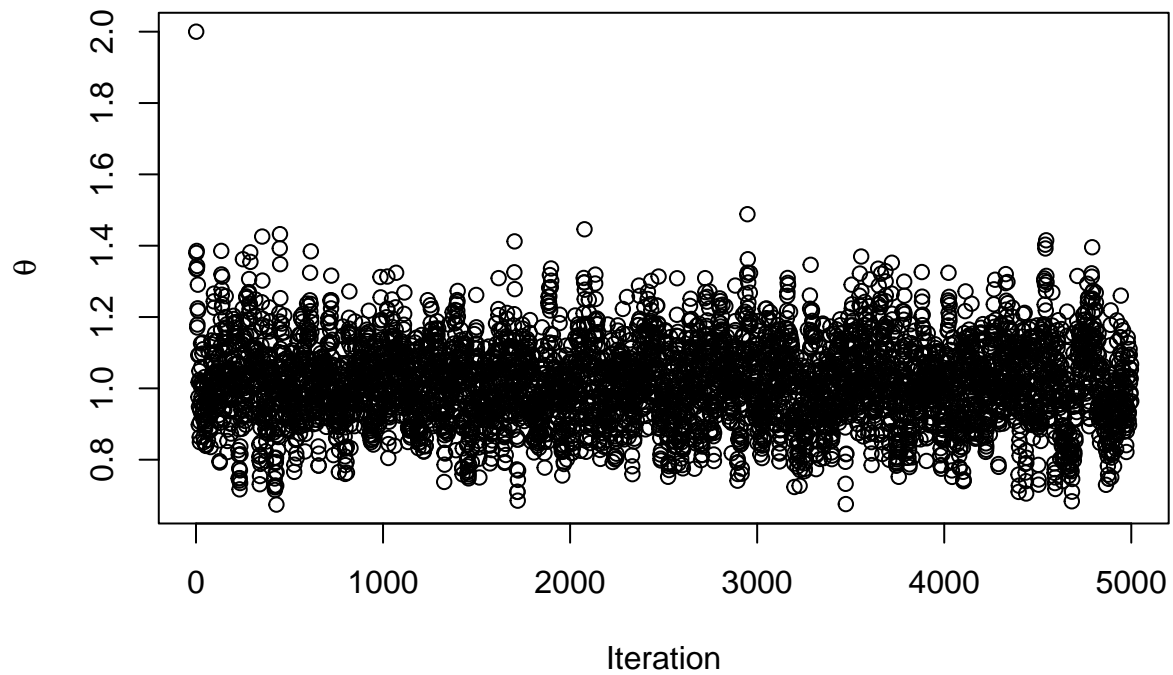
We can now set up the Gibbs sampler implementation.

```
set.seed(465)
for (k in 2:S){
  THETA[k] <- rgamma(1, shape = a_theta + sA + sB,
                    rate = b_theta + nA + nB*GAMMA[k-1])
  GAMMA[k] <- rgamma(1, shape = a_gamma + sB,
                    rate = b_gamma + nB * THETA[k])
}
```

We can now proceed with MCMC diagnostic plots.

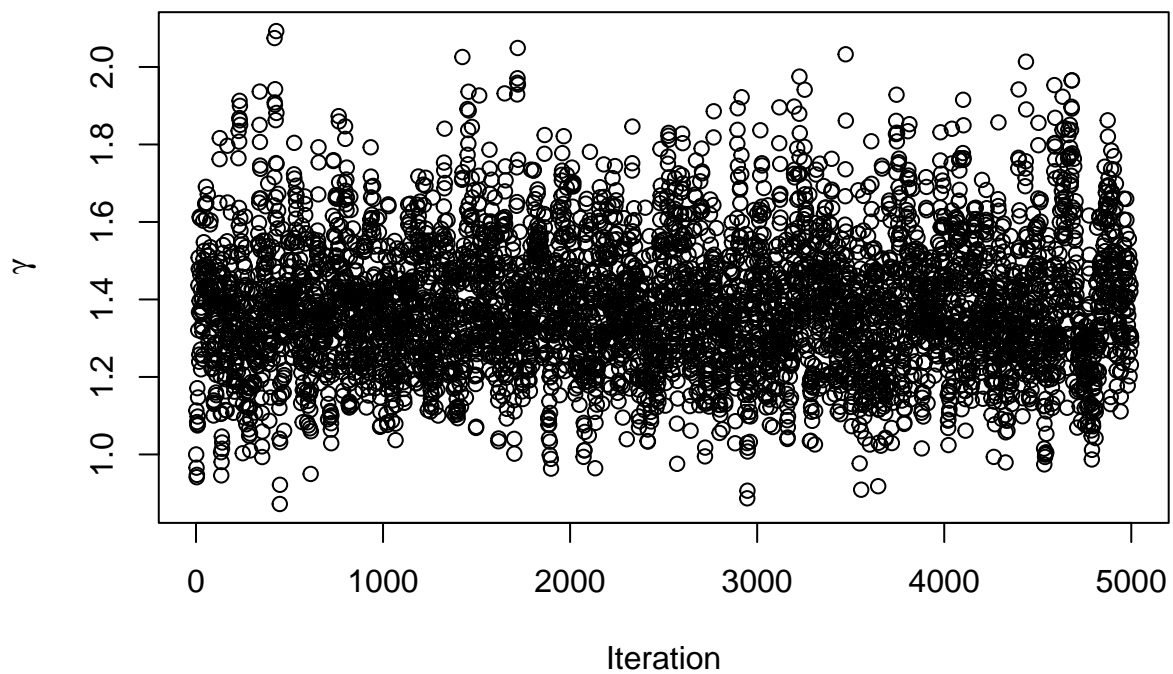
```
plot(THETA,
     main=expression(paste("Traceplot for ", theta)),
     xlab="Iteration", ylab=expression(theta))
```

Traceplot for θ



```
plot(GAMMA,
     main=expression(paste("Traceplot for ", gamma)),
     xlab="Iteration", ylab=expression(gamma))
```

Traceplot for γ

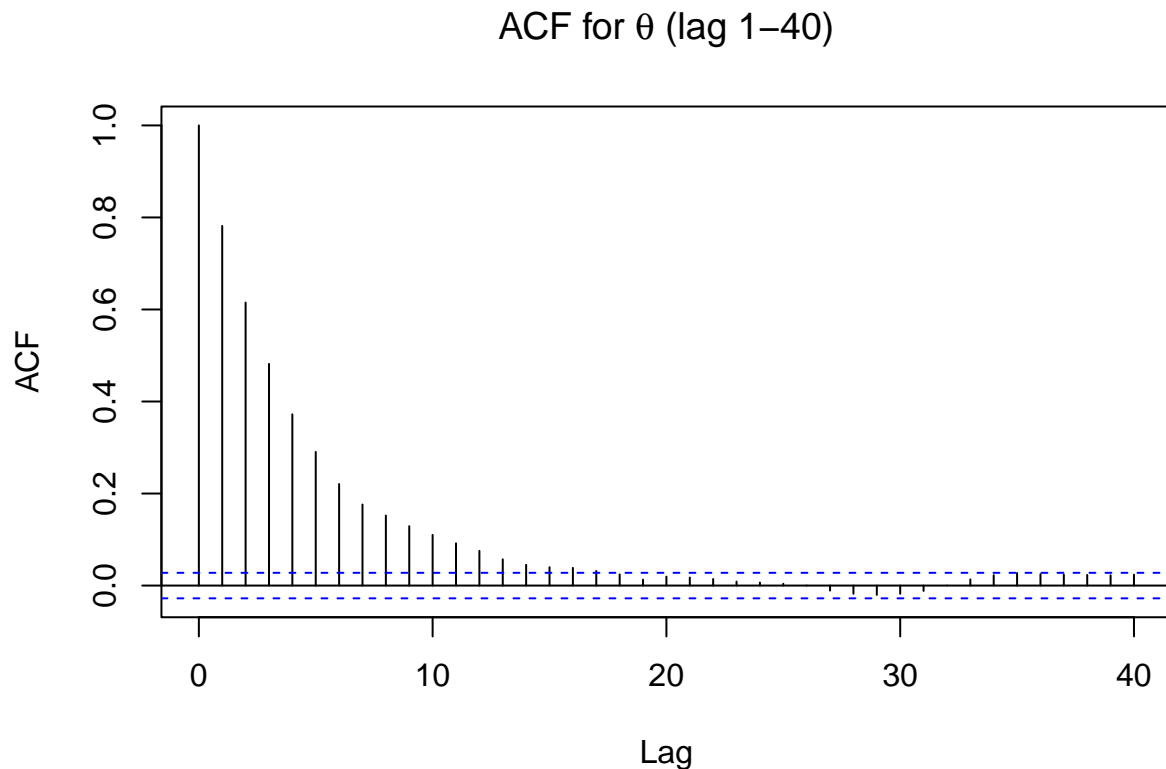


Both traceplots exhibit the characteristics of a well-mixing Markov chain, since they both show fluctuating,

dense, and fuzzy plots. The values for both θ and γ do not appear to be stuck in any particular region and are consistently sampling from the same distribution throughout the 5000 iterations. There is no significant 'burn-in' phase where the chain is moving from a poor starting value, with the exception of the very minimal transient phase for θ in the first few iterations, after which the plot enters the stable region, indicating that both achieve stationarity quickly. The samples show high variability between consecutive iterations, as visible in the fuzzy nature of the plot, indicating that the chain is efficiently exploring the joint posterior distribution. There are no long, smooth or sticky periods where the chain stays in one small region for hundreds of iterations. The range of the region is pretty wide, so there is low autocorrelation and good amounts of mixing. Based on these traceplots, the Gibbs sampler appears to be performing well. The chain has converged and is mixing efficiently, suggesting that the generated samples are a reliable basis for approximating posterior quantities.

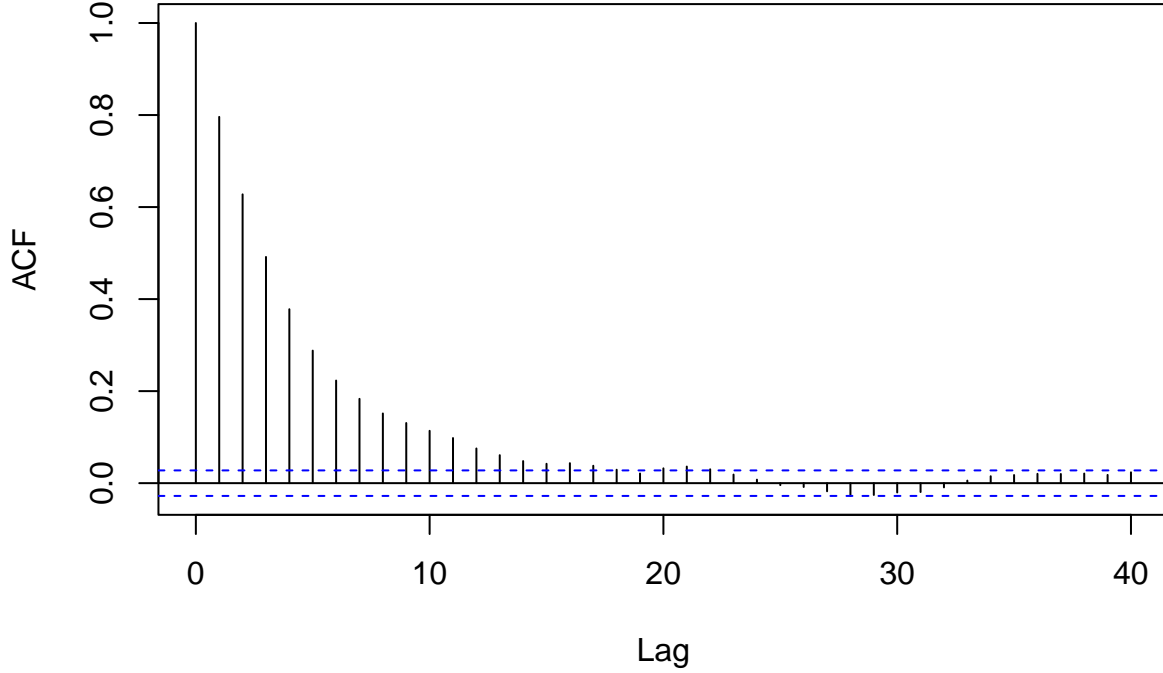
Now, I will create the autocorrelation function plots.

```
acf(THETA, lag.max=40,
    main=expression(paste('ACF for ', theta, ' (lag 1-40)')))
```



```
acf(GAMMA, lag.max=40,
    main=expression(paste('ACF for ', gamma, ' (lag 1-40)')))
```

ACF for γ (lag 1–40)



Both ACF plots show positive autocorrelation at small lags, indicating that these consecutive samples are highly correlated, but the autocorrelation decays relatively quickly towards zero, which is desirable for an MCMC algorithm.

For θ and γ both, the lag-1 autocorrelation was very high and almost close to 1.0, but dropped close to zero between lag-15 and lag 20. The chain exhibits some dependency, but the correlation is not excessive, and is generally low. The chain is mixing reasonably well, but the dependency in the beginning may lead to a higher variance in the final estimate, compared to independent samples, which we have less than 5000 of. This slightly reduces the precision of the posterior estimates.

The plots confirm that the Gibbs sampler is generating a dependent sequence of samples, which is expected for MCMC. The quick decay suggests that the chain has acceptable efficiency. If higher precision were required, one can run more iterations try sub-sampling. Since the precision is acceptable, we can now proceed with approximating the posterior expectation of the difference in θ_B and θ_A .

In order to do this, we have to sample a large number of values for θ and γ , which we have already done via MCMC approximation, to leverage the law of large numbers. We then apply prior knowledge that $\theta_A = \theta$ and $\theta_B = \theta\gamma$ to transform the existing samples to the required values. We then compute the mean of these samples. We can also technically discard a small burn-in period, but that may seem unnecessary considering the traceplots shows very minimal burn-in. This discards initial dependence on starting values. The calculated sample average remains a valid, albeit slightly less efficient, estimator, due to the effective independent samples being less than the total number of iterations. An approximation equation is provided below.

$$E(\theta_B - \theta_A \mid y^A, y^B) \approx \frac{1}{S} \sum_{k=1}^S \left(\theta^{(k)} \gamma^{(k)} - \theta^{(k)} \right)$$

The transformation needed is provided below:

$$\theta_B - \theta_A = \theta\gamma - \theta = \theta(\gamma - 1)$$

```
mean(THETA*(GAMMA-1))
```

```
## [1] 0.3706946
```

Thus, the posterior expectation is about 0.37.

Alternatively, we can also use a small burn in period (b) to get a more efficient estimate.

$$E(\theta_B - \theta_A \mid y^A, y^B) \approx \frac{1}{S-b} \sum_{k=1+b}^S \left(\theta^{(k)} \gamma^{(k)} - \theta^{(k)} \right)$$

Let's take $b = 200$ as a conservative estimate.

```
b <- 200  
mean(THETA[(b+1):S]*(GAMMA[(b+1):S]-1))
```

```
## [1] 0.3714289
```

The value does not change by much, despite the fact that I exceeded the point at which the chain started fluctuating as expected. This indicates that this is a reliable model.
