

Math 341 / 650 Spring 2017

Final Examination

Solutions

Professor Adam Kapelner

Thursday, May 25, 2017

Full Name _____

Code of Academic Integrity

Since the college is an academic community, its fundamental purpose is the pursuit of knowledge. Essential to the success of this educational mission is a commitment to the principles of academic integrity. Every member of the college community is responsible for upholding the highest standards of honesty at all times. Students, as members of the community, are also responsible for adhering to the principles and spirit of the following Code of Academic Integrity.

Activities that have the effect or intention of interfering with education, pursuit of knowledge, or fair evaluation of a student's performance are prohibited. Examples of such activities include but are not limited to the following definitions:

Cheating Using or attempting to use unauthorized assistance, material, or study aids in examinations or other academic work or preventing, or attempting to prevent, another from using authorized assistance, material, or study aids. Example: using an *unauthorized* cheat sheet in a quiz or exam, altering a graded exam and resubmitting it for a better grade, etc.

I acknowledge and agree to uphold this Code of Academic Integrity.

signature

date

Instructions

This exam is seventy five minutes and closed-book. You are allowed **three** pages (front and back) of "cheat sheets." You may use a graphing calculator of your choice. Please read the questions carefully. If the question reads "compute," this means the solution will be a number otherwise you can leave the answer in *any* widely accepted mathematical notation which could be resolved to an exact or approximate number with the use of a computer. I advise you to skip problems marked "[Extra Credit]" until you have finished the other questions on the exam, then loop back and plug in all the holes. I also advise you to use pencil. The exam is 100 points total plus extra credit. Partial credit will be granted for incomplete answers on most of the questions. Box in your final answers. Good luck!

Distribution of r.v.	Quantile Function	PMF / PDF function	CDF function	Sampling Function
beta	qbeta(p, α, β)	d-(x, α, β)	p-(x, α, β)	r-(α, β)
betabinomial	qbetabinom(p, n, α, β)	d-(x, n, α, β)	p-(x, n, α, β)	r-(n, α, β)
betanegativebinomial	qbeta_nbinom(p, r, α, β)	d-(x, r, α, β)	p-(x, r, α, β)	r-(r, α, β)
binomial	qbinom(p, n, θ)	d-(x, n, θ)	p-(x, n, θ)	r-(n, θ)
exponential	qexp(p, θ)	d-(x, θ)	p-(x, θ)	r-(θ)
gamma	qgamma(p, α, β)	d-(x, α, β)	p-(x, α, β)	r-(α, β)
geometric	qgeom(p, θ)	d-(x, θ)	p-(x, θ)	r-(θ)
inversegamma	qinvgamma(p, α, β)	d-(x, α, β)	p-(x, α, β)	r-(α, β)
negative-binomial	qnbinom(p, r, θ)	d-(x, r, θ)	p-(x, r, θ)	r-(r, θ)
normal (univariate)	qnorm(p, θ, σ)	d-(x, θ, σ)	p-(x, θ, σ)	r-(θ, σ)
normal (multivariate)		dmvnorm(x, μ, Σ)		r-(μ, Σ)
poisson	qpois(p, θ)	d-(x, θ)	p-(x, θ)	r-(θ)
T (standard)	qt(p, ν)	d-(x, ν)	p-(x, ν)	r-(ν)
T (nonstandard)	qt.scaled(p, ν, μ, σ)	d-(x, ν, μ, σ)	p-(x, ν, μ, σ)	r-(ν, μ, σ)
uniform	qunif(p, a, b)	d-(x, a, b)	p-(x, a, b)	r-(a, b)

Table 1: Functions from R (in alphabetical order) that can be used on this exam. The hyphen in columns 3, 4 and 5 is shorthand notation for the full text of the r.v. which can be found in column 2.

Problem 1 Assume a binomial likelihood with a known sample size n with a beta prior with parameters α and β .

(a) [3 pt / 3 pts] What is the kernel of $\mathbb{P}(\theta)$?

$$p(\theta) = \text{beta}(\alpha, \beta) \propto \theta^{\alpha-1} (1-\theta)^{\beta-1}$$

(b) [3 pt / 6 pts] What is the kernel of $\mathbb{P}(\theta | X, n)$?

$$p(\theta | X, n) = \text{beta}(\alpha + x, \beta + n - x) \propto \theta^{\alpha+x-1} (1-\theta)^{\beta+n-x-1}$$

(c) [4 pt / 10 pts] What is the kernel of $\mathbb{P}(\theta, X | n)$?

$$p(\theta, X | n) = \frac{p(X|\theta) \cdot p(\theta)}{p(X)} = \frac{\binom{n}{x} \theta^x (1-\theta)^{n-x}}{\frac{n!}{x!(n-x)!}} \left(\frac{1}{\text{beta}(\alpha, \beta)} \theta^{\alpha-1} (1-\theta)^{\beta-1} \right) \propto \frac{1}{x!(n-x)!} \theta^{x+\alpha-1} (1-\theta)^{n-x+\beta-1}$$

- (d) [5 pt / 15 pts] Find $\int \mathbb{P}(\theta, X | n) d\theta$ explicitly (i.e. saying the answer is e.g. $\mathbb{P}(\theta | X)$ is unacceptable. You can use any notation we used in class. Hint: nothing from Table 1 is the answer.

$$\int P(\theta, X | n) d\theta = P(X | n) = \text{betabinom}(n, \alpha, \beta)$$

- (e) [6 pt / 21 pts] Create a Gibbs sampler that will sample from $\mathbb{P}(\theta, X | n)$. Outline the steps of the sampler and describe explicitly how you do each step using functions from Table 1. You do not need to write about burning and thinning.

Step 1: Initialize $\theta_0 = 0.5, x_0 = 1$ i.e. arbitrary legal starting values

Step 2: Draw x_1 from $P(X | \theta = \theta_0) = \text{binom}(n, \theta_0)$

via `rbinom(n, θ_0)` etc

draw θ_1 from $P(\theta | X = x_1) = \text{beta}(\alpha + x_1, \beta + n - x_1)$

via `rbeta($\alpha + x_1, \beta + n - x_1$)`

Step 3: Repeat Step 2 many times.

Problem 2 The table below displays an excerpt of daily historical data of the S&P500 index for American equities going back to 1950, a total of $n = 16,428$ days. You had a similar dataset for the homework, but that was yearly returns.

Day	Ending Price	Gain or Loss (%)
1/4/1950	16.85	1.14
1/5/1950	16.93	0.47
1/6/1950	16.98	0.29
1/9/1950	17.08	0.58
1/10/1950	17.03	-0.29
\vdots		

For the daily percent gains (or losses) we calculate the sample average and sample standard deviation:

$$\bar{x} = 0.03\%$$

$$s = 0.97\%$$

We also assume that the daily percent gains (or losses) is modeled via

$$X_1, \dots, X_n \mid \theta, \sigma^2 \stackrel{iid}{\sim} \mathcal{N}(\theta, \sigma^2)$$

For the remainder of this problem, you are to answer explicitly using numbers for all arguments. If you use any substitutions, make it clear as to the value of or how the substitutions are calculated. They can only be a function of the x_i 's.

- (a) [4 pt / 25 pts] Assume σ^2 was known to be $0.97\%^2$, what is the distribution of θ given the data? Assume $\theta \sim \mathcal{N}(0\%, 100\%^2)$. Compute explicitly for this problem and the remainder of this question.

$$\begin{aligned}
 \theta \mid X, \sigma^2 &\sim \mathcal{N}\left(\frac{\frac{\bar{x}_n}{\sigma^2} + \frac{\mu_0}{\tau^2}}{\frac{n}{\sigma^2} + \frac{1}{\tau^2}}, \frac{1}{\frac{n}{\sigma^2} + \frac{1}{\tau^2}}\right) \\
 &= \mathcal{N}\left(\frac{\frac{0.03 \cdot 16428}{0.97^2}}{\frac{16428}{0.97^2} + \frac{1}{100^2}}, \frac{1}{\frac{16428}{0.97^2} + \frac{1}{100^2}}\right) \\
 &= \mathcal{N}(0.030, 5.72 \times 10^{-5})
 \end{aligned}$$

- (b) [4 pt / 29 pts] Assume μ was known to be 0.03%, what is the distribution of σ^2 given the data? Assume $\mathbb{P}(\sigma^2) \propto 1/\sigma^2$.

$$P(\sigma^2 | x, \theta) = \text{InverseGamma}\left(\frac{4-1}{2}, \frac{4\hat{\sigma}^2}{2}\right) = \text{InverseGamma}\left(\frac{4-1}{2}, \frac{\sum_{i=1}^4 (x_i - 0.03)^2}{2}\right) = \text{InverseGamma}\left(\frac{16428}{2}, \frac{\sum (x_i - 0.03)^2}{2}\right) \\ = \text{InverseGamma}\left(8213, \frac{1}{2} \sum_{i=1}^4 (x_i - 0.03)^2\right)$$

- (c) [4 pt / 33 pts] Assume neither μ nor σ^2 was known, what is the distribution of θ given the data assuming the Jeffrey's prior for the parameters?

$$P(\theta | x) = T_{4-1}\left(\bar{x}, \frac{s}{\sqrt{n}}\right) = T_{16428-1}\left(0.03, \frac{0.97}{\sqrt{16428}}\right) = T_{16427}(0.03, 0.0076)$$

- (d) [4 pt / 37 pts] Assume neither μ nor σ^2 was known, what is the distribution of σ^2 given the data assuming the Jeffrey's prior for the parameters?

$$P(\sigma^2 | x) = \text{InverseGamma}\left(\frac{4-1}{2}, \frac{(4-1)s^2}{2}\right) = \text{InverseGamma}\left(\frac{16428-1}{2}, \frac{(16428-1)0.97^2}{2}\right) = \text{InverseGamma}(8213.5, 7728.1)$$

- (e) [4 pt / 41 pts] Assume neither μ nor σ^2 was known, what is the distribution of tomorrow's return given the data assuming the Jeffrey's prior for the parameters?

$$P(x^a | x) = T_{4-1}\left(\bar{x}, s\sqrt{\frac{n+1}{n}}\right) = T_{16428-1}\left(0.03, 0.97\sqrt{\frac{16428+1}{16428}}\right) = T_{16427}(0.03, 0.97)$$

- (f) [6 pt / 47 pts] [Extra Credit] Assume neither μ nor σ^2 was known and assume the priors in (a) and (b). Explain how you would sample pairs $[\theta, \sigma^2]$ from $\mathbb{P}(\theta, \sigma^2 | X)$. Use functions from Table 1.

- (g) [4 pt / 51 pts] Using your answer from (d), provide a 99% credible region for σ^2 . Use functions from Table 1.

$$CR_{99\%, \sigma^2} = \left[\text{qinvgamma}(0.005, 8213.5, 7728.1), \text{qinvgamma}(0.995, 8213.5, 7728.1) \right]$$

- (h) [4 pt / 55 pts] Using your answer from (e), provide a 95% credible region for tomorrow's return. Use functions from Table 1.

$$CR_{95\%, X^*} = \left[\text{qt.scaled}(0.025, 16427, 0.03, 0.97), \text{qt.scaled}(0.975, 16427, 0.03, 0.97) \right]$$

$$\approx \left[\text{qnorm}(0.025, 0.03, 0.97), \text{qnorm}(0.975, 0.03, 0.97) \right]$$

- (i) [4 pt / 59 pts] We now want to make sure we picked a realistic likelihood model and realistic priors on its parameters (the uninformative Jeffrey's prior). We plot the true data and seven replicates using $\mathbb{P}(X^* | X)$ where the dimension of X^* is the same as of X , i.e. $n = 16428$. You can see the result in Figure 1 on page 10. Was our choice for the likelihood (and priors) a good choice to model this data? Yes / no. If not, where exactly did we go wrong?

*No. Most likely it is the model assumption of normality
this is the culprit.*

Problem 3 A call center at a large company receives many calls per day. In each minute, calls are approximately Poisson distributed as they are the result of millions of potential callers with a minuscule probability each can call in. For a particular day, the manager is trying to locate the "slightly busier period" between the calm period of the early morning and the calm period of the late evening. Figure 2 on page 11 shows a plot of a day's data where each point is the number of calls in one minute for all 24 hours:

You can barely see there is a busy period sandwiched between two calm periods. We will now build a Bayesian model to detect all parameters:

- m_1 — the first change point i.e. ^{last} the minute of the day that indicates the transition from the calm of the morning to the busy of the day ^{of calm}
- m_2 — the second change point i.e. ^{last} the minute of the day that indicates the transition from the busy of the day to the calm of the evening ^{of busy}
- λ_c — the mean number of calls during the calm of the morning and evening
- λ_b — the mean number of calls during the busy part of the day

We will assume each caller is independent. Let x_1, x_2, \dots, x_n represent the data (where $n = 1440$, the number of minutes in a 24hr day). For example, ~~x_{20}~~ is the number of calls in the 20th minute of the day i.e. from 12:19:00AM to 12:19:59AM. x_{20}

- (a) [6 pt / 65 pts] Write the likelihood of the model. Answer as explicitly as possible.

$$P(x_1, \dots, x_n | \lambda_c, \lambda_b, m_1, m_2) = \prod_{i=1}^{m_1} \frac{e^{-\lambda_c} \lambda_c^{x_i}}{x_i!} \prod_{i=m_1+1}^{m_2} \frac{e^{-\lambda_b} \lambda_b^{x_i}}{x_i!} \prod_{i=m_2+1}^n \frac{e^{-\lambda_c} \lambda_c^{x_i}}{x_i!}$$

- (b) [3 pt / 68 pts] Assume the ~~Haldane~~ prior of $\text{Gamma}(0, 0)$ for both λ_c and λ_b and assume a discrete uniform prior for m_1 and m_2 . Write each prior formally and indicate if each of them are proper or improper.

$$P(\lambda_c) = \text{Gamma}(0, 0) \propto \lambda_c^{-1} \quad (\text{improper})$$

$$P(\lambda_b) = \text{Gamma}(0, 0) \propto \lambda_b^{-1} \quad (\text{improper})$$

$$P(m_1) = U(1, 2, \dots, n_3) \quad (\text{proper})$$

$$P(m_2) = U(1, 2, \dots, n_3) \quad (\text{proper})$$

- (c) [6 pt / 74 pts] Assume the likelihood model from (a) and the priors from (b). Find the kernel of the posterior.

$$\begin{aligned} P(\lambda_c, \lambda_b, m_1, m_2 | x_1, \dots, x_n) &\propto P(x_1, \dots, x_n | \lambda_c, \lambda_b, m_1, m_2) P(\lambda_c) P(\lambda_b) P(m_1) P(m_2) \\ &= [\text{given from (a)}] (\lambda_c^{-1}) (\lambda_b^{-1}) (1) (1) \\ &\propto e^{-m_1 \lambda_c} \lambda_c^{\sum_{i=1}^{m_1} x_i} e^{-(m_2 - m_1) \lambda_b} \lambda_b^{\sum_{i=m_1+1}^{m_2} x_i} e^{-(n - m_2) \lambda_c} \lambda_c^{\sum_{i=m_2+1}^n x_i} \lambda_c^{-1} \lambda_b^{-1} \\ &= e^{-(m_1 + n - m_2) \lambda_c} \lambda_c^{\sum_{i=1}^{m_1} x_i + \sum_{i=m_2+1}^n x_i - 1} e^{-(m_2 - m_1) \lambda_b} \lambda_b^{\sum_{i=m_1+1}^{m_2} x_i - 1} \end{aligned}$$

- (d) [6 pt / 80 pts] Assume your expression for proportional posterior from (c). Find $P(\lambda_b | x_1, \dots, x_n, \lambda_c, m_1, m_2)$.

$$\propto e^{-(m_2 - m_1) \lambda_b} \lambda_b^{\sum_{i=m_1+1}^{m_2} x_i - 1} \propto \text{Gamma}\left(\sum_{i=m_1+1}^{m_2} x_i, m_2 - m_1\right)$$

- (e) [4 pt / 84 pts] Using the conditional from (d) and all other conditionals, we build a Gibbs sampler. ^{with 1000 iterations} Figure 3 on page 12 shows the first 80 iterations of each chain. It seems to burn in right away, at about the 10th iteration. Figure 4 on page 13 shows the autocorrelation of the chains. After thinning the burned-in chains, how many samples are we left with from the posterior?

$$\frac{1000 - 10}{3} = \frac{990}{3} = 330$$

- (f) [2 pt / 86 pts] Are these "left over samples" considered $\overset{iid}{\sim}$? Yes / no.
- (g) [4 pt / 90 pts] Figure 5 on page 14 shows the histograms from the burned and thinned chains. Provide an approximate 95% credible region for m_2 .

[1060, 1180]

- (h) [3 pt / 93 pts] Approximate the MMSE Bayesian estimate for λ_e .

5.02

- (i) [3 pt / 96 pts] Approximate the MAE Bayesian estimate for λ_b .

6.02

- (j) [4 pt / 100 pts] Test if m_1 is greater than 500.

$$H_0: m_1 \leq 500$$

$$H_a: m_1 > 500$$

$$p_{\text{val}} := P(H_0 | X) = P(m_1 \leq 500 | X) \approx 98\% > 5\%$$

\Rightarrow Fail to reject H_0

$\Rightarrow m_1 \neq 500$

- (k) [6 pt / 106 pts] Devise a means to test ~~H_0~~ if there was no busy period to begin with i.e. for the whole day there was only one homogeneous Poisson.

Assume $\alpha = 5\%$.

$$H_0: \lambda_1 - \lambda_2 = 0$$

$$H_a: \lambda_1 - \lambda_2 \neq 0$$

Use the Gibbs chain from (e) to compute a derived parameter $\{\lambda_{1,t} - \lambda_{2,t}\}$ for all t after burning and thinning. Then approximate a 95% $= 1 - \alpha$ CR.

If $0 \notin CR_{95\%, \lambda_1 - \lambda_2} \Rightarrow$ Reject H_0 (at fail to reject).

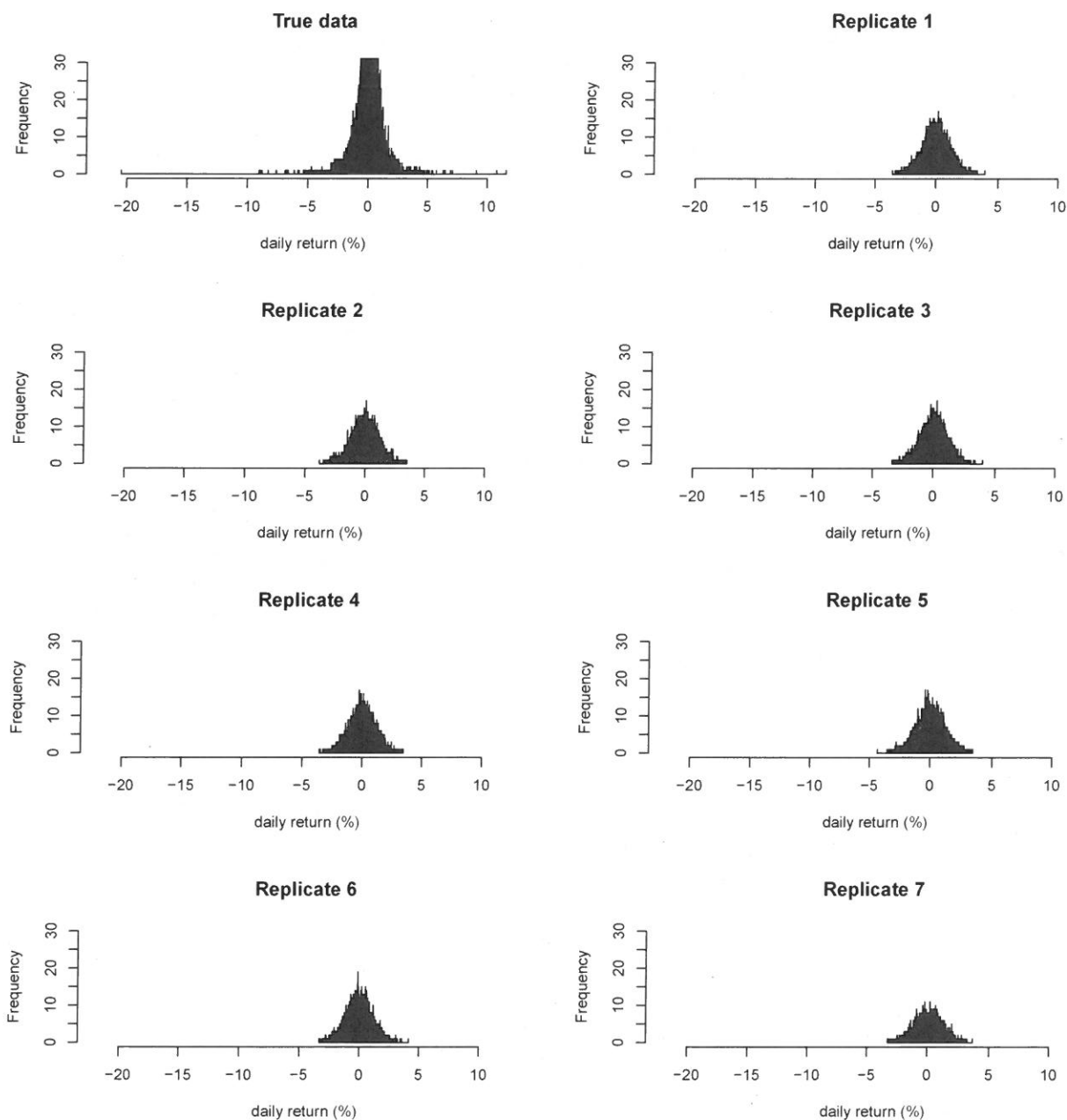


Figure 1: Replicates of the original data assuming the model and the prior.

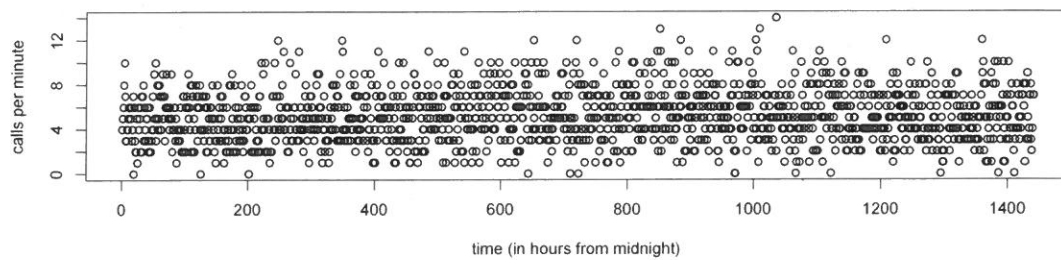


Figure 2: The call data.

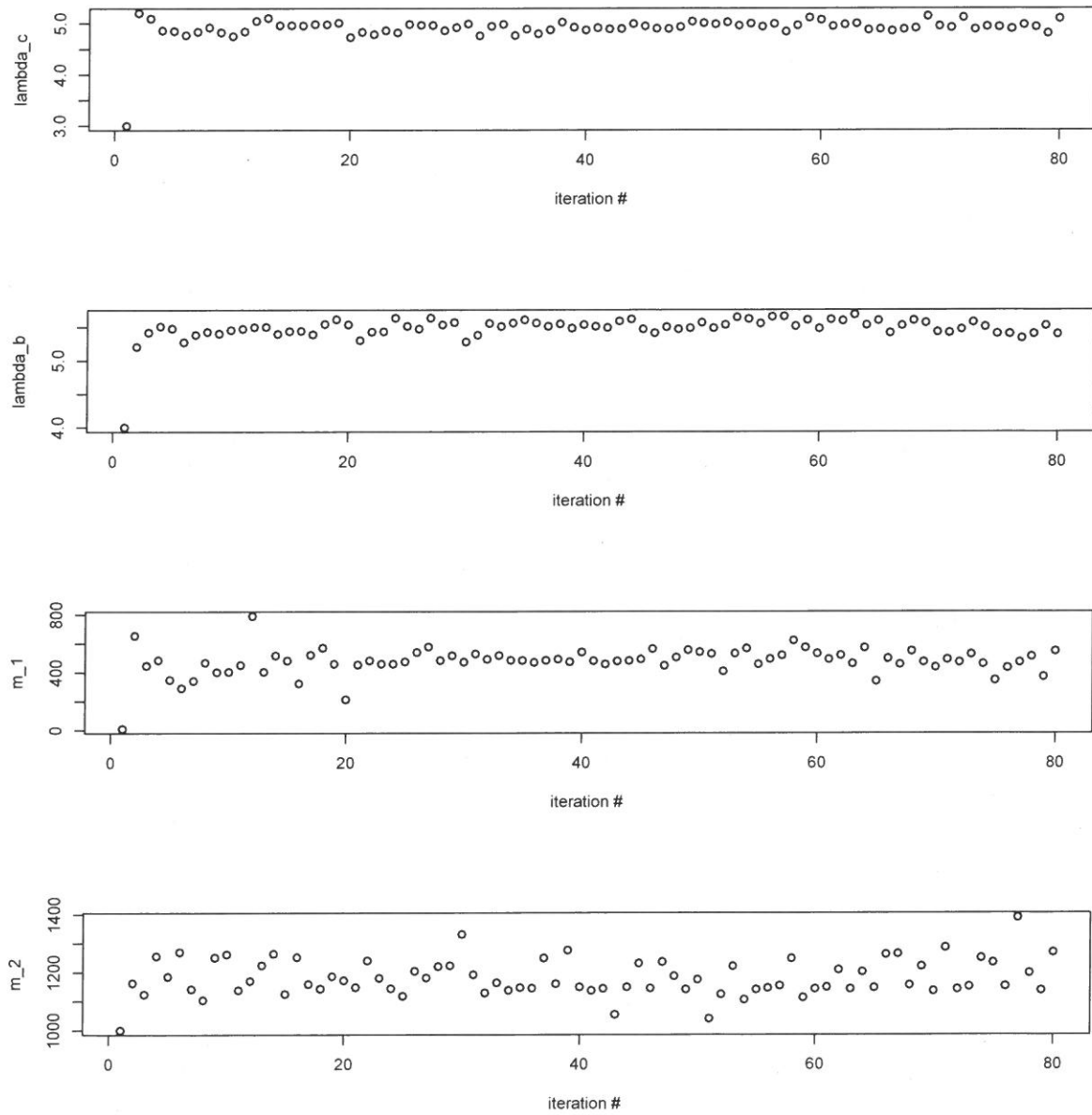


Figure 3: First 80 iterations of the Gibbs sampler in Problem 3.

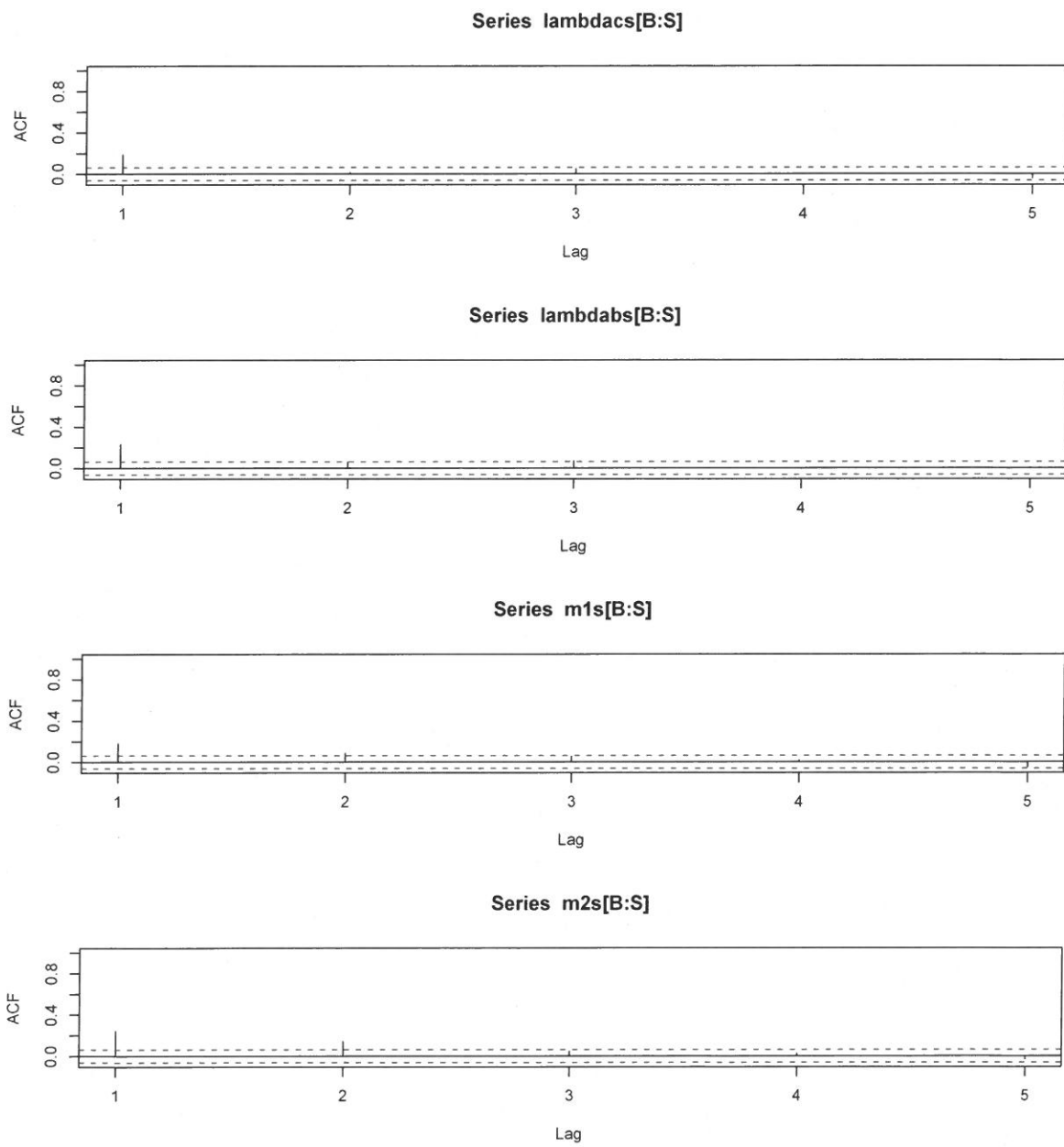


Figure 4: Autocorrelation estimates for the Gibbs sampler after burn-in in Problem 3.

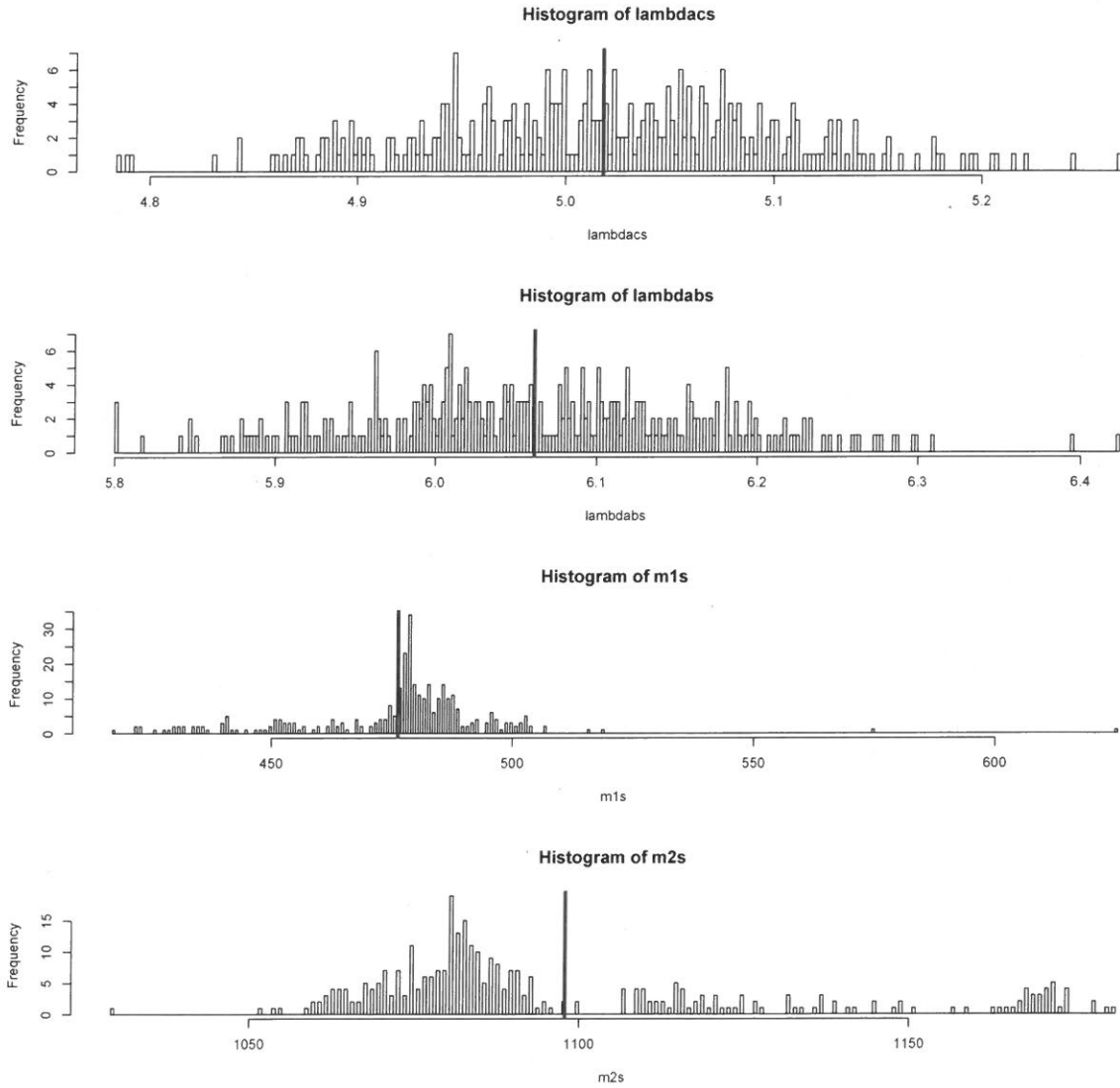


Figure 5: Histograms for the Gibbs sampler after burn-in and thinning in Problem 3.