Math 341 / 650 Spring 2019
Final Examination

Solutions

Professor Adam Kapelner

Thursday, May 16, 2019

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 120 minutes and closed-book. You are allowed **three** pages (front and back) of a "cheat sheet." 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. $\boxed{\text{Box}}$ 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)
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)
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)
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 assignments and exams. The hyphen in colums 3, 4 and 5 is shorthand notation for the full text of the r.v. which can be found in column 2.

Problem 1 Consider a call center that experiences calls from a large population of which there is very small probability each customer will call in each day. Let $X_1, \ldots, X_n$ denote the number of calls in a period of n days. However, there are two types of customers: A and B (since there are two main products the company offers). ~~Thus, we model the call data with a mixture of two Poissons with two different rate parameters:~~

And either customer from A call in ~~or~~ a given day or customer B calls.

$$X_1, \ldots, X_n \mid \lambda_1, \lambda_2, \rho \overset{iid}{\sim} \cancel{\rho \frac{e^{-\lambda_1}\lambda_1^x}{x!} + (1-\rho)\frac{e^{-\lambda_2}\lambda_2^x}{x!}} \quad \begin{cases} \frac{e^{-\lambda_1}\lambda_1^x}{x!} & \text{w.p. } \rho \\ \frac{e^{-\lambda_2}\lambda_2^x}{x!} & \text{w.p. } 1-\rho \end{cases}$$

(a) [6 pt / 6 pts] Provide uninformative priors for the three parameters.

$P(\lambda_1) \propto 1$

$P(\lambda_2) \propto 1$

$P(\rho) = U(0,1) \propto 1$

We now introduce the new parameters $I_1, \ldots, I_n$ where each is defined as:

$$I_i := \begin{cases} 1 & \text{if the calls from day } i \text{ come from customer type A} \\ 0 & \text{if the calls from day } i \text{ come from customer type B} \end{cases}$$

This is an idea known as "data augmentation". The likelihood now becomes:

$$\mathbb{P}(X_1, \ldots, X_n \mid \lambda_1, \lambda_2, \rho, I_1, \ldots, I_n) = \prod_{i=1}^{n} \left(\frac{e^{-\lambda_1}\lambda_1^{x_i}}{x_i!}\right)^{I_i} \left(\frac{e^{-\lambda_2}\lambda_2^{x_i}}{x_i!}\right)^{1-I_i} \rho^{I_i}(1-\rho)^{1-I_i}$$

(b) [5 pt / 11 pts] Denote $\sum I_i := \sum_{i=1}^{n} I_i$. By using the likelihood above and your priors from part (a), Show all your work to prove that the kernel of $\mathbb{P}(\lambda_1, \lambda_2, \rho, I_1, \ldots, I_n \mid X_1, \ldots, X_n)$, the posterior is

$$e^{-(\sum I_i)\lambda_1} \lambda_1^{\sum I_i x_i} e^{-(n-\sum I_i)\lambda_2} \lambda_2^{\sum (1-I_i) x_i} \rho^{\sum I_i} (1-\rho)^{n-\sum I_i}$$

$$P(\lambda_1, \lambda_2, \rho, I_1, \ldots, I_n \mid X_1, \ldots, X_n) \propto \prod e^{-\lambda_1 I_i} \lambda_1^{X_i I_i} e^{-\lambda_2 (1-I_i)} \lambda_2^{X_i(1-I_i)} \rho^{I_i} (1-\rho)^{1-I_i}$$

$$= e^{-\lambda_1 \Sigma I_i} \lambda_1^{\Sigma X_i I_i} e^{-\lambda_2 (n-\Sigma I_i)} \lambda_2^{\Sigma X_i (1-I_i)} \rho^{\Sigma I_i} (1-\rho)^{n-\Sigma I_i}$$

(c) [2 pt / 13 pts] Is this a kernel from a known distribution? Yes / no.

(d) [4 pt / 17 pts] Name two strategies that can be used to creates samples $[\lambda_1, \lambda_2, \rho, I_1, \ldots, I_n]$ from the posterior $\mathbb{P}(\lambda_1, \lambda_2, \rho, I_1, \ldots, I_n \mid X_1, \ldots, X_n)$.

① grid sampling ② gibbs sampling

(e) [4 pt / 21 pts] Which strategy would be better and why?

Gibbs sampling would be better. Grid sampling will suffer from not knowing the grid and it will be incorrect in high dimensions as we have here.

(f) [4 pt / 25 pts] Find the conditional distribution $\mathbb{P}(\lambda_1 \mid X_1, \ldots, X_n, \lambda_2, \rho, I_1, \ldots, I_n)$. If the PMF / PDF are from a known, brand name random variable, name the distribution and provide its parameter(s). Show all work.

$$P(\lambda_1 \mid -) \propto e^{-\lambda_1 \Sigma I_i} \lambda_1^{\Sigma X_i I_i} \propto \text{Gamma}\left(\Sigma X_i I_i + 1,\ \Sigma I_i\right)$$

(g) [4 pt / 29 pts] Find the conditional distribution $\mathbb{P}(\lambda_2 \mid X_1, \ldots, X_n, \lambda_1, \rho, I_1, \ldots, I_n)$. If the PMF / PDF are from a known, brand name random variable, name the distribution and provide its parameter(s). Show all work.

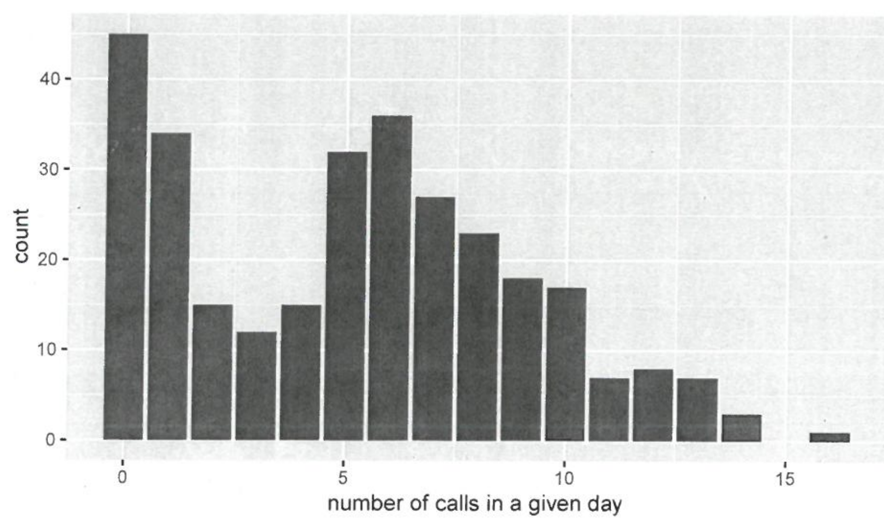
$$P(\lambda_2 \mid -) \propto e^{-\lambda_2 (n-\Sigma I_i)} \lambda_2^{\Sigma X_i (1-I_i)} \propto \text{Gamma}\left(\Sigma X_i (1-I_i) + 1,\ n-\Sigma I_i\right)$$

(h) [6 pt / 35 pts] Find the conditional distribution $\mathbb{P}(\rho \mid X_1, \ldots, X_n, \lambda_1, \lambda_2, I_1, \ldots, I_n)$. If the PMF / PDF is from a known, brand name random variable, name the distribution and provide its parameter(s). Show all work.

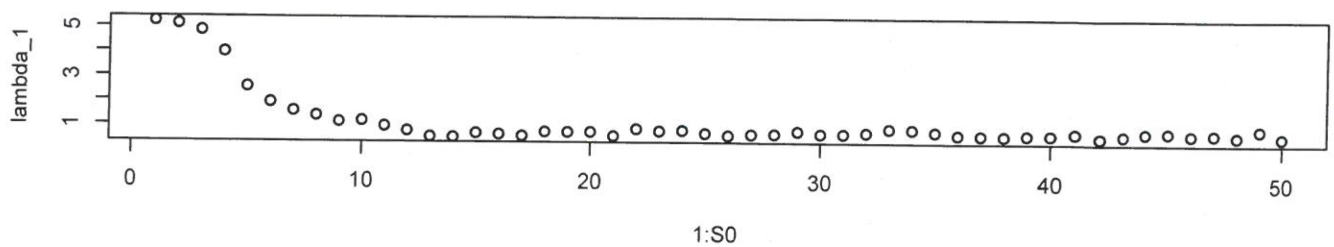
$P(\rho \mid -) \propto \rho^{\sum I_i} (1-\rho)^{n - \sum I_i} \propto \text{Beta}(\sum I_i + 1, n - \sum I_i + 1)$

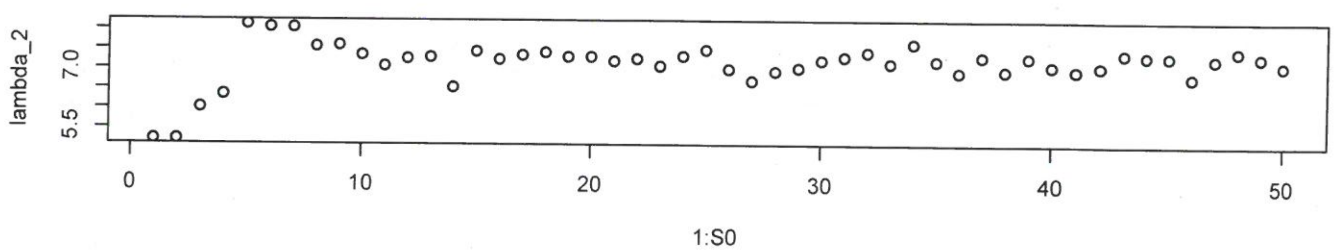
(i) [5 pt / 40 pts] [Extra Credit] Find the conditional distribution for I_1 given the data and the other parameters. If the PMF / PDF is from a known, brand name random variable, name the distribution and provide its parameter(s). Show all work.

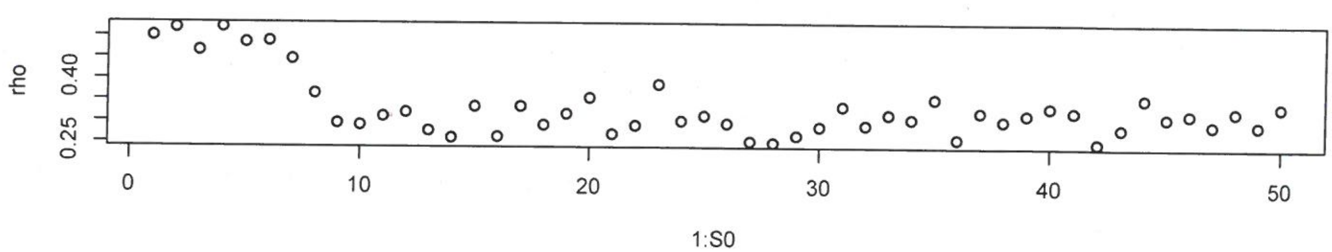
Below is a bar plot of our data for $n = 300$ days:



We now build a Gibbs Sampler for all parameters $\lambda_1, \lambda_2, \rho, I_1, \ldots, I_n$ based on your answers from (a)-(h) and initially sample 10,000 times. Below is the first 50 samples from the chains for $\lambda_1, \lambda_2, \rho$:



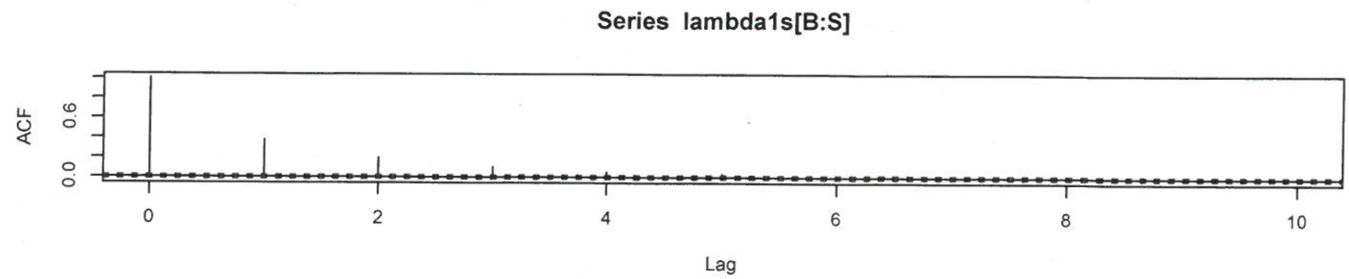


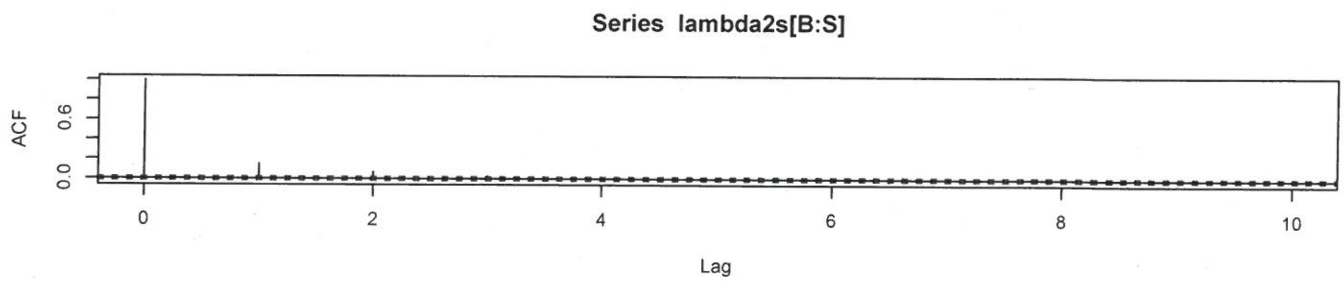


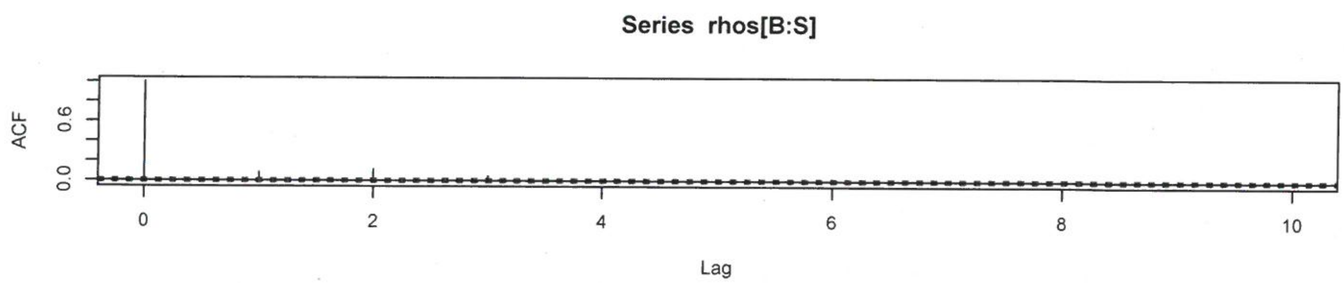
(j) [3 pt / 43 pts] How many of the initial samples should be thrown away?

11-15

After removing the initial part of the chain according to your answer in the previous question, below is an autocorrelation plot of the chains for $\lambda_1, \lambda_2, \rho$ with maximum lag 10:







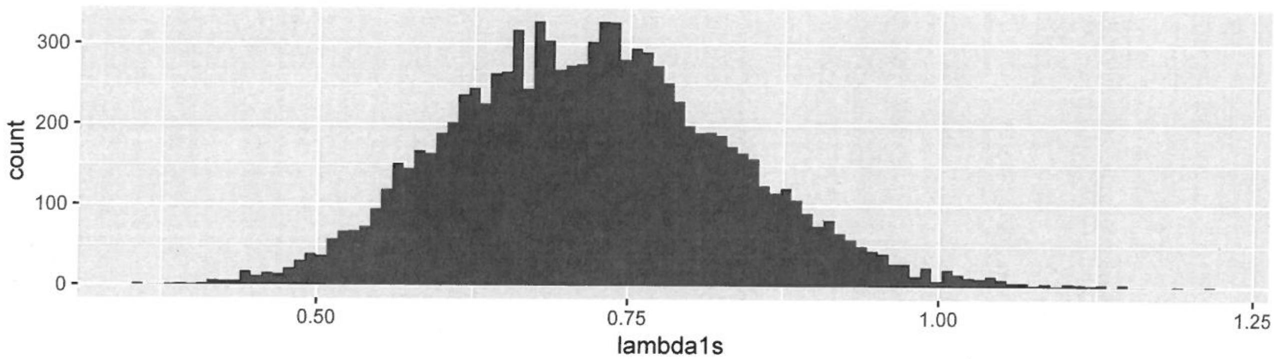
(k) [3 pt / 46 pts] How many iterations do you think should be skipped between samples in the burned-in chains if you want $\overset{iid}{\sim}$ samples from the posterior?

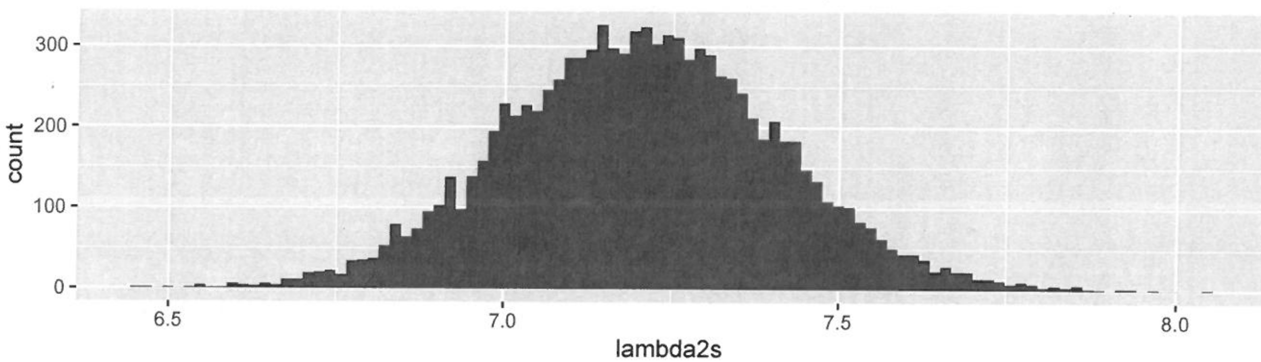
4 or 5

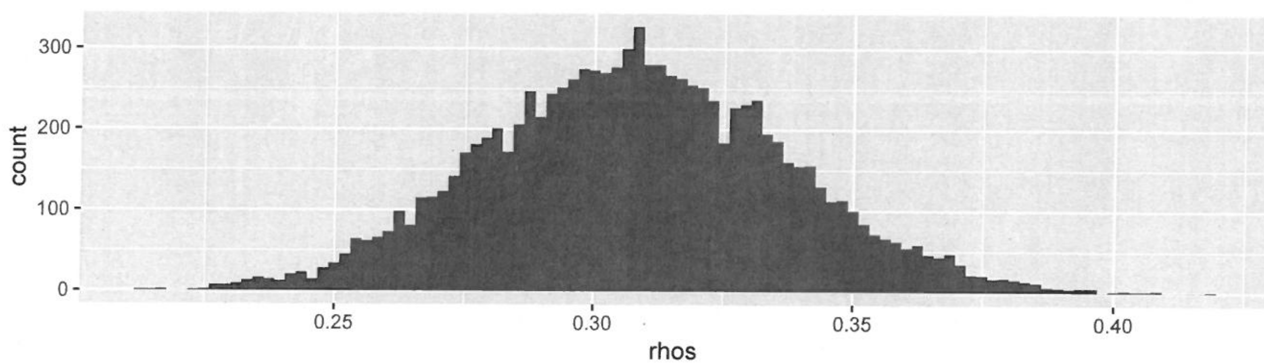
(l) [4 pt / 50 pts] How many $\overset{iid}{\sim}$ samples from the posterior do you have now?

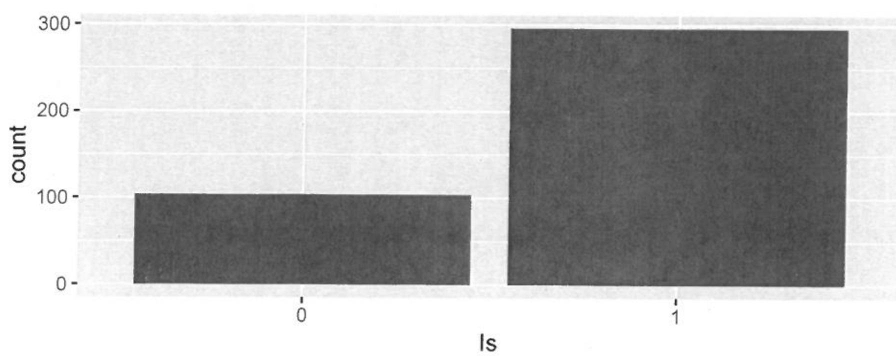
$$\frac{10{,}000 - 11}{4} = 2497.25 \Rightarrow 2497$$

The next page shows four different histograms: one for each of the burned and thinned chains for $\lambda_1, \lambda_2, \rho$ and I_{31}.


count
300
200
100
0
0.50
0.75
1.00
1.25
lambda1s


count
300
200
100
0
6.5
7.0
7.5
8.0
lambda2s


count
300
200
100
0
0.25
0.30
0.35
0.40
rhos


count
300
200
100
0
0
1
ls

(m) [6 pt / 56 pts] Provide an approximate $CR_{\lambda_1, 90\%}$.

$[0.58, 0.95]$

(n) [5 pt / 61 pts] Provide an approximate Bayesian point estimate for λ_2. Indicate which point estimate you are providing by specifying the subscript to $\hat{\theta}$.

$\hat{\theta}_{MAE} = 7.25$

(o) [8 pt / 69 pts] Test $\rho \neq 0.3$ using $\delta = 0.05$. List the hypotheses, pick an α level and provide an approximate Bayesian p-value and give a one sentence interpretation of the test result.

$H_0: \rho \in [0.25, 0.35]$

$H_a: \rho \notin [0.25, 0.35]$

$\alpha = 5\%$

eyeballed from histogram on p 7

$p_{val} = P(\rho \in [0.25, 0.35] \mid X_1, \ldots, X_n) \approx 90\% \not< \alpha = 5\%$

$\Rightarrow$ Fail to reject H_0. We cannot reject the claim that $\rho = 0.3$ ~~and~~ at a margin of equivalence $\delta = 0.05$.

(p) [4 pt / 73 pts] In your estimation, what type of customer dominates day 31's phone calls?

type A

(q) [8 pt / 81 pts] Explain how you would provide *one* prediction for the next day's number of phone calls. Clearly write out all your steps. Make sure the company can run your code. The functions you use from Table 1 must have explicit numeric values. Below are 10 random samples from the burned and thinned chains for parameters $\lambda_1, \lambda_2, \rho$:

```
           [,1]  [,2]  [,3]  [,4]  [,5]  [,6]  [,7]  [,8]  [,9] [,10]
lambda1s  0.731 0.573 0.861 0.770 0.797 0.683 0.597 0.649 0.670 0.784
lambda2s  7.312 7.203 7.259 7.077 7.328 7.489 7.006 6.709 6.950 7.036
rhos      0.356 0.271 0.291 0.388 0.301 0.314 0.296 0.254 0.251 0.260
```

Your answer must use some of the above numbers somehow.

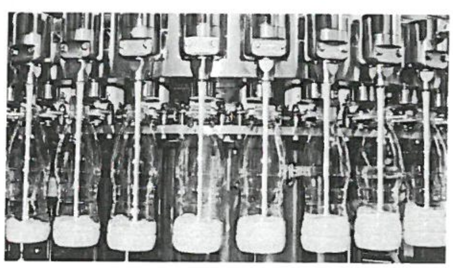
① We first sample one $[\lambda_1, \lambda_2, \rho]$ from the posterior as approximated by the burned and thinned Gibbs chains. Let's consider sample #7: $\lambda_1 = 0.597$, $\lambda_2 = 7.006$, $\rho = 0.296$

② Now we sample u from runif(0,1). If $u \leq \rho = 0.296$ then proceed to step 3. If not, proceed to step 4.

③ Draw X^*_{samp} from rpois(0.597)

④ Draw X^*_{samp} from rpois(7.006)

Problem 2 This question is about building a model to understand the accuracy of this beverage-filling machine



which fills 12oz plastic bottles. We decide to do an experiment and select $n = 20$ bottles at random and measure the amount of liquid in each bottle. Here are the volumes (in oz):

```
10.76 11.02 11.62 10.20 12.03 12.18 12.06 11.42 11.93 10.68
11.81 11.27 11.68 11.31 11.29 12.37 12.00 10.74 12.04 10.69
```

and sample statistics: $\bar{x} = 11.6$oz and $s = 0.53$oz. We will consider these measurements realizations from $\mathcal{F} : X_1, \ldots, X_n \mid \theta, \sigma^2 \overset{iid}{\sim} \mathcal{N}(\theta, \sigma^2)$ where both θ and σ^2 are unknown.

(a) [3 pt / 84 pts] If the posterior of interest is $\mathbb{P}(\theta, \sigma^2 \mid X_1, \ldots, X_n)$, provide a non-informative conjugate prior for both θ and σ^2 below. It does not need to be proper.

$$P(\theta, \sigma^2) \propto \frac{1}{\sigma^2}$$

(b) [5 pt / 89 pts] Using the prior from (a) and function(s) from Table 1, provide an exact $CR_{\theta,95\%}$.

$$P(\theta \mid X) = T_{n-1}\left(\bar{x}, \frac{s}{\sqrt{n}}\right) = T_{19}\left(11.6, \underbrace{\frac{0.53}{\sqrt{20}}}_{0.12}\right)$$

$$CR_{\theta,95\%} = [\text{qt.scaled}(0.025, 19, 11.6, 0.12), \text{qt.scaled}(0.975, 19, 11.6, 0.12)]$$

(c) [5 pt / 94 pts] Using the prior from (a), find $\hat{\sigma}^2_{MMSE}$.

$$P(\sigma^2 \mid X) = \text{InvGamma}\left(\frac{n-1}{2}, \frac{(n-1)s^2}{2}\right) = \text{InvGamma}\left(\overset{9.5}{\frac{19}{2}}, \overset{2.67}{\frac{19 \cdot 0.53^2}{2}}\right)$$

$$\hat{\sigma}^2_{MMSE} = \underset{\overset{\shortparallel}{b/(\alpha-1)}}{E[\sigma^2 \mid X]} = \frac{2.67}{9.5-1} = 0.31$$

(d) [6 pt / 100 pts] Using the prior from (a) and function(s) from Table 1, find the probability that the next milk bottle will have less than 12oz of milk.

$$P(X^* \mid X) = T_{n-1}\left(\bar{x}, \sqrt{\frac{n+1}{n}}\, s\right) = T_{19}\left(11.6, \underbrace{\sqrt{\frac{21}{20}}\, 0.53}_{0.54}\right)$$

$$P(X^* \leq 12\text{oz} \mid X) = \text{pt.scaled}(12, 19, 11.6, 0.54)$$

(e) [5 pt / 105 pts] Compute the following integral as a function of a, b and fundamental constants. To get full credit you must show and justify all steps.

$$\int_{\mathbb{R}} e^{ax-bx^2} dx = \sqrt{\frac{\pi}{b}}\, e^{a^2/4b}$$

(See midterm 2 solutions)