

STAT 544 Mid-term Exam
Thursday 13 March 8:00-9:20

Instructor: Jarad Niemi

INSTRUCTIONS

Please check to make sure you have 4 pages with writing on the front and back (some pages are marked ‘intentionally left blank’). Feel free to remove the last page, i.e. the one with R code.

On the following pages you will find short answer questions related to the topics we covered in class for a total of 50 points. Please read the directions carefully.

You are allowed to use a calculator and one $8\frac{1}{2} \times 11$ sheet of paper with writing on both front and back. A non-exhaustive list of items you are not allowed to use are **cell phones, laptops, PDAs, and textbooks**. Cheating will not be tolerated. Anyone caught cheating will receive an automatic F on the exam. In addition the incident will be reported, and dealt with according to University’s Academic Dishonesty regulations. Please refrain from talking to your peers, exchanging papers, writing utensils or other objects, or walking around the room. All of these activities can be considered cheating. **If you have any questions, please raise your hand.**

You will be given only the time allotted for the course; no extra time will be given.

Good Luck!

Please print your name below:

Student Name: _____

Confidence vs credible intervals (6 pts this page)

When I was a child my mother used to occasionally surprise me by ordering a jar of chocolate-chip cookies to be delivered by mail. The delivery supplier stocked four different kinds of cookie jars – type A, type B, type C, and type D, and they were all on the same truck and you were never sure what type you would get. Each jar had exactly 100 cookies, but the feature that distinguished the different cookie jars was their respective distributions of chocolate chips per cookie. If you reached into a jar and took out a single cookie uniformly at random, the table below provides the probability distribution for number of chips on a cookie by jar type. For example, a type-A cookie jar has 8 cookies with 0 chips, 72 cookies with two chips each, and 20 cookies with 3 chips. The goal of this problem is to construct confidence and credible sets for jar type when a single cookie is observed.

Chips	Jar Type			
	A	B	C	D
0	2	3	2	20
1	0	5	5	2
2	18	6	0	1
3	5	5	0	1
4	0	6	18	1

1. The list below contains 80% confidence sets in parentheses. For each jar type, calculate the coverage, i.e. the probability the true jar is in the set, conditional on the jar type. (4 pts)
 - type A (0,2) 80%
 - type B (0,2,3,4) 80%
 - type C (0,4) 80%
 - type D (0) 80%
2. If one cookie is observed and it has 1 chocolate chip, what is odd about this confidence set? (1 pt)
The confidence set contains no values.
3. If one cookie is observed and it has 3 chocolate chips, what is odd about this confidence set? (1 pt)
Only one of two jars with the same proportion of 3-chip cookies is in the set.

Confidence vs credible intervals continued (14 pts this page)

On this page, you will fill out Table 1. At minimum, show your work for the 0 chip row in the corresponding question number.

Table 1: Joint distribution for cookie and jar type.

Chips	Jar				marginal probability	credibility
	A	B	C	D		
0	2	3	2	20	27	81
1	0	5	5	2	12	83
2	18	6	0	1	25	96
3	5	5	0	1	11	91
4	0	6	18	1	25	96
total	25%	25%	25%	25%		

4. Assuming the delivery man picks jar types randomly and uniformly, fill in Table 1 for the joint distribution of chips and jar types. A (5 pts)

Divide all cells in Table 1 by 100.

5. Fill in the table above with the marginal probability for each number of chips? (2 pts)

$$\frac{20+2}{27} > 0.8$$

6. In Table 1, circle the cells that create a 80% credible sets, i.e. for any observed number of chips, the probability of the jars are at least 80%. (4 pts)

Credible sets are

0. C,D
1. B,C
2. A,B
3. A,B
4. B,C

7. For each row, calculate the actual credibility, i.e. the probability of the set of jars conditional on the number of chips. At minimum, show your work for the 0 chip row. (3 pts)

The actual credibility is

0. $(20+2)/27 = 81\%$
1. $(5+5)/12 = 83\%$
2. $(18+6)/25 = 96\%$
3. $(5+5)/11 = 91\%$
4. $(6+18)/25 = 96\%$

Gamma model (10 pts total)

Suppose $Y_i \stackrel{iid}{\sim} Ga(a, \beta)$ with a known and β unknown where the probability density function is

$$p(y|\beta) = \frac{\beta^a}{\Gamma(a)} y^{a-1} e^{-\beta y}.$$

1. Derive Jeffreys prior for β . (4 pts)

$$\begin{aligned} \log L(\beta) &= a \log(\beta) - \log \Gamma(a) - (a-1) \log(y) - \beta y \\ \frac{\partial}{\partial \beta} \log L(\beta) &= a/\beta - y \\ \frac{\partial^2}{\partial \beta^2} \log L(\beta) &= -a/\beta^2 \\ \mathcal{I}(\beta) &= -E \left[\frac{\partial^2}{\partial \beta^2} \log L(\beta) \right] = a/\beta^2 \\ p(\beta) &\propto \sqrt{|\mathcal{I}(\beta)|} = 1/\beta \end{aligned}$$

2. Derive the posterior for β under Jeffreys prior. Make sure to name the distribution. (2 pts)

$$\begin{aligned} p(\beta|y) &\propto p(y|\beta)p(\beta) \\ &= \left[\prod_{i=1}^n \frac{\beta^a}{\Gamma(a)} y_i^{a-1} e^{-\beta y_i} \right] \frac{1}{\beta} \\ &\propto \beta^{na-1} e^{-\beta n\bar{y}} \end{aligned}$$

where \bar{y} is the sample mean of the y_i . This is the kernel of a $Ga(na, n\bar{y})$.

3. Consider the conjugate prior $\beta \sim Ga(a_0, b_0)$, derive the posterior under this prior. (2 pts)

$$\begin{aligned} p(\beta|y) &\propto p(y|\beta)p(\beta) \\ &\propto \left[\prod_{i=1}^n \frac{\beta^a}{\Gamma(a)} y_i^{a-1} e^{-\beta y_i} \right] \beta^{a_0-1} e^{-b_0\beta} \\ &\propto \beta^{a_0+na-1} e^{-\beta[b_0+n\bar{y}]} \end{aligned}$$

This is the kernel of a $Ga(a_0 + na, b_0 + n\bar{y})$.

4. What is the interpretation for a_0 ? (1 pt)

Prior number of observations times a .

5. What is the interpretation for b_0 ? (1 pt)

Prior sum of those observations.

Failure times (10 pts total)

A manufacturing facility utilizes a large number of identical O-rings. Due to the large number, the facility purchases these O-rings from a variety of suppliers. The company has recorded the failure time in **months** for O-rings as well as the supplier and is interested in understanding which supplier makes longer lasting O-rings. The JAGS code on page 8 provides an analysis of their data. Please use this code and output to answer the following questions.

1. Write down the model that was used to analyze the failure times including any priors used. Make sure to define any notation you introduce, e.g. if you use Y_i , then define what it means. (5 pts)

Let Y_i be the failure time and $s[i]$ be the supplier id for the i^{th} O-ring. The model is

$$\begin{aligned}Y_i &\overset{ind}{\sim} \text{Exp}(\rho_{s[i]}) \\ \rho_s &\overset{ind}{\sim} \text{Ga}(\mu^2/\sigma^2, \mu/\sigma^2) \\ \mu &\sim \text{Exp}(1/100) \\ \sigma &\sim \text{Exp}(1/100)\end{aligned}$$

with μ and σ being independent and $\text{Exp}(\rho)$ is an exponential random variable with a mean of $1/\rho$.

2. Provide 95% credible intervals (1 pt each) for

The answers given here are based on the attached output and differ from the actual exam given.

- (a) μ (0.60,1.46)
- (b) ρ_1 , the rate of failures for O-rings for company 1 (0.46,1.73)
- (c) the **mean failure time** for O-rings from company 1 This is just the inverse of the previous, i.e. $(1/1.73, 1/0.46) = (0.58, 2.17)$.

3. Describe why the median for ρ for the 2nd supplier is so low compared to the others. (1 pt)

This company has the longest lasting O-rings.

4. Describe why the 95% credible interval for ρ is so much shorter for supplier 10 compared to the others (excluding supplier 3). (1 pt)

This company had many more observations than the others.

Failure times continued (10 pts total, 2 pts each)

5. Why is this model more reasonable than a model that assumes a common failure rate for all suppliers?

There is no reason to believe that all suppliers have the same failure rate.

6. What is an advantage of this model over a non-hierarchical model that treats each supplier completely independently?

This model allows us to borrow information across suppliers.

7. Describe how to obtain samples from the posterior predictive distribution of O-ring failure times for supplier s .

Let ρ_s^j be the j th Monte Carlo draw from JAGS, then $\tilde{Y}_s^j \stackrel{\text{ind}}{\sim} \text{Exp}(\rho_s^j)$ is a draw from the predictive distribution of O-ring failure times for supplier s .

8. For a single machine needing a single O-ring, describe how to determine the expected cost for replacing O-rings over one year if the O-rings come from supplier s at a cost of C_s .

Use draws from the predictive distribution, i.e. \tilde{Y}_s^j from the previous problem, and sum them up until they are over 12 (since there are 12 months in a year). The number that are needed is a draw from the predictive distribution for how many O-rings will be needed over a year. Repeat this process a large number of times and take the average. Multiply this average by C_s to obtain an estimate of the expected cost.

Alternatively (and preferably), the number of O-rings needed have a Poisson distribution with mean $12\rho_s$. Thus the expected cost is $E[12C_s\rho_s] = 12C_sE[\rho_s]$ and we can use the estimated posterior mean from the JAGS output for each ρ_s .

9. Describe how you would use the expected costs from the previous problem to determine which supplier you will buy O-rings from.

Take the estimated costs (and their Monte Carlo uncertainty) and choose the company with the lowest expected cost.

(intentionally left blank- use for scratch paper)

JAGS code - failure time

The JAGS result is not reproducible and therefore the actual exam version differs from this version.

```
library(rjags)
dat = list(n = nrow(d),
          y = d$failure_time,
          supplier = d$supplier,
          n_suppliers = max(d$supplier))

model = "
model {
  for (i in 1:n) {
    y[i] ~ dexp(rho[supplier[i]])
  }

  for (s in 1:n_suppliers) {
    rho[s] ~ dgamma(mu^2/sigma^2, mu/sigma^2)
  }

  sigma ~ dexp(0.01)
  mu ~ dexp(0.01)
}"
```

```
m = jags.model(textConnection(model), dat, quiet=TRUE,
               inits=list(.RNG.name="base:Wichmann-Hill", .RNG.seed=1))
r = coda.samples(m, c("rho", "mu", "sigma"), 10000, thin=10)
```

```
summary(r)

##
## Iterations = 1010:11000
## Thinning interval = 10
## Number of chains = 1
## Sample size per chain = 1000
##
## 1. Empirical mean and standard deviation for each variable,
##    plus standard error of the mean:
##
##           Mean      SD Naive SE Time-series SE
## mu      0.9352 0.2500 0.007904      0.012010
## rho[1]  0.9619 0.3521 0.011134      0.011776
## rho[2]  0.3233 0.1281 0.004051      0.004758
## rho[3]  0.9060 0.2988 0.009448      0.010049
## rho[4]  0.9880 0.3519 0.011127      0.012919
## rho[5]  0.9710 0.3370 0.010655      0.012335
## rho[6]  0.7810 0.2586 0.008177      0.007746
## rho[7]  0.8525 0.2837 0.008970      0.008970
## rho[8]  0.8212 0.2703 0.008547      0.008547
## rho[9]  0.9651 0.3392 0.010728      0.010728
## rho[10] 0.9272 0.1300 0.004110      0.004110
## sigma  0.4653 0.2994 0.009467      0.016222
##
## 2. Quantiles for each variable:
##
##           2.5%    25%    50%    75%   97.5%
## mu      0.6110 0.7730 0.8831 1.0366 1.5644
## rho[1]  0.4294 0.7284 0.8935 1.1552 1.8202
## rho[2]  0.1275 0.2250 0.3105 0.4033 0.6182
## rho[3]  0.4538 0.6943 0.8535 1.0756 1.5634
## rho[4]  0.4619 0.7523 0.9258 1.1715 1.8261
## rho[5]  0.4552 0.7445 0.9222 1.1382 1.7793
## rho[6]  0.3490 0.6001 0.7492 0.9306 1.3478
## rho[7]  0.4150 0.6626 0.8123 0.9929 1.5115
## rho[8]  0.3843 0.6294 0.7959 0.9864 1.3857
## rho[9]  0.4332 0.7298 0.9234 1.1322 1.7731
## rho[10] 0.7026 0.8406 0.9167 1.0096 1.1946
## sigma  0.1392 0.2763 0.3887 0.5649 1.1867
```