Student Name: ____

STAT 544 Mid-term Exam
Thursday 8 March 2018, 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!

1. In a location that rains only 5 days out of the year, a meteorologist has predicted rain for tomorrow. The meteorologist correctly predicts rain 95% of the time when it does rain and correctly predicts no rain 90% of the time when it does not rain. What is the probability it will rain tomorrow? (20 points)

   Answer: Let

   - R be the event that it rains and
   - F be the event that rain is forecasted.

   From the above information, we know

   - $P(R) = 5/365$
   - $P(F|R) = 0.95$ and
   - $P(F^C|R^C) = 0.90$

   From this information, we can calculate the probability it will rain using Bayes Rule.

$$\begin{aligned} P(R|F) &= \frac{P(F|R)P(R)}{P(F|R)P(R)+P(F|R^C)P(R^C)} \\ &= \frac{P(F|R)P(R)}{P(F|R)P(R)+[1-P(F^C|R^C)][1-P(R)]} \end{aligned}$$

```
prev = 5/365
sens = .95
spec = 0.9

prob = (sens*prev)/(sens*prev + (1-spec)*(1-prev)); prob

## [1] 0.1165644
```

   Thus, based on this information, there is only a 12% probability it will rain tomorrow.

2. A Pareto distribution for the random variable Y has a probability density function

$$p(y|\alpha,\beta) = \frac{\alpha\beta^\alpha}{y^{\alpha+1}}\mathrm{I}(\beta \le y)$$

for shape $\alpha > 0$ and scale $\beta > 0$. For the following, assume you have observed $y_1, \ldots, y_n$ independent observations from the same Pareto distribution.

(a) Suppose α is known and β is given a Pareto prior with shape a and scale b. Derive the posterior for β. (10 points)

Answer:

$$\begin{aligned} p(\beta|y) &\propto p(y|\beta)p(\beta) \\ &\propto \beta^{n\alpha}\mathrm{I}(\beta < \min(y_i))\tfrac{1}{\beta^{a+1}}\mathrm{I}(b < \beta) \\ &= \tfrac{1}{\beta^{a-n\alpha+1}}\mathrm{I}(b < \beta < \min(y_i)) \end{aligned}$$

where $\min(y_i)$ is the minimum of the data. This appears to be a Pareto distribution with shape $a - n\alpha$ and scale b, but truncated to be less than $\min(y_i)$.

(b) Suppose β is known and $\alpha \sim Ga(a,b)$, i.e. $p(\alpha) \propto \alpha^{a-1}e^{-b\alpha}$. Derive the posterior for α. (10 points)

Answer: Let $m = \prod_{i=1}^n y_i$.

$$\begin{aligned} p(\alpha|y) &\propto p(y|\alpha,\beta)p(\alpha) \\ &\propto \left[\prod_{i=1}^n \tfrac{\alpha\beta^\alpha}{y_i^\alpha}\right]\alpha^{a-1}e^{-b\alpha} \\ &= \tfrac{\alpha^n\beta^{n\alpha}}{m^\alpha}\alpha^{a-1}e^{-b\alpha} \\ &= \alpha^{a+n-1}e^{-b\alpha+n\alpha\log\beta+\alpha\log m} \\ &= \alpha^{a+n-1}e^{-\alpha(b-n\log\beta+\log m)} \end{aligned}$$

This is the kernel of $Ga(a+n, b - n\log\beta + \log m)$.

3. Assume the model $Y_i \stackrel{ind}{\sim} N(\mu, \sigma^2)$ $i = 1, \ldots, n$ with prior where the probability density function is $p(y) = (2\pi\sigma^2)^{-1/2} \exp(-[y-\mu]^2/2\sigma^2)$. Assume the prior

$$p(\mu, \sigma^2) \propto \exp\left(-\frac{1}{2C}(\mu - m)^2\right)(\sigma^2)^{-a-1}e^{-b/\sigma^2}.$$

with $a > 0$, $b > 0$, and $C > 0$. Let $y = (y_1, \ldots, y_n)$

(a) Derive the conditional posterior for μ, i.e. $p(\mu|\sigma^2, y)$. (10 points)

Answer: This is exactly the same as having normal data with an unknown mean, but with a known variance equal to σ^2. The prior is independent, i.e. $p(\mu, \sigma^2) = p(\mu)p(\sigma^2)$.

$$\begin{aligned}
p(\mu|\sigma^2, y) &\propto p(y|\mu, \sigma^2)p(\mu)p(\sigma^2) \\
&\propto \exp\left(-\tfrac{1}{2\sigma^2}\textstyle\sum_{i=1}^n (y_i - \mu)^2\right)\exp\left(-\tfrac{1}{2C}(\mu - m)^2\right) \\
&= \exp\left(-\tfrac{1}{2\sigma^2}\textstyle\sum_{i=1}^n (y_i^2 - 2\mu y_i + \mu^2)\right)\exp\left(-\tfrac{1}{2C}(\mu^2 - 2\mu m + m^2)\right) \\
&\propto \exp\left(-\tfrac{1}{2}\left[\left(\tfrac{n}{\sigma^2} + \tfrac{1}{C}\right)\mu^2 + \left(\tfrac{n}{\sigma^2}\overline{y} + \tfrac{1}{C}m\right)\mu\right]\right)
\end{aligned}$$

This is the kernel of normal and thus

$$\mu|\sigma^2, y \sim N(m', C')$$

with

$$\begin{aligned}
C' &= [1/C + n/\sigma^2]^{-1} \\
m' &= C'[m/C + n\overline{y}/\sigma^2].
\end{aligned}$$

(b) Derive the conditional posterior for σ^2, i.e. $p(\sigma^2|\mu, y)$. (10 points)

Answer: This is like a normal distribution with an unknown variance and a known mean μ.

$$\begin{aligned}
p(\sigma^2|\mu, y) &\propto p(y|\mu, \sigma^2)p(\mu)p(\sigma^2) \\
&\propto (\sigma^2)^{-n/2}\exp\left(-\tfrac{1}{2\sigma^2}\textstyle\sum_{i=1}^n (y_i - \mu)^2\right)(\sigma^2)^{-a-1}\exp\left(-b/\sigma^2\right) \\
&\quad (\sigma^2)^{-a-n/2-1}\exp\left(-1/\sigma^2\left[b + \tfrac{1}{2}\textstyle\sum_{i=1}^n (y_i - \mu)^2\right]\right)
\end{aligned}$$

This is the kernel of an inverse gamma and thus

$$\sigma^2|\mu, y \sim IG\left(a + n/2, b + b + \frac{1}{2}\sum_{i=1}^n (y_i - \mu)^2\right).$$

4. For the following questions, please refer to the pages titled R/Stan code and R/Stan output.

(a) Write down the model that is being fit including priors. (14 points)
Answer: The model is

$$\begin{aligned} Y_i &\stackrel{ind}{\sim} N(\theta_{g[i]}, \sigma^2_{g[i]}) \\ \theta_g &\stackrel{ind}{\sim} N(\mu, \tau^2) \\ \sigma_g &\stackrel{ind}{\sim} Ga(\eta\beta, \beta) \end{aligned}$$

for $i = 1, \ldots, n$, $g[i]$ giving the group number for observation i, and $g = 1, \ldots, G$. The priors on the hierarchical parameter are independent with marginal distributions

$$\begin{aligned} p(\mu) &\propto 1 \\ \tau &\sim Ca^+(0, 1) \\ \eta &\sim Ca^+(0, 1) \\ \beta &\sim Ca^+(0, 1) \end{aligned}$$

(b) For **group 1**, provide estimates for the following quantities to 2 decimal places:

i. posterior expectation for the group mean (1 point)
Answer: -0.38

ii. equal-tail 95% credible interval for the group mean (1 point)
Answer: (-1.07,0.31)

iii. posterior expectation for the group standard deviation (2 points)
Answer: 1.38

iv. equal-tail 95% credible interval for the group standard deviation (2 points)
Answer: (0.92,2.14)

5. Consider the hierarchical model

$$Y_{gi} \overset{ind}{\sim} Po(\lambda_g) \quad \text{and} \quad \lambda_g \overset{ind}{\sim} Ga(\alpha, \beta)$$

for groups $g = 1, \ldots, G$ and individuals within a group $i = 1, \ldots, n_i$. Assume some reasonable prior on α and β such that you can obtain posterior samples $\lambda_1^{(m)}, \ldots, \lambda_G^{(m)}, \alpha^{(m)}, \beta^{(m)}$ from the joint posterior $p(\lambda_1, \ldots, \lambda_G, \alpha, \beta|y)$ for $m = 1, \ldots, M$.

(a) The predictive distribution for a new, independent observation from group 1, $\tilde{y}_1$, is

$$p(\tilde{y}_1|y) = \int p(\tilde{y}_1|\lambda_1)p(\lambda_1|y)d\lambda_1.$$

Describe an algorithm to simulate values from this marginal distribution. (10 points)

Answer: For $m = 1, \ldots, M$, sample $\tilde{y}_1^{(m)} \sim Po(\lambda_1^{(m)})$.

(b) State an integral equation (like that in the previous question) that would allow you to find the predictive distribution for a new observation from a new group, $\tilde{y}_*$. (10 points)

Answer:

$$p(\tilde{y}_*|y) = \int \int \int p(\tilde{y}_*|\lambda_*)p(\lambda_*|\alpha, \beta)p(\alpha, \beta|y)d\lambda_* d\alpha d\beta$$

R/Stan code

```
d %>% group_by(group) %>% summarize(n=n(), mean=mean(y), sd=sd(y))
## # A tibble: 10 x 4
##    group     n    mean     sd
##    <fct> <int>   <dbl>  <dbl>
##  1 1        12 -0.528  1.30
##  2 2        10  0.429  0.525
##  3 3         9  0.0559 0.403
##  4 4        11 -0.763  1.73
##  5 5        14 -1.08   0.186
##  6 6        12  1.60   4.78
##  7 7        15  0.204  0.387
##  8 8         7  0.720  0.413
##  9 9        13  0.632  0.0738
## 10 10        9 -0.560  2.76
```

```
library("rstan")

model = "
data {
  int n;
  int G;
  int<lower=1,upper=G> group[n];
  real y[n];
}
parameters {
  real mu;
  real theta[G];
  real<lower=0> eta;
  real<lower=0> tau;
  real<lower=0> beta;
  real<lower=0> sigma[G];
}
model {
  tau ~ cauchy(0,1); eta ~ cauchy(0,1); beta ~ cauchy(0,1);
  theta ~ normal(mu,tau);
  sigma ~ gamma(eta*beta, beta);
  y ~ normal(theta[group],sigma[group]);
}
"

dat = list(y = d$y, group = as.numeric(d$group), n = nrow(d), G = nlevels(d$group))
m = stan_model(model_code = model)
```

R/Stan output

```
res = sampling(m, dat, seed=20180305)
```

```
res
## Inference for Stan model: ee8cacd250b98fbfb7adaa1c54a21039.
## 4 chains, each with iter=2000; warmup=1000; thin=1;
## post-warmup draws per chain=1000, total post-warmup draws=4000.
##
##              mean se_mean   sd   2.5%    25%    50%    75%  97.5% n_eff Rhat
## mu           0.02    0.01 0.27  -0.52  -0.14   0.02   0.19   0.55  2843    1
## theta[1]    -0.38    0.01 0.35  -1.07  -0.62  -0.39  -0.15   0.31  4000    1
## theta[2]     0.40    0.00 0.18   0.02   0.28   0.40   0.51   0.76  4000    1
## theta[3]     0.06    0.00 0.15  -0.24  -0.04   0.05   0.15   0.37  4000    1
## theta[4]    -0.46    0.01 0.45  -1.35  -0.75  -0.47  -0.15   0.45  4000    1
## theta[5]    -1.07    0.00 0.05  -1.18  -1.11  -1.07  -1.04  -0.96  4000    1
## theta[6]     0.42    0.01 0.67  -0.81  -0.04   0.40   0.82   1.87  3186    1
## theta[7]     0.20    0.00 0.11  -0.02   0.13   0.20   0.27   0.41  4000    1
## theta[8]     0.67    0.00 0.19   0.28   0.56   0.67   0.79   1.02  3184    1
## theta[9]     0.63    0.00 0.02   0.59   0.62   0.63   0.65   0.68  3787    1
## theta[10]   -0.20    0.01 0.58  -1.35  -0.57  -0.20   0.16   0.94  3261    1
## eta          1.31    0.01 0.44   0.68   1.00   1.23   1.53   2.40  2434    1
## tau          0.71    0.01 0.23   0.41   0.55   0.67   0.82   1.31  2047    1
## beta         0.79    0.01 0.37   0.26   0.52   0.73   1.00   1.67  3101    1
## sigma[1]     1.38    0.01 0.32   0.92   1.16   1.32   1.55   2.14  3334    1
## sigma[2]     0.59    0.00 0.16   0.37   0.48   0.56   0.66   0.98  2791    1
## sigma[3]     0.46    0.00 0.13   0.28   0.37   0.43   0.52   0.77  3184    1
## sigma[4]     1.81    0.01 0.40   1.21   1.52   1.74   2.03   2.80  4000    1
## sigma[5]     0.20    0.00 0.04   0.14   0.17   0.20   0.23   0.31  3424    1
## sigma[6]     4.55    0.01 0.89   3.19   3.92   4.42   5.04   6.70  4000    1
## sigma[7]     0.42    0.00 0.09   0.29   0.36   0.40   0.47   0.63  4000    1
## sigma[8]     0.50    0.00 0.18   0.28   0.38   0.46   0.57   0.93  2058    1
## sigma[9]     0.08    0.00 0.02   0.06   0.07   0.08   0.09   0.13  2905    1
## sigma[10]    2.71    0.01 0.64   1.79   2.26   2.62   3.03   4.28  3151    1
## lp__       -26.24    0.10 3.76 -34.46 -28.55 -25.96 -23.56 -19.81  1448    1
##
## Samples were drawn using NUTS(diag_e) at Wed Mar  7 16:47:32 2018.
## For each parameter, n_eff is a crude measure of effective sample size,
## and Rhat is the potential scale reduction factor on split chains (at
## convergence, Rhat=1).
```