Spring 2015 **Statistics 544** **Midterm Exam**
(100 points)

Please write your first and last name here:

Name ____

Instructions:

- Do NOT open the exam until instructed to do so!
- Please check to make sure you have 3 pages with writing on the front and back (pages may be marked 'intentionally left blank'). Feel free to remove (and keep) pages after number 6.
- On the following pages you will find short answer questions related to the topics we covered in class for a total of 100 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.
- You will be given only the time allotted for the course; no extra time will be given.

1. A hit-and-run taxi accident occurs at dusk in a city that has 90% black taxis, 5% green taxis, and 5% red taxis. The only eyewitness has told investigators that he believes the taxi was red. Investigators test the eyewitness's vision in this low light scenario and find the following identification probabilities, e.g. if the taxi was actually green, the eyewitness identified it as being black 20% of the time.

	Truth		
Identification	Black	Green	Red
Black	0.8	0.2	0.1
Green	0.1	0.7	0.4
Red	0.1	0.1	0.5

(a) Calculate the probability the car was black. (5 pts)

Let B, G, R indicate the car was truly blue, green, or red, respectively. Then let I_R indicate that the eyewitness identified the car as red. First calculate

$$P(I_R) = P(I_R|B)P(B) + P(I_R|G)P(G) + P(I_R|R)P(R) = .1 \times .9 + .1 \times .05 + .5 \times .05 = 0.12$$

$$P(B|I_R) = \frac{P(I_R|B)P(B)}{P(I_R)} = .1 \times .9/.12 \approx 0.75$$

(b) Calculate the probability the car was green. (5 pts)

$$P(G|I_R) = \frac{P(I_R|G)P(G)}{P(I_R)} = .1 \times .05/.12 \approx 0.04$$

(c) Calculate the probability the car was red. (5 pts)

$$P(R|I_R) = \frac{P(I_R|R)P(R)}{P(I_R)} = .5 \times .05/.135 \approx 0.21$$

(d) Suppose there was a different eyewitness, fill in their identification probabilities so that, if they said the taxi was red, the probability that the car was red is greater than the probabilities that the car was green or black. (5 pts)

The easiest way is to make the eyewitness perfect, i.e. $P(I_B|B) = P(I_G|G) = P(I_R|R) = 1$.

	Truth		
Identification	Black	Green	Red
Black			
Green			
Red			

2. Let $Y \sim Geo(\theta)$, i.e. a geometric random variable with probability of success θ and $E[y] = \frac{1}{\theta}$. The geometric probability mass function is

$$p(y|\theta) = (1-\theta)^{y-1}\theta, \quad y \in \{1, 2, 3, \ldots\}, \, 0 < \theta < 1$$

(a) Derive Jeffreys prior for θ and determine whether this prior is proper. (10 pts)
The geometric is an exponential family, thus we have

$$\begin{aligned}
\log p(y|\theta) &= (y-1)\log(1-\theta) + \log(\theta) \\
\frac{d}{d\theta}\log p(y|\theta) &= -\frac{y-1}{1-\theta} + \frac{1}{\theta} \\
\frac{d^2}{d\theta^2}\log p(y|\theta) &= -\frac{y-1}{(1-\theta)^2} - \frac{1}{\theta^2} \\
\mathcal{I}(\theta) &= -\,E_y\left[\frac{d^2}{d\theta^2}\log p(y|\theta)\right] \\
&= -\frac{E_y[y]-1}{(1-\theta)^2} - \frac{1}{\theta^2} \\
&= \frac{\frac{1}{\theta}-1}{(1-\theta)^2} + \frac{1}{\theta^2} \\
&= \theta^{-2}(1-\theta)^{-1} \\
p(\theta) &\propto \sqrt{\mathcal{I}(\theta)} = \theta^{-1}(1-\theta)^{-1/2}
\end{aligned}$$

as this would be a $Be(0, 1/2)$ it is not proper.

(b) Derive the posterior under Jeffreys prior with n independent observations $(y_1, \ldots, y_n)$ and determine when it is proper. (10 pts)

$$\begin{aligned}
p(\theta|y) &\propto p(y|\theta)p(\theta) \\
&= \left[\prod_{i=1}^{n}(1-\theta)^{y-1}\theta\right]\theta^{-1}(1-\theta)^{-1/2} \\
&= \theta^{n-1}(1-\theta)^{n(\overline{y}-1)+\frac{1}{2}-1}
\end{aligned}$$

This is the kernel of a $Be(n, n(\overline{y}-1)+1/2)$ which is proper so long as $n > 0$ since $\overline{y} \geq 1$.

(c) Derive the posterior predictive mass function for a new single observation $\tilde{y} \sim Geo(\theta)$, conditionally independent of the previous observations. Make sure to indicate the support. (10 pts)

The support is $\tilde{y} \in \{1, 2, 3, \ldots\}$.

$$\begin{aligned} p(\tilde{y}|y) & \int p(\tilde{y}|\theta)p(\theta|y)d\theta \\ &= \int (1-\theta)^{\tilde{y}-1}\theta \frac{\theta^{n-1}(1-\theta)^{n(\overline{y}-1)+\frac{1}{2}-1}}{B(n, n(\overline{y}-1)+1/2)} d\theta \\ &= \frac{1}{B(n, n(\overline{y}-1)+1/2)} \int \theta^{n+1-1}(1-\theta)^{n(\overline{y}-1)+\tilde{y}-\frac{1}{2}-1} d\theta \\ &= \frac{B(n+1, n(\overline{y}-1)+\tilde{y}-1/2)}{B(n, n(\overline{y}-1)+1/2)} \end{aligned}$$

where

$$\frac{1}{B(\alpha,\beta)} = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}$$

This is the beta negative binomial distribution.

(d) Based on the data $y = (y_1, \ldots, y_n)$, describe a procedure to simulate K replicate data sets to be used in a posterior predictive check. (10 pts)

Note that

$$p(y^{rep}|y) = \int \left[\prod_{i=1}^{n} p(y_i^{rep}|\theta)\right] p(\theta|y)d\theta$$

So the following is a an appropriate procedure:

1. For $k = 1, \ldots, K$,
   a. Simulate $\theta^{(k)} \sim p(\theta|y)$, if we use Jeffreys prior, it is $Be(n, n(\overline{y}-1)+1/2)$.
   b. For $i = 1, \ldots, n$, simulate $\tilde{y}_i^{(k)} \sim Geo\left(\theta^{(k)}\right)$.

3. For the following questions, please refer to the two pages on "Stan".

(a) Using statistical notation, write down model 1. (5 pts)

$$Y_i \overset{ind}{\sim} N(\mu, \sigma^2),\ i = 1, \ldots, n \qquad p(\mu, \sigma^2) \propto IG(\sigma^2; 1, 1)$$

(b) Using statistical notation, write down model 2. (5 pts)

$$Y_i \overset{ind}{\sim} N(\mu, \sigma^2),\ i = 1, \ldots, n \qquad p(\mu, \sigma) \propto Ca^+(\sigma; 0, 1)$$

(c) Provide a 95% credible interval for the data mean and standard deviation from model 1. (5 pts)

- mean (0.69, 1.29)
- standard deviation (0.31,0.71)

(d) Provide a 95% credible interval for the data mean and standard deviation from model 2. (5 pts)

- mean (0.93,1.04)
- standard deviation (0.05,0.14)

(e) Which prior for the standard deviation is more reasonable and why? (5 pts)

The inverse gamma distribution has a region near zero that has extremely low density. This is causing severe bias in estimation for σ in model 1 and thus model 2 is preferred.

(f) Derive the conditional posterior for the mean given the standard deviation, i.e. $p(\mu|\sigma, y)$ where $y = (y_1, \ldots, y_n)$. Name the distribution family and its parameters. (10 pts)

This is just normal data with an unknown mean but known variance.

$$\begin{aligned} p(\mu|\sigma, y) &= \left[\prod_{i=1}^{n} N(y_i|\mu, \sigma^2)\right] \times 1 \\ &\propto \exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i - \mu)^2\right) \\ &= \exp\left(-\frac{1}{2\sigma^2}\left[n\mu^2 - 2\mu n\overline{y}\right]\right) \\ &= \exp\left(-\frac{1}{2\sigma^2/n}\left[\mu^2 - 2\mu\overline{y}\right]\right) \end{aligned}$$

This is the kernel of a normal distribution with mean $\overline{y}$ and variance σ^2/n, thus $\mu|\sigma^2, y \sim N(\overline{y}, \sigma^2/n)$.

(g) Based on this conditional posterior, explain why the credible interval for the mean in model 1 is so much wider than in model 2. (5 pts)

In model 1, σ is estimated to be much larger than in model 2. Thus the credible interval for μ, which is $\overline{y} \pm 1.96\sigma^2/n$ (if σ^2 were known), will be wider. (Of course, the marginal uncertainty for μ will incorporate the uncertainty in σ. But the uncertainty in σ is small compared with the estimation bias.)

A Stan

```
model1 = "
data {
  int<lower=1> n;
  real y[n];
}
parameters {
  real mu;
  real<lower=0> tau;
}
transformed parameters {
  real<lower=0> sigma;
  sigma <- 1/sqrt(tau);
}
model {
  tau ~ gamma(1,1);
  y ~ normal(mu,sigma);
}
"

model2 = "
data {
  int<lower=1> n;
  real y[n];
}
parameters {
  real mu;
  real<lower=0> sigma;
}
model {
  sigma ~ cauchy(0,1);
  y ~ normal(mu,sigma);
}
"

m1 = stan_model(model_code=model1)
m2 = stan_model(model_code=model2)
```

Stan (continued)

```
dat = list(n=length(y), y=y)

r1 = sampling(m1, dat, c("mu","sigma"), seed=1)
r2 = sampling(m2, dat, c("mu","sigma"), seed=1)
```

```
r1

## Inference for Stan model: e3b63a84f14810e53fdeb3aac91e6aec.
## 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    0.93    0.00 0.15 0.63 0.84 0.93 1.03  1.24  1963    1
## sigma 0.47    0.00 0.11 0.31 0.39 0.45 0.52  0.74  1525    1
## lp__  3.44    0.03 1.11 0.39 3.03 3.78 4.21  4.50  1091    1
##
## Samples were drawn using NUTS(diag_e) at Mon Jan 11 11:47:52 2016.
## 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).

r2

## Inference for Stan model: afd6240a58bdfffc484662e4455f9b83.
## 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     0.93    0.00 0.04  0.86  0.91  0.93  0.95  1.00  1607    1
## sigma  0.11    0.00 0.03  0.07  0.09  0.10  0.12  0.19  1481    1
## lp__  15.77    0.03 1.13 12.59 15.38 16.11 16.55 16.85  1146    1
##
## Samples were drawn using NUTS(diag_e) at Mon Jan 11 11:47:52 2016.
## 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).
```

B Distributions

The table below provides the details of some common distributions. In all cases, the random variable is θ.

Distribution	Density or mass function	Moments
$\theta \sim N(\mu, \sigma^2)$	$\frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2}(\theta-\mu)^2\right)$	$E[\theta] = \mu$
		$V[\theta] = \sigma^2$
$\theta \sim Ga(\alpha, \beta)$	$\frac{\beta^\alpha}{\Gamma(\alpha)}\theta^{\alpha-1}e^{-\beta\theta}, \theta > 0$	$E[\theta] = \alpha/\beta$
		$V[\theta] = \alpha/\beta^2$
$\theta \sim IG(\alpha, \beta)$	$\frac{\beta^\alpha}{\Gamma(\alpha)}\theta^{-(\alpha+1)}e^{-\beta/\theta}, \theta > 0$	$E[\theta] = \beta/(\alpha-1), \alpha > 1$
		$V[\theta] = \beta^2/[(\alpha-1)^2(\alpha-2)], \alpha > 2$
$\theta \sim Be(\alpha, \beta)$	$\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\theta^{\alpha-1}(1-\theta)^{\beta-1}, 0 < \theta < 1$	$E[\theta] = \alpha/(\alpha+\beta)$
		$V[\theta] = \alpha\beta/[(\alpha+\beta)^2(\alpha+\beta+1)]$
$\theta \sim t_\nu(\mu, \sigma^2)$	$\frac{\Gamma([\nu+1]/2)}{\Gamma(\nu/2)}\left(1 + \frac{1}{\nu}\left[\frac{\theta-\mu}{\sigma}\right]^2\right)^{-(\nu+1)/2}$	$E[\theta] = \mu, \nu > 1$
		$V[\theta] = \frac{\nu}{\nu-2}\sigma^2, \nu > 2$
$\theta \sim Geo(\pi)$	$(1-\pi)^{\theta-1}\pi, \theta \in \{1, 2, 3, \ldots\}$	$E[\theta] = 1/\pi$
		$V[\theta] = (1-\pi)/\pi^2$

(intentionally left blank)