Student Name: ____

STAT 544 Mid-term Exam
Thursday 14 March 2019, 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 and distributions.

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. A machine learning algorithm has been set up to distinguish between benign and malignant tumors, but due to difficulties obtaining training data a third category called *unknown* is included.

Table 1: Prevalence and sensitivity/specificity for the algorithm.

		Probability		
Truth	Prevalence	benign	malignant	unknown
benign	0.7	0.9	0.1	0.0
malignant	0.1	0.1	0.8	0.1
unknown	0.2	0.2	0.2	0.6

Table 1 provides the prevalence for the three cancer types as well as the effectiveness of the algorithm to identify the correct tumor type. For example, if the tumor is truly benign, the algorithm will indicate benign 90% of the time, malignant 10% of the time, and will never indicate unknown.

What is the probability a tumor is actually malignant if the algorithm indicates the tumor is malignant? (20 points)

Answer: Let

- T_m, T_b, and T_u indicate the tumor is malignant, benign, and unknown respectively
- A_m, A_b, and A_u indicate the algorithm indicates malignant, benign, and unknown respectively

$$
\begin{aligned}
P(T_m|A_m) &= \frac{P(A_m|T_m)p(T_m)}{P(A_m|T_m)p(T_m)+P(A_m|T_b)p(T_b)+P(A_m|T_u)p(T_u)} \\
&= \frac{0.8\times 0.1}{0.8\times 0.1+0.1\times 0.7+0.2\times 0.2} \\
&= 0.4210526
\end{aligned}
$$

2. In point counts for birds, you know how many birds you detected, but not the number of birds present. Let Y be the number of birds detected and assume $Y \sim Bin(\eta, p)$ where the number of birds present η is unknown and the probability of success p is known. Assume $\eta \sim Po(m)$.

   (a) Derive the posterior for η. (10 points)

   Answer:

$$\begin{aligned} p(\eta|y) &\propto p(y|\eta)p(\eta) \\ &\propto \frac{\eta!}{(\eta-y)!}(1-p)^{\eta-y}\frac{m^\eta}{\eta!} \\ &\propto \frac{[m(1-p)]^{\eta-y}}{(\eta-y)!} && \text{since } (1p)^{-y} \text{ doesn't depend on } \eta \end{aligned}$$

   Since the binomial requires $y \leq \eta$, η has support on integers greater or equal to y. Thus, we will perform the transformation $\mu = \eta - y$ which has a Jacobian of 1 and then μ will have support on the non-negative integers.

$$p(\mu|y) \propto \frac{[m(1-p)]^\mu}{\mu!}$$

   This is the kernel of a Poisson random variable with mean $m(1-p)$. Thus, $\eta = \mu + y$ where $\mu \sim Po(m[1-p])$.

   (b) Describe how to sample from the posterior for η. (2 points)

   Answer: Sample $\mu \sim Po(m[1-p])$ and calculate $\eta = y + \mu$.

   (c) Suppose $p = 1$, state the posterior for η. (2 points)

   Answer: When $p = 1$, $P(\eta = y|y) = 1$.

   (d) Determine the posterior probability that the number of birds present is less than 5 when $y = 3$, $p = 0.7$, and $m = 10$. (6 points)

   Answer: Calculate $m(1-p) = 10 * (1 - 0.7) = 3$.

$$\begin{aligned} P(\eta < 5|y) &= P(\mu < 2|y) = P(\mu \leq 1|y) \\ &= e^{-m(1-p)}\left(1 + m[1-p]\right) \\ &= e^{-3}(1+3) \\ &= 0.1991483 \end{aligned}$$

3. The internet movie database (IMDb) provides users the following `star rating`:

```
star rating (SR) = (v / (v+m)) * R + (m / (v+m)) * C

where:

R = average for the movie (mean) = (Rating)
v = number of votes for the movie = (votes)
m = minimum votes required to be listed in the Top 250 (currently 25000)
C = the mean vote across the whole report (currently 7.1)
```

Let y_i be an individual users rating and assume $y_i \stackrel{ind}{\sim} N(\mu, v^2)$ with the prior $\mu \sim N(a, b^2)$.

(a) Relate the following quantities in the `star rating` formula to the associated quantities in this normal model. (2 points each)

- `v`=
  Answer: n/v^2

- `m`=
  Answer: $1/b^2$

- `R`=
  Answer: $\overline{y}$

- `C`=
  Answer: a

- `SR`=
  Answer: $E[\mu|y]$

(b) Let $y_{im} \in \{1, 2, 3, 4, 5\}$ be the rating for individual i for movie m for $i = 1, 2, \ldots, n_m$ and $m = 1, \ldots, M$. Assume each individual only rates one movie. Construct a hierarchical model for these data that has the proper support for y_{im}. (10 points)

Answer: A variety of answers could work here. One possibility is to let $z_{im} = y_{im} - 1$ and make the following assumptions

$$\begin{aligned} z_{im} &\stackrel{ind}{\sim} Bin(4, \theta_{im}) \\ \theta_{im} &\stackrel{ind}{\sim} Be(\alpha, \beta) \\ p(\alpha, \beta) &\propto (\alpha + \beta)^{-5/2} \end{aligned}$$

4. Determine the Bayes estimator, $\hat{\theta}$, for *scaled* squared error loss, $L(\theta,\hat{\theta}) = c(\theta-\hat{\theta})^2$ for some $c>0$ and $\theta \in \mathbb{R}$. (10 points)

   Answer: Intuitively, multiplying the squared error loss by a constant will have no affect on the estimator that minimizes the loss, so $\hat{\theta} = E[\theta|y]$ will minimize scaled squared error loss. To show this

$$\begin{aligned}
E[L(\theta,\hat{\theta})|y] &= E[c\theta^2 - 2c\hat{\theta}\theta + c\hat{\theta}^2|y] \\
&= cE[\theta^2|y] - 2c\hat{\theta}E[\theta|y] + c\hat{\theta}^2 \\
\frac{d}{d\hat{\theta}}E[L(\theta,\hat{\theta})|y] &= -2cE[\theta|y] + 2c\hat{\theta} \\
\frac{d^2}{d\hat{\theta}^2}E[L(\theta,\hat{\theta})|y] &= 2c
\end{aligned}$$

   Setting the first derivative to 0 results in $\hat{\theta} = E[\theta|y]$ and the second derivative being positive means this estimate minimizes the scaled squared error loss.

5. Let $Y_i \stackrel{ind}{\sim} Ber(\theta)$ for $1,\ldots,n$ with $\theta \sim Be(a,b)$. Let $\tilde{Y} \sim Ber(\theta)$ independent of $Y_1,\ldots,Y_n$. Determine the Bayes estimator in (0,1) for squared error loss of $\tilde{Y}$. (10 points)

   Answer: We know the posterior expectation minimizes squared error loss (see above). Thus, using iterated expectations, we have

$$E[\tilde{Y}|y] = E[E[\tilde{Y}|\theta]|y] = E[\theta|y] = \frac{a+n\overline{y}}{a+b+n}.$$

6. The Boeing 737 MAX aircraft has 9 generations. The Federal Aviation Administration (FAA) has grounded generations 8 and 9 due to two recent crashes of the Boeing 737 MAX 8. It appears that in both of these crashes, the plane stalled and the software to automatically reverse the stall failed. To assess the rate of stalls, Boeing builds a hierarchical model for the first stall (in months) for each aircraft in the 737 MAX fleet. Use the `R/Stan Code` on the following page to answer the following questions.

   (a) Write down the model that is being fit including priors. (12 points)

   Answer: Let Y_i be the time to first stall for observation i and $g_i \in \{1, 2, \dots, 9\}$ be the generation number for observation i. The model is

$$\begin{aligned} Y_i &\stackrel{ind}{\sim} Exp(\lambda_{g_i}) \\ \lambda_g &\stackrel{ind}{\sim} Ga(\mu\beta, \beta) \\ p(\mu, \beta) &= p(\mu)p(\beta) \\ \mu &\sim Exp(0.1) \\ \beta &\sim Exp(1) \end{aligned}$$

   (b) For **generation 8**, provide estimates for the following quantities to 2 decimal places:

      i. posterior expectation for the first stall rate (1 point)
      Answer: 0.16

      ii. equal-tail 95% credible interval for first stall rate (2 points)
      Answer: (0.09,0.24)

      iii. equal-tail 95% credible interval for first stall **mean** (2 points)
      Answer: We just need to invert the endpoints of the previous interval, so (1/0.24,1/0.09) = (4.17,11.11)

   (c) Suppose you have samples $\lambda_1^{(m)}, \dots, \lambda_9^{(m)}$ for $m = 1, \dots, M$ from the joint posterior for the generation first stall rates. Describe how you would estimate the posterior probability the **mean** time to first stall for generation 8 is smaller than for generation 7. (3 points)

   Answer: Let $\mu_g = 1/\lambda_g$ be the mean time to first stall for generation g. Then

$$\begin{aligned} P(\mu_8 < \mu_7|y) &= P(1/\lambda_8 < 1/\lambda_7|y) \\ &= P(\lambda_8 > \lambda_7|y) \\ &\approx \frac{1}{M}\mathrm{I}\left(\lambda_8^{(m)} > \lambda_7^{(m)}\right) \end{aligned}$$

R/Stan Code

```
m = "
data {
  int N;
  int n_generations;
  int<lower=1, upper=n_generations> generation[N];
  real<lower=0> first_stall[N];
}
parameters {
  real<lower=0> mu;
  real<lower=0> beta;
  real<lower=0> lambda[n_generations];
}
transformed parameters {
  real<lower=0> alpha;
  alpha = mu*beta;
}
model {
  mu ~ exponential(0.1);
  beta ~ exponential(1);
  lambda ~ gamma(alpha, beta);
  first_stall ~ exponential(lambda[generation]);
}
"
```

```
Inference for Stan model: b5d64f40602493924205dde199093bd0.
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	0.28	0.00	0.15	0.13	0.20	0.25	0.33	0.61	1765	1
beta	3.55	0.03	1.88	0.81	2.16	3.24	4.60	8.05	3305	1
lambda[1]	0.09	0.00	0.03	0.05	0.07	0.09	0.11	0.16	4923	1
lambda[2]	0.11	0.00	0.03	0.06	0.09	0.11	0.13	0.18	4912	1
lambda[3]	0.24	0.00	0.08	0.12	0.18	0.23	0.29	0.42	4957	1
lambda[4]	0.08	0.00	0.02	0.04	0.06	0.07	0.09	0.13	5401	1
lambda[5]	0.06	0.00	0.02	0.03	0.05	0.06	0.07	0.10	4906	1
lambda[6]	0.22	0.00	0.07	0.11	0.18	0.22	0.26	0.38	5483	1
lambda[7]	0.07	0.00	0.02	0.04	0.06	0.07	0.09	0.13	5621	1
lambda[8]	0.16	0.00	0.04	0.09	0.13	0.15	0.18	0.24	4789	1
lambda[9]	0.13	0.00	0.03	0.07	0.11	0.13	0.15	0.20	5527	1
alpha	0.85	0.01	0.33	0.35	0.61	0.80	1.04	1.63	3540	1
lp__	-348.99	0.06	2.41	-354.68	-350.37	-348.63	-347.26	-345.22	1835	1

```
Samples were drawn using NUTS(diag_e) at Thu Mar 14 06:45:26 2019.
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).
```

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
$\theta \sim Exp(\lambda)$	$\lambda e^{-\lambda\theta}$	$E[\theta] = 1/\lambda$ $V[\theta] = 1/\lambda^2$
$\theta \sim N(\mu, \sigma^2)$	$\frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2}(\theta-\mu)^2\right)$	$E[\theta] = \mu$ $V[\theta] = \sigma^2$
$\theta \sim Ga(\alpha, \beta)$	$\frac{\beta^\alpha}{\Gamma(\alpha)}\theta^{\alpha-1}e^{-\beta\theta}, \theta > 0$	$E[\theta] = \alpha/\beta$ $V[\theta] = \alpha/\beta^2$
$\theta \sim IG(\alpha, \beta)$	$\frac{\beta^\alpha}{\Gamma(\alpha)}\theta^{-(\alpha+1)}e^{-\beta/\theta}, \theta > 0$	$E[\theta] = \beta/(\alpha-1), \alpha > 1$ $V[\theta] = \beta^2/[(\alpha-1)^2(\alpha-2)], \alpha > 2$
$\theta \sim Be(\alpha, \beta)$	$\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\theta^{\alpha-1}(1-\theta)^{\beta-1}, 0 < \theta < 1$	$E[\theta] = \alpha/(\alpha+\beta)$ $V[\theta] = \alpha\beta/[(\alpha+\beta)^2(\alpha+\beta+1)]$
$\theta \sim t_\nu(\mu, \sigma^2)$	$\frac{\Gamma([\nu+1]/2)}{\Gamma(\nu/2)}\left(1 + \frac{1}{\nu}\left[\frac{\theta-\mu}{\sigma}\right]^2\right)^{-(\nu+1)/2}$	$E[\theta] = \mu, \nu > 1$ $V[\theta] = \frac{\nu}{\nu-2}\sigma^2, \nu > 2$
$\theta \sim Geo(\pi)$	$(1-\pi)^{\theta-1}\pi, \theta \in \{1, 2, 3, \ldots\}$	$E[\theta] = 1/\pi$ $V[\theta] = (1-\pi)/\pi^2$
$\theta \sim Bin(n, \pi)$	$\frac{n!}{(n-\theta)!\theta!}\pi^\theta(1-\pi)^{n-\theta}, \theta \in \{0, 1, \ldots, n\}$	$E[\theta] = n\pi$ $V[\theta] = n\pi(1-\pi)$
$\theta \sim Po(\lambda)$	$\frac{e^{-\lambda}\lambda^\theta}{\theta!}, \theta \in \{0, 1, 2, \ldots\}$	$E[\theta] = \lambda$ $V[\theta] = \lambda$