This Exam consists of 11 questions. Do at least 10 of the 11 parts of the main exam. I will score your best 10 answers at 10 points apiece (making 100 points possible).

There is also on the last page of the Exam an "Extra Credit" question that will be scored out of 10 points. Any Extra Credit obtained will be recorded and used at the end of the course at Vardeman's discretion in deciding borderline grades. DO NOT spend time on this question until you are done with the entirety of the regular exam.
1. Below are three pdfs for $X$, $f(x|1)$, $f(x|2)$, and $f(x|3)$. Use them in the rest of this question.

<table>
<thead>
<tr>
<th>$x$</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>.20</td>
<td>.06</td>
<td>.05</td>
<td>.15</td>
<td>.15</td>
<td>.10</td>
<td>.05</td>
</tr>
<tr>
<td>2</td>
<td>.10</td>
<td>.03</td>
<td>.05</td>
<td>.25</td>
<td>.05</td>
<td>.20</td>
<td>.15</td>
</tr>
<tr>
<td>3</td>
<td>.25</td>
<td>.15</td>
<td>.05</td>
<td>.25</td>
<td>.15</td>
<td>.05</td>
<td>.10</td>
</tr>
</tbody>
</table>

\[ P(x) = .4f(x|1) + .3f(x|2) + .3f(x|3) \]

b) Identify a most powerful size $\alpha = .15$ test of $H_0: \theta = 1$ vs $H_1: \theta = 2$.

\[ \phi(x) = \begin{cases} 1 & \text{if } x = 3,6 \\ .5 & \text{if } x = 7 \\ 0 & \text{if } x = 1,2,4,5 \end{cases} \]

\[ E\phi(x) = .5(.05+.05) + \frac{1}{2}(.10) = .15 \text{ as desired} \]
c) Find a 0-1 loss Bayes test of $H_0 : \theta = 1$ vs $H_1 : \theta = 2$ or 3 for a prior distribution with 
$g(1) = .4, g(2) = .3,$ and $g(3) = .3$. (Give all 7 values of $\phi(x)$.)

$$
\phi(x) = \begin{cases} 
1 & \text{if } x=3,5,6,7 \\
0 & \text{if } x=1,2,4 
\end{cases}
$$

See the computations on the table. Compare $\frac{4f(x|1)}{1} \text{ to } \frac{.3f(x|2) + .3f(x|3)}{1}$

2. In this problem we'll use the $\text{Exp}(\lambda)$ distribution with pdf $f(x | \lambda) = \lambda \exp(-\lambda x) I[x > 0]$. You may use without proof the facts that

- if $X \sim \text{Exp}(\lambda)$ and $t > 0$ then $P[X > t] = \exp(-\lambda t)$,
- if $X_i \sim \text{Exp}(\lambda)$ independent of $X_2 \sim \text{Exp}(\lambda_2)$ then $Y \equiv \min (X_1, X_2) \sim \text{Exp}(\lambda_1 + \lambda_2)$

In a so-called "competing risks" context, an individual or item has a lifetime $Z = \min (U, V)$ where $U$ and $V$ are positive times to failure/death from two different causes.

a) For $i = 1, \ldots, n$ model $U_i \sim \text{Exp}(\lambda_i)$ and $V_i \sim \text{Exp}(\lambda_2)$ with all $U$'s and $V$'s independent. Suppose that what is observed are the iid pairs $W_i = (Z_i, I[Z_i = U_i])$. (Note that $I[Z_i = U_i] = 1$ means that what is observed is the value of $U_i$ and the fact that $U_i < V_i$.) Give likelihood terms $f(w | \lambda_1, \lambda_2)$ for observed $w_i = (z_i, 1)$ and $w_i = (z_i, 0)$.

$$f((z_i, 1)| \lambda_1, \lambda_2) = \lambda_1 \exp(-\lambda_1 z_i) \exp(-\lambda_2 z_i)$$

$$f((z_i, 0)| \lambda_1, \lambda_2) = \exp(-\lambda_1 z_i) \lambda_2 \exp(-\lambda_2 z_i)$$

(Note, e.g. That $P[U < V] = \int_0^\infty \int_0^{\infty} \lambda_1 \exp(-\lambda_1 u) \lambda_2 \exp(-\lambda_2 v) dv \; du = \frac{\lambda_1}{\lambda_1 + \lambda_2}$

So $P[U < t | U < V] = \int_0^t \exp(-\lambda_2 u) \lambda_1 \exp(-\lambda_1 u) \; du / \left( \frac{\lambda_1}{\lambda_1 + \lambda_2} \right)$

$\frac{d}{dt} \left( \text{above} \right) = \frac{\lambda_1 \exp(-\lambda_1 \lambda_2 t)}{\left( \frac{\lambda_1}{\lambda_1 + \lambda_2} \right)}$ and so the likelihood term for the $w_2 = 1$ case is this time $\left( \frac{\lambda_1}{\lambda_1 + \lambda_2} \right)$.
Sometimes, a cause of failure may not be recorded and thus only \( Z_i \) (and not \( W_i \)) is known. Suppose that information on \( n = 5 \) individuals/items is \( W_1 = (3, 1), W_2 = (7, 1), W_3 = 2, W_4 = (3, 0) \), and \( W_5 = (1, 0) \).

Suppose further that a Bayesian uses a prior for \( (\lambda_1, \lambda_2) \) that is one of independence with both \( \lambda_1 \) and \( \lambda_2 \) \( a \ priori \) \( \text{Exp}(1) \) distributed. Carefully describe a Gibbs sampling algorithm for generating triples \( \left( (\lambda_1)_{j+1}, (\lambda_2)_{j+1}, (W_{3,2})_{j+1}^* \right) \) (iterates for the 2 rates and the unobserved indicator). If it is possible to name a distribution from which a given update must be sampled, do so. At a minimum, give a form for each univariate update distribution up to a multiplicative constant.

If one has available \( W_{3,2} = \mathbb{I}[Z_3 = Y_3] \) the likelihood is

\[
\prod_{i=1}^{5} f(W_i | \lambda_1, \lambda_2) = \lambda_1^{W_1} \exp(-3\lambda_1) \times \lambda_2^{W_2} \exp(-3\lambda_2) \\
\times \frac{1}{\lambda_1^{W_1} \lambda_2^{W_2}} \exp(-7\lambda_1) \times \exp(-7\lambda_2) \\
\times \frac{1}{\lambda_1^{W_1} \lambda_2^{W_2}} \exp(-2\lambda_1) \times \exp(-2\lambda_2) \\
\times \lambda_1^{W_3} \exp(-3\lambda_1) \times \lambda_2^{W_4} \exp(-3\lambda_2) \\
\times \lambda_1^{W_5} \exp(-1\lambda_1) \times \exp(-1\lambda_2) \\
\times \lambda_1^{W_5} \exp(-1\lambda_1) \times \exp(-1\lambda_2)
\]

and \( g(\lambda_1, \lambda_2) = \exp(-\lambda_1 - \lambda_2) \), so the posterior is proportional to

\[
h(\lambda_1, \lambda_2 | W) = \lambda_1^{2 + W_{3,2}} \lambda_2^{3 + (1 - W_{3,2})} \exp(-16\lambda_1) \exp(-16\lambda_2)
\]

So Gibbs updates (say in the order \( \lambda_1, \lambda_2, W \)) are

\[
(\lambda_1)_{j+1}^* \sim \Gamma(3 + (W_{3,2})_{j+1}^*, 17) \\
(\lambda_2)_{j+1}^* \sim \Gamma(4 - (W_{3,2})_{j+1}^*, 17)
\]

and \( (W_{3,2})_{j+1}^* \) is \( \text{Ber} \) where the "success probability" \( p \) has

\[
p = \frac{\lambda_1^*_{j+1}}{1 - p} = \frac{\lambda_1^*_{j+1}}{(\lambda_2)_{j+1}^*} \quad \text{i.e.} \quad p = \frac{(\lambda_1)_{j+1}^*}{(\lambda_1)_{j+1}^* + (\lambda_2)_{j+1}^*}
\]

i.e.

\[
(W_{3,2})_{j+1}^* \sim \text{Ber} \left( \frac{(\lambda_1)_{j+1}^*}{(\lambda_1)_{j+1}^* + (\lambda_2)_{j+1}^*} \right)
\]
c) Completely describe an EM algorithm that can be used to find an MLE of \((\lambda_1, \lambda_2)\) based on the data used in part b). (It is not really necessary to resort to EM here, as the calculus problem is fairly easy. But for purposes of the exam, write out the EM algorithm.)

The likelihood based on all \(w_i\)'s is as before

\[
\prod_{i=1}^{n} \frac{2 + w_{3,i} - w_{3,i}}{\lambda_1^{2+w_{3,i}} \lambda_2^{3-w_{3,i}}} \exp(-16\lambda_1) \exp(-16\lambda_2)
\]

with logarithm

\[
(2 + w_{3,i}) \log \lambda_1 + (3 - w_{3,i}) \log \lambda_2 - 16\lambda_1 - 16\lambda_2
\]

For fixed \(\lambda_1, \lambda_2\)

\[
E_{\lambda_1, \lambda_2}(W_{3,i}) = \sum_{u<v} \left[ \log \lambda_1 \exp(-\lambda_1 u + \lambda_2 v) \right]
\]

so the expectation step is to replace \(w_{3,i}\) above with

\[
(2 + \frac{\lambda_{ij}}{\lambda_1 + \lambda_2}) \log \lambda_1 + (3 - \frac{\lambda_{ij}}{\lambda_1 + \lambda_2}) \log \lambda_2 - 16\lambda_1 - 16\lambda_2
\]

The maximization step can be accomplished by taking partials and setting equal to 0. This is

\[
\frac{1}{\lambda_1} \left(2 + \frac{\lambda_{ij}}{\lambda_1 + \lambda_2}\right) - 16 = 0 \text{ i.e. } \lambda_{ij}^{t+1} = \frac{(2 + \lambda_{ij}^{t} + \lambda_{2j}^{t})}{16}\lambda_{ij}^{t}\n\]

and

\[
\frac{1}{\lambda_2} \left(3 - \frac{\lambda_{ij}}{\lambda_1 + \lambda_2}\right) - 16 = 0 \text{ i.e. } \lambda_{2ij}^{t+1} = \frac{(3 - \lambda_{ij}^{t} + \lambda_{2j}^{t})}{16}\lambda_{ij}^{t}\n\]
3. Suppose that $X_1, X_2, \ldots, X_n$ are iid $\text{Ber}(p)$. Let $S_m = \sum_{i=1}^{m} X_i$. A statistician expecting to have only $n-1$ observations $X_i$ available for inference develops an estimator $\delta(S_{n-1})$ for $p$ under SEL. (This estimator may well be a biased estimator.) In fact, $n$ observations will be available. Find another estimator of $p$, say $\delta^*(S_n)$, that you are sure will have smaller $\text{MSE}_p$ than $\delta(S_{n-1})$ no matter what is the value of $p \in (0,1)$.

Rao-Blackwellize the estimator $\delta(S_{n-1})$. Conditional on $S_n = s$, the probability that $S_{n-1} = S_n$ is the (conditional) probability that $X_n = 0$, i.e., $1 - \frac{s}{n}$. (The conditional probability that $S_{n-1} = S_n - 1$ is then $\frac{s}{n}$.)

So the Rao-Blackwellized estimator is

$$\delta^*(S_n) = \frac{s_n}{n} \delta(S_{n-1}) + \left(1 - \frac{s_n}{n}\right) \delta(S_n)$$
4. Suppose that $X_1, X_2, \ldots, X_n$ are iid with marginal pdf $f(x | \alpha) = \alpha x^{\alpha-1}I[0 < x < 1]$.

a) Find a lower bound for the variance of any unbiased estimator $\delta(X)$ of $\gamma(\alpha) = \sin \alpha$ (based on the vector of $n$ observations).

For a single $X$, 
\[
\log f(x | \alpha) = \log \alpha + (\alpha-1) \log x
\]

\[
\frac{d}{d \alpha} \log f(x | \alpha) = \frac{1}{\alpha} + \log x
\]

\[
\frac{d^2}{d \alpha^2} \log f(x | \alpha) = -\frac{1}{\alpha^2}
\]

So the information in $X$ is $\frac{n}{\alpha^2}$. Then, since $\frac{d}{d \alpha} \sin \alpha = \cos \alpha$

The CR lower bound is
\[
\frac{(\cos \alpha)^2}{\frac{n}{\alpha^2}} = \frac{\alpha^2 \cos^2 \alpha}{n}
\]

b) Do you expect there to exist an unbiased estimator of $\gamma(\alpha) = \sin \alpha$ achieving your bound from a)? Explain!

No, I do not expect to achieve the CR lower bound. The only cases where it is achieved is that when

1) $f(x | \theta)$ is an exponential family and
2) what is being estimated is the mean of the natural sufficient statistic

Here $f(x | \alpha) = \exp(\log \alpha + \alpha \log x - \log x)$. 

\[
E_{\alpha} \log x = \int_0^1 (\log x) x^{\alpha-1} dx = x^\alpha \log x \bigg|_0^1 - \int_0^1 \frac{1}{\alpha} x^\alpha dx
\]

\[
= 0 - \frac{1}{\alpha} x^\alpha \bigg|_0^1
\]

\[
= -\frac{1}{\alpha}
\]

And clearly $\sin \alpha = -\frac{1}{\alpha}$
5. Suppose that $X_1, X_2, \ldots, X_n$ are iid $\text{Bi}(m, p)$ and (perhaps for "acceptance sampling" purposes)

$$
\gamma(p) = P_p[X > 0] = 1 - (1 - p)^m
$$

is of interest. **Find** an UMVUE for this quantity and **say** why you know your estimator is UMVU.

(Hint: You may find it useful to think of the $X_i$'s as $\sum_{j=1}^{m} Y_{i,j}$ for $m \cdot n$ independent variables $Y_{i,j}$ each $\text{Ber}(p)$.)

In this exponential family $\Sigma X_i$ is the natural sufficient statistic and the conditional expected value of any unbiased estimator given $\Sigma X_i$ is UMVU. Clearly

$$
1 - I[X_i = 0]
$$

is unbiased for $\gamma(p)$. The conditional probability that $X_1 = 0$ given $\Sigma X_i$ is the conditional probability that $Y_{11} = Y_{12} = \cdots = Y_{1m} = 0$ given $\sum_{i=1}^{n} \sum_{j=1}^{m} Y_{i,j} = \sum Y_i$.

This is a hypergeometric probability

$$
\binom{nm - \Sigma X_i}{m} \binom{\Sigma X_i}{0} = \frac{(nm - \Sigma X_i)!}{m!(n - 1)m!} = \frac{(nm - \Sigma X_i)!}{nm!} \frac{(n - 1)m!}{(n - 1)m!}
$$

and the desired UMVUE is $1 - \text{This}$. 
Another (probably better) solution to this provided by several students is that the desired conditional probability is

\[
1 - \frac{P\left[X_1 = 0 \text{ and } \sum_{i=1}^{n} X_i = t\right]}{P\left[\sum_{i=1}^{n} X_i = t\right]}
\]

\[
= 1 - \frac{P\left[X_1 = 0 \text{ and } \sum_{i=2}^{n} X_i = t\right]}{P\left[\sum_{i=1}^{n} X_i = t\right]}
\]

\[
= 1 - \frac{(1-p)^m \binom{mn-m}{t} p^t (1-p)^{mn-m-t}}{\binom{mn}{t} p^t (1-p)^{mn-t}}
\]

\[
= 1 - \frac{\binom{mn-m}{t}}{\binom{mn}{t}}
\]

(which works out to be exactly as before).
6. Argue carefully that you could use iid double exponential observations (i.e. ones with marginal pdf $\frac{1}{2}\exp(-|x|)$ on $\mathbb{R}$) to generate a standard normal random variable via the rejection algorithm, but that you could NOT use iid standard normal random variables to generate a double exponential random variable via the rejection algorithm.

In order to use observations from $h$ to generate an observation from a dsn with density proportional to $f(x)$ we have to have a bound $M$ so that

$$M h(x) > f(x)$$

i.e. $M > f(x)/h(x)$ anywhere $f(x) > 0$

The ratio of normal to exponential density is

$$\frac{2}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2} + |x| \right)$$

Since for $x > 0$, $-\frac{x^2}{2} + x$ has derivative $-x+1$, the ratio decreases in $|x| > 1$ and since it is constant is bounded on $[-1,1]$. Thus it is bounded on $\mathbb{R}$.

On the other hand, since $-\frac{x^2}{2} + |x| \to -\infty$ as $|x| \to \infty$ the reciprocal of the ratio goes to $\infty$ as $|x| \to \infty$ and is thus unbounded.

So there is an $M$ for generating a normal from exponentials but not the other way around.
7. (EXTRA CREDIT ONLY) Consider the two marginal pdfs

\[ f(x | 0) = I[0 < x < 1] \quad \text{and} \quad f(x | 1) = \frac{2^x}{\ln(2)} I[0 < x < 1] \]

and iid observations \( X_1, X_2, \ldots \) from one of these distributions (specified by \( f(x | \theta) \)). Argue carefully that there exists a non-randomized UMP test of \( H_0 : \theta = 0 \) vs \( H_1 : \theta = 1 \) for any size \( \alpha \in (0, 1) \) based on \( X_1, \ldots, X_n \). Give an explicit large \( n \) approximate form for such a test for \( \alpha = .05 \).

The likelihood ratio is

\[
\frac{\prod_{i=1}^{n} f(x_i | 1)}{\prod_{i=1}^{n} f(x_i | 0)} = \left( \frac{1}{1 \ln(2)} \right)^n 2 \Sigma x_i
\]

and a MP test rejects for large \( \Sigma x_i \). Since the \( \theta = 0 \) dsn for each \( X_i \) is continous, so is the dsn of \( \frac{1}{n} \sum_{i=1}^{n} X_i \) and by choosing

\[ \phi_n(\Delta) = I\left[ \frac{1}{n} \sum_{i=1}^{n} X_i > \text{upper } \alpha \text{ pt of the } \theta = 0 \text{ dsn of } 2X_i \right] \]

we can have a size \( \alpha \) non-randomized test.

The mean of \( X_1 \) under \( \theta = 0 \) is \( \frac{1}{2} \) and the variance is \( \frac{1}{12} \). The CLT promises that for large \( n \)

\[ \frac{\sqrt{n}(\bar{X}_n - \frac{1}{2})}{\frac{1}{\sqrt{12}}} \xrightarrow{d} N(0, 1) \]

so if we reject \( H_0 \) for \( \bar{X} > \frac{1}{2} + 1.645 \frac{\sqrt{\frac{1}{12}}}{\sqrt{n}} \)

i.e.

\[ \Sigma x_i > \frac{n}{2} + 1.645 \sqrt{\frac{n}{12}} \]

we have a MP test of size approximately \( .05 \).