Stat 513
Emerson, Win 16
Final Examination Key
March 14, 2016

Instructions:

- **This exam is closed book, closed notes. No use of calculators, computers, or cell phones is permitted.**
- **Write answers to the following questions on separate sheets of paper, starting each problem at the top of a new page. Use only the front side of each page. Be sure to write your name on the top of each page.**
- **In order to receive full credit, you must make clear how you derived the answers to the problems.**
- **You are allowed 110 minutes for this exam. When time is called, you must put down your pencils**
- **<u>Write out</u> and sign the following pledge for the in-class exam:**
  "On my honor, I have neither given nor received any unauthorized aid on this examination."
  If for any reason you can not honestly sign the pledge, please discuss this with me.

1. (25 points) Let $X_1, X_2, \ldots, X_n$ be a sequence of i.i.d. random variables having probability density function for $\theta > 2$

$$f_X(x) = \theta x^{-(\theta+1)} \mathbf{1}_{[x>1]}.$$

   (a) Find the maximum likelihood estimate $\hat{\theta}$ and show that it is biased. (Hint: Use Jensen's inequality)
   <u>Ans</u>: In this regular probability model we find

$$\begin{aligned}
L_i(\theta \mid X_i) &= \theta X_i^{-(\theta+1)} \mathbf{1}_{[X_i>1]} \\
\mathcal{L}_i(\theta \mid X_i) &= \log(\theta) - (\theta+1)\log(X_i) + \log(\mathbf{1}_{[X_i>1]}) \\
\mathcal{U}_i(\theta \mid X_i) &= \frac{\partial}{\partial\theta}\mathcal{L}_i(\theta \mid X_i) = \frac{1}{\theta} - \log(X_i) \\
J_i(\theta \mid X_i) &= -E\left[\frac{\partial}{\partial\theta}\mathcal{U}_i(\theta \mid X_i)\right] = \frac{1}{\theta^2}
\end{aligned}$$

We then find the MLE according to

$$\mathcal{U}(\hat{\theta}\,|\,\vec{X}) = \sum_{i=1}^{n}\left(\frac{1}{\hat{\theta}} - \log(X_i)\right) = \frac{n}{\hat{\theta}} - \sum_{i=1}^{n}\log(X_i) = 0$$
$$\Rightarrow \qquad \hat{\theta} = \frac{n}{\sum_{i=1}^{n}\log(X_i)}$$

And we find the distribution of $W = \log(X_i)$ as

$$F_X(x|\theta) = \int_1^x \theta u^{-(\theta+1)}du = \int_1^x \theta e^{-(\theta+1)\log(u)}du = \int_0^{\log(x)} \theta e^{-(\theta+1)v}e^v dv$$
$$= \int_0^{\log(x)} \theta e^{-\theta v}dv = \left(1 - e^{-\theta\log(x)}\right)\mathbf{1}_{[x>1]}$$
$$F_W(w|\theta) = Pr(W \le w|\theta) = Pr(e^W \le e^w|\theta) = Pr(X \le e^w) = \left(1 - e^{-\theta w}\right)\mathbf{1}_{[w>0]}$$

So $W = \log(X_i) \sim \mathcal{E}(\theta)$ with $W \sim (\frac{1}{\theta}, \frac{1}{\theta^2})$. By linearity of integration, we know

$$\frac{1}{\hat{\theta}} = \frac{1}{n}\sum_{i=1}^{n}\log(X_i) \;\sim \left(\frac{1}{\theta},\, \frac{1}{n\theta^2}\right),$$

and because the function $\frac{1}{x}$ is concave, by Jensen's inequality we have

$$E\left[\hat{\theta}\right] = E\left[\frac{1}{1/\hat{\theta}}\right] > \frac{1}{E\left[1/\hat{\theta}\right]} = \frac{1}{1/\theta} = \theta$$

and $\hat{\theta}$ is biased.

(b) What is the asymptotic distribution of $\hat{\theta}$?
Ans: By the theorem we have on asymptotic distributions of MLEs in regular problems,

$$\sqrt{n}\left(\hat{\theta} - \theta\right) \to_d \mathcal{N}(0, J_i^{-1}(\theta)) = \mathcal{N}\left(0, \theta^2\right).$$

(c) Find the uniform minimum variance unbiased estimator of $g(\theta) = 1/\theta$ and derive its distribution.
Ans: We note that the score function is

$$\mathcal{U}(\theta\,|\,\vec{X}) = \frac{n}{\theta} - \sum_{i=1}^{n}\log(X_i)$$

which is of the form $\mathcal{U}(\theta) = A(\theta)(T(x) - g(\theta))$ for linear functions of $g(\theta) = \frac{1}{\theta}$. So our MLE of $\frac{1}{\theta}$ is BRUE, and thus must also be UMVUE. As found in part a, the MLE is the sample mean of exponential $\mathcal{E}(\theta)$ random variables, so

$$n\frac{1}{\hat{\theta}} \;\sim\; \Gamma(n, \theta)$$

in the rate parameterization of the gamma distribution.

(d) Find the efficiency of the UMVUE you found in part c.
Ans: Because the UMVUE is BRUE, the efficiency is 1.

2. (20 points) Let $X_1, X_2, \ldots, X_n$ be a sequence of i.i.d. random variables having probability density function for $\theta > 2$

$$f_X(x) = \theta x^{-(\theta+1)} \mathbf{1}_{[x>1]}.$$

(a) Find the uniformly most powerful level α (UMP-α) test (including critical value) of

$$H_0 : \theta \leq \theta_0 \qquad \text{versus} \qquad H_1 : \theta > \theta_0.$$

Ans: For $\theta_1 > \theta_0$, we find the likelihood ratio

$$\begin{aligned}\Lambda &= \frac{\theta_1^n \left(\prod_{i=1}^n X_i\right)^{-(\theta_1+1)}}{\theta_0^n \left(\prod_{i=1}^n X_i\right)^{-(\theta_0+1)}} \\ &= \left(\frac{\theta_1}{\theta_0}\right)^n \left(\prod_{i=1}^n X_i\right)^{(\theta_0-\theta_1)}\end{aligned}$$

and we see that Λ is monotonically decreasing in $\prod_{i=1}^n X_i$, and is thus monotonically increasing in $\left(\prod_{i=1}^n X_i\right)^{-1}$. Hence, the Karlin-Rubin theorem says that the UMP-α test will be of the form

$$\textit{reject } H_0 \qquad \Leftrightarrow \qquad \frac{1}{\prod_{i=1}^n X_i} > c$$

which is equivalent to decisions of the form

$$\textit{reject } H_0 \qquad \Leftrightarrow \qquad \prod_{i=1}^n X_i < c$$

or

$$\textit{reject } H_0 \qquad \Leftrightarrow \qquad \sum_{i=1}^n \log(X_i) < c^*.$$

In problem 1a and 1c, we noted that

$$\sum_{i=1}^n \log(X_i) \;\sim\; \Gamma(n, \theta),$$

so our critical value should be the α quantile of the $\Gamma(n, \theta_0)$ distribution:

$$\textit{reject } H_0 \qquad \Leftrightarrow \qquad \sum_{i=1}^n \log(X_i) < \Gamma(\alpha; n, \theta_0).$$

Using the asymptotic distribution derived from the CLT (or likelihood theory as given in problem 1), we could also use the asymptotic test

$$\textit{reject } H_0 \qquad \Leftrightarrow \qquad Z = \sqrt{n}\frac{\left(\frac{1}{n}\sum_{i=1}^{n}\log(X_i) - \frac{1}{\theta_0}\right)}{\frac{1}{\theta_0}} < z_\alpha.$$

An alternative solution would use the NP lemma for testing simple hypotheses $H_0^*: \theta = \theta_0$ versus $H_1^*: \theta = \theta_1$ for an arbitrary $\theta_1 > \theta_0$, and find the MP-α test of the form

$$\begin{aligned}
\textit{reject } H_0 \qquad \Leftrightarrow \qquad \left(\frac{\theta_1}{\theta_0}\right)^n \left(\prod_{i=1}^{n} X_i\right)^{(\theta_0-\theta_1)} &> c \\
\left(\prod_{i=1}^{n} X_i\right)^{(\theta_0-\theta_1)} &> c_1 = c\left(\frac{\theta_0}{\theta_1}\right)^n \\
(\theta_0 - \theta_1)\sum_{i=1}^{n}\log(X_i) &> c_2 = \log(c_1) \\
\sum_{i=1}^{n}\log(X_i) &< c_3 = \frac{c_2}{(\theta_0-\theta_1)}
\end{aligned}$$

where we find $c_3 = \Gamma(\alpha; n, \theta_0)$. Then, noting that we achieved this same result for all $\theta_1 > \theta_0$, we have that this test is UMP-α for H_0^* versus $H_1: \theta > \theta_0$. Furthermore, owing to the fact that in this parameterization of the gamma distribution, θ is a scale parameter in the $\Gamma(n, \theta)$ family such that $\theta\Gamma(n, \theta) \sim \Gamma(n, 1)$, we know that for $\theta^* < \theta_0$,

$$\begin{aligned}
Pr(\Gamma(n, \theta^*) < c) &= Pr\left(\frac{\theta^*}{\theta_0}\Gamma(n, \theta^*) < \frac{\theta^*}{\theta_0}c\right) \\
&= Pr\left(\Gamma(n, \theta_0) < \frac{\theta^*}{\theta_0}c\right) \\
&< Pr(\Gamma(n, \theta_0) < c) \qquad \text{since } \tfrac{\theta^*}{\theta_0} < 1.
\end{aligned}$$

Hence, the probability of rejecting H_0 is lower for all $\theta^* < \theta_0$, and our test is UMP-α for the composite H_0 vs H_1.

(b) Find the uniformly most powerful level α (UMP-α) test (including critical value) of

$$H_0: \theta \geq \theta_0 \qquad \text{versus} \qquad H_1: \theta < \theta_0.$$

Ans: In part a, the only way we used information about the alternative was to determine the direction of the inequality in the critical function. Hence the UMP-α test in this setting will be

$$\textit{reject } H_0 \qquad \Leftrightarrow \sum_{i=1}^{n}\log(X_i) > \Gamma(1-\alpha; n, \theta_0).$$

(Explicitly (using cut and paste with minor modifications), I can write:
For $\theta_1 < \theta_0$, we find the likelihood ratio

$$\Lambda = \frac{\theta_1^n \left(\prod_{i=1}^n X_i\right)^{-(\theta_1+1)}}{\theta_0^n \left(\prod_{i=1}^n X_i\right)^{-(\theta_0+1)}} = \left(\frac{\theta_1}{\theta_0}\right)^n \left(\prod_{i=1}^n X_i\right)^{(\theta_0-\theta_1)}$$

and we see that Λ is monotonically increasing in $\prod_{i=1}^n X_i$, and is thus monotonically decreasing in $\left(\prod_{i=1}^n X_i\right)^{-1}$. Hence, the Karlin-Rubin theorem says that the UMP-α test will be of the form

$$reject\ H_0 \qquad \Leftrightarrow \qquad \frac{1}{\prod_{i=1}^n X_i} < c$$

which is equivalent to decisions of the form

$$reject\ H_0 \qquad \Leftrightarrow \qquad \prod_{i=1}^n X_i > c$$

or

$$reject\ H_0 \qquad \Leftrightarrow \qquad \sum_{i=1}^n \log(X_i) > c^*.$$

In problem 1a and 1c, we noted that

$$\sum_{i=1}^n \log(X_i) \ \sim\ \Gamma(n, \theta),$$

so our critical value should be the $1-\alpha$ quantile of the $\Gamma(n, \theta_0)$ distribution:

$$reject\ H_0 \qquad \Leftrightarrow \qquad \sum_{i=1}^n \log(X_i) > \Gamma(1-\alpha; n, \theta_0).$$

Using the asymptotic distribution derived from the CLT (or likelihood theory as given in problem 1), we could also use the asymptotic test

$$reject\ H_0 \qquad \Leftrightarrow \qquad Z = \sqrt{n}\frac{\left(\frac{1}{n}\sum_{i=1}^n \log(X_i) - \frac{1}{\theta_0}\right)}{\frac{1}{\theta_0}} > z_{1-\alpha}.$$

An alternative solution would use the NP lemma for testing simple hypotheses $H_0^* : \theta = \theta_0$ versus $H_1^* : \theta = \theta_1$ for an arbitrary $\theta_1 < \theta_0$, and find the MP-α test of the form

$$
\begin{aligned}
\textit{reject } H_0 \qquad \Leftrightarrow \qquad \left(\frac{\theta_1}{\theta_0}\right)^n \left(\prod_{i=1}^n X_i\right)^{(\theta_0-\theta_1)} &> c \\
\left(\prod_{i=1}^n X_i\right)^{(\theta_0-\theta_1)} &> c_1 = c\left(\frac{\theta_0}{\theta_1}\right)^n \\
(\theta_0-\theta_1)\sum_{i=1}^n \log(X_i) &> c_2 = \log(c_1) \\
\sum_{i=1}^n \log(X_i) &> c_3 = \frac{c_2}{(\theta_0-\theta_1)}
\end{aligned}
$$

where we find $c_3 = \Gamma(1-\alpha; n, \theta_0)$. Then, noting that we achieved this same result for all $\theta_1 < \theta_0$, we have that this test is UMP-α for H_0^* versus $H_1 : \theta < \theta_0$. Furthermore, owing to the fact that in this parameterization of the gamma distribution, θ is a scale parameter in the $\Gamma(n, \theta)$ family such that $\theta\Gamma(n, \theta) \sim \Gamma(n, 1)$, we know that for $\theta^* > \theta_0$,

$$
\begin{aligned}
Pr(\Gamma(n,\theta^*) > c) &= Pr\left(\frac{\theta^*}{\theta_0}\Gamma(n,\theta^*) > \frac{\theta^*}{\theta_0}c\right) \\
&= Pr\left(\Gamma(n,\theta_0) > \frac{\theta^*}{\theta_0}c\right) \\
&< Pr(\Gamma(n,\theta_0) > c) \qquad\qquad \text{since } \tfrac{\theta^*}{\theta_0} > 1.
\end{aligned}
$$

Hence, the probability of rejecting H_0 is lower for all $\theta^* > \theta_0$, and our test is UMP-α for the composite H_0 vs H_1.

(c) Show that there is no uniformly most powerful level α (UMP-α) test of

$$H_0 : \theta = \theta_0 \qquad \text{versus} \qquad H_1 : \theta \neq \theta_0.$$

Ans: : Note that the MP-α test for $\theta_1 > \theta_0$ (as derived in part a) is different from the MP-α test for $\theta_1 < \theta_0$ (as derived in part b), so there is no UMP-α test for the two-sided hypothesis.

3. (25 points) Let $X_1, X_2, \ldots, X_n$ be a sequence of i.i.d. random variables having probability density function

$$f_X(x) = \frac{\theta}{x^2}\mathbf{1}_{[x>\theta]}.$$

(a) Show that $W_i = X_i/\theta$ has a distribution independent of θ.
Ans: We first note that this is only a probability distribution for $\theta > 0$. We then

find the cdf for X_i as

$$F_X(x\,|\,\theta) = \int_{-\infty}^{x} \frac{\theta}{u^2}\mathbf{1}_{[u>\theta]}\,du = \left[\int_{\theta}^{x} \frac{\theta}{u^2}\,du\right]\mathbf{1}_{[x>\theta]}$$
$$= \left[-\frac{\theta}{u}\right]_{u=\theta}^{u=x}\mathbf{1}_{[x>\theta]} = \left[1-\frac{\theta}{x}\right]\mathbf{1}_{[x>\theta]}$$

So the cdf for W_i can be found by

$$F_W(w) = Pr(W_i < w) = Pr\left(\frac{X_i}{\theta} < w\right) = Pr\left(X_i < w\theta\right) = F_X(w\theta\,|\,\theta)$$
$$= \left[1-\frac{\theta}{w\theta}\right]\mathbf{1}_{[w\theta>\theta]} = \left[1-\frac{1}{w}\right]\mathbf{1}_{[w>1]}$$

which cdf does not involve θ.

(b) Find the maximum likelihood estimator $\hat{\theta}$ and derive its distribution.
Ans: This distribution does not have common support, so the likelihood will not be differentiable everywhere. The likelihood is given by

$$L(\theta\,|\,\vec{X}) = \prod_{i=1}^{n} \frac{\theta}{X_i^2}\mathbf{1}_{[X_i>\theta]} = \frac{\theta^n}{\prod_{i=1}^{n} X_i^2}\mathbf{1}_{[X_{(1)}>\theta]}.$$

By inspection, $L(\theta)$ is increasing in θ for $\theta < X_{(1)}$, and then $L(\theta) = 0$ for $\theta > X_{(1)}$. So the likelihood is maximized at $\hat{\theta} = X_{(1)}$, the minimum of $(X_1, X_2, \ldots, X_n)$. We thus derive its distribution (pdf and cdf) using the survivor function as

$$S_{\hat{\theta}}(z\,\theta) = Pr(X_{(1)} > z\,|\,\theta) = \prod_{i=1}^{n} Pr(X_i > z\,|\,\theta) = \left[Pr(X_i > z\,|\,\theta)\right]^n$$
$$= \left[1 - F_X(z\,|\,\theta)\right]^n = \frac{\theta^n}{z^n}\mathbf{1}_{[z>\theta]} + \mathbf{1}_{[z\le\theta]}$$
$$f_{\hat{\theta}}(z\,\theta) = -\frac{\partial}{\partial z}S_{\hat{\theta}}(z\,\theta) = n\frac{\theta^n}{z^{n+1}}\mathbf{1}_{[z>\theta]}$$
$$F_{\hat{\theta}}(z\,\theta) = 1 - S_{\hat{\theta}}(z\,\theta) = \left[1-\frac{\theta^n}{z^n}\right]\mathbf{1}_{[z>\theta]}$$

(c) Find the uniform minimum variance unbiased estimator of θ, if possible.
Ans: By the factorization theorem, we can inspect the joint density for $\vec{X}$ as given in the likelihood function in part a and see that it factors into

$$f_{\vec{X}}(\vec{x}\,|\,\theta) = \prod_{i=1}^{n} \frac{\theta}{X_i^2}\mathbf{1}_{[X_i>\theta]}$$
$$= \frac{\theta^n}{\prod_{i=1}^{n} X_i^2}\mathbf{1}_{[X_{(1)}>\theta]}$$
$$= g(X_{(1)}\,|\,\theta)\,h(\prod_{i=1}^{n} X_i^2).$$

so $T = X_{(1)}$ is minimal sufficient for θ. We now consider the completeness of its probability distribution. Suppose that for some function g, $E[g(T)\,|\theta] \equiv 0$ for all $\theta > 0$. Then using the density derived in part b we find

$$E[g(T)\,|\theta] = \int_\theta^\infty g(t) n \frac{\theta^n}{t^{n+1}}\, dt \equiv 0.$$

Taking the derivative with respect to θ we find

$$\begin{aligned} 0 = \frac{d}{d\theta} E[g(T)\,|\theta] &= \frac{d}{d\theta} \int_\theta^\infty g(t) \frac{n\theta^n}{t^{n+1}}\, dt \\ &= g(t)\frac{n\theta^n}{\theta^{n+1}} - \left(\frac{d}{d\theta}\theta^n\right) \int_\theta^\infty g(t)\frac{n}{t^{n+1}}\, dt \\ &= g(t)\frac{n}{\theta} - n\theta^{n-1} \int_\theta^\infty g(t)\frac{n}{t^{n+1}}\, dt \\ &= \frac{n}{\theta}\left[g(t) - E[g(T)\,|\theta]\right] = \frac{n}{\theta} g(t) \end{aligned}$$

and we have $g(t) \equiv 0\ \forall t$, so T is a complete sufficient statistic.
Now we find $E[T\,|\,\theta]$ as

$$\begin{aligned} E[T\,|\,\theta] &= \int_\theta^\infty f_{\hat\theta}(z\,\theta) = \int_\theta^\infty zn\frac{\theta^n}{z^{n+1}}\, dz = \int_\theta^\infty n\theta^n z^{-n}\, dz = \left[\frac{n}{(-n+1)}\theta^n z^{-n+1}\right]_\theta^\infty \\ &= \frac{n}{n-1}\theta. \end{aligned}$$

So estimator $\tilde\theta = \frac{(n-1)}{n} X_{(1)}$ is an unbiased function of the complete sufficient statistic, and, by Lehmann-Scheffé, the UMVUE.

(d) Derive a formula for a $100(1-\alpha)\%$ confidence interval for θ.
Ans: We base the CI on the MLE $T = X_{(1)}$. For observation $T = t$, we use the survivor function for T (because it is an easier form) to find (θ_L, θ_U) such that

$$\left.\begin{aligned} Pr(T > t\,|\theta_L) = \tfrac{\theta_L^n}{t^n} = \tfrac{\alpha}{2} \\ Pr(T > t\,|\theta_U) = \tfrac{\theta_U^n}{t^n} = 1 - \tfrac{\alpha}{2} \end{aligned}\right\} \quad \Rightarrow \quad \left\{\begin{aligned} \theta_L = t\left(\tfrac{\alpha}{2}\right)^{\frac{1}{n}} \\ \theta_U = t\left(1 - \tfrac{\alpha}{2}\right)^{\frac{1}{n}} \end{aligned}\right.$$

4. (20 points) Suppose $X_i, i = 1, \ldots, n$ are independent and identically distributed random variables which, conditional upon a parameter $\theta > 0$, have the Poisson distribution $X_i \sim \mathcal{P}(\theta)$. Consider a prior distribution for θ according to the gamma distribution $\theta \sim \Gamma(\alpha, \beta)$, with density $\lambda(\theta) = \beta^\alpha\theta^{\alpha-1}e^{-\beta\theta}/\Gamma(\alpha)$ and mean α/β and variance α/β^2.

(a) Show that the above prior distribution is the conjugate prior for this problem.
Ans: We consider the kernel (i.e., the multiplicative terms that involve θ) of $p_{X|\theta}(x\,|\theta)\,\lambda_\theta(\theta)$:

$$\begin{aligned} \lambda(\theta\,|\,X) \propto p_{X|\theta}(x\,|\theta)\lambda_\theta(\theta) &= \frac{e^{-\theta}\theta^x}{x!}\beta^\alpha\theta^{\alpha-1}e^{-\beta\theta}/\Gamma(\alpha) \\ &\propto \theta^{\alpha-1+x}e^{-(\beta+1)\theta} \end{aligned}$$

which we see is the kernel of a $\Gamma(\alpha + x,\ \beta + 1)$ distribution. By induction, we then know

$$\theta \,|\, \vec{X} \ \sim \Gamma(\alpha + \sum_{i=1}^{n} X_i,\ \beta + n)$$

having posterior density

$$\lambda_{\theta|\vec{X}}(\theta \,|\, \vec{X}) = \frac{(\beta+n)^{\alpha+\sum X_i}\, \theta^{\alpha-1+\sum X_i}\, e^{-(\beta+n)\theta}}{\Gamma(\alpha + \sum X_i)}$$

(b) Find the Bayes estimator for squared error loss, i.e., $L(\theta, \hat{\theta}) = (\theta - \hat{\theta})^2$.
Ans: With squared error loss, the Bayes estimator is the mean of the posterior distribution. Hence,

$$\hat{\theta}_{Bayes} = \frac{\alpha + \sum_{i=1}^{n} X_i}{\beta + n}.$$

(c) Derive a $100(1-\alpha)\%$ credible interval for θ.
Ans: For notational convenience (and to help differentiate the unfortunate duplicate use of α as a parameter of the prior and α as the level of significance), define $\alpha^* = \alpha + \sum_{i=1}^{n} X_i$ and $\beta^* = \beta + n$. Then a credible interval based on equal tail probabilities would be the $\frac{\alpha}{2}$ and $1 - \frac{\alpha}{2}$ quantiles of the $\Gamma(\alpha^*, \beta^*)$ distribution:

$$\theta_L : \int_0^{\theta_L} \frac{(\beta^*+n)^{\alpha^*+\sum X_i}\, \theta^{\alpha^*-1+\sum X_i}\, e^{-(\beta^*+n)\theta}}{\Gamma(\alpha^* + \sum X_i)}\, d\theta = \frac{\alpha}{2}$$

$$\theta_U : \int_0^{\theta_U} \frac{(\beta^*+n)^{\alpha^*+\sum X_i}\, \theta^{\alpha^*-1+\sum X_i}\, e^{-(\beta^*+n)\theta}}{\Gamma(\alpha^* + \sum X_i)}\, d\theta = 1 - \frac{\alpha}{2}$$

(Note that if random variable $Z \sim \Gamma(\alpha^*, \beta^*)$ in this parameterization, then $2Z\beta^* \sim \chi^2_k$, a chi squared distribution with $k = \alpha^*/2$ degrees of freedom. Because of this, you sometimes see people convert the above critical values to the critical values of a chi squared distribution, however I see little advantage because in either case I use a computer to find the critical values.)

5. (30 points) Suppose $Y_1, Y_2, \ldots, Y_n$ are i.i.d. exponential random variables with $Y_i \sim \mathcal{E}(\lambda)$, and $X_1, X_2, \ldots, X_m$ are i.i.d. exponential random variables with $X_i \sim \mathcal{E}(\mu)$, where hazards $\lambda > 0$ and $\mu > 0$. We are interested in testing $H_0 : \lambda \leq \mu$ versus $H_1 : \lambda > \mu$.

   (a) Derive expressions for the log likelihood, efficient score vector, and information matrix under the unrestricted (full) model.
   Ans: Defining sample means $\overline{Y} = \frac{1}{n} \sum_{i=1}^{n} Y_i$ and $\overline{X} = \frac{1}{m} \sum_{i=1}^{m} X_i$ for notational

convenience, we find in this regular probability model

$$
\begin{aligned}
L(\lambda, \mu \,|\, \vec{Y}, \vec{X}) &= \lambda^n \, e^{-n\lambda\overline{Y}} \mu^m \, e^{-m\mu\overline{X}} \\
\mathcal{L}(\lambda, \mu \,|\, \vec{Y}, \vec{X}) &= n\log(\lambda) - n\lambda\overline{Y} + m\log(\mu) - m\mu\overline{X} \\
\mathcal{U}_1(\lambda, \mu \,|\, \vec{Y}, \vec{X}) &= \frac{\partial}{\partial\lambda}\mathcal{L}(\lambda, \mu \,|\, \vec{Y}, \vec{X}) = \frac{n}{\lambda} - n\overline{Y} \\
\mathcal{U}_2(\lambda, \mu \,|\, \vec{Y}, \vec{X}) &= \frac{\partial}{\partial\mu}\mathcal{L}(\lambda, \mu \,|\, \vec{Y}, \vec{X}) = \frac{m}{\mu} - m\overline{X} \\
J_{11}(\lambda, \mu \,|\, \vec{Y}, \vec{X}) &= -E\left[\frac{\partial}{\partial\lambda}\mathcal{U}_1(\lambda, \mu \,|\, \vec{Y}, \vec{X})\right] = \frac{n}{\lambda^2} \\
J_{12}(\lambda, \mu \,|\, \vec{Y}, \vec{X}) &= -E\left[\frac{\partial}{\partial\mu}\mathcal{U}_1(\lambda, \mu \,|\, \vec{Y}, \vec{X})\right] = 0 = J_{21}(\lambda, \mu \,|\, \vec{Y}, \vec{X}) \\
J_{22}(\lambda, \mu \,|\, \vec{Y}, \vec{X}) &= -E\left[\frac{\partial}{\partial\mu}\mathcal{U}_2(\lambda, \mu \,|\, \vec{Y}, \vec{X})\right] = \frac{m}{\mu^2}
\end{aligned}
$$

yielding

$$
\vec{\mathcal{U}}(\lambda, \mu \,|\, \vec{Y}, \vec{X}) = \begin{pmatrix} \frac{n}{\lambda} - n\overline{Y} \\ \\ \frac{m}{\mu} - m\overline{X} \end{pmatrix} \qquad\qquad \mathbf{J} = \begin{pmatrix} \frac{n}{\lambda^2} & 0 \\ \\ 0 & \frac{m}{\mu^2} \end{pmatrix}
$$

(b) Find the MLEs $\hat{\lambda}$ and $\hat{\mu}$ in the unrestricted (full) model, along with their joint asymptotic distribution.
<u>Ans</u>: Setting $\vec{\mathcal{U}}(\hat{\lambda}, \hat{\mu} \,|\, \vec{Y}, \vec{X}) = \vec{0}$ yields

$$
\hat{\lambda} = \frac{1}{\overline{Y}} \qquad\qquad \hat{\mu} = \frac{1}{\overline{X}}.
$$

In this regular probability model, we know that based on asymptotic results, the MLEs have approximate distribution

$$
\begin{pmatrix} \hat{\lambda} \\ \\ \hat{\mu} \end{pmatrix} = \begin{pmatrix} \frac{1}{\overline{Y}} \\ \\ \frac{1}{\overline{X}} \end{pmatrix} \quad \dot{\sim} \quad \mathcal{N}_2\left(\begin{pmatrix} \lambda \\ \\ \mu \end{pmatrix}, \; \mathbf{J}^{-1}(\lambda, \mu \,|\, \vec{Y}, \vec{X}) = \begin{pmatrix} \frac{\lambda^2}{n} & 0 \\ \\ 0 & \frac{\mu^2}{m} \end{pmatrix} \right).
$$

(c) Find the MLEs $\hat{\lambda}_0$ and $\hat{\mu}_0$ in the restricted (null) model having $\lambda = \mu$.
<u>Ans</u>: Arbitrarily using λ for the common hazard, we rewrite the log likelihood and score for the restricted model as

$$
\begin{aligned}
\mathcal{L}^{(0)}(\lambda, \mu \,|\, \vec{Y}, \vec{X}) &= n\log(\lambda) - n\lambda\overline{Y} + m\log(\lambda) - m\lambda\overline{X} \\
\mathcal{U}_1^{(0)}(\lambda, \mu \,|\, \vec{Y}, \vec{X}) &= \frac{\partial}{\partial\lambda}\mathcal{L}^{(0)}(\lambda, \mu \,|\, \vec{Y}, \vec{X}) = \frac{n+m}{\lambda} - n\overline{Y} - m\overline{X}
\end{aligned}
$$

which, when setting the score to 0, yields restricted MLEs

$$
\hat{\lambda}_0 = \hat{\mu}_0 = \frac{m+n}{n\overline{Y} + m\overline{X}}.
$$

(d) Provide a formula for the Wald statistic testing H_0, along with its asymptotic distribution. (More points are given for simplified formulas.)
<u>Ans</u>: We take the linear contrast of the full model MLEs and use properties of the normal distribution to find

$$\begin{pmatrix}1 & -1\end{pmatrix}\begin{pmatrix}\hat{\lambda}\\ \hat{\mu}\end{pmatrix} = \frac{1}{\overline{\overline{Y}}} - \frac{1}{\overline{\overline{X}}} \;\dot{\sim}\; \mathcal{N}\left(\lambda - \mu,\; \frac{\lambda^2}{n} + \frac{\mu^2}{m}\right).$$

We then form a Z statistic by using the full model MLEs in the inverse information matrix

$$Z = \frac{\frac{1}{\overline{Y}} - \frac{1}{\overline{X}}}{\sqrt{\frac{\hat{\lambda}^2}{n} + \frac{\hat{\mu}^2}{m}}} = \sqrt{mn}\frac{\overline{X} - \overline{Y}}{\sqrt{m\overline{X}^2 + n\overline{Y}^2}} \;\dot{\sim}_{H_0}\; \mathcal{N}(0,1)$$

(Note that I could also have expressed this as a chi squared statistic by creating the quadratic form.)

(e) Provide a formula for the score statistic testing H_0, along with its asymptotic distribution. (More points are given for simplified formulas.)
<u>Ans</u>: We create the quadratic form using the MLEs from the restricted (null) model

$$Q = \vec{\mathcal{U}}^T(\hat{\lambda}_0, \hat{\mu}_0 \,|\, \vec{Y}, \vec{X})\mathbf{J}^{-1}(\hat{\lambda}_0, \mu_0 \,|\, \vec{Y}, \vec{X})\vec{\mathcal{U}}(\hat{\lambda}_0, \hat{\mu}_0 \,|\, \vec{Y}, \vec{X})$$

which yields

$$\begin{aligned}
Q &= \begin{pmatrix}\left(\frac{n}{\hat{\lambda}_0} - n\overline{Y}\right) & \left(\frac{m}{\hat{\mu}_0} - m\overline{X}\right)\end{pmatrix}\begin{pmatrix}\frac{\hat{\lambda}_0^2}{n} & 0\\ 0 & \frac{\hat{\mu}_0^2}{m}\end{pmatrix}\begin{pmatrix}\left(\frac{n}{\hat{\lambda}_0} - n\overline{Y}\right)\\ \left(\frac{m}{\hat{\mu}_0} - m\overline{X}\right)\end{pmatrix}\\
&= \left(\frac{n}{\hat{\lambda}_0} - n\overline{Y}\right)^2\frac{\hat{\lambda}_0^2}{n} + \left(\frac{m}{\hat{\mu}_0} - m\overline{X}\right)^2\frac{\hat{\mu}_0^2}{m} = n\left(1 - \overline{Y}\hat{\lambda}_0\right)^2 + m\left(1 - \overline{X}\hat{\lambda}_0\right)^2\\
&= n\left(1 - \frac{(m+n)\overline{Y}}{n\overline{Y} + m\overline{X}}\right)^2 + m\left(1 - \frac{(m+n)\overline{X}}{n\overline{Y} + m\overline{X}}\right)^2\\
&= mn(m+n)\frac{(\overline{X} - \overline{Y})^2}{(n\overline{Y} + m\overline{X})^2} \;\dot{\sim}_{H_0}\; \chi_1^2
\end{aligned}$$

(Note, we can rearrange this last statistic to look like

$$Q = \left[\frac{(\overline{X} - \overline{Y})}{\sqrt{\left(\frac{n\overline{Y} + m\overline{X}}{m+n}\right)^2\left(\frac{1}{n} + \frac{1}{m}\right)}}\right]^2 = \left[\frac{(\overline{X} - \overline{Y})}{\sqrt{\left(\frac{1}{\hat{\lambda}_0}\right)^2\left(\frac{1}{n} + \frac{1}{m}\right)}}\right]^2,$$

which looks like the difference in sample means normalized by the standard error estimated under the restricted (null) hypothesis.)

(f) Provide a formula for the likelihood ratio statistic testing H_0, along with its asymptotic distribution. (More points are given for simplified formulas.)
<u>Ans</u>: We use both the full model and restricted model MLEs:

$$\begin{aligned}
2\log(\Lambda) &= 2\left(\mathcal{L}(\hat{\lambda}, \hat{\mu} \,|\, \vec{Y}, \vec{X}) - \mathcal{L}(\hat{\lambda}_0, \hat{\mu}_0 \,|\, \vec{Y}, \vec{X})\right) \\
&= 2\left(n\log(\hat{\lambda}) - n\hat{\lambda}\overline{Y} + m\log(\hat{\mu}) - m\hat{\mu}\overline{X} - (n+m)\log(\hat{\lambda}_0) + (n\overline{Y} + m\overline{X})\hat{\lambda}_0\right) \\
&= -2\Big(n\log(\overline{Y}) + n + m\log(\overline{X}) + m - (n+m)\log(n\overline{Y} + m\overline{X}) + \\
&\qquad (n+m)\log(n+m) - (m+n)\Big) \\
&= -2\Big(n\log(\overline{Y}) + m\log(\overline{X}) - (n+m)\log(n\overline{Y} + m\overline{X}) + (n+m)\log(n+m)\Big) \\
&\dot{\sim}_{H_0}\ \chi^2_1
\end{aligned}$$