APPM 5720

Solutions to Review Problems for Final Exam

1. The pdf is

$$f(x;\alpha,\beta) = \frac{1}{\Gamma(\alpha)}\beta^\alpha x^{\alpha-1}e^{-\beta x}\,I_{(0,\infty)}(x).$$

The joint pdf is

$$f(\vec{x};\alpha,\beta) = \frac{1}{[\Gamma(\alpha)]^n}\beta^{n\alpha}[\prod x_i]^{\alpha-1}e^{-\beta\sum x_i}\,\prod I_{(0,\infty)}(x_i)$$

$$= \underbrace{\frac{1}{[\Gamma(\alpha)]^n}\beta^{n\alpha}}_{a(\theta)}\underbrace{\prod I_{(0,\infty)}(x_i)}_{b(\vec{x})}\exp\left[\underbrace{(\alpha-1)}_{c_1(\theta)}\underbrace{\sum \ln x_i}_{d_1(\vec{x})} - \underbrace{\beta}_{c_2(\theta)}\underbrace{\sum x_i}_{d_2(\vec{x})}\right].$$

So, by "two-parameter exponential family"

$$S = (d_1(\vec{X}), d_2(\vec{X})) = (\sum \ln X_i, \sum X_i)$$

is complete and sufficient for this model.

---

2. Since σ^2 is fixed and known, we can use the one-parameter exponential family factorization to show that $S = \sum X_i$ is complete and sufficient for μ.

Since X_1 is an unbiased estimator of μ, the Rao-Blackwell Theorem gives us that $\mathsf{E}[X_1|\sum X_i]$ is still unbiased for μ. Also, because it is a function of the complete and sufficient statistic $S = \sum X_i$, it is the UMVUE.

We know that the UMVUE here is $\overline{X}$. Since UMVUEs are unique, we have that

$$\mathsf{E}[X_1|\sum X_i] = \overline{X}.$$

---

3. **Informally**, a minimal sufficient statistics is a sufficient statistic of lowest dimension. So, we will try to use the Factorization Criterion to find a low dimensional sufficient statistic. However, we will then have to prove that this really is minimal sufficient.

The joint pdf is

$$\begin{aligned} f(\vec{x};\theta) &= e^{-\sum_{i=1}^n (x_i-\theta)}\prod_{i=1}^n I_{(\theta,\infty)}(x_i) \\ &= \underbrace{e^{-\sum x_i}}_{h(\vec{x})}\underbrace{e^{n\theta}I_{(\theta,\infty)}(x_{(1)})}_{g(s(\vec{x};\theta))} \end{aligned}$$

By the Factorization Criterion for sufficiency, we see that $S = X_{(1)}$ is a sufficient statistic for this model.

To show that $X_{(1)}$ is minimal sufficient, you might try using our result that says, if we have

$$\frac{f(\vec{x};\theta)}{f(\vec{y};\theta)} \text{ is } \theta\text{-free} \quad \Leftrightarrow \quad s(\vec{x}) = s(\vec{y}) \tag{1}$$

then $S = s(\vec{X})$ is minimal sufficient.

However, if you try this you will see that we do not have (1) holding for this model!

Instead, we will use the fact that a complete and sufficient statistic is minimal sufficient.

That is, we will show completeness of $S = X_{(1)}$.

We will need the pdf of $S = X_{(1)}$. The cdf is

$$\begin{aligned} F_S(s) &= P(S \leq s) = P(X_{(1)} \leq s) \\ &= 1 - P(X_{(1)} > s) \\ &\stackrel{iid}{=} 1 - [P(X_1 > s)]^n \\ &= 1 - [e^{-(s-\theta)}]^n \\ &= 1 - e^{-n(s-\theta)} \end{aligned}$$

So, the pdf for $S = X_{(1)}$ is

$$f_S(s) = \frac{d}{ds} F_S(s) = \frac{d}{ds}[1 - e^{-n(s-\theta)}] = ne^{-n(s-\theta)}.$$

Since the minimum lives on (θ, ∞), we can complete this with an indicator:

$$f_S(s) = ne^{-n(s-\theta)}\, I_{(\theta,\infty)}(s).$$

So, the minimum of n shifted rate 1 exponentials is a shifted exponential with rate n. (This is not surprising and you could have just said that from the begining without showing it.)

We are ready to show completeness. Take any function g such that

$$\mathsf{E}[g(S)] = 0 \;\; \forall\theta.$$

Then

$$\begin{aligned} 0 &= \mathsf{E}[g(S)] = \int_{-\infty}^{\infty} g(s) f_S(s)\, ds \\ &= \int_{\theta}^{\infty} g(s) ne^{-n(s-\theta)}\, ds \\ &= ne^{n\theta} \int_{\theta}^{\infty} g(s) e^{-ns}\, ds \end{aligned}$$

This implies that

$$\int_{\theta}^{\infty} g(s) e^{-ns}\, ds = 0 \;\; \forall\theta$$

and therefore that

$$\int_{\infty}^{\theta} g(s) e^{-ns}\, ds = -\int_{\theta}^{\infty} g(s) e^{-ns}\, ds = 0 \;\; \forall\theta$$

Taking the derivative of both sides with respect to θ gives us

$$g(\theta)e^{-n\theta} = 0 \ \ \forall\theta$$

Since e to a power is never 0, this implies that

$$g(\theta) = 0 \ \ \forall\theta.$$

θ is just acting as a variable, to be clear, we have that

$$g(x) = 0 \ \ \forall x.$$

Thus, $g(S) = g(X_{(1)}) = 0$ with probability 1 and we have shown that $S = X_{(1)}$ is complete for this model.

---

4. Since the indicator is 1 everywhere, we don't need it. I just wanted to be clear about the domain of the problem.

   The joint pdf is

   $$f(\vec{x};\theta) = \frac{1}{2^n} e^{-\sum |x_i - \theta|}.$$

   There are many ways to rewrite this sum, for example, putting in indicators indicating whether each x_i is below of above θ. However, to show sufficiency of the order statistics $S = (X_{(1)}, X_{(2)}, \ldots, X_{(n)})$, it is enough to note that

   $$\sum |x_i - \theta| = \sum |X_{(i)} - \theta|.$$

   Thus, we may write

   $$f(\vec{x};\theta) = \frac{1}{2^n} e^{-\sum |x_{(i)} - \theta|}$$

   and use the Factorization Criterion for sufficiency. You can take $h(\vec{x})$ to be $1/2^n$ or even identically 1. Either way, the data does not appear in "a single clump" (or 2 or 3). In order to evaluate $e^{-\sum |x_{(i)} - \theta|}$, we need all of $X_{(1)}, X_{(2)}, \ldots, X_{(n)}$. Thus, by the Factorization Criterion, $S = (X_{(1)}, X_{(2)}, \ldots, X_{(n)})$ is sufficient for this model.

   Because of the high-dimensional nature of S, showing minimal sufficiency by showing completeness is hard. Instead, we will appeal to our result that if

   $$\frac{f(\vec{x};\theta)}{f(\vec{y};\theta)} \text{ is } \theta\text{-free} \ \ \Leftrightarrow \ \ s(\vec{x}) = s(\vec{y}) \tag{2}$$

   then $S = s(\vec{X})$ is minimal sufficient.

   For this part, it will help to rewrite the exponent as

   $$\begin{aligned} -\sum |x_{(i)} - \theta| &= -\sum_{\{i:x_{(i)}<\theta\}} |x_{(i)} - \theta| - \sum_{\{i:x_{(i)}\geq\theta\}} |x_{(i)} - \theta| \\ &= -\sum_{\{i:x_{(i)}<\theta\}} [-(x_{(i)} - \theta)] - \sum_{\{i:x_{(i)}\geq\theta\}} (x_{(i)} - \theta) \\ &= \sum_{\{i:x_{(i)}<\theta\}} x_{(i)} - \sum_{\{i:x_{(i)}\geq\theta\}} x_{(i)} + \theta[(\#x_i \geq \theta) - (\#x_i < \theta)] \end{aligned}$$

Note that

$$\frac{f(\vec{x};\theta)}{f(\vec{y};\theta)} = \frac{e^{-\sum |x_{(i)}-\theta|}}{e^{-\sum |y_{(i)}-\theta|}} = \exp[-\sum |x_{(i)} - \theta| + \sum |y_{(i)} - \theta|]$$

and that the exponent is

$$\sum_{\{i:x_{(i)}<\theta\}} x_{(i)} - \sum_{\{i:x_{(i)}\geq\theta\}} x_{(i)} - \sum_{\{i:y_{(i)}<\theta\}} y_{(i)} + \sum_{\{i:y_{(i)}\geq\theta\}} y_{(i)} + \theta[(\#x_i \geq \theta) - (\#x_i < \theta) - (\#y_i \geq \theta) + (\#y_i < \theta)]$$

For notational simplicity, let $\alpha(\theta) := (\#x_i \geq \theta)$ and $\beta(\theta) := (\#y_i \geq \theta)$. Note that $(\#x_i < \theta) = n - \alpha(\theta)$ and $(\#y_i < \theta) = n - \beta(\theta)$.

The exponent is now

$$\left[\sum_{\{i:x_{(i)}<\theta\}} x_{(i)} - \sum_{\{i:x_{(i)}\geq\theta\}} x_{(i)} - \sum_{\{i:y_{(i)}<\theta\}} y_{(i)} + \sum_{\{i:y_{(i)}\geq\theta\}} y_{(i)} \right] + 2\theta(\alpha(\theta) - \beta(\theta)) \qquad (3)$$

We want to show that this is "θ-free", or constant in θ, if and only if $(x_{(1)}, x_2, \ldots, x_{(n)}) = (y_{(1)}, y_{(2)}, \ldots, y_{(n)})$. This is clearly true if $(x_{(1)}, x_2, \ldots, x_{(n)}) = (y_{(1)}, y_{(2)}, \ldots, y_{(n)})$. So, let's assume that (3) is constant in θ and try to show that this forces equality of the order statistics.

Consider evaluating (3) over an interval of θ's that does not contain any of the x_i or y_i. The first term (in the square brackets) will remain constant over this interval. The second term will be constant on this interval if and only if $\alpha(\theta) = \beta(\theta)$ for all θ in the interval.

Now, this will be true <u>for all such intervals</u> if and only if the order statistics for the x's are the same as the order statistics for the y's.

Thus, we have (2) holding when $s(\vec{x}) = (x_{(1)}, x_{(2)}, \ldots, x_{(n)})$. By our result, this gives us that

$$S = s(\vec{X}) = (X_{(1)}, X_{(2)}, \ldots, X_{(n)})$$

is minimal sufficient.

---

5. Note that, for the $N(\theta, 1)$ distribution, θ is a location parameter. Also note that $Y = X_{(2)} - X_{(1)}$ is a location invariant statistic since adding a constant c to all data points would produce order statistics $X_{(1)} + c$, $X_{(2)} + c$, ..., $X_{(n)} + c$ and $(X_{(2)} + c) - (X_{(1)} - c) = X_{(2)} - X_{(1)}$. Thus, Y is an ancillary statistic.

   On the other hand, it is easy to show, by one-parameter exponential family, that $\sum X_i$ is complete and sufficient for this model which implies that the one-to-one transformation to $\overline{X}$ is complete and sufficient.

   Thus, by Basu's Theorem, we have that $\overline{X}$ is independent of $X_{(2)} - X_{(1)}$.

---

6. (a) For the exponential distribution, λ is a scale parameter. The statistic S is scale-invariant since

   $$\frac{cX_n}{\sum_{i=1}^n cX_i} = \frac{cX_n}{c\sum_{i=1}^n X_i} = \frac{X_n}{\sum_{i=1}^n X_i} = S.$$

   So, S in ancillary for this model.

By the exponential family factorization, it is easy to see that $T = \sum_{i=1}^n X_i$ is complete and sufficient for the model.

By Basu's Theorem, we then have that S and T are independent.

(b) Note that S times T is X_n. So, $\mathsf{E}[ST] = \mathsf{E}[X_n] = 1/\lambda$.

On the other hand, $\mathsf{E}[T] = n\mathsf{E}[X_1] = n/\lambda$.

By part (a), we know that S and T are independent, so we have $\mathsf{E}[ST] = \mathsf{E}[S]\mathsf{E}[T]$ and therefore

$$\mathsf{E}[S] = \frac{\mathsf{E}[ST]}{\mathsf{E}[T]} = \frac{1/\lambda}{n/\lambda} = \frac{1}{n}.$$

---

7. We already know that $S = X_{(n)}$ is complete and sufficient for this model. We want to find a function of $X_{(n)}$ that is unbiased for θ^p.

We can show that the pdf for $X_{(n)}$ is

$$f_{X_{(n)}}(x) = \frac{n}{\theta^n} x^{n-1} I_{(0,\theta)}(x).$$

Let's try

$$\begin{aligned} \mathsf{E}[X_{(n)}] &= \textstyle\int_0^\theta x \cdot \frac{n}{\theta^n} x^{n-1}\, dx \\ &= \textstyle\frac{n}{n+1}\theta \end{aligned}$$

From that integral, we can see that we will get θ^p if we compute

$$\mathsf{E}[X_{(n)}^p] = \int_0^\theta x^p \cdot \frac{n}{\theta^n} x^{n-1}\, dx = \frac{n}{n+p}\,\theta^p.$$

Therefore, the UMVUE for $\tau(\theta) = \theta^p$ is

$$\widehat{\tau(\theta)} = \frac{n+p}{n}\, X_{(n)}^p.$$

---

8. (a) First note that, when the parameter is in the indicator like this, the exponential family factorization for find a complete and sufficient statistic will never work. That factorization is about complete separation of the x's and θ ($a(\theta)$, $b(\vec{x})$, $c(\theta)$, $d(\vec{x})$) but they are stuck together in the indicator.

First, we need to find a sufficient statistic. We'll use the Factorization Criterion:

$$f(\vec{x};\theta) = \prod_{i=1}^n f(x_i;\theta) = \cdots = e^{-\sum x_i + n\theta} I_{(\theta,\infty)}(x_{(1)}) = \underbrace{e^{-\sum x_i}}_{h(\vec{x})} \underbrace{e^{n\theta} I_{(\theta,\infty)}(x_{(1)})}_{g(s(\vec{x});\theta)}$$

Thus, we see that $S = X_{(1)}$ is sufficient for θ.

To show that S is complete, we need to find the pdf for the minimum. I am running out of time and need to get these solutions posted, so I am omitting the details, but the pdf for the minimum is

$$f_{X_{(1)}}(x) = ne^{-n(x-\theta)} I_{(\theta,\infty)}(x)$$

To show completeness, assume that g is any function such that $\mathsf{E}[g(X_{(1)})] = 0$ for all θ. Then

$$0 = \mathsf{E}[g(X_{(1)})] = \int_\theta^\infty g(x)\, n\, e^{-n(x-\theta)}\, dx = ne^{n\theta} \int_\theta^\infty g(x)\, e^{-nx}\, dx$$

for all θ. This implies that

$$\int_\theta^\infty g(x)\, e^{-nx}\, dx = 0$$

or, equivalently,

$$-\int_\infty^\theta g(x)\, e^{-nx}\, dx = 0$$

and thus

$$\int_\infty^\theta g(x)\, e^{-nx}\, dx = 0$$

for all θ.

Taking the derivative of both sides with respect to θ gives

$$g(\theta)e^{-n\theta} = 0$$

for all θ. Since $e^{-n\theta} \neq 0$, we get that $g(\theta)$ must be zero for all θ. Thus, $g(X_{(1)}) = 0$ and we have that $S = X_{(1)}$ is complete for θ.

(b) We need to find a function of $X_{(1)}$ that is unbiased for θ. We consider $X_{(1)}$ itself.

$$\begin{aligned}
\mathsf{E}[X_{(1)}] &= \int_{-\infty}^\infty x f_{X_{(1)}}(x)\, dx \\
&= \int_\theta^\infty xne^{-n(x-\theta)}\, dx \\
&= ne^{n\theta} \int_\theta^\infty xe^{-nx}\, dx \\
&= e^{n\theta}[\theta e^{-n\theta} + \tfrac{1}{n} e^{-n\theta}] \\
&= \theta + \tfrac{1}{n}
\end{aligned}$$

So, $\widehat{\theta} = X_{(1)} - 1/n$.

---

9. (a) We first consider the simple versus simple hypotheses

$$H_0 : \sigma^2 = \sigma_0^2 \qquad H_1 : \sigma^2 = \sigma_1^2$$

for some fixed $\sigma_1^2 > \sigma_0^2$.

The joint pdf is

$$f(\vec{x}; \sigma^2) = (2\pi\sigma^2)^{-n/2} e^{-\frac{1}{2\sigma^2} \sum x_i^2}.$$

The likelihood ratio is

$$\begin{aligned}
\lambda(\vec{x}; \sigma_0^2, \sigma_1^2) &= \frac{f(\vec{x};\sigma_0^2)}{f(\vec{x};\sigma_1^2)} \\
&= \frac{(2\pi\sigma_0^2)^{-n/2} e^{-\frac{1}{2\sigma_0^2}\sum x_i^2}}{(2\pi\sigma_1^2)^{-n/2} e^{-\frac{1}{2\sigma_1^2}\sum x_i^2}} \\
&= (\sigma_1^2/\sigma_0^2)^{n/2} \cdot e^{-\frac{1}{2}\left(\frac{1}{\sigma_0^2} - \frac{1}{\sigma_1^2}\right)\sum x_i^2}
\end{aligned}$$

Setting this less than or equal to k and starting to move things, we get

$$e^{-\frac{1}{2}\left(\frac{1}{\sigma_0^2}-\frac{1}{\sigma_1^2}\right)\sum x_i^2} \leq (\sigma_0^2/\sigma_1^2)^{n/2}k$$

$$-\frac{1}{2}\left(\frac{1}{\sigma_0^2}-\frac{1}{\sigma_1^2}\right)\sum x_i^2 \leq \ln\left[(\sigma_0^2/\sigma_1^2)^{n/2}k\right]$$

$$\sum x_i^2 \geq \frac{\ln\left[(\sigma_0^2/\sigma_1^2)^{n/2}k\right]}{-\frac{1}{2}\left(\frac{1}{\sigma_0^2}-\frac{1}{\sigma_1^2}\right)}$$

since $\sigma_1^2 > \sigma_0^2$.

So, the best test of

$$H_0 : \sigma^2 = \sigma_0^2 \qquad H_1 : \sigma^2 = \sigma_1^2$$

for some fixed $\sigma_1^2 > \sigma_0^2$ will be to reject H_0 if

$$\sum X_i^2 \geq k_1$$

where k_1 is chosen to give a size α test.

Now let's find k_1.

$$\alpha = P\left(\sum X_i^2 \geq k_1; H_0\right)$$

Since, under H_0, $X_i \sim N(0, \sigma_0^2)$ so $X_i/\sigma_0^2 \sim N(0,1)$. Squaring a $N(0,1)$ gives a χ^2 random variable. Adding independent χ^2-random variables gives another χ^2 with all the degrees of freedom added up.

So,

$$\frac{\sum_{i=1}^n X_i^2}{\sigma_0^2} = \sum_{i=1}^n \frac{X_i^2}{\sigma_0^2} = \sum_{i=1}^n \left(\frac{X_i}{\sigma_0}\right)^2 \sim \chi^2(n)$$

So,

$$\begin{aligned}\alpha &= P\left(\sum X_i^2 \geq k_1; H_0\right) \\ &= P\left(\frac{\sum X_i^2}{\sigma_0^2} \geq k_1/\sigma_0^2; H_0\right) \\ &= P(W > k_1/\sigma_0^2)\end{aligned}$$

where $W \sim \chi^2(n)$.

So, we have that k_1/σ_0^2 is the $\chi^2(n)$ critical value that cuts off area α to the right. Our notation for this is $\chi^2_\alpha(2n)$. So

$$k_1 = \sigma_0^2\, \chi^2_\alpha(n).$$

So, the best test of size α of

$$H_0 : \sigma^2 = \sigma_0^2 \qquad H_1 : \sigma^2 = \sigma_1^2$$

for some fixed $\sigma_1^2 > \sigma_0^2$ will be to reject H_0 if

$$\sum X_i^2 \geq \sigma_0^2\, \chi^2_\alpha(n).$$

This test does not depend on the specific chosen value of σ_1^2 (with the exception that the form of the test depends on the fact that $\sigma_1^2 > \sigma_0^2$). So, this is a UMP test of size α for

$$H_0 : \sigma^2 = \sigma_0^2 \text{ versus } H_1 : \sigma^2 > \sigma_0^2.$$

(b) The power function is

$$\begin{aligned} \gamma(\sigma^2) &= P(\text{ Reject } H_0\,; \sigma^2) \\ &= P(\sum X_i^2 \geq \sigma_0^2\,\chi^2_\alpha(n); \sigma^2) \end{aligned}$$

---

10. (a) The ratio for the Neyman-Pearson test is

$$\lambda(\vec{x};\theta_0,\theta_1) = \frac{\frac{1}{\theta_0^n}\, I_{(0,\theta_0)}(x_{(n)}) \cdot I_{(0,x_{(n)})}(x_{(1)})}{\frac{1}{\theta_1^n}\, I_{(0,\theta_1)}(x_{(n)}) \cdot I_{(0,x_{(n)})}(x_{(1)})} = \left(\frac{\theta_1}{\theta_0}\right)^n \frac{I_{(0,\theta_0)}(x_{(n)})}{I_{(0,\theta_1)}(x_{(n)})} \overset{set}{\leq} k$$

The k should be something non-negative since λ is a ratio of pdfs and therefore is always non-negative. Note that if the indicator in the numerator is zero if $x_{(n)} > \theta_0$. In this case, we absolutely know that H_0 is not true since it states that all values in the sample will be between 0 and θ_0. This is reflected in the fact that $x_{(n)} > \theta_0 \Rightarrow \lambda = 0$ which is less than or equal to any valid k, so we will always reject.

On the other hand, if the indicator in the denominator is zero, this means that $x_{(n)} > \theta_1$. The N-P ratio λ becomes infinite (in a sense) which makes it NOT less than or equal to any cut-off k, so we would never reject H_0. This makes sense because $x_{(n)} > \theta_1$ implies that H_1 could not possibly be true since it says that all values in the sample are between 0 and θ_1.

All of these comments aside, this test is garbage if $x(n)$ is greater than both θ_0 and θ_1 since, in hypothesis testing, the assumption is that one of the two hypotheses is true. Since $\theta_1 < \theta_0$, and the sample came from either the $unif(0,\theta_0)$ or $unif(0,\theta_1)$ distribution, we must have that $x_{(n)} < \theta_0$, and so the indicator in the numerator is one. Thus, we have

$$\left(\frac{\theta_1}{\theta_0}\right)^n \frac{1}{I_{(0,\theta_1)}(x_{(n)})} \leq k$$

$$\Rightarrow \qquad \frac{1}{I_{(0,\theta_1)}(x_{(n)})} \leq \left(\frac{\theta_0}{\theta_1}\right)^n k$$

$$\Rightarrow I_{(0,\theta_1)}(x_{(n)}) \geq k_1$$

Now the indictor will be "large" (ie: 1) if $x_{(n)}$ is small, so this is equivalent to

$$X_{(n)} \leq k_2$$

for some k_2 such that

$$P(X_{(n)} \leq k_2; \theta_0) = \alpha$$

ie:

$$\left(\frac{k_2}{\theta_0}\right)^n = \alpha$$

$$\Rightarrow \qquad k_2 = \theta_0 \alpha^{1/n}$$

So, the best test of

$$H_0 : \theta = \theta_0 \qquad \text{versus} \qquad H_1 : \theta = \theta_1$$

is to reject H_0 if $X_{(n)} \leq \theta_0 \alpha^{1/n}$.

(b) Since the test from part (a) does not involve θ_1 (only that $\theta_1 < \theta_0$), it is UMP for

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

(c) The composite null hypothesis will only change the way the level of significance is defined

$$\begin{aligned} \alpha &= \max_{\theta \geq \theta_0} P(X_{(n)} \leq k_2; \theta) \\ &= \max_{\theta \geq \theta_0} \left(\tfrac{k_2}{\theta}\right)^n = \left(\tfrac{k_2}{\theta_0}\right)^n \\ &\Rightarrow \qquad k_2 = \theta_0 \alpha^{1/n} \end{aligned}$$

So, a UMP test of size α of

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

is to reject H_0 if $X_{(n)} \leq \theta_0 \alpha^{1/n}$.

---

11. The pdf is

$$f(x;\theta) = \theta e^{-\theta x} \, I_{(0,\infty)}(x).$$

This is also the "joint" pdf for our sample of size 1.

A likelihood is

$$L(\theta) = \theta e^{-\theta x}.$$

The log-likelihood is

$$\ell(\theta) = \ln L(\theta) = \ln \theta - \theta x.$$

Maximizing this with respect to θ gives the MLE

$$\widehat{\theta} = 1/X_1.$$

As for the restricted MLE (a rough sketch would be helpful here),

- If $\theta_0 \leq 1/X_1$, then $\widehat{\theta}_0 = 1/X_1$.
- If $\theta_0 > 1/X_1$, then $\widehat{\theta}_0 = \theta_0$.

So, the GLR is

$$\begin{aligned} \lambda(X_1) &= \frac{L(\widehat{\theta}_0)}{L(\widehat{\theta})} \\ &= \begin{cases} 1 & , \ \text{if } \theta_0 \leq 1/X_1 \\ \frac{\theta_0 e^{-\theta_0 X_1}}{\frac{1}{X_1} e^{-(1/X_1)X_1}} & , \ \text{if } \theta_0 > 1/X_1 \end{cases} \\ &= \begin{cases} 1 & , \ \text{if } X_1 \leq 1/\theta_0 \\ \theta_0 X_1 e^{-\theta_0 X_1 + 1} & , \ \text{if } X_1 > 1/\theta_0 \end{cases} \end{aligned}$$

As for the actual GLRT, it turned out waaaay harder than intended if you do it "directly" (Even for this sample of size 1!) For the record, here's how you might proceed if you want to do it "directly". (**Alternatively**, skip down to "***".)

Set

$$\alpha = \max_{\theta \geq \theta_0} P(\lambda(X_1) \leq k; \theta).$$

In order to compute this, we would first need to compute the probability. (i.e. Ignore the max for now.) Because writing that end "semicolon θ" will be cumbersome, I'll leave it out.

$$P(\lambda(X_1) \leq k) = P(\lambda(X_1) \leq k, X_1 \leq 1/\theta_0) + P(\lambda(X_1) \leq k, X_1 > 1/\theta_0)$$

$$= P(\lambda(X_1) \leq k | X_1 \leq 1/\theta_0) P(X_1 \leq 1/\theta_0) + P(\lambda(X_1) \leq k | X_1 > 1/\theta_0) P(X_1 > 1/\theta_0)$$

$$P(1 \leq k | X_1 \leq 1/\theta_0) P(X_1 \leq 1/\theta_0) + P(\theta_0 X_1 e^{-\theta_0 X_1 + 1} \leq k | X_1 > 1/\theta_0) P(X_1 > 1/\theta_0)$$

We can easily compute $P(X_1 \leq 1/\theta_0)$ and $P(X_1 > 1/\theta_0)$ for the exponential rate θ distribution.

The first term, $P(1 \leq k | X_1 \leq 1/\theta_0)$ is either 0 or 1, depending on the value of k.

The term $P(\theta_0 X_1 e^{-\theta_0 X_1 + 1} \leq k | X_1 > 1/\theta_0)$ can be thought of a little more simply as $P(\theta_0 Y e^{-\theta_0 Y + 1} \leq k)$ where Y is en exponential rate θ random variable with pdf restricted to $y > 1/\theta_0$ and renormalized so that it integrates to 1.

However, even for a "usual" exponential distribution starting at 0, this probability is hard to compute. I would suggest moving the extraneous terms to the other side of the inequality, taking the log of both sides, and looking into the "Lambert W" function. Yuck!

*** **A much simpler alternative** is to look at the GLR on the bottom of the previous page and note that it is a non-increasing function of X_1. So, having $\lambda(X_1) \leq k$ is equivalent to having $X_1 \geq k_1$ for some k_1. (!)

Thus, we have to solve

$$\begin{aligned} \alpha &= \max_{\theta \geq \theta_0} P(\lambda(X_1) \leq k; \theta) \\ &= \max_{\theta \geq \theta_0} P(X_1 \geq k_1; \theta) \\ &= \max_{\theta \geq \theta_0} e^{-\theta k_1} e^{-\theta_0 k_1} \end{aligned}$$

which implies that $k_1 = (-1/\theta_0) \ln \alpha$.

Thus, the GRLT of size α is to

$$\text{Reject } H_0 \text{ if } X_1 \geq (-1/\theta_0) \ln \alpha.$$

---

12. The joint pdf for X and Y is

$$f_{X,Y}(x, y) = \binom{n_1}{x} p_1^x (1 - p_1)^{n_1 - x} \cdot \binom{n_2}{y} p_2^y (1 - p_2)^{n_2 - y}$$

(a) The resctricted MLE:

We assume that $p_1 = p_2$ and denote the common value denoted simply by p. Then

$$f_{X,Y}(x, y) = \binom{n_1}{x}\binom{n_2}{y} p^{x+y}(1-p)^{n_1+n_2-(x+y)}$$

$$\Rightarrow \qquad L(p) = p^{x+y}(1-p)^{n_1+n_2-(x+y)}$$

$$\ln L(p) = (x+y)\ln p + (n_1+n_2-(x+y))\ln(1-p)$$

$$\frac{\partial}{\partial p}\ln L(p) = \frac{x+y}{p} - \frac{n_1+n_2-(x+y)}{1-p} \overset{set}{=} 0$$

$$\Rightarrow \qquad \hat{p}_0 = \frac{x+y}{n_1+n_2}$$

where $\hat{p}_0$ denotes the restricted MLE for p.

The unrestricted MLE's for p_1 and p_2:

Recall that the joint pdf for X and Y is

$$f_{X,Y}(x, y) = \binom{n_1}{x} p_1^x(1-p_1)^{n_1-x} \cdot \binom{n_2}{y} p_2^y(1-p_2)^{n_2-y}$$

So, a likelihood function is

$$L(p_1, p_2) = p_1^x(1-p_1)^{n_1-x} \cdot p_2^y(1-p_2)^{n_2-y}$$

and the log is

$$\ln L(p_1, p_2) = x\ln p_1 + (n_1-x)\ln(1-p_1) + y\cdot\ln p_2 + (n_2-y)\ln(1-p_2)$$

$$\frac{\partial}{\partial p_1}\ln L(p_1, p_2) = \frac{x}{p_1} - \frac{n_1-x}{1-p_1} \overset{set}{=} 0$$

$$\frac{\partial}{\partial p_2}\ln L(p_1, p_2) = \frac{y}{p_2} - \frac{n_2-y}{1-p_2} \overset{set}{=} 0$$

$$\Rightarrow \qquad \hat{p}_1 = \frac{x}{n_1}, \quad \hat{p}_2 = \frac{y}{n_2}$$

So, the GLR is

$$\lambda(\vec{x}) = \frac{\left(\frac{x+y}{n_1+n_2}\right)^{x+y}\left(1-\frac{x+y}{n_1+n_2}\right)^{n_1+n_2-(x+y)}}{\left(\frac{x}{n_1}\right)^x(1-\left(\frac{x}{n_1}\right))^{n_1-x}\cdot\left(\frac{y}{n_2}\right)^y(1-\left(\frac{y}{n_2}\right))^{n_2-y}}$$

(b) The approximate large sample GLRT of size α is to reject H_0 if

$$-2\ln\lambda(\vec{X}) \geq \chi^2_\alpha(2)$$

since the parameter space is $\{(p_1, p_2) : 0 \leq p_1 \leq 1, 0 \leq p_2 \leq 1\}$ is two-dimensional and the point $\{(p, p)\}$ is zero-dimensional. (The degrees of freedom for the χ^2 is then $2-0=2$.

13. (a) The pdf for Y_i is

$$f_{Y_i}(y_i;m) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}}(y_i - mx_i)^2.$$

The joint pdf is

$$f(\vec{y};m) \stackrel{indep}{=} f_{Y_i}(y_i;m) = (2\pi\sigma^2)^{-n/2} e^{-\frac{1}{2\sigma^2}\sum_{i=1}^n (y_i - mx_i)^2}.$$

A likelihood is

$$L(m) = e^{-\frac{1}{2\sigma^2}\sum_{i=1}^n (y_i - mx_i)^2}.$$

The log-likelihood is

$$\ln(m) = -\frac{1}{2\sigma^2}\sum_{i=1}^n (y_i - mx_i)^2.$$

Now,

$$\begin{aligned} \tfrac{d}{dm}\ln(m) &= -\tfrac{1}{2\sigma^2}\textstyle\sum_{i=1}^n 2(y_i - mx_i)(-x_i) \\ &= \tfrac{1}{\sigma^2}\textstyle\sum_{i=1}^n x_i(y_i - mx_i) \stackrel{set}{=} 0 \end{aligned}$$

Solving for m, we get the MLE

$$\widehat{m} = \frac{\sum x_i Y_i}{\sum x_i^2}.$$

The restricted MLE is $\widehat{m}_0 = m_0$, so the GLR is

$$\lambda(\vec{Y}) = \frac{L(\widehat{m}_0)}{L(\widehat{m})} = \frac{\sum_{i=1}^n (Y_i - m_0 x_i)^2}{\sum_{i=1}^n (Y_i - [(\sum_j x_j Y_j)/(\sum_j x_j^2)]x_i)^2}$$

(b) We will use Wilks' Theorem which says that $-2\ln\lambda(\vec{Y}) \stackrel{d}{\to} (1)$. (Althought it was not explicitly given, the slope m is assumed to be any real number. Since $\mathbb{R}$ is a one-dimensional space and the singleton point $\{m_0\}$ is considered a zero dimensional space, the degrees of freedom in Wilks' chi-squared is $1 - 0 = 1$.)

So, we have

$$\begin{aligned} \alpha &= P(\text{ Reject } H_0 \text{ when true }) \\ &= P(\lambda(\vec{Y}) \leq k; m_0) \\ &= P(-2\ln\lambda(\vec{Y}) \geq k_1; m_0) \\ &\approx P(W \geq k_1; m_0) \end{aligned}$$

where $W \sim \chi^2(1)$. Thus, $k_1 = \chi^2_{\alpha,1}$ and the approximate large sample GLRT is to reject H_0 if

$$-2\ln\lambda(\vec{Y}) \geq \chi^2_{\alpha,1}.$$