

JOE SEIDEL

STAT 244

HOMEWORK 6

Question 1 Rice 8.52 (b)-(d)

Let X_1, \dots, X_n be i.i.d. random variables with the density function

$$f(x | \theta) = (\theta + 1)x^\theta, 0 \leq x \leq 1$$

1. Find the mle of θ .

$$\begin{aligned} L(\theta) &= f(x_1 | \theta) \cdots f(x_n | \theta) \\ &= \prod_{i=1}^n (\theta + 1)x_i^\theta \\ &= (\theta + 1)^n \prod_{i=1}^n x_i^\theta \end{aligned}$$

Taking the log of $L(\theta)$

$$\log L(\theta) = n \log(\theta + 1) + \theta \sum_{i=1}^n \log(x_i).$$

Then

$$\frac{d}{d\theta} \log L(\theta) = \frac{n}{\theta + 1} + \sum_{i=1}^n \log(x_i).$$

Setting the above equal to 0

$$\hat{\theta} = -n \frac{1}{\sum_{i=1}^n \log(x_i)} - 1$$

2. Find the asymptotic variance of the mle.

Recall

$$\frac{1}{\tau^2(\theta)} = -E\left[\frac{d^2}{d\theta^2} \log L(\theta)\right]$$

Where we can substitute $f(x_i | \theta)$ with the log likelihood function, $L(\theta)$.

I will verify this is a maximum in the next part of the question

Also, showing $\hat{\theta}$ is maximized.

$$\frac{d^2}{d\theta^2} \log L(\theta) = -\frac{n}{(\theta + 1)^2}$$

Plugging this back into the formula for asymptotic variance, given above

$$\begin{aligned} \frac{1}{\tau^2} &= -E\left[-\frac{n}{(\theta + 1)^2}\right] \\ &= \frac{n}{(\theta + 1)^2} \end{aligned}$$

Therefore

$$\tau^2(\theta) = \frac{(\hat{\theta} + 1)^2}{n}$$

3. Find a sufficient statistic for θ .

Recall

$$L(\theta) = (\theta + 1)^n \prod_{i=1}^n x_i^\theta$$

Since $0 \leq x \leq 1$, the above can be factorized to

$$g(t, \theta) = (\theta + 1)^n t^\theta$$

Where $t = \prod_{i=1}^n x_i$ is sufficient statistic for θ .

Question 3 Rice 8.60 (a)-(e)

Let X_1, \dots, X_i be an i.i.d. sample from an exponential distribution with the density function

$$f(x | \tau) = \frac{1}{\tau} e^{-x/\tau}, 0 \leq x < \infty$$

1. Find the mle of τ .

The likelihood function

$$\begin{aligned} L(\tau) &= f(x_1 | \tau) \cdots f(x_i | \tau) \\ &= \prod_{i=1}^n \frac{1}{\tau} e^{-x_i/\tau} \\ &= \frac{1}{\tau^n} \prod_{i=1}^n e^{-x_i/\tau} \end{aligned}$$

Taking the log

$$\log L(\tau) = -n \log(\tau) - \frac{1}{\tau} \sum_{i=1}^n x_i.$$

Differentiating

$$\frac{d}{d\tau} \log L(\tau) = -\frac{n}{\tau} + \frac{\sum_{i=1}^n x_i}{\tau^2}.$$

Setting the above equal to 0

$$\hat{\tau} = \frac{1}{n} \sum_{i=1}^n x_i = \bar{X}$$

Furthermore, verify this is a maximum by checking the sign of the second derivative.

$$\begin{aligned}\frac{d^2}{d\tau^2}L(\tau) &= \frac{n}{\tau^2} - \frac{2\sum_{i=1}^n x_i}{t^3} \\ &= \frac{1}{\hat{\tau}^2} \left(n - \frac{2\bar{X}n}{\hat{\tau}} \right) \\ &= \frac{1}{\hat{\tau}^2} (n - 2n) \\ &= -\frac{n}{\hat{\tau}^2}\end{aligned}$$

Which is clearly negative.

2. What is the exact sample distribution of the mle?

LET

$$S = X_1 + X_2 + \dots + X_n$$

and find the mgf of S to be

$$\left(\frac{1/\tau}{1 - 1/\tau} \right)^n$$

combined with the reproductive property of distributions, the result is $S \sim \Gamma(n, \frac{1}{\tau})$. It remains to find the pdf of $\bar{X} = \frac{S}{n}$.

FOLLOWING Stigler's notes, (1.35), let $s = g(x) = nx$.

$$\begin{aligned}f_{\bar{X}} &= f_X(g(x)) \cdot |g'(x)| \\ &= \frac{s^{n-1}}{\tau^n \Gamma(n)} e^{-s/\tau} \cdot n \\ &= \frac{n^n x^{n-1}}{\tau^n \Gamma(n)} e^{-\frac{nx}{\tau}}, x > 0\end{aligned}$$

Which is the pdf of $\Gamma(n, \frac{n}{\tau})$ distribution.

3. Use the central limit theorem to find a normal approximation to the sample distribution.

SINCE X_i are i.i.d. with $E(X_i) = \tau$ and $\text{Var}(X_i) = \tau^2$, $X \sim N(\tau, \frac{\tau^2}{n})$ when n is large as a direct result of the CLT.

4. Show that the mle is unbiased, and find it's exact variance.

$$\begin{aligned}
 B(\hat{\tau}) &= E(\hat{\tau}) - \tau \\
 &= E(\bar{X}) - \tau \\
 &= \tau - \tau \\
 &= 0
 \end{aligned}$$

$$\begin{aligned}
 \text{Var}(\hat{\tau}) &= \text{Var}(\bar{X}) \\
 &= \text{Var}\left(\frac{1}{n} \sum_{i=1}^n x_i\right) \\
 &= \frac{1}{n^2} \sum_{i=1}^n \text{Var}(X_i) \\
 &= \frac{1}{n^2} \sum_{i=1}^n \tau^2 \\
 &= \frac{\tau^2}{n}
 \end{aligned}$$

5. Is there any other unbiased estimate with smaller variance?

ONCE MORE, recall

$$I(\tau) = -E\left[\frac{d^2}{d\tau^2} \log L(\tau)\right]$$

In part (1) of this this question, I found

$$\frac{d^2}{d\tau^2} \log L(\tau) = \frac{1}{\tau^2} (n - 2n).$$

So

$$I(\tau) = \frac{n}{\tau^2}$$

Cramer-Rao states that

$$\text{Var}(T) \geq \frac{1}{I(\tau)}$$

where $T = t(X_1, \dots, X_n)$.

Since I've found $\text{Var}(\bar{X}) = \frac{\tau^2}{n}$, \bar{X} attains the Cramer-Rao lower-bound and therefore is the smallest unbiased estimator.

I omit the n in the denominator since I defined $I(\tau)$ using likelihood function, and not $f(X_1 | \tau)$.

Question 3 Rice 8.68

Let X_1, \dots, X_i be an i.i.d. sample from a Poisson distribution with mean λ , and let $T = \sum_{i=1}^n X_i$.

1. Show that the distribution of X_1, \dots, X_i given T is independent of λ , and conclude that T is sufficient for λ .

In the last homework, I found the distribution of sum of i.i.d. Poisson and can generalize it to $T = \sum_{i=1}^n X_i \sim \text{Poi}(n\lambda)$.

In the last homework, it was shown for λ_1 and λ_2 , in this case X_i all have the same λ parameter

Next, find

$$\begin{aligned} f(x_1, \dots, x_i \mid T = t) &= \frac{\Pr(X_1 = x_1, \dots, X_i = x_i)}{\Pr(T = t)} \\ &= \frac{\prod_{i=1}^n e^{-\lambda} \frac{\lambda^{x_i}}{x_i!}}{e^{-n\lambda} \frac{(n\lambda)^t}{t!}} \\ &= \frac{t!}{n^t} \cdot \prod_{i=1}^n \frac{1}{x_i!} \end{aligned}$$

Which is not dependent on λ and therefore a sufficient statistic.

2. Show that X_1 is not sufficient.

$$\begin{aligned} f(x_1, \dots, x_n \mid X_1 = x_1) &= \frac{\Pr(X_1 = x_1, \dots, X_i = x_i)}{\Pr(X_1 = x_1)} \\ &= \prod_{i=2}^n e^{-\lambda} \frac{\lambda^{x_i}}{x_i!} \\ &= e^{-(n-1)\lambda} \cdot \lambda^{\sum_{i=2}^n x_i} \prod_{i=2}^n \frac{1}{x_i!} \end{aligned}$$

Which is dependent on λ .

3. Use Theorem A of Section 8.8.1 to show that T is sufficient. Identify the functions g and h of that theorem.

WANT TO SHOW we can factorize

$$f(x_1, \dots, x_i \mid \lambda) = g[T(x_1, \dots, x_n), \lambda] h(x_1, \dots, x_i).$$

$$\begin{aligned}
f(x_1, \dots, x_i \mid \lambda) &= \prod_{i=1}^n e^{-\lambda} \frac{\lambda^{x_i}}{x_i!} \\
&= e^{-n\lambda} \prod_{i=1}^n \frac{\lambda^{x_i}}{x_i!} \\
&= e^{-n\lambda} \lambda^{\sum_{i=1}^n x_i} \prod_{i=1}^n \frac{1}{x_i!}
\end{aligned}$$

Let $t = \sum_{i=1}^n x_i$ and $h(x) = \prod_{i=1}^n \frac{1}{x_i!}$. Then we see that $f(x_1, \dots, x_i \mid \lambda)$ can be written as the factorization of

$$e^{-n\lambda} \lambda^t \prod_{i=1}^n \frac{1}{x_i!} = g(T(x_1, \dots, x_n), \lambda) h(x_1, \dots, x_i)$$

Where

$$g(t, \lambda) = e^{-n\lambda} \lambda^t$$

and

$$h(x_1, \dots, x_i) = \prod_{i=1}^n \frac{1}{x_i!}.$$

Question 4 Rice 8.70

Use the factorization theorem to find a sufficient statistic for the exponential distribution.

THE EXPONENTIAL distribution with parameter λ is

$$f(x) = \lambda e^{-\lambda x}$$

Suppose X_1, \dots, X_n are i.i.d. random variables from such a distribution. Then

$$f(x_1, \dots, x_n \mid \lambda) = \prod_{i=1}^n \lambda e^{-\lambda x_i}$$

Which can be rewritten

$$f(x_1, \dots, x_n \mid \lambda) = \lambda^n e^{-\lambda(\sum_{i=1}^n x_i)}$$

Where $f(x \mid \lambda)$ depends only on x_1, \dots, x_i through the sufficient statistic $t = \sum_{i=1}^n x_i$ and $f(x \mid \lambda)$ is of the form

$$g\left(\sum_{i=1}^n x_i, \lambda\right) h(x)$$

Where $h(x) = 1$ and

$$g(t, \lambda) = \lambda^n e^{-\lambda(nt)}$$

Question 5

Suppose we face a pattern recognition problem, where the data consist of a single set of pixels X (where there are 16 possible pixel patterns), and there are two possible pattern θ , "o" and "6". The model is that X has the probability function $p(x | \theta)$ depending on θ , given by the following table. Find the best test for "o" versus "6" for which the chance of making the error of "6" when the pattern is "o" is no greater than 0.10. What is the power of this test?

FIRST add the likelihood ratios $\frac{p(x|\theta_0)}{p(x|\theta_6)}$ to the table and re-order the table based on the ratios.

Pixel	1	7	8	15	9	14	3	11	16	12	10	13	5	6	4	2
$p(x \theta_0)$	0	0.02	0.02	0.02	0.08	0.23	0.02	0.02	0.15	0.22	0.12	0.02	0.02	0.03	0.03	0
$p(x \theta_6)$	0	0	0	0	0.02	0.11	0.01	0.01	0.08	0.17	0.2	0.04	0.08	0.12	0.13	0.03
$\frac{p(x \theta_0)}{p(x \theta_6)}$	undef	∞	∞	∞	4	2.1	2	2	1.9	1.3	0.6	0.5	0.25	0.25	0.23	0

Let $H_0 = \theta_0$: "o" and $H_1 = \theta_6$: "6". Next find a critical value, C , to use in the likelihood ratio test.

$$\frac{p(x | H_0)}{p(x | H_1)} > C$$

The problem specifies that $Pr(\text{Reject } H_0 | H_0) = .1$. Looking at the table, it is clear that the test should be specified.

$$\frac{p(x | H_0)}{p(x | H_1)} > .5$$

Which means accept H_0 if the likelihood ratio is greater than $C = .5$.

Using the table we can also compute $Pr(\text{Accept } H_0 | H_1) = .6$ which is probability of type II error, β . The power of the test is

$$\pi = 1 - \beta = .4$$

Question 6

Suppose X has a $N(\mu, \sigma^2)$ distribution.

1. Find the Most Powerful test for testing at level $\alpha = 0.05$ the hypothesis $H_0: \mu = 6$ and $\sigma^2 = 4$ versus $H_1: \mu = 9$ and $\sigma^2 = 4$.

THE GENERAL form of the likelihood function for $X \sim N(\mu, \sigma^2)$ distributions is

$$L(\theta) = L(\mu, \sigma^2) = 2(\pi)^{-\frac{n}{2}} \sigma^{-n} e^{-\frac{1}{2\sigma^2} \sum_{i=1}^n (X_i - \mu)^2}$$

So the likelihood ratio given H_0 and H_1 can be found

See Stigler 6-6 for the algebra on likelihood ratio $\frac{L(\theta_1)}{L(\theta_0)}$

$$\begin{aligned}
\frac{L(\theta_1)}{L(\theta_0)} &= \frac{2(\pi)^{-\frac{n}{2}} \sigma_1^{-n} e^{-\frac{1}{2\sigma_1^2} \sum_{i=1}^n (X_i - \mu_1)^2}}{2(\pi)^{-\frac{n}{2}} \sigma_0^{-n} e^{-\frac{1}{2\sigma_0^2} \sum_{i=1}^n (X_i - \mu_0)^2}} \\
&= e^{\frac{1}{\sigma_0^2} [(\mu_1 - \mu_0) \sum_{i=1}^n X_i] - \frac{n}{2\sigma_0^2} [\mu_1^2 - \mu_0^2]}
\end{aligned}$$

Set up the test so that the model for $\bar{X} \sim N(\mu, \frac{\sigma^2}{n})$ and instead of dealing directly with the likelihood ratio above, reject H_0 when $\bar{X} > C$. To find C , first consider

$$\alpha = P(\bar{X} > C \mid H_0) \quad (1)$$

$$= P\left[\left(\frac{\bar{X} - \mu_0}{\frac{\sigma_0}{\sqrt{n}}}\right) > \left(\frac{C - \mu_0}{\frac{\sigma_0}{\sqrt{n}}}\right)\right] \quad (2)$$

$$= P\left[Z > \left(\frac{C - \mu_0}{\frac{\sigma_0}{\sqrt{n}}}\right)\right] \quad (3)$$

Notice that in (2) the left hand side of the inequality is the standardized normal, so $Z \sim N(0, 1)$. Then let $z_{1-\alpha}$ be the $(1 - \alpha)$ th percentage point of the standard normal, i.e $P(Z \leq z_{1-\alpha}) = 1 - \alpha$. Then it is possible to find $P(Z > z_{1-\alpha}) = \alpha$ using a "Z table". Then it is possible to find,

$$C = \mu_0 + z_{1-\alpha} \left(\frac{\sigma_0}{\sqrt{n}}\right).$$

In the context of this problem, $\alpha = .05$, $n = 1$, $\mu_0 = 6$ and $\sigma^2 = 4$. Furthermore, from the table in the back of Rice, $z_{.95} = 1.645$. These values give us the result,

$$C = 6 + 1.645(2) = 9.29.$$

In general, the model is set up so H_0 is rejected when $\bar{X} > C$. By Neyman-Pearson Lemma, no test with the same or lower α has a lower β than the likelihood ratio with the given $\alpha = .05$.

Evaluating this specific case,

$$\bar{X} = 6 < C = 9.29$$

and we therefore fail reject H_0 .

2. Find the power of this test.

THE POWER of the test is also determined using C .

$$\begin{aligned}
 \beta &= P(\bar{X} \leq C \mid H_1) \\
 &= P(\bar{X} \leq 9.92 \mid H_1) \\
 &= P\left(Z \leq \left(\frac{9.29 - 9}{2}\right)\right) \\
 &= P(Z \leq .145) \\
 &= .5576
 \end{aligned}$$

Therefore the power is given

$$\pi = 1 - \beta = .4424$$

3. Suppose that instead of the above H_1 , we have $H_1: \mu = \mu_1$ and $\sigma^2 = 4$, where $\mu_1 > 6$. Find and graph the power function .

$$\Phi\left(\frac{x-\mu}{\sigma}\right) = \frac{1}{2}\left[1 + \operatorname{erf}\left(\frac{x-\mu}{\sigma\sqrt{2}}\right)\right]$$

$$\pi = f(\mu_1) = 1 - \frac{1}{2}\left[1 + \operatorname{erf}\left(\frac{9.29 - \mu_1}{2\sqrt{2}}\right)\right], \mu_1 > 6$$

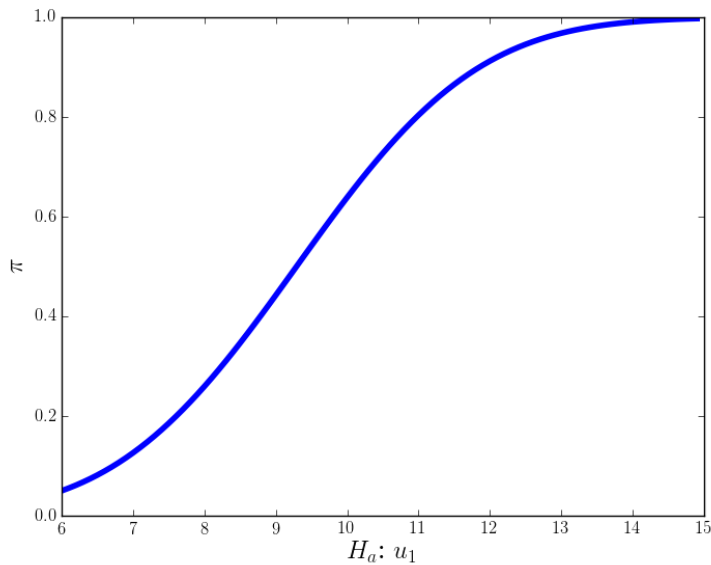


Figure 1: This function approaches 1 as μ_1 approaches ∞ .