MTH375: Mathematical Statistics - Homework #3

Cason Konzer

Key Concepts: Likelihood function, sufficient statistic, Fisher-Neyman Lemma.

1. Let $X_1, \ldots, X_n$ be a sample of *iid* $Gamma(\theta, 1)$ random variables with $\theta \in (0, \infty)$.

(a) Determine the likelihood function $L(\theta|x_1, \ldots x_n)$.

Solution:

We know that the *pdf* for an $X \sim Gamma(\alpha, \beta)$ is $f_X(x) = \dfrac{\beta^\alpha}{(\alpha-1)!} x^{\alpha-1} e^{-\beta x}$.

We thus have the *pdf* for our $X_i \sim Gamma(\theta, 1) = f_{X_i}(x_i) = \dfrac{1}{(\theta-1)!} x_i^{\theta-1} e^{-x_i}$.

We can now obtain our likelihood function $L(\theta|x_1, \ldots x_n)$.

- $L(\theta|x_1, \ldots x_n) = \dfrac{1}{(\theta-1)!} x_1^{\theta-1} e^{-x_1} \cdot \dfrac{1}{(\theta-1)!} x_2^{\theta-1} e^{-x_2} \cdots \dfrac{1}{(\theta-1)!} x_n^{\theta-1} e^{-x_n}$
- $L(\theta|x_1, \ldots x_n) = \dfrac{x_1^{\theta-1} \cdot x_2^{\theta-1} \cdots x_n^{\theta-1}}{((\theta-1)!)^n} e^{-x_1-x_2-.-.-.-x_n} = \dfrac{x_1^\theta x_1^{-1} \cdot x_2^\theta x_2^{-1} \cdots x_n^\theta x_n^{-1}}{((\theta-1)!)^n} e^{-\sum_{i=1}^n x_i}$
- $L(\theta|x_1, \ldots x_n) = \dfrac{x_1^\theta \cdot x_2^\theta \cdots x_n^\theta}{((\theta-1)!)^n (x_1 \cdot x_2 \cdots x_n)} e^{-\sum_{i=1}^n x_i} = \dfrac{\prod_{i=1}^n x_i^\theta}{((\theta-1)!)^n \prod_{i=1}^n x_i} e^{-\sum_{i=1}^n x_i}$

(b) Use the Fisher–Neyman factorization lemma to determine a sufficient statistic S for θ.

Solution:

From the Fisher–Neyman factorization lemma we will factorize the likelihood function into $g(S, \theta) \cdot h(X_1, \ldots, X_n)$.

- $g(S, \theta) = \dfrac{\prod_{i=1}^n x_i^\theta}{((\theta-1)!)^n \prod_{i=1}^n x_i}$
- $h(X_1, \ldots, X_n) = e^{-\sum_{i=1}^n x_i}$

We can now see that $S = \displaystyle\prod_{i=1}^n x_i$ is a sufficient statistic for θ.

2. Let $X_1, \dots, X_n$ be a sample of *iid* $Beta(4, \theta)$ random variables with $\theta \in (0, \infty)$. A $Beta(a, b)$ random variable X has *pdf* ...

$$f_X(x) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} x^{a-1}(1-x)^{b-1} \text{ for } 0 \le x \le 1.$$

(a) Determine the likelihood function $L(\theta|x_1, \dots x_n)$.

Solution:

We know the *pdf* for our $X_i \sim Beta(4, \theta)$ is $f_{X_i}(x_i) = \dfrac{\Gamma(4+\theta)}{\Gamma(4)\Gamma(\theta)} x_i^3 (1-x_i)^{\theta-1}$ for $0 \le x \le 1$.

We can now obtain our likelihood function $L(\theta|x_1, \dots x_n)$.

- $L(\theta|x_1, \dots x_n) = \dfrac{\Gamma(4+\theta)}{\Gamma(4)\Gamma(\theta)} x_1^3 (1-x_1)^{\theta-1} \cdot \dfrac{\Gamma(4+\theta)}{\Gamma(4)\Gamma(\theta)} x_2^3 (1-x_2)^{\theta-1} \cdots \dfrac{\Gamma(4+\theta)}{\Gamma(4)\Gamma(\theta)} x_n^3 (1-x_n)^{\theta-1}$
- $L(\theta|x_1, \dots x_n) = \dfrac{\big(\Gamma(4+\theta)\big)^n}{\big(\Gamma(4)\Gamma(\theta)\big)^n} (x_1^3 \cdot x_2^3 \cdots x_n^3)((1-x_1)^{\theta-1} \cdot (1-x_2)^{\theta-1} \cdots (1-x_n)^{\theta-1})$
- $L(\theta|x_1, \dots x_n) = \dfrac{\big(\Gamma(4+\theta)\big)^n}{\big(\Gamma(4)\Gamma(\theta)\big)^n} \left(\prod_{i=1}^{n} x_i^3\right) \left(\dfrac{(1-x_1)^\theta}{1-x_1} \cdot \dfrac{(1-x_2)^\theta}{1-x_2} \cdots \dfrac{(1-x_n)^\theta}{1-x_n}\right)$
- $L(\theta|x_1, \dots x_n) = \dfrac{\big(\Gamma(4+\theta)\big)^n}{\big(\Gamma(4)\Gamma(\theta)\big)^n} \left(\prod_{i=1}^{n} x_i^3\right) \left(\dfrac{\prod_{i=1}^n (1-x_i)^\theta}{\prod_{i=1}^n (1-x_i)}\right)$

(b) Use the Fisher–Neyman factorization lemma to determine a sufficient statistic S for θ.

Solution:

From the Fisher–Neyman factorization lemma we will factorize the likelihood function into $g(S, \theta) \cdot h(X_1, \dots, X_n)$.

- $g(S, \theta) = \dfrac{\big(\Gamma(4+\theta)\big)^n \prod_{i=1}^n (1-x_i)^\theta}{\big(\Gamma(4)\Gamma(\theta)\big)^n \prod_{i=1}^n (1-x_i)}$
- $h(X_1, \dots, X_n) = \prod_{i=1}^{n} x_i^3$

We can now see that $S = \prod_{i=1}^{n} (1-x_i)$ is a sufficient statistic for θ.

3. Let $X_1, \ldots, X_n$ be a sample of *iid* $Beta(\theta_1, \theta_2)$ random variables with $\theta \in \mathbb{R}^+ \times \mathbb{R}^+$. Use the Fisher–Neyman factorization lemma to determine a sufficient statistic S for $\overrightarrow{\theta}$.

Solution:

We know the *pdf* for our $X_i \sim Beta(\theta_1, \theta_2)$ is $f_{X_i}(x_i) = \dfrac{\Gamma(\theta_1+\theta_2)}{\Gamma(\theta_1)\Gamma(\theta_2)} x^{\theta_1 - 1}(1-x)^{\theta_2 - 1}$ for $0 \leq x \leq 1$.

We can now obtain our likelihood function $L(\theta | x_1, \ldots x_n)$.

- $L(\theta | x_1, \ldots x_n) = \dfrac{\Gamma(\theta_1+\theta_2)}{\Gamma(\theta_1)\Gamma(\theta_2)} x_1^{\theta_1 - 1}(1-x_1)^{\theta_2 - 1} \cdots \dfrac{\Gamma(\theta_1+\theta_2)}{\Gamma(\theta_1)\Gamma(\theta_2)} x_n^{\theta_1 - 1}(1-x_n)^{\theta_2 - 1}$
- $L(\theta | x_1, \ldots x_n) = \dfrac{\big(\Gamma(\theta_1+\theta_2)\big)^n}{\big(\Gamma(\theta_1)\Gamma(\theta_2)\big)^n} (x_1^{\theta_1 - 1} \cdots x_n^{\theta_1 - 1})((1-x_1)^{\theta_2 - 1} \cdots (1-x_n)^{\theta_2 - 1})$
- $L(\theta | x_1, \ldots x_n) = \dfrac{\big(\Gamma(\theta_1+\theta_2)\big)^n}{\big(\Gamma(\theta_1)\Gamma(\theta_2)\big)^n} \left(\dfrac{x_1^{\theta_1}}{x_1} \cdots \dfrac{x_n^{\theta_1}}{x_n}\right)\left(\dfrac{(1-x_1)^{\theta_2}}{(1-x_1)} \cdots \dfrac{(1-x_n)^{\theta_2}}{(1-x_n)}\right)$
- $L(\theta | x_1, \ldots x_n) = \dfrac{\big(\Gamma(\theta_1+\theta_2)\big)^n}{\big(\Gamma(\theta_1)\Gamma(\theta_2)\big)^n} \left(\dfrac{\prod_{i=1}^n x_i^{\theta_1}}{\prod_{i=1}^n x_i}\right)\left(\dfrac{\prod_{i=1}^n (1-x_i)^{\theta_2}}{\prod_{i=1}^n (1-x_i)}\right)$

From the Fisher–Neyman factorization lemma we will factorize the likelihood function into $g(S, \theta) \cdot h(X_1, \ldots, X_n)$.

- $g(S, \theta) = \dfrac{\big(\Gamma(\theta_1+\theta_2)\big)^n \big(\prod_{i=1}^n x_i^{\theta_1}\big)\big(\prod_{i=1}^n (1-x_i)^{\theta_2}\big)}{\big(\Gamma(\theta_1)\Gamma(\theta_2)\big)^n \big(\prod_{i=1}^n x_i\big)\big(\prod_{i=1}^n (1-x_i)\big)}$
- $h(X_1, \ldots, X_n) = 1$

We can now see that $S = \left(\displaystyle\prod_{i=1}^n x_i,\ \prod_{i=1}^n (1-x_i)\right)$ is a sufficient statistic for $\overrightarrow{\theta} = (\theta_1, \theta_2)$.

4. Let $X_1, \dots, X_n$ be a sample of *iid* random variables with *pdf*: $f_X(x) = \theta x^{\theta-1}$ for $x \in (0,1)$ and $\theta \in (0, \infty)$. Use the Fisher–Neyman factorization lemma to determine a sufficient statistic S for θ.

Solution:

We know the *pdf* for our X_i is $f_{X_i}(x_i) = \theta x^{\theta-1}$ for $x \in (0,1)$ and $\theta \in (0,\infty)$.

We can now obtain our likelihood function $L(\theta|x_1, \dots x_n)$.

- $L(\theta|x_1,\dots x_n) = \theta x_1^{\theta-1} \cdot \theta x_2^{\theta-1} \cdots \theta x_n^{\theta-1} = \theta^n \left(\frac{x_1^\theta}{x_1} \cdot \frac{x_2^\theta}{x_2} \cdots \frac{x_n^\theta}{x_n}\right)$
- $L(\theta|x_1,\dots x_n) = \dfrac{\theta^n \prod_{i=1}^n x_i^\theta}{\prod_{i=1}^n x_i}$

From the Fisher–Neyman factorization lemma we will factorize the likelihood function into $g(S,\theta) \cdot h(X_1, \dots, X_n)$.

- $g(S,\theta) = \dfrac{\theta^n \prod_{i=1}^n x_i^\theta}{\prod_{i=1}^n x_i}$
- $h(X_1, \dots, X_n) = 1$

We can now see that $S = \displaystyle\prod_{i=1}^n x_i$ is a sufficient statistic for θ.

5. Consider the family of distributions with pmf: $p_X(x) = \begin{cases} p & \text{if } x = -1 \\ 2p & \text{if } x = 0 \\ 1-3p & \text{if } x = 1 \end{cases}$

Here p is an unknown paramater and $0 \le p \le \frac{1}{3}$.

Let $X_1, \dots, X_n$ be iid with common pmf a member of this family. Consider the statistics:

$$\begin{aligned} A = &\quad \text{the number of } i \text{ with } X_i = -1, \\ B = &\quad \text{the number of } i \text{ with } X_i = 0, \\ C = &\quad \text{the number of } i \text{ with } X_i = 1. \end{aligned}$$

(i) Write down the joint pmf of $X_1, \dots, X_n$. This is most easily done using the statistics A, B, C.

Solution:

$p_{X_1,\dots,X_n}(x_1, \dots, x_n) = p^A \cdot (2p)^B \cdot (1-3p)^C = p^A \cdot 4p^B \cdot (1-3p)^C = 4p^{A+B} \cdot (1-3p)^C.$

(ii) Use the Fisher-Neyman lemma to show that C is a sufficient statistic for p. (Hint: Use the fact that $A + B + C = n$.)

Solution:

We know that $A + B = n - C$.

Thus we have $L(p|x_1, \dots x_n) = p_{X_1,\dots,X_n}(x_1, \dots, x_n) = 4p^{n-C} \cdot (1-3p)^C = \dfrac{4p^n \cdot (1-3p)^C}{p^C}$.

From the Fisher–Neyman factorization lemma we will factorize the likelihood function into $L(p|x_1, \dots x_n) = \Big(g(S,p) = \dfrac{4p^n \cdot (1-3p)^C}{p^C}\Big) \cdot \Big(h(X_1, \dots, X_n) = 1\Big)$.

We can now see that $S = C$ is a sufficient statistic for p.

(iii) Does it appear that A is also sufficient for p? Explain why or why not.

Solution:

A does not appear to be sufficient, as without knowing C we are unable to determine the expansion of $(1-3p)^C$.

6. Let X_1, X_2, X_3 be a sample of iid, $Bin(1,p)$. Let $T = X_1 + X_2 + 2X_3$. The purpose of this problem is to determine whether T is a sufficient statistic for p. Recall that the definition says that T is sufficient for p if for all $p \in [0,1]$:

$$p_{X_1,X_2,X_3|T}(x_1, x_2, x_3|T = x_1 + x_2 + 2x_3) \text{ does not depend on } p.$$

Let's examine one particular instance of this definition.

Compute $p_{X_1,X_2,X_3|T}(0,0,1|T = 2)$. Does it depend on p? Is T a sufficient statistic for p?

Solution:

We know that $p_{X_1,X_2,X_3|T}(x_1, x_2, x_3|T = x_1 + x_2 + 2x_3) = \dfrac{p_{X_1,X_2,X_3}(x_1, x_2, x_3)}{p_T(T = x_1 + x_2 + 2x_3)}$

It follows that $p_{X_1,X_2,X_3|T}(0,0,1|T = 2) = \dfrac{p_{X_1,X_2,X_3}(0,0,1)}{p_T(T = x_1 + x_2 + 2x_3 = 2)}$

We can see that there are only two sums such that $T = 2$.

- $(x_1 = 1, x_2 = 1, x_3 = 0)$.
- $(x_1 = 0, x_2 = 0, x_3 = 1)$.

Thus we have $p_T(T = x_1 + x_2 + 2x_3 = 2) = p_{X_1,X_2,X_3}(1,1,0) + p_{X_1,X_2,X_3}(0,0,1)$

- $p_{X_1,X_2,X_3}(1,1,0) = p^{1+1+0} \cdot (1-p)^{0+0+1} = p^2(1-p)^1$.
- $p_{X_1,X_2,X_3}(0,0,1) = p^{0+0+1} \cdot (1-p)^{1+1+0} = p^1(1-p)^2$.
- $p_T(T = x_1 + x_2 + 2x_3 = 2) = p^2(1-p)^1 + p^1(1-p)^2 = p(1-p)(p+1-p) = p(1-p)$.

We can now solve for $p_{X_1,X_2,X_3|T}(0,0,1|T = 2)$.

$$\frac{p_{X_1,X_2,X_3}(0,0,1)}{p_T(T = x_1 + x_2 + 2x_3 = 2)} = \frac{p(1-p)^2}{p(1-p)} = 1 - p.$$

We can see that this example of the conditional pmf does depend on p and thus T is not a sufficient statistic for p.