Statistics: Tutorial Week 3 - Solutions to Practice Exercises

Practice Exercises

Exercise 1. We have the statistical model $\{\text{Poisson}(\lambda) \mid \lambda > 0\}$.

a. Show that $T(\boldsymbol{X}) = \sum_{i=1}^n X_i$ is sufficient by calculating the conditional distribution of $\boldsymbol{X}$ given $T(\boldsymbol{X})$.

b. Show that $T(\boldsymbol{X}) = \sum_{i=1}^n X_i$ is sufficient by using the factorization theorem.

Solution.

a. Recall that the sum of n iid Poisson(λ) random variables is distributed Poisson($n\lambda$). Therefore

$$P_\lambda(\boldsymbol{X} = \boldsymbol{x} \mid T(\boldsymbol{X}) = t) = \frac{P_\lambda(\boldsymbol{X} = \boldsymbol{x}; T(\boldsymbol{X}) = t)}{P_\lambda(T(\boldsymbol{X}) = t)}.$$

The top probability is zero if $\sum_{i=1}^n x_i \neq t$. If $\sum_{i=1}^n x_i = t$, then

$$\frac{P_\lambda(\boldsymbol{X} = \boldsymbol{x}; T(\boldsymbol{X}) = t)}{P_\lambda(T(\boldsymbol{X}) = t)} = \frac{P_\lambda(\boldsymbol{X} = \boldsymbol{x})}{P_\lambda(T(\boldsymbol{X}) = t)} = \frac{\prod_{i=1}^n e^{-\lambda} \frac{\lambda^{x_i}}{x_i!}}{e^{-n\lambda} \frac{(n\lambda)^t}{t!}} = \frac{t!}{n^t \prod_{i=1}^n x_i!}.$$

The final expression does not depend on λ anymore, so T is sufficient.

b. We have

$$f(\boldsymbol{x} \mid \lambda) = \prod_{i=1}^n e^{-\lambda} \frac{\lambda^{x_i}}{x_i!} = e^{-n\lambda} \lambda^{\sum_{i=1}^n x_i} \prod_{i=1}^n \frac{1}{x_i!}.$$

We have $g(T(\boldsymbol{x}) \mid \lambda) = e^{-n\lambda} \lambda^{\sum_{i=1}^n x_i}$ and $h(\boldsymbol{x}) = \prod_{i=1}^n \frac{1}{x_i!}$, so $T(\boldsymbol{X})$ is sufficient by the factorization theorem.

Exercise 2. Let $X_1, \ldots, X_n$ be iid distributed with cdf

$$P(X \leq x) = \begin{cases} 0 & x < 0 \\ (x/\beta)^\alpha & 0 \leq x \leq \beta \\ 1 & x > \beta \end{cases}$$

a. Find the pdf of X_1.

b. Find a two-dimensional sufficient statistic for (α, β).

c. Is this statistical model an exponential family?

d. What system of equations do you have to solve to find the method of moment estimators for α and β? Note, don't actually solve them!

SOLUTION.

a. Let $\boldsymbol{\theta} = (\alpha, \beta)$. Then

$$g(x \mid \boldsymbol{\theta}) = \frac{d}{dx}\left(\frac{x}{\beta}\right)^{\alpha} = \left(\frac{x}{\beta}\right)^{\alpha}\frac{\alpha}{x} \qquad \text{for } 0 \leq x \leq \beta.$$

Note that the domain depends on parameters, so we include the indicator function

$$g(x \mid \boldsymbol{\theta}) = \left(\frac{x}{\beta}\right)^{\alpha}\frac{\alpha}{x}\mathbb{1}_{[0,\beta]}(x).$$

b. We have

$$\begin{aligned} f(\boldsymbol{x} \mid \boldsymbol{\theta}) &= \prod_{i=1}^{n}\left(\frac{x_i}{\beta}\right)^{\alpha}\frac{\alpha}{x_i}\mathbb{1}_{[0,\beta]}(x) \\ &= \left(\frac{\alpha}{\beta^{\alpha}}\right)^{n}\left(\prod_{i=1}^{n}x_i^{\alpha-1}\right)\left(\prod_{i=1}^{n}\mathbb{1}_{[0,\infty)}(x_i)\right)\left(\prod_{i=1}^{n}\mathbb{1}_{(-\infty,\beta]}(x_i)\right) \\ &= \left(\frac{\alpha}{\beta^{\alpha}}\right)^{n}\left(\prod_{i=1}^{n}x_i\right)^{\alpha-1}\mathbb{1}_{[0,\infty)}(x_{(1)})\mathbb{1}_{(-\infty,\beta]}(x_{(n)}). \end{aligned}$$

We find that $T(\boldsymbol{X}) = \left(\prod_{i=1}^{n} X_i, X_{(n)}\right)$ is a sufficient statistic by the factorization theorem.

c. No, the support of the density depends on β, which means that we have to include indicator functions in the pdf. These indicator functions cannot be rewritten to the required form for exponential families.

d. The parameter $\boldsymbol{\theta}$ is two dimensional, so we have to use two moments.

$$\begin{aligned} \mathbb{E}_{\theta}(X_1) &= \int_0^{\beta} xg(x \mid \theta)dx = \int_0^{\beta} x\left(\frac{x}{\beta}\right)^{\alpha}\frac{\alpha}{x}dx = \frac{\alpha}{\beta^{\alpha}}\int_0^{\beta} x^{\alpha}dx = \frac{\alpha\beta}{\alpha+1}. \\ \mathbb{E}_{\theta}(X_1^2) &= \int_0^{\beta} x^2 g(x \mid \theta)dx = \int_0^{\beta} x^2\left(\frac{x}{\beta}\right)^{\alpha}\frac{\alpha}{x}dx = \frac{\alpha}{\beta^{\alpha}}\int_0^{\beta} x^{\alpha+1}dx = \frac{\alpha\beta^2}{\alpha+2}. \end{aligned}$$

Therefore we have to solve the following system of two equations with two unknowns:

$$\begin{aligned} \frac{\alpha\beta}{\alpha+1} &= \overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i. \\ \frac{\alpha\beta^2}{\alpha+2} &= \overline{X^2} = \frac{1}{n}\sum_{i=1}^{n} X_i^2. \end{aligned}$$

Actually solving the system would give the expressions

$$\widehat{\alpha}_{MOM} = \frac{\left(\sqrt{\overline{X^2}}\right)}{\sqrt{\overline{X^2} - \overline{X}^2}} - 1.$$

$$\widehat{\beta}_{MOM} = \sqrt{\overline{X^2}} \frac{\overline{X}}{\sqrt{\overline{X^2}} - \sqrt{\overline{X^2} - \overline{X}^2}}.$$

Exercise 3. Let $X_1, \ldots, X_n$ be a random sample from a population belonging to the exponential family

$$g(x \mid \theta) = h(x)c(\theta)e^{\sum_{j=1}^m w_j(\theta)t_j(x)}.$$

Prove that $T(\boldsymbol{X}) = (\sum_{i=1}^n t_1(X_i), \ldots, \sum_{i=1}^n t_m(X_i)$ is a sufficient statistic for θ_0 by using the factorization theorem.

Solution. We can rewrite

$$\begin{aligned} f(\boldsymbol{x} \mid \theta) &= \prod_{i=1}^n h(x_i)c(\theta)e^{\sum_{j=1}^m w_j(\theta)t_j(x_i)} \\ &= \left(\prod_{i=1}^n h(x_i)\right) c(\theta)^n e^{\sum_{i=1}^n \sum_{j=1}^m w_j(\theta)t_j(x_i)} \\ &= \left(\prod_{i=1}^n h(x_i)\right) c(\theta)^n e^{\sum_{j=1}^m w_j(\theta)\left(\sum_{i=1}^n t_j(x_i)\right)}. \end{aligned}$$

We have $g(T(\boldsymbol{x}) \mid \theta) = c(\theta)^n e^{\sum_{j=1}^m w_j(\theta)\left(\sum_{i=1}^n t_j(x_i)\right)}$ and $h(x) = (\prod_{i=1}^n h(x_i))$. The result follows by the factorization theorem.

Exercise 4. Suppose we have the statistical model $\{g(x \mid \theta) \mid \theta > 0\}$, where

$$g\left(x \mid \theta\right) = \theta x^{\theta - 1} \qquad \text{if } 0 \le x \le 1.$$

a. Find a sufficient statistic for θ.

b. Find the moment estimator for θ_0.

c. Is the moment estimator based on a sufficient statistic? What does this tell us?

Solution.

a. We can rewrite the pdf in the form

$$g(x \mid \theta) = \theta \exp((\theta - 1)\log(x)),$$

which is clearly a member of the exponential family with $h(x) = 1$, $c(\theta) = \theta$, $w(\theta) = (\theta - 1)$ and $t(x) = \log(x)$. Then, the sufficient statistic is given by $T(\boldsymbol{X}) = \sum_{i=1}^n \log(X_i)$.

b. This is the pdf of the Beta(θ,1) distribution, with $\mathbb{E}(X) = \frac{\theta}{\theta+1}$. Alternatively, by integration

$$\mathbb{E}(X) = \theta \int_0^1 x^\theta dx = \theta \left[\frac{1}{\theta+1} x^{\theta+1} \right]_{x=0}^1 = \frac{\theta}{\theta+1}.$$

Then,

$$\begin{aligned} \bar{X} &\overset{s}{=} \frac{\theta}{\theta+1} \\ &\Rightarrow \bar{X}(\hat{\theta}+1) = \hat{\theta} \\ &\Rightarrow (\bar{X}-1)\hat{\theta} = -\bar{X} \\ &\Rightarrow \hat{\theta} = \frac{\bar{X}}{1-\bar{X}}. \end{aligned}$$

c. No, it is not based on a sufficient statistic. This indicates that the estimator does not include all possible information, and it might be possible to find a better estimator.