BIOSTAT 602 Biostatistical Inference
Homework 02

Ashton Baker

Thursday January 19, 2017

1. Let $X_1, \dots, X_n$ be *i.i.d.* random variables from the probability density function of the following form:

$$f_X(x|\mu,\sigma) = \frac{1}{\sigma} e^{-(x-\mu)/\sigma}, \quad \mu < x < \infty, \quad 0 < \sigma < \infty.$$

   (a) Assuming that μ is known, find a one-dimensional sufficient statistic for σ.
   Solution. The pdf can be rewritten

$$f(x|\mu,\sigma) = \frac{1}{\sigma} e^{-(x-\mu)/\sigma} 1_{\{x>\mu\}}$$

   so the joint pdf of $\mathbf{X}$ is

$$\begin{aligned} f_{\mathbf{X}}(\mathbf{x}|\mu,\sigma) &= \prod_{i=1}^{n} \frac{1}{\sigma} e^{-(x_i-\mu)} 1_{\{x_i>\mu\}} \\ &= \frac{e^{n\mu/\sigma}}{\sigma^n} e^{-\frac{1}{\sigma}\sum_{i=1}^{n} x_i} 1_{\{x_{(1)}>\mu\}} \end{aligned}$$

   So if $T(\mathbf{X}) = \sum_{i=1}^{n} X_i$, then we can construct

$$\begin{aligned} g\left(T(\mathbf{x}),\sigma\right) &= \frac{e^{n\mu/\sigma}}{\sigma^n} e^{-\frac{1}{\sigma}\sum_{i=1}^{n} x_i} \\ h(\mathbf{x}) &= 1_{\{x_{(1)}>\mu\}} \end{aligned}$$

   such that $f_{\mathbf{X}}(\mathbf{x}|\mu,\sigma) = g\left(T(\mathbf{x}),\sigma\right) h(\mathbf{x})$, proving that $T(\mathbf{x})$ is sufficient for σ.

   (b) Assuming that σ is known, find a one-dimensional sufficient statistic for μ.
   Solution. For $T(\mathbf{X}) = X_{(1)}$, we can construct

$$\begin{aligned} g(T(\mathbf{x}),\mu) &= 1_{\{x_{(1)}>\mu\}} \\ h(\mathbf{x}) &= \frac{e^{n\mu/\sigma}}{\sigma^n} e^{-\frac{1}{\sigma}\sum_{i=1}^{n} x_i} \end{aligned}$$

   such that $f_{\mathbf{X}}(\mathbf{x}|\mu,\sigma) = g\left(T(\mathbf{x}),\mu\right) h(\mathbf{x})$, proving that $T(\mathbf{x})$ is sufficient for μ.

   (c) Assuming that both parameters are unknown, find a two-dimensional sufficient statistic for (μ,σ).
   Solution. Simply combine the answers for (a) and (b) to obtain the two-dimensional statistic $T(\mathbf{X}) = \left(\sum_{i=1}^{n} X_i, X_{(1)}\right)$. Then construct

$$\begin{aligned} g(T(\mathbf{x}),(\sigma,\mu)) &= \frac{e^{n\mu/\sigma}}{\sigma^n} e^{-\frac{1}{\sigma}\sum_{i=1}^{n} x_i} 1_{\{x_{(1)}>\mu\}} \\ h(\mathbf{x}) &= 1 \end{aligned}$$

   such that $f_{\mathbf{X}}(\mathbf{x}|\mu,\sigma) = g\left(T(\mathbf{x}),(\sigma,\mu)\right) h(\mathbf{x})$, proving that $T(\mathbf{x})$ is sufficient for (σ,μ).

2. Let $X_1, \dots, X_n$ be *i.i.d* random variables from $N\left(0,\sigma^2\right)$

$$f_X(x|\sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{x^2}{2\sigma^2}\right), \quad -\infty < x < \infty, \quad \sigma^2 > 0$$

   (a) Apply the Factorization Theorem to show that $T(\mathbf{X}) = \sum_{i=1}^{n} X_i^2$ is a sufficient statistic for the parameter σ^2.
   Solution.

$$\begin{aligned} f_{\mathbf{X}}(\mathbf{x}|\sigma^2) &= \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{x_i^2}{2\sigma^2}\right) \\ &= \left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^n \exp\left(\sum_{i=1}^{n} -\frac{x_i^2}{2\sigma^2}\right) \\ &= \left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^n \exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^{n} x_i^2\right) \cdot 1 \end{aligned}$$

So if $g(\sum x_i^2, \sigma) = f_{\mathbf{X}}(x|\sigma^2)$ and $h(\mathbf{x}) = 1$, then we have factored $f_{\mathbf{X}}(x|\sigma^2) = g(\sum x_i^2, \sigma)h(\mathbf{x})$ and by the Factorization Theorem $\sum_{i=1}^n X_i^2$ is sufficient for σ^2.

(b) Is $\sum_{i=1}^n X_i^2$ also a minimal sufficient statistic for σ^2? Justify your answer.

Solution. The ratio

$$\begin{aligned}\frac{f_{\mathbf{X}}(\mathbf{x}|\sigma^2)}{f_{\mathbf{Y}}(\mathbf{y}|\sigma^2)} &= \frac{\left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^n \exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^n x_i^2\right)}{\left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^n \exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^n y_i^2\right)} \\ &= \frac{\exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^n x_i^2\right)}{\exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^n y_i^2\right)} \\ &= \exp\frac{1}{2\sigma^2}\left(\sum_{i=1}^n y_i^2 - \sum_{i=1}^n x_i^2\right)\end{aligned}$$

is constant with respect to σ^2 if and only if $T(\mathbf{x}) = T(\mathbf{y})$, so yes, T is a minimal sufficient statistic for σ^2.

3. Let $X_1, \ldots, X_n$ be *i.i.d.* random variables from a Poisson distribution whose probability mass function is given by

$$f_X(x|\lambda) = \frac{e^{-\lambda}\lambda^x}{x!}, \quad x = 0, 1, 2, \ldots, \quad \lambda > 0.$$

(a) Find a one-dimensional sufficient statistic for the parameter λ.

Solution. We begin by finding the joint probability mass function of $\mathbf{X} = \{X_1, X_2, \ldots, X_n\}$.

$$\begin{aligned}f_{\mathbf{X}}(\mathbf{x}|\lambda) &= \prod_{i=1}^n \frac{e^{-\lambda}\lambda^{x_i}}{x_i!} \\ &= \left(\prod_{i=1}^n \frac{1}{x_i!}\right)\left(e^{-n\lambda}\lambda^{\sum_{i=1}^n x_i}\right)\end{aligned}$$

So $T(\mathbf{X}) = \sum_{i=1}^n X_i$ is sufficient for λ.

(b) Show that your answer in (a) is also a minimal sufficient statistic.

Solution. For two sample points $\mathbf{X} = \{X_1, \ldots, X_m\}$ and $\mathbf{Y} = \{Y_1, \ldots, Y_n\}$, the ratio

$$\begin{aligned}\frac{f_{\mathbf{X}}(\mathbf{x}|\lambda)}{f_{\mathbf{Y}}(\mathbf{y}|\lambda)} &= \left(\prod_i \frac{1}{x_i}\right)\left(\prod_j \frac{1}{y_j}\right)\left(\frac{e^{-n\lambda}}{e^{-n\lambda}}\right)\left(\frac{\lambda^{\sum_i x_i}}{\lambda^{\sum_j y_j}}\right) \\ &= \left(\prod_i \frac{1}{x_i}\right)\left(\prod_j \frac{1}{y_j}\right)\lambda^{\sum_i x_i - \sum_j y_j}\end{aligned}$$

is constant as a function of λ if and only if $\sum_i x_i = \sum_j y_j$, so T is a minimal sufficient statistic.

4. Let $X_1, \ldots, X_n$ be a random sample from Beta(α, β). Find a joint sufficient statistic for (α, β).

Solution. The probability distribution function for $\beta\alpha, \beta$ is

$$f(x|\alpha, \beta) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}x^{\alpha-1}(1-x)^{\beta-1},$$

so the joint pdf of $\mathbf{X}$ is

$$f_{\mathbf{X}}(\mathbf{x}|\alpha, \beta) = \prod_{i=1}^n \left(\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}x_i^{\alpha-1}(1-x_i)^{\beta-1}\right)$$

So the ratio

$$\frac{f_{\mathbf{X}}(\mathbf{x}|\alpha,\beta)}{f_{\mathbf{Y}}(\mathbf{y}|\alpha,\beta)} = \frac{\prod_{i=1}^n \left(\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x_i^{\alpha-1}(1-x_i)^{\beta-1}\right)}{\prod_{i=1}^n \left(\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} y_i^{\alpha-1}(1-y_i)^{\beta-1}\right)}$$
$$= \frac{\prod_{i=1}^n \left(x_i^{\alpha-1}\right)\prod_{i=1}^n \left((1-x_i)^{\beta-1}\right)}{\prod_{i=1}^n \left(y_i^{\alpha-1}\right)\prod_{i=1}^n \left((1-y_i)^{\beta-1}\right)}$$

is constant with respect to (α, β) if and only if

$$\left(\prod_{i=1}^n x_i^{\alpha-1}, \prod_{i=1}^n (1-x_i)^{\beta-1}\right) = \left(\prod_{i=1}^n y_i^{\alpha-1}, \prod_{i=1}^n (1-y_i)^{\beta-1}\right)$$

so

$$T(\mathbf{X}) = \left(\prod_{i=1}^n X_i^{\alpha-1}, \prod_{i=1}^n (1-X_i)^{\beta-1}\right)$$

is minimally sufficient (and therefore sufficient) for (α, β).

5. Let $X_1, \ldots, X_n$ be a random sample from Cauchy$(\theta, 1)$. Find a minimal sufficient statistic for θ.

   Solution. The pdf of Cauchy$(\theta, 1)$ is

   $$f(x|\theta) = \frac{1}{\pi(1+(x-\theta)^2)}.$$

   So the ratio of the joint pdfs of **X** and **Y** is

   $$\frac{f_{\mathbf{X}}(\mathbf{x}|\theta)}{f_{\mathbf{Y}}(\mathbf{y}|\theta)} = \prod_{i=1}^n \frac{(1+(y_i-\theta)^2)}{(1+(x_i-\theta)^2)}$$

   which is clearly constant in θ if $(x_{(1)}, x_{(2)}, \ldots, x_{(n)}) = (y_{(1)}, y_{(2)}, \ldots, y_{(n)})$, which is to say if **x** and **y** are the same values, up to a permutation. And if $T(\mathbf{x}) \neq T(\mathbf{y})$, then the ratio forms a quotient of two polynomials whose roots are a function of θ. So the statistic $T(X) = (X_{(1)}, X_{(2)}, \ldots, X_{(n)})$ is minimally sufficient for θ.

6. Let $X_1, \ldots, X_n$ be a random sample from $U(-\theta, \theta)$. Find a minimal sufficient statistic for θ.

   Solution. The probability density function of $U{-\theta}, \theta$ is

   $$f(x|\theta) = \begin{cases} (2\theta)^{-1} & x \in (-\theta, \theta) \\ 0 & \text{else,} \end{cases}$$
   $$= (2\theta)^{-1} 1_{\{x \in (-\theta,\theta)\}}$$

   so the joint pdf of **X** is

   $$f_{\mathbf{X}}(\mathbf{x}|\theta) = \prod_{i=1}^n (2\theta)^{-1} 1_{\{x_i \in (-\theta,\theta)\}}$$
   $$= (2\theta)^{-n} 1_{\{\min(x_i) \in (-\theta,\theta)\}} 1_{\{\max(x_i) \in (-\theta,\theta)\}}$$

   Consider the two-dimensional statistic $T(\mathbf{X}) = (\min(\mathbf{X}), \max(\mathbf{X}))$. The ratio

   $$\frac{f_{\mathbf{X}}(\mathbf{x}|\theta)}{f_{\mathbf{Y}}(\mathbf{y}|\theta)} = \frac{1_{\{\min(x_i) \in (-\theta,\theta)\}} 1_{\{\max(x_i) \in (-\theta,\theta)\}}}{1_{\{\min(y_i) \in (-\theta,\theta)\}} 1_{\{\max(y_i) \in (-\theta,\theta)\}}}$$

   is clearly constant (equal to 1, where it is defined) in θ when $T(\mathbf{x}) = T(\mathbf{y})$. When $T(\mathbf{x}) \neq T(\mathbf{y})$, the function will achieve values of 0 and 1. So T is sufficient for θ.