Statistical Theory: Exercise Sheet 3 — Corrections

October 16, 2018; 8:15-10:00 in MA 12.

Tomáš Rubín, tomas.rubin@epfl.ch

Remark. When we say that $X_1, \ldots, X_n$ is a sample from the distribution F it means that $X_1, \ldots, X_n$ are iid with the distribution F.

Exercise 1. Let $X_1, \ldots, X_n$ be a sample from the uniform distribution on $(0, \theta)$ $(\theta > 0)$.

(a) Show that $S = X_{(n)}$ is a complete sufficient statistic for θ.

(b) Show that $T = X_{(1)}/X_{(n)}$ is ancillary for θ.

(c) Use Basu's theorem to show that S and T are independent.

Solution to Exercise 1. The sufficiency of S follows from the factorization lemma (see the lecture slides). The density of S is $f_S(s) = ns^{n-1}/\theta^n 1_{(0,\theta)}(s)$.

Completeness of S: Suppose that $h(s)$ is a function such that $\mathsf{E}_\theta\, h(S) = \int_0^\theta h(s) ns^{n-1}/\theta^n ds$ is equal to 0 for all $\theta > 0$. This implies that $\int_0^\theta h(s)s^{n-1}ds = 0$, thus its derivative (w.r.t. θ) $h(\theta)\theta^{n-1}$ equals zero, which is possible only if $h(\theta) = 0$ for all $\theta > 0$. Hence, S is complete.

Ancillarity of T: For any θ, X_i has the same distribution as θU_i where U_i are iid $\mathrm{U}(0, 1)$. Thus $T = X_{(1)}/X_{(n)} \stackrel{d}{=} (\theta U_{(1)})/(\theta U_{(n)}) = U_{(1)}/U_{(n)}$. Therefore, the distribution of T does not depend on θ and T is ancillary. The application of Basu's theorem is straightforward.

Exercise 2. Let $X_1, \ldots, X_n$ be a sample from the uniform distribution on $(-\theta, \theta)$ $(\theta > 0)$.

(a) Show that $S = (X_{(1)}, X_{(n)})^\top$ is a sufficient statistic for θ.

(b) Show that S is not a complete statistic.

(c) Suggest a different statistics that is complete.

Solution to Exercise 2. S is sufficient: As the joint density of $X_1, \ldots, X_n$ is $f(x_1, \ldots, x_n; \theta) = \frac{1}{2\theta} 1\{-\theta < x_{(1)} \leq x_{(n)} < \theta\}$, the sufficiency of S follows by the Neyman–Fisher factorization lemma.

S is not complete: Consider the function $h(u, v) = u + v$. It is easily computed that $\mathsf{E}_\theta\, h(X_{(1)}, X_{(n)}) = \mathsf{E}_\theta(X_{(1)} + X_{(n)}) = 0$ for all θ. However, $h(X_{(1)}, X_{(n)}) = X_{(1)} + X_{(n)}$ is not identically equal to zero (almost surely). Therefore, S is not complete.

The join density of $X_1, \dots, X_n$ can be rewritten as $f(x_1, \dots, x_n; \theta) = \frac{1}{(2\theta)^n} \mathbb{1}_{[-\theta < x_{(1)} \le x_{(n)} < \theta]} = \frac{1}{(2\theta)^n} \mathbb{1}_{[-X_{(1)} < \theta]} \mathbb{1}_{[X_{(n)} < \theta]} = \frac{1}{(2\theta)^n} \mathbb{1}_{[\max\{-X_{(1)}, X_{(n)}\} < \theta]}$. Thus it makes sense to consider $T = \max(-X_{(1)}, X_{(n)})$ which is clearly a sufficient statistics.

Now derive the density of T and get $f_T(t) = n\frac{t^{n-1}}{\theta^n}$. The proof of the completeness is identical to Exercise 1, (a).

Exercise 3. Let $X_1, \dots, X_n$ be a sample from the distribution with density $f(x; \mu) = e^{-(x-\mu)} 1_{(\mu,\infty)}(x)$ with a parameter $\mu \in \mathbb{R}$.

(a) Show that $X_{(1)}$ is a complete sufficient statistic for μ.

(b) Use Basu's theorem to show that $X_{(1)}$ and $\frac{1}{n}\sum_{i=1}^n (X_i - \bar{X})^2$ are independent.

Solution to Exercise 3. Sufficiency of $X_{(1)}$: By the factorization lemma. Completeness of $X_{(1)}$: Notice that $(X_1, \dots, X_n)^\mathsf{T} \stackrel{d}{=} \mu + (E_1, \dots, E_n)^\mathsf{T}$ where E_i are iid Exp(1). We know (immediately, or from Exercise 3 in Set 1) that $E_{(1)} \sim \text{Exp}(n)$, hence $X_{(1)} \stackrel{d}{=} \mu + E_{(1)}$. Therefore, $\mathsf{E}_\mu\, h(X_{(1)}) = \int_\mu^\infty h(u) n \exp\{-n(u-\mu)\} du$, which is zero for all μ if an only if $\int_\mu^\infty h(u) \exp\{-nu\} du = 0$ for all μ. Differentiate with respect to μ to see that this happens only if h is zero.

Ancillarity of $\frac{1}{n}\sum_{i=1}^n (X_i - \bar{X})^2$: The sample variance does not depend on the location shift μ, thus its distribution does not depend on μ. Therefore, it is ancillary. The independence of $X_{(1)}$ and $\frac{1}{n}\sum_{i=1}^n (X_i - \bar{X})^2$ follows directly from Basu's theorem.

Exercise 4. Let $Y_1, \dots, Y_n$ follow the normal linear regression model, that is, they are independent with distribution $\mathrm{N}(\beta_0 + \beta_1 x_i, \sigma^2)$, where $\sigma > 0$ is known and x_i are some fixed known constants.

Find a minimal sufficient statistic for the parameter $(\beta_0, \beta_1)^\mathsf{T}$ (find a sufficient statistic using the factorization lemma and then prove its minimality).

Solution to Exercise 4. The joint density of $Y_1, \dots, Y_n$ is

$$
\begin{aligned}
f(y_1, \dots, y_n; \beta_0, \beta_1) &= \frac{1}{(\sqrt{2\pi}\sigma)^n} \exp\left\{-\frac{\sum_{i=1}^n (y_i - \beta_0 - \beta_1 x_i)^2}{2\sigma^2}\right\} \\
&= \frac{1}{(\sqrt{2\pi}\sigma)^n} \exp\left\{-\frac{\sum_{i=1}^n y_i^2}{2\sigma^2}\right\} \exp\left\{\frac{2\beta_0 \sum_{i=1}^n y_i + 2\beta_1 \sum_{i=1}^n y_i x_i - \sum_{i=1}^n (\beta_0 + \beta_1 x_i)^2}{2\sigma^2}\right\}.
\end{aligned}
$$

Therefore, by the factorization lemma, the statistic $T(Y) = (\sum_{i=1}^n Y_i, \sum_{i=1}^n Y_i x_i)^\mathsf{T}$ is sufficient for $(\beta_0, \beta_1)^\mathsf{T}$. By observing that $f(y_1, \dots, y_n; \beta_0, \beta_1)/f(z_1, \dots, z_n; \beta_0, \beta_1)$ does not depend on $(\beta_0, \beta_1)^\mathsf{T}$ if and only if $T(y) = T(z)$, we find that T is minimal sufficient.

Exercise 5. Let $X_1, \dots, X_n$ be a sample from the exponential distribution with intensity $\lambda > 0$. Show that $T = \sum_{i=1}^n X_i$ is a complete and sufficient statistic for λ. Show that $R = X_1/(\sum_{i=1}^n X_i)$ is ancillary for λ. Conclude (using Basu's theorem) that T and R are independent. (*Useful fact:* If $X_1 \sim \Gamma(a, p_1)$, $X_2 \sim \Gamma(a, p_2)$ are independent, then $X_1 + X_2 \sim \Gamma(a, p_1 + p_2)$.)

Solution to Exercise 5. Sufficiency of T: Immediately by the factorization criterion. Completeness of S: We know that $T \sim \Gamma(\lambda, n)$. $\mathsf{E}_\lambda\, h(T) = \lambda^n/\Gamma(n) \int_0^\infty h(t) e^{-\lambda t} t^{n-1} dt$ is zero for all $\lambda > 0$ if and only if the integral is zero for all $\lambda > 0$. The integral, as a function of λ, is the Laplace transform of the function $h(t)t^{n-1}$. If the Laplace transform is zero, the function $h(t)t^{n-1}$ is zero (almost everywhere), thus $h(t) = 0$ (a.e.) which proves that T is complete. Ancillarity of R: Notice that $(X_1, \ldots, X_n)^\mathsf{T} \stackrel{d}{=} (E_1, \ldots, E_n)^\mathsf{T}/\lambda$ where E_i are iid Exp(1), and proceed similarly as in Exercise 1.

Exercise 6. Suppose that $X_1, \ldots, X_n$ are independent random variables, each uniformly distributed on $(0, \theta)$, where $\theta > 0$ is a parameter. For each of the variables $T_1, \ldots, T_6$ below, decide whether it is a sufficient statistic or not.

$$\begin{aligned}
T_1 &= (X_1, \ldots, X_n), \\
T_2 &= \big(X_{(1)}, \ldots, X_{(n)}\big), \\
T_3 &= \big(X_{(1)}, X_{(n)}\big), \\
T_4 &= X_{(1)} \\
T_5 &= X_{(n)} \\
T_6 &= \frac{n+1}{n} \times X_{(n)}
\end{aligned}$$

Is any of them minimal sufficient? If so, is it also complete? (Notice the dimensions.)

Solution to Exercise 6. The joint density of $X_1, \ldots, X_n$ is

$$f_{X_1,\ldots,X_n}(x_1, \ldots, x_n) = \frac{1}{\theta^n} \mathcal{I}\{x_{(n)} < \theta\} \times \mathcal{I}\{x_{(1)} > 0\},$$

where $\mathcal{I}$ is the indicator function. By the factorisation criterion it follows that all the statistics except for T_4 are sufficient.

All the sufficient statistics given in the exercise are functions of T_5 and T_6, but T_5 is is a function of neither of T_1, T_2 and T_3. It follows that T_1, T_2 and T_3 are not minimal. On the other hand, T_5 and T_6 are minimal sufficient. This follows from the theorem on Slide 13 of the lecture notes. Statistics T_5 and T_6 are also complete. We show this for T_5, the result for T_6 then follows. The density of $T_5 = \max(X_1, \ldots, X_n)$ is

$$f_{T_5}(x) = \frac{n}{\theta^n} \times x^{n-1}\, \mathcal{I}\{x < \theta\}.$$

Suppose that $h(x)$ is a function such that

$$\mathsf{E}_\theta\, h(T_4) = \frac{n}{\theta^n} \int_0^\theta h(x)\, x^{n-1} dx = 0 \quad \text{for every } \theta > 0.$$

Then

$$\int_{\theta-\varepsilon}^{\theta+\varepsilon} h(x)x^{n-1}dx = \int_0^{\theta+\varepsilon} h(x)x^{n-1}dx - \int_0^{\theta-\varepsilon} h(x)x^{n-1}dx = 0 - 0 = 0$$

for every $\theta > 0$ and $0 < \varepsilon < \theta$. It follows that

$$\lim_{\varepsilon \to 0} \frac{1}{2\varepsilon} \int_{\theta-\varepsilon}^{\theta+\varepsilon} h(x)x^{n-1}dx = 0,$$

and the Lebesgue differentiation theorem gives that $h(\theta) = 0$ for a.e. $\theta > 0$.

Exercise 7. Suppose that $X_1, \ldots, X_n$ are independent random variables, each with normal distribution $\mathrm{N}(\mu, \sigma^2)$, where σ^2 is a known constant and μ is an unknown parameter. Decide whether any of the variables T_1, ..., T_4 below is an ancillary statistic.

$$
\begin{aligned}
T_1 &= (X_1, \ldots, X_n), \\
T_2 &= (X_1 - X_2, X_2 - X_3, \ldots, X_{n-1} - X_n), \\
T_3 &= \frac{1/n \sum_{i=1}^{n} X_i - \mu}{\sigma/\sqrt{n}}. \\
T_4 &= \frac{1/n \sum_{i=1}^{n-1} (X_i - X_{i+1})^2}{\mathrm{var}\left(1/n \sum_{i=1}^{n-1} (X_i - X_{i+1})^2\right)}.
\end{aligned}
$$

Solution to Exercise 7. T_2 and T_3 are ancillary, T_1 is not (the distribution depends on the parameter), ~~T_4 is not a statistic (it is a function of the parameter)~~. The variance of the term in the denominator in T_4 is in fact independent of the unknown parameter μ, therefore it is a statistic. The distribution of T_4 is independent of μ thus T_4 is ancillary.