Statistics: Tutorial Week 5 - Solutions to Practice Exercises

Practice Exercises

Exercise 1. We are given the statistical model $\{\text{Bernoulli}(p) \mid p \in [0,1]\}$, that is

$$g(x \mid p) = p^x(1-p)^{1-x}.$$

Both the moment and maximum likelihood estimator are given by $\overline{X}$.

a. Show that the Bernoulli statistical model is an exponential family.

b. Show that $\overline{X}$ is the UMVU estimator for p_0.

SOLUTION.

a. We have

$$\begin{aligned} g(x \mid p) = p^x(1-p)^{1-x} &= (1-p)\left(\frac{p}{1-p}\right)^x \\ &= (1-p)e^{\log\left(\frac{p}{1-p}\right)^x} = (1-p)e^{x(\log(p)-\log(1-p))}. \end{aligned}$$

We therefore have an exponential family with $h(x) = 1, c(p) = 1-p, t_1(x) = x$ and $w_1(p) = \log(p) - \log(1-p)$.

b. We calculate the Cramér-Rao lower bound and use the fact that the Bernoulli distribution belongs to the exponential family.

$$\begin{aligned} \log g(x \mid p) &= \log\left(p^x(1-p)^{1-x}\right) = x\log p + (1-x)\log(1-p) \\ \frac{d}{dp}\log g(x \mid p) &= \frac{x}{p} - \frac{1-x}{1-p} \\ \frac{d^2}{dp^2}\log g(x \mid p) &= -\frac{x}{p^2} - \frac{1-x}{(1-p)^2} \\ i_p &= -\mathbb{E}_p\left(\frac{d^2}{dp^2}\log g(X_1 \mid p)\right) = \mathbb{E}_p\left(\frac{X_1}{p^2} + \frac{1-X_1}{(1-p)^2}\right) \\ &= \frac{p}{p^2} + \frac{1-p}{(1-p)^2} = \frac{1}{p(1-p)}. \end{aligned}$$

We obtain for the Cramér-Rao lower bound

$$\frac{\tau'(p)^2}{ni_p} = \frac{1}{n\frac{1}{p(1-p)}} = \frac{p(1-p)}{n} = \mathbb{V}\text{ar}_p(\overline{X}).$$

Exercise 2. In this exercise we study an iid random sample $X_1, \ldots, X_n$ from a population in the statistical model $\{g(x \mid \theta) \mid \theta \in \Theta\}$.

a. Prove that the set of order statistics $T = (X_{(1)}, \ldots, X_{(n)})$ is sufficient for θ_0.

b. We are interested in finding an unbiased estimator for $\tau(\theta) = \mathbb{P}_\theta(X_1 \leq 2)$. Show that $W(\boldsymbol{X}) = \mathbb{1}_{\{X_1 \leq 2\}}$ is unbiased.

c. Use Rao-Blackwellisation to find a better unbiased estimator for $\tau(\theta)$.

d. Show that the obtained estimator converges to $\tau(\theta_0)$ as $n \to \infty$.

Solution.

a. To show this result we go back to the original definition of sufficiency and check that the distribution of $\boldsymbol{X} \mid T$ does not depend on θ_0. This follows almost immediately. Note that each observation must be equal to exactly one of the order statistics. We therefore get a discrete distribution for $\boldsymbol{X} \mid T$ with equal probability for all $n!$ possible orderings:

$$P(X_1 = X_{(1)}, \ldots, X_n = X_{(n)} \mid T) = \ldots = P(X_1 = X_{(n)}, \ldots, X_n = X_{(1)} \mid T) = \frac{1}{n!}.$$

This distribution clearly does not depend on θ_0, and thus we conclude that T is sufficient. Intuitively this result is also clear, as our observations are independent, the order in which we observe them contains no information about θ_0.

b. This follows immediately from

$$\mathbb{E}_\theta W(\boldsymbol{X}) = \mathbb{E}_\theta \mathbb{1}_{\{X_1 \leq 2\}} = 0 + 1 \times P_\theta(\mathbb{1}_{\{X_1 \leq 2\}} = 1) = P_\theta(X_1 \leq 2) = \tau(\theta).$$

c. We have an unbiased estimator $W(\boldsymbol{X})$ and a sufficient statistic $T(\vec{X})$. It remains to derive

$$\phi(T) = \mathbb{E}(W(\boldsymbol{X}) \mid T) = \mathbb{E}(\mathbb{1}_{\{X_1 \leq 2\}} \mid T).$$

The distribution of $X_1 \mid T$ is uniformly discrete on the n random variables $X_{(1)}, \ldots, X_{(n)}$, which is the same as the discrete uniform distribution on $X_1, \ldots, X_n$. It follows that $\mathbb{E}(X_1 \mid T) = \overline{X}$ and in a similar fashion we obtain

$$\phi(T) = \mathbb{E}(\mathbb{1}_{\{X_1 \leq 2\}} \mid T) = \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}_{\{X_i \leq 2\}} = \frac{\#\{1 \leq i \leq n : X_i \leq 2\}}{n}.$$

d. This follows directly from the law of large numbers:

$$\phi(T) = \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}_{\{X_i \leq 2\}} \overset{n \to \infty}{\to} \mathbb{E}(\mathbb{1}_{\{X_1 \leq 2\}}) = \mathbb{P}(X_1 \leq 2) = \tau(\theta_0).$$