

STA 522 Sample Exam 1 Solutions

Problem 1

Part (a): Since X_1, X_2, \dots, X_n are iid, the cdf of each X_i is given by:

$$F(x) = \begin{cases} 0 & \text{if } x \leq -1 \\ \frac{x+1}{2} & \text{if } -1 < x < 1 \\ 1 & \text{if } x \geq 1 \end{cases}$$

Therefore,

$$\begin{aligned} P(X_{(1)} > 0.25 \text{ and } X_{(n)} \leq 0.8) &= P(X_i > 0.25 \text{ for all } i \text{ and } X_i \leq 0.8 \text{ for all } i) \\ &= P(0.25 < X_i \leq 0.8 \text{ for all } i) \\ &= \{P(0.25 < X_1 \leq 0.8)\}^n & (\text{iid}) \\ &= \{F(0.8) - F(0.25)\}^n \\ &= \left\{ \frac{0.8+1}{2} - \frac{0.25+1}{2} \right\}^n = (0.55/2)^n = \boxed{(0.275)^n}. \end{aligned}$$

Part (b): Yes, it does. Fix $\varepsilon > 0$. We have

$$\begin{aligned} P(|X_{(n)} - 1| \geq \varepsilon) &= P(X_{(n)} - 1 \geq \varepsilon) + P(X_{(n)} - 1 < -\varepsilon) \\ &= P(X_{(n)} \geq 1 + \varepsilon) + P(X_{(n)} < 1 - \varepsilon) \\ &= 0 + P(X_i < 1 - \varepsilon, \text{ all } i) \\ &= \{P(X_1 < 1 - \varepsilon)\}^n & (\text{iid}) \\ &= \begin{cases} \left(\frac{1-\varepsilon+1}{2}\right)^n = \left(1 - \frac{\varepsilon}{2}\right)^n & \text{if } \varepsilon < 2 \\ 0 & \text{if } \varepsilon \geq 2 \end{cases} \\ &\rightarrow 0 \quad \text{as } n \rightarrow \infty \end{aligned}$$

which means $X_{(n)} \xrightarrow{P} 1$.

Problem 2

This is from Lecture 2: see lecture notes. To verify that $F(v) = e^{-1/v} I(v > 0)$ is a cdf, observe:

- (i) $F(-\infty) = \lim_{v \rightarrow -\infty} F(v) = 0$ and $F(+\infty) = \lim_{v \rightarrow \infty} F(v) = 1$.
- (ii) F is continuous (actually differentiable everywhere except $v = 0$), and hence right continuous.
- (iii) $\frac{d}{dv} e^{-1/v} = e^{-1/v} \left(-\frac{1}{v^2}\right) (-1) = e^{-1/v} \frac{1}{v^2} > 0$ for all v . Hence, $F(v)$ is increasing.

These three collectively imply that F is a cdf.

Problem 3

Here $X_i \sim \text{iid Bernoulli}(\theta)$ for $i = 1, \dots, n = 10$.

Part (a): The likelihood of θ is

$$L(\theta | \underline{x}) = P(\underline{X} = \underline{x} | \theta) = \theta^{\sum_{i=1}^{10} x_i} (1 - \theta)^{10 - \sum_{i=1}^{10} x_i}$$

Part (b): Given that $\sum_{i=1}^{10} X_i = 6$. The likelihood is:

$$L(\theta | \sum_{i=1}^{10} X_i = 6) = \theta^6 (1 - \theta)^4$$

Therefore

$$L(\theta = 0.2 | \sum_{i=1}^{10} X_i = 6) = (0.2)^6 (0.8)^4 = \boxed{2.61 \times 10^{-5}}$$

and

$$L(\theta = 0.8 | \sum_{i=1}^{10} X_i = 6) = (0.8)^6 (0.2)^4 = \boxed{4.19 \times 10^{-4}}.$$

This shows that $\theta = 0.8$ has a higher likelihood. Intuitively, the observed data of 6 successes are more compatible with the configuration where population probability of success $\theta = 0.8$ than with $\theta = 0.2$.

Problem 4

The joint density of \underline{X} is:

$$f(\underline{x} | \theta, \gamma) = \prod_{i=1}^n \left(\frac{\gamma}{\theta} x_i^{\gamma-1} e^{-x_i/\theta} \right) = \left(\frac{\gamma}{\theta} \right)^n \prod_{i=1}^n x_i^{\gamma-1} \exp \left(-\frac{1}{\theta} \sum_{i=1}^n x_i \right)$$

Part (a): If $\gamma > 0$ is known, then

$$f(\underline{x} | \theta) = \underbrace{\left(\frac{\gamma}{\theta} \right)^n \exp \left(-\frac{1}{\theta} \sum_{i=1}^n x_i \right)}_{=g(T(\underline{x})|\theta)} \underbrace{\prod_{i=1}^n x_i^{\gamma-1}}_{=h(\underline{x})}$$

Therefore, by the Factorization theorem, $T(\underline{X}) = \sum_{i=1}^n X_i$ is sufficient for θ .

Part (b) It follows from the joint density above that a factorization based on a univariate or bivariate or a lower (than n) dimensional) statistic is not feasible when both θ and γ are unknown. Therefore, by the necessity half of the factorization theorem (must have factorization for a lower dimensional sufficient statistic to exist) it follows that no univariate sufficient statistic exist in this case. A non-trivial sufficient statistic would be the order statistics: $(X_{(1)}, \dots, X_{(n)})$.

Problem 5

Part (a): Because X_1, \dots, X_n are iid from the scale family $\frac{1}{\sigma} f(x/\sigma)$, we can construct iid observations Z_1, \dots, Z_n from the density $f(x)$ (the standard density of the family which is free of σ) such that $Z_i = X_i/\sigma$, i.e., $X_i = \sigma Z_i$.

Note that the sample median is:

$$\begin{aligned}
M(X_1, X_2, \dots, X_n) &= \begin{cases} X_{(\frac{n+1}{2})} & n \text{ is odd} \\ \frac{X_{(\frac{n}{2})} + X_{(\frac{n}{2}+1)}}{2} & n \text{ is even} \end{cases} \\
&= \begin{cases} \sigma Z_{(\frac{n+1}{2})} & n \text{ is odd} \\ \sigma \frac{Z_{(\frac{n}{2})} + Z_{(\frac{n}{2}+1)}}{2} & n \text{ is even} \end{cases} \\
&= \sigma M(Z_1, \dots, Z_n)
\end{aligned}$$

Again,

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i = \frac{1}{n} \sum_{i=1}^n (\sigma Z_i) = \sigma \bar{Z}$$

Hence

$$\log M - \log \bar{X} = \log \left(\frac{M(X_1, X_2, \dots, X_n)}{\bar{X}} \right) = \log \left(\frac{\sigma M(Z_1, \dots, Z_n)}{\sigma \bar{Z}} \right) = \log \left(\frac{M(Z_1, \dots, Z_n)}{\bar{Z}} \right)$$

where the RHS contains random variables whose distribution does not depend on the parameter σ . Hence $\log M - \log \bar{X}$ is an ancillary statistic.

Part(b): [NOTE: Recall that for Uniform(0, θ), $T = X_{(n)}$ is sufficient, and not $\sum_{i=1}^n X_i$. So we can't use Basu's theorem directly. However, for this specific type of problem, there is a much quicker way, shown as follows.]

As suggested in the hint, the random variables $\frac{X_i}{X_1 + \dots + X_n}$ all have the same distribution and hence mean as X_i 's are iid.

Therefore, for some constant k , $E \left[\frac{X_i}{X_1 + \dots + X_n} \right] = k$ for all $i = 1, \dots, n$.

So,

$$\begin{aligned}
E \left[\frac{X_1}{X_1 + \dots + X_n} \right] + \dots + E \left[\frac{X_n}{X_1 + \dots + X_n} \right] &= \underbrace{k + \dots + k}_{n \text{ many}} = nk \\
\Rightarrow E \left[\underbrace{\frac{X_1}{X_1 + \dots + X_n} + \dots + \frac{X_n}{X_1 + \dots + X_n}}_{=E \left[\frac{X_1 + \dots + X_n}{X_1 + \dots + X_n} \right] = E(1) = 1} \right] &= nk \\
\text{i.e., } nk = 1 &\Rightarrow k = \frac{1}{n}
\end{aligned}$$

Therefore, $E \left[\frac{X_n}{X_1 + \dots + X_n} \right] = k = \frac{1}{n}$.