EE601: Statistical Signal Analysis
Quiz # 7
Date: 10/11/2024 Time: 10:30 - 1

Q.1 Let $X_1, \ldots, X_n \sim$ iid Bernoulli(p) for $p \in (0,1)$.
Let $\psi(p) = p(1-p)$, and $T(\vec{x}) = \sum_{k=1}^{n} x_k$.

(a) Find an unbiased estimator δ for $\psi(p)$.

(b) Find $E[\delta(\vec{x}) \mid T(\vec{x})]$.

Q.2 Find minimal sufficient statistic for iid samples $X_1, \ldots, X_n$ with marginal given by

(a) $f_\alpha(x) = \frac{1}{B(\alpha)} x^{\alpha-1}(1-x)$ for $x \in (0,1)$ | Parameter $\alpha > 0$
$= 0$ o.w. and

$B(\alpha) = \frac{2\Gamma(\alpha)}{\Gamma(2+\alpha)}$. $\Gamma(\cdot)$ is std. γ. function.

(b) $f_\lambda(\theta) = \frac{\lambda e^{-\lambda\theta}}{1 - e^{-2\pi\lambda}}$, for $\theta \in [0, 2\pi]$

$= 0$ o.w.

Parameter $\lambda > 0$.

Note: show sufficiency and minimality both.

Solutions

$\boxed{Q.1}$ Define, $\delta(\vec{x}) = x_1(1-x_2)$

Note that $E_p[\delta(\vec{x})] = E_p[x_1(1-x_2)]$

$$= E_p[x_1] \cdot E_p[(1-x_2)] \quad \because x_1 \perp\!\!\!\perp x_2$$

$$= p(1-p). \quad \forall p \in (0,1)$$

Thus $\delta(\vec{x})$ is an unbiased estimator for $\psi(p) = p(1-p)$.

Note: This $\delta(\cdot)$ is not a unique solution. Other estimators that are shown to be unbiased, must be accepted.

Now consider,

$E[\delta(\vec{x}) \mid T(\vec{x})]$.

To obtain this we solve for $t \in \{0, \dots, n\}$,

$$E[\delta(\vec{x}) \mid T(\vec{x}) = t]$$

$$= E[x_1(1-x_2) \mid T(\vec{x}) = t]$$

$$= P(x_1(1-x_2) = 1 \mid T(\vec{x}) = t) \quad \because x_1(1-x_2) \in \{0,1\}$$

$$= P(x_1 = 1, x_2 = 0 \mid T(\vec{x}) = t)$$

$$= \frac{P(x_1 = 1, x_2 = 0, T(\vec{x}) = t)}{P(T(\vec{x}) = t)}$$

Note that

$$P(T(\vec{x}) = t) = \binom{n}{t} p^t (1-p)^{n-t}$$

Now, numerator

$$P(X_1=1, X_2=0, T(\vec{x})=t)$$

$$= 0 \quad \text{if} \quad t=0 \text{ or } n$$

$$= P(X_1=1, X_2=0, \sum_{k=3}^{n} X_k = t-1)$$

$$= P(X_1=1) \cdot P(X_2=0) \cdot P(\sum_{k=3}^{n} X_k = t-1)$$

as X_k's are independent

$$= p(1-p)\binom{n-2}{t-1} p^{t-1}(1-p)^{n-t-1}$$

$$= \binom{n-2}{t-1} p^t (1-p)^{n-t}.$$

Thus,

$$E[\delta(\vec{x}) \mid T(\vec{x})=t] = \frac{\binom{n-2}{t-1} \cancel{p^t (1-p)^{n-t}}}{\binom{n}{t} \cancel{p^t (1-p)^{n-t}}}$$

$$= \frac{t(n-t)}{n(n-1)} \quad \forall t \notin \{0, n\}$$

But note that

$$\frac{t(n-t)}{n(n-1)} = 0 \quad \text{if } t=0 \text{ or } n.$$

Thus,

$$E[\delta(\vec{x}) \mid T(\vec{x}) = t] = \frac{t(n-t)}{n(n-1)} \quad \forall t \in \{0, \ldots, n\}.$$

This implies

$$E[\delta(\vec{x}) \mid T(\vec{x})] = \frac{T(\vec{x})(n - T(\vec{x}))}{n(n-1)}.$$

Going further,

$$\frac{T(\vec{x})(n-T(\vec{x}))}{n(n-1)} = \frac{\Sigma X_k (n - \Sigma X_k)}{n(n-1)}$$

$$= \frac{n \Sigma X_k - (\Sigma X_k)^2}{n(n-1)}$$

$$= \frac{n \sum x_k^2 - 2\left(\sum x_k\right)^2 + \left(\sum x_k\right)^2}{n(n-1)}$$

$$= \frac{\sum_{k=1}^{n} x_k^2 - \frac{2}{n} \sum_{k=1}^{n} x_k \sum_{i=1}^{n} x_i + \frac{1}{n^2} \sum_{k=1}^{n} \left(\sum x_i\right)^2}{(n-1)}$$

$$= \frac{\sum_{k=1}^{n} \left(x_k^2 - \frac{2}{n} x_k \sum_{i=1}^{n} x_i + \left(\frac{1}{n} \sum x_i\right)^2 \right)}{(n-1)}$$

$$= \frac{\sum_{k=1}^{n} \left(x_k - \frac{1}{n} \sum x_i \right)^2}{(n-1)}$$

$$= \frac{1}{(n-1)} \sum_{k=1}^{n} \left(x_k - \bar{x} \right)^2.$$

In statistics / data science this is a most used estimator for variance. And $p(1-p)$ is the variance of Bernoulli(p) r.v.

$\boxed{Q.2}$ $f_\alpha(x) = \frac{1}{B(\alpha)} x^{1-\alpha}(1-x) \quad x \in (0,1).$

Joint density,

$$f_\alpha(\vec{x}) = \frac{1}{[B(\alpha)]^n} \left(\prod_{i=1}^{n} x_i \right)^{1-\alpha} \prod_{i=1}^{n} (1-x_i)$$

Define, $h(\vec{x}) = \prod_{i=1}^{n} (1-x_i)$, and

$T(\vec{x}) = \prod_{i=1}^{n} x_i$, and

$$g_\alpha(T(\vec{x})) = \frac{1}{[B(\alpha)]^n} \cdot \left(T(\vec{x}) \right)^{1-\alpha}$$

Thus, $T(\vec{x}) = \prod_{i=1}^{n} x_i$ is sufficient statistic by factorization criterion.

Now, note that

$$\frac{f_\alpha(\vec{x})}{f_\alpha(\vec{y})} = \left[\frac{\prod x_i}{\prod y_i}\right]^{1-\alpha} \cdot \prod_{i=1}^{m} \frac{(1-x_i)}{(1-y_i)}$$

Thus, $\dfrac{f_\alpha(\vec{x})}{f_\alpha(\vec{y})}$ is independent of α, $\forall \alpha$ iff $\prod x_i = \prod y_i$.

Also, note that $T(\vec{x}) = T(\vec{y})$ iff $\prod x_i = \prod y_i$.

Thus, $\mathcal{D}(\vec{x}) = \mathcal{D}_T(\vec{x}) \Rightarrow T(\vec{x}) = \prod x_i$ is minimal sufficient statistic.

(b) $$f_\lambda(\theta) = \frac{\lambda e^{-\lambda\theta}}{1-e^{-2\pi\lambda}}$$

Joint density

$$f_\lambda(\bar{\theta}) = \underbrace{\left[\frac{\lambda}{1-e^{-2\pi\lambda}}\right]^n e^{-\lambda\Sigma\theta_i}}_{g_\lambda(\Sigma\theta_i)} \cdot \underbrace{1}_{h(\bar{\theta})}$$

$T(\bar{\theta})$ - sufficient statistic by factorization criterion.

Also note that

$\dfrac{f_\lambda(\bar{\theta})}{f_\lambda(\bar{\theta}')}$ iff $\Sigma\theta_i = \Sigma\theta_i'$.

Moreover, $T(\bar{\theta}) = T(\bar{\theta}')$ iff

$$\sum \theta_i = \sum \theta_i'.$$

This proves minimality.