STAT6201: Theoretical Statistics I

Satvik Saha

Homework 1

1. We have $X_1, \dots, X_n \overset{iid}{\sim} \text{Bernoulli}(p)$, and $T = \sum_{i=1}^n iX_i$.

   (a) Let $n = 2$, so $T = X_1 + 2X_2$. Note that

   $$\begin{aligned} T(x_1, x_2) = 0 &\iff x_1 = 0, x_2 = 0, \\ T(x_1, x_2) = 1 &\iff x_1 = 1, x_2 = 0, \\ T(x_1, x_2) = 2 &\iff x_1 = 0, x_2 = 1, \\ T(x_1, x_2) = 3 &\iff x_1 = 1, x_2 = 1. \end{aligned}$$

   This immediately follows from the fact that T is the binary number corresponding to the digits (x_2, x_1). Thus, the map $(X_1, X_2) \mapsto T(X_1, X_2)$ is a bijection; since (X_1, X_2) is (trivially) sufficient for p, so is T.

   (b) Let $n = 3$, so $T = X_1 + 2X_2 + 3X_3$. Now, $T(x_1, x_2, x_3) = 3$ if and only if $(x_1, x_2, x_3) \in \{(1, 1, 0), (0, 0, 1)\}$. Thus,

   $$P(X_1 = x_1, X_2 = x_2, X_3 = x_3 \mid T = 3) = \begin{cases} p, & \text{if } x_1 = 1, x_2 = 1, x_3 = 0, \\ 1 - p, & \text{if } x_1 = 0, x_2 = 0, x_3 = 1, \\ 0, & \text{otherwise.} \end{cases}$$

   This is clearly not free of p, hence T is not sufficient for p.

   Remark: The statistic $S = \sum_{i=1}^n 2^i X_i$ is sufficient for p regardless of n, being a bijection of $(X_1, \dots, X_n)$.

2. (a) Let $\mathcal{P} = \{f(\boldsymbol{x}; \theta) \colon \theta \in \Theta\}$ be a parametric family of pmf's/pdf's on a sample space $\mathcal{X}$. Furthermore, let T satisfy for all $\boldsymbol{x}, \boldsymbol{y} \in \mathcal{X}$,

   $$T(\boldsymbol{x}) = T(\boldsymbol{y}) \iff \frac{f(\boldsymbol{x}; \theta)}{f(\boldsymbol{y}; \theta)} \text{ is free of } \theta, \text{ for all } \theta \in \Theta.$$

   Note that the distribution of T can be described via

   $$g_T(t; \theta) = \int_{\{\boldsymbol{x} \in \mathcal{X} \colon T(\boldsymbol{x}) = t\}} f(\boldsymbol{x}; \theta)\, d\boldsymbol{x}.$$

   Thus,

   $$f(\boldsymbol{y}; \theta \mid T = t) = \frac{f(\boldsymbol{y}; \theta)}{g_T(t; \theta)} \mathbf{1}(T(\boldsymbol{y}) = t) = \left(\int_{\{\boldsymbol{x} \in \mathcal{X} \colon T(\boldsymbol{x}) = t\}} \frac{f(\boldsymbol{x}; \theta)}{f(\boldsymbol{y}; \theta)}\, d\boldsymbol{x} \right)^{-1} \mathbf{1}(T(\boldsymbol{y}) = t).$$

   This vanishes when $T(\boldsymbol{y}) \neq t$; otherwise, the ratio in the integral is free of θ by the condition imposed on T. As a result, $f(\,\cdot\,, \theta \mid T = t)$ is free of θ, whence T is sufficient for θ.

Next, suppose that S is sufficient for θ. Use Neyman-Fisher factorization to write

$$f(\boldsymbol{x};\theta) = h(\boldsymbol{x})g(S(\boldsymbol{x});\theta).$$

To show that T is some measurable function of S, it is enough to show that $S(\boldsymbol{x}) = S(\boldsymbol{y}) \implies T(\boldsymbol{x}) = T(\boldsymbol{y})$. This would allow us to construct a map sending each $S^{-1}(s) = \{\boldsymbol{x} \in \mathcal{X} \colon S(\boldsymbol{x}) = s\}$ to $T(\boldsymbol{x}_s)$, where $\boldsymbol{x}_s \in S^{-1}(s)$; this is well-defined because the aforementioned property means that T is constant on $S^{-1}(s)$! Indeed, $S(\boldsymbol{x}) = S(\boldsymbol{y})$ gives

$$\frac{f(\boldsymbol{x};\theta)}{f(\boldsymbol{y};\theta)} = \frac{h(\boldsymbol{x})g(S(\boldsymbol{x});\theta)}{h(\boldsymbol{y})g(S(\boldsymbol{y});\theta)} = \frac{h(\boldsymbol{x})}{h(\boldsymbol{y})}$$

which is free of θ for all $\theta \in \Theta$, forcing $T(\boldsymbol{x}) = T(\boldsymbol{y})$ as desired.

(b) Let $X_1, \ldots, X_n \overset{iid}{\sim} \text{Cauchy}(\theta, 1)$, and let $T(\boldsymbol{X}) = (X_{(1)}, \ldots, X_{(n)})$. Then,

$$\frac{f(\boldsymbol{x};\theta)}{f(\boldsymbol{y};\theta)} = \prod_{i=1}^{n} \frac{1 + (y_i - \theta)^2}{1 + (x_i - \theta)^2} = \prod_{i=1}^{n} \frac{1 + (y_{(i)} - \theta)^2}{1 + (x_{(i)} - \theta)^2}.$$

Clearly, $T(\boldsymbol{x}) = T(\boldsymbol{y})$ forces this ratio to be identically 1 hence free of θ for all $\theta \in \mathbb{R}$. Conversely, suppose that this ratio is free of θ for all $\theta \in \mathbb{R}$. Fix $\boldsymbol{x}, \boldsymbol{y} \in \mathcal{X}$, and call the resulting ratio r, so that

$$\prod_{i=1}^{n}(1 + (y_i - \theta)^2) = r\prod_{i=1}^{n}(1 + (x_i - \theta)^2) \quad \text{for all } \theta \in \mathbb{R}.$$

In fact, we can say that the two polynomials, being identically equal on $\mathbb{R}$, must have the precisely the same coefficients; for instance, collecting and equating coefficients of θ^{2n} forces $r = 1$. Thus, this equality must hold for all $\theta \in \mathbb{C}$. The right hand side has $2n$ roots $\{x_i \pm i\}_{i=1}^n$, while the left has roots $\{y_i \pm i\}_{i=1}^n$ (with multiplicity). Since the polynomials are identically equal, they must share the same roots, so $\{x_i\}_{i=1}^n = \{y_i\}_{i=1}^n$, i.e. the x_i's and y_i's are permutations of each other. This immediately gives $T(\boldsymbol{x}) = T(\boldsymbol{y})$. Together, we have T minimal sufficient for θ.

3. Let $\mathcal{P} = \{f(x;\theta) \colon \theta \in \Theta\}$ be a parametric family of pmf's/pdf's on a sample space $\mathcal{X}$, and let $\mathcal{X}' = \{x \in \mathcal{X} \colon f(x;\theta) > 0\}$ be independent of θ. Let ν be the common dominating measure for $\mathcal{P}$ on $\mathcal{X}$. Further let $\Theta_0 \subseteq \Theta$, and let $\mathcal{P}_0 = \{f(x;\theta) \colon \theta \in \Theta_0\} \subseteq \mathcal{P}_0$. For $X_1, \ldots, X_n \overset{iid}{\sim} f(x;\theta)$, suppose that $T(X)$ is sufficient for $\mathcal{P}$, and minimal sufficient for $\mathcal{P}_0$.

Let S be sufficient for $\mathcal{P}$, hence for $\mathcal{P}_0$. Then, minimal sufficiency of T for $\mathcal{P}_0$ means that there exists a measurable function q such that

$$T(x) = q(S(x)) \quad f(\,\cdot\,;\theta)\text{-a.s. for all } \theta \in \Theta_0.$$

We claim that this equality also holds $f(\,\cdot\,;\theta)$-a.s. for all $\theta \in \Theta$. Indeed, consider the event $A = \{x \in \mathcal{X} \colon T(x) \neq q(S(x))\}$. By fixing $\theta_0 \in \Theta_0$, we must have $P_{\theta_0}(A) = 0$, i.e.

$$\int_A f(x;\theta_0)\, d\nu(x) = \int_{A\cap\mathcal{X}'} f(x;\theta_0)\, d\nu(x) = 0.$$

Since $f(\,\cdot\,,\theta_0) > 0$ on $A \cap \mathcal{X}'$, we must have $\nu(A \cap \mathcal{X}') = 0$, hence $P_\theta(A \cap \mathcal{X}') = P_\theta(A) = 0$ for all $\theta \in \Theta$.

4. (a) Note that the given Poisson family has common support. The joint density can be written as

$$f(\boldsymbol{x};\theta) = \prod_{i=1}^{n} \frac{\theta^{x_i}}{x_i!}e^{-\theta} = \frac{\theta^{n\bar{x}}}{\prod_{i=1}^{n} x_i!}e^{-n\theta},$$

whence $\bar{X} = \frac{1}{n}\sum_{i=1}^n X_i$ is sufficient for $\theta > 0$ by Neyman-Fisher factorization. We can use the Bahadur construction to obtain a minimal sufficient statistic for $\theta \in \{1, 2\}$, as

$$\frac{f(\boldsymbol{x}; 2)}{f(\boldsymbol{x}; 1)} = 2^{n\bar{x}} e^{-n},$$

hence $\bar{X}$ is also minimal sufficient for $\theta \in \{1, 2\}$. By the result in (3), $\bar{X}$ is minimal sufficient for $\theta > 0$.

(b) Write the joint density as

$$f(\boldsymbol{x}; \theta) = \prod_{i=1}^n \mathbf{1}(\theta < x_i < \theta + 1) = \mathbf{1}(\theta < x_{(1)} \leq x_{(n)} < \theta + 1).$$

Thus, $(X_{(1)}, X_{(n)})$ is sufficient for $\theta \in \mathbb{R}$. We now use the criterion from (2); examine the ratio

$$\frac{f(\boldsymbol{x}; \theta)}{f(\boldsymbol{y}; \theta)} = \frac{\mathbf{1}(x_{(n)} - 1 < \theta < x_{(1)})}{\mathbf{1}(y_{(n)} - 1 < \theta < y_{(1)})}.$$

Clearly, this is identically 1 (hence independent of θ) when $(x_{(1)}, x_{(n)}) = (y_{(1)}, y_{(n)})$. Conversely, for this ratio to be free of θ, the two indicator functions must describe precisely the same intervals, forcing $(x_{(1)}, x_{(n)}) = (y_{(1)}, y_{(n)})$. Thus, $(X_{(1)}, X_{(n)})$ is minimal sufficient for $\theta \in \mathbb{R}$.

(c) Observe that

$$\boldsymbol{Y} \sim N(\boldsymbol{X}\boldsymbol{\beta}, \sigma^2 \boldsymbol{I}_n), \qquad \boldsymbol{X} = \begin{bmatrix} 1 & x_1 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix}, \quad \boldsymbol{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix}.$$

Thus,

$$\begin{aligned} f(\boldsymbol{y}; \boldsymbol{\beta}, \sigma) &= (2\pi)^{-n/2} \sigma^{-n} \exp\left(-\frac{1}{2\sigma^2} \|\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}\|^2\right) \\ &= (2\pi)^{-n/2} \sigma^{-n} \exp\left(-\frac{1}{2\sigma^2} \|\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}\|^2\right). \end{aligned}$$

Examine

$$\frac{f(\boldsymbol{y}; \boldsymbol{\beta}, \sigma)}{f(\boldsymbol{z}; \boldsymbol{\beta}, \sigma)} = \exp\left(-\frac{1}{2\sigma^2}\left[\|\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}\|^2 - \|\boldsymbol{z} - \boldsymbol{X}\boldsymbol{\beta}\|^2\right]\right).$$

Note that because $\sigma > 0$ is allowed to vary, this is free of θ if and only if

$$\|\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}\|^2 = \|\boldsymbol{z} - \boldsymbol{X}\boldsymbol{\beta}\|^2 \quad \text{for all } \boldsymbol{\beta} \in \mathbb{R}^2.$$

Equivalently,

$$\boldsymbol{y}^\top \boldsymbol{y} - 2\boldsymbol{y}^\top \boldsymbol{X}\boldsymbol{\beta} = \boldsymbol{z}^\top \boldsymbol{z} - 2\boldsymbol{z}^\top \boldsymbol{X}\boldsymbol{\beta} \quad \text{for all } \boldsymbol{\beta} \in \mathbb{R}^2.$$

Thus, the density ratio is independent of θ if and only if $(\boldsymbol{y}^\top \boldsymbol{y}, \boldsymbol{X}^\top \boldsymbol{y}) = (\boldsymbol{z}^\top \boldsymbol{z}, \boldsymbol{X}^\top \boldsymbol{z})$. It follows that $(\sum_{i=1}^n Y_i, \sum_{i=1}^n Y_i^2, \sum_{i=1}^n X_i Y_i)$ is minimal sufficient for θ.

5. Note that applying Hölder's inequality with $p = 1 + 1/\alpha$, $q = 1 + \alpha$ yields

$$\int f^\alpha g \leq \left(\int f^{1+\alpha}\right)^{\alpha/(1+\alpha)} \left(\int g^{1+\alpha}\right)^{1/(1+\alpha)}.$$

Set $x = \log \int f^{1+\alpha}$, $y = \log \int g^{1+\alpha}$, $z = \log \int f^\alpha g$, and $\beta = \alpha/(1+\alpha) \in (0,1)$. Then, the above inequality becomes

$$z \leq \beta x + (1-\beta) y.$$

Now, observe that

$$\varphi\left(\int f^{1+\alpha}\right) = \psi(x), \qquad \varphi\left(\int g^{1+\alpha}\right) = \psi(y), \qquad \varphi\left(\int f^\alpha g\right) = \psi(z).$$

The fact that ψ is strictly increasing and convex gives

$$\psi(z) \leq \psi(\beta x + (1-\beta) y) \leq \beta \psi(x) + (1-\beta)\psi(y).$$

Dividing by β, this immediately gives

$$\mathrm{FDPD}_{\varphi,\alpha}(g,f) = \psi(x) - \frac{1}{\beta}\psi(z) + \left(\frac{1}{\beta} - 1\right)\psi(y) \geq 0.$$

For equality, we must have had equality in Hölder's, so $f^{1+\alpha} \propto g^{1+\alpha}$ μ-a.e. The normalizing condition that $\int f = \int g = 1$ will force $f = g$ μ-a.e. Conversely, $f = g$ gives $x = y = z$, hence $\mathrm{FDPD}_{\varphi,\alpha}(g,f) = 0$.

Remark: Note that x, y so defined must be finite via the normalizing conditions $\int f = \int g = 1$. If $z = -\infty$, the proof still holds.