1. Let $\{X_i\}_{i=1,\dots,n}$ be a random sample iid from F,

   (a) If F is $\mathcal{N}(\mu, \sigma^2)$ with μ and σ^2 unknown, please find a sufficient statistic for (μ, σ^2).

The joint density of $\{X_i\}_{i=1,\dots,n}$ is

$$\begin{aligned} f(x_1,\dots,x_n) &= \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x_i-\mu)^2}{2\sigma^2}\right) \\ &= \left(\frac{1}{\sqrt{2\pi}\sigma}\right)^n \exp\left(-\frac{\sum_{i=1}^n (x_i-\mu)^2}{2\sigma^2}\right) \\ &= \left(\frac{1}{\sqrt{2\pi}\sigma}\right)^n \exp\left(-\frac{\sum_{i=1}^n x_i^2 - 2\mu\sum_{i=1}^n x_i + n\mu^2}{2\sigma^2}\right) \\ &= \left(\frac{1}{\sqrt{2\pi}\sigma}\right)^n \exp\left(-\frac{n\mu^2}{2\sigma^2}\right) \exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^n x_i^2\right) \exp\left(\frac{\mu}{\sigma^2}\sum_{i=1}^n x_i\right). \end{aligned}$$

Let $T_1(\boldsymbol{X}) = \sum_{i=1}^n X_i^2$, $T_2(\boldsymbol{X}) = \sum_{i=1}^n X_i$,

$$g_\theta(T_1(\boldsymbol{x}), T_2(\boldsymbol{x})) = \left(\frac{1}{\sqrt{2\pi}\sigma}\right)^n \exp\left(-\frac{n\mu^2}{2\sigma^2}\right) \exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^n x_i^2\right) \exp\left(\frac{\mu}{\sigma^2}\sum_{i=1}^n x_i\right),$$

and $h(\boldsymbol{x}) = 1$, by Neyman-Fisher Factorization Criterion, we know that $(\sum_{i=1}^n X_i^2, \sum_{i=1}^n X_i)$ is a sufficient statistic for (μ, σ^2).

   (b) If F is Uniform$(\theta - 1/2, \theta + 1/2]$, please find the sufficient statistic for θ.

The joint density of $\{X_i\}_{i=1,\dots,n}$ is

$$\begin{aligned} f(x_1,\dots,x_n) &= \prod_{i=1}^{n} \mathbb{1}\left(X_i \in \left(\theta - \frac{1}{2}, \theta + \frac{1}{2}\right]\right) \\ &= \mathbb{1}\left(X_{(1)} \in \left(\theta - \frac{1}{2}, \theta + \frac{1}{2}\right]\right) \mathbb{1}\left(X_{(n)} \in \left(\theta - \frac{1}{2}, \theta + \frac{1}{2}\right]\right), \end{aligned}$$

where $X_{(k)}$ is the k-th order statistic of the sample. Let $T_1(\boldsymbol{X}) = X_{(1)}$, $T_2(\boldsymbol{X}) = X_{(n)}$,

$$g_\theta(T_1(\boldsymbol{x}), T_2(\boldsymbol{x})) = \mathbb{1}\left(X_{(1)} \in \left(\theta - \frac{1}{2}, \theta + \frac{1}{2}\right]\right) \mathbb{1}\left(X_{(n)} \in \left(\theta - \frac{1}{2}, \theta + \frac{1}{2}\right]\right),$$

and $h(\boldsymbol{x}) = 1$, by Neyman-Fisher Factorization Criterion, we know that $(X_{(1)}, X_{(n)})$ is a sufficient statistic for θ.

2. Let $\{X_i\}_{i=1,\dots,n}$ be a random sample iid from F, where $f = F'$ is continuous. For $\tau \in (0,1)$, denote ξ_τ as the τ-th quantile of the distribution (i.e., $F(\xi_\tau) = \tau$), and $f(\xi_\tau) > 0$, then please show that $X_{(k)} \xrightarrow{\text{p}} \xi_\tau$ where $X_{(k)}$ is the k-th order statistic of the sample and $k = [n\tau]$.

By Theorem 8.18 in *Theoretical Statistics – Topics of a Core Course* (Keener, 2010), we know that

$$\sqrt{n}(X_{(k)} - \xi_\tau) \xrightarrow{\text{d}} \mathcal{N}\left(0, \frac{\tau(1-\tau)}{[F'(\xi_\tau)]^2}\right).$$

Consider that fact that $1/\sqrt{n} \to 0$ and apply Slutsky's Theorem, we have

$$\sqrt{n}(X_{(k)} - \xi_\tau)\frac{1}{\sqrt{n}} = X_{(k)} - \xi_\tau \xrightarrow{\text{d}} 0,$$

which is a constant. Hence,

$$X_{(k)} - \xi_\tau \xrightarrow{\text{p}} 0,$$

and by Slusky's Theorem again,

$$(X_{(k)} - \xi_\tau) + \xi_\tau \xrightarrow{\text{p}} 0 + \xi_\tau,$$

i.e., $X_{(k)} \xrightarrow{\text{p}} \xi_\tau$.

3. Let X be one observation from a $\mathcal{N}(0, \sigma^2)$ population. Is $|X|$ a sufficient statistic?

The density of the sample is

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{x^2}{2\sigma^2}\right).$$

Let $T(X) = X^2$, $g_\theta(T(x)) = f(x)$, $h(x) = 1$, by Neyman-Fisher Factorization Criterion, we know that X^2 is a sufficient statistic for σ. Since 1-1 transformation preserves sufficiency, we apply the 1-1 function $m(\cdot) = \sqrt{\cdot}$ to X^2 and obtain that $\sqrt{X^2} = |X|$ is also a sufficient statistic.

4. Let $X_1, \ldots, X_n$ be a random sample from the pdf

$$f(x|\mu, \sigma) = \frac{1}{\sigma} e^{-(x-\mu)/\sigma}, \qquad \mu < x < \infty, \qquad 0 < \sigma < \infty.$$

Find a two-dimensional sufficient statistic for (μ, σ).

The joint density of $\{X_i\}_{i=1,\ldots,n}$ is

$$\begin{aligned} f(x_1, \ldots, x_n) &= \prod_{i=1}^{n} \frac{1}{\sigma} \exp\left(-\frac{x_i - \mu}{\sigma}\right) \mathbb{1}(x_i \in (\mu, \infty)) \\ &= \frac{1}{\sigma^n} \exp\left(-\frac{\sum_{i=1}^{n} x_i - n\mu}{\sigma}\right) \prod_{i=1}^{n} \mathbb{1}(x_i \in (\mu, \infty)) \\ &= \frac{1}{\sigma^n} \exp\left(\frac{n\mu}{\sigma}\right) \exp\left(-\frac{1}{\sigma}\sum_{i=1}^{n} x_i\right) \mathbb{1}(x_{(1)} \in (\mu, \infty)). \end{aligned}$$

Let $T_1(\boldsymbol{X}) = \sum_{i=1}^{n} X_i$, $T_2(\boldsymbol{X}) = X_{(1)}$,

$$g_\theta(T(\boldsymbol{x})) = \frac{1}{\sigma^n} \exp\left(\frac{n\mu}{\sigma}\right) \exp\left(-\frac{1}{\sigma}\sum_{i=1}^{n} x_i\right) \mathbb{1}(x_{(1)} \in (\mu, \infty))$$

and $h(\boldsymbol{x}) = 1$, by Neyman-Fisher Factorization Criterion, we know that $(\sum_{i=1}^{n} X_i, X_{(1)})$ is a two-dimensional sufficient statistic for (μ, σ).

5. Let $X_1, \ldots, X_n$ be iid observations from a pdf or pmf $f(x|\boldsymbol{\theta})$ that belongs to an exponential family given by

$$f(x|\boldsymbol{\theta}) = h(x)c(\boldsymbol{\theta}) \exp\left(\sum_{i=1}^{k} w_i(\boldsymbol{\theta})t_i(x)\right),$$

where $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_d), d \leq k$. Prove that $T(\boldsymbol{X}) = (\sum_{j=1}^{n} t_1(X_j), \ldots, \sum_{j=1}^{n} t_k(X_j))$ is a sufficient statistic for $\boldsymbol{\theta}$ where $\boldsymbol{X} = (X_1, \ldots, X_n)$.

The joint density of $\{X_i\}_{i=1,\ldots,n}$ is

$$\begin{aligned} f(x_1, \ldots, x_n) &= \prod_{j=1}^{n} \left(h(x_j)c(\boldsymbol{\theta}) \exp\left(\sum_{i=1}^{k} w_i(\boldsymbol{\theta})t_i(x_j)\right)\right) \\ &= \left(\prod_{j=1}^{n} h(x_j)\right) \left(\prod_{j=1}^{n} c(\boldsymbol{\theta})\right) \exp\left(\sum_{j=1}^{n}\sum_{i=1}^{k} w_i(\boldsymbol{\theta})t_i(x_j)\right) \\ &= \left(\prod_{j=1}^{n} h(x_j)\right) (c(\boldsymbol{\theta}))^n \exp\left(\sum_{i=1}^{k} w_i(\boldsymbol{\theta})\sum_{j=1}^{n} t_i(x_j)\right). \end{aligned}$$

Let $T_i(\boldsymbol{X}) = \sum_{j=1}^{n} t_i(X_j)$ for $i = 1, 2, \ldots, k$,

$$g_{\boldsymbol{\theta}}(T_1(\boldsymbol{x}), T_2(\boldsymbol{x}), \ldots, T_k(\boldsymbol{x})) = (c(\boldsymbol{\theta}))^n \exp\left(\sum_{i=1}^{k} w_i(\boldsymbol{\theta})\sum_{j=1}^{n} t_i(x_j)\right),$$

and $H(\boldsymbol{x}) = \prod_{j=1}^{n} h(x_j)$, by Neyman-Fisher Factorization Criterion, we know that $(\sum_{j=1}^{n} t_1(X_j), \sum_{j=1}^{n} t_2(X_j), \ldots, \sum_{j=1}^{n} t_k(X_j))$ is a sufficient statistic for $\boldsymbol{\theta}$.

6. Let $X_1, \dots, X_n$ be independent random variables with pdfs

$$f(x_i|\theta) = \begin{cases} 1/2i\theta & \text{if } -i(\theta-1) < x_i < i(\theta+1) \\ 0 & \text{otherwise} \end{cases},$$

where $\theta > 0$. Find a two-dimensional sufficient statistic for θ.

The joint density of $\{X_i\}_{i=1,\dots,n}$ is

$$\begin{aligned} f(x_1, \dots, x_n) &= \prod_{i=1}^n \frac{1}{2i\theta} \mathbb{1}(x_i \in (-i(\theta-1), i(\theta+1))) \\ &= \frac{1}{(2\theta)^n \prod_{i=1}^n i} \prod_{i=1}^n \mathbb{1}(x_i/i \in (-\theta+1, \theta+1)) \\ &= \frac{1}{(2\theta)^n \prod_{i=1}^n i} \mathbb{1}(\min_i(x_i/i) \in (-\theta+1, \theta+1)) \mathbb{1}(\max_i(x_i/i) \in (-\theta+1, \theta+1)). \end{aligned}$$

Let $T_1(\boldsymbol{X}) = \min_i(X_i/i)$, $T_2(\boldsymbol{X}) = \max_i(X_i/i)$,

$$g_\theta(T(\boldsymbol{x})) = \frac{1}{(2\theta)^n \prod_{i=1}^n i} \mathbb{1}(\min_i(x_i/i) \in (-\theta+1, \theta+1)) \mathbb{1}(\max_i(x_i/i) \in (-\theta+1, \theta+1)),$$

and $h(\boldsymbol{x}) = 1$, by Neyman-Fisher Factorization Criterion, we know that $(\min_i(X_i/i), \max_i(X_i/i))$ is a two-dimensional sufficient statistic for θ.

7. Let $f(x, y|\theta_1, \theta_2, \theta_3, \theta_4)$ be a bivariate pdf for the uniform distribution on the rectangle with lower left corner (θ_1, θ_2) and upper right corner (θ_3, θ_4) in $\mathbb{R}^2$. The parameters satisfy $\theta_1 < \theta_3$ and $\theta_2 < \theta_4$. Let $(X_1, Y_1), \dots, (X_n, Y_n)$ be a random sample from this pdf. Find a four-dimensional sufficient statistic for $(\theta_1, \theta_2, \theta_3, \theta_4)$.

From the given condition we can write out the bivariate pdf of $\boldsymbol{p} = (x, y)$ as

$$f(\boldsymbol{p}|\theta_1, \theta_2, \theta_3, \theta_4) = \frac{1}{(\theta_3 - \theta_1)(\theta_4 - \theta_2)} \mathbb{1}(x \in [\theta_1, \theta_3]) \mathbb{1}(y \in [\theta_2, \theta_4]).$$

The joint density of $(X_1, Y_1), \dots, (X_n, Y_n)$ is therefore

$$\begin{aligned} & f(\boldsymbol{p}_1, \dots, \boldsymbol{p}_n|\theta_1, \theta_2, \theta_3, \theta_4) \\ =& \prod_{i=1}^n \frac{1}{(\theta_3 - \theta_1)(\theta_4 - \theta_2)} \mathbb{1}(x_i \in [\theta_1, \theta_3]) \mathbb{1}(y_i \in [\theta_2, \theta_4]) \\ =& \frac{1}{(\theta_3 - \theta_1)^n (\theta_4 - \theta_2)^n} \prod_{i=1}^n \mathbb{1}(x_i \in [\theta_1, \theta_3]) \prod_{i=1}^n \mathbb{1}(y_i \in [\theta_2, \theta_4]) \\ =& \frac{1}{(\theta_3 - \theta_1)^n (\theta_4 - \theta_2)^n} \mathbb{1}(x_{(1)} \in [\theta_1, \theta_3]) \mathbb{1}(x_{(n)} \in [\theta_1, \theta_3]) \mathbb{1}(y_{(1)} \in [\theta_2, \theta_4]) \mathbb{1}(y_{(n)} \in [\theta_2, \theta_4]). \end{aligned}$$

Let $T_1 = X_{(1)}$, $T_2 = X_{(n)}, T_3 = Y_{(1)}$, $T_4 = Y_{(n)}$,

$$g_\theta(T_1, T_2, T_3, T_4) = \frac{1}{(\theta_3 - \theta_1)^n (\theta_4 - \theta_2)^n} \mathbb{1}(x_{(1)} \in [\theta_1, \theta_3]) \mathbb{1}(x_{(n)} \in [\theta_1, \theta_3]) \mathbb{1}(y_{(1)} \in [\theta_2, \theta_4]) \mathbb{1}(y_{(n)} \in [\theta_2, \theta_4]),$$

and $h(\boldsymbol{x}, \boldsymbol{y}) = 1$, by Neyman-Fisher Factorization Criterion, we know that $(X_{(1)}, X_{(n)}, Y_{(1)}, Y_{(n)})$ is a four-dimensional sufficient statistic for $(\theta_1, \theta_2, \theta_3, \theta_4)$.

8. Consider an exponential family whose density is given by

$$p(x|\eta) = \exp\left\{\sum_{i=1}^{s} \eta_i T_i(x) - A(\eta)\right\} h(x) \tag{1}$$

which natural parameter space Θ. Show that Θ is convex. (Hint: See Lemma 2.7.1 of Lehmann and Romano (2005).)

> Let $\boldsymbol{\eta} = (\eta_1, \cdots \eta_s)$ and $\boldsymbol{\eta}' = (\eta_1', \cdots, \eta_s')$ be two parameter points for which the integral of Equation 1 is finite. Note that
>
> $$\Theta = \left\{\boldsymbol{\eta} \in \mathbb{R}^s : 0 < \int \exp\left\{\sum_{i=1}^{s} \eta_i T_i(x)\right\} h(x)\, d\mu(x) < \infty\right\}$$
>
> Then by Hölder's inequality,
>
> $$\begin{aligned} A(\alpha\boldsymbol{\eta} + (1-\alpha)\boldsymbol{\eta}') &= \int \exp\left\{\sum_{i=1}^{s} [\alpha\eta_i + (1-\alpha)\eta_i'] T_i(x)\right\} h(x)\, d\mu(x) \\ &= \int \left(\exp\left\{\sum_{i=1}^{s} \eta_i T_i(x)\right\}\right)^{\alpha} \left(\exp\left\{\sum_{i=1}^{s} \eta_i' T_i(x)\right\}\right)^{1-\alpha} h(x)\, d\mu(x) \\ &\leq \left(\int \exp\left\{\sum_{i=1}^{s} \eta_i T_i(x)\right\} h(x)\, d\mu(x)\right)^{\alpha} \left(\int \exp\left\{\sum_{i=1}^{s} \eta_i' T_i(x)\right\} h(x)\, d\mu(x)\right)^{1-\alpha} \\ &< \infty \end{aligned}$$
>
> for any $0 < \alpha < 1$. Therefore, Θ is convex.

9. [Rao-Blackwell Theorem]. Let X be a random observable with distribution $P_\theta \in \mathcal{P} = \{P_{\theta'} : \theta' \in \Theta\}$, and let T be sufficient for $\mathcal{P}$. Let δ be an estimator of an estimand $g(\theta)$, and let the loss function $L(\theta, d)$ be a strictly convex function of d. If δ has finite expectation and risk,

$$R(\theta, \delta) = \mathbb{E}\{L(\theta, \delta(X))\} < \infty$$

and if

$$\eta(t) = \mathbb{E}\{\delta(X)|t\}$$

then the risk of the estimator $\eta(T)$ satisfies

$$R(\theta, \eta) < R(\theta, \delta)$$

unless $\delta(X) = \eta(T)$ with probability 1.
Proof

> Jensen's inequality with expectations against the conditional distribution of $\delta(X)$ given T gives
>
> $$L(\theta, \eta(T)) \leq \mathbb{E}_\theta\{L(\theta, \delta(X))|T\}$$
>
> Taking expectation gives
>
> $$R(\theta, \eta) < R(\theta, \delta)$$
>
> unless $\delta(X) = \eta(T)$ with probability 1.

10. Let $X_1 \leq X_2 \leq \dots \leq X_n$ be iid according to the exponential distribution $\operatorname{Exp}(a,b)$, i.e., X_i has density

$$f_X(x) = \frac{1}{b} e^{-(x-a)/b} \mathbb{1}(x \geq a), \qquad a \in \mathbb{R}, b > 0.$$

Note let $X_{(1)}, X_{(2)}, \dots, X_{(n)}$ be the corresponding order statistic of the sample and let $T_1 = X_{(1)}$, $T_2 = \sum_i (X_i - X_{(1)})$. Show that (T_1, T_2) are independently distributed as $\mathrm{E}(a, b/n)$ and $\frac{1}{2} b \chi^2_{2n-2}$ respectively, and there are jointly sufficient and complete.

(i) Sufficiency: Let $\theta = (a,b)$. Note that the joint pdf of $\{X_i\}_{i=1,\dots,n}$ can be wriiten as

$$\begin{aligned} f(x_1,\dots,x_n) = \prod_{i=1}^n \frac{1}{b} e^{-(x_i-a)/b} \mathbb{1}(x_i \geq a) &= \frac{1}{b^n} e^{-\sum_{i=1}^n (x_i-a)/b} \mathbb{1}(x_{(1)} \geq a) \\ &= \frac{1}{b^n} e^{-\sum_{i=1}^n (x_i - x_{(1)})/b} e^{-(x_{(1)}-a)/(b/n)} \mathbb{1}(x_{(1)} \geq a). \end{aligned}$$

Let $T_1(X) = X_{(1)}$, $T_2(X) = \sum_{i=1}^n (X_i - X_{(1)})$,

$$g_\theta(T_1(x), T_2(x)) = \frac{1}{b^n} e^{-T_2(x)/b} e^{-(T_1(x)-a)/(b/n)} \mathbb{1}(T_1(x) \geq a),$$

and $h(x) = 1$, by Neyman-Fisher Factorization Criterion, we know that $(T_1(X), T_2(X))$ is a sufficient statistic.

(ii) Independence: The joint pdf of all the order statistic $X_{(1)}, \dots, X_{(n)}$ is given by

$$\widetilde{f}(x_1,\dots,x_n) = n! f(x_1,\dots,x_n) = \frac{n!}{b^n} e^{-\sum_{i=1}^n (x_i-a)/b} \mathbb{1}(a \leq x_1 \leq x_2 \leq \dots \leq x_n).$$

Define $M_1 = X_{(1)}, M_2 = X_{(2)} - X_{(1)}, \dots, M_n = X_{(n)} - X_{(n-1)}$, then $X_{(1)} = M_1, X_{(2)} = M_1 + M_2, \dots, X_{(n)} = M_1 + M_2 + \dots M_n$, and the Jacobian matrix of such transform is

$$\left(\frac{\partial X_{(i)}}{\partial M_j} \right)_{1 \leq i,j \leq n} = \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 \\ 1 & 1 & 0 & \cdots & 0 \\ 1 & 1 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & 1 & \cdots & 1 \end{bmatrix}$$

with determinant 1, so the pdf of $M_1, \dots, M_n$ is given by

$$\begin{aligned} & f_M(m_1,\dots,m_n) \\ =& \frac{n!}{b^n} e^{-(nm_1 + (n-1)m_2 + \dots + m_n - na)/b} \mathbb{1}(m_1 \geq a, m_2 \geq 0, m_3 \geq 0, \dots, m_n \geq 0) \\ =& \left(\frac{1}{b/n} e^{-(m_1-a)/(b/n)} \mathbb{1}(m_1 \geq a) \right) \left(\frac{1}{b/(n-1)} e^{-m_2/(b/(n-1))} \mathbb{1}(m_2 \geq 0) \right) \cdots \left(\frac{1}{b/1} e^{-m_n/(b/1)} \mathbb{1}(m_n \geq 0) \right), \end{aligned}$$

so $M_1, M_2, \dots, M_n$ are independent, and hence $M_1 = T_1(X)$ is independent of $(n-1)M_2 + (n-2)M_3 \dots + M_n = T_2(X)$.

(iii) Distribution: From (ii) we immediately know that

$$f_{M_1}(m_1) = \frac{1}{b/n} e^{-(m_1-a)/(b/n)} \mathbb{1}(m_1 \geq a),$$

comparing with the density of $\mathrm{E}(a,b)$, we know that M_1, i.e., $X_{(1)}$, is $\mathrm{E}(a, b/n)$ distributed.

From (ii) we immediately know also that

$$f_{M_i}(m_i) = \frac{1}{b/(n+1-i)} e^{-m_i/(b/(n+1-i))} \mathbb{1}(m_i \geq 0),$$

for $i = 2, 3, \dots, n$, so comparing with the density of $\mathrm{E}(0, b/(n+1-i))$, we know that M_i is $\mathrm{E}(0, b/(n+1-i))$ distributed. From the form of density of $f_{M_i}(m_i)$, we know that M_i is a scale family, so we know $N_i := M_i/(b/(n+1-i))$ for each $i = 2, 3, \dots, n$ is exponentially distributed with rate 1, which is also a $\frac{1}{2}\chi^2_2$ distribution. Due to the independence of $M_2, \dots, M_n$, we know that $N_2, N_3, \dots, N_n$ are also independent, and using the property of independent chi-square random variables, we know

$$N_2 + N_3 + \dots + N_n = \frac{(n-1)M_2}{b} + \frac{(n-2)M_3}{b} + \dots + \frac{M_n}{b} \sim \frac{1}{2} \chi^2_{2(n-1)},$$

i.e.,

$$T_2(X) = (n-1)M_2 + (n-2)M_3 + \ldots + M_n \sim \frac{1}{2}b\chi^2_{2n-2}.$$

(iv) Completeness: Let $\mathbb{E}_\theta(h(T_1, T_2)) = 0$ for all θ. Note that the density of T_2 does not involve a, by the Tower Property we can write

$$0 = \mathbb{E}_{a,b}(h(T_1,T_2)) = \mathbb{E}_{a,b}[\mathbb{E}_{a,b}(h(T_1,T_2)|T_1)] = \mathbb{E}_{a,b}[\mathbb{E}_b(h(T_1,T_2)|T_1)].$$

Define $g_b(T_1) = \mathbb{E}_b(h(T_1,T_2)|T_1)$, we have for fixed b, (after multiplying b/n,)

$$0 = \int_a^\infty g_b(t_1)e^{-(t_1-a)/(b/n)}dt_1$$

for all a, so we must have $g_b(t_1) = 0$ almost everywhere in the support of T_1. Interchange the role of t_1 and b, we know that for almost every t_1, $g_b(t_1) = 0$ almost everywhere in $\mathbb{R}^+$; combining with the fact that $g_b(t_1)$ is continuous in b, we further conclude that for almost every t_1, $g_b(t_1) = \mathbb{E}_b(h(t_1,T_2)|t_1) = 0$ for all $b \in \mathbb{R}^+$. Hence, for all fixed $b \in \mathbb{R}^+$ and almost every fixed t_1, (after dividing $b/2$ and multiplying $2^{n-1}\Gamma(n-1)$,)

$$0 = \int_0^\infty h(t_1,x)x^{n-2}e^{-x/2}dx,$$

so we must have $h(t_1,x) = 0$ almost everywhere in the support of T_2. Hence, $h(t_1,t_2) = 0$ for almost every (t_1,t_2), and we finished the proof of completeness.

11. Let $X_1, X_2, \cdots, X_n$ be *i.i.d.* according to the logistic distribution $L(\theta, 1)$, *i.e.*, X_i has density

$$f_X(x) = \frac{e^{-(x-\theta)}}{\{1+e^{-(x-\theta)}\}^2}, \qquad \theta \in \mathbb{R} \tag{2}$$

Consider a subfamily $\mathcal{P}_0$ consisting of the distribution 2 with $\theta_0 = 0$ and $\theta_1, \cdots, \theta_{n+1}$. Show that the order statistic $T(X) = (X_{(1)}, X_{(2)}, \cdots, X_{(n)})$ is minimal sufficient for $\mathcal{P}_0$.

Consider the ratio of two samples $\mathbf{x}$ and $\mathbf{y}$ joint p.d.f.s

$$\begin{aligned}\frac{p(\mathbf{x}|\theta)}{p(\mathbf{y}|\theta)} = \frac{\prod_{i=1}^n \frac{e^{-(x_i-\theta)}}{\{1+e^{-(x_i-\theta)}\}^2}}{\prod_{i=1}^n \frac{e^{-(y_i-\theta)}}{\{1+e^{-(y_i-\theta)}\}^2}} &= \frac{\exp\{-(\sum_{i=1}^n x_i - \theta)\}}{\exp\{-(\sum_{i=1}^n y_i - \theta)\}} \times \frac{\prod_{i=1}^n\{1+e^{-(y_i-\theta)}\}^2}{\prod_{i=1}^n\{1+e^{-(x_i-\theta)}\}^2} \\ &= \exp\left\{-\left(\sum_{i=1}^n x_i - y_i\right)\right\}\left(\prod_{i=1}^n \frac{1+e^{-(y_i-\theta)}}{1+e^{-(x_i-\theta)}}\right)^2\end{aligned}$$

which does not depend on θ *iff* $T(\mathbf{x}) = T(\mathbf{y})$, *i.e.* having the same order statistics. Hence, by Lehmann-Scheffè Theorem, $T(X) = (X_{(1)}, X_{(2)}, \cdots, X_{(n)})$ is a minimal sufficient for $\theta \in \mathbb{R}$. Note that the minimal sufficient statistics for $\mathcal{P}_0$ is defined as:

$$T_{\mathcal{P}_0}(X) = (T_1(X), \cdots, T_n(X)) = \left(\frac{p(X|\theta_1)}{p(X|\theta_0)}, \cdots, \frac{p(X|\theta_n)}{p(X|\theta_0)}\right)$$

where

$$T_j(X) = \frac{p_{\theta_j}(X)}{p_{\theta_0}(X)} = \prod_{i=1}^n \frac{\frac{e^{-(x_i-\theta_j)}}{\{1+e^{-(x_i-\theta_j)}\}^2}}{\frac{e^{-x_i}}{\{1+e^{-x_i}\}^2}} = e^{n\theta_j}\prod_{i=1}^n\left(\frac{1+e^{-x_i}}{1+e^{-(x_i-\theta_j)}}\right)^2$$

As we have proof that $T(X) = (X_{(1)}, X_{(2)}, \cdots, X_{(n)})$ is a minimal sufficient for $\theta \in \mathbb{R}$, $T(X) = (X_{(1)}, X_{(2)}, \cdots, X_{(n)})$ is also minimal sufficient for $\mathcal{P}_0$

12. [Problem 3.27 of Keener (2010)] Let $X_1, \cdots, X_n$ be *i.i.d.* from a uniform distribution on $(-\theta, \theta)$, where $\theta > 0$ is an unknown parameter.

(a) Find a minimal sufficient statistic T.

(b) Define

$$V = \frac{\bar{X}_n}{\max_{1\le i\le n} X_i - \min_{1\le i\le n} X_i}$$

where $\bar{X}_n = \frac{1}{n}\sum_{i=1}^n X_i$ denotes the sample average. Show that T and V are independent.

(a) Consider the ratio of two samples $\mathbf{x}$ and $\mathbf{y}$ joint p.d.f.s

$$\frac{p(\mathbf{x}|\theta)}{p(\mathbf{y}|\theta)} = \frac{\prod_{i=1}^n \mathbb{1}\{-\theta < x_i < \theta\}}{\prod_{i=1}^n \mathbb{1}\{-\theta < y_i < \theta\}} = \frac{\mathbb{1}\{\min_{1\le i\le n} x_i > -\theta\}\mathbb{1}\{\max_{1\le i\le n} x_i < \theta\}}{\mathbb{1}\{\min_{1\le i\le n} y_i > -\theta\}\mathbb{1}\{\max_{1\le i\le n} y_i < \theta\}}$$

which does not depend on θ *iff* $T(\mathbf{x}) = T(\mathbf{y})$. Hence, by Lehmann-Scheffè Theorem, $T(X) = (\min_{1\le i\le n} X_i, \max_{1\le i\le n} X_i)$ is a minimal sufficient for θ.

(b) Let $Y_i \sim X_i/\theta \sim U(-1,1)$, which does not depend on θ. Note that

$$\bar{X}_n = \theta\bar{Y}_n, \qquad \min_{1\le i\le n} X_i = \theta \min_{1\le i\le n} Y_i, \qquad \max_{1\le i\le n} X_i = \theta \max_{1\le i\le n} Y_i$$

Then

$$\begin{aligned} V &= \frac{\bar{X}_n}{\max_{1\le i\le n} X_i - \min_{1\le i\le n} X_i} = \frac{\theta\bar{Y}_n}{\theta\max_{1\le i\le n} Y_i - \theta\min_{1\le i\le n} Y_i} \\ &= \frac{\bar{Y}_n}{\max_{1\le i\le n} Y_i - \min_{1\le i\le n} Y_i} \end{aligned}$$

which does not depend on θ. Hence V is ancillary. Note that in 1., we prove that $T(X) = (\min_{1\le i\le n} X_i, \max_{1\le i\le n} X_i)$ is a minimal sufficient for θ, then

$$\mathbb{P}(T \le t) = \mathbb{P}(X_{(1)} \ge -t, X_{(n)} \le t) = \mathbb{P}(-t \le X_1, \cdots, X_n \le t) = \{\mathbb{P}(-t \le X_1 \le t)\}^n = \left(\frac{t}{\theta}\right)^n$$

$$f_T(t) = F_T'(t) = \frac{nt^{n-1}}{\theta^n}$$

i.e. $T \sim U(0,\theta)$. Then,

$$\begin{aligned} & \forall\theta \in \mathbb{R}^+, && 0 = \mathbb{E}_\theta\{g(T)\} = \int_0^\theta g(t)\frac{nt^{n-1}}{\theta^n}dt \\ \Rightarrow\ & \forall\theta \in \mathbb{R}^+, && 0 = \int_0^\theta g(t)t^{n-1}dt \\ \Rightarrow\ & \forall\theta \in \mathbb{R}^+, && 0 = \frac{\partial}{\partial\theta}\left\{\int_0^\theta g(t)t^{n-1}dt\right\} = g(\theta)\theta^{n-1} \\ \Rightarrow\ & \forall\theta \in \mathbb{R}^+, && 0 = g(\theta) \qquad (\theta > 0) \\ \Rightarrow\ & \forall\theta \in \mathbb{R}^+, && \mathbb{P}_\theta\{g(T) = 0\} = 1 \end{aligned}$$

Hence, T is complete and minimal sufficient. Finally, by Basu's Theorem, T and V are independent.