Indian Statistical Institute, Kolkata

Large Sample Statistical Methods

End Semestral Examinations 2019-2020

Subhrajyoty Roy
Roll: MB1911

July 27, 2020

Contents

Note: $X, X_n, Y_n (n \geq 1)$ *denote random variables. For any limiting distribution we assume* $n \to \infty$.

1 Problem 1

Problem. Let $X, X_n, n \geq 1$, be random variables such that

$$\liminf_{n \to \infty} \mathbb{E}\left[g(X_n)\right] \geq \mathbb{E}\left[g(X)\right]$$

for every bounded continuous function $g : (-\infty, \infty) \to [0, \infty)$. Show that, $X_n \xrightarrow{d} X$. [It is known that $X_n \xrightarrow{d} X$ if and only if $\lim_{n \to \infty} \mathbb{E}\left[g(X_n)\right] = \mathbb{E}\left[g(X)\right]$ for every bounded continuous function $g : \mathbb{R} \to \mathbb{R}$.]

Solution:
We are given $X, X_n, n \geq 1$, be random variables such that

$$\liminf_{n \to \infty} \mathbb{E}\left[g(X_n)\right] \geq \mathbb{E}\left[g(X)\right] \tag{1.1}$$

for every bounded continuous function $g : (-\infty, \infty) \to [0, \infty)$.

Let, $g : \mathbb{R} \to [0, \infty)$ be any bounded and continuous function. Since g is bounded, there must exist $M > 0$ such that, $g(x) \leq M \quad \forall x \in \mathbb{R}$. Consider the function $g'(x) = M - g(x)$, clearly, $g' : \mathbb{R} \to [0, \infty)$, and is continuous and is also bounded, (as $g'(x) \leq M$ since $g(x) \geq 0$). Therefore, by applying eq. (1.1) on g' instead of g it follows that:

$$\liminf_{n \to \infty} \mathbb{E}\left[g'(X_n)\right] \geq \mathbb{E}\left[g'(X)\right]$$
$$\implies \quad \liminf_{n \to \infty} \mathbb{E}\left[M - g(X_n)\right] \geq \mathbb{E}\left[M - g(X)\right]$$
$$\implies \quad M - \limsup_{n \to \infty} \mathbb{E}\left[g(X_n)\right] \geq M - \mathbb{E}\left[g(X)\right]$$
$$\implies \quad \mathbb{E}\left[g(X)\right] \geq \limsup_{n \to \infty} \mathbb{E}\left[g(X_n)\right]$$

Thus, combining this and eq. (1.1) yields,

$$\liminf_{n \to \infty} \mathbb{E}\left[g(X_n)\right] \geq \mathbb{E}\left[g(X)\right] \geq \limsup_{n \to \infty} \mathbb{E}\left[g(X_n)\right]$$

but as limit superior of a sequence must be atleast as large as the limit inferior, equality must hold throughout and the limit of the sequence $\mathbb{E}\left[g(X_n)\right]$ must exists.

Thus we obtain that for any such bounded and continuous function $g : \mathbb{R} \to [0, \infty)$, the limit infimum in eq. (1.1) can actually be replaced by limit while the inequality can be replaced by an equality, i.e.

$$\lim_{n \to \infty} \mathbb{E}\left[g(X_n)\right] = \mathbb{E}\left[g(X)\right] \tag{1.2}$$

holds for any bounded and continuous positive valued function g.

We know that, $X_n \xrightarrow{d} X$ if and only if eq. (1.2) holds for every bound and continuous function $g : \mathbb{R} \to \mathbb{R}$, hence it is enough to show that eq. (1.2) holds for all such functions.

Consider such a function $h : \mathbb{R} \to \mathbb{R}$ such that $\lim_{n \to \infty} \mathbb{E}[h(X_n)] = \mathbb{E}[h(X)]$. Define the positive and negative part of the function as follows:

$$
h_+(x) = \begin{cases} h(x) & \text{if } h(x) \geq 0 \\ 0 & \text{if } h(x) < 0 \end{cases}
$$

and

$$
h_-(x) = \begin{cases} -h(x) & \text{if } h(x) \leq 0 \\ 0 & \text{otherwise} \end{cases}
$$

Now observe that, $h_+ : \mathbb{R} \to [0, \infty)$, $h_- : \mathbb{R} \to [0, \infty)$ and $h(x) = h^+(x) - h^-(x) \quad \forall x \in \mathbb{R}$. Also note that, $h_+(x) = \max\{h(x), 0\}$ and $h_-(x) = -\min\{h(x), 0\}$ for any $x \in \mathbb{R}$. Since, h is chosen to be bounded and continuous, so is h_+ and h_-.

Hence, as h_+ and h_- are positive valued function, by the limit in eq. (1.2) we get that,

$$
\lim_{n \to \infty} \mathbb{E}[h_+(X_n)] = \mathbb{E}[h_+(X)] \quad \text{and} \quad \lim_{n \to \infty} \mathbb{E}[h_-(X_n)] = \mathbb{E}[h_-(X)]
$$

Note that, the above expectations exist because it is assumed that $\mathbb{E}(h(X))$ and $\mathbb{E}(h(X_n))$ exist and thus $\mathbb{E}[|h(X)|]$ and $\mathbb{E}[|h(X_n)|]$ exist.

Finally we have,

$$
\begin{aligned}
\lim_{n \to \infty} \mathbb{E}[h(X_n)] &= \lim_{n \to \infty} \mathbb{E}[h_+(X_n) + h_-(X_n)] \\
&= \lim_{n \to \infty} [\mathbb{E}[h_+(X_n)] + \mathbb{E}[h_-(X_n)]] \quad \text{by linearity of expectation} \\
&= \lim_{n \to \infty} \mathbb{E}[h_+(X_n)] + \lim_{n \to \infty} \mathbb{E}[h_-(X_n)] \\
&= \mathbb{E}[h_+(X)] + \mathbb{E}[h_-(X)] \quad \text{by } eq. \text{ (1.1)} \\
&= \mathbb{E}[h_+(X) + h_-(X)] \\
&= \mathbb{E}[h(X)]
\end{aligned}
$$

$$
\therefore \lim_{n \to \infty} \mathbb{E}[h(X_n)] = \mathbb{E}[h(X)] \tag{1.3}
$$

Since, eq. (1.3) holds for any $h : \mathbb{R} \to \mathbb{R}$ being bounded and continuous function, by Portmanteau's theorem, it follows that $X_n \xrightarrow{d} X$.

This completes the proof.

■

2 Problem 2

Problem. If $X_n \xrightarrow{p} 0$, show that for any median M_n of X_n, $M_n \to 0$.

Solution:

We are given that, $X_n \xrightarrow{P} 0$. We consider any median M_n of X_n. We shall prove that $M_n \to 0$ as $n \to \infty$, by method of contradiction.

For the sake of contradiction, we start by assuming that $M_n \nrightarrow 0$. This means, there exists $\epsilon > 0$ such that, for any $n \in \mathbb{N}$, there is an even bigger integer $a_n > n$ such that $|M_{a_n}| > \epsilon$.

Now consider the sequence $\{a_n\}_{n=1}^{\infty}$. As M_{a_n}'s are medians of respective X_{a_n}'s, we have;

$$\mathbb{P}(X_{a_n} \geq M_{a_n}) \geq \frac{1}{2} \quad \text{and} \quad \mathbb{P}(X_{a_n} \leq M_{a_n}) \geq \frac{1}{2} \tag{2.1}$$

Now we consider two cases separately.

1. **Case 1:** $M_{a_n} > 0$. Since, a_n's are such that $|M_{a_n}| > \epsilon$, we must have $M_{a_n} > \epsilon$. Now,

   $$\{w : X_{a_n}(w) \geq M_{a_n}\} \subseteq \{w : X_{a_n}(w) \geq \epsilon\} \subseteq \{w : |X_{a_n}(w)| \geq \epsilon\}$$

   Therefore using eq. (2.1),

   $$\frac{1}{2} \leq \mathbb{P}(X_{n_N} \geq M_{n_N}) \leq \mathbb{P}(|X_{n_N}| > \epsilon) \tag{2.2}$$

2. **Case 2:** $M_{a_n} < 0$. Similar to the previous case, $M_{a_n} < 0 \implies M_{a_n} < -\epsilon$. Hence,

   $$\{w : X_{a_n}(w) \leq M_{a_n}\} \subseteq \{w : X_{a_n}(w) \leq -\epsilon\} \subseteq \{w : |X_{a_n}(w)| \geq \epsilon\}$$

   Therefore using eq. (2.1),

   $$\frac{1}{2} \leq \mathbb{P}(X_{n_N} \leq M_{n_N}) \leq \mathbb{P}(|X_{n_N}| > \epsilon) \tag{2.3}$$

So, now by combining eq. (2.2) and eq. (2.3), we get,

$$\mathbb{P}(|X_{n_N}| > \epsilon) \geq \frac{1}{2} \quad \forall a_n, n \in \mathbb{N}$$

i.e.

$$\mathbb{P}(|X_{n_N}| > \epsilon) \nrightarrow 0$$

which contradicts the fact that $X_n \xrightarrow{P} 0$. ∎

3 Problem 3

Problem. Let $X_1, X_2, \cdots, X_n$ be a random sample from a distribution with a density $f(x, \theta)$ where θ is an unknown real parameter. Give an example to show that there may exist a consistent estimator T_n of θ for which $\mathbb{E}(T_n) \to \theta + 1$ as $n \to \infty$. [**Hint**: Consider an example where there is an unbiased consistent estimator $\hat{\theta}_n$ and a set A_n, independent of $\hat{\theta}_n$, with probability tending to one and with some other property so that $\hat{\theta}_n$ can be modified with the help of A_n to construct T_n.]

Solution:

Let us consider a specific density, namely the density of the normal distribution. Let $X_1, X_2, \cdots, X_n$ be independent and identically distributed random sample from a normal distribution $\mathcal{N}(\theta, 1)$, where θ is an unknown and $\theta \in \mathbb{R}$.

Now, denoting $\bar{X}_n$ as the sample mean, i.e. $\overline{X_n} = \frac{1}{n} \sum_{i=1}^{n} X_i$, consider the set $A_n = \{w : |X_1 - \bar{X}_n| \le \sqrt{n}\}$. Clearly, $\bar{X}_n$ is independent of the statistic $|X_1 - \bar{X}_n|$, hence in effect, the set A_n becomes independent of the estimator $\bar{X}_n$, as suggested by the hint given in the question.

Consider the estimator which modifies the estimator $\bar{X}_n$ with the help of the set A_n appropriately,

$$T_n = \bar{X}_n \mathbf{1}_{A_n} + (\bar{X}_n + \delta_n) \mathbf{1}_{A_n^c}$$

where $\mathbf{1}_A$ is the indicator function of A and

$$\delta_n = \frac{1}{2} \left[1 - \Phi \left(\frac{n}{\sqrt{(n-1)}} \right) \right]^{-1} \quad \forall n \ge 2 \qquad \text{and} \qquad \delta_1 = 1$$

As the set A_n depends on the sample $X_1, \ldots X_n$ rather than some randomization device, it belongs to the same measure space as the space of functions of the samples. Thus, the given probability measure can be used to calculate appropriate probabilities and expectations regarding T_n without any problem.

First, note that, $\bar{X}_n \xrightarrow{P} \theta$, by Weak Law of Large Numbers, hence is consistent for θ.

Now, we shall show that $(T_n - \bar{X}_n) \xrightarrow{P} 0$. To show this, choose any sufficiently small $\epsilon > 0$, and consider the following probability;

$$
\begin{aligned}
\mathbb{P}_\theta(|T_n - \bar{X}_n| > \epsilon) &\le \mathbb{P}_\theta(T_n \ne \bar{X}_n) \\
&= \mathbb{P}_\theta(A_n^c) \qquad \text{, since } \delta_n > 0 \quad \forall n \ge 1 \text{ and } \lim_{n \to \infty} \delta_n > 0 \\
&= \mathbb{P}_\theta \left(|X_1 - \bar{X}_n| > \sqrt{n} \right) \\
&= 2\mathbb{P}_\theta \left((X_1 - \bar{X}_n) > \sqrt{n} \right) \qquad \text{, since } X_1 - \bar{X}_n \text{ has a symmetric distribution about } 0 \\
&= 2\mathbb{P}_\theta \left(\frac{\sqrt{n}}{\sqrt{(n-1)}} (X_1 - \bar{X}_n) > \frac{n}{\sqrt{(n-1)}} \right)
\end{aligned}
$$

$$= 2 \left(1 - \Phi \left(\frac{n}{\sqrt{(n-1)}} \right) \right) \qquad \text{, since } (X_1 - \bar{X}_n) \sim \mathcal{N}(0, 1 - 1/n)$$

$$\text{, where } \Phi(\cdot) \text{ is the cdf of standard normal distribution}$$

$$\to 0, \quad \text{as } n \to \infty$$

where the limit in the last line follows from the fact that $\Phi(\cdot)$ is a continuous function and $\Phi(x) \to 1$ as $x \to \infty$. Therefore, we have $(T_n - \bar{X}_n) \xrightarrow{P} 0$, and hence,

$$T_n = (T_n - \bar{X}_n) + \bar{X}_n \xrightarrow{P} \theta$$

Thus, T_n is a consistent estimator for θ.
On the other hand,

$$
\begin{aligned}
\mathbb{E}_\theta(T_n) &= \mathbb{E}_\theta \left(\bar{X}_n \mathbf{1}_{A_n} + (\bar{X}_n + \delta_n) \mathbf{1}_{A_n^c} \right) \\
&= \mathbb{E}_\theta \left(\bar{X}_n (\mathbf{1}_{A_n} + \mathbf{1}_{A_n^c}) + \delta_n \mathbf{1}_{A_n^c} \right) \\
&= \mathbb{E}_\theta \left(\bar{X}_n + \delta_n \mathbf{1}_{A_n^c} \right) \qquad \text{, since } \mathbf{1}_{A_n} + \mathbf{1}_{A_n^c} = 1 \\
&= \mathbb{E}_\theta \left(\bar{X}_n \right) + \delta_n \mathbb{E}_\theta \left(\mathbf{1}_{A_n^c} \right) \quad \text{, by linearity of expectation} \\
&= \theta + \delta_n \mathbb{P}_\theta(A_n^c) \quad \text{, since } \mathbb{E}(\mathbf{1}_A) = \mathbb{P}(A) \\
&= \theta + 2\delta_n \left(1 - \Phi \left(\frac{n}{\sqrt{(n-1)}} \right) \right) \qquad \text{, by the previously shown derivations} \\
&= \theta + 1 \quad \forall n \geq 2
\end{aligned}
$$

where the last equality follows from the definition of δ_n, which was chosen in a particular way to ensure this multiplication becomes 1. Thus, $\mathbb{E}_\theta(T_n) \to \theta + 1$ as $n \to \infty$.
Thus we have obtained a consistent estimator T_n of θ for which $\mathbb{E}(T_n) \to \theta + 1$. ∎

4   Problem 4

Problem. Suppose that $(X_i, Y_i), i \geq 1$ are i.i.d. bivariate random vectors with $\mathbb{E}(X_1) = \mu_x, \mathbb{E}(Y_1) = \mu_y, \operatorname{Var}(X_1) = \sigma_x^2, \operatorname{Var}(Y_1) = \sigma_y^2$ and $\operatorname{Cor}(X_1, Y_1) = \rho$. If X_i and Y_i are positive random variables, using univariate Central Limit Theorem (CLT) show that

$$\sqrt{n} \left(\frac{\sum_{i=1}^n X_i}{\sum_{i=1}^n Y_i} - \frac{\mu_x}{\mu_y} \right)$$

converges in distribution to a normal random variable with mean 0 and variance $\left(\mu_y^2 \sigma_x^2 + \mu_x^2 \sigma_y^2 - 2\rho \mu_x \mu_y \sigma_x \sigma_y \right) / \mu_y^4$. (Note that one can use multivariate CLT and delta method to find the asymptotic distribution. You have been asked to use univariate CLT).

Solution:

We are given $(X_i, Y_i), i \geq 1$ are i.i.d. bivariate random vectors with $\mathbb{E}(X_1) = \mu_x, \mathbb{E}(Y_1) = \mu_y, \operatorname{Var}(X_1) = \sigma_x^2, \operatorname{Var}(Y_1) = \sigma_y^2$ and $\operatorname{Cor}(X_1, Y_1) = \rho$.

We start by simplying the given quantity;

$$
\begin{aligned}
\frac{\sum_{i=1}^n X_i}{\sum_{i=1}^n Y_i} - \frac{\mu_x}{\mu_y} &= \frac{1}{\sum_{i=1}^n Y_i} \left(\sum_{i=1}^n X_i - \frac{\mu_x}{\mu_y} \sum_{i=1}^n Y_i \right) \\
&= \frac{\sum_{i=1}^n \left(X_i - \frac{\mu_x}{\mu_y} Y_i \right)}{\sum_{i=1}^n Y_i} \\
&= \frac{\sum_{i=1}^n Z_i}{\sum_{i=1}^n Y_i} \quad , \text{ where } Z_i = X_i - \frac{\mu_x}{\mu_y} Y_i, \quad \forall i = 1, 2, \ldots n \\
&= \frac{\sum_{i=1}^n Z_i/n}{\sum_{i=1}^n Y_i/n} \\
&= \frac{\overline{Z_n}}{\overline{Y_n}} \quad , \text{ where } \overline{Z_n} = \frac{1}{n} \sum_{i=1}^n Z_i \text{ and similarly } \overline{Y_n} = \frac{1}{n} \sum_{i=1}^n Y_i
\end{aligned}
$$

Now since X_i, Y_i's are i.i.d. bivariate normal random variables, being a linear combination of them, Z_i's are also i.i.d. normal random variables. Note that,

$$
\begin{aligned}
\mathbb{E}(Z_i) &= \mathbb{E}(X_i) - \frac{\mu_x}{\mu_y} \mathbb{E}(Y_i) \\
&= \mu_x - \frac{\mu_x}{\mu_y} \mu_y \\
&= 0
\end{aligned}
$$

and,

$$\text{Var}(Z_i) = \text{Var}(X_i) - \frac{\mu_x^2}{\mu_y^2}\text{Var}(Y_i) - 2\frac{\mu_x}{\mu_y}\text{Cov}(X_i, Y_i)$$

$$= \sigma_x^2 - \frac{\mu_x^2}{\mu_y^2}\sigma_y^2 - 2\frac{\mu_x}{\mu_y}\rho\sigma_x\sigma_y$$

$$= \frac{1}{\mu_y^2}\left[\mu_y^2\sigma_x^2 + \mu_x^2\sigma_y^2 - 2\rho\mu_x\mu_y\sigma_x\sigma_y\right]$$

$$= \frac{S}{\mu_y^2} \quad \left[\text{Let, } S = \mu_y^2\sigma_x^2 + \mu_x^2\sigma_y^2 - 2\rho\mu_x\mu_y\sigma_x\sigma_y\right]$$

Now, using Univariate Central Limit Theorem (CLT) on these Z_i's, we obtain,

$$\sqrt{n}(\overline{Z_n} - 0) \xrightarrow{d} \mathcal{N}\left(0, \frac{S}{\mu_y^2}\right)$$

$$\text{i.e.,} \quad \sqrt{n}\bar{Z}_n \xrightarrow{d} \mathcal{N}\left(0, \frac{S}{\mu_y^2}\right)$$

Again, by using Weak Law of Large Numbers, it follows that,

$$\frac{1}{n}\sum_{i=1}^{n}Y_i \xrightarrow{P} \mu_y$$

$$\text{i.e.,} \quad \overline{Y_n} \xrightarrow{P} \mu_y, \text{ which is a constant.}$$

So, by Slutsky's Theorem, we have,

$$\sqrt{n}\left(\frac{\overline{Z_n}}{\overline{Y_n}}\right) \xrightarrow{d} \frac{W}{\mu_y} \quad , \text{ where } W \sim \mathcal{N}\left(0, \frac{S}{\mu_y^2}\right)$$

$$\text{i.e.,} \quad \sqrt{n}\left(\frac{\sum_{i=1}^{n}X_i}{\sum_{i=1}^{n}Y_i} - \frac{\mu_x}{\mu_y}\right) \xrightarrow{d} \mathcal{N}\left(0, \frac{\mu_y^2\sigma_x^2 + \mu_x^2\sigma_y^2 - 2\rho\mu_x\mu_y\sigma_x\sigma_y}{\mu_y^4}\right)$$

This is what we intended to show.

$\blacksquare$

5   Problem 5

> **Problem.** Let $X_1, X_2, \cdots, X_n$ be a random sample from a distribution with mean μ, variance σ^2 and finite 4th moment μ_4. A common test for $H_0 : \sigma^2 = 1$ versus $H_1 : \sigma^2 > 1$ rejects H_0 when
>
> $$nS_n^2 = \sum_{i=1}^{n} \left(X_i - \bar{X}_n \right)^2 > \chi^2_{\alpha, n-1}$$
>
> where $\chi^2_{\alpha, n-1}$ is the upper α point of a central chi-square distribution with $(n-1)$ degrees of freedom. Show that $\mathbb{P}_{H_0} \left(nS_n^2 > \chi^2_{\alpha, n-1} \right)$ converges to α as $n \to \infty$ only if the value of the kurtosis $\kappa = \dfrac{\mu_4}{\sigma^4} - 3$ is equal to zero. [**Hint**: First show (using the CLT and Polya's Theorem) that $\dfrac{\chi^2_{\alpha, n-1} - (n-1)}{\sqrt{2n-2}}$ converges to the upper α point z_α of the standard normal distribution.]

Solution: Let, $Y_1, Y_2, \ldots Y_{(n-1)}$ be i.i.d. random sample from a standard normal distribution. Then, $T_n = \sum_{i=1}^{(n-1)} Y_i^2$ has a chi-squared distribution with $(n-1)$ degrees of freedom.

Note that, $\mathbb{E}(T_n) = \sum_{i=1}^{(n-1)} \mathbb{E}(Y_i^2) = \sum_{i=1}^{(n-1)}(\text{Var}(Y_i^2) + \mathbb{E}(Y_i)^2) = (n-1)$. Also,

$$\text{Var}(T_n) = \sum_{i=1}^{(n-1)} \text{Var}(Y_i^2) = \sum_{i=1}^{(n-1)} (\mathbb{E}(Y_i^4) - \mathbb{E}(Y_i^2)^2) = (n-1)(3-1) = (2n-2)$$

Therefore, an application of Central Limit Theorem yields,

$$\frac{T_n - (n-1)}{\sqrt{2n-2}} \xrightarrow{d} Z \sim N(0, 1)$$

Now, choose any $\epsilon > 0$. Then by using Polya's theorem, it follows that for large n, say $\forall n \geq N_1$,

$$\left| \mathbb{P}\left(\frac{T_n - (n-1)}{\sqrt{2n-2}} \leq \frac{\chi^2_{\alpha,(n-1)} - (n-1)}{\sqrt{2n-2}} \right) - \mathbb{P}\left(Z \leq \frac{\chi^2_{\alpha,(n-1)} - (n-1)}{\sqrt{2n-2}} \right) \right| < \epsilon \qquad (5.1)$$

However, since $\chi^2_{\alpha,(n-1)}$ is the upper α-quantile of chi-squared distribution with $(n-1)$ degrees of freedom and T_n is a random variable following chi-squared distribution with $(n-1)$ degrees of freedom, we have;

$$\mathbb{P}\left(\frac{T_n - (n-1)}{\sqrt{2n-2}} \leq \frac{\chi^2_{\alpha,(n-1)} - (n-1)}{\sqrt{2n-2}} \right) = \mathbb{P}(T_n \leq \chi^2_{\alpha,(n-1)}) = (1 - \alpha) \qquad (5.2)$$

Since $(1 - \alpha) = \mathbb{P}(Z \leq z_\alpha)$, combining eq. (5.1) and eq. (5.2), we obtain;

9

$$\left| \mathbb{P}\left(Z \leq \frac{\chi^2_{\alpha,(n-1)} - (n-1)}{\sqrt{2n-2}} \right) - \mathbb{P}(Z \leq z_\alpha) \right| < \epsilon \qquad \forall n \geq N_1 \tag{5.3}$$

Letting $\Phi(\cdot)$ denote the cumulative distribution function of a standard normal distribution, eq. (5.3) means that,

$$\Phi\left(\frac{\chi^2_{\alpha,(n-1)} - (n-1)}{\sqrt{2n-2}} \right) \xrightarrow{n \to \infty} \Phi(z_\alpha)$$

Now, since $\Phi(\cdot)$ is strictly increasing and continuous, thus $\Phi^{-1}(\cdot)$ exists and is continuous. Applying $\Phi^{-1}(\cdot)$ to both sides of the above limit thus yields,

$$\frac{\chi^2_{\alpha,(n-1)} - (n-1)}{\sqrt{2n-2}} \xrightarrow{n \to \infty} z_\alpha \tag{5.4}$$

Let us now consider the centered samples, $Z_i = (X_i - \mu)$. Then,

$$\mathbb{E}(Z_i) = \mathbb{E}(X_i) - \mu = 0$$
$$\mathbb{E}(Z_i^2) = \mathrm{Var}(Z_i) + \mathbb{E}^2(Z_i) = \mathrm{Var}(X_i) + 0 = \sigma^2$$
$$\mathrm{Var}(Z_i^2) = \mathbb{E}(Z_i^4) - \mathbb{E}^2(Z_i^2) = \mu_4 - \sigma^4$$

Therefore, by using Central Limit Theorem (CLT) it follows that;

$$\sqrt{n} \frac{n^{-1} \sum_{i=1}^n Z_i^2 - \sigma^2}{\sqrt{\mu_4 - \sigma^4}} \xrightarrow{d} \mathcal{N}(0,1)$$

However, note that, $\sum_{i=1}^n Z_i^2 = nS_n^2 + n(\bar{X}_n - \mu)^2$. Hence, the above can be rewritten as;

$$\sqrt{n} \left[\frac{S_n^2 + (\bar{X}_n - \mu)^2 - \sigma^2}{\mu_4 - \sigma^4} \right] \xrightarrow{d} \mathcal{N}(0,1) \tag{5.5}$$

Also, by Chebyshev's inequality,

$$\mathbb{P}_{H_0}(\sqrt{n}(\bar{X}_n - \mu)^2 > \epsilon) = \mathbb{P}_{H_0}\left(|\bar{X}_n - \mu| > \frac{\epsilon^{1/2}}{n^{1/4}} \right)$$
$$\leq \frac{\mathrm{Var}_{H_0}(\bar{X}_n - \mu)}{\left(\epsilon^{1/2}/n^{1/4} \right)^2}$$
$$= \frac{\sigma^2}{n} \times \frac{\sqrt{n}}{\epsilon}$$
$$= \frac{\sigma^2}{\epsilon\sqrt{n}}$$
$$\to 0 \qquad \text{as, } n \to \infty$$

10

Thus, $\sqrt{n}(\bar{X}_n - \mu)^2 \xrightarrow{P} 0$ under H_0. From eq. (5.5), by an application of Slutsky's theorem, we obtain;

$$\sqrt{n}\left[\frac{S_n^2 + (\bar{X}_n - \mu)^2 - \sigma^2}{\sqrt{\mu_4 - \sigma^4}}\right] - \left[\frac{\sqrt{n}(\bar{X}_n - \mu)^2}{\sqrt{\mu_4 - \sigma^4}}\right] = \sqrt{n}\left[\frac{S_n^2 - \sigma^2}{\sqrt{\mu_4 - \sigma^4}}\right] \xrightarrow{d} \mathcal{N}(0,1)$$

i.e.

$$\sqrt{n}\frac{((S_n^2/\sigma^2) - 1)}{\sqrt{\kappa + 2}} \xrightarrow{d} \mathcal{N}(0,1) \tag{5.6}$$

where $\kappa = \left(\dfrac{\mu_4}{\sigma^4} - 3\right)$ is the population kurtosis.

Now, under $H_0 : \sigma^2 = 1$,

$$\frac{(nS_n^2 - (n-1))}{\sqrt{2(n-1)}} = \sqrt{n}\frac{((S_n^2/\sigma^2) - 1)}{\sqrt{\kappa + 2}} \times \frac{\sqrt{n}\sqrt{\kappa + 2}}{\sqrt{2(n-1)}} + \frac{1}{\sqrt{2(n-1)}} \tag{5.7}$$

As, $\dfrac{\sqrt{n}\sqrt{\kappa+2}}{\sqrt{2(n-1)}} \to \sqrt{\dfrac{\kappa+2}{2}}$ and $\dfrac{1}{\sqrt{2(n-1)}} \to 0$, combining eq. (5.6) and eq. (5.7), it follows that

$$\frac{(nS_n^2 - (n-1))}{\sqrt{2(n-1)}} \xrightarrow{d} \mathcal{N}\left(0, \frac{\kappa+2}{2}\right) \qquad \text{under } H_0 \tag{5.8}$$

Hence,

$$\mathbb{P}_{H_0}\left(nS_n^2 > \chi^2_{\alpha,(n-1)}\right) = \mathbb{P}_{H_0}\left(\frac{(nS_n^2 - (n-1))}{\sqrt{2(n-1)}} > \frac{(\chi^2_{\alpha,(n-1)} - (n-1))}{\sqrt{2(n-1)}}\right)$$

$$\to \mathbb{P}_{H_0}(Z > z_\alpha) \qquad \text{where } Z \sim \mathcal{N}\left(0, \frac{\kappa+2}{2}\right)$$

the limit follows from Polya's theorem and eq. (5.4)

$$= \mathbb{P}_{H_0}\left(\sqrt{\frac{2}{\kappa+2}}Z > \sqrt{\frac{2}{\kappa+2}}z_\alpha\right), \quad \text{where } \sqrt{\frac{2}{\kappa+2}}Z \sim \mathcal{N}(0,1)$$

Since the limit of a sequence of real numbers is unique, thus $P_{H_0}(nS_n^2 > \chi^2_{\alpha,(n-1)})$ converges to α as $n \to \infty$ only if $\sqrt{\dfrac{2}{\kappa+2}}z_\alpha$ is the upper α quantile of standard normal distribution. However, definition z_α is the upper α quantile and as standard normal distribution function is strictly increasing, each quantile is unique. Thus, the limit holds only if $\sqrt{\dfrac{2}{\kappa+2}} = 1$, i.e. $\kappa = 0$, i.e the kurtosis is equal to 0.

Therefore, it follows that $\mathbb{P}_{H_0}\left(nS_n^2 > \chi^2_{\alpha,(n-1)}\right) \to \alpha$ only if the kurtosis is equal to 0. ∎

6 Problem 6

> **Problem.** *Simulation experiment.*
>
> (a) For $n = 20$ and 30, draw 1000 random samples of size n from a bivariate normal distribution $\mathcal{N}_2(0, 0, 1, 1, \rho)$ for $\rho = 0.2, 0.4, 0.6, 0.8$. Tabulate the estimated coverage probabilities of the (approximate) confidence intervals for ρ formed using the $\tanh^{-1}$ transformation of the correlation coefficient.
>
> (b) Plot the histogram of the 1000 sample correlation coefficents r_n's for $n = 30$ and $\rho = 0.4$ and also the histogram of the corresponding $\tanh^{-1} r_n$'s.

Solution:

(a) We know that based on Fisher's $\tanh^{-1}$ transformation, an asymptotic distribution of the transformed sample correlation coefficient can be obtained as;

$$\sqrt{n}(\tanh^{-1}(r_n) - \tanh^{-1}(\rho)) \xrightarrow{d} \mathcal{N}(0, 1)$$

where r_n is the sample correlation coefficent based on a sample of size n. Thus, an approximate asymptotic 95% confidence interval is;

$$\left(\tanh\left(\tanh^{-1}(r_n) - \frac{z_{0.975}}{\sqrt{n}} \right), \tanh\left(\tanh^{-1}(r_n) + \frac{z_{0.975}}{\sqrt{n}} \right) \right)$$

To obtain the coverage probabilities, we consider B monte carlo simulations, and for each simulation, we generate random samples from bivariate random sample and check whether the above confidence interval contains the population correlation coefficent ρ.

First, we shall create a function for computing $\tanh^{-1}(\cdot)$.

```
tanhinv <- function(r) {
    return(1/2 * log((1+r)/(1-r)))
}
```

Next, we create a function that performs $B = 1000$ monte carlo simulations to estimate the coverage probabilities. For reproducibility, I have set a seed `1911` with `R version 3.6.3` and OS platform `x86_64-w64-mingw32` in Windows 10.

```
estimate.coverage <- function(n, rho) {
# set a seed for reproducibility
set.seed(1911)

# create the parameters
mu_vec <- c(0, 0)
sigma_mat <- matrix(c(1, rho, rho, 1), nrow = 2, byrow = TRUE)
coverage <- logical(1000)  # a logical array to see if the
coverage happens

# loop through each simulation
```

```
11      for (i in 1:1000) {
12          data <- MASS::mvrnorm(n = n, mu = mu_vec, Sigma = sigma_mat)
13          r <- cor(data[, 1], data[, 2])    # sample correlation
    coefficient
14          z <- tanhinv(r)  # obtain tanh inverse transformation of
    sample correlation
15          z.upper <- tanh(z + qnorm(0.975) / sqrt(n))
16          z.lower <- tanh(z - qnorm(0.975) / sqrt(n))
17
18          if (rho <= z.upper & rho >= z.lower) {
19              coverage[i] <- TRUE
20          }
21      }
22
23      # finally output the proportion of simulation where the coverage
    occurs
24      return(sum(coverage) / 1000)
25 }
```

The estimated coverage probabilities are shown in the following table.

Sample size n	$\rho = 0.2$	$\rho = 0.4$	$\rho = 0.6$	$\rho = 0.8$
20	0.936	0.935	0.938	0.936
30	0.932	0.931	0.933	0.931

(b) The following function (very similar to above) can be used to save all sample correlations r_n for a given n and ρ.

```
1 gen.r <- function(n, rho) {
2      # set a seed for reproducibility
3      set.seed(1911)
4
5      # create the parameters
6      muVec <- c(0, 0)
7      sigmaMat <- matrix(c(1, rho, rho, 1), nrow = 2, byrow = TRUE)
8
9      r <- sapply(1:1000, FUN = function(i) {
10          data <- MASS::mvrnorm(n = n, mu = muVec, Sigma = sigmaMat)
11          return(cor(data[, 1], data[, 2]))
12      })
13
14      return(r)
15 }
16
17 # now, call the above function for n = 30 and rho = 0.4 to get sample
    correlations
18 r <- gen.r(30, 0.4)
19 z <- tanhinv(r)   # get tanh inverse transformation of those sample
    correlations
```

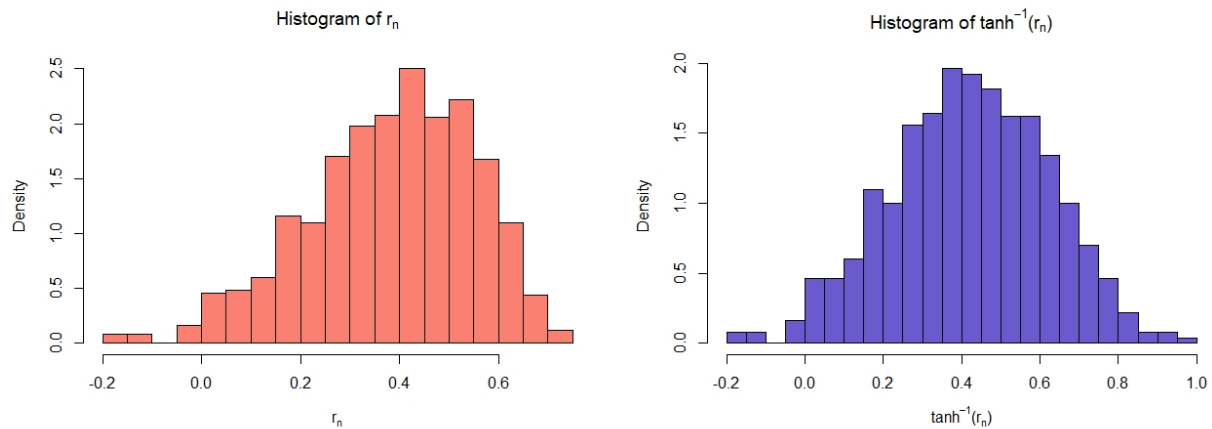Now, plotting the histograms result in the fig. 1.
```

13

Figure 1: Histogram (density based, not frequency based) of sample correlation coefficent and its transformation

As it seems from fig. 1, the distribution of sample correlation coefficient $r_n$ is more positively skewed, while the distribution of the transformed sample correlation coefficent $\tanh^{-1}(r_n)$ is more symmetric in nature, thus resembling the normal distribution better than the untransformed one.

∎

# 7 Problem 7

> **Problem.** Let $X_1, X_2, \cdots, X_n$ be a random sample from a double exponential distribution with density
> $$f(x, \theta) = \frac{1}{2} e^{-|x-\theta|}, \quad -\infty < x < \infty$$
> where $\theta \in \mathbb{R}$ is an unknown location parameter.
>
> (a) Find the maximum likelihood estimator (MLE) of $\theta$ and find the asymptotic (non-degenerate) distribution of properly normalized MLE.
>
> (b) Find the limiting non-degenerate distribution of the sample interquartile range (i.e., the difference between the third and first sample quartiles).

*Solution:*

(a) Note that the likelihood function is given by;

$$\mathcal{L}_n(\theta; X_1, X_2, \ldots X_n) = \prod_{i=1}^{n} f(X_i, \theta) = \frac{1}{2^n} e^{-\sum_{i=1}^{n} |X_i - \theta|}$$

Therefore, to maximize the likelihood $\mathcal{L}_n$, is same as maximizing the log-likelihood $\ell_n = \log \mathcal{L}_n = -n \log 2 - \sum_{i=1}^{n} |X_i - \theta|$, i.e. same as minimizing $\sum_{i=1}^{n} |X_i - \theta|$. It is known that, this minimization happens if $\widehat{\theta}_n = X_M$, the sample median[1].

Specifically, the minimization does not lead to a unique minima in case when $n$ is even. In that case, any value between $X_{([n/2])}$ and $X_{([(n+1)/2])}$ is MLE, where $X_{(k)}$ is the $k$-th order statistic.

Now, to find the asymptotic non-degenerate distribution of properly normalized MLE, we use the result on asymptotic distribution of sample quantile along with Representation theorem.

Let us choose, $p_n \in (0, 1)$ such that, $\left[\frac{n}{2}\right]/n \leq p_n \leq \left[\frac{n+1}{2}\right]/n$. Let, $p = 1/2$. Then, $|p_n - p| \leq \frac{1}{n} = O(n^{-1/2})$. Let, $X'_{p_n}$ be the $p_n$-th sample quantile. Here, the distribution function $F$ is continuous, differentiable and has a density $f(x, \theta)$ as specified in the question. Also note that,

$$\int_{-\infty}^{\theta} f(x, \theta) = \frac{1}{2} \int_{-\infty}^{\theta} e^{-|x-\theta|} dx = \frac{1}{2} \int_{-\infty}^{0} e^{-z} dz = \frac{1}{2}$$

Thus, the population median is $\xi_{1/2} = \theta$. Finally, at $x = \xi_{1/2}$, the density takes value $f(\xi_{1/2}, \theta) = \frac{1}{2} e^{-|\theta-\theta|} = \frac{1}{2}$, which is positive.

---

[1]This is true as one can group observations together like $\sum_{k=1}^{[n/2]} |X_{(k)} - \theta| + |X_{(n-k)} - \theta|$ and for each such pair the sum is minimum if $\theta \in (X_{(k)}, X_{(n-k)})$ by triangle inequality

Therefore, by the result on asymptotic distribution of sample quantile if follows that;

$$\sqrt{n}(X'_{p_n} - a_n) \xrightarrow{d} \mathcal{N}\left(0, \frac{1/4}{f(\xi_{1/2})^2}\right) \equiv \mathcal{N}(0, 1) \tag{7.1}$$

where $a_n = \xi_{1/2} + \dfrac{(p_n - 1/2)}{f(\xi_{1/2})}$. However, $a_n = \theta + 2\left[(p_n - 1/2)\right] \to \theta$, as $|p_n - 1/2| \le 1/n \to 0$ with $n \to \infty$. Therefore, by an application of Polya's theorem, it follows from eq. (7.1) that,

$$\sqrt{n}(X'_{p_n} - \theta) \xrightarrow{d} \mathcal{N}(0, 1) \tag{7.2}$$

Since it holds for any $p_n$ satisfying $\left[\dfrac{n}{2}\right]/n \le p_n \le \left[\dfrac{n+1}{2}\right]/n$, it holds for any MLE.

Thus, the properly normalized MLE $\sqrt{n}(\widehat{\theta}_n - \theta)$ is asymptotically normally distributed with mean 0 and variance 1.

This completes the first part.

(b) Let us first obtain the population first and third quantiles, denoted by $\xi_1$ and $\xi_3$ respectively.

Considering $\int_{-\infty}^{\xi_1} f(x, \theta)dx = 1/4$ yields,

$$\int_{-\infty}^{\xi_1} \frac{1}{2}e^{-|x-\theta|}dx = \frac{1}{4}$$
$$\Rightarrow \int_{-\infty}^{(\xi_1-\theta)} e^{-|z|}dz = \frac{1}{2} \quad \text{with } z = (x - \theta)$$
$$\Rightarrow e^{\xi_1-\theta} = 1/2$$
$$\Rightarrow \xi_1 = \theta - \log 2$$

Since the given distribution is symmetric with respect to location parameter $\theta$, $\xi_3 = \theta + \log 2$. Note that,

$$f(\xi_1, \theta) = f(\xi_3, \theta) = \frac{1}{2}e^{-|\log 2|} = \frac{1}{2}e^{\log(1/2)} = \frac{1}{4}$$

By using multivariate CLT and Representation theorem, we have the following result.

Let, $0 < p_1 < p_2 < \cdots < p_k$ and let $p_{in} = p_i + O\left(n^{-1/2}\right)$, with $F'(\xi_{p_i}) > 0 \quad \forall i$, where $\xi_p$ is the $p$-th quantile of the distribution function $F$. Then,

$$U_n = \left(\sqrt{n}(Y_{p_{1n},n} - a_{1n}), \sqrt{n}(Y_{p_{2n},n} - a_{2n}), \ldots \sqrt{n}(Y_{p_{kn},n} - a_{kn})\right)$$

16

is asymptotically normal with mean vector 0 and dispersion matrix $\Sigma = ((\sigma_{ij}))_{i,j=1}^k$ where $\sigma_{ij} = \sigma_{ji} = \dfrac{p_i(1-p_j)}{F'(\xi_{p_i})F'(\xi_{p_j})}$ for $i \leq j$, where $Y_{p,n}$ is the sample $p$-th quantile based on a sample of size $n$.

Using this result for $p_1 = 1/4$ and $p_2 = 3/4$ only, and denoting the first sample quartile as $Q_{1,n}$ and the third sample quartile as $Q_{3,n}$, we get that,

$$\sqrt{n}\left(\begin{bmatrix} Q_{1,n} \\ Q_{3,n} \end{bmatrix} - \begin{bmatrix} \xi_1 \\ \xi_3 \end{bmatrix}\right) \xrightarrow{d} \mathcal{MVN}\left(\mathbf{0}_2, \Sigma\right) \tag{7.3}$$

where,

$$\Sigma = \begin{bmatrix} \dfrac{1/4 \times 3/4}{f(\xi_1)^2} & \dfrac{(1/4)^2}{f(\xi_1)f(\xi_3)} \\[4mm] \dfrac{(1/4)^2}{f(\xi_1)f(\xi_3)} & \dfrac{3/4 \times 1/4}{f(\xi_3)^2} \end{bmatrix}$$

$$= \begin{bmatrix} (3/16)/(1/16) & (1/16)/(1/16) \\ (1/16)/(1/16) & (3/16)/(1/16) \end{bmatrix}$$

$$= \begin{bmatrix} 3 & 1 \\ 1 & 3 \end{bmatrix}$$

Now, consider the function $g(x,y) = (y - x)$, with $g : \mathbb{R}^2 \to \mathbb{R}$ being a differentiable function. Note that, $g\left(\begin{bmatrix} Q_{1,n} \\ Q_{3,n} \end{bmatrix}\right) = Q_{3,n} - Q_{1,n}$, the sample interquartile range, while $g\left(\begin{bmatrix} \xi_1 \\ \xi_3 \end{bmatrix}\right) = \xi_3 - \xi_1 = 2\log 2 = \log 4$. Thus an application of multivariate delta method to eq. (7.3) yields;

$$\sqrt{n}((Q_{3,n} - Q_{1,n}) - \log 4) \xrightarrow{d} \mathcal{N}\left(0, (\nabla g(\xi_1, \xi_3))^\intercal \Sigma (\nabla g(\xi_1, \xi_3))\right) \tag{7.4}$$

With $g(x,y) = (y - x)$, $\nabla g(x,y) = (-1, 1)^\intercal$ for any $x, y \in \mathbb{R}$. Therefore,

$$(\nabla g(\xi_1, \xi_3))^\intercal \Sigma (\nabla g(\xi_1, \xi_3)) = (3 - 2 + 3) = 4$$

Finally denoting $(Q_{3,n} - Q_{1,n}) = IQR_n$ we obtain from eq. (7.4) that,

$$\sqrt{n}(IQR_n - \log 4) \xrightarrow{d} \mathcal{N}(0, 4)$$

This completes part (b).

$\blacksquare$

17

# 8   Problem 8

> **Problem.** Let $X_1, X_2, \cdots, X_n$ be a random sample from $\mathcal{N}(\theta, 1)$. Find the joint asymptotic distribution of suitably normalized sample mean and sample median.

*Solution:*   Let, $\bar{X}_n = \dfrac{1}{n}\sum_{i=1}^n X_i$, be the sample mean. Since, $X_i \sim \mathcal{N}(\theta, 1)$, hence $\bar{X}_n \sim \mathcal{N}(\theta, 1/n)$.

Let us denote the sample median of $X_i$'s based on as $X_{Me,n}$.

Also, $\bar{X}_n$ is a minimal sufficent statistic for $\theta$, while $Y_i = (X_i - \bar{X}_n)$ are ancillary for $\theta$, in case of normal distribution. Thus, by Basu's theorem, it follows that $\bar{X}_n$ is independent of $Y_i$'s. Thus, $\bar{X}_n$ is also independent of the median of $Y_i$'s, say $Y_{Me,n}$. However, the sample median of $X_i$'s is simply $X_{Me,n} = Y_{Me,n} + \bar{X}_n$.

Therefore, we shall proceed by finding the joint asymptotic distribution of $\bar{X}_n$ and $Y_{Me,n}$ first. Since, $Y_i$'s are not independent of each other, it is not possible to obtain the asymptotic distribution of sample median of $Y_i$'s simply by direct use of the theorem regarding asymptotic distribution of sample quantiles.

We shall use **Levy's Continuity Theorem**[2] to find this joint asymptotic distribution.

Note that, by the result on asymptotic distribution of sample median, it follows that $\sqrt{n}(X_{Me,n} - \theta)$ is asymptotically normally distributed with mean 0 and variance $\dfrac{1/4}{\phi(\theta)^2}$, where $\phi(\cdot)$ is the density function of the normal distribution with mean $\theta$ and variance 1. Clearly, $\phi(\theta) = \dfrac{1}{\sqrt{2\pi}}$. Thus we have,

$$\sqrt{n}(X_{Me,n} - \theta) \xrightarrow{d} \mathcal{N}\left(0, \frac{1/4}{1/2\pi}\right) \equiv \mathcal{N}\left(0, \frac{\pi}{2}\right) \tag{8.1}$$

It is known that the characteristic function of a random variable normally distributed with mean $\theta$ and variance $\sigma^2$ is given as $\exp\left[i\mu t - \dfrac{\sigma^2 t^2}{2}\right]$. Therefore, it follows by Levy's Continuity Theorem,

$$\lim_{n \to \infty} \psi_{\sqrt{n}(X_{Me,n} - \theta)}(t) = e^{-\frac{\pi}{4}t^2} \tag{8.2}$$

On the other hand, $\sqrt{n}(X_{Me,n} - \theta) = \sqrt{n}(\bar{X}_n - \theta) + \sqrt{n}Y_{Me,n}$ and as noted above, these two quantities are independent. Thus, $\psi_{\sqrt{n}(X_{Me,n} - \theta)}(t) = \psi_{\sqrt{n}(\bar{X}_n - \theta)}(t) \times \psi_{\sqrt{n}Y_{Me,n}}(t)$ for any $t \in \mathbb{R}$. Since, $\sqrt{n}(\bar{X}_n - \theta) \sim \mathcal{N}(0, 1)$, it follows that $\psi_{\sqrt{n}(\bar{X}_n - \theta)}(t) = e^{-t^2/2}$. Therefore,

---

[2]**Statement of the theorem:** A sequence of random variables $X_n$ converges in distribution to the random variable $X$ if and only if the sequence of its characteristic function $\psi_{X_n}(t)$ converges pointwise to a function $\psi(t)$ which is continuous at $t = 0$, and in that case $\psi$ is the characteristic function of $X$

$$\lim_{n \to \infty} \psi_{\sqrt{n} Y_{Me,n}}(t) = \lim_{n \to \infty} \frac{\psi_{\sqrt{n}(X_{Me,n}-\theta)}(t)}{\psi_{\sqrt{n}(\bar{X}_n-\theta)}(t)}$$

$$= \frac{\lim_{n \to \infty} \psi_{\sqrt{n}(X_{Me,n}-\theta)}(t)}{e^{-t^2/2}} \qquad \text{, since the individual limit exists}$$

and denominator is nonzero

$$= e^{-t^2(\frac{\pi}{2}-1)/2}$$

Clearly, this is the characteristic function of a normally distributed random variable with mean 0 and variance $\left(\frac{\pi}{2} - 1\right)$, and this is continuous at $t = 0$. Hence, by application of Levy's Continuity theorem, it follows that;

$$\sqrt{n} Y_{Me,n} \xrightarrow{d} \mathcal{N}\left(0, \left(\frac{\pi}{2} - 1\right)\right)$$

Now since $\bar{X}_n$ is independent of $Y_{Me,n}$, hence any linear combination, namely for any $a, b \in \mathbb{R}$, we have,

$$a\sqrt{n}(\bar{X}_n - \theta) + b\sqrt{n} Y_{Me,n} \xrightarrow{d} \mathcal{N}\left(0, a + b\left(\frac{\pi}{2} - 1\right)\right) \tag{8.3}$$

Thus applying Cramer Wold theorem to eq. (8.3), we obtain the required joint asymptotic distribution as;

$$\sqrt{n}\left(\begin{bmatrix} \bar{X}_n \\ Y_{Me,n} \end{bmatrix} - \begin{bmatrix} \theta \\ 0 \end{bmatrix}\right) \xrightarrow{d} \mathcal{MVN}\left(\mathbf{0}_2, \begin{bmatrix} 1 & 0 \\ 0 & (\pi/2 - 1) \end{bmatrix}\right) \tag{8.4}$$

Now, consider the function $g : \mathbb{R}^2 \to \mathbb{R}^2$ such that $g(x, y) = (x, x+y)$. Then, the jacobian of the function is $J_g(x, y) = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$. Also, $g(\bar{X}_n, Y_{Me,n}) = (\bar{X}_n, \bar{X}_n + Y_{Me,n}) = (\bar{X}_n, X_{Me,n})$, and $g(\theta, 0) = (\theta, \theta)$.

Therefore, by an application of multivariate delta method to eq. (8.4), it follows that;

$$\sqrt{n}\left(\begin{bmatrix} \bar{X}_n \\ X_{Me,n} \end{bmatrix} - \begin{bmatrix} \theta \\ \theta \end{bmatrix}\right) \xrightarrow{d} \mathcal{MVN}\left(\mathbf{0}_2, J_g(\theta, 0)^\top \begin{bmatrix} 1 & 0 \\ 0 & (\pi/2 - 1) \end{bmatrix} J_g(\theta, 0)\right)$$

Note that,

$$J_g(\theta, 0)^\top \begin{bmatrix} 1 & 0 \\ 0 & (\pi/2 - 1) \end{bmatrix} J_g(\theta, 0) = \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & (\pi/2 - 1) \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$$

$$= \begin{bmatrix} 1 & 0 \\ 1 & (\pi/2 - 1) \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} \qquad = \begin{bmatrix} 1 & 1 \\ 1 & \pi/2 \end{bmatrix}$$

Therefore, a suiatably transformed and scaled joint asymptotic distribution of sample mean and sample median is given by,

$$\sqrt{n}\left(\begin{bmatrix} \bar{X}_n \\ X_{Me,n} \end{bmatrix} - \begin{bmatrix} \theta \\ \theta \end{bmatrix}\right) \xrightarrow{d} \mathcal{MVN}_2\left(\mathbf{0}_2, \begin{bmatrix} 1 & 1 \\ 1 & \pi/2 \end{bmatrix}\right)$$

∎

# 9   Problem 9

> **Problem.** Let $X_1, X_2, \cdots, X_n$ be a random sample from $\mathcal{N}(\theta, 1)$ where $\theta$ is an un-known integer. Find the maximum likelihood estimator (MLE) $\hat{\theta}_n$ of $\theta$. Prove that it is not possible to find a sequence of real constants $a_n$ such that $a_n\left(\hat{\theta}_n - \theta\right)$ converges to a non-degenerate limit distribution.

*Solution:*

We start by writing the likelihood function,

$$\mathcal{L}(\theta) = \frac{1}{(2\pi)^{(n/2)}} \exp\left[-\frac{1}{2}\sum_{i=1}^n (X_i - \theta)^2\right] \mathbf{1}_{\{\theta \in \mathbb{Z}\}}$$

where $\mathbf{1}_A$ is the indicator function of the set $A$ and $\mathbb{Z}$ is the set of integers. Note that, the likelihood can be expressed as,

$$\mathcal{L}(\theta) = \frac{1}{(2\pi)^{(n/2)}} \exp\left[-\frac{1}{2}\sum_{i=1}^n (X_i - \bar{X}_n)^2 - \frac{n}{2}(\bar{X}_n - \theta)^2\right] \mathbf{1}_{\{\theta \in \mathbb{Z}\}}$$

The likelihood function takes nonzero value if and only if $\theta \in \mathbb{Z}$ and is decreasing in the magnitude of $|\bar{X}_n - \theta|$ restricted to $\theta \in \mathbb{Z}$. Thus, the MLE of $\theta$ would be equal to the nearest integer of $\bar{X}_n$, i.e.

$$\widehat{\theta}_n = \begin{cases} \lceil \bar{X}_n \rceil & \text{if } \{\bar{X}_n\} > 1/2 \\ \lfloor \bar{X}_n \rfloor & \text{if } \{\bar{X}_n\} \le 1/2 \end{cases}$$

where $\lceil \cdot \rceil$ is the ceiling function, $\lfloor \cdot \rfloor$ is the floor function and $\{x\}$ is the fractional part of $x$. In other words, $\widehat{\theta}_n$ is the unique integer in the interval $[\bar{X}_n - 1/2, \bar{X}_n + 1/2)$.

Now let us consider the probability under $\theta$ that the MLE is not same as the true $\theta \in \mathbb{Z}$,

$$
\begin{aligned}
\mathbb{P}_\theta(\widehat{\theta}_n \ne \theta) &= \mathbb{P}_\theta\left(\text{Closest integer to } \bar{X}_n \ne \theta\right) \\
&= \mathbb{P}_\theta\left(\theta \notin [\bar{X}_n - 1/2, \bar{X}_n + 1/2)\right) \\
&= \mathbb{P}_\theta\left(\{\bar{X}_n > \theta + 1/2\} \cup \{\bar{X}_n \le \theta - 1/2\}\right) \\
&= \mathbb{P}_\theta(\bar{X}_n > \theta + 1/2) + \mathbb{P}_\theta(\bar{X}_n \le \theta - 1/2), \text{ as two events are disjoint} \\
&= 2\mathbb{P}_\theta(\sqrt{n}(\bar{X}_n - \theta) > \sqrt{n}/2) \\
&= 2\left(1 - \Phi(\sqrt{n}/2)\right), \qquad \text{since } \sqrt{n}(\bar{X}_n - \theta) \sim \mathcal{N}(0, 1)
\end{aligned}
$$

where $\Phi(\cdot)$ is the cumulative distribution function of a standard normal random variable.

Now, we shall prove a tail bound for this probability. Let, $Z$ be a standard normal random variable and $\phi(\cdot)$ be the density function for standard normal distribution. Then, for $t > 0$,

$$\mathbb{E}(e^{t(Z-a)}) = \int_{-\infty}^{\infty} e^{t(z-a)}\phi(z)dz$$

$$\geq \int_a^{\infty} e^{t(z-a)}\phi(z)dz$$

$$\geq \int_a^{\infty} \phi(z)dz, \text{ since } e^{t(z-a)} \geq 1 \text{ for } z \geq a, t > 0$$

$$= (1 - \Phi(a))$$

Therefore,

$$(1 - \Phi(a)) \leq \inf_{t>0} \mathbb{E}(e^{t(Z-a)}) = \inf_{t>0} e^{-at}\mathbb{E}(e^{tZ}) = \inf_{t>0} e^{-at}M_Z(t)$$

where $M_Z(t)$ is the moment generating function for $Z$. However, it is known that for standard normal distribution, the moment generating function $M_Z(t) = e^{t^2/2}$. Thus, it follows that;

$$(1 - \Phi(a)) \leq \inf_{t>0} e^{t^2/2 - at} = e^{-a^2/2} \tag{9.1}$$

with the infimum being achieved at $t = a$. Therefore, applying eq. (9.1) as a tail bound, we obtain that;

$$\mathbb{P}_\theta(\widehat{\theta}_n \neq \theta) \leq 2\exp\left[-\frac{(\sqrt{n}/2)^2}{2}\right] = 2\exp\left[-\frac{n^2}{8}\right] \tag{9.2}$$

Hence, it follows from eq. (9.2),

$$\sum_{n=1}^{\infty} \mathbb{P}_\theta(\widehat{\theta} \neq \theta) \leq \sum_{n=1}^{\infty} 2\exp\left[-\frac{n^2}{8}\right]$$

$$\leq 2\sum_{n=1}^{\infty} \exp\left[-\frac{n}{8}\right] \qquad \text{as, } e^{-n^2/8} \geq e^{-n/8} \quad \forall n \geq 1$$

$$= 2\sum_{n=1}^{\infty} \left(e^{-1/8}\right)^n$$

$$= 2\frac{e^{-1/8}}{1 - e^{-1/8}} < \infty \qquad \text{as } e^{-1/8} < 1$$

Therefore, by Borel Cantelli Lemma, the event has $\widehat{\theta}_n \neq \theta$ happens infinitely many often has probability equal to 0. In other words, the event $\widehat{\theta}_n = \theta$ happens almost surely after some finite time i.e. for all $n \geq N$ for some large $N$. This essentially means, for all $n \geq N$,

$$\mathbb{P}_\theta\left(a_n(\widehat{\theta}_n - \theta) = 0\right) \leq \mathbb{P}_\theta\left((\widehat{\theta}_n - \theta) = 0\right) = 1$$

i.e. it is not possible to find a sequence of real constants $a_n$ such that $a_n(\widehat{\theta}_n - \theta)$ converges to a non-degenerate limit distribution. ∎

# 10 Problem 10

> **Problem.** Let $X_1, X_2, \cdots, X_n$ be i.i.d. with a common density $f(x, \theta)$, where $\theta \in \Theta$ and $\Theta$ consists of only finitely many real numbers. Assume also that the set $\{x : f(x, \theta) > 0\}$ is the same for all $\theta \in \Theta$ and that the distributions under different $\theta$'s are different. If $\theta_0$ is the true value of $\theta$, prove that with probability tending to one (under $\theta_0$) as $n \to \infty$, the likelihood function will be maximized at the value $\theta = \theta_0$.

*Solution:* We are given that $X_1, X_2, \cdots, X_n$ be i.i.d. with a common density $f(x, \theta)$, where $\theta \in \Theta$ and $\Theta$ consists of only finitely many real numbers. Let us assume, $\Theta = \{\theta_0, \theta_1, \ldots \theta_k\}$ among which $\theta_0$ is the true parameter.

We start by considering the likelihood ratio,

$$
\begin{aligned}
LR(\theta, \theta_0) &= \frac{\mathcal{L}(\theta; X_1, X_2, \ldots X_n)}{\mathcal{L}(\theta_0; X_1, X_2, \ldots X_n)} \\
&= \frac{\prod_{i=1}^n f(X_i, \theta)}{\prod_{i=1}^n f(X_i, \theta_0)} \\
&= \prod_{i=1}^n \frac{f(X_i, \theta)}{f(X_i, \theta_0)} \\
\Rightarrow \quad \log LR(\theta, \theta_0) &= \sum_{i=1}^n \log\left(\frac{f(X_i, \theta)}{f(X_i, \theta_0)}\right)
\end{aligned}
$$

Since, the set $A = \{x : f(x, \theta) > 0\}$ is same for all $\theta \in \Theta$, and the ratios $f(x, \theta)/f(x, \theta_0)$ and its logarithm make sense for any $x \in A$, the above individual logarithms make sense as $X_i \in A$ for all $i = 1, 2, \ldots n$.

Now, since $X_i$'s are i.i.d., so is the ratio $f(X_i, \theta)/f(X_i, \theta_0)$ and its natural logarithm. Hence, by Weak Law of Large Numbers,

$$
\frac{1}{n}\sum_{i=1}^n \log\left(\frac{f(X_i, \theta)}{f(X_i, \theta_0)}\right) \xrightarrow{P} \mathbb{E}_{\theta_0}\left(\log\left(\frac{f(X_1, \theta)}{f(X_1, \theta_0)}\right)\right) \qquad \text{as, } \theta_0 \text{ is assumed to be true parameter}
$$

Now, by Jensen's inequality,

$$
\begin{aligned}
\mathbb{E}_{\theta_0}\left[\log\left(\frac{f(X_1, \theta)}{f(X_1, \theta_0)}\right)\right] &\leq \log\left(\mathbb{E}_{\theta_0}\left[\frac{f(X_1, \theta)}{f(X_1, \theta_0)}\right]\right) \\
&= \log\left(\int_{x \in A} \frac{f(x, \theta)}{f(x, \theta_0)} f(x, \theta_0) dx\right) \\
&= \log 1 \quad \text{since, for } x \in A, f(x, \theta_0) \neq 0 \\
&= 0
\end{aligned}
$$

with equality if and only if $f(x, \theta_0) = f(x, \theta)$ for all $x \in A$, where $A$ is the common support of both of these distributions. However, as the distributions under different $\theta$'s are different (i.e. identifiability condition), this yields that in the above inequality, equality will hold if and only if $\theta = \theta_0$.

Thus, for any $\theta \neq \theta_0$,

$$\mathbb{P}_{\theta_0}\left(\frac{1}{n}\log LR(\theta, \theta_0) < 0\right) \to 1$$

$$\Rightarrow \quad \mathbb{P}_{\theta_0}\left(LR(\theta, \theta_0) < 1\right) \to 1$$

$$\Rightarrow \quad \mathbb{P}_{\theta_0}\left(\mathcal{L}_n(\theta) < \mathcal{L}_n(\theta_0)\right) \to 1 \qquad \text{where, } \mathcal{L}_n(\theta) = \mathcal{L}(\theta; X_1, X_2, \ldots X_n)$$

$$\Rightarrow \quad \mathbb{P}_{\theta_0}\left(\mathcal{L}_n(\theta) \geq \mathcal{L}_n(\theta_0)\right) \to 0$$

Therefore, for a given $\epsilon > 0$, for $\theta_i$, $i = 1, 2, \ldots k$, there exists $N_i$ such that for all $n \geq N_i$,

$$\mathbb{P}_{\theta_0}(\mathcal{L}_n(\theta_i) \geq \mathcal{L}_n(\theta_0)) < \frac{\epsilon}{k} \tag{10.1}$$

Hence, for all $n \geq N = \max\{N_1, N_2, \ldots N_k\}$ we must have;

$$
\begin{aligned}
\mathbb{P}_{\theta_0}\left(\text{Likelihood is maximized at } \theta = \theta_0\right) &= \mathbb{P}_{\theta_0}\left(\mathcal{L}_n(\theta_0) > \mathcal{L}_n(\theta_i); \forall i = 1, 2, \ldots k\right) \\
&= \mathbb{P}_{\theta_0}\left(\cap_{i=1}^{k}\{\mathcal{L}_n(\theta_0) > \mathcal{L}_n(\theta_i)\}\right) \\
&= 1 - \mathbb{P}_{\theta_0}\left(\cup_{i=1}^{k}\{\mathcal{L}_n(\theta_0) \leq \mathcal{L}_n(\theta_i)\}\right) \\
&\geq 1 - \sum_{i=1}^{k} \mathbb{P}_{\theta_0}\left(\mathcal{L}_n(\theta_0) \leq \mathcal{L}_n(\theta_i)\right) \\
&\geq 1 - \sum_{i=1}^{k} \epsilon/k \qquad \text{by eq. (10.1)} \\
&= (1 - \epsilon)
\end{aligned}
$$

Therefore, $\mathbb{P}_{\theta_0}\left(\text{Likelihood is maximized at } \theta = \theta_0\right) \to 1$ as $n \to \infty$.

■

# 11 Problem 11

**Problem.** Let $X_1, X_2, \cdots, X_n$ be i.i.d. with a common density $f(x, \theta)$ given by,

$$f(x, \theta) = \frac{e^{-(x-\theta)}}{\left[1 + e^{-(x-\theta)}\right]^2}, \quad -\infty < x < \infty,$$

where $\theta \in \mathbb{R}$ is unknown. Assume that Conditions (A1)–(A7) of the section on MLE hold.

(a) Does there exist a consistent maximum likelihood estimator of $\theta$ for this problem? Justify your answer.

(b) Can you suggest any sequence of functions $g_n(\cdot)$ (depending on $n$) of $\bar{X}_n$ such that $g_n(\bar{X}_n)$ (after suitable centering and scaling) converges in distribution to $\mathcal{N}\left(0, \frac{1}{I(\theta)}\right)$ where $I(\theta)$ is the Fisher Information? Justify your answer.

*Solution:*

(a) We start by considering the likelihood;

$$\mathcal{L}_n(\theta) = \prod_{i=1}^{n} f(X_i, \theta) = \prod_{i=1}^{n} \frac{e^{-(X_i-\theta)}}{\left[1 + e^{-(X_i-\theta)}\right]^2}$$

Therefore, the log-likelihood is;

$$\log \mathcal{L}_n(\theta) = \ell_n(\theta) = \sum_{i=1}^{n} (\theta - X_i) - 2\log\left[1 + e^{-(X_i-\theta)}\right]$$

Since the log-likelihood is a continuous function of $\theta$ for all $\theta \in \mathbb{R}$, the maximum likelihood estimator can be easily obtained by setting the derivative of the log-likelihood equal to 0, thus considering a likelihood equation.

$$\frac{d\ell_n}{d\theta} = 0 \implies n - 2\sum_{i=1}^{n} \frac{e^{-(X_i-\theta)}}{1 + e^{-(X_i-\theta)}} = 0$$

$$\implies n - 2\sum_{i=1}^{n}\left[1 - \frac{1}{1 + e^{-(X_i-\theta)}}\right] = 0$$

$$\implies n - 2n + 2\sum_{i=1}^{n} \frac{1}{1 + e^{-(X_i-\theta)}} = 0$$

$$\implies \sum_{i=1}^{n} \frac{1}{1 + e^{-(X_i-\theta)}} = \frac{n}{2}$$

In the last equation, the left hand side is sum of functions of type $1/(1+e^{-(x-\theta)})$, which is a strictly decreasing, continuous function in $\theta$. Also, as $\theta \to \infty$, $1/(1+e^{-(x-\theta)}) \to 0$ for any $x \in \mathbb{R}$, thus the left hand side goes to 0. On the other hand, as $\theta \to (-\infty)$, $1/(1+e^{-(x-\theta)}) \to 1$ for any $x \in \mathbb{R}$, thus the left hand side goes to $n$. Therefore, for any $n \in \mathbb{N}$, the left hand side strictly decreases continuously from $n$ to 0 as $\theta$ moves from $(-\infty)$ to $\infty$, showing that there must exists a unique $\widehat{\theta}_n$ such that the above equation holds, i.e. there must exist a unique solution to the likelihood equation.

Clearly, this unique solution must be the unique MLE for the given setup. Now, since assumptions (A1)-(A3) hold, by theorem 9.1 of the notes it follows that, for any $\delta > 0$, where the true parameter is $\theta_0$,

$$\mathbb{P}_{\theta_0}\left[\frac{\partial \log \mathcal{L}_n}{\partial \theta} = 0 \text{ has a solution in } (\theta_0 - \delta, \theta_0 + \delta)\right] \to 1 \qquad (11.1)$$

However, since the solution to the likelihood equation is found to be unique, it must correspond to $\widehat{\theta}_n$. Therefore, by eq. (11.1) it follows that for any $\delta > 0$,

$$\mathbb{P}_{\theta_0}\left[|\widehat{\theta}_n - \theta_0| < \delta\right] \to 1$$

i.e.

$$\mathbb{P}_{\theta_0}\left[|\widehat{\theta}_n - \theta_0| \geq \delta\right] \to 0$$

This shows that the MLE is consistent in this case, i.e. $\widehat{\theta}_n \xrightarrow{P} \theta$.

(b) Starting from the quantity $\dfrac{d\ell_n}{d\theta}$, we have;

$$\frac{d^2\ell_n}{d\theta^2} = -2\sum_{i=1}^{n} \frac{d}{d\theta}\left[1 - \frac{1}{(1+e^{-(X_i-\theta)})}\right] = -2\sum_{i=1}^{n} \frac{e^{-(X_i-\theta)}}{(1+e^{-(X_i-\theta)})^2}$$

Therefore,

$$\mathcal{I}_n(\theta) = \mathbb{E}\left[-\frac{d^2\ell_n}{d\theta^2}\right]$$

$$= 2n\mathbb{E}\left[\frac{e^{-(X_1-\theta)}}{(1+e^{-(X_1-\theta)})^2}\right] \qquad \text{since } X_i\text{'s are i.i.d.}$$

$$= 2n\int_{-\infty}^{\infty} \frac{e^{-2(x-\theta)}}{(1+e^{-(x-\theta)})^4}dx$$

Substituting $z = (1+e^{-(x-\theta)})$ yields,

25

$$\mathcal{I}_n(\theta) = 2n \int_1^\infty \frac{(z-1)}{z^4} dz \qquad\qquad = 2n \left[ \int_1^\infty z^{-3} dz - \int_1^\infty z^{-4} dz \right]$$

$$= 2n \left[ \frac{1}{2} - \frac{1}{3} \right]$$

$$= 2n \times \frac{1}{6} = \frac{n}{3}$$

which is independent of $\theta$. In other words, $\mathcal{I}(\theta) = 1/3$, for a single sample from the distribution.

Now, we know that as $X_i$ has the same density function as a logisitic distribution with location parameter $\theta$ and scale parameter 1, it follows that $\mathbb{E}(X_i) = \theta$ and $\text{Var}(X_i) = \pi^2/3$. Therefore, by Central Limit theorem, it follows that

$$\sqrt{n} \frac{(\bar{X}_n - \theta)}{\pi/\sqrt{3}} \xrightarrow{d} \mathcal{N}(0,1) \tag{11.2}$$

Thus if we consider, $g : \mathbb{R} \to \mathbb{R}$ such that $g(x) = \frac{3}{\pi} x$, then

$$\sqrt{n}(g(\bar{X}_n) - g(\theta)) \xrightarrow{d} \mathcal{N}\left(0, \frac{\pi^2}{3} \times \frac{9}{\pi^2}\right) \equiv \mathcal{N}(0,3) \equiv \mathcal{N}\left(0, 1/\mathcal{I}(\theta)\right) \tag{11.3}$$

Therefore, we can consider the constant sequence of functions, namely $g_n(x) = \frac{3}{\pi} x$ such that the sequence of random variables $g_n(\bar{X}_n)$, after suitable centering and scaling as shown in eq. (11.3), converges in distribution to $\mathcal{N}(0, 1/\mathcal{I}(\theta))$.

∎

# 12   Problem 12

**Problem.** Let $X_1, X_2, \cdots, X_n$ be i.i.d. with a common density $f(x, \theta)$ given by,

$$f(x, \theta) = e^{-(x-\theta)} \quad x \geq \theta, \quad -\infty < \theta < \infty,$$

(a) Find the MLE $\hat{\theta}_n$ of $\theta$ and the limiting non-degenerate distribution of suitably normalized $\left(\hat{\theta}_n - \theta\right)$.

(b) Given that $\theta \leq 1$, find the MLE $\tilde{\theta}_n$ of $\theta$ and also find the limiting non-degenerate distribution of suitably normalized $\left(\tilde{\theta}_n - \theta\right)$ under $\theta = \theta_0 < 1$.

*Solution:*

(a) The likelihood function is given as follows:

$$
\begin{aligned}
\mathcal{L}(\theta) &= \prod_{i=1}^{n} f(X_i, \theta) \\
&= \prod_{i=1}^{n} e^{-(X_i - \theta)} \mathbf{1}_{\{X_i \geq \theta\}} \qquad \text{where, } \mathbf{1}_A \text{ is the indicator function of } A \\
&= e^{-\sum_{i=1}^{n}(X_i - \theta)} \mathbf{1}_{\{X_{(1)} \geq \theta\}} \\
&= e^{(n\theta - \sum_{i=1}^{n} X_i)} \mathbf{1}_{\{X_{(1)} \geq \theta\}}
\end{aligned}
$$

Now, it is clear that this function takes value $0$ if $\theta > X_{(1)}$, and is monotonically increasing in $\theta$ before $\theta = X_{(1)}$. Hence, the likelihood function is maximized at $\theta = X_{(1)}$.

Therefore, the MLE is $\widehat{\theta}_n = X_{(1)}$. Now note that,

$$
\begin{aligned}
\mathbb{P}_\theta(X_{(1)} > x) &= \prod_{i=1}^{n} \mathbb{P}_\theta(X_i > x) \text{since, the former event is equivalent to the later event} \\
&= \prod_{i=1}^{n} \int_{x}^{\infty} e^{-(u-\theta)} du \\
&= \prod_{i=1}^{n} e^{-(x-\theta)} \\
&= e^{-n(x-\theta)}
\end{aligned}
$$

Therefore,

$$\mathbb{P}_\theta(n(\widehat{\theta}_n - \theta) \leq x) = \mathbb{P}_\theta(n(X_{(1)} - \theta) \leq x)$$
$$= \mathbb{P}_\theta\left(X_{(1)} \leq \theta + \frac{x}{n}\right)$$
$$= 1 - \mathbb{P}_\theta\left(X_{(1)} > \theta + \frac{x}{n}\right)$$
$$= 1 - \exp\left[-n\left(\theta + \frac{x}{n} - \theta\right)\right] \qquad \text{by previous derivation}$$
$$= 1 - e^{-x}$$

which is the cumulative distribution function of exponential distribution with mean 1. Therefore, the limiting non-degenerate distribution of suitably normalized quantity regarding the MLE is: $n(\widehat{\theta}_n - \theta) \xrightarrow{d} \mathrm{Exp}(1)$, where $\mathrm{Exp}(1)$ is the exponential distribution with scale parameter 1.

(b) In case it is also given that $\theta \leq 1$, the likelihood function is given by;

$$\mathcal{L}(\theta) = e^{(n\theta - \sum_{i=1}^n X_i)}\mathbf{1}_{\{X_{(1)} \geq \theta\}}\mathbf{1}_{\{\theta \leq 1\}}$$

Therefore, the likelihood function only takes positive value if $\theta \leq \min\{X_{(1)}, 1\}$ and in the region it takes positive value, the likelihood is monotonically increasing as before. Therefore, the MLE $\widetilde{\theta}_n$ in this case is given by, $\widetilde{\theta}_n = \min\{X_{(1)}, 1\}$.

Now, under $\theta = \theta_0 < 1$, similar to before,

$$\mathbb{P}_{\theta_0}(n(\widetilde{\theta}_n - \theta_0) \leq x) = \mathbb{P}_{\theta_0}(n(\min\{X_{(1)}, 1\} - \theta_0) \leq x)$$
$$= 1 - \mathbb{P}_{\theta_0}\left(n(\min\{X_{(1)}, 1\} - \theta_0) > x\right)$$
$$= 1 - \mathbb{P}_{\theta_0}(n(X_{(1)} - \theta_0) > x)\mathbf{1}_{\{n(1-\theta_0)>x\}},$$
$$\text{where } \mathbf{1}_A \text{ is the indicator function}$$
$$= 1 - \mathbb{P}_{\theta_0}\left(X_{(1)} > \theta_0 + \frac{x}{n}\right)\mathbf{1}_{\left\{(1-\theta_0)>\frac{x}{n}\right\}}$$

Now note that, since $\theta_0 < 1$, $(1-\theta_0) > 0$, and as $x/n \to 0$ for $n \to \infty$, for large enough $n$, we must have $(1 - \theta_0) > x/n$, hence there exists a large $N \in \mathbb{N}$ such that for all $n \geq N$,

$$\mathbb{P}_{\theta_0}(n(\widetilde{\theta}_n - \theta_0) \leq x) = 1 - \mathbb{P}_{\theta_0}\left(X_{(1)} > \theta_0 + \frac{x}{n}\right) = 1 - e^{-x}$$

Hence, $n(\widetilde{\theta}_n - \theta_0) \xrightarrow{d} \mathrm{Exp}(1)$, which shows that the asymptotic distribution of properly normalized $\widetilde{\theta}_n$ i.e. $n(\widetilde{\theta}_n - \theta_0)$ is an exponential distribution with scale parameter equal to 1.

■

# 13 Problem 13

*Solution:* Let us first develop the method of scoring in case of Cauchy distribution. Let, $X_1, X_2, \ldots X_n$ be i.i.d. random sample from a Cauchy distribution with location parameter $\theta$, and scale parameter 1, where $\theta$ is unknown and to be estimated from the data.

The likelihood is given by;

$$\mathcal{L}(\theta) = \prod_{i=1}^{n} \frac{1}{\pi\left[1+(X_i-\theta)^2\right]} = \frac{1}{\pi^n}\prod_{i=1}^{n}\left[1+(X_i-\theta)^2\right]^{-1}$$

Hence, the log-likelihood is given by;

$$\ell(\theta) = -n\log\pi - \sum_{i=1}^{n}\log\left[1+(X_i-\theta)^2\right]$$

Differentiating this log-likelihood with respect to $\theta$ yields,

$$\frac{\partial\ell}{\partial\theta} = \sum_{i=1}^{n}\frac{2(X_i-\theta)}{\left[1+(X_i-\theta)^2\right]}$$

$$\frac{\partial^2\ell}{\partial\theta^2} = 2\sum_{i=1}^{n}\frac{(X_i-\theta)^2-1}{\left[1+(X_i-\theta)^2\right]^2}$$

Finally, the Fisher's Information turns out to be;

$$\mathcal{I}(\theta) = \mathbb{E}_\theta\left[-\frac{\partial^2\ell}{\partial\theta}\right]$$

$$= (-2)\sum_{i=1}^{n}\mathbb{E}_\theta\left[\frac{(X_i-\theta)^2-1}{\left[1+(X_i-\theta)^2\right]^2}\right]$$

$$= (-2)\sum_{i=1}^{n}\int_{-\infty}^{\infty}\frac{(x-\theta)^2-1}{\pi\left[1+(x-\theta)^2\right]^3}dx$$

$$= (-2)\sum_{i=1}^{n}\left[\int_{-\infty}^{\infty}\frac{(x-\theta)^2+1}{\pi\left[1+(x-\theta)^2\right]^3}dx - \int_{-\infty}^{\infty}\frac{2}{\pi\left[1+(x-\theta)^2\right]^3}dx\right]$$

Letting $y = \tan^{-1}(x-\theta)$, we obtain;

$$
\begin{aligned}
\mathcal{I}(\theta) &= \frac{(-2)}{\pi} \sum_{i=1}^{n} \left[ \int_{-\pi/2}^{\pi/2} \frac{\sec^2 y}{(\sec^2 y)^2} dy - 2 \int_{-\pi/2}^{\pi/2} \frac{\sec^2 y}{(\sec^2 y)^3} dy \right] \\
&= \frac{(-2)}{\pi} \sum_{i=1}^{n} \int_{-\pi/2}^{\pi/2} (\cos^2 y - 2 \cos^4 y) dy \\
&= \frac{(-2)}{\pi} \sum_{i=1}^{n} \int_{-\pi/2}^{\pi/2} \left( \frac{(1 + \cos 2y)}{2} - \frac{(1 + \cos 2y)^2}{2} \right) dy \\
&= \frac{(-1)}{\pi} \sum_{i=1}^{n} \int_{-\pi/2}^{\pi/2} (1 + \cos 2y - 1 - 2\cos 2y - \cos^2 2y) dy \\
&= \frac{(-1)}{\pi} \sum_{i=1}^{n} \int_{-\pi/2}^{\pi/2} \left( -\cos 2y - \frac{(1 + \cos 4y)}{2} \right) \\
&= \frac{1}{\pi} \sum_{i=1}^{n} \frac{\pi}{2} \\
&= \frac{n}{2}
\end{aligned}
$$

Therefore, the Fisher's scoring method for $n = 25$ samples for the given setup is based on the iteration;

$$
\theta^{(t+1)} = \theta^{(t)} + \left[ \mathcal{I}(\theta^{(t)}) \right]^{-1} \left[ \frac{\partial \ell}{\partial \theta} \right]_{\theta = \theta^{(t)}}
$$

i.e.

$$
\theta^{(t+1)} = \theta^{(t)} + \frac{2}{25} \sum_{i=1}^{25} \frac{2(X_i - \theta^{(t)})}{[1 + (X_i - \theta^{(t)})^2]} \tag{13.1}
$$

Now, we generate a random sample of size 25 from the above mentioned Cauchy distribution with true $\theta = 1$. Again, my roll number is set as a seed to ensure reproducibility.

```
1 set.seed(1911)
2 samp <- rcauchy(25, location = 1, scale = 1)
```

The obtained samples are as follows:

−0.88420300	25.42372020	1.79721325	3.74698929	2.13951320
0.08265225	1.17629455	−0.95525771	0.95430173	1.13180492
1.44572412	1.78815468	16.92143196	0.31327702	3.00195967
0.96317277	0.97412383	−3.68335105	0.35682861	−7.84890081
−0.73825877	2.88164737	−3.62805420	−4.34463180	0.63972766

Now we use eq. (13.1) to estimate the unknown parameter θ starting from a initial estimate $\theta^{(0)}$. This initial estimate $\theta^{(0)}$ is chosen to be equal to 0. The following code performs iteration of the Fisher's scoring algorithm unless the error between subsequent estimates are very small.

```
1  update.theta <- function(init.theta) {
2
3     theta_t <- init.theta   # initialize
4     error <- Inf    # initialize error to some large value
5     n_iter <- 0     # store number of iterations
6
7     while(error > 1e-5) {
8         # update theta by iteration
9         new_theta <- theta_t + (2/25) * sum( 2 * (samp - theta_t)/(1 + (
    samp - theta_t)^2 ) )
10
11         error <- (new_theta - theta_t) / theta_t    # calculate relative
    error
12         n_iter <- n_iter + 1    # increase counter of iteration
13
14         theta_t <- new_theta   # looping
15     }
16
17     return(list(MLE = theta_t, iteration = n_iter))
18  }
19  update.theta(init.theta = 0)
```

The output turns out to be as follows:

```
1  $MLE
2  [1] 0.9303578
3
4  $iteration
5  [1] 6
```

Therefore, the MLE of θ turns out to be equal to 0.9303578 by Fisher's scoring algorithm, which converges in 6 steps.

∎

14 Problem 14

> **Problem.** Consider the set up for the chi-square test for independence of two attributes (see the document "Large Sample Chi-square test 2" sent to you by email). Note that under independence the $k \times l$ cell probabilities can be expressed as functions of $(k-1) + (l-1)$ independent parameters. Find the maximum likelihood estimates of these parameters under independence.

Solution: First, the setup is described as follows:

There are two attributes A and B, with k levels $A_1, A_2, \ldots A_k$ and l levels $B_1, B_2, \ldots B_l$ respectively. The individuals in a population can be described by belonging to one of the k level with respect to attribute A and one of the l levels with respect to attribute B. Let, the probability that a person will belong to class $A_i B_j$ is π_{ij}, where $i = 1, 2, \ldots k; j = 1, 2, \ldots l$ with $\sum_{i=1}^{k} \sum_{j=1}^{l} \pi_{ij} = 1$.

From a sample of n individuals in total, their attribute levels are recorded and the observed number of individuals belonging to class $A_i B_j$ is n_{ij} with $\sum_{i=1}^{k} \sum_{j=1}^{l} n_{ij} = n$.

Let us use the following notations to denote the marginals.

$$n_{i0} = \sum_{j=1}^{l} n_{ij} \quad n_{0j} = \sum_{i=1}^{k} n_{ij}$$

$$\pi_{i0} = \sum_{j=1}^{l} \pi_{ij} \quad \pi_{0j} = \sum_{i=1}^{k} \pi_{ij}$$

Note that, the number of observations belonging to the class $A_i B_j$, i.e. n_{ij}, conditional on the total number of observations n, follows a multinomial distribution with cell probabilities π_{ij}. Thus, the likelihood is given as;

$$\mathcal{L}(\pi_{ij}; n_{ij} \quad \forall i, j) = \frac{n!}{\prod_{i=1}^{k} \prod_{j=1}^{l} n_{ij}!} \prod_{i=1}^{k} \prod_{j=1}^{l} \pi_{ij}^{n_{ij}} \tag{14.1}$$

Under the setup of independence, $\pi_{ij} = \pi_{i0} \pi_{0j} \quad \forall i, j$. Under this setup, all probabilities π_{ij} can be expressed as functions of $\pi_{10}, \pi_{20}, \ldots \pi_{(k-1)0}$ and $\pi_{01}, \pi_{02}, \ldots \pi_{0(l-1)}$, i.e. in terms of $(k-1) + (l-1)$ many independent parameters, while $\pi_{k0} = 1 - \sum_{i=1}^{(k-1)} \pi_{i0}$ and $\pi_{0l} = 1 - \sum_{j=1}^{(l-1)} \pi_{0j}$. Thus, under independence the likelihood given in eq. (14.1) turns into;

$$\mathcal{L}(\pi_{i0}, \pi_{0j}; n_{ij} \quad \forall i, j) = \frac{n!}{\prod_{i=1}^{k} \prod_{j=1}^{l} n_{ij}!} \prod_{i=1}^{(k-1)} \prod_{j=1}^{(l-1)} (\pi_{i0} \pi_{0j})^{n_{ij}} \times \prod_{j=1}^{(l-1)} \left(\left(1 - \sum_{i=1}^{(k-1)} \pi_{i0} \right) \pi_{0j} \right)^{n_{kj}}$$

$$\times \prod_{i=1}^{(k-1)} \left(\pi_{i0} \left(1 - \sum_{j=1}^{(l-1)} \pi_{0j} \right) \right)^{n_{il}} \times \left(\left(1 - \sum_{i=1}^{(k-1)} \pi_{i0} \right) \left(1 - \sum_{j=1}^{(l-1)} \pi_{0j} \right) \right)^{n_{kl}}$$

This can be simplified as;

$$\mathcal{L}(\pi_{i0}, \pi_{0j}; n_{ij}\forall i, j) = \frac{n!}{\prod_{i=1}^{k}\prod_{j=1}^{l} n_{ij}!} \prod_{i=1}^{(k-1)} \pi_{i0}^{\sum_{j=1}^{l} n_{ij}} \prod_{j=1}^{(l-1)} \pi_{0j}^{\sum_{i=1}^{k} n_{ij}} \times$$

$$\left(1 - \sum_{i=1}^{(k-1)} \pi_{i0}\right)^{\sum_{j=1}^{l} n_{kj}} \left(1 - \sum_{j=1}^{(l-1)} \pi_{0j}\right)^{\sum_{i=1}^{k} n_{il}}$$

$$= \frac{n!}{\prod_{i=1}^{k}\prod_{j=1}^{l} n_{ij}!} \prod_{i=1}^{(k-1)} \pi_{i0}^{n_{i0}} \prod_{j=1}^{(l-1)} \pi_{0j}^{n_{0j}} \left(1 - \sum_{i=1}^{(k-1)} \pi_{i0}\right)^{n_{k0}} \left(1 - \sum_{j=1}^{(l-1)} \pi_{0j}\right)^{n_{0l}}$$

Now, note that, if any $\pi_{i0} = 0$ or if $\sum_{i=1}^{(k-1)} \pi_{i0} = 1$, then due to independence assumption, the corresponding probabilities of that row will be 0, then we can simply drop that level of attribute A from consideration. Thus, without loss of generality, we can assume, the parameters belong to the set

$$A = \left\{ (\pi_{10}, \dots \pi_{(k-1)0}, \pi_{01}, \dots \pi_{0(l-1)}) : \pi_{i0} > 0 \quad \forall i = 1, \dots (k-1), \right.$$

$$\left. \pi_{0j} > 0 \quad \forall j = 1, \dots (l-1), \sum_{i=1}^{(k-1)} \pi_{i0} < 1, \sum_{j=1}^{(l-1)} \pi_{0j} < 1 \right\}$$

Therefore, assuming the parameters belonging to A, we can consider the log-likelihood,

$$\ell = \text{constant} + \sum_{i=1}^{(k-1)} n_{i0} \log(\pi_{i0}) + \sum_{j=1}^{(j-1)} n_{0j} \log(\pi_{0j}) + n_{k0} \log\left(1 - \sum_{i=1}^{(k-1)} \pi_{i0}\right) + n_{0l} \log\left(1 - \sum_{j=1}^{(l-1)} \pi_{0j}\right)$$

and note that, since the log-likelihood is differentiable on A, to find the MLE (i.e. to maximize the likelihood), it is enough to set partial derivatives of ℓ equal to 0 and obtain the solutions. Hence,

$$\frac{\partial \ell}{\partial \pi_{i0}} = 0$$

$$\Rightarrow \frac{n_{i0}}{\pi_{i0}} - \frac{n_{k0}}{1 - \sum_{i'=1}^{(k-1)} \pi_{i'0}} = 0 \tag{14.2}$$

$$\Rightarrow \frac{n_{i0}}{\pi_{i0}} = \frac{n_{k0}}{1 - \sum_{i'=1}^{(k-1)} \pi_{i'0}}$$

Similarly, setting $\frac{\partial \ell}{\partial \theta_{0j}} = 0$ yields,

$$\frac{n_{0j}}{\pi_{0j}} = \frac{n_{0l}}{1 - \sum_{j'=1}^{(l-1)} \pi_{0j'}} \tag{14.3}$$

Since, eq. (14.2) holds for any $i = 1, 2, \ldots (k-1)$, we then have,

$$\frac{n_{10}}{\pi_{10}} = \frac{n_{20}}{\pi_{20}} = \cdots = \frac{n_{(k-1)0}}{\pi_{(k-1)0}} = \frac{n_{k0}}{1 - \sum_{i'=1}^{(k-1)} \pi_{i'0}} = \frac{n_{10} + \cdots + n_{(k-1)0} + n_{k0}}{\pi_{10} + \cdots + \pi_{(k-1)0} + 1 - \sum_{i'=1}^{(k-1)} \pi_{i'0}} = \frac{n}{1}$$

Thus, $\widehat{\pi_{i0}} = n_{i0}/n$ for all $i = 1, 2, \ldots (k-1)$. In an exactly similar way, starting from eq. (14.3) we get $\widehat{\pi_{0j}} = n_{0j}/n$ for all $j = 1, 2, \ldots (l-1)$. Therefore, the MLE of those parameters are simply as follows:

$$\widehat{\pi_{i0}} = \frac{n_{i0}}{n} \quad \forall i$$

$$\widehat{\pi_{0j}} = \frac{n_{0j}}{n} \quad \forall j$$

Hence, $\widehat{\pi_{ij}} = \frac{n_{i0} n_{0j}}{n^2} \quad \forall i, j \quad$, under independence

∎

Thank you