Probability Theory and Statistics 3, Assignment 1

Jiyeon Kim [*], Younjoo Mo [†]

October 30, 2020

Exercise 1

a. ..

Let $Y_n = \sqrt{n} \cdot X_{1:n}$ and $G_n(y)$ is the CDF of Y_n.

$$
\begin{aligned}
G_n(y) &= P[Y_n \leq y] && (1)\\
&= P\left[X_{1:n} \leq \frac{y}{\sqrt{n}}\right] = 1 - P\left[all\, X_i \geq \frac{y}{\sqrt{n}}\right] && (2)\\
&\overset{\text{i.i.d.}}{=} 1 - \left(P\left[X_i \geq \frac{y}{\sqrt{n}}\right]\right)^n && (3)\\
&= 1 - \left(1 - F_X\left(\frac{y}{\sqrt{n}}\right)\right)^n, (0 \leq y \leq \sqrt{n}\theta) && (4)
\end{aligned}
$$

$$
\begin{aligned}
F_X(x) &= P[X \leq x] && (5)\\
&= \int_0^x f_X(u)du = \int_0^x \frac{2u}{\theta^2}du && (6)\\
&= \frac{x^2}{\theta^2}\ ,\ (0 \leq x \leq \theta) && (7)
\end{aligned}
$$

[*] Student number: 12470236

[†] Studentnumber: 12475440

Substitute (7) to (4)

$$G_n(y) = 1 - \left(1 - \frac{1}{\theta^2} \cdot \frac{y^2}{n}\right)^n \tag{8}$$

$$= 1 - \left(1 + \frac{-\theta^{-2}y^2}{n}\right)^n \ , \ (0 \leq y \leq \sqrt{n}\theta) \tag{9}$$

Find the limit of the parentheses part of (9)

$$\lim_{n\to\infty} \left(1 + \frac{-\theta^{-2}y^2}{n}\right)^n \overset{7.2.7}{=} \exp\left(-\frac{y^2}{\theta^2}\right) \tag{10}$$

Substitute (10) to (9)

$$G(y) = \lim_{n\to\infty} G_n(y) = 1 - \exp\left(-\frac{y^2}{\theta^2}\right) \ , \ (0 \leq y) \tag{11}$$

The pdf of the limiting distribution

$$g(y) = \frac{d}{dy}G(y) = \frac{d}{dy}\left[1 - \exp\left(-\frac{y^2}{\theta^2}\right)\right] \tag{12}$$

$$= \frac{2}{\theta^2} \cdot y \cdot e^{-(y/\theta)^2} \ , \ (0 \leq y) \tag{13}$$

$$\therefore Y \sim WEI(\theta, 2) \tag{14}$$

b. ..

Substitute $\theta = 1$, $n = 50$ to (9)

$$G_{50}(y) = 1 - \left(1 - \frac{y^2}{50}\right)^{50} \ , \ (0 \leq y \leq \sqrt{50}) \tag{15}$$

Find the pdf

$$g_{50}(y) = \frac{d}{dy}G_{50}(y) = \frac{d}{dy}\left[1 - \left(1 - \frac{y^2}{50}\right)^{50}\right] \tag{16}$$

$$= 2y\left(1 - \frac{y^2}{50}\right)^{49} \ , \ (0 \leq y \leq \sqrt{50}) \tag{17}$$

Compute $E[\sqrt{50}X_{1:50}] = E[Y_{50}]$

$$\begin{aligned}
E[\sqrt{50}X_{1:50}] &= E[Y_{50}] && (18)\\
&= \int_0^{\sqrt{50}} y \cdot g_{50}(y)dy && (19)\\
&= \int_0^{\sqrt{50}} 2y^2\left(1-\frac{y^2}{50}\right)^{49} dy && (20)\\
&= 100\int_0^{\sqrt{50}} \frac{y^2}{50}\left(1-\frac{y^2}{50}\right)^{49} dy && (21)\\
&\text{Use the given formula} && (22)\\
&= 100 \cdot \frac{\sqrt{\pi}\sqrt{50}}{4} \cdot \frac{\Gamma(50)}{\Gamma(50+\frac{3}{2})} \overset{\text{Rstudio}}{\approx} 0.8796 && (23)\\
& && (24)
\end{aligned}$$

c. ..

Compute $E[\sqrt{n}X_{1:n}] \approx E[Y]$

$$\begin{aligned}
&\text{From the subquestion (a)}, \;\; Y \sim WEI(\theta, 2) && (25)\\
&E[\sqrt{n}X_{1:n}] \approx E[Y] \overset{\text{App. B}}{=} \theta\Gamma\left(1+\frac{1}{2}\right) \;\; , \; (\theta > 0) && (26)
\end{aligned}$$

When $\theta = 1$, $n = 50$

$$E[\sqrt{50}X_{1:50}] \approx E[Y] = \Gamma\left(1+\frac{1}{2}\right) = \frac{1}{2}\sqrt{\pi} \approx 0.886226925 \qquad (27)$$

Exercise 2

a. ..

$f_X(x) = \frac{1}{\pi(1+x^2)}, x \in R$
The question requires to show that mean and variance of the standard Cauchy distribution is not well defined
First, in order to get the mean, we are going to use general formula
$\int_{-\infty}^{\infty} x f_X(x)dx = \int_{-\infty}^{\infty} \frac{x}{\pi(1+x^2)}dx$
We used integral here since Cauchy distribution is continuous probability distribution
$\int_{-\infty}^{\infty} \frac{x}{\pi(1+x^2)}dx = \int_{-\infty}^{0} \frac{x}{\pi(1+x^2)}dx + \int_{0}^{\infty} \frac{x}{\pi(1+x^2)}dx$
Let's put $u = 1+x^2, du = 2xdx$
$\int_{0}^{\infty} \frac{x}{\pi(1+x^2)}dx = \int_{1}^{\infty} \frac{1}{2\pi(u)}du = [\frac{lnu}{2\pi}]_{1}^{\infty}$
$\int_{-\infty}^{0} \frac{x}{\pi(1+x^2)}dx = \int_{-\infty}^{1} \frac{1}{2\pi(u)}du = [\frac{lnu}{2\pi}]_{-\infty}^{1}$
$\int_{-\infty}^{\infty} \frac{x}{\pi(1+x^2)}dx = \frac{1}{\pi}[\frac{1}{2}ln(1+x^2)]_{-\infty}^{0} + \frac{1}{\pi}[\frac{1}{2}ln(1+x^2)]_{0}^{\infty}$
For the integral to exist, values need to be finite. But as shown as follows both of the separate values, higher half and lower half are infinite.
$\therefore$ We got the conclusion that mean is undefined.
The general formula to get variance is $E(X^2) - E(X)^2$.
But since E(X)is undefined and $E(x^2) = \frac{1}{\pi}\int \frac{x^2}{1+x^2}dx = \infty$, we can also prove that the variance of the standard Cauchy distribution are not well- defined.

b. ..

We solved the question b by two ways. First by using the hint and second by using the convolution formula.

[Solution 1]

Make the joint pdf of X_1 and X_2

$$f_{X_1,X_2}(x_1,x_2) \overset{\text{i.i.d.}}{=} \frac{1}{\pi^2} \cdot \frac{1}{x_1^2+1} \cdot \frac{1}{x_2^2+1} , \quad (x_1, x_2 \in \mathbb{R}) \tag{28}$$

Let $J = \bar{X_2} = \frac{X_1+X_2}{2}$, $\quad (j \in \mathbb{R})$

Use the CDF-method to find the pdf of J

$$
\begin{aligned}
F_J(j) &= P(J \leq j) && (29)\\
&= P\left(\frac{X_1+X_2}{2} \leq j\right) = P(X_1+X_2 \leq 2j) && (30)\\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{2j-x_1} \frac{1}{\pi^2}\cdot\frac{1}{x_1^2+1}\cdot\frac{1}{x_2^2+1}dx_2dx_1 && (31)\\
&= \frac{1}{\pi^2}\int_{-\infty}^{\infty}\frac{1}{x_1^2+1}\left(\int_{-\infty}^{2j-x_1}\frac{1}{x_2^2+1}dx_2\right)dx_1 && (32)\\
&= \frac{1}{\pi^2}\int_{-\infty}^{\infty}\frac{1}{x_1^2+1}\left(\arctan(2j-x_1)+\frac{\pi}{2}\right)dx_1 && (33)\\
&= \frac{1}{\pi^2}\left\{\int_{-\infty}^{\infty}\frac{\arctan(2j-x_1)}{x_1^2+1}dx_1+\int_{-\infty}^{\infty}\frac{\pi}{2(x_1^2+1)}dx_1\right\} && (34)\\
&= \frac{1}{\pi^2}\int_{-\infty}^{\infty}\frac{\arctan(2j-x_1)}{x_1^2+1}dx_1+\frac{1}{\pi^2}\cdot\frac{\pi}{2}\Big[\arctan(x_1)\Big]_{-\infty}^{\infty} && (35)\\
&\text{Use the given formula for the first term} && (36)\\
&= \frac{1}{\pi}arctan(j)+\frac{1}{2}\ , \quad (j\in\mathbb{R}) && (37)
\end{aligned}
$$

The pdf of J

$$
\begin{aligned}
f_J(j) &= \frac{d}{dj}F_J(j) = \frac{d}{dj}\left[\frac{1}{\pi}\arctan(j)+\frac{1}{2}\right] && (38)\\
&= \frac{1}{\pi(j^2+1)}\ , \quad (j\in\mathbb{R}) && (39)
\end{aligned}
$$

$\therefore J = \bar{X_2} = \frac{X_1+X_2}{2} \sim Cauchy(0,1)$

[Solution 2]

The sample mean is average of the values of random sample.

$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$

We are going to use convolution formula here that we learned in PTS2

If X and Y are two independent continuous random variables with pdfs $f_X(x)$ and $f_Y(y)$, then the pdf of $S = X+Y$ is

$$f_S(s) = \int_{-\infty}^{\infty} f_X(t) f_Y(s-t) dt \tag{40}$$

$$\tag{41}$$

Before using this formula , we are going to prove this shortly.

From the PTS2 reader theorem 6.3, we can see that Jacobian from (X,Y) to (S,T) is 1.

$$f_{T,S}(t,s) = f_{x,y}(t,s-t) \tag{42}$$

$$f_S(s) = \int_{-\infty}^{\infty} f_X(t) f_Y(s-t) dt \tag{43}$$

$\because$ We assumed X,Y are independent variables.

Now, we are going to prove that the $\bar{X_2} = \frac{X_1+X_2}{2}$ is also Cauchy distributed

Here, since using variables X_1, X_2 could be confusing, we put the variables X_1, X_2 each as P, Q and sum of them $X_1 + X_2 = 2\bar{X_2}$ as S

We first assumed that P and Q each of them is an independent Cauchy random variable,

$$P \sim Cauchy(0,1) \ , \quad Q \sim Cauchy(0,1) \ , \quad (p,q,s \in \mathbb{R}) \tag{44}$$

$$f_P(p) = \frac{1}{\pi(1+p^2)} \ , \quad f_Q(q) = \frac{1}{\pi(1+q^2)} \tag{45}$$

If we use the convolution formula that we proved, the equation becomes as follows

$$f_S(s) = \int_{-\infty}^{\infty} f_P(p) f_Q(s-p) dt \tag{46}$$

$$f_S(s) = \int_{-\infty}^{\infty} \frac{1}{\pi(1+p^2)} \frac{1}{\pi(1+(s-p)^2)} dw \tag{47}$$

$$= \frac{1}{2\pi} \frac{1}{(1+(\frac{s}{2})^2)} \tag{48}$$

Here, if we put $S = P + Q = X + Y = X_1 + X_2 = 2\bar{X_2}$We get the result that the sum of the two independent identical distributed random variables is also Cauchy distributed.

$\therefore \frac{S}{2} = \bar{X_2} = \frac{X_1+X_2}{2} \sim Cauchy(0,1)$

c. ..

Assume a random variable X follows a certain distribution.

According to the Central Limit Theorem(CLT), if n is sufficiently large, $\bar{X}$ is approximately normally distributed although X is not normally distributed. However, the sample mean of the Cauchy distribution always follows the Cauchy distribution no matter how large n is. At a glance, these two facts seem to contradict each other.

But actually, CLT can be used only when the distribution of X satisfies the following two conditions.

1. The distribution has a finite mean.
2. The distribution has a non-zero variance.

As we proved in the subquestion (a), the Cauchy distribution does not have its mean and variance. Therefore CLT cannot be applied to $\bar{X}$ of the Cauchy distribution in the first place.