Theoretical Statistics : Homework 1

Syamantak Kumar, UT EID : sk52827

March 19, 2023

1. (a) Consider the sequence of random variables $Z_n := X_n - Y_n$. Then,

$$\begin{aligned}\mathbb{E}\left[Z_n^2\right] &= \mathbb{E}\left[X_n^2\right] + \mathbb{E}\left[Y_n^2\right] - 2\mathbb{E}\left[X_n Y_n\right] \\ &= 1 + 1 - 2Corr\left(X_n, Y_n\right) \\ &= 2\left(1 - Corr\left(X_n, Y_n\right)\right)\end{aligned}$$

We note that $\mathbb{E}\left[Z_n^2\right] \to 0$ as $Corr\left(X_n, Y_n\right) \to 1$. Therefore,

$$Z_n \xrightarrow{\text{q.m}} 0$$

which implies $Z_n \xrightarrow{\text{p}} 0$ and completes our proof for the first part.

Next, consider the sequences of random variables -

$$X_n' := \frac{X_n - \mathbb{E}\left[X_n\right]}{\sqrt{Var\left(X_n\right)}} \text{ and}$$
$$Y_n' := \frac{Y_n - \mathbb{E}\left[Y_n\right]}{\sqrt{Var\left(Y_n\right)}}$$

Then,

$$\mathbb{E}\left[X_n'\right] = 0, Var\left(X_n'\right) = 1 \text{ and } \mathbb{E}\left[Y_n'\right] = 0, Var\left(Y_n'\right) = 1$$

Using the result from the first part, since

$$Corr\left(X_n, Y_n\right) = Corr\left(X_n', Y_n'\right) \to 1$$

Then, $X_n' - Y_n' \xrightarrow{\text{p}} 0$. From the Theorem on Page 23, Lecture 1 of Lecture Notes, we have, that

$$X_n' \xrightarrow{\text{d}} X \text{ and } X_n' - Y_n' \xrightarrow{\text{p}} 0 \implies Y_n' \xrightarrow{\text{d}} X$$

Hence proved.

(b) No it is not necessarily true that $Y_n \xrightarrow{\text{d}} X$. Consider the following sequences of random variables - (the joint distribution only represents the values where the probability is non-zero)

X_n	$P(X_n)$
n^2	$\frac{1}{2n}$
$-n^2$	$\frac{1}{2n}$
-2	$\frac{1}{2} - \frac{1}{2n}$
2	$\frac{1}{2} - \frac{1}{2n}$

Y_n	$P(Y_n)$
n^2	$\frac{1}{2n}$
$-n^2$	$\frac{1}{2n}$
$-2\sqrt{2}$	$\frac{1}{4} - \frac{1}{4n}$
$2\sqrt{2}$	$\frac{1}{4} - \frac{1}{4n}$
0	$\frac{1}{2} - \frac{1}{2n}$

X_n	Y_n	$P(X_n, Y_n)$
n^2	n^2	$\frac{1}{2n}$
$-n^2$	$-n^2$	$\frac{1}{2n}$
-2	$2\sqrt{2}$	$\frac{1}{4} - \frac{1}{4n}$
-2	0	$\frac{1}{4} - \frac{1}{4n}$
2	$-2\sqrt{2}$	$\frac{1}{4} - \frac{1}{4n}$
2	0	$\frac{1}{4} - \frac{1}{4n}$

Then,

$$\mathbb{E}\left[X_n\right] = 0, \mathbb{E}\left[X_n^2\right] = n^3 + 4 - \frac{4}{n}$$
$$\mathbb{E}\left[Y_n\right] = 0, \mathbb{E}\left[Y_n^2\right] = n^3 + 4 - \frac{4}{n}$$
$$\mathbb{E}\left[X_n, Y_n\right] = n^3 - 2\sqrt{2} + \frac{2\sqrt{2}}{n}$$

and

$$\begin{aligned} Corr\left(X_n, Y_n\right) &= \frac{\mathbb{E}\left[X_n, Y_n\right] - \mathbb{E}\left[X_n\right]\mathbb{E}\left[Y_n\right]}{\sqrt{Var\left(X_n\right)}\sqrt{Var\left(Y_n\right)}} \\ &= \frac{n^3 - 2\sqrt{2} + \frac{2\sqrt{2}}{n} - 0}{n^3 + 4 - \frac{4}{n}} \\ &= \frac{n^3 - 2\sqrt{2} + \frac{2\sqrt{2}}{n}}{n^3 + 4 - \frac{4}{n}} \end{aligned}$$

Therefore, $Corr\left(X_n, Y_n\right) \to 1$ as $n \to \infty$. However, $X_n \xrightarrow{\text{p}} X$ and $Y_n \xrightarrow{\text{p}} Y$ (and consequently $X_n \xrightarrow{\text{d}} X, Y_n \xrightarrow{\text{d}} Y$) given as -

X	$P(X)$
-2	$\frac{1}{2}$
2	$\frac{1}{2}$

Y	$P(Y)$
$-2\sqrt{2}$	$\frac{1}{4}$
$2\sqrt{2}$	$\frac{1}{4}$
0	$\frac{1}{2}$

This can be verified by noting that

$$P(|X_n - X| \geq \epsilon) \leq P(X_n = n^2) + P(X_n = -n^2) = \frac{1}{n} \to 0$$
$$P(|Y_n - Y| \geq \epsilon) \leq P(Y_n = n^2) + P(Y_n = -n^2) = \frac{1}{n} \to 0$$

Therefore we have shown a counterexample where X_n and Y_n have the same variance and $Corr(X_n, Y_n) \to 1$, but $X_n \xrightarrow{d} X$ does not imply $Y_n \xrightarrow{d} X$.

2. To complete Y_i's PMF, we can define it as :

$$\forall k \in \mathbb{N}, \;\; Y_i = k \;\; \text{if} \sum_{j=0}^{k-1} \frac{e^{-p_i} p_i^j}{j!} \leq U_i < \sum_{j=0}^{k} \frac{e^{-p_i} p_i^j}{j!}$$

Along with the definition of $Y_i = 0$ for $U_i < 0$, we get $Y_i \sim \mathcal{P}(p_i)$. Therefore, S_n and $\sum_i X_i$ follow the same distribution, and similarly Z and $\sum_i Y_i$ follow the same distribution. This follows from the fact that sum of independent poisson random variables with parameters p_i is a poisson random variable with parameter $\sum_i p_i$. Now, for the next part, we note that

$$|P(S_n \in A) - P(Z \in A)| = \left| P\left(\sum_i X_i \in A\right) - P\left(\sum_i Y_i \in A\right) \right|$$

Denoting $\sum_i X_i := C$, $\sum_i Y_i := D$, we have,

$$\begin{aligned}
|P(S_n \in A) - P(Z \in A)| &= |P(C \in A) - P(D \in A)| \\
&= |P(C \in A, C = D) + P(C \in A, C \neq D) - P(D \in A, C = D) - P(D \in A, C \neq D)| \\
&= |P(C \in A, C \neq D) - P(D \in A, C \neq D)| \text{ since } P(C \in A, C = D) = P(D \in A, C = D) \\
&\leq P(C \in A, C \neq D) + P(D \in A, C \neq D) \\
&\leq P(C \neq D) \\
&= P\left(\sum_i X_i \neq \sum_i Y_i\right) \\
&\leq P(\exists i \text{ such that } X_i \neq Y_i) \\
&\leq \sum_i P(X_i \neq Y_i) \text{ using the Union Bound}
\end{aligned}$$

Let's calculate $P(X_i \neq Y_i)$. Figure 1 helps simplify this calculation. Therefore,

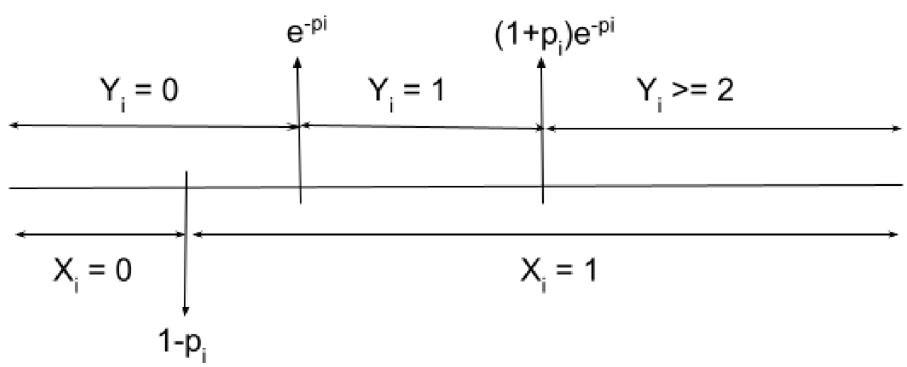

Figure 1: Joint probability space of X_i and Y_i

$$\begin{aligned}
P(X_i \neq Y_i) &= P(X_i = 1, Y_i = 0) + P(Y_i \geq 2) \\
&= e^{-p_i} - (1 - p_i) + 1 - (1 + p_i) e^{-p_i} \\
&= p_i \left(1 - e^{-pi}\right) \\
&\leq p_i^2 \text{ since } e^{-pi} \geq 1 - p_i
\end{aligned}$$

which completes our proof.

3. Consider the random variables

$$Y_i := \frac{z_{ni}}{\sqrt{\sum_i z_{ni}^2}} \frac{X_i - \mu}{\sigma}$$

Then, we note that

$$\forall i, \quad \mathbb{E}\left[Y_i\right] = 0 \text{ and } Var\left(Y_i\right) = \frac{z_{ni}^2}{\sum_i z_{ni}^2}$$

Therefore, $\mathbb{E}\left[\sum_i Y_i\right] = 0$ and $B_n^2 := Var\left(\sum_i Y_i\right) = 1$ since Y_i's are independent. Further, we note that

$$\forall i, \quad \frac{Var\left(Y_i\right)}{Var\left(\sum_i Y_i\right)} = \frac{z_{ni}^2}{\sum_i z_{ni}^2} \to 0 \text{ as } n \to \infty$$

Therefore, the Lindeberg condition is necessary and sufficient to prove convergence of

$$\frac{\sum_i Y_i}{B_n} = \frac{T_n - \mu_n}{\sigma_n} \to \mathcal{N}\left(0,1\right)$$

To verify Lindeberg condition, we need to prove that $\forall \epsilon > 0$,

$$\underbrace{\frac{1}{B_n^2} \sum_{i=1}^{n} \mathbb{E}\left[Y_i^2 \mathbb{1}\left(|Y_i| \geq \epsilon B_n\right)\right]}_{\mathcal{L}} \to 0 \text{ as } n \to \infty$$

$\mathcal{L}$ can be simplified as :

$$\begin{aligned}
\mathcal{L} &= \frac{1}{B_n^2} \sum_{i=1}^{n} \mathbb{E}\left[Y_i^2 \mathbb{1}\left(|Y_i| \geq \epsilon B_n\right)\right] \\
&= \frac{1}{\sigma^2} \frac{z_{ni}^2}{\sum_i z_{ni}^2} \sum_{i=1}^{n} \mathbb{E}\left[\left(X_i - \mu\right)^2 \mathbb{1}\left(|X_i - \mu| \geq \epsilon\sigma\left(\frac{\sum_i z_{ni}^2}{z_{ni}^2}\right)\right)\right]
\end{aligned}$$

Now consider the quantity,

$$Q := \left(X_i - \mu\right)^2 \mathbb{1}\left(|X_i - \mu| \geq \epsilon\sigma\left(\frac{\sum_i z_{ni}^2}{z_{ni}^2}\right)\right)$$

We note that $|Q| \leq \left(X_i - \mu\right)^2$ with $\mathbb{E}\left[\left(X_i - \mu\right)^2\right] = \sigma^2 < \infty$. Further, since $\frac{z_{ni}^2}{\sum_i z_{ni}^2} \to 0$, $\epsilon\sigma\left(\frac{\sum_i z_{ni}^2}{z_{ni}^2}\right) \to \infty$ as $n \to \infty$ since $\epsilon, \sigma > 0$. Therefore, $Q \xrightarrow{a.s} 0$ since for a fixed i, ϵ after some $n > n_0$, $|X_i - \mu| < \epsilon\sigma\left(\frac{\sum_i z_{ni}^2}{z_{ni}^2}\right)$ and therefore, $\mathbb{1}\left(|X_i - \mu| \geq \epsilon\sigma\left(\frac{\sum_i z_{ni}^2}{z_{ni}^2}\right)\right)$ would be 0 infinitely often. Therefore, we can use the dominated convergence theorem to claim that

$$\mathbb{E}\left[Q\right] = \mathbb{E}\left[\left(X_i - \mu\right)^2 \mathbb{1}\left(|X_i - \mu| \geq \epsilon\sigma\left(\frac{\sum_i z_{ni}^2}{z_{ni}^2}\right)\right)\right] \to 0 \text{ as } n \to \infty$$

This implies that $\mathcal{L} \to 0$ as $n \to \infty$, which means that the the Lindeberg condition is satisfied and

$$\frac{\sum_i Y_i}{B_n} = \frac{T_n - \mu_n}{\sigma_n} \to \mathcal{N}\left(0,1\right)$$

Hence proved

4. (a) This is not necessarily true since $g(x)$ is not continuous at $x = 0$. Consider the sequence of random variables

$$X_n = X + \frac{2}{n}$$

Note that $X_n \xrightarrow{p} X$ (and consequently $X_n \xrightarrow{d} X$) since

$$\begin{aligned}\forall \epsilon > 0,\ P\left(|X - X_n| \geq \epsilon\right) &= P\left(\left|\frac{2}{n}\right| \geq \epsilon\right) \\ &\leq \frac{\mathbb{E}\left[\frac{2}{n}\right]}{\epsilon} \text{ by Markov's Inequality} \\ &= \frac{2}{n\epsilon} \to 0 \text{ as } n \to \infty\end{aligned}$$

However, since $X \sim \mathcal{P}\left(\lambda\right)$, therefore $X \geq 0$. Therefore, $\forall n \geq 1$, $X_n > 0$. Therefore,

$$g(X_n) = \begin{cases} 1 \text{ for } X < 10 - \frac{2}{n} \\ 0 \text{ otherwise} \end{cases}$$

and

$$g(X) = \begin{cases} 1 \text{ for } X < 10 \text{ and } X \geq 1 \\ 0 \text{ otherwise} \end{cases}$$

Therefore,

$$\mathbb{E}\left[g(X_n)\right] = P(X < 10 - \frac{2}{n}) \to P\left(X < 10\right)$$

but

$$\mathbb{E}\left[g\left(X\right)\right] = P(X < 10) - P(X = 0) = P(X < 10) - e^{-\lambda}$$

(b) This is true by Portmanteau Theorem since $0 \leq e^{-x^2} \leq 1$ and e^{-x^2} is a continuous function.

(c) This is again true by Portmanteau Theorem since $g(x) = sgn(cos(x))$ is a bounded function, and is discontinuous at points $x = \frac{2n+1}{2}\pi, n \in \mathbb{Z}$ but the measure of the random variable X at these points is 0 since the Poisson random variable only takes non-negative integral values.

(d) This is not necessarily true since $g(x) = x$ is not a bounded function. Consider the sequence of random variables :

$$X_n = \begin{cases} 2n \text{ with probability } \frac{1}{n} \\ X \text{ with probability } \frac{n-1}{n} \end{cases}$$

Note that $X_n \xrightarrow{p} X$ (and consequently $X_n \xrightarrow{d} X$) since

$$\begin{aligned}\forall \epsilon > 0,\ P\left(|X - X_n| \geq \epsilon\right) &\leq P\left(X_n = 2n\right) \\ &= \frac{1}{n} \to 0 \text{ as } n \to \infty\end{aligned}$$

However,

$$\mathbb{E}\left[g\left(X_n\right)\right] = 2 + \frac{n-1}{n}\mathbb{E}\left[X\right] \to 2 + \mathbb{E}\left[X\right]$$

but

$$\mathbb{E}\left[g(X)\right] = \mathbb{E}\left[X\right]$$

5. (a) Let's compute the probability of $X_{n,k} = t$. For this event to occur, we note that for the first $t-1$ trials, one of $k-1$ possible items should be selected, and for the the t^{th} trial, one of the remaining new $n-k+1$ items should be selected. Therefore,

$$\begin{aligned} P\left(X_{n,k} = t\right) &= \left(\frac{k-1}{n}\right)^{t-1} \frac{n-k+1}{n} \\ &= \left(1 - \frac{n-k+1}{n}\right)^{t-1} \frac{n-k+1}{n} \end{aligned}$$

Therefore, $p_{n,k} = \frac{n-k+1}{n}$.

(b) We note that

$$T_n = \tau_n^n = \sum_{k=1}^{n} X_{n,k}$$

Therefore,

$$\begin{aligned} \mathbb{E}\left[T_n\right] &= \mathbb{E}\left[\sum_{k=1}^{n} X_{n,k}\right] \\ &= \sum_{k=1}^{n} \mathbb{E}\left[X_{n,k}\right], \text{ using linearity of expectation} \\ &= \sum_{k=1}^{n} \frac{n}{n-k+1}, \text{ since expectation of geometric random variable with parameter } p = \frac{1}{p} \\ &= n \sum_{k=1}^{n} \frac{1}{k} \end{aligned}$$

Similarly,

$$\begin{aligned} Var\left(T_n\right) &= Var\left(\sum_{k=1}^{n} X_{n,k}\right) \\ &= \sum_{k=1}^{n} Var\left(X_{n,k}\right), \text{ using independence of } X_{n,k} \\ &\leq \sum_{k=1}^{n} \left(\frac{n}{n-k+1}\right)^2, \text{ using bound provided in problem} \\ &= n^2 \sum_{k=1}^{n} \frac{1}{k^2} \\ &\leq n^2 \sum_{k=1}^{\infty} \frac{1}{k^2} \\ &= n^2 \frac{\pi^2}{6} \end{aligned}$$

(c) As provided in the hint, first consider the random variable

$$Z_n := \frac{T_n - \mathbb{E}\left[T_n\right]}{n \ln\left(n\right)}$$

Then,

$$\mathbb{E}\left[(Z_n - 0)^2\right] = \mathbb{E}\left[Z_n^2\right] = \mathbb{E}\left[\left(\frac{T_n - \mathbb{E}\left[T_n\right]}{n\ln(n)}\right)^2\right]$$
$$= \frac{Var\left(T_n\right)}{n^2 \ln(n)^2}$$
$$\leq \frac{\pi^2}{6}\frac{1}{\ln(n)^2} \to 0 \text{ as } n \to \infty$$

Therefore, $Z_n \xrightarrow{q.m} 0$ and consequently, $Z_n \xrightarrow{p} 0$. Then, we consider the random variable

$$Y_n := \frac{\mathbb{E}\left[T_n\right]}{n\ln(n)}$$

We prove that $Y_n \xrightarrow{p} 1$ as $n \to \infty$. This can be shown by noting that

$$\forall \epsilon > 0, \;\; P\left(|Y_n - 1| > \epsilon\right) = P\left(\left|\frac{\mathbb{E}\left[T_n\right]}{n\ln(n)} - 1\right| > \epsilon\right)$$
$$= P\left(\left|\sum_{i=1}^{n}\frac{1}{i} - \ln(n)\right| > \epsilon\ln(n)\right)$$

Using the bounds given in the problem, we have that

$$\sum_{i=1}^{n}\frac{1}{i} \geq \ln(n) \geq \sum_{i=1}^{n}\frac{1}{i} - 1$$

which implies that

$$0 \leq \sum_{i=1}^{n}\frac{1}{i} - \ln(n) \leq 1$$

Therefore,

$$P\left(\left|\sum_{i=1}^{n}\frac{1}{i} - \ln(n)\right| > \epsilon\ln(n)\right) \leq P\left(1 > \epsilon\ln(n)\right)$$
$$\leq \frac{1}{\epsilon\ln(n)} \to 0 \text{ using Markov's inequality}$$

Now, using the theorem on Page 23 of lecture 1, we have, since

$$Z_n \xrightarrow{p} 0 \text{ and } Y_n \xrightarrow{p} 1 \implies (Z_n, Y_n) \xrightarrow{p} (0, 1)$$

Therefore, using the continuous mapping theorem using the function $g\left(Z_n, Y_n\right) = Z_n + Y_n$ which is continuous everywhere, we have

$$\frac{T_n}{n\ln(n)} = Z_n + Y_n \xrightarrow{p} 0 + 1 = 1$$

Hence proved.