homework03

zhewei xie

2024-07-30

Problem 1: CDFs and PDFs

Part A

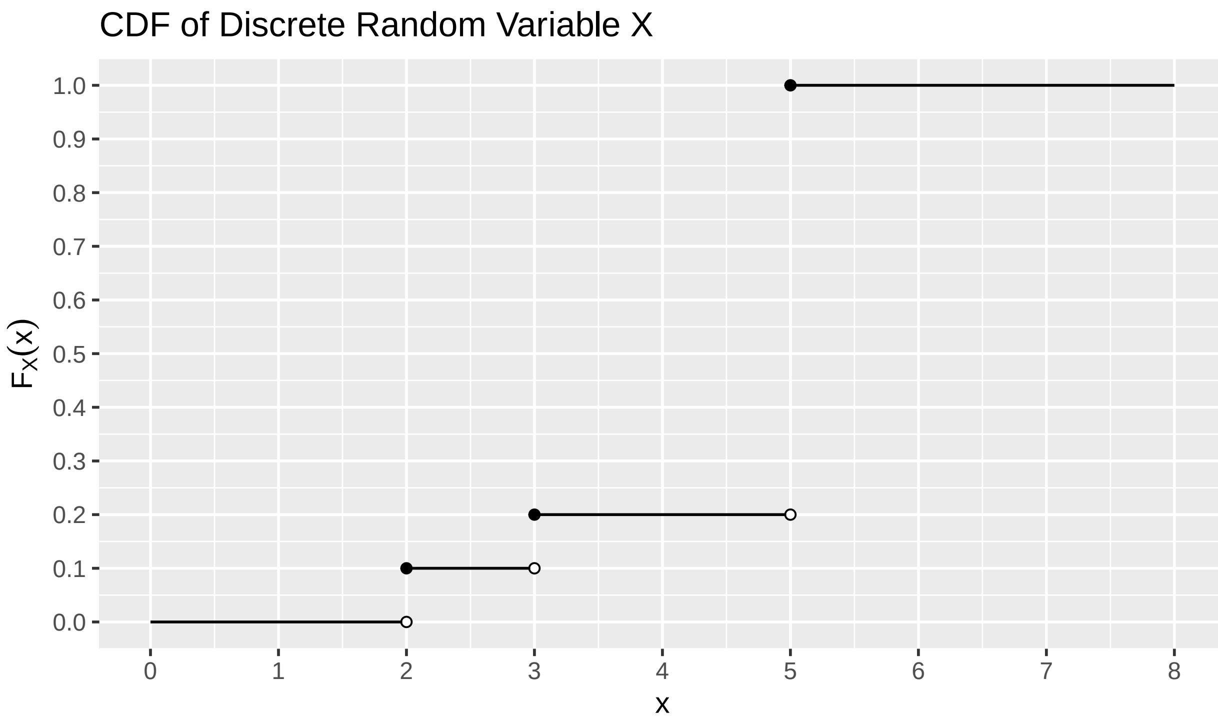

Figure 1. The cumulative distribution function, or CDF, of a discrete random variable X is defined as:
X=2 with probability 0.1
X=3 with probability 0.1
X=5 with probability 0.8

Given the definition in the Part A, the CDF of X is:

$$\mathrm{CDF}_X = P(X \leq x) = \begin{cases} 0, \text{ for } x < 2 \\ 0.1, \text{ for } 2 \leq x < 3 \\ 0.2, \text{ for } 3 \leq x < 5 \\ 1, \text{ for } 5 \leq x \end{cases}$$

therefore, it is clear that:

$P(2 < X \leq 4.5) = P(X \leq 4.5) - P(X \leq 2) = 0.2 - 0.1 = 0.1$

$P(2 \leq X < 4.5) = P(X < 4.5) - P(X < 2) = 0.2 - 0 = 0.2$

Part B

Given the definition in the Part B ($X\ U(0,1)$), the CDF of X is:

$$\text{CDF}_X = P(X \leq x) = \begin{cases} 0, & \text{for } x < 0 \\ x, & \text{for } 0 \leq x \leq 1 \\ 1, & \text{for } x > 1 \end{cases}$$

i. Compute $P(X^2 \leq 0.25)$

$P(X^2 \leq 0.25) = P(X \leq 0.5) = 0.5$

ii. For any number a, compute $P(X^2 \leq a)$

$$P(X^2 \leq a) = P(X \leq \sqrt{a}) = \begin{cases} \sqrt{a}, & \text{for } 0 \leq a \leq 1 \\ 1, & \text{for } a > 1 \end{cases}$$

iii. From (ii), find the PDF of the random variable $Y = X^2$

From (ii), we have the CDF of $Y = X^2$:

$$F_Y(y) = P(Y \leq y) = P(X^2 \leq y) = P(X \leq \sqrt{y}) = \begin{cases} \sqrt{y}, & \text{for } 0 \leq y \leq 1 \\ 1, & \text{for } y > 1 \end{cases}$$

therefore, it is clear that:

$\text{PDF}_Y = f_Y(y) = \frac{d}{dy}F_Y(y) = \frac{1}{2}y^{\frac{-1}{2}}$, for $0 < y \leq 1$

iv. Compute $E(Y)$ and $var(Y)$ directly from the PDF

Given (iii), it is clear that:

$E(Y) = \int_{-\infty}^{+\infty} y \cdot f_Y(y)dy = \int_0^1 \frac{1}{2}y^{\frac{1}{2}}dy = \frac{1}{3}$

For $var(Y)$, given the formula $var(Y) = E(Y^2) - (E(Y))^2$, it is necessary to calculate $E(Y^2)$ firstly:

$E(Y^2) = \int_{-\infty}^{+\infty} y^2 \cdot f_Y(y)dy = \int_0^1 \frac{1}{2}y^{\frac{3}{2}}dy = \frac{1}{5}$

Therefore, $var(Y) = \frac{1}{5} - (\frac{1}{3})^2 = \frac{4}{45}$

Problem 2: practice with expected value

Part A

For each standard normal random variable Z_i picked from $Z_1,\ldots,Z_d$ is said to follow $Z_i \sim N(0,1)$.

$f(Z_i) = \frac{1}{\sqrt{2\pi}}e^{-\frac{Z_i^2}{2}}$

$E(Z_i^2) = \int_{-\infty}^{+\infty} Z_i^2 \frac{1}{\sqrt{2\pi}}e^{-\frac{Z_i^2}{2}}dZ_i$

Because of

$(\frac{1}{\sqrt{2\pi}}e^{-\frac{Z_i^2}{2}})' = -Z_i(\frac{1}{\sqrt{2\pi}}e^{-\frac{Z_i^2}{2}})$,

and

$(\frac{1}{\sqrt{2\pi}}e^{-\frac{Z_i^2}{2}})'' = Z_i^2(\frac{1}{\sqrt{2\pi}}e^{-\frac{Z_i^2}{2}}) - (\frac{1}{\sqrt{2\pi}}e^{-\frac{Z_i^2}{2}})$

therefore,

$E(Z_i^2) = \int_{-\infty}^{+\infty}[(\frac{1}{\sqrt{2\pi}}e^{-\frac{Z_i^2}{2}})'' + (\frac{1}{\sqrt{2\pi}}e^{-\frac{Z_i^2}{2}})]dZ_i$

which is equal to:

$E(Z_i^2) = \int_{-\infty}^{+\infty}(\frac{1}{\sqrt{2\pi}}e^{-\frac{Z_i^2}{2}})''dZ_i + \int_{-\infty}^{+\infty}(\frac{1}{\sqrt{2\pi}}e^{-\frac{Z_i^2}{2}})dZ_i$

so,

$E(Z_i^2) = 0 + 1 = 1$

It also is clear that $E(Z_i) = 0$ and $var(Z_i) = 1$.

Because the formula $var(Z_i) = E(Z_i^2) - (E(Z_i))^2$, we also can find:

$E(Z_i^2) = var(Z_i) + (E(Z_i))^2 = 1 + 0^2 = 1$

For $X \stackrel{D}{=} Z_1^2 + \cdots + Z_d^2$, we can compute $E(X)$ as:

$E(X) = E(Z_1^2) + \cdots + E(Z_d^2) = \underbrace{1 + \cdots + 1}_{d} = d$

Optional Practice Exercise

Let's denote the PDF of standard normal distribution as $\phi(z)$.

Because of

$\phi'(z) = -z\phi(z)$

and

$\phi''(z) = z^2\phi - \phi$

and

$\phi^{(3)}(z) = 2z\phi(z) - z^3\phi + z\phi$

and

$\phi^{(4)}(z) = z^4\phi - 6z^2\phi + 3\phi$

therefore,

$E(Z_i^4) = \int_{-\infty}^{+\infty} Z_i^4\phi(Z_i)dZ_i = \int_{-\infty}^{+\infty}[\phi^{(4)}(Z_i) + 6Z_i^2\phi(Z_i) - 3\phi(Z_i)]dZ_i$

which is equal to:

$E(Z_i^4) = \int_{-\infty}^{+\infty} \phi^{(4)}(Z_i)dZ_i + 6\int_{-\infty}^{+\infty} Z_i^2\phi(Z_i)dZ_i - 3\int_{-\infty}^{+\infty} \phi(Z_i)dZ_i = 0 + 6 - 3 = 3$

So $var(Z_i^2) = E(Z_i^4) - (E(Z_i^2))^2 = 3 - 1^2 = 2$

Since X is the sum of the squares of d independent standard normal random variables, the variance of X is the sum of the variances of Z_i^2:

$var(X) = \underbrace{var(Z_1^2) + \cdots + var(Z_d^2)}_{d} = 2d$

Part B

Given the situation that Markov faces, it is easily to calculate the expected time both by walk and by scooters:

$T_{walk} = \frac{2}{5}$

$T_{scooter} = \frac{2}{10}$

Therefore, the expected time T that it takes him to get to class on a random winter day is:

$E(T) = T_{walk} \times 0.4 + T_{scooter} \times (1 - 0.4) = 0.16 + 0.12 = 0.28$

It is clear that $E(T) = 0.28 \neq \frac{2}{8}$, which implies that the calculation of Markov is wrong. The reason is $E(2/V) \neq 2/E(V)$.

Problem 3: inverse CDF

$\because U \sim U(0, 1)$

$\therefore f_U(u) = \begin{cases} 1, & \text{for } 0 \leq u \leq 1 \\ 0, & \text{otherwise} \end{cases}$

$\therefore F_U(u) = \begin{cases} 0, & \text{for } u < 0, \\ u, & \text{for } 0 \leq u \leq 1 \\ 1, & \text{for } u > 1 \end{cases}$

$\because X = F^{-1}(U)$

$\therefore F_X(x) = P(F^{-1}(U) \leq x)$

$\therefore F_X(x) = P(U \leq F(x))$

$\therefore F_X(x) = F(x)$

$\therefore f_X(x) = \frac{d}{dx}F_X(x) = \frac{d}{dx}F(x) = f(x)$

Problem 4: simulation

Part A

$\because X_n \sim \text{Binominal}(N, P)$

$\therefore E(X_N) = NP$ and $var(X_N) = NP(1 - P)$

Let $\hat{p}N = X_N/N$ denote the proportion of observed successes.

Based on the formulas:

$E(aX) = aE(X)$ and $var(aX) = a^2 var(X)$

it it clear that:

$E(\hat{p}_N) = E(X_N/N) = \frac{1}{N}E(X_N) = \frac{1}{N}NP = P$

$var(\hat{p}_N) = var(X_N/N) = (\frac{1}{N})^2 var(X_N) = \frac{1}{N^2}NP(1 - P) = \frac{P(1-P)}{N}$

$\therefore sd(\hat{p}_N) = \sqrt{\frac{P(1-P)}{N}}$

Part B

Table 1: This is the result (rounded to 3 decimal places) of a Monte Carlo simulation of 1000 realizations of the random variable $\hat{p}_5$, assuming that the true $P = 0.5$.

mean	standard deviation
0.502	0.224

Based on the Part A, it is clear that:

the theoretical mean is $E(\hat{p}_N) = P = 0.5$

and the theoretical standard deviation is $sd(\hat{p}_N) = \sqrt{\frac{P(1-P)}{N}} = \sqrt{\frac{0.5\times0.5}{5}} \approx 0.2236$.

Therefore, the Monte Carlo mean and standard deviation of the simulated $\hat{p}_5$'s agree, at least approximately, with the theoretical mean and standard deviation computed from the result in (A).

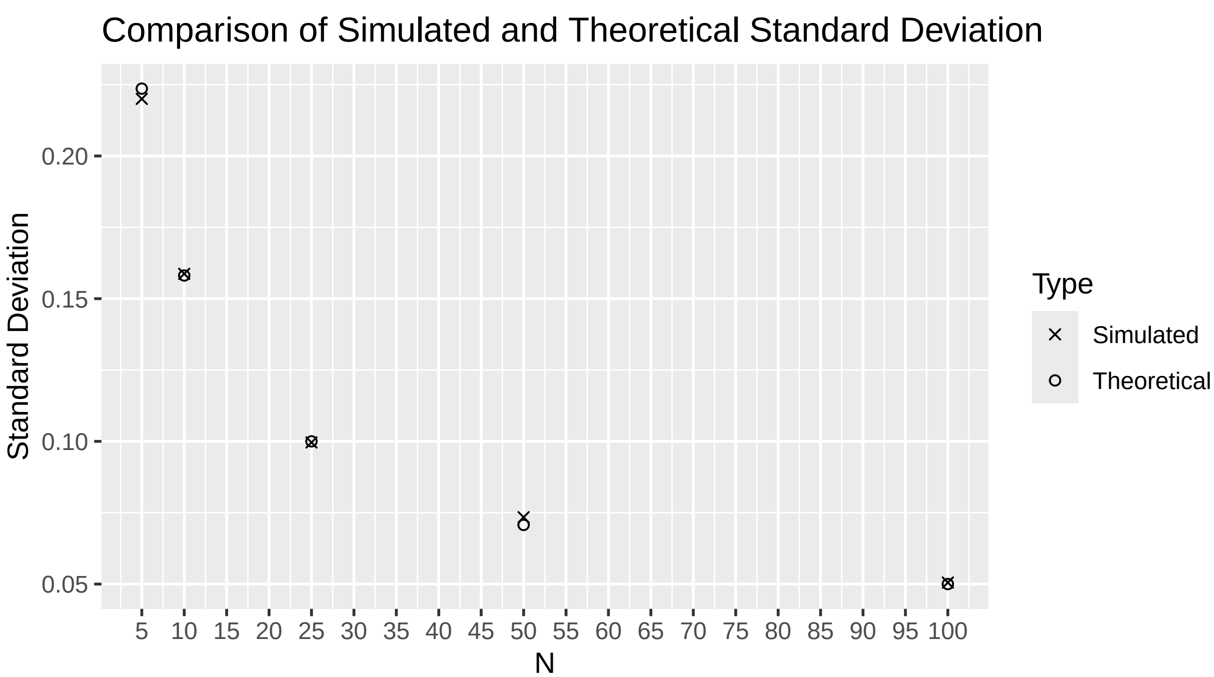

Figure 2. The plot shows comparison of Monte Carlo simulated and theoretical standard deviations. Through this figure, the Monte Carlo standard deviation of the simulation, which simulated 1000 realizations, agrees, at least approximately, with the theoretical standard deviation computed from the result in (A). This is a verification of the Law of Large Numbers. Besides, there is a pattern that the standard deviation decreases by the N increases, which demonstrates that as the sample size increases, the mean of the observed values will gradually converge to the expected probability. This is a verification of the Central Limit Theorem.

Problem 5: more PDF/CDF practice

For each standard normal random variable X_i picked from $X_1, \cdots, X_N$ is said to follow $X_i \sim Exp(\lambda)$ for $x \geq 0$, and $f(x) = 0$ otherwise.

So the cumulative distribution function of X is $F_X(x) = \int_0^x \lambda e^{-\lambda t} dt = 1 - e^{-\lambda x}$.

Given $Y_N = \max\{X_1, \cdots, X_N\}$, it is clear that:

$F_{Y_N}(y) = P(Y \leq y) = P(X_1 \leq y, \cdots, X_N \leq y)$

Because $X_1, \cdots, X_N$ are a set of N independent samples, it implies that:

$F_{Y_N}(y) = P(X_1 \leq y) \times \cdots \times P(X_N \leq y)$

Then,

$F_{Y_N}(y) = \underbrace{(1 - e^{-\lambda y}) \times \cdots \times (1 - e^{-\lambda x})}_{N} = (1 - e^{-\lambda y})^N$

Therefore, the PDF of Y_N for fixed N is:

$f_{Y_N}(y) = \frac{d}{dy}(1 - e^{-\lambda y})^N = \lambda N(1 - e^{-\lambda y})^{N-1}e^{-\lambda y}$