Preparation to Midterm Next week:

1. Sample Midterm
2. Homework 1 − 4
3. Section Notes Week 2 − 5

Study Guide + Solution

Please check the messages in Inbox on Canvas for the past few weeks.

Ⓘ Discrete Random Variable

<u>Def</u>: A Random Variable is a function from Ω to $\mathbb{R}$.

In other words,

$$X : \Omega \to \mathbb{R},$$

$\forall \omega \in \Omega$, $X(\omega) = x \in \mathbb{R}$, then X is a R.V.

For a discrete R.V. X:

1. Support of X, denoted by S_X, refers to the set of values in which X can take.

2. Probability Mass Function (P.M.F) of X:

$$P_X(k) \overset{Def}{=} \mathbb{P}(X=k) \quad \text{for } k \in S_X$$

satisfies $\sum_{k \in S_X} \mathbb{P}(X=k) = 1$

3. Cumulative Distribution Function (C.D.F.) of X:

$$F_X(x) \overset{Def}{=} \mathbb{P}(X \le x) = \sum_{\substack{k \le x \\ k \in S_X}} \mathbb{P}(X=k)$$

e.g. Consider a R.V. X with P.M.F.

$$\mathbb{P}(X=k) = c\left(\frac{1}{3}\right)^k \qquad k = 0, 1, 2, \cdots$$

where c is a constant.

1. Support: $S_X = \{0, 1, 2, \cdots\} = \{0\} \cup \mathbb{N}$

2. P.M.F: $p_X(k) = c\left(\frac{1}{3}\right)^k \qquad k \in S_X$

$$1 = \sum_{k=0}^{\infty} p_X(k) = \sum_{k=0}^{\infty} c\left(\frac{1}{3}\right)^k = c\sum_{k=0}^{\infty}\left(\frac{1}{3}\right)^k$$

Geometric Series: $\displaystyle\sum_{k=0}^{\infty} p^k = \begin{cases} \dfrac{1}{1-p} & |p|<1 \\ & |p| \geq 1. \end{cases}$

$$1 = c\sum_{k=0}^{\infty}\left(\frac{1}{3}\right)^k \overset{p=1/3}{=} c\cdot\frac{1}{1-1/3} \quad \Rightarrow \quad c = \frac{2}{3}$$

3. C.D.F: given that $x \geq 0$,

$$\mathbb{P}(X \leq x) = \mathbb{P}(X \leq \lfloor x \rfloor) + \mathbb{P}(\lfloor x \rfloor < X \leq x)$$

$\lfloor x \rfloor$: the largest integer less than or equal to x

e.g. $x = 3.5$, $\lfloor x \rfloor = 3$; $y = -3.5$, $\lfloor y \rfloor = -4$.

Note that no integer in the interval $(\lfloor x \rfloor, x]$

$\Rightarrow \mathbb{P}(\lfloor x \rfloor < X \leq x) = 0$

partial sums for Geometric Series: $\displaystyle\sum_{n=0}^{N} p^n = \begin{cases} \dfrac{1-p^{N+1}}{1-p} & p \neq 1 \\ N+1 & p = 1 \end{cases}$

$$F_X(x) = \mathbb{P}(X \le \lfloor x \rfloor)$$

$$= \sum_{k=0}^{\lfloor x \rfloor} c\left(\frac{1}{3}\right)^k = \frac{2}{3} \cdot \frac{1-\left(\frac{1}{3}\right)^{\lfloor x \rfloor + 1}}{1 - 1/3} = 1 - \left(\frac{1}{3}\right)^{\lfloor x \rfloor + 1}$$

//

② Special Distribution for Discrete R.V.

Brian Wainwright is shooting a basketball, and all shoots are independent and the probability that he makes the shot is p $(0<p<1)$.

(1) In one shoot, event $A = \{\text{Brian makes the shoot}\}$

$$1_A = \begin{cases} 1 & \text{if } A \text{ happens} \\ 0 & \text{if } A^c \text{ happens} \end{cases}$$ (Indicator function for the event A).

$$\mathbb{P}(1_A = 1) = \mathbb{P}(\text{Brian makes the shoot}) = p$$

$$\mathbb{P}(1_A = 0) = 1 - p$$

$$\Rightarrow 1_A \sim \text{Bernoulli}(p)$$

(2) Brian shoots the ball for n times,

X = the number of shots he make. For $k = 0, 1, \cdots, n$

$$\mathbb{P}(X=k) = \mathbb{P}(\text{Brian makes } k \text{ out of } n \text{ shots})$$

$$= \binom{n}{k} p^k (1-p)^{n-k}$$

$X \sim \text{Binomial}(n, p)$ $\quad S_X = \{0, 1, \cdots, n\}$

(3) Y = the number of shots taken to get the first shot made.

$$\mathbb{P}(Y=k) = \underbrace{(1-p)^{k-1}}_{\text{failure in the first } (k-1) \text{ shots}} \, p \rightarrow \text{success in the } k\text{-th shot}$$

$Y \sim$ Geometric (p) $\quad S_Y = \{0\} \cup \mathbb{N}$

(4) Z = the number of shots needed to get the k-th shot. For $n \geq k$,

$\{Z=n\} = \{k-1$ success in the first $(n-1)$ shots, and success in the n-th shots$\}$

$k-1$ success / $n-k$ failures $\quad$ k-th success

$$\mathbb{P}(Z=n) = \binom{n-1}{k-1} p^{k-1} (1-p)^{n-k} \cdot p$$

$$= \binom{n-1}{k-1} p^{k} (1-p)^{n-k} \qquad S_Z = \{k, k+1, \cdots\}$$

(5) Suppose Brian continues shooting the ball for two hours, R = the number of shots he makes in 2 hrs.

λ = the mean number of shots he make.

$R \sim \text{Poi}(\lambda)$ and $\mathbb{P}(R=k) = e^{-\lambda} \frac{\lambda^k}{k!}$ $\qquad S_R = \{0\} \cup \mathbb{N}$.

e.g. Unreliable COVID Test

- A person $\begin{cases} H = \{\text{healthy}\} & \mathbb{P}(H) = .95 \\ H^c = \{\text{COVID}\} & \mathbb{P}(H^c) = .05 \end{cases}$

- Given H : $\begin{cases} \text{misclassify as COVID w.p. } .2 \\ \text{correctly classify as healthy w.p. } 1-(.2) = .8 \end{cases}$

  Given H^c: misclassify as healthy w.p. .15

- Doing the COVID test for three times.

  N = # of tests indicating he is healthy

  $N \mid H \sim \text{Bin}(3, .8)$, $N \mid H^c \sim \text{Bin}(3, .15)$

(a) $$\mathbb{P}(N=2) = \underbrace{\mathbb{P}(N=2 \mid H)}_{\binom{3}{2}(.8)^2(.2)^1} \cdot \underbrace{\mathbb{P}(H)}_{(.95)} + \underbrace{\mathbb{P}(N=2 \mid H^c)}_{\binom{3}{2}\cdot(.15)^2(.85)^1} \cdot \underbrace{\mathbb{P}(H^c)}_{(.05)}$$

(b) Given $N=2$, what is the prob. that he is healthy? Bayes' thm

$$\mathbb{P}(H \mid N=2) = \frac{\mathbb{P}(N=2 \mid H)\,\mathbb{P}(H)}{\mathbb{P}(N=2)}$$

$$= \frac{\binom{3}{2}(.8)^2(.2)^1(.95)}{\binom{3}{2}(.8)^2(.2)^1(.95) + \binom{3}{2}(.15)^2(.85)^1(.05)} = .9921$$

図

Remark: Why is it NOT TRUE to calculate in this way:

for each test, define $A = \{\text{one test indicates healthy}\}$

and $p = \mathbb{P}(A) = \mathbb{P}(A|H)\,\mathbb{P}(H) + \mathbb{P}(A|H^c)\,\mathbb{P}(H^c)$

$= (.8)(.95) + (.15)(.05)$ (Law of total prob.)

and $N \sim \text{Bin}(3, p)$

Solution: Define the following two events:

$A_1 = \{\text{The first test indicates healthy}\}$

$A_2 = \{\text{the second test indicates healthy}\}$

$\mathbb{P}(A_1) = \mathbb{P}(A_2) = (.8)(.95) + (.15)(.05) = .7675$

but $\mathbb{P}(A_1 A_2)$

$= (.8)^2(.95) + (.15)^2(.05) \neq \mathbb{P}(A_1)\,\mathbb{P}(A_2)$

In other words, the results of the two tests are conditionally independent given H or H^c, but NOT independent, so it contradicts with the assumption of Binomial Distribution that each test is independent.