Lecture 9: Discrete Random Variables - Part III

Satyajit Thakor
IIT Mandi

Discrete random variables

Ex: let $n=3$, $k=1$ → $\binom{3}{1}$ sequences with 1s:
$(S,F,F), (F,S,F), (F,F,S)$

consider (S,F,F):
$P(X_1=1, X_2=0, X_3=0)$
$= P(X_1=1)\cdot P(X_2=0)\cdot P(X_3=0) = p(1-p)^2$

Let $X_i \sim \text{Bern}(p)$, $S=1$, $F=0$
success failure

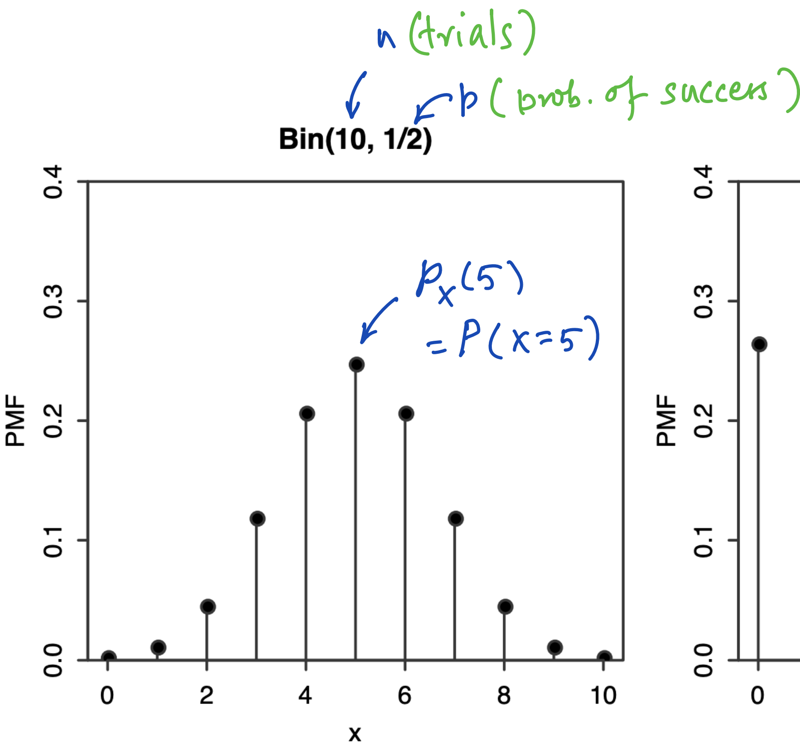
- **Theorem:** If $X \sim \text{Bin}(n, p)$, then the PMF of X is

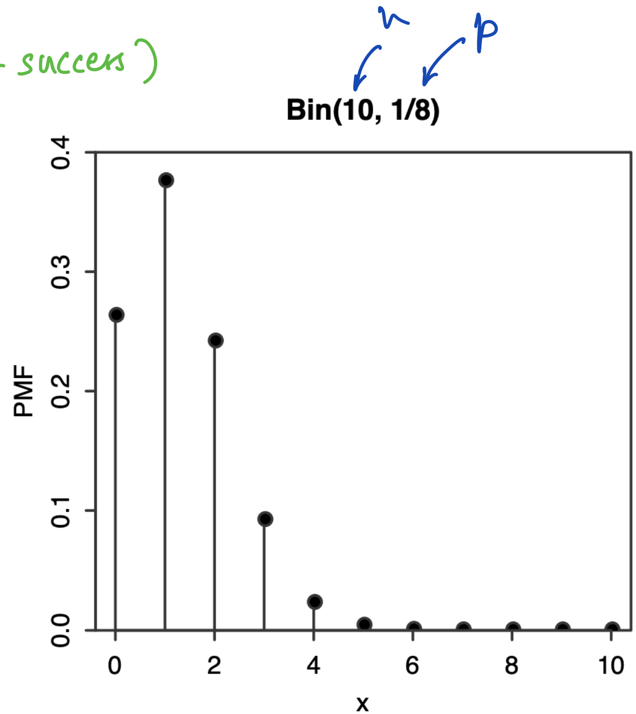
$$p_X(k) = P(X = k) = \binom{n}{k} p^k (1-p)^{n-k}$$

for $k = 0, 1, \ldots, n$ and $p_X(k) = P(X = k) = 0$ otherwise.

- Prob. that $X=k$ is the prob. that in n Bernoulli trials we have k successes (and $n-k$ failures)
- There are $\binom{n}{k}$ such ways (sequences/strings) such that k successes can occur in n Bernoulli trials.
- Prob. of occurence of such a sequence is $p^k \cdot (1-p)^{n-k}$
- Hence, $P(X=k) = \binom{n}{k} p^k (1-p)^{n-k}$ for $k=0,1,\ldots,n$ and
$= 0$ otherwise

Discrete random variables





- Note that the PMF of the $\text{Bin}(10, 1/2)$ distribution is symmetric about 5, but when the success probability is not $1/2$, the PMF is skewed.

Discrete random variables

i.e., $0 < p < 1$

"open interval"

$(0,1) \triangleq \{x \in \mathbb{R} : 0 < x < 1\}$

- Consider a sequence of independent Bernoulli trials, each with the same success probability $p \in (0, 1)$, with trials performed until a success occurs. Let X be the number of failures before the first successful trial. Then X has the Geometric distribution with parameter p; denoted $X \sim \text{Geom}(p)$.
- Theorem: If $X \sim \text{Geom}(p)$, then the PMF of X is

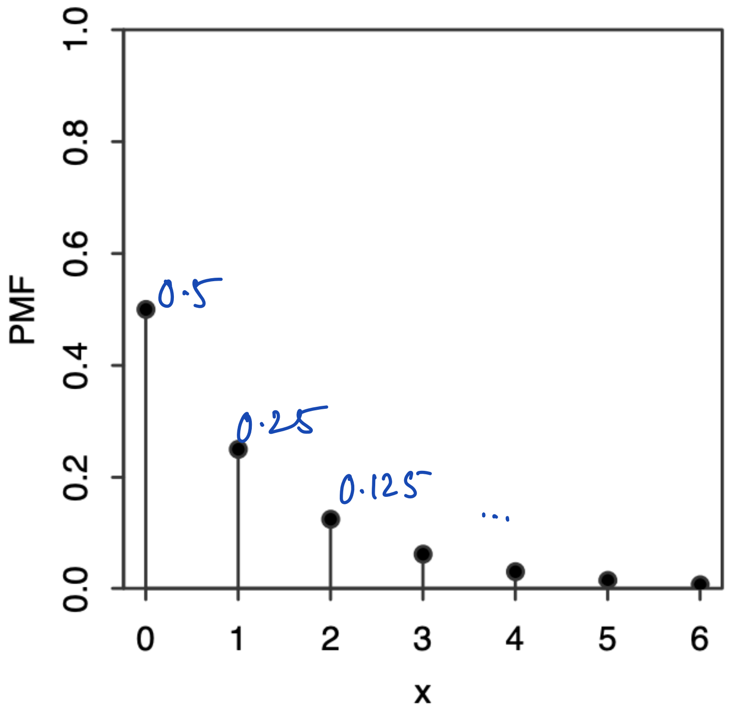
$$P(X = k) = (1-p)^k p, \qquad \text{for } k = 0, 1, 2, \ldots$$

- Note that, the support of a geometric r.v. has infinite cardinality.

– For $0 < p < 1$, the value $(1-p)^k \cdot p$ is always a positive number in the set $(0,1)$.

Discrete random variables

- Geom(1/2):



- Theorem: Geom(p) is a valid PMF.

Proof: We need to verify whether the "prob." function is non-negative and sums up to 1.

Discrete random variables

- Note that $p_X(k) \geq 0$. Now we only need to verify:

$$\sum_{k \in \mathbb{R}} p_X(k) = 1.$$

$$\sum_{k \in \mathbb{R}} p_X(k) = \sum_{k=0}^{\infty} p_X(k)$$

$$= \sum_{k=0}^{\infty} p(1-p)^k$$

$$= p \underbrace{\sum_{k=0}^{\infty} (1-p)^k}_{\text{geometric series.}}$$

$$= p \cdot \frac{1}{(1-1+p)}$$

$$= 1.$$

Geometric series:

Let $S_n = 1 + y + y^2 + \ldots + y^n$, $y \in (0,1)$

$\Rightarrow y \cdot S_n = \quad y + y^2 + \ldots + y^n + y^{n+1}$

$\Rightarrow S_n(1-y) = 1 - y^{n+1}$

$\Rightarrow S_n = \dfrac{1-y^{n+1}}{1-y}$

As $n \to \infty$, we have $y^{n+1} \to 0$ since $y \in (0,1)$.

$\Rightarrow \quad S_\infty = \dfrac{1}{1-y} = \displaystyle\sum_{k=0}^{\infty} y^k$

Discrete random variables

- Let C be a finite, non-empty set of numbers. Choose one of these numbers uniformly at random (i.e., all values in C are equally likely). Call the chosen number X. Then X is said to have the Discrete Uniform distribution with parameter C; we denote this by $X \sim \mathrm{DUnif}(C)$.
- The PMF of $X \sim \mathrm{DUnif}(C)$ is

$$P(X = x) = \frac{1}{|C|}$$

- Note that, for any $A \subseteq C$,

$$P(X \in A) = \frac{|A|}{|C|}.$$

- Example: Imagine an n-sided fair dice.

Discrete random variables

- Many random phenomena observed can be modelled as Poisson distribution, e.g., the number of customers arriving at a shopping centre within a fixed interval.
- An r.v. X has the Poisson distribution with parameter λ, where $\underbrace{\lambda > 0}_{\text{any positive real no.}}$, if the PMF of X is

$$P(X = k) = \frac{e^{-\lambda}\lambda^k}{k!}, \quad k = 0, 1, 2, \ldots$$

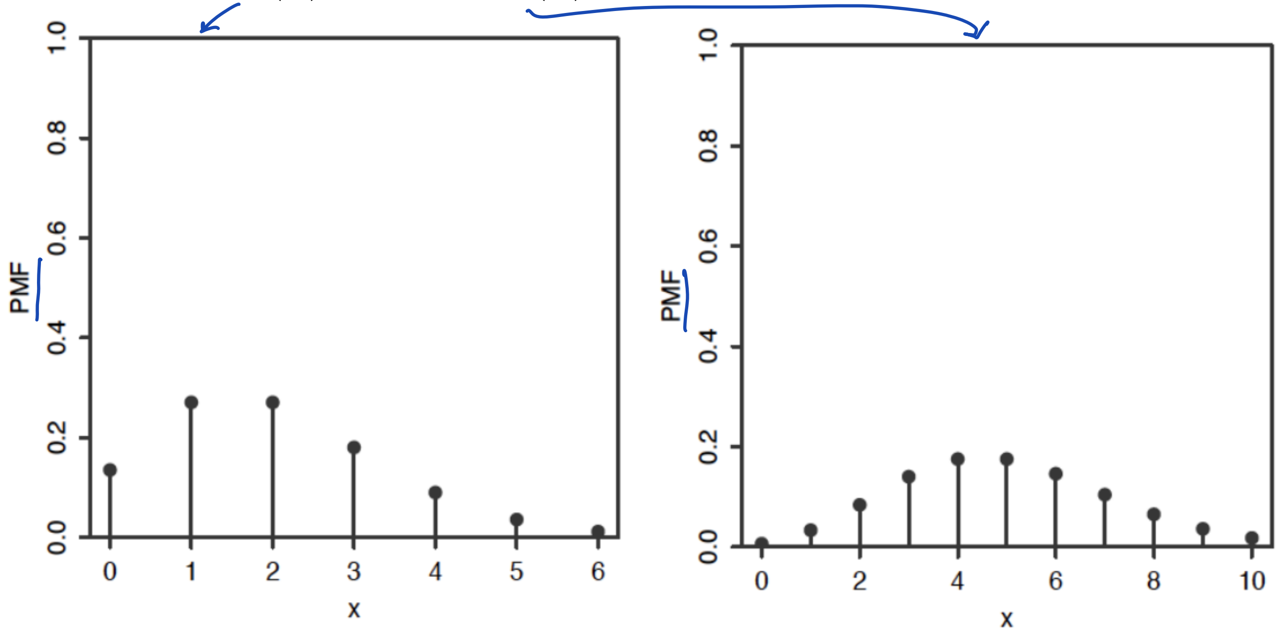
and 0 otherwise.
- Written as $X \sim \mathrm{Pois}(\lambda)$.

Other examples of Poisson dist.: no. of mutations in DNA / unit lenght.
no. of stars in a unit space.
no. of photons reaching a telescope surface.

Discrete random variables

λ λ

- Example: Pois(2) and Pois(5).



- Theorem: The Poisson distribution is a valid distribution.

Proof: First note that $P(X=k) = \dfrac{e^{-\lambda}\lambda^k}{k!}$ is non-negative for all $k = 0, 1, 2, \ldots$

Discrete random variables

- Now, $\sum_{k \in \mathbb{R}} p_X(k) = \sum_{k=0}^{\infty} \frac{e^{-\lambda} \lambda^k}{k!}$

$$= e^{-\lambda} \underbrace{\sum_{k=0}^{\infty} \frac{\lambda^k}{k!}}_{\text{Taylor series (expansion) of } e^{\lambda}.} = e^{-\lambda} \cdot e^{\lambda} = 1.$$

Recall: Taylor series of a infinitely differentiable function $f(x)$ centered at a:

$$f(x) = f(a) + \frac{f^{(1)}(a)}{1!}(x-a) + \frac{f^{(2)}(a)}{2!}(x-a)^2 + \dots$$

$$= \sum_{n=0}^{\infty} \frac{f^{(n)}(a)}{n!}(x-a)^n \text{ where } f^{(n)}(a) = \left.\frac{d^n f(x)}{dx^n}\right|_{x=a}$$

Let $f(x) = e^x$, $a = 0$. Then

$e^x = e^a = e^0 = 1$

$$e^x = \sum_{n=0}^{\infty} \frac{1}{n!} x^n$$

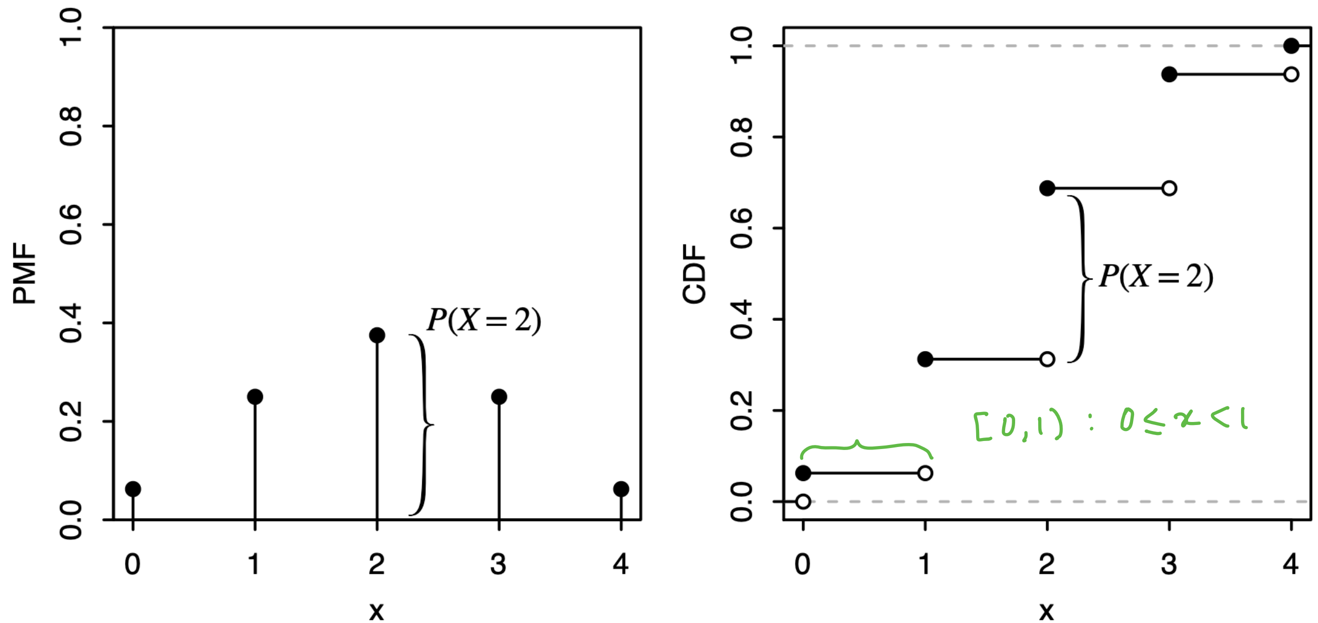
Discrete random variables

- Another function that describes the distribution of an r.v. is the cumulative distribution function (CDF).
- Unlike the PMF, which only discrete r.v.s have, the CDF is defined for all r.v.s.

- The cumulative distribution function (CDF) of an r.v. X is the function F_X given by $F_X(x) = P(X \leq x)$.

Discrete random variables

- Example: $\mathrm{Bin}(4, 1/2)$



- From PMF to CDF $F_X(1.5) = P(X \le 1.5) = P(X=0) + P(X=1) + P(1 < X \le 1.5)$

$$= (1/2)^4 + 4(1/2)^4 + 0 = 5/16$$

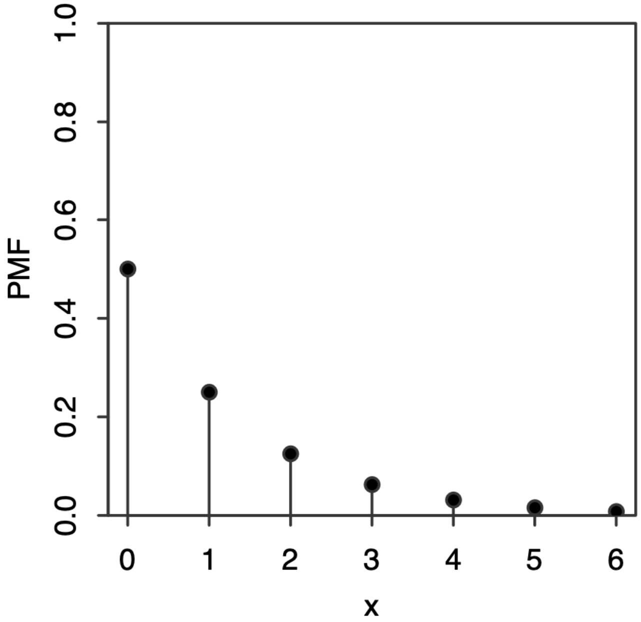
- From CDF to PMF $p_X(1) = P(X=1) = P(X \le 1) - \underbrace{P(X < 1)}_{=P(X=0)+P(0<X<1) \; = 1/16 + 0}$

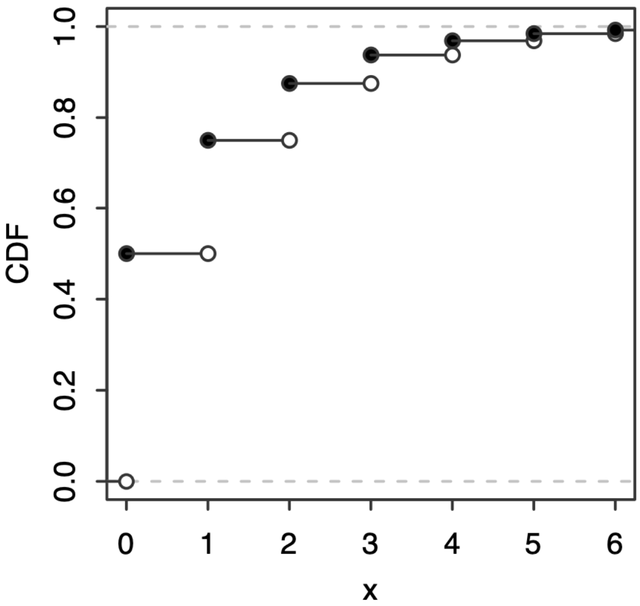
$$= \frac{5}{16} - \frac{1}{16} = \frac{4}{16}$$

Discrete random variables

- Example: Geom(1/2)

$p_X(k) = (1-p)^k \cdot p, \quad k = 0, 1, 2, \dots$





Note

- Source of figures: reference books