Probability and Random Variables (ECE313/317)

Fatima Taousser

Departement of Electrical Engineering and Computer Sciences, UTK
ftaousse@utk.edu

Discrete Random Variables
Fall 2023

Probability: Discrete Random variables

- Mathematical background
- Discrete Random variables - Probability mass function (PMF).
  - Uniform
  - Bernoulli
  - Binomial
  - Geometric
  - Poisson
- The cumulative distribution function (CDF)

Mathematical background:

- Arithmetic sequence: An arithmetic sequence, or an arithmetic progression $\{u_n\}_{n\in\mathbb{N}}$, is a set of numbers where the difference between two consecutive terms is a constant number r, such that $u_{n+1} = u_n + r$. For example, the sequence: $1, 4, 7, 10, 13, 16, 19, \ldots$ is an arithmetic sequence since the difference between each consecutive numbers is $r = 3$.
Formally, the general term of an arithmetic sequence is:

$$u_{n+1} = u_n + r; \qquad u_n = u_0 + nr; \quad n = 0, 1, 2, \ldots$$

- Arithmetic series: is the sum of an arithmetic sequence: $\sum_{i=0}^{n} u_i$

$$\begin{aligned} \sum_{i=0}^{n} u_i &= u_0 + u_1 + u_2 + u_3 + u_4 + \ldots + u_{n-1} + u_n \quad ---> (1) \\ \sum_{i=0}^{n} u_i &= u_n + u_{n-1} + u_{n-2} + u_{n-3} + u_{n-4} + \ldots + u_1 + u_0 \quad ---> (2) \end{aligned}$$

By adding (1) to (2), we get

$$\begin{aligned} 2\sum_{i=0}^{n} u_i &= (u_0 + u_n) + (u_1 + u_{n-1}) + (u_2 + u_{n-2}) + \ldots + (u_{n-1} + u_1) + (u_n + u_0) \\ &= [u_0 + u_n] + [(u_0 + r) + (u_n - r)] + [(u_0 + 2r) + (u_n - 2r)] + \ldots + [u_0 + u_n] \\ &= (n+1)(u_0 + u_n) \end{aligned}$$

$$\Rightarrow \sum_{i=0}^{n} u_i = \frac{n+1}{2}(u_0 + u_n) = \frac{\#\text{terms}}{2}(\text{first term} + \text{last term})$$

Mathematical background:

• Geometric sequence: A geometric sequence, or a geometric progression $\{u_n\}_{n\in\mathbb{N}}$, is a set of numbers where the ratio r of two consecutive terms is constant, such that $u_{n+1} = r \times u_n$. For example, the sequence: $1, 2, 4, 8, 16, 32, 64\ldots$ is a geometric sequence with ratio $r = 2$. Formally, the general term of a geometric sequence is:

$$u_{n+1} = u_n \times r; \qquad u_n = u_0 \times r^n; \quad n = 0, 1, 2, \ldots$$

• Geometric series: is the sum of a geometric sequence: $\sum_{i=0}^{n} u_i$

$$\begin{aligned}\sum_{i=0}^{n-1} u_i &= u_0 + u_1 + u_2 + u_3 + u_4 + \ldots + u_{n-2} + u_{n-1} \\ &= u_0 + u_0 r + u_0 r^2 + u_0 r^3 + u_0 r^4 + \ldots + u_0 r^{n-2} + u_0 r^{n-1} \quad ---> (1)\end{aligned}$$

$$r\sum_{i=0}^{n-1} u_i = u_0 r + u_0 r^2 + u_0 r^3 + u_0 r^4 + u_0 r^5 + \ldots + u_0 r^{n-1} + u_0 r^n \quad ---> (2)$$

By subtracting (2) from (1), we get

$$(1-r)\sum_{i=0}^{n} u_i = u_0 - u_0 r^n = u_0(1-r^n) \Rightarrow \sum_{i=0}^{n-1} u_i = u_0\frac{1-r^n}{1-r} = (\text{first term}) \times \left(\frac{1-r^{\#\text{terms}}}{1-r}\right)$$

Mathematical background:

- Sum of squares: $1^2 + 2^2 + 3^2 + 4^2 + 5^2 + \ldots + n^2 = ?$

- We have: $a^3 - b^3 = (a-b)(a^2 + ab + b^2)$.

$$\Rightarrow\ n^3 - (n-1)^3 = [n - (n-1)][n^2 + n(n-1) + (n-1)^2] = 1 \times [3n^2 - 3n + 1]$$

Consider the following process now:

$$\begin{aligned}
n^3 - (n-1)^3 &= 3n^2 - 3n + 1 \\
(n-1)^3 - (n-2)^3 &= 3(n-1)^2 - 3(n-1) + 1 \\
(n-2)^3 - (n-3)^3 &= 3(n-2)^2 - 3(n-2) + 1 \\
\vdots &= \vdots \\
2^3 - 1^3 &= 3(2)^2 - 3(2) + 1 \\
1^3 - 0^3 &= 3(1)^2 - 3(1) + 1
\end{aligned}$$

By adding all the above equations, we get

$$n^3 = 3\sum_1^n n^2 - 3\sum_1^n n + \sum_1^n 1 = 3\sum_1^n n^2 - 3\frac{n(n+1)}{2} + n \Rightarrow \sum_1^n n^2 = \frac{1}{3}\left(n^3 + 3\frac{n(n+1)}{2} - n\right)$$

$$\Rightarrow\ \sum_1^n n^2 = \frac{n(n+1)(2n+1)}{6}$$

Mathematical background:

- Combination properties:
  - $C_k^{n-1} + C_{k-1}^{n-1} = C_k^n$
  - $kC_k^n = nC_{k-1}^{n-1}$
  - $k^2 C_k^n = n(n-1)C_{k-2}^{n-2} + nC_{k-1}^{n-1}$

- Proof:

- $$\begin{aligned} C_k^{n-1} + C_{k-1}^{n-1} &= \frac{(n-1)!}{k!(n-1-k)!} + \frac{(n-1)!}{(k-1)!((n-1)-(k-1))!} \\ &= \frac{(n-1)!}{k!(n-1-k)!} + \frac{(n-1)!}{(k-1)!(n-k)!} \\ &= \frac{(n-1)!}{k(k-1)!(n-k-1)!} + \frac{(n-1)!}{(k-1)!(n-k)(n-k-1)!} \\ &= \frac{(n-1)!(n-k)}{k(k-1)!(n-k)(n-k-1)!} + \frac{(n-1)!k}{k(k-1)!(n-k)(n-k-1)!} \\ &= \frac{(n-1)![(n-k)+k]}{k!(n-k)!} = \frac{n(n-1)!}{k!(n-k)!} = \frac{n!}{k!(n-k)!} \\ &= C_k^n \end{aligned}$$

Mathematical background:

- $kC_k^n = k\frac{n!}{k!(n-k)!} = k\frac{n(n-1)!}{k(k-1)!(n-k)!} = \frac{n(n-1)!}{(k-1)!(n-k)!}$

$$= \frac{n(n-1)!}{(k-1)!(n-k-1+1)!} = n\frac{(n-1)!}{(k-1)!((n-1)-(k-1))!} = nC_{k-1}^{n-1}$$

- $k^2C_k^n = k[kC_k^n] = knC_{k-1}^{n-1} = n(k-1+1)C_{k-1}^{n-1} = n[(k-1)C_{k-1}^{n-1} + C_{k-1}^{n-1}]$

$$= n[(n-1)C_{k-2}^{n-2} + C_{k-1}^{n-1}] = n(n-1)C_{k-2}^{n-2} + nC_{k-1}^{n-1}$$

Mathematical background:

- $e^x = \sum_{n=0}^{\infty} \frac{x^n}{n!} = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \ldots + \frac{x^n}{n!} + \ldots$

- $\lim_{n\to\infty} \left(1 + \frac{1}{n}\right)^n = e?$

$$\lim_{n\to\infty} \left(1 + \frac{1}{n}\right)^n = \lim_{n\to\infty} e^{\log\left(1+\frac{1}{n}\right)^n} = \lim_{n\to\infty} e^{n\log\left(1+\frac{1}{n}\right)} = \lim_{n\to\infty} e^{\frac{\log\left(1+\frac{1}{n}\right)}{\frac{1}{n}}}$$

$$= e^{\lim_{n\to\infty} \frac{\log\left(1+\frac{1}{n}\right)}{\frac{1}{n}} = \frac{0}{0} \text{ undefined limit} \rightarrow \text{we can use l'Hopital's law}}$$

$$= e^{\lim_{n\to\infty} \frac{[\log\left(1+\frac{1}{n}\right)]'}{[\frac{1}{n}]'} = \lim_{n\to\infty} \frac{\frac{-1}{n^2(1+\frac{1}{n})}}{\frac{-1}{n^2}} = \lim_{n\to\infty} \frac{1}{1+\frac{1}{n}}} = e^1 = e$$

- $\lim_{n\to\infty} \left(1 + \frac{a}{n}\right)^n = e^a$

\- Let $\frac{a}{n} = \frac{1}{x} \Rightarrow \lim_{n\to\infty} \left(1 + \frac{a}{n}\right)^n = \lim_{x\to\infty} \left(1 + \frac{1}{x}\right)^{ax} = \lim_{x\to\infty} \left(\left(1 + \frac{1}{x}\right)^x\right)^a = e^a$

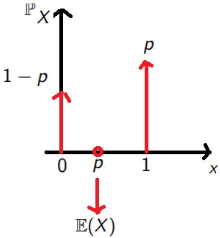
Discrete Random Variables

- Bernoulli with parameter $p \in [0,1]$: Model a trials that results in Success/Failure, Heads/Tails, etc.

$$X = \begin{cases} 1; & \text{with probability } p \\ 0; & \text{with probability } 1-p \end{cases}$$

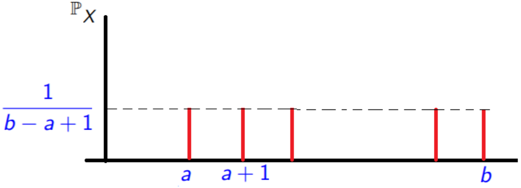
$$\mathbb{P}_X(0) = \mathbb{P}(X=0) = 1-p, \qquad \mathbb{P}_X(1) = \mathbb{P}(X=1) = p$$

- $\mathbb{E}(X) = (0 \times \mathbb{P}(X=0)) + (1 \times \mathbb{P}(X=1)) = p$
- $V(X) = [(0^2 \times \mathbb{P}(X=0)) + (1^2 \times \mathbb{P}(X=1))] - \mathbb{E}^2(X) = p - p^2 = p(1-p)$
- $\sigma(X) = \sqrt{p(1-p)}$



Discrete Random Variables

- Uniform discrete random variable with parameters {a, b} $\rightarrow \mathcal{U}(a, b)$
- **Parameters:** Integers a, b: $a \leq b$
- **Experiment:** Pick one of $a, a+1, \ldots, b$ at random; all equally likely.
- **Sample set:** $\Omega = \{a, a+1, \ldots, b\} \rightarrow$ The number of elements $= b - a + 1$
- **Random variable:** $X(\omega) = x = \omega \rightarrow \mathbb{P}(X = \omega) = \dfrac{1}{b-a+1}$



Discrete Random Variables

- $\mathbb{E}(X) = \sum_i a_i \ \mathbb{P}(X = a_i) = a \ \mathbb{P}(X = a) + (a+1) \ \mathbb{P}(X = a+1) + \ldots + b \ \mathbb{P}(X = b)$
$= a\left(\dfrac{1}{b-a+1}\right) + (a+1)\left(\dfrac{1}{b-a+1}\right) + \cdots + b\left(\dfrac{1}{b-a+1}\right)$
$= [a + (a+1) + \ldots + b] \left(\dfrac{1}{b-a+1}\right) = \left[\left(\dfrac{b-a+1}{2}\right)(a+b)\right]\left(\dfrac{1}{b-a+1}\right) = \dfrac{a+b}{2}$

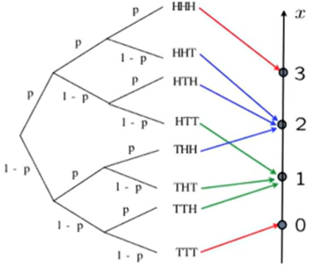
- $V(X) = \sum_i [a_i^2 \ \mathbb{P}(X = a_i)] - \mathbb{E}^2(X) = [a^2 + (a+1)^2 + \ldots + b^2] \left(\dfrac{1}{b-a+1}\right) - \dfrac{(a+b)^2}{4}$
$= \dfrac{1}{6}[(b-a+1)((b-a+1)+1)(2(b-a+1)+1)] \left(\dfrac{1}{b-a+1}\right) - \dfrac{(a+b)^2}{4}$
$= \dfrac{1}{6}\left[2(b-a+1)^2 + 3(b-a+1) + 1\right] - \dfrac{(a+b)^2}{4} = \dfrac{(b-a+1)^2 - 1}{12}$
- In a more general formula: $V(X) = \dfrac{(\#\text{terms})^2 - 1}{12}$

- $\sigma(X) = \sqrt{\dfrac{(b-a+1)^2 - 1}{12}} = \sqrt{\dfrac{(\#\text{terms})^2 - 1}{12}}$

Discrete Random Variables

- Binomial random variable with parameters (n, p) $\rightarrow$ $\mathcal{B}(n, p)$
- **Parameters:** Integers n, p: $p \in [0, 1]$, $n > 0$
- **Experiment:** n independent tosses of a coin with $\mathbb{P}(H) = p$.
- **Sample set:** Set of sequences of H and T of length n
- **Random variable:** X: number of Heads observed.

- Let $n = 3 \Rightarrow X = \{0, 1, 2, 3\}$



Discrete Random Variables

- $\mathbb{P}(X = 0) = \mathbb{P}(TTT) = (1-p)^3 = C_0^3\ (1-p)^3$
- $\mathbb{P}(X = 1) = \underbrace{\mathbb{P}(HTT)}_{p(1-p)^2} + \underbrace{\mathbb{P}(THT)}_{p(1-p)^2} + \underbrace{\mathbb{P}(TTH)}_{p(1-p)^2} = 3p(1-p)^2 = C_1^3\ p(1-p)^2$
- $\mathbb{P}(X = 2) = \underbrace{\mathbb{P}(HHT)}_{p^2(1-p)} + \underbrace{\mathbb{P}(THH)}_{p^2(1-p)} + \underbrace{\mathbb{P}(HTH)}_{p^2(1-p)} = 3p^2(1-p) = C_2^3\ p^2(1-p)$
- $\mathbb{P}(X = 3) = \mathbb{P}(HHH) = p^3 = C_3^3\ p^3$

- The **binomial** distribution with parameters $(n, p) \rightarrow \mathcal{B}(n, p)$ is given by

$$\boxed{\mathrm{P}_X(k) = \mathbb{P}(X = k) = C_k^n\ p^k(1-p)^{n-k}; \quad \text{for} \quad k = 0, 1, \ldots, n.}$$

Discrete Random Variables

- $\mathbb{E}(X) = \ \sum_i x_i \ \mathbb{P}(X = x_i) = \sum_k k \ C_k^n p^k (1-p)^{n-k}$

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

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

  $= n \ p \ [p + (1-p)]^2 = n \ p$

- $V(X) = \ \sum_i x_i^2 \ \mathbb{P}(X = x_i) - \mathbb{E}^2(X) = [\sum_k k^2 \ C_k^n p^k (1-p)^{n-k}] - (n \ p)^2$

  $= [\sum_k n(n-1) \ C_{k-2}^{n-2} \ p^k (1-p)^{n-k} + n \ C_{k-1}^{n-1} \ p^k (1-p)^{n-k}] - n^2 p^2$

  $= n(n-1) \ p^2 \sum_k C_{k-2}^{n-2} p^{k-2} \ (1-p)^{n-k}$

  $+ n \ p \sum_k C_{k-1}^{n-1} p^{k-1} \ (1-p)^{n-k} - n^2 p^2 = n \ p(1-p)$

Discrete Random Variables

- Geometric random variable with parameters $0 < p \leq 1 \rightarrow \mathcal{G}(p)$
- **Parameters:** Integers $0 < p \leq 1$
- **Experiment:** infinitely many independent tosses of a coin with $\mathbb{P}(H) = p$.
- **Sample set:** Set of infinite sequences of H and T: $\{TTTTHHT\dots\}$
- **Random variable:** X: number of tosses until the first Heads $\rightarrow \underbrace{\{TTTTH\}}_{X=5}$
- **Model of:** waiting times for something to happen, number of trials until getting a success, etc....

$$\mathbb{P}(X = k) = \mathbb{P}(\underbrace{TTTT\dots}_{k-1}H) = [\underbrace{(1-p)(1-p)\dots(1-p)}_{k-1 \text{ times}}]\, p = (1-p)^{k-1}\, p; \quad k = 1, 2, \dots$$

$$\boxed{\mathrm{P}_X(k) = \mathbb{P}(X = k) = (1-p)^{k-1}\, p}$$

Discrete Random Variables

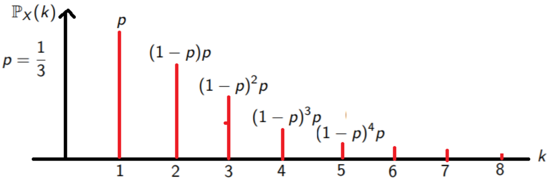
- Example: Infinitely many independent tosses of a coin with $\mathbb{P}(H) = p = \frac{1}{3}$.
- Random variable: X: number of tosses until getting the first Heads $\rightarrow \underbrace{\{TTTTH\}}_{X=5}$

$$\mathbb{P}(X = 1) = \mathbb{P}(\underbrace{H}_{k=1}) = p$$

$$\mathbb{P}(X = 2) = \mathbb{P}(\underbrace{T}_{k=2}H) = (1-p)p$$

$$\vdots$$

$$\mathbb{P}(X = k) = \mathbb{P}(\underbrace{TTTT\ldots}_{k-1 \text{ of } T}H) = \underbrace{[(1-p)(1-p)\ldots(1-p)]}_{k-1 \text{ times}}\, p = (1-p)^{k-1}\, p; \quad k = 1, 2, \ldots$$



Discrete Random Variables

- $\mathbb{E}(X) = \sum_i x_i \; \mathbb{P}(X = x_i) = \sum_{k=1}^{+\infty} k \; (1-p)^{k-1} \; p$

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

$$= p \sum_{k=1}^{+\infty} \frac{-d[(1-p)^k]}{dp}$$

$$= p \; \frac{-d \; \left(\sum_{k=1}^{+\infty} (1-p)^k\right)}{dp} = p \; \frac{-d \; \left[(1-p)\left(\dfrac{1-(1-p)^{\infty}}{1-(1-p)}\right)\right]}{dp}$$

$$= p \; \frac{-d \; \left[\dfrac{1-p}{p}\right]}{dp}$$

$$= p \; \left(\frac{1}{p^2}\right) = \frac{1}{p}$$

Discrete Random Variables

- $V(X) = \sum_i x_i^2 \ \mathbb{P}(X = x_i) - \mathbb{E}^2(X) = [\sum_{k=1}^{+\infty} k^2 \ (1-p)^{k-1} \ p] - \dfrac{1}{p^2}$

$$= p \ [\sum_{k=1}^{+\infty} k^2 \ (1-p)^{k-1} - k \ (1-p)^{k-1} + k \ (1-p)^{k-1}] - \frac{1}{p^2}$$

$$= p \ [\sum_{k=1}^{+\infty} k(k-1) \ (1-p)^{k-1}] + p \ [\sum_{k=1}^{+\infty} k \ (1-p)^{k-1}] - \frac{1}{p^2}$$

$$= p \ [\sum_{k=1}^{+\infty}[k(k-1) \ (1-p)^{k-2} \ (1-p)] + \frac{1}{p} - \frac{1}{p^2}$$

$$= p(1-p) \ \sum_{k=1}^{+\infty} \frac{-d}{dp}[k \ (1-p)^{k-1}] + \frac{1}{p} - \frac{1}{p^2}$$

$$= p(1-p) \ \left(\frac{-d}{dp} \sum_{k=1}^{+\infty}[k \ (1-p)^{k-1}] \right) + \frac{1}{p} - \frac{1}{p^2}$$

$$= p(1-p) \ \left(\frac{-d}{dp} \left[\frac{1}{p^2} \right] \right) + \frac{1}{p} - \frac{1}{p^2}$$

$$= p(1-p) \ \left(\frac{2}{p^3} \right) + \frac{1}{p} - \frac{1}{p^2}$$

$$= \frac{2(1-p)}{p^2} + \frac{1}{p} - \frac{1}{p^2} = \frac{2 - 2p + p - 1}{p^2} = \frac{1-p}{p^2}$$

Discrete Random Variables

- Geometric random variable with parameters $0 < p \leq 1 \;\rightarrow\; \mathcal{G}(p)$

▷ **There is another form of a geometric random variable such that:**

- **Experiment:** infinitely many independent tosses of a coin with $\mathbb{P}(H) = p$.
- **Sample set:** Set of infinite sequences of H and T: $\{TTTTHHT\ldots\}$
- **Random variable:** X: number of tosses before the first Heads $\rightarrow \{\underbrace{TTTT}_{X=4}H\}$

$$\mathbb{P}(X = k) = \mathbb{P}(\underbrace{TTTT\ldots}_{k}H) = [\underbrace{(1-p)(1-p)\ldots(1-p)}_{k \text{ times}}]\, p = (1-p)^k\, p; \quad k = 0, 1, 2, \ldots$$

$$\boxed{\mathrm{P}_X(k) = \mathbb{P}(X = k) = (1-p)^k\, p}$$

Discrete Random Variables

- $\mathbb{E}(X) = \sum_i x_i \; \mathbb{P}(X = x_i) = \sum_{k=0}^{+\infty} k \; (1-p)^k \; p$

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

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

$$= \frac{1-p}{p}$$

- $V(X) = \sum_i x_i^2 \; \mathbb{P}(X = x_i) - \mathbb{E}^2(X) = [\sum_{k=0}^{+\infty} k^2 \; (1-p)^k \; p] - \frac{(1-p)^2}{p^2}$

$$= (1-p) \; [\sum_{k=1}^{+\infty} k^2 \; (1-p)^{k-1} p - -\frac{(1-p)^2}{p^2}$$

$$= (1-p) \left[\frac{2(1-p)}{p^2} + \frac{1}{p} \right] - \frac{(1-p)^2}{p^2} = \frac{2(1-p)^2 + p(1-p) - (1-p)^2}{p^2} = \frac{1-p}{p^2}$$

Discrete Random Variables

- Summary:

1) Binomial Distribution $\rightarrow$ $\mathcal{B}(n, p)$:

- There are n identical and independent trials of a common procedure.
- There are exactly two possible outcomes for each trial, "success" and "failure."
- The probability of success $\mathbb{P}(S) = p$.
- The discrete random variable $X = \{\text{the number of successes in the n trials}\}$.

2) Bernoulli Distribution $\rightarrow$ $\mathcal{B}(1, p)$:

- There are exactly two possible outcomes for each trial, "success" and "failure."
- The probability of success $\mathbb{P}(S) = p$.
- The discrete random variable $X = 1$ or $X = 0$.
- Here are some examples:

*) You take a pass-fail exam. You either pass (resulting in X=1) or fail (resulting in X=0).
*) You toss a coin. The outcome is either heads or tails.
*) A child is born. The gender is either male or female.
*) In a communication network. You are receiving the signal or not.

Discrete Random Variables

- Summary:

3) Uniform Distribution $\rightarrow \mathcal{U}(a,b)$:

- There are $b-a+1$ equally likely elements.
- The discrete random variable $X = \{\text{getting one element}\} = \{a, a+1, a+2, \dots, b\}$.
- The probability of $X = x_i$ is given by $\mathbb{P}(X = x_i) = \dfrac{1}{b-a+1}$.

4) Geometric Distribution $\rightarrow \mathcal{G}(p)$:

- There are exactly two possible outcomes for each trial, "success" and "failure."
- The probability of success $\mathbb{P}(S) = p$.
- The discrete random variable $X = k$ means that the first success is at the k^{th} trial.
- Here are some examples:

*) Waiting time for something to happen.

Discrete Random Variables

- Example1: A multiple-choice test has 15 questions, each question has 5 choices of answers. A student taking the test answers each of the questions randomly by choosing an arbitrary answer from the five provided. Suppose X denotes the number of answers that the student gets right.

1) Which kind of probability distribution follows X?
2) What is the probability that $X = 10$?
3) Compute the expectation, the variance and the standard deviation.

Discrete Random Variables

- Example1: A multiple-choice test has 15 questions, each question has 5 choices of answers. A student taking the test answers each of the questions randomly by choosing an arbitrary answer from the five provided. Suppose X denotes the number of answers that the student gets right.

1) Which kind of probability distribution follows X?
2) What is the probability that $X = 10$?
3) Compute the expectation, the variance and the standard deviation.

- Solution:

1) We have a sample of 15 answers with two possibilities: {Correct answer, False answer}. Since, for each question there 4 false answers and 1 correct, so $\mathbb{P}$(The answer is correct)$=\frac{1}{5}$. The variable X={The number of correct answers} $\rightarrow$ is a binomial random variable with parameters n=15 and $p = \frac{1}{5} = 0.20, \rightarrow \; X \rightarrow \mathcal{B}(15, 0.2)$.

2) The probability $\mathbb{P}(X = 10) = C_{10}^{15}(1-p)^{15-10}p^{10} = \frac{15!}{10!5!}\left(\frac{4}{5}\right)^5\left(\frac{1}{5}\right)^{10} = 0.000176$

3) The expectation: $\mathbb{E}(X) = np = (15)(0.2) = 3 \rightarrow 3$ expected correct answers.
The variance: $V(X) = np(1-p) = (3)(0.8) = 2.4 \Rightarrow \; \sigma = \sqrt{V(X)} = 1.55$

Discrete Random Variables

- Remark:

We can compute the probability distribution of the random variable X={The number of correct answers} with $\mathbb{P}$(The answer is correct)$=\dfrac{1}{5} = 0.2 \Rightarrow X \to \mathcal{B}(15, 0.2)$.

$$\mathbb{P}(X = k) = C_k^{15}(1-p)^{15-k}p^k = \frac{15!}{k!(15-k)!}(0.2)^k(0.8)^{15-k}$$

x_i	0	1	2	3	4	5	6	7	8
$\mathbb{P}(X = x_i)$	0.035	0.1319	0.2309	0.2501	0.1976	0.1032	0.043	0.013	0.003

x_i	9	10	11	12	13	14	15
$\mathbb{P}(X = x_i)$	0.00067	0.00071	$1.14(10^{-5})$	$9.5(10^{-7})$	$5.5(10^{-8})$	$9.8(10^{-10})$	$3.2(10^{-11})$

Discrete Random Variables

- Example2:

You play a game of chance that you can either win or lose **until** you lose. Your probability of losing is $p = 0.57$.

1) What is the probability that it takes five games until you lose?

2) What is the probability that it takes five games before you lose?

3) Compute the expectation, the variance and the standard deviation for both cases.

Discrete Random Variables

- Example2:

You play a game of chance that you can either win or lose **until** you lose. Your probability of losing is $p = 0.57$.

1) What is the probability that it takes five games until you lose?

2) What is the probability that it takes five games before you lose?

3) Compute the expectation, the variance and the standard deviation for both cases.

- Solution:

Let $X = \{\text{the number of games you play until you lose}\} \rightarrow X$ follows a geometric distribution with parameter $p = 0.57 \Rightarrow X \rightarrow \mathcal{G}(0.57)$.

1) $\mathbb{P}(X = 5) = \mathbb{P}(\text{w, w, w, w, lose}) = (1-p)^4 P = (1-0.57)^4(0.57) = 0.019.$

$$\mathbb{E}(X) = \frac{1}{p} = \frac{1}{0.57} = 1.754, \quad V(X) = \frac{1-p}{p^2} = \frac{1-0.57}{(0.57)^2} = 1.32 \Rightarrow \sigma = \sqrt{V(X)} = 1.15.$$

2) $\mathbb{P}(X = 5) = \mathbb{P}(\text{w, w, w, w, w, lose}) = (1-p)^5 P = (1-0.57)^5(0.57) = 0.0084.$

$$\mathbb{E}(X) = \frac{1-p}{p} = \frac{1}{0.57} = 0.7544,$$

$$V(X) = \frac{1-p}{p^2} = \frac{1-0.57}{(0.57)^2} = 1.32 \Rightarrow \sigma = \sqrt{V(X)} = 1.15.$$

Discrete Random Variables

- Example3:

Suppose that you are looking for a student at your college who lives within five miles of you. You know that 55% of the 25,000 students do live within five miles of you. You randomly contact students from the college until one says he or she lives within five miles of you. What is the probability that you need to contact four persons?

Discrete Random Variables

- Example3:

Suppose that you are looking for a student at your college who lives within five miles of you. You know that 55% of the 25,000 students do live within five miles of you. You randomly contact students from the college until one says he or she lives within five miles of you. What is the probability that you need to contact four persons?

Solution:

This is a geometric distribution because you may have a number of failures before you have the success you desire.

• $X =$ The number of students to contact until you will find a person who lives within 5 miles of you $\rightarrow\ \mathcal{G}(p)$

• The probability of a success is $p = 0.55$

• The probability to contact 4 person to get a success

$$\mathbb{P}(X = 4) = \mathbb{P}(F, F, F, S) = (1 - p)^3 p = (1 - 0.55)^3(0.55) = 0.0501$$

$$\mathbb{E}(X) = \frac{1}{p} = \frac{1}{0.55} = 1.81$$

$$V(X) = \frac{1 - p}{p^2} = \frac{1 - 0.55}{(0.55)^2} = 1.48 \Rightarrow\ \sigma = \sqrt{V(X)} = 1.21$$

Discrete Random Variables

- Poisson random variable with parameters $\lambda \rightarrow \mathcal{P}(\lambda)$

- The Poisson distribution is a count of the number of occurrences of an event in a given unit of time, distance, area, etc.....

- **Random variable:** X: The number of events in a fixed unit of time.

- The number of accidents in a day
- The number of cars pass in an hour
- The number of customers arriving in an hour

- **Parameters:** $\lambda = \mathbb{E}(X)$ (the expectation of X, which is an estimated value)

- **The PMF:** $\mathbb{P}_X(k) = \mathbb{P}(X = k) = \dfrac{\lambda^k \ e^{-\lambda}}{k!}$

Discrete Random Variables

- The relationship between the Poisson and the Binomial distributions:

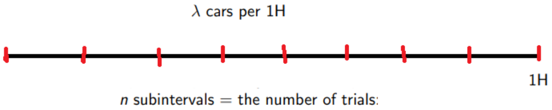
▷ The binomial distribution works when we have a fixed number of events or trials n with probability of a success p.

▷ Imagine we don't know the number of trials. Instead, we only know **the average number of successes** λ **per time period** $\rightarrow$ we know the rate of successes per day, but not the number of trials n or the probability of success p that led to that rate.

- Example of cars passing in 1H:

- Let λ be the average number of cars passing in 1H (In average λ cars per 1H). We want to compute the probability that k cars pass in 1H.

- Let n be the number of trials: it can be in mins, in seconds, or in smaller time units



$\rightarrow$ If a car pass it is a success and if no car pass it is a failure.

- To be sure that only one car can pass in each trial (or each subintervale) we need to minimize this duration by taking $n \rightarrow \infty$

Discrete Random Variables

▷ We want to compute the probability that k cars pass in 1H $\to$ k successes

$$\lambda = np = 1H.\frac{\lambda}{1H} = 60\ mins.\frac{\lambda}{60\ mins} = 3600s.\frac{\lambda}{3600s} = \ldots \Rightarrow p = \frac{\lambda}{n}$$

$$\mathbb{P}(X = k) = \lim_{n\to\infty} C_k^n (1-p)^{n-k} p^k = \lim_{n\to\infty} \frac{n!}{k!(n-k)!}(1-\frac{\lambda}{n})^{n-k}(\frac{\lambda}{n})^k$$

$$= \frac{\lambda^k}{k!}\lim_{n\to\infty}[\frac{n!}{(n-k)!}(\frac{1}{n^k})(1-\frac{\lambda}{n})^n(1-\frac{\lambda}{n})^{-k}] = \frac{\lambda^k}{k!}\lim_{n\to\infty}[\frac{n(n-1)\ldots(n-k+1)}{n^k}(1-\frac{\lambda}{n})^n(1-\frac{\lambda}{n})^{-k}]$$

$$= \frac{\lambda^k}{k!}\lim_{n\to\infty}[\underbrace{\left(\frac{n}{n}\right)\left(\frac{n-1}{n}\right)\ldots\left(\frac{n-k+1}{n}\right)}_{=1}\underbrace{(1-\frac{\lambda}{n})^n}_{=e^{-\lambda}}\underbrace{(1-\frac{\lambda}{n})^{-k}}_{=1}] = \frac{\lambda^k}{k!}e^{-\lambda}$$

Discrete Random Variables

- Computation of the Expectation and Variance of the poisson distribution

- $\mathbb{E}(X) = \sum_i x_i \; \mathbb{P}(X = x_i) = \sum_{k=0}^{\infty} k \; \mathbb{P}(X = k) = \sum_{k=0}^{\infty} k \; \dfrac{\lambda^k \; e^{-\lambda}}{k!}$

$$= e^{-\lambda} \sum_{k=0}^{\infty} k \frac{\lambda^k}{k!} = e^{-\lambda} \sum_{k=1}^{\infty} \frac{\lambda^k}{(k-1)!} = \lambda \; e^{-\lambda} \sum_{k=1}^{\infty} \frac{\lambda^{k-1}}{(k-1)!} = \lambda \; e^{-\lambda} \; e^{\lambda} = \lambda$$

- $V(X) = \sum_i (x_i - \mathbb{E}(X))^2 \; \mathbb{P}(X = x_i) = \mathbb{E}(X^2) - \mathbb{E}^2(X)$

$$= [\sum_{k=0}^{\infty} k^2 \, \mathbb{P}(X = k)] - \lambda^2 = [\sum_{k=0}^{\infty} k^2 \, \frac{\lambda^k \; e^{-\lambda}}{k!}] - \lambda^2 = e^{-\lambda} [\sum_{k=0}^{\infty} k^2 \frac{\lambda^k}{k!}] - \lambda^2 = e^{-\lambda} [\sum_{k=1}^{\infty} k \frac{\lambda^k}{(k-1)!}] - \lambda^2$$

$$= \; e^{-\lambda} [\sum_{k=1}^{\infty} (k-1) \frac{\lambda^k}{(k-1)!} + \frac{\lambda^k}{(k-1)!}] - \lambda^2 = \; e^{-\lambda} [\sum_{k=2}^{\infty} \frac{\lambda^2 \; \lambda^{k-2}}{(k-2)!} + \sum_{k=1}^{\infty} \frac{\lambda \; \lambda^{k-1}}{(k-1)!}] - \lambda^2$$

$$= \lambda^2 \; e^{-\lambda} (e^{\lambda}) + \lambda \; e^{-\lambda} (e^{\lambda}) - \lambda^2 = \lambda$$

Discrete Random Variables

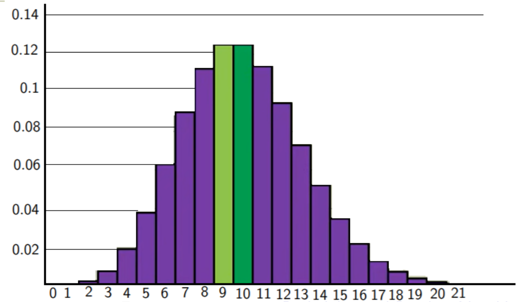
- Example1: Let's say for a cashier at Walmart, it is 4:30pm and he shift ends at 5:00pm. The store policy is to close your checkout line 15 minutes before your shift ends (in this case 4:45 pm).
By examining overhead cameras, store data indicates that between 4:30pm and 4:45pm each weekday, an average of 10 customers enter any given checkout line.
1) What is the probability that exactly 7 customers enter your line between 4:30 and 4:45?
2) Plot this probability distribution.

Discrete Random Variables

- Example1: Let's say for a cashier at Walmart, it is 4:30pm and he shift ends at 5:00pm. The store policy is to close your checkout line 15 minutes before your shift ends (in this case 4:45 pm).
By examining overhead cameras, store data indicates that between 4:30pm and 4:45pm each weekday, an average of 10 customers enter any given checkout line.
1) What is the probability that exactly 7 customers enter your line between 4:30 and 4:45?
2) Plot this probability distribution.
- Solution:
- Outcomes = # number of customer.
- λ = # customers/per specified interval = 10 customers/per 15 minutes $\Rightarrow$ $\lambda = 10$
- $\mathbb{P}(X = 7) = \frac{\lambda^7}{7!} e^{-\lambda} = \frac{10^7}{7!} e^{-10} = 0.0901.$

Discrete Random Variables

- $\mathbb{P}(X=0)=\frac{10^0}{0!}e^{-10}=4.5\times10^{-5}$, $\mathbb{P}(X=1)=\frac{10^1}{1!}e^{-10}=4.5\times10^{-4}$, $\mathbb{P}(X=2)=\frac{10^2}{2!}e^{-10}=0.0023$, $\mathbb{P}(X=3)=\frac{10^3}{3!}e^{-10}=0.0076$, $\mathbb{P}(X=4)=\frac{10^4}{4!}e^{-10}=0.0189$, $\mathbb{P}(X=5)=\frac{10^3}{3!}e^{-10}=0.0378,\ldots\mathbb{P}(X=10)=\frac{10^{10}}{10!}e^{-10}=0.1252$, $\mathbb{P}(X=11)=\frac{10^{11}}{11!}e^{-10}=0.1137$, $\mathbb{P}(X=12)=\frac{10^{12}}{12!}e^{-10}=0.0948,\ldots\mathbb{P}(X=21)=\frac{10^{21}}{21!}e^{-10}=8.8\times10^{-4}$



Discrete Random Variables

- Example2:

Suppose that a fast food restaurant can expect 2 customers every 3 minutes, on average. What is the probability that 4 patrons will enter the restaurant in a 9 minute period?

Discrete Random Variables

- Example2:

Suppose that a fast food restaurant can expect 2 customers every 3 minutes, on average. What is the probability that 4 patrons will enter the restaurant in a 9 minute period?

- Solution:

- The expected value λ is assumed to be constant through the experiment.
- If $\lambda = 2$ customers/3 minutes $\Rightarrow$ $\lambda = 6$ customers/9 minutes

$$\mathbb{P}(X = 4) = \frac{6^4}{4!} e^{-6} = 0.1339.$$

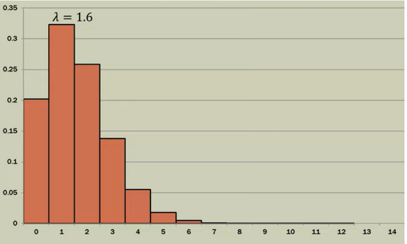
Discrete Random Variables

- Example3: A bank is interested in studying the number of people who use the ATM located outside its office late at night. On average, 1.6 customers walk up to the ATM during any 10 minute interval between 9pm and midnight.
- What is λ for this problem?
- What is the probability of exactly 3 customers using the ATM during any 10 minute?
- Plot this PMF.

Discrete Random Variables

- Example3: A bank is interested in studying the number of people who use the ATM located outside its office late at night. On average, 1.6 customers walk up to the ATM during any 10 minute interval between 9pm and midnight.
- What is λ for this problem?
- What is the probability of exactly 3 customers using the ATM during any 10 minute?
- Plot this PMF.
- Solution:

- $\lambda = 1.6$
- $\mathbb{P}(X = 3) = \frac{(1.6)^3}{3!} e^{-1.6} = 0.1378.$



Discrete Random Variables

- Table of Discrete Distributions and Properties:

Distribution	range X	PMF $\mathbb{P}_X(x)$	mean $\mathbb{E}(X)$	Variance $V(X)$
Bernoulli (p)	$0,\ 1$	$\mathbb{P}(0) = 1 - p,\ \mathbb{P}(1) = p$	p	p (1-p)
Binomial (n,p)	$0, 1, 2, \ldots, n$	$\mathbb{P}(X = k) = C_k^n (1 - p)^{n-k} p^k$	np	np(1-p)
Uniform (n)	$1, 2, \ldots, n$	$\mathbb{P}(X = k) = \frac{1}{n}$	$\frac{n+1}{2}$	$\frac{n^2-1}{12}$
Geometric (p)	$1, 2, \ldots\ldots$	$\mathbb{P}(X = k) = (1 - p)^{k-1}\ p$	$\frac{1}{p}$	$\frac{1-p}{p^2}$
Geometric (p)	$0, 1, 2, \ldots\ldots$	$\mathbb{P}(X = k) = (1 - p)^{k}\ p$	$\frac{1-p}{p}$	$\frac{1-p}{p^2}$
Poisson (λ)	$0, 1, 2, \ldots\ldots$	$\mathbb{P}(X = k) = \frac{\lambda^k e^{-\lambda}}{k!}$	λ	λ

Discrete Random Variables

- Cumulative Distribution Function (CDF): The Cumulative Distribution Function (PDF) gives the probability of seeing an outcome less than or equal to a particular value of the random variable. CDFs are used to check how the probability has added up to a certain point. For example, if $\mathbb{P}(X = 5)$ is the probability that the number of heads on flipping a coin is 5 then, $\mathbb{P}(X \leq 5)$. The cumulative probability of obtaining 0 to 5 heads, it is equal to

$$\mathbb{P}(X \leq 5) = \mathbb{P}(X = 0) + \mathbb{P}(X = 1) + \mathbb{P}(X = 2) + \mathbb{P}(X = 3) + \mathbb{P}(X = 4) + \mathbb{P}(x = 5)$$

$$F(x) = \mathbb{P}(X \leq x) = \sum_{x_i \leq x} \mathbb{P}(x = x_i)$$

- $\mathbb{P}(X > x) = 1 - \mathbb{P}(X \leq x) = 1 - F(X)$
- $\mathbb{P}(x_1 < X \leq x_2) = \mathbb{P}(X \leq x_2) - \mathbb{P}(X \leq x_1) = F(x_2) - F(X_1)$
- $\mathbb{P}(x_1 \leq X \leq x_2) = \mathbb{P}(X \leq x_2) - \mathbb{P}(X < x_1)$
- Example: the probability that you will win \$3 or less in this game

x_i	\$ 2	\$ 3	\$ -4
$\mathbb{P}(X = x_i)$	$\frac{1}{2}$	$\frac{1}{6}$	$\frac{1}{3}$

$$\mathbb{P}(X \leq 3) = \mathbb{P}(X = 2) + \mathbb{P}(X = 3) = \tfrac{1}{2} + \tfrac{1}{6} = \tfrac{2}{3}$$

Discrete Random Variables

- Example:

A student takes a ten-question true/false exam.

1) Compute the expected value, the variance and the standard deviation.

2) Find the probability that the student gets exactly 6 of the questions right simply by guessing the answer on every question.

3) Find the probability that the student will obtain a passing grade by guessing at least 6 right simply by guessing.

Discrete Random Variables

- Solution:

1) The random variable $X =$ The number of right answers from 10 questions.
$\rightarrow$ It is Binomial with parameters $n = 10$ and $p = 0.5$; $q = 1 - p = 0.5 \rightarrow \mathcal{B}(10, 0, 5)$.

$$\mathbb{E}(X) = np = 10 \times 0.5 = 5, \quad V(X) = np(1-p) = 10 \times 0.5 \times 0.5 = 2.5, \quad \sigma(X) = \sqrt{2.5} = 1.58$$

2) $\mathbb{P}(X = k) = C_k^n (1-p)^{n-k} p^k \Rightarrow \mathbb{P}(X = 6) = C_6^{10} (0.5)^4 (0.5)^6 = \dfrac{10!}{6!4!} (0.5)^4 (0.5)^6 = 0.205.$

x_i	0	1	2	3	4	5	6	7	8	9	10
$\mathbb{P}_X(x_i)$	$9.7.\ 10^{-4}$	0.009	0.04	0.11	0.20	0.24	0.20	0.11	0.04	0.009	$9.7\ 10^{-4}$

3) $\mathbb{P}(X \geq 6) = \mathbb{P}(X = 6) + \mathbb{P}(X = 7) + \mathbb{P}(X = 8) + \mathbb{P}(X = 9) + \mathbb{P}(X = 10)$
$= 0.2 + 0.11 + 0.04 + 0.009 + 0 = 0.359$

- Method2:

$\mathbb{P}(X \geq 6) = 1 - \mathbb{P}(X \leq 5)$
$= 1 - [\mathbb{P}(X = 5) + \mathbb{P}(X = 4) + \mathbb{P}(X = 3) + \mathbb{P}(X = 3) + \mathbb{P}(X = 1) + \mathbb{P}(X = 0)]$
$= 1 - [0.24 + 0.2 + 0.11 + 0.04 + 0.009 + 0] = 0.359$

Discrete Random Variables

- Example:

The literacy rate for a nation measures the proportion of people aged 15 and over who can read and write. The literacy rate for women is 12%. Let X be the number of women you ask until one says that she is literate.

1) What is the probability distribution of X ?

2) What is the probability that you must ask 10 women?

3) What is the probability that you ask at least 5 women before one says she is literate?

Discrete Random Variables

- Solution:

1) This is a geometric distribution $\rightarrow \mathcal{G}(p)$ with parameter $p = 0.12$

2) $\mathbb{P}(X = k) = (1 - p)^{k-1} p \Rightarrow \mathbb{P}(X = 10) = (1 - 0.12)^9 \times (0.12) = 0.038.$

3) $\mathbb{P}(X \geq 5) = 1 - \mathbb{P}(X < 5)$

$$= 1 - [\mathbb{P}(X = 1) + \mathbb{P}(X = 2) + \mathbb{P}(X = 3) + \mathbb{P}(X = 4)] = 1 - (0.4003) = 0.6$$

x_i	1	2	3	4
$\mathbb{P}_X(x_i)$	0.12	0.1056	0.0929	0.0818

Discrete Random Variables

- Example:

Suppose a fast food restaurant can expect 2 customers every 3 minutes on average. What is the probability that at least 4 patrons will enter the restaurant in a 9 minute period?

- Solution:

- This is a poisson distribution: $X =$ The number of customer that will enter the restaurant in a 9 minute period.
- The expected value λ is $\lambda = 2$ customers/3 minutes $\Rightarrow$ $\lambda = 6$ customers/9 minutes

$$\mathbb{P}(X = k) = \frac{\lambda^k}{k!}e^{-\lambda} = \frac{6^k}{k!}e^{-6}$$

$$\mathbb{P}(X \geq 4) = 1 - \mathbb{P}(X < 4) = 1 - [\mathbb{P}(X = 0) + \mathbb{P}(X = 1) + \mathbb{P}(X = 2) + \mathbb{P}(X = 3)]$$

x_i	0	1	2	3
$\mathbb{P}_X(x_i)$	0.0025	0.0149	0.0446	0.0892

$$= 1 - 0.1512 = 0.8488.$$

Discrete Random Variables

- Example: $X \to \mathcal{P}(\lambda)$ with $\lambda = 6$ in a 9 minute period.

$$\mathbb{P}(X = k) = \frac{\lambda^k}{k!}e^{-\lambda} = \frac{6^k}{k!}e^{-6} \to \mathbb{E}[X] = \lambda = 6 \to \sigma(X) = \sqrt{V(X)} = \sqrt{6} = 2.45$$

x_i	0	1	2	3	4	5	6	7	8	9
$\mathbb{P}_X(x_i)$	0.0025	0.0149	0.0446	0.0892	0.133	0.16	0.16	0.13	0.1	0.06

10	11	12	13	14
0.04	0.022	0.011	0.0052	0.0022

Discrete Random Variables

- Example:

Let's say for a cashier at Walmart, it is 4:30 pm and he shift ends at 5:00pm. The store policy is to close your checkout line 15 minutes before your shift ends (in this case 4:45 pm).
By examining overhead cameras, store data indicates that between 4:30pm and 4:45pm each weekday, an average of 10 customers enter any given checkout line.
- What is the probability that more than 10 people arrive? (Which means you will probably be on shift later than 5:00pm).

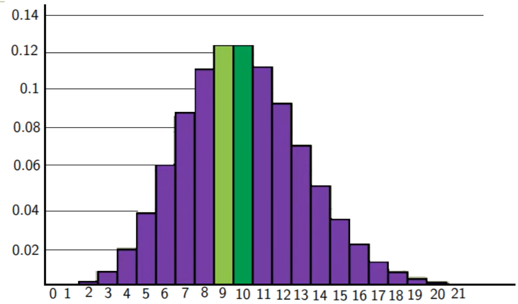
Discrete Random Variables

- Solution:

- This is a Poisson distribution: X = # number of customers enter between 4:30pm and 4:45pm.

$\lambda = \#$ customers/per specified interval = 10 customers/per 15 minutes $\Rightarrow$ $\lambda = 10$

- $\mathbb{P}(X = k) = \dfrac{\lambda^k}{k!} e^{-\lambda} = \dfrac{10^k}{k!} e^{-10}$



$\mathbb{P}(X > 10) = 1 - \mathbb{P}(X \leq 10)$

Discrete Random Variables

- Solution:

$\mathbb{P}(X > 10) = 1 - \mathbb{P}(X \leq 10)$

x_i	0	1	2	3	4	5	6	7	8	9	10
$\mathbb{P}_X(x_i)$	4.10^{-4}	4.10^{-5}	0.002	0.007	0.018	0.037	0.063	0.09	0.11	0.12	0.12

$\mathbb{P}(X > 10) = 1 - 0.5824 = 0.4176.$