ECE 203 – Section 3
Random variables

- Random variables
- Cumulative Distribution Function
- Probability Mass Function
- Mean and variance of random variables
- Functions of random variables
- Variance and higher order moments

The slides have been prepared based on the lecture notes of Prof. Patrick Mitran.

- After an experiment is done, we are often interested in a *function* of the outcome (e.g., sum of two dice rolls or $\#H$ after flipping 10 coins).
- A function that maps each outcome $s \in S$ to a real number is called a *random variable* (often abbreviated as *rv*).
- For example, let $S = \{(s_1, s_2) \mid 1 \leq s_1 \leq 6, 1 \leq s_2 \leq 6\}$ be outcomes of two dice rolls. For $s = (a, b)$, if we define $X(s) = a + b$, then $X(s)$ is a random variable.
- We often write X instead of $X(s)$, since s and S are either clear from context or do not matter.

Example

- Suppose we toss 3 coins. Let $X = \#H$. Then X is an *rv* that can only take values 0, 1, 2 or 3.

$$\{X = 0\} = \{TTT\}$$

$$\{X = 1\} = \{TTH, THT, HTT\}$$
$$\{X = 2\} = \{THH, HHT, HTH\}$$
$$\{X = 3\} = \{HHH\}$$

- In this case,

$$P[X = 0] = P[X = 3] = 1/8$$
$$P[X = 1] = P[X = 2] = 3/8$$

- Note that since $\{X = 0\}, \{X = 1\}, \{X = 2\}, \{X = 3\}$ are disjoint and cover all possible outcomes for X, we have

$$P\left[\bigcup_{i=0}^{3}\{X = i\}\right] = \sum_{i=0}^{3} P[X = i] = 1$$

Another example

- Let E and F be independent events with

$$P[E] = 0.1 \quad \text{and} \quad P[F] = 0.2$$

- Let $Y = \#$ events that have occurred. Then

$$P[Y = 0] = P[E^c F^c] = P[E^c]P[F^c] = 0.9 \cdot 0.8$$

$$P[Y = 1] = P[EF^c \cup E^c F] = P[EF^c] + P[E^c F] = 0.1 \cdot 0.8 + 0.9 \cdot 0.2$$

$$P[Y = 2] = P[EF] = P[E]P[F] = 0.1 \cdot 0.2$$

Yet another example

- A flipped coin has probability p of being heads. We flip the coin until a head occurs, up to a max of n flips. Let $Z = \#$ of flips. Then

$$
\begin{aligned}
P[Z = 1] &= P[H] = p \\
P[Z = 2] &= P[TH] = (1 - p) \cdot p \\
P[Z = 3] &= P[TTH] = (1 - p)^2 \cdot p \\
P[Z = n - 1] &= P[n - 2\ T\text{s followed by } H\text{s}] = (1 - p)^{n-2} \cdot p \\
P[Z = n] &= P[n - 1\ T\text{s followed by anything}] = (1 - p)^{n-1}
\end{aligned}
$$

Cumulative Distribution Function

- In each example above, we enumerated the probability of each possible outcome of X (or Y or Z), i.e., $P[X = 1], P[X = 2], \ldots$

- Instead, we could have enumerated $P[X \leq x]$. Particularly, let

$$F_X(x) = P[X \leq x]$$

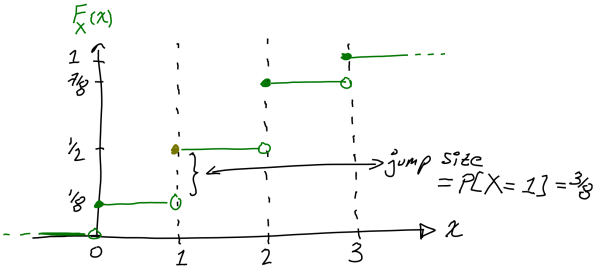
- $F_X(x)$ is called the *Cumulative Distribution Function* (CDF) of X.

- **Important:** $F_X(x)$ is a *continuously-defined* function of x.

- In fact, it will be a *piecewise-constant* function with discontinuities at $x = 0, 1, 2, \cdots$.

- Thus, for the previous example, we have

$$F_X(0) = P[X \leq 0] = \frac{1}{8}$$
$$F_X(1) = P[X \leq 1] = \frac{1}{8} + \frac{3}{8}$$
$$F_X(2) = P[X \leq 2] = \frac{1}{8} + \frac{3}{8} + \frac{3}{8}$$
$$F_X(3) = P[X \leq 3] = 1$$



- **Definition:** A random variable that can take at most a *countable number of possible outcomes* is called a *discrete random variable*.

- **Definition:** For a discrete random variable X, we define its *Probabillity Mass Function (PMF)* $p_X(a)$ as

$$p_X(a) = P[X = a]$$

- Let $\mathcal{X} = \{x_1, x_2, ...\}$ be all the possible outcomes that X takes. Then

$$\begin{aligned} p_X(x) &\geq 0, \quad \text{if } x \in \mathcal{X} \\ p_X(x) &= 0, \quad \text{otherwise} \end{aligned}$$

  and, since X must take one of its possible values, we also have

$$\sum_{x \in \mathcal{X}} p_X(x) = 1$$

Example

- Suppose the PMF of the random variable X is defined as

$$p_X(k) = C\frac{\lambda^k}{k!}, \qquad \text{for } k = 0, 1, 2, \ldots$$

where $\lambda > 0$ is given.

- **Question A:** *Define C in terms of λ.*
Solution: $1 = \sum_{k=0}^{\infty} p_X(k) = C\sum_{k=0}^{\infty} \lambda^k/k! = Ce^{\lambda}$, where we used the fact that $\sum_{k=0}^{\infty} \lambda^k/k! = e^{\lambda}$. Therefore, $C = e^{-\lambda}$.

- **Question B:** Find $P[X = 0]$.
Solution: $P[X = 0] = p_X(0) = C\lambda^0/0! = e^{-\lambda}$.

- **Question C:** Find $P[X > 2]$.
Solution:

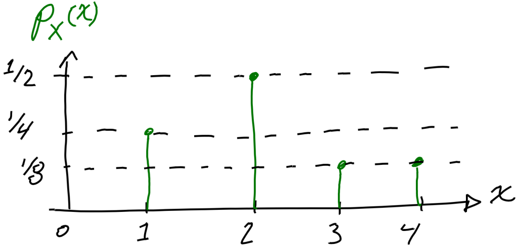
$$P[X > 2] = 1 - P[X = 0] - P[X = 1] - P[X = 2] =$$

$$= 1 - e^{-\lambda} - \lambda e^{-\lambda} - \frac{\lambda^2}{2}e^{-\lambda}.$$

- Let X be such that

$$p_X(1) = 1/4, \quad p_X(2) = 1/2, \quad p_X(3) = 1/8, \quad p_X(4) = 1/8$$

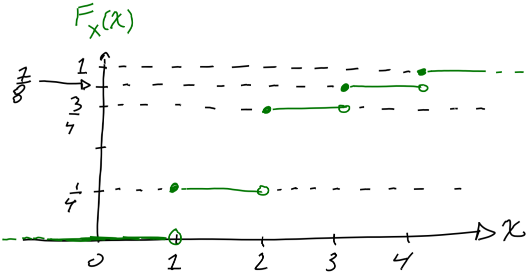
- First, we want to plot the PMF and CDF of X.



Another example (cont.)

- Recall that

$$F_X(x) = P[X \le x] = \sum_{a:a\le x} P[X = a]$$



- Note that the size of the “jump” at $x = a$ is equal to $P[X = a]$.
- At $x = a$, the function is open on the left and closed on the right.

Expected (Mean) Value

- **Definition:** The *expected (or mean) value* of a random variable X is defined as

$$E[X] = \sum_{x \in \mathcal{X}} x \, p_X(x)$$

- The expected (mean) value of X is an "average" where each outcome is weighted by probability that X assumes that outcome.

- For example, if $p_X(0) = 1/2$ and $p_X(1) = 1/2$, then

$$E[X] = 0 \cdot 1/2 + 1 \cdot 1/2 = 1/2$$

- Similarly, if $p_X(0) = 1/3$ and $p_X(1) = 2/3$, then

$$E[X] = 0 \cdot 1/3 + 1 \cdot 2/3 = 2/3$$

Example

- Let $A \subset S$ be an event. The *indicator function of A* is defined as

$$I = \begin{cases} 1, & \text{if } A \text{ occurs} \\ 0, & \text{if } A \text{ does not occur} \end{cases}$$

(We often write I_A or 1_A for this kind of random variable.)

- In this case, the mean value of I can be computed as

$$E[I] = 0 \cdot P[I = 0] + 1 \cdot P[I = 1] = 0 \cdot P[A^c] + 1 \cdot P[A] = P[A]$$

- 120 students are driven in 3 buses with 36, 40 and 44 students each to an event. One of the 120 students is chosen randomly.
- Let $X = \#$ students on the bus with the randomly chosen student.
- *What is $E[X]$?*
- To find the answer, let $\mathcal{X} = \{36, 40, 44\}$ and note that

$$P[X = 36] = 36/120$$
$$P[X = 40] = 40/120$$
$$P[X = 44] = 44/120$$

- Consequently,

$$E[X] = 36 \cdot \frac{36}{120} + 40 \cdot \frac{40}{120} + 44 \cdot \frac{44}{120} \approx 40.267$$

- Note that $E[X]$ is not an integer.

Functions of a Random Variable

- Suppose we know X and $p_X(x)$.
- Now let $Y = g(X)$ for some function $g(\cdot)$.
- Since X is a function of the outcome $s \in S$ and Y is a function of X, Y is a function of the outcome $s \in S$. Therefore, Y is a random variable.
- Y has a PMF $p_Y(y)$, which can determined from $p_X(x)$.

Example

- Let X be a random variable such that

$$P[X=-1]=0.1, \quad P[X=0]=0.3, \quad P[X=1]=0.6$$

  and let $Y=X^2$. What are $E[X]$ and $E[Y]$?

- To solve the problem, we first compute

$$E[X]=-1\cdot 0.1+0\cdot 0.3+1\cdot 0.6=0.5$$

$$P[Y=0]=P[X^2=0]=P[X=0]=0.3$$

$$P[Y=1]=P[X^2=1]=P[\{X=1\}\cup\{X=-1\}]=0.1+0.6=0.7$$

- Consequently, we have

$$E[X^2]=E[Y]=0\cdot 0.3+1\cdot 0.7=0.7$$

- **Note:** $(E[X])^2=(0.5)^2\neq 0.7=E[X^2]$. So, in general, we have

$$E[g(X)]\neq g(E[X])$$

- **Proposition:** If X is an *rv* with possible values $\mathcal{X} = \{x_1, x_2, \ldots\}$, then

$$E[g(X)] = \sum_{i \geq 1} g(x_i) p_X(x_i)$$

- To see that, let $\mathcal{Y} = \{y_1, y_2, \ldots\}$ be all possible values of Y.

$$\begin{aligned}
\sum_{i \geq 1} g(x_i) p_X(x_i) &= \sum_{j \geq 1} \sum_{i : g(x_i) = y_j} g(x_i) p_X(x_i) \\
&= \sum_{j \geq 1} \sum_{i : g(x_i) = y_j} y_j p_X(x_i) \\
&= \sum_{j \geq 1} y_j \sum_{i : g(x_i) = y_j} p_X(x_i) \\
&= \sum_{j \geq 1} y_j P[g(X) = y_j] \\
&= E[g(X)]
\end{aligned}$$

- **Corollary:** If a and b are constants, then

$$E[aX+b] = aE[X] + b$$

This is because

$$\begin{aligned} E[aX+b] &= \sum_{x\in\mathcal{X}} (ax+b)p_X(x) \\ &= a\sum_{x\in\mathcal{X}} xp_X(x) + b\sum_{x\in\mathcal{X}} p_X(x) \\ &= aE[X] + b \end{aligned}$$

- For example, say $E[X] = 3$. Then $E[10\cdot X + 4] = 10\cdot 3 + 4 = 34$.

- $E[X]$ is called the *mean of X* and is often denoted as $\mu_X = E[X]$. $E[X^n]$ is called the *n-th moment of X*.

Variance

- Given X, it is useful to summarize some essential properties of X.
- Thus, for instance, $E[X]$ tells us about the "centre" of how X is distributed.
- For example, suppose we have

$$P[W = 0] = 1$$
$$P[Y = 1] = P[Y = -1] = \frac{1}{2}$$
$$P[Z = 100] = P[Z = -100] = \frac{1}{2}$$

  In this case, $E(W) = E(Y) = E(Z) = 0$.
- However, the values of the random variables are not equally spread.

Variance (cont.)

- **Definition:** The *variance* of X is defined as

$$Var[X] = E\left[(X - E[X])^2\right] = E\left[(X - \mu_X)^2\right]$$

- We often write $\sigma_X^2 = Var[X]$, where σ_X is called *standard deviation*.
- **Note:** Since $(X - \mu_X)^2 \geq 0$, then $Var[X] \geq 0$.

Variance (cont.)

- We also have

$$
\begin{aligned}
Var[X] &= E[\,(X-\mu_X)^2\,] \\
&= \sum_{x\in\mathcal{X}} (x-\mu_X)^2\, p_X(x) \\
&= \sum_{x\in\mathcal{X}} \left(x^2 - 2\mu_X x + \mu_X^2\right) p_X(x) \\
&= \sum_{x\in\mathcal{X}} x^2 p_X(x) - 2\mu_X \sum_{x\in\mathcal{X}} x p_X(x) + \mu_X^2 \sum_{x\in\mathcal{X}} p_X(x) \\
&= E[X^2] - 2\,\underbrace{\mu_X\, E[X]}_{\mu_X^2} + \mu_X^2 \\
&= E[X^2] - (E[X])^2
\end{aligned}
$$

- Thus, we can conclude that

$$
E[X^2] \geq (E[X])^2 \quad \text{and} \quad \frac{E[X^2]}{E[X]} \geq E[X]
$$

Example

- Let X be the the outcome of a dice roll. What is $Var[X]$?

$$E[X] = 1 \cdot \frac{1}{6} + 2 \cdot \frac{1}{6} + \cdots + 6 \cdot \frac{1}{6} = \frac{7}{2}$$

$$E[X^2] = 1 \cdot \frac{1}{6} + 2^2 \cdot \frac{1}{6} + \cdots + 6^2 \cdot \frac{1}{6} = \frac{91}{6}$$

- Therefore,

$$Var[X] = E[X^2] - (E[X])^2 = \frac{91}{6} - \left(\frac{7}{2}\right)^2 = \frac{35}{12}$$

- Also,

$$E\left[(X - E[X])^2\right] =$$

$$= \left(1 - \frac{7}{2}\right)^2 \cdot \frac{1}{6} + \left(2 - \frac{7}{2}\right)^2 \cdot \frac{1}{6} + \cdots + \left(6 - \frac{7}{2}\right)^2 \cdot \frac{1}{6} = \frac{35}{12}$$

Another example

- The distance from Vancouver, BC to Boston, MA is 4200 km. If the wind is good (this happens with probability 0.7), the speed of a plane for the trip is $V = 700$ km/h. If the wind is not good (with probability 0.3), the speed of the plane is $V = 600$ km/h. *What is the average flight time?*

- **Solution:** If the wind is good, the flight time T is $4200/700 = 6$ hours. Alternatively, if the wind is not good, the flight time T is $4200/600 = 7$ hours. Thus, with

$$P[T = 6] = 0.7 \quad \text{and} \quad P[T = 7] = 0.3$$

  we have

$$E[T] = 6 \cdot 0.7 + 7 \cdot 0.3 = 6.3$$

- Note, that this is not the same as computing the average speed

$$E[V] = 700 \cdot 0.7 + 600 \cdot 0.3 = 670 \text{ km/h}$$

  and then computing $4200/670 \approx 6.27$ hours.

- In other words, even though $T = 4200/V$, $E[T] \neq 4200/E[V]$.

"Friendship Paradox"

- Suppose there are n people named $1, 2, \dots, n$.
- Person i has $f(i)$ friends, and we let $m = \sum_{i=1}^{n} f(i)$.
- Now, let X be a random person, equally likely to be any of the n people. Also, Let $Z = f(X)$ (i.e. Z is # of friends of random person).
- Then we have

$$E[Z] = \sum_{i=1}^{n} f(i) \underbrace{P[X = i]}_{1/n} = \frac{m}{n}$$

$$E[Z^2] = \sum_{i=1}^{n} (f(i))^2 \, P[X = i] = \frac{1}{n} \sum_{i=1}^{n} (f(i))^2$$

- Now, each person writes the names of their friends on a sheet of paper (one sheet per friend).
- There are m sheets, and one sheet is drawn at random, each sheet being equally likely to be chosen.

- Let Y be the name of the friend on the drawn sheet and let $W = f(Y)$.
- Note that, in this case,
$$P[Y = i] = \frac{f(i)}{m}$$
as opposed to $1/n$.
- Thus, we have
$$E[W] = E[f(Y)] = \sum_i f(i)\,P[Y = i] = \sum_i f(i) \cdot \frac{f(i)}{m} =$$
$$= \frac{n}{m} \cdot \frac{1}{n} \sum_i (f(i))^2 = \frac{E[Z^2]}{E[Z]} \geq E[Z]$$
since $E[Z^2] \geq (E[Z])^2$.
- The inequality $E[Z] \leq E[W]$ suggests that the expected number of friends of a random person is smaller than the expected number of friends of a random friend.

- There are n days in a year.
- Persons 1, 2 and 3 are independently born on day r with probability p_r, for $r = 1, 2, \ldots, n$.
- Let $A_{i,j} = \{\text{persons } i \text{ and } j \text{ born on same day}\}$.
- **Question (a):** What is $P[A_{1,3}]$?

$$
\begin{aligned}
P[A_{1,3}] &= P[\cup_r \{1 \text{ and } 3 \text{ both born on day } r\}] = \\
&= \sum_r P[\{1 \text{ and } 3 \text{ both born on day } r\}] = \\
&= \sum_r P[\{1 \text{ born on day } r\}] P[\{3 \text{ born on day } r\}] = \sum_r p_r^2
\end{aligned}
$$

- **Question (b):** What is $P[A_{1,3} \mid A_{1,2}]$?

$$P[A_{1,3} \mid A_{1,2}] = \frac{P[A_{1,3}A_{1,2}]}{P[A_{1,2}]} =$$

$$= \frac{P[\{1,\ 2 \text{ and } 3 \text{ born on same day}\}]}{P[\{1 \text{ and } 2 \text{ born on same day}\}]} = \frac{\sum_r p_r^3}{\sum_r p_r^2}$$

- **Question (c):** Show that $P[A_{1,3} \mid A_{1,2}] \geq P[A_{1,3}]$.

We want to show that $\sum_r p_r^3 / \sum_r p_r^2 \geq \sum_r p_r^2$. Let X be a random variable that is equal to p_r with probability p_r. In this case,

$$E[X] = \sum_r p_r P[X = p_r] = \sum_r p_r^2$$

$$E[X^2] = \sum_r p_r^2 P[X = p_r] = \sum_r p_r^3$$

and the result follows from $E[X^2]/E[X] \geq E[X]$.

- We had $E[aX+b] = aE[X]+b$. What about $Var[aX+b]$?

$$Var[aX+b] = E\left[(aX+b-E[aX+b])^2\right] =$$

$$= E\left[(aX+b-aE[X]-b])^2\right] = E\left[(aX-aE[X])^2\right] =$$

$$= E\Big[a^2 \underbrace{(X-E[X])^2}_{Y}\Big] = E\left[a^2 Y\right] = a^2 E\left[Y\right] =$$

$$= a^2 E\left[(X-E[X])^2\right] = a^2 Var[X]$$

- Therefore, we have

$$Var[aX+b] = a^2 Var[X]$$

Remarks (cont.)

- If X has units of, say, kg, then $E[X]$ has units of kg, while $Var[X]$ has units of kg^2.
- We also define $SD[X] = \sqrt{Var[X]}$, called *standard deviation*.
- $SD[X]$ has units of kg again.
- If we write $Var[X] = \sigma_X^2$, then $SD[X] = \sigma_X$.

Bernoulli Random Variable

- Let

$$p_X(k) = \begin{cases} 1-p, & \text{if } k=0 \\ p, & \text{if } k=1 \end{cases}$$

with $0 \leq p \leq 1$.

- X is called **Bernoulli** with parameter p, denoted $X \sim \text{Bernoulli}(p)$.

- This random variable models binary conditions (e.g., state of a connection, preference of a person for/against politician, etc).

- Consider n independent trials of Bernoulli(p), and let $X = \#$ of ones in the n trials.

- Then X is called **binomial** with parameters n and p, denoted $X \sim \text{Binomial}(n, p)$.

- Note that $\text{Bernoulli}(p) = \text{Binomial}(1, p)$.

- For $0 \leq k \leq n$, there are $\binom{n}{k}$ ways to get k ones from n Bernoulli trials, where each has probability $p^k(1-p)^{n-k}$. Thus,

$$p_X(k) = \begin{cases} \binom{n}{k} p^k (1-p)^{n-k}, & 0 \leq k \leq n \\ 0, & \text{else} \end{cases}$$

- Note that, since X must be between 0 and n, we have

$$1 = \sum_{k=0}^{n} p_X(k) = \sum_{k=0}^{n} \binom{n}{k} p^k (1-p)^{n-k}$$

Example

- A company sells screw in packs of 10. If each screw has a probability of 0.01 of being defective, and there is a money-back guarantee if more than 1 screw is defective, what is the proportion of packs that will be replaced?

- In this case, $X \sim \text{Binomial}(10, 0.01)$. So, we have

$$P[\text{not replacing a pack}] = P[X = 0] + P[X = 1] =$$

$$= \binom{10}{0}(0.01)^0(0.99)^{10} + \binom{10}{1}(0.01)^1(0.99)^9 \approx 0.996$$

- Therefore,

$$P[\text{replace a pack}] = 1 - P[\text{not replacing a pack}] \approx 0.004$$

Another example

- A system consists of n components, each of which independently functions with probability p. The system functions if at least half of its components function.

- For what value of p is a 5-component system more likely to function than a 3-component system?

$$\begin{aligned} a &= P[\text{5 component system functions}] \\ &= \binom{5}{3}p^3(1-p)^2 + \binom{5}{4}p^4(1-p)^1 + \binom{5}{5}p^5 \\ &= 10p^3(1-p)^2 + 5p^4(1-p)^1 + p^5 \end{aligned}$$

$$\begin{aligned} b &= P[\text{3 component system functions}] \\ &= \binom{3}{2}p^2(1-p)^1 + \binom{3}{3}p^3 = 3p^2(1-p)^1 + p^3 \end{aligned}$$

- Now, if we set $a > b$ and substitute the above expressions, then after some algebra, we get $p > 1/2$.

Poisson random variable

- We say X is **Poisson** with parameter $\lambda > 0$, denoted $X \sim \text{Poisson}(\lambda)$, if

$$p_X(k) = \begin{cases} \frac{\lambda^k}{k!} e^{-\lambda}, & \text{for } k = 0, 1, 2, \dots \\ 0, & \text{else} \end{cases}$$

- We note that

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

- The Poisson random variable is an approximation of the binomial random variable when
  1. n is large,
  2. p is small, and
  3. $\lambda = n\,p$ is moderate.

Example

- Say $n = 100$, $p = 0.01$, resulting in $\lambda = 1$. Then

$$p_X(5) = \frac{100!}{95!\,5!}(0.01)^5(0.99)^{95} \approx 0.00290$$

and

$$\frac{1^5}{5!}e^{-1} \approx 0.00306$$

- However, if we repeat with $n = 1000, p = 0.001$, then $\lambda = 1$ again. In this case,

$$p_X(5) = \frac{1000!}{995!\,5!}(0.001)^5(0.999)^{995} \approx 0.00305$$

Poisson Random Variable (cont.)

- It can be shown that, if $X \sim \text{Binomial}(n, p)$ and $Y \sim \text{Poisson}(np)$, then

$$\sum_{k=0}^{\infty} |P[X = k] - P[Y = k]| \leq 4p$$

- The examples of where Poisson distribution should be a good approximation include
  1. # of wrong numbers dialled in a day
  2. # of oranges sold in a day at a store
  3. # number of alpha particles emitted by a radioactive substance in 1 second
  4. # of dead pixels in an LCD display

- Intuitively, suppose $X \sim \text{Binomial}(n, p)$ with large n, small p and $\lambda = np$.

- In this case, we have:

$$E[X] = np = \lambda$$
$$Var[X] = np(1-p) = \lambda(1-p) \approx \lambda$$

- More precisely, let $X \sim \text{Poisson}(\lambda)$. Then

$$E[X] = \sum_{k=0}^{\infty} k \frac{\lambda^k}{k!} e^{-\lambda} = \sum_{k=1}^{\infty} k \frac{\lambda^k}{k!} e^{-\lambda} = \sum_{k=1}^{\infty} \frac{\lambda^k}{(k-1)!} e^{-\lambda} =$$

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

Mean and Variance of Poisson RV (cont.)

- At the same time, we have

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

$$= \sum_{\ell=0}^{\infty} \frac{(\ell+1)\lambda^{\ell+1}}{\ell!} e^{-\lambda} = \lambda \Big(\underbrace{\sum_{\ell=0}^{\infty} \frac{\ell\, \lambda^\ell}{\ell!} e^{-\lambda}}_{\lambda} + \underbrace{\sum_{\ell=0}^{\infty} \frac{\lambda^\ell}{\ell!} e^{-\lambda}}_{1} \Big) = \lambda(1+\lambda)$$

- Consequently, the variance of X can then be computed as

$$Var[X] = E[X^2] - (E[X])^2 = \lambda(1+\lambda) - (\lambda)^2 = \lambda$$

- Thus, in the case of Poisson *r.v.* we have

$$E[X] = Var[X] = \lambda$$

Example

- A radioactive substance with a large number of atoms emits 3.2 alpha particles per second on average. What is the probability that no more than 2 alpha particles are emitted in a 1 second interval?

- *Solution:* If X denotes the number of emitted particles in one second, then X is Poisson with $E[X] = 3.2 = \lambda$. In this case,

$$P[X \leq 2] = P[X = 0] + P[X = 1] + P[X = 2] =$$

$$= e^{-3.2} + 3.2\, e^{-3.2} + \frac{(3.2)^2}{2!} e^{-3.2} \approx 0.3799$$

Geometric RV

- Consider an infinite sequence of independent Bernoulli(p) trials.
- Let X be the *trial number* of the first outcome that is a one. X is called **geometric** with parameter p, denoted $X \sim \text{Geometric}(p)$. Its PMF is given by

$$p_X(k) = P[(k-1) \text{ zeros followed by a one}] =$$

$$= \begin{cases} (1-p)^{k-1}p & k \geq 1 \\ 0 & \text{else} \end{cases}$$

with $k = 1, 2, \ldots$

Example

- A bag contains 2 white balls and 3 black balls. Balls are drawn randomly from the bag until a black ball is drawn. If each selected ball is replaced before the next is drawn,
  1. What is the probability that exactly n draws are needed?
  2. What is the probability that at least k draws are needed?

- *Solution:* In each draw, the probability of getting a black ball is equal to $3/5 = 0.6$. If X denotes the number of draws until a black ball, then $X \sim \text{Geometric}(p)$ with $p = 0.6$.

- Consequently,

$$P[X = n] = \left(1 - \frac{3}{5}\right)^{n-1} \cdot \frac{3}{5} = \frac{3}{5}\left(\frac{2}{5}\right)^{n-1}$$

while

$$P[X \geq k] = \sum_{n=k}^{\infty} P[X = n] = \frac{3}{5} \cdot \sum_{n=k}^{\infty} \left(\frac{2}{5}\right)^{n-1} =$$

$$= \frac{3}{5} \cdot \left(\frac{2}{5}\right)^{k-1} \sum_{n=0}^{\infty} \left(\frac{2}{5}\right)^{n} = \frac{3}{5} \cdot \left(\frac{2}{5}\right)^{k-1} \frac{1}{1 - 2/5}$$

Mean and Variance of Geometric RV

- If $X \sim \text{Geometric}(p)$, then

$$E[X] = \sum_{k=1}^{\infty} k(1-p)^{k-1}p = \sum_{k=1}^{\infty}(k-1)(1-p)^{k-1}p + \underbrace{\sum_{k=1}^{\infty}(1-p)^{k-1}p}_{1} =$$

$$= \sum_{\ell=0}^{\infty} \ell(1-p)^{\ell}p + 1 = \sum_{\ell=1}^{\infty} \ell(1-p)^{\ell}p + 1 = (1-p)\underbrace{\sum_{\ell=1}^{\infty} \ell(1-p)^{\ell-1}p}_{E[X]} + 1$$

$$= (1-p)E[X] + 1$$

Therefore, $E[X] = 1/p$.

- One can also show that

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

- Finally,

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

- Recall, a random variable X is really $X(s)$ – a function of the outcome s of a random experiment.
- We can have two functions of the same outcome s, say $X(s)$ and $Y(s)$.
- For example, let $X = \#$ heads in first 3 flips and $Y = \#$ heads in last 2 flips.
- Since X and Y are numbers, we can add them: $Z(s) = X(s) + Y(s)$. In other words, Z is also a random variable.
- Here, $Z = \#$ of heads in all 5 flips.

Expected Values of Sums of RVs (cont.)

- Now, for each $s \in S$, let $p(s) = P[\{s\}]$. Then $P[A] = \sum_{s \in A} p(s)$.
- Let $X \in \mathcal{X} = \{x_1, \ldots, x_n\}$ and $A_k = \{s \in S \mid X(s) = x_k\}$. Then,

$$E[X] = \sum_{k=1}^{n} x_k P[X = x_k] = \sum_{k=1}^{n} x_k P[A_k] = \sum_{k=1}^{n} x_k \sum_{s \in A_k} p(s) =$$

$$= \sum_{k=1}^{n} \sum_{s \in A_k} x_k p(s) = \sum_{k=1}^{n} \sum_{s \in A_k} X(s) p(s) = \sum_{s \in S} X(s) p(s)$$

Example

- Two independent flips of a fair coin are made. Let $X = \#$ heads. Then,

$$P[X=0]=1/4, \quad P[X=1]=1/2, \quad P[X=2]=1/4$$

  and thus

$$E[X] = 0\cdot 1/4 + 1\cdot 1/2 + 2\cdot 1/4 = 1$$

- Also, $S = \{tt, th, ht, hh\}$, and each outcome has probability of $1/4$. So,

$$E[X] = X(tt)\times 1/4 + X(th)\times 1/4 + X(ht)\times 1/4 + X(hh)\times 1/4 =$$

$$= 0\cdot 1/4 + 1\cdot 1/4 + 1\cdot 1/4 + 2\cdot 1/4 = 1$$

- **Proposition:** For random variables $X_1, X_2, \ldots, X_n$, one has

$$E[X_1 + \ldots + X_n] = E[X_1] + \ldots + E[X_n]$$

- To see this, let $Z = X_1 + \ldots + X_n$. Then

$$E[Z] = \sum_{s \in S} Z(s)p(s) = \sum_{s \in S} (\,X_1(s) + \ldots + X_n(s))p(s) =$$

$$= \sum_{s \in S} X_1(s)p(s) + \ldots + \sum_{s \in S} X_n(s)p(s) = E[X_1] + \ldots + E[X_n]$$

- Let $X \sim \text{Binomial}(n, p)$. Then

$$X = X_1 + \ldots + X_n$$

where each $X_k \sim \text{Bernoulli}(p)$ is the outcome of an independent trial.

- In this case,

$$E[X] = E[X_1 + \ldots + X_n] = E[X_1] + \ldots + E[X_n] = p + \ldots + p = np$$

while

$$E[X^2] = E\left[\left(\sum_{k=1}^{n} X_k\right)\left(\sum_{\ell=1}^{n} X_\ell\right)\right] = E\left[\sum_{k=1}^{n}\left(\sum_{\ell=1}^{n} X_k X_\ell\right)\right] =$$

$$= E\left[\sum_{k=1}^{n}\left(X_k X_k + \sum_{\substack{\ell=1 \\ \ell \neq k}}^{n} X_k X_\ell\right)\right] = E\left[\sum_{k=1}^{n} X_k^2 + \sum_{k=1}^{n}\sum_{\substack{\ell=1 \\ \ell \neq k}}^{n} X_k X_\ell\right] =$$

$$= \sum_{k=1}^{n} E\left[X_k^2\right] + \sum_{k=1}^{n}\sum_{\substack{\ell=1 \\ \ell \neq k}}^{n} E\left[X_k X_\ell\right]$$

Example (cont.)

- Now,

$$P[X_k^2 = 1] = P[X_k = 1] = p$$

and

$$P[X_k\, X_\ell = 1] = P[X_k = 1, X_\ell = 1] = P[X_k = 1]P[X_\ell = 1] = p^2$$

- So, we finally obtain

$$E[X^2] = np + n(n-1)p^2$$

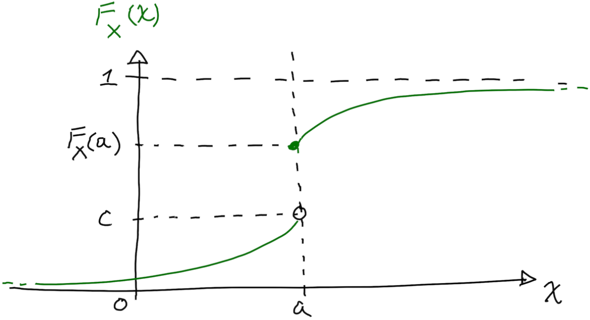
Properties of CDFs

- Recall that $F_X(x) = P[X \leq x]$. Therefore, $0 \leq F_X(x) \leq 1$.
- If $a < b$ then $\{X \leq a\} \subset \{X \leq b\}$. This suggests that

$$P[X \leq a] \leq P[X \leq b] \quad \Longrightarrow \quad F_X(a) \leq F_X(b)$$

- This shows that $F_X(x)$ is *non-decreasing* in x.
- It can also be shown that
  1. $$\lim_{b \to \infty} F_X(b) = 1, \quad \lim_{b \to -\infty} F_X(b) = 0$$
  2. $F_X(x)$ is continuous from the right (i.e., if $b_n \downarrow b$ then $\lim_{n \to \infty} F_X(b_n) = F_X(b)$)
  3. $F_X(x)$ has left limits (i.e., if $b_n \uparrow b$ then $\lim_{n \to \infty} F_X(b_n)$ exists)
- A function with these two limit properties is called *càdlàg*.

Example



- Here:

$$\lim_{b \downarrow a} F_X(b) = F_X(a)$$

$$\lim_{b \uparrow a} F_X(b) = c \neq F_X(a)$$