Chapter 3
Discrete Random Variables

Random Variables

Defn: A **random variable (r.v.)** is a real-valued function of the outcome of an experiment involving randomness.

Example: Experiment: Roll two dice

Q: Here are some r.v.s. What values can these take on?

X = sum of the rolls
Y = difference of the rolls
Z = max of the rolls
W = value of the first roll

We can now ask, "What is $\boldsymbol{P}\{X = 11\}$?"

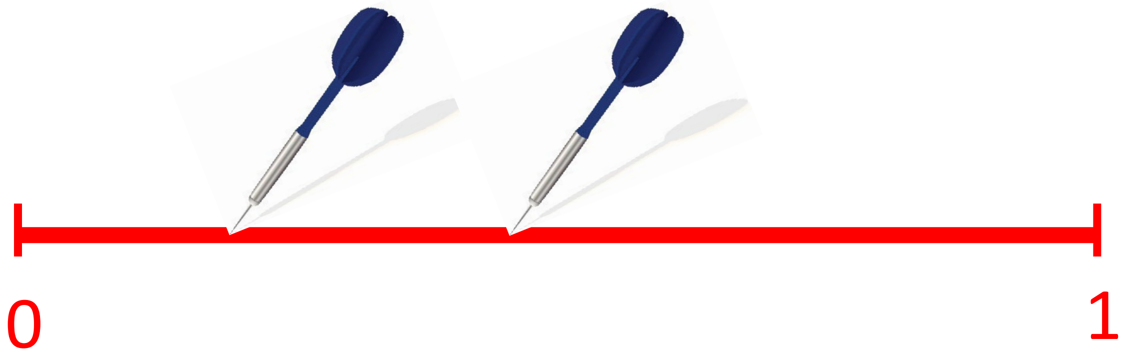
Random Variables

<u>Defn</u>: A **random variable (r.v.)** is a real-valued function of the outcome of an experiment involving randomness.

<u>Example</u>: Throw 2 darts uniformly at random at unit interval

Here are some random variables:

D = difference in location of the 2 darts
L = location of leftmost dart



Q: Can you define some more r.v.s?

Random Variables

Defn: A **discrete random variable** can take on at most a countably infinite number of possible values, whereas a **continuous random variable** can take on an uncountable set of possible values.

Q: Which of these random variables is discrete and which is continuous?

- ❑ The sum of the rolls of two dice
- ❑ The number of arrivals at a website by time t
- ❑ The time until the next arrival at a website
- ❑ The CPU time requirement of an HTTP request

From Random Variables to Events

We use CAPITAL letters to denote random variables.

When we set a random variable (r.v.) equal to a value, we get an event, and all the theorems we learned about events and their probabilities now apply.

Random Variable (R.V.)	**Event**	**Probability of Event**
X = sum of 2 rolls of a die	$X = 7$	$\frac{1}{6}$
N = number arrivals to a website within the next hour	$N > 10$	$\boldsymbol{P}\{N > 10\} = \boldsymbol{P}\{N > 10 \mid \text{weekday}\} \cdot \frac{5}{7} + \boldsymbol{P}\{N > 10 \mid \text{weekend}\} \cdot \frac{2}{7}$

Discrete Random Variables

Defn: A **discrete r.v.** takes on a countable number of values, each with some probability.

A discrete r.v. is associated with a **discrete distribution** that represents the likelihood of each of these values occurring. We sometimes define a r.v. by its associated distribution.

Defn: For a discrete r.v. X, the **probability mass function** of X is:

$$p_X(a) = \mathbf{P}\{X = a\}$$

The **cumulative distribution function** of X is:

$$F_X(a) = \mathbf{P}\{X \leq a\} = \sum_{x \leq a} p_X(x)$$

The **tail** of X is:

$$\bar{F}_X(a) = \mathbf{P}\{X > a\} = 1 - F_X(a)$$

Q: What is this?

$$\sum_x p_X(x)$$

Common Discrete R.V.s / Distributions

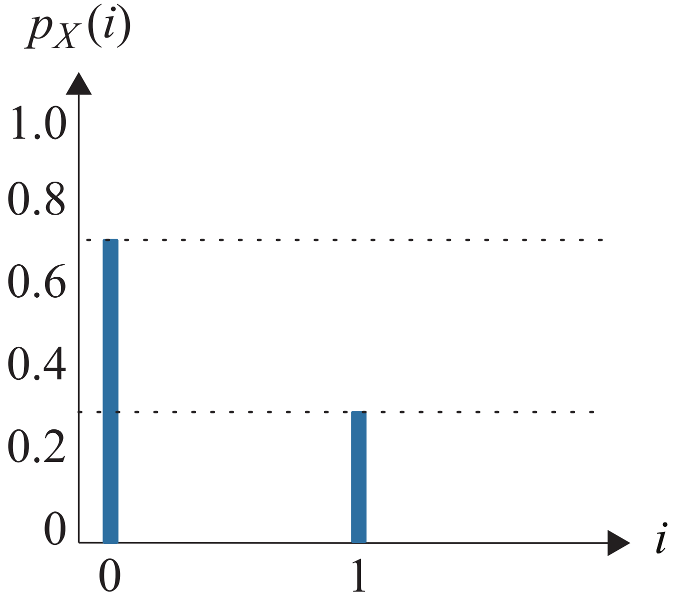
Bernoulli(p)

Experiment: Flip a single coin, with probability p of Heads.

Random Variable X = value of the coin flip

<u>Defn</u>: $X \sim Bernoulli(p)$:

$$X = \begin{cases} 1 & \text{w.p. } p \\ 0 & \text{w.p. } 1-p \end{cases}$$

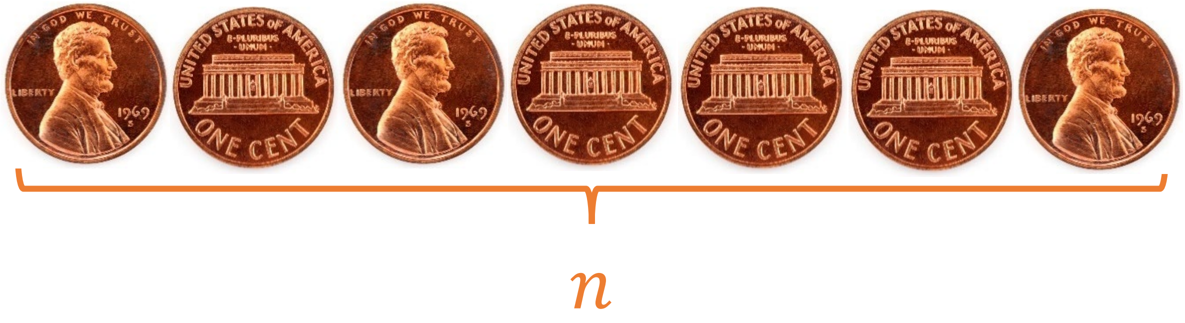


Q: What distribution is shown above, with what parameter?

Binomial(n, p)

Experiment: Flip a coin, with probability p of Heads, n times

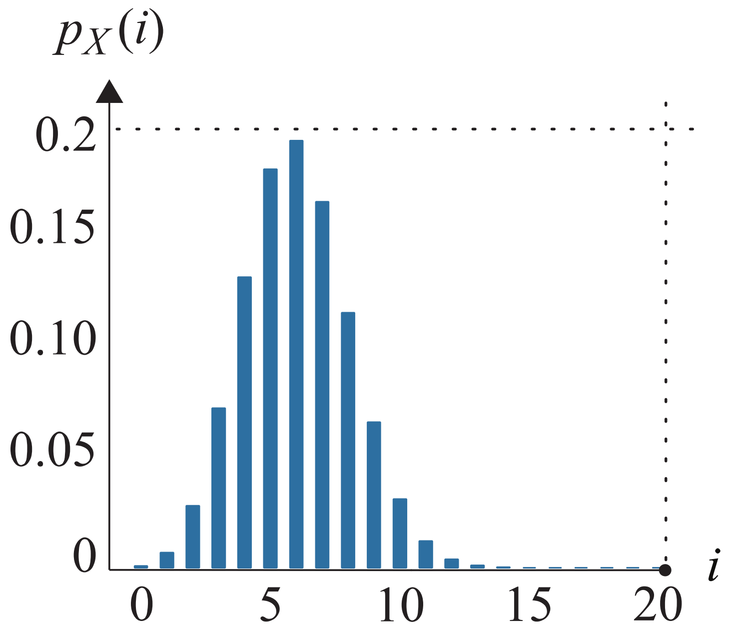
Random Variable X = number of heads



<u>Defn</u>: $X \sim Binomial(n, p)$:

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

where $i = 0, 1, 2, \ldots, n$



Binomial$(n = 20, p = 0.3)$

Q: What is this?

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

(Hint: binomial expansion)

Geometric(p)

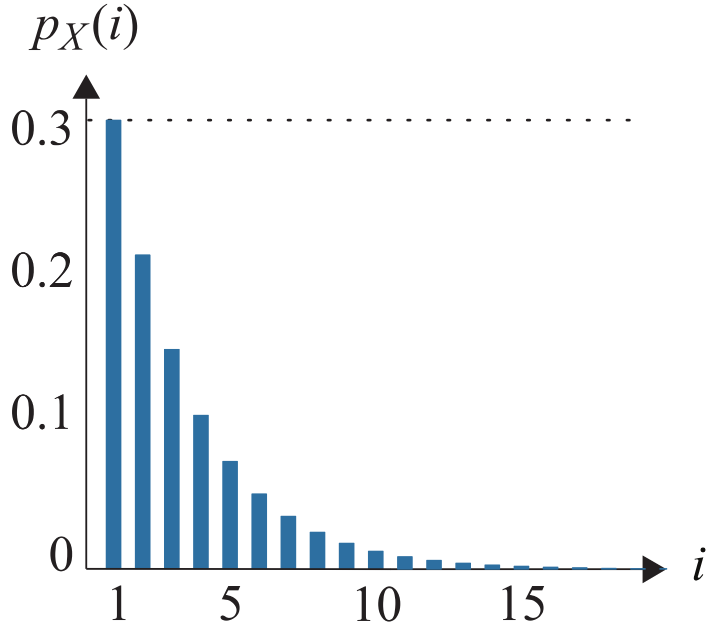
Experiment: Flip a coin, with probability p of Heads, until see first head

Random Variable X = number flips until first head

Defn: $X \sim Geometric(p)$:

$$p_X(i) = (1-p)^{i-1} \cdot p$$

where $i = 1, 2, 3, \ldots$



Geometric($p = 0.3$)

Q: What is:

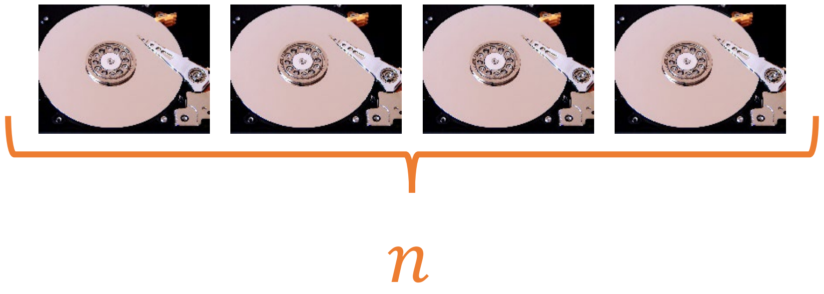
$$\bar{F}_{\boldsymbol{X}}(i) = \boldsymbol{P}\{X > i\}?$$

Q: What is this?

$$\sum_{i=1}^{\infty} (1-p)^{i-1} \cdot p$$

Pop Quiz

Q: You have a room of n disks.
Each disk independently dies with probability p.
How are the following quantities distributed?



a) The number of disks that die in the first year Binomial(n, p)
b) The number of years until a particular disk dies Geometric(p)
c) The state of a particular disk after one year Bernoulli(p)

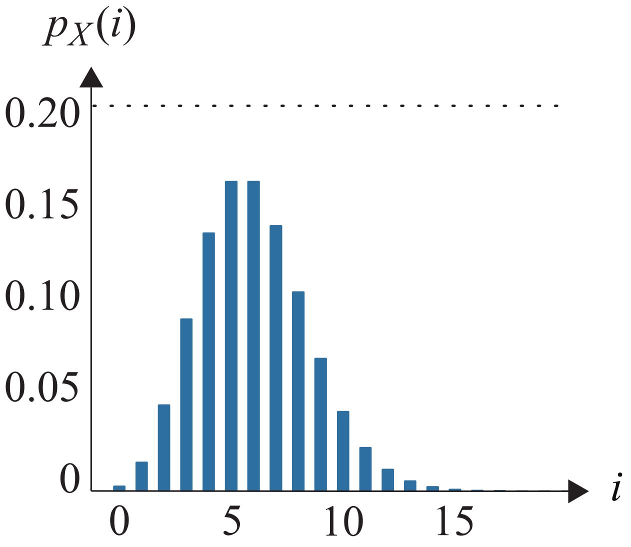
Poisson(λ)

The Poisson distribution occurs naturally when looking at a mixture of a large number of independent sources.

<u>Defn</u>: $X \sim Poisson(\lambda)$:

$$p_X(i) = \frac{e^{-\lambda} \cdot \lambda^i}{i!}$$

where $i = 0, 1, 2, 3, \ldots$



Poisson($\lambda = 6$)

Q: What is this?

$$\sum_{i=0}^{n} \frac{e^{-\lambda} \cdot \lambda^i}{i!}$$

(Hint: Taylor series of e^{λ})

Q: Does the shape of the Poisson p.m.f. remind you of another distribution?

Two Random Variables

Defn: The **joint probability mass function** between discrete r.v.'s X and Y is:

$$p_{X,Y}(x,y) = \boldsymbol{P}\{X = x \ \& \ Y = y\}$$

or equivalently, $\boldsymbol{P}\{X = x\ , Y = y\}$ or $\boldsymbol{P}\{X = x \cap \ Y = y\}$,
where, by definition:

$$\sum_x \sum_y p_{X,Y}(x,y) = 1.$$

Marginal Probability Mass Function

How is $p_X(x)$ related to $p_{X,Y}(x,y)$?

Table shows $p_{X,Y}(x,y)$

	$X=0$	$X=1$	$X=2$
$Y=0$	0.4	0.05	0.05
$Y=1$	0.05	0.05	0.1
$Y=2$	0.1	0.2	0

$p_Y(1) = 0.2$

$$p_Y(y) = \sum_x p_{X,Y}(x,y)$$

$p_X(0) = 0.55$

$$p_X(x) = \sum_y p_{X,Y}(x,y)$$

Called "**marginal probabilities**" because written in the margins.

Independence

<u>Defn</u>: Discrete random variables X and Y are **independent** (written $X \perp Y$) if :

$$p_{X,Y}(x, y) = p_X(x) \cdot p_Y\,(y), \qquad \forall x, y$$

Q: If X and Y are independent, what does this say about $\boldsymbol{P}\{X = x \mid Y = y\}$?

Independence

Defn: Discrete random variables X and Y are **independent** (written $X \perp Y$) if :

$$p_{X,Y}(x, y) = p_X(x) \cdot p_Y\,(y), \qquad \forall x, y$$

Q: If X and Y are independent, what does this say about $\boldsymbol{P}\{X = x \mid Y = y\}$?

$$\boldsymbol{P}\{X = x \mid Y = y\} = \frac{\boldsymbol{P}\{X = x \,\&\, Y = y\}}{\boldsymbol{P}\{Y = y\}}$$

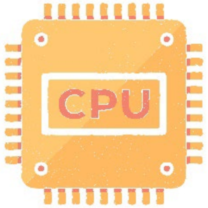
$$= \frac{\boldsymbol{P}\{X = x\} \cdot \boldsymbol{P}\{Y = y\}}{\boldsymbol{P}\{Y = y\}}$$

$$= \boldsymbol{P}\{X = x\}$$

Who Fails First?

You have a disk with probability p_1 of failing each day, and a CPU which independently has probability p_2 of failing each day.

$\boldsymbol{p_1}$



$\boldsymbol{p_2}$

Q: What is the probability that the disk fails *before* the CPU?

Intuitively, what answer makes sense?

Who Fails First?

You have a disk with probability p_1 of failing each day, and a CPU which independently has probability p_2 of failing each day.

Q: What is the probability that the disk fails *before* the CPU?

$X_1 =$ days until disk fails $\sim Geometric(p_1)$

$X_2 =$ days until CPU fails $\sim Geometric(p_2)$

$X_1 \perp X_2$

$$\boldsymbol{P}\{X_1 < X_2\} = \sum_{k=1}^{\infty} \sum_{k_2=k+1}^{\infty} p_{X_1,X_2}(k, k_2) = \sum_{k=1}^{\infty} \sum_{k_2=k+1}^{\infty} p_{X_1}(k) \cdot p_{X_2}(k_2)$$

$$= \sum_{k=1}^{\infty} \sum_{k_2=k+1}^{\infty} (1-p_1)^{k-1} p_1 \cdot (1-p_2)^{k_2-1} p_2$$

Who Fails First?

You have a disk with probability p_1 of failing each day, and a CPU which independently has probability p_2 of failing each day.

Q: What is the probability that the disk fails *before* the CPU?

$X_1 =$ days until disk fails $\sim Geometric(p_1)$

$X_2 =$ days until CPU fails $\sim Geometric(p_2)$

$X_1 \perp X_2$

$$\boldsymbol{P}\{X_1 < X_2\} = \sum_{k=1}^{\infty} \sum_{k_2=k+1}^{\infty} (1-p_1)^{k-1} p_1 \cdot (1-p_2)^{k_2-1} p_2 = \ldots$$

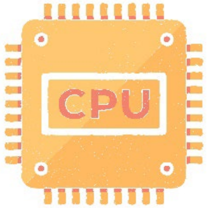
after some algebra

$$= \frac{p_1(1-p_2)}{1-(1-p_2)(1-p_1)}$$

Who Fails First?

You have a disk with probability p_1 of failing each day, and a CPU which independently has probability p_2 of failing each day.

$\boldsymbol{p_1}$



$\boldsymbol{p_2}$

$$\boldsymbol{P}\{\text{disk fails before CPU fails}\} = \frac{p_1(1 - p_2)}{1-(1-p_2)(1-p_1)}$$

But WHY?

Who Fails First?

You have a disk with probability p_1 of failing each day, and a CPU which independently has probability p_2 of failing each day.

$$\boldsymbol{P}\{\text{disk fails before CPU fails}\} = \frac{p_1(1-p_2)}{1-(1-p_2)(1-p_1)}$$

Intuition: Think about flipping 2 coins each day.
There may be many days where both coins are tails.
We only care about the *first day where the coins are not both tails*.

Given that both coins are not tails, what's the probability that coin 1 is H and coin 2 is T?

$$\boldsymbol{P}\{\text{coin 1 is H \& coin 2 is T} \mid \text{not both tails}\} = \frac{\boldsymbol{P}\{\text{coin 1 is H \& coin 2 is T}\}}{\boldsymbol{P}\{\text{not both tails}\}} = \frac{p_1(1-p_2)}{1-(1-p_2)(1-p_1)}$$

Law of Total Probability

Theorem: **[Law of Total Probability for Discrete R.V.s]**
Let E be an event. Let Y be a discrete r.v.

$$\boldsymbol{P}\{E\} = \sum_y \boldsymbol{P}\{E \cap Y = y\} = \sum_y \boldsymbol{P}\{E \mid Y = y\} \cdot \boldsymbol{P}\{Y = y\}$$

For a discrete r.v. X :

$$\boldsymbol{P}\{X = k\} = \sum_y \boldsymbol{P}\{X = k \cap Y = y\} = \sum_y \boldsymbol{P}\{X = k \mid Y = y\} \cdot \boldsymbol{P}\{Y = y\}$$

Proof: Follows immediately from Law of Total Probability for Events, if we realize that $Y = y$ represents an event and the set of events $Y = y$ over all y form a partition.

Who Fails First?

Disk with prob. p_1 of failing each day, and a CPU with indpt. prob. p_2 of failing each day.

Q: What is the probability that the disk fails *before* the CPU? (Redo using conditioning!)

X_1 = days until disk fails $\sim Geometric(p_1)$ X_2 = days until CPU fails $\sim Geometric(p_2)$

$X_1 \perp X_2$

$$\begin{aligned}
\boldsymbol{P}\{X_1 < X_2\} &= \sum_{k=1}^{\infty} \boldsymbol{P}\{X_1 < X_2 \mid X_1 = k\} \cdot \boldsymbol{P}\{X_1 = k\} \\
&= \sum_{k=1}^{\infty} \boldsymbol{P}\{k < X_2 \mid X_1 = k\} \cdot \boldsymbol{P}\{X_1 = k\} \\
&= \sum_{k=1}^{\infty} \boldsymbol{P}\{X_2 > k\} \cdot \boldsymbol{P}\{X_1 = k\} \\
&= \sum_{k=1}^{\infty} (1 - p_2)^k \cdot (1 - p_1)^{k-1} \cdot p_1 = \frac{p_1(1 - p_2)}{1 - (1 - p_2)(1 - p_1)}
\end{aligned}$$