CS 237: Probability in Computing

Wayne Snyder
Computer Science Department
Boston University

Lecture 12:

- Review: Binomial, Geometric, Pascal
- Expectation of Random Variables
- Fair Games
- Properties of Expectation
    - Linearity of Expectation
    - Expectation of Sum of Independent Random Variables

Binomial Distribution

The Binomial Distribution occurs when you count the number of successes in N independent and identically distributed Bernoulli Trials (i.e., p is the same each time).

Formally, if Y ~ Bernoulli(p), and

$$X = \text{"The number of successes in N trials of Y"} = \overbrace{Y + Y + \ldots + Y}^{\text{N times}}$$

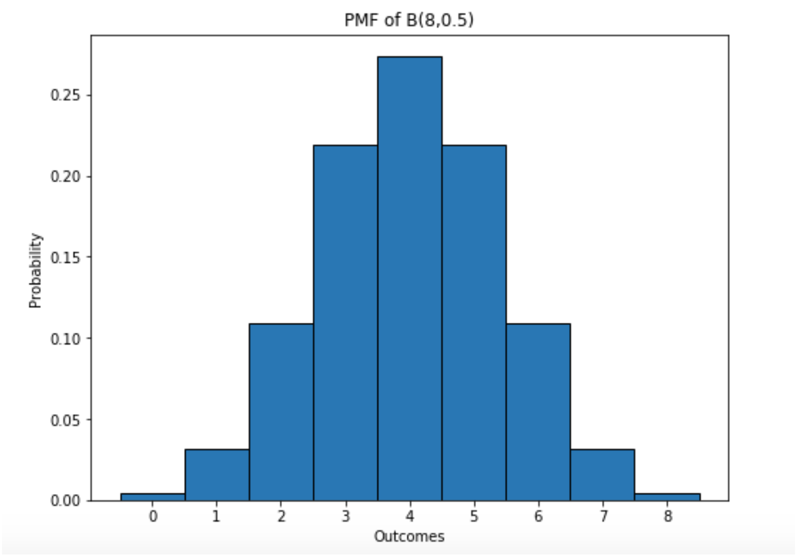
then we say that X is distributed according to the Binomial Distribution with parameters N and p, and write this as:

$$X \sim B(N, p)$$

Where

$$R_X = \{\, 0, \ldots, N \,\}$$

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



Note: k successes and N-k failures SSS…FFFF…

has probability $p^k (1\text{-}p)^{N\text{-}k}$ and there are $\binom{N}{k}$ such sequences.

Geometric Distribution

The Geometric Distribution occurs when you count the number of independent and identically distributed Bernoulli trials until the first success.

Formally, if Y ~ Bernoulli(p), and

X = "The number of trials of Y until the first success"

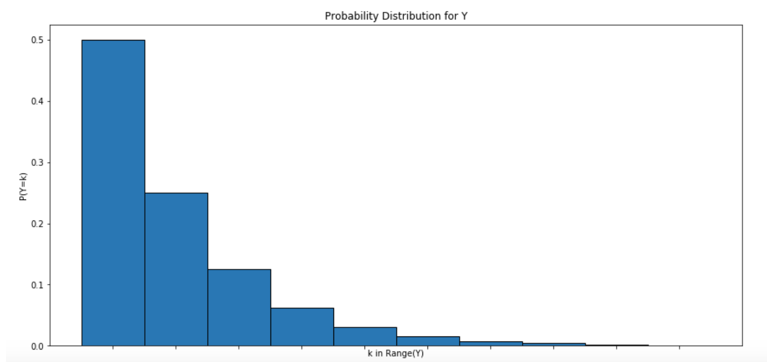
then we say that X is distributed according to the Geometric Distribution with parameter p, and write this as:

$$X \sim G(p)$$

where

$$R_X = \{ 1, 2, 3, \dots \}$$

$$P_X(k) = (1-p)^{k-1} p$$



For k, we have k-1 failures and 1 success (FFF... FS), which has probability $(1-p)^{k-1}p$.

Negative Binomial (Pascal) Distribution

The Negative Binomial is simply an "iterated" version of the Geometric.

Formally, if Y ~ Bernoulli(p) and

X = "The number of trials of Y until m successes occur"

$$= \underbrace{Y + \ldots + Y}_{\text{m times}}$$

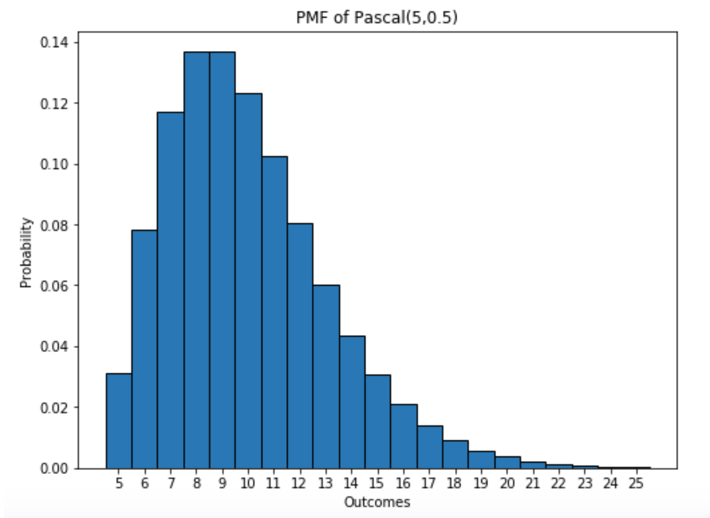
then we say that X is distributed according to the Pascal Distribution with parameters m and p, and write:

$$X \sim Pascal(m, p)$$

where

$$R_X = \{m, m+1, m+2, \ldots\}$$

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



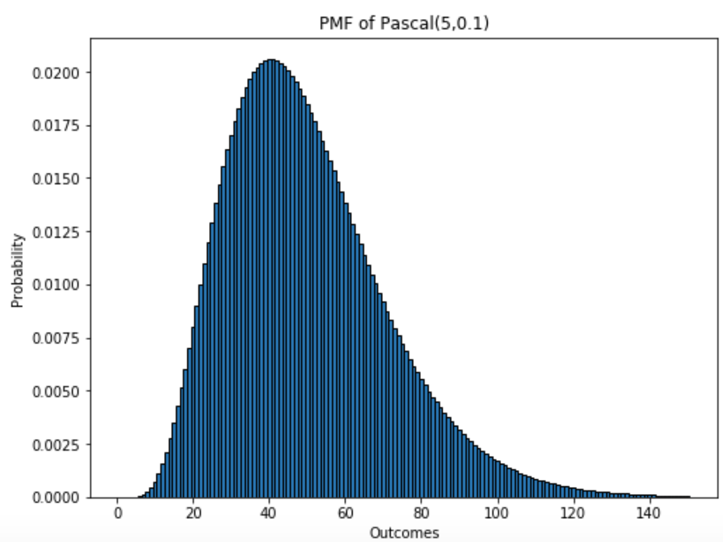
Pascal Distribution

Example

Suppose you are throwing darts at a target for practice, and you decide that you will keep throwing until you hit the bull's eye 5 times. Suppose your probability of hitting the bullseye is 10%. What is the probability it takes exactly 10 throws?

Solution:

$$P_X(10) = \binom{9}{4} 0.1^5 \cdot 0.9^5 = 7.44 \times 10^{-4}$$



C(9,4) * 0.1^(5) * 0.9 ^(5)

Extended Keyboard   Upload

Assuming "C" is a math function | Use as a

Input:

$\binom{9}{4} \times 0.1^5 \times 0.9^5$

Result:

0.0007440174

```
# Just a more convenient syntax:

from scipy.special import comb

def C(N,K):
    return comb(N,K,exact=True)
```

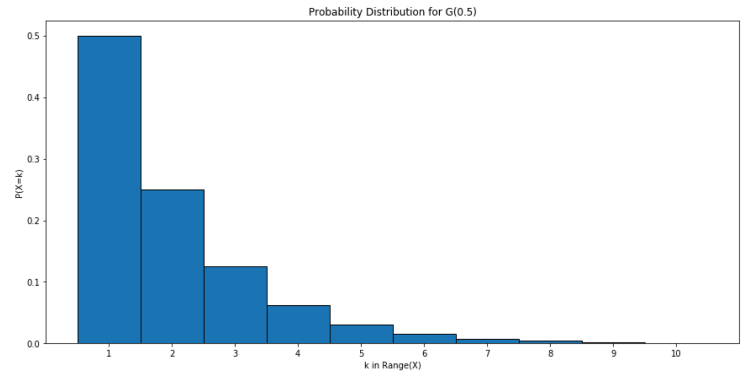
In [6]:

```
C(9,4)*0.1**5 * 0.9 ** 5
```

Out[6]: 0.0007440174000000003

Geometric Distribution: The Memoryless Property

The Geometric is one of two distributions that has the Memoryless Property, which we have already discussed informally before now as "the coin doesn't remember its past flips" but which is actually a much stronger statement about the entire distribution.



Intuitive Version: Suppose we are about to start flipping a coin for which the probability of heads is p. Then the probability distribution of X = "how many flips until the first head?" is G(p).

Now suppose that the first k flips are tails. Then the probability distribution of Y = "how many **more** flips until the first head?" is still G(p).

In other words, it doesn't matter when you start to count or what the past history is: the exact theoretical distribution is always the same.

Intuitive Proof: Suppose you come into the room while someone is flipping the coin. How do you know how many flips have occurred before you came in, and why would it matter? The distribution would be exactly the same regardless of the past.

Geometric Distribution: The Memoryless Property

Theorem A random variable X is called memoryless if, for any n, m ≥ 0,

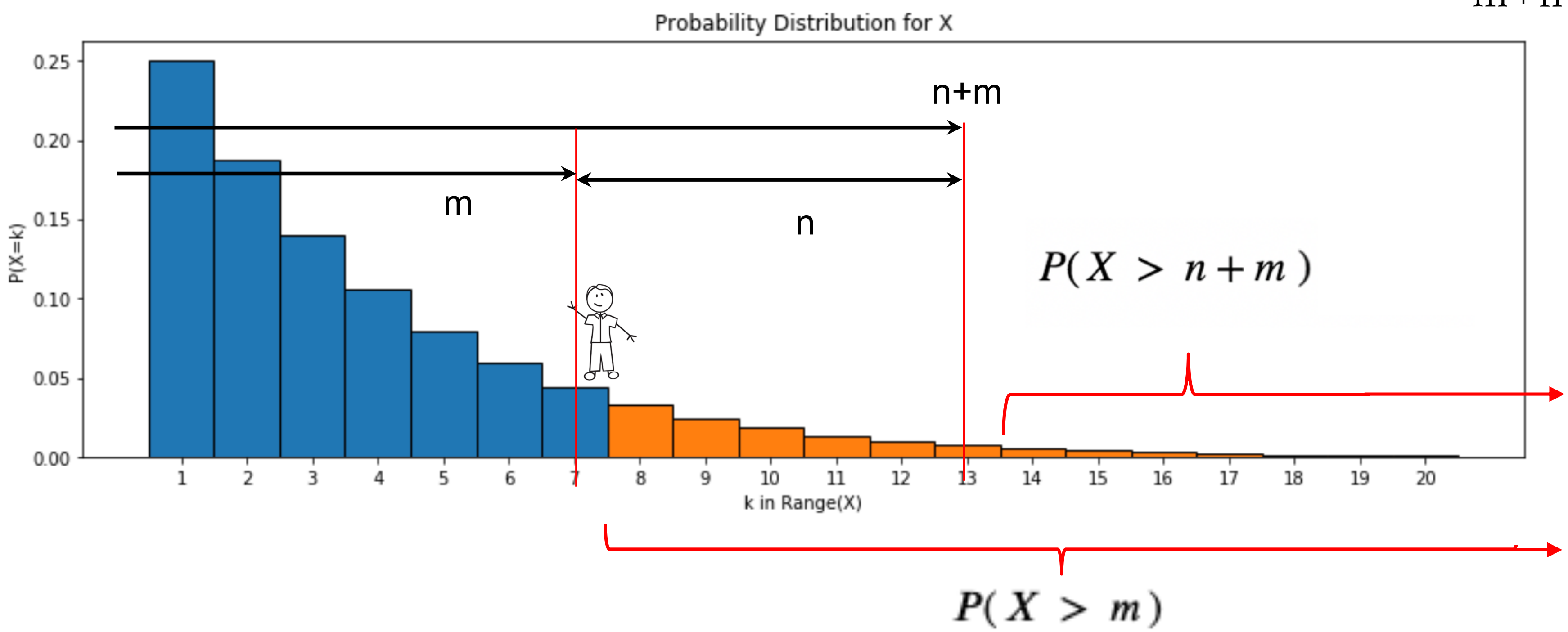
$$P(X > n+m \mid X > m) = P(X > n)$$

Example:

m = 7 n = 6

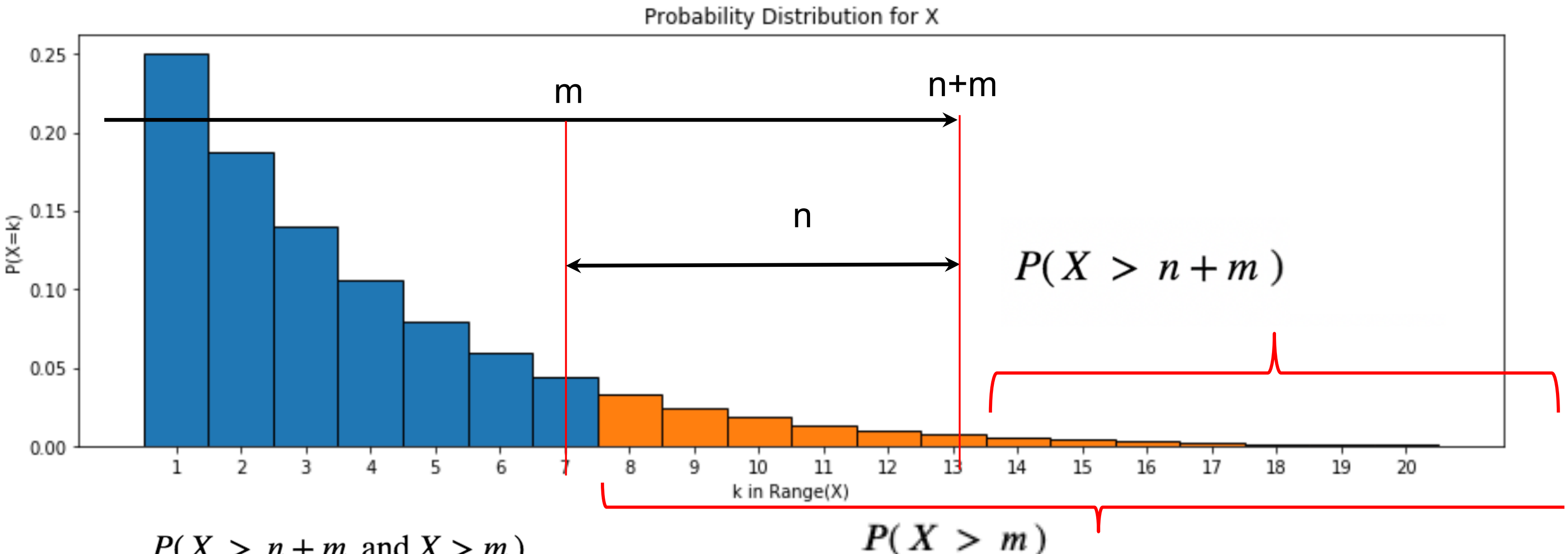
m+n = 13

Fact: For any probability p, X ~ G(p) has the memoryless property.



(In fact, the Geometric is the only discrete distribution with this property; a continuous version of the Geometric, called the Exponential, is the other one.)

Geometric Distribution: The Memoryless Property



Proof:

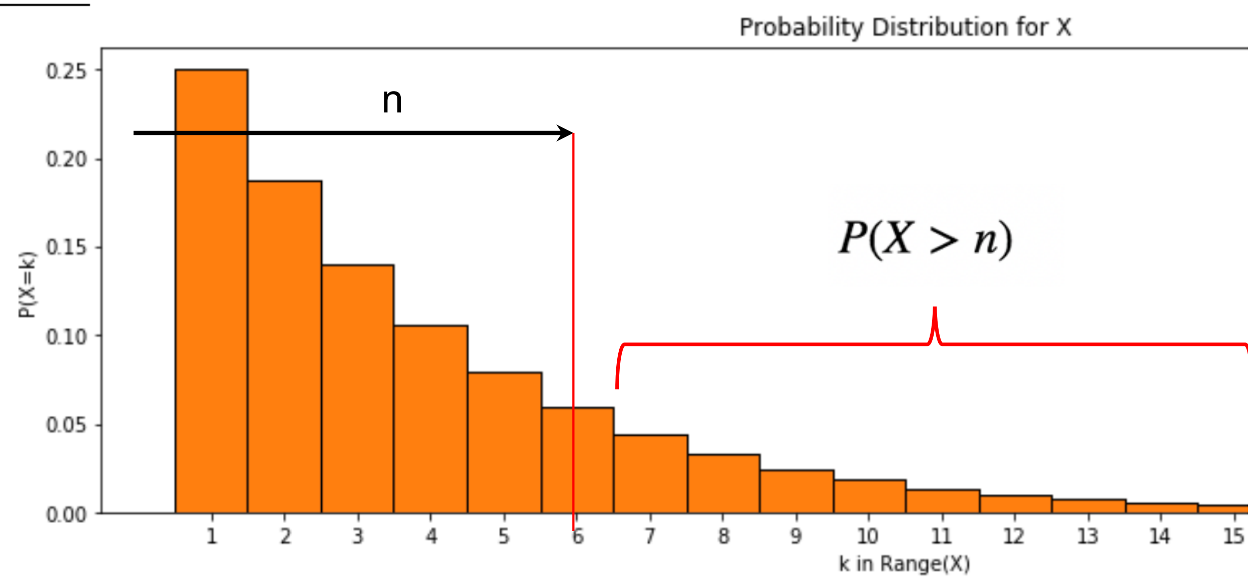
$$P(X > n+m \mid X > m) = \frac{P(X > n+m \text{ and } X > m)}{P(X > m)}$$

$$= \frac{P(X > n+m)}{P(X > m)}$$

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

$$= (1-p)^n$$

$$= P(X > n)$$



Geometric Distribution: The Memoryless Property

Why is this important? It simplies conditional probability problems!

Example 1: Suppose in Park Street a train arrives every 15 minutes, and the probability that an arriving train is for the Green line is 1/5.

(A) Suppose you have just arrived at Park Street and are waiting for a Green Line train. When you arrive, you just miss a Red Line train. What is the probability that you will have to wait at least 45 minutes (3 train arrivals) for a Green Line train?

(B) Suppose you have just arrived at Park Street and are waiting for a Green Line train. When you arrive, you just miss a Green Line train. What is the probability that you will have to wait at least 45 minutes (3 train arrivals) for a Green Line train?

(C) Suppose you are in Park Street, and 10 trains have arrived and left, and none of them was for the Green Line. What is the probability that you will have to wait at least 45 minutes (3 train arrivals) for the next Green Line train?"

Geometric Distribution: The Memoryless Property

Discrete Random Variables: Expected Value

A fundamental way of characterizing a collection of real numbers is the average or mean value of the collection:

Example: The mean/average of { 2, 4, 6, 9 } = 21/4 = 5.7

The corresponding notion for a random variable X is the Expected Value:

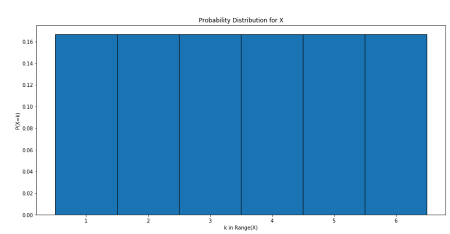
$$E(X) = \sum_{k \in R_X} k \cdot P(X = k)$$

Example: X = “the number of dots showing on a single thrown die”

$$R_X = \{1, 2, 3, 4, 5, 6\}$$

$$f_X = \{ \tfrac{1}{6}, \tfrac{1}{6}, \tfrac{1}{6}, \tfrac{1}{6}, \tfrac{1}{6}, \tfrac{1}{6} \}$$

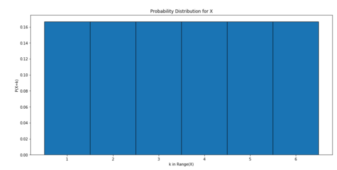
$$E(X) = \sum_{k \in R_X} \frac{k}{6} = \frac{1+2+3+4+5+6}{6} = \frac{21}{6} = 3.5$$



Discrete Random Variables: Expected Value

$R_X = \{1, 2, 3, 4, 5, 6\}$

$f_X = \{ \frac{1}{6}, \frac{1}{6}, \frac{1}{6}, \frac{1}{6}, \frac{1}{6}, \frac{1}{6} \}$

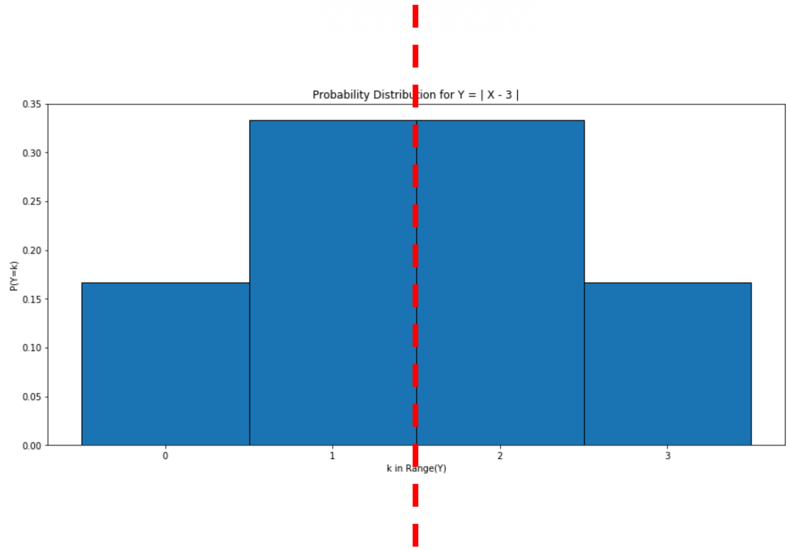


Example: Y = | X – 3 |

$E(Y)$

$R_Y = \{ 0, 1, 2, 3 \}$

$f_Y = \{ \frac{1}{6}, \frac{2}{6}, \frac{2}{6}, \frac{1}{6} \}$



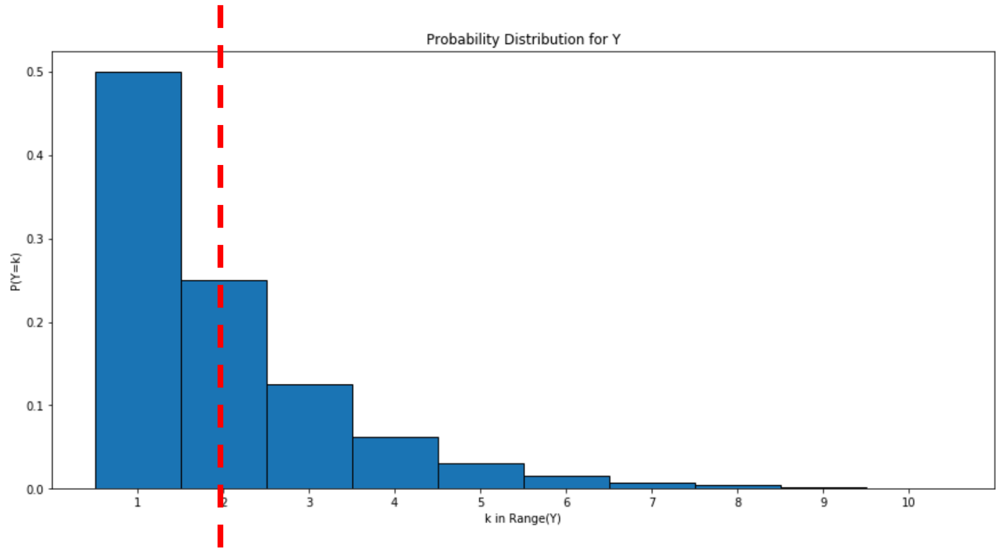
$$E(Y) = \sum_{k \in R_Y} k * f(k) = \frac{0}{6} + \frac{1}{3} + \frac{2}{3} = \frac{3}{6} = 1.5$$

Discrete RandomVariables: Expected Value

Example: Y = "tosses of a fair coin until a heads appears"

$$R_Y = \{\ 1, 2, 3,\ \dots\ \}$$

$$f_Y = \{\ \tfrac{1}{2}, \tfrac{1}{4}, \tfrac{1}{8},\ \dots \}$$



$$E(Y) = \sum_{k \in R_Y} k * f(k) = \frac{1}{2} + \frac{2}{4} + \frac{3}{8} + \frac{4}{16} \dots = 2.0$$

Expected Value and Fair Games

The expected value of a random variable is a **one-number summary of its behavior**. It describes what you can expect as the limiting behavior over many trials.

A good example of what this means occurs with games in which you win or lose money on each round or trial. Such a game can be modeled by a random variable: X = "the amount you win (+) or lose (-)." A game is **fair** if E(X) = 0.

Example: The rules of "**Chuck-a-luck**" are as follows. The player makes a bet on any number 1 through 6 and then three dice are thrown. If 1, 2, or 3 dice show the same number as the player's choice, then he or she wins back the original bet plus 1, 2, or 3 times the original bet.

- So if you bet \$1 on 4 and the dice roll 2, 4, and 6, you get back 1+1=2 for a net win of \$1.
- If you bet \$1 on 2 and the dice roll 2, 6, and 2, you get back 1 + 2 = 3 for a net win of \$2.
- If you bet \$1 on 5 and no 5's show and you lose \$1.

Expected Value and Fair Games

To analyze Chuck-a-Luck, suppose the player always bets \$1 on each round (or trial) of the game. Let

X = "net win or loss for one round." (net = winnings – loss/cost)

Then $R_X = \{ -1, 1, 2, 3 \}$

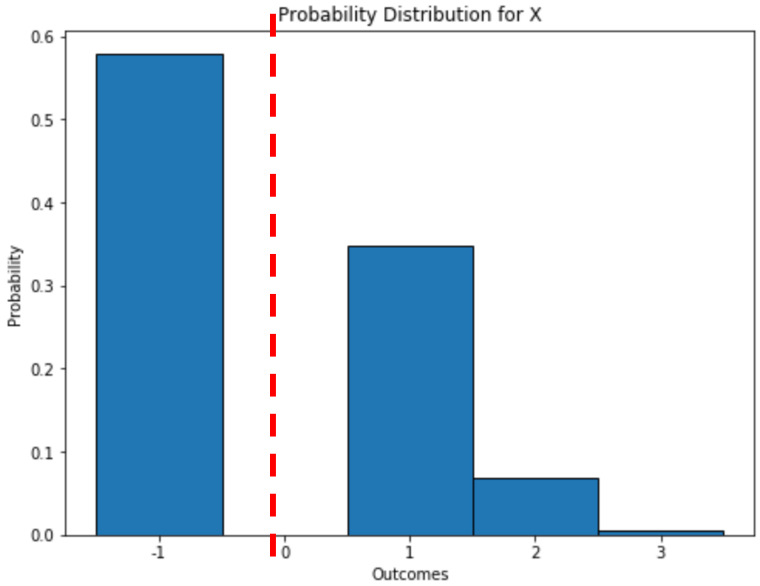
$$P(X = -1) = \binom{3}{0} * \frac{5}{6} * \frac{5}{6} * \frac{5}{6} = 0.5787$$

$$P(X = 1) = \binom{3}{1} * \frac{1}{6} * \frac{5}{6} * \frac{5}{6} = 0.3472$$

$$P(X = 2) = \binom{3}{2} * \frac{1}{6} * \frac{1}{6} * \frac{5}{6} = 0.0694$$

$$P(X = 3) = \binom{3}{3} * \frac{1}{6} * \frac{1}{6} * \frac{1}{6} = 0.0047$$

$$f_X = \{ 0.5787,\ 0.3472,\ 0.0694,\ 0.0047 \}$$



On average you lose about 8 cents per round, so Chuck-a-luck is not fair:

$$E(X) = \sum_{k \in R_X} k * f_X(k) = -0.5787 + 0.3472 + 2 * 0.0694 + 3 * 0.0047 = -0.0786$$

Expected Value: Basic Properties

Theorem (Linearity of Expectation)

For any random variable **X** and real numbers **a** and **b**,

$$E(a * X + b) = a * E(X) + b$$

Proof:

$$\begin{aligned} E(aX + b) &= \sum_{k \in R_X} (a * k + b) * P_X(k) \\ &= \sum_{k \in R_X} (a * k * P_X(k)) + (b * P_X(k)) \\ &= \sum_{k \in R_X} (a * k * P_X(k)) + \sum_{k \in R_X} (b * P_X(k)) \\ &= a * \sum_{k \in R_X} (k * P_X(k)) + b * \sum_{k \in R_X} P_X(k) \\ &= a * E(X) + b * 1.0 \\ &= a * E(X) + b \end{aligned}$$

This will make many calculations involving expected value MUCH easier!

(Obvious) Corollary: For any constant b, E(b) = b.

Expected Value: Basic Properties

Theorem (Expectation of Sums of Random Variables):

If X and Y are two discrete random variables (not necessarily independent), then:

$$
\begin{aligned}
E(X+Y) &= \sum_{j \in R_X} \sum_{k \in R_Y} (j+k) \cdot P(X=j, Y=k) \\
&= \sum_{j \in R_X} \sum_{k \in R_Y} j \cdot P(X=j, Y=k) + k \cdot P(X=j, Y=k) \\
&= \sum_{j \in R_X} \sum_{k \in R_Y} j \cdot P(X=j, Y=k) + \sum_{j \in R_X} \sum_{k \in R_Y} k \cdot P(X=j, Y=k) \\
&= \sum_{j \in R_X} j \cdot P(X=j) \ + \sum_{k \in R_Y} k \cdot P(X=j) \\
&= E(X) \ + \ E(Y)
\end{aligned}
$$

where in the second-to-last step, we used the Law of Total Probability:

If $S_1, \ldots, S_n$ is a partition of the sample space S, and A is an event, then $A \cap S_1, S \cap S_2, \ldots, S \cap S_n$ is a partition of the even A, and

$$\$P(A) \ = \sum_{1 \le i \le n} P(A, S_i)$$

(This is essentially case analysis, breaking A up into n disjoint cases.)

Expected Value: Basic Properties

Theorem (Expectation of Product of Independent Random Variables):

If X and Y are two *independent* discrete random variables, then:

$$\begin{aligned}
E(X \cdot Y) &= \sum_{j \in R_X} \sum_{k \in R_Y} j \cdot k \cdot P(X = j, Y = k) \\
&= \sum_{j \in R_X} \sum_{k \in R_Y} j \cdot k \cdot P(X = j) \cdot P(Y = k) \\
&= \sum_{j \in R_X} j \cdot P(X = j) \cdot \left(\sum_{k \in R_Y} k \cdot P(Y = k) \right) \\
&= \sum_{j \in R_X} j \cdot P(X = j) \cdot E(Y) \\
&= E(Y) \cdot \sum_{j \in R_X} j \cdot P(X = j) \\
&= E(Y) \cdot E(X) \\
&= E(X) \cdot E(Y)
\end{aligned}$$

where in the second step, we used the independence of X and Y.

Expected Value: Basic Properties

Note: This theorem is not true if X and Y are dependent:

It is easy to see that this result is *not* true for dependent variables: Consider the following. Flip a coin and let X count the number of heads and Y count the number of tails. Clearly X and Y are *not* indepedent, and in fact $Y = 1 - X$. Clearly

$$E(X) \cdot E(Y) = \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4}.$$

But we have (showing the product and then the probability in parentheses):

XY	0	1
0	0 (0)	0 (1/2)
1	0 (1/2)	1 (0)

So

$$E(X \cdot X) = 0 \cdot 0 + 0 \cdot \frac{1}{2} + 0 \cdot \frac{1}{2} + 1 \cdot 0 = 0$$

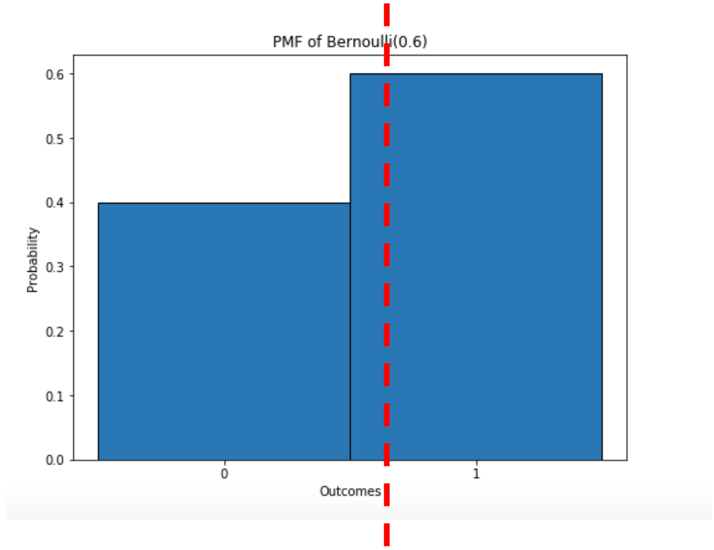
Expected Value of the Standard Distributions

Expected Value of Bernoulli

$$X \sim \text{Bernoulli}(p)$$

$$R_X = \{0, 1\}$$

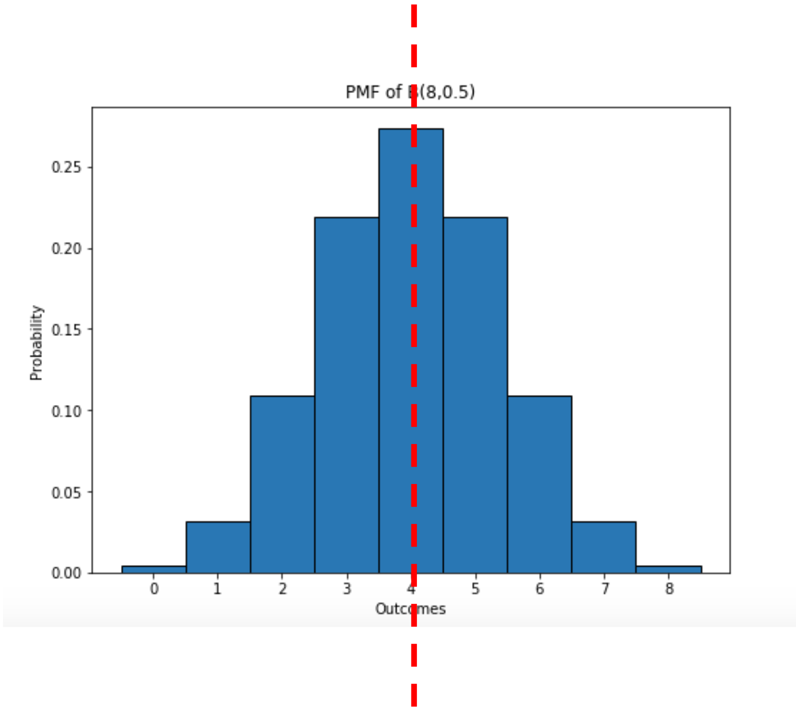
$$P_X = \{1 - p, p\}$$



$$E(X) = 1 \cdot p + 0 \cdot (1 - p) = p$$

Expected Value of the Standard Distributions

Expected Value of Binomial



$$X \sim B(N, p)$$

$$R_X = \{\, 0, \ldots, N \,\}$$

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

Formally, if Y ~ Bernoulli(p), and

$$X = \text{"The number of successes in N trials of Y"} = \overbrace{Y + Y + \ldots + Y}^{\text{N times}}$$

By the expectation of sums of independent RVs we immediately have:

E(X) = N * p

Geometric Distribution: Expected Value

To derive the expected value, we can use the fact that X ~ G(p) has the memoryless property and break into two cases, depending on the result of the first Bernoulli trial. Let

X_S = "result of X when there is a success on the first trial"
X_F = "result of X when there is a failure on the first trial"

Clearly,

- $E(X_S) = 1$
- $E(X_F) = 1 + E($ X for the remaining trials $)$

  $= 1 + E(X)$

  By the memoryless property!

Thus we have:

$$\begin{aligned} \mu_X &= 1\,p + (1-p)\,(1+\mu_X) \\ &= p + 1 - p + \mu_X - p\,\mu_X \\ &= 1 + \mu_X - p\,\mu_X \end{aligned}$$

$$\begin{aligned} 0 &= 1 - p\,\mu_X \\ p\,\mu_X &= 1 \\ \mu_X &= 1/p \end{aligned}$$

Geometric Distribution

Example

Suppose you draw cards WITH replacement until you get an Ace. How many draws would you expect it to take?

Solution: This is G(1/13). E(X) = 13

On average, how many independent games of poker are required until a particular player is dealt a **Royal Flush**?

Solution: This is G(0.00000154). E(X) = 1/ 0.00000154 = 649,350.6493

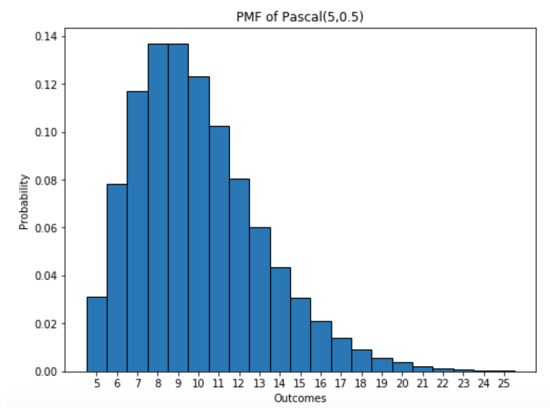
Pascal Distribution: Expected Value

Since the Pascal is simply an "iterated" version of the Geometric, we can use the linearity of expectation again!

Formally, if $Y \sim$ Bernoulli(p) and

X = "The number of trials of Y until m successes occur"

$$= \underbrace{Y_1 + \ldots + Y_m}_{\text{m times}}$$



Then

$$X \sim Pascal(m, p)$$

and by the linearity of expectation we have

$$E(X) = E(Y_1) + \cdots + E(Y_m) = m \cdot E(Y) = m \cdot p$$