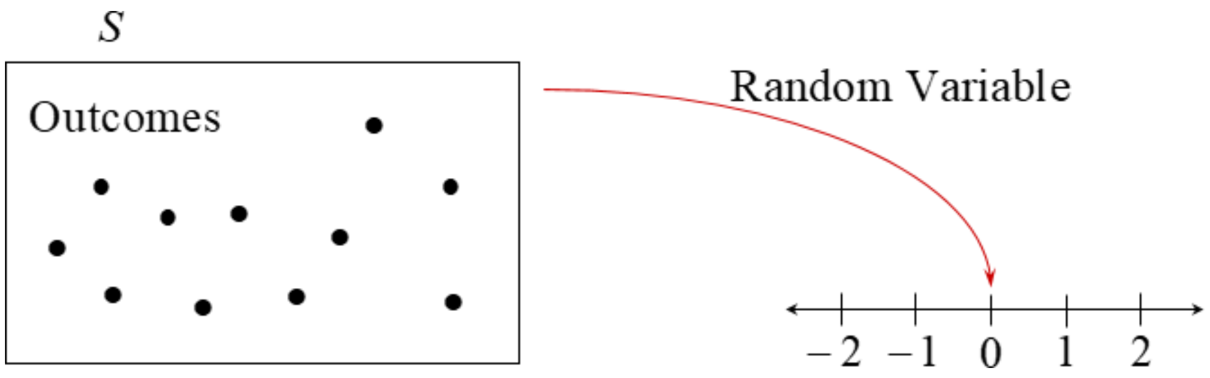
L04: Random Variables – Discrete

Yiming QIN
Assistant Professor
School of Energy and Environment
City University of Hong Kong

Random Variable

An event is defined as a subset of outcomes from the sample space. Indeed we are more interested in numerical values rather than the events themselves, which leads to the notion of a random variable (rv).



Definition: A random variable X: $S \rightarrow R$ is a numerical valued function defined on a sample space. That is, a number $X(a)$ is assigned to an outcome a in S.

$$P(X = x) = P(\{X = x\}) = P(\{a \in S |\, X(a) = x\})$$

Random Variable

Remarks

- Random variables are often denoted by capital letters, say *X, Y, Z*, and their possible numerical values (or called ***realizations***) denoted by the same lowercase letters, say x,y,z.
- Note that the random variable *X* is a random quantity **BEFORE** the experiment is performed and its realization x is the value of the random quantity **AFTER** the experiment has been performed. The word ***RANDOM*** reminds us of the fact that we cannot predict the outcomes of the experiment and consequently its associated numerical value beforehand.
- Please always keep in mind that a rv *X* is a function rather than a number. The value of *X* depends on an outcome. More rigorously, we would write the event $\{ X=x \}$ to represent $\{a\in S \mid X(a)=x\}$, the event $\{ X\geq x \}$ to represent $\{a\in S \mid X(a)\geq x\}$, etc.
- By the notion of rv, now we have a NEW interpretation of DATA:
- **DATA ARE THE ACTUAL VALUES (REALIZATIONS) OF** THE CORRESPONDING **RANDOM VARIABLE**.

Random Variables

Define **New Random Variables**

Let X, Y : S ➔ R be random variables, and a ϵ R. Define new random variables as follows:

Linear Transformation	1. $(X+Y)(s) = X(s) + Y(s)$ *for all* s *in* S.
Scalar Multiplication	2. $(aX)(s) = a(X(s))$ *for all* s *in* S.
Product of Random Variables	3. $(XY)(s) = (X(s))(Y(s))$ *for all* s *in* S.

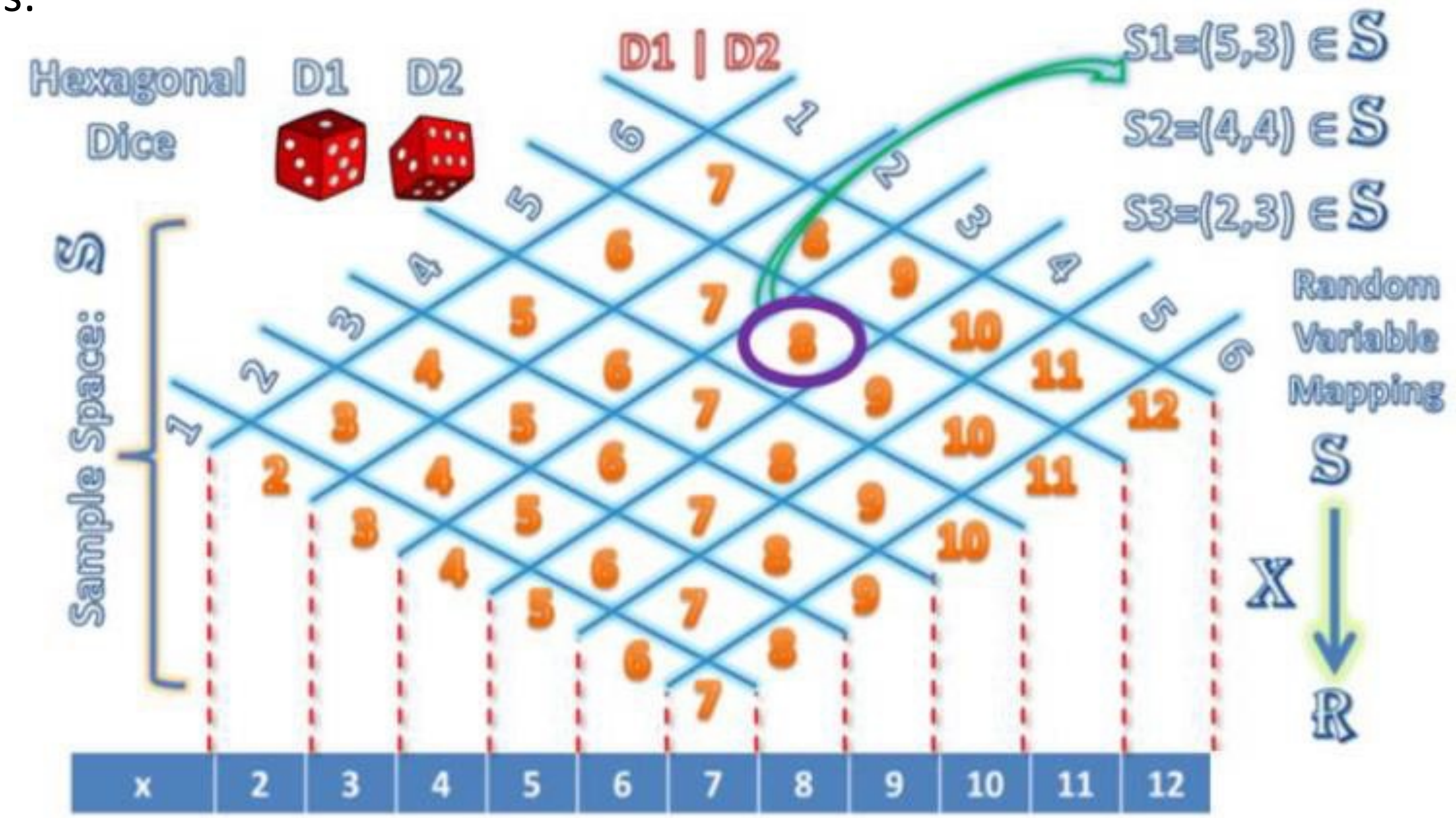
Examples:

In the experiment of rolling a die for twice. X, Y are the numbers appeared on the first and second roll respectively. *T* is the sum of the numbers appeared. Then, *X + Y = T.*

Example

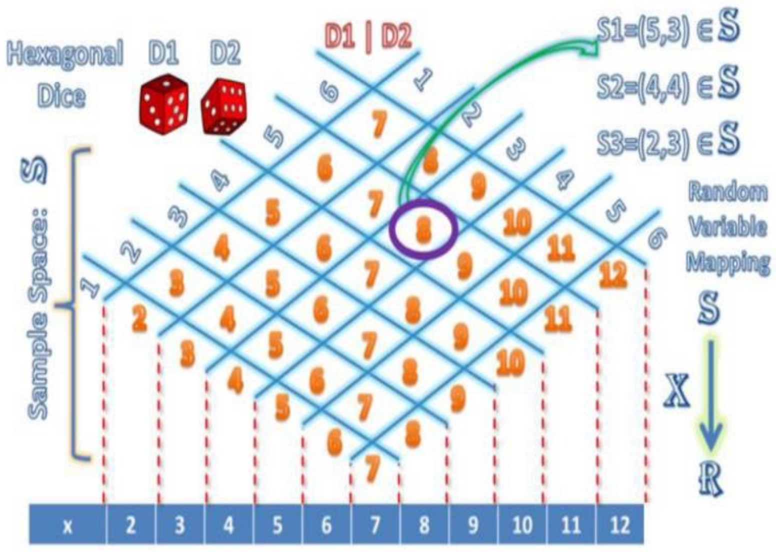
Rolling a **pair of** fair hexagonal dices

The sample space and the rv X mapping from the sample space (S) into the set (R) of all real numbers:



Example

Rolling a pair of fair hexagonal dices



The same value can be assigned to different elementary outcomes in S, e.g. $X(\{1,3\}) = X(\{2,2\}) = X(\{3,1\}) = 4$.

$$\boxed{P(X = x) = P(\{X = x\}) = P(\{a \in S \mid X(a) = x\})}$$

Thus, the probability of X=4 and X≤4:

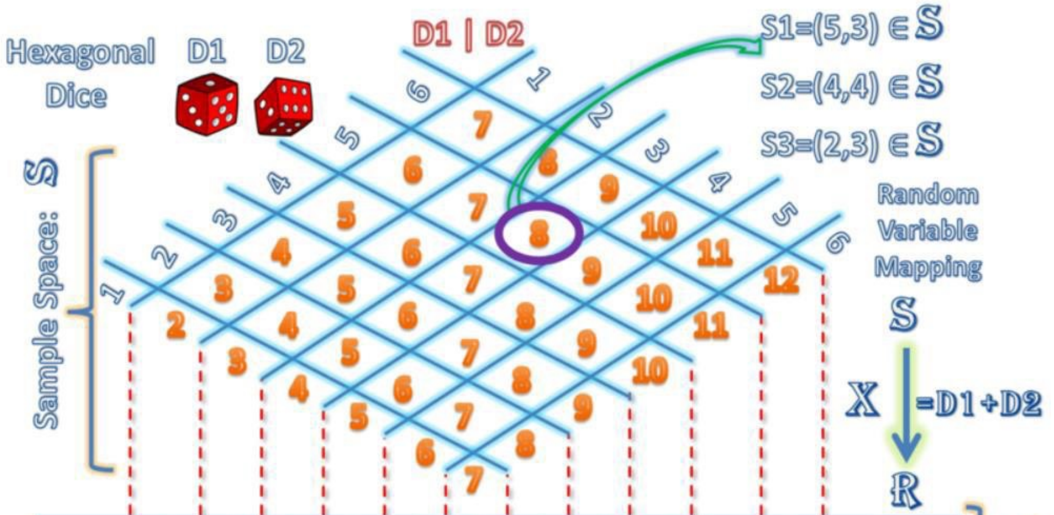
$P(X=4) = P(\{a \in S \mid X(a)=4\}) = P(\{1,3\}) + P(\{2,2\}) + P(\{3,1\}) = 3/36 = 1/12$

$P(X \leq 4) = P(\{a \in S \mid X(a) \leq 4\}) = P(\{a \in S \mid X(a)=2, 3 \text{ or } 4\} = P(\{1,1\},\{1,2\},\{2,1\},\{1,3\},\{2,2\},\{3,1\}) = 6/36 = 1/6$

Probability Distribution of a Random Variable

Note that a rv has its own probability law ➔ a rule that assigns probabilities to the different values of the r.v. Such a probability law of the probability assignment is often called a **probability distribution**.

In other words, **a (probability) distribution** of a rv is the collection of all values it can take on along with the probability of each value. So, a rv can be specified by its distribution.



x	2	3	4	5	6	7	8	9	10	11	12	R.V.
P(X=x)	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{3}{36}$	$\frac{4}{36}$	$\frac{5}{36}$	$\frac{6}{36}$	$\frac{5}{36}$	$\frac{4}{36}$	$\frac{3}{36}$	$\frac{2}{36}$	$\frac{1}{36}$	$\sum_{x=2}^{12} P(x) = 1$

This is a tabular form of showing the distribution of a (*discrete*) rv, where the *discrete* rv will be defined and discussed more later in this Chapter.

Probability Distributions of a Random Variable

Probability Distribution: A distribution of the probabilities associated with each of the values of a random variable. The probability distribution is a theoretical distribution; it is used to represent populations.

Note:

1. The probability distribution tells you everything you need to know about the random variable.
2. The probability distribution may be presented in the form of a table, chart, function, etc.

Probability Function: A rule that assigns probabilities to the values of the random variable.

Probability Distributions of a Random Variable

For the previous example, the following tabular form shows the probability of X:

x	2	3	4	5	6	7	8	9	10	11	12
P(X=x)	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{3}{36}$	$\frac{4}{36}$	$\frac{5}{36}$	$\frac{6}{36}$	$\frac{5}{36}$	$\frac{4}{36}$	$\frac{3}{36}$	$\frac{2}{36}$	$\frac{1}{36}$

Now that the distribution of X is known, we can find the probability of X directly and easily:

$$P(X \leq 4) = P(X = 2, 3 \; or \; 4) = P(X = 2) + P(X = 3) + P(X = 4)$$
$$= \frac{1}{36} + \frac{2}{36} + \frac{3}{36} = \frac{1}{6}$$

Definition: A cumulative distribution function $F(x)$ is defined as $F(x) = P(X < x)$ for all real values x.

The probability that X falls in an open and close interval: $P(a < X \leq b) = F(b) - F(a)$

Two Types of Random Variables

- **The range of a rv, denoted by χ, is the collection of all possible values it can take on.** For instance:
  - $\boldsymbol{X} \rightarrow \{\mathbf{0}, \mathbf{1}, \ldots, \boldsymbol{n}\}$; $\boldsymbol{Y} \rightarrow \{\mathbf{1}, \mathbf{2}, \mathbf{3}, \ldots,\}$; $\boldsymbol{Z} \rightarrow [\mathbf{0}, \infty]$.
  - We then can use the range of rv to classify it to be a Discrete rv, or Continuous rv.
  - Note that there exists a rv being both discrete and continuous (mixed rv), but we will not discuss it in this course.

➢ **Discrete Random Variable:**

It is a rv that has a finite or countable range (the number of defective items, the number of sales...).

➢ **Continuous Random Variable:**

It is a rv whose range is an interval over the real line (weight of an item, time until failure of a component, length of an object…).

Two Types of Random Variables

Discrete Random Variable: A quantitative random variable that can assume a countable number of values.

Intuitively, a discrete random variable can assume values corresponding to isolated points along a line interval. That is, there is a gap between any two values.

Note: Usually associated with counting.

Continuous Random Variable: A quantitative random variable that can assume an uncountable number of values.

Intuitively, a continuous random variable can assume any value along a line interval, including every possible value between any two values.

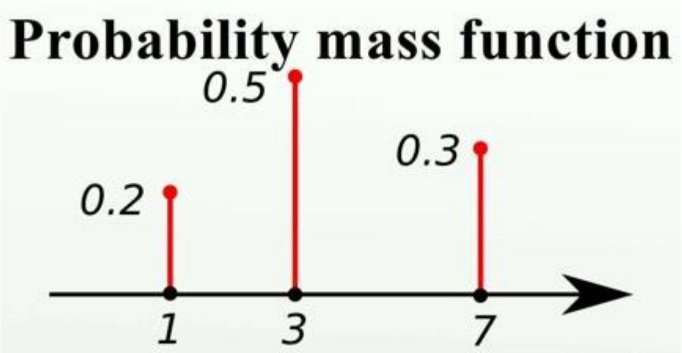
Note: Usually associated with a measurement.

Probability Mass Function for Discrete rv

➢ **Probability Mass Function and Distribution Function**

Probability Mass Function (PMF)

The probability MASS function of a DISCRETE rv X, denoted by $p(x)$, is a function that gives us the probability of occurrence for each possible value x of X. It is valid for all possible values x of X.



PMF

A discrete random variable can be characterized by a probability mass function (PMF)

$$p(x_i) = P\{X = x_i\}$$

✓ **<u>Conditions for a PMF:</u>**

- $0 < p(x) \leq 1$, for all x in the range of X.
- $\sum_{x \in \chi} p(x) = 1$

Question:
Is the function $p(x)$=$x/6$, for x in χ={1,2,3}, a valid PMF?

Example

A coin is tossed five times. Let the random variable x be the number of heads. The probability distribution is given in various forms below.

x	0	1	2	3	4	5
$P(x)$	1/32	5/32	10/32	10/32	5/32	1/32

The Probability Mass Function is

$$P(x) = \frac{{}_5C_x}{2^5}.$$

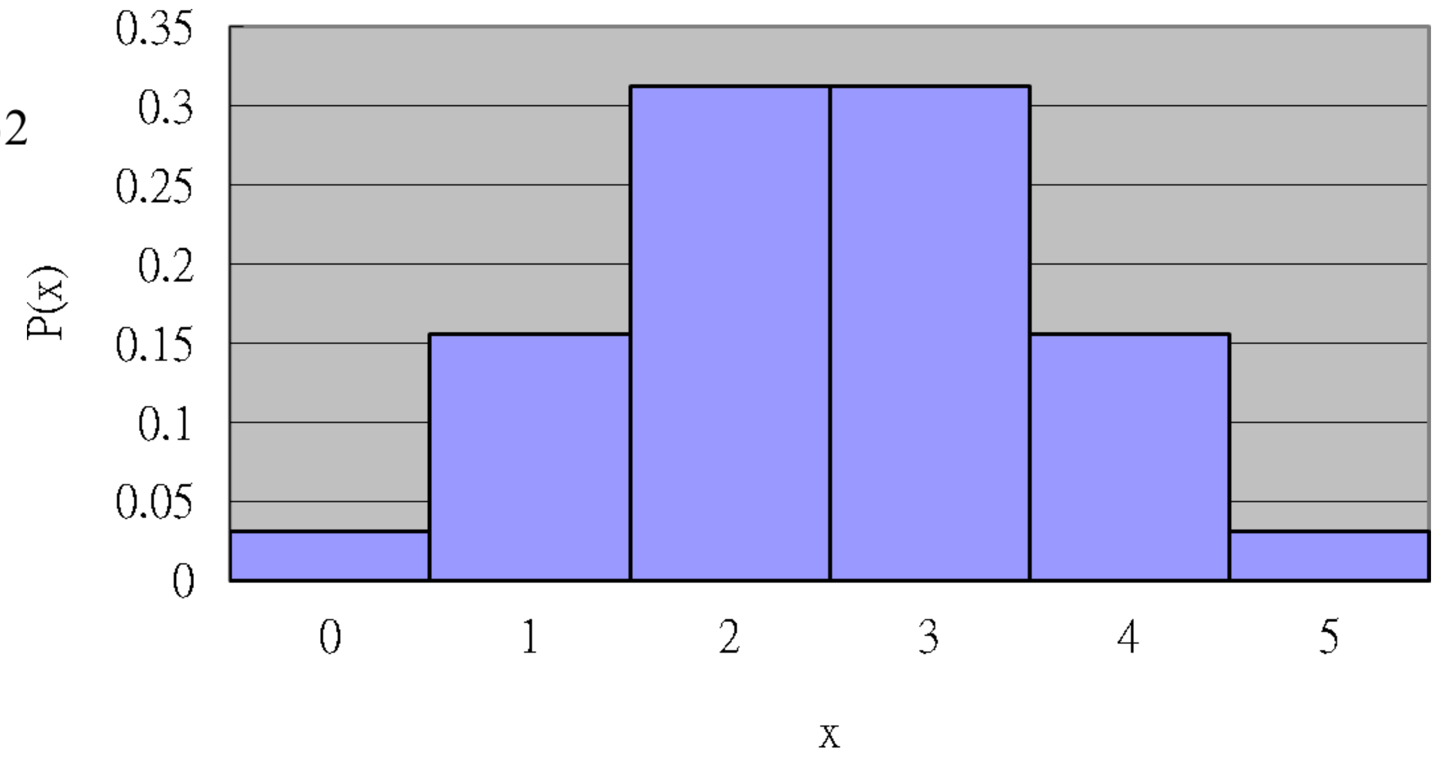
5 choose x

Example

A **histogram** may be used to present a probability distribution.

Probability distribution of number of heads:

x	0	1	2	3	4	5
$P(x)$	1/32	5/32	10/32	10/32	5/32	1/32



Cumulative Distribution Function

➤ **Probability Mass Function and Distribution Function**

Cumulative Distribution Function, (CDF)

- $F(a) = P(X \leq a) = \sum_{x \leq a} p(x)$, for all real values a.

Referring to the PMF in the previous question:

Question:
Is the function $p(x)$=$x/6$, for x in χ={1,2,3}, a valid pmf?

$$F(1) = P(X \leq 1) = P(X = 1) = p(1) = 1/6$$

$$F(1.5) = P(X \leq 1.5) = P(X = 1) = 1/6$$

$$F(1.26) = P(X \leq 1.26) = P(X = 1) = 1/6$$

$$F(2) = P(X \leq 2) = P(X = 1) + P(X = 2) = p(1) + p(2) = 1/2$$

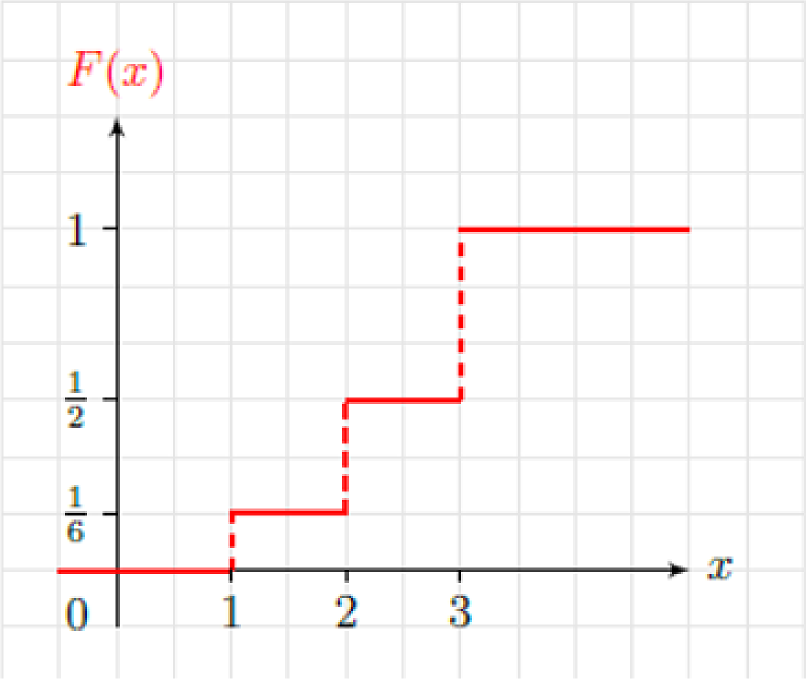
$$F(3) = P(X \leq 3) = p(1) + p(2) + p(3) = 1$$

As can be seen above, the cdf of a discrete random variable would be ***a step-function*** with the size $p(a)$ of the jumps at the possible value a.

If the range of the discrete rv X is expressed by $\{x_1, x_2, x_3, \ldots\}$, then we have

$$p(x_1) = F(x_1) \ and\ p(x_j) = F(x_j) - F(x_{j-1})$$

where j=2,3,....



Example

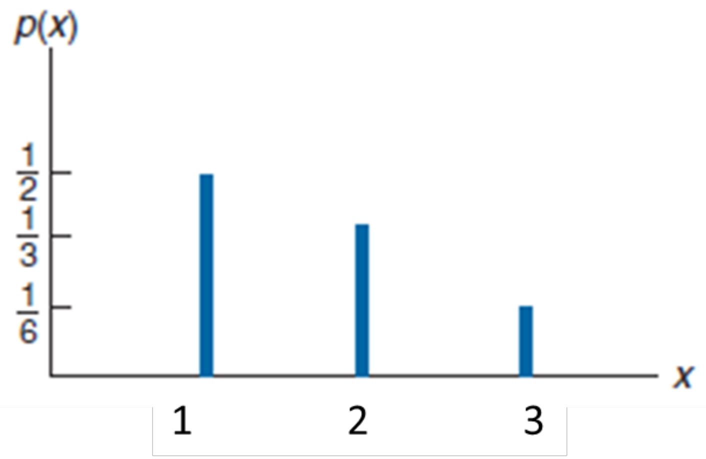
Suppose X has a PMF

$$p(1) = \frac{1}{2}, \quad p(2) = \frac{1}{3}, \quad p(3) = \frac{1}{6}$$

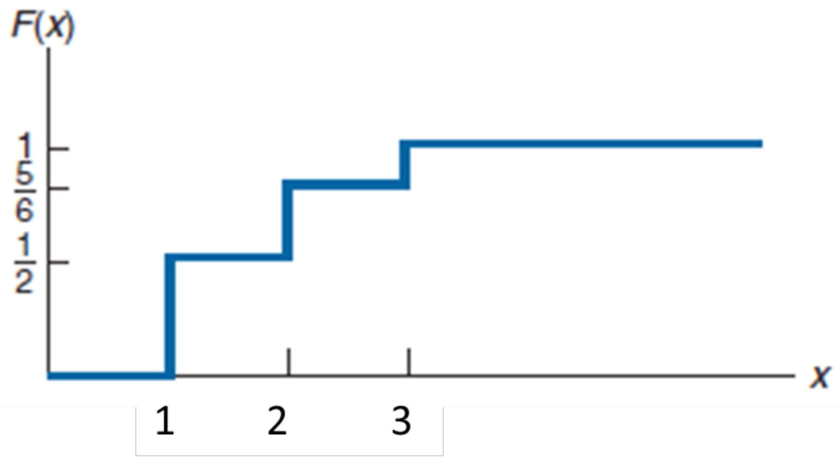
then the CDF is

$$F(a) = \begin{cases} 0 & a < 1 \\ 1/2 & 1 \le a < 2 \\ 5/6 & 2 \le a < 3 \\ 1 & 3 \le a \end{cases}$$

Probability Distribution Function



Probability Cumulative Distribution Function



Example

In the experiment that a fair die is rolled, compute the probability mass function and cumulative distribution function for the number appeared on the die.

Solution:

Let X be the number appeared on the die. f is the probability mass function of X. Then, $f(1) = P(X = 1) = 1/6$. Similarly, $f(1) = f(2) = f(3) = f(4) = f(5) = f(6) = \frac{1}{6}$ and $f(x) = 0$ otherwise.

If F is the cumulative distribution function of X,

$$F(x) = \begin{cases} 0 & x < 1 \\ \frac{1}{6} & 1 \leq x < 2 \\ \frac{2}{6} & 2 \leq x < 3 \\ \frac{3}{6} & 3 \leq x < 4 \\ \frac{4}{6} & 4 \leq x < 5 \\ \frac{5}{6} & 5 \leq x < 6 \\ 1 & 6 \leq x \end{cases}$$

Expectation

- The expectation (expected value, mean) of a random variable X is denoted by $E[X]$.
- In the discrete case, where X takes on the possible values
  $x_1, x_2, \ldots, x_n$ with probability $p(x_1), p(x_2), \ldots\ldots, p(x_n)$
- The expectation is the **weighted average** of all possible values.

EXPECTATION

- *The expectation of a* $\boxed{\textit{discrete random variable}}$ *is defined as*

$$E[X] = \sum_i x_i p(x_i)$$

Example

The expectation of the outcome of rolling a die.

Since

$$p(1) = p(2) = \cdots = p(6) = \frac{1}{6}$$

The expectation is

$$1 \times \frac{1}{6} + 2 \times \frac{1}{6} + \cdots + 6 \times \frac{1}{6} = \frac{7}{2}$$

Note:
the expectation should not be interpreted as the value that we expect *X* to be, but rather as **the average of *X* in a large number of repetitions of the experiment**
Suppose we roll the dice 6000 times, we may get roughly 1000 1's, 1000 2's, . . .

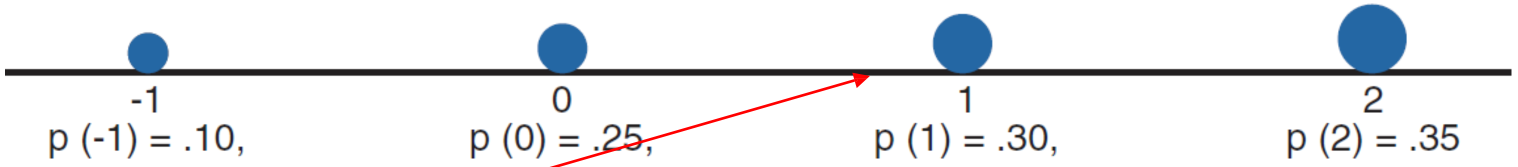
Expectation

Physical explanation of the mean (or expectation):

Consider a random variable that takes value -1; 0; 1 and 2 with probability 0.10; 0.25; 0.30 and 0.35, respectively.

The expectation is:

$$E[X] = -1 \times .10 + 0 \times .25 + 1 \times .30 + 2 \times .35 = .9$$



Centre of gravity = 0.9

Example

$$E[X] = \sum_i x_i p(x_i)$$

Suppose X has the following PMF,
p(0) = 0.2; p(1) = 0.5; p(2) = 0.3
What is E[X^2]?

Let $Y = X^2$, we can compute the PMF of Y

$$p_Y(0) = P\{Y = 0^2\} = P\{X = 0\} = .2$$
$$p_Y(1) = P\{Y = 1^2\} = P\{X = 1\} = .5$$
$$p_Y(4) = P\{Y = 2^2\} = P\{X = 2\} = .3$$

So,

$$E[Y] = 0^2 \times .2 + 1^2 \times .5 + 2^2 \times .3 = 1.7$$

CityU

Expectation

- **A general question: If we know the PMF (or CDF) of *X*, how can we compute the expectation of g(X)?**

✓ Law of the Unconscious Statistician (LOTUS):

If X is discrete:

$$E[g(X)] = \sum_i g(x_i)p(x_i)$$

The n^{th} moment of random variable X is defined as:

$$E[X^n] = \sum_i x_i^n p(x_i)$$

Linearity of Expectation

The expectation of linear combination equals to the linear combination of the expectation

LINEARITY

For any constants a, b

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

LINEARITY – MULTI

For any constants $a_1, a_2, \ldots, a_K$ and b

$$E[a_1X_1 + a_2X_2 + \cdots a_KX_K + b] = a_1E[X_1] + a_2E[X_2] + \cdots + a_KE[X_K] + b$$

This linearity property applies regardless of whether the random variables X and Y are independent or dependent.

Independence of Random Variables

INDEPENDENCE

Random variables X and Y are independent means for all x, y

$$P(X \leq x, Y \leq y) = P(X \leq x)P(Y \leq y)$$

In the discrete case, this implies

INDEPENDENCE - DISCRETE

- *If X and Y are discrete, independence means for all x_i and y_j*

$$P(X = x_i, Y = y_j) = P(X = x_i)P(Y = y_j)$$

CityU

Expectation of Product given Independence

If X and Y are independent, then

$$E[XY] = E[X]E[Y]$$

Example

(a) What is the expected value of the sum of two fair dices?

Sum of two fair dices

	1	2	3	4	5	6
1	2	3	4	5	6	7
2	3	4	5	6	7	8
3	4	5	6	7	8	9
4	5	6	7	8	9	10
5	6	7	8	9	10	11
6	7	8	9	10	11	12

(b) What is the expected value of the product of two fair dices?

Product of two fair dice

	1	2	3	4	5	6
1	1	2	3	4	5	6
2	2	4	6	8	10	12
3	3	6	9	12	15	18
4	4	8	12	16	20	24
5	5	10	15	20	25	30
6	6	12	18	24	30	36

Example

(a) What is the expected value of the sum of two fair dices?

Solution:

Expected value of one fair dice: $E(X) = \frac{1}{6} \times 1 + \frac{1}{6} \times 2 + \frac{1}{6} \times 3 + \frac{1}{6} \times 4 + \frac{1}{6} \times 5 + \frac{1}{6} \times 6 = 3.5$

Expected value of the sum of two fair dices: $E(X + X) = E(X) + E(X) = 7$

Sum of two fair dices

	1	2	3	4	5	6
1	2	3	4	5	6	7
2	3	4	5	6	7	8
3	4	5	6	7	8	9
4	5	6	7	8	9	10
5	6	7	8	9	10	11
6	7	8	9	10	11	12

Example

(a) What is the expected value of the sum of two fair dices?

Solution:

Expected value of one fair dice: $E(X) = \frac{1}{6} \times 1 + \frac{1}{6} \times 2 + \frac{1}{6} \times 3 + \frac{1}{6} \times 4 + \frac{1}{6} \times 5 + \frac{1}{6} \times 6 = 3.5$

Expected value of the sum of two fair dices: $E(X+X) = E(X) + E(X) = 7$

Alternative:

Expected value of the sum of two fair dices: $E(X+X) \boxed{= \frac{1}{36} \times 2 + \frac{2}{36} \times 3 + \frac{3}{36} \times 4 + \cdots + \frac{1}{36} \times 12} = 7$

Sum of two fair dices

	1	2	3	4	5	6
1	2	3	4	5	6	7
2	3	4	5	6	7	8
3	4	5	6	7	8	9
4	5	6	7	8	9	10
5	6	7	8	9	10	11
6	7	8	9	10	11	12

Example

(a) What is the expected value of the sum of two fair dices?

Solution:

Expected value of one fair dice: $E(X) = \frac{1}{6} \times 1 + \frac{1}{6} \times 2 + \frac{1}{6} \times 3 + \frac{1}{6} \times 4 + \frac{1}{6} \times 5 + \frac{1}{6} \times 6 = 3.5$

Expected value of the sum of two fair dices: $E(X + X) = E(X) + E(X) = 7$

Sum of two fair dices

	1	2	3	4	5	6
1	2	3	4	5	6	7
2	3	4	5	6	7	8
3	4	5	6	7	8	9
4	5	6	7	8	9	10
5	6	7	8	9	10	11
6	7	8	9	10	11	12

Alternative:

Expected value of the sum of two fair dices: $E(X + X) \boxed{= \frac{1}{36} \times 2 + \frac{2}{36} \times 3 + \frac{3}{36} \times 4 + \cdots + \frac{1}{36} \times 12} = 7$

(b) What is the expected value of the product of two fair dices?

Solution:

Expected value of the product of two fair dices: $E(X \times X) = E(X) \times E(X) = 12.25$

Product of two fair dice

	1	2	3	4	5	6
1	1	2	3	4	5	6
2	2	4	6	8	10	12
3	3	6	9	12	15	18
4	4	8	12	16	20	24
5	5	10	15	20	25	30
6	6	12	18	24	30	36

Example

(a) What is the expected value of the sum of two fair dices?

Solution:

Expected value of one fair dice: $E(X) = \frac{1}{6} \times 1 + \frac{1}{6} \times 2 + \frac{1}{6} \times 3 + \frac{1}{6} \times 4 + \frac{1}{6} \times 5 + \frac{1}{6} \times 6 = 3.5$

Expected value of the sum of two fair dices: $E(X + X) = E(X) + E(X) = 7$

Alternative:

Expected value of the sum of two fair dices: $E(X + X) \boxed{= \frac{1}{36} \times 2 + \frac{2}{36} \times 3 + \frac{3}{36} \times 4 + \cdots + \frac{1}{36} \times 12} = 7$

Sum of two fair dices

	1	2	3	4	5	6
1	2	3	4	5	6	7
2	3	4	5	6	7	8
3	4	5	6	7	8	9
4	5	6	7	8	9	10
5	6	7	8	9	10	11
6	7	8	9	10	11	12

(b) What is the expected value of the product of two fair dices?

Solution:

Expected value of the product of two fair dices: $E(X \times X) = E(X) \times E(X) = 12.25$

Alternative:

Expected value of the product of two fair dices: $E(X \times X)$

$= \boxed{\frac{1}{36} \times 1 + \frac{2}{36} \times 2 + \frac{2}{36} \times 3 + \cdots + \frac{1}{36} \times 36} = 12.25$

Product of two fair dice

	1	2	3	4	5	6
1	1	2	3	4	5	6
2	2	4	6	8	10	12
3	3	6	9	12	15	18
4	4	8	12	16	20	24
5	5	10	15	20	25	30
6	6	12	18	24	30	36

Variance

If *X* is a random variable with mean , the variance of *X*, denoted by Var(*X*), is defined by

$$\boxed{Var(X) = E[(X-\mu)^2]}$$

Variance

If X is a random variable with mean , the variance of X, denoted by Var(X), is defined by

$$\boxed{Var(X) = E[(X-\mu)^2]}$$

Note that

$$\begin{aligned} E[(X-\mu)^2] &= E[X^2 - 2\mu X + \mu^2] \\ &= E[X^2] - 2\mu E[X] + \mu^2 \\ &= E[X^2] - 2\mu\mu + \mu^2 \\ &= E[X^2] - \mu^2 \end{aligned}$$

Variance

If X is a random variable with mean , the variance of X, denoted by Var(X), is defined by

$$\boxed{Var(X) = E[(X-\mu)^2]}$$

Note that

$$\begin{aligned} E[(X-\mu)^2] &= E[X^2 - 2\mu X + \mu^2] \\ &= E[X^2] - 2\mu E[X] + \mu^2 \\ &= E[X^2] - 2\mu\mu + \mu^2 \\ &= E[X^2] - \mu^2 \end{aligned}$$

So, alternatively,

$$\boxed{Var(X) = E[X^2] - (E[X])^2}$$

You may choose the one convenient for computation.

Example

Compute Var(X) when X represents the outcome of rolling a die:

Example

Compute Var(X) when X represents the outcome of rolling a die:

Since: $Var(X) = E[X^2] - (E[X])^2$ *and*:

$$\mathrm{E}(\mathrm{X}^2) = 1^2 \times \frac{1}{6} + 2^2 \times \frac{1}{6} + 3^2 \times \frac{1}{6} + 4^2 \times \frac{1}{6} + 5^2 \times \frac{1}{6} + 6^2 \times \frac{1}{6} = \frac{91}{6}$$

Example

Compute Var(X) when X represents the outcome of rolling a die:

Since: $Var(X) = E[X^2] - (E[X])^2$ *and*:

$$\mathrm{E}(\mathrm{X}^2) = 1^2 \times \frac{1}{6} + 2^2 \times \frac{1}{6} + 3^2 \times \frac{1}{6} + 4^2 \times \frac{1}{6} + 5^2 \times \frac{1}{6} + 6^2 \times \frac{1}{6} = \frac{91}{6}$$

And having computed $E(X) = \frac{7}{2}$ in a previous slide: $\leftarrow$ $1 \times \frac{1}{6} + 2 \times \frac{1}{6} + \cdots + 6 \times \frac{1}{6} = \frac{7}{2}$

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

Property of Variance

Linearity with constant

For any constants b

$$\boxed{Var\ (X + b) = Var(X)}$$

Property of Variance

Linearity with constant

For any constants b

$$\boxed{\mathrm{Var}(X+b) = \mathrm{Var}(X)}$$

Given: $\mathrm{Var}(X) = E[(X-\mu)^2]$

Now, let's *Y=X+b, the mean of Y is* $\mu_Y = E[Y] = E[X+b] = E[X] + E[b] = \mu + b$

The variance of Y is then:

$$\mathrm{Var}(Y) = E[(Y-\mu_Y)^2]$$

Substituting $Y = X + b$ and $\mu_Y = \mu + b$, we get:

$$\mathrm{Var}(Y) = E[((X+b)-(\mu+b))^2]$$

Notice that b cancels out:

$$\mathrm{Var}(Y) = E[(X-\mu)^2]$$

Which is just:

$$\mathrm{Var}(Y) = \mathrm{Var}(X)$$

Property of Variance

Scalar with constant

For any constants a

$$\boxed{Var\ (aX) = a^2 Var(X)}$$

Property of Variance

Scalar with constant

For any constants a

$$\boxed{\text{Var }(aX) = a^2\text{Var}(X)}$$

The mean of Y is:

$$\mu_Y = E[Y] = E[aX] = aE[X] = a\mu$$

The variance of Y is then:

$$\text{Var}(Y) = E[(Y - \mu_Y)^2]$$

Substituting $Y = aX$ and $\mu_Y = a\mu$, we get:

$$\text{Var}(Y) = E[((aX) - (a\mu))^2]$$

Now we can factor out the constant a (since a^2 is a constant):

$$\text{Var}(Y) = E[a^2(X - \mu)^2] = a^2E[(X - \mu)^2]$$

Since $E[(X - \mu)^2]$ is the variance of X:

$$\text{Var}(Y) = a^2\text{Var}(X)$$

Property of Variance

For any constants b

$$\boxed{\text{Var }(X+b) = \text{Var}(X)}$$

For any constants a

$$\boxed{\text{Var }(aX) = a^2\text{Var}(X)}$$

For any constants a and b,

$$\boxed{Var(aX + b) = a^2 Var(X)}$$

Property of Variance

If X and Y are $\boxed{\text{independent}}$, then

$$Var(X+Y) = Var(X) + Var(Y)$$

However, if X and Y are not independent, you must also consider the covariance

Example

The number of standby passengers who get seats on a daily commuter flight from Boston to New York is a random variable, *x*, with the probability distribution given below (in extension table).

Find the mean, variance and standard deviation.

$$\mu = E[X] = xp(x)$$

$$\sigma^2 = Var(X) = E[X^2] - (\mathrm{E}[\mathrm{X}])^2 = x^2p(x) - [xp(x)]^2$$

	x	$P(x)$	$xP(x)$	x^2	$x^2P(x)$
	0	0.30	0.00	0	0.00
	1	0.25	0.25	1	0.25
	2	0.20	0.40	4	0.80
	3	0.15	0.45	9	1.35
	4	0.05	0.20	16	0.80
	5	0.05	0.25	25	1.25
Totals		1.00	1.55		4.45
		$\sum P(x)$ (check)	$\sum[xP(x)]$		$\sum[x^2P(x)]$

Example

The number of standby passengers who get seats on a daily commuter flight from Boston to New York is a random variable, *x*, with the probability distribution given below (in extension table).

Find the mean, variance and standard deviation.

$$\mu = E[X] = xp(x)$$

$$\sigma^2 = Var(X) = E[X^2] - (\mathrm{E}[X])^2 = x^2p(x) - [xp(x)]^2$$

x	$P(x)$	$xP(x)$	x^2	$x^2P(x)$
0	0.30	0.00	0	0.00
1	0.25	0.25	1	0.25
2	0.20	0.40	4	0.80
3	0.15	0.45	9	1.35
4	0.05	0.20	16	0.80
5	0.05	0.25	25	1.25
Totals	1.00	1.55		4.45
	$\sum P(x)$ (check)	$\sum[xP(x)]$		$\sum[x^2P(x)]$

Solution:

Using the formulas for mean, variance, and standard deviation:

$$\mu = \sum[xP(x)] = 1.55$$

Note: 1.55 is *not* a value of the random variable (in this case). It is only what happens on average.

$$\sigma^2 = \sum[x^2P(x)] - \left\{\sum[xP(x)]\right\}^2$$

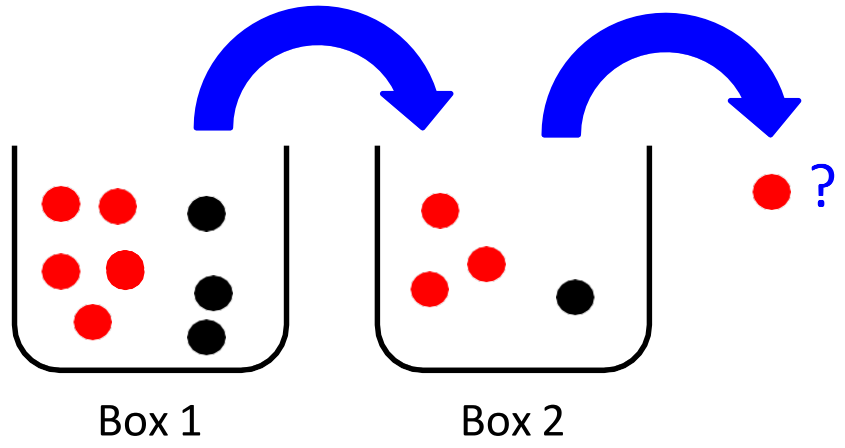
$$= 4.45 - (1.55)^2 = 4.45 - 2.4025 = 2.0475$$

$$\sigma = \sqrt{\sigma^2} = \sqrt{2.0475} \approx 1.43$$

Example

Box 1 contains 5 red balls and 3 black balls. Box 2 contains 3 red balls and 1 black ball. A ball is selected at random from Box 1 and is put into Box 2. Then, a ball is selected at random from Box 2. Let x be the number of red balls selected from the boxes.

1. Find the probability distribution for x.
2. What are the mean and standard deviation of x?
3. Find the probability that the ball selected from Box 1 is red, given that the ball selected from Box 2 is black.



Binomial Probability Distribution

Binomial Probability Distribution

Binomial Probability Experiment:

An experiment that is made up of repeated trials that possess the following properties:

1. There are n repeated independent trials.
2. Each **trial** has two possible outcomes (success, failure).
3. $P(\text{success}) = p$, $P(\text{failure}) = q$, and $p + q = 1$
4. The **binomial random variable** x is the count of the number of successful trials that occur; x may take on any integer value from zero to n.

Binomial Probability Distribution

A **Binomial distribution** is related to a random experiment with the following features:

- Fixed **finite** number of **identical** trials, say $n < \infty$.
- Trials are **independent**.
- Trials result in two possible outcomes denoted by **success and failure**.
- The probability of success ***p*** **is constant** across trials.

Here are some typical examples for Binomial distributions

Trial	Success (the outcome of our interest)	Failure
Tossing a coin	Head	Tail
Pure guess in M.C.	Correct	Wrong
Randomly select a product	Non-defective	Defective

Binomial Probability Distribution

If X is the discrete rv of the number of successes in n trials, then we can use a **Binomial distribution** to characterize its random behavior.

More details about the Binomial distribution

- Notation:

$$\boxed{X \sim Binomial(n, p)}$$

- pmf:

$$\boxed{\begin{array}{l} p(x) = P(X = x) = \binom{n}{x} p^x (1-p)^{n-x}, for\ x = 0, 1, \dots, n; \\ p(x) = 0, otherwise. \end{array}}$$

- Population mean and population variance:

$$\boxed{E(X) = np,\ Var(X) = np(1-p).}$$

Binomial Probability Distribution

REMARK:

The notation $\binom{n}{x}$ (or C_x^n pronounced n C x) is defined as

$$\binom{n}{x} = \frac{n!}{x!\,(n-x)!}\ ,$$

where the notation $k!$ (pronounced k-*factorial*) is defined as, if k is a positive integer,

$$k! = k(k-1)(k-2)\cdots(2)(1)\ and\ 0! = 1.$$

For instance, $5! = 5 \times 4 \times 3 \times 2 \times 1 = 120 \quad and \quad 3! = 3 \times 2 \times 1 = 6.$

In the binomial pmf, $\binom{n}{x}$ means the total number of ways x success occur in *n* trials.

Binomial Probability Function

For a binomial experiment, let p represent the probability of a "success" and q represent the probability of a "failure" on a single trial; then $P(x)$, the probability that there will be exactly x successes on n trials is

$$P(x) = \binom{n}{x}(p^x)(q^{n-x}), \quad \text{for } x = 0,\ 1,\ 2,\ \ldots,\ \text{or } n$$

Note:

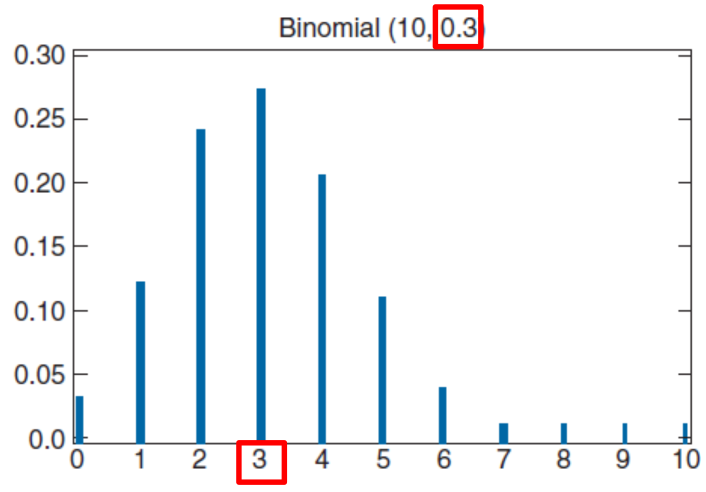
1. The number of ways that exactly x successes can occur in n trials: $\binom{n}{x}$
2. The probability of exactly x successes: p^x
3. The probability that failure will occur on the remaining $(n - x)$ trials: q^{n-x}

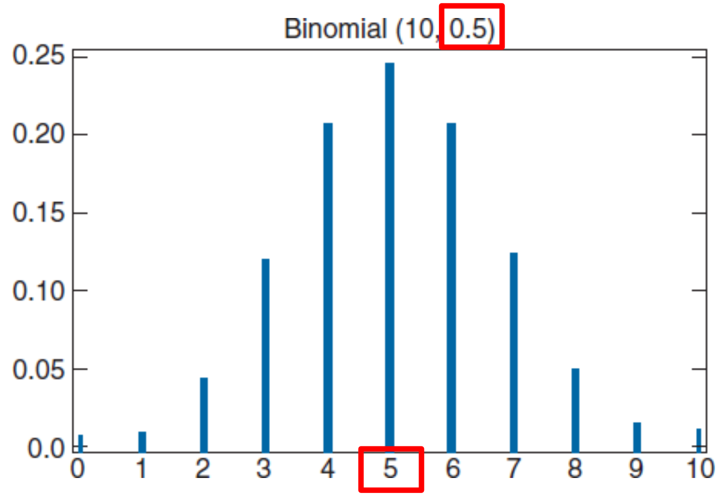
Binomial Random Variable

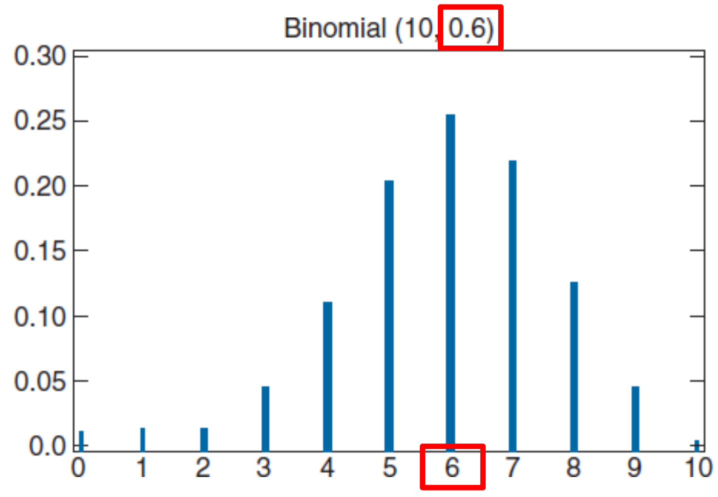
There are 10 independent trials

The probability of success on each trial

Graphic View







Skewed with long tail at right side if p<0.5

Symmetry: When p=0.5, with symmetric around its mean np

Skewed with long tail at left side if p>0.5

Binomial Random Variable

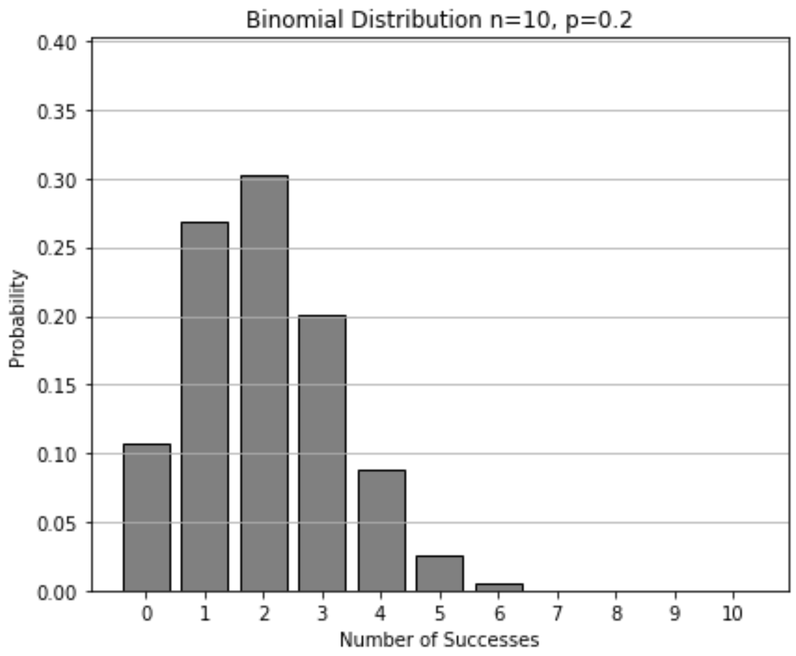
Graphic View

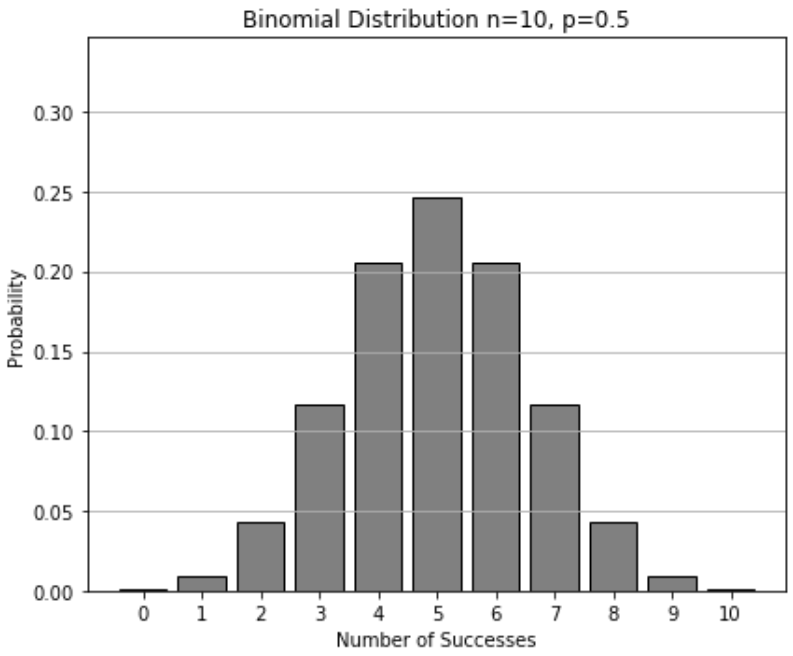
There are 10 (small number of trials) independent trials

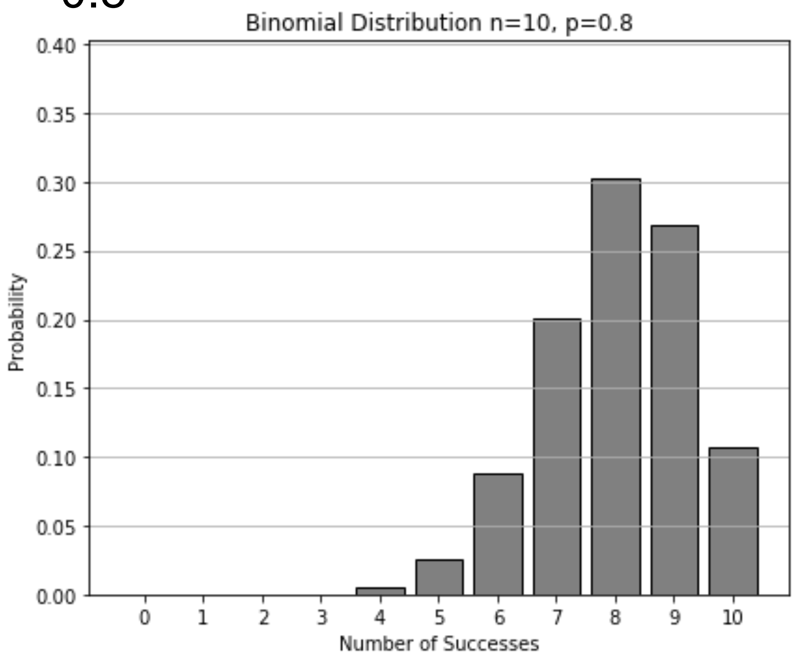
The probability of success = 0.2

The probability of success = 0.5

The probability of success = 0.8







Skewed with long tail at right side if p<0.5

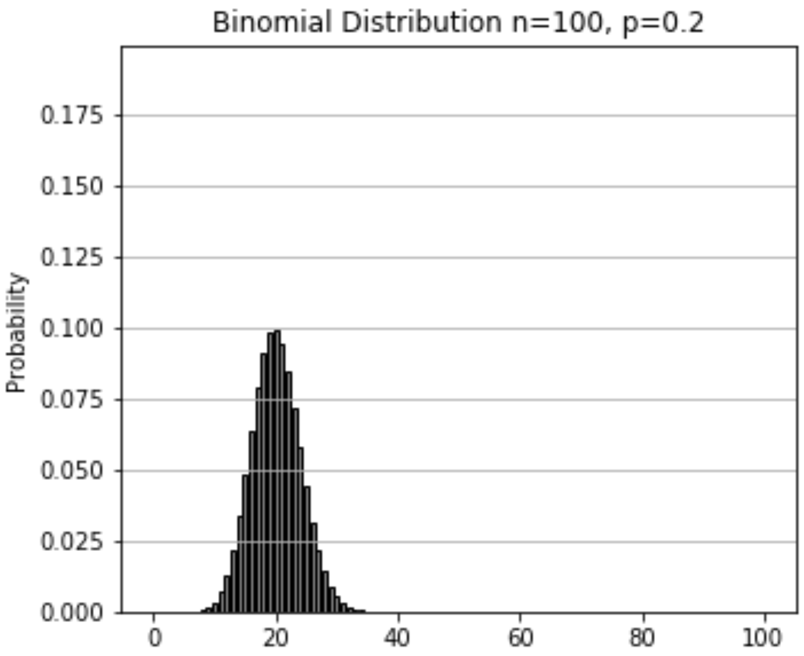
Symmetry: When p=0.5, with symmetric around its mean np

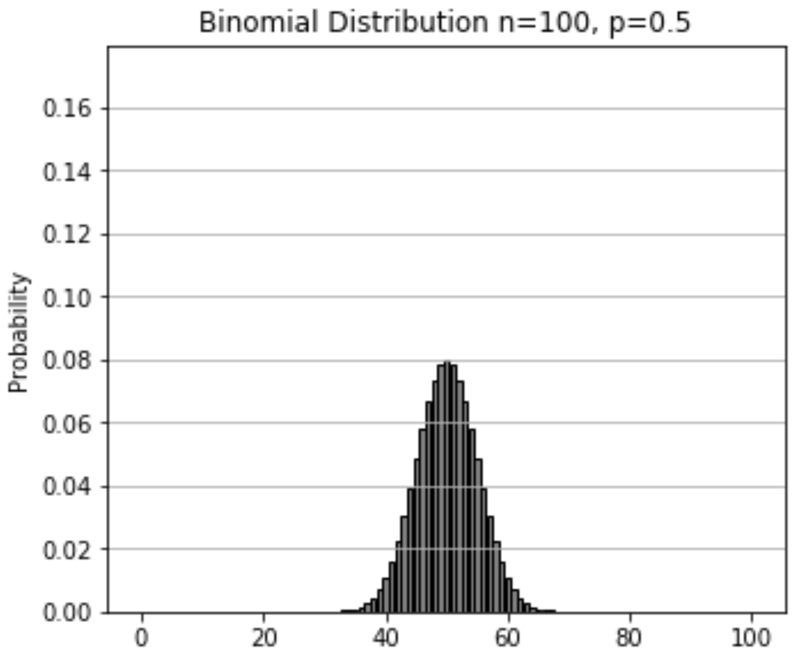
Skewed with long tail at left side if p>0.5

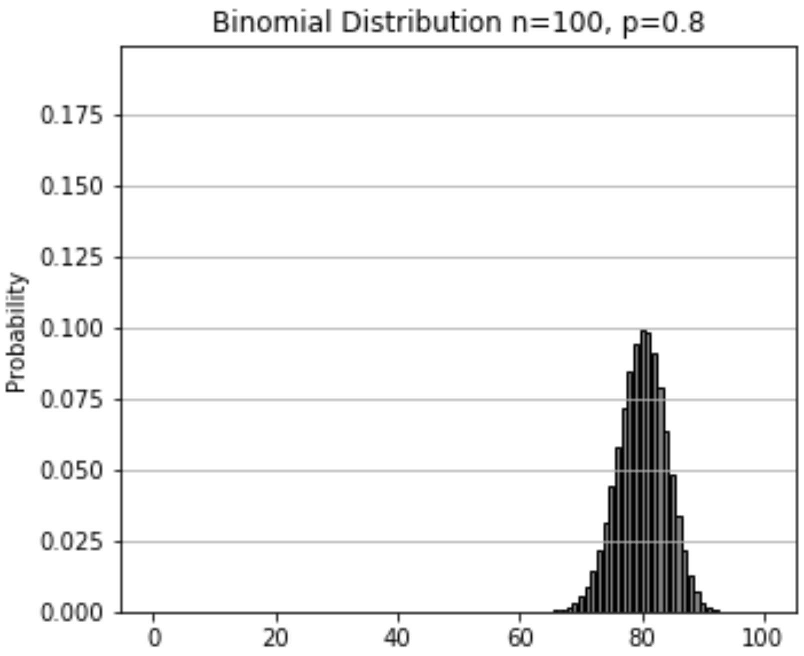
Binomial Random Variable

Graphic View

There are 100 trials







Law of Large Numbers: With more trials, the actual ratio of successes to total trials is more likely to be close to p, making extreme results (like getting all successes or all failures) less likely.

Central Limit Theorem: For large n, the distribution of the sample mean will approach a normal distribution, regardless of the shape of the original distribution, as long as the original distribution has a finite variance.

Example

It is known that 40% of all graduating seniors on campus have taken a statistics class. Five seniors are selected at random and asked if they have taken a statistics class.

Example

It is known that 40% of all graduating seniors on campus have taken a statistics class. Five seniors are selected at random and asked if they have taken a statistics class.

1. A trial is asking one student, repeated 5 times. The trials are independent since the probability of taking a statistics class for any one student is not affected by the results from any other student.

Example

It is known that 40% of all graduating seniors on campus have taken a statistics class. Five seniors are selected at random and asked if they have taken a statistics class.

1. A trial is asking one student, repeated 5 times. The trials are independent since the probability of taking a statistics class for any one student is not affected by the results from any other student.
2. Two outcomes on each trial:
   taken a statistics class (success),
   not taken a statistics class (failure)
3. $p = P(\text{taken a statistics class}) = 0.4$
   $q = P(\text{not taken a statistics class}) = 0.6$

Example

It is known that 40% of all graduating seniors on campus have taken a statistics class. Five seniors are selected at random and asked if they have taken a statistics class.

1. A trial is asking one student, repeated 5 times. The trials are independent since the probability of taking a statistics class for any one student is not affected by the results from any other student.
2. Two outcomes on each trial:
   taken a statistics class (success),
   not taken a statistics class (failure)
3. *p* = *P*(taken a statistics class) = 0.4
   q = *P*(not taken a statistics class) = 0.6
4. *x* = number of students who have taken a statistics class

$$P(X = k) = \binom{n}{k} p^k q^{n-k} = \binom{5}{k} (0.4)^k (0.6)^{n-k}$$

$X \sim Binomial\,(n=5, p=0.4)$.

Example

Three students are asked to randomly pick one number between 0 and 9 inclusively. Let X be the rv of the number of students who pick the number “7”.

Example

Three students are asked to randomly pick one number between 0 and 9 inclusively. Let *X* be the rv of the number of students who pick the number "7".

n = 3 students are asked to randomly pick a number in the same range.

- **3 identical trails, Trials are independent:**

Example

Three students are asked to randomly pick one number between 0 and 9 inclusively. Let X be the rv of the number of students who pick the number "7".

n = 3 students are asked to randomly pick a number in the same range.

- **3 identical trails, Trials are independent:**

A student's choice does not affect the choice of another student.

- **Trials result in two possible outcomes denoted by success/failure:**

Success: the number 7 is picked; Failure: another number (not 7) is picked.

- **The probability of success *p* is constant across trials:**

Example

Three students are asked to randomly pick one number between 0 and 9 inclusively. Let *X* be the rv of the number of students who pick the number "7".

n = 3 students are asked to randomly pick a number in the same range.

- **3 identical trails, Trials are independent:**

A student's choice does not affect the choice of another student.

- **Trials result in two possible outcomes denoted by success/failure:**

Success: the number 7 is picked; Failure: another number (not 7) is picked.

- **The probability of success *p* is constant across trials:**

If randomly choose, a student has a probability *p* = 1/10 of picking "7".

Thus, $X \sim Binomial\,(n=3, p=0.1)$.

$$P(X=k) = \binom{n}{k} p^k q^{n-k} = \binom{3}{k} \left(\frac{1}{10}\right)^k \left(\frac{9}{10}\right)^{n-k}$$

Example

$$P(X=k) = \binom{n}{k} p^k q^{n-k} = \binom{3}{k} (\frac{1}{10})^k (\frac{9}{10})^{n-k}$$

- When $k=0$: No student picks "7".

$P(X=0) = \binom{3}{0}\left(\frac{1}{10}\right)^0\left(\frac{9}{10}\right)^3 = 1 \times 1 \times \left(\frac{9}{10}\right)^3$

$P(X=0) = \left(\frac{9}{10}\right)^3$

$P(X=0) = \frac{729}{1000}$

- When $k=1$: One student picks "7".

$P(X=1) = \binom{3}{1}\left(\frac{1}{10}\right)^1\left(\frac{9}{10}\right)^2 = 3 \times \frac{1}{10} \times \left(\frac{9}{10}\right)^2$

$P(X=1) = 3 \times \frac{1}{10} \times \frac{81}{100}$

$P(X=1) = \frac{243}{1000}$

- When $k=2$: Two students pick "7".

$P(X=2) = \binom{3}{2}\left(\frac{1}{10}\right)^2\left(\frac{9}{10}\right)^1 = 3 \times \left(\frac{1}{10}\right)^2 \times \frac{9}{10}$

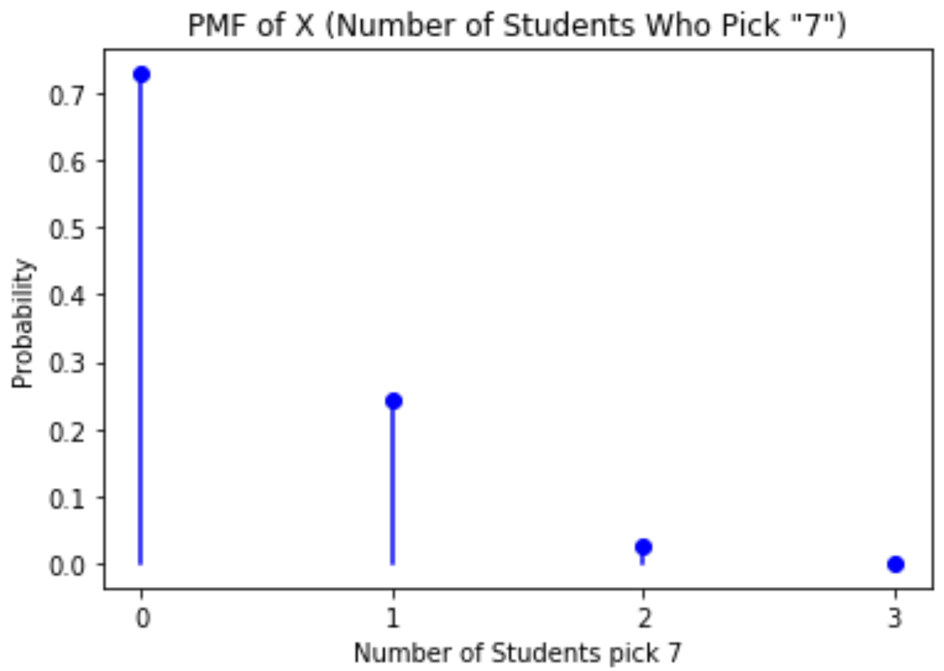
$P(X=2) = 3 \times \frac{1}{100} \times \frac{9}{10}$

$P(X=2) = \frac{27}{1000}$

- When $k=3$: All three students pick "7".

$P(X=3) = \binom{3}{3}\left(\frac{1}{10}\right)^3\left(\frac{9}{10}\right)^0 = 1 \times \left(\frac{1}{10}\right)^3 \times 1$

$P(X=3) = \frac{1}{1000}$



Example

If we toss a fair dice 5 times, then what is the probability of getting two "6"?
Let *X* be the rv of the number of "6" we would get.

- **5 identical trials:**

Toss the SAME dice $n = 5$ times.

- **Trials are independent:**

An outcome does not affect another outcome.

- **Trials result in two possible outcomes denoted by success/failure:**

Success: the number "6" is obtained; Failure: another number (not "6") is obtained.

- **The probability of success *p* is constant across trials:**

The probability of getting "6"in a trial is fixed to be $p = 1/6$.

$$\boxed{P(x) = \binom{n}{x}(p^x)(q^{n-x}), \quad \text{for } x = 0,\ 1,\ 2,\ \ldots,\ \text{or } n}$$

Thus, $X \sim Binomial\left(5, \frac{1}{6}\right)$ and $P(X = 2) = p(2) = \binom{5}{2}\left(\frac{1}{6}\right)^2\left(1 - \frac{1}{6}\right)^{5-2}$

Example

According to a recent study, 65% of all homes in a certain county have high levels of radon gas leaking into their basements. Four homes are selected at random and tested for radon. The random variable *x* is the number of homes with high levels of radon (out of the four).

Example

According to a recent study, 65% of all homes in a certain county have high levels of radon gas leaking into their basements. Four homes are selected at random and tested for radon. The random variable *x* is the number of homes with high levels of radon (out of the four).

Properties:

1. There are 4 repeated trials: $n = 4$. The trials are independent.
2. Each test for radon is a trial, and each test has two outcomes: *radon* or *no radon*.
3. $p = P(\text{radon}) = 0.65$, $q = P(\text{no radon}) = 0.35$

   $p + q = 1$
4. *x* is the number of homes with high levels of radon, possible values: 0, 1, 2, 3, 4

The binomial probability function:

$$P(x) = \binom{4}{x}(.65)^x(.35)^{4-x}, \quad \text{for } x = 0,\ 1,\ 2,\ 3,\ 4$$

$$P(0) = \binom{4}{0}(.65)^0(.35)^4 = (1)(1)(0.0150) = 0.0150$$

$$P(1) = \binom{4}{1}(.65)^1(.35)^3 = (4)(0.65)(0.0429) = 0.1115$$

$$P(2) = \binom{4}{2}(.65)^2(.35)^2 = (6)(0.4225)(0.1225) = 0.3105$$

$$P(3) = \binom{4}{3}(.65)^3(.35)^1 = (4)(0.2746)(0.35) = 0.3845$$

$$P(4) = \binom{4}{4}(.65)^4(.35)^0 = (1)(0.1785)(1) = 0.1785$$

Example

In a certain automobile dealership, 70% of all customers purchase an extended warranty with their new car. For 15 customers selected at random:

1. Find the probability that exactly 12 will purchase an extended warranty.
2. Find the probability at most 13 will purchase an extended warranty.

Example

Disks produced by a certain company will be defective with probability 0.01 independently of each other. The company sells the disks in packages of 10 and offers a money-back guarantee that at most 1 of the 10 disks is defective. What proportion of packages is returned?

Example

If we consider a random experiment of tossing a fair coin in 5 trials with X of the total number of heads and want to find $P(X>3)$, then we need to know the distribution of X. Later, we would study some well-known distributions and then would know that such a rv X follows a ***Binomial Distribution***.

According to this distribution, we know that

Now, we are able to find its exact value.

Example

If we consider a random experiment of tossing a fair coin in 5 trials with X of the total number of heads and want to find $P(X>3)$, then we need to know the distribution of X. Later, we would study some well-known distributions and then would know that such a rv X follows a ***Binomial Distribution***.

According to this distribution, we know that

Now, we are able to find its exact value.

Note that

- **5 identical trials:**

Flip the SAME coin $n = 5$ times.

- **Trials are independent:**

An outcome does not affect another outcome.

- **Trials result in two possible outcomes denoted by success/failure:**

Success: head; Failure: tail.

- **The probability of success *p* is constant across trials:**

The probability of getting a head in a trial is fixed to be $p = 1/2$.

Example

If we consider a random experiment of tossing a fair coin in 5 trials with *X* of the total number of heads and want to find $P(X>3)$, then we need to know the distribution of *X*. Later, we would study some well-known distributions and then would know that such a rv *X* follows a ***Binomial Distribution***.

According to this distribution, we know that

Thus, $X \sim Binomial\left(5,\ \frac{1}{2}\right)$ and

$$P(X > 3) = P(X = 4) + P(X = 5) = p(4) + p(5)$$

$$= \binom{5}{4}\left(\frac{1}{2}\right)^4\left(\frac{1}{2}\right)^{5-4} + \binom{5}{5}\left(\frac{1}{2}\right)^5\left(\frac{1}{2}\right)^{5-5}$$

= 3/16 = 0.1875

Now, we are able to find its exact value.

Note that

- **5 identical trials:**

Flip the SAME coin $n = 5$ times.

- **Trials are independent:**

An outcome does not affect another outcome.

- **Trials result in two possible outcomes denoted by success/failure:**

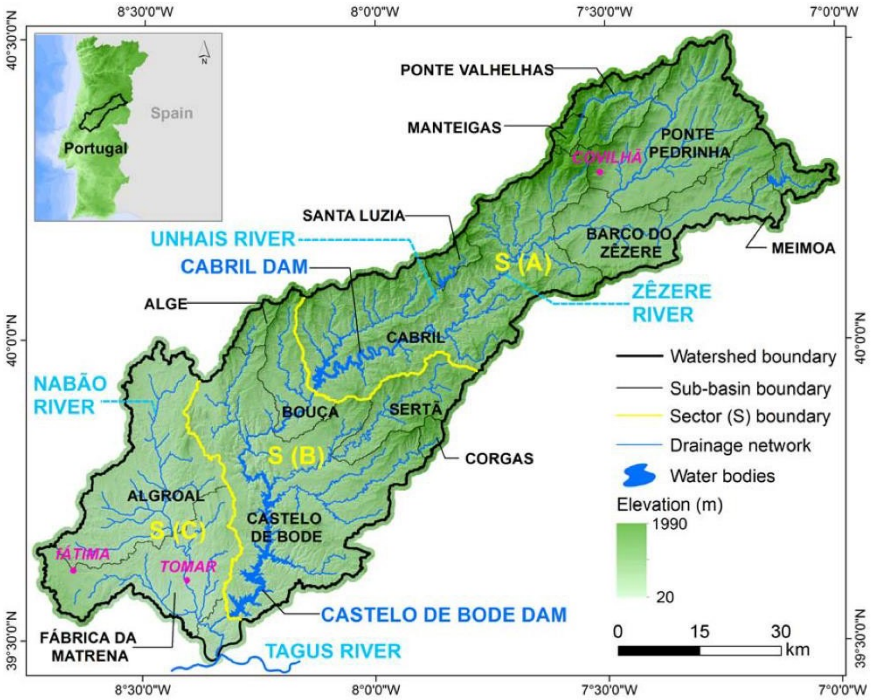
Success: head; Failure: tail.

- **The probability of success *p* is constant across trials:**

The probability of getting a head in a trial is fixed to be $p = 1/2$.

Example

Dams in Portugal go under careful inspection every 5 years for major renovation. 12 % of the dams go under renovation due to leakage or rusting. There are 10 major dams in Portugal.



a) No more than 2 dams renovated

b) At least 2 dams renovated

c) Plot the probability distribution

Mean and Standard Deviation of the Binomial Distribution

The **mean (expectation)** and **standard deviation** of a theoretical binomial distribution can be found by using the following two formulas:

$$\mu = np$$

$$\sigma = \sqrt{npq}$$

n: number of trials
p: the probability of success
q: the probability of failure (q =1-p)

Note:

1. Mean is intuitive: number of trials multiplied by the probability of a success.
2. The variance of a binomial probability distribution is:

$$\sigma^2 = \left(\sqrt{npq}\right)^2 = npq$$

Example

Find the mean and standard deviation of the binomial distribution when n = 18 and p = 0.75.

Solution:

$n = 18$, $p = 0.75$, $q = 1 - 0.75 = 0.25$

$$\mu = np = (18)(.75) = 13.5$$

$$\sigma = \sqrt{npq} = \sqrt{(18)(.75)(.25)} = \sqrt{3.375} \approx 1.8371$$

CityU

Example

Find the mean and standard deviation of the binomial distribution when n = 18 and p = 0.75.

The probability function is

$$P(x)=\binom{18}{x}(.75)^x(.25)^{18-x} \quad \text{for } x=0,\ 1,\ 2,\ \ldots,\ 18$$

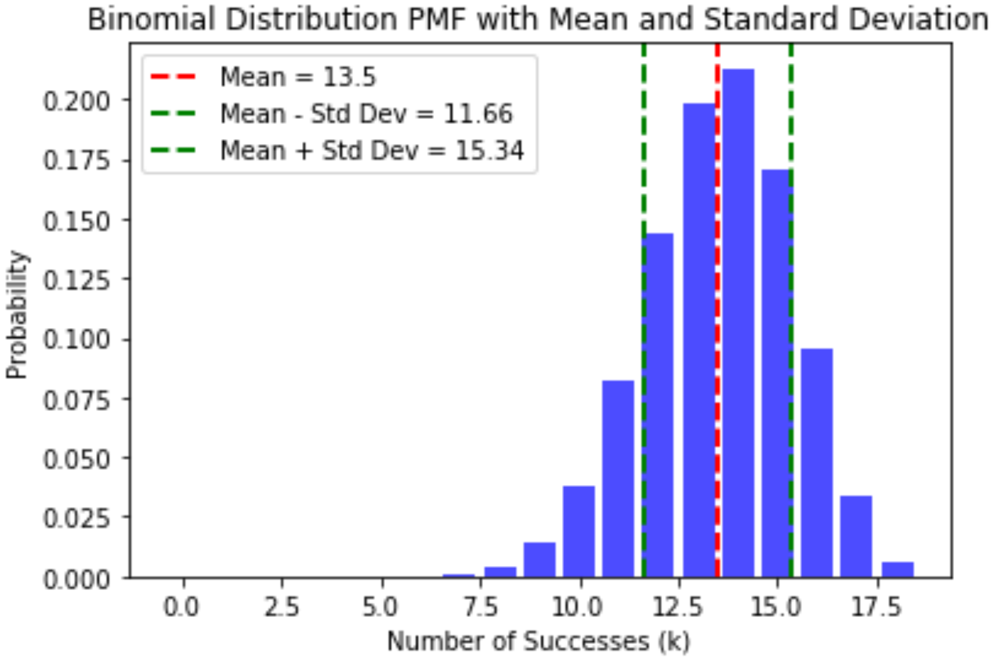


Table of values and probabilities:

x	P(X = x)
0.00	0.0000
1.00	0.0000
2.00	0.0000
3.00	0.0000
4.00	0.0000
5.00	0.0000
6.00	0.0002
7.00	0.0010
8.00	0.0042
9.00	0.0139
10.00	0.0376
11.00	0.0820
12.00	0.1436
13.00	0.1988
14.00	0.2130
15.00	0.1704
16.00	0.0958
17.00	0.0338
18.00	0.0056

Example

Find the mean and standard deviation of the binomial distribution when *n* = 18 and *p* = 0.75.

Solution:

n = 18, *p* = 0.75, *q* = 1 – 0.75 = 0.25

$$\mu = np = (18)(.75) = 13.5$$

$$\sigma = \sqrt{npq} = \sqrt{(18)(.75)(.25)} = \sqrt{3.375} \approx 1.8371$$

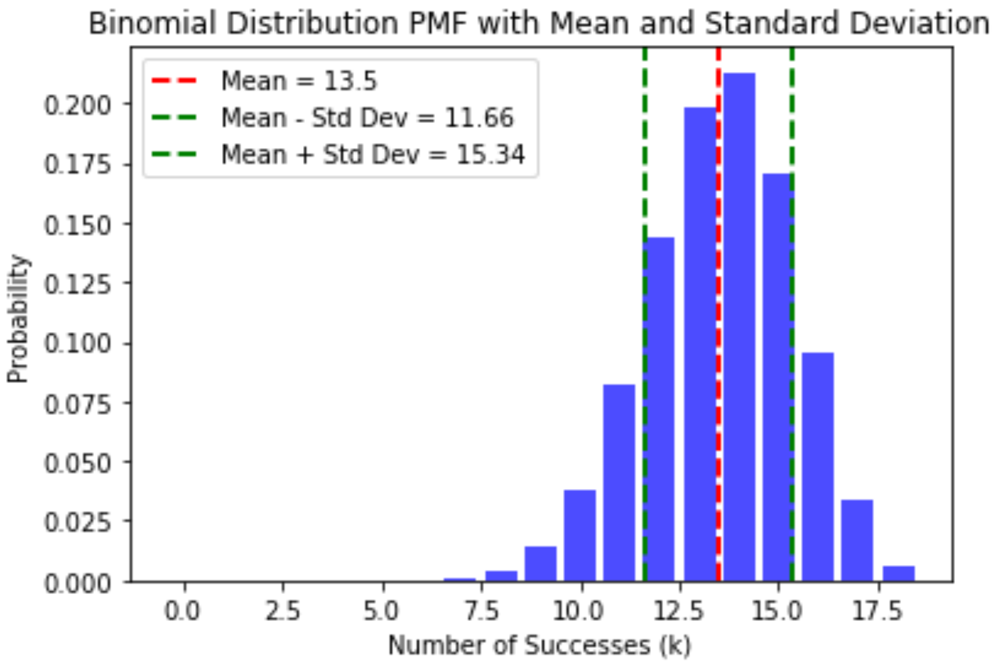


Table of values and probabilities:

x	P(X = x)
0.00	0.0000
1.00	0.0000
2.00	0.0000
3.00	0.0000
4.00	0.0000
5.00	0.0000
6.00	0.0002
7.00	0.0010
8.00	0.0042
9.00	0.0139
10.00	0.0376
11.00	0.0820
12.00	0.1436
13.00	0.1988
14.00	0.2130
15.00	0.1704
16.00	0.0958
17.00	0.0338
18.00	0.0056

Python code

Implement from scratch by using the formula for the PMF of a binomial distribution

```python
# Parameters for the binomial distribution
n = 18
p = 0.75

# Function to calculate the factorial of a number
def factorial(num):
    if num <= 1:
        return 1
    return num * factorial(num - 1)

# Function to calculate the binomial coefficient (n choose k)
def binomial_coefficient(n, k):
    return factorial(n) / (factorial(k) * factorial(n - k))

# Function to calculate the PMF of the binomial distribution
def binomial_pmf(n, k, p):
    return binomial_coefficient(n, k) * (p ** k) * ((1 - p) ** (n - k))

# Compute PMF for all possible values of k (0, 1, ..., n)
pmf_values = [binomial_pmf(n, k, p) for k in range(n + 1)]

# Print the PMF values
for k, pmf in enumerate(pmf_values):
    print(f"P(X = {k}) = {pmf}")
```

Output:

```
P(X = 0) = 1.4551915228366852e-11
P(X = 1) = 7.858034223181e-10
P(X = 2) = 2.0037987269461155e-08
P(X = 3) = 3.2060779631137785e-07
P(X = 4) = 3.6068377085030008e-06
P(X = 5) = 3.0297436751425266e-05
P(X = 6) = 0.00019693333888426423
P(X = 7) = 0.0010128000285476446
P(X = 8) = 0.004177800117759034
P(X = 9) = 0.013926000392530113
P(X = 10) = 0.037600201059831306
P(X = 11) = 0.0820368023123592l
P(X = 12) = 0.1435644040466286 2
P(X = 13) = 0.1987814825261011 7
P(X = 14) = 0.212980159849394l
P(X = 15) = 0.170384127879515 3
P(X = 16) = 0.0958410719322273 5
P(X = 17) = 0.0338262606819625 94
P(X = 18) = 0.00563771011366043 2
```

Python code

Use the scipy.stats

```
from scipy.stats import binom
# Parameters for the binomial distribution
n = 18
p = 0.75

# Compute PMF for all possible values of k (0, 1, ..., n)
pmf_values = [binom.pmf(k, n, p) for k in range(n + 1)]

# Print the PMF values
for k, pmf in enumerate(pmf_values):
    print(f"P(X = {k}) = {pmf:.8f}")
```

```
# Calculate the mean and standard deviation
mean, var = binom.stats(n, p)
std_dev = np.sqrt(var)
```

Output:

```
P(X = 0) = 0.00000000
P(X = 1) = 0.00000000
P(X = 2) = 0.00000002
P(X = 3) = 0.00000032
P(X = 4) = 0.00000361
P(X = 5) = 0.00003030
P(X = 6) = 0.00019693
P(X = 7) = 0.00101280
P(X = 8) = 0.00417780
P(X = 9) = 0.01392600
P(X = 10) = 0.03760020
P(X = 11) = 0.08203680
P(X = 12) = 0.14356440
P(X = 13) = 0.19878148
P(X = 14) = 0.21298016
P(X = 15) = 0.17038413
P(X = 16) = 0.09584107
P(X = 17) = 0.03382626
P(X = 18) = 0.00563771
```

Binomial Random Variable

The two random variables have the same probability of success

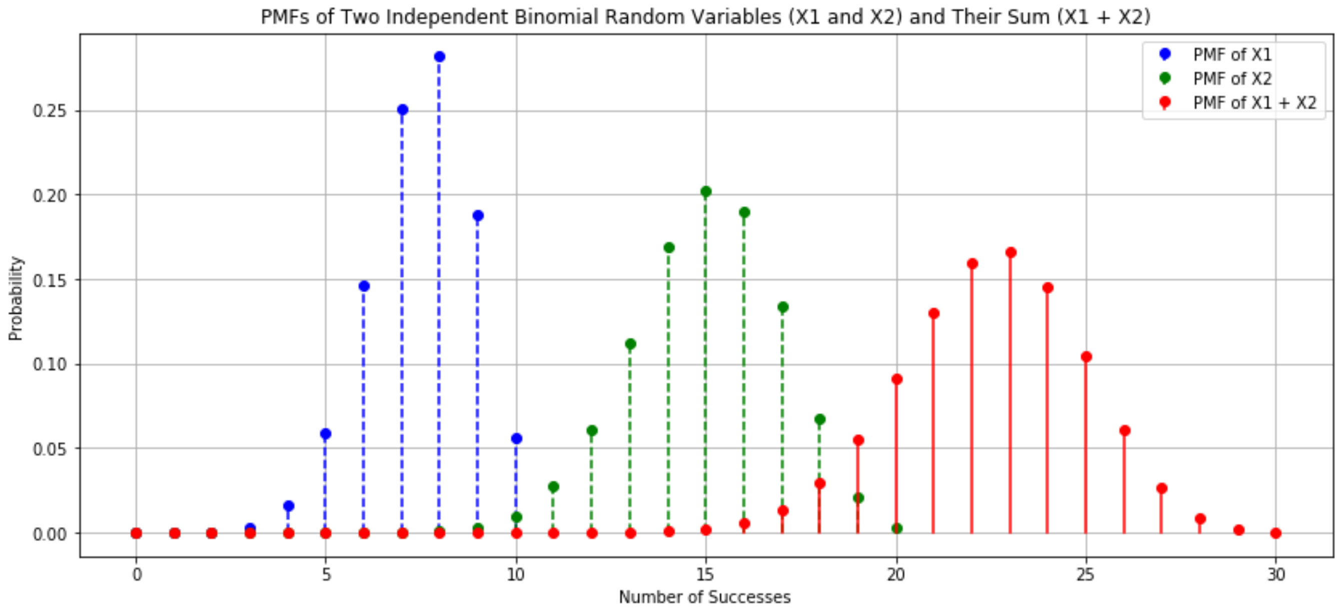
If

$$X_1 \sim \text{Binomial}(n_1, p), \quad X_2 \sim \text{Binomial}(n_2, p)$$

and X_1 and X_2 are $\boxed{\text{independent}}$, then

$$X_1 + X_2 \sim \text{Binomial}(n_1 + n_2, p)$$

Let's assume $n1$=10, n2 = 20 and $p1$=$p2$=0.75



Binomial Random Variable

If

The two random variables have the same probability of success

$$X_1 \sim \text{Binomial}(n_1, p), \quad X_2 \sim \text{Binomial}(n_2, p)$$

and X_1 and X_2 are independent, then

$$X_1 + X_2 \sim \text{Binomial}(n_1 + n_2, p)$$

Consider $I_1, \ldots, I_{n_1}$ and $I_{n_1+1}, \ldots, I_{n_1+n_2}$

$$X_1 = \sum_{k=1}^{n_1} I_k, \quad X_2 = \sum_{k=n_1+1}^{n_1+n_2} I_k$$

because the sum of the number of successes in n1+*n*2 trials is simply the total number of successes

Thus

$$\boxed{X_1 + X_2 = \sum_{k=1}^{n_1+n_2} I_k}$$

Hypergeometric Random Variable

A hypergeometric random variable models the number of successes in a **sequence of draws without replacement from a finite population**. It's used in situations where the binomial distribution isn't suitable because the draws are not independent.

Hypergeometric Random Variable

Randomly pick *n* batteries from a bin of mixed batteries

- *N* are of acceptable quality.
- *M* are defective.

Let *X* denote the number of acceptable batteries.

How to compute P{*X*=i}?

How many ways to choose *n* from *N*+*M*: $\binom{N+M}{n}$

How many ways to choose *i* from *N* and (*n*-*i*) from *M*: $\binom{N}{i}\binom{M}{n-i}$

Thus, the PMF is:

$$P\{X=i\} = \frac{\binom{N}{i}\binom{M}{n-i}}{\binom{N+M}{n}}$$

We say that *X* follows the Hypergeometric law with parameters (*N*, *M*, *n*).

X ~ Hypergeometric (*N*, *M*, *n*)

Hypergeometric Random Variable

The probability mass function of a hypergeometric random variable X is :

$$P(X = k) = \frac{\binom{N}{k}\binom{M}{n-k}}{\binom{M+N}{n}}$$

where:

- $\binom{N}{k}$ represents the number of ways to choose k acceptable quality items from the N available.
- $\binom{M}{n-k}$ represents the number of ways to choose $n-k$ defective items from the M available.
- $\binom{M+N}{n}$ is the total number of ways to choose n items from the entire population of $M+N$.

Example

The components of a 6-component system are to be randomly chosen from a bin of 20 used components. The resulting system will be functional if at least 4 of its 6 components are in working condition. If 15 of the 20 components in the bin are in working condition, what is the probability that the resulting system will be functional?

Solution: Let X be the number of working components, then

$$X \sim \text{Hypergeometric}(N, M, n)$$

$$P\{X = i\} = \frac{\binom{N}{i}\binom{M}{n-i}}{\binom{N+M}{n}}$$

X ~ Hypergeometric (15; 5; 6)

$$P\{X \geq 4\} = P\{X = 4\} + P\{X = 5\} + P\{X = 6\}$$
$$= \frac{\binom{15}{4}\binom{5}{2} + \binom{15}{5}\binom{5}{1} + \binom{15}{6}\binom{5}{0}}{\binom{20}{6}}$$
$$= .8687$$

Hypergeometric Random Variable

How to find the mean for Hypergeometric (N, M, n)?

Label the *n* selected batteries by *i*=1, 2, 3,…n

$$X_i = \begin{cases} 1 & \text{if battery } i \text{ is acceptable} \\ 0 & \text{otherwise} \end{cases}$$

Then the number of acceptable batteries is:

$$X = \sum_{i=1}^{n} X_i$$

We can apply the similar method as for binomial random variables, but be careful

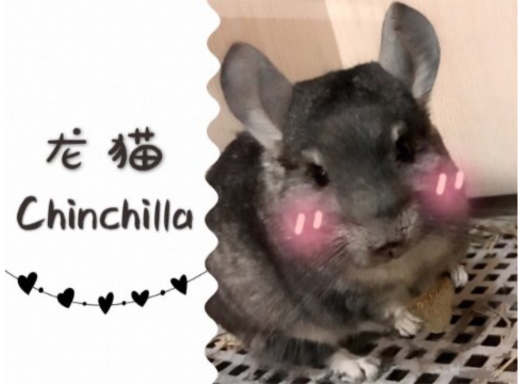
Since each battery is picked from a pool of *N* + *M*,

$$P\{X_i = 1\} = \frac{N}{N+M}$$

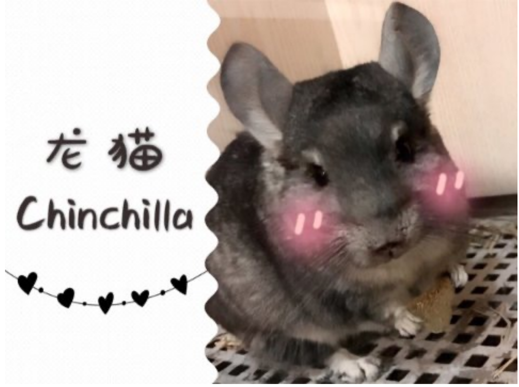
$$\boxed{E[X] = \sum_{i=1}^{n} \mathbb{E}[X_i] = \sum_{i=1}^{n} P\{X_i = 1\} = \frac{nN}{N+M}}$$

Example

Example: An unknown number, say Z, of chinchillas inhabit a natural reserve. How could ecologists estimate the population?



Example

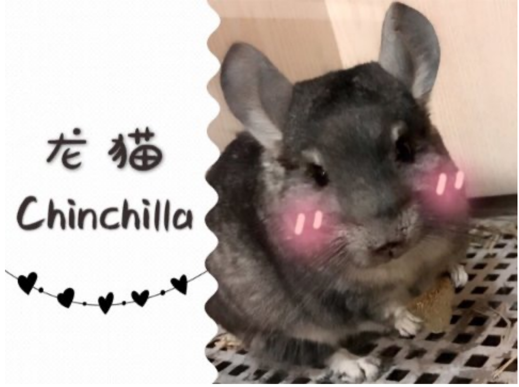


Example: An unknown number, say Z, of chinchillas inhabit a natural reserve.
How could ecologists estimate the population?

Since it is impossible to capture all of them, we often perform the following experiment:

1. First randomly catch a number, say $r = 100$, of these animals, mark them.
2. Release them to let them disperse the region.
3. Make a new catch of size, say $n = 200$, and release them.

Example



Example: An unknown number, say Z, of chinchillas inhabit a natural reserve. How could ecologists estimate the population?

Since it is impossible to capture all of them, we often perform the following experiment:

1. First randomly catch a number, say $r = 100$, of these animals, mark them.
2. Release them to let them disperse the region.
3. Make a new catch of size, say $n = 200$, and release them.

We can repeat <u>step 3</u> to get more data. Suppose we make 8 catches

20; 30; 25; 20; 36; 18; 28; 23

The sample mean is $\overline{X} = 25$.

How would you use this experiment to estimate Z?

Example

Estimate the number of Chinchillas Z.

Since r of them are marked,

Marked	Unmarked
r	$Z - r$

Capture n, X denote the number of marked, so

$$X \sim \text{Hypergeometric}(r, Z\text{-}r, n)$$

$$E[X] = \frac{nN}{N+M}$$

The expectation is E[X] = nr/Z.

Since $\overline{X}$ is a good approximation of E[X], ($\overline{X} \approx$ nr/Z),

$$Z \approx \frac{nr}{\overline{X}}$$

Example: An unknown number, say Z, of chinchillas inhabit a natural reserve. How could ecologists estimate the population?

Since it is impossible to capture all of them, we often perform the following experiment:

1. First randomly catch a number, say r = 100, of these animals, mark them.
2. Release them to let them disperse the region.
3. Make a new catch of size, say n = 200, and release them.

We can repeat step 3 to get more data. Suppose we make 8 catches

20; 30; 25; 20; 36; 18; 28; 23

The sample mean is $\overline{X}$ = 25.

How would you use this experiment to estimate Z?

Plugin the numbers,

Z = (200 X 100)/25 = 800

Poisson Distribution

Poisson Distribution

Another discrete distribution that we will consider in this section is the **Poisson distribution**. It can be used to determine the probability of counts of the occurrence of an event over time (or space).

- Notation:

$$X \sim Poisson(\lambda),$$

where $\lambda \in (0, \infty)$ is the rate of occurrences of an event per unit time (or space) or the average number of occurrences of the event per unit time (or space).

Here are some typical examples for Poisson distributions.

1. The number of traffic accidents occurring on a highway in a day.
2. The number of people joining a line in an hour.
3. The number of typos per page of an essay.

Poisson Process

The Poisson distribution models the number of events in a fixed interval of time or space for a process where events occur continuously and independently at a constant average rate.

The key assumptions of a Poisson process are:

1. Events are independent of each other.
2. **The average rate (or intensity) λ at which events occur is constant**.
3. Two events cannot occur at exactly the same instant.
4. The probability of exactly one event occurring in a sufficiently small interval is proportional to the length of the interval.

Poisson Process

Given these assumptions, the formula:

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

the probability of observing exactly *k* events in the interval, where λ is the expected number of events

$e^{-\lambda}$ is the probability that no events occur, *e* is Euler's number (~2.71828).

λ^k represents the weight of the probability for *k* occurrences.

The division by *k*! accounts for the fact that the order of occurrences does not matter (it's the combination, not the permutation).

Poisson Distribution

- Population mean and population variance:

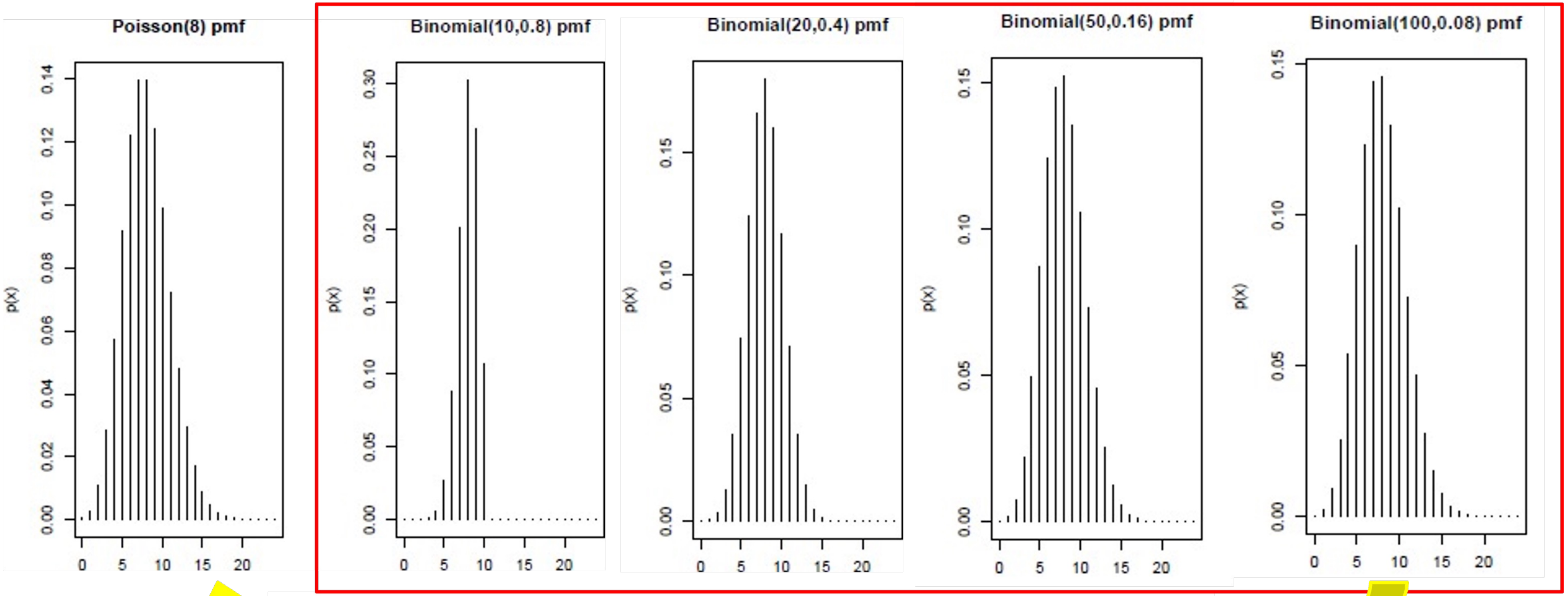
$$E(X) = \lambda,\ Var(X) = \lambda.$$

λ represents the average rate of occurrence of the events in the given interval.

This rate λ also describes the variance in the number of events.

The Connection Between the Poisson and Binomial Distributions

- The Poisson distribution is actually a limiting case of a Binomial distribution when the number of trials, n, gets very large and p, the probability of success, is small.



Consider Poisson(λ) and Binomial(n, p) with

$$\boxed{\lambda = np}$$

Poisson(λ) and Binomial(n, p) becomes "close" when n becomes large and p becomes small, but keep $\lambda = np$.

A rule of thumb:
- n **≥100 and np≤10**;
- Poisson distribution taking **λ=np**

The Connection Between the Poisson and Binomial Distributions

Proof Let us denote the expected value of the binomial distribution np by λ

$$p = \frac{\lambda}{n} \qquad q = 1 - \frac{\lambda}{n}$$

Rewrite Binomial distribution P(x) in $P(x) = {}_nC_x \left(\frac{\lambda}{n}\right)^x \left(1 - \frac{\lambda}{n}\right)^{n-x}$

$$P(x) = \frac{n(n-1)(n-2)\cdots(n-x+1)}{x!} \cdot \frac{\lambda^x}{n^x}\left(1 - \frac{\lambda}{n}\right)^{n-x}$$

$$P(x) = \frac{n}{n} \cdot \frac{n-1}{n} \cdots \frac{n-x+1}{n} \cdot \frac{\lambda^x}{x!}\left(1 - \frac{\lambda}{n}\right)^{n-x}$$

$$P(x) = \frac{n}{n} \cdot \frac{n-1}{n} \cdots \frac{n-x+1}{n} \cdot \frac{\lambda^x}{x!}\left(1 - \frac{\lambda}{n}\right)^{n}\left(1 - \frac{\lambda}{n}\right)^{-x}$$

n→∞, the **first term and the third term** approach 1,
and the last term approaches $e^{-\lambda}$ by definition of the exponential function

So:

$$\lim_{n\to\infty} P(x) = \frac{e^{-\lambda}\lambda^x}{x!}$$

~ PMF of the Poisson distribution with parameter λ

Example

Suppose the probability that an item produced by a certain machine will be defective is 0.1, independent of other items. Find the probability that a sample of 10 items will contain at most one defective item.

Let X be the number of defects, then X Binomial $\sim$ (10, 0.1).

So the probability is:

$$P\{X=0\}+P\{X=1\}=\binom{10}{0}(.1)^0(.9)^{10}+\binom{10}{1}(.1)^1(.9)^9=\boxed{.7361}$$

Example

Suppose the probability that an item produced by a certain machine will be defective is 0.1, independent of other items. Find the probability that a sample of 10 items will contain at most one defective item.

Let X be the number of defects, then X Binomial ~ (10, 0.1).

So the probability is:

$$P\{X=0\}+P\{X=1\}=\binom{10}{0}(.1)^0(.9)^{10}+\binom{10}{1}(.1)^1(.9)^9=\boxed{.7361}$$

You can use Y ~ Poisson (1) to approximate it:

$$p(x)=P(X=x)=\frac{\lambda^x e^{-\lambda}}{x!}, for\ x=0,1,\dots;$$

$$\lambda=np=(10)\ (0.1)=1$$

$$P\{Y=0\}+P\{Y=1\}=e^{-1}\frac{1^0}{0!}+e^{-1}\frac{1^1}{1!}=2e^{-1}=\boxed{0.7358}$$

CityU

Poisson Random Variable

Poisson random variables can be used to approximate (or model) the following

- The number of misprints on a page (or a group of pages) of a book.
- The number of people in a community living to 100 years of age.
- The number of wrong telephone numbers that are dialed in a day.
- The number of transistors that fail on their first day of use.
- The number of customers entering a post office on a given day.
- …

Example

If the average number of claims handled daily by an insurance company is 5. What proportion of days have less than (<) 3 claims?

$$p(x) = P(X = x) = \frac{\lambda^x e^{-\lambda}}{x!}, for\ x = 0, 1, \ldots;$$

$\lambda = 5$

Solution:

Because the company probability insures:

- a large number of clients
- each having a small probability of making a claim

It is reasonable to suppose that the number of claims handled daily, call it X, is Poisson(5).

$$\begin{aligned} P\{X < 3\} &= P\{X = 0\} + P\{X = 1\} + P\{X = 2\} \\ &= e^{-5} + \frac{5^1}{1!}e^{-5} + \frac{5^2}{2!}e^{-5} \\ &= \frac{37}{2}e^{-5} \end{aligned}$$

CityU

Example

Suppose that the average number of accidents occurring weekly on a particular stretch of a highway equals 3. Calculate the probability that there is at least one accident this week.

We model the number of accidents X ~ Poisson(3):

$$P\{X \geq 1\} = \sum_{i=1}^{\infty} P\{X = i\} = \sum \text{..... too complicated}$$

$$\begin{aligned} P\{X \geq 1\} &= 1 - P\{X = 0\} \\ &= 1 - e^{-3}\frac{3^0}{0!} = 1 - e^{-3} \approx .9502 \end{aligned}$$

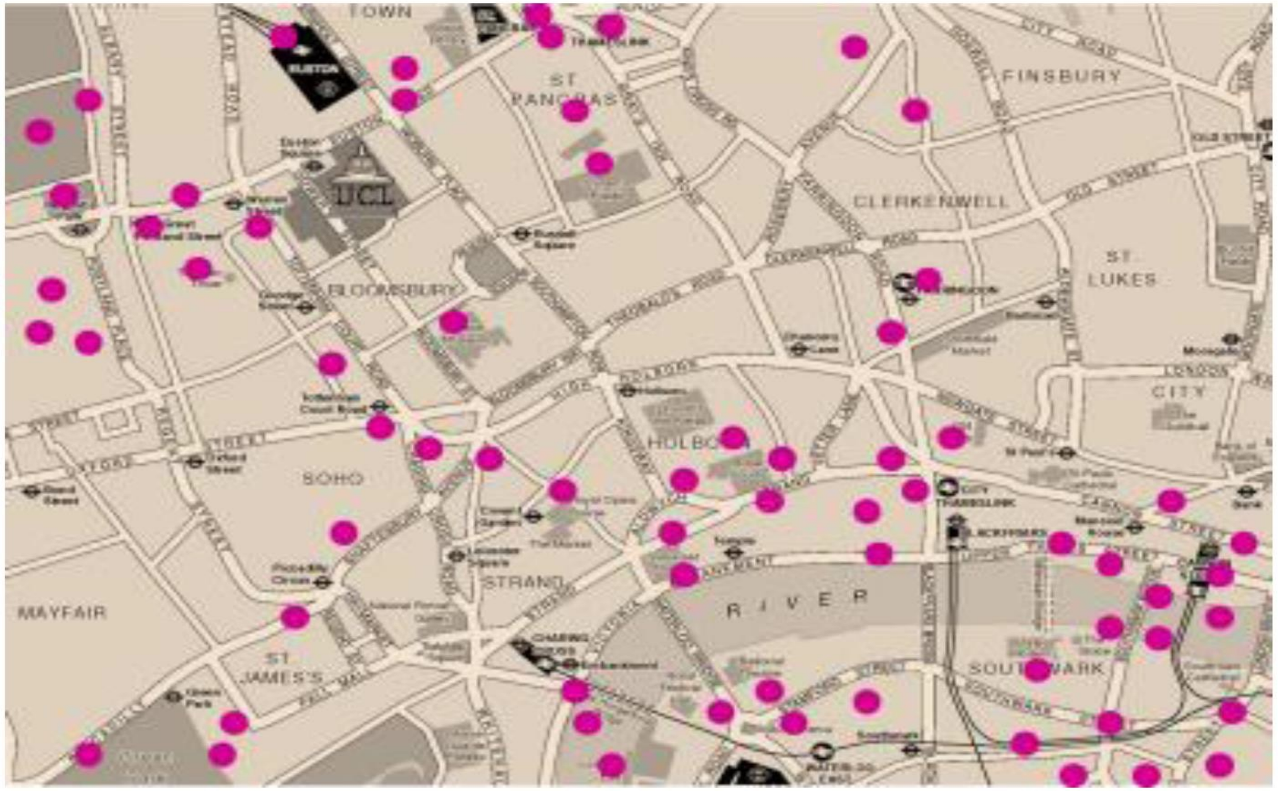
Example

From June 1944 to March 1945 during World War II, Germany launched a total of 9,251 V-1 flying bombs — "buzz" bombs — against England. Of these, only 2,419 made it to their intended target areas and, of these, 537 struck South London.

Shortly after World War II, a British statistician named *R.D. Clarke* took a 12 km x 12 km heavily bombed region of South London, and sliced it up in to a grid. In all, he divided it into 576 squares (or regions), each about the size of 25 city blocks. Next, he counted the number of regions with 0 bombs dropped, 1 bomb dropped, 2 bombs dropped, and so on.

The red dots show where the flying bombs landed in South London

Clarke then showed that the pattern of "hits" would follow a Poisson distribution!



Example

Now based on his result, find the probability that a randomly selected region was hit exactly twice.

First, since 537 bombs struck 576 regions, the mean number of hits per region is

$$\lambda = \frac{537}{576} = 0.9323$$

Let X be the rv of the number of hits in the selected region. Thus,

$$X \sim Poisson(0.9323)$$

and the required probability of the selected region being hit exactly twice is

$$P(X = 2) = \frac{(0.9323)^2 e^{-0.9323}}{2!} = 0.1711.$$

Example

We are studying the earthquakes in California with a reading over 6.7 on the Richter scale. Suppose that ***on average***, there are 1.5 earthquakes with a reading over 6.7 on the Richter scale in California per year.

In this case, $\lambda=1.5$ is called the rate of the occurrences of the earthquakes above 6.7 on the Richter scale, and the time unit is 1 year.
Let X be the rv of the number of earthquakes above 6.7 on the Richter scale in the upcoming year, then

$$X \sim Poisson(1.5).$$

Thus, the probability that there will be 5 earthquakes with a reading over 6.7 on the Richter scale in the upcoming year is

$$P(X = 5) = \frac{(1.5)^5 e^{-1.5}}{5!} = 0.0141.$$

Example

How about the probability that there will be 5 earthquakes with a reading over 6.7 on the Richter scale ***in the next 4 years***?

To answer this question, we need the following result:

When we study the count of occurrences of an event over a period of *t* units of time with the rate λ of the occurrences per unit time, use

$$Yt \sim Poisson(\lambda t),$$

where *Yt* is the random variable of the count of occurrences of the event over a period of time *t*.

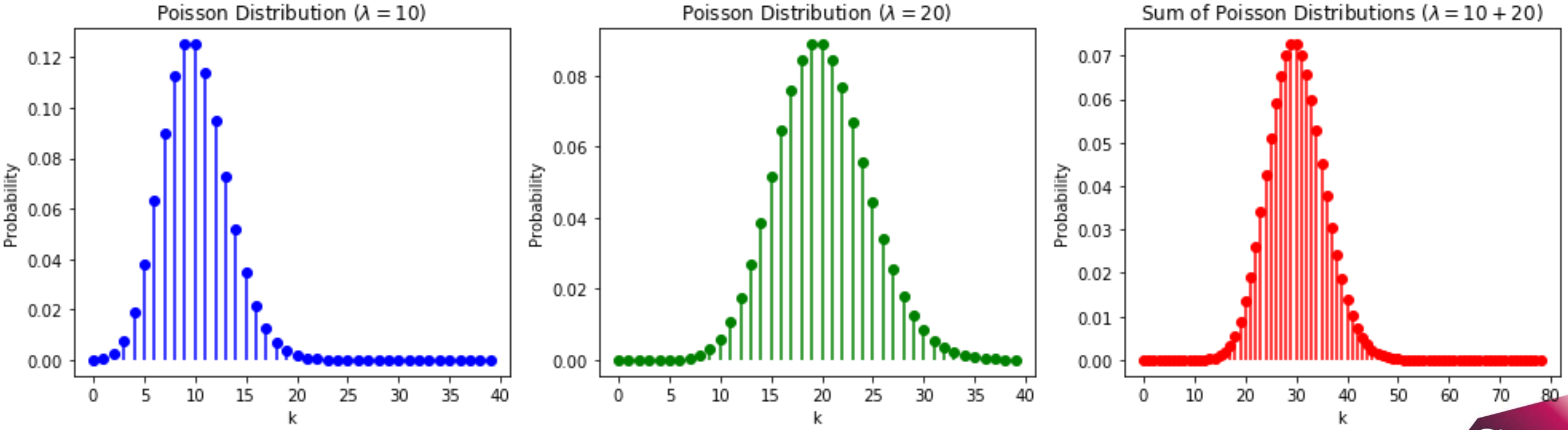
Thus, the required probability is

$$P(Y_4 = 5) = \frac{(\lambda t)^5 e^{-(\lambda t)}}{5!} = \frac{(1.5 \times 4)^5 e^{-(1.5 \times 4)}}{5!} = 0.1606.$$

Poisson Random Variable

If $X_1 \sim$ Poisson(λ_1) and $X_2 \sim$ Poisson(λ_2), and X_1 and X_2 are independent, then;

$$X_1+X_2 \sim \text{Poisson}(\lambda_1+\lambda_2)$$



Poisson Random Variable

Proof

$$P(X=k)=\frac{e^{-\lambda_1}\lambda_1^k}{k!}$$
$$P(Y=m)=\frac{e^{-\lambda_2}\lambda_2^m}{m!}$$

When X and Y are independent, the probability mass function of their sum $Z=X+Y$ is given by the convolution of their individual PMFs:

$$P(Z=n)=P(X+Y=n)=\sum_{k=0}^{n}P(X=k)P(Y=n-k)$$

Using the PMFs of X and Y, we have:

$$P(Z=n)=\sum_{k=0}^{n}\left(\frac{e^{-\lambda_1}\lambda_1^k}{k!}\frac{e^{-\lambda_2}\lambda_2^{n-k}}{(n-k)!}\right)$$

$$P(Z=n)=e^{-\lambda_1}e^{-\lambda_2}\sum_{k=0}^{n}\frac{\lambda_1^k\lambda_2^{n-k}}{k!(n-k)!}$$

$$P(Z=n)=e^{-(\lambda_1+\lambda_2)}\frac{1}{n!}\sum_{k=0}^{n}\binom{n}{k}\lambda_1^k\lambda_2^{n-k}$$

Now, recognize that the summation is the binomial expansion of $(\lambda_1+\lambda_2)^n$:

$$P(Z=n)=e^{-(\lambda_1+\lambda_2)}\frac{(\lambda_1+\lambda_2)^n}{n!}$$

Example

It has been estimated that the number of defective stereos produced daily at a certain plant is Poisson distributed with mean 4. Over a 2-day span, what is the probability that the number of defective stereos does not exceed 3?

Let X_1 and X_2 be the number of defectives produced in day 1 and day 2, respectively.

Example

It has been estimated that the number of defective stereos produced daily at a certain plant is Poisson distributed with mean 4. Over a 2-day span, what is the probability that the number of defective stereos does not exceed 3?

Let X_1 and X_2 be the number of defectives produced in day 1 and day 2, respectively.

$$X_1 \sim \text{Poisson}(4), \quad X_2 \sim \text{Poisson}(4)$$

$$P\{X_1 + X_2 \leq 3\} = \sum_{i=0}^{3} P\{X_1 + X_2 = i\}$$

$$X_1 + X_2 \text{ is Poisson with } (4+4)$$

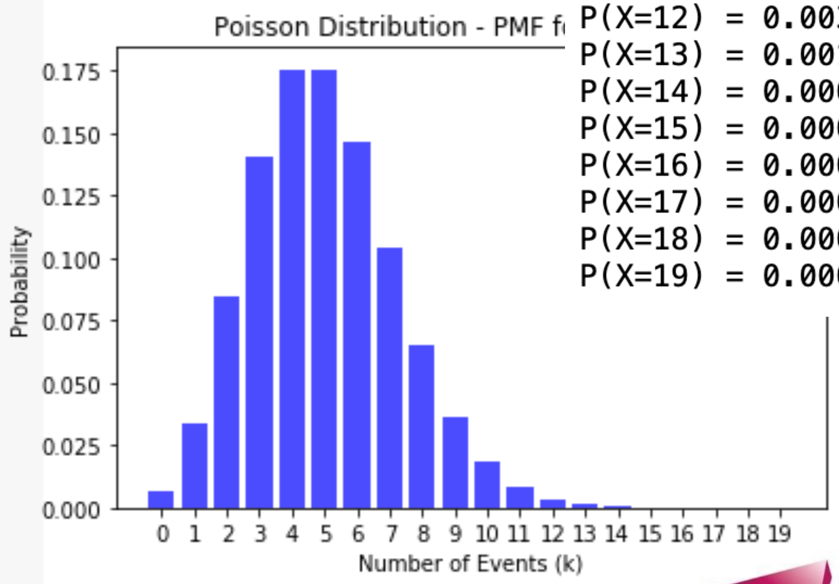
$$= \sum_{i=0}^{3} e^{-8} \frac{8^i}{i!} = \frac{379}{3} e^{-8} \approx .04238$$

Python code for Poisson distribution

From Scratch

```
import math
import matplotlib.pyplot as plt

# Poisson PMF calculation
def poisson_pmf(lmbda, k):
    return (lmbda**k * math.exp(-lmbda)) / math.factorial(k)

# Parameters
lmbda = 5  # Average rate (lambda)
k_values = range(0, 20)  # Range of k values for which to calculate PMF

# Calculate PMF values
pmf_values = [poisson_pmf(lmbda, k) for k in k_values]

# Output the PMF values
for k, pmf in zip(k_values, pmf_values):
    print(f"P(X={k}) = {pmf:.5f}")

# Plotting the PMF
plt.bar(k_values, pmf_values, color='blue', alpha=0.7)
plt.title('Poisson Distribution - PMF for lambda=5')
plt.xlabel('Number of Events (k)')
plt.ylabel('Probability')
plt.xticks(k_values)
plt.show()
```

```
P(X=0) = 0.00674
P(X=1) = 0.03369
P(X=2) = 0.08422
P(X=3) = 0.14037
P(X=4) = 0.17547
P(X=5) = 0.17547
P(X=6) = 0.14622
P(X=7) = 0.10444
P(X=8) = 0.06528
P(X=9) = 0.03627
P(X=10) = 0.01813
P(X=11) = 0.00824
P(X=12) = 0.00343
P(X=13) = 0.00132
P(X=14) = 0.00047
P(X=15) = 0.00016
P(X=16) = 0.00005
P(X=17) = 0.00001
P(X=18) = 0.00000
P(X=19) = 0.00000
```

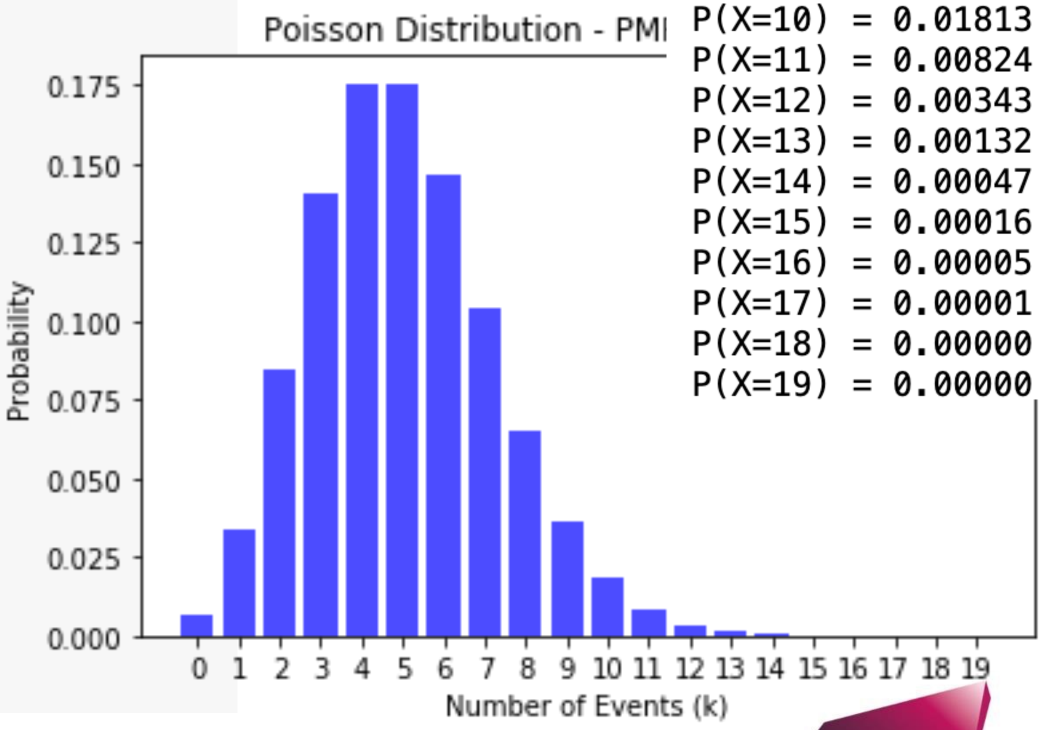


Python code for Poisson distribution

Using SciPy

```python
import matplotlib.pyplot as plt
from scipy.stats import poisson

# Define the average rate (lambda) for the Poisson distribution
lambda_ = 5

# Define the range of k values for which we want to calculate PMF values
k_values = range(20)

# Calculate the PMF values using scipy's poisson.pmf function
pmf_values = poisson.pmf(k_values, lambda_)

# Print the PMF values
for k, pmf in zip(k_values, pmf_values):
    print(f"P(X={k}) = {pmf:.5f}")

# Plot the PMF using a bar chart
plt.bar(k_values, pmf_values, color='blue', alpha=0.7)
plt.title('Poisson Distribution - PMF for lambda=5')
plt.xlabel('Number of Events (k)')
plt.ylabel('Probability')
plt.xticks(k_values)
plt.grid(axis='y', alpha=0.75)
plt.show()
```

```
P(X=0) = 0.00674
P(X=1) = 0.03369
P(X=2) = 0.08422
P(X=3) = 0.14037
P(X=4) = 0.17547
P(X=5) = 0.17547
P(X=6) = 0.14622
P(X=7) = 0.10444
P(X=8) = 0.06528
P(X=9) = 0.03627
P(X=10) = 0.01813
P(X=11) = 0.00824
P(X=12) = 0.00343
P(X=13) = 0.00132
P(X=14) = 0.00047
P(X=15) = 0.00016
P(X=16) = 0.00005
P(X=17) = 0.00001
P(X=18) = 0.00000
P(X=19) = 0.00000
```



More efficient if using SciPy

```python
import math
import time
from scipy.stats import poisson

def poisson_pmf_scratch(k, lambda_):
    return (lambda_ ** k * math.exp(-lambda_)) / math.factorial(k)

# Define the lambda and k_values
lambda_ = 5
k_values = range(100)  # Larger range of k_values

# Time the scratch implementation
start_time_scratch = time.time()
pmf_values_scratch = [poisson_pmf_scratch(k, lambda_) for k in k_values]
end_time_scratch = time.time()
runtime_scratch = end_time_scratch - start_time_scratch

# Time the scipy implementation
start_time_scipy = time.time()
pmf_values_scipy = poisson.pmf(k_values, lambda_)
end_time_scipy = time.time()
runtime_scipy = end_time_scipy - start_time_scipy

# Output the runtimes
print(f"Runtime for scratch implementation: {runtime_scratch:.6f} seconds")
print(f"Runtime for scipy implementation: {runtime_scipy:.6f} seconds")
```

```
Runtime for scratch implementation: 0.000783 seconds
Runtime for scipy implementation: 0.002831 seconds
```