Exp 3a-Binomial and Poisson Distributions

The Binomial distribution

Consider the following circumstances (binomial scenario):

1. There are n trials.

2. The trials are independent.

3. On each trial, only two things can happen. We refer to these two events as success and failure.

4. The probability of success is the same on each trial. This probability is usually called p.

5. We count the total number of successes. This is a discrete random variable, which we denote by X, and which can take any value between 0 and n (inclusive).

• The random variable X is said to have a binomial distribution with parameters n and p; abbreviated

$$X \sim \text{Bin}(n, p)$$

- It is easy to show that if $X \sim \text{Bin}(n, p)$ then

$$\text{P}[X = k] = \binom{n}{k} p^k (1-p)^{n-k}$$

for $k = 0, 1, \ldots, n$.

- $\binom{n}{k}$ is the *binomial coefficient* and is the number of sequences of length n containing k successes.

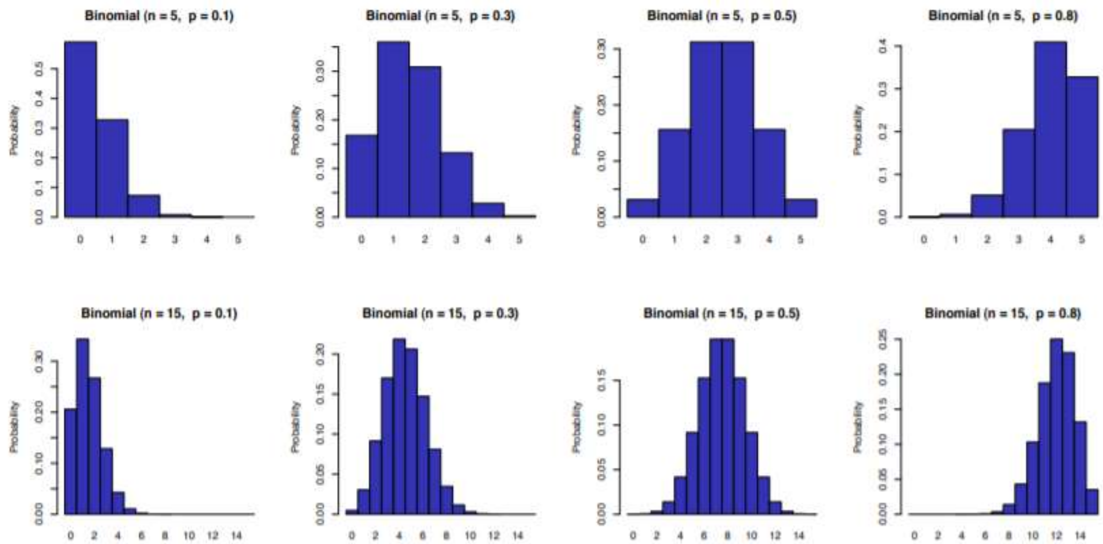
$$\binom{n}{k} = \frac{n!}{(n-k)!k!}$$

- The expectation and variance of X are given by

$$\text{E}[X] = np$$
$$\text{Var}[X] = np(1-p)$$

The Binomial Distribution: Example

The shape of the distribution depends on n and p.



R has four in-built functions to generate binomial distribution. They are described below.

```
dbinom(x, size, prob)
pbinom(x, size, prob)
qbinom(p, size, prob)
rbinom(n, size, prob)
```

Following is the description of the parameters used −

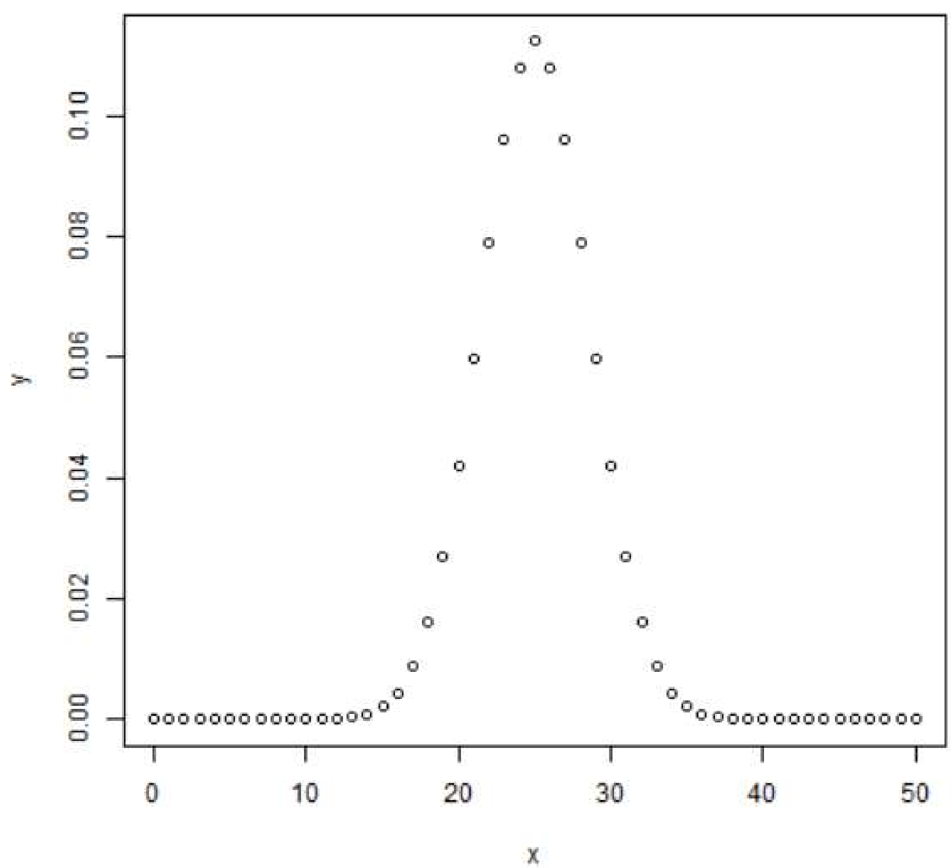
- **x** is a vector of numbers.
- **p** is a vector of probabilities.
- **n** is number of observations.
- **size** is the number of trials.
- **prob** is the probability of success of each trial.

dbinom()

This function gives the probability density distribution at each point.

```
# Create a sample of 50 numbers which are incremented by 1.
x <- seq(0,50,by = 1)

# Create the binomial distribution.
y <- dbinom(x,50,0.5)
```

```
# Give the chart file a name.
png(file = "dbinom.png")

# Plot the graph for this sample.
plot(x,y)

# Save the file.
dev.off()
```



pbinom()

This function gives the cumulative probability of an event. It is a single value representing the probability.

```
# Probability of getting 26 or less heads from 51 tosses of a coin.
x <- pbinom(26,51,0.5)

print(x)
```

```
[1] 0.610116
```

qbinom()

This function takes the probability value and gives a number whose cumulative value matches the probability value.

```
# How many heads will have a probability of 0.25 will come out when a coin
 is tossed 51 times.

x <- qbinom(0.25,51,1/2)
print(x)
[1] 23
```

rbinom()

This function generates required number of random values of given probability from a given sample.

```
# Find 8 random values from a sample of 150 with probability of 0.4.

x <- rbinom(8,150,.4)

print(x)

[1] 58 61 59 66 55 60 61 67
```

Example:

1. Let X~Bin(5,0.9). Find (a) $P(X \leq 4)$ and $P(X = 4)$

```
(a)> sum(dbinom(0:4,5,0.9))
 [1] 0.40951

(b) > dbinom(4,5,0.9)
   [1] 0.32805
```

2. The proportion of students wearing spectacles is 40%. Let X be the number of students wearing spectacles in a random sample of 10 students. Find

(a) $P(X \leq 2)$; (b) $P(2 \leq X < 5)$; (c) $P(X > 2)$:

```
(a) > sum(dbinom(0:2,10,0.4))
   [1] 0.1672898
Or
  >pbinom(2,10,0.4)
```

[1] 0.1672898

(b)> sum(dbinom(2:4,10,0.4))
[1] 0.5867459

(c) $P(X > 2) = 1 - P(X \leq 2)$
> 1-pbin
om(2,10,0.4)
[1] 0.8327102

3. If a committee has 7 members, find the probability of having more female me mbers than male members given that the probability of having a male or a femal e member is equal.

Sol: The probability of having a female member = 0.5
The probability of having a male member = 0.5
To have more female members, the number of females should be greater than or equal to 4.

```
> 1-pbinom(3,7,0.5)
[1] 0.5
```

4. In a box of switches it is known 10% of the switches are faulty. A technician is wiring 30 circuits, each of which needs one switch. What is the probability that (a) all 30 work, (b) at most 2 of the circuits do not work?

(a) Probability that all 30 work is $P(X = 30) = {}^{30}C_{30}(0.9)^{30}(0.1)^0 = 0.04239$

(b) The statement that "at most 2 circuits do not work" implies that 28, 29 or 30 work. That is $X \geq 28$

$$\begin{aligned} P(X \geq 28) &= P(X = 28) + P(X = 29) + P(X = 30) \\ P(X = 30) &= {}^{30}C_{30}(0.9)^{30}(0.1)^0 = 0.04239 \\ P(X = 29) &= {}^{30}C_{29}(0.9)^{29}(0.1)^1 = 0.14130 \\ P(X = 28) &= {}^{30}C_{28}(0.9)^{28}(0.1)^2 = 0.22766 \end{aligned}$$

Hence $P(X \geq 28) = 0.41135$

```
> dbinom(30,30,0.9)
[1] 0.04239116
```

```
> 1-pbinom(27,30,0.9)
[1] 0.4113512
```

5. If 10% of the Screws produced by an automatic machine are defective, find the probability that out of 20 screws selected at random, there are

(i) Exactly 2 defective (ii) At least 2 defectives

(iii) Between 1 and 3 defectives (inclusive)

(i) # Exactly 2 defective

```
dbinom(2,20,0.10)
[1] 0.2851798
```

(ii) At least 2 defectives

```
1-pbinom(2,20,0.10)
[1] 0.3230732
```

(iii) Between 1 and 3 defectives (inclusive)

```
sum(dbinom(1:3,20,0.10))
[1] 0.74547
```

Poisson Distribution in R

We call it the distribution of rare events., a Poisson process is where DISCRETE events occur in a continuous, but finite interval of time or space in R

The following conditions must apply:

- For a small interval, the probability of the event occurring is proportional to the size of the interval.
- The probability of more than one occurrence in the small interval is negligible.
- Each occurrence must be independent of others and must be at random.
- The events are often defects, accidents or unusual natural happenings, such as an earthquake.
- The parameter for the Poisson distribution is a lambda. It is average or mean of occurrences over a given interval.
- The probability function is: for *x= 0,1.2,3* ....

Difference between Binomial and Poisson Distribution in R

Binomial Distribution:

- Fixed no. of Trials (n) [10 pie throws], although, only two possible outcomes are possible.
- A probability of success is constant(p).
- Each trial is independent.
- Also, it predicts no.s of successes within a set no. of trials.
- We use it to test for independence.

Poisson Distribution

- Infinite no. of trials.
- Also, it has unlimited no. of outcomes possible.
- The mean of the distribution is the same for all intervals.
- No. of occurrence in any given interval independent of others.
- Also, it predicts no. of occurrences per unit, time, space.
- We use it to test for independence.

R-Code

- dpois(x, lambda) # **the probability of x successes in a period when the expected number of events is lambda**
- ppois(q, lambda) # **the cumulative probability of less than or equal to q successes**
- qpois(p, lambda) # **returns the value (quantile) at the specified cumulative probability (percentile) p**
- rpois(n, lambda) # **returns n random numbers from the Poisson distribution**

Practice problems:

1. What is P(X = 4) with lambda 2.6?

```
> dpois(4, lambda = 2.6)
[1] 0.1414218
```

2. What is $P(X \geq 2)$ with lambda 3?

```
> 1-ppois(2,3)
```

[1] 0.5768099

2. Consider a computer system with Poisson job-arrival stream at an `
average of 2 per minute. Determine the probability that in any one-m
inute interval there will be

(i) 0 jobs
(ii) Exactly 3 jobs
(iii) at most 3 arrivals

Solution:

Job arrivals lambda =2

(i) No job arrivals

```
> dpois(0,2)
[1] 0.1353353
```

(ii) Exactly 3 jobs

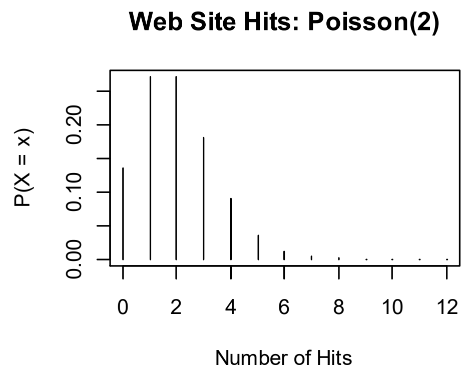
```
> dpois(3,2)
[1] 0.180447
```

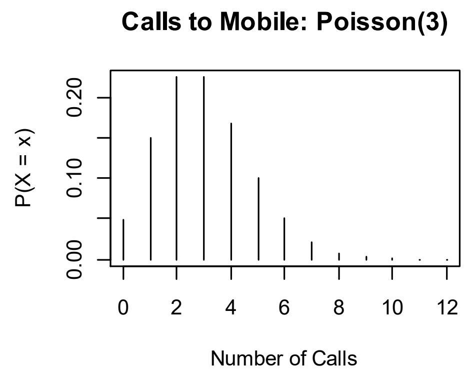
(iii) Atmost 3 job arrivals

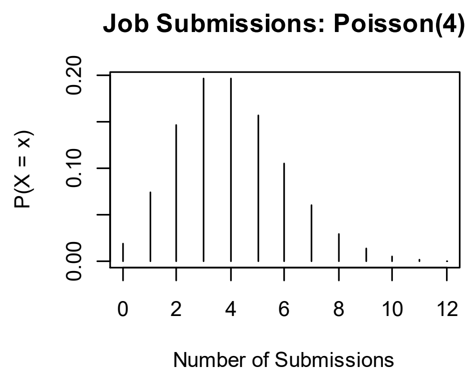
```
> ppois(3,2)
[1] 0.8571235
```

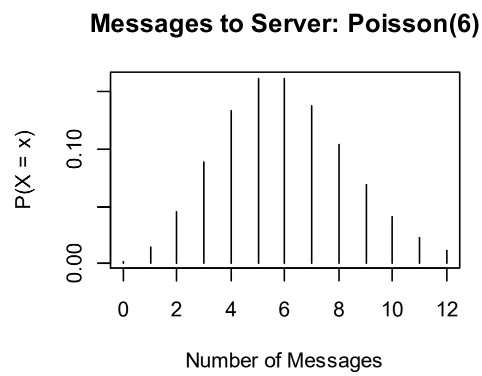
Poisson Probability Density Functions

```
par(mfrow = c(2,2))
# multiframe
x<-0:12 #look at the first 12 probabilities
plot (x, dpois(x, 2), xlab = "Number of Hits", ylab = "P(X = x)", type = "h
", main= "Web Site Hits: Poisson(2)")
plot (x, dpois(x, 3), xlab = "Number of Calls", ylab = "P(X = x)", type = "
h", main= "Calls to Mobile: Poisson(3)")
plot (x, dpois(x, 4), xlab = "Number of Submissions", ylab = "P(X = x)", t
ype = "h", main= "Job Submissions: Poisson(4)")
plot (x, dpois(x, 6), xlab = "Number of Messages", ylab = "P(X = x)", typ
e = "h", main= "Messages to Server: Poisson(6)")
```



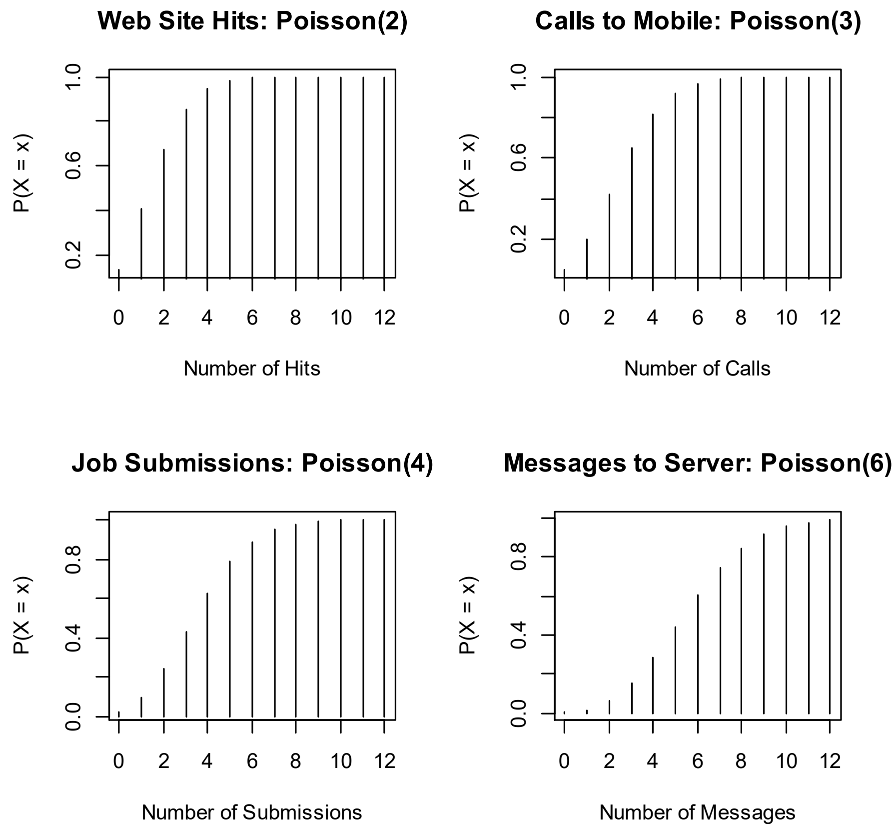






Poisson Cumulative Distribution Functions

```
par(mfrow = c(2,2))

# multiframe

x<-0:12   #look at the first 12 probabilities

plot (x, ppois(x, 2), xlab = "Number of Hits", ylab = "P(X = x)", type =
"h", main= "Web Site Hits: Poisson(2)")

plot (x, ppois(x, 3), xlab = "Number of Calls", ylab = "P(X = x)", type =
"h", main= "Calls to Mobile: Poisson(3)")

plot (x, ppois(x, 4), xlab = "Number of Submissions", ylab = "P(X = x)",
type = "h", main= "Job Submissions: Poisson(4)")

plot (x, ppois(x, 6), xlab = "Number of Messages", ylab = "P(X = x)",
type = "h", main= "Messages to Server: Poisson(6)")
```



Practice problems:

1. A recent national study showed that approximately 55.8% of college students have used Google as a source in at least one of their term papers. Let X equal the number of students in have used Google as a source:

a) Find the probability that X is equal to 17
b) Find the probability that X is at most 13.
c) Find the probability that X is bigger than 11.
d) Find the probability that X is at least 15.
e) Find the probability that X is between 16 and 19,
f) Give the mean of X
g) Give the variance of X.
h) Find $E(4X + 51.324)$