

## 10. Probability distributions

These notes consider the Chapter 10 of the [handbook](#) on various probability distributions.

### Discrete distributions

For random variables with countable number of possible values.

#### Binomial distribution

See page 200 for proper definition.

Useful when a random variable  $X$  has exactly two exclusive possible outcomes (e.g. success/fail) with known probabilities  $p \in [0, 1]$  (success) and  $q = 1 - p$  (fail). For  $n \in \mathbb{N}$  trials with  $k = 0, 1, 2, \dots, n$  successes

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

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

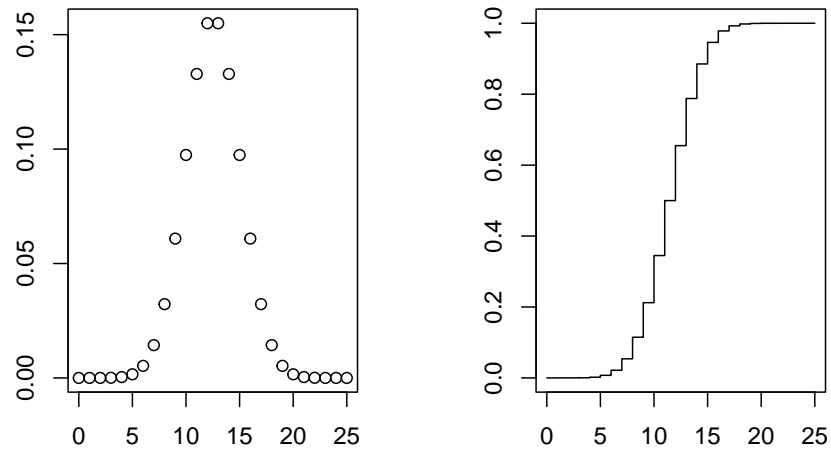
$$F(k) = P(X \leq k) = \sum_{i=0}^k \binom{n}{i} p^i q^{n-i}$$

In the following code examples, equal probabilities are assumed,  $n = \text{size}$ ,  $p = \text{prob} = 0.5$ .

```
# Sequence for visualization
binomial_seq <- seq(0, 25, by=1)

# Functions
binomial_density_function <- dbinom(x = binomial_seq, size = 25, prob = 0.5)
binomial_distribution_function <- pbinom(q = binomial_seq, size = 25, prob = 0.5)

# Plot
par(mfrow=c(1,2))
plot(binomial_seq, binomial_density_function, ann = FALSE)
plot(binomial_seq, binomial_distribution_function, type = "S", ann = FALSE)
```



```
# Probability for exactly 3 successes out of 10 trials
dbinom(x = 3, size = 10, prob = 0.5)
```

```
## [1] 0.1171875
```

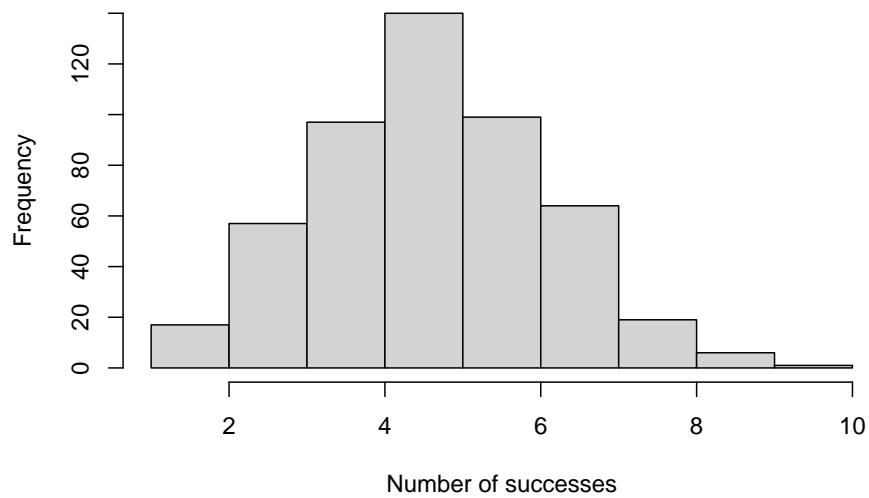
```
# Probability for up to 3 successes out of 10 trials
pbinom(q = 3, size = 10, prob = 0.5)
```

```
## [1] 0.171875
```

```
# Simulate 10 times how many successes there is using random numbers
rbinom(n = 10, size = 10, prob = 0.5)
```

```
## [1] 4 5 6 5 3 8 4 5 3 6
```

```
# With large enough n, expected value (np = 5) should become visible
rbinom500 <- rbinom(n = 500, size = 10, prob = 0.5)
hist(rbinom500, xlab = "Number of successes", ylab = "Frequency", main = NULL)
```



```
summary(rbinom(500))
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##      1.000  4.000   5.000   5.092  6.000   10.000
```

## Poisson distribution

See page 204 for proper definition.

Useful when estimating amounts in random processes where the expected value ( $\lambda$ ) is known

$$X \sim Poi(\lambda)$$

$$P(X = k) = \frac{\lambda^k e^{-\lambda}}{k!}, \quad k = 0, 1, 2, \dots, n$$

Note that the Poisson distribution can be used to approximate the binomial distribution when  $n$  is large and  $p$  is small. In this case,  $\lambda = np$  and

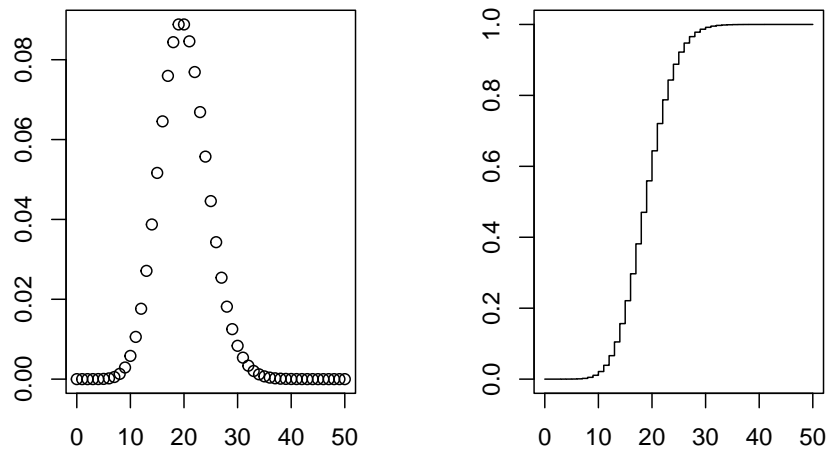
$$P(X = k) = \frac{(np)^k e^{-np}}{k!}$$

```
# Sequence for visualization
Poisson_seq <- seq(0, 50, by=1)
```

```
# Functions
```

```
Poisson_density_function <- dpois(x = Poisson_seq, lambda = 20)
Poisson_distribution_function <- ppois(q = Poisson_seq, lambda = 20)
```

```
# Plot
par(mfrow=c(1,2))
plot(Poisson_seq, Poisson_density_function, ann = FALSE)
plot(Poisson_seq, Poisson_distribution_function, type = "S", ann = FALSE)
```



```
# A bridge is crossed by 12 people per minute, on average.
# What is the probability that exactly 16 people crosses it in a minute?
dpois(x = 16, lambda = 12)
```

```
## [1] 0.05429334
```

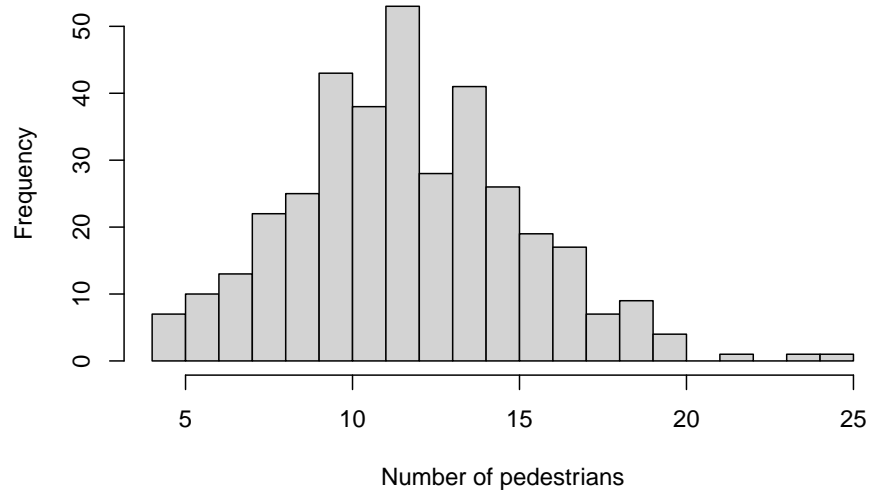
```
# A bridge is crossed by 12 people per minute, on average.
# What is the probability that up to 16 people crosses it in a minute?
ppois(q = 16, lambda = 12)
```

```
## [1] 0.898709
```

```
# Simulate amount of people per minute 10 times
rpois(n = 10, lambda = 12)
```

```
## [1] 17 10 13 17 7 12 11 5 10 15
```

```
# By repeating the observation once every day for a year,
# the expected value (lambda = 12) should become visible
Poisson_pedestrians <- rpois(n = 365, lambda = 12)
hist(Poisson_pedestrians, breaks = 20,
     xlab = "Number of pedestrians", ylab = "Frequency", main = NULL)
```



```
summary(Poisson_pedestrians)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##      4.00   10.00   12.00   12.16   14.00   25.00
```