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 $k = 0, 1, 2, \ldots, n$ successes

$$X \sim Bin(n, p)$$

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

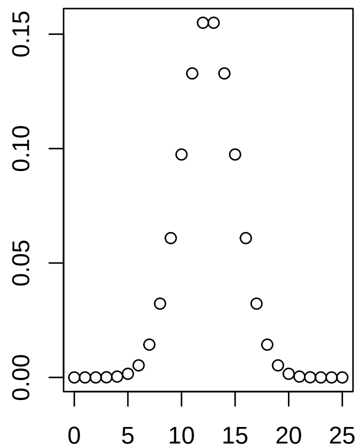
$$\mathrm{F}(k) = \mathrm{P}(X \leqslant k) = \sum_{i=0}^{k} \binom{n}{i} p^i q^{n-i}$$

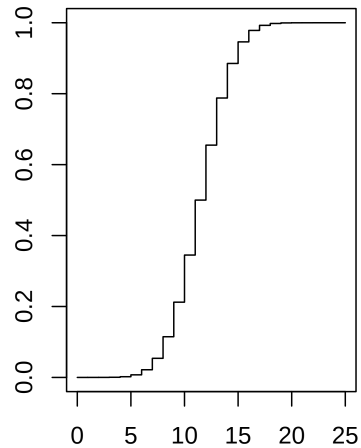
In the following code examples, equal probabilities are assumed, $n =$ **size**, $p =$ **prob** $= 0.5$.

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

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

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






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

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

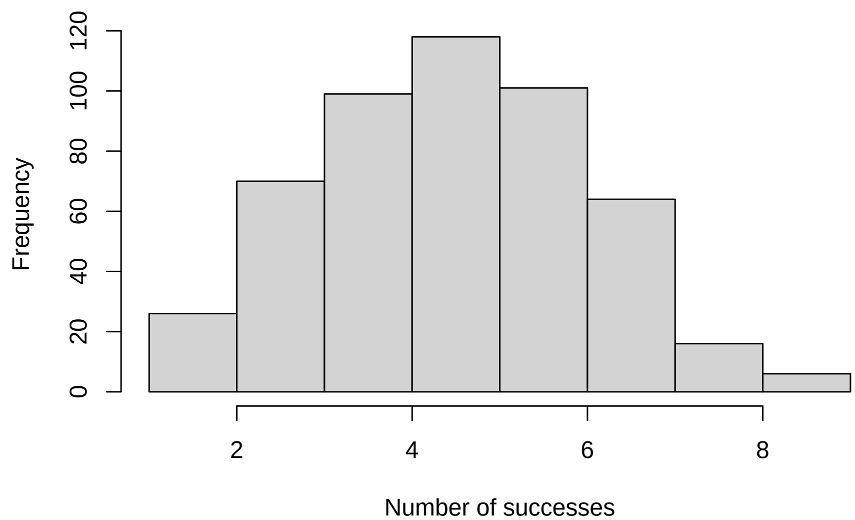
```r
# Probability for up to 3 successes out of 10 trials
pbinom(q = 3, size = 10, prob = 0.5)
```

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

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

```
##  [1] 7 8 4 7 4 4 4 5 7 5
```

```r
# 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)
```




```r
summary(rbinom500)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max. 
##    1.00    4.00    5.00    4.96    6.00    9.00
```

Poisson distribution

See page 204 for proper definition.

Useful when estimating amounts in random processes where the expected value (λ) is known

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

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

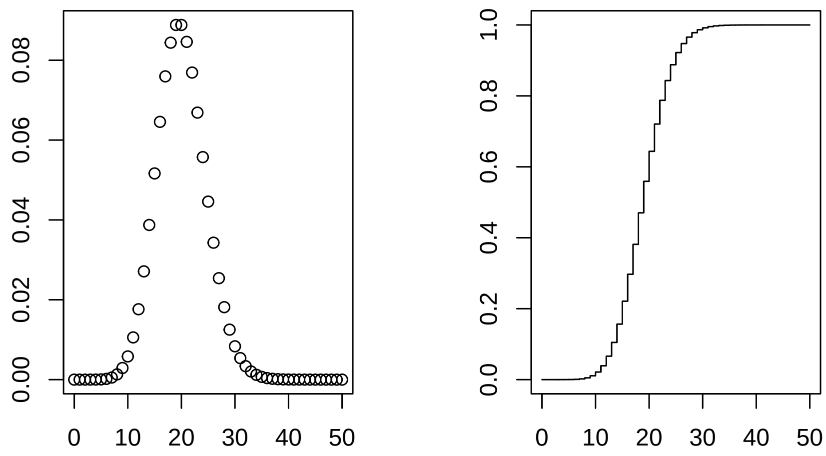
Note that the Poisson distribution can be used to approximate the binomial distribution when n is large and p is small. In this case, $\lambda = np$ and

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

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

# Functions
poisson_pmf <- dpois(x = poisson_seq, lambda = 20)
poisson_cdf <- ppois(q = poisson_seq, lambda = 20)
```

```r
# Plot
par(mfrow = c(1, 2))
plot(poisson_seq, poisson_pmf, ann = FALSE)
plot(poisson_seq, poisson_cdf, type = "S", ann = FALSE)
```




```r
# 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
```

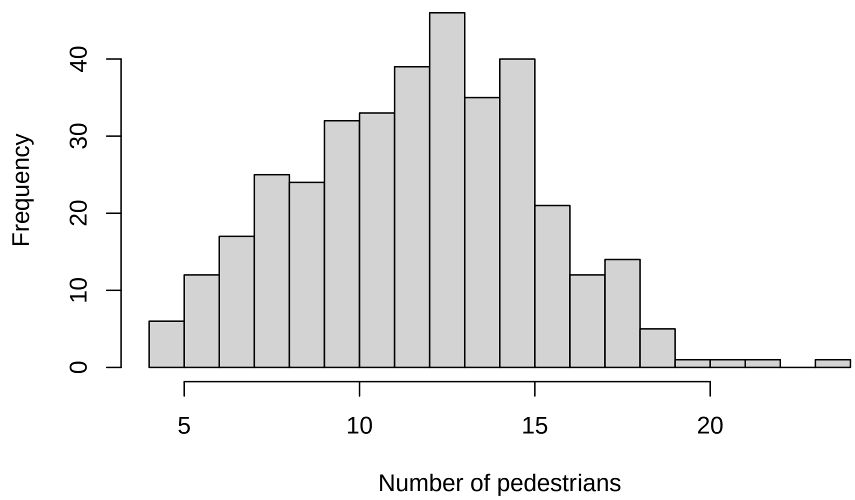
```r
# A bridge is crossed by 12 people per minute, on average.
# What is the probability that up to 16 people crosses it in a minute?
ppois(q = 16, lambda = 12)
```

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

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

```
##  [1] 11 15  7 11  6 10 13  8 19  9
```

```r
# 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)
```




```r
summary(poisson_pedestrians)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##    4.00   10.00   12.00   12.18   15.00   24.00
```