Review of PDF and R

Emanuele Coradin

March 2024

Contents

1 Pairing and ordering objects

We can identify different useful tools to compute the numbers of different pairings as orderings that we can apply to a sequence of length r taken from a dictionary of n objects.

2 Probability distributions in R

R provides almost all the standard PDFs that we could wish. The name convention prescribes: * d for the pdf * p for the cumulative density function (cdf) * q for the quantile function * r to sample a random number from the distribution.

Now we will experiment with them with same simple exercise.

2.1 Binomial distribution

The Binomial probability function is described by: $P(X = k) = \binom{n}{k} \cdot p^k \cdot (1 - p)^{n-k}$

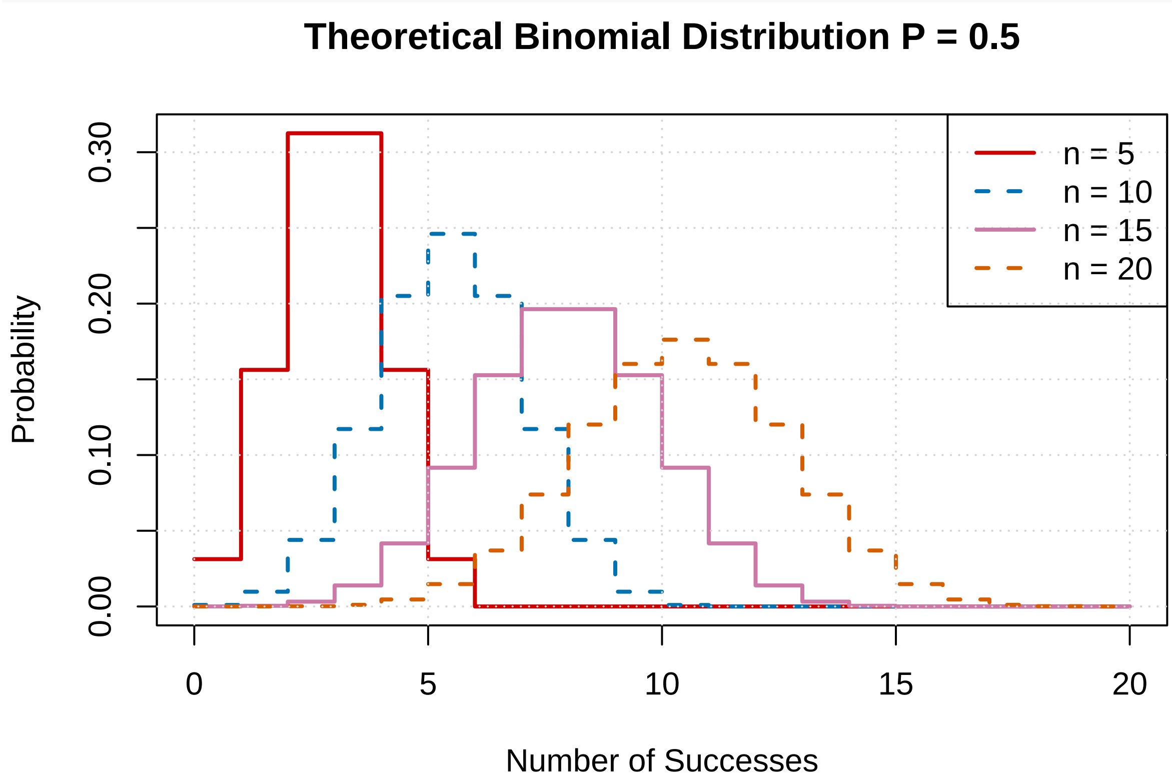
It's the probability followed by the sum of n indipendent Bernoulli trials

First of all let's define a palette.

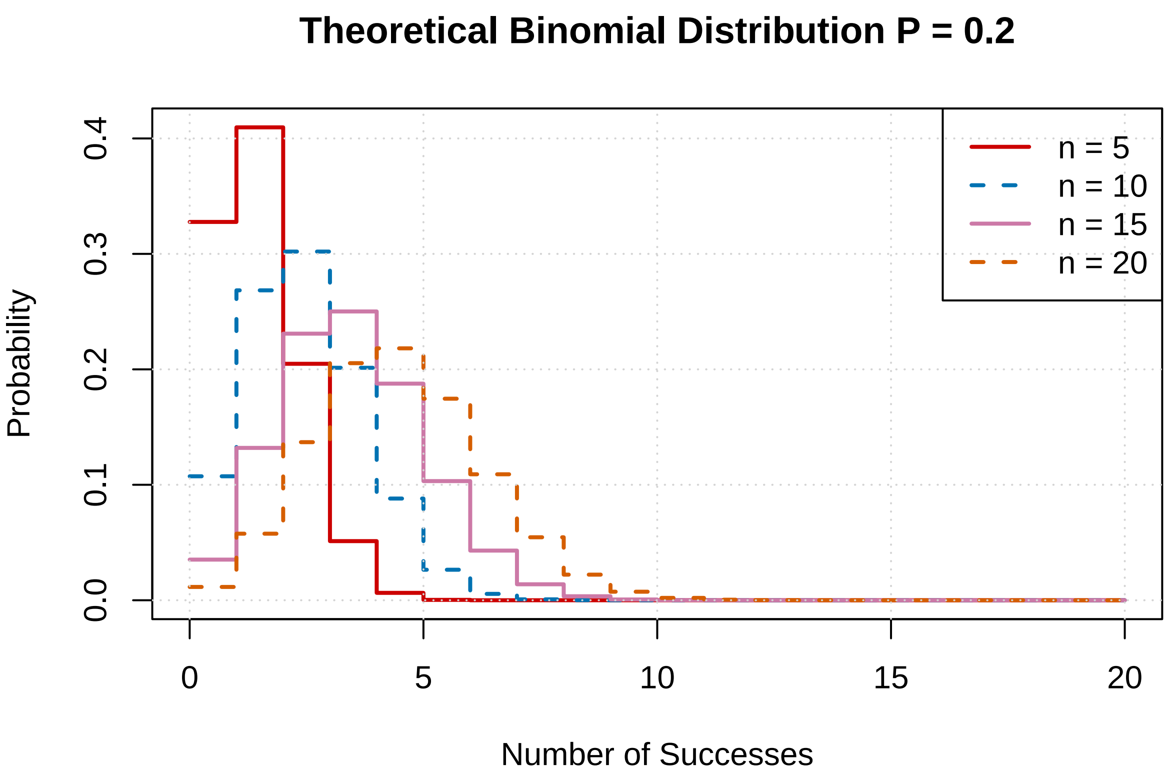
```
color_vector <- c("#CC0000",   # Wine-like color
                  "#0072B2",   # Strong blue
                  "#CC79A7",   # Muted purple
                  "#D55E00",   # Vermilion
                  "#009E73",   # Bluish green
                  "#56B4E9",   # Sky blue
                  "#E69F00")   # Yellow-orange
```

```
# Parameters for the binomial distribution

n <- c(5, 10, 15, 20)  # number of trials
linetype <- 2-(1:length(n))%%2
```

```r
p <- 1./2. # probability of success

# Generate values for x (number of successes)
x <- 0:20

#define a function to prepare the plots
plot_pdf <- function(N) {dbinom(x, N, p)}
binomial_pdf <- sapply(n, plot_pdf)


# create the first straight line plot
plot(x,binomial_pdf[,1], type = "s", lwd = 2, col = color_vector[1],
     xlab = "Number of Successes", ylab = "Probability", lty = linetype[1])

# create all the others
for (i in 2:length(n)) {
  lines(x, binomial_pdf[, i], col=color_vector[i], type='s', lwd=2, lty = linetype[i])

}

grid()  # Add grid

# Create legend labels with strings "n = "
legend_labels <- paste("n =", n)
# Add a legend
legend("topright", legend = legend_labels, col = color_vector[1:length(n)], lty = linetype, lwd = 2)
# Add a title
title("Theoretical Binomial Distribution P = 0.5")
```



```r
# Parameters for the binomial distribution

n <- c(5, 10, 15, 20)  # number of trials
linetype <- 2-(1:length(n))%%2
p <- 1./5. # probability of success

# Generate values for x (number of successes)
x <- 0:20

#define a function to prepare the plots
plot_pdf <- function(N) {dbinom(x, N, p)}
binomial_pdf <- sapply(n, plot_pdf)


# create the first straight line plot
plot(x,binomial_pdf[,1], type = "s", lwd = 2, col = color_vector[1],
     xlab = "Number of Successes", ylab = "Probability", lty = linetype[1])

# create all the others
for (i in 2:length(n)) {
  lines(x, binomial_pdf[, i], col=color_vector[i], type='s', lwd=2, lty = linetype[i])

}

grid()  # Add grid

# Create legend labels with strings "n = "
legend_labels <- paste("n =", n)
# Add a legend
legend("topright", legend = legend_labels, col = color_vector[1:length(n)], lty = linetype, lwd = 2)
# Add a title
title("Theoretical Binomial Distribution P = 0.2")
```



2.2 Geometric Distribution

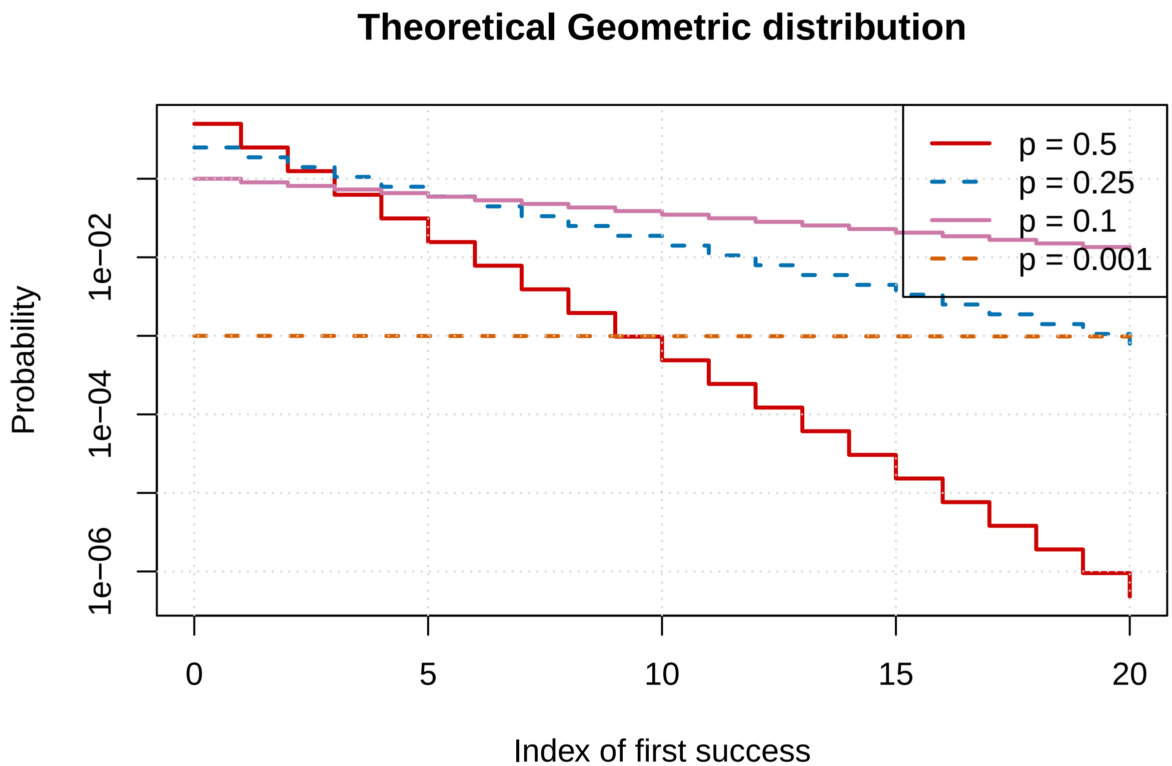
The Geometric distribution can be described by: $P(X = k) = (1 - p)^{k-1} \cdot p$

It gives us the number of trials to get the first success

Now let's visualize it.

```
# Parameters for the binomial distribution

p <- c(1./2., 1/4, 1/10, 10**(-3)) # probability of success
linetype <- 2-(1:length(p))%%2

# Generate values for x (number of successes)
x <- 0:20

#define a function to prepare the plots
plot_pdf <- function(P) {dgeom(x, P)}
geom_pdf <- sapply(p, plot_pdf)


# create the first straight line plot
plot(x,geom_pdf[,1], type = "s", lwd = 2, col = color_vector[1],
     xlab = "Index of first success", ylab = "Probability", lty = linetype[1], log="y")

# create all the others
for (i in 2:length(p)) {
  lines(x, geom_pdf[, i], col=color_vector[i], type='s', lwd=2, lty = linetype[i])

}
```

```r
grid()  # Add grid

# Create legend labels with strings "p = "
legend_labels <- paste("p =", p)
# Add a legend
legend("topright", legend = legend_labels, col = color_vector[1:length(p)], lty = linetype, lwd = 2)
# Add a title
title("Theoretical Geometric distribution")
```



2.3 Poisson Process

2.4 Poisson Distribution

It' described by:

It can be derived by the Binomial distribution when the rate of success is close to 0.

```r
# Parameters for the binomial distribution

lambda <- c(5, 10, 15, 20)  # number of trials
linetype <- 2-(1:length(lambda))%%2

# Generate values for x (number of successes)
x <- 0:20

#define a function to prepare the plots
plot_pdf <- function(L) {dpois(x, L)}
poisson_pdf <- sapply(lambda, plot_pdf)


# create the first straight line plot
```

```r
plot(x,poisson_pdf[,1], type = "s", lwd = 2, col = color_vector[1],
     xlab = "Number of Events", ylab = "Probability", lty = linetype[1])

# create all the others
for (i in 2:length(lambda)) {
  lines(x, poisson_pdf[, i], col=color_vector[i], type='s', lwd=2, lty = linetype[i])

}

grid()  # Add grid

# Create legend labels with strings
legend_labels <- paste("lambda = ", lambda)
# Add a legend
legend("topright", legend = legend_labels, col = color_vector[1:length(lambda)], lty = linetype, lwd = 
# Add a title
title("Theoretical Poisson Distribution")
```