

The screenshot shows the RStudio interface with the following details:

- Toolbar:** Includes standard icons for file operations (New, Open, Save, Print, Find, Copy, Paste), Go to file/function, and Addins.
- Project Explorer:** Shows two files: Exercise1-polya.R and Exercise2-polya.R (the current active file).
- Code Editor:** Displays the R code for a Polya random walk. The code uses color-coded syntax highlighting where blue represents numbers and functions, green represents comments, and black represents variables and operators.
- Toolbars:** Includes Source on Save, a magnifying glass for search, and a pencil for edit.
- Buttons:** Run, Source, and other standard RStudio buttons.

```
1 # POLYA RANDOM WALK
2 n <- 1000 # number of iterations
3 alpha0 <- 1 # the parameter alpha0 of the Polya process
4 alpha1 <- 10 # the parameter alpha1 of the Polya process
5 alpha <- alpha0+alpha1 # the parameter alpha of the Polya process
6 m0 <- 0 # the initial number of failures
7 m1 <- 0 # the initial number of successes
8 Y <- rep(0,times=n)# definition of the r.v. Y
9 # start the main cycle
10 for (m in 0:(n-1)){
11   if ( runif(1) < (alpha1+m1)/(alpha+m)) { # comparison with the predictive probability of success
12     Y[m+1] <- +1 # upward jump
13     m1 <- m1+1 # update total number of successes
14   } # end if
15   else { # else a failure
16     Y[m+1] <- -1 # downward jump
17     m0 <- m0+1 # update total number of failures
18   } # end else
19 } # end for
20 S <- cumsum(Y) # implementing Eq.(46) of notes Scalas2 (definition of Polya random walk)
21 t <- c(1:n) # time steps
22 plot(t,S,type="l",xlab="m",ylab="S(m)",main="Polya random walk")
```

The screenshot shows the RStudio interface with the following details:

- Toolbar:** Standard RStudio icons for file operations (New, Open, Save, Print, etc.), Go to file/function, and Addins.
- File Tabs:** Two tabs are open: "Exercise1-polya.R" and "Exercise2-polya.R".
- Code Editor:** The "Exercise1-polya.R" tab contains the following R code:

```
1 - #####
2 # PROGRAM POLYA COMPARISON WITH BINOMIAL #
3 #####
4
5 #####
6 # Parameters #
7 #####
8 N<-10000 #number of realizations
9 m<-10 #total number of individuals/observations
10 alpha0<-100 #parameter of category 0 (failure)
11 alpha1<-100 #parameter of category 1 (success)
12 alpha<-alpha0+alpha1 #parameter alpha
13 #
14 M1<-rep(c(0),times=(m+1)) #vector storing MC results
15 #
16 for(i in 1:N){ #cycle of realizations
17   m0<-0 #initialization of the number of failures
18   m1<-0 #initialization of the number of successes
19   #
20   for(k in 1:m){#cycle on individuals/observations
21     if ( runif(1) < (alpha1+m1)/(alpha+k) ) { #predictive probability for success
22       m1<-m1+1 #update of successes number
23     }#end if
24     else {
25       m0<-m0+1 #update of failures number
26     }#end else
27   }#end for (k)
28   #
29   M1[m1+1]<-M1[m1+1]+1 #storing the number of successes
```

The code implements a Monte Carlo simulation to compare Polya's urn model with a binomial distribution. It initializes parameters (N=10000, m=10, alpha0=100, alpha1=100, alpha=200), creates a vector M1 of length m+1, and then iterates over N cycles. Each cycle consists of m observations. For each observation k, it generates a random number between 0 and 1. If the number is less than the predictive probability for success ($\frac{\alpha_1 + m_1}{\alpha + k}$), it increments the success count m1. Otherwise, it increments the failure count m0. Finally, it updates the total number of successes in M1.

```
#M1[k+1] at the end of the for cycle gives the number of iterations in which k successes happened
}#end for (i)
M1[1:(m+1)] #displays simulation values
k<-seq(0,10) #creates vector of integers from 0 to 10
N*dbinom(k,m,0.5) # displays binomial prediction

#####
# Plots of the values #
#####
Propor<-(N)*dbinom(k,m,0.5) #proportion of the binomial distribution having parameters m and 0.5
plot(k,M1[1:11], ylab="number of successes", xlab="number of observations") #plot of the number of successes
points(k,Propor,type="p",pch=3) #plot of the mean of the binomial distribution
legend(cex=1,"topright",c("Polya", "Binomial"), pch=c(1,3))
```

Birthday problem

lunedì 14 giugno 2021 13:25

There are n persons in a room, $n < 365$.

Ob: What is the prob. that at least two of them have a common birthday?

$$\begin{cases} n = \text{number of persons in the room} \\ g = 365 \end{cases}$$

$X_i = x_i \leftrightarrow$ the i -th person is born on the x_i -th day of the year

For ex. $X_i = 1 \leftrightarrow$ the i -th person is born on Jan 1st

$X_i = 365 \leftrightarrow$ the i -th person is born on Dec. 31st

$\rightarrow (X_1 = x_1, \dots, X_m = x_m)$ 365^m possible individual descriptions
 $\underbrace{}_{365 \text{ possibilities}}$ $\underbrace{}_{365 \text{ possibilities}}$

$\rightarrow (Y_1 = m_1, \dots, Y_g = m_g)$ $Y_i = m_i \leftrightarrow$ there are m_i persons born
on the i -th day of
the year
FREQUENCY VECTOR

$$\rightarrow (Z_0 = z_0, \dots, Z_m = z_m)$$

$Z_i = z_i \leftrightarrow$ there are z_i days with i birthdays

PARTITION VECTOR

m_0 common birthday

complementary situation (simpler to analyse)

$$\longleftrightarrow (Z_0 = 365 - m, Z_1 = m, Z_2 = 0, \dots, Z_m = 0)$$

From eq. (6) of the notes Scalars

$$(*) P(Y_1 = m_1, \dots, Y_g = m_g) = \frac{m!}{m_1! \dots m_g!} \cdot P(X_1 = x_1, \dots, X_m = x_m)$$

In a similar way, one can show that

$$(**) P(Z_0 = z_0, Z_1 = z_1, \dots, Z_m = z_m) = \frac{g!}{z_0! z_1! \dots z_m!} \cdot P(Y_1 = m_1, \dots, Y_g = m_g)$$

From eqs (*) and (**)

$$(1) P(Z_0 = z_0, \dots, Z_m = z_m) = \frac{g!}{z_0! z_1! \dots z_m!} \cdot \frac{m!}{m_1! \dots m_g!} \cdot \underbrace{P(X_1 = x_1, \dots, X_m = x_m)}$$

let's compute this probability

We assume that X_1, \dots, X_m are i.i.d. random variables and that all the days are equiprobable, so that $P(X_i = x_i) = \frac{1}{365}$ for any x_i .

Hence

$$P(X_1 = x_1, \dots, X_m = x_m) = \prod_{i=1}^m P(X_i = x_i) = \left(P(X_i = x_i) \right)^m = \frac{1}{365^m}$$

↓
because of the
independence

↓
because of the
identical distribution,
for a fixed $i \in \{1, \dots, m\}$

So from (1)

$$\begin{aligned} P(Z_0 = 365-m, Z_1 = m, Z_2 = 0, \dots, Z_m = 0) &= \frac{365!}{(365-m)! \underbrace{m! \cdot 0! \cdots 0!}_{\substack{m \text{ times}}}} \cdot \frac{\overbrace{m!}^{\substack{m \text{ times}}}}{\underbrace{1! \cdots 1! \cdot 0! \cdots 0!}_{\substack{365-m \text{ times}}}} \cdot \frac{1}{365^m} \\ &= \frac{365!}{(365-m)! \cdot 365^m} \end{aligned}$$

Hence, the probability of having at least two common birthdays is

$$p = 1 - P(Z_0 = 365-n, Z_1=n, Z_2=0, \dots, Z_m=0) = \\ = 1 - \frac{365!}{(365-n)! 365^n}.$$

If $n \geq 23$, then $p > 0.5$ (See the screen below...)

