Resolução dos desafios da semana 02

Arthur Klemenchuk[*], Victor Tanaka[†], Igor Bragaia[‡]
Instituto Tecnológico de Aeronáutica,
12228-900 São José dos Campos, São Paulo, Brasil

14 de agosto de 2017

Resumo

No presente trabalho, solucionamos os desafios propostos na semana 02 do curso de MOQ-13. Primeiramente, resolvemos ambos os problemas de modo analítico usando o conceito clássico de probabilidade. Em seguida, utilizando o software R, construímos simulações computacionais para solucionar os problemas. Ambos os resultados obtidos foram comparados.
Palavras-chave: Probabilidade, simulação, R.

1 Introdução

O conceito de probabilidade possui diversas interpretações, as quais diferem-se nas hipóteses acerca dos eventos tratados e no método de cálculo utilizado. Dentre elas, destacam-se as definições clássica (a priori) e empírica (a posteriori). A primeira afirma que a probabilidade de um evento A é dada por

$$P_N(A) = \frac{n_A}{N}, \tag{1}$$

em que n_A é o número de resultados favoráveis e, N, o número total de resultados possíveis. Estes devem ser equiprováveis, mutuamente exclusivos e coletivamente exaustivos. Nota-se que, segundo essa abordagem, probabilidades podem ser calculadas por meio da determinação teórica de n_A e N.

Diferentemente, a definição empírica abandona tais hipóteses e interpreta a probabilidade do evento A pelo quociente

$$P_N(A) = lim_{N\to\infty}\frac{n_A}{N}, \tag{2}$$

sendo n_A o número de vezes em que A foi observado e, N, o número de repetições do experimento. Com isso, faz-se necessário realizar uma simulação ou um experimento para determinar a probabilidade de um evento pela definição em questão. O experimento deve ser repetido muitas vezes a fim de obter resultados uniformes, o que constitui o *princípio da regularidade estatística*.

Diante das diferenças entre as definições apresentadas, este trabalho visa calcular probabilidades segundo ambos os métodos e comparar os resultados obtidos. Para isso, 2 problemas, apresentados na seção 2, serão resolvidos.

[*] `a.klemenchuk98@gmail.com`
[†] `victorheiji@gmail.com`
[‡] `igor.bragaia@gmail.com`

2 Enunciado dos problemas

O primeiro problema resolvido baseia-se na seguinte obervação feita por Galileu Galilei: "*Embora o número de triplas de números inteiros de 1 a 6 com soma 9 seja igual ao número de tais triplas com soma 10, quando três dados são lançados, parece ser menos frequente observar um total de 9 do que um total de 10 – na experiência de jogadores*". Deve-se escrever um programa em R para simular o lançamento de três dados. O experimento deve ser repetido várias vezes, pelo princípio da regularidade estatística. A frequência relativa para observação da soma dos resultados ser 9 ou 10 deve ser anotada. A partir da simulação, deve-se analisar se a conjectura dos jogadores estava correta.

O segundo problema envolve a seguinte situação: Romeu e Julieta combinaram um encontro secreto entre as 12h e 13h, escolhendo aleatoriamente o tempo de chegada. Romeu chega, espera 10 minutos e vai embora. Julieta faz o mesmo, mas espera 12 minutos antes de ir embora. Deve-se escrever um programa em R para simular o panorama apresentado e determinar a probabilidade de Romeu e Julieta se encontrarem.

Além dos resultados obtidos através das simulações em R, também apresentamos os resultados analíticos obtidos pela definição clássica de probabilidade.

3 Método de solução

3.1 Desafio 1

Apresentamos a solução por meio da definição clássica de probabilidades, para um espaço amostral de resultados igualmente prováveis, mutuamente exclusivos e coletivamente exaustivos.

Vamos determinar os casos favoráveis:

- Soma das faces dos três dados igual a 9:
  O evento será

$$\Omega_9 = \{(1,2,6),(1,3,5),(1,4,4),(2,2,5),(2,3,4),(3,3,3)\}. \tag{3}$$

  Para cada um dos casos acima incluem-se também todas as permutações na ordem dos dados. Assim, a quantidade de casos será:

$$n_9 = 3 \cdot 3! + 2 \cdot \frac{3!}{2!} + 1\frac{3!}{3!} = 25. \tag{4}$$

  Finalmente,

$$P(\Omega_9) = \frac{25}{6^3}. \tag{5}$$

- Soma das faces dos três dados igual a 10:
  Temos o evento

$$\Omega_{10} = \{(1,3,6),(1,4,5),(2,2,6),(2,3,5),(2,4,4),(3,3,4)\}. \tag{6}$$

  Para cada um dos casos acima incluem-se também todas as permutações na ordem dos dados. Assim, a quantidade de casos será:

$$n_{10} = 3 \cdot 3! + 3 \cdot \frac{3!}{2!} = 27. \tag{7}$$

Finalmente,

$$P(\Omega_{10}) = \frac{27}{6^3}. \tag{8}$$

A simulação computacional consistiu num programa que simula o lançamento de três dados. Para isso, sorteiam-se 3 números inteiros aleatórios entre 1 e 6 e repete-se o mesmo experimento quantas vezes forem desejadas. Também verificou-se a quantidade de vezes em que a soma dos resultados foi 9 ou 10. A partir dessa informação, determinou-se a frequência relativa para os dois casos favoráveis apontados acima. Esses resultados foram plotados e comparados com (5) e (8). Como exemplo, simulou-se a repetição do experimento 50000 vezes. O código em R utilizado é:

```
dice <- function(n,plot=TRUE){

#
# Input:
# n    = Maximum amount of meetup tries
# prob = probabily of getting dices sum equal to 9 or 10
#
# Output:
# Chart: Relative frequency vs. Amount of tries
#

gameTheoretical9 <- 25/6^3
gameTheoretical10 <- 27/6^3

freq9 <- vector(length = n)
game9 <- vector(length = n)

freq10 <- vector(length = n)
game10 <- vector(length = n)

for(i in 1:n) {
game9[i] <- if (sum(sample(1:6,3,replace=TRUE)) == 9 ) 1 else 0
freq9[i] <- sum(game9[1:i])/i

game10[i] <- if (sum(sample(1:6,3,replace=TRUE)) == 10 ) 1 else 0
freq10[i] <- sum(game10[1:i])/i
}

if (plot == TRUE){
plot(freq9, main ="",
cex.lab = 2,
cex.axis = 2,
xlab = "amount of tries",
ylab = "relative frequency",
ylim = c(0.1,0.3),
type = "l", lwd = 2, col = "red")
lines(freq10, lwd = 2, col = "blue")

abline(h = gameTheoretical9, lty = "dashed", col = "red")
abline(h = gameTheoretical10, lty = "dashed", col = "blue")
```

```
# Legend
leg.tex <- c(paste("P[sum == 9]=", round(gameTheoretical9,
   digits=4)),
paste("P[sum == 10]=", round(gameTheoretical10, digits=4)))
legend("topright", leg.tex, col = c(2,4), lty = c(1,1), bty = "n",
   inset = 0.1, cex = 2.5)
mtext("dices", line = 1, adj =0, cex = 2)
}
}

## EXAMPLES:

# 50000 tries
dice(50000)
```

3.2 Desafio 2

Para a solução analítica, vamos denotar por R e J os instantes de tempo de chegada de Romeu e Julieta, respectivamente. Por simplicidade, consideramos a origem de tempo às 12 h, de modo que R e j pertencem ao intervalo $[0, 60]$ (em minutos). Temos os seguintes casos favoráveis:

- Romeu chega primeiro:
  Após sua chegada, ele espera por mais 10 min. Logo, para haver encontro, devemos ter
$$J - R \leq 10 \Rightarrow R \geq J - 10. \tag{9}$$
- Julieta chega primeiro:
  Como seu tempo de espera é 12 min, para haver encontro devemos ter:
$$R - J \leq 12 \Rightarrow R \leq 12 + J. \tag{10}$$

Representando R e J num plano cartesiano, o sistema de inequações (9) e (10) corresponde à área entre as retas $R = J - 10$ e $R = 12 + J$, representadas, respectivamente, em azul claro e laranja na Figura 1. Como $0 \leq R, J \leq 60$, o espaço amostral corresponde ao quadrado $[0, 60] \times [0, 60]$, mostrado em azul na Figura 1. Diante desses vínculos, a região favorável ao encontro é a área vermelha na Figura 1.

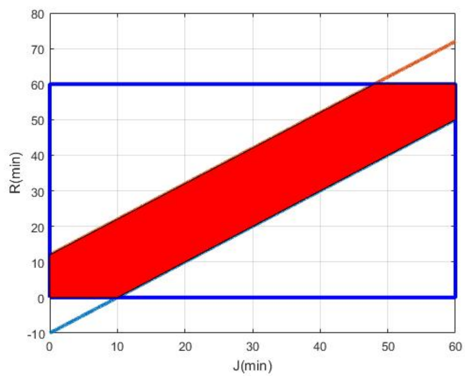

Figura 1: Representação geométrica do desafio 2.

Por fim, a probabilidade de ocorrer o encontro é a razão entre a área vermelha e a do quadrado azul:

$$P(E) = \frac{60^2 - \frac{48^2}{2} - \frac{50^2}{2}}{60^2} = 33,27\%. \tag{11}$$

A solução empírica consistiu num programa que sorteia dois números reais entre 0 e 60, um correspondente ao horário de chegada de Romeu e, outro, ao de Julieta. Em seguida, o programa verifica se houve encontro, testando as condições relativas ao tempo de espera de cada personagem. Repetindo-se o experimento, determina-se a frequência relativa de ocorrer encontro. Os resultados são plotados e comparados com (11). O código em R utilizado é

```
meetup <- function(n,plot=TRUE){

#
# Input:
# n    = Maximum amount of meetup tries
# prob = probabily of meeting
#
# Output:
# Chart: Relative frequency vs. Amount of tries
#

meetupTheoretical <- (60^2-48^2/2-50^2/2)/60^2

freq <- vector(length = n)
meetup <- vector(length = n)
for(i in 1:n){
timeRomeo <- runif(1, 0, 60)
timeJuliet <- runif(1, 0, 60)
meetup[i]<- if (timeRomeo>timeJuliet) ( if
    (abs(timeRomeo-timeJuliet)<=10) 1 else 0) else (if
    (abs(timeRomeo-timeJuliet)<=12) 1 else 0)

freq[i]<-sum(meetup[1:i])/i
}

if (plot == TRUE){
plot(freq, main ="",
cex.lab = 2,
cex.axis = 2,
xlab = "amount of tries",
ylab = "relative frequency",
ylim = c(0,1),
type = "l", lwd = 2, col = "red")

abline(h = meetupTheoretical, lty = "dashed", col = "red")

# Legend
```

```
leg.tex <- c(paste("P[meetup]=", round(meetupTheoretical, digits=4)))
legend("topright", leg.tex, col = c(2,4), lty = c(1,1), bty = "n",
    cex=2, inset = 0.1)
mtext("Romeo and Juliet meet each other!", line = 1, adj =0, cex = 2)
}
}


## EXAMPLES:

# 50000 tries
meetup(50000)
```

4 Resultados e discussões

4.1 Desafio 1

Pelos cálculos (5) e (8), tem-se que a probabilidade de obter a soma dos três dados igual a 10 é maior que a soma 9, logo a observação de Galileu Galilei estava correta. O problema que Galileu se propôs a resolver surge, pois, embora exista uma mesma quantidade de ternos (i, j, k) com soma 9 ou 10, eles não são equiprováveis. Logo, segundo a abordagem de Galileu, não é possível utilizar a definição clássica de probabilidade. Com o cálculo apresentado neste trabalho, o espaço amostral foi escolhido de modo a conter eventos equiprováveis, possibilitando o uso daquela definição.

Os resultados da simulação em R para 50000 tentativas está na Figura 2. O gráfico mostra a frequência relativa para a soma dos dados igual a 9 (linha em vermelho) e igual a 10 (linha em azul). As linhas tracejadas correspondem ao resultado clássico para a probabilidade dos respectivos eventos. Nota-se que, para um número pequeno de lançamentos (menor que 20000), as frequências relativas variam muito e não aparentam convergir. Porém, observa-se que, após 30000 tentativas, os resultados gráficos obtidos tornam-se essencialmente constantes e convergem para o resultado clássico, comprovando o princípio da regularidade estatística. Deste modo, a conjectura de Galileu estava, de fato, correta.

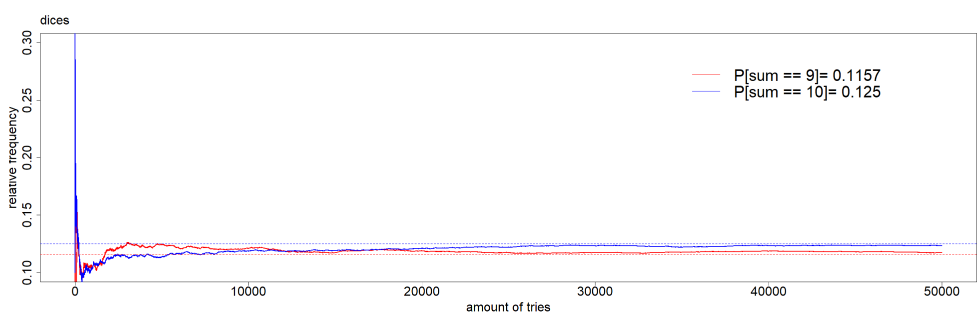

Figura 2: Frequência relativa vs. Número de lançamentos para três dados.

5 Desafio 2

De acordo com os cálculos analíticos, a probabilidade de Romeu e julieta se encontrarem é $33,27\%$. A Figura 3 expõe os resultados da simulação para 50000 tentativas. No gráfico, tem-se a frequência relativa em que acontece o encontro entre Romeu e Julieta, para 50000 tentativas. A linha tracejada corresponde à probabilidade de encontro prevista de modo analítico. Novamente, observa-se que a frequência relativa oscila muito quando o número de tentativas é pequeno (inferior a 5000 nesse caso). Entretanto, no experimento em questão, a convergência dos resultados gráficos para a solução que fez uso de probabilidade clássica ($33,27\%$) se deu de maneira muito mais rápida. A partir de 10000 repetições não observa-se nenhuma oscilação no gráfico. Este fato ressalta a imprecisão em definir o que é um número suficiente grande de repetições, o qual é uma noção necessária ao conceito empírico de probabilidade.

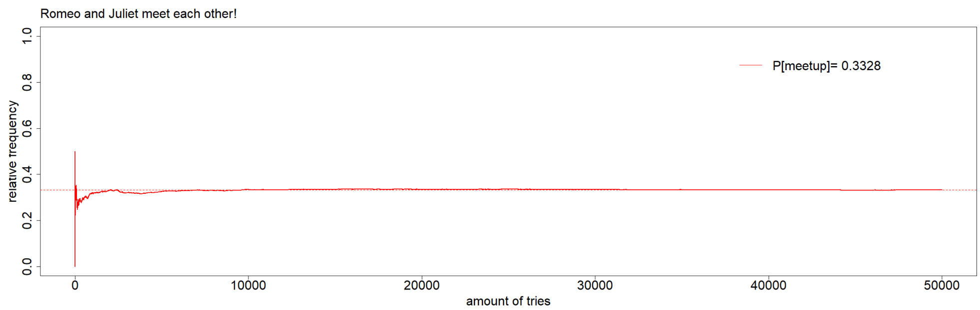

Figura 3: Frequência relativa vs. Número de tentativas de encontros.

6 Conclusões

Este trabalho analisou de forma analítica e empírica dois problemas de probabilidade. Para ambos os problemas, notou-se que os resultados clássicos estão de acordo com os obtidos através de simulações computacionais. O princípio da regularidade estatística foi observado, corroborando a necessidade de repetir-se um experimento várias vezes a fim de obter resultados probabilísticos corretos.