Cálculo aproximado de probabilidade

Neste exemplo a probabilidade de um evento é estimada usando simulação de Monte Carlo implementada em linguagem R.

Neste exemplo a probabilidade de um evento é estimada usando simulação de Monte Carlo implementada em linguagem R.

Problema. Em uma eleição com dois candidatos C1 e C2, uma urna contém cédulas com n_1 votos de C1 e n_2 votos de C2, sem votos nulos ou em branco e $n_1 > n_2$. Os votos são sorteados e apurados um a um e a cada retirada a contagem de votos de cada candidato é atualizada. Calcule a probabilidade de que em alguma etapa da apuração ocorra igual número de votos para os dois candidatos.

Neste exemplo a probabilidade de um evento é estimada usando simulação de Monte Carlo implementada em linguagem R.

Problema. Em uma eleição com dois candidatos C1 e C2, uma urna contém cédulas com n_1 votos de C1 e n_2 votos de C2, sem votos nulos ou em branco e $n_1 > n_2$. Os votos são sorteados e apurados um a um e a cada retirada a contagem de votos de cada candidato é atualizada. Calcule a probabilidade de que em alguma etapa da apuração ocorra igual número de votos para os dois candidatos.

Por exemplo, para $n_1 = 3$ e $n_2 = 2$, na sequência (C1, C1, C2, C1, C2) em nenhuma retirada ocorre empate. Na sequência (C1, C2, C1, C2, C1) ocorrem empates na segunda e na quarta retiradas.

Neste exemplo a probabilidade de um evento é estimada usando simulação de Monte Carlo implementada em linguagem R.

Problema. Em uma eleição com dois candidatos C1 e C2, uma urna contém cédulas com n_1 votos de C1 e n_2 votos de C2, sem votos nulos ou em branco e $n_1 > n_2$. Os votos são sorteados e apurados um a um e a cada retirada a contagem de votos de cada candidato é atualizada. Calcule a probabilidade de que em alguma etapa da apuração ocorra igual número de votos para os dois candidatos.

Por exemplo, para $n_1 = 3$ e $n_2 = 2$, na sequência (C1, C1, C2, C1, C2) em nenhuma retirada ocorre empate. Na sequência (C1, C2, C1, C2, C1) ocorrem empates na segunda e na quarta retiradas.

Nota 1. Prove que a resposta é $2n_2/(n_1 + n_2)$.

A solução exata poderia ser obtida por enumeração de todas as sequências de $n_1 + n_2$ elementos em que n_1 elementos são C1 e n_2 elementos são C2. O número destas sequências é $K = (n_1 + n_2)!/(n_1!n_2!)$ e cada uma delas tem probabilidade $1/K$.

A solução exata poderia ser obtida por enumeração de todas as sequências de $n_1 + n_2$ elementos em que n_1 elementos são C1 e n_2 elementos são C2. O número destas sequências é $K = (n_1 + n_2)!/(n_1!n_2!)$ e cada uma delas tem probabilidade $1/K$.

Se m denota o número de sequências em que ocorre pelo menos um empate na apuração, a probabilidade é m/K. Se $n_1 = 35$ e $n_2 = 20$, K é da ordem de 10^{14}, de modo que a enumeração pode ser inviável. Neste exemplo, em R, o valor de K é calculado como `choose(55, 20)`.

A solução exata poderia ser obtida por enumeração de todas as sequências de $n_1 + n_2$ elementos em que n_1 elementos são C1 e n_2 elementos são C2. O número destas sequências é $K = (n_1 + n_2)!/(n_1!n_2!)$ e cada uma delas tem probabilidade $1/K$.

Se m denota o número de sequências em que ocorre pelo menos um empate na apuração, a probabilidade é m/K. Se $n_1 = 35$ e $n_2 = 20$, K é da ordem de 10^{14}, de modo que a enumeração pode ser inviável. Neste exemplo, em R, o valor de K é calculado como `choose(55, 20)`.

Uma estimativa da probabilidade é calculada com base em simulações de Monte Carlo. Por conveniência, os candidatos C1 e C2 são representados por 1 e -1, respectivamente.

A solução exata poderia ser obtida por enumeração de todas as sequências de $n_1 + n_2$ elementos em que n_1 elementos são C1 e n_2 elementos são C2. O número destas sequências é $K = (n_1 + n_2)!/(n_1!n_2!)$ e cada uma delas tem probabilidade $1/K$.

Se m denota o número de sequências em que ocorre pelo menos um empate na apuração, a probabilidade é m/K. Se $n_1 = 35$ e $n_2 = 20$, K é da ordem de 10^{14}, de modo que a enumeração pode ser inviável. Neste exemplo, em R, o valor de K é calculado como `choose(55, 20)`.

Uma estimativa da probabilidade é calculada com base em simulações de Monte Carlo. Por conveniência, os candidatos C1 e C2 são representados por 1 e -1, respectivamente. Desta forma, se em uma determinada retirada a soma dos votos apurados (`cumsum` em R) for igual a 0, significa que ocorreu empate.

```r
# Número de repetições
R <- 1e3

# Número de votos (n1 > n2)
n1 <- 35
n2 <- 20
n12 <- n1 + n2
```

No trecho de código abaixo, o vetor `aa` com a sequência dos votos é criado com todos os elementos iguais a 1 (C1). Em seguida são sorteadas `n2` posições para receber o valor -1 (C2). Com a função `any` testamos a ocorrência de pelo menos um empate (evento "sucesso").

No trecho de código abaixo, o vetor `aa` com a sequência dos votos é criado com todos os elementos iguais a 1 (C1). Em seguida são sorteadas `n2` posições para receber o valor -1 (C2). Com a função `any` testamos a ocorrência de pelo menos um empate (evento "sucesso").

```
set.seed(9012)  # semente
nsuc <- 0  # número de "sucessos" (empates)
for (i in 1:R) {
   aa <- rep(1, n12)
   ind2 <- sample(n12, n2, replace = FALSE) # C2
   aa[ind2] <- -1
   if (any(cumsum(aa) == 0)) {
      nsuc <- nsuc + 1
   }
}
```

No trecho de código abaixo, o vetor `aa` com a sequência dos votos é criado com todos os elementos iguais a 1 (C1). Em seguida são sorteadas `n2` posições para receber o valor -1 (C2). Com a função `any` testamos a ocorrência de pelo menos um empate (evento "sucesso").

```
set.seed(9012)  # semente
nsuc <- 0  # número de "sucessos" (empates)
for (i in 1:R) {
   aa <- rep(1, n12)
   ind2 <- sample(n12, n2, replace = FALSE) # C2
   aa[ind2] <- -1
   if (any(cumsum(aa) == 0)) {
      nsuc <- nsuc + 1
   }
}
```

Nota 2. No código acima não é necessário testar o último elemento de `cumsum(aa)`. Por quê?

```r
# Resultados
tetac <- nsuc / R
epMC <- sqrt(tetac * (1 - tetac) / R)
emax <- qnorm(0.975) * epMC
cat("\n Valor exato:",
       2 * n2, "/", n12, "=", 2 * n2 / n12,
 "\n Número de repetições:", R,
 "\n Estimativa (e.p. MC):", tetac, "(", epMC, ")",
 "\n IC de 95% aproximado: (", tetac - emax, ",",
    tetac + emax, ")")
```

```
##
##  Valor exato: 40 / 55 = 0.7272727
##  Número de repetições: 1000
##  Estimativa (e.p. MC): 0.725 ( 0.01412002 )
##  IC de 95% aproximado: ( 0.6973253 , 0.7526747 )
```