Řešení dom.cvič. 1-3 (úterní skupina) v R-ku

Příklad 1.14 – Dvě kostky a podmíněná pravděpodobnost

První přístup – samplování

Pravděpodobnost nespočítáme přesně, ale odhadneme tím, že pokus mnohokrát zopakujeme. Hodíme kostkami N-krát a budeme počítat četnosti jednotlivých jevů. Všimněte si, jak podmíněná pravděpodobnost znamená vlastně to, že se omezíme (v čitateli i ve jmenovateli) na ty souřadnice, kde platí podmiňující jev.

```
N = 10^6

kostka1 = sample(1:6, N, replace=TRUE)
kostka2 = sample(1:6, N, replace=TRUE)
soucet = kostka1 + kostka2
SD = soucet==10
PS = kostka1==6
NS = kostka1==6 | kostka2==6

cat("\nP(SD)=", sum(SD)/N)
```

```
##
## P(SD)= 0.082954
```

```
cat("\nP(PS)=", sum(PS)/N)
```

```
##
## P(PS)= 0.166355
```

```
cat("\nP(NS)=", sum(NS)/N)
```

```
##
## P(NS)= 0.305513
```

```
cat("\nP(PS|SD)=", (sum(PS & SD)/N)/(sum(SD)/N))
```

```
##
## P(PS|SD)= 0.3315693
```

```
cat("\nP(NS|SD)=", sum(NS & SD)/sum(SD))
```

```
##
## P(NS|SD)= 0.6673578
```

```
cat("\nP(PS|NS)=", sum(PS & NS)/sum(NS))
```

```
##
## P(PS|NS)= 0.5445104
```

```
cat("\nP(SD|NS)=", sum(SD & NS)/sum(NS))
```

```
##
## P(SD|NS)= 0.1812034
```

```r
cat("\nP(NS|PS)=", sum(NS & PS)/sum(PS))
```

```
##
## P(NS|PS)= 1
```

```r
cat("\nP(SD|PS)=", sum(SD & PS)/sum(PS))
```

```
##
## P(SD|PS)= 0.1653392
```

```r
cat("\n")
```

```r
c(6/11, 1/6, 1/12, 1/3, 2/3, 2/11, 11/36)
```

```
## [1] 0.54545455 0.16666667 0.08333333 0.33333333 0.66666667 0.18181818 0.30555556
```

Druhý přístup – přesný

Projdeme celý pravděpodobnostní prostor. Funguje jen pro malé prostory, jinak trvá moc dlouho!

```r
Omega = expand.grid(k1=1:6,k2=1:6)
kostka1 = Omega$k1; kostka1
```

```
##  [1] 1 2 3 4 5 6 1 2 3 4 5 6 1 2 3 4 5 6 1 2 3 4 5 6 1 2 3 4 5 6 1 2 3 4 5 6
```

```r
kostka2 = Omega$k2; kostka2
```

```
##  [1] 1 1 1 1 1 1 2 2 2 2 2 2 3 3 3 3 3 3 4 4 4 4 4 4 5 5 5 5 5 5 6 6 6 6 6 6
```

```r
soucet = kostka1 + kostka2
N = length(kostka1)
SD = soucet==10

Omega$soucet = soucet
Omega$SD = SD
Omega
```

```
##    k1 k2 soucet    SD
## 1   1  1      2 FALSE
## 2   2  1      3 FALSE
## 3   3  1      4 FALSE
## 4   4  1      5 FALSE
## 5   5  1      6 FALSE
## 6   6  1      7 FALSE
## 7   1  2      3 FALSE
## 8   2  2      4 FALSE
## 9   3  2      5 FALSE
## 10  4  2      6 FALSE
## 11  5  2      7 FALSE
## 12  6  2      8 FALSE
## 13  1  3      4 FALSE
## 14  2  3      5 FALSE
## 15  3  3      6 FALSE
## 16  4  3      7 FALSE
## 17  5  3      8 FALSE
## 18  6  3      9 FALSE
## 19  1  4      5 FALSE
## 20  2  4      6 FALSE
```

```
## 21  3  4      7 FALSE
## 22  4  4      8 FALSE
## 23  5  4      9 FALSE
## 24  6  4     10  TRUE
## 25  1  5      6 FALSE
## 26  2  5      7 FALSE
## 27  3  5      8 FALSE
## 28  4  5      9 FALSE
## 29  5  5     10  TRUE
## 30  6  5     11 FALSE
## 31  1  6      7 FALSE
## 32  2  6      8 FALSE
## 33  3  6      9 FALSE
## 34  4  6     10  TRUE
## 35  5  6     11 FALSE
## 36  6  6     12 FALSE
```

```r
SD
```

```
##  [1] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## [13] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE  TRUE
## [25] FALSE FALSE FALSE FALSE  TRUE FALSE FALSE FALSE FALSE  TRUE FALSE FALSE
```

```r
PS = kostka1==6
NS = kostka1==6 | kostka2==6

cat("\nP(SD)=", sum(SD)/N)
```

```
##
## P(SD)= 0.08333333
```

```r
cat("\nP(PS)=", sum(PS)/N)
```

```
##
## P(PS)= 0.1666667
```

```r
cat("\nP(NS)=", sum(NS)/N)
```

```
##
## P(NS)= 0.3055556
```

```r
cat("\nP(PS|SD)=", sum(PS & SD)/sum(SD))
```

```
##
## P(PS|SD)= 0.3333333
```

```r
cat("\nP(NS|SD)=", sum(NS & SD)/sum(SD))
```

```
##
## P(NS|SD)= 0.6666667
```

```r
cat("\nP(PS|NS)=", sum(PS & NS)/sum(NS))
```

```
##
## P(PS|NS)= 0.5454545
```

```r
cat("\nP(SD|NS)=", sum(SD & NS)/sum(NS))
```

```
##
## P(SD|NS)= 0.1818182
```

```r
cat("\nP(NS|PS)=", sum(NS & PS)/sum(PS))
```

```
##
## P(NS|PS)= 1
```

```r
cat("\nP(SD|PS)=", sum(SD & PS)/sum(PS))
```

```
##
## P(SD|PS)= 0.1666667
```

```r
cat("\n")
```

Příklad 2.16 – Mince v truhle

Zkusíme pro ověření příklad "nasamplovat". Experiment zopakujeme N-krát a pravděpodobnost "změříme". $N = 10^4$ funguje příjemně rychle, pro přesnější odpověď můžeme zvýšit třeba na $N = 10^6$.

```r
N = 10^4
mince = sample(1:100, N, replace=TRUE)
so = integer(N)  # vektor N nul
for (i in 1:N){
  if (mince[i] == 1) {
    so[i] = 1
  } else {
    so[i] = sample(c(0,1), size=1, prob=c(1-1/2^6,1/2^6))
  }
}
sum(mince==1)/sum(so==1)
```

```
## [1] 0.4092827
```

```r
sum(mince==1 & so==1)/sum(so==1)
```

```
## [1] 0.4092827
```

```r
# druhá formulka je přesně podle definice, ale ze smyslu je vidět,
# že to dá totéž a jenom dělá zbytečný výpočet navíc.
```

A pro srovnání totéž napsané vektorově, bez for cyklu. (Cca 10x rychlejší.)

```r
N = 10^6
mince = sample(1:100, N, replace=TRUE)
hody = rbinom(N,1,1/2^6)  # N nasamplovaných hodnot Bin(1,1/2^6)
so = (mince==1) | hody    # pokud jsme vybrali první minci, tak vektor hody ignorujeme.
sum(mince==1)/sum(so==1)
```

```
## [1] 0.3945151
```

Příklad 3.9 – Dvanáctistěnná kostka

První přístup – přesný

Napřed použijeme R-ko jenom jako kalkulačku, počítáme podle základního vzorce, z pravděpodobnostní funkce (x jsou hodnoty, p je odpovídající pravděpodobnostní funkce).

```r
x = c(1,2,4,5)
p = c(1,2,4,5)/12; sum(p)  # pro kontrolu
```

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

```r
E = sum(x*p)
V = sum((x-E)^2*p)

E                # spočteno podle definice
```

```
## [1] 3.833333
```

```r
V                # spočteno podle definice
```

```
## [1] 1.805556
```

```r
sum(x^2*p)-E^2   # věta o rozptylu
```

```
## [1] 1.805556
```

```r
sqrt(V)          # definice
```

```
## [1] 1.34371
```

Teď podle druhého vzorce – průměr přes pravděpodobnostní prostor.

```r
hody = c(1,2,2,4,4,4,4,5,5,5,5,5); length(hody)  # pro kontrolu
```

```
## [1] 12
```

```r
E = sum(hody)/12
V = sum((hody-E)^2)/12

E #střední hodnota
```

```
## [1] 3.833333
```

```r
V #rozptyl
```

```
## [1] 1.805556
```

```r
sqrt(V) #směrodatná odchylka
```

```
## [1] 1.34371
```

Existují i funkce mean, var, sd. Mean dělá to, co by člověk čekal, ale var, sd počítají něco trochu jiného (místo počtem měření se dělí o jedna nižším číslem). Více vysvětlíme v části o statistice.

```r
mean(hody)
```

```
## [1] 3.833333
```

```r
var(hody)
```

```
## [1] 1.969697
```

```r
sd(hody)
```

```
## [1] 1.403459
```

Druhý přístup – samplování

Zde je správné místo pro funkce mean, var, i sd.

```r
N = 10^7
```

```r
mnoho_hodu = sample(hody,N,replace = TRUE)
mean(mnoho_hodu)
```

```
## [1] 3.833569
```

```r
var(mnoho_hodu)
```

```
## [1] 1.805486
```

```r
sd(mnoho_hodu)
```

```
## [1] 1.343684
```