4.1 Pravděpodobnostní prostor, podmíněná pravděpodobnost, nezávislost jevů

Už víme, jak v programu R nalézt a vypsat množinu elementárních jevů Ω, jak pomocí funkce `subset` vypsat jen některé elementární jevy, např. příslušící nějakému náhodnému jevu A. Aby ale pravděpodobnostní prostor byl kompletní, musíme umět přiřadit elemtárním jevům jejich pravděpodobnost. Nejjednodušší případ je tzv. „Equally Likely Model“, který všem elementárním jevům přiřadí stejnou pravděpodobnost. Tedy pokud náhodný pokus může skončit n různými výsledky (máme n různých elementárních jevů), pravděpodobnost každého z nich bude $1/n$.

▷ **Příklad 4.1.** U funkcí `tosscoin(), rolldie(), cards()`, dostaneme pravděpodobnostní prostor jednoduše tím, že nastavíme parametr `makespace=TRUE`.

```
> DF <- tosscoin(2,makespace=TRUE)
> DF
  toss1 toss2 probs
1     H     H  0.25
2     T     H  0.25
3     H     T  0.25
4     T     T  0.25
```

U funkce `urnsamples()` parametr chybí a důvod je nasnadě. Protože parametry výběru můžeme měnit, obecně nemůže být zaručeno, že množina elementárních jevů má uniformní distribuci. Pokud chceme neuniformní distrubuci simulovat, můžeme k tomu použít funkci `probspace(x,probs)`, kde `x` je množina elementárních jevů a `probs` je vektor pravděpodobností. Pokud vektor pravděpodobností neuvedeme, považuje se za uniformní.

▷ **Příklad 4.2.** Uvažujme, že máme fale šnou kostku a že 2 padá dvakrát časteji než ostatní čísla. Pak pravděpodobnostní prostor můžeme vytvořit následovně:

```
> outcomes<-rolldie(1)
> p<-c(1/7,2/7,1/7,1/7,1/7,1/7)
> S<-probspace(x=outcomes,probs=p)
> S
  X1     probs
1  1 0.1428571
2  2 0.2857143
3  3 0.1428571
4  4 0.1428571
5  5 0.1428571
6  6 0.1428571
```

Pokud ale budeme házet falešnou kostkou/mincí vícekrát, bude situace komplikovanější. V našem případě je jasné, že výsledek (2,2) má větší pravděpodobnost než výsledek (1,2) a obecně tedy neplatí model ELM. Protože si ale kostka „nepamatuje“, co bylo hozeno dříve, jsou jednotlivé hody nezávislé náhodné pokusy a můžeme proto použít funkci `iidspace(x,ntrials,probs)`, kde `x` je množina elementárních jevů (musí být *vektor*) pro jeden pokus (jeden hod), `ntrials` udává počet, kolikrát náhodný pokus opakujeme (kolikrát hodíme kostkou) a `probs` je vekor pravděpodobností pro elementární jevy v `x`.

▷ **Příklad 4.3.** Jak bude vypadat pravděpodobností prostor pro naši falešnou kostku, jestliže hodíme dvakrát? Nejpve si uvědomte, že `outcomes` z předchozího příkladu je data.frame.

```
> x<-outcomes[,1]
> iidspace(x,ntrials=2, probs=p)
  X1 X2      probs
1  1  1 0.02040816
2  2  1 0.04081633
3  3  1 0.02040816
...................
8  2  2 0.08163265
9  3  2 0.04081633
...................
34 4  6 0.02040816
35 5  6 0.02040816
36 6  6 0.02040816
```

Pomocí funkce `prob(x,event,given)` můžeme určit pravděpodobnost výskytu nějakého náhodného jevu (parametr `event`) a to dokonce za předpokladu, že nastal nějaký jiný jev, tj. podmíněnou pravděpodobnost (parametr `given`). Prvním argumentem je samotný pravděpodobnostní prostor nebo jeho podmožina.

▷ **Příklad 4.4.** Chceme zjistit, jaká je pravděpodobnost, že padne číslo menší než 4.

```
> prob(S, event=X1<4)
[1] 0.5714286
```

Předpokládejme, že nám někdo řekl, že padlo sudé číslo. Jaká je pravděpodobnost, že padla 2?

```
> prob(S,event=X1==2,given=(X1==2|X1==4|X1==6))
[1] 0.5
```

▷ **Příklad 4.5.** Třikrát hodíme kostkou. Jaká je pravděpodobnost, že hodíme právě dvě čtverky. Jaká je pravděpodobnost, že hodíme minimálně dvě čtverky? Množina všech možných elementárních jevů (všech možných trojic) má 216 prvků. Všechny možnosti, jak mohly padnout právě dvě čtverky, můžeme zapsat takto:
`((X1==4 & X2==4)|(X1==4 & X3==4)|(X2==4 & X3==4))&!(X1==X2 & X2==X3))`
Poslední podmínkou jsme vyloučili případy, že padnou tři čtverky. Vidíme ale, že podmínka je příliš komplikovaná, naštěstí jde zapsat jednodušeji a to pomocí funkce `isrep(x,vals,nrep)`, kde `x` je vekor (nebo dataframe) pro který chceme zjišťovat, jestli se v něm `nrep`-krát opakuje hodnota `vals`. Pro nalezení odpovědi u druhé otázky můžeme využít již dříve zmiňovanou funkci `isin()`. Řešení je nyní již zřejmé:

```
> L<-rolldie(3, makespace=TRUE)
> A<-subset(L, isrep(L,4,2))
> prob(A)
[1] 0.06944444
> B<-subset(L, isin(L,c(4,4,4)))
> prob(B)
[1] 0.00462963
```

Následující příklady byste již měli být schopni vyřešit sami.

► **Příklad 4.1.** Náhodně taháme karty z balíčku 52 hracích karet. Určete pravděpodobnost, že vytáhnete srdcovou kartu nebo jakékoliv eso? [0.3077]

► **Příklad 4.2.** V prodejně mají žárovky od 3 různých firem A, B, C: 45% od firmy A, 30% od firmy B, 25% od firmy C. Jaká je pravděpodobnost, že zákazník koupí žárovku, která bude mít životnost deklarovanou na obalu, když tento požadavek splňuje pouze 90/80/70% procent žárovek od firmy A/B/C? Zákazník si žárovku vybírá náhodně. [0.82]

► **Příklad 4.3.** Čtyři firmy A, B, C, D se ucházejí o státní zakázku. Na základě výsledků z podobných výběrových řízení jsme schopni stanovit pravděpodobnosti přidělení zakázky:
$P(A) = 0.4$, $P(B) = 0.3$, $P(C) = 0.2$, $P(D) = 0.1$
Firma A se rozhodla z výběrového řízení odstoupit? Jaké jsou nyní pravděpodobnosti úspěchu u zbývajících firem? [1/2; 1/3; 1/6]

► **Příklad 4.4.** Dvakrát hodíme kostkou. Určete pravděpodobnost, že padl součet ostře větší než 10, víte-li, že padla alespoň jedna šestka. Určete pravděpodobnost, že padla alespoň jedna šestka, víte-li, že padl součet ostře větší než 10. Jak zdůvodníte druhý výsledek? [0.27273;1]

► **Příklad 4.5.** V urně je 10 bílých a 10 černých balónků. Náhodně taháme balónky jeden po druhém z urny a vždy po vytažení vracíme. Jaká je pravděpodobnost, že čtvrtý bílý balónek bude vytažený jako šestý v pořadí? [0.15625]
Jak se změní situace, pokud balónky nevracíme? [0.162338]

► **Příklad 4.6.** Čtyřikrát hodíme čtyřstěnnou kostkou. Pokud na k-tý hod vidíme k ok, označíme tento hoxd za „shodu“. Celý pokus považujeme za úspěšný, došlo-li alespoň jednou ke shodě. Určte pravděpodobnost, že pokus skončil úspěchem. [0.68359]

► **Příklad 4.7.** Uvažujme dva hody jednou kostkou. Jev A_1 nastane, pokud v prním hodu padne sudé číslo. Jev A_2 nastane, pokud ve druhém hodu padne sudé číslo. Jev A_3 nastane, pokud v prvním a druhém hodu padne stejné číslo.

a) Určete $P(A_1)$, $P(A_2)$, $P(A_3)$. [1/2;1/2;1/6]

b) Určete $P(A_1 \cap A_2)$, $P(A_1 \cap A_3)$, $P(A_1 \cap A_2 \cap A_3)$. [1/4;1/12;1/12]

c) Určete $P(A_1|A_2)$, $P(A_3|A_1)$. [1/2;1/6]

d) Jsou jevy A_1 a A_3 nezávislé? [ano]

e) Jsou jevy A_1 a A_2 nezávislé? [ano]

f) Jsou jevy A_1, A_2 a A_3 nezávislé? [ne]

► **Příklad 4.8.** Máme balíček 32 karet a vybíráme karty jednu po druhé. Karty do balíčku nevracíme. Jaká je pravděpodobnost, že druhá vybraná karta je eso? [1/8]

► **Příklad 4.9.** Mějme n náhodných jevů $A_1, \ldots, A_n$. Kolik podmínek musíme otestovat, chceme-li ukázat, že jevy jsou vzájemně nezávislé? [$2^n - n - 1$]

► **Příklad 4.10.** Podle statistik kouří v České republice 26% populace. 18% populace tvoří lidé, kteří chtějí přestat kouřit. Určete pravděpodobnost, že náhodně svolený kuřák chce přestat kouřit. [0.69]

► **Příklad 4.11.** Čerpací stanice si eviduje zákazníky a proto ví, že 40% zákazníků tankuje naftu, 35% tankuje Natural 95 a 25% tankuje Natural 98. Z těch zákazníků, kteří tankují naftu/Natural 95/Natural 98 jich 30/60/50% tankuje do plna. Pokud následující zákazník takoval do plna, jaká je pravděpodobnost, že tankova naftu/Natural 95/Natural 98? [0.2637/0.4615/0.2747]

► **Příklad 4.12.** V oblasti chemického průmyslu se sleduje znečištění řek. Uvažujme následující náhodné jevy:
A . . . . . . náhodně zvolená řeka je skutečně znečištěna,
B . . . . . . testovaný vzorek vody byl vyhodnocen jako znečištěný,
C . . . . . . u náhodně zvolené řeky je povoleno rybaření.
Víte, že $P(A) = 0.3$, $P(B|A) = 0.75$, $P(B|A') = 0.2$. Dále víte, že $P(C|A \cap B) = 0.2$, $P(C|A' \cap B) = 0.15$, $P(C|A \cap B') = 0.8$, $P(C|A' \cap B') = 0.9$.
Určete $P(B)$, $P(A \cap B \cap C)$, $P(B' \cap C)$, $P(C)$. [0.365, 0.045, 0.564, 0.630]

Reference

[1] Devore J. L.: Probability and Statistics for Engineering and the Sciences
Duxbury Press, 7. vydání 2008, ISBN 978-0-495-55744-9.

[2] Kerns G. J.: Elementary Probability on Finite Sample Spaces, 2009, reference manual package prob, available from: http://CRAN.R-project.org/package=prob

[3] Kerns G. J: Introduction to Probability and Statistics Using R, First Edition
http://cran.r-project.org/web/packages/IPSUR/vignettes/IPSUR.pdf