

Tvorba diagramů s českými popisky a matematickou notací v prostředí R

Lubomír Štěpánek

14. srpna 2018

Obsah

1 Ukázkový příklad	1
2 Kód pro vytvoření všech tří diagramů v R	4
3 Závěr	9

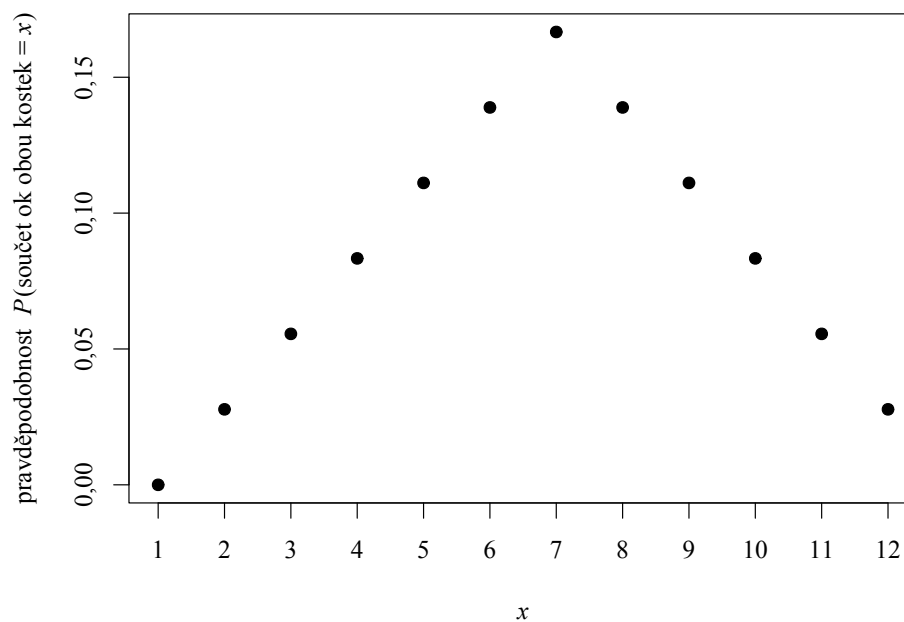
1 Ukázkový příklad

Uvažujme dvě nebiasované laplaceovské kostky vždy o počtu ok 1 až 6, kterými současně hodíme. Hledíme pravděpodobnost, že součet ok na obou kostkách bude roven číslu $x \in \mathbb{R}$.

Součet ok, která padnou na obou kostkách, vnímejme jako náhodnou veličinu a označme ji X . Hledáme tedy $P(X = x)$. Po troše uvažování a základní aritmetiky můžeme odvodit, že pravděpodobnostní funkce náhodné veličiny X je $P_X(x) = P(X = x)$ rovna

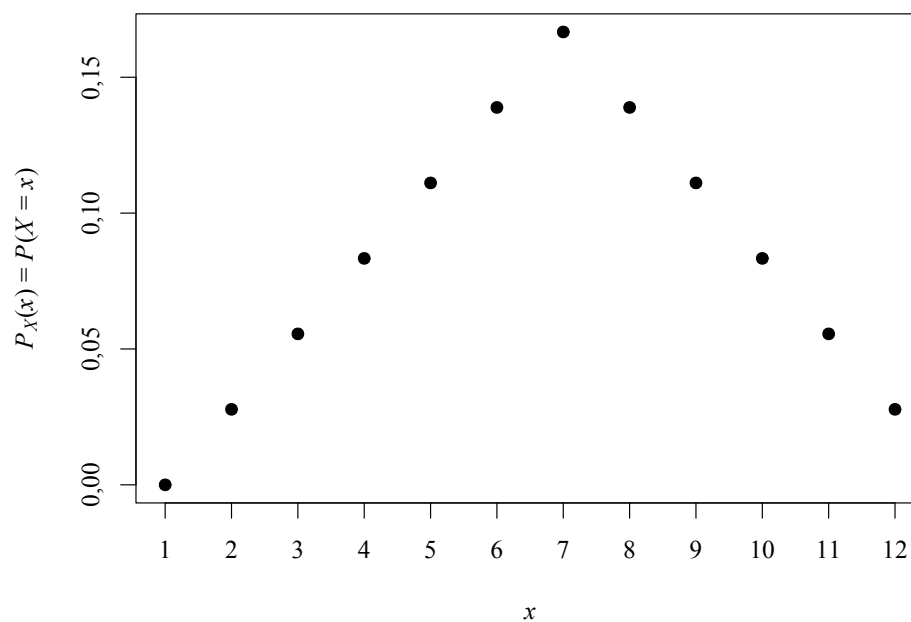
$$P_X(x) = P(X = x) = \begin{cases} \frac{6-|x-7|}{36}, & x \in \{2, 3, \dots, 12\} \\ 0, & \text{jinak.} \end{cases}$$

Na obrázku 1 je pak grafická prezentace pravděpodobnostní funkce $P_X(x)$ pro rozumné hodnoty x .



Obrázek 1: Pravděpodobnosti $P(\text{součet ok obou kostek} = x)$ pro jednotlivé hodnoty x , kterých může součet ok dvou hozených kostek nabýt.

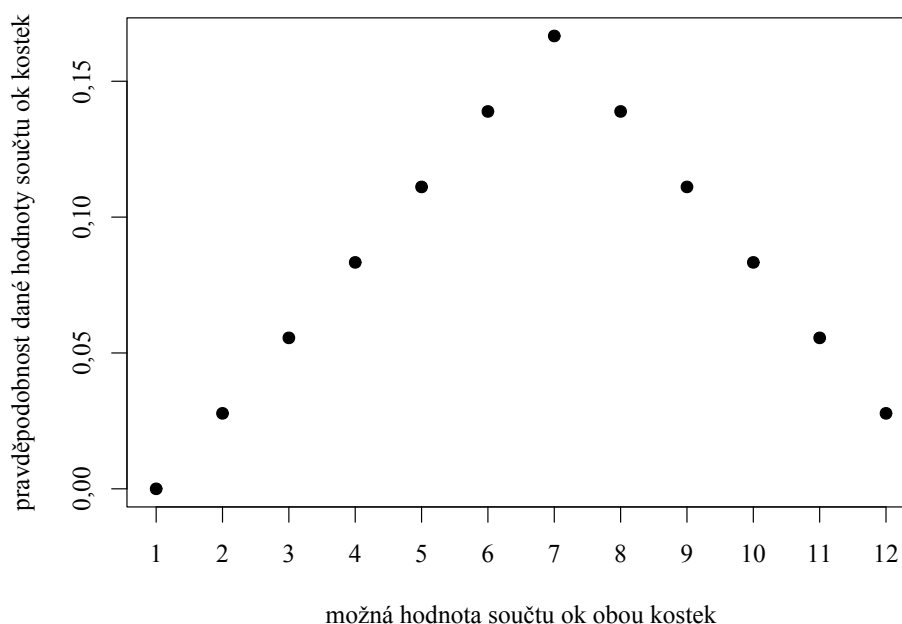
Na obrázku 2 pak tatáž pravděpodobnostní funkce $P_X(x)$ pro rozumné hodnoty x , avšak s formálnějším popiskem svislé osy.



Obrázek 2: Pravděpodobnostní funkce $P_X(x) = P(X = x)$, kde $x \in \{1, 2, 3, \dots, 12\}$, pro náhodnou veličinu X odpovídající součtu ok dvou hozených laplaceovských kostek.

Do třetice a pro cvičné účely je stejný diagram i na obrázku 3, kde je v popiscích os zcela upuštěno od matematické notace¹. Popisky však obsahují bohatou českou diakritiku.

¹Všechny tři diagramy byly vytvořeny pouze s využitím jazyka a prostředí R.



Obrázek 3: Pravděpodobnosti jednotlivých hodnot, kterých může součet ok obou hozených kostek nabýt.

2 Kód pro vytvoření všech tří diagramů v R

Nakonec se ještě podívejme, jaký kód v prostředí R takové diagramy vytvoří.

```

1  # -----
2
3  ## tabulka s pravděpodobnostmi, že při hodu dvěma
4  ## laplaceovými kostkami bude součet ok obou
5  ## kostek roven číslu x
6
7  my_table <- list(
8    "x" = as.numeric(1:12),
9    "pravdepodobnost_x" = unlist(
10     lapply(
11       1:12,
12       function(x){
13         sum(outer(1:6, 1:6, "+") == x) / 36
14       }
15     )
16   )
17 )
18
19

```

```

20  # -----
21
22  ## diagram součet x vs. pravděpodobnost, že obě hozené kostky
23  ## mají součet ok roven číslu x
24
25  cairo_ps(
26    file = "pravdepodobnost_souctu_x.eps",
27    width = 8,
28    height = 5.4,
29    pointsize = 16,
30    family = "serif"      # patkový font kvůli matematické notaci
31  )
32
33  par(mar = c(4.1, 4.1, 0.1, 0.1))
34
35  plot(
36    x = my_table[["x"]],
37    y = my_table[["pravdepodobnost_x"]],
38    xlab = expression(italic(x)),
39    ylab = "",
40    xaxt = "n",      # zakazují vykreslení vodorovné osy
41    yaxt = "n",      # zakazují vykreslení svislé osy
42    pch = 19
43  )
44
45  # všimněme si, že má-li popisek osy obsahovat současně písmeno
46  # (či písmena) s diakritikou (ě ve slově pravděpodobnost)
47  # a matematickou notaci ( $P(\text{součet ok obou kostek} = x)$ ),
48  # není možné použít obvyklý xlab či ylab, ale popisek se musí
49  # sestavit pomocí funkce pro text kopírující osu diagramu mtext();
50  # ani engine cairo_pdf() či cairo_ps() oboje najednou neumí
51
52  mtext(
53    text = "pravděpodobnost      (součet ok obou kostek =
54    )",
55    side = 2,
56    line = 3
57  ) # počet mezer jsem prostě jen odhadl metodou error-trial,
58    # celkem "fishy", není to moc elegantní a zabere to čas,
59    # ale dle mě a mnohých fór asi jediné řešení pro
60    # češtinu & matematiku současně v diagramech v R
61
62  mtext(
63    text = expression(italic(P)),
64    side = 2,      # side = 2 pro svislou osu
65    line = 3,      # vzdálenost od canvasu diagramu
66    at = 0.0655    # v jednotkách vodorovné osy
67                  # opět, pozici dle jednotky svislé
68                  # osy odhaduji od oka
69  )

```

```

69
70 mtext(
71   text = expression(italic(x)),
72   side = 2,
73   line = 3,
74   at = 0.168      # same story here
75 )
76
77 axis(
78   side = 1,
79   at = 1:12,
80   labels = as.character(1:12)
81 )
82
83 axis(
84   side = 2,
85   at = seq(0, 0.15, 0.05),
86   labels = gsub(
87     "\\.",
88     ",",
89     format(
90       round(
91         seq(0, 0.15, 0.05),
92         digits = 2
93       ),
94       nsmall = 2
95     )
96   ) # nahrazuju desetinnou tečku za čárku, must-do v česky
97   # mluvících diagramech
98 )
99
100 dev.off()
101
102
103 # -----
104
105 ## stejný diagram, ale pro cvičné účely tentokrát v popiscích
106 ## použijeme jen matematickou notaci, tvorba popisku se tím
107 ## zjednoduší
108
109 cairo_ps(
110   file = "pravdepodobnost_souctu_x_formalneji.eps",
111   width = 8,
112   height = 5.4,
113   pointsize = 16,
114   family = "serif"      # patkový font kvůli matematické notaci
115 )
116
117 par(mar = c(4.1, 4.1, 0.1, 0.1))
118

```

```

119 plot(
120   x = my_table[["x"]],
121   y = my_table[["pravdepodobnost_x"]],
122   xlab = expression(italic(x)),
123   ylab = expression(
124     paste(
125       italic(P)[italic(X)],
126       "(",
127       italic(x),
128       ") = ",
129       italic(P),
130       "(",
131       italic(X),
132       " = ",
133       italic(x),
134       ") ",
135       sep = " "
136     )
137   ),
138   xaxt = "n",      # zakazuji vykreslení vodorovné osy
139   yaxt = "n",      # zakazuji vykreslení svislé osy
140   pch = 19
141 )
142
143 axis(
144   side = 1,
145   at = 1:12,
146   labels = as.character(1:12)
147 )
148
149 axis(
150   side = 2,
151   at = seq(0, 0.15, 0.05),
152   labels = gsub(
153     "\\.",
154     ",",
155     format(
156       round(
157         seq(0, 0.15, 0.05),
158         digits = 2
159       ),
160       nsmall = 2
161     )
162   ) # nahrazuji desetinnou tečku za čárku, must-do v česky
163     # mluvících diagramech
164 )
165
166 dev.off()
167
168

```

```

169 | # -----
170 |
171 | ## do třetice stejný diagram, ale bez matematické notace
172 | ## v popiscích
173 |
174 | cairo_ps(
175 |   file = "pravdepodobnost_souctu_x_neformalne.eps",
176 |   width = 8,
177 |   height = 5.4,
178 |   pointsize = 16,
179 |   family = "serif"      # patkový font kvůli matematické notaci
180 | )
181 |
182 | par(mar = c(4.1, 4.1, 0.1, 0.1))
183 |
184 | plot(
185 |   x = my_table[["x"]],
186 |   y = my_table[["pravdepodobnost_x"]],
187 |   xlab = "možná hodnota součtu ok obou kostek",
188 |   ylab = "pravděpodobnost dané hodnoty součtu ok kostek"
189 |   ,
190 |   xaxt = "n",      # zakazují vykreslení vodorovné osy
191 |   yaxt = "n",      # zakazují vykreslení svislé osy
192 |   pch = 19
193 | )
194 |
195 | axis(
196 |   side = 1,
197 |   at = 1:12,
198 |   labels = as.character(1:12)
199 | )
200 |
201 | axis(
202 |   side = 2,
203 |   at = seq(0, 0.15, 0.05),
204 |   labels = gsub(
205 |     "\\.",
206 |     ",",
207 |     format(
208 |       round(
209 |         seq(0, 0.15, 0.05),
210 |         digits = 2
211 |       ),
212 |       nsmall = 2
213 |     )
214 |   ) # nahrazuju desetinnou tečku za čárku, must-do v česky
215 |   # mluvících diagramech
216 | )
217 | dev.off()

```


218 ||
219 ||

3 Závěr

Základním ideou tvorby diagramů s českými popisky je použití enkapsulovaného postscriptu (`.eps`) a (nejen) R-kového enginu `cairo`, který narozdíl od běžných základních enginů (`pdf`, ...) pro sázení vektorových obrázků používá znakovou sadu extended ASCII, tedy minimálně 512 znaků, což pokryje i česká písmena s diakritikou. V R jde o funkci `cairo_ps()`, která spustí vývojářský mód, vytvoří podle kódu diagram, zrenderuje ho a nakonec uloží ve formátu enkapsulovaného postscriptu (přípona `.eps`). Obdobná funkce `cairo_pdf()` vrátí koncový formát `.pdf`. Obě funkce jsou součástí balíčku `grDevices`, který se inicializuje automaticky po spuštění R, takže není nutné nic navíc instalovat. Obdobné funkce a možná trochu lepší uživatelské možnosti nabízí R-kový balíček `Cairo`, nicméně pro uvedené tři diagramy postačila jen základní funkce `cairo_ps()`.

Přestože není dobrou praxí vkládat do \TeX -ového dokumentu obrázky jako `.pdf`, reálně by mělo vše fungovat i se zachováním české diakritiky. Formát postscriptu je preferován proto, že jde ve skutečnosti o vlastní jazyk (takže obrázek s příponou `.eps` nelze otevřít v běžném editoru obrázků, ale naopak lze otevřít v textovém editoru). Při znalosti kódu postscriptu (je podobný jazyku PROLOG) lze dokonce tímto způsobem dělat postprocessingové úpravy obrázků pouhou úpravou jeho kódu (bez nutnosti obrázků znovu generovat).

Kombinování českých písmen s diakritikou a matematické notace v jednom popisku je svízelné, ale možné. Jedním z řešení² je tvorba takového popisku „ve dvou vrstvách“, postupně, tedy nejdříve vytvoření textového podkladu – ten může obsahovat diakritiku, a poté až umístění matematické notace obalené v R-kové funkci `expression()`. Obě „vrstvy“ popisku lze vytvořit pomocí R-kové funkce `mtext()`. Naopak známé grafické argumenty `xlab`, `ylab` samy o sobě v tomto duchu fungovat nebudou (engine `cairo` pravděpodobně neumí zpracovat výstup jednoho generického grafického příkazu v R tak, aby zrenderoval současně diakritiku i matematickou notaci, a v ten okamžik přestane kód diagramu dále renderovat, aniž by vrátil chybovou hlášku – výsledkem je tak „diagram“ s chybějícími komponentami). Příkladem kombinace matematické notace a diakritiky v jednom popisku je obrázek 1.

²Osobně žádné jiné bohužel neznám a po dlouhém (řádově dny v roce 2017) pátrání po všemožných Q&A fórech ani nic nenasevňuje, že by jiné existovalo.