Zadání příkladů – Statistická inference I – 2016

Příklad 20 (standardizované normální rozdělení). Vypočítejte kritické hodnoty $u(\alpha)$ rozdělení $N(0,1)$, kde $\alpha = 0.1$, 0.05, 0.01, 0.025 a 0.005.
Na výpočet pravděpodobnosti pod kvantilem se používá funkce `pnorm(Q)`. Na výpočet pravděpodobnosti nad kritickou hodnotou se používá funkce `1-pnorm(Q)`. Jelikož je standardizované normální rozdělení symetrické okolo nuly, $u(\alpha) = u(1-\alpha)$.

Příklad 21 (Studentovo t-rozdělení). Vypočítejte kritické hodnoty Studentova t-rozdělení se stupni volnosti $df = 10$, tj. $t_{df}(\alpha)$, kde $\alpha = 0.1$, 0.05, 0.01, 0.025 a 0.005.

Na výpočet pravděpodobnosti pod kvantilem se využívá funkce `pt(Q,df)`. Na výpočet pravděpodobnosti nad kritickou hodnotou se využívá funkce `1-pt(Q,df)`. Jelikož Studentovo t-rozdělení je symetrické okolo nuly, $t_{df}(\alpha) = t_{df}(1-\alpha)$. Nějaký kvantil standardizovaného normálního rozdělení je přibližně rovný kvantilu t-rozdělení (resp. pravděpodobnosti nad kritickými hodnotami jsou přibližně stejné) až pro velmi vysoké stupně volnosti. Např. `1-pnorm(1.644869)` $\approx$ `1-pt(1.644869, 100 000)=0.05`. Avšak např. `1-pt(1.644869, 100)` $\doteq$ 0.052.

Příklad 22 (χ^2 rozdělení). Vypočítejte kritické hodnoty χ^2-rozdělení se stupni volnosti $df = 10$, tj. $\chi^2_{df}(\alpha)$, kde $\alpha = 0.1$, 0.05, 0.01, 0.025 a 0.005.

Na výpočet pravděpodobnosti pod kvantilem anebo nad kritickou hodnotou se používá funkce `pchisq(Q,df)`. Jelikož χ^2-rozdělení není symetrické, $\chi^2_{df}(\alpha) \neq \chi^2_{df}(1-\alpha)$.

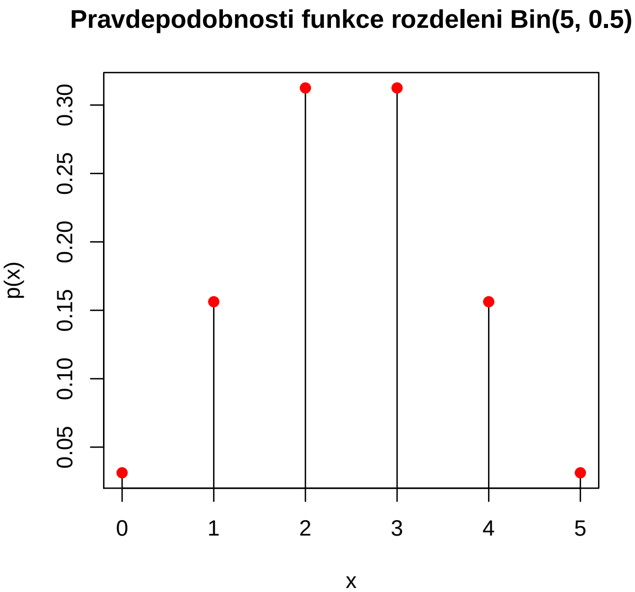
Příklad 23 (F-rozdělení). Vypočítejte kritické hodnoty F-rozdělení se stupni volnosti $df_1 = 20$ a $df_2 = 20$, tj. $F_{df_1,df_2}(\alpha)$, kde $\alpha = 0.1$, 0.05, 0.01, 0.025 a 0.005.

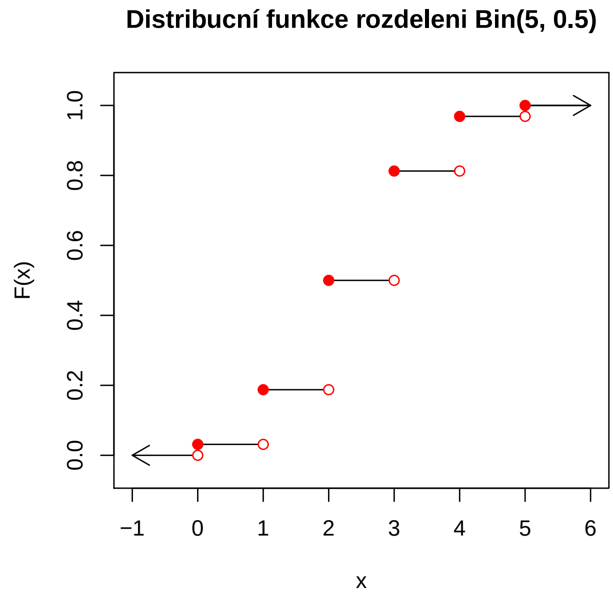
Na výpočet pravděpodobnosti pod kvantilem se používá funkce `pf(Q, df1, df2)`. Na výpočet pravděpodobnosti nad kritickou hodnotou se používá funkce `1-pt(Q, df1, df2)`. Jelikož F-rozdělení není symetrické, $F_{df1,df2}(\alpha) \neq F_{df1,df2}(1-\alpha)$.

Příklad 24 (binomické rozdělení, binomický experiment). Experiment sestávající z fixního počtu Bernoulliho experimentů (ozn. N) se nazývá binomický experiment. Pravděpodobnost úspěchu označme p, pravděpodobnost neúspěchu $q = 1-p$. Náhodná proměnná X je počet pozorovaných úspěchů po dobu experimentu. Pravděpodobnost $X = x$ za podmínky, že X pochází z binomického rozdělení $Bin(N,p)$, píšeme jako

$$\Pr(X = x) = \binom{N}{x} p^x (1-p)^{N-x}, x = 0, 1, \ldots, N \tag{1}$$

(Ugarte a kol. 2008). Střední hodnota $E[X] = Np$ a rozptyl $Var[X] = Np(1-p)$. Naprogramujte a zobrazte v R pravděpodobnostní funkci a (kumulativní) distribuční funkci pro $Bin(5, 0.5)$.





Příklad 25 (**podíl chlapců a dívek v rodinách).** Nechť X představuje početnost chlapců mezi dětmi v rodinách. Zde můžeme předpokládat, že $X \nsim Bin(N, p)$, tj. rodina může mít vychýlený poměr pohlaví dětí ve směru k chlapcům nebo k dívkám. V realitě tedy můžeme mít velmi mnoho rodin jen s chlapci nebo jen s děvčaty a nemáme dostatek rodin s poměrem pohlaví blízkým 51 : 49 (poměr chlapců ku dívkám). Z toho nám vyplývá, že rozptyl početnosti chlapců bude ve skutečnosti větší než rozptyl předpokládaný binomickým rozdělením $Bin(n, P)$.

Příklad 26 (**overdispersion v binomickém modelu).** V klasické studii poměru pohlaví u lidí z roku 1889 na základě záznamů z nemocnic v Sasku (více informací viz Lindsey a Altham, (1998)) zaznamenal Geissler (1889) rozdělení počtu chlapců v rodinách. Mezi $M = 6115$ rodinami s $N = 12$ dětmi pozoroval následující početnosti chlapců (n jsou početnosti chlapců a m_n početnosti rodin s n chlapci).

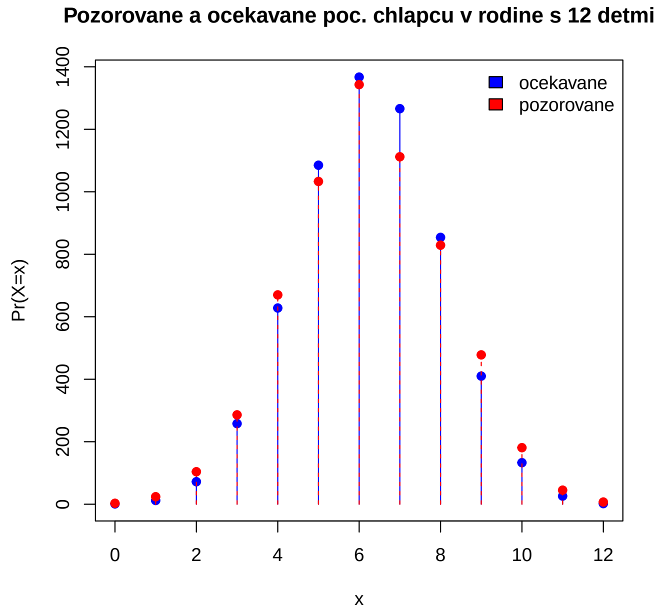
n	0	1	2	3	4	5	6	7	8	9	10	11	12
m_n	3	24	104	286	670	1033	1343	1112	829	478	181	45	7

Vypočítejte m_n za předpokladu, že početnosti chlapců X v rodinách mají binomické rozdělení s parametry

$$\pi = \frac{\sum_{n=0}^{N} n m_n}{NM} = 0.5192 \tag{2}$$

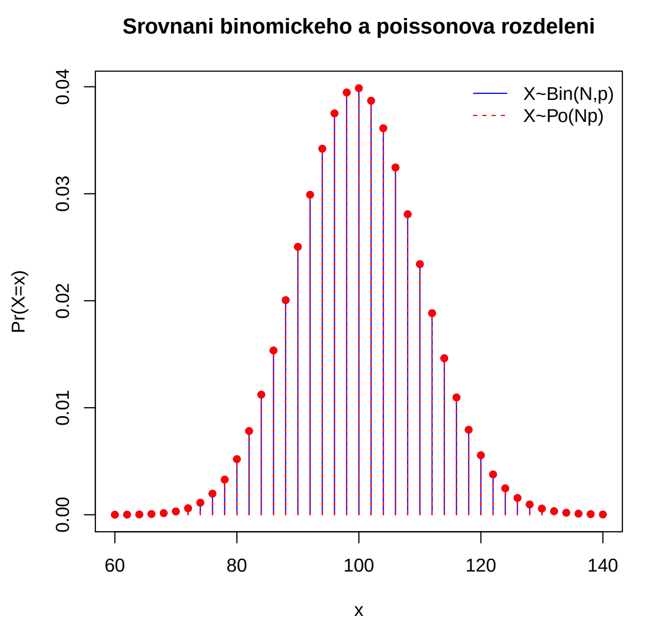
a $N = 12$, ozn. $X \sim Bin(N, \pi)$.

```
##             0  1   2   3   4    5    6    7   8   9  10 11 12
## pozorovane 3 24 104 286 670 1033 1343 1112 829 478 181 45  7
## ocekavane  1 12  72 258 628 1085 1367 1266 854 410 133 26  2
```




```
## [1] 3.489269
## [1] 2.993832
```

Příklad 27 (Poissonovo rozdělení; počet havárií za týden). Pokud každý z 50 milionů lidí řídí v Itálii řídí auto následující týden nezávisle, potom pravděpodobnost smrti při autonehodě bude 0.000002, kde počet úmrtí má binomické rozdělení $\text{Bin}(50mil, 0.000002)$ anebo limitní Poissonovo rozdělení s parametrem $\lambda = 50mil \times 0.000002 = 100$.



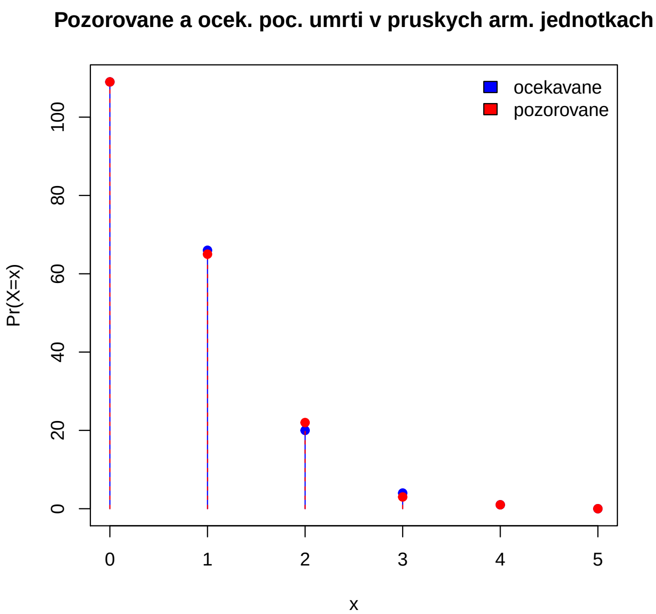
Příklad 28 (Poissonovo rozdělení; pruské armádní jednotky). Nechť početnosti úmrtí X jako následek kopnutí koněm v Pruských armádních jednotkách (Bortkiewicz, 1898) mají Poissonovo rozdělení s parametrem λ, tj. $X \sim Poiss(\lambda)$. Pravděpodobnost, že někdo bude smrtelně zraněný v daném dni, je extrémně malá. Mějme

10 vojenských jednotek za 20-letou periodu s rozsahem $M = 200$ ($200 = 10 \times 20$), kde, při početnostech úmrtí $n = 1, 2, 3, 4, 5+$ v dané jednotce a v daném roce, zaznamenáváme také početnosti vojenských jednotek m_n při daném n, kde $M = \sum m_n$ (viz tabulka). Vypočítejte očekávané početnosti, za předpokladu $X \sim Poiss(\lambda)$, kde

$$\lambda = \frac{\sum_n nm_n}{\sum_n m_n}. \tag{3}$$

n	0	1	2	3	4	5+
m_n	109	65	22	3	1	0

```
## [1] 0.61
##              0  1  2 3 4 5+
## pozorovane 109 65 22 3 1  0
## ocekavane  109 66 20 4 1  0
```



Příklad 29 (overdispersion v Poissonově modelu). Mějme početnosti úrazů n mezi dělníky v továrně, kde početnosti dělníků m_n při daném n (viz tabulka) (Greenwood a Yule (1920)).

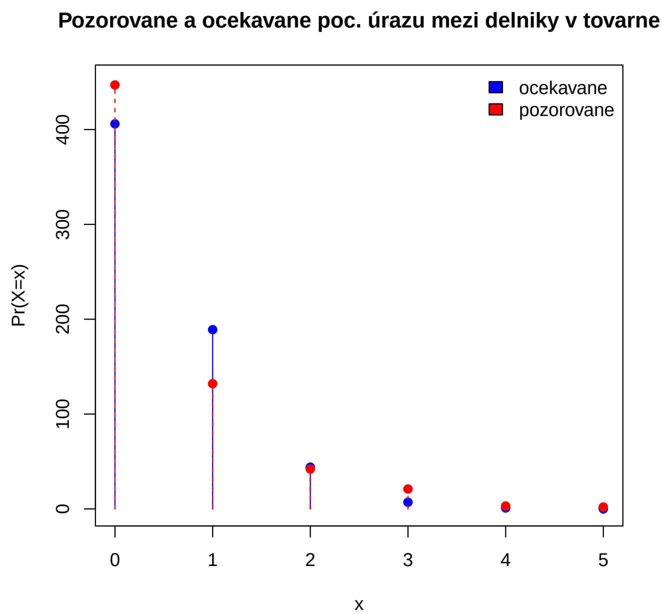
n	0	1	2	3	4	≥ 5
m_n	447	132	42	21	3	2

Vypočítejte očekávané početnosti dělníků za předpokladu, že početnosti úrazů na dělníka X mají Poissonovo rozdělení s parametrem

$$\lambda = \frac{\sum_n nm_n}{\sum_n m_n} = 0.47. \tag{4}$$

Ozn. $X \sim Poiss(\lambda)$.

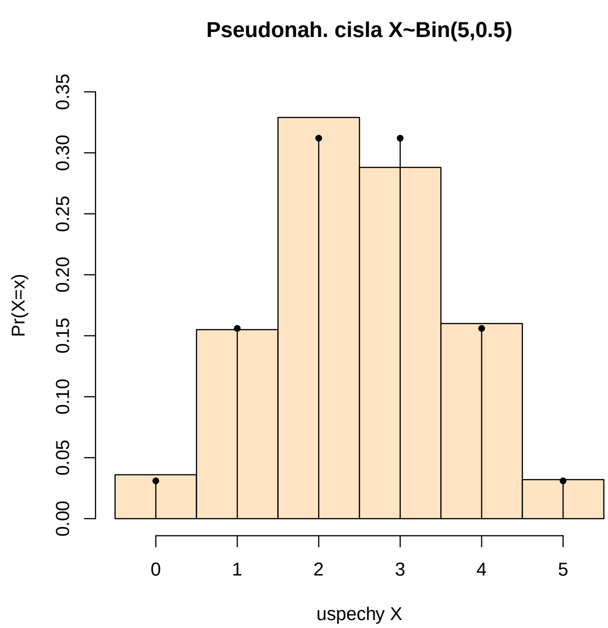
```
##              0   1  2  3 4 5+
## pozorovane 447 132 42 21 3  2
## ocekavane  406 189 44  7 1  0
```




```
## [1] 0.6908308
## [1] 0.4683702
```

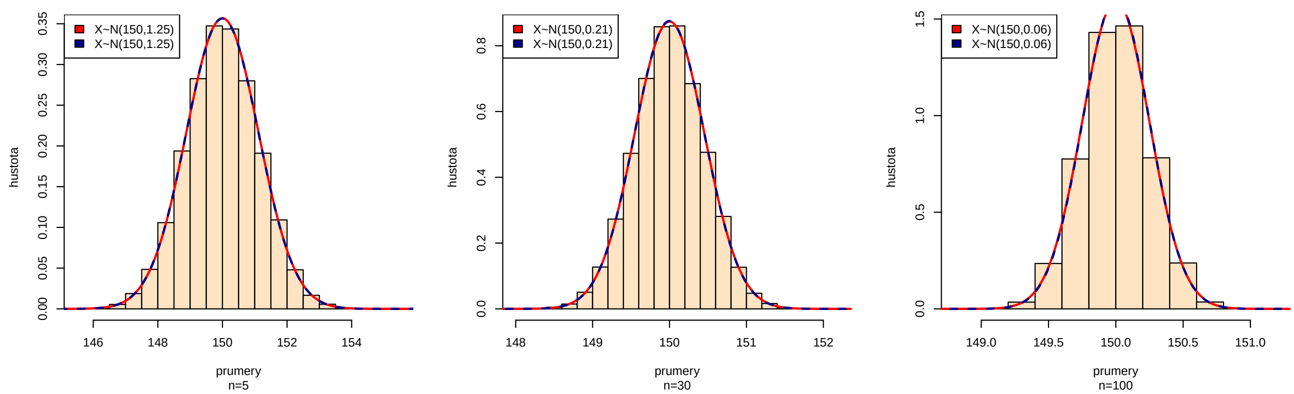
Příklad 30 (binomické rozdělení, simulační studie). Vygenerujte pseudonáhodná čísla X (početnosti úspěchů) opakovaná M-krát ($M = 1000$) z $Bin(N, p)$, kde $N = 5$ a $p = 0.5$. Vytvořte tabulku vygenerovaných (simulovaných) i teoretických relativních početností (pro $n = 0, 1, \ldots, 5$). Superponujte histogram vygenerovaných pseudonáhodných čísel s teoretickou pravděpodobnostní funkcí.

```
##                0     1     2     3     4     5
## simulovane 0.036 0.155 0.329 0.288 0.160 0.032
## teoreticke 0.031 0.156 0.312 0.312 0.156 0.031
```



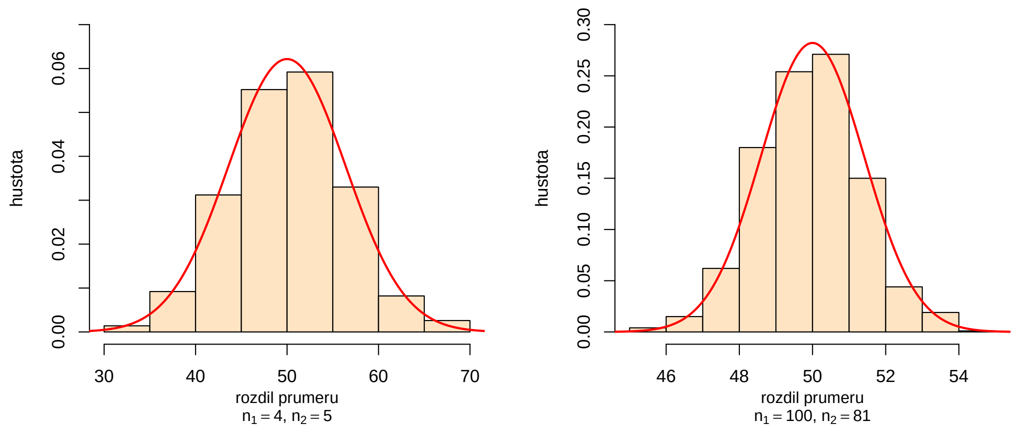
Příklad 31 (normální rozdělení, simulační studie). Na základě simulační studie prověřte, že pokud $X \sim N(150, 6.25)$, potom $\bar{X}_n \sim N(150, \frac{6.25}{n})$. Použijte $n = 30$. Pro každou simulaci X vypočítejte aritmetické průměry $\bar{x}_m$, $m = 1, 2, \ldots, M$, kde $M = 500\,000$. Superponujte je histogramem v relativní škále s teoretickou křivkou hustoty pro $\bar{X}_n$. Vypočítejte $\Pr(\bar{X}_n > 151)$ ze simulovaných dat a porovnejte tento výsledek s teoretickou (očekávanou) pravděpodobností. Řešení viz obrázek **??**.

```
## [1] "teoreticka:" "0.1855"
## [1] "simulovana:" "0.18616"
## [1] "teoreticka:" "0.0142"
## [1] "simulovana:" "0.01374"
## [1] "teoreticka:" "0"
## [1] "simulovana:" "2e-05"
```



Příklad 32 (normální rozdělení, simulační studie). Nechť $X \sim N(\mu_1, \sigma_1^2)$ a $Y \sim N(\mu_2, \sigma_2^2)$. Potom $\bar{X}_{n_1} - \bar{Y}_{n_2} \sim N(\mu_1 - \mu_2, \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2})$. Generujte pseudonáhodná čásla X a Y rozdělení $N(\mu_j, \sigma_j^2)$, $j = 1, 2$, kde $\mu_1 = 100$, $\sigma_1 = 10$, $\mu_2 = 50$, $\sigma_2 = 9$ při (a) $n_1 = 4$, $n_2 = 5$, (b) $n_1 = 100$, $n_2 = 81$. Pro každou simulaci X a Y vypočítejte rozdíl $\bar{x}_m - \bar{y}_m$, $m = 1, 2, \ldots M$, kde $M = 1000$. Superponujte histogram těchto rozdílů v relativní škále s teoretickou křivkou hustoty rozdílu $\bar{X}_{n_1} - \bar{Y}_{n_2}$. Pro případ (a) i (b) vypočítejte $\Pr(\bar{X}_{n_1} - \bar{Y}_{n_2} \leq 52)$ na základě empirického (vygenerovaného) a teoretického rozdělení $\bar{X}_{n_1} - \bar{Y}_{n_2}$.

```
## [1] "teoreticke" "0.622"
## [1] "simulovane" "0.618"
## [1] "teoreticke" "0.921"
## [1] "simulovane" "0.936"
```



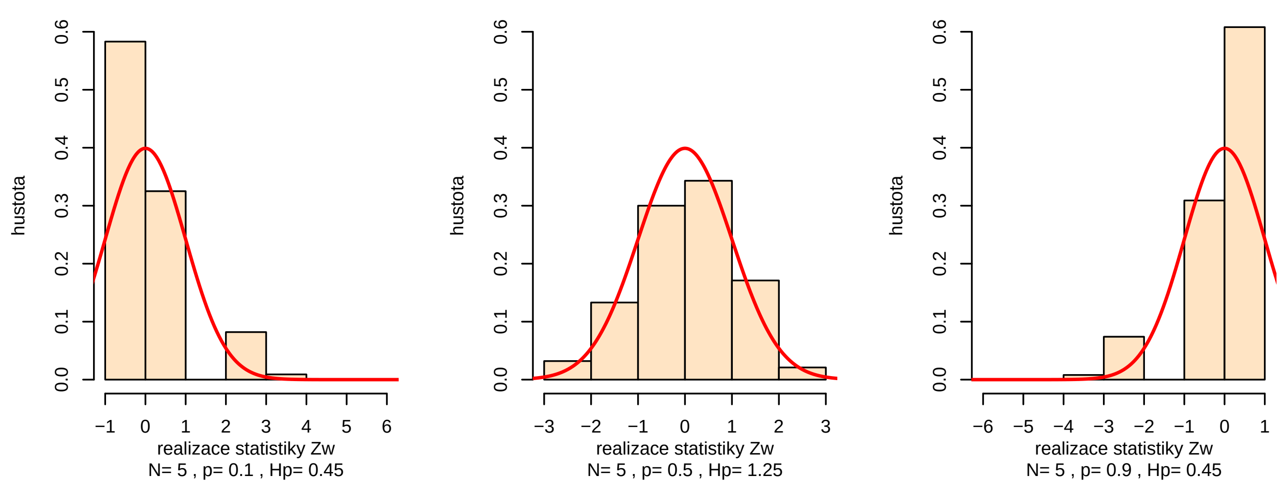
Příklad 33 (statistika). Mějme náhodný výběr $(X_1, X_2, \ldots, X_n)^T$, kde $X_i \in \mathbb{R}, i = 1, 2, \ldots, n$, potom příklady statistik jsou:

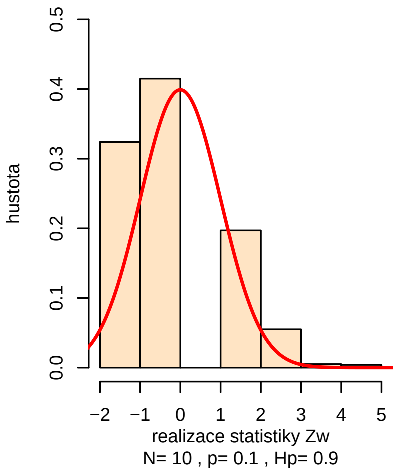
- $T_1 = \sum_{i=1}^{n} X_i \in \mathbb{R}$,
- $T_2 = \sum_{i=1}^{n} X_i^2 \in \mathbb{R}^+ \cup \{0\}$,
- $T_3 = (\sum_{i=1}^{n} X_1, \sum_{i=1}^{n} X_1^2) \in \mathbb{R}^2$.

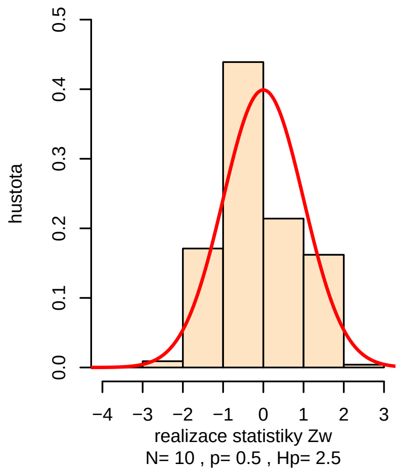
Příklad 34 (testovací statistika, simulační studie). Na základě simulační studie prověřte, že pokud náhodná proměnná X má asymptoticky binomické rozdělení $Bin(N, p)$, potom testovací statistika

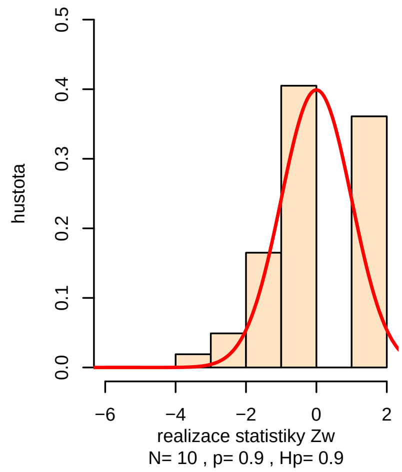
$$Z_W = \frac{X/N - p}{\sqrt{p(1-p)/N}}$$

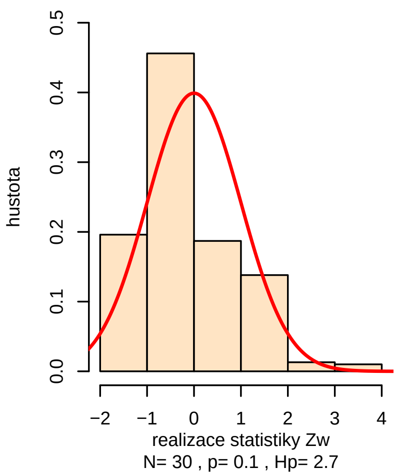
má asymptoticky normální rozdělení $N(0, 1)$. Použijte $p = 0.1$, 0.5, 0.9 a 1, a $N = 5$, 10, 30, 50 a 100. Okomentujte výsledky ve spojitosti s Haldovou podmínkou $Np(1-p) > 9$. Pro každou simulaci X vypočítejte $z_{W,m}$, $m = 1, 2, \ldots, M$, kde $M = 1000$. Superponujte histogram vygenerovaných testovacích statistik v relativní škále s teoretickou křivkou hustoty Z_W.

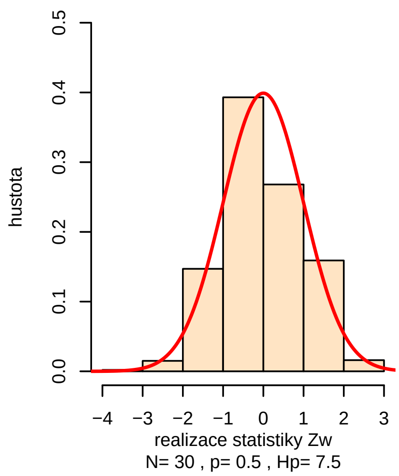


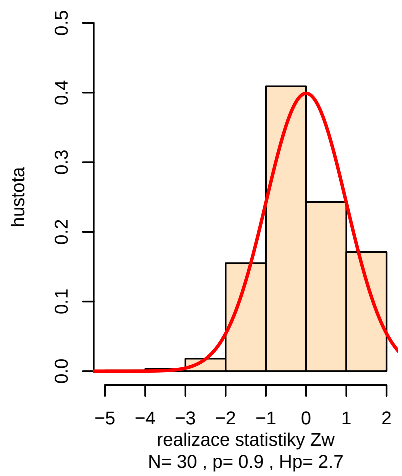

hustota
realizace statistiky Zw
N= 10 , p= 0.1 , Hp= 0.9

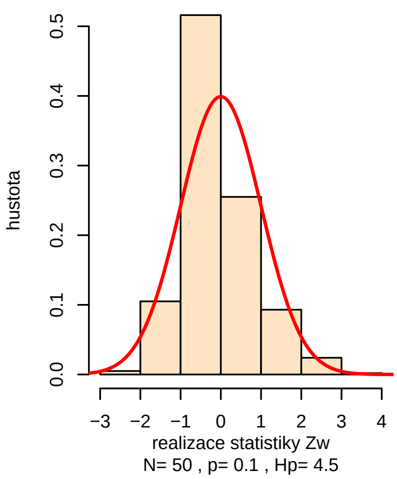

hustota
realizace statistiky Zw
N= 10 , p= 0.5 , Hp= 2.5

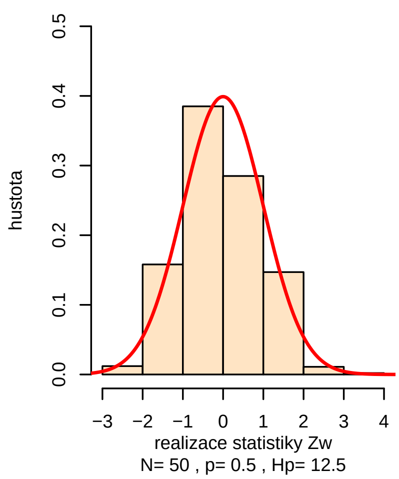

hustota
realizace statistiky Zw
N= 10 , p= 0.9 , Hp= 0.9

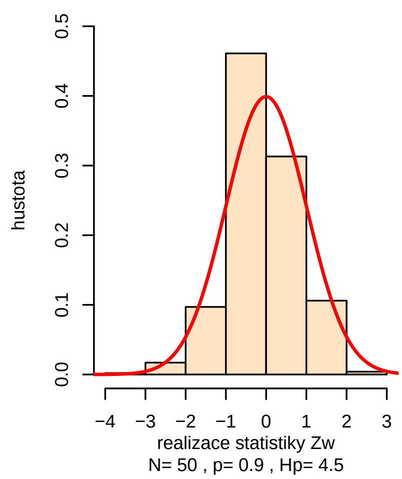

hustota
realizace statistiky Zw
N= 30 , p= 0.1 , Hp= 2.7

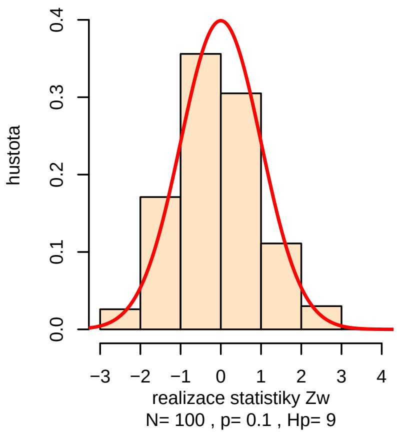

hustota
realizace statistiky Zw
N= 30 , p= 0.5 , Hp= 7.5

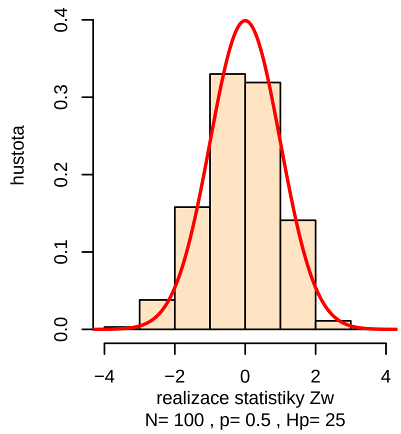

hustota
realizace statistiky Zw
N= 30 , p= 0.9 , Hp= 2.7

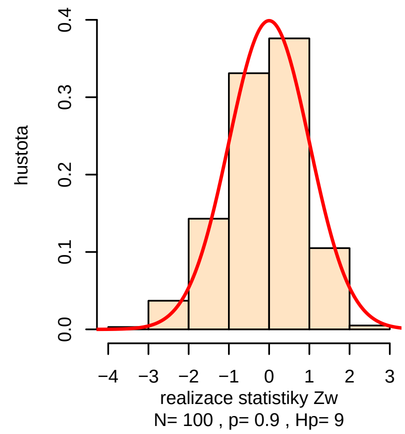

hustota
realizace statistiky Zw
N= 50 , p= 0.1 , Hp= 4.5


hustota
realizace statistiky Zw
N= 50 , p= 0.5 , Hp= 12.5


hustota
realizace statistiky Zw
N= 50 , p= 0.9 , Hp= 4.5







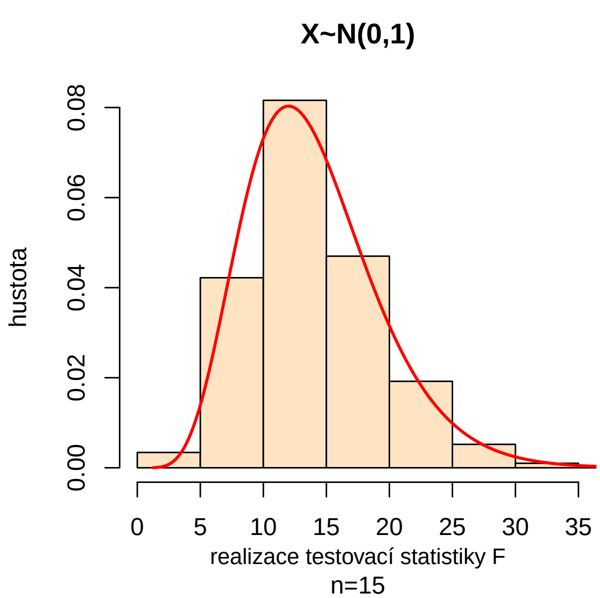
Příklad 34 mluví o použití jednovýběrové testovací statistiky pro parametr binomického rozdělení (pravděpodobnost) pro různé pravděpodobnosti a různé početnosti. Pokud není Haldova odmínka splněná,není možné testovací statistiku použít.

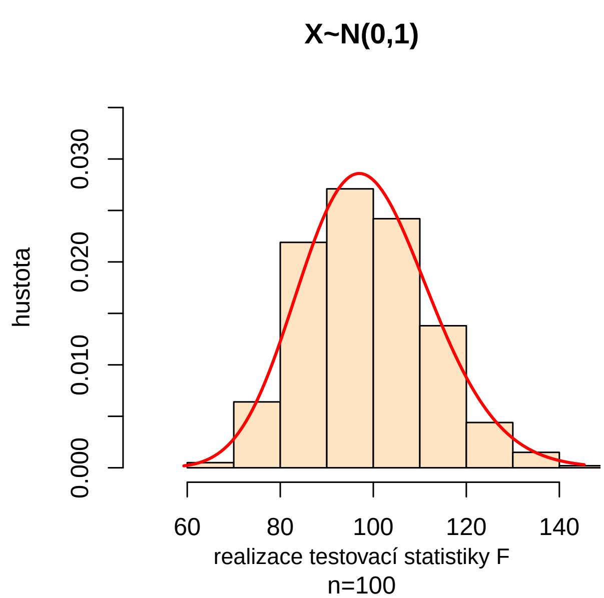
Příklad 35 (testovací statistika, simulační studie)**.** Na základě simulační studie prověřte, že pokud

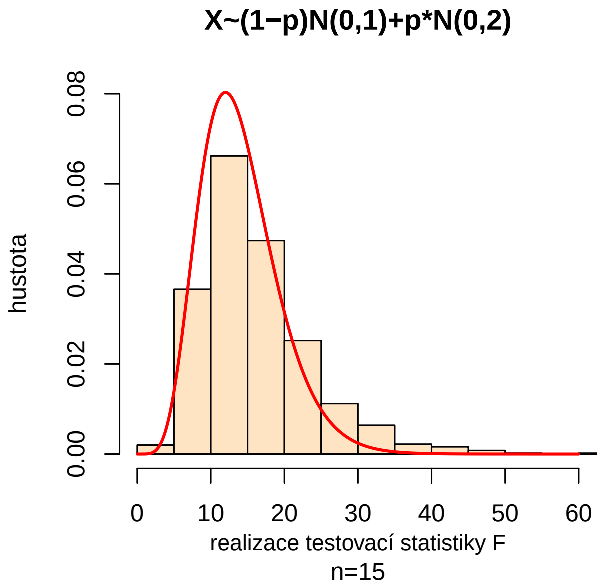
a) $X \sim N(\mu, \sigma^2)$, kde $\mu = 0$, $\sigma^2 = 1$;

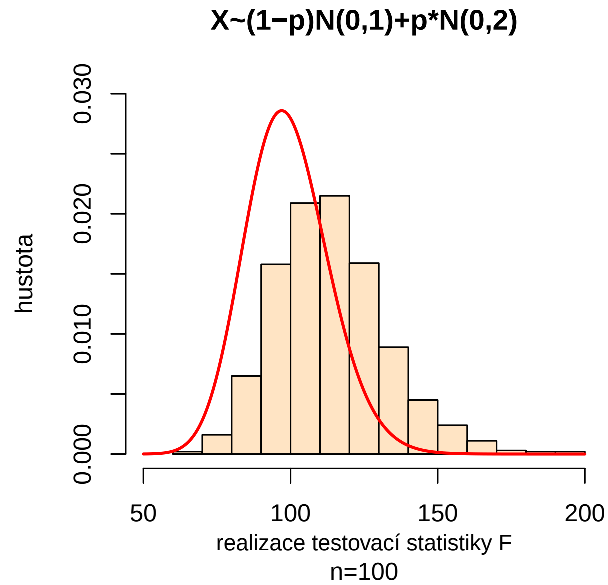
b) $X \sim [(1-p)N(\mu, \sigma^2) + pN(\mu, \sigma_1^2)]$, kde $\mu = 0$, $\sigma^2 = 1$, $p = 0.05$, $\sigma_1^2 = 2$,

potom testovací statistika $F = \frac{(n-1)S^2}{\sigma^2}$ má asymptoticky χ^2_{n-1} rozdělení o $n-1$ stupních volnosti. Použijte rozsahy náhodných výběrů $n = 15$ a $n = 100$. Pro každou simulaci X vypočítejte $F_{poz,m}$, $m = 1, 2, \ldots, M$, kde $M = 1000$. Superponujte histogram vygenerovaných testovacích statistik v relativní škále s teoretickou křivkou hustoty F.









Příklad 36 (**hypergeometrické rozdělení**). Koupili jsme 10 cibulek červených tulipánů a 5 cibulek žlutých tulipánů. Zasadili jsme 8 náhodně vybraných cibulek.

a) Jaká je pravděpodobnost, že žádná nebude cibulka žlutých tulipánů?

b) Jaká je pravděpodobnost, že jsme zasadili všech 5 cibulek žlutých tulipánů?

c) Jaká je pravděpodobnost, že aspoň dvě budou cibulky žlutých tulipánů?

Příklad 37 (**hypergeometrické rozdělení**). Dítě dostalo sáček, v němž bylo 5 červených a 5 žlutých bonbónů. Dítě náhodně vybralo ze sáčku 6 bonbónů. Jaká je pravděpodobnost, že mezi vybranými bonbóny budou právě 2 červené?

Příklad 38 (**multinomické rozdělení – definice**). Nechť N je počet nezávislých identických pokusů a v každém z nich může nastat $J \geq 2$ navzájem disjunktních událostí s možnými odpověďmi $X_{ij} = 1$ (událost nastala) nebo $X_{ij} = 0$ (událost nenastala), kde $i = 1, 2, \ldots, N$ a $j = 1, 2, \ldots, J$. Potom $X_j = \sum_{i=1}^{N} X_{ij}$. Pravděpodobnost nastání j-té události v i-tém pokuse $\Pr(X_{ij} = 1) = p_j$, $\sum_{j=1}^{J} p_j = 1$. Náhodná proměnná $\mathbf{X} = (X_1, \ldots, X_J)^T$ má (J-rozměrné) multinomické rozdělení s parametry N a $\mathbf{p}$, t.j. $\mathbf{X} \sim \text{Mult}_J(N, \mathbf{p})$. Pravděpodobnost, že X_j je rovné nějakému číslu n_j zapisujeme jako

$$\Pr(X_1 = x_1, \ldots, X_J = x_J) = \frac{N!}{x_1! x_2! \ldots x_J!} p_1^{x_1} p_2^{x_2} \ldots p_j^{x_J} = \frac{N!}{\prod_j x_j!} \prod_{j=1}^{J} p_j^{x_j},$$

kde $N = \sum_{j=1}^{J} X_j$, $X_j \geq 0$ a $x_j = n_j$ jsou realizace X_j. Potom $\mathbf{n} = (n_1, n_2, \ldots, n_J)^T$. Pro marginální rozdělení píšeme $X_j \sim \text{Bin}(N, p_j)$, kde střední hodnota $E[X_j] = Np_j$, rozptyl $\text{Var}[X_j] = Np_j(1-p_j)$, kovariance $\text{Cov}[X_i, X_j] = -Np_ip_j$, korelační koeficient $\text{Cor}[X_i, X_j] = (-p_ip_j)/\sqrt{p_i(1-p_i)p_j(1-p_j)}$. Střední hodnota $E[\mathbf{X}] = N\mathbf{p}$ a kovarianční matice $\text{Var}[\mathbf{X}] = N(\mathbf{D_p} - \mathbf{pp}^T)$, kde $\mathbf{D_p} = \text{diag}(\mathbf{p})$ a

$$(\mathbf{D_p} - \mathbf{pp}^T)_{ij} = \begin{cases} p_i(1-p_i) & \text{pokud i=j} \\ -p_ip_j & \text{pokud } i \neq j. \end{cases}$$

Příklad 39 (multinomické rozdělení). Mějme proměnnou barva vlasů (blond – BlH, hnědá – BrH, zrzavá – RH) a proměnnou barva očí (modrá – BlE, hnědá – BrE, zelená – GE). Jejich interakce jsou uspořádané v tabulce jako X_1 (BlH-BlE), X_2 (BlH-BrE), X_3 (BlH-GE), X_4 (BrH-BlE), X_5 (BrH-BrE), X_6 (BrH-GE), X_7 (RH-BlE), X_8 (RH-BrE), X_9 (RH-GE). Předpokládejme, že máme náhodný výběr s rozsahem $N = 100$. Pravděpodobnosti p_j, $j = 1, \dots, 9$ viz následující tabulka.

Barva vlasů / barva očí	modrá (BlE)	hnědá (BrE)	zelená (GE)
blond (BlH)	0.12	0.15	0.03
hnědá (BrH)	0.22	0.34	0.04
zrzavá (RH)	0.06	0.01	0.03

Vypočítejte $E[X_2]$, $E[X_8]$, $\mathrm{Var}[X_2]$, $\mathrm{Var}[X_8]$, $\mathrm{Cov}[X_2, X_8]$ a $\mathrm{Cor}[X_2, X_8]$.

Příklad 40 (součinové multinomické rozdělení – definice). Nechť N_k je počet nezávislých identických pokusů a v každém z nich může nastat $J \geq 2$ navzájem disjunktních událostí s možnými odpověďmi $X_{kji} = 1$ (událost nastala) anebo $X_{kji} = 0$ (událost nenastala), kde $i = 1, 2, \dots, N_k$, $k = 1, 2, \dots, K$ a $j = 1, 2, \dots, J$. Nechť $X_{kj} = \sum_{i=1}^{N_k} X_{kji}$ a $\sum_{k=1}^{K} N_k = N$. Pravděpodobnost nastání (j)-té události v i-tém pokuse k-té skupinyje $\Pr(X_{kji} = 1) = p_{kj} = p_{j|k}$, $\sum_{j=1}^{J} p_{kj} = 1$. Náhodná proměnná $\mathbf{X}_k = (X_{k1}, X_{k2}, \dots, X_{kJ})^T$ má (J-rozměrné) multinomické rozdělení s parametry N_k a $\mathbf{p}_k = (p_{k1}, \dots, p_{kJ})^T$, t.j. $\mathbf{X}_k \sim \mathrm{Mult}_J(N_k, \mathbf{p}_k)$.Realizace náhodné proměnné $\mathbf{X}_k$ označujeme jako $\mathbf{x}_k$. Potom $x_{kj} = n_{kj}$ a navíc $\mathbf{n}_k = (n_{k1}, n_{k2}, \dots, n_{kJ})^T$. Nechť $\mathbf{X}_k$ jsou nezávislé, potom $\mathbf{X} = (\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_K)^T$ má součinové multinomické rozdělení s parametry $\boldsymbol{\theta}_k = \mathbf{p}_k$, $k = 1, 2, \dots, K$.

Příklad 41 (součinové multinomické rozdělení).

1. Mějme data z příkladu 39 a náhodný výběr s rozsahy $N_1 = 30$ pro blond barvu vlasů, $N_2 = 60$ pro hnědou barvu vlasů a $N_3 = 10$ pro zrzavou barvu vlasů. Označme interakce proměnných následovně: $X_{11} = X_{1|1}$ (BlH-BlE), $X_{12} = X_{2|1}$ (BlH-BrE), $X_{13} = X_{3|1}$ (BlH-GE), $X_{21} = X_{1|2}$ (BrH-BlE), $X_{22} = X_{2|2}$ (BrH-BrE), $X_{23} = X_{3|2}$ (BrH-GE), $X_{31} = X_{1|3}$ (RH-BlE), $X_{32} = X_{2|3}$ (RH-BrE), $X_{33} = X_{3|3}$ (RH-GE), kde $\mathbf{X}_1 = (X_{11}, X_{12}, X_{13})^T$, $\mathbf{X}_2 = (X_{21}, X_{22}, X_{230})^T$ a $\mathbf{X}_3 = (X_{31}, X_{32}, X_{33})^T$. Potom $\mathbf{X} = (\mathbf{X}_1, \mathbf{X}_2, \mathbf{X}_3)^T$ má součinové multinomické rozdělení s $K = 3$, $N_1 = 30$, $J_1 = 3$, $N_2 = 60$, $J_2 = 3$ a $N_3 = 10$, $J_3 = 3$. Zápis s $X_{j|k}$, kde $j = 1, 2, 3$ a $k = 1, 2, 3$ zvýrazňuje fakt, že rozdělení je podmíněno barvou vlasů, t.j. rozdělení ve sloupcích tabulky je podmíněné jejím řádkem. Realizace $x_{j|k}$ značíme jako $n_{j|k}$, pravděpodobnosti ekvivalentní $X_{j|k} = x_{j|k}$ značíme jako $p_{j|k} = p_{kj}$. Vypočítejte podmíněné pravděpodobnosti $p_{j|k}$, očekávané početnosti $N_k p_{kj}$, $\mathrm{Var}[X_{22}]$, $\mathrm{Var}[X_{32}]$, $\mathrm{Var}[X_{23}]$, $\mathrm{Cov}[X_{22}, X_{32}]$, $\mathrm{Cov}[X_{22}, X_{32}]$ a $\mathrm{Cor}[X_{22}, X_{32}]$, $\mathrm{Cor}[X_{22}, X_{23}]$.

2. Celý postup zopakujte pro $N_1 = 20$, $N_2 = 30$ a $N_3 = 50$.

Příklad 42 (součinové multinomické rozdělení). Mějme proměnnou barva vlasů (blond – BlH, hnědá – BrH, zrzavá – RH) a proměnnou barva očí (modrá – BlE, hnědá – BrE, zelená – GE). Jejich interakce jsou uspořádané v tabulce jako X_1 (BlH-BlE), X_2 (BlH-BrE), X_3 (BlH-GE), X_4 (BrH-BlE), X_5 (BrH-BrE), X_6 (BrH-GE), X_7 (RH-BlE), X_8 (RH-BrE), X_9 (RH-GE). Jim odpovídají pravděpodobnosti p_j, $j = 1, \dots, 9$

Barva vlasů / barva očí	modrá (BlE)	hnědá (BrE)	zelená (GE)
blond (BlH)	0.12	0.15	0.03
hnědá (BrH)	0.22	0.34	0.04
zrzavá (RH)	0.06	0.01	0.03

$\mathbf{X} = (X_1, X_2, \dots, X_9)^T \sim \mathrm{Mult}_9(N, \mathbf{p})$. Transformujte multinomický model na součinový multinomický model následovně:

1. vypočítejte řádkově marginální pravděpodobnosti $p_{j.}$;

2. vypočítejte sloupcově marginální pravděpodobnosti $p_{.k}$;

3. podmíněné pravděpodobnosti $p_{j|k} = p_{kj}$;

Jakému číslu jsou rovné sumy $\sum_{j=1}^{3} p_{j|k}$ pro každé k?