Písemka 10.1.07

Cvičení 0.0.1: Fotbalisté B, C, D promění penaltu s pravděpodobnostmi po řadě $b = 0.9$, $c = 0.8$, $d = 0.6$. Určili střelce losem (všichni měli stejnou pravděpodobnost). Jaká je pravděpodobnost proměnění penalty? Pokud víme, že se penalta nezdařila, jaká je pro jednotlivé hráče pravděpodobnost, že ji kopali?

Řešení: Pravděpodobnost úspěchu je $(b + c + d)/3 \doteq 0.767$, neúspěchu $q \doteq 1 - 0.767 = 0.233$. Podmíněné pravděpodobosti střelců neúspěšného pokusu vypočítáme z Bayesova vzorce: $\frac{(1-b)/3}{q} = 0.143$, $\frac{(1-c)/3}{q} = 0.286$, $\frac{(1-d)/3}{q} = 0.572$.

Cvičení 0.0.2: Náhodná veličina X nabývá hodnot z intervalu $\langle 0, 100\rangle$, její medián je $q_X(1/2) = 70$. Co lze říci o její střední hodnotě a rozptylu?

Řešení: Jelikož $X \leq 70$ s pravděpodobností aspoň $1/2$ a ve zbývajících případech je $X \leq 100$, musí být $EX \leq (70 + 100)/2 = 85$. Podobně dostaneme dolní odhad $EX \geq 70/2 = 35$. Přehledněji lze tuto úvahu provést, chápeme-li střední hodnotu jako integrál kvantilové funkce, její hodnoty jsou omezeny po částech konstantními funkcemi s uvedenými integrály.

Rozptyl bude zřejmě maximální, bude-li náhodná veličina nabývat co možno nejčastěji jen své extrémní hodnoty, 0 a 100, a to s pravděpodobnostmi danými střední hodnotou. Podle vzorce pro rozptyl alternativního rozdělení (zde vynásobeného 100) je bez omezení mediánu rozptyl maximálně $100^2 (1 - EX) EX$, největší může být pro $EX = 1/2$, a to $100^2/4 = 2500$. V tom případě by podmínka na medián nebyla splněna, ta však má jen lokální vliv. Můžeme vzít např. rozdělení s distribuční funkcí

$$F_X(u) = \begin{cases} 0 & \text{pro } u < 0, \\ \frac{1 - \left(\frac{70-u}{70}\right)^k}{2} & \text{pro } 0 \leq u < 70, \\ \frac{1 + \left(\frac{u-70}{30}\right)^k}{2} & \text{pro } 70 \leq u < 100, \\ 1 & \text{pro } u \geq 100. \end{cases}$$

Zvolíme-li kladnou konstantu k dostatečně velkou, rozptyl bude libovolně blízký vypočtené mezi, vždy však $DX < 2500$.

Cvičení 0.0.3: Pacient si měřil teplotu vždy současně dvěma teploměry, hodnoty jsou v tabulce:

teplota1	37.5	38.2	38.3	38.6	38.2	37.6	37.2	36.9	36.6
teplota2	37.7	38.3	38.5	38.6	38.3	37.7	37.4	36.7	36.5

Posuďte na hladině významnosti 5% hypotézu, že mezi průměrnými údaji teploměrů není rozdíl. Uveďte použité předpoklady.

Řešení: Rozdíly teplot jsou $\delta = (-0.2, -0.1, -0.2, 0, -0.1, -0.1, -0.2, 0.2, 0.1)$, $n = 9$, $\bar{\delta} = -0.0667$, $s_\Delta^2 \doteq 0.02$, $s_\Delta \doteq 0.141$,

$$t = \frac{\bar{\delta}}{s_\Delta}\sqrt{n} \doteq -1.41,$$

porovnáme s kvantilem $q_{t(8)}(0.025) = -q_{t(8)}(0.975) \doteq -2.306$ [1] a nulovou hypotézu nezamítáme. Předpoklady pro párový pokus: střední hodnoty náhodných veličin v obou výběrech kolísají stejně, odchylky od nich mají normální rozdělení a jsou nezávislé.

Písemka 22.1.07

Cvičení 0.0.4: Náhodné veličiny X, Y jsou nezávislé, mají diskrétní rovnoměrná rozdělení; X na $\{0, 1\}$, Y na $\{0, 1, 2\}$. Popište a znázorněte rozdělení náhodných veličin (a) $X/2 + Y/2$, (b) $\mathrm{Mix}_{1/2}(X, Y)$, (c) $X/2 + EY/2$.

[1] V Rogalewiczově skriptu je na tomto místě tabulky překlep.

Řešení: (a) $X/2+Y/2$ nabývá hodnot $0, 1/2, 1, 3/2$ s pravděpodobnostmi po řadě $1/6, 1/3, 1/3, 1/6$ (pouze zde potřebujeme nezávislost), (b) $\mathrm{Mix}_{1/2}(X,Y)$ nabývá hodnot $0, 1, 2$ s pravděpodobnostmi po řadě $5/12, 5/12, 1/6$, (c) jelikož $EY = 1$, $X/2 + EY/2$ nabývá hodnot $1/2, 1$ s pravděpodobností $1/2$.

Cvičení 0.0.5: Opisujeme data. Při čtení se z chybných položek ve vstupních datech odhalí a opraví 80%. Při zápisu se do 1% položek dostanou nové chyby. Při jakém výskytu chyb se opsáním jejich počet v průměru sníží?

Řešení: Apriorní pravděpodobnost chybné položky označme c, pravděpodobnost správné položky $1 - c$. Změnu pravděpodobností čtením a zápisem kvantifikuje násobení maticemi podmíněných pravděpodobností (předpokládáme, že chybná položka se chybou při zápisu nemůže opravit):

$$\begin{bmatrix} c & 1-c \end{bmatrix} \begin{bmatrix} 0.2 & 0.8 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0.01 & 0.99 \end{bmatrix} = \begin{bmatrix} c & 1-c \end{bmatrix} \begin{bmatrix} 0.208 & 0.792 \\ 0.01 & 0.99 \end{bmatrix} = \begin{bmatrix} 0.198\,c + 0.01 & -0.198\,c + 0.99 \end{bmatrix} .$$

Pravděpodobnost (nikoli nutně četnost!) chybných položek se sníží, pokud $0.198\,c + 0.01 < c$, tj. přibližně $c > 0.0125 = 1.25\%$.

Cvičení 0.0.6: Náhodná veličina X nabývá hodnot $1, 2, 3, 4, 5, 6$. Její realizací jsme dostali následující četnosti výsledků:

hodnota	1	2	3	4	5	6
pozorovaná četnost	2	3	5	13	10	12

Posuďte na hladině významnosti 5% hypotézu, že náhodná veličina X má rozdělení, při němž pravděpodobnost je úměrná hodnotě, tj. $P[X = k] = ck$, kde c je konstanta.

Řešení: Konstantu c určíme z rovnice (nikoli z dat):

$$1 = \sum_k P[X = k] = 21c\,,$$

$c = 1/21$. Sloučíme první dvě skupiny, jinak by jejich teoretické četnosti byly příliš malé:

hodnota	$1-2$	3	4	5	6	celkem
pozorovaná četnost	5	5	13	10	12	45
teoretická pravděpodobnost	1/7	1/7	4/21	5/21	2/7	1
teoretická četnost	6.429	6.429	8.5711	10.714	12.857	45
příspěvek ke kritériu	0.317	0.317	2.288	0.048	0.057	3.028

Porovnáme hodnotu kritéria 3.028 s kvantilem $q_{\chi^2(4)}(0.95) \doteq 9.488$ a hypotézu nezamítáme.

Písemka 29.1.07

Cvičení 0.0.7: Náhodné veličiny $X_i, i = 1, 2, 3$, jsou nezávislé a mají parametry $EX_i = i$, $\sigma_{X_i} = i$. Určete střední hodnotu a směrodatnou odchylku náhodné veličiny $Y = X_1 + 2X_2 - X_3$.

Řešení: $EY = EX_1 + 2EX_2 - EX_3 = 2$, $\sigma_Y = \sqrt{DY} = \sqrt{DX_1 + 2^2 DX_2 + DX_3} = \sqrt{1 + 2^2 4 + 9} = \sqrt{26} \doteq 5.099$.

Cvičení 0.0.8: V letadle je $j = 216$ míst pro cestující. V průměru $q = 5\%$ cestujících se k odletu nedostaví. Kolik letenek lze prodat, aniž by riziko, že se do letadla nevejdou, nebylo větší než $\alpha = 2\%$? Posuďte použité předpoklady.

Řešení: Z m cestujících přijde počet, daný náhodnou veličinou X s binomickým rozdělením $Bi(m, 1-q)$. Potřebujeme splnit nerovnost $q_X(1-\alpha) \le j$. Exaktní výpočet by vyžadoval určit m zkusmo. Místo toho použijeme centrální limitní větu a s rozdělením náhodné veličiny X budeme pracovat, jako by bylo normální, s parametry $EX = m\,(1-q)$, $DX = mq\,(1-q)$. Znormováním X dostaneme

$$\frac{j - EX}{\sigma_X} \ge \Phi^{-1}(1-\alpha)\ ,$$

$$j \ge EX + \sigma_X\,\Phi^{-1}(1-\alpha) = m\,(1-q) + \sqrt{mq\,(1-q)}\,\Phi^{-1}(1-\alpha)\ .$$

Z nerovnice

$$216 \geq 0.95\,m + \sqrt{0.05\,0.95\,m}\,\Phi^{-1}\,(0.98) \doteq 0.95\,m + 2.05375\,\sqrt{0.05\,0.95\,m}$$

pro kladná m dostáváme $m \leq 220.\,374$, tedy celočíselná řešení $m \leq 220$. Použili jsme předpoklady centrální limitní věty, tedy nezávislost alternativních náhodných veličin, vytvářejících binomické rozdělení, a jejich dostatečně velký počet. Obojí lze zpochybnit. Zrušení letu dvou cestujících nemusí být nezavislé, pokud např. chtěli letět na společnou dovolenou. Počet $m = 220$ se zdá být dostatečný na to, abychom směli použít přibližné řešení pomocí centrální limitní věty, avšak z něj by v průměru zrušilo let jen $m\,(1-q) = 11$ cestujících. Je tedy žádoucí výsledek ověřit pomocí binomického rozdělení. Pro $m = 220$ vyjde

$$\sum_{k=217}^{m} \binom{m}{k}\,(1-q)^k\,q^{m-k} \doteq 4.\,2 \times 10^{-3}\,,$$

riziko je příliš velké; pro $m = 219$ vyjde $1.\,040 \times 10^{-3}$, což vyhovuje.

Cvičení 0.0.9: Náhodná veličina nabývá hodnot $j \in \{0, 1, 2\}$ s pravděpodobnostmi $a + bj$. Odhadněte parametry a, b z četností dle tabulky:

hodnota	0	1	2
četnost	17	15	8

Řešení: Jelikož součet pravděpodobností musí být jednotkový, dostáváme rovnici $3a + 3b = 1$, tj. $b = 1/3 - a$. Zbývá odhadnout jediný parametr.

Metoda momentů: Střední hodnota je

$$(a+b) + 2\,(a+2b) = 3a + 5b = \frac{5}{3} - 2a\,,$$

její odhad – výběrový průměr

$$\frac{1}{17+15+8}\,(15 + 8 \cdot 2) = \frac{31}{40}\,.$$

Rovnice $\frac{5}{3} - 2\hat{a} = \frac{31}{40}$ má řešení $\hat{a} = \frac{107}{240} \doteq 0.445\,833$, $\hat{b} = \frac{-9}{80} \doteq -0.112\,5$.

Metoda maximální věrohodnosti:

$$\ell(a) = 17 \ln a + 15 \ln (a+b) + 8 \ln (a+2b) = 17 \ln a + 15 \ln \frac{1}{3} + 8 \ln \left(\frac{2}{3} - a\right)\,,$$

derivace

$$\frac{\partial \ell(a)}{\partial a} = \frac{17}{a} - \frac{8}{\frac{2}{3} - a}$$

je nulová pro $\hat{a} = \frac{34}{75} \doteq 0.453\,333$, $\hat{b} = \frac{-3}{25} = -0.12$.

Písemka 16.2.07

Cvičení 0.0.10: Mezi velkým množstvím hracích kostek je 1/2 falešných, na nichž padá šestka s pravděpodobností 1/3. Hodili jsme dvěma náhodně vybranými kostkami, padly dvě šestky. Jaká je pravděpodobnost, že obě kostky jsou falešné?

Řešení: Jevy: F = falešná kostka, S = padla šestka. Dáno: $P\,(S|F) = 1/3$, $P\,(S|\neg F) = 1/6$, $P\,(F) = 1/2$. Pravděpodobnost, že jedna kostka, na níž padla šestka, je falešná, je

$$P\,(F|S) = \frac{P\,(S|F)\;P\,(F)}{P\,(S|F)\;P\,(F) + P\,(S|\neg F)\;P\,(\neg F)} = \frac{c\,f}{c\,f + a\,(1-f)} = \frac{c}{c - a + a/f} = \frac{2}{3}\,.$$

Pravděpodobnost, že nastanou dva nezávislé jevy s touto pravděpodobností, je

$$(P\,(F|S))^2 = \frac{4}{9}\,.$$

Cvičení 0.0.11: Nezávislé náhodné veličiny X, Y mají spojité rovnoměrné rozdělení na intervalu $\langle 0,2\rangle$, resp. $\langle 0,3\rangle$. Určete střední hodnotu a rozptyl veličiny $Y - 2\,X$.

Řešení: $E\,(Y-2\,X) = EY - 2\,EX$, $D\,(Y-2\,X) = DY + 2^2\,DX$.

veličina	X	Y	$Y-2X$
střední hodnota	1	$\frac{3}{2}$	$\frac{-1}{2}$
rozptyl	$\frac{1}{3}$	$\frac{3}{4}$	$\frac{25}{12}$

Cvičení 0.0.12: Ve vzorku je 10^6 atomů radioaktivního izotopu. Určete symetrický 99%-ní intervalový odhad počtu atomů, které se rozpadnou za dvojnásobek poločasu rozpadu.

Řešení: Odhadujeme náhodnou veličinu X s rozdělením $Bi\,(n,p)$, $n = 10^6$, $p = 1-(1/2)^2 = 3/4$ (=pravděpodobnost, že se atom v daném čase rozpadne), $EX = n\,p = 750\,000$, $DX = n\,p\,(1-p) = 187\,500$, $\sigma_X = \sqrt{n\,p\,(1-p)} = 250\sqrt{3} \doteq 433$. Při aproximaci normálním rozdělením vyjdou meze $EX \pm \sigma_X\,\Phi^{-1}\,(0.995) \doteq 750\,000 \pm 433 \cdot 2.576 \doteq 750\,000 \pm 1\,115$, interval přibližně $\langle 748\,885, 751\,115\rangle$.

Písemka 21.2.07

Cvičení 0.0.13: Diskrétní náhodné veličiny X, Y jsou nezávislé, X má rovnoměrné rozdělení na množině $\{1,2,4\}$, Y nabývá hodnot z $\{0,1\}$, $P\,[Y=1] = 2/3$. Určete střední hodnotu a rozptyl veličiny $2\,Y - X$.

Řešení: $E\,(2\,Y - X) = 2\,EY - EX$, $D\,(2\,Y - X) = 2^2\,DY + DX$.

veličina	X	Y	$2Y-X$
střední hodnota	$\frac{7}{3}$	$\frac{2}{3}$	-1
rozptyl	$\frac{14}{9}$	$\frac{2}{9}$	$\frac{22}{9}$

Cvičení 0.0.14: Opilí řidiči způsobili 10% nehod. Při kontrolách bylo 1.5% řidičů opilých. Porovnejte rizika havárie opilého a střízlivého řidiče.

Řešení: Jevy: A = opilý, H = havárie. Dáno: $P\,(A|H) = 0.1\,P\,(H)$, $P\,(A) = 0.015$. Pravděpodobnost havárie není uvedena, ale lze vypočítat, kolikrát větší je riziko havárie opilého řidiče než střízlivého:

$$P\,(H|A) = \frac{P\,(A|H)\;P\,(H)}{P\,(A)}\,,$$
$$P\,(H|\neg A) = \frac{P\,(\neg A|H)\;P\,(H)}{P\,(\neg A)}\,,$$
$$\frac{P\,(H|A)}{P\,(H|\neg A)} = \frac{P\,(A|H)\;P\,(\neg A)}{P\,(\neg A|H)\;P\,(A)} = \frac{P\,(A|H)\;(1-P\,(A))}{(1-P\,(A|H))\;P\,(A)} = \frac{0.1\cdot 0.985}{0.9\cdot 0.015} \doteq 7.3\,.$$

Cvičení 0.0.15: Realizací náhodné veličiny X jsme dostali následující četnosti výsledků:

hodnota	0	1	2	3	4	5	6
pozorovaná četnost	30	13	10	5	2	3	1

Posuďte na hladině významnosti 5% hypotézu, že náhodná veličina X má geometrické rozdělení, $P\,[X=k] = c\,2^{-k}$, kde c je konstanta a $k = 0,1,2,3,\ldots$.

Řešení: Konstantu c určíme z rovnice (nikoli z dat):

$$1 = \sum_k P\,[X=k] = 2\,c\,,$$

$c = 1/2$. Sloučíme výsledky „3 a více“do jedné skupiny, aby teoretické četnosti nebyly příliš malé:

hodnota	0	1	2	3 a více	celkem
pozorovaná četnost	30	13	10	1	64
teoretická pravděpodobnost	1/2	1/4	1/8	1/8	1
teoretická četnost	32	16	8	8	64
příspěvek ke kritériu	0.125	0.5625	0.5	1.125	2.3125

Porovnáme hodnotu kritéria 2.3125 s kvantilem $q_{\chi^2(3)}\,(0.95) \doteq 7.81$ a hypotézu nezamítáme.