

## Písemka 2. února 2017: vzor řešení

**Příklad 1** (8 bodů). Cyril a Dana hrají jednoduchou hru. Pokud na kostce padne nejvýše 4, zaplatí Cyril Daně 1 dolar. Padne-li 5 nebo 6, zaplatí Dana dolar Cyrilovi. Na začátku má Cyril 2 dolary a Dana 1 dolar. Hra končí ve chvíli, kdy jeden z hráčů nemá nic.

- (a) S jakou pravděpodobností vyhraje Cyril?
- (b) Jaká je střední hodnota počtu hodů do konce hry?
- (c) Jaká je podmíněná střední hodnota počtu hodů za podmínky, že vyhrál Cyril?
- (d)\* S jakou pravděpodobností vyhraje Cyril, pokud na začátku má Cyril 4 dolary a Dana 2 dolary?

**Řešení 1.** Označme  $C$  Cyrilovy peníze. Na začátku je  $C = 2$ . Zřejmě je jsou jen dva stavy, kdy hra není u konce, konkrétně  $C = 2$  a  $C = 1$ . Ze stavu  $C = 2$  s pravděpodobností  $1/3$  Cyril vyhraje a s pravděpodobností  $2/3$  přejde do stavu  $C = 1$ . Ze stavu  $C = 1$  s pravděpodobností  $1/3$  přejdeme do počátečního stavu  $C = 2$  a s pravděpodobností  $2/3$  hra skončí Cyrilovou prohrou. Cyril tedy vyhraje s pravděpodobností

$$p = \frac{1}{3} + \frac{2}{3}p \Rightarrow p = \frac{3}{7}.$$

Tento iterační postup je často užitečný a dá se využít i v bodě d), kde je třeba označit  $p_i$  pravděpodobnost výhry Cyrila za podmínky, že má  $i$  dolarů. Pak dostaneme soustavu pěti rovnic o pěti neznámých  $p_1, p_2, \dots, p_5$  odpovídajících pěti stavům hry, která ještě neskončila.

Hra skončí v lichém kole s pořadím  $2k + 1$  a v sudém kole s pořadím  $2k + 2$  s pravděpodobností

$$p_{2k+1} = \frac{1}{3} \left( \frac{2}{3} \frac{1}{3} \right)^k, \quad p_{2k+2} = \frac{4}{9} \left( \frac{2}{3} \frac{1}{3} \right)^k.$$

Odtud je střední počet her

$$EX = \sum_{k=0}^{\infty} (2k+1)p_{2k+1} + \sum_{k=0}^{\infty} (2k+2)p_{2k+2},$$

což lze dále upravovat.

Cyrl může vyhrát jen v lichém kole, čili z předchozího použijeme první sumu. Nesmíme ale zapomenout na podmiňování! Střední počet her za předpokladu, že vyhrál Cyril proto je

$$E(X|C) = \sum_{k=0}^{\infty} (2k+1)p_{2k+1} \times \frac{7}{3},$$

kde jsme dělili pravděpodobností  $3/7$  výhry Cyrila.

**Příklad 2** (5 bodů). Zformulujte a dokažte pro diskrétní náhodný vektor  $(X, Y)$  větu, že

$$E(E(X|Y)) = EX.$$

Nezapomeňte na vhodné předpoklady této věty!

**Řešení 2.** Aby vůbec mělo smysl hovořit o střední hodnotě, je nutné předpokládat  $E|X| < \infty$ . Předpokládejme, že  $S$  je množina hodnot  $Y$  taková, aby  $P[Y \in S] = 1$  a pro  $s \in S$  platilo  $P[Y = s] > 0$  a  $K$  je množina hodnot  $X$ . Pak

$$\begin{aligned} E(E(X|Y)) &= \sum_{s \in S} E(X|Y = s)P[Y = s] = \sum_{s \in S} \sum_{k \in K} kP[X = k|Y = s]P[Y = s] \\ &= \sum_{s \in S} \sum_{k \in K} kP[X = k, Y = s] = \sum_{k \in K} k \sum_{s \in S} P[X = k, Y = s] \\ &= \sum_{k \in K} kP[X = k] = EX. \end{aligned}$$

Jiné předpoklady jsme cestou při výpočtech nepotřebovali.

**Příklad 3** (6 bodů).  $X$  a  $Y$  jsou nezávislé náhodné veličiny s exponenciálním rozdělením s parametrem  $\lambda$ , tedy s hustotou

$$f(x) = \begin{cases} \lambda e^{-\lambda x} & x > 0 \\ 0 & \text{jinak.} \end{cases}$$

Určete

- (a) Střední hodnotu a momentovou vytvořující funkci  $X$ .
- (b) Rozdělení součtu  $X + Y$ , střední hodnotu a rozptyl  $X + Y$ .
- (c)\* Podmíněnou hustotu  $f_{X+Y|X}$  náhodné veličiny  $X + Y$  za podmínky  $X = x$ .

**Řešení 3.** Pro střední hodnotu platí

$$\begin{aligned} EX &= \int_0^{\infty} \lambda x \exp(-\lambda x) dx = \text{substituce } \lambda x = y, \text{ meze se nemění} = \frac{1}{\lambda} \int_0^{\infty} y \exp(-y) \\ &= \text{per partes} = \frac{1}{\lambda}. \end{aligned}$$

Momentová vytvořující funkce je

$$\begin{aligned} \psi_X(t) &= E \exp(tX) = \int_0^{\infty} \lambda \exp(tx) \exp(-\lambda x) dx = \lambda \int_0^{\infty} \exp(x(t - \lambda)) \\ &= \text{integrál je konečný jen pro } t < \lambda = \frac{\lambda}{\lambda - t} \end{aligned}$$

pro  $t < \lambda$ .

Pro hustotu součtu můžeme použít větu o konvoluci, pro distribuční funkci v bodě  $z$  stačí integrovat sdruženou hustotu přes množinu  $\{(x, y); x > 0, y > 0, x + y < z\}$ .

Hustota:

$$\begin{aligned} f_Z(z) &= \int_0^{\infty} f_{(X,Y)}(x, z-x) dx = z \text{ nezávislosti a tvaru hustoty} = \int_0^z \lambda \exp(-\lambda x) \lambda \exp(-\lambda(z-x)) dx \\ &= \lambda^2 \int_0^z \exp(-\lambda z) dx \text{ integrujeme podle } x = \lambda^2 z \exp(-\lambda z). \end{aligned}$$

Střední hodnota součtu  $Z = X + Y$  je součet středních hodnot, rozptyl součtu je **pro nezávislé**  $X$  a  $Y$  součet rozptylů.

$$EX^2 = \frac{\partial^2 \psi_X}{\partial t^2}(0) = \frac{\partial^2 \psi_X}{\partial t^2} \lambda (\lambda - t)^{-1} |_{t=0} = 2\lambda (\lambda - t)^{-3} |_{t=0} = \frac{2}{\lambda^2}.$$

Proto  $\text{var}(X) = 1/\lambda^2$  a  $\text{var}(Z) = 2/\lambda^2$ .

Bonusový příklad je trochu těžší, nicméně stačí si uvědomit, že rozdělení  $Z = X + Y$  za podmínky  $X = x$  je stejné jako rozdělení  $Y + x$ . Pak tedy

$$P[Z \leq z | X = x] = P[X + Y \leq z | X = x] = P[x + Y \leq z | X = x] = P[Y \leq z - x | X = x].$$

Protože  $Y$  a  $X$  jsou nezávislé, platí

$$P[Z \leq z] = P[Y \leq z - x] = \begin{cases} 0 & \text{pro } z < x \\ F_Y(z - x) = 1 - \exp(-\lambda(z - x)) & z \geq x. \end{cases}$$

Derivací podle  $z$  můžeme dostat hustotu  $Z$ .

**Příklad 4** (6 bodů). Nezáporná náhodná veličina  $X$  má diskrétní rozdělení na přirozených číslech (tedy množině  $\{0, 1, 2, \dots\}$ ). Najděte

- (a) Nestranný a konzistentní odhad  $\hat{p}_1$  pravděpodobnosti  $P[X = 1]$ . Ukažte nestrannost i konzistenci  $\hat{p}_1$ .
- (b) Odvoďte asymptotickou normalitu odhadu  $\hat{p}_1$ .

**Řešení 4.** Máme-li cokoliv odhadnout, musíme předpokládat, že máme k dispozici nějaká pozorování. Zde je na místě předpokládat náhodný výběr  $X_1, X_2, \dots, X_n$  z rozdělení stejného jako má  $X$ .

Protože se jedná o diskrétní rozdělení, je úkolem odhadnout  $P[X = 1] = F_X(1) - F_X(0)$  ( $X$  nabývá jen celočíselných hodnot). Nestranným a konzistentním odhadem  $F_X(x)$  je empirická distribuční funkce

$$\widehat{F}_n(x) = \frac{1}{n} \sum_{i=1}^n \chi[X_i \leq x],$$

a proto

$$\widehat{p}_1 = \widehat{F}_n(1) - \widehat{F}_n(0) = \frac{1}{n} \sum_{i=1}^n (\chi[X_i \leq 1] - \chi[X_i \leq 0]) = \frac{1}{n} \sum_{i=1}^n \chi[X_i = 1].$$

Nestrannost a konzistence plyne z nestrannosti a konzistence empirické distribuční funkce.

Protože

$$\widehat{p}_1 = \frac{1}{n} \sum_{i=1}^n \chi[X_i = 1],$$

kde  $E\chi[X_i = 1] = P[X = 1]$  a  $\text{var} \chi[X_i = 1] = P[X = 1](1 - P[X = 1])$ , lze využít centrální limitní větu k ověření

$$P \left[ \sqrt{n} \frac{\widehat{p}_1 - P[X = 1]}{\sqrt{P[X = 1](1 - P[X = 1])}} \leq x \right] \rightarrow \Phi(x).$$

**Příklad 5** (6 bodů). Uvažujte náhodný vektor  $(X, Y)$  s konečnými druhými momenty.

- Dokažte, že varianční matice  $\text{Var}(X, Y)$  je symetrická a pozitivně semidefinitní.
- Dokažte, že  $|\text{corr}(X, Y)| = 1$  právě tehdy, existují-li konstanty  $a \neq 0$  a  $b$  takové, že  $Y = aX + b$ .

**Řešení 5.** Symetrie varianční matice plyne ze symetrie  $\text{cov}(X, Y) = E(X - EX)(Y - EY) = E(Y - EY)(X - EX) = \text{cov}(Y, X)$ .

Protože

$$a^T \text{Var}(X, Y)a = a_1^2 \text{var}(X) + 2a_1a_2 \text{cov}(X, Y) + a_2^2 \text{var}(Y) = \text{var}(a_1X + a_2Y) \geq 0,$$

ověřili jsme pozitivní semidefinitnost matice.

Pokud  $Y = aX + b$ , pak  $\text{var}(Y) = a^2 \text{var}(X)$  a  $\text{cov}(X, Y) = a \text{cov}(X, X) = a \text{var}(X)$ . Z definice

$$\text{corr}(X, Y) = \frac{\text{cov}(X, Y)}{\sqrt{\text{var}(X) \text{var}(Y)}} = \frac{a \text{var}(X)}{\sqrt{a^2 \text{var}(X) \text{var}(X)}} = \frac{a}{|a|} \in \{-1, 1\}.$$

Opačná implikace je trochu těžší. Nechť například  $\text{corr}(X, Y) = 1$ . Uvědomíme si, že v tom případě je  $\text{cov}(X, Y) = \sqrt{\text{var}(X) \text{var}(Y)}$  a tedy varianční matice má tvar

$$\text{Var}(X, Y) = \begin{pmatrix} \sigma_1^2 & \sigma_1\sigma_2 \\ \sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}.$$

Tato matice má nulový determinant, její řádky jsou lineárně závislé. Existuje tedy vektor  $a = (1, a_2)$  takový, že

$$\text{var}(X + a_2Y) = a^T \text{Var}(X, Y)a = 0$$

a to je možné jen pokud je  $X + a_2Y$  konstanta (náhodná veličina degenerovaná do jediné hodnoty).

**Příklad 6** (5 bodů). Zformulujte centrální limitní větu. Uvažujte nezávislé bernoulliovské pokusy  $X_1, X_2, \dots$  s pravděpodobností úspěchu  $p \in (0, 1)$ . Použijte centrální limitní větu k určení

- minimálního počtu pokusů, které potřebujete k tomu, aby  $|\overline{X}_n - p| \leq \sqrt{p(1-p)}/100$  nastalo s pravděpodobností alespoň 0.98 (asymptoticky).
- intervalového odhadu pro neznámý parametr  $p$ , provedeme-li  $n$  pokusů.

**Řešení 6.** Nezapomínejte na předpoklady:

- Náhodný výběr, tedy nezávislé a stejně rozdělené náhodné veličiny.
- Konečná střední hodnota  $\mu = EX$ .
- Konečný a nenulový rozptyl  $\sigma^2 = \text{var}(X)$ .

Protože  $EX = p$  a  $\text{var}(X) = p(1-p)$ , z CLV platí

$$P \left[ |\bar{X}_n - p| \leq \sqrt{p(1-p)}/100 \right] = P \left[ \sqrt{n} \frac{|\bar{X}_n - p|}{\sqrt{p(1-p)}} \leq \frac{\sqrt{n}}{100} \right] \rightarrow \Phi \left( \frac{\sqrt{n}}{100} \right) - \Phi \left( -\frac{\sqrt{n}}{100} \right) = 2\Phi \left( \frac{\sqrt{n}}{100} \right) - 1.$$

Poslední výraz se má rovnat 0,98 a proto  $\sqrt{n} \geq 100u_{0,99}$ , kde  $\Phi(u_{0,99}) = 0,99$ , tedy jde o patřičný kvantil normálního rozdělení.

Protože platí

$$P \left[ \sqrt{n} \frac{|\bar{X}_n - p|}{\sqrt{p(1-p)}} \leq u_{1-\alpha/2} \right] \doteq \Phi(u_{1-\alpha/2}) - \Phi(-u_{1-\alpha/2}) = 1 - \alpha$$

mohli bychom odvodit intervalový odhad pro  $p$  z těchto nerovností. Protože ale není snadné „osvobodit“  $p$  ze jmenovatele, je možné pomoci si Sluckého větou a faktem

$$\bar{X}_n(1 - \bar{X}_n) \xrightarrow{P} p(1-p)$$

k tomu, že

$$P \left[ \sqrt{n} \frac{|\bar{X}_n - p|}{\sqrt{\bar{X}_n(1 - \bar{X}_n)}} \leq u_{1-\alpha/2} \right] \doteq \Phi(u_{1-\alpha/2}) - \Phi(-u_{1-\alpha/2}) = 1 - \alpha.$$

Toto už je mnohem snazší k provedení několika algebraických úprav na levé straně a k dovození

$$P \left[ \bar{X}_n - u_{1-\alpha/2} \frac{\sqrt{\bar{X}_n(1 - \bar{X}_n)}}{n} \leq p \leq \bar{X}_n + u_{1-\alpha/2} \frac{\sqrt{\bar{X}_n(1 - \bar{X}_n)}}{n} \right] \doteq \Phi(u_{1-\alpha/2}) - \Phi(-u_{1-\alpha/2}) = 1 - \alpha.$$

Tím jsme hotovi.