Úloha I.S ... náhodná

10 bodů; průměr 7,04; řešilo 45 studentů

a) *Zkuste vlastními slovy popsat, co je to náhodná veličina a jaké má vlastnosti (postačí vlastními slovy objasnit následující pojmy: náhodná veličina, rozdělení náhodné veličiny, realizace náhodné veličiny, střední hodnota, rozptyl, histogram).*
b) *Vygenerujte grafy hustot pravděpodobnosti (případně pravděpodobností nabývání jednotlivých hodnot) všech v seriálu popsaných rozdělení náhodných veličin pro různé typy parametrů daného rozdělení a popište, jaký má změna parametru/ů vliv na tvar hustoty pravděpodobnosti (případně pravděpodobností nabývání jednotlivých hodnot).*
c) *Vygenerujte z přiložených dat histogramy a pokuste se určit, ze kterého rozdělení tato data pocházejí.*
d) *Definujme si náhodnou veličinu X jako výsledek hodu „férovou“ šestistěnnou kostkou (všechna čísla padají se stejnou pravděpodobností). Určete rozdělení náhodné veličiny X a dále spočítejte $\mathrm{E}X$ a $\mathrm{var}X$.*

Bonus *Uveďte příklad dvou náhodných veličin, které mají stejnou střední hodnotu i stejný rozptyl, ale mají různá rozdělení.*

Pro práci s daty a vykreslování grafů použijte výpočetní prostředí R. *Pro vyřešení těchto úkolů postačí drobně upravit přiložený skript, ve kterém je pomocí komentářů v kódu vysvětlena potřebná syntaxe jazyka* R. *Michal stanovil zadání úlohy náhodně, snad nebude moc těžká.*

a) Detailní odpověď na tuto otázku dostanete přečtením 1. dílu seriálu, na tomto místě si vysvětlíme jen ty opravdu nejdůležitější pojmy. Náhodné veličiny se používají k matematickému modelování situací, u kterých nemůžeme dopředu s jistotou znát přesný výsledek. Náhodnou veličinu si můžeme představovat jako proměnnou, která ale nemá jednu pevnou hodnotu (ať už známou či neznámou), ale může nabývat (s různými pravděpodobnostmi) různých hodnot.

Rozdělení náhodné veličiny je potom to, čím je náhodná veličina určená (podobně jako je klasická proměnná určená svojí hodnotou). Rozdělení náhodné veličiny může být buď diskrétní (potom je toto rozdělení určeno pravděpodobnostmi nabývání jednotlivých hodnot), nebo spojité (potom je toto rozdělení určeno hustotou pravděpodobnosti).

Realizace náhodné veličiny je hodnota, kterou dostaneme v konkrétním případě, kdy budeme náhodnou veličinu měřit.

Střední hodnota a rozptyl náhodné veličiny představují jakýsi zjednodušený popis náhodné veličiny (plný popis je právě rozdělení náhodné veličiny), jsou určeny následujícími vzorci

$$\mathrm{E}X = \sum_{k=1}^{\infty} k\,\mathrm{P}(X=k)\,,$$

$$\mathrm{var}X = \sum_{k=1}^{\infty} k\,\mathrm{P}(X=k)\,(k-\mathrm{E}X)^2\,,$$

pokud se jedná o diskrétní náhodnou veličinu, případně vzorci

$$\mathrm{E}X = \int_{-\infty}^{\infty} x f(x) \mathrm{d}x \,,$$

$$\mathrm{var}X = \int_{-\infty}^{\infty} (x - \mathrm{E}X)^2 f(x) \mathrm{d}x$$

v případě, že se jedná o spojitou náhodnou veličinu s hustotou f. Střední hodnota vyjadřuje, jakou hodnotu bychom měli v průměru očekávat při konkrétní realizaci náhodné veličiny a rozptyl vyjadřuje míru rozptýlenosti náhodné veličiny kolem její střední hodnoty.

Histogram je potom typ grafu, který vykresluje naměřená data (tedy realizace nějaké náhodné veličiny) a slouží k odhadování typu rozdělení, ze kterého naměřená data pocházejí.

b) V seriálu byly uvedeny celkem čtyři nejčastěji se vyskytující rozdělení náhodných veličin (normální, exponenciální, Poissonovo a rovnoměrné rozdělení). Budeme postupně vykreslovat hustoty pravděpodobnosti (případně pravděpodobnosti nabývání jednotlivých hodnot) pro tato čtyři rozdělení.

Začneme normálním rozdělením, které, jak bylo uvedeno v seriálu, závisí na dvou parametrech $\mu \in \mathbb{R}, \sigma^2 > 0$. Hustota pravděpodobnosti normálního rozdělení je dána vztahem

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \mathrm{e}^{-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}} \,.$$

Grafy hustot normálního rozdělení pro různé hodnoty parametrů μ, σ^2 můžeme vidět na obrázcích 1 a 2. Jak je z těchto obrázků patrné, parametr μ ovlivňuje pouze posun po ose x a parametr σ^2 škáluje graf hustoty v horizontálním směru (laicky řečeno graf rozšiřuje nebo zužuje).

Další v seriálu uvedené rozdělení je exponenciální rozdělení, které závisí na jednom parametru $\lambda > 0$. Hustota pravděpodobnosti exponenciálního rozdělení je určená vztahem

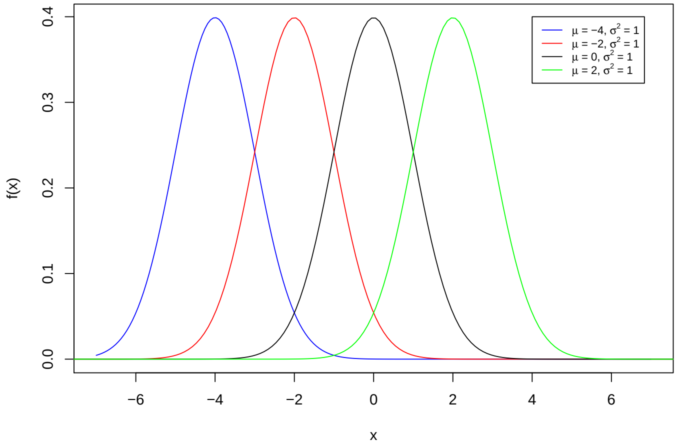
$$f(x) = \begin{cases} \lambda \mathrm{e}^{-\lambda x} & \text{pro } x > 0 \\ 0 & \text{jinak} \,. \end{cases}$$

Grafy hustot exponenciálního rozdělení pro různé hodnoty parametru λ jsou vykresleny na obrázku 3. Jak je na obrázku vidět, tvar hustoty pravděpodobnosti exponenciálního rozdělení zůstává stále stejný, parametr λ je zodpovědný pouze za škálování v horizontálním směru.

Posledním v seriálu uvedeným spojitým rozdělením je rovnoměrné rozdělení, které závisí na dvou parametrech $-\infty < a < b < \infty$. Hustota pravděpodobnosti rovnoměrného rozdělení je určena vztahem

$$f(x) = \begin{cases} \frac{1}{b-a} & \text{pro } x \in (a, b) \\ 0 & \text{jinak} \,. \end{cases}$$

Grafy hustot rovnoměrného rozdělení pro různé hodnoty parametrů a, b si můžeme prohlédnout na obrázku 4. Jak je z obrázku vidět, hustota pravděpodobnosti rovnoměrného rozdělení je nulová mimo interval (a, b) a na intervalu (a, b) nabývá takové konstantní hodnoty, aby platilo $\mathrm{P}(a \leq X \leq b) = 1$.



Obr. 1: Graf hustoty pravděpodobnosti normálního rozdělení pro různé hodnoty parametru μ

Jediným v seriálu uvedeným diskrétním rozdělím je Poissonovo rozdělení, které je určeno jedním parametrem $\lambda > 0$. Pravděpodobnosti nabývání jednotlivých hodnot jsou určeny vztahem

$$\mathrm{P}(X = k) = \mathrm{e}^{-\lambda}\frac{\lambda^k}{k!}, k \in \mathbb{N}.$$

Grafy pravděpodobností nabývání jednotlivých hodnot pro různé hodnoty parametru můžeme vidět na obrázku 5. Opět je vidět, že se tvar závislosti pravděpodobností nabývání jednotlivých hodnot při změně parametru λ nemění, ale pouze škáluje v horizontálním směru.

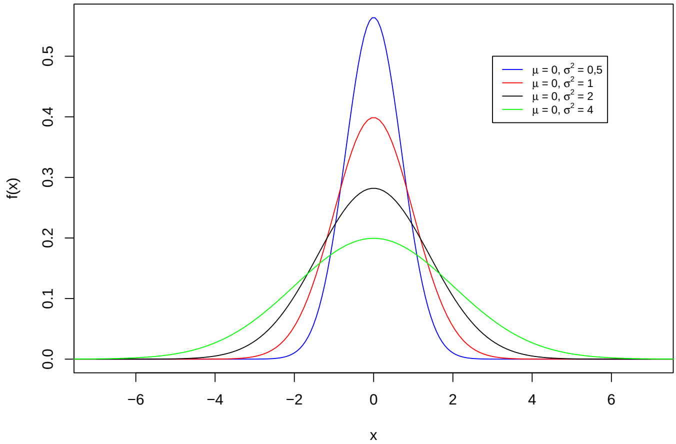
c) Vygenerované histogramy můžeme vidět na obrázcích 6, 7, 8 a 9. Z těchto obrázků lze odhadovat, že první vzorek dat pochází z rovnoměrného rozdělení, druhý vzorek z Poissonova rozdělení, třetí vzorek z normálního rozdělení a čtvrtý vzorek z exponenciálního rozdělení. Je nutno poznamenat, že toto jsou pouze odhady, nelze to chápat jako rigorózní matematický postup určování typu rozdělení.

d) Takováto náhodná veličina může nabývat jen konečně mnoha hodnot (jen hodnot $1, 2, 3, 4, 5$ nebo 6), proto se bude jednat o diskrétní náhodnou veličinu. Rozdělení náhodné veličiny X tedy bude určeno výčtem pravděpodobností nabývání těchto jednotlivých hodnot, jelikož podle zadání budou všechny výsledky hodu touto kostkou stejně pravděpodobné, musí platit

$$\mathrm{P}(X = 1) = \frac{1}{6}, \mathrm{P}(X = 2) = \frac{1}{6}, \ldots, \mathrm{P}(X = 6) = \frac{1}{6}.$$

Toto je rozdělení náhodné veličiny X.
Pro výpočet střední hodnoty a rozptylu pouze dosadíme do vzorců uvedených výše. Dostáváme, že střední hodnota X je rovna

$$\mathrm{E}X = 1 \cdot \frac{1}{6} + 2 \cdot \frac{1}{6} + \cdots + 6 \cdot \frac{1}{6} = 3{,}5.$$



Obr. 2: Graf hustoty pravděpodobnosti normálního rozdělení pro různé hodnoty parametru σ^2

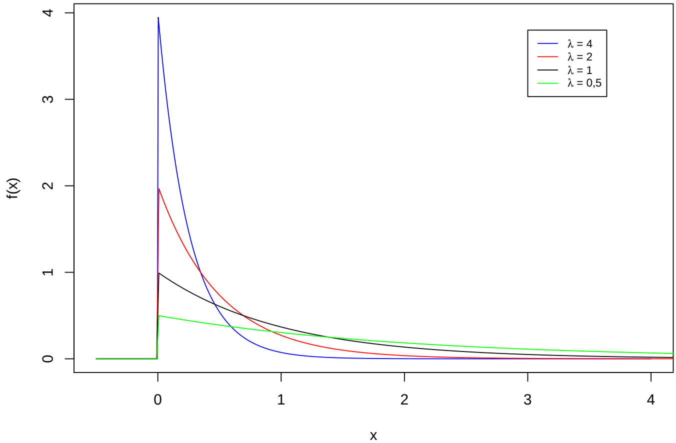
Když nyní známe střední hodnotu náhodné veličiny X, můžeme spočítat i její rozptyl, který je roven

$$\mathrm{var}X = (1-3,5)^2 \cdot \frac{1}{6} + (2-3,5)^2 \cdot \frac{1}{6} + \dots + (6-3,5)^2 \cdot \frac{1}{6} = \frac{35}{12}$$

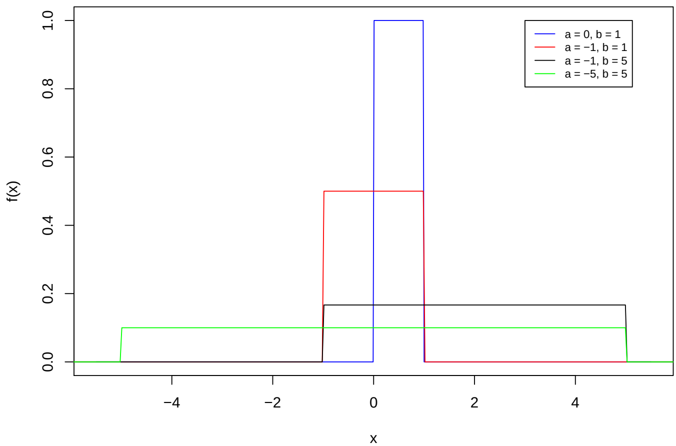
Bonus: Tento úkol je velice jednoduchý, stačí ze seznamu nejčastěji se vyskytujících rozdělení náhodných veličin vybrat dvě vhodná rozdělení (různá) a vhodně navolit jejich parametry tak, aby se jejich střední hodnota i rozptyl rovnaly. Jako příklad můžeme uvést náhodné veličiny X a Y, kde X bude mít exponenciální rozdělení $Exp(2)$ a náhodná veličina Y bude mít rozdělení $N(\frac{1}{2}, \frac{1}{4})$. Pomocí vzorců uvedených v seriálu si každý může ověřit, že střední hodnota obou těchto náhodných veličin jsou rovny $\frac{1}{2}$ a rozptyl obou těchto náhodných veličin je roven $\frac{1}{4}$. Takovýchto příkladů lze samozřejmě vymyslet mnoho, pro udělení bonusových bodů stačí uvést jeden.

Michal Nožička
`nozicka@fykos.cz`

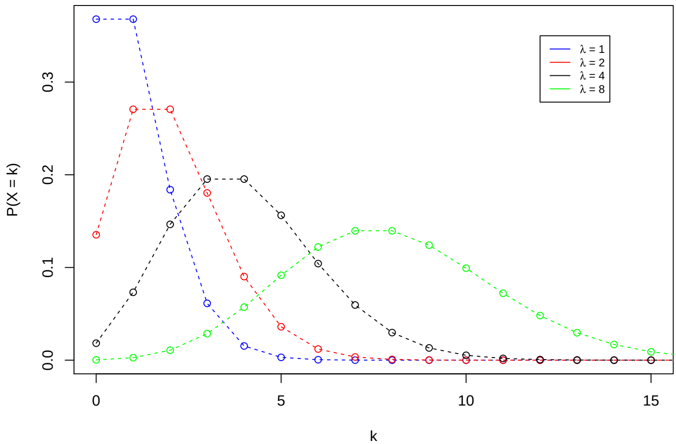
Fyzikální korespondenční seminář je organizován studenty MFF UK. Je zastřešen Oddělením pro vnější vztahy a propagaci MFF UK a podporován Ústavem teoretické fyziky MFF UK, jeho zaměstnanci a Jednotou českých matematiků a fyziků.



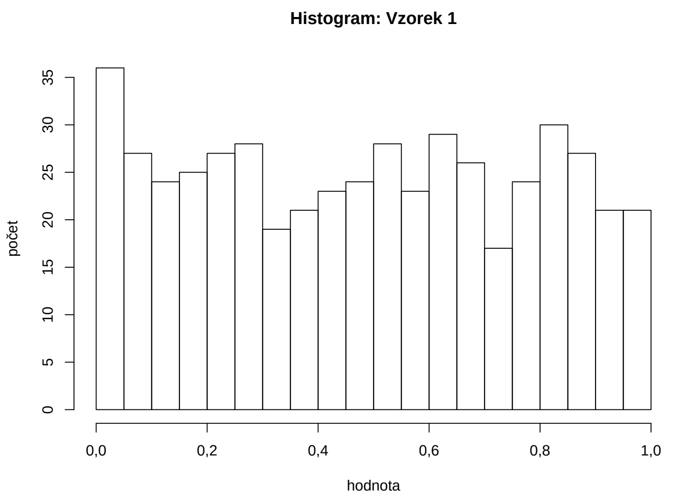
Obr. 3: Graf hustoty pravděpodobnosti exponenciálního rozdělení pro různé hodnoty parametru λ.



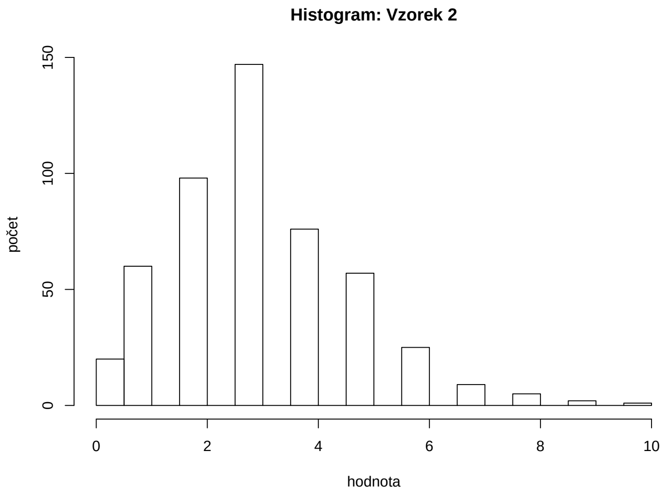
Obr. 4: Grafy hustot pravděpodobnosti rovnoměrného rozdělení pro různé hodnoty parametrů a, b.



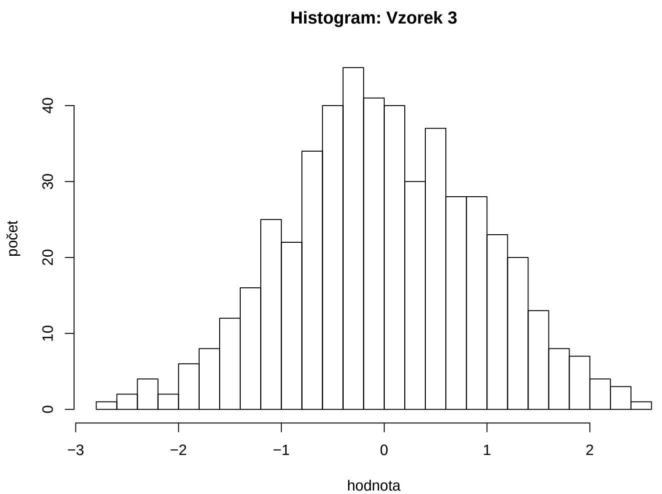
Obr. 5: Pravděpodobnosti nabývání jednotlivých hodnot poissonova rozdělení pro různé hodnoty parametru λ.



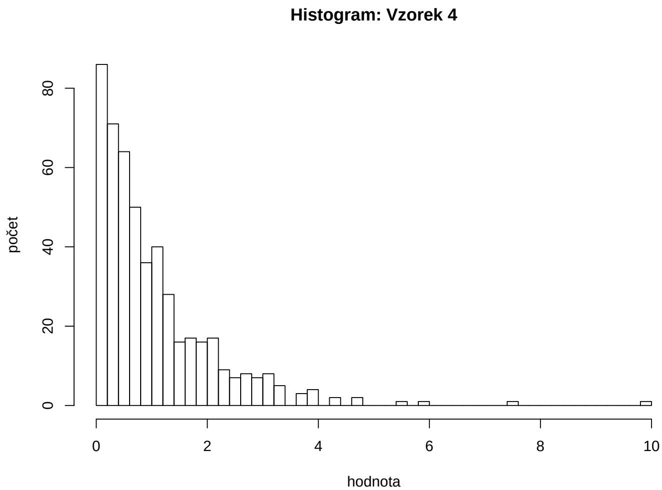
Obr. 6: Histogram prvního vzorku dat.



Obr. 7: Histogram druhého vzorku dat.



Obr. 8: Histogram třetího vzorku dat.



Obr. 9: Histogram čtvrtého vzorku dat.