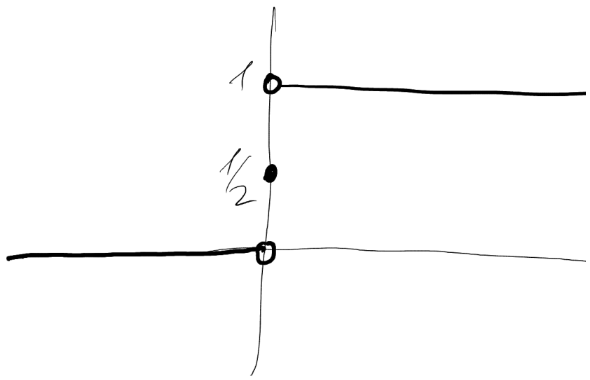
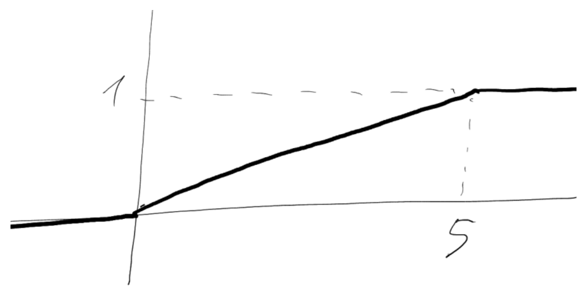
1. zkoušková písemka NMAI059 Pravd. a Stat. 1 – řešení – 16.6.2021

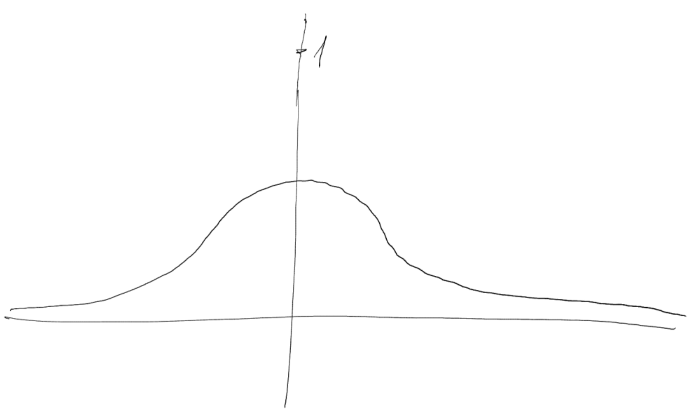
1. (10 bodů) (a) Rozhodněte, který z obrázků popisuje distribuční funkci nějaké náhodné veličiny. Další dvě části provádějte jen u těch obrázků, které distribuční funkci popisují.

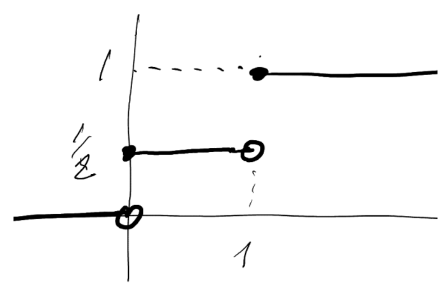
(b) Odhadněte střední hodnotu.

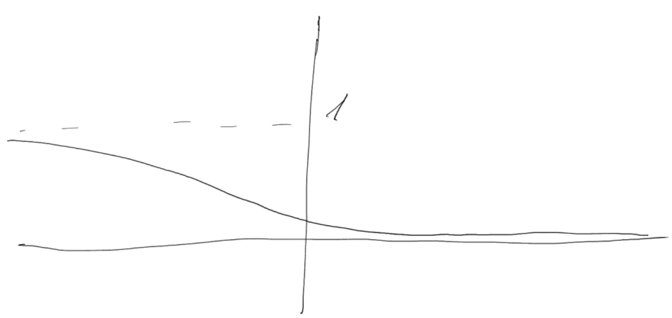
(c) Seřaďte podle hodnoty rozptylu.

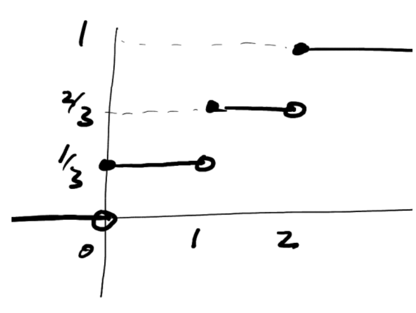


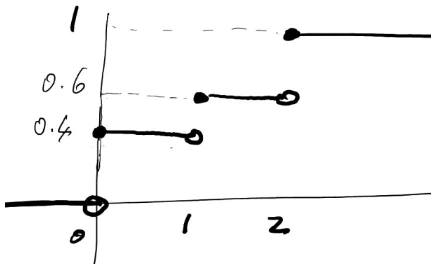












Řešení:

(a) Třetí a pátá funkce nejsou distribuční funkce žádné náhodné veličiny, protože nejsou neklesající. První také není, protože není zprava spojitá. (Neboli: příslušná náhodná veličina by musela s pravděpodobností 1/2 generovat čísla, která jsou kladná, ale menší než každé kladné reálné číslo.) Všechny ostatní funkce splňují všechny požadavky z přednášky (neklesající, zprava spojitá, správné limity v $\pm\infty$), jedná se tedy o distribuční funkce. Druhá funkce odpovídá uniformnímu rozdělení $U(0,5)$, čtvrtá, šestá a sedmá odpovidají diskrétní náhodné veličině, která nabývá jen hodnot $\{0,1\}$, resp. $\{0,1,2\}$; pravděpodobnost nějakého bodu je rovna velikosti skoku distribuční funkce v tom bodě. Označíme dále X_i náhodnou veličinu, jejíž distribuční funkce je na i-tém obrázku.

(b) $X_2 \sim U(0,5)$, tedy $\mathbb{E}(X_2) = 5/2$. (Pokud si to nepamatujeme, můžeme spočítat jako $\int_0^5 x\frac{1}{5} = \frac{1}{5}\frac{5^2}{2}$.)

$X_4 \sim Ber(1/2)$, tedy $\mathbb{E}(X_4) = 1/2$

Pro $X = X_6, X_7$ platí, že $P(X=0) = P(X=2)$, tedy je $\mathbb{E}(X_6) = \mathbb{E}(X_7) = 1$. (Pro kontrolu: $P(X_6 = 0) = 1/3$, $P(X_7 = 0) = 0.4$.)

(c) $X_2 \sim U(0,5)$, tedy $\text{var}(X_2) = 5^2/12$.

$X_4 \sim Ber(1/2)$, tedy $\text{var}(X_4) = 1/4$

Pro $X = X_6, X_7$ platí, že $P(X=0) = P(X=2)$, a $\mathbb{E}(X) = 1$. Je tedy $\text{var}(X) = \mathbb{E}((X-1)^2) = 1 \cdot P(X \neq 1)$.

Závěr: $\text{var}(X_4) < \text{var}(X_6) < \text{var}(X_7) < \text{var}(X_2)$.

---

2. (10 bodů) (a) Dva hráči, Adam a Božena, opakovaně hází kostkou, v pořadí ABABAB... Jaká je pravděpodobnost, že šestka padne první Adamovi?

(b) Do hry se přidá ještě Cecil, hází v pořadí ABCABCABC... Pravděpodobnost, že šestka padne napřed Adamovi, pak Boženě, a nakonec Cecilovi je 216/1001. Zdůvodněte. (Pokud padne Adamovi šestka vícekrát, a až pak Boženě, tak je to také v pořádku, jde nám jen o pořadí časů, kdy poprvé hodí šestku.)

Řešení:

(a) Označme hledanou pravděpodobnost p. Použijeme větu o úplné pravděpodobnosti. Pro požadované pořadí padnutí šestky jsou dvě možnosti: buď Adamovi padne šestka napoprvé (pravděpodobnost 1/6) a dál je jedno co se stane, nebo Adamovi ani Boženě nepadne šestka (pravděpodobnost $(5/6)^2$) a jsme ve stejné situaci jako na začátku, tj. pravděpoobnost úspěchu je p. Celkově tedy víme, že

$$p = \frac{1}{6} + \left(\frac{5}{6}\right)^2 \cdot p$$

Odsud snadnou úpravou zjistíme, že $p = 6/11$.

Alternativně můžeme označit jako A_i jev, že Adam hodí šestku v $(i+1)$-ním hodu, a předchozích $2i$ hodů nepadla nikdy šestka. Zjevně je $P(A_i) = \frac{1}{6}(5/6)^{2i}$ a zajímá nás

$$p = P\Big(\bigcup_{i=0}^{\infty} A_i\Big) = \frac{1}{6}\sum_{i=0}^{\infty}\big(\frac{5}{6}\big)^{2i}.$$

Součet geometrické řady nám dá týž výsledek 6/11.

(b) Budeme postupovat podobně jako v minulé části, hledanou pravděpodobnost tentokrát označíme q. Opět máme možnost úpěchu, kdy Adamovi, Boženě, ani Cecilovi nepadne šestka (pravděpodobnost $(5/6)^3$) a jsme v situaci jako na začátku, tj. máme pak pravděpodobnost úspěchu q. Pokud Adamovi napoprvé padne šestka, nezáleží pak už na tom, co mu bude dále padat, můžeme ho tedy z úvah vyřadit. Božena a Cecil pak hrají hru z části (a), kde je pravděpodobnost úspěchu $p = 6/11$. Celkově tedy máme

$$q = \frac{1}{6} \cdot \frac{6}{11} + \left(\frac{5}{6}\right)^3 \cdot q$$

Po úpravě zjistíme, že $q = 216/1001$.

---

3. (10 bodů) Paretovo rozdělení s parametrem $\alpha > 1$ má hustotu $f(x) = \frac{\alpha}{x^{\alpha+1}}$ pro $x \in [1, \infty)$ (a nula jinde).

(a) Ověřte, že se jedná o hustotu.

(b) Z tohoto rozdělení jsme nasamplovali hodnoty 5, 2, 3. Odvoďte odhad $\hat{\alpha}$ pomocí metody maximální věrohodnosti.

(c) Nechť X má výše popsané Paretovo rozdělení, tj. $f_X = f$. Spočtěte $\mathbb{E}(X)$.

(d) Najděte odhad $\hat{\alpha}$ metodou momentů.

Řešení:

(a) Potřebujeme ověřit, že integrál přes všecha x je 1:

$$\int_{-\infty}^{\infty} f(x)dx = \int_{1}^{\infty} \alpha x^{-\alpha-1} dx = \big[\frac{\alpha x^{-\alpha}}{-\alpha}\big]_1^{\infty} = 0 - (-1) = 1.$$

(b) Věrohodnost trojice $(5, 2, 3)$ je

$$L(5, 3, 2; \alpha) = \frac{\alpha}{5^{\alpha+1}} \frac{\alpha}{2^{\alpha+1}} \frac{\alpha}{3^{\alpha+1}}.$$

Tuto funkci chceme tedy maximalizovat. Pro zjednodušení ji napřed zlogaritmujeme, pak použijeme derivaci:

$$\ell(\alpha) = \big(\log\alpha - (\alpha+1)\log 5\big) + \big(\log\alpha - (\alpha+1)\log 2\big) + \big(\log\alpha - (\alpha+1)\log 3\big)$$
$$\ell'(\alpha) = \frac{3}{\alpha} - \log(5\cdot 2\cdot 3)$$

Maximální věrohodnost má hodnota α, pro kterou je $\ell'(\alpha) = 0$ (vidíme, že pro menší α je derivace kladná, pro větší záporná, jedná se tedy opravdu o maximum). Klademe tedy $\hat\alpha = \frac{1}{3}\log 30 \doteq 1.13$. (Při písemce netřeba vyčíslovat.)

(c) Podle definice je

$$\begin{aligned}\mathbb{E}(X) &= \int_{-\infty}^{\infty} x\cdot f(x)dx = \int_{1}^{\infty} \alpha x^{-\alpha}dx \\ &= \Big[\frac{\alpha}{-\alpha+1}x^{-\alpha+1}\Big]_1^\infty = 0 - \Big(\frac{\alpha}{-\alpha+1}\Big) \\ &= \frac{\alpha}{\alpha-1}\end{aligned}$$

Podotkněme, že 0 na druhém řádku je $\lim_{x\to\infty} x^{-\alpha+1}$. Zde tedy využíváme toho, že $\alpha > 1$.

(d) Použijeme předchozí část. První moment dané distribuce je $\alpha/(\alpha-1)$, ten položíme rovný prvnímu výběrovému momentu $\bar x_3 = (5+2+3)/3 = 10/3$. Vyřešíme tedy lehkou rovnici $\alpha/(\alpha-1) = \bar x_3$, odsud $1 - 1/\alpha = 1/\bar x_3$, a tedy odhad získaný momentovou metodou je

$$\hat\alpha = \frac{1}{1-1/\bar x_3} = \frac{\bar x_3}{\bar x_3 - 1} = \frac{10}{7} \doteq 1.43.$$

---

4. (10 bodů) (a) Definujte pojem sdružená distribuční funkce.
(b) Popište, jak se spočítá empirická distribuční funkce.

Řešení:

(a) Pro náhodné veličiny $X_1,\dots,X_n$ je jejich sdružená distribuční funkce definována vzorcem

$$F_{X_1,\dots,X_n}(x_1,\dots,x_n) = P(X_1 \le x_1 \,\&\dots\&\, X_n \le x_n).$$

(Stačila i definice pro dvě náhodné veličiny, tj. $F_{X,Y}(x,y) = P(X \le x \,\&\, Y \le y)$.)

(b) Máme-li náhodný výběr $X_1,\dots,X_n$, tak k nim příslušná empirická distribuční funkce je

$$\hat F_n(x) = \frac{\sum_{i=1}^n I(X_i \le x)}{n},$$

kde $I(X_i \le x) = 1$ pokud $X_i \le x$ a 0 jinak.

Ekvivalentně řečeno, pokud naměříme hodnoty $x_1,\dots,x_n$ (realizaci náhodného výběru), tak empirická distribuční funkce $\hat F(x)$ jedna n-tina z počtu i takových, že $x_i \le x$.

---

5. (10 bodů) Vyslovte Centrální limitní větu. Vysvětlete, k čemu se hodí.

Řešení: Nechť $X_1, X_2, \dots$ jsou stejně rozdělené náhodné veličiny se střední hodnotou μ a rozptylem σ^2. Označme

$$Y_n = \frac{X_1 + \cdots + X_n - n\mu}{\sigma\sqrt{n}}.$$

Pak Y_n konvergují v distribuci k $N(0,1)$, neboli pro každé $x \in \mathbb{R}$ platí

$$\lim_{n\to\infty} P(Y_n \le x) = \Phi(x).$$

Význam věty je v tom, že nám umožňuje aproximovat součet mnoha náhodných veličin pomocí jednoho, dobře známého rozdělení. Takové součty se často vyskytují – na příklad $Bin(n,p)$ je součet n veličin $Ber(p)$, ve statistice takové součty často potkáme a mnoho fyzikálních jevů je (přibližně) popsáno jako součet nezávislých náhodných veličin. To vysvětluje proč binomická čísla, stejně jako mnoho v praxi se vyskytujících veličin, mají přibližně normální rozdělení.

---

6. (10 bodů) Vyslovte a dokažte větu o střední hodnotě součtu náhodných veličin (důkaz stačí pro diskrétní náhodné veličiny).

Řešení:

Věta: $\mathbb{E}(X+Y) = \mathbb{E}(X)+\mathbb{E}(Y)$, má-li pravá strana smysl. (Tj. jsou-li obě střední hodnoty definovány a nejedná se o případ $\infty - \infty$.)

Důkaz: Použijeme pravidlo LOTUS pro funkci $g(x,y) = x + y$. Podle něj je

$$\begin{aligned}
\mathbb{E}(g(X,Y)) &= \sum_{x\in ImX} \sum_{y\in ImY} g(x,y) \cdot P(X = x \,\&\, Y = y) \\
&= \sum_{x\in ImX} \sum_{y\in ImY} (x+y) \cdot P(X = x \,\&\, Y = y) \\
&= \sum_{x\in ImX} \sum_{y\in ImY} x \cdot P(X = x \,\&\, Y = y) + \sum_{x\in ImX} \sum_{y\in ImY} y \cdot P(X = x \,\&\, Y = y) \\
&= \sum_{x\in ImX} x \cdot P(X = x) + \sum_{y\in ImY} y \cdot P(Y = y) \\
&= \mathbb{E}(X) + \mathbb{E}(Y)
\end{aligned}$$

Využíváme zde toho, že $\sum_{y\in ImY} P(X = x \,\&\, Y = y) = P(X = x)$, obdobně pro prohozené X a Y.