

UNIVERZITA PALACKÉHO V OLMOUCI
PŘÍRODOVĚDECKÁ FAKULTA
KATEDRA MATEMATICKÉ ANALÝZY A APLIKACÍ MATEMATIKY

BAKALÁŘSKÁ PRÁCE

Charakteristiky rozdělení pravděpodobností



Vedoucí diplomové práce:
RNDr. PhDr. Ivo Müller, Ph.D.
Rok odevzdání: 2013

Vypracoval:
Petr Zatloukal
Aplikovaná statistika, III. ročník

Prohlášení

Prohlašuji, že jsem bakalářskou práci zpracoval samostatně pod vedením pana RNDr. PhDr. Ivo Müllera, Ph.D. a s použitím uvedené literatury.

V Olomouci dne 24. dubna 2013

Poděkování

Rád bych na tomto místě poděkoval vedoucímu mojí bakalářské práce panu RNDr. PhDr. Ivo Müllerovi, Ph.D. za obětavou spolupráci i za čas, který mi věnoval při konzultacích. Dále si zaslouží poděkování můj počítač, který vydržel mé pracovní tempo, a typografický systém $\text{T}_{\text{E}}\text{X}$, kterým je práce vysázena.

Obsah

Úvod	6
1 Základní pojmy	7
1.1 Pokus	7
1.2 Náhodný jev	7
1.3 Pravděpodobnost náhodného jevu	8
1.4 Náhodná veličina	9
1.5 Rozdělení pravděpodobností náhodné veličiny	10
1.6 Distribuční funkce náhodné veličiny	11
2 Diskrétní rozdělení pravděpodobností	13
2.1 Alternativní rozdělení	13
2.2 Binomické rozdělení	13
2.3 Poissonovo rozdělení	14
2.4 Geometrické rozdělení	15
3 Spojitá rozdělení pravděpodobností	16
3.1 Rovnoměrné rozdělení	16
3.2 Normální rozdělení	16
3.3 Exponenciální rozdělení	17
3.4 Cauchyho rozdělení	17
4 Číselné charakteristiky náhodné veličiny	19
4.1 Charakteristiky centrální tendence (polohy)	19
4.1.1 Střední hodnota	20
4.1.2 Modus	21
4.1.3 Medián	21
4.2 Charakteristiky variability	22
4.2.1 Rozptyl	23
4.2.2 Střední absolutní odchylka	23
4.2.3 Rozpětí	24
4.2.4 Mezikvartilové rozpětí	24
4.3 Charakteristiky šikmosti a špičatosti	24
4.3.1 Šikmost	24
4.3.2 Špičatost	25

5	Výpočet empirických charakteristik	26
5.1	Charakteristiky polohy	26
5.1.1	Střední hodnota	26
5.1.2	Výběrový modus	26
5.1.3	Výběrový medián	27
5.2	Charakteristiky variability	28
5.2.1	Výběrový rozptyl	28
5.2.2	Střední absolutní odchylka	28
5.2.3	Výběrové rozpětí	28
5.2.4	Mezikvartilové rozpětí	28
5.3	Charakteristiky šikmosti a špičatosti	29
5.3.1	Výběrová šikmost	29
5.3.2	Výběrová špičatost	29
6	Výpočet teoretických charakteristik	30
6.1	Alternativní rozdělení	32
6.2	Binomické rozdělení	34
6.3	Poissonovo rozdělení	37
6.4	Geometrické rozdělení	39
6.5	Rovnoměrné rozdělení	41
7	Empirické charakteristiky pro simulovaná data	44
7.1	Alternativní rozdělení	46
7.2	Binomické rozdělení	50
7.3	Poissonovo rozdělení	53
7.4	Geometrické rozdělení	55
7.5	Rovnoměrné rozdělení	57
7.6	Normální rozdělení	60
7.7	Shrnutí vlivu poruch	62
8	Empirické charakteristiky pro reálná data	65
8.1	Alternativní rozdělení	65
8.2	Binomické rozdělení	66
8.3	Poissonovo rozdělení	67
8.4	Geometrické rozdělení	68
8.5	Rovnoměrné rozdělení	69
8.6	Normální rozdělení	71
8.7	Exponenciální rozdělení	72
	Závěr	73

9 Přílohy	74
9.1 Simulované hodnoty	74
9.2 Reálné hodnoty	76

Úvod

Cílem této bakalářské práce je zpracovat základní typy číselných charakteristik pro diskrétní i spojitý případ rozdělení náhodné veličiny a poukázat na jejich vztah s parametry daného rozdělení. Číselné charakteristiky nám podávají velice rozsáhlou informaci o každém výběrovém souboru s napozorovanými hodnotami. V práci, nazvané *Charakteristiky rozdělení pravděpodobností*, se zaměříme na nejdůležitější a nejpoužívanější typy rozdělení pravděpodobností v praxi a k nim odpovídající číselné charakteristiky.

V první části, teoretické, která pokrývá kapitoly 1 - 4, si v úvodu vysvětlíme základní pojmy z teorie pravděpodobnosti, s nimiž budeme dále pracovat, následované po čtyřech typech rozdělení z diskrétního i spojitého případu. Teoretickou část uzavře uvedení základních charakteristik, jejich vlastnosti a použití. Praktickou část začneme výčtem vzorců pro jednotlivé charakteristiky. Dále budeme napočítávat tyto charakteristiky pro konkrétní rozdělení. Vzhledem k rozsahu bakalářské práce se omezíme na charakteristiky pro prvních 5 rozdělení z našeho výčtu, u dalších zmíněných rozdělení již také velmi značně stoupá náročnost samotných výpočtů. Závěr této části bude patřit výpočtům charakteristik ze simulovaných i reálně napozorovaných dat a jejich souvislost s parametry.

Doufáme, že tato práce čtenáři osvětlí problematiku základů teorie pravděpodobnosti a matematické statistiky a zvýší zájem o danou problematiku. Práce je psána ve snaze o pochopení i pro čtenáře bez rozsáhlejších znalostí statistiky.

1. Základní pojmy

Na začátek si vysvětlíme několik základních pojmů, se kterými budeme v dalším textu pracovat.

1.1. Pokus

Základem statistického pozorování je *pokus*. Pokusem rozumíme jisté uskutečnění pevně daného systému podmínek. S nejrůznějšími typy pokusů se setkáváme v přírodních vědách, jako např. ve fyzice, chemii, biologii. Jiné pokusy, složitěji popisované, se objevují v teorii pravděpodobnosti. Rozlišujeme dva typy pokusů, a sice podle výsledků, jež mohou nastat. První si uvedeme pokus s výsledkem, který můžeme očekávat (předpovědět) s naprostou jistotou. Takový pokus nazýváme *nenáhodný (deterministický)*. Jmenujme např. změnu skupenství vody z kapalného na pevné při dosažení 0°C či nižší teploty. V teorii pravděpodobnosti se ovšem budeme zajímat o ty pokusy, jejichž výsledek je *náhodný (stochastický)*. Tyto nazýváme *náhodné* a končí nastoupením právě jednoho výsledku ω z množiny všech možných výsledků Ω , přičemž tato množina může být *nejvýše spočetná* nebo *nespočetná*. Může se jednat o hod kostkou, kde je možný jeden ze šesti výsledků ($\Omega = \{\omega_1, \omega_2, \omega_3, \omega_4, \omega_5, \omega_6\}$).

Pro lepší pochopení náhodného pokusu jsou vhodné zejména ty spojené s hazardními hrami, protože tyto pokusy je možno opakovat teoreticky nekonečně mnohokrát za stejných podmínek. Náhodný pokus může být ovšem i reakce nemocného organismu na právě podaný lék.

1.2. Náhodný jev

Prvky množiny Ω se nazývají *elementární jevy*. Jsou to základní výsledky pokusu, z nichž při každém jeho opakování nastává jen a pouze jeden výsledek. Můžeme s nimi provádět operace jako *sjednocení*, *průnik*, *rozdíl* či *doplňek*. Náhodnému pokusu přísluší širší množina výsledků, jež získáme množinovými operacemi nad základními výsledky čili elementárními jevy. O každém výsledku přitom můžeme

určit, zda nastal, či nenastal.

Definice 1. Nechť $\Omega \neq \emptyset$ je libovolná množina. Neprázdný systém \mathcal{A} podmnožin množiny Ω se nazývá σ -algebra, platí-li

- a) $A \in \mathcal{A} \Rightarrow A^c \in \mathcal{A}$,
- b) $A_n \in \mathcal{A}, n= 1,2, \dots \Rightarrow \bigcup_1^\infty A_n \in \mathcal{A}$.

V případě, že uvažujeme neprázdnou množinu všech možných výsledků $\Omega \neq \emptyset$, se prvky $A \in \mathcal{A}$ nazývají *náhodné jevy* a \mathcal{A} se nazývá též *jevové pole*. V případě opakovatelného náhodného pokusu hovoříme o *hromadných náhodných jevech*. Vztahy mezi těmito jevy se dají nejlépe popsat graficky, především pomocí Vennových diagramů. Mezi náhodné jevy patří i jevy elementární.

1.3. Pravděpodobnost náhodného jevu

Pravděpodobnost náhodného jevu A je číslo $P(A)$, které vysvětlíme jako míru (šanci, pravděpodobnost) nastoupení daného jevu A . Teorie pravděpodobnosti tuto míru definuje jako funkci, která každému náhodnému jevu přiřazuje reálné číslo, které charakterizuje, s jakou pravděpodobností tento jev nastane.

Definice 2. Nechť je dána neprázdná množina Ω a na ní jevové pole \mathcal{A} . *Pravděpodobností* nazveme každou reálnou funkci P definovanou na \mathcal{A} , která vyhovuje následujícím axiomům :

- $a_1)$ $P(\Omega) = 1$,
- $a_2)$ $P(A) \geq 0, \forall A \in \mathcal{A}$,
- $a_3)$ pro libovolnou posloupnost $A_n \in \mathcal{A}, n= 1,2, \dots$, neslučitelných náhodných jevů platí

$$P\left(\bigcup_1^\infty A_n\right) = \sum_1^\infty P(A_n).$$

Funkci P též nazýváme *pravděpodobnostní mírou*. Uspořádanou trojici (Ω, \mathcal{A}, P)

nazýváme *pravděpodobnostní prostor* či *Kolmogorovovo pravděpodobnostní pole*. Známe několik typů pravděpodobnostních modelů. Uvedeme si zde pouze jejich výčet. Jedná se o klasickou, neklasickou, geometrickou, podmíněnou pravděpodobnost či pravděpodobnost v případě, že Ω je nekonečná spočetná.

1.4. Náhodná veličina

Při řešení různých pravděpodobnostních úloh nahrazujeme náhodné jevy jistými hodnotami proměnlivé veličiny, označované jako *náhodná veličina*. Setkáváme se i s variantou tohoto pojmu jako *náhodná proměnná*. Náhodná veličina je jedním z nejdůležitějších a stěžejních pojmů teorie pravděpodobnosti. Nyní uvažujme prostor elementárních jevů, respektive jevové pole, které přísluší určitému náhodnému pokusu. Bude nás zajímat reálná funkce elementárních jevů, respektive náhodných jevů, která každému takovému jevu přiřadí reálné číslo.

Náhodná veličina je tedy funkcí, jejíž výsledné hodnoty jsou ovlivněny náhodou, tj. nelze je dopředu odhadnout se 100% jistotou. Např. uvažujeme výšku studentů vybraných k testování či velikost reakce organismu při zkoušce léků. Každému výsledku lze přiřadit číslo, i v tom případě, že výsledek má kvalitativní charakter, např. vyrobený výrobek je vadný - přiřadíme 0, výrobek je bez vady - přiřadíme 1. Náhodnou veličinu plně charakterizuje rozdělení pravděpodobnosti či distribuční funkce (viz kapitoly 1.5. a 1.6.). Náhodné veličiny se zpravidla značí velkými písmeny z konce abecedy, např. X, Y, Z . Jejich realizace se naopak označují malými písmeny z konce abecedy, např. x, y, z . Náhodné veličiny dělíme na diskrétní a spojité.

Definice 3. Nechť je dán pravděpodobnostní prostor (Ω, \mathcal{A}, P) . Reálnou funkci $X : \Omega \rightarrow \mathbb{R}^1$ nazveme *náhodnou veličinou*, jestliže pro každé $x \in \mathbb{R}^1$ platí

$$\{\omega \in \Omega : X(\omega) \leq x\} \in \mathcal{A}.$$

Reálná funkce $X(\omega), \omega \in \Omega$, splňující uvedený vztah, se nazývá *\mathcal{A} -měřitelná*. Množinu $M \subset \mathbb{R}^1$ všech hodnot náhodné veličiny X nazýváme *obor hodnot náhodné veličiny X* .

1.5. Rozdělení pravděpodobností náhodné veličiny

Jedním z úkolů teorie pravděpodobnosti je zkonstruovat matematický nástroj, který by každému popisovanému jevu přiřadil pravděpodobnost. Hovoříme tedy o typu zobrazení, kde každému jevu přiřadíme určité reálné číslo, které bude vyjadřovat, s jakou pravděpodobností tento jev nastane. Např. konstatovat, že pravděpodobnost toho, že výsledek se realizuje v intervalu $(10,25)$, je rovna 0,93. Náhodná veličina X zobrazuje výsledky pokusů do oboru reálných čísel, elementárním jevům přiřazuje čísla, náhodným jevům pak množiny čísel. Na reálné přímce jsou typickými číselnými množinami intervaly, např. $\langle a, b \rangle$, $(-\infty, b)$ či $\langle a, \infty)$. Nejmenší σ -algebra obsahující tyto intervaly se nazývá borelovská a značíme ji \mathcal{B}_1 . Náhodná veličina tedy zobrazuje původní prostor (Ω, \mathcal{A}, P) na nový prostor $(\mathbb{R}^1, \mathcal{B}_1, P)$, který opět lze chápat jako prostor pravděpodobnostní. Pravděpodobnostní míra $P=P_X$ na \mathbb{R}^1 se nazývá indukovaná a splňuje

$$P(B) = P(X \in B) = P(\omega \in \Omega : X(\omega) \in B) \quad \forall B \in \mathcal{B}_1.$$

Podle kontextu chápeme P jako míru buď na Ω , nebo v \mathbb{R}^1 .

Definice 4. *Rozdělení pravděpodobností náhodné veličiny X je indukovaná pravděpodobnostní míra, tedy množinová funkce $P_X(B) : \mathcal{B}_1 \rightarrow \mathbb{R}^1$ definovaná vztahem*

$$P_X(B) = P(X \in B), \quad B \in \mathcal{B}_1.$$

Pro lepší pochopení si uvedeme příklad s hodem mincí. Množina možných výsledků $\Omega = \{\text{líc}, \text{rub}\}$. Náhodná veličina je zobrazení do množiny $\{1,0\}$, kde jsme si nastoupení jevu „padne líc“ označili jako 1 a naopak „padne rub“ jako 0. Symbolicky zapíšeme $X : \Omega \rightarrow \{1,0\}$. Nastoupení jednoho či druhého jevu je přiřazena stejná pravděpodobnost: $P(1) = P(0) = \frac{1}{2}$.

Definice 5. Předpokládejme, že existuje konečná nebo spočetná prostá posloup-

nost reálných čísel $\{x_n\}$ taková, že

$$\sum_n P(X = x_n) = 1.$$

Označme $p_n = P(X = x_n)$. Posloupnost $\{x_n\}$ hodnot, kterých nabývá náhodná veličina X , a posloupnost $\{p_n\}$ pravděpodobností, s nimiž náhodná veličina svých hodnot nabývá, určují tzv. *diskrétní rozdělení pravděpodobností náhodné veličiny* X . Náhodná veličina, která má diskrétní rozdělení pravděpodobností, se nazývá diskrétní, resp. diskrétního typu. Posloupnost $\{x_n\}$ se nazývá též *pravděpodobnostní funkce*.

Definice 6. Řekneme, že náhodná veličina X má *spojité rozdělení pravděpodobností* (má rozdělení spojitého typu, je spojitá), existuje-li nezáporná borelovsky měřitelná funkce $f_X(x) : \mathbb{R}^1 \rightarrow \mathbb{R}^1$ taková, že

$$F_X(x) = \int_{-\infty}^x f_X(t) dt, \quad \forall x \in \mathbb{R}^1.$$

Funkce f_X se nazývá *hustota* (rozdělení pravděpodobností) náhodné veličiny X .

1.6. Distribuční funkce náhodné veličiny

Jednou z možností, jak plně určit rozdělení pravděpodobností náhodné veličiny X , je užít tzv. *distribuční funkci náhodné veličiny*. Je to reálná funkce reálného argumentu a s takovou funkcí se dobře pracuje.

Definice 7. Nechť X je náhodná veličina definovaná na pravděpodobnostním prostoru (Ω, \mathcal{A}, P) . Reálná funkce F_X definovaná na \mathbb{R}^1 předpisem

$$F_X(x) = P(X \leq x), \quad x \in \mathbb{R}^1,$$

se nazývá *distribuční funkce náhodné veličiny* X .

Dvě odlišné náhodné veličiny mohou mít tutéž distribuční funkci. Uvažujme např.

$\Omega = \{\omega_1, \omega_2\}$, \mathcal{A} je systém všech podmnožin množiny Ω , tedy $\mathcal{A} = \{\emptyset, \Omega, \omega_1, \omega_2\}$, $P(\omega_1) = \frac{1}{2}$, $P(\omega_2) = \frac{1}{2}$. Definujeme-li dvě náhodné veličiny $X(\omega_1) = 1$, $X(\omega_2) = 0$, $Y(\omega_1) = 0$, $Y(\omega_2) = 1$, dostaneme

$$F_X(x) = F_Y(x) = \begin{cases} 0, & x < 0, \\ \frac{1}{2}, & 0 \leq x < 1, \\ 1, & x \geq 1. \end{cases}$$

Distribuční funkce diskrétní náhodné veličiny je dána vztahem

$$F_X(x) = \sum_{n: x_n \leq x} p_n, \quad \forall x \in \mathbb{R}^1.$$

Tato funkce se nazývá *diskrétní distribuční funkce*.

Náhodnou veličinu X můžeme detailně popsat pomocí distribuční funkce $F_X(x)$, $x \in \mathbb{R}$, která nám o této veličině řekne plnou informaci. Znalost distribuční funkce je ekvivalentní se znalostí rozdělení pravděpodobností $P_X(B)$, $B \in \mathcal{B}_1(\mathbb{R}^1)$.

Kapitola 1 byla zpracována s pomocí [1] a [6].

2. Diskrétní rozdělení pravděpodobností

U diskrétní náhodné veličiny užíváme znalost pravděpodobnostní funkce $\{p_k\}$, kde $p_k = P(X = x_k)$. V této kapitole uvedeme některé nejdůležitější zástupce diskrétních rozdělení.

2.1. Alternativní rozdělení

Nejprve si uveďme situaci, ve které nás zajímá pouze to, zda náhodný jev, označíme ho jako A , nastal či nenastal. Zadanou situaci nejlépe popisuje *alternativní* neboli také *nula-jedničkové rozdělení*. Náhodná veličina X nabývá pouze dvou hodnot, a to $X = 1$, pokud jev A nastal, nebo $X = 0$, pokud tento jev nenastal. Jevy mohou nastat s pravděpodobnostmi

$$P(A) = P(X = 1) = p_1, \quad P(\neg A) = P(X = 0) = p_0, \quad \text{kde } 0 < p_0, p_1 < 1.$$

Pravděpodobnost $p_1 = p$ se týká situace, kdy jev A nastane. Náhodnou veličinu mající alternativní rozdělení krátce označujeme $X \sim \text{Alt}(p)$. Distribuční funkce je ve tvaru

$$F(x) = \begin{cases} 0, & \text{pro } x < 0, \\ 1 - p, & \text{pro } 0 \leq x < 1, \\ 1, & \text{pro } x \geq 1. \end{cases}$$

Alternativní rozdělení má například náhodná veličina, která udává počet šestek, které padnou při jednom hodu kostkou.

2.2. Binomické rozdělení

Nyní si promysleme jinou situaci. Obdobně jako v 2.1 označme $X_i = 1$ nastoupení jevu A v i -tém pokusu. Náhodný pokus n -krát zopakujeme. Pokud pravděpodobnost nastoupení jevu A v každém pokusu nezávisí na výsledcích pokusů předcházejících (pokusy jsou tedy nezávislé), potom jsou náhodné veličiny $X_i, i = 1, 2, \dots, n$, nezávislými alternativními veličinami.

Dále předpokládejme, že pro všechny pokusy je pravděpodobnost p nastoupení náhodného jevu A stejná. To znamená, že náhodné veličiny $X_i, i = 1, 2, \dots, n$,

mají všechny alternativní rozdělení s parametrem p . Potom definujeme náhodnou veličinu

$$X = \sum_{i=1}^n X_i$$

a její rozdělení nazveme *binomické rozdělení* s parametry n a p . Toto rozdělení je dáno pravděpodobnostní funkcí

$$p_k = P(X = k) = \binom{n}{k} p^k (1-p)^{n-k}, \quad k = 0, 1, \dots, n,$$

kde $n \in \mathbb{N}$ a $p \in (0,1)$ jsou výše zmíněné parametry tohoto rozdělení. Krátce označujeme $X \sim \text{Bi}(n, p)$. Uveďme si opět distribuční funkci rozdělení

$$F(x) = \begin{cases} 0, & \text{pro } x < 0, \\ \sum_{k \leq x} \binom{n}{k} p^k (1-p)^{n-k}, & \text{pro } 0 \leq x < n, \\ 1, & \text{pro } x \geq n. \end{cases}$$

Jako příklad binomického rozdělení uvedeme počet nekvalitních výrobků mezi n výrobky nebo počet šestek, které padnou v n hodech na hrací kostce.

2.3. Poissonovo rozdělení

Dalším diskrétním rozdělením, které si uvedeme, je *Poissonovo rozdělení*. Bude nás zajímat, kolikrát se jev A vyskytne v jistém časovém nebo objemovém intervalu, přičemž se jedná o počet výskytů řídkce se vyskytujícími náhodnými jevy. Pravděpodobnostní funkce bude ve tvaru

$$p_k = P(X = k) = \frac{\lambda^k}{k!} e^{-\lambda}, \quad k = 0, 1, \dots,$$

kde $\lambda > 0$ je parametr tohoto rozdělení. Krátce označujeme $X \sim \text{Po}(\lambda)$. Dále si uveďme distribuční funkci Poissonova rozdělení

$$F(x) = \begin{cases} 0, & \text{pro } x < 0, \\ \sum_{k \leq x} \frac{\lambda^k}{k!} e^{-\lambda}, & \text{pro } x \geq 0. \end{cases}$$

Než se přesuneme k poslednímu diskrétnímu rozdělení v našem výčtu, uvedeme si zajímavý vztah mezi binomickým a Poissonovým rozdělením. Pro $n > 30$ a $p < 0,1$ se doporučuje rozdělení $\text{Bi}(n, p)$ aproximovat rozdělením $\text{Po}(\lambda)$, kde $\lambda = np$.

2.4. Geometrické rozdělení

Posledním rozdělením, které si v kategorii diskrétních uvedeme, bude *geometrické rozdělení*. Nastiňme si opět vzorovou situaci. Předpokládáme, že provádíme nezávisle na sobě pokusy a že pravděpodobnost nastoupení sledovaného jevu v jednom pokuse je stále stejná. Náhodná veličina X označuje počet neúspěšných pokusů, které předcházejí prvnímu úspěšnému, přičemž pravděpodobnost úspěchu je p . Pravděpodobnost toho, že k úspěchu dojde až v $(k + 1)$ -ním pokusu, je vzhledem k nezávislé povaze pokusů

$$p_k = P(X = k) = (1 - p)^k p, \quad k = 0, 1, 2, \dots,$$

kde $p \in (0, 1)$ je parametr tohoto rozdělení. Označujeme krátce $X \sim \text{Ge}(p)$. Uvedeme si opět distribuční funkci

$$F(x) = \begin{cases} 0, & \text{pro } x < 0, \\ \sum_{k \leq x} p(1 - p)^k, & \text{pro } x \geq 0. \end{cases}$$

Geometrické rozdělení má náhodná veličina, která je rovna počtu pokusů, které předcházejí padnutí čísla 6 na hrací kostce.

3. Spojitá rozdělení pravděpodobnosti

U diskrétní náhodné veličiny jsme si uvedli, že k jejímu popisu používáme znalosti pravděpodobnostní funkce p_k . Ve spojitém případě nám poslouží znalost její hustoty f_X , někdy označované také jako frekvenční funkce. Dále si uvedeme několik nejdůležitějších případů spojitého rozdělení náhodné veličiny.

3.1. Rovnoměrné rozdělení

Nejjednodušším případem spojitého rozdělení je *rovnoměrné rozdělení*. Toto rozdělení má náhodná veličina X , jejíž hustota pravděpodobnosti je konstantní pro všechna $x \in (a, b)$. Pak tato hustota vypadá

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

Označujeme krátce $X \sim \text{Ro}(a, b)$, kde $a, b \in \mathbb{R}^1, a < b$, jsou parametry tohoto rozdělení. Distribuční funkce je rovna

$$F(x) = \begin{cases} 0, & \text{pro } x \leq a, \\ \int_a^x \frac{1}{b-a} dt = \frac{x-a}{b-a}, & \text{pro } a < x < b, \\ 1, & \text{pro } x \geq b. \end{cases}$$

Rovnoměrné rozdělení má například chyba při stanovení času z digitálních hodin.

3.2. Normální rozdělení

(Obecné) normální rozdělení má v oblasti teorie pravděpodobnosti a matematické statistice zásadní význam. Jeho síla je především v tom, že za jistých okolností, které formuluje centrální limitní věta, velmi dobře aproximuje celou řadu jiných, i nespojitých rozdělení pravděpodobností. Hustota náhodné veličiny X , která má normální rozdělení, vypadá

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

Označujeme krátce $X \sim \text{N}(\mu, \sigma^2)$, kde $\mu \in \mathbb{R}^1, \sigma^2 > 0$ jsou parametry. Distribuční funkce je rovna

$$F(x) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{(t-\mu)^2}{2\sigma^2}} dt, \quad x \in \mathbb{R}^1.$$

Ovšem pro určení hodnot (obecného) normálního rozdělení stačí tabelovat hodnoty distribuční funkce standardního normálního rozdělení ($\mu = 0, \sigma^2 = 1$), značíme jako $\Phi(x)$, protože platí

$$F(x) = \Phi\left(\frac{x-\mu}{\sigma}\right), \quad x \in \mathbb{R}^1.$$

Normální rozdělení mají například chyby, které vznikají při opakovaném měření fyzikálních veličin, nebo výška obyvatel.

3.3. Exponenciální rozdělení

Dalším příkladem spojitého rozdělení je *exponenciální rozdělení*. S tímto rozdělením se můžeme často setkat v teorii spolehlivosti a životnosti. Za náhodnou veličinu X potom můžeme považovat dobu do nastoupení sledovaného jevu, např. selhání. Exponenciální rozdělení s parametrem $\lambda > 0$ má náhodná veličina, která má hustotu

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

Označujeme krátce $X \sim E(\lambda)$. Distribuční funkce je rovna

$$F(x) = \begin{cases} 1 - e^{-\lambda x}, & \text{pro } x > 0, \\ 0, & \text{jinak.} \end{cases}$$

Příkladem exponenciálního rozdělení může být časový interval mezi příchozími lidmi do obchodu.

3.4. Cauchyho rozdělení

Posledním zástupcem spojitých rozdělení, kterého si uvedeme, je *Cauchyho rozdělení*. Je to uměle vytvořené rozdělení. Cauchyho rozdělení s parametry a, b má hustotu ve tvaru

$$f(x) = \frac{1}{\pi} \frac{b}{b^2 + (x-a)^2}, \quad x \in \mathbb{R}^1.$$

Označujeme krátce $X \sim C(a, b)$, kde $a \in \mathbb{R}$, $b > 0$ jsou parametry tohoto rozdělení. Toto rozdělení charakterizuje podíl dvou nezávislých náhodných veličin X, Y s normálním rozdělením. Pokud $X \sim N(0, 1), Y \sim N(0, 1)$ jsou nezávislé náhodné veličiny, potom $\frac{X}{Y} \sim C(0, 1)$.

Kapitoly 2. a 3. s výčtem různých typů rozdělení byly čerpány z [1] a [3]. Informace o exponenciálním rozdělení jsou převzaty [7].

4. Číselné charakteristiky náhodné veličiny

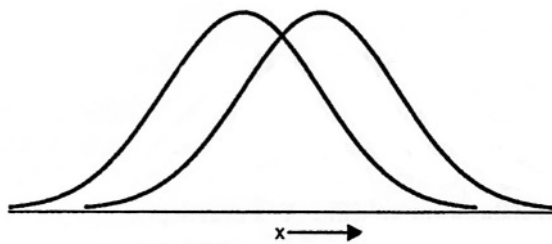
Distribuční funkce jednoznačně charakterizuje pravděpodobnostní chování náhodné veličiny, tudíž nám podává o náhodné veličině úplnou informaci. Občas může ale tato informace být poněkud nepřehledná. Pro lepší a názornější interpretaci lze tuto informaci shrnout pouze do několika čísel. Taková čísla výstižně popisují základní vlastnosti rozdělení náhodné veličiny. Nazýváme je *charakteristikami náhodné veličiny* či *charakteristikami příslušného rozdělení*. Níže si uvedeme nejznámější a nejpoužívanější z nich. Jejich výpočet pro vybraná rozdělení provedeme v kapitole 6.

4.1. Charakteristiky centrální tendence (polohy)

Mějme náhodnou veličinu X a k ní příslušnou distribuční funkci $F(x)$. Potom náhodná veličina $Y = X + a$ má distribuční funkci $G(x) = P(Y \leq x) = P(X + a \leq x) = P(X \leq x - a) = F(x - a)$. Pokud chceme získat graf distribuční funkce $G(x)$ náhodné veličiny Y z grafu distribuční funkce $F(x)$ náhodné veličiny X , tak tento graf stačí posunout pouze o hodnotu a . Rozdělení těchto dvou náhodných veličin se tedy od sebe liší jen polohou, viz obrázek 1. Charakteristikou polohy nazýváme takovou charakteristiku $\xi(X)$, která je ekvivariantní vůči posunutí, tedy

$$\xi(X + a) = \xi(X) + a.$$

Takové vlastnosti mají střední hodnota, modus či medián.



Obrázek 1: Rozdělení lišící se polohou

4.1.1. Střední hodnota

Střední (neboli očekávaná) hodnota je číselnou charakteristikou polohy hodnot náhodné veličiny X na reálné ose. Známe-li tedy střední hodnotu $E(X)$, víme, kde jsou hodnoty náhodné veličiny koncentrovány a kde je s největší pravděpodobností můžeme očekávat. Anglický překlad slova „očekávaný“ vidíme i v symbolickém značení střední hodnoty: $E =$ „expected“.

Pro diskrétní náhodnou veličinu X s rozdělením $\{x_k\}, \{p_k\}$ se střední hodnota definuje jako

$$E(X) = \sum_k x_k p_k = \sum_k x_k P(X = x_k). \quad (4.1)$$

V tomto případě každou realizaci náhodné veličiny násobíme její pravděpodobností, tzn. že náhodná veličina této realizace nabude. Laicky řečeno, každou hodnotu náhodné veličiny „vážíme její šancí“. Hodnoty náhodné veličiny mohou být kladná ale i záporná čísla. Pro existenci střední hodnoty budeme vždy požadovat absolutní konvergenci příslušných řad.

Nechť X je spojitá náhodná veličina s hustotou $f_X(x)$. Její střední hodnota se definuje jako

$$E(X) = \int_{-\infty}^{\infty} x f_X(x) dx. \quad (4.2)$$

Opět předpokládáme absolutní konvergenci integrálu. Dále si na tomto místě definujeme pojem *moment*, který budeme využívat v dalším textu.

Definice. Nechť X je náhodná veličina definovaná na pravděpodobnostním prostoru (Ω, \mathcal{A}, P) a nechť $r = 1, 2, \dots$, potom

$E(X^r)$ se nazývá *r-tý (počáteční nebo obecný) moment*,
 $E[X - E(X)]^r$ se nazývá *r-tý centrální moment*.

Pro diskrétní náhodnou veličinu X je *r-tý obecný moment* definován jako

$$E(X^r) = \sum_k x_k^r p_k.$$

Pro spojitou náhodnou veličinu X je r -tý obecný moment definován jako

$$E(X^r) = \int_{-\infty}^{\infty} x^r f(x) dx.$$

Ze všech obecných momentů má největší význam ten první, $E(X^1) = E(X)$, který tedy nazýváme *střední hodnotou*.

4.1.2. Modus

Další charakteristikou polohy, kterou si uvedeme, je *modus* (nazývaný také jako typická hodnota). Označujeme $\text{mod}(X)$. V případě diskrétní náhodné veličiny X je modus taková hodnota x_k^* , která pro všechny hodnoty x_k náhodné veličiny X splňuje podmínku

$$P(X = x_k^*) \geq P(X = x_k), \quad k = 1, 2, \dots$$

Naopak pro spojitou náhodnou veličinu s hustotou $f(x)$ je modus taková hodnota x^* určená podmínkou

$$f(x^*) \geq f(x), \quad -\infty < x < \infty.$$

Modus ovšem nemusí být jednoznačně určen, tzn. může se vyskytovat více hodnot s nejvyšší frekvencí. Rozdělení pravděpodobností, kde se vyskytuje jeden modus, se nazývá *jednovrcholové (unimodální)*. Tam, kde se vyskytují dva mody, se rozdělení nazývá *dvouvrcholové (bimodální)*. Nespornou výhodou modu je to, že ho lze použít i pro data, která nejsou číselná. Např. když pozorujeme a zaznamenáváme počty aut s různou barvou laku.

4.1.3. Medián

Třetí a poslední charakteristikou polohy náhodné veličiny, kterou si uvedeme, je *medián*. Medián je nejvýznamější kvantilovou charakteristikou. Abychom si ujasnili význam mediánu, definujeme si pojem *kvantil*.

Definice. Nechť $\alpha \in (0, 1)$. α -kvantil náhodné veličiny X je takové reálné číslo x_α , pro které platí

$$P(X \leq x_\alpha) \geq \alpha \quad \text{a současně} \quad P(X \geq x_\alpha) \geq 1 - \alpha.$$

Nejužívanějším typem kvantilu je právě medián, čili 50%-ní kvantil, tedy číslo $x_{0,5}$, které nám dělí obor hodnot náhodné veličiny na dvě stejně pravděpodobné části. Nadále jej budeme označovat jako $\text{med}(X)$.

V případě diskrétního rozdělení pravděpodobností platí

$$\sum_{k: x_k \leq \text{med}(X)} p_k \geq \frac{1}{2} \quad \text{a} \quad \sum_{k: x_k \geq \text{med}(X)} p_k \geq \frac{1}{2}.$$

Naopak při spojitém rozdělení pravděpodobností pro medián platí

$$P(X \leq \text{med}(X)) = P(X \geq \text{med}(X)) = \frac{1}{2} \quad \text{neboli} \quad \int_{-\infty}^{\text{med}(X)} f(x) dx = 0,5.$$

Je-li spojitě unimodální rozdělení symetrické kolem nějakého bodu a , pak tento bod je zároveň střední hodnotou, modem i mediánem. Takovou vlastnost například splňuje obecné normální rozdělení $N(\mu, \sigma^2)$, kde $\mu = E(X) = \text{mod}(X) = \text{med}(X)$.

4.2. Charakteristiky variability

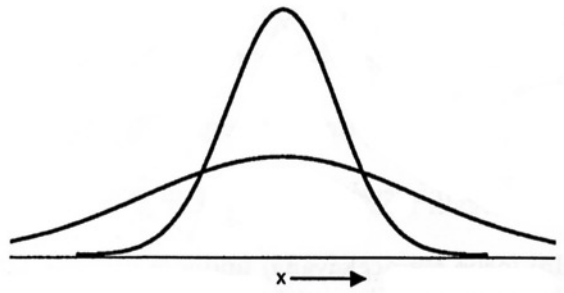
Do této kategorie zahrnujeme charakteristiky $\eta(X)$, u kterých nemá vliv změna polohy, tj. jsou invariantní vůči posunutí, tedy posunutím zůstává taková charakteristika nezměněna:

$$\eta(X + a) = \eta(X),$$

a pro které dále platí

$$\eta(a + bX) = b^2\eta(X), \quad \text{nebo} \quad \eta(a + bX) = |b|\eta(X), \quad \forall a, b \in \mathbb{R}^1.$$

V dalším textu si z této kategorie uvedeme rozptyl, mezikvartilové rozpětí, rozpětí a střední absolutní odchylku.



Obrázek 2: Rozdělení lišící se rozptylem

4.2.1. Rozptyl

Druhý centrální moment se nazývá *rozptyl*, neboli *variance* či *disperse*. Rozptyl je po střední hodnotě druhou nejdůležitější číselnou charakteristikou náhodné veličiny. Existence rozptylu je ovšem podmíněna existencí střední hodnoty, bez ní bychom nebyli schopni rozptyl spočítat. Díky rozptylu si můžeme udělat obrázek, jako moc nebo málo jsou hodnoty náhodné veličiny koncentrovány kolem střední hodnoty, viz obrázek 2. V případě náhodné veličiny X jej značíme jako $\text{var } X$ a definujeme

$$\text{var}(X) = E[X - E(X)]^2. \quad (4.3)$$

Druhé odmocnině z rozptylu, $\sqrt{\text{var}(X)}$, říkáme *směrodatná (standardní) odchylka* náhodné veličiny X .

4.2.2. Střední absolutní odchylka

Další charakteristikou, kterou si uvedeme, je *střední absolutní odchylka*. Její teoretická hodnota je definována jako

$$E|X - \text{med}(X)|. \quad (4.4)$$

V literatuře se můžeme setkat s modifikací tohoto vzorce v podobě $E|X - E(X)|$ či $\text{med}|X - \text{med}(X)|$.

4.2.3. Rozpětí

Předposlední charakteristikou v tomto seznamu je *teoretické rozpětí*. Jedná se o šířku oboru hodnot náhodné veličiny. Například u $Ro(a, b)$ je to vzdálenost $b - a$ či u $N(\mu, \sigma^2)$, kde oborem hodnot je celá reálná osa, je rozpětí nekonečné. Jinak řečeno, rozpětí je rozdíl mezi maximální a minimální hodnotou náhodné veličiny

$$R = \max(X) - \min(X). \quad (4.5)$$

4.2.4. Mezikvartilové rozpětí

Poslední charakteristikou variability, kterou si na tomto místě uvedeme, je *mezikvartilové rozpětí*. To je definováno jako rozdíl mezi třetím a prvním kvantilem, tedy

$$R_Q = x_{0.75} - x_{0.25}. \quad (4.6)$$

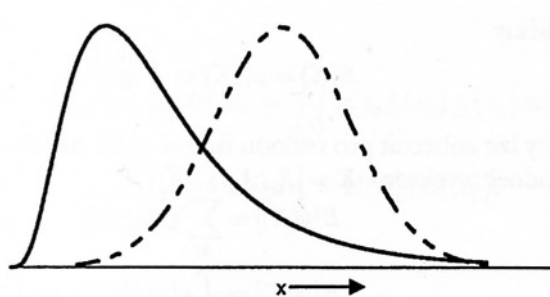
4.3. Charakteristiky šikmosti a špičatosti

4.3.1. Šikmost

Momentový koeficient šikmosti (zkráceně *šikmost*) je nejčastěji užívanou charakteristikou asymetrie, viz obrázek 3. Zavádí se podle vzorce

$$\alpha_3(X) = \frac{E[(X - E(X))^3]}{(\sqrt{\text{var}(X)})^3}. \quad (4.7)$$

V případě, že $\alpha_3(X) = 0$, je rozdělení naprosto symetrické. Když $\alpha_3(X) > 0$, je rozdělení protáhlejší směrem napravo a naopak, pokud je $\alpha_3(X) < 0$, je rozdělení protáhlejší směrem nalevo. Pro záporně zešikmené rozdělení platí $E(X) < \text{med}(X) < \text{mod}(X)$, pro symetrické rozdělení platí $E(X) = \text{med}(X) = \text{mod}(X)$. Tato rovnost platí pouze pro spojitě unimodální rozdělení. Pro kladně zešikmené rozdělení platí $E(X) > \text{med}(X) > \text{mod}(X)$.



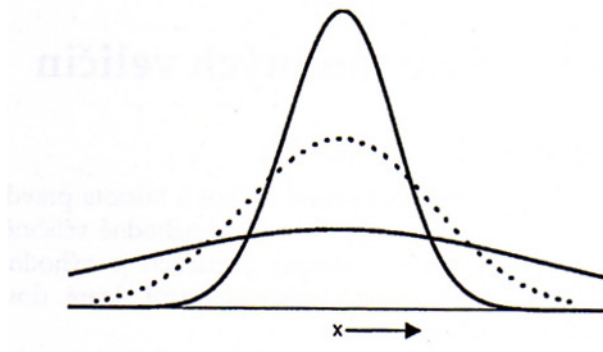
Obrázek 3: Rozdělení lišící se šikmostí

4.3.2. Špičatost

Momentový koeficient špičatosti (zkráceně *špičatost*) je nejčastěji užívanou charakteristikou plochosti. Platí pro ni obecný vzorec

$$\alpha_4(X) = \frac{E[(X - E(X))^4]}{(\sqrt{\text{var}(X)})^4}. \quad (4.8)$$

V případě, že $\alpha_4(X) = 3$, potom se jedná o normální rozdělení. Když $\alpha_4(X) > 3$, potom to znamená, že pravděpodobnostní funkce p_n nebo hustota f_X je na svých koncích větší než hustota normálního rozdělení se stejnou střední hodnotou i rozptylem. Naopak, v případě, že $\alpha_4(X) < 3$, potom je pravděpodobnostní funkce p_k nebo hustota f_X menší než hustota normálního rozdělení opět se stejnou střední hodnotou a rozptylem, viz obrázek 4. Špičatost je vždy kladná.



Obrázek 4: Rozdělení lišící se špičatostí

5. Výpočet empirických charakteristik

Nyní si uvedeme vzorce pro výpočet charakteristik v *jednorozměrném statistickém souboru*. Tímto souborem nazýváme množinu *statistických jednotek*. Tyto jednotky mohou představovat osoby, organizace či vozidla. Jedná se o výběr (zpravidla náhodný) ze širší populace. Na těchto jednotkách pozorujeme pouze jeden *statistický znak* X (hodnotu náhodné veličiny X), proto tento soubor nazýváme *jednorozměrným*. Při výpočtu charakteristik pracujeme s napozorovanými daty x_1, \dots, x_n , čili se *souborem hodnot*. Hodnoty v tomto souboru se mohou opakovat.

5.1. Charakteristiky polohy

5.1.1. Střední hodnota

Výběrová střední hodnota \bar{x} se vypočte jako obyčejný aritmetický průměr, tedy

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i. \quad (5.1)$$

Tato hodnota nám určuje pomyslný střed, kolem kterého se hodnoty seskupují. Dále si uvedeme jednu důležitou vlastnost vyplývající z podstaty výběrové střední hodnoty (aritmetického průměru), a to tu, že součet odchylek naměřených hodnot od tohoto průměru se rovná nule

$$\sum_{i=1}^n (x_i - \bar{x}) = 0.$$

5.1.2. Výběrový modus

Výběrový modus \hat{x} je takovou variantou znaku, která má nejvyšší četnost v rámci celého souboru hodnot. Smysl jej má používat tehdy, je-li počet vzájemně různých variant znaku ve statistickém souboru výrazně menší než rozsah souboru. Určíme jej jako nejčastější hodnotu vyskytující se v souboru. Označme x_{i_1}, \dots, x_{i_m} různé varianty sledovaného znaku, jejichž počet je m , n je počet pozorování (rozsah

souboru). Definujme n_j četnost x_{i_j} , $j = 1, \dots, m$. Označme n_{j_0} ,

$$\text{tj. } n_{j_0} = \max_j n_j,$$

největší četnost. Pro výběrový modus \hat{x} platí vztah

$$\hat{x} = x_{i_{j_0}}. \quad (5.2)$$

V případě dat ze spojitého rozdělení se modus zpravidla nepoužívá.

5.1.3. Výběrový medián

Výběrový medián \tilde{x} je charakteristikou, která dělí statistický soubor na dvě stejně početné poloviny. Jedná se o výběrový 0,5 kvantil, tedy $\tilde{x} = \tilde{x}_{0,5}$. K jeho určení stačí soubor dat seřadit od nejmenší po největší hodnotu. V dalším textu budeme užívat značení $x_{(.)}$, kde výraz v závorce značí pořadí (pozici) prvku v seřazeném souboru. Při lichém počtu dat je mediánem prostřední hodnota, tedy

$$\tilde{x} = x_{(\frac{n+1}{2})}. \quad (5.3)$$

Při sudém počtu hodnot je mediánem průměr dvou prostředních hodnot, tedy

$$\tilde{x} = \frac{x_{(\frac{n}{2})} + x_{(\frac{n+1}{2})}}{2}. \quad (5.4)$$

Výhodou mediánu je to, že není ovlivněn extrémními hodnotami. U šikmých rozdělení je například nevhodné použití aritmetického průměru, ale naopak medián lépe vystihuje tento soubor. Ukázkovým příkladem na rozdíl mezi aritmetickým průměrem a mediánem může být často v médiích zmiňovaný průměrný měsíční plat 1 obyvatele v ČR. S použitím aritmetického průměru a dat z Českého statistického úřadu vychází průměrná měsíční hrubá mzda ke dni 24.4.2013 na 27 170,- Kč [8]. Při použití mediánu by se ale tato hranice hrubé průměrné měsíční mzdy pohybovala podstatně níže.

5.2. Charakteristiky variability

5.2.1. Výběrový rozptyl

Jak jsme již uvedli dříve, rozptyl udává, jak jsou data více či méně koncentrovány kolem střední hodnoty. Pro výběrový rozptyl užíváme vztah

$$s^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2. \quad (5.5)$$

Odhad rozptylu počítaný dle tohoto vzorce je nevychýlený a konzistentní. Můžeme ovšem použít i jiný vzorec, a to

$$(s^2)' = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2. \quad (5.6)$$

V tomto případě má odhad nižší hodnotu, ale je vychýlený.

5.2.2. Střední absolutní odchylka

Dále si uvedeme vzorec pro střední absolutní odchylku d , pro tu platí

$$d = \frac{1}{n} \sum_{i=1}^n |x_i - \bar{x}|. \quad (5.7)$$

5.2.3. Výběrové rozpětí

Tato hodnota udává rozdíl mezi maximální a minimální hodnotou, tedy

$$\hat{R} = \max_i x_i - \min_i x_i. \quad (5.8)$$

5.2.4. Mezikvartilové rozpětí

Tuto hodnotu definujeme jako rozdíl mezi horním a dolním výběrovým kvantilem, tedy

$$\hat{R}_Q = \tilde{x}_{0.75} - \tilde{x}_{0.25}. \quad (5.9)$$

5.3. Charakteristiky šikmosti a špičatosti

5.3.1. Výběrová šikmost

Pro výběrovou šikmost používáme vzorec

$$\hat{\alpha}_3 = \frac{1}{n-1} \frac{\sum_{i=1}^n (x_i - \bar{x})^3}{s^3}, \quad (5.10)$$

kde výraz s^3 je 3. mocnina výběrové směrodatné odchylky

$$s^3 = \left[\sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2} \right]^3. \quad (5.11)$$

5.3.2. Výběrová špičatost

Pro výběrovou špičatost používáme vzorec

$$\hat{\alpha}_4 = \frac{1}{n-1} \frac{\sum_{i=1}^n (x_i - \bar{x})^4}{s^4}, \quad (5.12)$$

kde výraz s^4 je 4. mocnina výběrové směrodatné odchylky (druhá mocnina výběrového rozptylu)

$$s^4 = (s^2)^2. \quad (5.13)$$

V případě výběrové šikmosti i špičatosti dělíme celý výraz $n-1$, nikoliv pouze n . Činíme tak ze skutečnosti, že s použitím $n-1$ bude odhad nevychýlený, naproti vychýlenosti výsledku po dělení n . Této myšlenky využijeme v kapitolách 7 a 8, abychom získali nevychýlené odhady příslušných charakteristik.

Kapitoly 4 a 5 byly čerpány z [1], [2], [4] a [5].

6. Výpočet teoretických charakteristik

V této části si předvedeme výpočty některých charakteristik pro vybraná rozdělení. Půjde především o charakteristiky diskrétních rozdělení, u většiny spojitých rozdělení bývají tyto výpočty příliš složité.

Na začátek si uvedeme pomocná odvození, jichž využijeme u výpočtů zejména šikmosti a špičatosti diskrétních rozdělení. Utvoříme binomický rozvoj výrazu $(a - b)^n$ pro $n = 2, 3, 4$

$$(a - b)^2 = a^2 - 2ab + b^2, \quad (6.1)$$

$$(a - b)^3 = a^3 - 3a^2b + 3ab^2 - b^3, \quad (6.2)$$

$$(a - b)^4 = a^4 - 4a^3b + 6a^2b^2 - 4ab^3 + b^4. \quad (6.3)$$

Vzorec (6.1) užijeme k výpočtu rozptylu

$$\begin{aligned} \text{var}(X) &= \text{E}[X - \text{E}(X)]^2 = \text{E}\{X^2 - 2X\text{E}(X) + [\text{E}(X)]^2\} = \\ &= \text{E}(X^2) - 2\text{E}(X)\text{E}(X) + [\text{E}(X)]^2 = \text{E}(X^2) - 2[\text{E}(X)]^2 + [\text{E}(X)]^2 = \\ &= \text{E}(X^2) - [\text{E}(X)]^2. \end{aligned} \quad (6.4)$$

V těch případech, kde nebude možné triviálně určit $\text{E}(X^2)$, si pomůžeme úpravou $\text{E}(X^2) = \text{E}(X^2) - \text{E}(X) + \text{E}(X) = \text{E}[X(X - 1)] + \text{E}(X)$, odkud

$$\text{var}(X) = \text{E}[X(X - 1)] + \text{E}(X) - [\text{E}(X)]^2. \quad (6.5)$$

Dále si vysvětlíme výpočet 3. resp. 4. centrálního momentu. Zde využijeme vztahů (6.2) resp. (6.3). Začneme s výpočtem 3. centrálního momentu

$$\begin{aligned} \mu_3(X) &= \text{E}[X - \text{E}(X)]^3 = \text{E}\{X^3 - 3X^2\text{E}(X) + 3X[\text{E}(X)]^2 - [\text{E}(X)]^3\} = \\ &= \text{E}(X^3) - 3\text{E}(X^2)\text{E}(X) + 3\text{E}(X)[\text{E}(X)]^2 - [\text{E}(X)]^3 = \\ &= \text{E}(X^3) - 3\text{E}(X^2)\text{E}(X) + 2[\text{E}(X)]^3. \end{aligned} \quad (6.6)$$

V tomto vzorci se vyskytuje výraz $\text{E}(X^3)$, který často nelze triviálně určit. Výjimku tvoří podoba tohoto výrazu u alternativního rozdělení. Analogicky jako v případě $\text{E}(X^2)$, avšak poněkud delším způsobem, provedeme úpravu i v případě

$E(X^3)$. Využijeme toho, že $E[(X(X-1)(X-2))] = E(X^3) - 3E(X^2) + 2E(X)$, a proto $E(X^3) = E[(X(X-1)(X-2))] + 3E(X^2) - 2E(X)$. Dosazením do (6.6) dostaneme

$$\begin{aligned}\mu_3(X) &= E[(X(X-1)(X-2))] + 3E(X^2) - 2E(X) - 3E(X^2)E(X) + 2[E(X)]^3 = \\ &= E[(X(X-1)(X-2))] + 3\{\text{var}(X) + [E(X)]^2\} - 2E(X) - 3\{\text{var}(X) + \\ &\quad + [E(X)]^2\}E(X) + 2[E(X)]^3 = E[(X(X-1)(X-2))] + 3\text{var}(X) - \\ &\quad - 3\text{var}(X)E(X) - [E(X)]^3 + 3[E(X)]^2 - 2E(X).\end{aligned}$$

Výsledný vzorec bude ve tvaru

$$\begin{aligned}\mu_3(X) &= E[(X(X-1)(X-2))] + 3\text{var}(X)[1 - E(X)] - \\ &\quad - [E(X)]^3 + 3[E(X)]^2 - 2E(X).\end{aligned}\tag{6.7}$$

Výpočty pro 4. centrální moment budou analogické, ovšem složitější. Dosazením do (6.3) dostaneme

$$\begin{aligned}\mu_4(X) &= E[X - E(X)]^4 = \\ &= E\{X^4 - 4X^3E(X) + 6X^2[E(X)]^2 - 4X[E(X)]^3 + [E(X)]^4\} = \\ &= E(X^4) - 4E(X^3)E(X) + 6E(X^2)[E(X)]^2 - 4E(X)[E(X)]^3 + [E(X)]^4 = \\ &= E(X^4) - 4E(X^3)E(X) + 6E(X^2)[E(X)]^2 - 3[E(X)]^4.\end{aligned}\tag{6.8}$$

Jelikož $E[X(X-1)(X-2)(X-3)] = E(X^4) - 6E(X^3) + 11E(X^2) - 6E(X)$, tak $E(X^4) = E[X(X-1)(X-2)(X-3)] + 6E(X^3) - 11E(X^2) + 6E(X)$. Dosazením do (6.8) a úpravou vzorců (6.4) a (6.6) do podoby $E(X^2) = \text{var}(X) + [E(X)]^2$ a $E(X^3) = \mu_3(X) + 3E(X^2)E(X) - 2[E(X)]^3$ dostaneme

$$\begin{aligned}\mu_4(X) &= E[X(X-1)(X-2)(X-3)] + E(X^3)[6 - 4E(X)] + \\ &\quad + E(X^2)\{6[E(X)]^2 - 11\} + 6E(X) - 3[E(X)]^4 = \\ &= E[X(X-1)(X-2)(X-3)] + 6\mu_3(X) + 18E(X^2)E(X) - 12[E(X)]^3 - \\ &\quad - 4\mu_3(X)E(X) - 12E(X^2)[E(X)]^2 + 8[E(X)]^4 + 6\text{var}(X)[E(X)]^2 + \\ &\quad + 6[E(X)]^4 - 11\text{var}(X) - 11[E(X)]^2 + 6E(X) - 3[E(X)]^4 =\end{aligned}$$

$$\begin{aligned}
&= \mathbb{E}[X(X-1)(X-2)(X-3)] + 6\mu_3(X) - 4\mu_3(X)\mathbb{E}(X) + 18\text{var}(X)\mathbb{E}(X) + \\
&\quad + 18[\mathbb{E}(X)]^3 - 12\text{var}(X)[\mathbb{E}(X)]^2 - 12[\mathbb{E}(X)]^4 + 8[\mathbb{E}(X)]^4 + \\
&\quad + 6[\mathbb{E}(X)]^4 - 3[\mathbb{E}(X)]^4 - 12[\mathbb{E}(X)]^3 + 6\text{var}(X)[\mathbb{E}(X)]^2 - \\
&\quad - 11\text{var}(X) - 11[\mathbb{E}(X)]^2 + 6\mathbb{E}(X).
\end{aligned}$$

Výsledný vzorec je ve tvaru

$$\begin{aligned}
\mu_4(X) &= \mathbb{E}[X(X-1)(X-2)(X-3)] + 6\mu_3(X) - 4\mu_3(X)\mathbb{E}(X) - 6\text{var}(X)[\mathbb{E}(X)]^2 + \\
&\quad + 18\text{var}(X)\mathbb{E}(X) - 11\text{var}(X) - [\mathbb{E}(X)]^4 + 6[\mathbb{E}(X)]^3 - 11[\mathbb{E}(X)]^2 + 6\mathbb{E}(X). \quad (6.9)
\end{aligned}$$

6.1. Alternativní rozdělení

Již v předchozím textu jsme uvedli, že alternativní rozdělení má pouze jeden parametr $p \in (0, 1)$, kde $p = \mathbb{P}(X = 1)$.

Střední hodnota

Střední hodnotu spočteme jako součet hodnot náhodné veličiny násobených pravděpodobnostmi, se kterými těchto hodnot nabývá

$$\mathbb{E}(X) = \sum_k x_k \cdot \mathbb{P}(X = x_k) = 0 \cdot (1 - p) + 1 \cdot p = p$$

Modus

Modus tohoto rozdělení odvodíme jako maximum posloupnosti $\{p_k\} = \mathbb{P}(X = k)$, kde $k = 0, 1$. Tedy $p_1 = p$ a $p_0 = 1 - p$.

$$\text{mod}(X) = \begin{cases} 1, & p > \frac{1}{2}, \\ 0, & p < \frac{1}{2}. \end{cases}$$

Při $p = \frac{1}{2}$ není hodnota modu jednoznačně určena a může nabývat hodnoty 0 nebo 1.

Medián

$$\text{med}(X) = \begin{cases} 1, & p > \frac{1}{2}, \\ 0, & p < \frac{1}{2}. \end{cases}$$

Při $p = \frac{1}{2}$ není medián jednoznačně určen a lze jej volit jako libovolnou hodnotu z intervalu $< 0, 1 >$.

Rozptyl

Dle (6.4) a s využitím $E(X^2) = 0^2(1-p) + 1^2p = p$ jednoduše spočteme rozptyl tohoto rozdělení

$$\text{var}(X) = p - p^2 = p(1-p).$$

Šikmost

Již víme, že $E(X^2) = p$ a $E(X) = p$, nyní ještě dopočteme $E(X^3) = 0^3(1-p) + 1^3p = p$. Dle (6.6) dopočítáme 3. centrální moment $\mu_3(X) = p - 3p \cdot p + 2p^3 = p - 3p^2 + 2p^3 = p(1-p)(1-2p)$. Dle vzorce (4.7) je šikmost

$$\alpha_3(X) = \frac{p(1-p)(1-2p)}{p(1-p)\sqrt{p(1-p)}} = \frac{1-2p}{\sqrt{p(1-p)}}.$$

Funkce je klesající na celém svém definičním oboru, tedy na $(0, 1)$. Funkce má dva extrémy: $+\infty$, když se přibližujeme k 0 zprava, a $-\infty$ v případě, že se přibližujeme k 1 směrem zleva. Pro $p > \frac{1}{2}$ je šikmost záporná, pro $p = \frac{1}{2}$ nulová a pro $p < \frac{1}{2}$ kladná.

Špičatost

Obdobně jako výše zbývá určit $E(X^4) = 0^4(1-p) + 1^4p = p$. Dle (6.8) vypočítáme 4. centrální moment, jako $\mu_4 = p - 4p \cdot p + 6p \cdot p^2 - 3p^4 = p - 4p^2 + 6p^3 - 3p^4 = p(1 - 4p + 6p^2 - 3p^3) = p(1-p)(1-3p+3p^2)$. Dle vzorce (4.8) pro špičatost dostáváme

$$\alpha_4(X) = \frac{p(1-p)(1-3p+3p^2)}{[p(1-p)]^2} = \frac{1-3p(1-p)}{p(1-p)} = \frac{1}{p(1-p)} - 3.$$

Funkce nabývá svého minima v $p = \frac{1}{2}$ a to hodnoty 1. Maxima, konkrétně $+\infty$, nabývá funkce, pokud p přibližujeme k 0 zprava a k 1 zleva, tedy k okrajům definičního oboru.

6.2. Binomické rozdělení

Binomické rozdělení má dva parametry, $n \in \mathbb{N}$ a $p \in (0, 1)$. Označme $p_k = P(X = k) = \binom{n}{k} p^k (1-p)^{n-k}$, $k = 0, 1, \dots, n$.

Střední hodnota

Pro výpočet střední hodnoty použijeme součet přes k s příslušnými pravděpodobnostmi podle definice pro diskrétní náhodnou veličinu

$$\begin{aligned} E(X) &= \sum_{k=0}^n k \binom{n}{k} p^k (1-p)^{n-k} = \sum_{k=1}^n \frac{n!}{(k-1)!(n-k)!} p^k (1-p)^{n-k} = \\ &= np \sum_{k=1}^n \binom{n-1}{k-1} p^{k-1} (1-p)^{n-k} = np \sum_{j=0}^{n-1} \binom{n-1}{j} p^j (1-p)^{n-1-j} = \\ &= np[p + (1-p)]^{n-1} = np. \end{aligned}$$

Modus

Opět budeme hledat maximum posloupnosti $\{p_k\}$. Je-li tato posloupnost rostoucí, tak

$$\begin{aligned} \frac{p_k}{p_{k-1}} &> 1 \\ \binom{n}{k} p^k (1-p)^{n-k} &> \binom{n}{k-1} p^{k-1} (1-p)^{n-k+1} \\ \frac{n!}{k!(n-k)!} p &> \frac{n!}{(k-1)!(n-k+1)!} (1-p) \\ \frac{1}{k} p &> \frac{1}{n-k+1} (1-p) \\ (n-k+1)p &> k(1-p) \\ (n+1)p &> k(1-p) + kp \\ (n+1)p &> k. \end{aligned}$$

Modus je tedy největší celé číslo k splňující $k < (n + 1)p$. Je-li $p_k = p_{k-1}$, tak $k = (n + 1)p$ a v případě, že $(n + 1)p$ je celé číslo, tak je modus určen nejednoznačně jako $(n + 1)p$ nebo $(n + 1)p - 1$. Celkem tedy platí vztah $(n + 1)p - 1 \leq \text{mod}(X) \leq (n + 1)p$.

Medián

Pro symetrické rozdělení, tj. $p = \frac{1}{2}$, a pro n sudé je medián roven střední hodnotě $\text{med}(X) = \frac{n}{2}$. Pro $p = \frac{1}{2}$ a n liché je medián nejednoznačně určen v intervalu $< \frac{n-1}{2}, \frac{n+1}{2} >$.

Rozptyl

K určení rozptylu si pomůžeme úpravou $E[X(X - 1)]$. Podobně jako u výpočtu střední hodnoty tedy určíme výraz $E[X(X - 1)] = \sum_{k=0}^n k(k-1) \binom{n}{k} p^k (1-p)^{n-k} = \sum_{k=2}^n \frac{n!}{(k-2)!(n-k)!} p^k (1-p)^{n-k} = n(n-1)p^2 \sum_{k=2}^n \frac{(n-2)!}{(k-2)!(n-k)!} p^{k-2} (1-p)^{n-k} = n(n-1)p^2 \sum_{j=0}^{n-2} \binom{n-2}{j} p^j (1-p)^{(n-2)-j} = n(n-1)p^2 [p + (1-p)]^{n-2} = n(n-1)p^2$.

Podle (6.5) je rozptyl binomického rozdělení

$$\begin{aligned} \text{var}(X) &= E[X(X - 1)] + E(X) - [E(X)]^2 = n(n-1)p^2 + np - (np)^2 = \\ &= n^2p^2 - np^2 + np - n^2p^2 = np(1-p). \end{aligned}$$

Šikmost

Pro snažší výpočet šikmosti si opět určíme výraz $E[X(X - 1)(X - 2)] = \sum_{k=0}^n k(k-1)(k-2) \binom{n}{k} p^k (1-p)^{n-k} = \sum_{k=3}^n \frac{n!}{(k-3)!(n-k)!} p^k (1-p)^{n-k} = n(n-1)(n-2) \sum_{k=3}^n \frac{(n-3)!}{(k-3)!(n-k)!} p^{k-3} (1-p)^{n-k} = n(n-1)(n-2) \sum_{j=0}^{n-3} \binom{n-3}{j} p^j (1-p)^{(n-3)-j} = n(n-1)(n-2)p^3 [p + (1-p)]^{n-3} = n(n-1)(n-2)p^3$.

Dle (6.7) jednoduše dopočítáme 3. centrální moment, tedy $\mu_3(X) = E[X(X - 1)(X - 2)] + 3\text{var}(X)[1 - E(X)] - [E(X)]^3 + 3[E(X)]^2 - 2E(X) = n(n-1)(n-2)p^3 + 3np(1-p)[1 - np] - (np)^3 + 3(np)^2 - 2np =$

$= n^3p^3 - 2n^2p^3 - n^2p^3 + 2np^3 + 3np - 3n^2p^2 - 3np^2 + 3n^2p^3 - n^3p^3 + 3n^2p^2 - 2np =$
 $= 2np^3 - 3np^2 + np = np(2p^2 - 3p + 1) = np(1 - 2p)(1 - p)$ a odtud opět dosazením
do (4.7) bude šikmost

$$\alpha_3(X) = \frac{np(1 - 2p)(1 - p)}{(\sqrt{np(1 - p)})^3} = \frac{1 - 2p}{\sqrt{np(1 - p)}}.$$

Funkce je klesající v p na celém svém definičním oboru, podobně jako u alternativního rozdělení. Funkce má maximum $+\infty$ v případě, že hodnotu p přibližujeme k 0 zprava. Pokud p přibližujeme zleva k 1, dostáváme minimum, $-\infty$. Pro $p > \frac{1}{2}$ je šikmost záporná, pro $p = \frac{1}{2}$ nulová a pro $p < \frac{1}{2}$ kladná. To byl případ, kdy máme pevné n . Naopak při pevném p a zvyšování faktoru n se bude funkční hodnota šikmosti blížit zprava ke svému minimu 0. Při snižování faktoru n se bude funkční hodnota blížit ke svému maximu $+\infty$.

Špičatost

Zde si opět určíme výraz $E[X(X - 1)(X - 2)(X - 3)]$, metoda výpočtu je stejná, tedy $E[X(X - 1)(X - 2)(X - 3)] = \sum_{k=0}^n k(k - 1)(k - 2)(k - 3) \binom{n}{k} p^k (1 - p)^{n-k} =$
 $= \sum_{k=4}^n \frac{n!}{(k-4)!(n-k)!} p^k (1 - p)^{n-k} = n(n - 1)(n - 2)(n - 3) \sum_{k=4}^n \frac{(n-4)!}{(k-4)!(n-k)!} p^{k-4} \cdot$
 $\cdot (1 - p)^{n-k} = n(n - 1)(n - 2)(n - 3) \sum_{j=0}^{n-4} \binom{n-4}{j} p^j (1 - p)^{(n-4)-j} =$
 $= n(n - 1)(n - 2)(n - 3)p^4 [p + (1 - p)]^{n-3} = n(n - 1)(n - 2)(n - 3)p^4.$

Dle (6.9) určíme 4. centrální moment $\mu_4(X) = E[X(X - 1)(X - 2)(X - 3)] +$
 $+ 6\mu_3(X) - 4\mu_3(X)E(X) - 6\text{var}(X)[E(X)]^2 + 18\text{var}(X)E(X) - 11\text{var}(X) - [E(X)]^4 +$
 $+ 6[E(X)]^3 - 11[E(X)]^2 + 6E(X) =$
 $= n(n - 1)(n - 2)(n - 3)p^4 + 12np^3 - 18np^2 - 6np - 8n^2p^4 + 12n^2p^3 - 4n^2p^2 -$
 $6np(1 - p)n^2p^2 + 18np(1 - p)np - 11np(1 - p) - n^4p^4 + 6n^3p^3 - 11n^2p^2 + 6np =$
 $= n^4p^4 - 5n^3p^4 + 6n^2p^4 - n^3p^4 + 5n^2p^4 - 6np^4 + 12np^3 - 18np^2 + 6np - 8n^2p^4 +$
 $12n^2p^3 - 4n^2p^2 - 6n^3p^3 + 6n^3p^4 + 18n^2p^2 - 18n^2p^3 - 11np + 11np^2 - n^4p^4 + 6n^3p^3 -$
 $11n^2p^2 + 6np =$
 $= 3n^2p^4 - 6np^4 - 6n^2p^3 + 12np^3 + 3n^2p^2 - 7np^2 + np = 3n^2p^2(p^2 - 2p + 1) +$

$+ np(-6p^3 - 12p^2 - 7p + 1) = 3n^2p^2(1-p)^2 + np(1-p)(6p^2 - 6p + 1) =$
 $= np(1-p)[3np(1-p) - 6p(1-p) + 1]$ a dosazením do (4.8) spočteme špičatost

$$\begin{aligned}\alpha_4(X) &= \frac{np(1-p)[3np(1-p) - 6p(1-p) + 1]}{[np(1-p)]^2} = \frac{1 - 6p(1-p)}{np(1-p)} + 3 = \\ &= \frac{1}{np(1-p)} + 3 - \frac{6}{n}.\end{aligned}$$

Speciálně pro $n = 1$ vzorec odpovídá špičatosti alternativního rozdělení. Nejprve jako pevný zvolíme parametr p a to konkrétně například jako $p = \frac{1}{2}$. Nyní se výsledný vzorec špičatosti binomického rozdělení zjednoduší na tvar $-\frac{2}{n} + 3$. Pokud nyní hodnotu faktoru n zvětšujeme do $+\infty$, blíží se funkční hodnota zleva ke svému maximu, hodnotě 3. V opačném případě, kdy n snižujeme směrem k 0 zprava, klesá funkční hodnota ke svému minimu $-\infty$. Funkce je při pevném p rostoucí. Nyní se budeme zabývat případem, kdy za pevný považujeme faktor n . V takovém případě špičatost nabývá svého minima pro $p = \frac{1}{2}$. Hodnota minima kolísá mezi hodnotami 1 a 3. Pokud zvolíme $n = 1$ bude mít minimum hodnotu 1. Naopak, budeme-li zvyšovat hodnotu faktoru n do $+\infty$, bude se hodnota minima blížit zleva k 3. Při pevném n bude funkce nabývat svého maxima, $+\infty$, pokud parametr p budeme přibližovat k okrajům definičního oboru funkce, tedy zleva k 0 a zprava k 1. Funkce je na intervalu $(0, \frac{1}{2})$ klesající, na intervalu $(\frac{1}{2}, 1)$ je naopak rostoucí.

6.3. Poissonovo rozdělení

Poissonovo rozdělení má jeden parametr, $\lambda > 0$. Označme $p_k = P(X = k) =$
 $= \frac{\lambda^k}{k!} e^{-\lambda}$.

Střední hodnota

Analogicky jako tomu bylo u binomického rozdělení, spočteme střední hodnotu jako součet hodnot přes k , kterých náhodná veličina X nabývá, násobených danou pravděpodobností, tedy

$$E(X) = \sum_{k=0}^{\infty} k \frac{\lambda^k}{k!} e^{-\lambda} = \lambda e^{-\lambda} \sum_{k=1}^{\infty} \frac{\lambda^{k-1}}{(k-1)!} = \lambda e^{-\lambda} \sum_{j=0}^{\infty} \frac{\lambda^j}{j!} = \lambda.$$

Modus

Opět budeme hledat maximum posloupnosti $\{p_k\}$. Je-li tato posloupnost rostoucí, tak

$$\begin{aligned} \frac{p_k}{p_{k-1}} &> 1 \\ \frac{\lambda^k}{k!} e^{-\lambda} &> \frac{\lambda^{k-1}}{(k-1)!} e^{-\lambda} \\ \frac{\lambda^k}{k} &> \lambda^{k-1} \\ \frac{\lambda}{k} &> 1 \\ \lambda &> k. \end{aligned}$$

Modus je tedy největší celé číslo k , pro které platí $k < \lambda$. Pokud $p_k = p_{k-1}$, tak $k = \lambda$ a je-li λ celé, modus je nejednoznačně určen jako λ nebo $\lambda - 1$. Platí tedy vztah $\lambda - 1 \leq \text{mod}(X) \leq \lambda$.

Rozptyl

Rozptyl spočteme s pomocí vzorce (6.5), nejprve tedy určíme $E[X(X-1)] = \sum_{k=0}^{\infty} k(k-1) \frac{\lambda^k}{k!} e^{-\lambda} = \lambda^2 e^{-\lambda} \sum_{k=2}^{\infty} \frac{\lambda^{k-2}}{(k-2)!} = \lambda^2 e^{-\lambda} \sum_{j=0}^{\infty} \frac{\lambda^j}{j!} = \lambda^2$ a po dosazení

$$\text{var}(X) = \lambda^2 + \lambda - \lambda^2 = \lambda.$$

Střední hodnota a rozptyl Poissonova rozdělení jsou totožné a rovnají se parametru λ .

Šikmost

K výpočtu šikmosti potřebujeme nejdříve určit výraz $E[X(X-1)(X-2)] = \sum_{k=0}^{\infty} k(k-1)(k-2) \frac{\lambda^k}{k!} e^{-\lambda} = \lambda^3 e^{-\lambda} \sum_{k=3}^{\infty} \frac{\lambda^{k-3}}{(k-3)!} = \lambda^3 e^{-\lambda} \sum_{j=0}^{\infty} \frac{\lambda^j}{j!} = \lambda^3$. Dosazením do (6.7) dopočítáme 3. centrální moment $\mu_3(X) = \lambda^3 + 3\lambda(1-\lambda) - \lambda^3 +$

$+ 3\lambda^2 - 2\lambda = \lambda$ a následně dle (4.7) celou šikmost

$$\alpha_3(X) = \frac{\lambda}{\sqrt{(\lambda^3)}} = \frac{1}{\sqrt{\lambda}}.$$

Jelikož $\lambda > 0$, tak šikmost Poissonova rozdělení bude vždy kladná a klesající v λ . Funkce dosahuje svého maxima, $+\infty$, když parametr λ přibližujeme zprava k 0. Naopak minima, konkrétně hodnoty 0, dosahuje funkce při růstu parametru λ do $+\infty$.

Špičatost

K výpočtu špičatosti nejdříve určíme výraz $E[X(X-1)(X-2)(X-3)] = \sum_{k=0}^{\infty} k(k-1)(k-2)(k-3) \frac{\lambda^k}{k!} e^{-\lambda} = \lambda^4 e^{-\lambda} \sum_{k=4}^{\infty} \frac{\lambda^{k-4}}{(k-4)!} = \lambda^4 e^{-\lambda} \sum_{j=0}^{\infty} \frac{\lambda^j}{j!} = \lambda^4$.

Dle (6.9) spočteme 4. centrální moment $\mu_4(X) = \lambda^4 + 6\lambda - 4\lambda^2 - 6\lambda^3 + 18\lambda^2 - 11\lambda - \lambda^4 + 6\lambda^3 - 11\lambda^2 + 6\lambda = 3\lambda^2 + \lambda$, a následně dle (4.8) celou špičatost

$$\alpha_4(X) = \frac{\mu_4(X)}{(\sqrt{\text{var}(X)})^4} = \frac{3\lambda^2 + \lambda}{\lambda^2} = 3 + \frac{1}{\lambda} = 3 + [\alpha_3(X)]^2.$$

Funkce $\alpha_4(X)$ je opět klesající na celém svém definičním oboru. Maximum $+\infty$ funkce dosahuje v případě, že parametr λ přibližujeme zprava k 0, minimum v podobě hodnoty 3 má funkce, když hodnota parametru λ roste do $+\infty$.

6.4. Geometrické rozdělení

Geometrické rozdělení má pouze jeden parametr $p \in (0, 1)$. Uvedeme si pomocnou funkci, díky které se nám budou lépe počítat faktoriální momenty geometrického rozdělení

$$F(s) = \sum_{k=0}^{\infty} (qs)^k = \frac{1}{1-qs}, \quad |qs| < 1.$$

Postupným derivováním zvlášť pravé a levé strany dostáváme

$$F'(s) = \sum_{k=0}^{\infty} k \cdot q^k \cdot s^{k-1} = \frac{q}{(1-qs)^2},$$

$$F''(s) = \sum_{k=0}^{\infty} k(k-1) \cdot q^k \cdot s^{k-2} = \frac{2q^2}{(1-qs)^3},$$

$$F'''(s) = \sum_{k=0}^{\infty} k(k-1)(k-2) \cdot q^k \cdot s^{k-3} = \frac{6q^3}{(1-qs)^4},$$

⋮

$$F^{(n)}(s) = \sum_{k=0}^{\infty} k(k-1) \cdots (k-n+1) \cdot q^k \cdot s^{k-n} = \frac{n!q^n}{(1-qs)^{(n+1)}}$$

a pro $s = 1$ dostáváme :

$$F^{(n)}(1) = \sum_{k=0}^{\infty} k(k-1) \cdots (k-n+1) \cdot q^k = \frac{n!q^n}{(1-q)^{(n+1)}} = \frac{n!q^n}{p^{n+1}}.$$

Jelikož pro geometrické rozdělení platí $p_k = P(X = k) = p \cdot q^k$, kde $q = 1 - p$, tak

$$EX(X-1) \cdots (X-n+1) = p \cdot F^{(n)}(1) = p \frac{n!q^n}{p^{n+1}} = \frac{n!q^n}{p^n}. \quad (6.10)$$

Střední hodnota

Dle (6.10) a pro $n = 1$ je střední hodnota rovna

$$E(X) = \frac{q}{p} = \frac{1-p}{p}.$$

Modus

Pro libovolné p je posloupnost $p_k = p(1-p)^k$ maximální v bodě $k = 0$, tedy v situaci, kdy prvnímu úspěšnému pokusu nepředchází žádný neúspěšný pokus.

Proto modus tohoto rozdělení je roven 0.

Rozptyl

K výpočtu rozptylu uijeme úpravy (6.10) a výpočtu $E[X(X-1)] = \frac{2q^2}{p^2}$. Dosazením do (6.5) získáme rozptyl

$$\text{var}(X) = \frac{2q^2}{p^2} + \frac{q}{p} - \frac{q^2}{p^2} = \frac{q(q+p)}{p^2} = \frac{q(1-p+p)}{p^2} = \frac{q}{p^2} = \frac{1-p}{p^2}.$$

Šikmost

Nejprve určíme $E[X(X-1)(X-2)] = \frac{6q^3}{p^3}$. Dosazením do (6.7) získáme 3. centrální moment $\mu_3(X) = \frac{6q^3}{p^3} + \frac{3q}{p^2}(1-\frac{q}{p}) - \frac{q^3}{p^3} + \frac{3q^2}{p^2} - \frac{2q}{p}$, po složitějších algebraických operacích se dostaneme k výrazu $\mu_3(X) = \frac{q}{p^3}(q+1)$. Dle vzorce (4.7) bude šikmost

$$\alpha_3(X) = \frac{q}{p^3}(q+1) \frac{p^3}{\sqrt{q^3}} = \frac{q+1}{\sqrt{q}} = \frac{2-p}{\sqrt{1-p}}.$$

Šikmost geometrického rozdělení je vždy kladná a rostoucí v p . Svého minima, hodnoty 2, dosahuje funkce, pokud hodnotu parametru p přibližujeme k 0 zprava. Naopak svého maxima, $+\infty$, funkce dosahuje, pokud parametr p přibližujeme k 1 zleva.

Špičatost

Nejprve si určíme $E[X(X-1)(X-2)(X-3)] = \frac{24q^4}{p^4}$. Dosazením do (6.9) získáme 4. centrální moment $\mu_4(X) = \frac{24q^4}{p^4} + \frac{6q(q+1)}{p^3} - \frac{4q(q+1)q}{p^3} - \frac{6q}{p^2} \cdot \frac{q^2}{p^2} + \frac{18q}{p^2} \cdot \frac{q}{p} - \frac{11q}{p^2} - \frac{q^4}{p^4} + \frac{6q^3}{p^3} - \frac{11q^2}{p^2} + \frac{6q}{p}$, po složitějších algebraických operacích se dostaneme k výrazu $\mu_4(X) = \frac{q}{p^4}(9q + p^2)$. Dle vzorce (4.8) a s využitím $[\alpha_3(X)]^2 = \left(\frac{2-p}{\sqrt{1-p}}\right)^2 = \left[\frac{2-(1-q)}{\sqrt{1-p}}\right]^2 = \frac{(1+q)^2}{1-p} = \frac{(1+q)^2}{q}$ bude vzorec pro špičatost vypadat

$$\begin{aligned}\alpha_4(X) &= \frac{q}{p^4}(9q + p^2) \frac{p^4}{q^2} = 9 + \frac{p^2}{q} = 9 + \frac{(1-q)^2}{q} = 9 + \frac{1-2q+q^2}{q} = \\ &= 5 + \frac{4q+1-2q+q^2}{q} = 5 + \frac{1+2q+q^2}{q} = 5 + \frac{(1+q)^2}{q} = 5 + [\alpha_3(X)]^2.\end{aligned}$$

Funkce $\alpha_4(X)$ je v parametru p rostoucí, tedy v parametru q klesající. Svého minima, hodnoty 9, funkce nabývá v případě, že hodnotu parametru p přibližujeme k 0 zprava. Naopak svého maximu, $+\infty$, funkce dosahuje pokud hodnotu parametru p přibližujeme k 1 zleva.

6.5. Rovnoměrné rozdělení

Rovnoměrné rozdělení bude jediné spojité rozdělení, u kterého si předvedeme výpočet teoretických charakteristik. U dalších spojitých rozdělení již značně stoupá náročnost a délka samotných výpočtů.

Střední hodnota

Střední hodnotu spojitě náhodné veličiny spočteme dle definice jako integrál ze

součinu hustoty a proměnné vyskytující se v hustotě

$$E(X) = \int_a^b x \frac{1}{b-a} dx = \frac{1}{b-a} \left[\frac{x^2}{2} \right]_a^b = \frac{1}{b-a} \frac{b^2 - a^2}{2} = \frac{1}{b-a} \frac{(b-a)(b+a)}{2} = \frac{b+a}{2}.$$

Modus

Ačkoliv je geometrické rozdělení symetrické, neplatí rovnost $\text{mod}(X) = \text{med}(X) = E(X)$, protože rozdělení není unimodální. Modus tedy volíme libovolně z intervalu (a, b) .

Medián

Medián určíme vyšetřením rovnice $\int_a^y \frac{1}{b-a} dx = \frac{1}{2}$, kde y je neznámá proměnná. Vypočteme integrál nalevo, tedy $\int_a^y \frac{1}{b-a} dx = \frac{1}{b-a} [x]_a^y = \frac{1}{b-a} (y - a)$ a položíme roven $\frac{1}{2}$, čímž dostáváme řešení

$$y = \text{med}(X) = \frac{a+b}{2}.$$

Rozptyl

Pro zjištění rozptylu bude potřeba pomocného kroku $E(X^2) = \int_a^b x^2 \frac{1}{b-a} dx = \frac{1}{b-a} \left[\frac{x^3}{3} \right]_a^b = \frac{1}{b-a} \frac{b^3 - a^3}{3} = \frac{1}{b-a} \frac{(b-a)(b^2 + ba + a^2)}{3} = \frac{b^2 + ba + a^2}{3}$. Dosazením do (6.4) jednoduše spočteme rozptyl

$$\text{var}(X) = \frac{b^2 + ba + a^2}{3} - \left(\frac{b+a}{2} \right)^2 = \frac{b^2 - 2ba + a^2}{12} = \frac{(b-a)^2}{12}.$$

Mezikvartilové rozpětí

Podobnými úpravami jako v případě výpočtu mediánu určíme 1. kvartil jako $\frac{3a+b}{4}$ a 3. kvartil jako $\frac{a+3b}{4}$. Odtud je již snadné dopočítat mezikvartilové rozpětí

$$R_Q = \frac{2b - 2a}{4}.$$

Šikmost

Pro výpočet šikmosti jsme používali vzorec (6.7), v tomto případě si připravíme

výraz $E(X^3) = \int_a^b x^3 \frac{1}{b-a} dx = \frac{1}{b-a} [\frac{x^4}{4}]_a^b = \frac{1}{b-a} (\frac{b^4-a^4}{4}) = \frac{1}{b-a} (\frac{(b^2+a^2)(b^2-a^2)}{4}) =$
 $= \frac{1}{b-a} (\frac{(b^2+a^2)(b-a)(b+a)}{4}) = \frac{b^3+ba^2+b^2a+a^3}{4}$. Nyní je již výpočet 3. centrálního mo-
mentu snadný, dosazením do (6.6) bude $\mu_3(X) = \frac{b^3+ba^2+b^2a+a^3}{4} - 3(\frac{b^2+ba+a^2}{3})(\frac{b+a}{2}) +$
 $+ 2(\frac{b+a}{2})^3$ a po několika základních algebraických úpravách dostaneme $\mu_3(X) = 0$,
tedy

$$\alpha_3(X) = \frac{0}{\frac{(b-a)^3}{12 \cdot 2\sqrt{3}}} = 0.$$

Šikmost rovnoměrného rozdělení je tedy za každé situace nulová. Rozdělení je zcela symetrické kolem středu.

Špičatost

Pro výpočet 4. centrálního momentu použijeme obdobný pomocný výpočet jako v případě šikmosti, tedy $E(X^4) = \int_a^b x^4 \frac{1}{b-a} dx = \frac{1}{b-a} [\frac{x^5}{5}]_a^b = \frac{1}{b-a} \cdot \frac{b^5-a^5}{5} =$
 $= \frac{1}{b-a} \cdot \frac{(b-a)(b^4+b^3a+b^2a^2+ba^3+a^4)}{5} = \frac{(b^4+b^3a+b^2a^2+ba^3+a^4)}{5}$. Dle (6.8) bude 4. centrální
moment $\mu_4(X) = \frac{(b^4+b^3a+b^2a^2+ba^3+a^4)}{5} - 4\frac{b^3+ba^2+b^2a+a^3}{4} \cdot \frac{b+a}{2} + 6\frac{b^2+ba+a^2}{3} \cdot (\frac{b+a}{2})^2 -$
 $- 4(\frac{b+a}{2})(\frac{b+a}{2})^3 + (\frac{b+a}{2})^4$. S využitím základních algebraických operací dojdeme k
výslednému tvaru v podobě $\mu_4(X) = \frac{(b-a)^4}{80}$. Špičatost rovnoměrného rozdělení je
konstantní, tj. nezávisí na parametrech, a je rovna

$$\alpha_4(X) = \frac{(b-a)^4}{80} \cdot \frac{144}{(b-a)^4} = \frac{9}{5}.$$

7. Empirické charakteristiky pro simulovaná data

V této kapitole bude naším úkolem zkoumat charakteristiky nasimulovaných dat. Budeme se zabývat zaprvé, porovnáváním odhadnutých charakteristik s těmi teoretickými, zadruhé, vlivem parametrů na tvar rozdělení, a zatřetí, vlivem poruch dat na charakteristiky. Data jsme generovali pro šest typů rozdělení, jednak pro pět, u kterých jsme určovali teoretické charakteristiky v kapitole 6, plus pro normální rozdělení, které má v teorii pravděpodobnosti velký význam. U diskrétních dat, konkrétně u binomického, Poissonova a geometrického rozdělení, jsme se snažili generovat data se stejnou (podobnou) střední hodnotou, proto jsou dále uvedeny tyto soubory mající střední hodnoty 1, 3, resp. 6. Pro generování dat jsme použili software R (<http://www.r-project.org/>). Nasimulovaná data jsou uvedena v příloze 9.1.

Pro každý typ rozdělení jsme počítali jednak teoretické hodnoty charakteristik (v tabulkách označené jako TH) podle vzorců uvedených v kapitole 6, a to vždy pro tři různé hodnoty parametrů. Dále jsme pro každý typ rozdělení a pro každou hodnotu parametrů generovali data počítačově, vždy o rozsahu $n = 50$. Pro získaná data jsme spočítali příslušné empirické (výběrové) charakteristiky dle vzorců z kapitoly 5:

$$\bar{x} \quad \dots \quad \text{výběrová střední hodnota} = \text{aritmetický průměr} \quad (5.1)$$

$$\hat{x} \quad \dots \quad \text{výběrový modus} \quad (5.2)$$

$$\tilde{x} \quad \dots \quad \text{výběrový medián} \quad (5.3), (5.4)$$

$$s^2 \quad \dots \quad \text{výběrový rozptyl} \quad (5.5)$$

$$d \quad \dots \quad \text{střední absolutní odchylka} \quad (5.7)$$

$$\hat{R} \quad \dots \quad \text{výběrové rozpětí} \quad (5.8)$$

$$\hat{R}_Q \quad \dots \quad \text{výběrové mezikvartilové rozpětí} \quad (5.9)$$

$$\hat{\alpha}_3 \quad \dots \quad \text{výběrová šikmost} \quad (5.10)$$

$$\hat{\alpha}_4 \quad \dots \quad \text{výběrová špičatost} \quad (5.12)$$

Stejné charakteristiky jsme počítali vždy i pro data různým způsobem upravená (porušená), přičemž jsme uvažovali následující varianty:

TH	...	teoretická hodnota rozdělení
ZV	...	základní generovaný výběr bez úprav,
1 OH	...	soubor s jednou odlehlou hodnotou,
2 OH	...	soubor s dvěma odlehlými hodnotami,
+1,5	...	soubor se všemi hodnotami zvětšenými o +1,5,
· 1,5	...	soubor se všemi hodnotami násobenými 1,5,
- 6 H	...	soubor zmenšený o 6 extrémních hodnot,
± 5%	...	soubor s porušenými daty o ±5%.

Kombinací charakteristik a datových sad vznikla pro každé rozdělení a pro každou hodnotu parametrů dvourozměrná tabulka. V každé tabulce jsme nejprve napočítali teoretické hodnoty daného rozdělení (TH) a dále charakteristiky pro základní (neupravený) výběr (ZV). Nepočítali jsme teoretickou hodnotu pro střední absolutní odchylku d . Standardně jsme zaokrouhlovali na 3 desetinná místa. Pokud jsou v tabulce za desetinnou čárkou uvedeny méně než 3 číslice, znamená to, že na zbylá vynechaná místa připadají nuly. Dále jsme počítali charakteristiky pro výběr s jednou odlehlou hodnotou (1 OH). Tato hodnota byla zvolena na levém konci číselné osy. Druhá odlehlá hodnota byla přidána na opačném konci osy (2 OH). Volba odlehlých hodnot probíhala subjektivně v závislosti na tom, jakých chyb či překlepů by se mohl přepisovatel dopustit při ručním zadávání dat do počítačové podoby. Pro každou trojici výběrů z jednoho typu rozdělení jsme zvolili stejné odlehlé hodnoty pro lepší porovnání jejich vlivu. Uvedeme si příklad u alternativního rozdělení, které je tvořeno sériemi 0 či 1. Zde se můžeme dopustit překlepu například v podobě zadání 10 či 11. Další modifikací je přičtení konstanty 1,5 ke každé nasimulované hodnotě (+ 1,5). Dále vynásobením konstantou 1,5 ($\cdot 1,5$). Předposlední úpravou bylo odstranění šesti extrémních hodnot, konkrétně po třech nejodlehlejších hodnotách na obou koncích datového souboru (- 6H). Této metodě potlačení extrémů se říká *winsorizace*. Při poslední modifikaci porušíme data o ±5%. Konkrétně bude tato metoda pracovat takto: původní hodnoty násobíme konstantou, kterou volíme náhodně buď 1,05 nebo 0,95 (±5%).

Ke grafické interpretaci byl u diskrétních hodnot použit graf rozdělení četností, u spojitých rozdělení histogram. Simulovaná data, doplněné odlehlé hodnoty a

příkazy k jejich generování jsou uvedeny v příloze. Dílčí postřehy a komentáře jsou uvedeny ihned za příslušným rozdělením, celkové závěry platné pro všechna rozdělení stejně jsou shrnuty v závěru kapitoly. Tyto celkové závěry se budou týkat zejména vlivu datových poruch na výběrové charakteristiky.

7.1. Alternativní rozdělení

Výsledky pro alternativní rozdělení jsou uvedeny v tabulkách 1-3. Tyto tři případy se mezi sebou liší volbou parametru p ($p = \frac{1}{4}, \frac{1}{2}, \frac{3}{4}$). U symetrického rozdělení (tabulka 2) není jednoznačně určen teoretický modus ani medián, jednoznačně také není určeno mezikvartilové rozpětí u nesymetrického rozdělení (tabulky 1 a 3). V první řadě zkusíme posoudit, jak přesně software vygeneroval hodnoty, tj. shodu teoretických charakteristik a charakteristik ze základního výběru (sloupce SH a ZV v tabulkách). U alternativního rozdělení, na rozdíl od dalších rozdělení, stačí k posouzení kvality vzorku pouze pohled na výběrovou střední hodnotu. Známe-li totiž počet hodnot a počet nul (nebo počet jedniček), pak známe všechny hodnoty ve výběru. Známe-li tedy rozsah výběru a aritmetický průměr, jsme schopni plně zrekonstruovat původní data. U ostatních rozdělení toto neplatí. Výběrová střední hodnota se spočte jako počet jedniček dělených počtem pozorování. Až na případ v tabulce 3 ($p = 0,75$) vidíme vcelku podařené simulované výběry. V této tabulce je u výběrové střední hodnoty největší rozdíl mezi TH a ZV, a sice 0,09, což je téměř 10% celého rozpětí. Velký rozdíl je i v porovnání TH a ZV pro šikmosti a špičatosti. Poslední možnou nesrovnalostí, kterou bychom se v tomto případě mohli zabývat, je špičatost 0,98 v tabulce 2. Teoretická špičatost, tedy hodnota ve sloupci TH, se skutečně rovná 1, my jsme ovšem pro výpočet ZV použili vzorec (5.12), kde jsme celý výraz dělili $n - 1$. Při dělení n by výběrová špičatost vyšla 1.

Na obrázku 5 vidíme, že rozdělení má tvar podle toho, kolik nul a jedniček jsme ze softwaru získali. Vidíme, že data z tabulky 1 jsou kladně zešikmena, protože větší seskupení dat je na levé straně. Z tabulky 3 jsou záporně zešikmena, protože naopak máme k dispozici více jedniček než nul. Z tabulky 2 vidíme, že

Tabulka 1. $\mathbf{X} \sim \text{Alt}(0,25)$, OH: -10, 11

Char.	TH	ZV	1 OH	2 OH	+ 1,5	· 1,5	- 6 H	±5%
\bar{x}	0,25	0,26	0,059	0,269	1,76	0,39	0,227	0,261
\hat{x}	0	0	0	0	1,5	0	0	0
\tilde{x}	0	0	0	0	1,5	0	0	0
s^2	0,188	0,196	2,257	4,514	0,196	0,442	0,18	0,198
d	-	0,26	0,451	0,654	0,26	0,39	0,227	0,261
\hat{R}	1	1	11	21	1	1,5	1	1,05
\hat{R}_Q	< 0; 1 >	0,5	0,5	1	0,5	0,75	0	0,475
$\hat{\alpha}_3$	1,155	1,083	-5,941	0,321	1,083	1,083	1,287	1,094
$\hat{\alpha}_4$	2,333	2,154	40,252	23,462	2,154	2,154	2,633	2,194

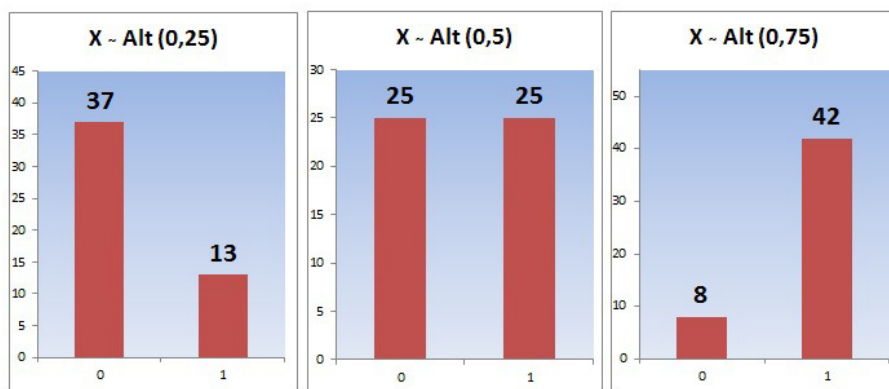
Tabulka 2. $\mathbf{X} \sim \text{Alt}(0,5)$, OH: -10, 11

Char.	TH	ZV	1 OH	2 OH	+ 1,5	· 1,5	- 6 H	±5%
\bar{x}	0,5	0,5	0,294	0,5	2	0,75	0,5	0,497
\hat{x}	{0; 1}	{0;1}	{0;1}	{0;1}	{1,5;2,5}	{0;1,5}	{0;1}	0
\tilde{x}	< 0; 1 >	0,5	0	0,5	2	0,75	0,5	0,475
s^2	0,25	0,255	2,412	4,569	0,255	0,574	0,256	0,253
d	-	0,5	0,686	0,885	0,5	0,75	0,5	0,497
\hat{R}	1	1	11	21	1	1,5	1	1,05
\hat{R}_Q	1	1	1	1	1	1,5	1	0,95
$\hat{\alpha}_3$	0	0	-5,781	0	0	0	0	0,015
$\hat{\alpha}_4$	1	0,98	38,634	22,84	0,98	0,98	0,977	1

Tabulka 3. $\mathbf{X} \sim \text{Alt}(0,75)$, OH: -10, 11

Char.	TH	ZV	1 OH	2 OH	+ 1,5	· 1,5	- 6 H	±5%
\bar{x}	0,75	0,84	0,628	0,827	2,34	1,26	0,886	0,838
\hat{x}	1	1	1	1	2,5	1,5	1	0,95
\tilde{x}	1	1	1	1	2,5	1,5	1	0,95
s^2	0,188	0,137	2,438	4,46	0,137	0,309	0,103	0,139
d	-	0,16	0,373	0,558	0,16	0,24	0,114	0,192
\hat{R}	1	1	11	21	1	1,5	1	1,05
\hat{R}_Q	< 0; 1 >	0	0	0	0	0	0	0,1
$\hat{\alpha}_3$	-1,155	-1,836	-6,304	-0,459	-1,836	-1,836	-2,407	-1,774
$\hat{\alpha}_4$	2,333	4,352	42,914	24,11	4,352	4,352	6,771	4,236

rozdělení je symetrické, jelikož máme k dispozici stejný počet nul i jedniček. Pokud je parametr $p < 0,5$ je rozdělení kladně zešikmené s větší koncentrací prvků na levé straně, tedy nul, naopak pokud je $p > 0,5$ je rozdělení záporně zešikmené, s větším počtem prvků na pravé straně, tedy jedniček.



Obrázek 5: Rozdělení četností u alternativního rozdělení

Nyní se podíváme na vliv jednotlivých poruch na výběrové charakteristiky. Začneme u odlehlých hodnot. Z charakteristik polohy mají odlehlé hodnoty nejmenší vliv na modus a medián. Tyto charakteristiky se dokonce nijak nemění. Pouze v tabulce 2, kdy $p = 0,5$, vidíme při 1 OH změnu mediánu, který se rovná 0. Je to dáno tím, že máme k dispozici 25 nul a 25 jedniček a po přidání -10 se změnil indexování prvků, kdy ten prostřední, tedy $x_{(26)}$, je 0. Největší vliv mají odlehlé hodnoty na výběrovou střední hodnotu. Odlehlé hodnoty mají vliv i na charakteristiky variability. Ten největší mají na rozpětí a rozptyl. Obě tyto charakteristiky po postupném přidávání odlehlých hodnot zvyšují svoji velikost. Jelikož jsme tyto hodnoty zvolili jako -10 resp. 11, představují to velmi velký zásah do výběru z alternativního rozdělení. Nejzajímavější je pro nás změna výběrového mezikvartilového rozpětí v tabulce 1. Jak je možné, že v tomto rozdělení při nabývaných hodnotách pouze 0 a 1 vyjde tato charakteristika jako neceločíselná, konkrétně 0,5? Vše vychází z definice kvantilů. Uvažujeme seřazený základní výběr. Protože máme sudý počet prvků ve výběru, v našem případě 50, spočítá

se třetí kvartil jako $50 \cdot 0,75 = 37,5$ tedy průměr prvků na pozicích 37 a 38. V tomto případě $x_{(37)}=0$ a $x_{(38)}=1$, tedy průměr je 0,5. Obecně, pokud $(n - 1)$ není dělitelné čtyřmi, bereme 3. kvartil jako aritmetický průměr dvou příslušných sousedních hodnot s pozicemi $1 + \lfloor \frac{3}{4}(n - 1) \rfloor$ a $1 + \lceil \frac{3}{4}(n - 1) \rceil$. Analogicky se bude 1. kvartil počítat jako $1 + \lfloor \frac{1}{4}(n - 1) \rfloor$ a $1 + \lceil \frac{1}{4}(n - 1) \rceil$. První kvartil počítaný pomocí tohoto postupu vychází 0. Mezikvartilové rozpětí je tedy $0,5 - 0 = 0,5$. Po přidání 1 OH se mezikvartilové rozpětí nemění. Změna přichází až po přidání 2 OH, kdy je $\hat{R}_Q = 1$, protože máme datový soubor o 52 prvcích s -10 na první pozici a 11 na poslední pozici. 3. kvartil je průměr hodnot na pozicích 39 a 40, tedy 1, 1. kvartil je průměr hodnot na pozicích 13 a 14, tedy 0. Výsledek je $1 - 0 = 1$. Střední absolutní odchylka roste po přidání 1 OH i 2 OH. Odlehlé hodnoty mají velký vliv také na šikmost a špičatost alternativního rozdělení. Po přidání první odlehlé hodnoty šikmost klesá, špičatost naopak ještě výraznějším způsobem stoupá. Druhá odlehlé hodnota má spíše korigující vliv na obě tyto charakteristiky, které se přibližují ke své původní hodnotě ZV. Více se ke své původní hodnotě blíží šikmost.

Při +1,5 modifikaci se hodnota všech charakteristik polohy zvětší právě o tuto konstantu, vše vychází z předpokladů zveřejněných v kapitole 4.1. Charakteristiky variability, šikmost i špičatost si zachovávají svoje hodnoty. Změna polohy nemá na tyto charakteristiky vliv, tak jak jsme popsali v kapitole 4.2. Při $\cdot 1,5$ modifikace se všechny charakteristiky polohy o 50% zvětší. Stejného navýšení se dočkáme i v případě rozpětí a mezikvartilového rozpětí. Rozptyl a střední absolutní odchylka se zvýší přesně o 2,25, tedy druhou mocninu čísla 1,5. Šikmost se špičatostí si opět ponechávají své hodnoty. Vypuštění šesti hodnot v případě alternativního rozdělení má největší vliv na šikmost a špičatost. Šikmost pokud v ZV vyšla kladná, mírně roste, pokud vyšla v ZV záporná, mírně klesá. Špičatost svoji hodnotu také zvětšuje. Pouze v symetrickém případě (tabulka 2) hodnota mírně klesla. Naopak nejmenší vliv má tato porucha na modus, medián, rozpětí a mezikvartilové rozpětí. Rozptyl až na symetrický případ klesá. Střední hodnota při vypuštění šesti hodnot roste (v případě záporného zešikmení) a klesá (v

případě kladného zešikmení). Porucha $\pm 5\%$ má podobný vliv na všechny charakteristiky, které se od svých původních hodnot v ZV nepatrně liší, ať už kladně, či záporně.

7.2. Binomické rozdělení

Výsledky pro binomické rozdělení jsou uvedeny v tabulkách 4-6. Tyto případy se mezi sebou liší volbou parametrů n a p . Nejprve si připomeňme postup při výpočtu teoretických hodnot modu a mediánu. Použijeme pravděpodobnostní funkci $p_k = P(X = k) = \binom{n}{k} p^k (1-p)^{n-k}$. Do ní budeme dosazovat hodnoty n, p a k . Modus určíme jako to k , pro něž je p_k maximální. Při určování mediánu budeme hodnoty p_k sčítat, až součet přesáhne 0,5. Za medián budeme považovat to číslo k , pro něž součet $p_0 + \dots + p_k$ předcházel přesáhnutí 0,5. Analogickým způsobem budeme postupovat i v případě výpočtu \hat{R}_Q , kde potřebujeme 0,25 a 0,75 kvantily. Obecně platí, že teoretický α -kvantil je takové největší přirozené číslo k^* , pro které platí $\sum_0^{k^*} p_k \leq \alpha$. Konkrétně si uvedeme výpočet modu, mediánu, 1. a 3.kvartilu pro Bi(8;0,75). Hodnoty pravděpodobnostní funkce p_k shrnuje následující tabulka:

k	0	1	2	3	4	5	6	7	8
p_k	0,0000	0,0004	0,0038	0,0231	0,0865	0,2076	0,3115	0,267	0,1001
$\sum_0^k p_k$	0,0000	0,0004	0,0042	0,0273	0,1138	0,3214	0,6329	0,8999	1

Z této tabulky získáme modus $\hat{x}=6$, medián $\tilde{x}=5$, 1. kvartil $x_{0,25}=4$ a 3. kvartil $x_{0,75}=6$. Mezikvartilové rozpětí $\hat{R}_Q=2$.

Jak již bylo řečeno u komentářů k alternativnímu rozdělení, jen střední hodnota k posouzení kvality výběru u dalších rozdělení stačit nebude. Nyní se budeme při porovnávání TH a ZV zaměřovat jednak na \bar{x} , tak na s^2 , $\hat{\alpha}_3$ a $\hat{\alpha}_4$, protože samotná výběrová střední hodnota k určení kvality výběru nestačí. V případě binomického rozdělení máme k dispozici více než pouhé 2 varianty znaku a proto musíme brát v potaz i další charakteristiky. Žádný z výběrů z tabulek 4-6 se nedá považovat za naprosto ideální, nejbliže k němu má ovšem výběr z tabulky 4. Zajímavý je pohled na rozdíl mezi TH a ZV u rozpětí. Dle definice je teoretickým

Tabulka 4. $\mathbf{X} \sim \mathbf{Bi}(4;0,25)$, OH: -4, 12

Char.	TH	ZV	1 OH	2 OH	+ 1,5	· 1,5	- 6 H	±5%
\bar{x}	1	0,98	0,882	1,096	2,48	1,47	0,909	0,987
\hat{x}	1	0	0	0	1,5	0	{0, 1}	0
\tilde{x}	1	1	1	1	2,5	1,5	1	0,95
s^2	0,75	0,877	1,346	3,697	0,877	1,974	0,643	0,902
d	-	0,74	0,824	1,019	0,74	1,11	0,636	0,759
\hat{R}	4	3	7	16	3	4,5	2	3,15
\hat{R}_Q	1	2	2	2	2	3	2	1,9
$\hat{\alpha}_3$	0,577	0,486	-1,078	3,224	0,486	0,486	0,161	0,521
$\hat{\alpha}_4$	2,833	2,134	7,274	21,36	2,134	2,134	1,572	2,207

Tabulka 5. $\mathbf{X} \sim \mathbf{Bi}(6;0,5)$, OH: -4, 12

Char.	TH	ZV	1 OH	2 OH	+ 1,5	· 1,5	- 6 H	±5%
\bar{x}	3	2,28	2,157	2,346	3,78	3,42	2,273	2,272
\hat{x}	3	2	2	2	3,5	3	2	2,1
\tilde{x}	3	2	2	2	3,5	3	2	2,1
s^2	1,5	0,777	1,535	3,368	0,777	1,749	0,482	0,74
d	-	0,64	0,745	0,923	0,64	0,96	0,5	0,626
\hat{R}	6	4	8	16	4	6	3	4,2
\hat{R}_Q	2	1	1	1	1	1,5	1	0,95
$\hat{\alpha}_3$	0	-0,032	-2,317	2,016	-0,032	-0,032	0,01	-0,134
$\hat{\alpha}_4$	2,667	2,872	12,945	17,966	2,872	2,872	2,637	3,013

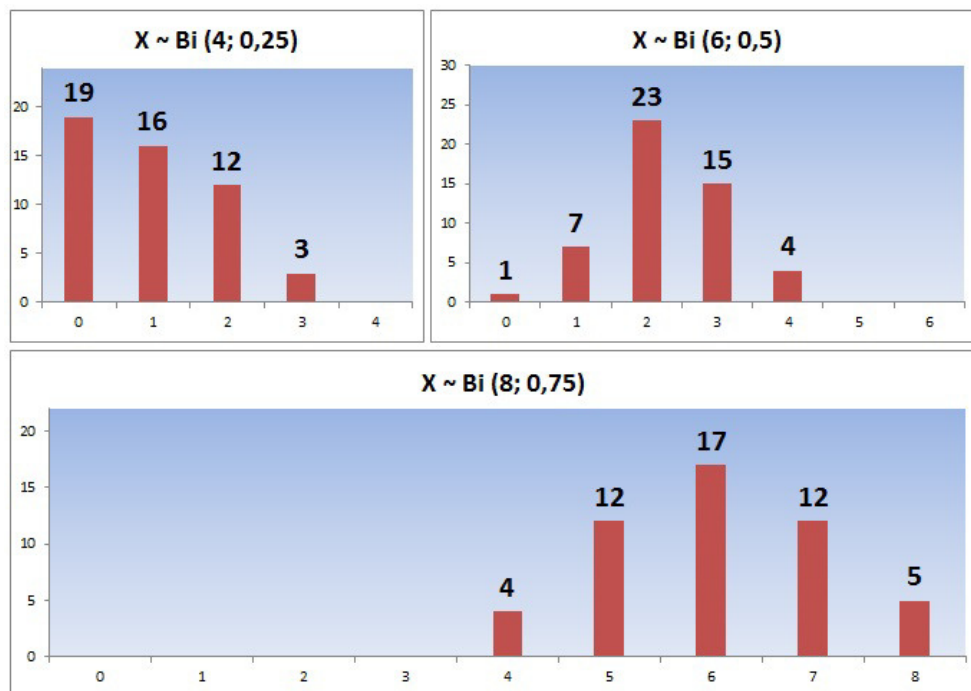
Tabulka 6. $\mathbf{X} \sim \mathbf{Bi}(8;0,75)$, OH: -4, 12

Char.	TH	ZV	1 OH	2 OH	+ 1,5	· 1,5	- 6 H	±5%
\bar{x}	6	6,04	5,843	5,961	7,54	9,06	6,046	6,038
\hat{x}	6	6	6	6	7,5	9	6	5,7
\tilde{x}	5	6	6	6	7,5	9	6	5,7
s^2	1,5	1,223	3,175	3,842	1,223	2,751	0,835	1,36
d	-	0,84	1,02	1,115	0,84	1,26	0,682	0,914
\hat{R}	8	4	12	16	4	6	4	4,2
\hat{R}_Q	2	2	2	2	2	3	2	1,4
$\hat{\alpha}_3$	-0,408	0,012	-3,242	-1,962	0,012	0,012	0,094	0,461
$\hat{\alpha}_4$	2,917	2,283	18,986	15,075	2,283	2,283	2,367	2,574

rozptětím hodnota n , u výběrů tomu ale tak není, záleží na velikosti rozptýlení hodnot kolem střední hodnoty. Pokrytí celého definičního oboru daty bychom se mohli dočkat pouze v případě, kdy by byl rozsah výběru velmi velký. V tabulce 5 se velmi podstatně liší rozptyl od své teoretické hodnoty a v tabulce 6 naopak moc přesně nevychází porovnání šikmosti a špičatosti.

Tvar rozdělení je opět určen rozmístěním dat. V případě, že zvětšujeme hodnotu parametru p , zvyšuje se koncentrace hodnot na pravé straně. Na obrázku 6 vidíme, že data pro $Bi(4;0,25)$ jsou kladně zešikmená, protože větší seskupení dat je na levé straně. Při pohledu na tvar dat z $Bi(6;0,5)$ a $Bi(8;0,75)$ by se mohlo zdát, že data jsou symetrická. Ovšem není tomu tak, protože v případě grafu $Bi(6;0,5)$ na obrázku 6 nejsou vůbec zastoupeny hodnoty 5,6 a v grafu $Bi(8;0,75)$ nejsou zastoupeny hodnoty 0,1,2 a 3.

Odlehle hodnoty mají zde malý vliv na charakteristiky polohy, oproti alter-



Obrázek 6: Rozdělení četností u binomického rozdělení

nativnímu rozdělení. Vysvětlení je takové, že odlehle hodnoty -4 a 12 nejsou pro $Bi(4;0,25)$, $Bi(6;0,5)$ a $Bi(8;0,75)$ tak moc odlehle jako -10 a 11 pro alternativní

rozdělení, které nabývá pouze hodnot 0 a 1. Naopak velký vliv mají na charakteristiky variability, hlavně na rozptyl a rozpětí. Střední absolutní odchylka se zvýší pouze nepatrně a mezikvartilové rozpětí se nezmění vůbec. Nárůst rozptylu s odhlednými hodnotami je o něco menší než u alternativního rozdělení. Opět je to dáno menší citlivostí binomického rozdělení na námi zvolené odlehle hodnoty. Největší vliv mají 1 OH a 2 OH na šikmost a špičatost. Modifikace + 1,5 a \cdot 1,5 jsou velice podobné těm z alternativního rozdělení. Naprosto stejně jako toto rozdělení se chovají charakteristiky polohy spolu se šikmostí a špičatostí. Poslední dvě modifikace, tedy - 6H a \pm 5%, nemají velký vliv na charakteristiky. Za zmínku pouze stojí mírné snížení hodnot šikmosti a špičatosti při vypuštění šesti odlehlejších hodnot.

7.3. Poissonovo rozdělení

Výsledky pro Poissonovo rozdělení jsou uvedeny v tabulkách 7-9. Tyto případy se mezi sebou liší volbou parametru λ . Nejprve se opět zastavíme u procesu generování dat a porovnání TH a ZV. Nejlépe svůj teoretický protějšek vystihují data z tabulky 9, tedy v případě, kdy máme Poissonovo rozdělení s parametrem 6. Tabulky 7 a 8 jsou nagenеровané méně přesně. Poměrně veliký rozdíl vidíme v případě špičatosti z tabulky 8. Veliký rozdíl mezi TH a ZV zasluhující upozornění je také, v té stejné tabulce, u modu, který se liší o 2. V tabulce 7 a sloupci ZV vidíme v řádku s hodnotou mediánu hodnotu 0,5. Je to opět neceločíselná hodnota ze souboru celočíselných dat. Postup výpočtu kvantilů, tedy i mediánu, je totožný jako v případě binomického rozdělení. 26. prvek seřazeného výběru je 1, 25. prvek je 0, odtud vidíme, že medián bude 0,5. Výběrové mezikvartilové rozpětí jsme počítali totožně.

Na obrázku 7 vidíme, že tvar rozdělení nejlépe odpovídá teorii hned v prvním případě. Již z definice víme, že parametr λ je vždy kladný, tedy šikmosti budou také kladné s větším naskupením hodnot na levé straně. Ve druhém a třetím případě vidíme, že máme k dispozici více dat nabývajících větších hodnot. Čím je větší λ , tím je menší šikmost. Tento jev vysvětlíme tak,

Tabulka 7. $\mathbf{X} \sim \mathbf{Po}(1)$, OH: -7, 18

Char.	TH	ZV	1 OH	2 OH	+ 1,5	· 1,5	- 6 H	±5%
\bar{x}	1	0,9	0,745	1,077	2,4	1,35	0,75	0,901
\hat{x}	0	0	0	0	1,5	0	0	0
\tilde{x}	1	0,5	0	0,5	2	0,75	0,5	0,475
s^2	1	1,439	2,634	8,308	1,439	3,237	0,843	1,424
d	-	0,9	1,02	1,346	0,9	1,35	0,75	0,901
\hat{R}	∞	5	12	25	5	7,5	3	4,75
\hat{R}_Q	1	1	1	1	1	1,5	1	1,05
$\hat{\alpha}_3$	1	1,469	-1,439	3,608	1,469	1,469	1,048	1,425
$\hat{\alpha}_4$	4	4,697	12	24,625	4,697	4,697	3,145	4,442

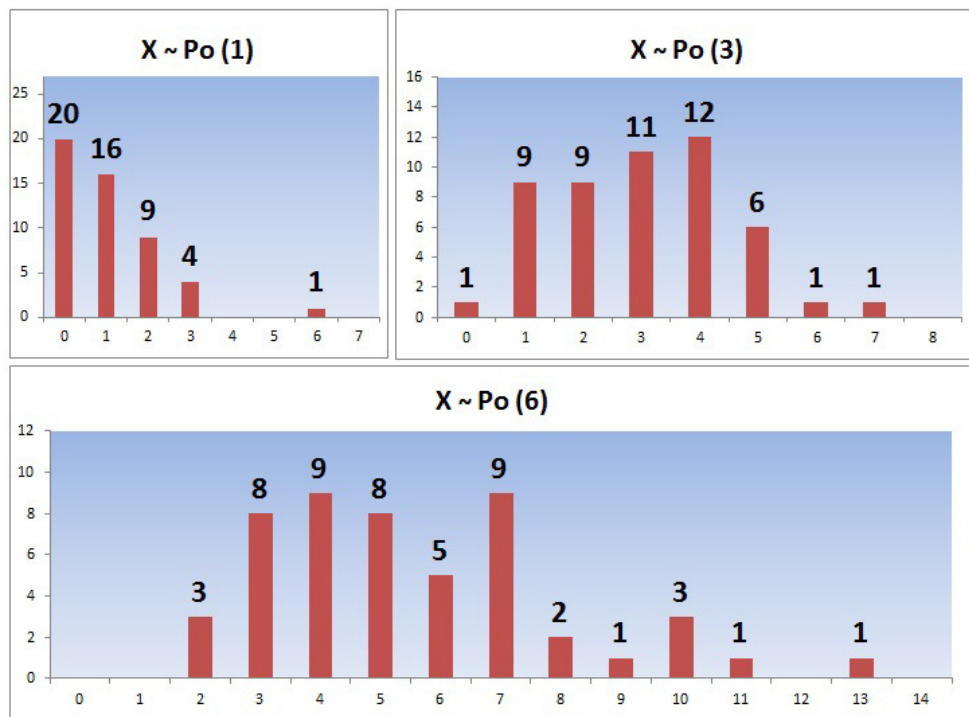
Tabulka 8. $\mathbf{X} \sim \mathbf{Po}(3)$, OH: -7, 18

Char.	TH	ZV	1 OH	2 OH	+ 1,5	· 1,5	- 6 H	±5%
\bar{x}	3	3,02	2,824	3,115	4,52	4,53	2,977	3,029
\hat{x}	2	4	4	4	5,5	6	4	4,2
\tilde{x}	3	3	3	3	4,5	4,5	3	3
s^2	3	2,347	4,268	8,614	2,347	5,28	1,604	2,412
d	-	1,22	1,392	1,654	1,22	1,83	1,023	1,271
\hat{R}	∞	7	14	25	7	10,5	4	7,35
\hat{R}_Q	2	2	2	2	2	3	2	2,3
$\hat{\alpha}_3$	0,577	0,205	-1,914	1,758	0,205	0,205	-0,095	0,238
$\hat{\alpha}_4$	3,333	2,505	11,026	15,915	2,505	2,505	1,916	2,636

Tabulka 9. $\mathbf{X} \sim \mathbf{Po}(6)$, OH: -7, 18

Char.	TH	ZV	1 OH	2 OH	+ 1,5	· 1,5	- 6 H	±5%
\bar{x}	6	5,56	5,314	5,558	7,06	8,34	5,409	5,52
\hat{x}	5	{4; 7}	{4; 7}	{4; 7}	{5, 5; 8, 5}	{6; 10, 5}	{4; 7}	(4,2;5,25)
\tilde{x}	6	5	5	5	6,5	7,5	5	5,25
s^2	6	6,17	9,14	12,055	6,17	13,882	3,689	5,667
d	-	1,92	2,118	2,327	1,92	2,88	1,546	1,83
\hat{R}	∞	11	20	25	11	16,5	7	10,45
\hat{R}_Q	4	3	3,5	3	3	4,5	3	2,85
$\hat{\alpha}_3$	0,408	0,835	-0,735	0,27	0,835	0,835	0,572	0,808
$\hat{\alpha}_4$	3,167	3,343	7,172	7,431	3,343	3,343	2,577	3,347

že při růstu parametru jsou hodnoty více rozmístěny po reálné ose, tedy tím pádem se šikmost zmenšuje. Již při pohledu vidíme, že špičatost je nejvyšší také v prvním případě.



Obrázek 7: Rozdělení četností u Poissonova rozdělení

Největší nárůst rozptylu, pokud uvažujeme druhou odlehlou hodnotu, vidíme v tabulce 7, naopak nejmenší v tabulce 9. Vysvětlení je takové, že extrémní hodnota 18 má větší vliv na data koncentrující se blízko nuly, tedy v tabulce 7, než na hodnoty koncentrující se více vpravo na číselné ose. Zůstaneme opět u rozptylu. Můžeme si všimnout, že je velmi velký nárůst této charakteristiky v tabulce 9 po vynásobení konstantou. Ovšem při detailnějším pohledu vidíme zvětšení rozptylu na zhruba dvojnásobek u ostatních dvou výběrů.

7.4. Geometrické rozdělení

Výsledky pro geometrické rozdělení jsou uvedeny v tabulkách 10-12. Tyto případy se mezi sebou liší volbou parametru p . Data ze všech třech tabulek jsou

Tabulka 10. $X \sim \text{Ge}(0,5)$, OH: -6, 100

Char.	TH	ZV	1 OH	2 OH	+ 1,5	· 1,5	- 6 H	±5%
\bar{x}	1	1,02	0,882	2,788	2,52	1,53	0,864	1,023
\hat{x}	0	0	0	0	1,5	0	0	0
\tilde{x}	-	1	1	1	2,5	1,5	1	0,95
s^2	2	1,653	2,586	191,195	1,653	3,719	0,911	1,712
d	-	0,94	1,059	3,942	0,94	1,41	0,773	0,947
\hat{R}	∞	6	12	106	6	9	3	6,3
\hat{R}_Q	0	2	2	2	2	3	1	1,9
$\hat{\alpha}_3$	2,121	1,577	-0,617	6,79	1,577	1,577	0,752	1,692
$\hat{\alpha}_4$	9,499	5,886	9,364	47,771	5,886	5,886	2,436	6,54

Tabulka 11. $X \sim \text{Ge}(0,25)$, OH: -6, 100

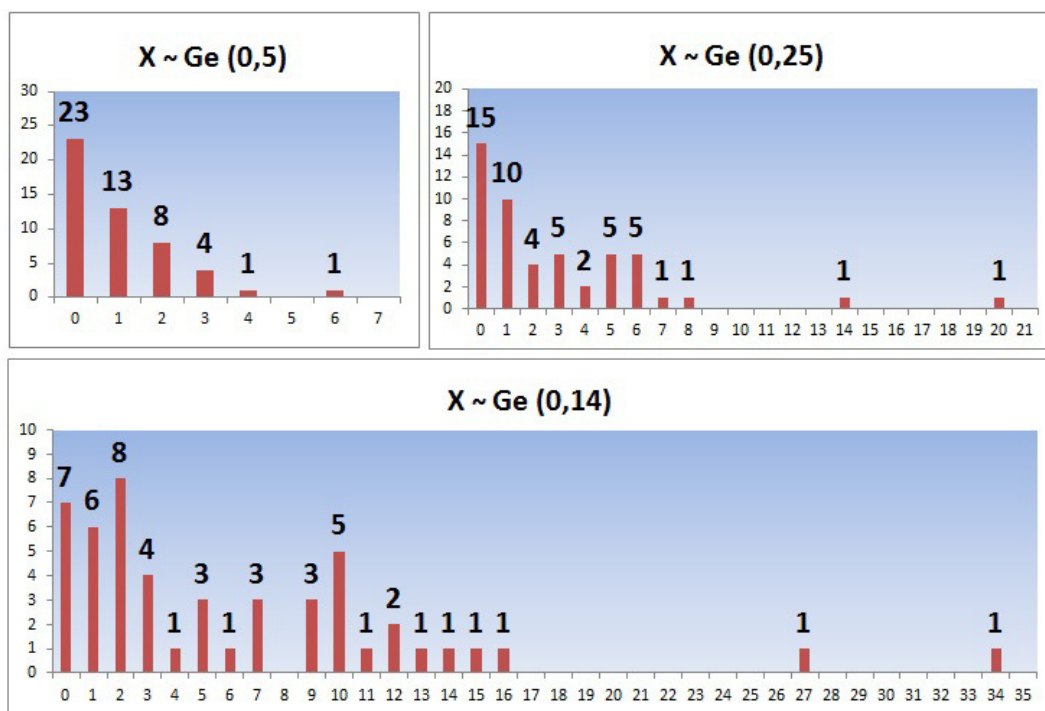
Char.	TH	ZV	1 OH	2 OH	+ 1,5	· 1,5	- 6 H	±5%
\bar{x}	3	2,9	2,726	4,596	4,4	4,35	2,341	2,911
\hat{x}	0	0	0	0	1,5	0	0	0
\tilde{x}	-	1,5	1	1,5	3	2,25	1,5	1,475
s^2	12	14,214	15,483	197,147	14,214	31,982	5,021	14,669
d	-	2,5	2,588	5,194	2,5	3,75	1,886	2,515
\hat{R}	∞	20	26	106	20	30	7	21
\hat{R}_Q	3	5	5	5	5	7,5	4	5
$\hat{\alpha}_3$	2,021	2,411	1,98	6,161	2,411	2,411	0,563	2,549
$\hat{\alpha}_4$	9,084	10,444	9,489	41,837	10,444	10,444	1,87	11,539

Tabulka 12. $X \sim \text{Ge}(0,14)$, OH: -6, 100

Char.	TH	ZV	1 OH	2 OH	+ 1,5	· 1,5	- 6 H	±5%
\bar{x}	6,143	6,22	5,98	7,788	7,72	9,33	5,318	6,215
\hat{x}	0	2	2	2	3,5	3	2	0
\tilde{x}	-	3,5	3	3,5	5	5,25	3,5	3,475
s^2	43,878	46,869	48,86	217,896	46,869	105,455	19,664	47,718
d	-	4,86	4,941	7,339	4,86	7,29	3,773	4,847
\hat{R}	∞	34	40	106	34	51	15	35,7
\hat{R}_Q	8	9	9	9	9	13,5	7	8,45
$\hat{\alpha}_3$	2,006	1,93	1,773	4,882	1,93	1,93	0,545	2,033
$\hat{\alpha}_4$	9,024	7,663	7,333	30,15	7,663	7,663	1,945	8,422

generována poměrně přesně. Ovšem i zde se najdou odchylky, za tu největší se dá považovat porovnání špičatosti u tabulky 10 či rozptyl u tabulky 11. V tabulce 12 vidíme rozdíl modu. Připomeňme, že podle definice je modulus geometrického rozdělení vždy 0, neboť právě v případě, kdy se $k = 0$, dosahuje posloupnost p_k svého maxima.

Na obrázku 8 lze vyčíst kladnou šikmost, tedy větší seskupení dat na levé straně. Této skutečnosti se dočkáme při libovolném nastavení parametru p .



Obrázek 8: Rozdělení četností u geometrického rozdělení

Očekávaný je pohled na velké zvýšení rozptylu po přidání druhé extrémní hodnoty, v tomto případě 100.

7.5. Rovnoměrné rozdělení

Výsledky pro rovnoměrné rozdělení jsou uvedeny v tabulkách 13-15. Tyto případy se mezi sebou liší volbou parametrů a, b . Data ze všech třech tabulek byla generována přesně, největší odchylky vidíme v porovnání špičatostí. Zajímavostí

Tabulka 13. $\mathbf{X} \sim \mathbf{Ro}(0,2)$, OH: -15,199, 22,16

Char.	TH	ZV	1 OH	2 OH	+ 1,5	· 1,5	- 6 H	±5%
\bar{x}	1	0,94	0,623	1,038	2,44	1,41	0,934	0,944
\hat{x}	< 0; 2 >	-	-	-	-	-	-	-
\tilde{x}	1	0,901	0,85	0,901	2,401	1,352	0,901	0,886
s^2	0,333	0,335	5,435	14,248	0,335	0,753	0,264	0,347
d	-	0,491	0,796	1,19	0,491	0,736	0,433	0,497
\hat{R}	2	1,937	17,138	37,359	1,937	2,906	1,652	2,034
\hat{R}_Q	1	0,979	1,007	1,024	0,979	1,469	0,842	0,952
$\hat{\alpha}_3$	0	0,14	-6,223	1,874	0,14	0,14	0,155	0,191
$\hat{\alpha}_4$	1,8	1,797	42,499	25,94	1,797	1,797	1,854	1,857

Tabulka 14. $\mathbf{X} \sim \mathbf{Ro}(0,6)$, OH: -15,199, 22,16

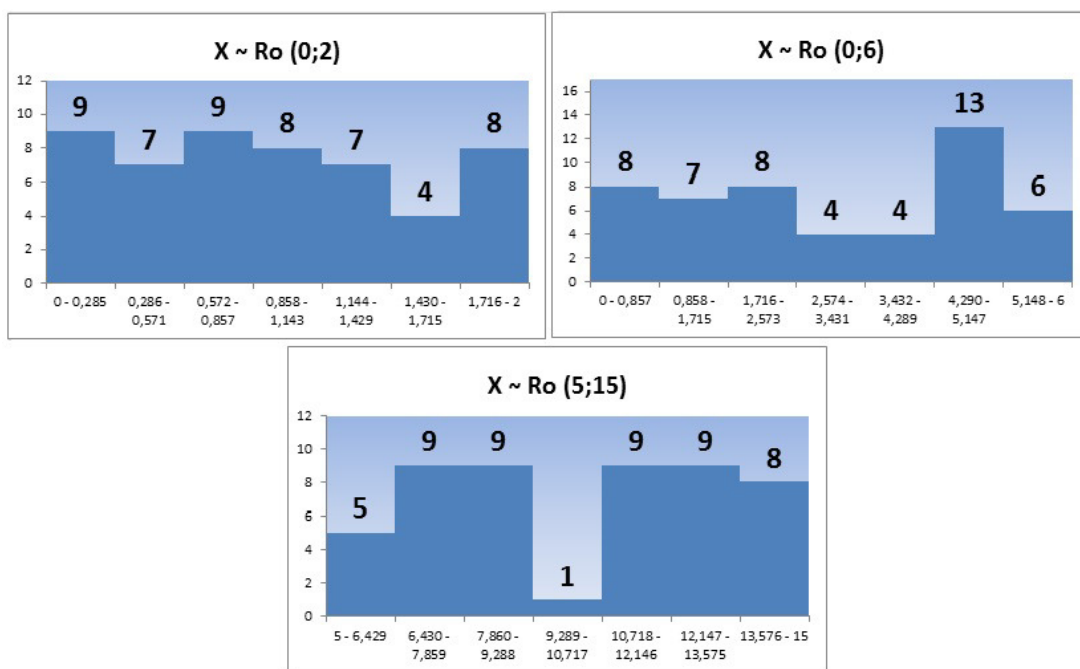
Char.	TH	ZV	1 OH	2 OH	+ 1,5	· 1,5	- 6 H	±5%
\bar{x}	3	3,018	2,661	3,036	4,518	4,527	3,026	3,03
\hat{x}	< 0; 6 >	-	-	-	-	-	-	-
\tilde{x}	3	2,977	2,936	2,977	4,477	4,465	2,977	2,864
s^2	3	3,615	10,05	17,165	3,615	8,134	2,948	3,682
d	-	1,695	2,018	2,349	1,695	2,543	1,532	1,721
\hat{R}	6	5,95	21,178	37,359	5,95	8,925	5,302	6,194
\hat{R}_Q	3	3,481	3,63	3,534	3,481	5,221	3,398	3,515
$\hat{\alpha}_3$	0	-0,088	-3,474	0,246	-0,088	-0,088	-0,11	-0,095
$\hat{\alpha}_4$	1,8	1,573	20,366	16,328	1,573	1,573	1,559	1,551

Tabulka 15. $\mathbf{X} \sim \mathbf{Ro}(5,15)$, OH: -15,199, 22,16

Char.	TH	ZV	1 OH	2 OH	+ 1,5	· 1,5	- 6 H	±5%
\bar{x}	10	10,274	9,774	10,013	11,774	15,411	10,279	10,27
\hat{x}	< 5; 15 >	-	-	-	-	-	-	-
\tilde{x}	10	10,965	10,95	10,965	12,465	16,447	10,965	10,715
s^2	8,333	8,79	21,337	23,868	8,79	19,777	7,063	8,985
d	-	2,641	3,102	3,258	2,641	3,962	2,375	2,704
\hat{R}	10	9,685	30,169	37,359	9,685	14,528	7,944	10,698
\hat{R}_Q	5	5,379	5,431	5,414	5,379	8,069	4,957	5,19
$\hat{\alpha}_3$	0	-0,025	-3,035	-2,342	-0,025	-0,025	-0,025	0,005
$\hat{\alpha}_4$	1,8	1,573	17,375	14,865	1,573	1,573	1,514	1,614

je, že v tabulkách 14 a 16 vyšly naprosto shodné špičatosti, konkrétně hodnota 1,573. Rozptýly i střední hodnoty vcelku kopírují své teoretické protějšky. Dále vidíme, že špičatost ze základního vzorku se nejvíce blíží své teoretické hodnotě hned v prvním výběru, tedy tabulce 13, kdy je rozdíl mezi těmito dvěma hodnotami nepatrný.

Histogramy z našich třech výběrů z rovnoměrného rozdělení vidíme na obrázku 9. Jejich tvar závisí na generovaných datech, ale také na zvoleném počtu

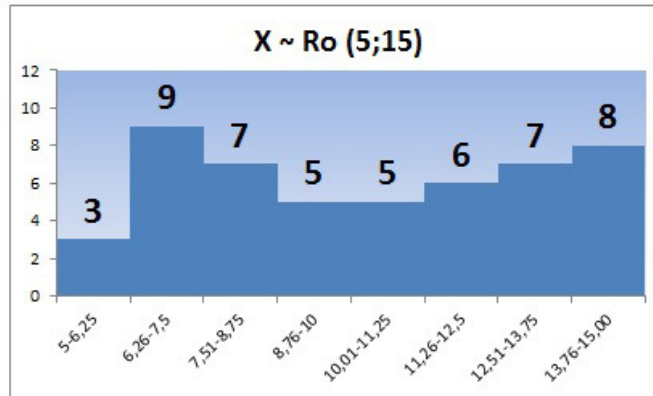


Obrázek 9: Histogramy z rovnoměrného rozdělení

buněk. Zde jsme použili *Sturgesovo pravidlo* pro optimální počet sloupců $1 + 3.3 \log_n$. Při $n=50$ vychází 6,607. Proto jsme u spojitých rozdělení uvažovali 7 buněk. Protože histogram pro $Ro(5;15)$ na obrázku 9 moc nepřipomíná rovnoměrné rozdělení, zkusili jsme přidat jednu buňku, na obrázku 10 jich je 8.

Vidíme, že po přidání jedné buňky histogram více vystihuje rovnoměrné rozdělení. *Sturgesovo pravidlo* tedy vždy nemusí podávat optimální doporučení, kolik sloupců bychom měli použít. Záleží ovšem také na vygenerovaných hodnotách.

Parametry rovnoměrného rozdělení mají největší vliv na rozpětí a jeho pro-



Obrázek 10: Upravený histogram z Ro(5;15)

střednictvím na další charakteristiky. S rostoucím rozdílem parametrů a, b se zvětšuje rozptyl. Zajímavostí je, že se rozptyl zvýší i při pětiprocentní poruše dat, ve všech třech případech. To u předešlých rozdělení pravidlem nebývalo.

7.6. Normální rozdělení

Výsledky pro normální rozdělení jsou uvedeny v tabulkách 16-18. Tyto případy se mezi sebou liší volbou parametrů μ, σ^2 . K normálnímu rozdělení jsme si praktické výpočty teoretických charakteristik neuváděli, v našem případě jsou převzaty z [1]. Postup při výpočtech mediánu a mezikvartilového rozpětí jsou analogické jako v případě rovnoměrného rozdělení v kapitole 6.

Na tvar rozdělení, tak jak jej vidíme na obrázku 11, mají vliv parametry μ, σ^2 . Nejvíce připomíná normální rozdělení histogram pro $N(3;1)$. Naopak histogram pro $N(0;4)$ normální rozdělení nepřipomíná. Ovšem při změně počtu sloupců bychom se mohli dočkat podobné změny v rozložení dat jako jsme toho dosáhli po přidání jednoho sloupce v předchozím případě u rovnoměrného rozdělení.

Zajímavostí je, že šikmost z dat pro $N(0;4)$ se po pětiprocentní poruše dat takřka nezměnila, pouze o 0,001. Podle předpokladů je při poruše - 6H nejvíce citlivé výběrové rozpětí. Pro data z $N(0;4)$ a vypuštění šesti nejextrémnějších hodnot se šikmost nejvíce blíží svojí teoretické hodnotě. Liší se pouze o 0,039.

Tabulka 16. $\mathbf{X} \sim \mathbf{N}(\mathbf{0};1)$, OH: -6,209, 7,702

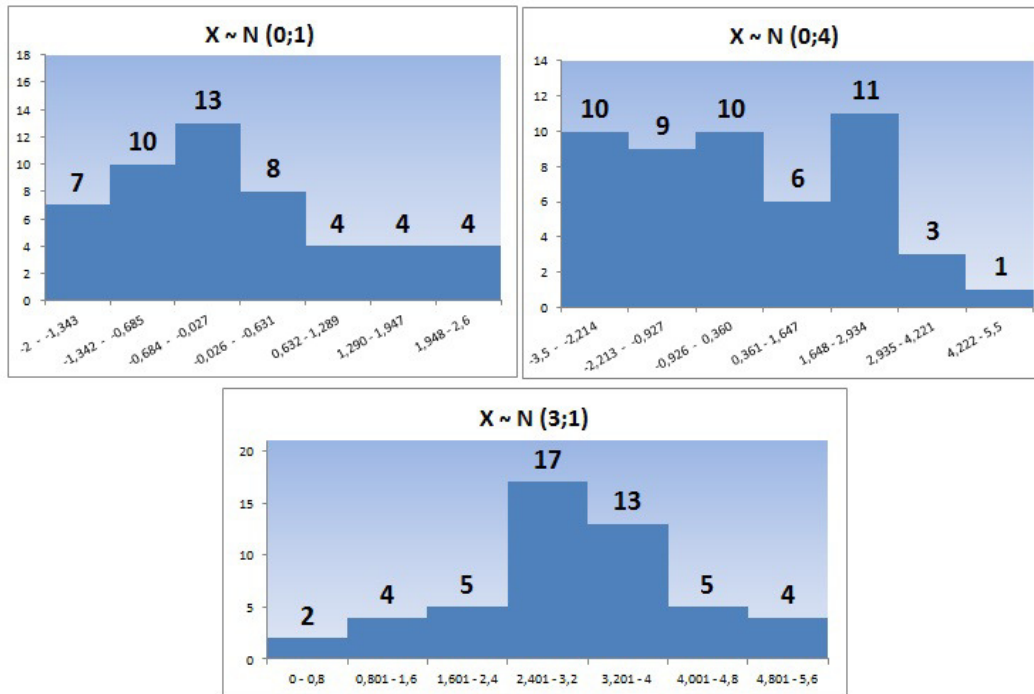
Char.	TH	ZV	1 OH	2 OH	+ 1,5	· 1,5	- 6 H	±5%
\bar{x}	0	-0,068	-0,189	-0,037	1,432	-0,102	-0,118	-0,077
\hat{x}	0	-	-	-	-	-	-	-
\tilde{x}	0	-0,24	-0,249	-0,24	1,26	-0,36	-0,24	-0,237
s^2	1	1,421	2,132	3,287	1,421	3,197	0,954	1,404
d	-	0,684	1,032	1,165	0,934	1,4	0,768	0,671
\hat{R}	∞	4,507	8,754	13,911	4,507	6,761	3,721	4,416
\hat{R}_Q	1,34	1,636	1,721	1,726	1,636	2,454	1,484	1,585
$\hat{\alpha}_3$	0	0,478	-0,985	0,86	0,478	0,478	0,377	0,464
$\hat{\alpha}_4$	3	2,435	6,952	9,569	2,435	2,435	2,36	2,43

Tabulka 17. $\mathbf{X} \sim \mathbf{N}(\mathbf{0};4)$, OH: -6,209, 7,702

Char.	TH	ZV	1 OH	2 OH	+ 1,5	· 1,5	- 6 H	±5%
\bar{x}	0	-0,098	-0,218	-0,066	1,402	-0,147	-0,184	-0,105
\hat{x}	0	-	-	-	-	-	-	-
\tilde{x}	0	-0,28	-0,283	-0,28	1,22	-0,42	-0,28	-0,28
s^2	4	4,552	5,193	6,298	4,552	10,242	3,067	4,706
d	-	1,756	1,838	1,956	1,756	2,634	1,478	1,772
\hat{R}	∞	8,933	11,653	13,911	8,933	13,399	6,461	9,379
\hat{R}_Q	2,68	3,31	3,392	3,462	3,31	4,965	2,731	3,22
$\hat{\alpha}_3$	0	0,35	0,053	0,473	0,35	0,35	0,039	0,351
$\hat{\alpha}_4$	3	2,436	2,863	3,717	2,436	2,436	1,889	2,569

Tabulka 18. $\mathbf{X} \sim \mathbf{N}(\mathbf{3};1)$, OH: -6,209, 7,702

Char.	TH	ZV	1 OH	2 OH	+ 1,5	· 1,5	- 6 H	±5%
\bar{x}	3	3,049	2,867	2,96	4,549	4,573	3,071	3,05
\hat{x}	3	-	-	-	-	-	-	-
\tilde{x}	3	3,121	3,095	3,121	4,622	4,682	3,122	3,064
s^2	1	1,29	2,944	3,336	1,29	2,902	0,702	1,356
d	-	0,832	0,998	1,068	0,832	1,248	0,629	0,861
\hat{R}	∞	5,365	11,793	13,911	5,365	8,047	3,786	5,633
\hat{R}_Q	1,34	1,063	1,175	1,161	1,063	1,595	0,896	1,116
$\hat{\alpha}_3$	0	-0,298	-2,907	-2,152	-0,299	-0,298	-0,286	-0,155
$\hat{\alpha}_4$	3	3,175	16,246	13,791	3,175	3,175	3,054	3,26



Obrázek 11: Histogramy z normálního rozdělení

7.7. Shrnutí vlivu poruch

Každá z tabulek 1-18 vykazuje podobné změny charakteristik při různě nastavených modifikacích dat. Na jednu stranu máme výběr ze šesti typů rozdělení, na stranu druhou máme šest různých modifikací dat (poruch). U každé výběrové charakteristiky budeme sledovat, jaký vliv na ni mají jednotlivé poruchy, který by se projevoval u všech rozdělení stejně.

Jako první si probereme charakteristiky polohy, tedy \bar{x} , \hat{x} , \tilde{x} . Výběrová střední hodnota \bar{x} se po přidání první odlehlé hodnoty na levém konci číselné osy mírně či více sníží, podle velikosti extrémní hodnoty, po přidání druhé odlehlé hodnoty na opačný konec se naopak výsledek více přiblíží původní hodnotě, protože vliv obou extrémních hodnot zároveň se lehce vyruší. Největší eliminace vlivu obou extrémních hodnot se projeví, pokud jsou obě extrémní hodnoty stejně vzdálené od střední hodnoty. Na modus \hat{x} a medián \tilde{x} extrémní hodnoty vliv nemají, pouze

v případě mediánu se může hodnota změnit, protože po přidání odlehlých hodnot se datový soubor o 1 resp. 2 hodnoty zvětší a tím pádem se změní indexování prvků a to může vyvolat změnu mediánu, pokud v seřazeném souboru jsou vedle sebe dva prvky různých hodnot. Změna však nebývá velká. Po přičtení +1.5 se tyto charakteristiky o přesně tuto hodnotu zvýší. Při násobení konstantou 1,5 se charakteristiky polohy zvětší o 50% na 1,5 násobek svojí původní hodnoty. Při vypuštění šesti extrémních hodnot a 5% poruše dat se střední hodnota nijak výrazně nezmění. Modus a medián se po vypuštění hodnot také nemění, při poruše dat vyjdou tyto charakteristiky o 5% více nebo méně než ty ze základního vzorku. Pokud uvažujeme pouze charakteristiky polohy, nejvíce se úpravy dat s odlehlými hodnotami projeví na střední hodnotu, nejméně na modus. Na střední hodnotu se nejméně projeví 5% porucha dat, tato porucha se naopak nejvíce projeví na medián.

Nyní se zastavíme u charakteristik variability, tedy s^2 , d , \hat{R} a \hat{R}_Q . První tři zástupci své výsledky po přidávání extrémních hodnot zvyšují, mezikvartilové rozpětí povětšinou zůstává stejné, protože 1 či 2 extrémní hodnoty nemají vliv na 1. resp. 3. kvartil. Po přičtení 1,5 jsou všechny 4 charakteristiky totožné jako u základního vzorku, protože na charakteristiky variability nemá posunutí v poloze vliv. V případě, že základní vzorek násobíme konstantou 1,5, se \hat{R} , \hat{R}_Q a d zvýší přesně o 50% na 1,5 násobek své původní hodnoty ze ZV. s^2 při násobení 1,5 navýší svoji velikost o 125%, tedy na 2,25 ($1,5^2$) násobek své původní hodnoty. Při odstranění 6ti nejdlejších pozorování se hodnoty s^2 a d sníží. Sníží se i \hat{R} a \hat{R}_Q , větší pokles z těchto dvou charakteristik vidíme u celkového rozpětí, protože mezikvartilové rozpětí je méně citlivé na extrémní hodnoty, protože počítáme s prvním, resp. třetím kvartilem. V případě charakteristik variability mají tedy extrémní hodnoty největší vliv na rozptyl a rozpětí. Porucha $\pm 5\%$ nemá velký vliv na změnu těchto charakteristik.

Na závěr si projdeme výsledky šikmosti a špičatosti. Odlehlé hodnoty mají naprosto zásadní vliv na šikmost. V případě přidání jedné odlehlé hodnoty, tedy na levou stranu číselné osy, se šikmost sníží. Po korekci, přidání druhé odlehlé

hodnoty na pravou stranu číselné osy se šikmost zvýší. U špičatosti vidíme, že ve většině případů se hned po přidání první odlehlé hodnoty hodnota této charakteristiky zvýší. Stejně tak v případě přidání druhé odlehlé hodnoty. V případě, že k hodnotám datového souboru přičítáme 1,5 či je násobíme touto hodnotou, hodnoty šikmosti a špičatost se nemění. Vysvětlení je nasnadě. Při pohledu na vzorce pro výpočet výběrové šikmosti a špičatosti vidíme, že čitatel se jmenovatelem se skládají z výrazů pro výpočet centrálních momentů, kde od sebe odečítáme hodnoty o 1,5 zvětšené a aritmetický průměr, který je ovšem také o 1,5 větší, tedy stejně jako bychom počítali šikmost a špičatost pro základní výběr. Násobení konstantou nemá vliv na tvar rozdělení. Podobně jako tomu bylo u charakteristik variability, se i zde po vypuštění 6 extrémních hodnot hodnoty šikmosti a špičatosti sníží v porovnání s těmi, jichž nabývaly u základního vzorku. Porucha dat o 5% nemá velký vliv na výsledné hodnoty a tvar šikmosti a špičatosti.

8. Empirické charakteristiky pro reálná data

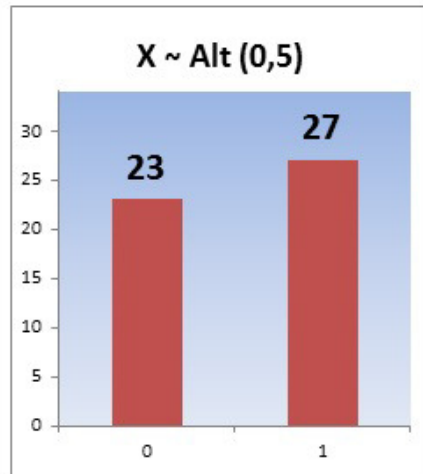
V této kapitole si předvedeme vypočtené hodnoty charakteristik pro skutečná data, která jsme získali sériemi pokusů či z externích zdrojů, počet dat je vždy $n = 50$. Zde budeme porovnávat především rozdíl mezi teoretickou hodnotou a tou ze základního vzorku pro zkoumání vhodnosti zvoleného reálného příkladu. Skutečná data následně pozměníme prostřednictvím šesti poruch (modifikací), které jsou stejné jako v kapitole 7. Výsledky jsou shrnuty v tabulkách, které jsou svojí strukturou totožné s tabulkami v kapitole 7. Data jsou uvedena v příloze 9.2.

8.1. Alternativní rozdělení

Pro reálný příklad alternativního rozdělení byl použit hod mincí. V případě, kdy padla hlava, byl výsledek pokusu označen jako 1, v opačném případě, když padl orel, byl označen jako 0. Pravděpodobnost nastoupení jevu byla tedy $p = 0,5$, symbolicky $X \sim \text{Alt}(0,5)$. Z výsledků vyplývajících z tabulky 19 a obrázku 12 vidíme, že zvolený pokus v hod mincí opravdu vystihuje alternativní rozdělení.

Tabulka 19. $X \sim \text{Alt}(0,5)$, OH: -10, 11

Char.	TH	ZV	1 OH	2 OH	+ 1,5	· 1,5	- 6 H	±5%
\bar{x}	0,5	0,54	0,333	0,539	2,04	0,81	0,545	0,545
\hat{x}	{0; 1}	1	1	1	2,5	1,5	1	0
\tilde{x}	< 0; 1 >	1	1	1	2,5	1,5	1	0,95
s^2	0,25	0,253	2,427	4,567	0,253	0,57	0,253	0,26
d	-	0,46	0,667	0,846	0,46	0,69	0,455	0,469
\hat{R}	1	1	11	21	1	1,5	1	1,05
\hat{R}_Q	1	1	1	1	1	1,5	1	1,05
$\hat{\alpha}_3$	0	-0,159	-5,8	-0,053	-0,159	-0,159	-0,180	-0,144
$\hat{\alpha}_4$	1	1,005	38,742	22,857	1,005	1,005	1,01	1,02



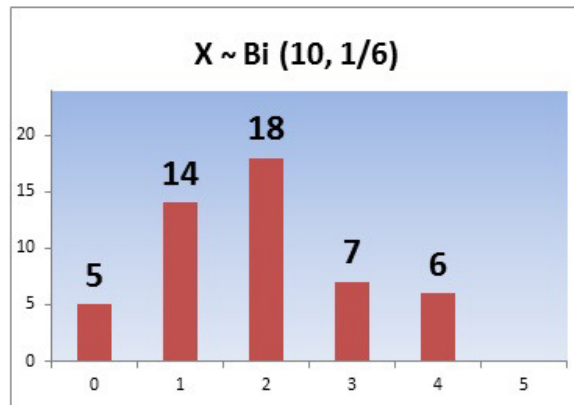
Obrázek 12: Rozdělení četností pro hod mincí

8.2. Binomické rozdělení

Reálná data pro binomické rozdělení jsme získali pomocí obyčejné hrací kostky. Počítali jsme, kolikrát nám z deseti pokusů padne číslo „6”. Symbolicky $X \sim \text{Bi}(10, \frac{1}{6})$. V tomto případě se nám mírně liší střední hodnota, šikmost a špičatost, ale i tento případ reálného pokusu se dá považovat za velmi věrný vzhledem k napodobení binomického rozdělení.

Tabulka 20. $X \sim \text{Bi}(10, \frac{1}{6})$, OH: -4, 12

Char.	TH	ZV	1 OH	2 OH	+ 1,5	· 1,5	- 6 H	±5%
\bar{x}	1,667	1,9	1,784	1,981	3,4	2,85	1,886	1,901
\hat{x}	1	2	2	2	3,5	3	2	2,1
\tilde{x}	1	2	2	2	3,5	3	2	2
s^2	1,389	1,316	1,973	3,941	1,316	2,962	0,94	1,29
d	-	0,86	0,961	1,135	0,86	1,29	0,705	0,885
\hat{R}	10	4	8	16	4	6	4	4,2
\hat{R}_Q	1	1,5	1,5	2	1,5	2,25	1	1,475
$\hat{\alpha}_3$	0,62	0,276	-1,088	1,997	0,276	0,276	0,379	0,228
$\hat{\alpha}_4$	3,12	2,371	6,864	14,588	2,371	2,371	2,752	2,368



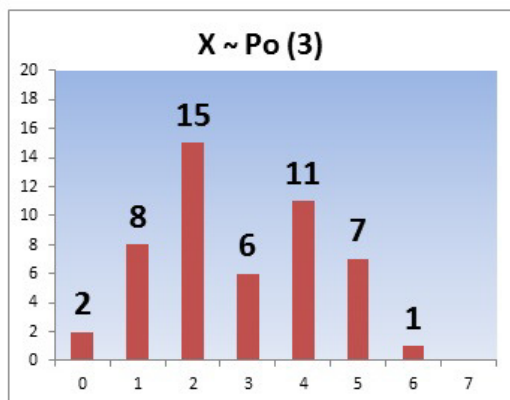
Obrázek 13: Rozdělení četností pro hod kostkou

8.3. Poissonovo rozdělení

Pro reálný příklad Poissonova rozdělení jsme použili počty vstřelených branek v úvodních 50ti zápasech české fotbalové Gambrinus ligy ročníku 2012/2013. (Zdroj dat: <http://www.sport.cz/fotbal/gambrinus-liga/#vysledky>). Z dlouhodobého hlediska je průměrný počet branek na zápas 3, proto budeme tuto hodnotu brát jako parametr teoretického rozdělení. Symbolicky $X \sim \text{Po}(3)$. Zde vidíme, že očekávaná střední hodnota se realizovala méně často než by se z dlouhodobého hlediska dalo předpokládat. Shodou okolností ale tato skutečnost nemá velký vliv na vypočtenou střední hodnotu, jelikož se liší pouze o 0,18. Střední hodnota vyšla poměrně přesně jen díky velikému zastoupení hodnot 2 a 4. Šikmost a špičatost moc neodpovídají teoretickým hodnotám. Tento reálný pokus tedy zhodnotíme jako méně kvalitní, ale i tak popisuje rozdělení řídce se vyskytujících se jevů.

Tabulka 21. $X \sim \text{Po}(3)$, OH: -7, 18

Char.	TH	ZV	1 OH	2 OH	+ 1,5	· 1,5	- 6 H	±5%
\bar{x}	3	2,82	2,628	2,923	4,32	4,23	2,818	2,813
\hat{x}	2	2	2	2	3,5	3	2	2,1
\tilde{x}	2	2,5	2	2,5	4	3,75	2,5	2,475
s^2	3	2,273	4,118	8,582	2,273	5,114	1,687	2,219
d	-	1,3	1,451	1,731	1,3	1,95	1,136	1,227
\hat{R}	∞	6	13	25	6	9	4	5,7
\hat{R}_Q	2	2	2	2	2	3	2	2,1
$\hat{\alpha}_3$	0,577	0,128	-1,929	1,901	0,128	0,128	0,21	0,093
$\hat{\alpha}_4$	3,333	1,986	10,772	16,470	1,986	1,986	1,781	1,974



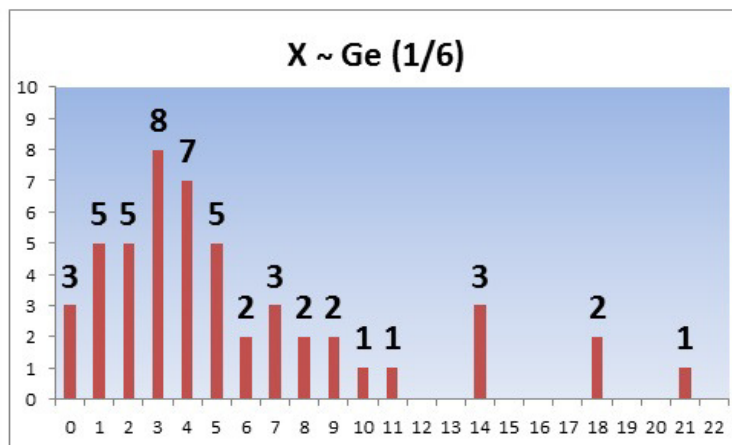
Obrázek 14: Rozdělení četností pro počet vstřelených branek

8.4. Geometrické rozdělení

Pro geometrické rozdělení jsme si opět připravili hrací kostku. Čekali jsme na situaci, než nám padne číslo „6“. Náhodná veličina tedy značí počet pokusů, kdy šestka nepadne. Symbolicky $X \sim \text{Ge}(\frac{1}{6})$. Již dříve jsme si řekli, že teoretický modus geometrického rozdělení je roven 0. V tomto případě nám jako nejčastější hodnota vyšlo číslo 3. I v případě, že hrajeme „Člověče nezlob se“, taky nám číslo 6 nepadne hned. I přes tento zádrhel ale data poměrně dobře kopírují geometrické rozdělení.

Tabulka 22. $X \sim \text{Ge}(\frac{1}{6})$, OH: -6, 100

Char.	SH	ZV	1 OH	2 OH	+ 1,5	· 1,5	- 6 H	±5%
\bar{x}	5	5,58	5,353	7,173	7,08	8,37	5,045	5,537
\hat{x}	0	3	3	3	4,5	4,5	3	{3, 15; 4, 2}
\tilde{x}	2	4	4	4	5,5	6	4	4,2
s^2	30	24,167	26,313	198,068	24,167	54,376	12,37	23,242
d	-	3,38	3,51	5,288	3,38	5,07	2,545	3,323
\hat{R}	∞	21	27	27	106	31,5	13	19,95
\hat{R}_Q	7	5	5	5	5	7,5	4	5,25
$\hat{\alpha}_3$	2,2	1,418	1,126	5,627	1,418	1,418	1,143	1,397
$\hat{\alpha}_4$	9,84	4,442	4,379	37,172	4,442	4,442	3,61	4,349



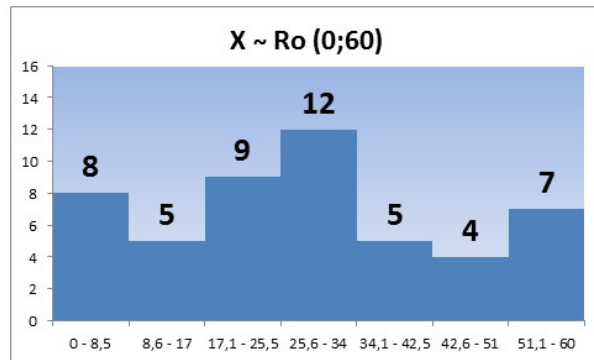
Obrázek 15: Rozdělení četností pro čekání na padnutí šestky

8.5. Rovnoměrné rozdělení

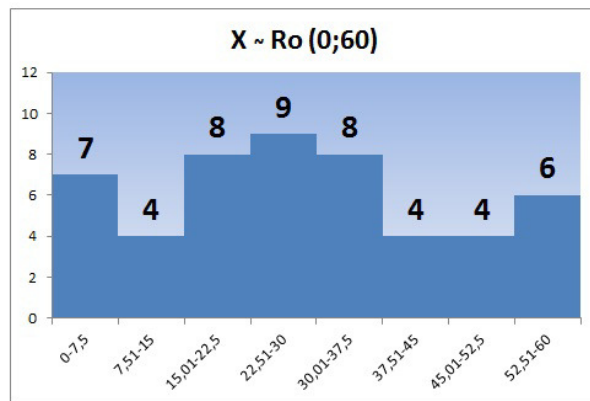
Pro reálný příklad rovnoměrného rozdělení jsme použili běžné ručičkové hodinky. Libovolně během dne jsme pozorovali ručičku, udávající počet sekund, a danou hodnotu jsme si zapsali. Zaokrouhlovali jsme na celé vteřiny. Můžeme označit jako $X \sim \text{Ro}(0,60)$. Tento pokus je také poměrně zdařilý. Z grafu lze vyčíst, že nejčasteji jsme ručičku zpozorovali mezi 17. a 34. vteřinou. Níže je uveden ještě druhý histogram s 8 buňkami pro lepší grafickou interpretaci.

Tabulka 23. $X \sim \text{Ro}(0,60)$, OH: -45, 127

Char.	TH	ZV	1 OH	2 OH	+ 1,5	· 1,5	- 6 H	$\pm 5\%$
\bar{x}	30,5	28,32	26,882	28,808	29,82	42,48	28,205	28,086
\hat{x}	< 1; 60 >	-	-	-	-	-	-	-
\tilde{x}	30,5	27,5	26	27,5	29	41,25	27,5	27,375
s^2	290,08	274,712	374,626	560,041	274,712	618,102	214,812	259,125
d	-	13,64	14,765	16,423	13,64	20,46	11,886	13,272
\hat{R}	60	55	102	172	55	82,5	51	54,6
\hat{R}_Q	29,5	24,5	25,5	25	24,5	36,75	18	24,95
$\hat{\alpha}_3$	0	0,152	-0,77	0,827	0,152	0,152	0,193	0,083
$\hat{\alpha}_4$	1,8	1,939	4,878	8,112	1,939	1,939	2,108	1,927



Obrázek 16: Histogram pro vteřinovou ručičku



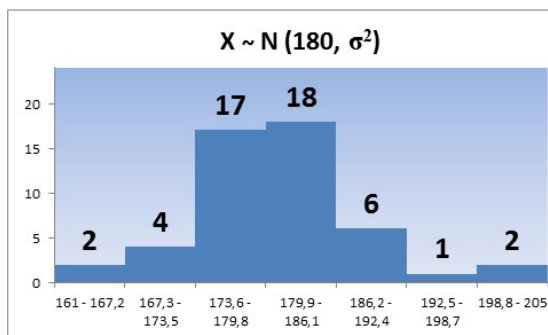
Obrázek 17: Upravený histogram pro vteřinovou ručičku

8.6. Normální rozdělení

Jako reálný příklad normálního rozdělení byla zjišťována výška mužské populace v Olomouci. Výzkum jsem prováděl na Horním náměstí v Olomouci u městské radnice 5.4.2013. Náhodně jsem oslovoval různé věkové skupiny mužů pro více reprezentativní alespoň věkově stratifikovaný vzorek. Jak je z grafu vidět, $\frac{2}{3}$ všech dotázaných uvedlo výšku mezi 173 až 186cm. V knize „Two centuries of growth among Czech children and youth” od autorů J.Vignerová, M.Brabec, P.Bláha, je uvedeno, že v roce 2001 byla průměrná výška české mužské populace 1,8031m. Náš pokus tedy velice dobře odpovídá tomuto předpokladu. Rozptyl je neznámý.

Tabulka 24. $X \sim N(180, \sigma^2)$, OH: 110, 472

Char.	TH	ZV	1 OH	2 OH	+ 1,5	· 1,5	- 6 H	±5%
\bar{x}	180	180,46	179,078	184,712	181,96	270,69	180,296	180,339
\hat{x}	180	-	-	-	-	-	-	-
\tilde{x}	180	180	180	180	181,5	270	180	182,55
s^2	-	64,213	160,274	1807,19	64,213	144,479	28,539	101,55
d	-	5,74	7	12,481	5,74	8,61	4,159	8,499
\hat{R}	∞	44	95	362	44	66	21	40
\hat{R}_Q	-	8	8,5	9	8	12	7	17,4
$\hat{\alpha}_3$	0	0,438	-3,01	5,937	0,438	0,438	0,414	-0,063
$\hat{\alpha}_4$	3	4,206	18,466	41,091	4,206	4,206	2,635	1,977



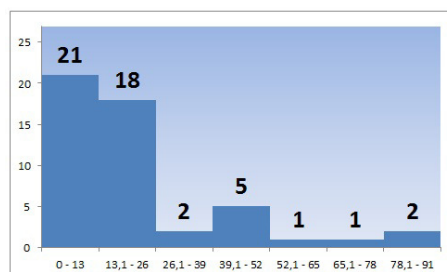
Obrázek 18: Histogram pro výšku mužů

8.7. Exponenciální rozdělení

Pro toto rozdělení jsem zkusil následující pokus: zjišťoval jsme časový interval v sekundách mezi příchozími lidmi do parku. Jako bod, ve kterém jsem anuloval stopky, byla brána (vstup do parku). Při vejíí prvního člověka do parku jsem začal měřit čas, stopnul, když vešel druhý člověk a zapsal první údaj. Takto jsem postupoval až do 51. osoby, která pro mne znamenala konečný 50. údaj. Můžeme pouze spekulovat, zda tento pokus vyhovuje exponenciálnímu rozdělení. Můžeme se tak domnívat domnívat pouze podle grafu, podle kterého by se o exponenciální rozdělení jednat mohlo. SH neuvádíme, protože jsme si je neuváděli v kapitole 6 a dále z důvodu složitosti exponenciálního rozdělení. Jediné, čím se můžeme řídit, je, že ve shodě s teorií, vychází rozptyl jako druhá mocnina střední hodnoty.[1]

Tabulka 25. $X \sim E$, OH: -10, 115

Char.	ZV	1 OH	2 OH	+ 1,5	· 1,5	- 6 H	±5%
\bar{x}	21,44	20,824	22,635	22,94	32,16	18,614	21,336
\hat{x}	-	-	-	-	-	-	-
\tilde{x}	15,5	15	15,5	17	23,25	15,5	15,75
s^2	447,884	458,308	619,884	447,884	1007,74	197,591	428,1
d	14,2	14,412	16,058	14,2	21,3	10,523	14,072
\hat{R}	90	101	125	90	135	54	85,5
\hat{R}_Q	18	18,5	19	18	27	15	18,675
$\hat{\alpha}_3$	1,713	1,645	1,868	1,713	1,713	1,008	1,606
$\hat{\alpha}_4$	5,759	5,667	6,476	5,759	5,759	3,28	5,239



Obrázek 19: Histogram pro časy mezi příchody

Závěr

V této práci jsme si uvedli a vysvětlili význam nejdůležitějších pojmů teorie pravděpodobnosti a matematické statistiky. Dále jsme zpracovali některá diskrétní a spojitá rozdělení náhodné veličiny. Teoretická část byla uzavřena výčtem číselných charakteristik a jejich vysvětlení pro teoretický i empirický případ. Hlavní část této práce se věnovala výpočtům charakteristik pro jednotlivá rozdělení. Uvažovali jsme ovšem pouze prvních 5 rozdělení, a to z důvodu rozsahu této bakalářské práce a náročností samotných výpočtů. Charakteristiky jsme počítali pro simulovaná i reálně napozorovaná data.

Při psaní této bakalářské práce jsem se naučil zpracovávat odborný text a vložit do něj svoje poznámky. Také jsem si prohloubil své znalosti v užívání softwarů R a Excel 2010.

Věřím, že úsilí, které jsem věnoval vytvoření této bakalářské bude mít za následek můj přínos v oboru.

9. Přílohy

9.1. Simulované hodnoty

$X \sim \text{Alt}(0,25)$

Data : 0,0,1,1,0,0,0,0,1,0,0,1,0,0,0,1,1,0,1,0,0,0,1,1,1,0,0,0,1,0,0,0,0,0,0,0,0,0,1,0,0,0,1,0,0,0,0,0

Odlehlé hodnoty : -10, 11; Příkaz v R : `\rbinom(50,1,0.25)`

$X \sim \text{Alt}(0,5)$

Data : 1,0,0,0,1,1,0,1,0,1,1,1,0,1,1,0,0,0,0,1,0,0,0,1,0,0,1,1,0,1,0,1,1,1,1,0,1,1,1,0,0,1,0,1,0,0,1,0,0

Odlehlé hodnoty : -10, 11; Příkaz v R : `\rbinom(50,1,0.5)`

$X \sim \text{Alt}(0,75)$

Data : 1,1,1,1,1,0,1,1,0,1,1,1,1,1,1,1,0,0,1,1,1,1,1,1,1,1,1,1,1,1,1,0,1,1,1,1,1,0,1,1,1,1,0,1,1,1,1,0

Odlehlé hodnoty : -10, 11; Příkaz v R : `\rbinom(50,1,0.75)`

$X \sim \text{Bi}(4;0,25)$

Data : 1,2,0,1,1,1,1,1,1,1,3,2,2,2,1,2,0,0,0,1,1,1,0,0,0,2,1,2,0,3,0,1,0,2,1,0,0,0,0,2,2,0,0,3,0,2,1,0,2,0

Odlehlé hodnoty : -4, 12; Příkaz v R : `\rbinom(50,4,0.25)`

$X \sim \text{Bi}(6;0,5)$

Data : 2,1,3,2,2,2,1,4,3,2,3,2,2,0,3,2,2,2,2,2,1,3,3,3,2,4,1,3,3,2,3,2,1,2,2,3,1,2,3,2,3,3,2,2,1,4,4,2,3,2

Odlehlé hodnoty : -4, 12; Příkaz v R : `\rbinom(50,6,0.5)`

$X \sim \text{Bi}(8;0,75)$

Data : 6,7,7,4,7,6,7,6,4,5,6,6,6,6,6,8,5,6,6,7,7,7,6,5,5,5,6,7,5,5,8,7,7,5,5,6,7,6,6,5,6,4,4,8,7,8,5,8,5,6

Odlehlé hodnoty : -4, 12; Příkaz v R : `\rbinom(50,8,0.75)`

$X \sim \text{Po}(1)$

Data : 1,1,5,0,0,0,1,2,1,4,0,0,0,1,0,1,3,0,1,3,0,3,3,1,0,0,0,0,2,0,2,1,0,0,0,1,0,2,0,2,1,0,0,0,0,1,1,0,0,1

Odlehlé hodnoty : -7, 18; Příkaz v R : `\rpois(50,1)`

$X \sim \text{Po}(3)$

Data : 4,3,1,5,4,1,3,5,4,1,3,3,5,4,2,3,1,4,1,3,3,4,5,3,3,2,1,5,4,4,4,7,0,2,2,6,3,3,2,5,1,4,4,4,2,2,2,2,1,1

Odlehlé hodnoty : -7, 18; Příkaz v R : `\rpois(50,3)`

$X \sim \text{Po}(6)$

Data : 5,5,4,6,3,4,4,6,3,5,6,5,5,7,13,7,4,9,10,7,10,3,4,10,6,2,7,7,3,4,4,3,5,3,5,7,5,7,7,8,11,2,4,8,7,4,6,3,3,2

Odlehlé hodnoty : -7, 18; Příkaz v R : `\rpois(50,6)`

$X \sim \text{Ge}(0,5)$

Data : 3,1,0,1,1,1,0,2,0,1,0,3,6,1,0,1,1,0,0,3,2,2,4,0,2,3,1,0,0,2,0,2,1,0,2,0,0,1,0,0,0,0,0,1,0,0,0,2,1

Odlehlé hodnoty : -6, 100; Příkaz v R : `\rgeom(50,0.5)`

$X \sim \text{Ge}(0,25)$

Data : 2,0,1,5,1,0,0,3,0,1,0,5,5,7,14,1,4,6,6,3,6,2,0,1,0,1,0,6,5,0,8,0,4,0,3,3,20,0,0,5,1,3,2,0,6,0,1,1,1,2

Odlehlé hodnoty : -6, 100; Příkaz v R : `\rgeom(50,0.25)`

$X \sim \text{Ge}(0,14)$

Data : 0,4,9,14,10,1,2,16,2,10,6,2,9,1,5,12,13,3,0,7,5,0,12,2,5,10,0,27,7,0,11,10,2,0,1,9,1,2,3,1,0,2,7,2,34,15,10,3,1,3

Odlehlé hodnoty : -6, 100; Příkaz v R : `\rgeom(50,0.14)`

$X \sim \text{Ro}(0;2)$

Data : 0.7884,0.8498,0.8224,0.6722,1.4276,0.4427,0.1473,1.1012,0.3866,1.5066,1.1273,0.9525,0.4867,0.5754,0.8351,1.0710,1.1272,0.4039,0.9920,1.5669,0.5521,0.2729,1.1469,0.0017,0.0526,0.1653,0.1695,1.3772,0.2216,1.7525,1.7953,1.5199,0.4668,0.3633,0.7892,0.2020,1.2527,1.8755,1.7755,1.0472,0.1562,1.3086,0.5903,1.8084,1.7607,1.9390,0.8123,1.0421,1.6095,1.8767

Odlehlé hodnoty : -15.199(-45), 22.16(127); Příkaz v R : `\runif(50,0,2)`

$X \sim \text{Ro}(0;6)$

Data : 0.2669,2.2374,1.0469,2.7257,0.0290,1.9287,1.6552,4.5613,1.0647,3.4299,0.0575,5.9249,4.3299,2.1921,4.5888,0.9950,4.8116,2.4034,2.9361,4.1739,5.9788,1.2723,1.7392,1.3634,0.3387,2.1021,0.4120,1.3725,4.7709,4.7914,2.1918,0.0921,4.8996,1.8408,4.9295,5.5158,4.9439,0.5247,5.1356,5.4678,4.4033,0.2729,3.8736,5.6444,4.8061,4.8271,5.5689,3.9291,3.0171,3.5031

Odlehlé hodnoty : -15.199(-45), 22.16(127); Příkaz v R : `\runif(50,0,6)`

$X \sim \text{Ro}(5;15)$

Data : 13.4637,8.6722,14.7747,14.0604,10.9495,5.8293,6.4893,12.9495,6.5763,8.0903,9.1569,12.0158,12.0248,10.9794,14.7466,6.7130,11.1041,14.0298,7.1362,6.3633,11.9002,5.8015,9.5780,8.9229,8.5981,8.7813,13.0751,14.1178,11.1240,12.8868,7.8096,7.4730,7.4001,14.9701,6.5602,7.8755,12.1882,5.2849,14.2873,7.9606,14.0862,13.4194,7.5028,6.3436,11.6089,11.1472,12.8474,8.9491,12.7669,12.3025

Odlehlé hodnoty : -15.199(-45), 22.16(127); Příkaz v R : `\runif(50,5,15)`

$X \sim \text{N}(0;1)$

Data : -0.6209,-1.9616,1.6617,-0.0505,-1.8417,-0.741,-1.1975,-0.2512,-1.7306,-0.2658,0.532,-0.4920,-1.3347,0.2453,2.3603,-0.2307,-0.5015,-0.211,-1.1377,2.4308,-1.1156,0.589,-1.7338,-0.9574,-0.9518,-0.0965,1.0312,-0.2494,-1.1272,-0.2655,0.3539,0.0329,-1.5844,0.7702,0.6099,2.5454,1.3949,1.306,0.212,-0.2495,-0.8586,-1.3949,1.9899,0.2104,0.836,-1.6753,1.9217,-0.8025,-0.0687,1.2631

Odlehlé hodnoty : -6.209, 7.702; Příkaz v R : `\rnorm(50,0,1)`

$X \sim N(0;4)$

Data : -1.525,2.3016,-0.1967,0.4844,1.646,0.8259,-0.2768,-2.1176,-0.4015,-3.1426,-1.0538,-2.9534,
-1.0589,1.8442,2.0117,3.4536,1.1839,-2.2493,-2.8466,1.9186,1.399,-1.9528,-0.3172,-0.2829,0.9896,
1.9931,4.0248,1.2045,-2.969,-3.16,-3.007,1.1364,-2.2275,-2.0633,5.4437,-0.8796,-1.4167,-1.48,
-0.6765,2.3827,1.0097,-2.6688,2.2533,-3.4889,0.0535,0.7657,-1.5467,3.5046,0.094,-0.8726

Odlehlé hodnoty : -6.209, 7.702; Příkaz v R : `\rnorm(50,0,2)`

$X \sim N(3;1)$

Data : 4.9969,3.5791,4.4487,3.1475,3.0684,1.0921,0.2188,2.8604,1.1785,2.5911,2.1174,5.0417,
1.8264,3.4742,2.3664,3.0158,0.5837,2.1366,2.951,3.4585,1.471,3.5009,4.8779,4.0352,3.0954,
2.7628,3.3505,2.8277,3.7984,2.0012,3.9294,2.8123,3.1936,2.7388,3.3311,3.0509,4.3693,0.8824,
2.7485,4.4867,3.6866,2.605,5.5837,3.2328,4.065,3.5558,3.4668,2.4178,3.256,3.1547

Odlehlé hodnoty : -6.209, 7.702; Příkaz v R : `\rnorm(50,3,1)`

9.2. Reálné hodnoty

Alternativní rozdělení

Data : 1,1,0,0,1,1,1,0,0,1,0,1,1,0,0,1,1,0,1,0,1,1,0,1,1,0,0,1,1,1,1,0,0,0,0,0,1,1,1,1,0,1,0,1,0,1,0,0

Odlehlé hodnoty : -10, 11

Binomické rozdělení

Data : 1,3,2,2,0,4,2,0,0,1,4,2,2,3,2,2,2,3,0,3,3,1,1,1,4,1,3,2,2,1,2,4,4,2,1,1,2,2,2,1,1,1,2,4,2,1,1,2,0,3

Odlehlé hodnoty : -4, 12

Poissonovo rozdělení

Data : 2,4,1,2,1,1,3,6,4,5,4,3,2,2,5,2,4,2,1,2,5,5,2,3,4,1,2,4,0,2,5,4,5,2,2,3,2,2,2,3,4,4,1,3,0,1,4,5,4,1

Odlehlé hodnoty : -7, 18

Geometrické rozdělení

Data : 3,0,18,1,0,14,2,4,1,4,4,3,1,18,1,4,4,9,3,11,3,7,2,6,2,5,5,14,7,5,2,7,6,3,4,9,4,21,10,2,0,3,3,1,5,
5,3,8,8,14

Odlehlé hodnoty : -6,100

Rovnoměrné rozdělení

Data : 36,54,32,32,22,52,2,54,11,29,53,5,16,18,45,29,18,3,42,24,33,18,51,31,18,22,8,6,3,5,
54,51,36,24,25,26,31,49,30,56,24,25,31,3,16,41,11,14,57,40

Odlehlé hodnoty : -45, 127

Normální rozdělení

Data : 178,184,169,177,181,193,180,179,182,174,188,182,174,177,181,191,163,205,184,181,
175,190,182,179,184,179,180,192,182,171,175,177,181,185,179,199,189,176,180,171,183,
161,175,183,172,174,191,182,177,176

Odlehlé hodnoty : 110,472

Exponenciální rozdělení

Data : 11,22,32,3,44,91,15,24,1,10,9,32,68,11,1,5,18,22,22,22,56,5,11,21,2,8,1,3,23,2,26,
50,6,42,13,4,17,16,91,15,2,4,10,15,18,23,14,41,50,20

Odlehlé hodnoty : -10, 115

Literatura

- [1] P.Kunderová : Základy pravděpodobnosti a matematické statistiky, UP Olomouc, 2004
- [2] J.Kohoutová, D.Bílková : Počet pravděpodobnosti, Nakladatelství VŠE, 2005
- [3] J.Anděl : Základy matematické statistiky, MatFyz press, 2007
- [4] <http://cs.wikipedia.org/wiki/Modus> 15.3.2013
- [5] <http://cs.wikipedia.org/wiki/Medián> 15.3.2013
- [6] K.Zvára, J.Štěpán : Pravděpodobnost a matematická statistika, MatFyz press, 2006
- [7] <http://home.zcu.cz/~friesl/hpsb/exp.html> 11.4.2013
- [8] <http://www.czso.cz> 11.3.2013