

PRAVDĚPODOBŇNOST A MATEMATICKÁ STATISTIKA I

RNDr. Tomáš Mrkvička, Ph.D.

16. března 2009

Literatura

- [1] J. Anděl: *Statistické metody*, Matfyzpress, Praha 1998
- [2] V. Dupač, M. Hušková: *Pravděpodobnost a matematická statistika*, Karolinum, Praha 1999
- [3] T. Mrkvička, V. Petrášková: *Úvod do teorie pravděpodobnosti*, PF JU, České Budějovice 2008
- [4] T. Mrkvička, V. Petrášková: *Úvod do statistiky*, PF JU, České Budějovice 2006

Kapitola 1

Jev, náhodný jev, pravděpodobnosti náhodného jevu

1.1 Axiomatická definice pravděpodobnosti

Každému náhodnému pokusu můžeme přiřadit množinu Ω , tj. množinu všech možných výsledků pokusu. Při hodu kostkou je $\Omega = \{1, 2, 3, 4, 5, 6\}$, při hodu mincí je $\Omega = \{\text{rub}, \text{líc}\}$, při hodu dvěma mincemi je $\Omega = \{\{\text{rub}, \text{rub}\}, \{\text{líc}, \text{líc}\}, \{\text{líc}, \text{rub}\}, \{\text{rub}, \text{líc}\}\}$.

Prvky $\omega \in \Omega$ nazýváme *elementárními jevy*,
podmnožiny množiny Ω *jevy*.

Jev jistý - Ω - je takový jev, který nastane při každé realizaci pokusu.

Jev nemožný - \emptyset - je takový jev, který nenastane při žádné realizaci pokusu.

Definice 1.1 *Nechť \mathcal{A} je neprázdný systém podmnožin množiny $\Omega \neq \emptyset$ takový, že*

a) $\emptyset \in \mathcal{A}$

b) je-li $A \in \mathcal{A}$, pak $\bar{A} \in \mathcal{A}$

c) jsou-li $A_i \in \mathcal{A}$, $i = 1, 2, \dots$, pak $\cup_{i=1}^{\infty} A_i \in \mathcal{A}$.

Pak \mathcal{A} nazýváme σ -algebrou.

zápis	pravděpodobnostní interpretace
$\omega \in A$	jev A nastal, výsledek ω náhodného pokusu je příznivý jevu A
$A \subset B$	A je podjev jevu B (jev A nastane, kdykoliv nastane jev B)
$B - A$	rozdíl jevu B a A (jev, který nastane právě tehdy, když nastane jev B a zároveň nenastane jev A)
$\bar{A} = \Omega - A$	doplňěk jevu A (jev, který nastane právě tehdy, když nenastane jev A)
$A \cup B$	sjednocení jevů A, B (jev, který nastane právě tehdy, nastane-li aspoň jeden z jevů A, B)
$A \cap B$	průnik jevů A, B (jev, který nastane právě tehdy, nastanou-li oba dva jevy současně)
$A \cap B = \emptyset$	jevy A, B nazveme disjunktní (nemohou nastat současně)
$A_i \cap A_j = \emptyset$ $\forall i \neq j$ $\bigcup_{i=1}^n A_i = C$	jevy $A_1 \dots A_n$ tvoří rozklad jevu C

Tabulka 1.1: Zápis základních pravděpodobnostních relací a operací.

σ -algebra \mathcal{A} je tedy množinový systém uzavřený vzhledem k doplňku a spočetnému sjednocení. Prvky σ -algebry nazýváme *náhodné jevy*. Uvedme si nyní některé další vlastnosti náhodných jevů:

1) Jev jistý je náhodný jev $\Omega \in \mathcal{A}$.

2) Rozdíl dvou náhodných jevů je náhodný jev

$$A, B \in \mathcal{A} \Rightarrow B - A \in \mathcal{A}.$$

3) Průnik spočetně mnoha náhodných jevů je náhodný jev

$$A_1, A_2, \dots \in \mathcal{A} \Rightarrow \bigcap_{i=1}^{\infty} A_i \in \mathcal{A}.$$

Dvojici (Ω, \mathcal{A}) nazýváme *jevové pole*.

Příklad 1.1 *Nechť $\Omega = \{1, \dots, n\}$. Pak potenční množina $P(\Omega)$ (tj. množina všech podmnožin Ω) je σ -algebra a neexistuje menší σ -algebra obsahující všechny*

elementární jevy $\{\omega\}$, $\omega \in \Omega$.

Příklad 1.2 *Nechť $\Omega = \mathbb{R}$. Pak potenční množina je také σ -algebra, ale existuje i menší σ -algebra, které dáváme přednost. Např. Borelovská σ -algebra (tj. nejmenší σ -algebra obsahující všechny otevřené podmnožiny \mathbb{R}). Borelovská σ -algebra obsahuje všechna spočetná sjednocení otevřených množin, ale také i všechny uzavřené podmnožiny \mathbb{R} .*

Definice 1.2 *Nechť $\Omega \neq \emptyset$, \mathcal{A} je σ -algebra definovaná na Ω . Pak pravděpodobností libovolného náhodného jevu A nazveme libovolnou reálnou funkcí P definovanou na \mathcal{A} , která splňuje*

a) $P(\Omega) = 1$, $P(\emptyset) = 0$,

b) $P(A) \geq 0 \forall A \in \mathcal{A}$,

c) pro každou posloupnost disjunktích jevů $\{A_n\}_{n=1}^{\infty}$ platí

$$P(\cup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} P(A_i).$$

Trojice (Ω, \mathcal{A}, P) se nazývá pravděpodobnostní prostor.

Některé vlastnosti pravděpodobnosti.

- 1) $P(\emptyset) = 0$,
- 2) P je konečně aditivní, tzn., jestliže $A_1, \dots, A_n \in \mathcal{A}$, $A_i \cap A_j = \emptyset \quad \forall i \neq j, i, j = 1, \dots, n \Rightarrow P(\cup_{i=1}^n A_i) = \sum_{i=1}^n P(A_i)$,
- 3) P je monotónní: $A, B \in \mathcal{A}, A \subset B \Rightarrow P(A) \leq P(B)$,
- 4) $A, B \in \mathcal{A}, A \subset B \Rightarrow P(B - A) = P(B) - P(A)$,
- 5) $P(\bar{A}) = 1 - P(A), \forall A \in \mathcal{A}$,

6) $P(A \cup B) = P(A) + P(B) - P(A \cap B)$ pro libovolné $A, B \in \mathcal{A}$,

7) $P(\cup_{i=1}^{\infty} A_i) \leq \sum P(A_i)$, pro libovolnou posloupnost $\{A_i\}$ v \mathcal{A} ,

8) $A_n \in \mathcal{A}, A_1 \subset A_2 \subset A_3 \dots, A = \cup_{n=1}^{\infty} A_n \Rightarrow P(A) = \lim_{n \rightarrow \infty} P(A_n)$,

9) $A_n \in \mathcal{A}, A_1 \supset A_2 \supset A_3 \dots, A = \cap_{n=1}^{\infty} A_n \Rightarrow P(A) = \lim_{n \rightarrow \infty} P(A_n)$,

Vlastnost 6) lze indukcí rozšířit na libovolný konečný počet jevů, a to následovně:

$$\begin{aligned}
 P(\cup_{i=1}^n A_i) &= \sum_{i=1}^n P(A_i) - \sum_{i=1}^{n-1} \sum_{j=i+1}^n P(A_i \cap A_j) + \\
 &+ \sum_{i=1}^{n-2} \sum_{j=i+1}^{n-1} \sum_{k=j+1}^n P(A_i \cap A_j \cap A_k) + \dots + (-1)^{n-1} P(\cap_{i=1}^n A_i).
 \end{aligned} \tag{1.1}$$

1.2 Klasický pravděpodobnostní prostor

Definice 1.3 *Pravděpodobnostní prostor (Ω, \mathcal{A}, P) nazveme klasickým pravděpodobnostním prostorem, jestliže*

- a) množina Ω je konečná o m prvcích a všechny možné výsledky jsou stejně pravděpodobné, tzn. označíme-li postupně p_1, \dots, p_m pravděpodobnosti jednotlivých výsledků elementárních jevů, pak $p_1 = p_2 = \dots = p_m = \frac{1}{m}$ (je-li možných výsledků m),
- b) za σ -algebru \mathcal{A} vezmeme systém všech podmnožin množiny Ω ,
- c) pravděpodobnost P náhodného jevu A je rovna

$$P(A) = \frac{m_A}{m},$$

kde m_A je počet výsledků příznivých jevů A a m je počet všech možných výsledků náhodného pokusu. Pravděpodobnost takto definovaná se nazývá kla-

sická pravděpodobnost.

Čtenář si sám ověří, že klasická pravděpodobnost je pravděpodobností ve smyslu definice 1.2, tzn. ověří všechny tři axiomy.

Příklad 1.3 *Házíme dvěma kostkami. Stanovte pravděpodobnost jevu A "na kostkách padne součet menší než 5".*

Jsou-li jevy disjunktní, jejich pravděpodobnosti se sčítají.

Příklad 1.4 *V urně máme 32 karet, z toho 4 esa. Dvakrát za sebou vytáhneme náhodně jednu kartu s tím, že po prvním tahu ji a) vrátíme zpět do urny, b) nevrátíme. Stanovte pravděpodobnost jevu A "alespoň jedna z vytažených karet je eso".*

Dvě reprezentace: 1) rozlišujeme pořadí, 2) nerozlišujeme pořadí.

Princip inkluze a exkluze. Použití doplňkového jevu.

Příklad 1.5 *Házíme jednou šesti kostkami, přičemž každou kostku si očíslováme jedním z čísel $1, \dots, 6$ (každá kostka bude očíslovaná jiným číslem). Jaká je pravděpodobnost, že alespoň na jedné kostce bude počet ok souhlasit s číslem, jímž jsme kostku označili?*

Příklad 1.6 *V osudí je n lístků očíslovaných čísly 1 až n ; r -krát po sobě vytáhneme po jednom lístku, přičemž každý lístek po tahu vracíme zpět. Jaká je pravděpodobnost, že v r tazích vyjde každé z n čísel aspoň jednou? (Předpokládejme, že $n < r$).*

1.3 Geometrická pravděpodobnost

O geometrické pravděpodobnosti mluvíme v případě, že

a) $\Omega \subset \mathbb{R}^d$.

- b) $\mathcal{A} = \mathcal{B}(\Omega)$ je Borelovská σ -algebra na Ω (tj. nejmenší σ -algebra obsahující všechny otevřené podmnožiny Ω).
- c) $P(A) = \frac{\mu^d(A)}{\mu^d(\Omega)}$, kde μ^d je d -rozměrná Lebesgueova míra. Pro naše účely postačí, pokud si pod $\mu^1(A)$ představíme délku množiny A , pod $\mu^2(A)$ obsah A a pod $\mu^3(A)$ objem A .

Geometrická pravděpodobnost je vhodným modelem tam, kde výsledkům pokusu lze jednoznačně přiřadit body $\omega \in \Omega \subset \mathbb{R}^d$ a kde žádným výsledkům nelze dát přednost před ostatními.

Příklad 1.7 *Autobusy přijíždějí na zastávku pravidelně v 10 minutových intervalech. Student přijde na zastávku v náhodném čase. Jaká je pravděpodobnost, že bude čekat déle než 5 minut?*

Příklad 1.8 *Dvě osoby (I, II) přijdou na místo schůzky mezi 12. a 13. hodinou. Doby příchodu osob jsou náhodné a nezávislé. Ten, kdo přijde na místo*

schůzky, čeká 20 minut a nedočká-li se druhého, odchází. Jaká je pravděpodobnost, že se osoby setkají?

1.4 Další příklady pravděpodobnostních prostorů

• Diskrétní

a) $\Omega = \{\omega_1, \omega_2, \dots\}$.

b) \mathcal{A} je množina všech podmnožin Ω .

c) Jsou dány pravděpodobnosti elementárních jevů $P(\omega_i)$, které splňují: $\sum_{i=1}^{\infty} P(\omega_i) = 1$.

1. Pak pravděpodobnost libovolného jevu je dána jednoznačně vztahem

$$P(A) = \sum_{\omega_i \in A} P(\omega_i).$$

• Spojitý

a) $\Omega = \mathbb{R}$.

b) $\mathcal{A} = \mathcal{B}(\mathbb{R})$ je Borelovská σ -algebra nad \mathbb{R} .

c) Je dána funkce $f: \mathbb{R} \rightarrow [0, \infty]$ taková, že $\int_{\mathbb{R}} f(x)dx = 1$. Pak pravděpodobnost libovolného jevu $A \in \mathcal{A}$ je dána jednoznačně vztahem

$$P(A) = \int_A f(x)dx.$$

Kapitola 2

Podmíněná pravděpodobnost, nezávislost

2.1 Podmíněná pravděpodobnost

Definice 2.1 *Nechť je dán pravděpodobnostní prostor (Ω, \mathcal{A}, P) a náhodné jevy A, B , kde $P(B) > 0$. Podmíněnou pravděpodobnost jevu A za podmínky,*

že nastal jev B , definujeme vztahem

$$P(A|B) = \frac{P(A \cap B)}{P(B)}. \quad (2.1)$$

Věta 2.1 *Nechť je dán pravděpodobnostní prostor (Ω, \mathcal{A}, P) a náhodný jev B , kde $P(B) > 0$. Potom pro libovolný jev $A \in \mathcal{A}$ platí:*

a) $P(A|B) \geq 0$,

b) $P(\Omega|B) = 1$,

c) $P(\cup_{n=1}^{\infty} A_n|B) = \sum_{n=1}^{\infty} P(A_n|B)$ pro každou posloupnost $\{A_n\}$ disjunktních jevů.

Poznámka 2.1 *Věta 2.1 říká, že podmíněná pravděpodobnost má všechny základní vlastnosti pravděpodobnosti nepodmíněné (definice 2.2), a tudíž je to také pravděpodobnost.*

Věta 2.2 (o násobení pravděpodobnosti):

Pro libovolnou posloupnost náhodných jevů A_1, A_2, \dots, A_n , takových, že $P(A_1 \cap A_2 \cap \dots \cap A_{n-1}) > 0$, platí

$$P(\cap_{i=1}^n A_i) = P(A_1)P(A_2|A_1)P(A_3|A_1 \cap A_2) \dots \quad (2.2) \\ \dots P(A_n|A_1 \cap A_2 \cap \dots \cap A_{n-1}).$$

Příklad 2.1 *Z urny, ve které je n bílých a n černých kuliček, náhodně vybereme n -krát po dvou kuličkách bez vrácení vytažených kuliček. Určete, jaká je pravděpodobnost, že vždy vytáhneme jednu bílou a jednu černou kuličku.*

2.2 Nezávislost

Uvažujme nyní dva náhodné jevy A a B . Jestliže pro ně platí

$$P(A|B) = P(A) \quad a \quad P(B|A) = P(B), \quad (2.3)$$

pak mluvíme o jejich vzájemné nezávislosti. Z (2.3) vidíme, že pravděpodobnost jevu A podmíněná jevem B nezávisí na jevu B a naopak. Z (2.3) a z definice podmíněné pravděpodobnosti pak dostáváme následující definici nezávislosti dvou náhodných jevů.

Definice 2.2 *Náhodné jevy A a B jsou nezávislé, jestliže platí*

$$P(A \cap B) = P(A) \cdot P(B). \quad (2.4)$$

Pojem nezávislosti můžeme rozšířit i na skupinu náhodných jevů.

Definice 2.3 *Nechť A_1, A_2, \dots, A_n jsou náhodné jevy. Řekneme, že jsou skupinově (totálně) nezávislé, jestliže pro libovolnou posloupnost indexů $\{k_1, k_2, \dots, k_r\} \subset \{1, \dots, n\}$, $r = 2, \dots, n$ platí*

$$P(A_{k_1} \cap A_{k_2} \cap \dots \cap A_{k_r}) = P(A_{k_1}) \cdot P(A_{k_2}) \cdot \dots \cdot P(A_{k_r}). \quad (2.5)$$

Definice 2.4 *Nechť A_1, \dots, A_n jsou náhodné jevy. Řekneme, že jsou po dvou nezávislé, jestliže jevy A_i, A_j jsou nezávislé pro všechna $i, j = 1, \dots, n$, $i \neq j$.*

Příklad 2.2 Při hodu dvěma mincemi uvažujeme tyto náhodné jevy:

$A_1 \dots$ jev spočívající v tom, že na 1. minci padne rub,

$A_2 \dots$ jev spočívající v tom, že na 2. minci padne líc,

$A_3 \dots$ jev spočívající v tom, že na obou mincích padne rub, nebo líc.

Zjistěte, zda dané jevy jsou skupinově nezávislé.

Příklad 2.3 Profesor zapomene deštník při každé návštěvě obchodu s pravděpodobností $\frac{1}{4}$. Jestliže navštívil čtyři obchody a přišel domů bez deštníku, jaká je pravděpodobnost, že jej zapomněl v posledním obchodě?

Pravděpodobnosti nezávislých jevů násobíme.

Věta 2.3 Nechť A, B jsou nezávislé náhodné jevy. Pak dvojice jevů (A, \bar{B}) , (\bar{A}, B) , (\bar{A}, \bar{B}) jsou nezávislé.

Věta 2.4 Nechť A_1, \dots, A_n jsou skupinově (totálně) nezávislé jevy. Potom

platí následující

$$P(\cup_{i=1}^n A_i) = 1 - \prod_{i=1}^n [1 - P(A_i)]. \quad (2.6)$$

Příklad 2.4 *Během dne se v porodnici narodilo 10 dětí. Pravděpodobnost narození chlapce je $p = 0,514$. Jaká je pravděpodobnost, že během tohoto dne se narodil v porodnici alespoň jeden chlapec?*

Věta 2.5 *Náhodné jevy A, B , kde $0 < P(B) < 1$, jsou nezávislé právě tehdy, když*

$$P(A|B) = P(A|\bar{B}). \quad (2.7)$$

Kapitola 3

Celková pravděpodobnost, Bayesův vzorec

Věta 3.1 (O celkové pravděpodobnosti) *Nechť A_1, A_2, \dots jsou náhodné jevy tvořící rozklad jevu jistého, tzn.*

$$A_i \cap A_j = \emptyset, \quad \forall i \neq j \quad \text{a} \quad \bigcup_{i=1}^{\infty} A_i = \Omega.$$

Nechť tyto náhodné jevy mají postupně pravděpodobnosti $P(A_1), P(A_2), \dots$, přičemž $P(A_i) > 0, \forall i = 1, 2, \dots$. Uvažujme libovolný náhodný jev B , u něhož

známe podmíněné pravděpodobnosti

$$P(B|A_i), \quad \forall i = 1, 2, \dots$$

Potom

$$P(B) = \sum_{i=1}^{\infty} P(A_i) \cdot P(B|A_i). \quad (3.1)$$

Příklad 3.1 *Ve městě jsou tři obchodní společnosti. Pod první obchodní společností spadá 20 obchodů, pod druhou 15 obchodů a pod třetí 10 obchodů. Při návštěvě obchodu první společnosti budete ošizen s pravděpodobností 0,15, v obchodě druhé společnosti 0,08 a u třetí společnosti s pravděpodobností 0,05. Jaká je pravděpodobnost, že při nákupu v tomto městě budete ošizen?*

Věta 3.2 (Bayesova věta) *Nechť jsou splněny předpoklady věty 3.1. Pak*

$$P(A_i|B) = \frac{P(B|A_i) \cdot P(A_i)}{\sum_{j=1}^{\infty} P(A_j) \cdot P(B|A_j)}, \quad i = 1, 2, \dots \quad (3.2)$$

Příklad 3.2 Při vyšetřování pacienta je podezření na 3 navzájem se vylučující nemoci. Pravděpodobnost výskytu první nemoci je 0,3, druhé 0,5 a třetí nemoci 0,2. Laboratorní zkouška dává pozitivní výsledek u 15% nemocných na první nemoc, u 30% nemocných na druhou nemoc a u 30% na třetí nemoc. Jaká je pravděpodobnost výskytu druhé nemoci, jestliže po vykonání laboratorní zkoušky je výsledek pozitivní?

Poznámka 3.1 Pravděpodobnosti $P(A_1), P(A_2), \dots$ v (4.4) se nazývají apriorní a jevy A_1, A_2, \dots se nazývají hypotézami. Pravděpodobnosti $P(A_i|B)$ nazývají aposteriorní.

Příklad 3.3 (AIDS).

Krevní test na pozitivní virus HIV nemusí vždy správně identifikovat chorobu. Mohou nastat dva druhy chyb.

1. Test špatně indikuje pozitivitu,

2. test špatně indikuje negativitu.

Statistickým pozorováním bylo zjištěno, že tento test je velmi spolehlivý, přesněji, je-li objekt infikován, bude test pozitivní s pravděpodobností 0,995. Neboli $P(\text{Poz}|\text{Inf}) = 0,995$, odtud dostáváme, že pravděpodobnost chyby 1. druhu je $P(\text{Neg}|\text{Inf}) = 0,005$. Podobně $P(\text{Neg}|\text{NeInf}) = 0,995$ a pravděpodobnost 2. chyby je $P(\text{Poz}|\text{NeInf}) = 0,005$.

Předpokládejme, že bude vydán zákon, který nařídí všem lidem provést tento "přesný" test, aby mohli být identifikováni všichni infikovaní lidé. Jestliže pak náhodně vybereme jednoho člověka s pozitivním výsledkem testu, ptáme se, jaká je pravděpodobnost, že skutečně má HIV. Neboli zajímá nás pravděpodobnost $P(\text{Inf}|\text{Poz})$, kterou určíme podle Bayesova vzorce:

$$P(\text{Inf}|\text{Poz}) = \frac{P(\text{Poz}|\text{Inf}) \cdot P(\text{Inf})}{P(\text{Poz}|\text{Inf}) \cdot P(\text{Inf}) + P(\text{Poz}|\text{NeInf}) \cdot P(\text{NeInf})}.$$

Zbývá nám určit apriorní pravděpodobnosti $P(\text{Inf})$, $P(\text{NeInf})$. Např. v USA

v roce 1996 bylo 293.433 infikovaných lidí, což vede k odhadům

$$P(Inf) = 0,001 \text{ a } P(NeInf) = 0,999.$$

Po dosazení dostáváme

$$P(Inf|Poz) = \frac{0,995 \cdot 0,001}{0,995 \cdot 0,001 + 0,005 \cdot 0,999} = 0,16.$$

Můžeme tedy mluvit o štěstí, že takový zákon nebyl nikdy vytvořen, protože pouze 16% pozitivních lidí by bylo skutečně infikovaných HIV.

Kapitola 4

Náhodná veličina

4.1 Definice náhodné veličiny

Definice 4.1 *Nechť (Ω, \mathcal{A}, P) je pravděpodobnostní prostor. Reálnou funkci X definovanou na Ω nazýváme náhodnou veličinou, jestliže X je měřitelné zobrazení $X : (\Omega, \mathcal{A}) \rightarrow (\mathbb{R}, \mathcal{B})$, tj.*

$$\{\omega \in \Omega : X(\omega) \in B\} \in \mathcal{A} \tag{4.1}$$

pro libovolnou borelovskou množinu $B \in \mathcal{B}$ (\mathcal{B} je σ -algebra borelovských podmnožin, tj. nejmenší σ -algebra obsahující systém všech otevřených podmnožin \mathbb{R}).

Poznámka 4.1 Náhodné veličiny budeme značit velkými písmeny: $X, Y, Z \dots$. Hodnoty, kterých mohou náhodné veličiny nabývat, budeme značit malými písmeny x, y, z .

Místo $\{\omega \in \Omega : X(\omega) \in B\}$ budeme zjednodušeně psát $\{X \in B\}$ a místo $\{\omega \in \Omega : X(\omega) < x\}$ budeme zjednodušeně psát $\{X < x\}$.

Poznámka 4.2 Součty, součiny a podíly náhodných veličin jsou náhodné veličiny, umocnění náhodné veličiny přirozeným číslem, násobení náhodné veličiny skalárem jsou opět náhodné veličiny.

Příklad 4.1 Nechť $\Omega = \{1, 2, 3, 4, 5, 6\}$ představuje prostor všech výsledků náhodného hodu kostkou. Za σ -algebru \mathcal{A} vezmeme systém všech podmnožin Ω .

Pravděpodobnost $P(A)$ náhodného jevu $A \in \mathcal{A}$ je rovna poměru příznivých jevů ke všem jevům. Toto odpovídá klasickému pravděpodobnostnímu prostoru. Nyní na tomto prostoru (Ω, \mathcal{A}, P) zkonstruujeme náhodnou veličinu X , která má hodnotu 1, padne-li 6, a hodnotu 0, padne-li něco jiného. Neboli $X(6) = 1$ a $X(i) = 0$; $i = 1, \dots, 5$.

$$X : (\{1, 2, 3, 4, 5, 6\}, \mathcal{A}, P) \rightarrow \{0, 1\},$$

$$P(\omega = 5) = P(5) = 1/6,$$

$$P(\omega \in \Omega : X(\omega) = 0) = P(X = 0) = 5/6, \quad P(\omega \in \Omega : X(\omega) = 1) = P(X = 1) = 1/6.$$

$$P(X < 2) = P(X = 0) + P(X = 1) = 1.$$

4.2 Distribuční funkce

Definice 4.2 *Nechť X je náhodná veličina. Její distribuční funkcí nazýváme reálnou funkci F_X reálné proměnné x definovanou*

$$F_X(x) = P(X < x) = P(\{\omega : X(\omega) < x\}). \quad (4.2)$$

Definice distribuční funkce má smysl, neboť z definice náhodné veličiny X víme, že $\{\omega : X(\omega) < x\} \in \mathcal{A}$, tzn. $\{\omega : X(\omega) < x\}$ je náhodný jev a každému náhodnému jevu můžeme přiřadit pravděpodobnost. Distribuční funkce je definovaná pro všechna $x \in \mathbb{R}$.

Příklad 4.2 *Distribuční funkce F náhodné veličiny definované v předchozím příkladu je pak definována takto:*

$$F(x) = 0, \text{ pokud } x \leq 0,$$

$$F(x) = 5/6, \text{ pokud } 0 < x \leq 1 \text{ a}$$

$$F(x) = 1, \text{ pokud } x > 1.$$

Distribuční funkce mají určité společné vlastnosti.

Věta 4.1 (vlastnosti distribuční funkce) *Distribuční funkce $F_X(x)$ náhodné veličiny X je*

a) *neklesající, tj. pro libovolné $a, b \in \mathbb{R}, a \leq b$, platí $F_X(a) \leq F_X(b)$,*

b) *zleva spojitá v libovolném bodě $x \in \mathbb{R}$,*

c) $\lim_{x \rightarrow -\infty} F_X(x) = 0, \lim_{x \rightarrow \infty} F_X(x) = 1$,

d) *má nejvýše spočetně bodů nespojitosti; tyto body nespojitosti jsou 1. druhu (tj. skoky) a velikost skoku v bodě x_0 je*

$$F_X(x_0^+) - F_X(x_0) = P(\{\omega \in \Omega : X(\omega) = x_0\}),$$

kde $F_X(x_0^+) = \lim_{x \rightarrow x_0^+} F_X(x)$.

Poznámka 4.3 *Je-li distribuční funkce $F_X(x)$ spojitá v bodě x_0 , pak velikost skoku v bodě x_0 je rovna nule tzn. $P(\{\omega : X(\omega) = x\}) = 0$.*

Tyto úvahy nás vedou k rozdělení náhodných veličin na dva základní typy, na diskrétní a absolutně spojité náhodné veličiny. Distribuční funkce diskrétní náhodné veličiny je konstantní až na spočetně mnoho bodů, ve kterých má skok. Distribuční funkce absolutně spojité náhodné veličiny je spojitá, a tudíž neobsahuje žádné skoky.

4.3 Diskrétní náhodné veličiny

Definice 4.3 *Náhodná veličina X se nazývá diskrétní, jestliže existuje posloupnost reálných čísel $\{x_n\}$ a odpovídající posloupnost nezáporných čísel $\{p_n\}$ taková, že*

$$\sum_{n=1}^{\infty} p_n = 1, \text{ kde } p_n = P(X = x_n). \quad (4.3)$$

Distribuční funkce diskrétní náhodné veličiny X má tvar

$$F_X(x) = P(X < x) = \sum_{\{n: x_n < x\}} P(X = x_n) = \sum_{\{n: x_n < x\}} p_n \quad (4.4)$$

a

$$P(a \leq X < b) = F_X(b) - F_X(a) = \sum_{\{n:a \leq x_n < b\}} P(X = x_n) = \sum_{\{n:a \leq x_n < b\}} p_n$$

pro libovolná reálná čísla a, b , kde $a \leq b$.

Poznámka 4.4 *Distribuční funkce je schodovitá funkce se skoky v bodech x_1, x_2, \dots a je konstantní na intervalech $(x_n, x_{n+1}]$. Velikost skoku v bodě x_n je $p_n = P(X = x_n)$.*

Příklad 4.3 *Uvažujme náhodnou veličinu X , jejíž hodnota udává počet telefonních výzev za 1 minutu. Distribuční funkce F ani pravděpodobnosti $\{p_n\}$ nejsou známy. Sledovali jsme 60 realizací této náhodné veličiny a zaznamenali*

výsledky.

3,	2,	2,	3,	1,	1,	0,	4,	2,	1
1,	4,	0,	1,	2,	3,	1,	2,	5,	2
3,	0,	2,	4,	1,	2,	3,	0,	1,	2
1,	3,	1,	2,	0,	7,	3,	2,	1,	1
4,	0,	0,	1,	4,	2,	3,	2,	1,	3
2,	2,	3,	1,	4,	0,	2,	1,	1,	5.

Jednotlivé realizace náhodné veličiny X jsou nezávislé, máme tedy k dispozici náhodný výběr (tj. X_1, \dots, X_{60} jsou nezávislé stejně rozdělené náhodné veličiny s distribuční funkcí F).

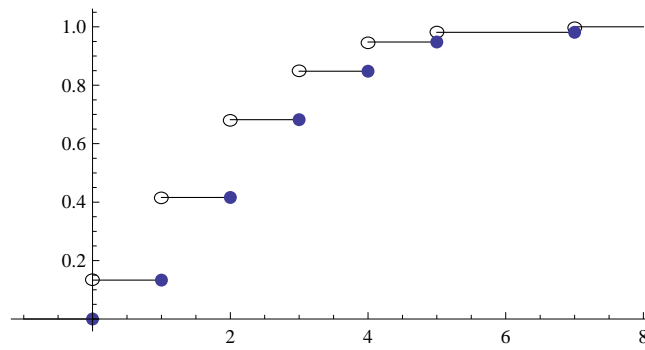
Vytvořme si tabulku absolutních a relativních četností výskytů jednotlivých

výsledků.

<i>Počet telefonních výzev za 1 min</i>	<i>Absolutní četnost</i>	<i>Relativní četnost</i>
<i>0</i>	<i>8</i>	<i>0,133</i>
<i>1</i>	<i>17</i>	<i>0,283</i>
<i>2</i>	<i>16</i>	<i>0,266</i>
<i>3</i>	<i>10</i>	<i>0,166</i>
<i>4</i>	<i>6</i>	<i>0,1</i>
<i>5</i>	<i>2</i>	<i>0,033</i>
<i>7</i>	<i>1</i>	<i>0,016</i>
<i>Celkem</i>	<i>60</i>	<i>1</i>

Relativní četnosti nám odhadují pravděpodobnosti p_n . Vzhledem k zákonu velkých čísel (viz věta 12.4) je tento odhad vhodný. Vezměme tedy tyto p_n jako skutečné pravděpodobnosti $p_n = P(X = n)$. Můžeme pak zakreslit distribuční funkci F .

Někdy se místo zobrazení distribuční funkce používá zobrazení relativních četností neboli histogram.

Obrázek 4.1: Distribuční funkce $F_X(x)$.

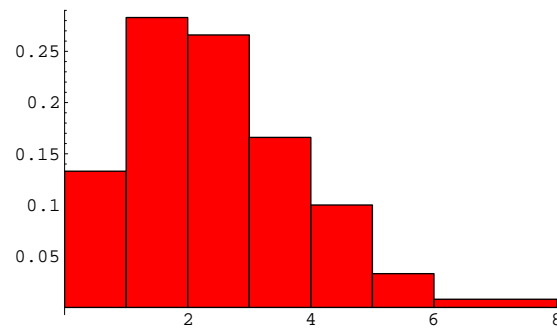
4.4 Absolutně spojitě náhodné veličiny

Definice 4.4 Náhodná veličina X se nazývá *absolutně spojitá*, jestliže existuje nezáporná integrovatelná funkce f_X taková, že platí

$$F_X(x) = P(X < x) = \int_{-\infty}^x f_X(t) dt, \quad x \in (-\infty, \infty). \quad (4.5)$$

Funkce f_X se nazývá *hustotou rozdělení pravděpodobnosti*.

Poznámka 4.5 Místo " $P[X \text{ má vlastnost } V] = 1$ " budeme říkat " X má vlast-

Obrázek 4.2: Histogram X .

nost V skoro jistě.” Často budeme užívat zkratku s.j.

Věta 4.2 (Vlastnosti hustoty) *Nechť $f_X(x)$ je hustota rozdělení pravděpodobnosti náhodné veličiny X . Pak platí:*

a) $f_X(x) = \frac{d}{dx}F_X(x)$ s. j.

b) $\int_{-\infty}^{\infty} f_X(x)dx = 1$

c) $P(a \leq X < b) = F_X(b) - F_X(a) = \int_a^b f_X(x)dx$ pro libovolná reálná čísla

a, b , kde $a \leq b$.

Příklad 4.4 *Uvažujeme hypotetickou populaci ryb. Je známo, že funkce umírání ryb závisí na kvadrátu délky života a že žádná ryba se nedožije více než 10 let. Neboli $F(x)$ je dána vztahem*

$$F(x) = \begin{cases} 0 & x \leq 0 \\ (c \cdot x)^2 & 0 < x \leq 10 \\ 1 & x > 10 \end{cases}$$

- a) určeme konstantu c tak, aby $F(x)$ byla distribuční funkce,
- b) spočtíme hustotu umírání v rybí populaci,
- c) spočtíme pravděpodobnost, že ryba zemře mezi 3. a 4. rokem života.

Příklad 4.5 Určete koeficient c tak, aby funkce

$$f(x) = \begin{cases} c \cdot x^2 \cdot e^x & 0 \leq x \leq 1 \\ 0 & \text{jinde} \end{cases}$$

byla hustotou nějaké náhodné veličiny.

Kapitola 5

Charakteristiky náhodných veličin

Definice 5.1 a) Necht' X je diskrétní náhodná veličina nabývající reálných hodnot x_1, x_2, x_3, \dots , tzn. taková, že $P(X = x_i) = p_i$. Pak střední hodnota $\mathbb{E}X$ náhodné veličiny X je tvaru

$$\mathbb{E}X = \sum_{i=1}^{\infty} x_i \cdot p_i, \quad (5.1)$$

pokud řada v (5.1) konverguje.

b) Nechť X je absolutně spojitá náhodná veličina s hustotou f_X . Pak střední hodnota náhodné veličiny X je

$$\mathbb{E}X = \int_{-\infty}^{\infty} x f_X(x) dx, \quad (5.2)$$

pokud integrál existuje.

Vlastnosti střední hodnoty. Nechť $X, Y, X_n, n = 1, 2, \dots$ jsou náhodné veličiny na pravděpodobnostním prostoru (Ω, \mathcal{A}, P) , a, b jsou reálné konstanty.

1) *střední hodnota konstanty je konstanta*

$$\mathbb{E}a = a$$

2) *absolutní integrovatelnost*

$$\mathbb{E}X < \infty \Leftrightarrow \mathbb{E}|X| < \infty$$

3) *linearita*

$$\mathbb{E}(aX + bY) = a\mathbb{E}X + b\mathbb{E}Y$$

4)

$$X \geq 0 \text{ s.j.} \Rightarrow \mathbb{E}X \geq 0$$

5) *monotonie*

$$X_1 \leq X \leq X_2 \text{ s.j.} \Rightarrow \mathbb{E}X_1 \leq \mathbb{E}X \leq \mathbb{E}X_2$$

6)

$$|\mathbb{E}X| \leq \mathbb{E}|X|$$

7)

$$|X| \leq Y \text{ s.j.}, \mathbb{E}Y < \infty \Rightarrow \mathbb{E}X < \infty$$

8) *integrace člen po členu*

$$\sum_{n=1}^{\infty} \mathbb{E}|X_n| < \infty \Rightarrow \mathbb{E}\left(\sum_{n=1}^{\infty} X_n\right) = \sum_{n=1}^{\infty} \mathbb{E}X_n$$

9) *Fatouovo lemma*

$$X_n \geq 0 \text{ s.j.} \Rightarrow \mathbb{E}\left(\liminf_{n \rightarrow \infty} X\right) \leq \liminf_{n \rightarrow \infty} \mathbb{E}X$$

Věta 5.1 *Nechť X je náhodná veličina a nechť $\phi : \mathbb{R} \rightarrow \mathbb{R}$. Pak platí:*

Má-li náhodná veličina X diskrétní rozdělení $\{x_n, p_n\}_{n \in \mathbb{N}_0}$, pak

$$\mathbb{E}\phi(X) = \sum_{n \in \mathbb{N}_0} \phi(x_n)p_n, \quad (5.3)$$

pokud jedna ze stran rovnosti existuje.

Má-li náhodná veličina X absolutně spojitě rozdělení s hustotou f , potom

$$\mathbb{E}\phi(X) = \int_{-\infty}^{\infty} \phi(x)f(x)dx, \quad (5.4)$$

pokud jeden z integrálů existuje.

Definice 5.2 *Nechť n je přirozené číslo, n -tý moment náhodné veličiny X je definován jako $\mathbb{E}(X^n)$; n -tý absolutní moment jako $\mathbb{E}(|X|^n)$; n -tý centrální moment jako $\mathbb{E}[(X - \mathbb{E}X)^n]$.*

Poznámka 5.1

- a) Z předešlé definice vidíme, že střední hodnota je první moment.
- b) První centrální moment je vždy roven nule, neboť

$$\mathbb{E}(X - \mathbb{E}X) = \mathbb{E}X - \mathbb{E}(\mathbb{E}X) = \mathbb{E}X - \mathbb{E}X = 0.$$

Definice 5.3 Druhý centrální moment náhodné veličiny X se nazývá rozptyl, označuje se obvykle $\text{var } X$ (z anglického "variance")

$$\text{var } X = \mathbb{E}(X - \mathbb{E}X)^2.$$

Rozptyl je druhou nejdůležitější charakteristikou náhodné veličiny. Z jeho definice vidíme, že existence střední hodnoty je nutnou podmínkou k existenci rozptylu. Číslo $\text{var } X$ je vždy nezáporné a rovná se nule právě tehdy, když $P(X = c) = 1$, c je konstanta. Vzhledem k tomu, že rozptyl udává variabilitu náhodné veličiny ve čtvercích jejích jednotek, používá se také často druhé odmocniny z rozptylu, tzv. *směrodatné odchytky*

$$\sigma = \sqrt{\text{var } X},$$

která měří variabilitu v původních jednotkách náhodné veličiny.

Nejdůležitější vlastnosti rozptylu náhodné veličiny X .

1) Nechť X je náhodná veličina, pak $\text{var } X$ počítáme nejčastěji pomocí vzorce:

$$\text{var } X = \mathbb{E}(X^2) - (\mathbb{E}X)^2.$$

2) Nechť c je konstanta. Pak $\text{var } c = 0$.

3) Nechť X je náhodná veličina, a a je reálné číslo. Pak

$$\text{var } (aX) = a^2 \text{var } X.$$

4) Nechť X je náhodná veličina a c je konstanta. Pak

$$\text{var } (X + c) = \text{var } X.$$

5) Nechť X je náhodná veličina, která má konečnou střední hodnotu a konečný nenulový rozptyl. Nechť

$$Y = \frac{X - \mathbb{E}X}{\sqrt{\text{var } X}}.$$

Pak $\mathbb{E}Y = 0$ a $\text{var } Y = 1$.

Lemma 5.2 *Nechť existuje n -tý moment náhodné veličiny X , $n > 0$. Pak pro libovolné $\varepsilon > 0$ platí:*

$$P[|X| \geq \varepsilon] \leq \frac{\mathbb{E}|X|^n}{\varepsilon^n}.$$

Věta 5.3 (Čebyševova nerovnost) *Nechť X je náhodná veličina s konečným rozptylem. Pak pro libovolné $\varepsilon > 0$ platí*

$$P[|X - \mathbb{E}X| \geq \varepsilon] \leq \frac{\text{var } X}{\varepsilon^2}.$$

Příklad 5.1 *Uvažujme absolutně spojitou náhodnou veličinu X , která má hustotu $f(x) = \frac{1}{\pi(1+x^2)}$ (Cauchyho rozdělení). Vypočtete $\mathbb{E}X$.*

V některých situacích, jako například v předchozím příkladu je vhodné používat k popisu rozdělení další charakteristiky, kterých je celá řada. Jednou z nich je tzv.

medián \tilde{x} . Je to číslo, pro které platí

$$P(X \leq \tilde{x}) \geq \frac{1}{2} \text{ a } P(X \geq \tilde{x}) \geq \frac{1}{2}.$$

Je nutné poznamenat, že medián není těmito podmínkami určen jednoznačně. Pro již zmíněné Cauchyho rozdělení má medián hodnotu $\tilde{x} = 0$. Další charakteristikou rozdělení může být *modus*, který se obvykle značí \hat{x} . Je-li diskrétní rozdělení soustředěno v bodech x_1, x_2, \dots , je \hat{x} ta hodnota, pro kterou platí

$$P(X = \hat{x}) \geq P(X = x_i), \quad \forall i = 1, 2, \dots$$

Je-li rozdělení absolutně spojitě, za modus bereme takovou hodnotu \hat{x} , pro kterou platí

$$f(\hat{x}) \geq f(x), \quad \forall x \in (-\infty, \infty).$$

Také modus nemusí být určen jednoznačně (najděte příklad).

Je-li F distribuční funkce, zavedme funkci F^{-1} předpisem

$$F^{-1}(u) = \inf\{x : F(x) \geq u\}, \quad 0 < u < 1.$$

Pak se F^{-1} nazývá *kvantilová funkce* odpovídající distribuční funkci F . Hodnotám funkce $F^{-1}(u)$ se říká *kvantily*. Tedy α -kvantilem budeme nazývat hodnotu $F^{-1}(\alpha)$. Pokud F je rostoucí a spojitá, pak kvantilová funkce je inverzní funkcí k F . Odtud pochází i označení F^{-1} . Kvantil $F^{-1}(0, 25)$, resp. $F^{-1}(0, 75)$ bývá zvykem nazývat *dolním*, resp. *horním kvantilem*. Kvantilové charakteristiky se používají zřídka a jsou užitečné zejména tehdy, kdy nelze užít momentů.

Příklad 5.2 *Podle úmrtnostních tabulek USA (1978 až 1979) je pravděpodobnost úmrtí 32 leté ženy během jednoho roku rovna 0,001819. Pojišťovna nabízí ženám tohoto věku, že při ročním pojistném 100 USD vyplatí pozůstalým v případě úmrtí pojištěnce 25 000 USD. Jaký zisk může pojišťovna očekávat, jestliže takovou pojistku uzavře 5 000 žen uvedeného věku?*

Příklad 5.3 *Označme dobu čekání rybáře na úlovek (v minutách) jako náhodnou veličinu X . Předpokládejme, že tato náhodná veličina má hustotu pravdě-*

podobnosti

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

Určete střední hodnotu a rozptyl doby čekání rybáře na úlovek.

Příklad 5.4 *Určete modus \hat{x} následujících náhodných veličin:*

1. diskrétní veličiny X s rozložením pravděpodobnosti

$$p_n = \begin{cases} (\frac{1}{2})^n & \text{pro } n = 1, 2, \dots \\ 0 & \text{jinak} \end{cases}$$

2. spojité náhodné veličiny s hustotou

$$f(x) = \frac{x^2 e^{-x}}{2}, \quad x \in (0, \infty), \quad f(x) = 0 \text{ jinde.}$$

Kapitola 6

Příklady diskrétních náhodných veličin

1. Nula - jedničkové (alternativní) rozdělení.

Tak budeme nazývat rozdělení náhodné veličiny X , která nabývá jen hodnot 0 a 1 s pravděpodobnostmi $1 - p$ a p . Číslo p se nazývá parametr alternativního rozdělení, $0 < p < 1$.

Distribuční funkce alternativního rozdělení je dána výrazem

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

Střední hodnota $\mathbb{E}X = p$. Rozptyl $\text{var } X = p(1 - p)$ (dokažte). Alternativní rozdělení s parametrem p budeme zkráceně označovat $A(p)$.

2. Binomické rozdělení.

Je to rozdělení náhodné veličiny X , která nabývá hodnot $k = 0, 1, 2, \dots, n$. Binomické rozdělení je jednoznačně určeno dvěma parametry: přirozeným číslem n a číslem $p \in (0, 1)$. Pro binomické rozdělení s parametry n, p budeme užívat zkráceného značení $Bi(n; p)$. Binomickým rozdělením se řídí např. náhodná veličina X , která je rovna počtu úspěchů v posloupnosti n nezávislých alternativních pokusů, kde pravděpodobnost úspěchu v každém pokusu je p , $0 < p < 1$.

Tedy

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

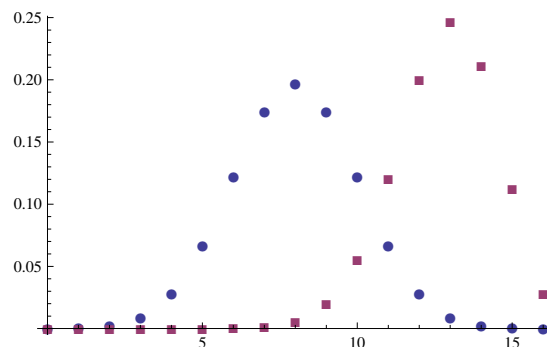
kde

$$X_i = \begin{cases} 1 & \text{pokud v } i\text{-tém pokuse nastal úspěch,} \\ 0 & \text{pokud úspěch nenastal.} \end{cases}$$

X je součtem n alternativních náhodných veličin. Vzhledem k nezávislosti X_i

je hledaná pravděpodobnost p_k tvaru

$$p_k = \binom{n}{k} p^k (1-p)^{n-k} \text{ pro } k = 0, 1, \dots, n.$$



Obrázek 6.1: Pravděpodobnosti p_k binomického rozdělení. $Bi(16; 0, 5)$ - kruhy, $Bi(16; 0, 8)$ - čtverce.

Pravděpodobnosti p_k splňují podmínky pro pravděpodobnostní rozdělení, neboť platí:

a) $p_k \geq 0, \forall k,$

b) $\sum_{k=0}^n p_k = \sum_{k=0}^n \binom{n}{k} p^k (1-p)^{n-k} = [(1-p) + p]^n = 1.$

Název binomické rozdělení vyplývá ze skutečnosti, že pravděpodobnost p_k je členem binomického rozvoje. Distribuční funkce $F(x)$ je tvaru

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

Rozdělení $Bi(n; p)$ má střední hodnotu np a rozptyl $np(1-p)$.

3. Poissonovo rozdělení

je rozdělení náhodné veličiny X , která nabývá hodnot $k = 0, 1, 2, \dots$ s pravděpodobnostmi

$$p_k = e^{-\lambda} \frac{\lambda^k}{k!}.$$

Číslo $\lambda > 0$ je parametr Poissonova rozdělení. Vidíme, že pro takto definované pravděpodobnosti p_k jsou splněny podmínky

a) $p_k \geq 0, \forall k = 0, 1, 2, \dots,$

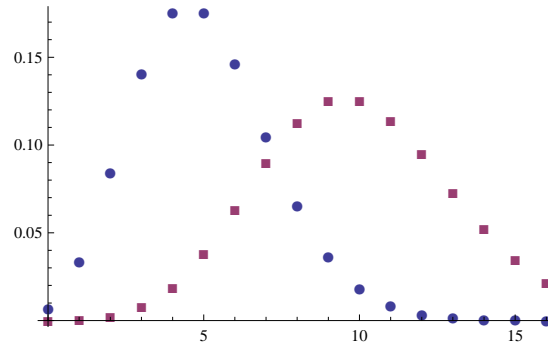
b) $\sum_{k=0}^{\infty} p_k = \sum_{k=0}^{\infty} \frac{\lambda^k e^{-\lambda}}{k!} = e^{-\lambda} \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} = 1,$

a tedy p_k je rozdělení pravděpodobnosti.

Distribuční funkce je tvaru

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

$$\mathbb{E}X = \lambda.$$



Obrázek 6.2: Pravděpodobnosti p_k Poissonova rozdělení. $Po(5)$ - kruhy, $Po(10)$ - čtverce.

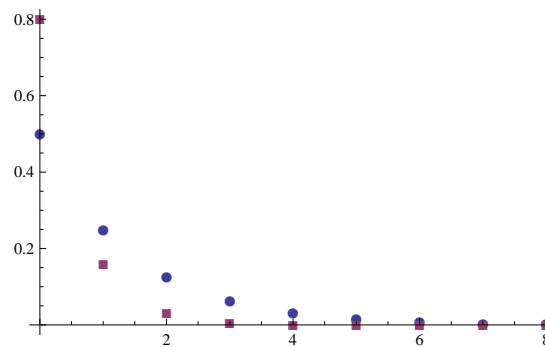
$$\text{var } X = \lambda.$$

Poissonovo rozdělení je limitním případem binomického rozdělení pro $n \rightarrow \infty, p \rightarrow 0, np \rightarrow \lambda$ (=konstanta).

4. Geometrické rozdělení

je rozdělení náhodné veličiny X , která nabývá hodnot $k = 0, 1, 2, \dots$ s pravděpodobnostmi $p_k = p(1 - p)^k$. Parametr p je z intervalu $(0,1)$. Je zřejmé, že všechna $p_k \geq 0$ a $\sum_{k=0}^{\infty} p_k = 1$, neboť

$$\sum_{k=0}^{\infty} p_k = \sum_{k=0}^{\infty} p(1 - p)^k = p \sum_{k=0}^{\infty} (1 - p)^k = p \frac{1}{1 - (1 - p)} = 1.$$



Obrázek 6.3: Pravděpodobnosti p_k geometrického rozdělení. $Geom(0, 5)$ - kruhy, $Geom(0, 8)$ - čtverce.

Distribuční funkce geometrického rozdělení je tvaru

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

$$\mathbb{E}X = \frac{1-p}{p}.$$

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

Příklad 6.1 *Jaká je pravděpodobnost, že mezi čtyřmi po sobě narozenými dětmi budou*

- a) *první dva chlapci, další dvě dívky*
- b) *právě dva chlapci, víme-li, že pravděpodobnost narození chlapce je 0,515?*
- c) *Zjistěte, kolik se musí narodit dětí, aby pravděpodobnost, že mezi nimi bude alespoň jeden chlapec, byla větší nebo rovna 0,99.*

Příklad 6.2 *Víme, že pravděpodobnost vypěstování zdravé sazenice ze semena je 0,62. Za náhodnou veličinu X budeme považovat počet zdravých sazenic vypěstovaných z 27 semen. Určete*

- a) *jaký je nejpravděpodobnější počet zdravých rostlin a jaká je jeho pravděpodobnost,*
- b) *střední hodnotu a rozptyl náhodné veličiny X .*

Příklad 6.3 Předpokládejme, že mladá dravá ryba přežije, pokud uloví rybu alespoň jednou za dva dny. Během dvou dnů podnikne 8 zápasů s pravděpodobností ulovení $p = 0,25$. Jaká je pravděpodobnost, že dravá ryba nezemře?

Příklad 6.4 Na nádraží mají být instalovány automaty na prodej jízdenek, které po vhození příslušné mince vydají během 10 sekund žádanou jízdenku. Předpokládejme, že v době největší frekvence bude chtít použít automat v průměru 6 osob za minutu. Kolik automatů je nutné instalovat, aby s pravděpodobností větší než 0,95 byl v době největší frekvence obslužen každý zájemce okamžitě?

Kapitola 7

Příklady spojitých náhodných veličin

1. Rovnoměrné rozdělení na intervalu $[a, b]$

je dáno hustotou

$$f(x) = \begin{cases} \frac{1}{b-a} & a \leq x \leq b, \\ 0 & x < a, \quad x > b. \end{cases}$$

Distribuční funkce je

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

Střední hodnota a rozptyl jsou:

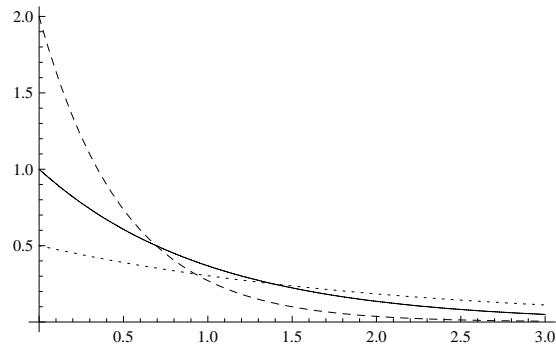
$$\mathbb{E}X = \frac{a+b}{2}, \quad \text{var } X = \frac{1}{12}(b-a)^2,$$

dokažte. Toto rozdělení budeme označovat $U(a, b)$ (z angl. "uniform").

2. Exponenciální rozdělení.

Hustota pravděpodobnosti exponenciálního rozdělení je

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



Obrázek 7.1: Graf hustoty exponenciálního rozdělení - plná čára $Exp(1)$, čárkovaná $Exp(1/2)$, čerchovaná $Exp(2)$.

Distribuční funkce je

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

kde λ je parametrem rozdělení. Ověříme nejprve, zda $f(x)$ je hustota. Vidíme, že $f(x) \geq 0$, $\forall x \in \mathbb{R}$.

$$\int_{-\infty}^{\infty} f(x) dx = \int_0^{\infty} \frac{1}{\lambda} e^{-\frac{x}{\lambda}} dx = 1.$$

Střední hodnota $\mathbb{E}X = \lambda$ a rozptyl $\text{var } X = \lambda^2$.

Věta 7.1 *Má-li náhodná veličina X exponenciální rozdělení, pak*

$$P(X > x + y | X > y) = P(X > x), \quad \forall x > 0, y > 0.$$

2. Normované normální rozdělení

je definováno hustotou

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2}, \quad -\infty < x < \infty.$$

Jeho distribuční funkce se tradičně značí písmenem Φ .

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{t^2}{2}} dt, \quad -\infty < x < \infty.$$

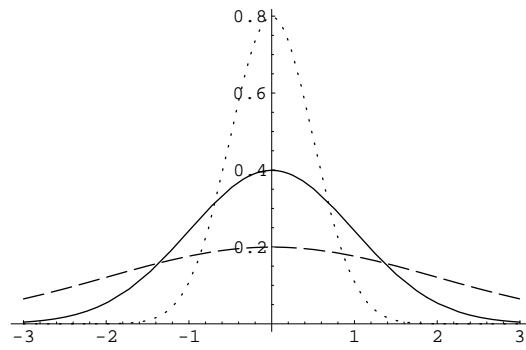
Střední hodnota $\mathbb{E}X = 0$ je zároveň mediánem i modusem tohoto rozdělení.
 $\text{var } X = 1$.

3. (Obecné) normální rozdělení.

Toto rozdělení je definováno hustotou

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

kde μ reálné a σ^2 kladné jsou parametry.



Obrázek 7.2: Graf hustoty normálního rozdělení - plná čára $N(0, 1)$, čárkovaná $N(0, 2)$, tečkovaná $N(0, 1/2)$.

Distribuční funkci

$$F(x) = \frac{1}{\sqrt{2\pi} \sigma} \int_{-\infty}^x e^{-\frac{(t-\mu)^2}{2\sigma^2}} dt, \quad -\infty < x < \infty \quad (7.1)$$

lze vyjádřit pomocí funkce Φ jako $\Phi\left(\frac{x-\mu}{\sigma}\right)$.

Příklad 7.1 *Prodejna očekává dodávku zboží v určitý den v době od 12 do 16 hodin. Podle sdělení dodavatele je uskutečnění dodávky stejně možné kdykoliv během tohoto časového intervalu. Jaká je pravděpodobnost, že zboží bude dodáno v době od jedné hodiny do půl druhé?*

Příklad 7.2 *Autobusy městské dopravy odjíždějí ze stanice v sedmiminutových intervalech. Cestující může přijít na stanici v libovolném okamžiku. Jaká je střední hodnota a rozptyl doby jeho čekání na odjezd autobusu ze stanice?*

Příklad 7.3 *Z dlouhodobých měření je známo, že radiomagnetofon Sony má poruchu v průměru jednou za 10 000 hodin. Předpokládejme, že "doba čekání na poruchu" je náhodná veličina X s exponenciálním rozdělením. Stanovme hodnotu t tak, aby pravděpodobnost, že radiomagnetofon bude pracovat delší dobu než t , byla 0,99.*

Příklad 7.4 *Do obchodu přijde průměrně 60 zákazníků za 1 hodinu. Jaká je pravděpodobnost, že do obchodu nepřijde žádný zákazník během $\frac{1}{2}$ min.,*

ve které je prodavač nepřítomen.

Příklad 7.5 *Víme, že populace určitého druhu květin dorůstá výšky X s normálním rozdělením $N(20, 16)$. Spočtěte pravděpodobnost, že náhodně vybraná květina má výšku*

- a) menší než 16,*
- b) větší než 20,*
- c) v mezích od 12 do 28,*
- d) menší než 12 nebo větší než 28,*
- e) rovnu 22.*

Kapitola 8

Zpracování statistického materiálu

8.1 Rozložení četností a jejich znázornění

Definice 8.1 *Mějme soubor dat o rozsahu n : x_1, \dots, x_n . Nechť a je minimální hodnota (popř. infimum) , b je maximální hodnota (popř. suprémum), tj. $x_{min} = a$, $x_{max} = b$.*

1. *Interval* $\langle a, b \rangle$ nazýváme **variačním oborem**
2. *Rozdíl* $x = b - a$ nazýváme **variačním rozpětím**.
3. *Variační obor* $\langle a, b \rangle$ rozkládáme na menší části, nazývané **třídy** (popř. *třídní intervaly*).
4. **Šířkou (délkou) h třídy příslušného třídního intervalu** $\langle a, b \rangle$ nazýváme číslo $h = b_k - a_k$. Číslo $\frac{1}{2} (a_k + b_k)$ nazýváme *středem třídy*, číslo a_k *dolní hranicí uvažované třídy*, číslo b_k *horní hranicí uvažované třídy*.
5. *Hodnotu x_k argumentu X , která je zpravidla dána středem k -té třídy a zastupuje všechny hodnoty patřící do této třídy, nazýváme* **třídním znakem k -té třídy**.

Při rozkladu variačního oboru $\langle a, b \rangle$ v třídy budeme dbát zpravidla těchto zásad:

1. Obsahuje-li soubor jen malý počet různých hodnot, volíme každou hodnotu x_k za samostatnou třídu. Pokud statistický soubor má značně velký počet různých

hodnot x_k (popř. je jich nekonečně mnoho), sdružujeme hodnoty argumentu v třídy. Přitom šířky tříd volíme obvykle stejně velké. Pro výpočet šířky h lze použít přibližného vzorce $h \approx \frac{8}{100} \cdot (b - a)$.

Při volbě počtu třídních intervalů se doporučuje, aby jich bylo 8 až 20. Záleží na rozsahu souboru a účelu statistické tabulky. Počet k třídních intervalů volíme např. $k \approx 3,3 \log(n)$ nebo $k \approx \sqrt{n}$. Dvě pozorování považujeme za ekvivalentní, jakmile padnou do téhož třídního intervalu.

2. Jestliže na hranici dvou sousedních tříd padne více hodnot argumentu, zařazujeme polovinu z nich do nižší třídy a druhou polovinu do třídy vyšší. Zbyla-li ještě jedna hodnota (toto odpovídá lichému počtu hodnot ležících na hranic), rozhodneme o její příslušnosti k dané třídě losem. Není vhodné zařazovat stereotypně takové hraniční hodnoty vždy do vyšší, popř. nižší třídy, neboť by se tím mohl zkreslit celkový obraz rozložení uvažovaného souboru ve prospěch vyšších, popř. nižších tříd.
4. Vyskytuje-li se v hraničních třídách velmi málo hodnot, je vhodné tyto třídy spojit se sousední třídou v třídu jedinou.

Definice 8.2 *Druhy četností:*

1. *Počet prvků souboru patřících do k -té třídy nazýváme **absolutní četností** argumentu v k -té třídě nebo **absolutní třídní četností** (stručně četností) k -té třídy. Značíme f_k .*
2. *Jeli f_k absolutní třídní četnost k -té třídy, n rozsah uvažovaného souboru, potom*
 - a) $\frac{f_k}{n}$ nazýváme **relativní četností** k -té třídy,
 - b) $100 * \frac{f_k}{n}$ nazýváme **procentní relativní četností** k -té třídy.
3. **Kumulativní (součtovou) absolutní četností** f_k k -té třídy nazýváme *součet všech četností f_k až do k -té třídy včetně, tj.*

$$F_k = \sum_{j=1}^k f_j.$$

4. Kumulativní relativní četnosti R_k k -té třídy nazýváme součet

$$R_k = \sum_{j=1}^k \frac{f_j}{n} = \frac{F_k}{n}.$$

Poznámka 8.1 Pro četnosti platí některé vlastnosti (uvažujeme statistický soubor rozsahu n , který je rozdělen do r tříd)

1.

$$\sum_{k=1}^r f_k = n$$

2.

$$F_r = n$$

3.

$$\sum_{k=1}^r \frac{f_k}{n} = 1$$

Definice 8.3 *Tabulkou rozložení četností daného statistického souboru nazýváme tabulku, v níž jsou uvedeny hodnoty s příslušnými absolutními, popř. relativními četnostmi.*

Příklad 8.1 *Na telefonní stanici zaznamenávali počet telefonních výzev za dobu 1 min. Během jedné hodiny bylo v určité denní době dosaženo těchto výsledků (v každém řádku jsou hodnoty získané během 10 minut):*

3,2,2,3,1,1,0,4,2,1
1,4,0,1,2,3,1,2,5,2
3,0,2,4,1,2,3,0,1,2
1,3,1,2,0,7,3,2,1,1
4,0,0,1,4,2,3,2,1,3
2,2,3,1,4,0,2,1,1,5.

Sestavte tabulku rozložení daného statistického souboru.

Počet telefonních výzev za 1 min	Absolutní četnost	Relativní četnost
0	8	0.133
1	17	0.283
2	16	0.266
3	10	0.166
4	6	0,1
5	2	0,033
7	1	0,016
Celkem	60	1

Tabulka 8.1: Tabulka rozložení četností

Argument statistického souboru představuje náhodnou veličinu X . Ze zákona velkých čísel plyne, že relativní četnost $\frac{f_k}{n}$ udává (přibližně) pravděpodobnost, že X padne do k -té třídy, takže platí $p_k = P(a_k \leq X \leq b_k) \approx \frac{f_k}{n}$, přičemž interval $\langle a_k, b_k \rangle$ je k -tou třídou.

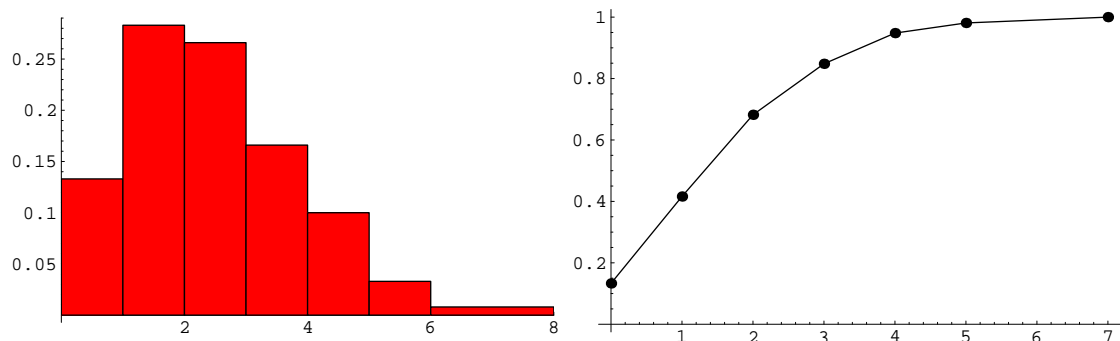
Definice 8.4 *Typy znázornění absolutních či relativních četností:*

1. Histogram rozložení absolutních (relativních) četností sestavíme tak, že na

osu x vyneseme středy jednotlivých tříd a nad každou úsečkou zobrazující určitou třídu (šířky h) sestrojíme pravoúhelník s výškou rovnou příslušné absolutní četnosti f_k , popř. relativní četnosti $\frac{f_k}{n}$. Horní obraz pravoúhelníka představuje histogram rozložení četností. Histogram relativních četností aproximuje hustotu rozdělení náhodné veličiny X .

2. **Úsečkový diagram** (nebo graf) rozložení absolutních (relativních) četností dostaneme, jestliže na ose x zobrazíme středy jednotlivých tříd a v každém z nich sestrojíme ve směru osy y úsečku o délce rovné příslušné absolutní četnosti f_k , popř. relativní četnosti $\frac{f_k}{n}$.
3. **Polygon** rozložení četností (spojnicový diagram) dostaneme, jestliže koncové body úsečkového diagramu rozložení četností spojíme úsečkami a vytvoříme tak lomenou čáru, která pak představuje hledaný polygon neboli spojnicový diagram.
4. Graf, polygon nebo histogram kumulativních četností dostaneme analogicky jako v bodech 1, 2 a 3.

5. Ogivní křivku (stručně **ogivu**) dostaneme, sestrojíme-li polygon kumulativních relativních četností. Ogiva aproximuje graf distribuční funkce uvažované náhodné veličiny X .



Obrázek 8.1: Histogram a ogiva dat z Příkladu 8.1

8.2 Charakteristiky polohy

Definice 8.5 Nechť je dán statistický soubor o hodnotách x_1, x_2, \dots, x_n , které jsou popř. roztrženy do r tříd, přičemž f_k značí absolutní četnost k -té třídy.

1. Aritmetický průměr \bar{X} je definován vztahy

$$\bar{X} = \frac{1}{n} \sum_{k=1}^n x_k = \frac{1}{n} \sum_{i=1}^r f_i x_i. \quad (8.1)$$

2. Geometrický průměr \bar{X}_g je definován vztahem

$$\bar{X}_g = \sqrt[n]{x_1 \cdot x_2 \cdot \dots \cdot x_n} \quad (8.2)$$

3. Harmonický průměr \bar{X}_h je definován vztahy

$$\bar{X}_h = \frac{1}{A}, \text{ kde } A = \frac{1}{n} \sum_{k=1}^n \frac{1}{x_k} = \frac{1}{n} \sum_{i=1}^r \frac{f_i}{x_i}. \quad (8.3)$$

Ve vztazích 8.1, 8.3 jsou uvedeny dva tvary. První tvar odpovídá souboru neroztříděnému a druhý tvar roztříděnému.

Věta 8.1

$$\bar{X}_h \leq \bar{X}_g \leq \bar{X}.$$

Nechť je dán statistický soubor o hodnotách x_1, x_2, \dots, x_n . Setřídíme-li hodnoty podle velikosti dostaneme tzv. setříděný statistický soubor

$$X_{(1)}, X_{(2)}, \dots, X_{(n)},$$

kde $X_{(1)}$ označuje nejmenší hodnotu, $X_{(2)}$ označuje druhou nejmenší hodnotu, ... Obecně $X_{(i)}$ označuje i -tou pořadovou hodnotu.

Definice 8.6 Medián je určen dvěma způsoby, v závislosti na počtu prvků statistického souboru. V případě lichého počtu hodnot vezmeme za medián \tilde{x} prostřední hodnotu

$$\tilde{x} = X_{(\lfloor \frac{n}{2} \rfloor + 1)}.$$

Pokud X má sudý počet hodnot, vezmeme za medián \tilde{x} aritmetický průměr prostředních dvou hodnot

$$\tilde{x} = \frac{X_{(\lfloor \frac{n}{2} \rfloor)} + X_{(\lfloor \frac{n}{2} \rfloor + 1)}}{2}.$$

Medián je speciálním případem *výběrového kvantilu*. Výběrovým kvantilem nazýváme hodnotu zvolenou tak, že pozorování, která jsou menší než tato hodnota, tvoří předepsaný díl výběru (např. 25% výběrový kvantil se nazývá dolní výběrový kvartil, 50% výběrový kvantil je medián a 75% výběrový kvantil se nazývá horní výběrový kvartil).

Definice 8.7 *Nechť argument statistického souboru může nabývat pouze konečně mnoha hodnot. Pak **Modus** je hodnota argumentu s největší absolutní četností. Modus nemusí být určen jednoznačně.*

Příklad 8.2 *Uvažujme následující hypotetický příklad. Ve firmě F existují 4 platové třídy, s platy uvedenými v následující tabulce. Počet zaměstnanců udává, kolik zaměstnanců je v dané platové třídě.*

Spočtěme některé charakteristiky polohy. Aritmetický průměr $\bar{X} = 13500$, geometrický průměr $\bar{X}_g = 12381.3$, harmonický průměr $\bar{X}_h = 11726.6$. Jelikož máme 44 hodnot, bude medián průměr 22. a 23. pořadové hodnoty, tedy

třída	zařazení	plat v Kč	počet zaměstnanců
1.	výkonná síla	10.000	30
2.	mistr	16.000	10
3.	náměstek	28.000	3
4.	ředitel	50.000	1

Tabulka 8.2: Tabulka četností příjmu zaměstnanců ve firmě F.

$\tilde{x} = 10.000$. Dolní výběrový kvartil bude průměr 11. a 12. pořadové hodnoty, tj. 10.000 a horní výběrový kvartil je 16.000.

Každá charakteristika polohy nám dává jen parciální informaci o statistickém souboru, zatímco grafy rozložení četností nám dávají úplnou informaci o statistickém souboru.

8.3 Charakteristiky variability

Definice 8.8 *Charakteristiky variability:*

1. Rozptylem (disperzí) *statického souboru s rozsahem n nazýváme aritmetický průměr kvadratických odchylek $(x_k - \bar{X})^2$ hodnot argumentu X od aritmetického průměru \bar{X}*

$$s^2 = \frac{1}{n} \sum_{k=1}^n (x_k - \bar{X})^2 = \frac{1}{n} \sum_{i=1}^r f_i (x_i - \bar{X})^2. \quad (8.4)$$

2. Směrodatnou odchylkou *nazýváme*

$$\sqrt{s^2} = s \geq 0. \quad (8.5)$$

3. Průměrnou odchylkou \bar{d} *nazýváme aritmetický průměr absolutních hodnot odchylek od aritmetického průměru \bar{X} , tj.*

$$\bar{d} = \frac{1}{n} \sum_{k=1}^n |x_k - \bar{X}| = \frac{1}{n} \sum_{i=1}^r f_i |x_i - \bar{X}|. \quad (8.6)$$

4. Variační koeficient v *statistického souboru je definován jako*

$$v = \frac{s}{\bar{X}}. \quad (8.7)$$

Poznámka 8.2 Rozptyl je definován vzorcem 8.4, pro jeho výpočet se však častěji používá vzorce

$$s^2 = \frac{1}{n} \sum_{k=1}^n (x_k^2) - \bar{X}^2 = \frac{1}{n} \sum_{i=1}^r f_i x_i^2 - \bar{X}^2. \quad (8.8)$$

Poznámka 8.3 Hodnoty statistického souboru jsou realizace nějaké náhodné veličiny. Např. Počet telefonních hovorů na ústředně za 1 minutu (viz. Příklad 8.1) je náhodná veličina, která má Poissonovo rozdělení $X \sim \text{Po}(\lambda)$. Všechny charakteristiky polohy aproximují střední hodnotu náhodné veličiny $\mathbb{E}X = \lambda$. Podobně rozptyl statistického souboru aproximuje rozptyl náhodné veličiny $\text{Var}X = \lambda$. Rozptyl uvedený ve vzorcích 8.4 a 8.8 rozptyl náhodné veličiny podhodnucuje, proto se k výpočtu rozptylu častěji používá vzorců:

$$S^2 = \frac{1}{n-1} \sum_{k=1}^n (x_k - \bar{X})^2 = \frac{1}{n-1} \sum_{i=1}^r f_i (x_i - \bar{X})^2, \quad (8.9)$$

$$S^2 = \frac{1}{n-1} \sum_{k=1}^n (x_k^2) - \frac{n}{n-1} \bar{X}^2 = \frac{1}{n-1} \sum_{i=1}^r f_i x_i^2 - \frac{n}{n-1} \bar{X}^2. \quad (8.10)$$

Tyto vzorce již teoretickou hodnotu nepodhodnocují.