Úvod do statistické analýzy

Ing. Josef Chudoba, Ph.D.
Josef.chudoba@tul.cz
Telefon: 48535 3844

Ústav nových technologií a aplikované informatiky
Fakulta mechatroniky, Technická univerzita v Liberci

Verze 1.2 – 5. 2. 2024

Učební text vychází především ze skript:

1) M. Litschmannová – Vybrané kapitoly z pravděpodobnosti, Ostrava 2011, VŠB-TU Ostrava
2) M. Litschmannová – Úvod do statistiky, Ostrava 2011, VŠB-TU Ostrava

Obsah

Kapitola 0 - Barevné označení textu

- Černý text /vzorce základní text
- Červený text /vzorce důležitý text
- Modrý text /vzorce rozšiřující text
- Zelený text příkazy v matlabu /octave

0 – Používání příkladů na cvičeních

- Příklady jsou uloženy v adresářích se stejným číslem jako čísla kapitol ve skriptech.
- Označení příkladů má strukturu Pxxyyzz, kde
  - xx – představuje číslo kapitoly ve skriptech
  - yy – číselné označení příkladu
  - zz – může mít následující označení
    - Res – řešení příkladu (skript)
    - Nic – vstupní data, generátor vstupních dat
- Například příklady:
  - P0312.mat vstupní data příkladu 12 ze 3. kapitoly
  - P0312res.m matlabovský skript řešení příkladu 12 ze 3. kapitoly
  - P0312.m generátor vstupu pro příklad 12 ze 3. kapitoly
  - P0405res.m matlabovský skript řešení příkladu 5 ze 4. kapitoly

0 – Požadavky na SW

- Matlab
  - Statistický toolbox
  - Symbolický toolbox
- Alternativa
  - Octave (není zajištěna shodnost příkazů s matlabem).
    - Výhoda octave volně šiřitelný
    - Statistický toolbox se nahraje příkazem pkg load statistics
  - R (příkazy nejsou shodné s matlabem)
    - Výhoda R je také volně šiřitelný
- Nedoporučuji Excel
  - neúplnost množiny používaných funkcí,
  - možná chybná implementace některých příkazů, zvláště při větším množství vstupních dat.

1. Kombinatorika

- 1.1 Faktoriál
- 1.2 Základní kombinatorická pravidla
- 1.3 Uspořádané výběry – variace, permutace
- 1.4 Neuspořádané výběry – kombinace
- 1.5 Základní příkazy

1.1 Faktoriál

- Faktoriál čísla *n* je definován pro nezáporná celá čísla:

$$n! = \prod_{i=1}^{n} i = n \cdot (n-1) \cdot \cdots \cdot 2 \cdot 1$$

$$0! = 1$$

- Funkce v matlabu: factorial(n)

- Výpočet pro velká *n* pomocí Stirlingova vzorce

$$n! \approx \sqrt{2\pi n}\left(\frac{n}{e}\right)^{n}\left(1 + \frac{1}{12n}\right)$$

- Pro výpočet velkých hodnot faktoriálů se doporučuje počítat jeho logaritmus.

$$\log(n!) \approx \frac{1}{2}\log(2\pi n) + nlog(n) - nlog(e)$$

- Rozšíření na reálná čísla $z! = \Gamma(z+1)$
  - $\Gamma(z)$ je Gamma funkce $\Gamma(z) = \int_0^\infty t^{z-1} e^{-t} dt$ gamma(z)
  - Například: factorial(4)=gamma(5)=24

1.2 Základní kombinatorická pravidla

- 1.2.1 - Kombinatorické pravidlo součinu
- 1.2.2 – Kombinatorické pravidlo součtu

1.2.1 Kombinatorické pravidlo součinu

- Počet všech uspořádaných *k*-tic, jejichž první člen lze vybrat n_1 způsoby, druhý člen po výběru prvního členu n_2 způsoby, až *k*-tý člen po výběru všech předcházejících členů n_k způsoby, přičemž jednotlivé členy se navzájem neovlivňují (jsou nezávislé), je roven $n_1 \cdot n_2 \cdot \dots \cdot n_k$.

- Př. Při dlouhé služební cestě projíždíte tři města, kde se určitě zastavíte na jídle. V prvním městě znáte 3 dobré restaurace, v druhém také 3 a ve třetím 4. Kolika druhy uspořádání lze vybrat restaurace?
  - Počet vybraných restaurací je nezávislý na volbě předchozí restaurace.
  - Restaurace v prvním městě označíme: 1 2 3, v druhém 1 2 3, ve třetím 1 2 3 4.
  - Můžeme zvolit následující uspořádání restaurací:

111	112	113	114	121	122	123	124
131	132	133	134				
211	212	213	214	221	222	223	224
231	232	233	234				
311	312	313	314	321	322	323	324
331	332	333	334				

  - Celkem tedy 36 uspořádání = $3 \cdot 3 \cdot 4$.

1.2.1 Kombinatorické pravidlo součinu

- Důležité v předpokladech: „se navzájem neovlivní“.
- Příklad: Při dlouhé služební cestě projíždíte tři města, kde se určitě zastavíte na jídle. V prvním městě znáte 3 dobré restaurace, v druhém také 3 a ve třetím 4.
  - Jestliže v prvním městě půjdete do restaurace 1, v dalších městech nemůžete jít do restaurace 2 nebo 3.
  - Kolika druhy uspořádání lze vybrat restaurace?
  - Můžeme zvolit následující uspořádání restaurací:

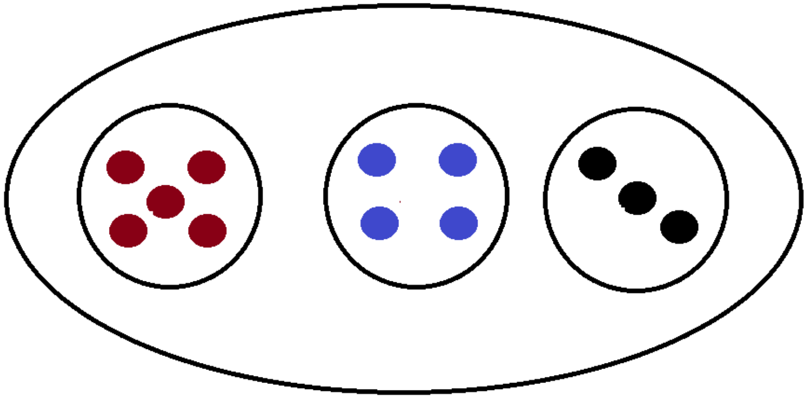
111	114						
211	212	213	214	221	222	223	224
231	232	233	234				
311	312	313	314	321	322	323	324
331	332	333	334				

  - Kombinatorické pravidlo součinu potom neplatí.

1.2.2 – Kombinatorické pravidlo součtu

- Jsou-li $A_1, A_2, \dots, A_n$ konečné množiny, které mají po řadě $n_1, n_2, \dots, n_n$ prvků, a jsou-li každé dvě disjunktní, pak počet prvků množiny $A_1 \cup A_2 \cup \dots \cup A_n$ je roven $n_1 + n_2 + \cdots + n_n$.

- Př. V prvním osudí je 5 koulí červených, 4 koule modré a 3 koule černé. Kolik je v osudí červených a modrých koulí?



1.3 Uspořádané výběry – variace, permutace

- 1.3.1 Variace *k* třídy bez opakování
- 1.3.2 Permutace bez opakování
- 1.3.3 Variace *k* třídy s opakováním
- 1.3.4 Permutace s opakováním

1.3.1 Variace *k* třídy bez opakování

- Nechť *M* je libovolná množina *n* prvků. Každá uspořádána *k*-tice (skupina k prvků) navzájem různých prvků množiny *M* se nazývá variace *k*-té třídy množiny *M* bez opakování. Počet variací *k*-té třídy množiny *M* bez opakování nazýváme variační číslo a značíme jej

$$V(n,k) = \frac{n!}{(n-k)!}$$

- Záleží na pořadí vylosovaných prvků.
- Matlab: výpočet přes funkci faktoriál

factorial(n)/factorial(n-k)

- U variací vždy záleží na pořadí v jakém se prvky vybírají.

- Př. V osudí máme koule očíslované 1 až 9. Vybíráme 3 koule, přičemž je nevracíme. Kolik různých čísel můžeme vybrat? (záleží na pořadí vybírání)
  - První kouli můžeme vybrat z 9 možností, druhou již z 8 (předchozí jsme nevrátili zpět) a poslední ze 7. Celkem to je $9 \cdot 8 \cdot 7$=504 způsobů.
  - $V(n,k) = \frac{n!}{(n-k)!} = \frac{9!}{6!} = 9 \cdot 8 \cdot 7 = 504$
  - Jedná se o kombinatorické pravidlo součinu.

1.3.2 Permutace bez opakování

- Permutace množiny *M* bez opakování je každé navzájem různé uspořádání množiny *M*. Počet permutací *n* prvkové množiny lze stanovit ze vztahu:

$$P(n) = V(n, n) = n!$$

- Př. V osudí máme koule očíslované 1 až 9. Postupně vybíráme všechny koule, přičemž je nevracíme. Kolik různých čísel můžeme vybrat? (záleží na pořadí vybírání)
  - P$(9) = n! = 9! = 362\,880$
  - První kouli vybereme z 9 možností, druhou z 8 atd. Kombinatorické pravidlo součinu.

1.3.3 Variace *k* třídy s opakováním

- Variací *k*-té třídy s opakováním se předpokládá každá uspořádaná k-tice prvků množiny *M*, v níž se jednotlivé prvky mohou opakovat.
$$V^*(n,k) = n^k$$
- Záleží na pořadí vylosovaných prvků.
- Vzorec vychází z kombinatorického pravidla součinu – viz kapitola 1.2.1, kdy každý pokus vybíráme z *n* prvků.
- Matlab/Octave n.^k

- Př. Při zápisu písmene do paměti počítače používáme dvoustavovou logiku – cifry 0 a 1. Kolik můžeme zapsat různých písmen, když pro zapsání písmene používáme 8 bitů.
  - Při zapsání slova záleží na pořadí cifer a cifry se mohou opakovat, proto variace s opakováním.
  - $n = \{0,1\}; k = 8 \text{ pokusů}$
  - $V^*(2,8) = 2^8 = 256$
  - Prvek na 1. pozici má dvě možnosti, na 2. pozici také dvě atd. Kombinatorické pravidlo součinu.

1.3.4 Permutace s opakováním

- Permutace s opakováním jsou permutace, kde se prvky ve výběru mohou opakovat. Počet permutací s opakováním je určen:

$$P^*(n_1, n_2, \dots, n_k) = \frac{n!}{n_1! \cdot n_2! \cdot \dots \cdot n_k!}$$

, přičemž mezi vybranými prvky je *k* skupin, které mají postupně n_1, n_2, …, n_k stejných prvků a musí platit $\mathrm{n} = \sum_{i=1}^{k} n_i$.

- Př. Mějme slovo „ABECEDABECE“. Kolika způsoby můžeme provést přehození jednotlivých písmen tohoto slova.
  - Písmeno A je v textu 2x; písmeno B 2x, písmeno C 2x, D 1x, E 4x
  - Počet písmen ve slově je 11 $n = 11$
  - $n_A = 2; n_B = 2; n_C = 2; n_D = 1; n_E = 4;$
  - $P^*(n_1, n_2, \dots, n_k) = \frac{11!}{2! \cdot 2! \cdot 2! \cdot 1! \cdot 4!} = 207900$

1.4 Neuspořádané výběry – kombinace

- 1.4.1 – Rozdíl mezi uspořádanými a neuspořádanými výběry
- 1.4.2 – Kombinace bez opakování
- 1.4.3 – Kombinace s opakováním

1.4.1 – Rozdíl mezi uspořádanými a neuspořádanými výběry

- Uspořádané výběry (variace a permutace)
  - Záleží na pořadí
  - Výběr $14785 \neq 58741$
- Neuspořádané výběry (kombinace)
  - Nezáleží na pořadí
  - Výběr $14785 = 58741$
  - Př. Sportka – je jedno zda námi uhodnuté číslo bylo vylosováno jako první, nebo jako páté.

1.4.2 – Kombinace bez opakování

- Kombinaci *k*-třídy z *n* prvků bez opakování nazveme každou *k* prvkovou podmnožinu z *n* prvkové množiny *M*. Počet různých kombinací značíme $C(n,k) = \binom{n}{k}$ a určíme dle vzorce

$$C(n,k) = \binom{n}{k} = \frac{n!}{(n-k)! \cdot k!}$$

- Výraz $\binom{n}{k}$ nazýváme kombinační číslo.

nchoosek(n,k)

- Pro výpočet kombinačních čísel s velkým číslem n, k lze použít výpočet pomocí Stirlingova vzorce, nebo výpočet provést pomocí logaritmů.

$$\log\bigl(C(n,k)\bigr) = \sum_{i=2}^{n} \log i - \sum_{i=2}^{n-k} \log i - \sum_{i=2}^{k} \log i$$

- Př. Kolika způsoby můžeme vyplnit tiket, kde vybereme 20 čísel z 80.
  - Nezáleží na pořadí, protože na odevzdaném tiketu nepoznáme, zda jsme číslo 53 vyplnili nejdříve nebo jako poslední. Jedná se proto o kombinace. Kdyby na pořadí záleželo, jednalo by se o variace.
  - $C(80,20) = \frac{80!}{(80-20)! \cdot 20!} = 3.5353 \cdot 10^{18}$

1.4.3 – Kombinace s opakováním

- Kombinaci *k*-třídy z *n* prvků s opakováním nazveme každou *k*-člennou skupinu sestavenou z prvků množiny *M* tak, že se prvky ve skupině mohou opakovat a přitom nezáleží na jejich pořadí. Počet kombinací s opakováním je dán vztahem:

$$C^*(n,k) = \binom{n+k-1}{k} = \frac{(n+k-1)!}{(n-1)!\cdot k!}$$

- V matlabu se počítá pomocí funkce nchoosek(n+k-1,k), pozor na zadávání správných parametrů.
- Př. V restauraci máme možnost si vybrat ze 3 druhů piv – Svijany, Konrád a Gambrinus. Návštěvník vypil 5 piv. Kolika způsoby mohl zkombinovat svoji konzumaci.
  - Jedná se o kombinace s opakováním, protože některých piv mohl (dokonce musel) vypít více a nezáleží na pořadí. Kdyby záleželo na pořadí v jakém vypije svoje pivo, jednalo by se o variace s opakováním.
  - n=3 – vybíráme ze 3 druhů piv   k=5 – zákazník vypil 5 piv
  - n je neměnné   k záleží na zákazníkovi, může vypít 1, 2, 3, … piv
  - Svijany označíme S, Konrád K a Gambrinus G

5xS					
4xS+K	4xS+G				
3xS+2xK	3xS+2xG	3xS+K+G			
2xS+3xK	2xS+2xK+G	2xS+K+2xG	2xS+3xG		
1xS+4xK	1xS+3xK+G	1xS+2xK+2xG	1xS+K+3xG	1xS+4xG	
5xK	4xK+1xG	3xK+2xG	2xK+3xG	1xK+4xG	5xG

  - $C^*(3{,}5) = \binom{3+5-1}{5} = \frac{7!}{5!\cdot 2!} = 21$

1.5 Základní příkazy v matlabu/octave

- Faktoriál factorial(n)
- Gamma funkce gamma(n)
- Kombinační číslo nchoosek(n,k)

2. Úvod do teorie pravděpodobnosti

- 2.1 Základní pojmy
- 2.2 Označování jevů a operace mezi jevy
- 2.3 Pravděpodobnost a její vlastnosti
- 2.4 Nezávislost pokusů, podmíněná pravděpodobnost

2.1 Základní pojmy

- Teorie pravděpodobnosti je matematická disciplína, jejíž logická struktura je budována axiomaticky. To znamená, že její základ tvoří několik tvrzení, která vyjadřují základní vlastnosti pravděpodobnosti a všechna další tvrzení jsou z nich odvozena deduktivně.
  - Př. Kolika způsoby jsme schopni …, kolik existuje kombinací …
  - Výsledky úloh jsou stejné, ať správně počítá Tonda nebo Jarda.

- Matematická statistika je věda zahrnující studium dat vykazujících náhodná kolísání, ať už jde o data získaná pečlivě připraveným pokusem provedeným pod stálou kontrolou experimentálních podmínek v laboratoři, či o data provozní, případně o data získaná počítačovými simulacemi (tzv. metodou Monte-Carlo).
  - Př. Tonda i Jarda měli každý 10 výrobků a zjišťovali jejich dobu do poruchy.
  - Oba mají odlišné výsledky.

2.1 Základní pojmy

- Náhodný pokus - je každý děj, jehož výsledek není předem jednoznačně určen podmínkami, za kterých probíhá.
  - Například hod mincí či kostkou, životnost výrobku, přesná velikost určitého rozměru.
- Množina možných výsledků $\{\omega\}$ pokusu, tzv. základní prostor Ω
  - Množina možných výsledků musí být volena tak, aby žádné z nich nemohly nastat současně.
  - Ω – rub, líc; hod kostkou 1,2,3,4,5,6
  - Ω – doba do poruchy R^+ počet poruch $Z^+ + \{0\}$
  - Základní prostor není – sjednocení množin lichých čísel a čísel dělitelných tří beze zbytku.

2.1 Základní pojmy

- Náhodný jev představuje každou podmnožinu A základního prostoru Ω.
  - Na kostce padne číslo 3.
  - Na kostce padne sudé číslo.
  - Výška člověka je vyšší než 180 cm.
- Elementární jev – jednoprvkové podmnožiny základního prostoru Ω.
  - Na kostce padne číslo 3 – náhodný jev již nelze dále rozdělit na dílčí náhodné jevy.
- Složený jev – víceprvkové podmnožiny základního prostoru Ω.
  - Na kostce padne sudé číslo. Jedná se o složený jev, protože mohou padnout čísla $\{2,4,6\}$. Čísla $\{2,4,6\}$ tvoří elementární jevy.
  - Životnost výrobku je mezi 2 a 3 roky.

2.2 Označování jevů a operace mezi jevy

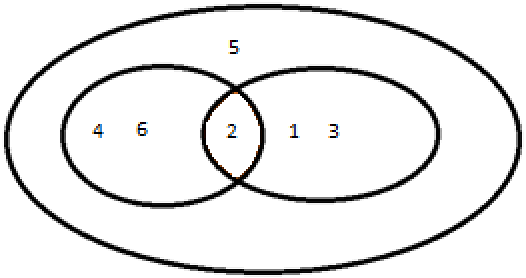
- Jistý jev – jev, který nastane vždy při každé realizaci náhodného pokusu.
  - Př. Při hodu mincí padne buď panna nebo orel.
- Nemožný jev – jev, který nemůže nikdy nastat.
  - Př. Výška člověka je menší než -5 m.
- Jev A je podjevem jevu B - nastal-li jev A, nastane vždy také jev B.
  - $A \subset B$
  - Př. Jev A – člověk je menší než 160 cm,
    Jev B – člověk je menší než 170 cm.
    Jev A je podjevem jevu B.

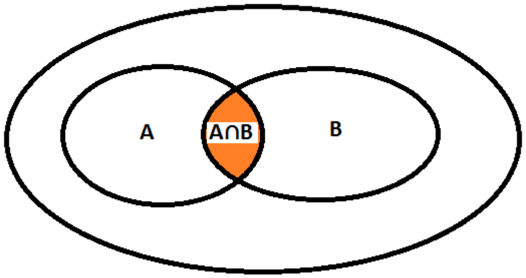
2.2 Označování jevů a operace mezi jevy

- Rovnost jevů A a B – nastal-li jev A, nastane vždy také jev B a naopak
  - $A = B$
  - Jev A je sudé číslo, jev B je číslo dělitelné dvěma beze zbytku.
- Disjunktní jevy A,B – dva jevy A a B náhodného pokusu nemohou nikdy nastat současně.
  – Hod kostkou disjunktní jevy jsou padnutí čísla 1 a 2.
- Doplněk jevu A – značí se $\bar{A}$ - Jev $\bar{A}$ nastane vždy, když nenastane jev A.
  – Při hodu mince je doplňkovým jevem k jevu „padne orel“ „padne panna“.

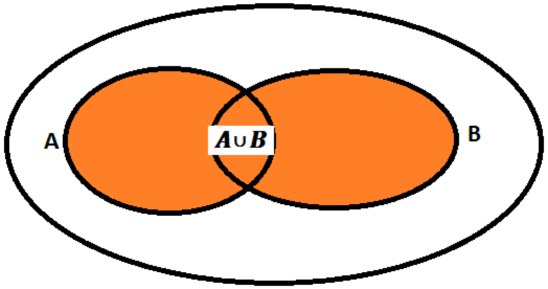
2.2 Označování jevů a operace mezi jevy

- Průnik jevů $A \cap B$ – nastane, jestliže jevy A a B vzniknou současně.
  - Př. Na šestistěnné kostce padne sudé číslo (jev A) a číslo menší rovno 3 (jev B). Pak průnikem $A \cap B = \{2\}$.





- Sjednocení jevů $A \cup B$ – nastane, jestliže výsledkem bude jev A nebo jev B
  - Př. Na šestistěnné kostce padne sudé číslo (jev A) a číslo menší rovno 3 (jev B). Pak sjednocením $A \cup B = \{1,2,3,4,6\}$.



2.2 Označování jevů a operace mezi jevy

- Základní pravidla pro operace s náhodnými jevy:
  - $A \cup B = B \cup A$ $\quad$ $A \cap B = B \cap A$
  - $A \cup (B \cup C) = (A \cup B) \cup C$
  - $A \cap (B \cap C) = (A \cap B) \cap C$
  - $A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$
  - $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$
  - $A \cup A = A$ $\quad$ $A \cap A = A$
  - $A \cup \Omega = \Omega$ $\quad$ $A \cap \Omega = A$
  - $\overline{A \cup B} = \bar{A} \cap \bar{B}$ $\quad$ 1. de Morganův zákon
  - $\overline{A \cap B} = \bar{A} \cup \bar{B}$ $\quad$ 2. de Morganův zákon

2.3 Pravděpodobnost a její vlastnosti

- 2.3.1 Pravděpodobnost
- 2.3.2 Klasická pravděpodobnost
- 2.3.3 Statistická pravděpodobnost
- 2.3.4 Geometrická pravděpodobnost
- 2.3.5 Kolmogorova pravděpodobnost
- 2.3.6 Vlastnosti pravděpodobnosti

2.3.1 Pravděpodobnost

- Výsledek náhodného jevu nelze s jistotou předpovědět. Při častějších opakování však některé náhodné pokusy vykazují zákonitosti a pravidelnost výskytu.
- Pravděpodobností označujeme míru očekávatelnosti výskytu náhodného jevu.
  S rostoucí pravděpodobností roste i šance, že jev nastane.
- Pravděpodobnost se obecně označuje číslem z intervalu $\langle 0,1 \rangle$.

2.3.2 Klasická pravděpodobnost

- Je-li základní prostor konečná neprázdná množina elementárních jevů, které mají stejnou šanci výskytu $\frac{1}{n}$, potom pravděpodobnost, že při realizaci náhodného pokusu jev A nastane je $P(A) = \frac{m}{n}$, kde m je počet výsledků příznivých jevů A, n počet všech možných výsledků.
- Př. Pravděpodobnost, že na šestistěnné kostce padne jedno z čísel 2,3,5 nebo 6.
  - $m = 4, n = 6$
  - Každé z čísel má stejnou pravděpodobnost výskytu
  - $P(A) = \frac{4}{6} = \frac{2}{3}$.

2.3.2 Klasická pravděpodobnost

- Př. V osudí je 10 černých a 5 bílých koulí. Jaká je pravděpodobnost, že při výběru 2 koulí budou obě dvě černé.
  - Černé koule si očíslujeme 1 až 10, bílé 11 až 15.
  - Celkem je v osudí 15 koulí a vybíráme 2 z nich. Počet způsobů výběru dvou koulí je kombinační číslo $\binom{15}{2} = \frac{15 \cdot 14}{2} = 105$.
  - Počet výběru dvou černých koulí je kombinační číslo $\binom{10}{2} = \frac{10 \cdot 9}{2} = 45$
  - Pravděpodobnost je $P(A) = \frac{45}{105} = \frac{3}{7}$.

2.3.2 Klasická pravděpodobnost

- Jaká je pravděpodobnost, že z balíčku 32 mariášových karet vylosujeme 8 karet a všechny budou buď spodci, svršci, králové nebo esa. Karty po vylosování nevracíme.
- $P = \frac{16}{32} \cdot \frac{15}{31} \cdot \frac{14}{30} \cdot \frac{13}{29} \cdot \frac{12}{28} \cdot \frac{11}{27} \cdot \frac{10}{26} \cdot \frac{9}{25} = 0.0122$
  - 1. člen – mám 16 příznivých karet ze 32 celkových
  - 2. člen – mám 15 příznivých karet ze 31
  - Kombinatorické pravidlo součinu
- $P = \frac{\binom{16}{8}}{\binom{32}{8}} = \frac{\frac{16!}{8! \cdot 8!}}{\frac{32!}{8! \cdot 24!}} = 0.0122$

2.3.3 Statistická pravděpodobnost

- Jestliže je náhodný pokus libovolněkrát opakovatelný za stejných podmínek, pak lze pravděpodobnost jevu odhadnout na základě počtu jevů příznivých výsledků pokusu.
- Provedeme-li n realizací náhodného pokusu, přičemž $n(A)$ je počet příznivých realizací, potom pravděpodobnost jevu A lze odhadnout poměrem:

$$P(A) = \frac{n(A)}{n}$$

  – Odhad je tím přesnější, čím větší je počet realizací náhodného pokusu.
  – Jedná se o odhad pravděpodobnosti, přesná hodnota se obdrží limitně při nekonečném počtu pokusu.

2.3.3 Statistická pravděpodobnost

- Antonín Novák jezdí v MHD tzv. „na černo“. Dělá si statistiku, za loňský rok jel 2000x a z toho byl 28x chycen. Jaká je pravděpodobnost, že je A. Novák chycen revizorem.

$$P(A) = \frac{n(A)}{n} = \frac{28}{2000} = 0.014$$

- Pokud si bude dělat statistiku i letos, určitě mu vyjde $P(A)$ odlišná od loňské hodnoty. Se zvyšujícím se počtem náhodných pokusů – jízd na „černo“ se pravděpodobnost bude blížit přesné hodnotě.

2.3.3 Statistická pravděpodobnost

- Na základě statistické pravděpodobnosti je odvozena „metoda Monte Carlo“, kdy se mnohonásobně opakuje náhodný pokus a sleduje se úspěšnost pokusů.
  - Náhodné pokusy mohou být fyzikální (např. hod kostkou, životnost výrobku)
  - Náhodné pokusy mohou být generovány pomocí počítače – for cyklus, funkce náhodné číslo (v matlabu funkce unifrnd).
- Př. Chceme zjistit, jaká je pravděpodobnost, že při hodu hrací kostkou hodíme 6. Učinili jsme náhodný pokus, kdy jsme 300x hodili hrací kostkou. Počítáme kolikrát padla 6. Celkem se to stalo 46x. Potom odhad pravděpodobnosti padnutí 6 je:

$$P(A) = \frac{n(A)}{n} = \frac{46}{300}$$

- Pokud náhodný pokus bude opakovat někdo jiný, může mu vyjít jiná pravděpodobnost.
- Přesná hodnota výsledku je $P(A) = \lim_{n \to \infty} \frac{n(A)}{n}$.

2.3.4 Geometrická pravděpodobnost

- Geometrickou pravděpodobnost lze použít v případech, kdy počet všech možných výsledků náhodného pokusu je nespočetný.
- Pravděpodobnost je založená na porovnání velikosti délek, plochy, objemů.
- Definice: V prostoru je dána určitá oblast Ω a v ní podoblast A. Pravděpodobnost výskytu každého bodu v oblasti Ω je shodná. Potom pravděpodobnost jevu A (náhodně zvolený bod leží v oblasti A) lze zjistit dle vzorce:

$$P(A) = \frac{|A|}{|\Omega|}$$

, kde $|A|$ a $|\Omega|$ jsou velikosti oblasti.

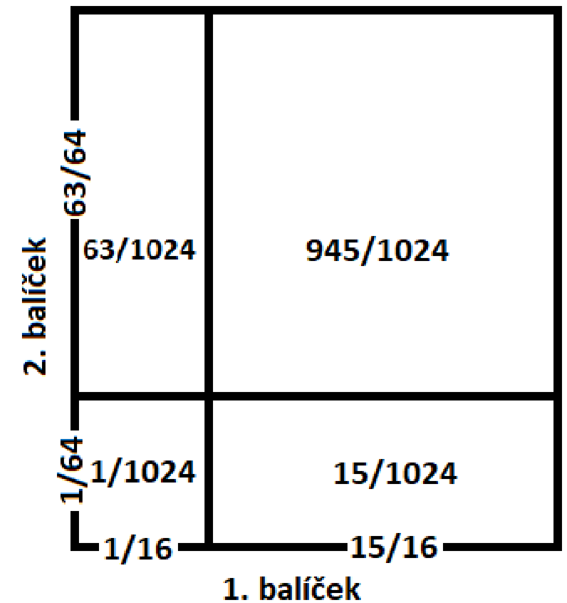
2.3.4 Geometrická pravděpodobnost

- Př. Máte krychli o délce strany a. V krychli je umístěna koule o průměru $\frac{3}{4}a$. Vypočtěte pravděpodobnost, že náhodně zvolený bod v krychli bude zároveň uvnitř koule.
  - Objem krychle je a^3. $|\Omega| = a^3$
  - Objem koule je $\frac{4}{3}\pi r^3$. Poloměr koule je $\frac{3}{8}a$. Objem koule je tedy $\frac{9}{128}\pi a^3$. $|A| = \frac{9}{128}\pi a^3$
  - Pravděpodobnost je:

$$P(A) = \frac{|A|}{|\Omega|} = \frac{\frac{9}{128}\pi a^3}{a^3} = \frac{9}{128}\pi$$

2.3.4 Geometrická pravděpodobnost

- Př. Máte dva oddělené balíčky mariášových karet. Spočtěte pravděpodobnost, že:
  - z prvního balíčku táhnete 4x a vytáhnete vždy spodka, filka, krále nebo eso;
  - z druhého balíčku táhnete 2x a vytáhnete v obou případech sedmičku.
  - (karty vracíte zpět do balíčku)
  - $P_1 = \frac{16}{32} \cdot \frac{16}{32} \cdot \frac{16}{32} \cdot \frac{16}{32} = \frac{1}{16}$
  - $P_2 = \frac{4}{32} \cdot \frac{4}{32} = \frac{1}{64}$
  - $P = P_1 \cdot P_2 = \frac{1}{16} \cdot \frac{1}{64} = \frac{1}{1024}$
  - Kombinatorické pravidlo součinu



2.3.5 Kolmogorova pravděpodobnost

- Kolmogorova pravděpodobnost je zobecnění definice pravděpodobnosti.
- Je-li $\mathcal{A}$ jevové pole, pak pravděpodobnost na jevovém poli $\mathbb{A}$ je reálná funkce, pro kterou platí Kolmogorovy axiomy pravděpodobnosti:
  1. Pravděpodobnost každého jevu $A \in \mathbb{A}$ je nezáporné reálné číslo $P(A) \geq 0$.
  2. Pravděpodobnost jistého jevu je rovna jedné $P(\Omega) = 1$.
  3. Pravděpodobnost sjednocení spočetného počtu vzájemně neslučitelných jevů je rovna součtu jejich pravděpodobností
     $A_i \in \mathbb{A}, \forall\, i \qquad A_i \cap A_j = \emptyset, \forall\, i, j, i \neq j$
     $P(\bigcup_i A_i) = \sum_i P(A_i)$

2.3.6 Vlastnosti pravděpodobnosti

- Jevy $A, B \in \mathbb{A}$
  - 1) Pravděpodobnost jevu je omezena 0 a 1 včetně
    $$0 \leq P(A) \leq 1$$
  - 2) Pravděpodobnost nemožného jevu je 0
    $$P(\emptyset) = 0$$
  - 3) Pravděpodobnost doplňku k jevu A je
    $$P(\bar{A}) = 1 - P(A)$$
  - 4) Jestliže A je podmnožinou B, potom pravděpodobnost A je menší nebo rovna B
    $$A \subset B \quad P(A) \leq P(B)$$
  - 5) Rozdíl jevů B-A je
    $$P(B - A) = P(B) - P(A \cap B)$$
  - 6) Pravděpodobnost sjednocení jevů A a B
    $$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$
  - 7) $P(A \cup B) = 1 - P(\bar{A} \cap \bar{B})$
  - 8) $P(A \cap B) = 1 - P(\bar{A} \cup \bar{B})$

- Porovnej se základními pravidly s náhodnými jevy.
- Možnost dokázat pomocí geometrické pravděpodobnosti.

2.4 Nezávislost pokusů, podmíněná pravděpodobnost

- 2.4.1 Nezávislost pokusů
- 2.4.2 Podmíněná pravděpodobnost
- 2.4.3 Věta o úplné pravděpodobnosti
- 2.4.4 Bayesův vzorec
- 2.4.5 Příklad

2.4.1 Nezávislost pokusů

- Provedeme několik pokusů. Jestliže pravděpodobnost jevu A při každém opakování nezávisí na výsledcích předchozích pokusů, potom jsou tyto pokusy nezávislými pokusy vzhledem k jevu A.
  - Např. opakované házení šestistěnnou hrací kostkou. Výsledek hodu není závislý na výsledku předchozího hodu.
  - Výběr ze souboru, kde prvky se vrací zpět. Neustále shodné podmínky pokusu.
- Pravděpodobnost jevu A je při všech pokusech stejná.

2.4.1 Nezávislost pokusů

- Nezávislé pokusy
  - Jaká je pravděpodobnost, že při dvou hodech mincí padne vždy orel.
  - Pravděpodobnost, že při jednom hodu padne orel je $\frac{1}{2}$.
  - Pravděpodobnost druhého pokusu, že padne orel je také $\frac{1}{2}$. Výsledek není závislý na výsledku prvního pokusu.
- Kombinatorické pravidlo součinu.
  - Pravděpodobnost, že u obou hodů padne orel je

$$P = \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4}$$

2.4.1 Nezávislost pokusů

- Závislé pokusy – pravděpodobnost nastoupení jevu v určitém pokusu je závislé na předchozích výsledcích.

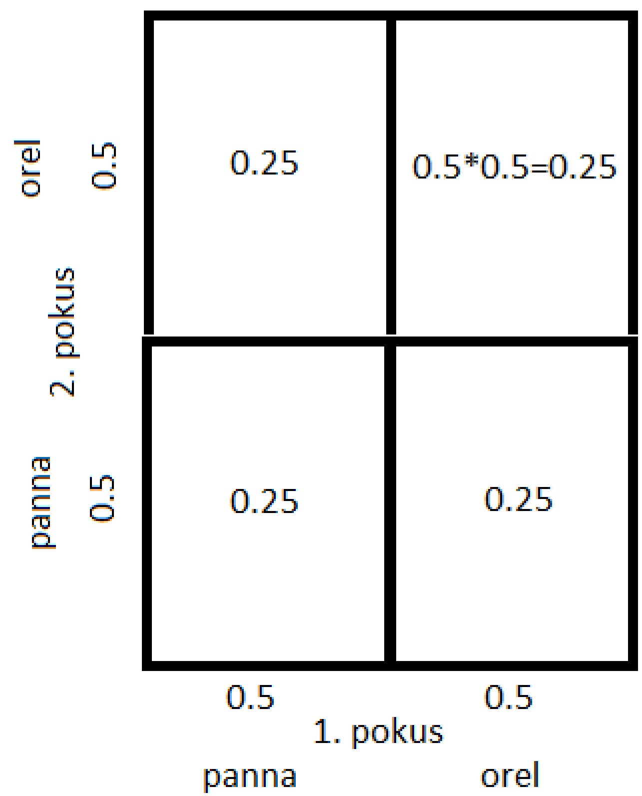
- Př. Losování z osudí bez vracení. Před prvním pokusem je v osudí n míčků, před druhým již jen $n-1$. Podmínky pokusu jsou odlišné.
- V osudí je 10 koulí černých a 5bílých. Vybereme 2 koule, které nevracíme. Jaká je pravděpodobnost, že …
  - První pokus je výběr z 15 koulí , druhý pokus již jen ze 14.

  - Při prvním pokusu byla vybrána bílá. Při druhém je $P_2(bíl\text{á}) = \frac{4}{14}$, $P_2(\check{c}ern\text{á}) = \frac{10}{14}$.
  - Při prvním pokusu byla vybrána černá. Při druhém je $P_2(bíl\text{á}) = \frac{5}{14}$, $P_2(\check{c}ern\text{á}) = \frac{9}{14}$.
  - Pravděpodobnosti pokusu P_2 jsou závislé na výsledku prvního pokusu, a tím na sobě závislé.

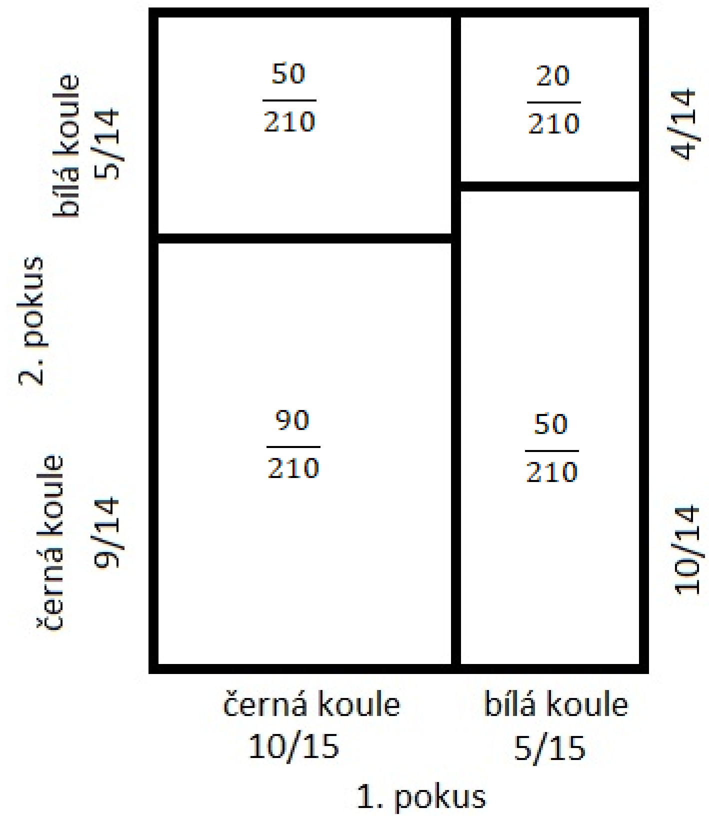
2.4.1 Nezávislost pokusů

- Spojitost s geometrickou pravděpodobností

Nezávislé jevy
obdélníky se protínají v kříži



Závislé jevy
obdélníky se neprotínají v kříži



2.4.2 Podmíněná pravděpodobnost

- Pravděpodobnost, že nastane jev A za podmínky, že nastal jev B se vypočte podle vzorce:

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

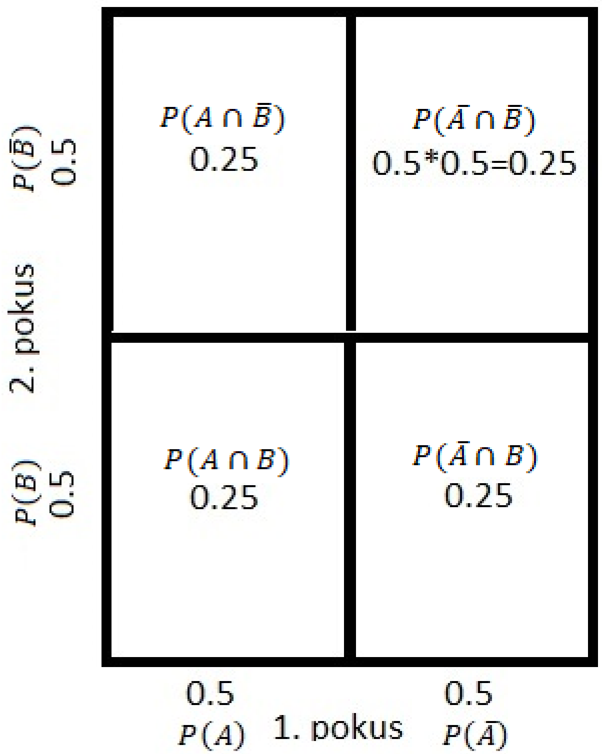
- Podmíněná pravděpodobnost se značí $P(A|B)$
- Ze vzorce lze odvodit pravděpodobnost průniku dvou jevů

$$P(A \cap B) = P(A|B) \cdot P(B)$$

- U nezávislých jevů platí, že $P(A|B) = P(A)$
  – Nastoupení jevu B nemá žádný vliv na jev A.

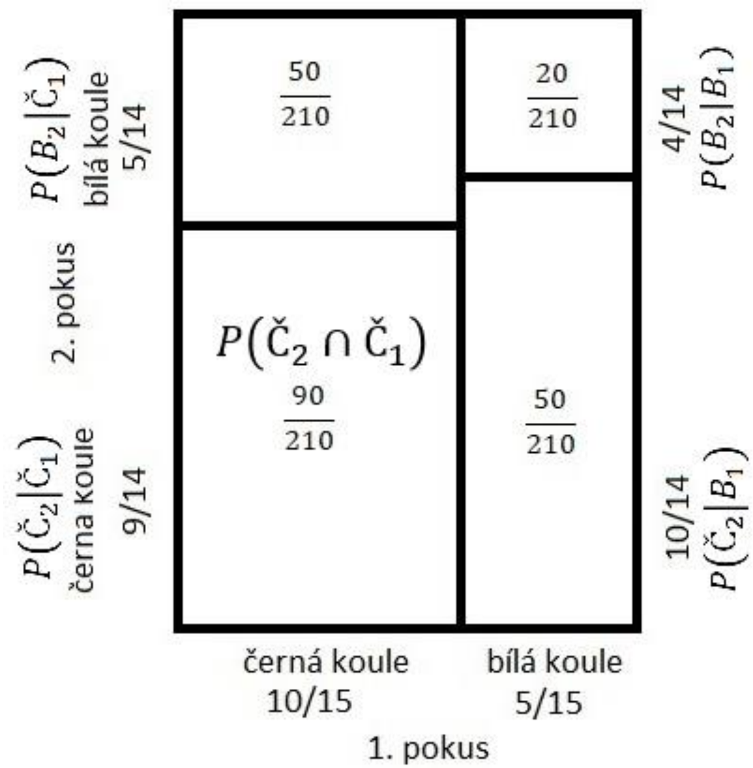
2.4.2 Podmíněná pravděpodobnost

- Nezávislé jevy



$$P(A \cap B) = P(A) \cdot P(B)$$

- Závislé jevy



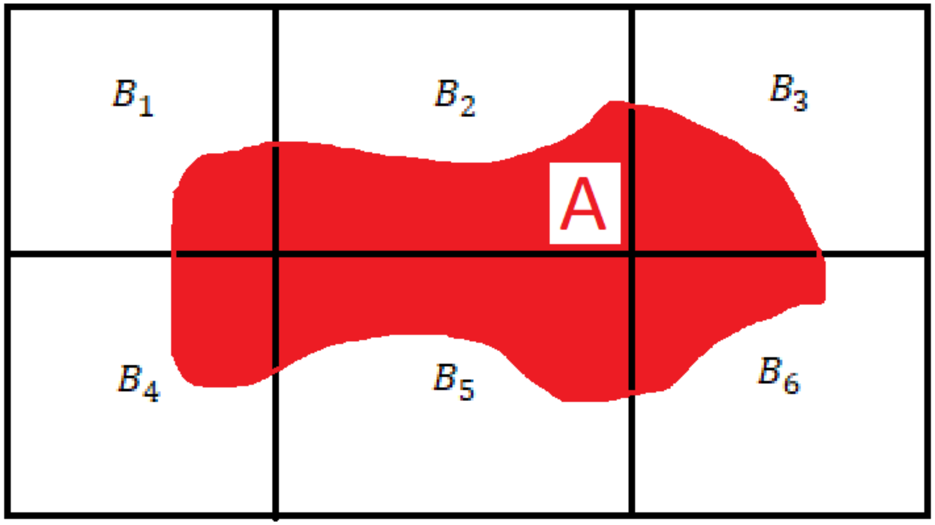
$$P(\check{C}_2 \cap \check{C}_1) = P(\check{C}_2|\check{C}_1) \cdot P(\check{C}_1)$$

2.4.3 Věta o úplné pravděpodobnosti

- Pravděpodobnost, že nastane jev A za podmínky, že nastal jev B_i se vypočte podle vzorce $\mathrm{P}(A|B_i) = \frac{P(A \cap B_i)}{P(B_i)}$.
- Odvozením obdržíme $P(A \cap B_i) = \mathrm{P}(A|B_i) \cdot \mathrm{P}(B_i)$
- Jestliže máme množinu všech možných jevů, které mohli nastat B_i, potom jevu A předcházel jeden z jevů B_i. Platí $\sum_i P(B_i) = 1$

- Uvažujme systém disjunktních jevů $B_1, B_2, \ldots, B_n$, které zcela pokrývají pravděpodobnostní prostor. Pak pravděpodobnost libovolného jevu A lze vypočítat z pravděpodobností $\mathrm{P}(B_i)$ a z podmíněných pravděpodobností $\mathrm{P}(A|B_i)$ podle vzorce:

$$P(A) = \sum_{i=1}^{n} \mathrm{P}(A|B_i) \cdot \mathrm{P}(B_i)$$



2.4.4 Bayesův vzorec

- Pomocí věty o úplné pravděpodobnosti lze stanovit pravděpodobnost jevu A. Předpokládejme, že $P(A) \neq 0$.
- Potom lze stanovit pravděpodobnost jevu $\mathrm{P}(B_k|A)$, tj. určit, který z jevů B_k vedl k nastoupení jevu A.

$$\mathrm{P}(B_k|A) = \frac{\mathrm{P}(A|B_k) \cdot \mathrm{P}(B_k)}{\sum_{i=1}^{n} \mathrm{P}(A|B_i) \cdot \mathrm{P}(B_i)}$$

2.4.4 Bayesův vzorec

- Podmíněná pravděpodobnost průniku $A \cap B_k$ lze stanovit (kapitola 2.4.2):

$$P(A \cap B_k) = P(A|B_k) \cdot P(B_k)$$

- Podmíněná pravděpodobnost průniku $B_k \cap A$ lze stanovit (kapitola 2.4.2):

$$P(B_k \cap A) = P(B_k|A) \cdot P(A)$$

- Přičemž platí:

$$P(A \cap B_k) = P(B_k \cap A)$$

$$P(A|B_k) \cdot P(B_k) = P(B_k|A) \cdot P(A)$$

$$P(B_k|A) = \frac{P(A|B_k) \cdot P(B_k)}{P(A)}$$

- Věta o úplné pravděpodobnosti, jmenovatel Bayesova vzorce je roven $P(A)$.

$$P(A) = \sum_{i=1}^{n} \mathrm{P}(A|B_i) \cdot \mathrm{P}(B_i)$$

- Bayesův vzorec:

$$\mathrm{P}(B_k|A) = \frac{\mathrm{P}(A|B_k) \cdot \mathrm{P}(B_k)}{\sum_{i=1}^{n} \mathrm{P}(A|B_i) \cdot \mathrm{P}(B_i)}$$

2.4.5 Příklad

- Pokračování příkladu na závislé pokusy
  - V osudí je 10 koulí černých a 5 bílých. Vybereme 2 koule, které nevracíme. Jaká je pravděpodobnost, že …
  - Viz kapitola 2.4.2
    - Při prvním pokusu byla vybrána bílá. Při druhém je pravděpodobnost $P_2(bílá) = \frac{4}{14}$ a $P_2(černá) = \frac{10}{14}$.
    - Při prvním pokusu byla vybrána černá. Při druhém je $P_2(bílá) = \frac{5}{14}$ a $P_2(černá) = \frac{9}{14}$.

    - $P(B_2|B_1) = \frac{4}{14}$ $\qquad$ $P\left(Č_2\middle|B_1\right) = \frac{10}{14}$
    - $P\left(B_2\middle|Č_1\right) = \frac{5}{14}$ $\qquad$ $P\left(Č_2\middle|Č_1\right) = \frac{9}{14}$

2.4.5 Příklad

$$P(B_2|B_1) = \frac{4}{14} \qquad P(\check{C}_2|B_1) = \frac{10}{14}$$
$$P(B_2|\check{C}_1) = \frac{5}{14} \qquad P(\check{C}_2|\check{C}_1) = \frac{9}{14}$$
$$P(B_1) = \frac{5}{15} \qquad P(\check{C}_1) = \frac{10}{15}$$

- Udělejme první pokus s tím, že nekoukneme na barvu koule. Při druhém z pokusů padla černá. Jaká je pravděpodobnost, že i při prvním pokusu padla černá.
- Chceme zjistit: $P(\check{C}_1|\check{C}_2)$
  - Pravděpodobnost, že při první pokusu padla černá, za předpokladu, že při druhém pokusu padla černá.

$$\mathrm{P}(B_k|A) = \frac{\mathrm{P}(A|B_k)\cdot\mathrm{P}(B_k)}{\sum_{i=1}^{n}\mathrm{P}(A|B_i)\cdot\mathrm{P}(B_i)}$$

- $P(\check{C}_1|\check{C}_2)=\frac{\mathrm{P}(\check{C}_2|\check{C}_1)\cdot\mathrm{P}(\check{C}_1)}{\sum_i \mathrm{P}(\check{C}_2|X_i)\cdot\mathrm{P}(X_i)}$, kde X_i je buď B_1 nebo $\check{C}_1$.
- $P(\check{C}_1|\check{C}_2)=\frac{\frac{9}{14}\cdot\frac{10}{15}}{\frac{10}{14}\cdot\frac{5}{15}+\frac{9}{14}\cdot\frac{10}{15}} = \frac{\frac{90}{210}}{\frac{140}{210}} = \frac{9}{14}$.
- $P(\check{C}_1|\check{C}_2)$ znamená, že jsme zatajili výsledek prvního tahu (na vytaženou kouli jsme se nepodívali) a zjistili jsme výsledek až druhého tahu. Pak jsme se zeptali, jaká je pravděpodobnost, že první tah byla také černá. Výsledek $\frac{9}{14}$ lze očekávat, protože výsledky tahů jsou záměnné.

3 Náhodná veličina

- 3.1 Distribuční funkce
- 3.2 Diskrétní náhodná veličina
- 3.3 Spojitá náhodná veličina
- 3.4 Funkce náhodné veličiny
- 3.5 Číselné charakteristiky náhodné veličiny
- 3.6 Charakteristiky numerických proměnných
- 3.7 Identifikace odlehlých měření
- 3.8 Příkazy v Matlabu

3 Náhodná veličina

- V této kapitole si ukážeme jakým způsobem lze popsat náhodnou veličinu.
  - Snažíme se o zpřehlednění výsledků náhodného pokusu pomocí tabulky, grafu, několika popisných hodnot.
  - Je přehlednější mít 1 graf či tabulku než vektor 1000 výsledků náhodného pokusu.
- Náhodná veličina - je libovolná reálná veličina (výsledek náhodného pokusu), kterou je možné opakovaně měřit u různých objektů, v různých místech či čase a její hodnoty podrobit zpracování metodami teorie pravděpodobnosti nebo statistiky.
  - Značí se X, Y, Z, …
  - Může být diskrétní nebo spojitá
- Distribuční funkce – F(x) nebo $F(X \leq x)$ – graf popisující pravděpodobnost, že náhodná veličina X nabude menší nebo rovno hodnotě x.
  - Pravděpodobnost, že životnost výrobku je menší nebo rovno t.
  - Distribuční funkce může být pro diskrétní i spojitou náhodnou veličinu.
- Pravděpodobnostní funkce – p(x) – pro diskrétní náhodnou veličinu – graf vyjadřující pravděpodobnost, že náhodná veličina bude nabývat přímo hodnoty x.
  - Pravděpodobnost, že při hodu šestistěnnou kostkou mi padne 3.
- Hustota pravděpodobnosti – f(x) – pro spojitou náhodnou veličinu – graf derivace distribuční funkce. Obdoba pravděpodobnostní funkce pro spojitou náhodnou veličinu.

- Bodové ukazatele náhodné veličiny - střední hodnota, rozptyl, medián
  - popis náhodné veličiny pomocí několika hodnot.

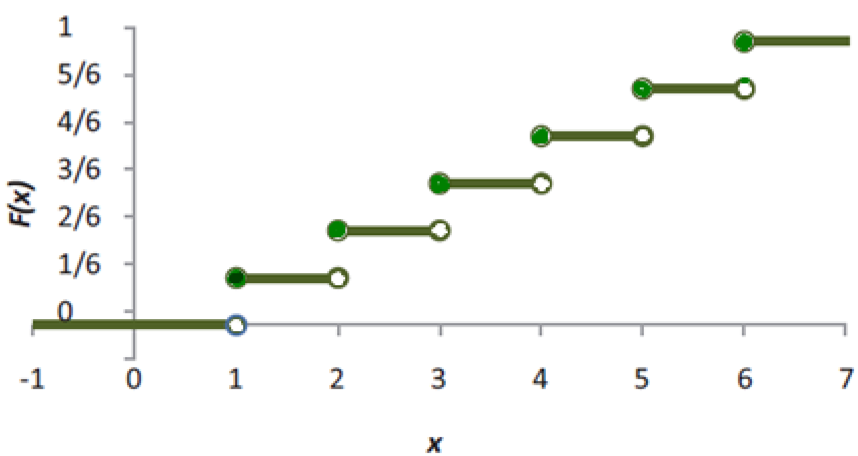
3.1 Distribuční funkce

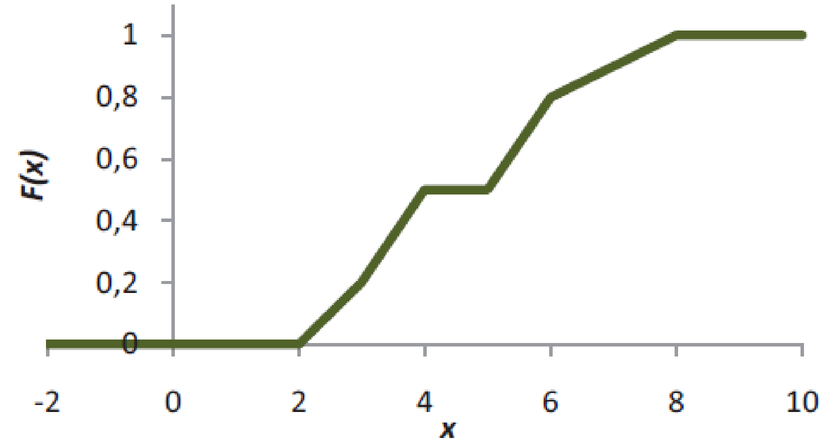
- Distribuční funkce – Nechť X je náhodná veličina. Funkci F(x) definovanou pro všechna reálná x vztahem:

$$F(x) = P(X \leq x)$$

nazýváme distribuční funkcí náhodné veličiny X.

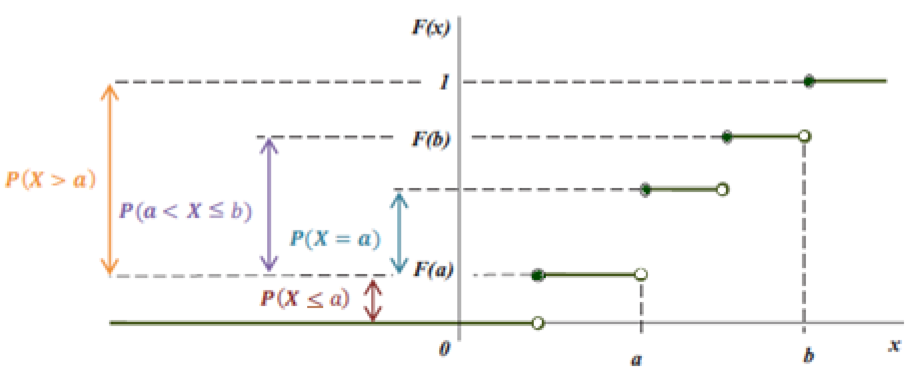
  – Velkým písmenem označujeme náhodnou veličinu; malým hodnotu.
  – Distribuční funkce přiřazuje každému reálnému x pravděpodobnost, že náhodná veličina bude nabývat hodnoty menší nebo rovno x.
  – Distribuční funkce může být buď diskrétní nebo spojitá.
  – Někdy je distribuční funkce definována i $F(x) = P(X < x)$

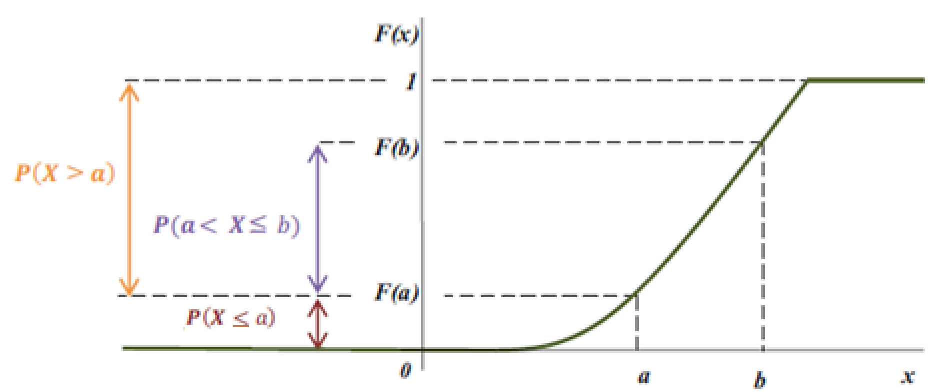




3.1 Distribuční funkce

- Vlastnosti distribuční funkce
  - $0 \leq F(x) \leq 1$ — Funkce je omezena
  - $\forall x_1 \forall x_2, x_1 < x_2, F(x_1) \leq F(x_2)$ — Funkce je neklesající
  - Funkce je zleva spojitá
  - $\lim_{x \to -\infty} F(x) = 0$ — Funkce je zleva omezená
  - $\lim_{x \to \infty} F(x) = 1$ — Funkce je zprava omezená
- Z vlastností distribuční funkce lze odvodit
  - $P(X \leq a) = F(a)$
  - $P(X > a) = 1 - F(a)$ — Doplněk distribuční funkce do 1
  - $P(a < X \leq b) = F(b) - F(a)$ — Rozdíl dvou pravděpodobností
  - $P(X = a) = F(a) - \lim_{x \to a^-} F(x)$ — Funkce je zleva spojitá





3.2 Diskrétní náhodná veličina

- Diskrétní náhodná veličina nastává, jestliže množina výsledků náhodného pokusu nabývá pouze diskrétních hodnot (může být konečný i spočetný počet).
  - Konečný počet: orel/panna po hodu mincí, počet bodů po hodu šestistěnné kostky,
  - Spočetný počet: počet dopravních nehod za jeden den, počet poruch zařízení za jeden rok.

3.2 Diskrétní náhodná veličina

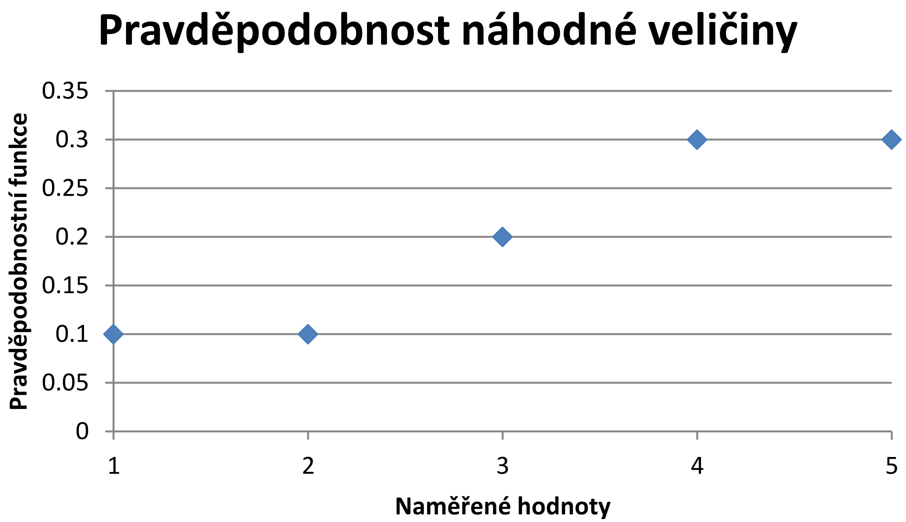
- Náhodná veličina X má diskrétní rozdělení pravděpodobnosti právě tehdy, když nabývá nejvýše spočetně mnoha hodnot $\{x_1, x_2, \ldots\}$ tak, že $P(X = x_i) \geq 0$ a $\sum_{i=1}^{\infty} P(X = x_i) = 1$.

- Funkci $P(X = x_i) = p(x_i)$ nazýváme pravděpodobnostní funkcí náhodné veličiny X.

3.2 Diskrétní náhodná veličina

- Pravděpodobnostní funkce může být popsána:
  - Matematickým předpisem: $\mathrm{P}(X = x) = \binom{5}{x} \cdot 0.2^{x} \cdot 0.8^{5-x}$
  - Tabulkou
  - Grafem

x	1	2	3	4	5
P(x)	0.1	0.1	0.2	0.3	0.3



- Distribuční funkce má vztah s pravděpodobnostní funkcí: $F(x) = \sum_{x_i \leq x} p(x_i)$

3.2 Diskrétní náhodná veličina

- Vizualizace diskrétních hodnot pomocí histogramu
  - x-ová osa intervaly naměřených hodnot
  - y-ová osa četnost dat
- histogram(x,M)
  - x naměřená data
  - M – skalár počet rovnoměrných sloupců v histogramu
  - M – vektor hraničních bodů
- Příklad

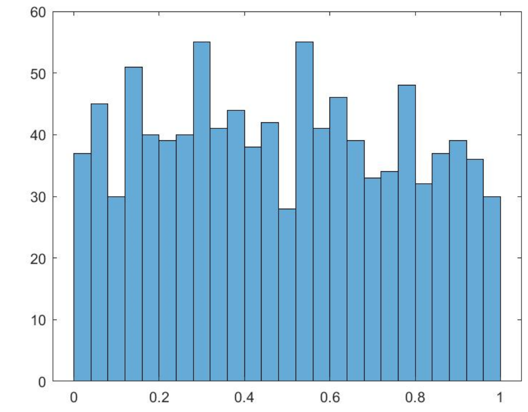
  x=rand(1,1000);

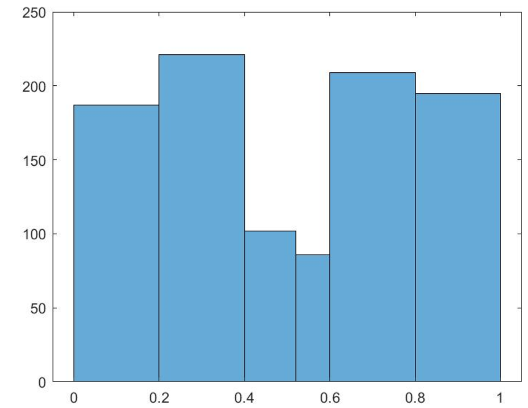
  histogram(x,25) (obrázek nahoře)

  edges=[0,0.2,0.4,0.52,0.6,0.8,1] (hraniční body)

  histogram(x,edges) (obrázek dole)

- Obvyklý počet sloupců lze odhadnout podle vzorců:
  - $M = \sqrt{n}$
  - $M = 3.3 \cdot \ln n$
  - Pokud M není uvedeno, předpokládá se, že M=10





3.2 Diskrétní náhodná veličina

- Další možnosti histogramů:
  - histogram(x)
    - ‘Normalization’
      - ‘count’ – na svislé ose počet jednotek
      - ‘probability’ – na svislé ose pravděpodobnost vzniku
      - ‘countdensity’ – počet položek / šířka intervalu
      - ‘pdf’ – odhad hustoty pravděpodobnosti
      - ‘cumcount’ – kumulativní počet měření
      - ‘cdf’ – odhad distribuční funkce
  - histcounts(x) – zjistí počet prvků v daném intervalu
  - histogram2(x,y) – 2D histogram
  - histcounts2(x,y) – zjistí počet prvků v 2D intervalu

3.3 Spojitá náhodná veličina

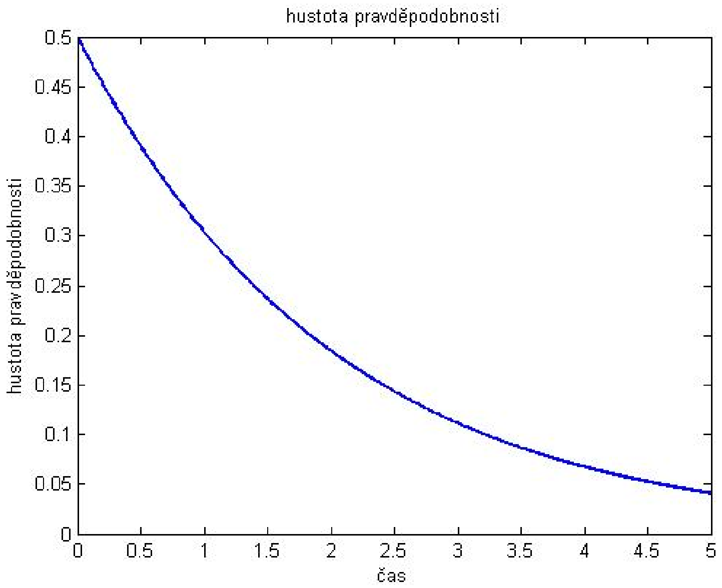
- Spojitá náhodná veličina nastává, jestliže množina výsledků náhodného pokusu nabývá všech hodnot z určitého intervalu (obecně nespočetný počet).
  - Například: životnost výrobku, náhodně vybrané reálné číslo, přesně naměřený rozměr.
- Náhodná veličina X má spojité rozdělení pravděpodobnosti právě tehdy, má-li spojitou distribuční funkci.
  - Pro popis nelze použít pravděpodobnostní funkci, protože pravděpodobnost intervalu o délce 0 je nulová.

3.3 Spojitá náhodná veličina

- Pro popis spojité náhodné veličiny se používá hustota pravděpodobnosti $f(t)$.
- Hustota pravděpodobnosti $f(t)$ spojité náhodné veličiny je reálná nezáporná funkce, která je definována:

$$f(x) = \frac{dF(x)}{dx} \qquad\qquad F(X = a) = \int_{-\infty}^{a} f(x)dx$$

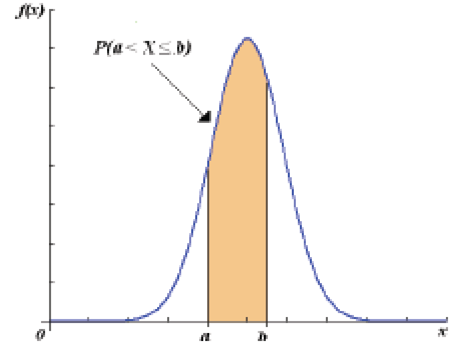
- Hustota pravděpodobnosti má následující vlastnosti:
  - $f(x) \geq 0$ hustota je nezáporná
  - $\int_{-\infty}^{\infty} f(x)dx = 1$ plocha pod křivkou je rovna 1
  - $\lim_{x\to-\infty} f(x) = \lim_{x\to\infty} f(x) = 0$



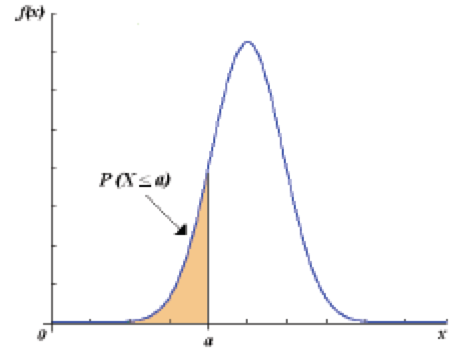
Poznámka: náhodná veličina, která nabývá hodnot v intervalu $\langle 0, \infty)$ označujeme t.

3.3 Spojitá náhodná veličina

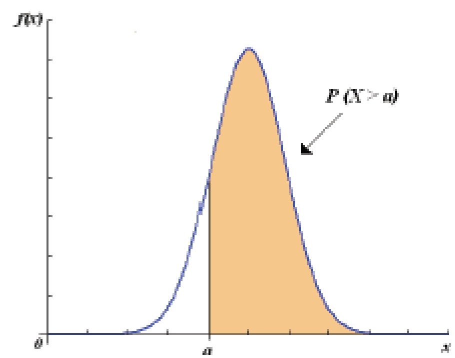
- Základní vlastnosti:
  - $P(a < X \leq b) = \int_a^b f(x)dx$
    - Např. pravděpodobnost, že se výrobek porouchá v intervalu $\langle a, b \rangle$.



  - $P(X \leq a) = \int_{-\infty}^a f(x)dx$
    - Pravděpodobnost, že se výrobek porouchá do času t.



  - $P(X > a) = \int_a^\infty f(x)dx = 1 - \int_{-\infty}^a f(x)dx$
    - Pravděpodobnost, že životnost výrobku je delší než a



3.3 Spojitá náhodná veličina

- Hustota pravděpodobnosti je definována funkcí $f(t) = 0.5e^{-0.5\cdot t}$. Zjistěte $\mathrm{P}(t > 3)$.
  - Například pravděpodobnost, že životnost výrobku je delší než 3 časové jednotky

  - $P(t > 3) = \int_3^\infty 0.5e^{-0.5\cdot t} = -e^{-0.5t} + c = [1 - e^{-0.5t}]_3^\infty =$
  - $P(t > 3) = 0.2231$

- Zjistěte pravděpodobnost pro stejnou hustotu pravděpodobnosti, ale v intervalu <1,3>.
  - $P(1 < t < 3) = \int_1^3 0.5e^{-0.5\cdot t} = [1 - e^{-0.5t}]_1^3$
  - $P(1 < t < 3)$ =0.3834

- Způsob výpočtu v matlabu

syms t	inicializace symbolických výpočtů v matlabu
f=0.5*exp(-0.5*t)	definice funkce
F=int(f,1,3)	meze od 1 do 3

3.3 Spojitá náhodná veličina

- Hustota pravděpodobnosti je definována pro nezáporná x funkcí $f(t) = 0.5e^{-0.5 \cdot t}$. Zjistěte distribuční funkci.
- Poznámka:

$$F(t) = \int_0^t f(t)dt = \int_0^t 0.5e^{-0.5 \cdot t}dt = \left|-e^{-0.5 \cdot t}\right|_0^t = -e^{-0.5 \cdot t} - (-1) = 1 - e^{-0.5 \cdot t}$$

- Zkontroluji limity $\lim_{t \to 0} F(t) = 0$ $\lim_{t \to \infty} F(t) = 1$, souhlasí proto posun $c = 0$.
- Způsob výpočtu v matlabu

syms t	inicializace symbolických výpočtů v matlabu
f=0.5*exp(-0.5*t)	definice funkce
F=int(f,0,t)	meze od 0 do t, F=1 - exp(-t/2)

- Distribuční funkce je dána pro nezáporná x funkcí $F(t) = 1 - e^{-0.5 \cdot t}$. Vypočtěte hustotu pravděpodobnosti.
- $f(t) = \frac{dF(t)}{dt} = (1 - \ e^{-0.5 \cdot t})' = 0.5e^{-0.5 \cdot t}$
- Způsob výpočtu v matlabu

syms t	inicializace symbolických výpočtů v matlabu
F=1-exp(-0.5*t)	definice distribuční funkce
f=diff(F,t)	

3.3 Spojitá náhodná veličina

- Spojitá náhodná veličina má hustotu pravděpodobnosti na intervalu $\langle -1,1\rangle$: $f(x) = \frac{3}{4}\cdot(1-x)\cdot(1+x)$. Mimo tento interval je nulová. Vypočtěte distribuční funkci.
  - Distribuční funkce je:

    F(x)=int(f,-1,x) $F(x) = \frac{1}{4}(-x^3+3x+2)+c$
  - c je nutno dopočítat, aby v bodě x=-1 byla F(x)=0, a zároveň v bodě x=1 byla F(x)=1.
  - Dosadím do výrazu $F(x) = \frac{1}{4}(-x^3+3x+2)$ levý kraj intervalu a zjistím, že $F(x=-1) = \frac{1}{4}(-x^3+3x+2) = 0$
  - $F(x=-1)$ se musí rovnat 0, proto $c=0$.

$$F(x) = \frac{1}{4}(-x^3+3x)+\frac{1}{2}$$

  - (platí pro distribuční funkci v intervalu <-1,1>)

3.4 Transformace náhodné veličiny

- Známe rozdělení náhodné veličiny X a chceme zjistit rozdělení náhodné veličiny Y, která je funkcí náhodné veličiny X: $Y = g(X)$.
  - Např. g(X) je $Y = \log(X)$, $Y = 5X + 6$, $Y = X^2$
- Pro diskrétní náhodnou veličinu X s pravděpodobnostní funkcí p_x má transformovaná náhodná veličina $Y = g(X)$ pravděpodobnostní funkci:

$$p(y) = \sum_{y=g(x)} p(x)$$

- Pro spojitou náhodnou veličinu X a spojitě diferencovatelnou funkci g je hustota pravděpodobnosti $f_Y(y)$ náhodné veličiny Y rovna:

$$f_Y(y) = f_X(g^{-1}(y)) \cdot \left| \frac{dg^{-1}(y)}{dy} \right|$$

3.4.1 Transformace diskrétní náhodné veličiny

- Příklad - transformace diskrétní náhodné veličiny
  - Pravděpodobnostní funkce náhodné veličiny X je:

x	-2	-1	0	1	2
$p(x)$	0,1	0,25	0,15	0,3	0,2

  - Vytvořme pravděpodobnostní funkci pro náhodnou veličinu $Y = X^2 - 1$
  - K tabulce přidáme řádek funkčních hodnot proměnné y po transformaci

y	3	0	-1	0	3
x	-2	-1	0	1	2
$p(x)$	0,1	0,25	0,15	0,3	0,2

  - Tabulku vzestupně setřídíme podle y a sloučíme sloupce se stejným y. Tím obdržíme pravděpodobnostní funkci proměnné Y

y	-1	0	0	3	3
x	0	-1	1	-2	2
$p(x)$	0,15	0,25	0,3	0,1	0,2

y	-1	0	3
p(y)	0.15	0.55	0.3

3.4.2 Transformace spojité náh. vel.

- Transformace spojité náhodné veličiny

$$f_Y(y) = f_X(g^{-1}(y)) \cdot \left|\frac{dg^{-1}(y)}{dy}\right|$$

- Co znamená transformace náhodné veličiny $Y = g(X)$?
  - Posun n. v. o a: $\mathrm{Y} = X - a$
  - Logaritmování n. v. : $\mathrm{Y} = \ln \mathrm{X}$

  Vynásobení konstantou n. v. c: $\mathrm{Y} = cX$

  Odmocnina n. v. : $Y = \sqrt{X}$

- Význam členů ve vzorci:
  - $g^{-1}(y)$ inverzní funkce k transformační funkci
  - $\left|\frac{dg^{-1}(y)}{dy}\right|$ zderivujeme funkci $g^{-1}(y)$ podle y a výsledek upravíme absolutní hodnotou.
  - $f_X(g^{-1}(y))$ každé x v původní hustotě pravděpodobnosti nahradíme funkcí $g^{-1}(y)$

- Inverzní funkce je definována v matlabu pomocí funkce g=finverse(f)
  - >> syms x
  - >> y=sqrt(x)-1;
  - >> g=finverse(y)
  - g =(x + 1)^2

3.4.2 Transformace spojité náh. vel.

- Transformace spojité náhodné veličiny

$$f_Y(y) = f_X(g^{-1}(y)) \cdot \left|\frac{dg^{-1}(y)}{dy}\right|$$

Příklad: Mějte hustotu pravděpodobnosti $f(x) = \frac{1}{8}x$ na intervalu $\langle 0,4 \rangle$. Mimo interval je $f(x)$ nulová.

Posuňte hustotu pravděpodobnosti: $Y = 2X - 5$.

1) Přetransformujeme definiční obor náhodné veličiny. Transformace je prostá funkce. Interval $\langle 0,4 \rangle$ se přetransformuje na $\langle -5,3 \rangle$.
2) Stanovení $g^{-1}(y)$
   - $g^{-1}(y)$ je inverzní funkce k transformační funkci
   - $g^{-1}(y)$ určíme, když upravíme transformační funkci $\mathrm{Y} = f(X)$ do tvaru $\mathrm{X} = f(Y)$
   - Např. $Y = \sqrt{X} - 1$ upravíme na $\mathrm{X} = (Y + 1)^2$
   - V našem příkladu to je $X = \frac{Y+5}{2}$

3.4.2 Transformace spojité náh. vel.

3) Určení $f_X(g^{-1}(y))$ dosadím za každé x v původní hustotě pravděpodobnosti funkci $g^{-1}(y)$

$$f_X\left(g^{-1}(y)\right) = \frac{1}{8} \cdot \frac{y+5}{2}$$

4) Zjištění $\left|\frac{dg^{-1}(y)}{dy}\right|$, derivace $g^{-1}(y)$ podle y a absolutní hodnota

$$\left|\frac{dg^{-1}(y)}{dy}\right| = \left(\frac{y+5}{2}\right)' = \frac{1}{2}$$

- Derivace je kladná, proto absolutní hodnota nemá žádný vliv.

5) Nová hustota pravděpodobnosti je:

$$g(y) = \frac{1}{8} \cdot \frac{y+5}{2} \cdot \frac{1}{2} = \frac{y+5}{32}$$

na intervalu $< -5{,}3)$, jinde je nulová.

6) Správnost hustoty lze ověřit pomocí výpočtu distribuční funkce a dosazení krajních intervalů.

$$\int_{-5}^{3} g(y) = 1$$

$$F(y) = \left[\frac{y^2}{64} + \frac{5y}{32}\right]_{-5}^{3} = \frac{9}{64} + \frac{15}{32} - \frac{25}{64} - \left(-\frac{25}{32}\right) = \frac{40}{32} - \frac{16}{64} = 1$$

3.4.2 Transformace spojité náh. vel.

Příklad 2: $$f_Y(y) = f_X(g^{-1}(y)) \cdot \left|\frac{dg^{-1}(y)}{dy}\right|$$

- Transformační funkce $\mathrm{Y} = \sqrt{X} - 1$; původní hustota pravděpodobnosti je: $f(x) = e^{-x}$, pro kladné x, jinde 0.

1) Určení nového definičního oboru:
   - $x = 0$ $y = -1$
   - $x = \infty$ $y = \infty$ Nový definiční obor je $\langle -1, \infty\rangle$

2) Stanovení $g^{-1}(y)$ je inverzní funkce k transformační funkci: $X = (Y+1)^2$
   - \>> syms x
   - \>> y = sqrt(x) - 1
   - \>> g = finverse(y)

3) Určení $f_X(g^{-1}(y))$ dosadím za každé x v původní hustotě pravděpodobnosti funkci $g^{-1}(y)$

$$f_X\left(g^{-1}(y)\right) = e^{-(y+1)^2}$$

4) Zjištění $\left|\frac{dg^{-1}(y)}{dy}\right|$, derivace $g^{-1}(y)$ podle y a absolutní hodnota

$$\left|\frac{dg^{-1}(y)}{dy}\right| = ((Y+1)^2\ )' = 2(y+1)$$

   - Derivace je kladná na intervalu $\langle -1, \infty\rangle$, proto absolutní hodnota nemá žádný vliv.

5) Nová hustota pravděpodobnosti je:

$$g(y) = e^{-(y+1)^2} \cdot 2(y+1),$$

na intervalu $< -1, \infty)$, jinde je nulová.

6) Správnost hustoty lze ověřit pomocí výpočtu distribuční funkce a dosazení krajních intervalů.
   - \>> syms x
   - \>> F=int(2*exp(-(x+1)*(x+1))*(x+1),-1,inf)
   - F = 1

3.5 Číselné charakteristiky náhodné veličiny

- 3.5.1 Úvod do číselných charakteristik
- 3.5.2 Obecné momenty
- 3.5.3 Střední hodnota
- 3.5.4 Centrální momenty
- 3.5.5 Rozptyl a směrodatná odchylka
- 3.5.6 Šikmost
- 3.5.7 Špičatost
- 3.5.8 Kvantily
- 3.5.9 Modus

3.5.1 Úvod do číselných charakteristik

- Rozdělení pravděpodobnosti je jednoznačně popsáno
  - distribuční funkcí $F(x)$
  - hustotou pravděpodobnosti $f(x)$ spojitá n.v.
  - pravděpodobnostní funkcí $\mathrm{p}(x)$ diskrétní n.v.
- Snahou je popsat rozdělení pomocí několika málo čísel, které popisují hlavní charakteristiky náhodné veličiny. Jsou to například:
  - Střední hodnota $E(X)$
  - Rozptyl $D(X)$
  - Směrodatná odchylka $\sigma(X)$
- Výše uvedené veličiny vycházejí z obecných a centrálních momentů r-tého řádu

3.5.2 Obecné momenty

- Obecný moment r-tého řádu:
  - Značí se $E(X^r)$, nebo μ_r
  - Pro diskrétní náhodnou veličinu je definován:

$$\mu_r = \sum_i x_i^r \cdot p(x_i)$$

  - Pro spojitou náhodnou veličinu je definován:

$$\mu_r = \int_{-\infty}^{\infty} x^r \cdot f(x) dx$$

- Velmi často se využívá 1. obecný moment – střední hodnota

3.5.3 Střední hodnota

- Střední hodnota - $E(X)$, 1. obecný moment
  - Pro diskrétní náhodnou veličinu je definován:
$$\mu_r = \sum_i x_i \cdot P(x_i)$$
  - Pro spojitou náhodnou veličinu je definován:
$$\mu_r = \int_{-\infty}^{\infty} x \cdot f(x)dx$$
- Interpretace (pro zjištěná data) – průměr všech realizací náhodné veličiny (při shodné váze každé realizace)
  - Aritmetický průměr výsledků, průměrný plat v ČR, střední doba do poruchy výrobku
- Vlastnosti střední hodnoty:
  - $E(aX + b) = aE(X) + b$ $\forall a, b$
  - $E(\sum_i X_i) = \sum_i E(X_i)$
  - $E(\prod_i X_i) = \prod_i E(X_i)$ jestliže X_i jsou navzájem nezávislé

3.5.3 Střední hodnota

- Střední hodnota v matlabu
  - >> syms x   definování proměnných
  - >> fx=0.25*exp(-x/4);   definovaní hustoty pravděpodobnosti
  - >> EX=int(x*fx,0,inf)   výpočet střední hodnoty, dolní a horní mez, kde je hustota nenulová
  - EX =4

- Například 3. obecný moment v matlabu
  - >> syms x
  - >> fx=0.25*exp(-x/4);
  - >> my3=int(x^3*fx,0,inf)
  - my3 =384

3.5.4 Centrální momenty

- Centrální moment r-tého řádu:
  - Značí se $\mu_r{}'$
  - $\mu_r' = E(X - E(X))^r$ pro r=2,3,4,…
  - Pro diskrétní náhodnou veličinu je definován:
$$\mu_r{}' = \sum_i (x_i - E(X))^r \cdot p(x_i)$$
  - Pro spojitou náhodnou veličinu je definován:
$$\mu_r{}' = \int_{-\infty}^{\infty} (x - E(X))^r \cdot f(x)dx$$
- Velmi často se využívá 2. centrální moment – rozptyl, nebo jeho odmocnina směrodatná odchylka
- Využívá se i 3. a 4. centrální moment – šikmost a špičatost.

3.5.5 Rozptyl a směrodatná odchylka

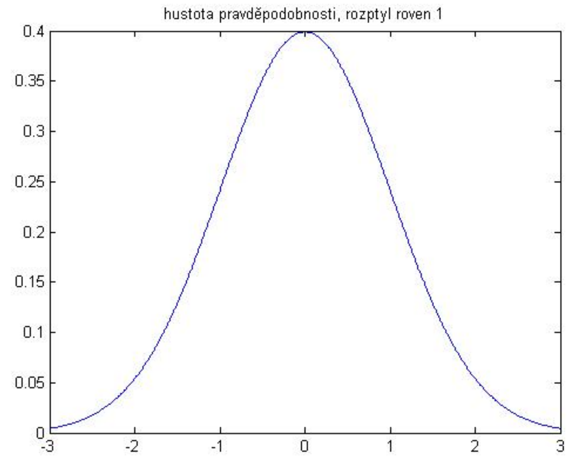
- Rozptyl - $\mathrm{D}(X)$ nebo také σ^2, 2. centrální moment
  - Pro diskrétní náhodnou veličinu je definován:

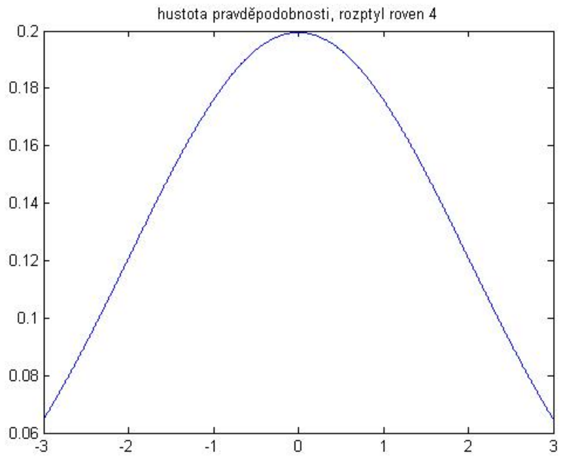
$$\sigma^2 = \sum_i (x_i - E(X))^2 \cdot p(x_i)$$

  - Pro spojitou náhodnou veličinu je definován:

$$\sigma^2 = \int_{-\infty}^{\infty} (x - E(X))^2 \cdot f(x) dx$$

- Rozptyl je parametr vyjadřující variabilitu (rozptýlenost) realizací od střední hodnoty.
  - Jednotkou rozptylu je kvadrát střední hodnoty. $[m^2, kg^2, °C^2]$ apod.





3.5.5 Rozptyl a směrodatná odchylka

- Vlastnosti rozptylu:
  - $D(aX + b) = a^2 D(X)$ $\forall a, b$
  - $D(\sum_i X_i) = \sum_i D(X_i)$ jestliže X_i jsou navzájem nezávislé
- Výpočet v matlabu:
  - >> syms x
  - >> fx=0.25*exp(-x/4); nutno definovat hustotu pravd.
  - >> EX=int(x*fx,0,inf)
  - >> DX=int((x-EX)^2*fx,0,inf)
- Z důvodu nevhodných jednotek se využívá odmocnina z rozptylu – směrodatná odchylka σ

$$\sigma = \sqrt{\sigma^2} = \sqrt{D(X)}$$

- Směrodatná odchylka je obdobně jako rozptyl parametr vyjadřující variabilitu dat od střední hodnoty

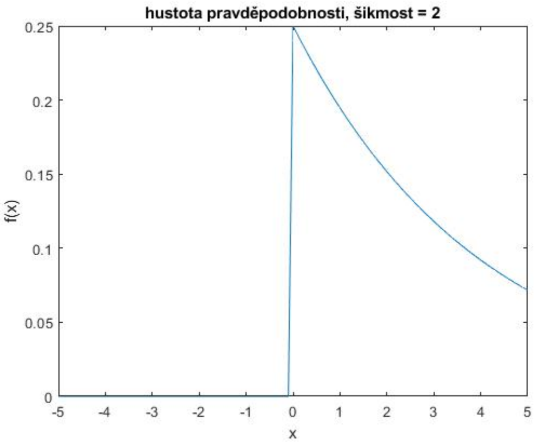
3.5.6 Šikmost

- Šikmost je mírou symetrie rozdělení pravděpodobnosti
- Definována je jako podíl 3. centrálního momentu a třetí mocniny směrodatné odchylky

$$\alpha_3 = \frac{\mu_3{}'}{\sigma^3}$$

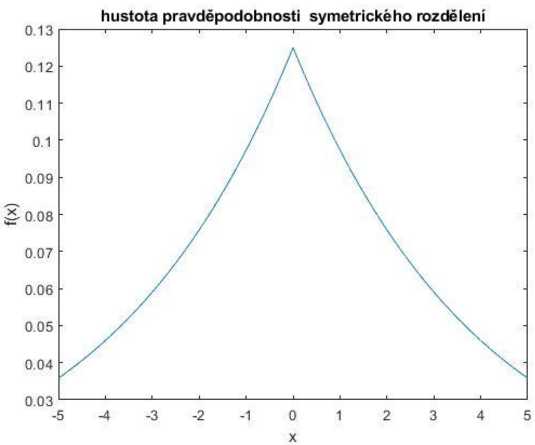
- $\alpha_3 = 0$ symetrické rozdělení
- $\alpha_3 < 0$ negativně zešikmené rozdělení
- $\alpha_3 > 0$ pozitivně zešikmené rozdělení

- Matlab:
  - >> syms x
  - >> fx=0.25*exp(-x/4); f(x) je zde defin. pouze pro nezáporná čísla
  - >> EX=int(x*fx,0,inf)
  - >> sikmost=int((x-EX)^3*fx,0,inf)/(int((x-EX)^2*fx,0,inf)^1.5)
  - sikmost = 2



- Příklad pro symetrické rozdělení (hustota z minulého příkladu je zrcadlena i do záporných čísel)

```
>> syms x
>> fx=0.125*exp(-abs(x)/4);
>> EX=int(x*fx,-inf,inf)
>> sikmost=int((x-EX)^3*fx,-inf,inf)/(int((x-EX)^2*fx,-inf,inf)^1.5)
sikmost = 0
```



3.5.7 Špičatost

- Špičatost je mírou koncentrace hodnot kolem střední hodnoty
- Definována je podílem čtvrtého centrálního momentu a čtvrté mocniny směrodatné odchylky.

$$\alpha_4 = \frac{\mu_4{}'}{\sigma^4} - 3$$

- $\alpha_4 = 0$ špičatost normálního rozdělení
- $\alpha_4 < 0$ menší špičatost než u normálního rozdělení
- $\alpha_4 > 0$ větší špičatost než u normálního rozdělení

- V některé literatuře je koeficient špičatosti definován odlišně $\alpha_4 = \frac{\mu_4{}'}{\sigma^4}$
- Matlab
  - >> syms x
  - >> fx=0.25*exp(-x/4); hustota je zde definována pouze pro nezáporná čísla
  - >> EX=int(x*fx,0,inf)
  - >> spicatost=int((x-EX)^4*fx,0,inf)/(int((x-EX)^2*fx,0,inf)^2)

3.5.8 Kvantily

- Značí se x_p
- Představuje takovou hodnotu, že pravděpodobnost, že náhodná veličina nabude hodnoty menší než x_p je 100*p procent.
- 50% kvantil je nazýván medián
- 25% a 75% kvantil je nazýván dolní / horní kvartil
- 1% kvantil je nazýván percentil

Způsob určení kvantilu z distribuční funkce

Matlab:

```
syms t
Fx=1-exp(-0.1*t)            %distribuční funkce
kvantil=solve(Fx==0.25) %dolní kvartil
```

3.5.9 Modus

- Značí se $\hat{x}$.
- Jedná se o nejčetnější hodnotu
- Pro diskrétní náhodnou veličinu

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

- Pro spojitou náhodnou veličinu

$$f(\hat{x}) \geq f(x), \forall x$$

- Náhodná veličina může mít i více modů.

3.6 Charakteristiky numerických proměnných

- 3.6.1 Základní charakteristiky
- 3.6.2 Aritmetický průměr
- 3.6.3 Harmonický a geometrický průměr
- 3.6.4 Modus a shorth
- 3.6.5 Kvantily, kvartily a medián
- 3.6.6 Výběrový rozptyl, směrodatná odchylka
- 3.6.7 MAD
- 3.6.8 Výběrová šikmost
- 3.6.9 Výběrová špičatost

3.6.1 Základní charakteristiky

- Příkazy pro stanovení základních charakteristik v matlabu
  - Maximum
    max(x) – vrátí maximum
    nanmax(x) – zjistí maximum pouze z čísel, ostatní ignoruje
  - Minimum
    min(x) – vrátí minimum
    nanmin(x) – zjistí minimum pouze z čísel, ostatní ignoruje
  - Rozdíl maxima a minima
    range(x) – $x_{max} - x_{min}$
  - Součet prvků ve vektoru
    sum(x)
    nansum(x) – zjistí součet z číselných hodnot, ostatní ignoruje
  - Rozsah výběru
    length(x) – počet prvků ve vektoru
    size (x) – velikost matice

3.6.2 Aritmetický průměr

- Vstupem jsou data: $\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n}$
  - x_i - vstupní data, n – počet dat
- Vstupem je četnostní tabulka:

$$\bar{x} = \frac{x_1 \cdot n_1 + x_2 \cdot n_2 + \cdots + x_k \cdot n_k}{n_1 + n_2 + \cdots + n_k}$$

  - n_i - představuje buď váhu, četnost, nebo procento výskytu
  - x_i - data
- Vstupem jsou pravděpodobnosti jevů $\bar{x} = x_1 \cdot p_1 + x_2 \cdot p_2 + \cdots + x_k \cdot p_k$
- $\sum_{i=1}^{k} p_i = 1$
- Matlab: mean(x)
  nanmean(x) – počítá průměr pouze z číselných hodnot

- Vlastnosti střední hodnoty:
  - Přičteme-li všem datům konstantu c, střední hodnota se zvýší o konstantu c.
  - Vynásobíme-li všechna data konstantou c, střední hodnota se c-krát zvýší.

3.6.3 Harmonický a geometrický průměr

- Pro výpočet průměru převrácených hodnot se využívá harmonický průměr:
- Pro data: $$\overline{x_H} = \frac{n}{\sum_{i=1}^{n} \frac{1}{x_i}}$$
- Pro četnostní tabulku $$\overline{x_H} = \frac{\sum_{i=1}^{k} n_i}{\sum_{i=1}^{k} \frac{n_i}{x_i}}$$
- Využití například pro úlohy s nepřímou úměrností (úlohy práce).
- Matlab: harmmean(x)

3.6.3 Harmonický a geometrický průměr

- Pro výpočet průměru představující relativní změny se využívá geometrický průměr:
- Pro data: $\overline{x_G} = \sqrt[n]{x_1 \cdot x_2 \cdot \dots \cdot x_n}$
- Pro četnostní tabulku

$$\overline{x_G} = \sqrt[n]{{x_1}^{n_1} \cdot {x_2}^{n_2} \cdot \dots \cdot {x_k}^{n_k}}$$

$$n = \sum_{i=1}^{k} n_i$$

- Matlab: geomean(x)

3.6.4 Modus a shorth

- Modus je nejčastější hodnota ze vstupního souboru.
- $\hat{x}$ modus
- Matlab: mode(x)
  - Pozn.: při více modů funkce vrací první

```
x=normrnd(0,1,1,1001);   %definování vstupního vektoru
delka=length(x);         %určení délky vektoru
x=sort(x);               %setřídění od min po max
nejmensi_rozdil=range(x);   %odhad maximálního rozdílu
if mod(1001,2)==0        %pro vektor sudé délky
    for i=1:delka/2
    rozdil=x(i+delka/2)-x(i);
    if rozdil<nejmensi_rozdil
        nejmensi_rozdil=rozdil;
        min=i;
        max=i+delka/2;
    end
    end
else                     %pro vektor liché délky
    delka=delka+1;
    for i=1:(delka/2-1)
    rozdil=x(i+delka/2)-x(i);
    if rozdil<nejmensi_rozdil
        nejmensi_rozdil=rozdil;
        min=i;
        max=i+delka/2;
    end
    end
end
vysledek_shorth=x(max)-x(min)
```

- Shorth je nejmenší interval, v němž leží alespoň 50 % vstupních dat.

3.6.5 Kvantily, kvartily a medián

- $100 \cdot p$ % kvantil proměnné x je hodnota, která rozděluje soubor na $100 \cdot p$ % menších hodnot od zbytku. Značíme jej x_p.
  – Kvantily tvoří inverzní funkci k distribuční funkci.
  – 50% kvantil se nazývá medián
    - rozděluje datový soubor na dvě skupiny takové, že v první je 50 % dat menších než medián a obdobně v druhé je 50 % dat větších než medián.
  – 25% kvantil se nazývá dolní kvartil
  – 75% kvantil se nazývá horní kvantil

3.6.5 Kvantily, kvartily a medián

- Určení z naměřených dat:
  - Setřídím data od min po max a stanovím jim pořadí $z_1, z_2, \dots, z_n$.
  - 100p % kvantil je roven proměnné s pořadím:

$$z_p = n \cdot p + 0.5$$

    - Není-li z_p celé číslo, pak kvantil určíme jako průměr prvků z pořadím $\lfloor z_p \rfloor$ a $\lceil z_p \rceil$.
- Interkvartilové rozpětí
  - Rozdíl mezi horním a dolním kvartilem
  - Pomocí interkvartilového rozpětí lze porovnávat variabilitu souboru.

3.6.5 Kvantily, kvartily a medián

- Matlab
  - median(x) $x_{0.5}$ určí medián z vektoru
  - nanmedian(x) určí medián z číselných hodnot ve vektoru, ostatní ignoruje
  - quantile(x,p) určí p kvantil z dat

    např. quantile(x,0.5) je 50% kvantil
  - prctile(x,p) x_p určí p% kvantil z dat

    např. prctile(x,50) je 50% kvantil
  - iqr(x) interkvartilové rozpětí

3.6.6 Výběrový rozptyl, směrodatná odchylka, variační koeficient

- Výběrový rozptyl a směrodatná odchylka popisují variabilitu dat kolem střední hodnoty.
- Výběrový rozptyl je definován:

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

  - Čitatel je součet kvadrátů odchylek od průměru.
  - Jmenovatel je zmenšen o 1, protože se ztrácí jeden stupeň volnosti odhadem střední hodnoty. Z jednoho čísla nelze vypočítat odchylku od střední hodnoty.
  - Ve statistice často, když odhadujeme k parametrů se zmenší počet stupňů volnosti o k a statistika je např. dělena $n-k$, kde n je počet naměřených dat.

3.6.6 Výběrový rozptyl, směrodatná odchylka, variační koeficient

- Základní vlastnosti výběrového rozptylu
  - Konstantní hodnoty mají nulový rozptyl
  - Přičtením konstanty ke všem datům se rozptyl nezmění
  - Vynásobíme-li data konstantou k, rozptyl se změní k^2 krát.
- Nevýhoda rozptylu je, že jednotky jsou v kvadrátu oproti střední hodnotě.
- Tuto nevýhodu odstraňuje směrodatná odchylka.

3.6.6 Výběrový rozptyl, směrodatná odchylka, variační koeficient

- Výběrová směrodatná odchylka s

$$s = \sqrt{rozptylu} = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1}}$$

- Matlab:
  - Určení libovolného centrálního momentu moment(x, řád momentu)
  - Výběrový rozptyl var(x)
  - Výběrová směrodatná odchylka std(x)
  - Výběrový rozptyl vypočtený z číselných hodnot, ostatní ignoruje nanvar(x)
  - Výběrová směrodatná odchylka z číselných hodnot, ostatní ignoruje nanstd(x)

- Variační koeficient
  - Vhodný pro porovnání míry variabilit mezi sebou
  - Například je více variabilní plat v ČR, nebo v Německu
  - Použitelný pouze pro kladně naměřené hodnoty (záporné výsledky nemohou nastat)

$$V_x = \frac{s}{\bar{x}}$$

3.6.6 Výběrový rozptyl, směrodatná odchylka, variační koeficient

	Rozptyl	Směrodatná odch.
• Vstupem jsou data:	$s^2 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1}$	$s = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1}}$

- Vstupem jsou data:
  - x_i - vstupní data, n – počet dat
- Vstupem je četnostní tabulka:

$$s^2 = \frac{\sum_{i=1}^{n} n_i \cdot (x_i - \bar{x})^2}{\left(\sum_{i=1}^{n} n_i\right) - 1} \qquad s = \sqrt{\frac{\sum_{i=1}^{n} n_i \cdot (x_i - \bar{x})^2}{\left(\sum_{i=1}^{n} n_i\right) - 1}}$$

  - n_i - představuje četnost
  - x_i - data
- Vstupem jsou pravděpodobnosti jevů (platí $\sum_{i=1}^{n} p_i = 1$):

$$s^2 = \sum_{i=1}^{n} p_i \cdot (x_i - \bar{x})^2 \qquad s = \sqrt{\sum_{i=1}^{n} p_i \cdot (x_i - \bar{x})^2}$$

- Rada ze zkušeností - vždy se snažte počítat statistické ukazatele přímo z naměřených dat. Můžete zjistit:
  - Opticky odlehlé hodnoty,
  - Data vznikla spojením různých souborů, mezi kterými je systematická chyba (posun středních hodnot),
  - Odhadnout rozptyl a symetričnost dat.
  - Víte s kolika naměřenými hodnotami pracujete.
- V praxi se Vám řada lidí bude snažit „usnadnit práci“ dodáním pouze středních hodnot a rozptylů a nikoliv naměřených dat. Toto bylo užitečné v době „logaritmických pravítek“. Dnes však SW vypočte statistické výsledky prakticky okamžitě.

3.6.7 MAD

- MAD –
  - a) Střední hodnota absolutních odchylek od střední hodnoty
  - b) Medián absolutních odchylek od mediánu
- Způsob výpočtu ad a:
  - Setřídíme hodnoty od min po max a určíme střední hodnotu
  - Vypočteme rozdíly v absolutní hodnotě naměřených dat od střední hodnoty
  - Z výsledků stanovíme střední hodnotu
- Matlab
  - mad(x,0) střední hodnota absolutních odchylek od střední hodnoty
  - mad(x,1) medián absolutních odchylek od mediánu

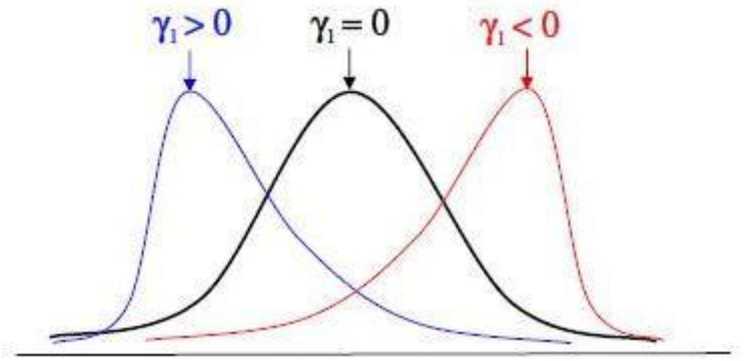
3.6.8 Výběrová šikmost

- Vyjadřuje symetrii rozložení hodnot kolem průměru

$$a = \frac{n}{(n-1)\cdot(n-2)} \cdot \frac{\sum_{i=1}^{n}(x_i - \bar{x})^3}{s^3}$$

  - $a = 0$ data jsou kolem průměru rozloženy symetricky

- Matlab: skewness(x)

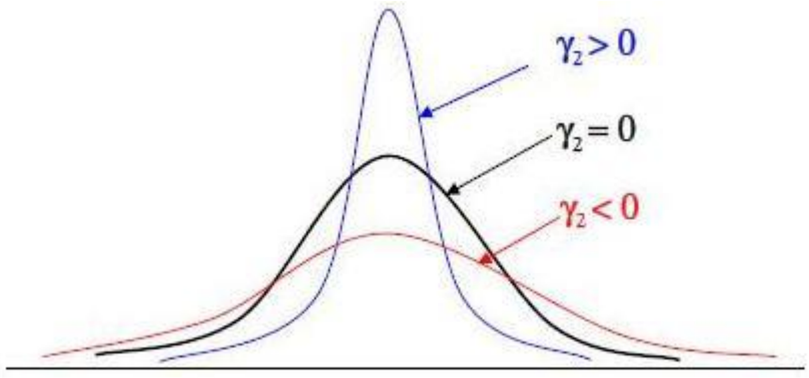


3.6.9 Výběrová špičatost

- Vyjadřuje koncentraci naměřených hodnot kolem jejího průměru a porovnává je s normálním rozdělením

$$b = \frac{n \cdot (n+1)}{(n-1) \cdot (n-2) \cdot (n-3)} \cdot \frac{\sum_{i=1}^{n}(x_i - \bar{x})^4}{s^4} - 3 \cdot \frac{(n-1)^2}{(n-2) \cdot (n-3)}$$

  - $b = 0$ špičatost odpovídá normálnímu rozdělení
- Matlab: kurtosis(x)



3.7 Identifikace odlehlých měření

- Identifikace odlehlých měření je důležité pro zjištění těch hodnot, které se mimořádně liší od ostatních a ovlivňují tím vypovídající hodnotu průměru.
- U všech identifikovaných dat je vhodné se zamyslet , zda nedošlo k chybě měření.
- Práce s odlehlým měřením
  - 1) byla zjištěna chyba při měření – data odstraníme
  - 2) nebyla zjištěna chyba měření – data ponecháme a zdůvodníme jejich ponechání
  - 3) nebyla zjištěna chyba měření – data odstraníme a zdůvodníme jejich odstranění a uvedeme důvod proč nebyly ponechány.

3.7 Identifikace odlehlých měření

- Způsob zjištění odlehlých měření
- 1) metoda z-souřadnice
  - Data z norm. rozdělení
  - Data z normálního rozdělení, symetrická
  - $x_{odlehle}$: $\left|\frac{x_i - \bar{x}}{s}\right| > 3$

```
n=length(x);
for i=1:n
    z=abs(x(i)-mean(x))/std(x);
    if z>3
        i
    end
end
```

- 2) metoda vnitřních hradeb
  - Data nesymetrická
  - $x_{odlehle} < x_{0.25} - 1.5 \cdot IQR$
  - $x_{odlehle} > x_{0.75} + 1.5 \cdot IQR$

```
n=length(x);
xdolni=quantile(x,0.25)-1.5*iqr(x);
xhorni=quantile(x,0.75)+1.5*iqr(x);
for i=1:n
    if x(i)<xdolni
        i
    end
    if x(i)>xhorni
        i
    end
end
```

3.7 Identifikace odlehlých měření

- Naměřená data: 4,7,10,11,12,12,12,13,13,15,22 (symetrická data)
- Základní číselné veličiny:
  - $\bar{x} = 11.8$ $s^2 = 22.62$ $s = 4.75$ $x_{0.5} = 12$
- Zjištění odlehlých hodnot
  - Metoda z-souřadnice
    - $x_{odlehle}$: $\left|\frac{x_i - \bar{x}}{s}\right| > 3$
    - $x_{odlehle} < \bar{x} - 3s = 11.8 - 3 \cdot 4.75 = -2.05$
    - $x_{odlehle} > \bar{x} + 3s = 11.8 + 3 \cdot 4.75 = 26.45$
  - Metoda vnitřních hradeb
    - $x_{0.25} = 10$ $x_{0.75} = 13$ $IQR = 3$
    - $x_{odlehle} < x_{0.25} - 1.5 \cdot IQR = 5.5$ $x_{odlehle} > x_{0.75} + 1.5 \cdot IQR = 17.5$

3.7 Identifikace odlehlých měření

- Naměřená data: 4,7,10,16,18,22,28,35,49,65,81,145
  (nesymetrická data)
- Základní číselné veličiny:
  - $\bar{x} = 40$ $s^2 = 1659.1$ $s = 40.73$ $x_{0.5} = 25$
-
- Zjištění odlehlých hodnot
  - Metoda z-souřadnice
    - $x_{odlehle}$: $\left|\frac{x_i - \bar{x}}{s}\right| > 3$
    - $x_{odlehle} < \bar{x} - 3s = 40 - 3 \cdot 40.73 = -82.2$
    - $x_{odlehle} > \bar{x} + 3s = 40 + 3 \cdot 40.73 = 162.2$
  - Metoda vnitřních hradeb
    - $x_{0.25} = 13$ $x_{0.75} = 57$ $IQR = 44$
    - $x_{odlehle} < x_{0.25} - 1.5 \cdot IQR = 13 - 66 = -53$
    - $x_{odlehle} > x_{0.75} + 1.5 \cdot IQR = 57 + 66 = 123$

3.8 Příkazy v Matlabu

- Symbolické výpočty
  - Definování proměnných syms x
  - Inverzní funkce finverse(x)
  - Derivace funkce diff(x)
  - Integrál funkce int(x)
  - Určitý integrál int(x,dolní mez, horní mez)
  - Stanovení kvantilu solve(x==hodnota)

3.8 Příkazy v Matlabu

- Základní výpočty
  - Maximum max(x) nanmax(x)
  - Minimum min(x) nanmin(x)
  - Rozdíl maxima a minima range(x)
  - Součet prvků ve vektoru sum(x) nansum(x)
  - Počet prvků ve vektoru length(x)
- Výpočty z naměřených dat
  - Aritmetický průměr mean(x) nanmean(x)
  - Harmonický průměr harmmean(x)
  - Geometrický průměr geomean(x)
  - Modus mode(x)
  - Medián median(x) nanmedian(x)
  - Kvantily quantile(x,p) prctile(x,p)
  - Interkvartilové rozpětí iqr(x)
  - Výběrový rozptyl var(x) nanvar(x)
  - Výběrová směrodatná odchylka std(x) nanstd(x)
  - MAD mad(x,0) mad(x,1)
  - Výběrová šikmost skewness(x)
  - Výběrová špičatost kurtosis(x)

3.8 Příkazy v Matlabu

- Grafy – záložka v horní liště plots
  - Histogram histogram(x)
  - Sloupcový graf bar(x,y)
  - Graf funkce plot(x,y)
  - Vykreslení bodů scatter(x,y)
  - Koláčový graf pie(x)
- Popisky:
  - Název grafu title(‘text’)
  - X-ová osa xlabel(‘text’)
  - Y-ová osa ylabel(‘text’)

4 Diskrétní rozdělení pravděpodobnosti

- 4.1 Hypergeometrické rozdělení
- 4.2 Binomické rozdělení
- 4.3 Rozdělení odvozená z binomického rozdělení
- 4.4 Multinomické rozdělení
- 4.5 Poissonovo rozdělení
- 4.6 Aproximace diskrétních rozdělení

4.1 Hypergeometrické rozdělení

- Hypergeometrické rozdělení se používá v situacích, kdy je třeba vypočítat pravděpodobnost určitého počtu úspěchů v n závislých pokusech.
- Popis pro závislé pokusy – při každém pokusu nastávají jeho odlišné podmínky
  - Závislé pokusy:
    - Pravděpodobnost, že při vytažení 3 karet (bez vracení) z balíčku budou 2 esa
    - Pokusy s losováním čísel z osudí bez jejich vracení (nemůže nastat dvakrát vylosování stejného čísla)
    - Tzv. pokusy bez vracení, podmínky pokusu jsou odlišné.
  - Nezávislé pokusy:
    - Pravděpodobnost, že při vytažení 3 karet (karty vracíme) z balíčku budou 2 esa
    - Pokusy s losováním čísel z osudí s jejich vracením (může nastat dvakrát vylosování stejného čísla)
    - Tzv. pokusy s vracením, podmínky pokusu jsou vždy stejné.

4.1 Hypergeometrické rozdělení

- Nechť soubor M prvků obsahuje K prvků s určitou vlastností a zbylých $M - K$ prvků tuto vlastnost nemá. Náhodně se ze souboru vybere N prvků, z nichž se žádný nevrací zpět. Potom pravděpodobnost, že z vylosovaných N prvků má právě x prvků danou vlastnost určíme pomocí hypergeometrického rozdělení dle vzorce:

$$P(X = x) = \frac{\binom{K}{x}\binom{M-K}{N-x}}{\binom{M}{N}}$$

4.1 Hypergeometrické rozdělení

$$P(X = k) = \frac{\overset{(1)}{\binom{K}{x}}\overset{(2)}{\binom{M-K}{N-x}}}{\underset{(3)}{\binom{M}{N}}}$$

- (1) – počet kombinací, jak lze vybrat x prvků s danou vlastností, z celkově K prvků s danou vlastností.
- (2) – počet kombinací, jak lze vybrat $(N - x)$ prvků bez dané vlastnosti z celkového množství $(M - K)$ prvků.
- (3) – počet kombinací, jak lze vybrat N-tici z M prvků.
- (1) a (2) – počet příznivých pokusů
- (3) – počet všech pokusů

4.1 Hypergeometrické rozdělení

- Distribuční funkce $P(X \leq x) = \sum_{i=0}^{x} \frac{\binom{K}{i}\binom{M-K}{N-i}}{\binom{M}{N}}$
- Střední hodnota $E(X) = N\frac{K}{M}$
- Rozptyl $D(X) = N\frac{K}{M}\left(1 - \frac{K}{M}\right)\left(\frac{M-N}{M-1}\right)$

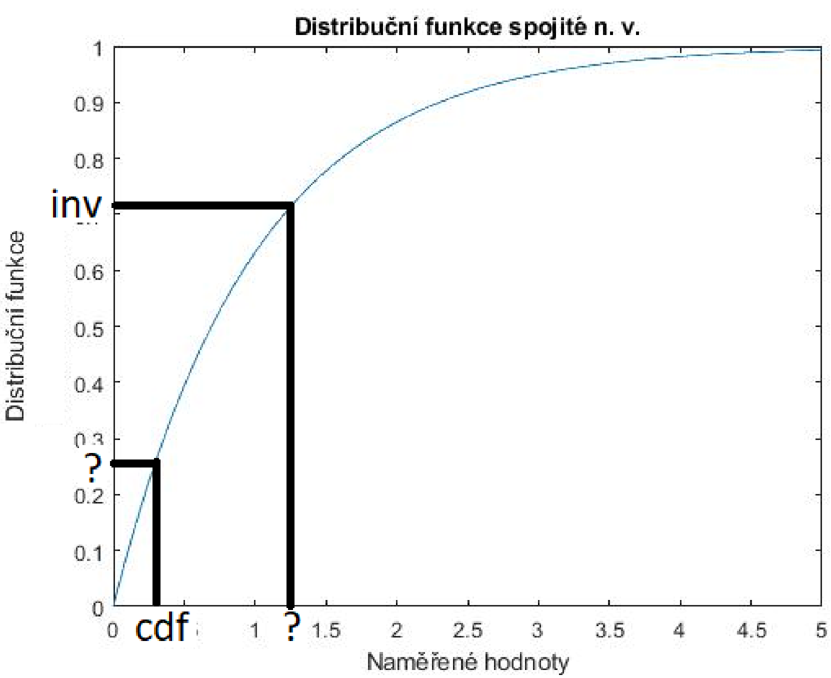
4.1 Hypergeometrické rozdělení

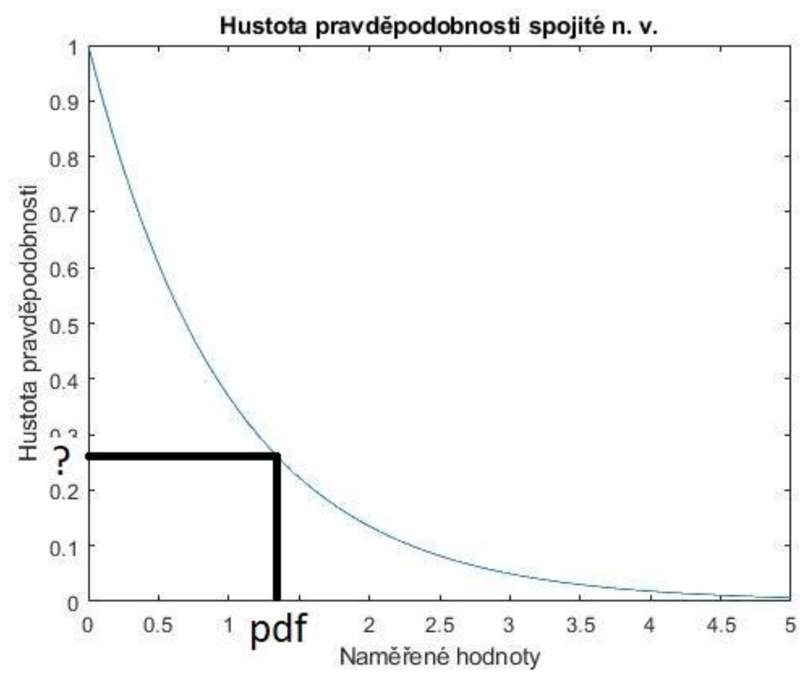
- x – počet vybraných prvků s danou vlastností M – celkový počet prvků
- K – celkový počet prvků s danou vlastností N – počet vybraných prvků

- Distribuční funkce
  - F=hygecdf (x,M,K,N)
- Pravděpodobnostní funkce
  - P=hygepdf (x,M,K,N)
- Inverzní distribuční funkce
  - x=hygeinv (pravd,M,K,N)
- Stanovení střední hodnoty a rozptylu
  - [MN,var]=hygestat(M,K,N)
- Náhodná čísla z hypergeometrického rozdělení
  - R=hygernd(M,K,N,m,n) m,n – matice náhodných čísel o rozměru $m \cdot n$

4.1 Hypergeometrické rozdělení

- Souvislost příkazů v matlabu jednotlivých rozdělení
  - Předpona – typ rozdělení
  - Přípona – co potřebujeme vypočítat
  - Např. hygecdf – hyge – hypergeometrické rozdělení,
    cdf (cumulative density function) hodnotu distribuční funkce.





4.1 Hypergeometrické rozdělení

- Př. V osudí je 10 černých míčků a 15 bílých. Z osudí vylosujeme 4 míčky, které nevracíme. Určete pravděpodobnost, že 3 míčky budou bílé a jeden černý.
- M=25 (celkem je v osudí 25 míčků)
- k=15 (15 bílých míčků, míčky s určitou vlastností)
- N=4 (losujeme 4 míčky)
- x=3 (3 míčky budou bílé)
- $P(X=3)=\frac{\binom{15}{3}\binom{25-15}{4-3}}{\binom{25}{4}}$=0.3597
- Matlab:
  - hygepdf(3,25,15,4)

4.1 Hypergeometrické rozdělení

- $P(X=0)=\dfrac{\binom{15}{0}\binom{25-15}{4-0}}{\binom{25}{4}}$=0.0166
- $P(X=1)=\dfrac{\binom{15}{1}\binom{25-15}{4-1}}{\binom{25}{4}}$=0.1423
- $P(X=2)=\dfrac{\binom{15}{2}\binom{25-15}{4-2}}{\binom{25}{4}}$=0.3735
- $P(X=3)=\dfrac{\binom{15}{3}\binom{25-15}{4-3}}{\binom{25}{4}}$=0.3597
- $P(X=4)=\dfrac{\binom{15}{4}\binom{25-15}{4-4}}{\binom{25}{4}}$=0.1079

- Všimněte si, že nejpravděpodobnější hodnota je vytažení 2 nebo 3 bílých míčků.
- Střední hodnota: $E(X)=n\frac{M}{N}=\frac{4\cdot 15}{25}=2.4$
- Nejpravděpodobnější stavy přibližně odpovídají střední hodnotě.

- Určete pravděpodobnost, že vylosujete 2 a méně bílých míčků
- hygecdf(2.5,25,15,4)=0.0166+0.1423+0.3735=0.5324
  - První parametr vzorce je 2.5. Inženýrský přístup, abychom měli jistotu, že vylosování 2 bílých míčků bude skutečně započítáno. Je tím odstraněn problém v rozdílných definicích distribuční funkce buď $F(x)=P(X\leq x)$ nebo $F(x)=P(X<x)$

4.2 Binomické rozdělení

- Binomické rozdělení se používá v situacích, kdy je třeba vypočítat pravděpodobnost určitého počtu úspěchů v n nezávislých pokusech.
  - Hypergeometrické rozdělení – závislé pokusy
    - Výsledek náhodného pokusu je závislý na předcházejících výsledcích.
    - Losování míčků označených čísly z osudí bez vracení. Nemůže nastat, že by byl dvakrát vylosován stejný.
  - Binomické rozdělení – nezávislé pokusy
    - Výsledky náhodného pokusu nejsou závislé na předcházejících výsledcích.
    - Losování míčků označených čísly z osudí s jejich vracením. Může nastat, že by byl dvakrát vylosován stejný.

4.2 Binomické rozdělení

- Pravděpodobnost úspěšného pokusu je p, který se opakuje n-krát, přičemž náhodný pokus je nezávislý. Potom pravděpodobnost, že náhodný pokus bude právě k-krát úspěšný lze stanovit podle vzorce:

$$P(X = k) = \binom{n}{k} p^k (1 - p)^{n-k}$$

- n – počet náhodných pokusů
- p – pravděpodobnost úspěšného pokusu
- k – počet úspěšných pokusů

4.2 Binomické rozdělení

$$P(X = k) = \overset{(1)}{\binom{n}{k}} \overset{(2)}{p^k} \overset{(3)}{(1 - p)^{n-k}}$$

- (1) – počet kombinací, kolika způsoby můžeme vytvořit z n prvků k-tice.
- (2) – pravděpodobnost úspěchu, který se opakuje k-krát
- (3) – pravděpodobnost neúspěchu, který se opakuje (n-k) krát

- Prvky (2 a 3) tvoří kombinatorické pravidlo součinu

4.2 Binomické rozdělení

- Distribuční funkce

$$P(X \leq k) = \sum_{i=0}^{k} \binom{n}{i} p^i (1-p)^{n-i}$$

- Střední hodnota $E(X) = np$
- Rozptyl $D(X) = np(1-p)$

4.2 Binomické rozdělení

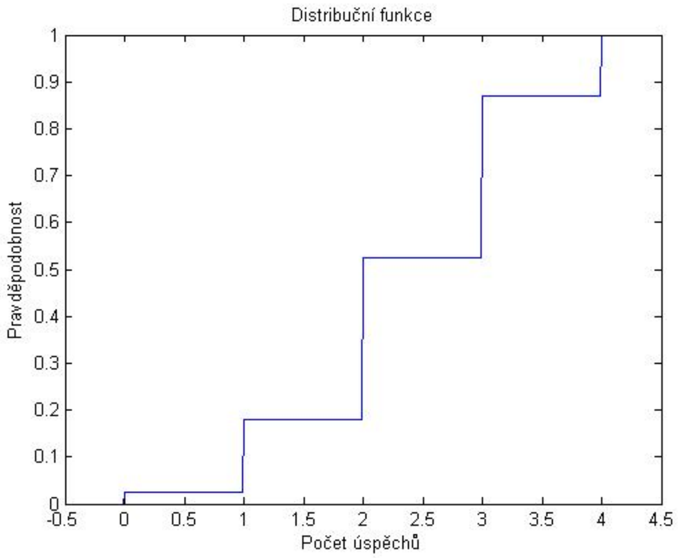
- k počet úspěšných pokusů n počet celkových pokusů
- p pravděpodobnost
- Matlab:
  - Distribuční funkce F=binocdf (k,n,p)
  - Pravděpodobnostní funkce P=binopdf (k,n,p)
  - Inverzní distribuční funkce x=binoinv (pravd,n,p)
  - Stanovení střední hodnoty a rozptylu [MN,var]=binostat(n,p)
  - Náhodná čísla z binomického rozdělení R=binornd(n,p)
  - Odhad parametrů rozdělení p=binofit(k,n)

4.2 Binomické rozdělení

- Př. V osudí je 10 černých míčků a 15 bílých. Z osudí vylosujeme 4 míčky, které vracíme. Určete pravděpodobnost, že 3 míčky budou bílé a jeden černý.
  - $p(bílý) = \frac{15}{25} = 0.6$
  - $P(X = 3) = \binom{4}{3} 0.6^3 (1 - 0.6)^{4-3} = 0.3456$
- Pro porovnání výsledků s hypergeometrickým rozdělením

Binomické	Hypergeometrické
$P(X = 0) = 0.0256$	$P(X = 0) = 0.0166$
$P(X = 1) = 0.1536$	$P(X = 1) = 0.1423$
$P(X = 2) = 0.3456$	$P(X = 2) = 0.3735$
$P(X = 3) = 0.3456$	$P(X = 3) = 0.3597$
$P(X = 4) = 0.1296$	$P(X = 4) = 0.1079$

- Matlab:
  - binopdf([0,1,2,3,4],4,0.6)



4.3 Rozdělení odvozená z binomického rozdělení

- 4.3.1 Alternativní rozdělení
- 4.3.2 Geometrické rozdělení
- 4.3.3 Negativně binomické rozdělení

4.3.1 Alternativní rozdělení

- Popisuje pravděpodobnost jednoho náhodného pokusu
- Binomické rozdělení, kde parametr n=1
  - P(X=1)=p P(X=0)=1-p
  - E(X)=p D(X)=p(1-p)
- Matlab:
  - k počet úspěšných pokusů n počet celkových pokusů
  - p pravděpodobnost

  - Distribuční funkce F=binocdf (k,1,p)
  - Pravděpodobnostní funkce P=binopdf (k,1,p)
  - Inverzní distribuční funkce x=binoinv (pravd,1,p)
  - Stanovení střední hodnoty a rozptylu [MN,var]=binostat(1,p)
  - Náhodná čísla z binomického rozdělení binornd(1,p)
  - Odhad parametrů rozdělení p=binofit(k,1)
- Binomické rozdělení je n-krát opakované alternativní rozdělení.

4.3.2 Geometrické rozdělení

- Popisuje počet neúspěšných pokusů před prvním úspěchem
  - Prvních n pokusů je bez úspěchu, právě (n+1) pokus je úspěšný.
  - Pravděpodobnost je součinem dvou binomických rozdělení (nezávislé pokusy). První určuje pravděpodobnost neúspěchů v n pokusech, druhé pravděpodobnost úspěchu v (n+1) pokusu (alternativní rozdělení)
  - Ad 1) Binomické rozdělení, kde $k=0$
  - Ad 2) Alternativní rozdělení $P(X=1)=p$

4.3.2 Geometrické rozdělení

- Pravděpodobnostní funkce

$$P(X = n) = (1 - p)^n \cdot p$$

- Distribuční funkce

$$P(X \leq n) = p \sum_{i=0}^{n} (1 - p)^i$$

- Střední hodnota $E(X) = \frac{1-p}{p}$
- Rozptyl $D(X) = \frac{(1-p)}{p^2}$

4.3.2 Geometrické rozdělení

- X je počet neúspěšných pokusů p pravděpodobnost
- Matlab
  - Distribuční funkce F=geocdf (X,p)
  - Pravděpodobnostní funkce P=geopdf (X,p)
  - Inverzní distribuční funkce x=geoinv (pravd,p)
  - Stanovení střední hodnoty a rozptylu [M,var]=geostat(p)
  - Náhodná čísla z geometrického rozdělení geornd(p)

4.3.2 Geometrické rozdělení

- Př. Určete pravděpodobnost, že do pátého hodu šestistěnou kostkou padne 6.
  - Pravděpodobnost úspěchu je 1/6
  - Pravděpodobnost, že maximálně 4x padne něco jiného a hned potom padne 6

Pokus	1	2	3	4	5
Neúspěch	0	1	2	3	4

- $P(X \leq 5) = \frac{1}{6}\sum_{i=0}^{4}\left(1-\frac{1}{6}\right)^{i}$=0.5981
- geocdf(4,1/6)

4.3.3 Negativně binomické rozdělení

- Popisuje počet neúspěšných pokusů n, které předchází k-tému výskytu události
  - Zobecnění Geometrického rozdělení.
  - Geometrické rozdělení má prvních (n-1) pokusů neúspěšných.
  - Neg. binomické rozdělení: prvních (n+k-1) pokusů má (k-1) úspěchů, právě (n+k) pokus je úspěšný.
  - Pravděpodobnost je součinem dvou binomických rozdělení (nezávislé pokusy). První určuje pravděpodobnost (k-1) úspěchů v (n+k-1) pokusech, druhé pravděpodobnost úspěchu v (n+k) pokusu (alternativní rozdělení)

4.3.3 Negativně binomické rozdělení

- Pravděpodobnostní funkce

$$P(X = n) = \binom{n + k - 1}{k - 1}(1 - p)^n \cdot p^k$$

- Distribuční funkce

$$P(X \leq n) = \sum_{i=0}^{n} \binom{i + k - 1}{k - 1}(1 - p)^i \cdot p^k$$

- Střední hodnota $\qquad E(X) = \frac{k(1-p)}{p}$
- Rozptyl $\qquad D(X) = \frac{k(1-p)}{p^2}$

4.3.3 Negativně binomické rozdělení

- n – počet neúspěšných pokusů
- k – počet úspěšných pokusů
- p – pravděpodobnost úspěšného pokusu

  – Distribuční funkce F=nbincdf (n,k,p)
  – Pravděpodobnostní funkce P=nbinpdf (n,k,p)
  – Inverzní distribuční funkce x=nbininv (pravd,k,p)
  – Stanovení střední hodnoty a rozptylu
    [M,var]=nbinstat(k,p)
  – Náhodná čísla z negativně binomického rozdělení
    nbinrnd(k,p)

4.3.3 Negativně binomické rozdělení

- Př. Hrajeme speciální „člověče nezlob se". Z domečku můžeme jít teprve, jestliže po třetí hodíme šestku. Určete pravděpodobnost, že se tak stane do 10. hodu.
  - n=0 až 7 k=3

Pokus	3	4	5	6	7	8	9	10	11
Neúspěch	0	1	2	3	4	5	6	7	8

  - $P(X = 10) = \sum_{i=0}^{n} \binom{i+k-1}{k-1} (1-p)^i \cdot p^k =$ $\sum_{i=0}^{7} \binom{i+3-1}{3-1} \left(\frac{5}{6}\right)^i \cdot \left(\frac{1}{6}\right)^3 = 0.2248$
  - nbincdf(7,3,1/6)

- Př. Určete minimální počet hodů, abyste na šestistěnné kostce s pravděpodobností 0.95 vyšli v tomto „člověče nezlob se" z domečku.
  - nbininv(0.95,3,1/6)=33 neúspěšných hodů.

4.4 Multinomické rozdělení

- Multinomické rozdělení se používá v situacích, kdy je třeba vypočítat pravděpodobnost určitého počtu více jevů v n nezávislých pokusech.
  - Multinomické rozdělení je rozšířením binomického o více druhů výsledků.
  - Není již ano/ne, ale více druhů výsledků (například na kostce 1, 2, 3, 4, 5, 6)
  - Předpoklady: pokusy jsou nezávislé, z jevů musí nastat právě jeden.

4.4 Multinomické rozdělení

- Pravděpodobnostní funkce

$$P(X = k) = \frac{n!}{x_1! \cdot x_2! \cdot \dots \cdot x_k!} p_1^{x_1} \cdot p_2^{x_2} \cdot \dots \cdot p_k^{x_k}$$

- mnpdf(četnost, pravděpodobnost)
  - Četnost – vektor s uvedením četností výsledků
  - Pravděpodobnost – vektor pravděpodobností jevů
- mnrnd(počet prvků, pravděpodobnost)
  - Počet prvků – počet vygenerovaných prvků
  - Pravděpodobnost – vektor pravděpodobností jevů

4.4 Multinomické rozdělení

- Házíte 6x šestistěnnou kostkou. Vypočtěte pravděpodobnost, že hodíte právě 1x každé číslo.
  - $P(X=k)=\frac{n!}{x_1!\cdot x_2!\cdot\cdots\cdot x_k!}p_1^{x_1}\cdot p_2^{x_2}\cdot\cdots\cdot p_k^{x_k}$
  - $P(X=k)=\frac{6!}{1!\cdot 1!\cdot 1!\cdot 1!\cdot 1!\cdot 1!}\left(\frac{1}{6}\cdot\frac{1}{6}\cdot\frac{1}{6}\cdot\frac{1}{6}\cdot\frac{1}{6}\cdot\frac{1}{6}\right)$=0.0154
  - mnpdf([1,1,1,1,1,1],[1/6,1/6,1/6,1/6,1/6,1/6])

- Házíte 21x šestistěnnou kostkou. Vypočtěte pravděpodobnost, že hodíte právě 1x jedničku, 2x 2, 3x 3, 4x 4, 5x 5 a 6x 6.
  - $P(X=k)=\frac{21!}{1!\cdot 2!\cdot 3!\cdot 4!\cdot 5!\cdot 6!}\left(\frac{1}{6}\cdot\frac{1}{6^2}\cdot\frac{1}{6^3}\cdot\frac{1}{6^4}\cdot\frac{1}{6^5}\cdot\frac{1}{6^6}\right)=9.3597\cdot 10^{-5}$
  - mnpdf([1,2,3,4,5,6],[1/6,1/6,1/6,1/6,1/6,1/6])

4.4 Multi-hypergeometrické rozdělení

- Multinomické rozdělení
  - Nezávislé pokusy
  - Vychází z binomického rozdělení
- Multi-hypergeometrické rozdělení
  - Závislé pokusy
  - Vychází z hypergeometrického rozdělení
- Pravděpodobnostní funkce
  - $P(X = k_1, k_2, \dots, k_l) = \frac{\prod_{i=1}^{l} \binom{M_i}{k_i}}{\binom{N}{n}}$
  - Platí: $\sum_{i=1}^{l} M_i = N, \sum_{i=1}^{l} k_i = n$

4.4 Multi-hypergeometrické rozdělení

- Př.: Máte balíček 32 karet, který obsahuje 4 karty 7, 4 karty 8, 9, 10, spodky, filky, krále a 4 esa. Z balíčku vyberete 8 karet. Jaká je pravděpodobnost, že vylosujete právě 2 spodky, 2 filky, 2 krále a 2 esa.
- Karty nevracíte.
- $P(X = 0,0,0,0,2,2,2,2) = \frac{\prod_{i=1}^{l} \binom{M_i}{k_i}}{\binom{N}{n}} = \frac{\binom{4}{0} \cdot \binom{4}{0} \cdot \binom{4}{0} \cdot \binom{4}{0} \cdot \binom{4}{2} \cdot \binom{4}{2} \cdot \binom{4}{2} \cdot \binom{4}{2}}{\binom{32}{8}} = 1.232 \cdot 10^{-4}$

4.5 Poissonovo rozdělení

- Poissonův proces popisuje počet náhodných událostí v nějakém pevném „časovém“ intervalu.
  - Termín „časovém“ lze podle typu úlohy nahradit za ks, vzdálenost apod.
- Předpoklady:
  - Pravděpodobnost, že nastane více jevů v limitně krátkém čase je nulová.
  - Pravděpodobnost výskytu jevu závisí pouze na délce intervalu, nikoliv na okamžiku jeho začátku.
  - Počty událostí ve vzájemně disjunktních intervalech jsou nezávislé.
  - Například u životnosti výrobků:
    - ad 1) nevzniknou dvě poruchy naráz (například vyšší hodnotou napětí v síti),
    - ad 2) průměrný počet poruch auta Škoda 120 nebude odlišný při najetí mezi <10000,20000> km a <350000,360000> km.
    - ad 3) nevznikají poruchy se společnou příčinou, kdy vznik jedné poruchy má za následek vznik dalších.

4.5 Poissonovo rozdělení

- Parametrem Poissonova procesu je intenzita λ náhodného jevu.
  - Počet jevů za jednotku času (vzdálenosti, počet poruch na 1000 výrobků apod.)
  - Základní jednotky intenzity hod^{-1}, km^{-1}, ks^{-1}
  - Součin $\lambda \cdot t$ je bezrozměrná veličina, která je základním parametrem Poissonova procesu.

4.5 Poissonovo rozdělení

- Jestliže se definuje náhodný pokus jako Poissonův proces s intenzitou λ, potom náhodnou veličinu X lze definovat jako počet výskytů události v časovém intervalu délky t. Náhodná veličina X se popisuje Poissonovým rozdělením s parametrem λt.
  - Lze stanovit například pravděpodobnost určitého počtu poruch v 1000 výrobcích.
    - $\lambda = 0.03\,\frac{poruchy}{výrobek}, t = 1000\ výrobků, \lambda t = 30$
  - Počet dopravních poruch na 8 km úseku silnice za 30 dní.
    - $\lambda = 10^{-7}\,\frac{nehody}{km\ den}, t = 8 \cdot 30 = 240\ km\ den, \lambda t = 2.4 \cdot 10^{-5} nehody$
  - Počet bakterií v 16 ml vody.
    - $\lambda = 400\,\frac{bakterii}{1\ ml\ vody}, t = 16\ ml, \lambda t = 6400$
- Př. Pravděpodobnost vadného výrobku je 1 %. Potom intenzita $\lambda = 0.01$.
  - Mějme v krabici t=50 výrobků. Potom Poissonovo rozdělení má parametr λt=0.01$\cdot$ 50=0.5.
  - Mějme v krabici t=200 výrobků. Potom Poissonovo rozdělení má parametr λt=0.01$\cdot$ 200=2.
- Chceme znát pravděpodobnost, že v krabici jsou právě 0, 1, 2, 3 porouchané výrobky.

4.5 Poissonovo rozdělení

- Pravděpodobnostní funkce

$$P(X = k) = \frac{(\lambda t)^k e^{-\lambda t}}{k!}$$

- Distribuční funkce

$$P(X \leq k) = \sum_{i=0}^{k} \frac{(\lambda t)^i e^{-\lambda t}}{i!}$$

- Střední hodnota $E(X) = \lambda t$
- Rozptyl $\mathrm{D}(X) = \lambda t$
- λt je jediný parametr Poissonova rozdělení

4.5 Poissonovo rozdělení

- X je počet pokusů, λ intenzita náhodného jevu
- Matlab:
  - Parametr lambda v matlabu představuje součin $\lambda \cdot t$
  - Distribuční funkce F=poisscdf (X,lambda)
  - Pravděpodobnostní funkce P= poisspdf (X,lambda)
  - Inverzní distribuční funkce x= poissinv (pravd,lambda)
  - Stanovení střední hodnoty a rozptylu
    [M,var]= poisstat(lambda)
  - Náhodná čísla z hypergeometrického rozdělení
    poissrnd(lambda)
  - Stanovení parametru lambda z dat
    poissfit(data)

4.5 Poissonovo rozdělení

- Př. Z dat o poruchách stroje se zjistilo, že za jednu osmihodinovou směnu jsou v průměru 2.4 poruchy. Stanovte pravděpodobnost, že za 1.5 hodiny budou na stroji zjištěny 2 poruchy.
  – Průměrný počet poruch za 8 hodin 2.4
  – Průměrný počet poruch za 1 hodinu 0.3
  – Parametr λt 0.3*1.5 0.45
  – Pravděpodobnost zjistím z pravděpodobnostní funkce

$$P(X=2) = \frac{(0.3 \cdot 1.5)^2 e^{-0.3 \cdot 1.5}}{2!} = 0.0646$$

  – poisspdf(2,0.45)

4.6 Aproximace diskrétních rozdělení

- Aproximace hypergeometrického rozdělení binomickým
  - Hypergeometrické – závislé pokusy, odlišné podmínky pokusu
    - x – počet vybraných prvků s danou vlastností
    - K – celkový počet prvků s danou vlastností
    - M – celkový počet prvků
    - N – počet vybraných prvků

$$P(X = x) = \frac{\binom{K}{x}\binom{M-K}{N-x}}{\binom{M}{N}}$$

  - Binomické – nezávislé pokusy, shodné podmínky pokusu
    - n – počet náhodných pokusů
    - p – pravděpodobnost úspěšného pokusu
    - k – počet úspěšných pokusů

$$P(X = k) = \binom{n}{k} p^k (1-p)^{n-k}$$

- Je-li $\frac{N}{M}$ menší než 0.05, lze hypergeometrické rozdělení nahradit binomickým s parametry: $n_{bin} = N_{hyp}, p_{bin} = \frac{K_{hyp}}{M_{hyp}}$

4.6 Aproximace diskrétních rozdělení

- V osudí je 20 bílých a 30 černých míčků. Z osudí se losuje 5 míčků. Jaká je pravděpodobnost, že se vylosují právě 2 černé.
- Hypergeometrické: M=50, K=30, N=5, x=2
  - $P(X=2)=\frac{\binom{30}{2}\binom{20}{3}}{\binom{50}{5}}$
  - hygepdf (x,M,K,N)=hygepdf(2,50,30,5)=0.2341
- Binomické: n=5, p=0.6, k=2
  - $P(X=2)=\binom{5}{2}0.6^2(1-0.6)^3$
  - binopdf (k,n,p)=binopdf(2,5,0.6)=0.2304
- Nejsou zcela splněny předpoklady přechodu z hypergeometrického rozdělení na binomické, ale výsledky pravděpodobností jsou podobné.
- Jestliže se úloha upraví a bílých míčků bude 200, černých 300 a budu losovat 5 míčků bez vracení, tak podmínky tahu budou skoro stejné, jako bychom míčky vraceli.
  - Hypergeometrické: M =500, K=300, N=5, x=2 hygepdf(2,500,300,5)=0.2308
  - Binomické: binopdf(2,5,0.6)=0.2304

4.6 Aproximace diskrétních rozdělení

- Aproximace binomického rozdělení Poissonovým
  - Binomické – nezávislé pokusy, shodné podmínky pokusu
    - n – počet náhodných pokusů
    - p – pravděpodobnost úspěšného pokusu
    - k – počet úspěšných pokusů

$$P(X = k) = \binom{n}{k} p^k (1-p)^{n-k}$$

  - Poissonovo – poissonův proces,
    - λt – intenzita náhodného jevu

$$P(X = k) = \frac{(\lambda t)^k e^{-\lambda t}}{k!}$$

Poissonovým rozdělením lze aproximovat binomické, jestliže n je větší než 30 a p<0.05

$$\lambda t_{Poiss} = n_{bin} \cdot p_{bin}$$

4.6 Aproximace diskrétních rozdělení

- Př. Při kontrole výrobků je v průměru 1 % výrobků chybných. Vypočtěte pravděpodobnost, že 1000 výrobků bude mít maximálně (včetně) 8 chybných výrobků.
- Binomické rozdělení
  - n=1000 p=0.01 k=8
  - $P(X \leq k) = \sum_{i=0}^{8} \binom{1000}{i} 0.01^i (1-0.01)^{n-i}$
  - P=binocdf(8.5,1000,0.01)=0.3317
    - První parametr je 8.5 – máme jistotu, že je započítáno i 8 poruch výrobku.
- Poissonovo rozdělení
  - λt=1000· $0.01 = 10$ k=8
  - $P(X \leq k) = \sum_{i=0}^{8} \frac{10^i e^{-10}}{i!}$
  - P=poisscdf(8.5,10)=0.3328

4.7 Přehled diskrétních rozdělení

Odvozená od binomického

Alternativní rozdělení

Geometrické rozdělení

Neg. binomické rozdělení

Multinomické rozdělení

4.8 Základní příkazy v matlabu/octave

	Hypergeometrické	Binomické	Poissonovo
Distribuční funkce	hygecdf (x,M,K,N)	binocdf (k,n,p)	poisscdf (X,lambda)
Pravděpodobnostní funkce	hygepdf (x,M,K,N)	binopdf (k,n,p)	poisspdf (X,lambda)
Inverzní distribuční funkce	hygeinv (pravd,M,K,N)	binoinv (pravd,n,p)	poissinv(pravd,lambda)
Stanovení střední hodnoty a rozptylu	hygestat(M,K,N)	binostat(n,p)	poisstat(lambda)
Náhodná čísla	hygernd(N,K,N)	binornd(n,p)	poissrnd(lambda)
Odhad parametrů		binofit(k,n)	poissfit(data)

5. Spojitá rozdělení pravděpodobnosti

- 5.1 Rovnoměrné rozdělení
- 5.2 Exponenciální rozdělení
- 5.3 Weibullovo rozdělení
- 5.4 Erlangovo (Gamma) rozdělení
- 5.5 Normální rozdělení
- 5.6 Normované normální rozdělení
- 5.7 Logaritmicko-normální rozdělení
- 5.8 Grafické ověření, že data pochází z určitého spojitého rozdělení

5.1 Rovnoměrné rozdělení

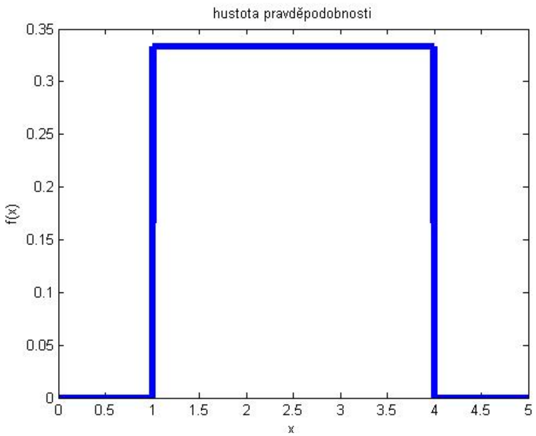
- Rovnoměrné rozdělení má konstantní hustotu pravděpodobnosti na intervalu $< a, b >$
  - Speciálním případem je funkce náhodné číslo, která je definována na intervalu $< 0{,}1 >$
  - Hustota pravděpodobnosti: $f(x) = \begin{cases} \frac{1}{b-a} & < a, b > \\ 0 & jinde \end{cases}$
  - Distribuční funkce

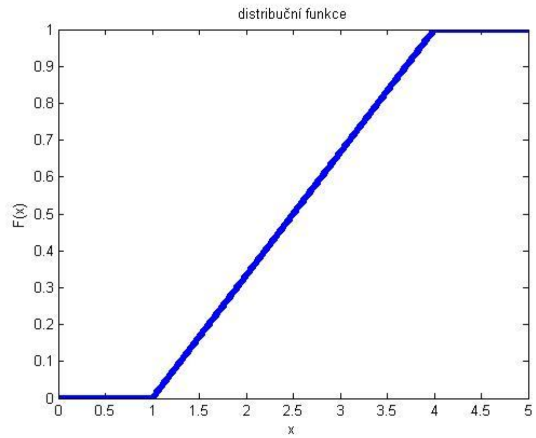
    $F(x) = \frac{x-a}{b-a} \qquad a < x < b$

    $F(x) = 0 \qquad x < a;$

    $F(x) = 1 \qquad x > b$
  - Střední hodnota $\mathrm{E}(X) = \frac{a+b}{2}$
  - Rozptyl $\mathrm{D}(X) = \frac{(a-b)^2}{12}$
  - Šikmost $\mathrm{a} = 0$ (je symetrické)
  - Špičatost $b = \frac{9}{5}$

- Rovnoměrné rozdělení: a=1 b=4





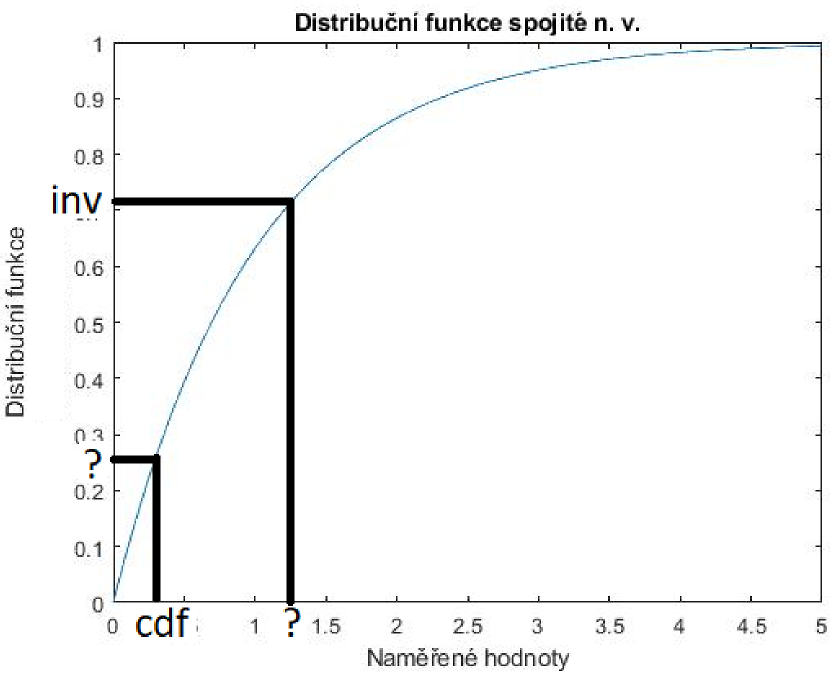
5.1 Rovnoměrné rozdělení

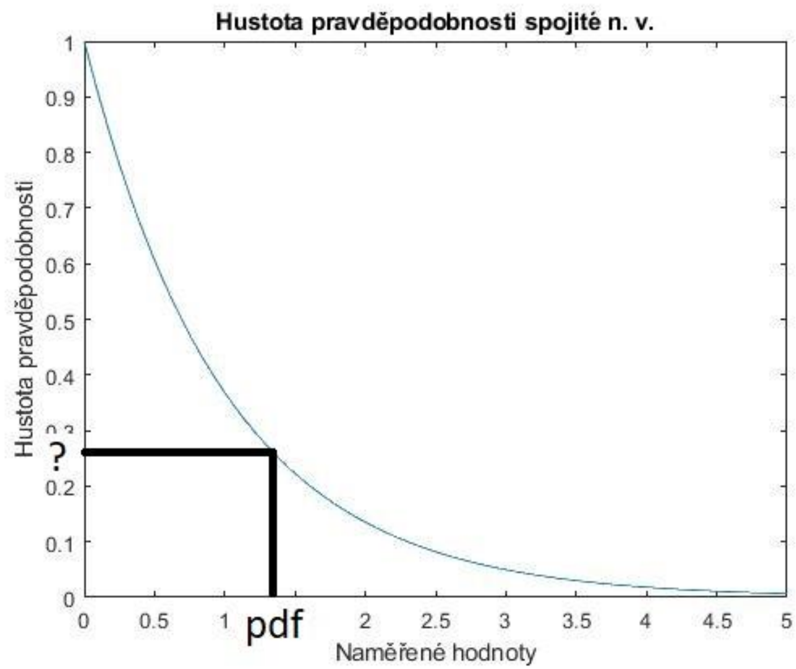
- x parametr a,b minimum a maximum rovnoměrného rozdělení

- Matlab:
  - Distribuční funkce F=unifcdf(x,a,b)
  - Hustota pravděpodobnosti f=unifpdf(x,a,b)
  - Inverze distribuční funkce x=unifinv(pravd,a,b)
  - Odhad parametrů [a,b]=unifit(data)
  - Stanovení střední hodnoty a rozptylů [m,v]=unifstat(a,b)
  - Náhodné číslo v matici o velikosti m x n R=unifrnd(a,b,m,n)

5.1 Rovnoměrné rozdělení

- Souvislost příkazů v matlabu jednotlivých rozdělení
  - Předpona – typ rozdělení
  - Přípona – co potřebujeme vypočítat
  - Např. unifinv unif – rovnoměrné rozdělení,
    inv – znám hodnotu kvantilu, neznám x.





5.1 Rovnoměrné rozdělení

- Př. Vygenerujte 100 náhodných čísel z intervalu <0,1>, vypočtěte jejich střední hodnotu, rozptyl, šikmost a špičatost a porovnejte s teoretickými hodnotami.
  - `>> x=unifrnd(0,1,1,100);`
  - `>> EX=mean(x)` EX = 0.4993
  - `>> DX=var(x)` DX = 0.0765
  - `>> a=skewness(x)` a = -0.1577
  - `>> b=kurtosis(x)` b = 1.8648
  - Teoretické hodnoty jsou: EX=0.5; DX=0.0833; a=0; b=1.8 (někde se uvádí i -1.2 po odečtení 3)

  - Při opakování výpočtů se výsledky budou lišit, protože nejsou počítány pomocí integrálů, ale z naměřených dat. Jsou tedy závislé na vygenerovaných náhodných číslech.
  - Čím více se vygeneruje náhodných čísel, tím více se střední hodnota, rozptyl, šikmost a špičatost bude blížit teoretickým hodnotám.

5.2 Exponenciální rozdělení

- Exponenciální rozdělení se používá pro popis doby do první události Poissonova procesu (viz Poissonovo rozdělení) s intenzitou náhodného jevu λ, nebo s její převrácenou hodnotou (střední hodnotou) μ.
- Souvislost s Poissonovým rozdělením
  - Poissonovo rozdělení – pravděpodobnost počtu událostí za dobu t.
    - Pravděpodobnost výskytu jevu závisí pouze na délce intervalu, nikoliv na okamžiku jeho začátku.
    - Počty událostí ve vzájemně disjunktních intervalech jsou nezávislé.
  - Exponenciální rozdělení – pravděpodobnost, že první událost nastane do doby t.
    - Například pro popis doby do poruchy nedegradujícího výrobku.

5.2 Exponenciální rozdělení

- Aplikovatelnost exponenciálního rozdělení:
  - Neměnná intenzita náhodného jevu na čase
  - Doba do poruchy nedegradujících výrobků
    - Střední doba do poruchy komponenty nezáleží, jak dlouho je v provozu. Naopak degradující výrobek – střední počet poruch za 1 rok se z délkou provozu zvyšuje.
  - Teorie front (doba čekání ve frontě).
- Obecné použití exponenciální funkce
  - Rozdělení bez paměti
  - Rozpad radioaktivních látek
  - Nabíjení vybíjení kondenzátoru
  - Řešení obyčejných diferenciálních rovnic 1. řádu

5.2 Exponenciální rozdělení

- Hustota pravděpodobnosti

$$\mathrm{f(t)} = \frac{e^{-\frac{t}{\mu}}}{\mu}, t \geq 0 \qquad f(t) = 0, t < 0$$

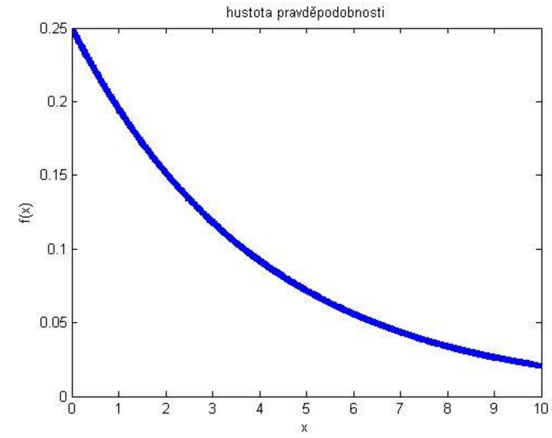
- Distribuční funkce

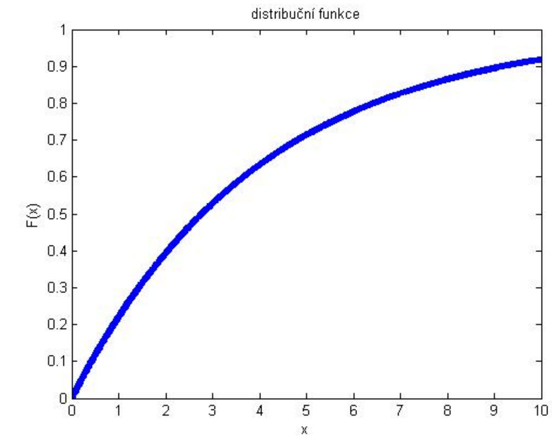
$$\mathrm{F(t)} = 1 - e^{-\frac{t}{\mu}}, t \geq 0 \qquad F(t) = 0, t < 0$$

- Střední hodnota $E(X) = \mu$
- Rozptyl $D(X) = \mu^2$
- Šikmost a = 2
- Špičatost $b = 6$

- Exponenciální rozdělení se často využívá ve spolehlivosti pro modelování životnosti nedegradujících výrobků.
  - Zde se využívá intenzita náhodného jevu $\lambda = \frac{1}{\mu}$
  - Intenzita náhodného jevu $\lambda(t) = \frac{f(t)}{1-F(t)} = \lambda$ není závislá na čase
  - Pokud je intenzita náhodného jevu konstantní, potom je rozdělení tzv. bez paměti a lze ho popsat pro životnost nedegradujících výrobků
  - Ve spolehlivosti $\lambda\Delta t$ označuje pravděpodobnost, že se výrobek porouchá v krátkém časovém intervalu Δt , za předpokladu, že na začátku intervalu byl v provozuschopném stavu.
  - Distribuční funkce představuje pravděpodobnost, že se výrobek porouchá do času t

$\mu = 4$





5.2 Exponenciální rozdělení

- μ je parametr exponenciálního rozdělení
- V teorii spolehlivosti se používá místo střední hodnoty μ její převrácená hodnota představující intenzitu poruch λ.
- Počítání s intenzitami má ve spolehlivosti výhodu při řazení komponent do větších funkčních celků (seriové, paralelní, m z n zálohování)

Distribuční funkce	F=expcdf(x,μ)
Hustota pravděpodobnosti	f=exppdf(x, μ)
Inverze distribuční funkce	x=expinv(pravd, μ)
Odhad parametrů	expfit(data)
Stanovení střední hodnoty a rozptylu	[m,v]=expstat(μ)
Náhodná čísla v matici m x n	R = exprnd(μ, m, n)

5.2 Exponenciální rozdělení

- Doba do poruchy výrobku je popsána exponenciálním rozdělením. Byly zjištěny následující doby do poruchy: t=[123,167,195,213,258,324,387,423,541,630] dní.
- Určete střední dobu do poruchy výrobku. Kolik procent výrobků se porouchá mezi 200 a 400 dny provozu.

- Řešení:
  - >> t=[123,167,195,213,258,324,387,423,541,630]
  - >> μ = expfit(t) μ = 326.1
  - >> pravd=expcdf(400, μ)-expcdf(200, μ)
  - pravd = 0.2483

- Střední doba do poruchy výrobku je 326.1 dne. Pravděpodobnost, že se výrobek porouchá mezi 200 a 400 dny provozu je 0.2483.

5.2 Exponenciální rozdělení

- Př. Doba do poruchy výrobku je popsána exponenciálním rozdělením. Bylo zjištěno, že do 24 měsíců se porouchá 10 % výrobků. Určete parametry exponenciálního rozdělení.
- Nelze jednoduše použít matlabovské funkce, protože neznáme přesný tvar distribuční funkce. Lze řešit buď iteračně (například půlením intervalu), nebo dosazením do distribuční funkce.
-
- Řešení – neznáme parametr lambda

$$F(24) = 0.1$$

$$1 - e^{-\frac{24}{\mu}} = 0.1$$

$$e^{-\frac{24}{\mu}} = 0.9$$

$$-\frac{24}{\mu} = \ln 0.9$$

$$\mu = -\frac{24}{\ln 0.9} = 228$$

- Střední doba do poruchy je $\mu = 228$ měsíců.

5.2 Exponenciální rozdělení

- Zjištění parametrů exponenciálního rozdělení pro data o poruchách výrobků
  - Zkouška je ukončena buď poruchou, nebo časem
- $E(X) = \mu = \frac{\sum_{i=1}^{n} t_i}{r}$
  - t_i je doba do poruchy, nebo doba do ukončení zkoušky
  - n počet výrobků
  - r počet poruch

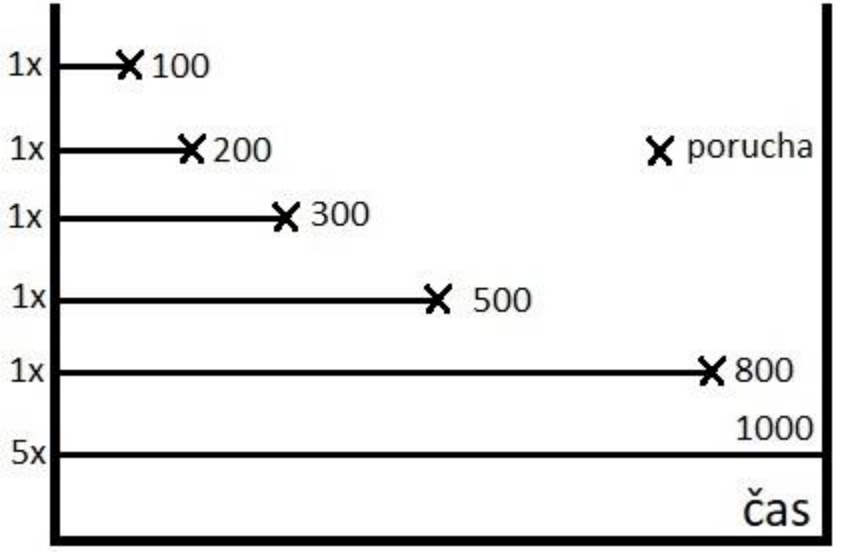
5.2 Exponenciální rozdělení

- Zjištění parametrů exponenciálního rozdělení
- EX=expfit(x,alpha,cens,freq)
  - EX vypočtená střední hodnota
  - x doba, kdy došlo k poruše, nebo ukončení zkoušky
  - alpha vysvětleno v kapitole 8, zadávejte 0.05 (hladina významnosti)
  - Cens způsob ukončení zkoušky, 0 porucha, 1 časem
  - Freq počet výskytů
- Vektory x, cens a freq musejí být stejně dlouhé.

5.2 Exponenciální rozdělení

- Máte 10 výrobků a chcete zjistit střední dobu do poruchy. Zkouška probíhá 1000 hodin. Za 1000 hodin se porouchalo 5 výrobků v časech 100, 200, 300, 500, 800 hodin. Po poruše nebyly nahrazeny. Zjistěte parametry exponenciálního rozdělení.
- $E(X) = \frac{\sum_{i=1}^{n} t_i}{r} = \frac{100+200+300+500+800+5\cdot 1000}{5} = \frac{6900}{5} = 1380\ h$
- $\lambda = \frac{1}{E(X)} = \frac{1}{1380} = 7.2 \cdot 10^{-4} h^{-1}$
- Matlab:
  - x=[100,200,300,500,800,1000];
  - cens=[0,0,0,0,0,1]
  - freq=[1,1,1,1,1,5];
  - EX=expfit(x,0.05,cens,freq)

  - EX = 1380 hodin
- Střední doba do poruchy je 1380 hodin



5.3 Weibullovo rozdělení

- Weibullovo rozdělení slouží (obdobně jako exponenciální) k modelování doby do poruchy zařízení. Na rozdíl od exponenciálního je však obecnější, protože popisuje i degradující komponenty a tím nevyžaduje konstantní intenzitu náhodného jevu.
  - Rozdělení obsahuje dva parametry
    - a – parametr měřítka b – parametr tvaru
- Hustota pravděpodobnosti

$$f(t) = \frac{bt^{b-1}}{a^b} e^{-\left(\frac{t}{a}\right)^b}, a > 0, b > 0, t \geq 0 \qquad f(t) = 0, t < 0$$

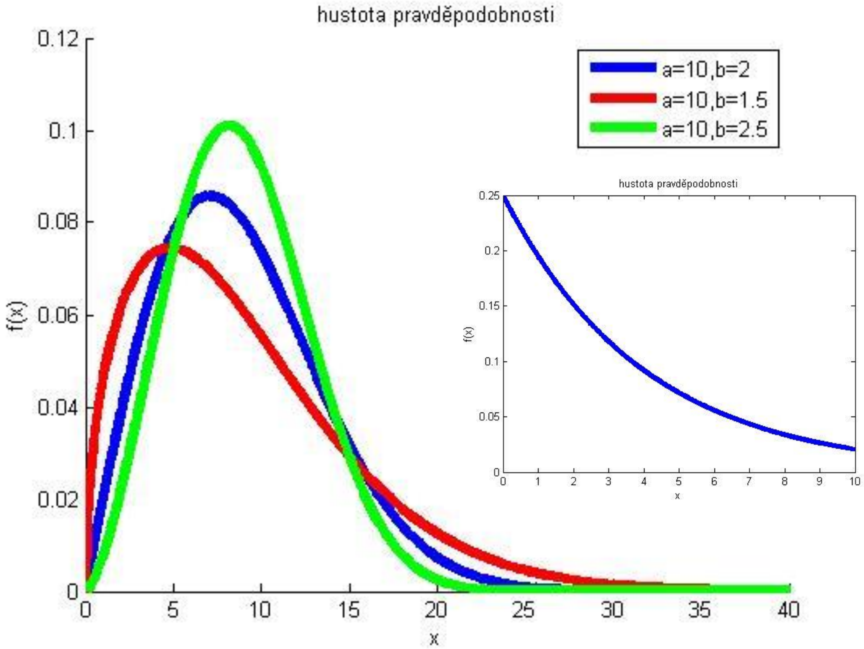
- Distribuční funkce

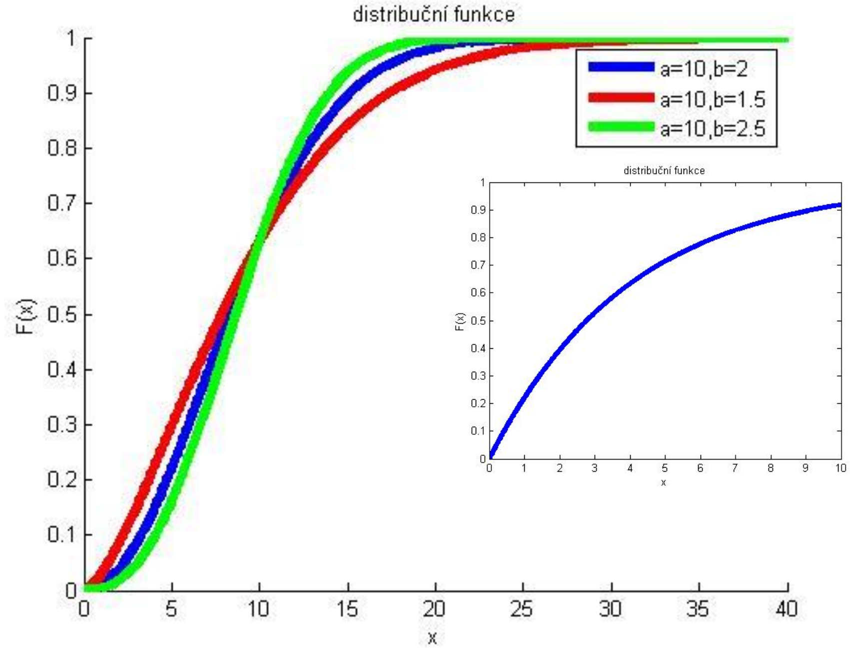
$$F(t) = 1 - e^{-\left(\frac{t}{a}\right)^b}, a > 0, b > 0, t \geq 0 \qquad F(t) = 0, t < 0$$

- $b = 1$, přechází Weibullovo rozdělení na exponenciální.
- $b < 1$, pro popis doby do poruchy výrobků, kde se projevují časné poruchy
- $b > 1$, pro popis doby do poruchy výrobků, kde se projevují poruchy z opotřebení.

5.3 Weibullovo rozdělení

- Čím větší je parametr b, tím pozvolněji dochází k nástupu poruch, a naopak rychleji dochází k poškozování výrobků v období degradace (porovnej zelenou a červenou čáru)
- Vložený malý graf je hustota a distribuční funkce exponenciálního rozdělení (Weibullovo s b=1)





5.3 Weibullovo rozdělení

- Střední hodnota $E(X) = a\left[\Gamma\left(1+\frac{1}{b}\right)\right]$
- Rozptyl $D(X) = a^2\left[\Gamma\left(1+\frac{2}{b}\right) - \left(\Gamma\left(1+\frac{1}{b}\right)\right)^2\right]$
- Γ představuje gamma funkci, jedná se o rozšíření funkce faktoriál do oboru reálných čísel.
- Pro stanovení střední hodnoty a rozptylu využíváme SW prostředků – v matlabu funkce wblstat.

5.3 Weibullovo rozdělení

- t parametr   a – parametr měřítka   b – parametr tvaru

- Matlab
  - Distribuční funkce F=wblcdf(t,a,b)
  - Hustota pravděpodobnosti f=wblpdf(t,a,b)
  - Inverze distribuční funkce x=wblinv(pravd,a,b)
  - Odhad parametrů a=wblfit(data)
  - Stanovení střední hodnoty a rozptylu [m,v]=wblstat(a,b)
  - Náhodné číslo wblrnd(a,b)

5.3 Weibullovo rozdělení

- Doba do poruchy výrobku je popsána Weibullovým rozdělením. Byly zjištěny následující doby do poruchy: t=[123,167,195,213,258,324,387,423,541,630] dní.
- Zjistěte optimální parametry Weibullova rozdělení, vypočtěte střední hodnotu a porovnejte výsledky s příkladem v kapitole 5.2.

- Výsledek:
  - >> t=[123,167,195,213,258,324,387,423,541,630]
  - >> a=wblfit(t)
  - a = 370.1311   2.2136   (a=370; b=2.21)
  - >> [m,v]=wblstat(a(1),a(2))
  - m = 327.8   střední hodnota
  - v =   2.4469e+04   rozptyl

- Parametr b je dosti odlišný od 1, u výrobků dochází k degradaci a pro popis doby do poruchy je vhodnější popis Weibullovým rozdělením. Střední doba do poruchy je 328 dní.
- Jedna z úloh statistiky je stanovit, zda parametr b je natolik odlišný od 1, že je vhodné pro popis doby do poruchy použít Weibullovo rozdělení. Řešeno v kapitole 8.

5.3 Weibullovo rozdělení

- Zjištění parametrů Weibullova rozdělení pro data o poruchách výrobků, kde zkouška je ukončena buď poruchou, nebo časem
  - t_i je doba do poruchy, nebo doba do ukončení zkoušky
  - n počet výrobků
  - r počet poruch
- par=wblfit(x,alpha,cens,freq)
  - par — vypočtené parametry Weibullova rozdělení
  - x — doba, kdy došlo k poruše, nebo ukončení zkoušky
  - alpha — vysvětleno v kapitole 8, zadávejte 0.05 (hladina významnosti)
  - Cens — způsob ukončení zkoušky, 0 porucha, 1 časem
  - Freq — počet výskytů
- Vektory x, cens a freq musejí být stejně dlouhé

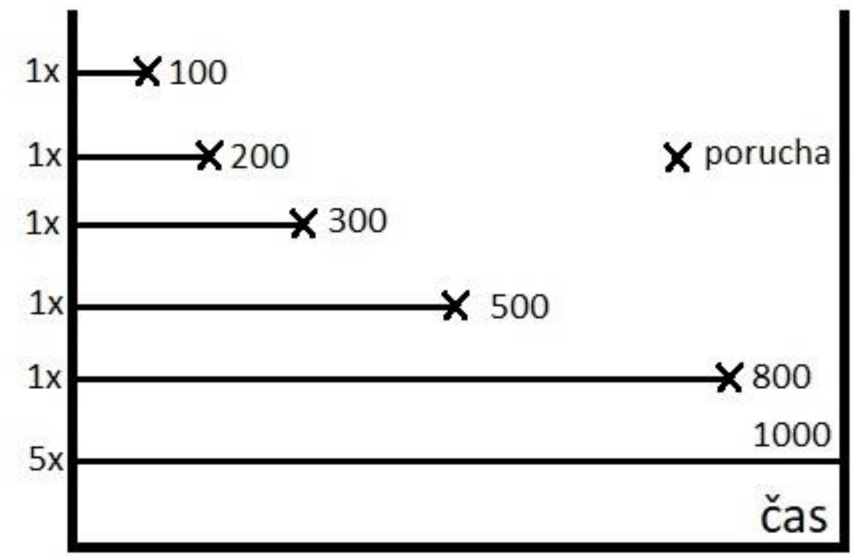
5.3 Weibullovo rozdělení

- Máte 10 výrobků a chcete zjistit parametry Weibullova rozdělení. Zkouška probíhá 1000 hodin. Za 1000 hodin se porouchalo 5 výrobků v časech 100, 200, 300, 500, 800 hodin. Po poruše nebyly nahrazeny. Zjistěte jeho parametry.

- Matlab:
  - x=[100,200,300,500,800,1000];
  - cens=[0,0,0,0,0,1]
  - freq=[1,1,1,1,1,5];
  - a=wblfit(x,0.05,cens,freq)

  - a(1) = 1378.8
  - a(2) = 1.0017
  - Parametry W. r. jsou:

    a=1378, b=1.0017



5.4 Erlangovo (Gamma) rozdělení

- Popisuje v Poissonově procesu dobu do k-té poruchy.
  - Je odvozeno z exponenciálního rozdělení
  - Obsahuje dva parametry:
    - 1/lambda – střední doba do poruchy
    - k – k-tý počet poruch
  - Hustotu pravděpodobnosti lze získat konvolucí exponenciálních rozdělení (složité)

- Hustota pravděpodobnosti

$$f(t) = \lambda e^{-\lambda t} \frac{(\lambda t)^{k-1}}{(k-1)!}, t \geq 0 \qquad f(t) = 0, t < 0$$

- Distribuční funkce

$$F(t) = 1 - e^{-\lambda t} \sum_{i=0}^{k-1} \frac{(\lambda t)^i}{i!}, t \geq 0 \qquad F(t) = 0, t < 0$$

- Střední hodnota $E(X) = \frac{k}{\lambda}$
- Rozptyl $D(X) = \frac{k}{\lambda^2}$

5.4 Erlangovo (Gamma) rozdělení

- t parametr k – kolikatá událost
- b – parametr měřítka $b = \frac{1}{\lambda}$

- Výpočet v Matlabu
  - Distribuční funkce F=gamcdf(t,k,b)
  - Hustota pravděpodobnosti f=gampdf(t, k,b)
  - Inverze distribuční funkce x=gaminv(pravd, k,b)
  - Odhad parametrů [k,b]=gamfit(data)
  - Stanovení střední hodnoty a rozptylů [m,v]=gamstat(k,b)
  - Náhodné číslo gamrnd(k,b)

5.4 Erlangovo (Gamma) rozdělení

- Př. Střední doba do poruchy nedegradujícího zařízení je 1000 hodin. Určete pravděpodobnost, že doba do 10. poruchy bude kratší než 8000 hodin.
- Pravděpodobnost se vypočte: F=gamcdf(t,k,b)
  - P=gamcdf(8000,10,1000)
  - P=0.2834

S pravděpodobností 28.3 % bude doba do 10. poruchy kratší než 8000 h.

5.5 Normální rozdělení

- Často je označováno jako Gaussovo rozdělení
- Jedná se o nejpoužívanější pravděpodobnostní rozdělení.
- Použití pro popis náhodných veličin, které lze interpretovat jako sumární výsledek mnoha nepatrných a vzájemně nezávislých vlivů.
- Za určitých podmínek lze pomocí normálního rozdělení aproximovat řadu jiných spojitých i nespojitých rozdělení.

5.5 Normální rozdělení

- Normální rozdělení má dva parametry
  - μ – střední hodnota
  - σ^2 – rozptyl
  - !!! Pozor v SW matlab se zadává odmocnina z rozptylu – směrodatná odchylka.
- Jestliže data jsou z normálního rozdělení, tak zapisujeme $N(\mu, \sigma^2)$
- Normální rozdělení je natolik významné, že jeho:
  - střední hodnotu označujeme μ
  - rozptyl σ^2

5.5 Normální rozdělení

- Hustota pravděpodobnosti

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

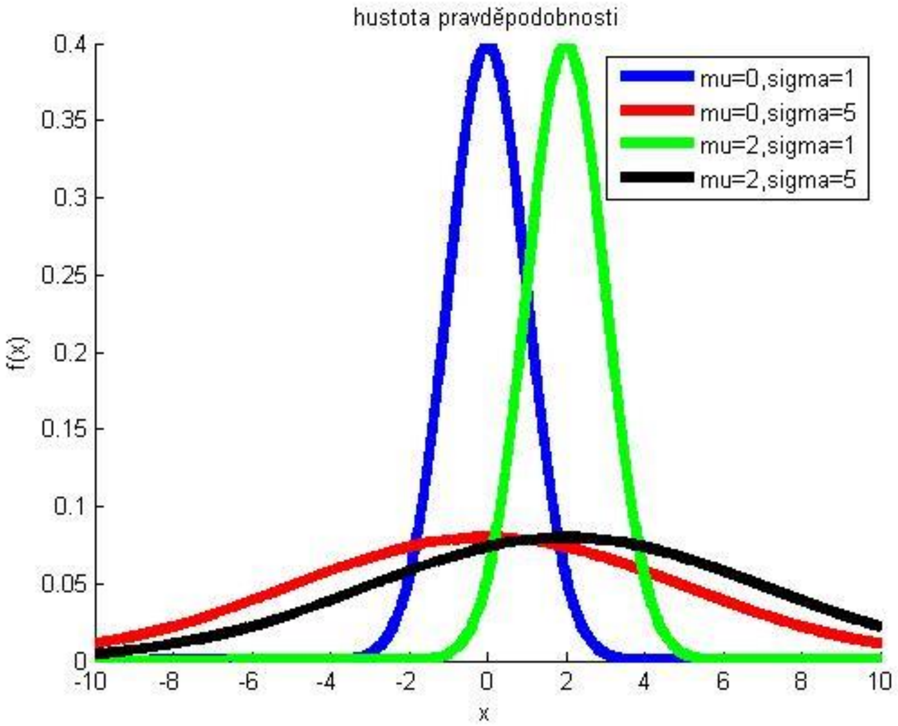
- Distribuční funkce

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

  - Hustotu pravděpodobnosti nelze analyticky integrovat, proto se využívá softwaru, který má tuto funkci implementovánu.
  - Historicky, v dobách kdy nebyla funkce implementována v počítači, se používaly tabulky a transformace normálního rozdělení na normální rozdělení se střední hodnotou 0 a rozptylem 1. (tzv. normované normální rozdělení)
- Střední hodnota $\mathrm{E}(X) = \mu$
- Rozptyl $D(X) = \sigma^2$
- Šikmost a=0 rozdělení je symetrické
- Špičatost b=3
  - Špičatost může nabývat pouze nezáporných hodnot, proto se často od výsledku odečítá 3, aby špičatost normálního rozdělení byla nulová.

5.5 Normální rozdělení

- Změna střední hodnoty posune hustotu rozdělení, bez změny jejího tvaru
- Změna směrodatné odchylky zvětší/zmenší hustotu rozdělení, bez změny střední hodnoty



5.5 Normální rozdělení

- x parametr
- μ – střední hodnota
- σ – směrodatná odchylka, pozor nezadává se rozptyl

- Matlab
  - Distribuční funkce F=normcdf(x, μ, σ)
  - Hustota pravděpodobnosti f=normpdf(x, μ, σ)
  - Inverze distribuční funkce x=norminv(pravd, μ, σ)
  - Odhad parametrů [μ, σ]=normfit(data)
  - Stanovení střední hodnoty a rozptylů [μ, σ]=normstat(μ, σ)
  - Náhodné číslo normrnd(μ, σ)

5.5 Normální rozdělení

- Př. 2: Máte naměřená data x=[1.34,1.36,1.42,1.44,1.45,1.48,1.52,1.57,1.57,1.59], zjistěte
  - a) parametry rozdělení,
  - b) pravděpodobnost, že naměřená data budou menší než 1;
  - c) pravděpodobnost, že data budou v intervalu <1.3,1.6>.

  - Výpočet
  - >> x=[1.34,1.36,1.42,1.44,1.45,1.48,1.52,1.57,1.57,1.59]
  - a) [mu,sigma]=normfit(x)

    mu = 1.4740

    sigma = 0.0880
  - b) p1=normcdf(1,a,b)

    p1 = 3.55e-08

    Pravděpodobnost je extrémně malá, protože 1 se ani zdaleka nepřibližuje nejmenšímu číslu
  - c) p2=normcdf(1.6,a,b)-normcdf(1.3,a,b)

    p2 = 0.9000

    Pravděpodobnost je vysoká, protože všechna naměřená data jsou mezi $\langle 1.3,1.6\rangle$

5.6 Normované normální rozdělení

- Normované normální rozdělení je speciální případ normálního rozdělení, kdy střední hodnota $\mu = 0$ a rozptyl $\sigma^2 = 1$.
- Normované normální rozdělení se využívá pro:
  - 1) jednoduchý převod z normálních rozdělení a jejich vzájemné porovnávání.
  - 2) funkce je tabelována a lze pomocí ní provést výpočet distribuční funkce s obecnými parametry.
  - 3) velký význam zejména ve statistice.

5.6 Normované normální rozdělení

- Hustota pravděpodobnosti:

$$f(z) = \frac{1}{\sqrt{2\pi}} e^{\frac{-z^2}{2}}$$

- Distribuční funkce:

$$F(z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} e^{\frac{-t^2}{2}} dt$$

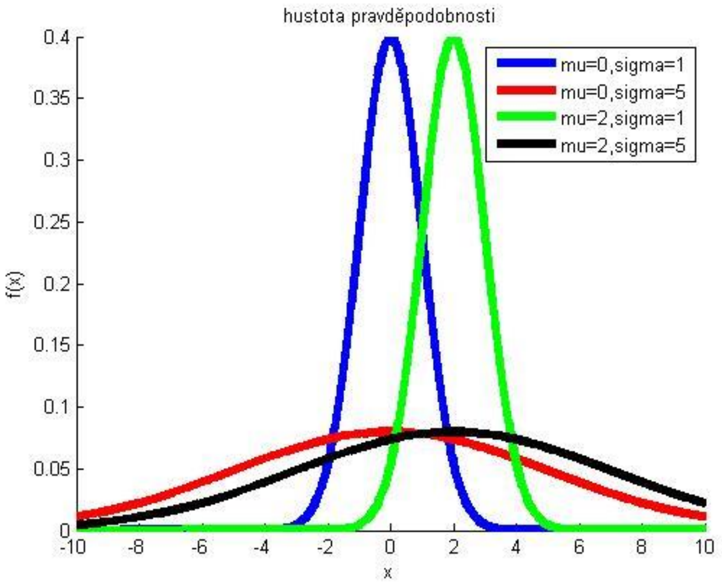
- Hustota pravděpodobnosti normovaného normálního rozdělení je někdy označována $\varphi(z)$.
- Distribuční funkce normovaného normálního rozdělení je někdy označována $\Theta(z)$

5.6 Normované normální rozdělení

- Normování normálního rozdělení
  - Náhodná veličina $\mathrm{X} \rightarrow N(\mu, \sigma^2)$ lze přetransformovat na náhodnou veličinu $\mathrm{Z} \rightarrow N(0,1)$ pomocí transformace

$$\mathrm{Z} = \frac{X - \mu}{\sigma}$$

- K transformaci dat na normované normální rozdělení se používá výše uvedené transformace, a výsledek se nazývá z-skóre.



5.6 Normované normální rozdělení

- Tabulka uvádí hodnotu distribuční funkce (označujeme Θ) normovaného normálního rozdělení
- Sloupec druhé desetinné místo
- Řádek celé číslo a první desetinné místo
- V tabulce z pouze kladné, $\Theta > 0.5$

- z záporné => zjistíme Θ pro z kladné a skutečné Θ=1- Θ odečtené z tabulky
- Θ < 0.5 => vypočteme 1- Θ a hledáme příslušné z v tabulce. Skutečné z =-z odečtené z tabulky

- Př: z=1.92 Θ =0.973
- Př:Θ=0.90 z=1.28
- Př: z=-1.92 Θ =1-0.973=0.027
- Př: Θ=0.10=1-0.9 z=-1.28

Distribuční funkce normovaného normálního rozdělení $\Theta(x)$ pro $x > 0$

$$\Theta(-z) = 1 - \Theta(z)$$

z	0	1	2	3	4	5	6	7	8	9
0,0	0,500	0,504	0,508	0,512	0,516	0,520	0,524	0,528	0,532	0,536
0,1	0,540	0,544	0,548	0,552	0,556	0,560	0,564	0,567	0,571	0,575
0,2	0,579	0,583	0,587	0,591	0,595	0,599	0,603	0,606	0,610	0,614
0,3	0,618	0,622	0,626	0,629	0,633	0,637	0,641	0,644	0,648	0,652
0,4	0,655	0,659	0,663	0,666	0,670	0,674	0,677	0,681	0,684	0,688
0,5	0,691	0,695	0,698	0,702	0,705	0,709	0,712	0,716	0,719	0,722
0,6	0,726	0,729	0,732	0,736	0,739	0,742	0,745	0,749	0,752	0,755
0,7	0,758	0,761	0,764	0,767	0,770	0,773	0,776	0,779	0,782	0,785
0,8	0,788	0,791	0,794	0,797	0,800	0,802	0,805	0,808	0,811	0,813
0,9	0,816	0,819	0,821	0,824	0,826	0,829	0,831	0,834	0,836	0,839
1,0	0,841	0,844	0,846	0,848	0,851	0,853	0,855	0,858	0,860	0,862
1,1	0,864	0,867	0,869	0,871	0,873	0,875	0,877	0,879	0,881	0,883
1,2	0,885	0,887	0,889	0,891	0,893	0,894	0,896	0,898	0,900	0,901
1,3	0,903	0,905	0,907	0,908	0,910	0,911	0,913	0,915	0,916	0,918
1,4	0,919	0,921	0,922	0,924	0,925	0,926	0,928	0,929	0,931	0,932
1,5	0,933	0,934	0,936	0,937	0,938	0,939	0,941	0,942	0,943	0,944
1,6	0,945	0,946	0,947	0,948	0,949	0,951	0,952	0,953	0,954	0,954
1,7	0,955	0,956	0,957	0,958	0,959	0,960	0,961	0,962	0,962	0,963
1,8	0,964	0,965	0,966	0,966	0,967	0,968	0,969	0,969	0,970	0,971
1,9	0,971	0,972	0,973	0,973	0,974	0,974	0,975	0,976	0,976	0,977
2,0	0,977	0,978	0,978	0,979	0,979	0,980	0,980	0,981	0,981	0,982
2,1	0,982	0,983	0,983	0,983	0,984	0,984	0,985	0,985	0,985	0,986
2,2	0,986	0,986	0,987	0,987	0,987	0,988	0,988	0,988	0,989	0,989
2,3	0,989	0,990	0,990	0,990	0,990	0,991	0,991	0,991	0,991	0,992
2,4	0,992	0,992	0,992	0,992	0,993	0,993	0,993	0,993	0,993	0,994
2,5	0,994	0,994	0,994	0,994	0,994	0,995	0,995	0,995	0,995	0,995
2,6	0,995	0,995	0,996	0,996	0,996	0,996	0,996	0,996	0,996	0,996
2,7	0,997	0,997	0,997	0,997	0,997	0,997	0,997	0,997	0,997	0,997
2,8	0,997	0,998	0,998	0,998	0,998	0,998	0,998	0,998	0,998	0,998
2,9	0,998	0,998	0,998	0,998	0,998	0,998	0,998	0,999	0,999	0,999
3,0	0,999	0,999	0,999	0,999	0,999	0,999	0,999	0,999	0,999	0,999
3,1	0,999	0,999	0,999	0,999	0,999	0,999	0,999	0,999	0,999	0,999
3,2	0,999	0,999	0,999	0,999	0,999	0,999	0,999	0,999	0,999	0,999
3,3	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000

5.6 Normované normální rozdělení

- Způsob výpočtu s tabulkami
  - Náh. vel. X Z Z

$$N(\mu,\sigma^2) \Leftrightarrow N(0,1) \quad \Leftrightarrow \quad tabulky$$

  - Zjistíme F(x) $\Theta(z)$ $\Theta(z)$

- Převedení libovolného normálního rozdělení na normované normální rozdělení:

$$\mathrm{Z} = \frac{X-\mu}{\sigma} \qquad X = z\sigma + \mu$$

5.6 Normované normální rozdělení

- Př. Střední hodnota je 5 a víte, že 90 % dat je menších než 8. Zjistěte pravděpodobnost, že data budou větší než 10. Data jsou z normálního rozdělení.
  - Známe $X\colon N(\mu, \sigma^2) = N(5, \sigma^2)$
    90 % dat je menších než 8
  - Neznáme rozptyl, proto nelze použít funkcí v Matlabu.

- Začneme z normovaného normálního rozdělení. 90% kvantilu odpovídá: z=norminv(0.9,0,1)=1.28. 90 % dat je vzdáleno o 1.28 σ od střední hodnoty.
- Převedeme na normální rozdělení a zjistíme směrodatnou odchylku

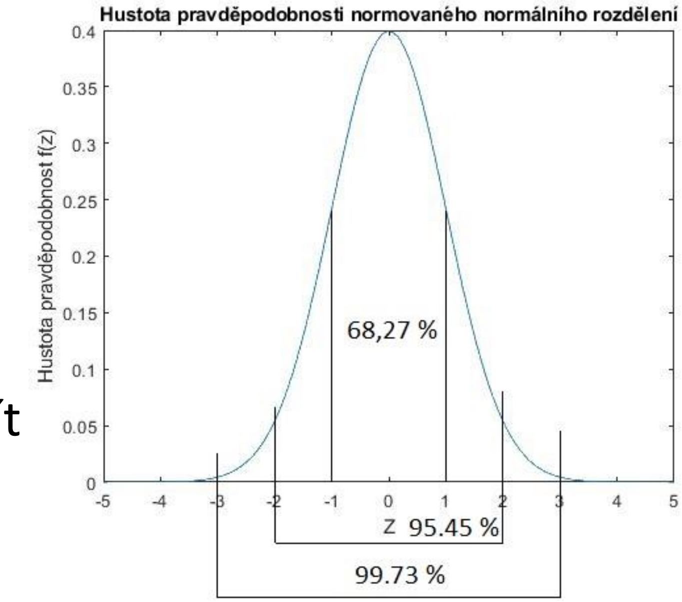
$$X = z \cdot \sigma + \mu$$

$$\sigma = \frac{X - \mu}{z} = \frac{8 - 5}{1.28} = 2.34$$

- Určíme pravděpodobnost, že data budou větší než 10
  - P=1-normcdf(10,5,2.34)=1.63 %

5.6 Normované normální rozdělení

- Vypočtěte pravděpodobnost, že naměřená data z normálního rozdělení jsou v rozmezí $\langle \mu - \sigma, \mu + \sigma \rangle$; $\langle \mu - 2\sigma, \mu + 2\sigma \rangle$; $\langle \mu - 3\sigma, \mu + 3\sigma \rangle$ a $\langle \mu - 4\sigma, \mu + 4\sigma \rangle$.
- Jakékoliv normální rozdělení lze přetransformovat na normované normální rozdělení.
  - $\mathrm{Z} = \frac{X - \mu}{\sigma}$
  - $\langle \mu - \sigma, \mu + \sigma \rangle$ $\mathrm{Z} = \frac{\mu - \sigma - \mu}{\sigma} = -1$ >> p1=normcdf(1,0,1)-normcdf(-1,0,1)
    p1 = 0.6827
  - $\langle \mu - 2\sigma, \mu + 2\sigma \rangle$ p2 = 0.9545
  - $\langle \mu - 3\sigma, \mu + 3\sigma \rangle$ p3 = 0.9973
  - $\langle \mu - 4\sigma, \mu + 4\sigma \rangle$ p4 = 0.9999



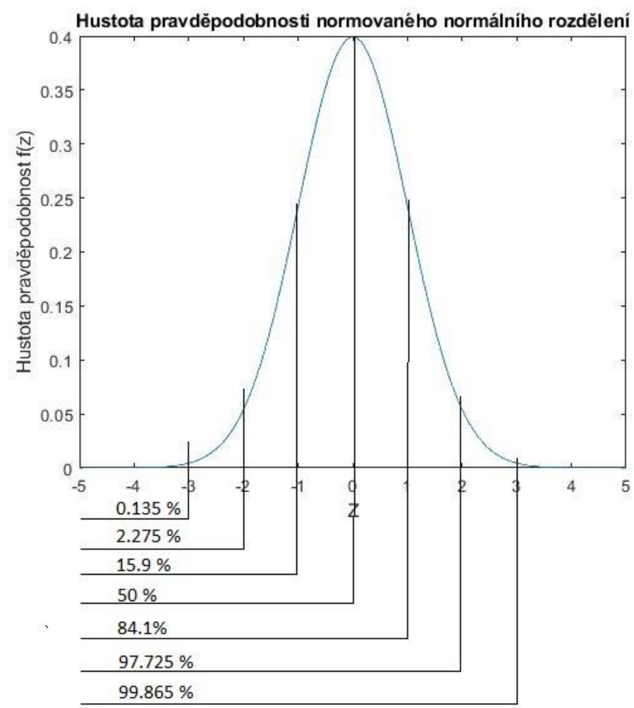
- Často se využívá pravidlo 6σ, které říká, že v daném intervalu je 99.73 % naměřených dat. (data musí být z normálního rozdělení)
- Znalosti pravděpodobností intrvalů $\langle \mu - \sigma, \mu + \sigma \rangle$; $\langle \mu - 2\sigma, \mu + 2\sigma \rangle$; $\langle \mu - 3\sigma, \mu + 3\sigma \rangle$ lze využít pro rychlé odhady správnosti výpočtů

5.6 Normované normální rozdělení

- Pro normální rozdělení platí, že vzdálenostem o $z\sigma$ odpovídá hodnota distribuční funkce:

-4σ	P=0.0000317	4σ	P=0.9999683
-3σ	P=0.00135	3σ	P=0.99865
-2σ	P=0.02275	2σ	P=0.97725
-1σ	P=0.159	1σ	P=0.841
	0σ	P=0.500	

P=0.5	z=0		
P=0.2	z=-0.8416	P=0.8	z=0.8416
P=0.1	z=-1.282	P=0.9	z=1.282
P=0.05	z=-1.645	P=0.95	z=1.645
P=0.02	z=-2.054	P=0.98	z=2.054
P=0.01	z=-2.326	P=0.99	z=2.326



- S jistou mírou nepřesnosti platí i pro symetrická rozdělení

5.7 Logaritmicko-normální rozdělení

- Logaritmicko-normální rozdělení vznikne, jestliže náhodnou veličinu X s normálním rozdělením transformujeme na $Y = e^X$.
  - Použití pro popis náhodných veličin, které lze interpretovat jako multiplikativní výsledek mnoha nepatrných a vzájemně nezávislých vlivů.
  - Jestliže data jsou z logaritmicko-normálního rozdělení, tak zapisujeme $LN(\mu, \sigma^2)$
- Často data zlogaritmujeme a potom se k nim chováme jako z normálního rozdělení.
  - Poznámka z vlastní zkušenosti – pomáhá to s přehledností dat

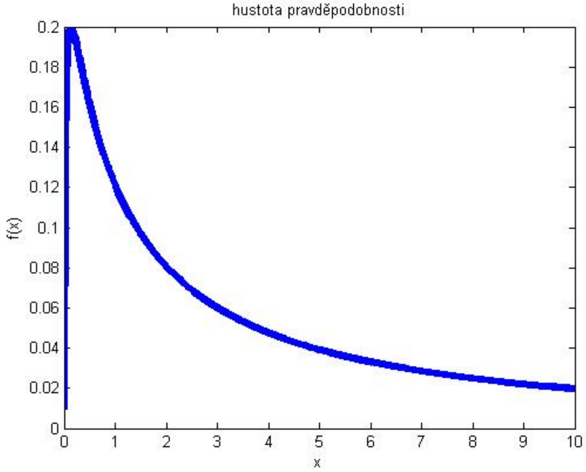
5.7 Logaritmicko-normální rozdělení

- Hustota pravděpodobnosti

$$f(x) = \frac{1}{x\sigma\sqrt{2\pi}} e^{\frac{-(\ln x-\mu)^2}{2\sigma^2}}, x > 0$$

- Střední hodnota $E(X) = e^{\mu+\frac{\sigma^2}{2}}$
- Rozptyl $\mathrm{D}(X) = e^{2\mu+\sigma^2}\left(e^{\sigma^2} - 1\right)$
- Rozdělení je značně nesymetrické: $a = \left(e^{\sigma^2} + 2\right)\sqrt{e^{\sigma^2} - 1}$

$\mu = 2$
$\sigma = 2$



5.7 Logaritmicko-normální rozdělení

- x parametr
- μ – střední hodnota norm. rozdělení σ – směrodatná odchylka norm. rozdělení

- Matlab:
  - Distribuční funkce F=logncdf(x, μ, σ)
  - Hustota pravděpodobnosti f=lognpdf(x, μ, σ)
  - Inverze distribuční funkce x=logninv(pravd, μ, σ)
  - Odhad parametrů [μ, σ]=lognfit(data)
  - Stanovení střední hodnoty a rozptylů [μ, σ]=lognstat(μ, σ)
  - Náhodné číslo lognrnd(μ, σ)

5.7 Logaritmicko-normální rozdělení

- Průměrná velikost zrn písku je 0.5 mm. Je popsaná log-normálním rozdělením. 90 % zrn má menší průměr než 0.8 mm. Určete pravděpodobnost, že zrno bude menší než 1.2 mm
- Data jsou z log normálního rozdělení, lépe se pracuje s jeho logaritmem, proto známá data zlogaritmujeme
  - Průměrná velikost zrn -3.301
  - 90 % zrn je menších než -3.0969
  - Chceme znát kolik procent zrn je menších než 1.2 mm -2.9208
- Střední hodnota je -3.301
- Směrodatnou odchylku neznáme, ale víme že 90 % je menších než -3.0969
- Zjistěte pravděpodobnost, že data budou menší než -2.9208

- Postupujeme dále viz příklad z kapitoly 5.6
  - 90 % kvantil normovaného normálního rozdělení odpovídá: z=norminv(0.9,0,1)=1.2816.
  - 90 % dat je vzdáleno o 1.28 σ od střední hodnoty
  - Převedeme na normální rozdělení a zjistíme směrodatnou odchylku $\sigma = \frac{X-\mu}{z} = \frac{-3.0969-(-3.301)}{1.28} = 0.1595$
  - Určíme pravděpodobnost, že náh. veličina bude menší než -2.9208
  - P=normcdf(-2.9208,-3.301,0.1595)=99.14 %

  - 99.1 % zrn je menších než 1.2 mm.

5.8 Grafické ověření, že data pochází z určitého spojitého rozdělení

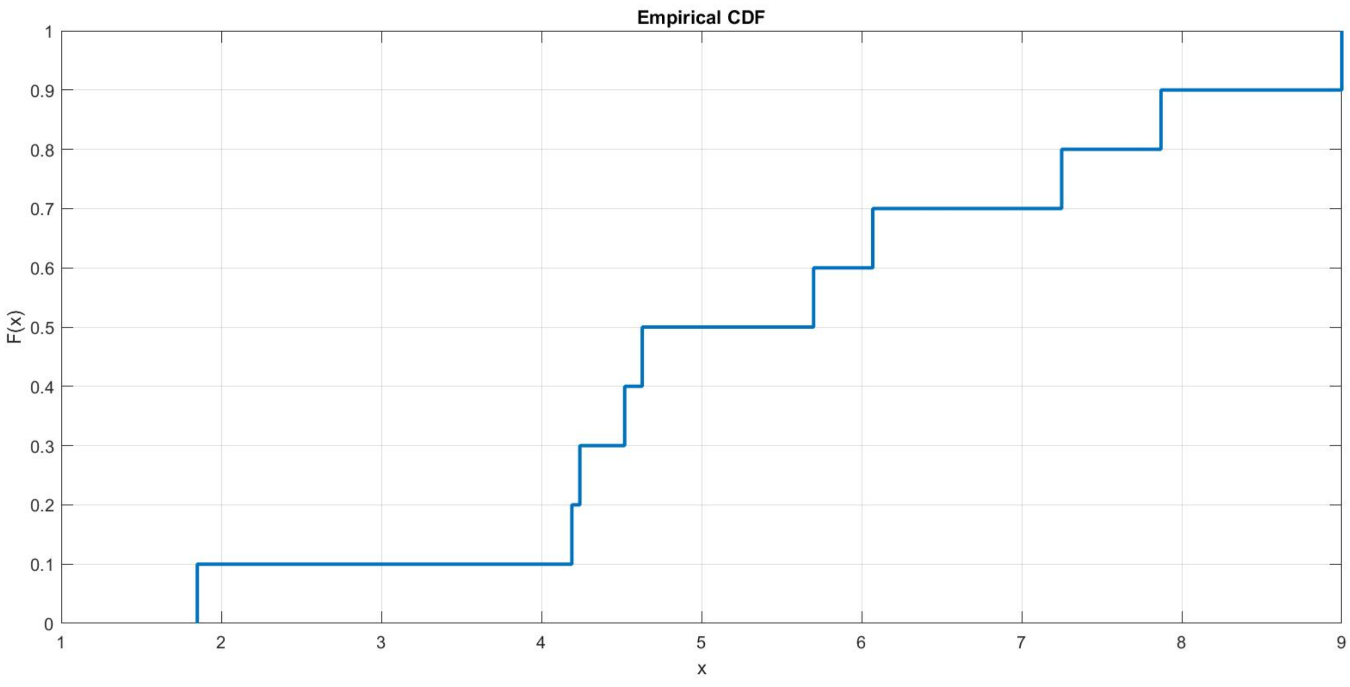
- 5.8.1 Empirická distribuční funkce
- 5.8.2 Weibullův pravděpodobnostní papír
- 5.8.3 Normální pravděpodobnostní papír
- 5.8.4 Obecný pravděpodobnostní papír
- 5.8.5 QQ plot
- 5.8.6 Krabicový graf

5.8.1 Empirická distribuční funkce

- Empirickou distribuční funkci lze použít:
  - Odečtení kvantilů náhodné proměnné
  - Ověření, zda data mohou být z určitého rozdělení
- Získání empirické distribuční funkce
  - 1) setřídíme naměřená data od min po max.
  - 2) $F(x) = \begin{cases} 0, x < x_1 \\ \frac{i}{n}, x_i \leq x < x_{i+1} \\ 1, x > x_n \end{cases}$
  - 3) empirickou distribuční funkci graficky zobrazíme
- Matlab cdfplot(x)
- Lze využít i funkce ecdf, která poskytuje i další textové výsledky.

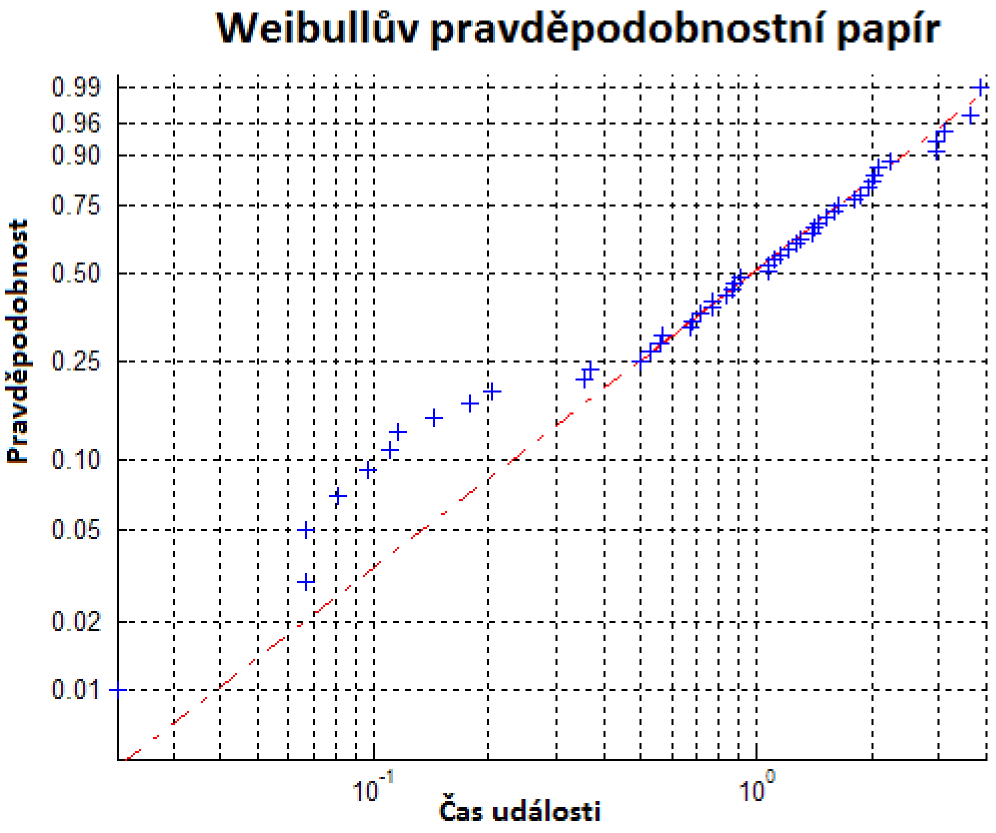
5.8.1 Empirická distribuční funkce

- Př. Vygenerujte 10 dat z normálního rozdělení s parametry $\mu = 5, \sigma^2 = 4$ a vykreslete empirickou distribuční funkci.
  - >> x=normrnd(5,2,1,10)
  - x = [1.85, 4.19, 4.24, 4.52, 4.63, 5.70, 6.07, 7.25, 7.87 9.00]
  - >> cdfplot(x)



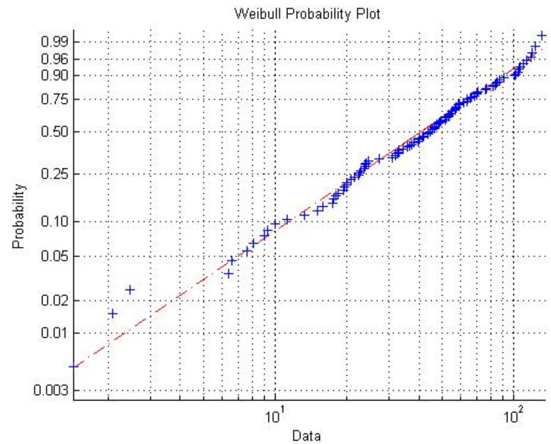
5.8.2 Weibullův pravděpodobnostní papír

- Umožňuje opticky určit, zda data mohou pocházet z exponenciálního nebo Weibullova rozdělení
- Základní princip:
  - Na vodorovné ose se zadávají časy události / naměřené hodnoty
  - Na svislé ose se zaznamenává pravděpodobnost hypotetické distribuční funkce
  - Křížky označují naměřené hodnoty
  - Červená čára představuje proložení křížků přímkou.
  - Podle sklonu a polohy červené čáry lze zjistit parametry daného rozdělení
  - Křížky leží v blízkosti červené čáry – data pocházejí z daného rozdělení
  - Křížky netvoří přímku (tvar S, konkávní nebo konvexní funkce) data nelze proložit daným typem rozdělení
- Matlab wblplot(data)



5.8.2 Weibullův pravděpodobnostní papír

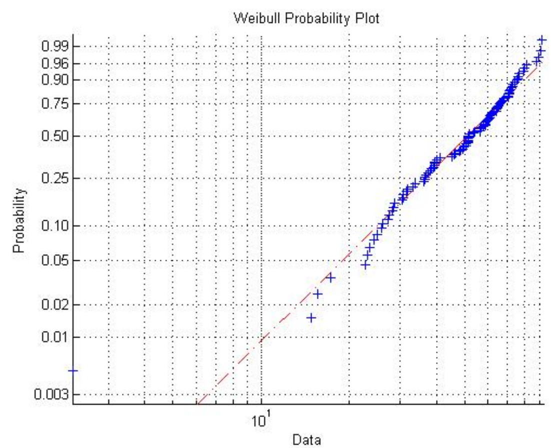
Matlab funkce: wblplot(data)

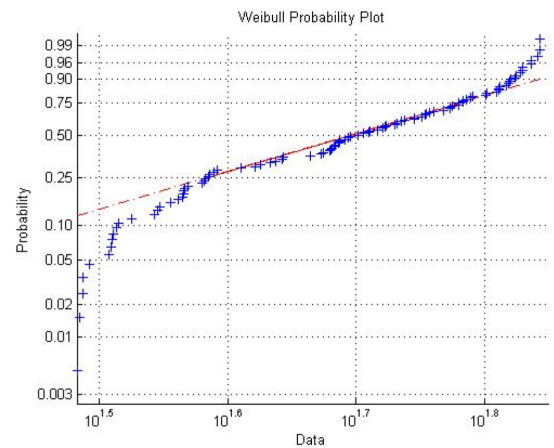


% horní obrázek, data z Weibullova rozdělení
data=wblrnd(50,1.5,100,1);
wblplot(data)
wblfit(data)

% levý dolní obrázek
% data nejsou z Weibullova rozdělení
data=normrnd(50,20,100,1);
wblplot(data)
wblfit(data)

% pravý dolní obrázek
% data nejsou z Weibullova rozdělení
data=unifrnd(30,70,100,1);
wblplot(data)
wblfit(data)

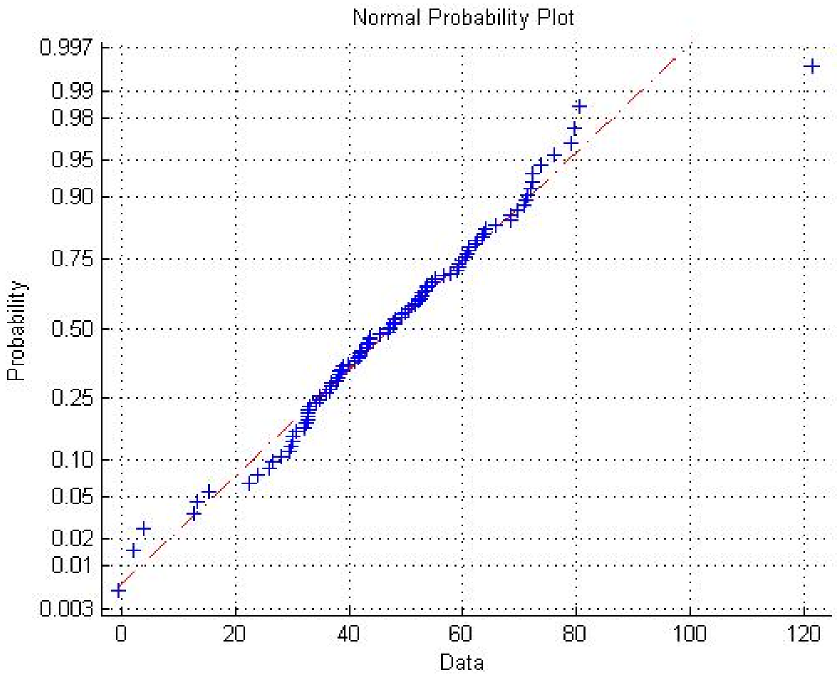




5.8.3 Normální pravděpodobnostní papír

- Obdobný princip jako Weibullův pravděpodobnostní papír
  - Matlab: normplot(data)

```
data=normrnd(50,20,100,1);
normplot(data)
normfit(data)
```

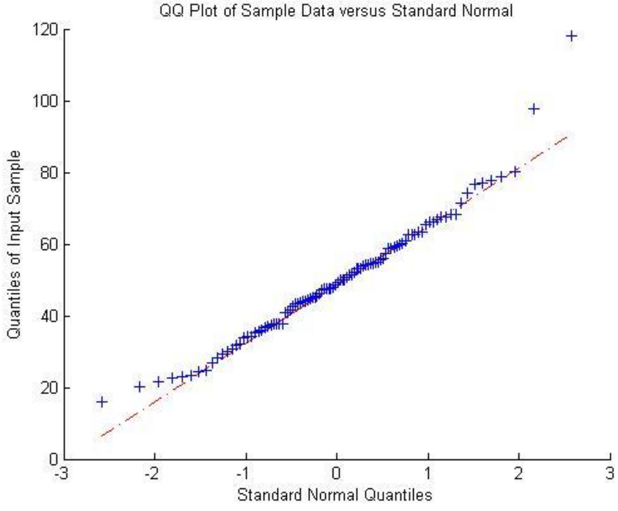


5.8.4 Obecný pravděpodobnostní papír

- Pravděpodobnostní papíry i pro další rozdělení
  – Matlab funkce: probplot(typ rozdělení,data)
  – Umožňuje výpočet pro následující typy rozdělení:
    - Exponenciální 'exponential'
    - Lognormální 'lognormal'
    - Normalní 'normal'
    - Weibullovo 'weibull'
  – Př: probplot('lognormal',data)

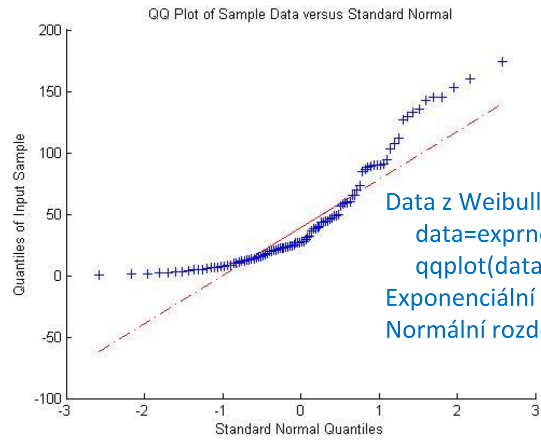
5.8.5 QQ plot

- QQ plot umožňuje stanovit, zda data pochází z normálního rozdělení.
  - Vodorovná osa – teoretické z-skore normovaného normálního rozdělení
  - Svislá osa – naměřená data
  - Matlab funkce: qqplot(data)
- Jestliže data jsou na přímce, jsou z normálního rozdělení

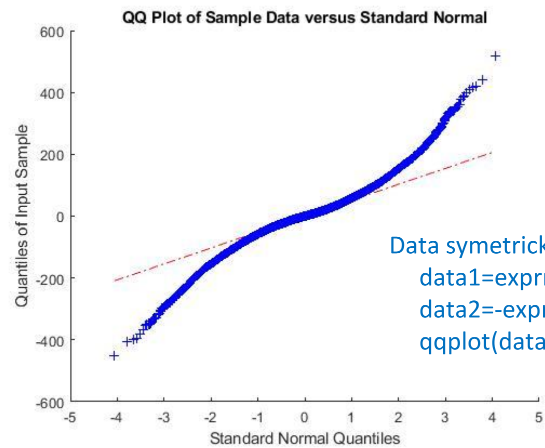


5.8.5 QQ plot

- Výsledky:
  - Normální rozdělení
    přímka
  - Normální rozdělení s odlehlými daty
    přímka s koncovými daty mimo tuto přímku
  - Kladná šikmost dat
    konvexní tvar
  - Záporná šikmost dat
    konkávní tvar
  - Vyšší špičatost než má normální rozdělení
    - konkávní a pak konvexní tvar
  - Nižší špičatost než má normální rozdělení
    - konvexní a pak konkávní tvar
    - Je vidět, pokud data mají nulovou nebo skoro nulovou šikmost



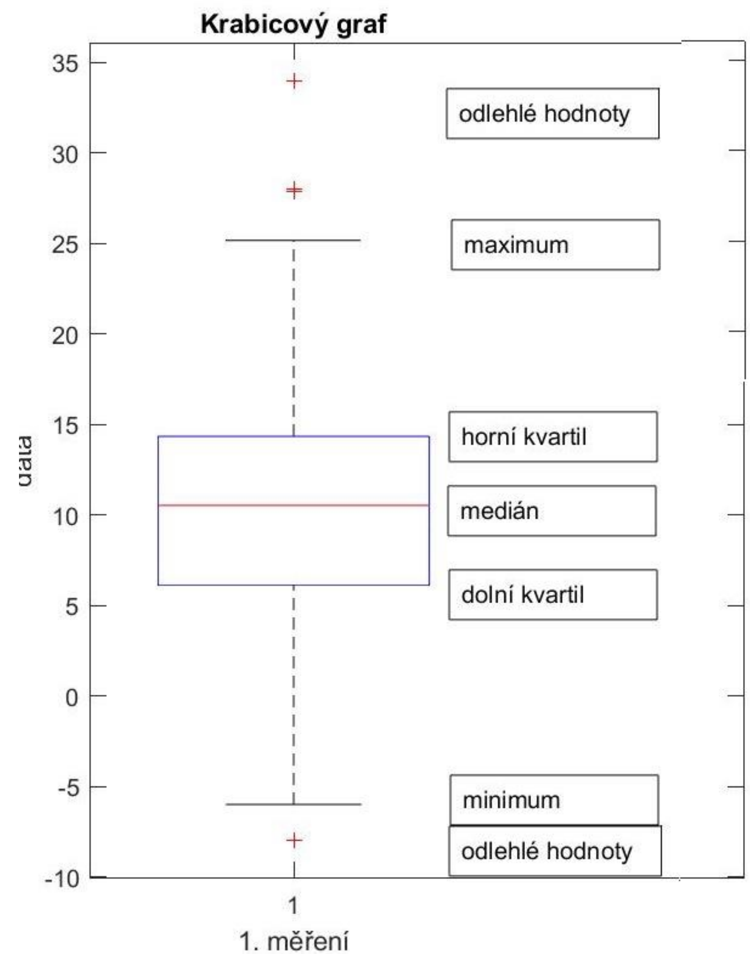
Data z Weibullova rozdělení
data=exprnd(50,100,1);
qqplot(data);
Exponenciální rozdělení: šikmost = 2 špičatost = 6
Normální rozdělení: šikmost= 0 špičatost = 3



Data symetrická s vyšší špičatostí (špičatost je 6)
data1=exprnd(50,10000,1);
data2=-exprnd(50,10000,1); data=[data1;data2]
qqplot(data);

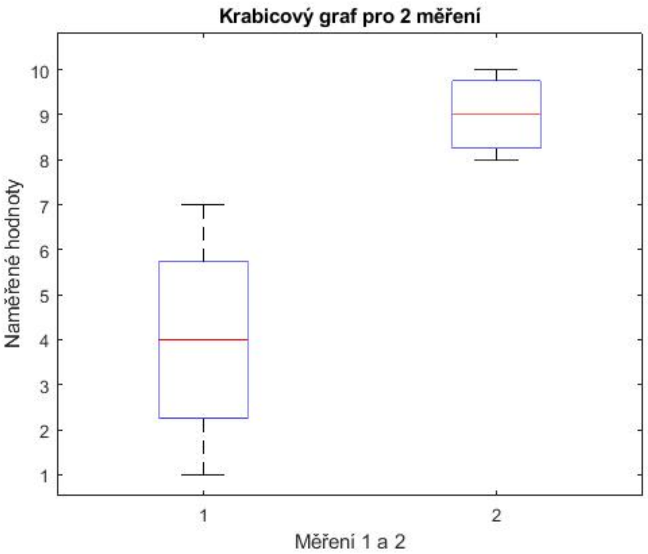
5.8.6 Krabicový graf

- Krabicový graf se využívá pro rychlé zobrazení:
  - Mediánu
  - Dolního a horního kvartilu
  - Minima a maxima
  - Odlehlých dat.
- Lze odhadnout
  - Velikost rozptylu – z interkvartilového rozpětí
  - Symetrii
    - Shoda průměru a mediánu
    - Rozdíl $x_{0.75} - x_{0.5} \approx x_{0.5} - x_{0.25}$
- Matlab – boxplot(x)



5.8.6 Krabicový graf

- Při více náhodných výběrů lze pomocí krabicového grafu stanovit trend středních hodnot a velikost rozptylu.
- boxplot(x,y)
  - x vektor naměřených hodnot
  - y vektor k jakému měření hodnoty patří
- Př: x=[1,2,3,4,5,6,7, 8,9,10]
  y=[1,1,1,1,1,1,1, 2,2,2]



Střední hodnota prvního výběru je menší než druhého
Rozptyl prvního výběru je větší než druhého

5.9 Základní příkazy v matlabu/octave

	Rovnom. rozdělení	Exponenc. rozdělení	Weibull. rozdělení	Erlang. rozdělení	Normální rozdělení	Lognorm. rozdělení
Distribuční funkce	unifcdf	expcdf	wblcdf	gamcdf	normcdf	logncdf
Hustota pravděp.	unifpdf	exppdf	wblpdf	gampdf	normpdf	lognpdf
Inverzní Funkce	unifinv	expinv	wblinv	gaminv	norminv	Logninv
Odhad parametrů	unifit	expfit	wblfit	gamfit	normfit	lognfit
Střed hod. a rozptyl	unifstat	expstat	wblstat	gamstat	normstat	lognstat
Náhod. číslo	unifrnd	exprnd	wblrnd	gamrnd	normrnd	lognrnd

5.9 Základní příkazy v matlabu/octave

• Empirická distribuční funkce	cdfplot
• Weibullův pravděpodobnostní papír	wblplot
• Normální pravděpodobnostní papír	normplot
• Obecný pravděpodobnostní papír	probplot
• QQ plot	qqplot
• Krabicový graf	boxplot

6 Výběrové charakteristiky

- 6.1 Výběrové charakteristiky
- 6.2 Výběrový průměr
- 6.3 Limitní věty
- 6.4 Rozdíl výběrových průměrů
- 6.5 Relativní četnost
- 6.6 Rozdíl výběrových četností
- 6.7 Spojitá rozdělení užívaná ve statistice

6.1 Výběrové charakteristiky

- Pravděpodobnost
  - Pravděpodobnost každého jevu je předem dána a je neměnná – konstantní hodnota
    - Př. Hod mincí 50 % orel, 50 % panna
    - Pravděpodobnost výhry v sázkových hrách
  - Výsledek – pravděpodobnost, že v 10 náhodných pokusech nastane úspěch právě k-krát
  - Charakteristiky: střední hodnota, rozptyl, směrodatná odchylka, medián, pravděpodobnost
- Statistika
  - Pravděpodobnost úspěchu jevu se snažíme zjistit za pomoci naměřených dat – stochastická hodnota
    - Střední doba do poruchy n výrobků – měříme dobu do poruchy a z dat vypočteme střední hodnotu. U jiného souboru shodných výrobků obdržíme odlišné výsledky.
    - Průměrný plat 20 pracujících v ČR.
  - Obdržíme výběrové charakteristiky náhodných veličin
    - Výběrová střední hodnota, výběrový rozptyl, výběrová směrodatná odchylka, výběrový medián, relativní četnost

6.1 Výběrové charakteristiky

Pravděpodobnost	Statistika
Střední hodnota $E(X)$, μ	Výběrový průměr $\bar{X}$
Rozptyl $\mathrm{D}(X)$, σ^2	Výběrový rozptyl s^2
Směrodatná odchylka σ	
	Výběrová směrodatná odchylka s
Pravděpodobnost jevu π	Relativní četnost p
Medián $x_{0.5}$	Výběrový medián $\widetilde{X_{0.5}}$

6.1 Výběrové charakteristiky

- Příklad:
  - 1) Zeptejte se 10 lidí zda preferují čaj nebo kávu. Obdržíte nějakou relativní četnost p_1.
  - 2) Zeptejte se 1000 lidí na stejnou otázku. Opět obdržíte nějakou relativní četnost p_2.
  - 3) Nechme hlasovat všechny lidi v ČR na stejnou otázku, obdržíme pravděpodobnost jevu π.
- Otázka:
  - Bude se (obecně) více odlišovat od pravděpodobnosti jevu π relativní četnost p_1, nebo relativní četnost p_2?
  - Pravděpodobně se více bude odlišovat relativní četnost p_1, protože se ptáme malého množství lidí a bude zde velký rozptyl výsledku od skutečnosti.

6.2 Součet náhodných veličin a výběrový průměr

- Vlastnosti střední hodnoty a rozptylu nezávislých náhodných veličin
  - $E(\sum_i X_i) = \sum_i E(X_i)$ výsledná střední hodnota je dána součtem nezávislých středních hodnot náhodných veličin
  - $D(\sum_i X_i) = \sum_i D(X_i)$ výsledný rozptyl je dán součtem rozptylů nezávislých náhodných veličin

----

  - $E(aX) = aE(X)$ vynásobení náhodné veličiny konstantou, se výsledná střední hodnota také vynásobí konstantou
  - $\mathrm{D}(aX) = a^2 D(X)$ vynásobení náhodné veličiny konstantou a, se výsledný rozptyl vynásobí a^2.

6.2 Součet náhodných veličin a výběrový průměr

- $E(\sum_i X_i) = \sum_i E(X_i)$ $\qquad$ $D(\sum_i X_i) = \sum_i D(X_i)$
- $E(aX) = aE(X)$ $\qquad$ $\mathrm{D}(aX) = a^2 D(X)$

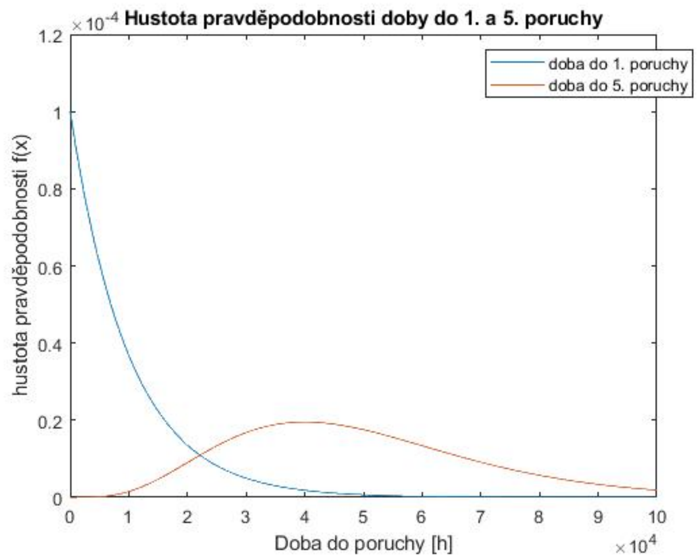
- Ověření na příkladu
  - Pravděpodobnost úspěchu v 1 tahu je p a lze popsat alternativním rozdělením
    - $E(X) = p;$ $\qquad$ $D(X) = p \cdot (1-p)$
  - Pravděpodobnost, že budu úspěšný právě k krát v n tazích lze popsat binomickým rozdělením.
    - $E(X) = n \cdot p;$ $\qquad$ $D(X) = n \cdot p \cdot (1-p)$
    - Hody jsou vzájemně nezávislé
    - Střední hodnota u binomického rozdělení je n krát vyšší než u alternativního. Stejně tak pro rozptyl.

  - Chceme znát rozdělení výběrového průměru, tj. máme n tahů (pokusů) a k krát jsme byli úspěšní. Chceme odhadnout pravděpodobnost úspěchu v průměrném tahu (násobíme $\frac{1}{n}$) a jak moc se můžeme zmýlit (rozptyl).
  - $E\left(\frac{X}{n}\right) = \frac{n \cdot p}{n} = p$ $\qquad$ $\mathrm{D}\left(\frac{X}{n}\right) = \frac{n \cdot p \cdot (1-p)}{n^2} = \frac{p \cdot (1-p)}{n}$

6.2 Součet náhodných veličin a výběrový průměr

- Příklad 2
- Doba do poruchy výrobku je popsána exponenciálním rozdělením a je $\mu =$10000 hodin, tj. $\lambda = 10^{-4}h^{-1}$.
  - Pro exponenciální rozdělení platí, že:
    - $E(X) = \mu = \frac{1}{\lambda} = 10000\ h$ $\qquad D(X) = \mu^2 = \frac{1}{\lambda^2} = 10^8 h^2$
  - Doba do 5. poruchy (poruchy se vzájemně neovlivňují, jsou nezávislé) výrobku by měla být:
    - $E(\sum_i X_i) = \sum_i E(X_i) = 50000\ h$
    - $D(\sum_i X_i) = \sum_i D(X_i) = 5 \cdot 10^8 h^2$

- Doba do páté poruchy je popsána Erlangovým rozdělením, které má:
  - Střední hodnotu $E(X) = \frac{k}{\lambda} = \frac{5}{\lambda} = 50000\ h$
  - Rozptyl $D(X) = \frac{k}{\lambda^2} = \frac{5}{\lambda^2} = 5 \cdot 10^8 h^2$
  - Směrodatná odchylka $\sigma(X) = 22300\ h$
- Poznámka: všimněte si, že rozdělení doby do 5. poruchy se stává více symetričtější než do 1. poruchy (exponenciálního rozdělení).



6.2 Součet náhodných veličin a výběrový průměr

- Pokračování příkladu 2: Doba do poruchy výrobku je popsána exponenciálním rozdělením a je 10000 hodin, tj. $\lambda = 10^{-4} h^{-1}$.
  - Pro exponenciální rozdělení platí, že:
    - $E(X) = \frac{1}{\lambda} = 10000\ h$ $\quad$ $D(X) = \frac{1}{\lambda^2} = 10^8 h^2$
  - Doba do 5. poruchy (poruchy se vzájemně neovlivňují, jsou nezávislé) výrobku by měla být:
    - $E(\sum_i X_i) = 50000\ h$ $\quad$ $D(\sum_i X_i) = 5 \cdot 10^8 h^2$

- Jak bude rozdělena střední doba do poruchy?
  - Střední doba do poruchy se vypočte $\frac{\sum_{i=1}^{n} t_i}{n}$, kde t_i jsou doby do poruchy a n je počet poruch
  - Ze znalosti vlastností střední hodnoty platí : $E(aX) = aE(X)$
  - Ze znalosti vlastností rozptylu platí : $\mathrm{D}(aX) = a^2 D(X)$

- Střední doba do poruchy zjištěná z 5 poruch bude mít:
  - $E\left(\frac{1}{n}X\right) = \frac{1}{n}E(X) = \frac{50000}{5} = 10000\ hod$ (porovnej pro 1 výrobek je to 10000 hod)
  - Rozptyl tohoto odhadu bude:
  - $\mathrm{D}\left(\frac{1}{n}X\right) = \frac{1}{n^2}D(X) = \frac{5 \cdot 10^8}{25} = 0.2 \cdot 10^8\ hod^2$ (porovnej pro 1 výrobek je to $10^8\ hod^2$)

6.2 Součet náhodných veličin a výběrový průměr

- Zobecněně: Mějme náhodný výběr $X_1, X_2, \dots X_n$ z náhodné veličiny X s distribuční funkcí $F(x)$. Označme μ_X střední hodnotu a σ_X směrodatnou odchylku náhodné veličiny X_i. Potom výběrový průměr náhodného výběru $X_1, X_2, \dots X_n$ rozumíme náhodnou veličinu: $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ mající $E(\bar{X}) = \mu_X$ a $D(\bar{X}) = \frac{\sigma_X^2}{n}$.
- Věta obsahuje ještě další předpoklady: Předpokládejme dále, že všechny dvojice náhodných veličin jsou nezávislé. A že všechny náhodné veličiny X_i mají stejnou a konečnou střední hodnotu i směrodatnou odchylku.

- Pochází-li náhodný výběr $X_1, X_2, \dots X_n$ z normálního rozdělení $N(\mu_X, \sigma_X^2)$, pak výběrový průměr má normální rozdělení s parametry $N(\mu_X, \frac{\sigma_X^2}{n})$.

6.2 Součet náhodných veličin a výběrový průměr

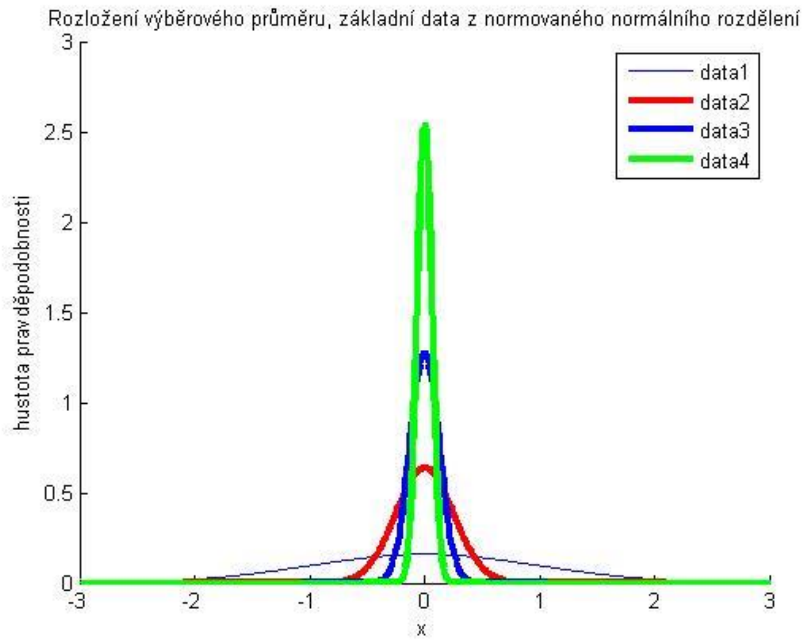
Úkol:

- Generujme 16 hodnot z normovaného normálního rozdělení.
- Vygenerované hodnoty sečtěme a jejich součet vydělme 16. (obdržíme 1 realizaci z výběrového průměru)
- Pokus opakujme mnohokrát.
- Vytvořme hustotu pravděpodobnosti.

- Totéž opakujte pro 64 a 256 hodnot.
- Obdržíme rozložení výběrového průměru.

Rozložení výběrového průměru pro (výběr je z normovaného normálního rozdělení):

1 hodnota – slabá modrá 16 hodnot – červená
64 hodnot – silná modrá 256 hodnot – zelená



6.3 Limitní věty

- 6.3.1 Zákon velkých čísel
- 6.3.2 Centrální limitní věta
- 6.3.3 Příklady na centrální limitní větu

6.3.1 Zákon velkých čísel

- Jestliže výběr pochází z normálního rozdělení, pak s rostoucím rozsahem výběru se výběrový průměr soustřeďuje kolem střední hodnoty. (vlastnost výběrového průměru).

- Mějme nekonečný náhodný výběr $X_1, X_2, \ldots$ z rozdělení se střední hodnotou μ_X a konečným rozptylem σ_X^2, kde $X_1, X_2, \ldots$ jsou nezávislé náhodné veličiny. Potom platí, že výběrový průměr $\overline{X_n}$ vypočítaný z prvních n pozorování se pro $n \to \infty$ blíží ke střední hodnotě μ_X.

$$\lim_{n\to\infty}\left[P(|\overline{X_n} - \mu_X| > \varepsilon)\right] = 0 \qquad \forall \varepsilon > 0$$

  – Všimněte si, že ve větě se vůbec nemluví o typu rozdělení

6.3.1 Zákon velkých čísel

- Pravděpodobnost úspěchu náhodného pokusu je 0.75. Tento náhodný pokus opakujeme n-krát (n je proměnné). Spočtěte pravděpodobnost, že relativní četnost náhodného jevu bude při n pokusů menší nebo rovna 0.7.
  - n krát opakovaný pokus lze popsat binomickým rozdělením.
  - $\lim_{n \to \infty} [P(|\overline{X_n} - \mu_X| > \varepsilon)] = 0 \qquad \forall \varepsilon > 0$
  - binocdf(0.7*n,n,0.75)

n	**pravděpodobnost**
10	0.4744
100	0.1495
1000	1.9359E-4
10000	6.1496E-30
100000	1.2049E-280

6.3.2 Centrální limitní věta

Součet náhodných veličin

- Centrální limitní věta rozšiřuje zákon velkých čísel o tvrzení, že za určitých podmínek lze součet náhodných veličin, nebo výběrový průměr popsat pomocí normálního rozdělení.
- Mějme X_i nezávislé náhodné veličiny ze stejného rozdělení, se stejnou střední hodnotou a s konečným rozptylem. Pak součet n náhodných veličin má při dostatečně velkém počtu pozorování přibližně normální rozdělení, ať X_i pochází z libovolného rozdělení. Normální rozdělení má parametry:

$$\sum_{i=1}^{n} X_i \sim N\left(n \cdot \mu_{X,} n \cdot \sigma_X^2\right)$$

nebo použitím transformace na normované normální rozdělení

$$\frac{\sum_{i=1}^{n} X_i - n \cdot \mu_X}{\sqrt{n} \cdot \sigma_x} \sim N(0,1)$$

  - Obvykle se za dostatečně velké označují výběry o rozsahu 30 a větší.
  - Pro výběrové rozdělení symetrické, unimodální, které neobsahuje odlehlá pozorování stačí výběr o rozsahu minimálně 15 dat.

6.3.2 Centrální limitní věta

Výběrový průměr

- Centrální limitní věta rozšiřuje zákon velkých čísel o tvrzení, že za určitých podmínek lze součet náhodných veličin, nebo výběrový průměr popsat pomocí normálního rozdělení.
- Mějme X_i nezávislé náhodné veličiny ze stejného rozdělení, se stejnou střední hodnotou a s konečným rozptylem. Pak výběrový průměr má při dostatečně velkém počtu pozorování přibližně normální rozdělení, ať X_i pochází z libovolného rozdělení. Normální rozdělení má parametry:

$$\bar{X} \sim N\left(\mu_{X,} \frac{\sigma_X^2}{n}\right)$$

nebo použitím transformace na normované normální rozdělení

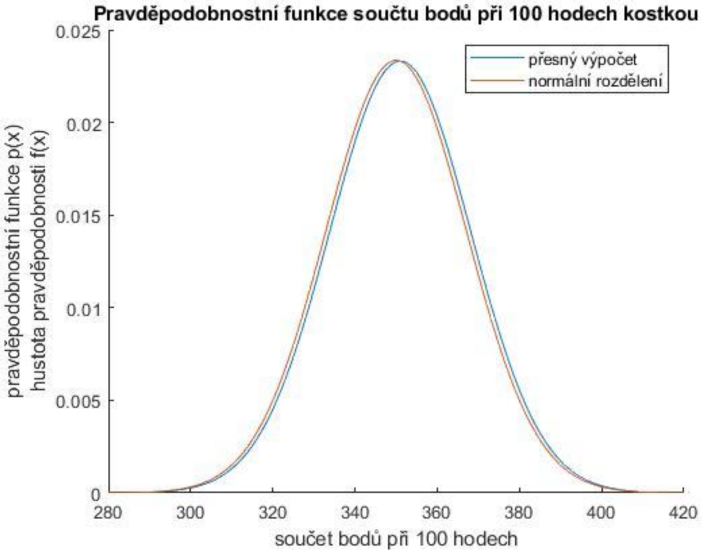
$$\frac{\bar{X} - \mu_X}{\sigma_x} \sqrt{n} \sim N(0{,}1)$$

  - Obvykle se za dostatečně velké označují výběry o rozsahu 30 a větší.
  - Pro výběrové rozdělení symetrické, unimodální, které neobsahuje odlehlá pozorování stačí výběr o rozsahu minimálně 15 dat.

6.3.2 Centrální limitní věta

- Ověření na příkladu
  - Budete házet šestistěnnou kostkou
    - $E(X) = 3.5$ $\quad D(X) = \frac{35}{12}$ $\quad \sigma(X) = 1.708$
  - Budete házet 100x a body sečtete. Jaká je střední hodnota a jaký rozptyl.
    - $E(X) = 350$ $\quad D(X) = \frac{3500}{12}$ $\quad \sigma(X) = 17.08$

Porovnejte „hustotu“ přesného rozdělení s normálním rozdělením s parametry $\left(\mu = 350, \sigma^2 = \frac{3500}{12}\right)$.



6.3.2 Centrální limitní věta

Ljapunovova věta - Součet náhodných veličin

- Mějme X_i nezávislé náhodné veličiny s konečnou střední hodnotou, rozptylem a nechť platí podmínka výsledné symetrie ($\lim_{n\to\infty} \frac{\left[\sum_{i=1}^{n} \mu_3{}'\right]^{\frac{1}{3}}}{\sqrt{\sum_{i=1}^{n} D(X_i)}} = 0$), pak při dostatečně velkém počtu má součet náhodných náhodných veličin přibližně normální rozdělení, které má parametry:

$$\sum_{i=1}^{n} X_i \sim N\left(\sum_{i=1}^{n} \mu_i\,, \sum_{i=1}^{n} \sigma_i^2\right)$$

nebo použitím transformace na normované normální rozdělení

$$\frac{X - \sum_{i=1}^{n} E(X_i)}{\sqrt{\sum_{i=1}^{n} D(X_i)}} \sim N(0,1)$$

  - Obvykle se za dostatečně velké označují výběry o rozsahu 30 a větší.
  - Pro výběrové rozdělení symetrické, unimodální, které neobsahuje odlehlá pozorování stačí výběr o rozsahu minimálně 15 dat.

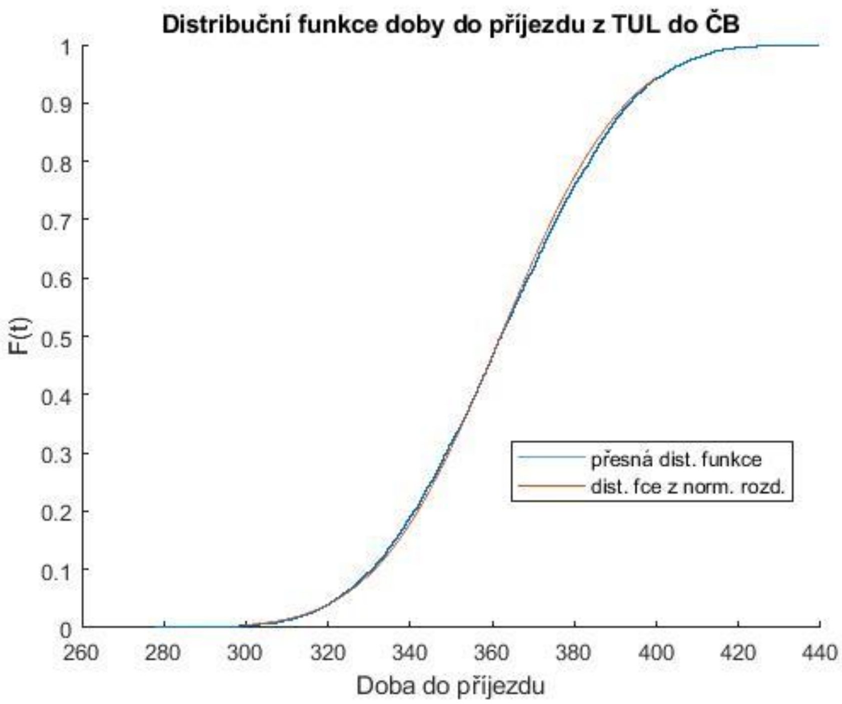
6.3.2 Centrální limitní věta

Ljapunovova věta - Součet náhodných veličin

Příklad

- Chceme se přesunout z TUL do Českých Budějovic. Celou cestu můžeme rozdělit na následující úseky. U každého je uvedena střední doba, rozptyl a typ rozdělení. Na příkladu ověřte platnost centrální limitní věty.
- Maximální rozdíl obou distribučních funkcí je 1.05%

Úsek	**Stř. doba [min]**	**Rozptyl [min^2]**	**rozdělení**
TUL – Fugnerova	20	100/12	Rovnoměrné (a=15, b=25)
Fugnerova čekání	15	400/12	Rovnoměrné (a=5, b=25)
Bus do Prahy	70	25	Normální (mu=70, sigma=5)
Čekání na metro	4	4	Normální (mu=4, sigma=2)
Metro B	20	1	Normální (mu=20, sigma=1)
Přestup na C	8	3	Rovnoměrné (a=5, b=11)
Metro C a výstup	5	4/12	Rovnoměrné (a=4, b=6)
Čekání na vlak	30	300	Rovnoměrné (a=0, b=60)
Jízda vlakem	150	100	Normální (mu=150, sigma=10)
Cesta od vlaku	40	100	Normální (mu=40, sigma = 10)
Součet	**362 min**	**575 min^2**	



6.3.3 Příklady na centrální limitní větu

Příklad 1:

Životnost zařízení má exponenciální rozdělení se střední hodnotou 5 let. Určete pravděpodobnost, že průměrná životnost 100 prodaných zařízení bude vyšší než 5.6 let.

- $\mu = 5\ roků$
- $E(X) = \mu$=5 roků $\qquad D(X) = \mu^2 = 25\ roků^2$

- Z CLV platí, že průměrná životnost bude popsána normálním rozdělením.

- $\bar{X} \sim N\left(\mu_{X,} \frac{\sigma_X^2}{n}\right) = N\left(\mu_X = 5\ roků, \frac{\sigma_X^2}{n} = \frac{25}{100} roků^2\right)$
  - $N(5\ roků, 0.25\ roků^2)$
  - Směrodatná odchylka je 0.5 roku.
  - Matlab: P=1-normcdf(5.6,5,0.5)
  - P=0.1151
  - Pravděpodobnost, že 100 zařízení bude mít větší střední hodnotu životnosti než 5.6 let je 0.1151.

6.3.3 Příklady na centrální limitní větu

- Ověřme výsledek pomocí metody Monte Carlo (statistická pravděpodobnost)

```
clear all
clc
iteraci = 1000000;
pocet = 0;
for i = 1:iteraci
    t = sum(exprnd(5,1,100));
    if t > 560
        pocet = pocet+1;
    end
end
pravd = pocet/iteraci
```

- Pravděpodobnost vyšla 0.1179, tedy podobná hodnota jako u výpočtu pomocí CLV.

6.3.3 Příklady na centrální limitní větu

Příklad 2

- Čas pro odhalení poruchy a její opravu trvá 1 hodinu se směrodatnou odchylkou 0.5 hod. Určete pravděpodobnost, že doba potřebná k odhalení a opravení 100 nezávislých poruch nepřekročí 110 hodin.
  - Známe střední hodnotu μ=1 hod a směrodatnou odchylkou σ=0.5 hod., $\sigma^2 = 0.25$ hod^2
  - Chceme znát rozdělení součtu 100 nezávislých poruch a pravděpodobnost P(X<110 hod)
  - Z CLV platí, že součet času k opravení 100 výrobků lze popsat normálním rozdělením .
  - Parametry rozdělení: $\mathrm{N}(100 \cdot \mu, 100 \cdot \sigma^2) = \mathrm{N}(100{,}25)$
  - Do Matlabu se zadává směrodatná odchylka
  - Matlab: $\mathrm{P}(\mathrm{X} < 110) = \mathrm{normcdf}(110{,}100{,}5)$
  - $\mathrm{P}(\mathrm{X} < 110)$=0.9772

6.3.3 Příklady na centrální limitní větu

- Příklad 3:
- Starý turistický most má nosnost 2000 kg. Kolik na most může maximálně vstoupit lidí, jestliže chceme aby pravděpodobnost překročení hmotnosti byla 1E-6. Člověk má průměrnou hmotnost 80 kg, směrodatná odchylka je 30 kg.
  - nejsou splněny všechny předpoklady CLV, protože první odhad počtu lidí na mostě 2000/80=25, což je méně než 30.
  - $\mu = 80$ kg $\sigma = 30$ kg $\sigma^2 = 900$ kg²
  - Z CLV platí, že součet hmotností lze popsat normálním rozdělením
  - $\mathrm{P(X > 2000) = 0.000\,001 \rightarrow N(n \cdot \mu, n \cdot \sigma^2)}$ maximalizujeme n
  - $\mathrm{P(X < 2000) = 0.999\,999 \rightarrow N(n \cdot \mu, n \cdot \sigma^2)}$
- Početní řešení – převedeme normální rozdělení na normované normální rozdělení a dále počítáme s ním
  - $X \sim N(n \cdot \mu_X, n \cdot \sigma_X^2)$
  - Pravděpodobnosti 0.999999 odpovídá u normovaného normálního rozdělení z-skore 4.75342, který zjistíme pomocí příkazu norminv(0.999999,0,1)
  - Převedeme parametry normálního rozdělení na normované normální rozdělení pomocí transformace $\frac{X-\mu}{\sigma} = N(0,1)$. $\frac{X-\mu}{\sigma} = 4.75342$
  - $\frac{2000 - n \cdot 80}{\sqrt{n \cdot 900}} = 4.75342$
  - Řešíme dále iracionální rovnici (pozor u řešení je nutnost provést zkoušku!!!)
  - Přesný výpočet vyjde 17.54

6.3.3 Příklady na centrální limitní větu

- Matlabovský skript
- Zvyšujeme počet lidí na mostě o 1 a zjišťujeme, kdy pravděpodobnost překročení hmotnosti se zvýší nad hranici 0.000 001.

```
i=0;    pravd=1;
while pravd>0.999999
      i=i+1;
      pravd=normcdf(2000,i*80,sqrt(i*900));
 end
 i-1
```

- 17 lidí na mostě s pravděpodobností 1E-6 nepřekročí hmotnost 2000 kg.

6.3.3 Příklady na centrální limitní větu

- Př. 4
- Vypočtěte pravděpodobnost, že při 60x opakovaném hodu šestistěnnou kostkou bude součet bodů v intervalu <190,230>.
- Pravděpodobnost hodu: $P(1) = P(2) = P(3) = P(4) = P(5) = P(6) = \frac{1}{6}.$
- Střední hodnota jednoho hodu je $E(X) = \frac{1+2+3+4+5+6}{6} = \frac{21}{6} = 3.5.$
- Rozptyl jednoho hodu je $D(X) = \frac{(1-3.5)^2+(2-3.5)^2+\cdots+(6-3.5)^2}{6} = \frac{35}{12}$

- 60x opakováním náhodného jevu můžeme využít centrální limitní větu pro součet náhodných veličin.

$$E(X) = n \cdot \mu_X \qquad D(X) = n \cdot \sigma_X^2 \qquad X \sim N(n \cdot \mu_X, n \cdot \sigma_X^2)$$

  - $N\big(60 \cdot E(X), 60 \cdot D(X)\big) = N(210{,}175)$
- Pravděpodobnost se určí pomocí Matlabu:
  - pravd=normcdf(230,210,sqrt(175))-normcdf(190,210,sqrt(175))
  - pravd = 0.8694

6.3.3 Příklady na centrální limitní větu

- Př. 5
- 600x házíte hrací kostkou. Určete pravděpodobnost, že hodíte šestku 110x a více
- Výpočet pomocí binomického rozdělení
  - P=1-binocdf(109.5, 600, 1/6)
  - P=0.1492
- Výpočet pomocí CLV
  - Binomické rozdělení má E(X)=np a D(X)=np(1-p)
  - E(X)=600*1/6=100   D(X)=600*1/6*5/6=500/6
  - Toto jsou parametry normálního rozdělení
  - P=1-normcdf(109.5,100,sqrt(500/6))
  - P=0.1490

6.4 Rozdíl výběrových průměrů

- Mějme náhodný výběr $X_{11}, X_{12}, \dots, X_{1n_1}$ z rozdělení se střední hodnotou μ_1 a náhodný výběr $X_{21}, X_{22}, \dots, X_{2n_2}$ z rozdělení se střední hodnotou μ_2. Při splnění požadavků uvedených níže, má rozdíl výběrových průměrů následující vlastnosti:

$$E(\overline{X_1} - \overline{X_2}) = \mu_1 - \mu_2$$

$$D(\overline{X_1} - \overline{X_2}) = \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}$$

$$(\overline{X_1} - \overline{X_2}) \sim N(\mu_1 - \mu_2, \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2})$$

- Předpoklady:
  - Výběry jsou navzájem nezávislé, tj. hodnoty pozorování z 1. populace nejsou ovlivněny pozorováním z 2. populace, a naopak
  - Platí předpoklady CLV, že každý z výběrů pochází z normálního rozdělení, nebo rozsah každého z výběrů je větší než 30.
- 3. vzorec lze přetransformovat na:

$$\frac{(\overline{X_1} - \overline{X_2}) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2})}} \sim N(0,1)$$

6.5 Relativní četnost

- Náhodný jev A se vyskytuje s pravděpodobností π. Předpokládejme, že provádíme opakovaná nezávislá pozorování tohoto jevu. Potom výběrový průměr $\bar{X}$ vypočtený z prvních n pozorování označujeme relativní četností a značíme ji p.

$$p = \frac{\sum_{i=1}^{n} X_i}{n}$$

  – Náhodný jev A je popsán alternativním rozdělením, n-krát opakováním jevu A přecházíme k binomickému rozdělení.

- Vlastnosti relativní četnosti: $E(p) = \pi \quad D(p) = \frac{\pi \cdot (1-\pi)}{n}$
- Z centrální limitní věty (požadavky uvedeny níže) lze odvodit, že:

$$\frac{p - \pi}{\sqrt{\frac{\pi \cdot (1-\pi)}{n}}} \sim N(0,1)$$

- Předpoklady:
  – Dostatečná velikost výběru: $n > \frac{9}{p \cdot (1-p)}$

6.6 Rozdíl výběrových četností

- Náhodný jev A se vyskytuje s pravděpodobností π_1. Náhodný jev B se vyskytuje s pravděpodobností π_2. Předpokládáme, že provádíme opakovaná nezávislá pozorování těchto jevů.
- Potom rozdíl relativních četností obou výběrů má následující vlastnosti:

$$E(p_1 - p_2) = \pi_1 - \pi_2$$

$$D(p_1 - p_2) = \frac{\pi_1 \cdot (1 - \pi_1)}{n_1} + \frac{\pi_2 \cdot (1 - \pi_2)}{n_2}$$

$$(p_1 - p_2) \sim N\left(\pi_1 - \pi_2, \frac{\pi_1 \cdot (1 - \pi_1)}{n_1} + \frac{\pi_2 \cdot (1 - \pi_2)}{n_2}\right)$$

$$\frac{(p_1 - p_2) - (\pi_1 - \pi_2)}{\sqrt{\frac{\pi_1 \cdot (1 - \pi_1)}{n_1} + \frac{\pi_2 \cdot (1 - \pi_2)}{n_2}}} \sim N(0,1)$$

- Předpoklady:
  - Výběry z obou populací jsou dostatečně velké $n_1 > \frac{9}{p_1 \cdot (1 - p_1)}$ a $n_2 > \frac{9}{p_2 \cdot (1 - p_2)}$.
  - Výběry musí být nezávislé, tj. nejsou vzájemně ovlivněny.

6.7 Spojitá rozdělení užívaná ve statistice

- 6.7.1 χ^2 rozdělení
- 6.7.2 Studentovo rozdělení
- 6.7.3 Fisher Snedecorovo rozdělení

6.7.1 χ^2 rozdělení

- Mějme nezávislé náhodné veličiny $Z_1, Z_2, \dots, Z_n$ z nichž každá má normované normální rozdělení. Součet čtverců těchto náhodných veličin – X má rozdělení χ^2 s n stupni volnosti.

$$X = \sum_{i=1}^{n} Z_i{}^2 \rightarrow \chi_n{}^2$$

- χ^2 rozdělení se čte „chí kvadrát rozdělení“, někdy též Pearsonovým rozdělením
- Veličina s tímto rozdělením může nabývat pouze nezáporných hodnot.
- Rozdělení je nesymetrické
- Má parametr n – počet stupňů volnosti, který vychází z počtu naměřených dat. Často pro n naměřených dat, porovnáváme výsledek s $n-1$ stupni volnosti.

- Rozdělení se velmi často využívá ve statistice při testování veličin.

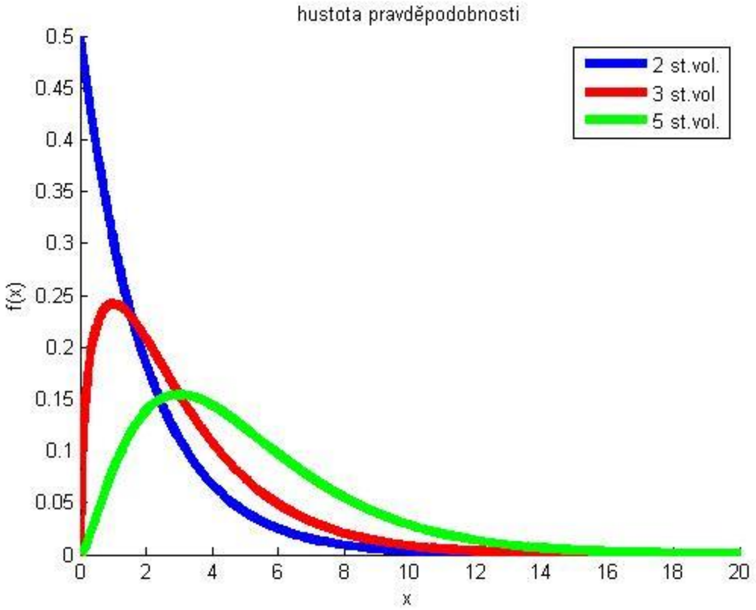
6.7.1 χ^2 rozdělení

- Základní vlastnosti
  - Hustota pravděpodobnosti je složitá.
  - Střední hodnota rozdělení χ^2 s n stupni volnosti je:
$$E(X) = n$$
  - Rozptyl rozdělení χ^2 s n stupni volnosti je:
$$D(X) = 2 \cdot n$$
  - Vlivem centrální limitní věty se se vzrůstajícím počtem stupňů volnosti rozdělení χ^2 s n stupni volnosti blíží normálnímu rozdělení s parametry: $N(\mu = n, \sigma^2 = 2 \cdot n)$

6.7.1 χ^2 rozdělení

- χ^2 rozdělení v matlabu
- n – počet stupňů volnosti

- Matlab
  - Distribuční funkce
    F=chi2cdf(x,n)
  - Hustota pravděpodobnosti
    f=chi2pdf(x,n)
  - Inverze distribuční funkce
    x=chi2inv(pravd, n)
  - Stanovení střední hodnoty a rozptylů
    [m,v]=chi2stat(n)
  - Náhodné číslo
    chi2rnd(n)



6.7.1 χ^2 rozdělení

- χ^2 rozdělení má uplatnění především při:
  - Odhad rozptylu základního souboru (předpoklad: data musejí být z normálního rozdělení).

  - Testování, zda rozptyl základního souboru je roven ${\sigma_0}^2$.
  - Testování nezávislosti proměnných (tzv. kontingenční tabulka)
  - Testování, zda náhodná veličina pochází z určitého rozdělení (zde data nemusí být z normálního rozdělení)

6.7.2 Studentovo rozdělení

- Uvažujme dvě nezávislé náhodné veličiny Z a V. Náhodná veličina Z má normované normální rozdělení, náhodná veličina V má χ^2 rozdělení s n stupni volnosti. Potom náhodná veličina T,

$$T = \frac{Z}{\sqrt{\frac{V}{n}}}$$

  má Studentovo rozdělení s n stupni volnosti.
- Studentovo rozdělení značíme t_n
- Rozdělení má jediný parametr n – počet stupňů volnosti.
- Se zvyšujícím se počtem stupňů volnosti (obvykle $n > 30$) se Studentovo rozdělení blíží normovanému normálnímu rozdělení.
- Rozdělení se velmi často využívá ve statistice při testování veličin.

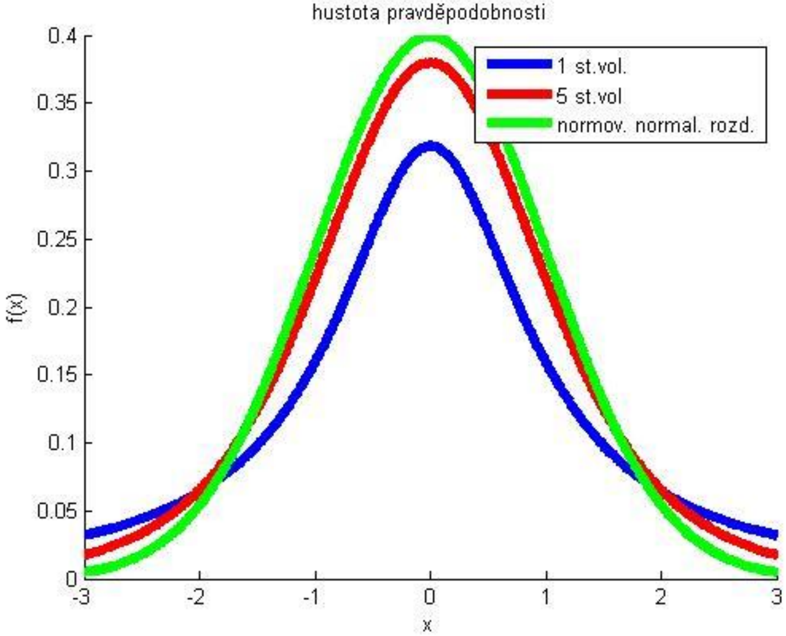
6.7.2 Studentovo rozdělení

- Základní vlastnosti
  - Hustota pravděpodobnosti je složitá
  - Střední hodnota Studentova rozdělení t s n stupni volnosti je:
$$E(X) = 0$$
  - Rozptyl Studentova rozdělení t s n stupni volnosti je:
$$D(X) = \frac{n}{n-2}, n > 2$$
  - Rozdělení je symetrické kolem 0.

6.7.2 Studentovo rozdělení

- t rozdělení v matlabu
- n – počet stupňů volnosti

- Distribuční funkce F=tcdf(x,n)
- Hustota pravděpodobnosti f=tpdf(x,n)
- Inverze distribuční funkce x=tinv(pravd, n)
- Stanovení střední hodnoty a rozptylů [m,v]=tstat(n)
- Náhodné číslo trnd(n)



6.7.2 Studentovo rozdělení

- Studentovo t rozdělení má uplatnění především při:
  - Odhadu střední hodnoty výběru, pokud je rozptyl základního souboru neznámý.
  - Testování hypotéz o střední hodnotě výběru, pokud je rozptyl výběru neznámý.
  - Testování hypotéz o shodě středních hodnot dvou nezávislých výběrů, pokud jsou rozptyly výběru neznámé.
  - Regresní analýza
- Předpoklad: Data musejí být z normálního rozdělení.

6.7.3 Fisher-Snedecorovo rozdělení

- Mějme dvě nezávislé náhodné veličiny V a W s rozdělením χ^2 . První má m stupňů volnosti, druhé má n stupňů volnosti (obecně odlišný). Pak náhodná veličina
$$F = \frac{\frac{V}{m}}{\frac{W}{n}}$$
má Fisherovo-Snedecorovo o m, n stupňů volnosti.
- Fisher-Snedecorovo rozdělení značíme $F_{m,n}$
- Rozdělení má dva parametry m a n.
- Rozdělení se velmi často využívá ve statistice při testování veličin.

6.7.3 Fisher-Snedecorovo rozdělení

- Základní vlastnosti
  - Střední hodnota Fisher Snedecorova rozdělení s m, n stupni volnosti je:

$$E(X) = \frac{n}{n-2}, n > 2$$

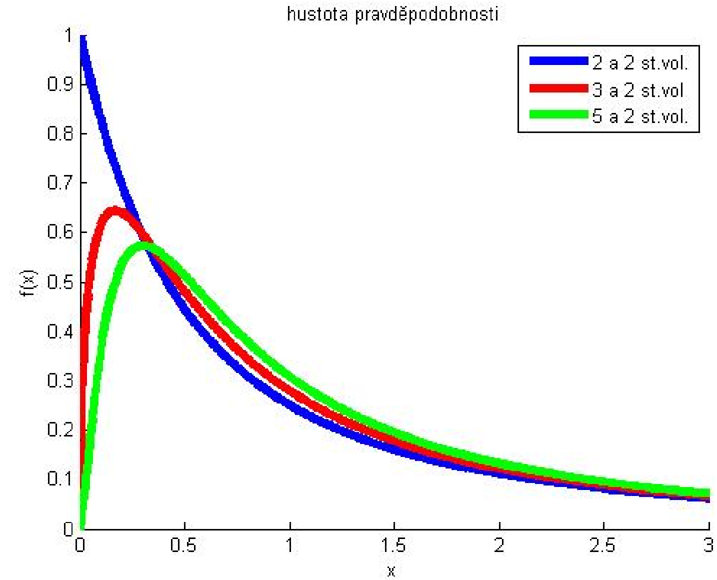
  - Rozptyl Fisher Snedecorova s m, n stupni volnosti je:

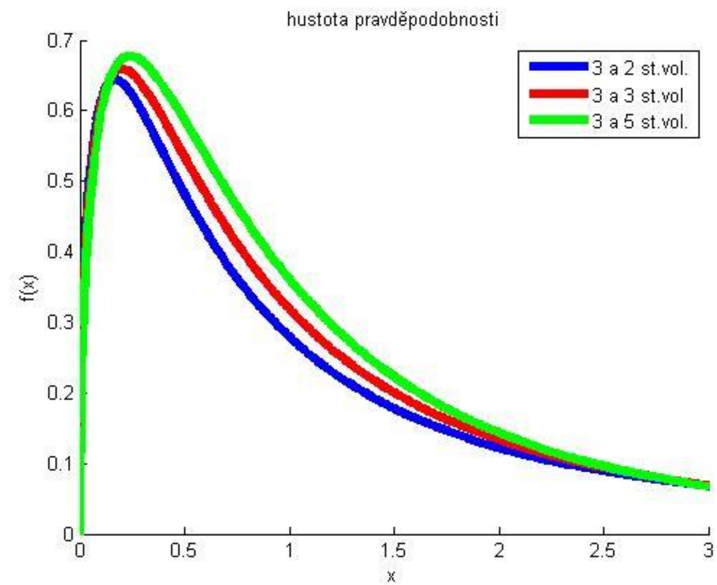
$$D(X) = \frac{2n^2 \cdot \left(1 + \frac{n-2}{m}\right)}{(n-2)^2 \cdot (n-4)}, n > 4$$

6.7.3 Fisher-Snedecorovo rozdělení

- F rozdělení v matlabu
- m,n – počet stupňů volnosti

- Výpočet v matlabu:
  - Distribuční funkce
    F=fcdf(x,m,n)
  - Hustota pravděpodobnosti
    f=fpdf(x,m,n)
  - Inverze distribuční funkce
    x=finv(pravd, m,n)
  - Stanovení střední hodnoty a rozptylů
    [m,v]=fstat(m,n)
  - Náhodné číslo
    frnd(m,n)





6.7.3 Fisher Snedecorovo rozdělení

- Fisher Snedecorovo rozdělení má uplatnění především při:
  - Testování shody rozptylů dvou základních souborů
  - Testování shody středních hodnot více než dvou základních souborů (analýza rozptylu)
  - Regresní analýza

- Předpoklad: Data musejí být z normálního rozdělení.

6.8 Základní příkazy v matlabu/octave

	χ^2 rozdělení	Studentovo rozdělení	Fisher-Snedecorovo rozdělení
Distribuční funkce	chi2cdf	tcdf	fcdf
Hustota pravděpodobnosti	chi2pdf	tpdf	fpdf
Inverzní funkce	chi2inv	tinv	finv
Střední hodnota a rozptyl	chi2stat	tstat	fstat
Náhodné číslo	chi2rnd	trnd	frnd