Cvičení ze statistiky - 6

Filip Děchtěrenko

Minule bylo..

- Probrali jsme základní charakteristiky pravděpodobnostních modelů a diskrétní modely
- Tyhle termíny by měly být známé:
  - Distribuční funkce
  - Střední hodnota
  - Rozptyl
  - Z rozdělení
  - Hypergeometrické rozdělení
  - Alternativní rozdělení
  - Binomické rozdělení
  - Kombinační číslo

Příklady na rozdělení

1. Lovec má 5 patron, střílí dokud netrefí (nebo nedojdou) a pst zásahu je 0.4. Popište rozdělení, střední hodnotu a rozptyl
2. Pravděpodobnost, že se narodí kluk, je 0,515. Kolik potřebujeme dětí, aby pst, že je v nich aspoň jeden kluk, je větší než 99%?
3. V urně je 5 černých a 3 zelené koule. Náhodně vyberu 2 koule, jaká je pravděpodobnost, že právě jedna bude zelená?
4. To samé jako 3, ale kouli tam po vytažení vrátím?
5. To samé jako 3, ale v urně je 5000 černých a 3000 zelených

Spojité rozdělení

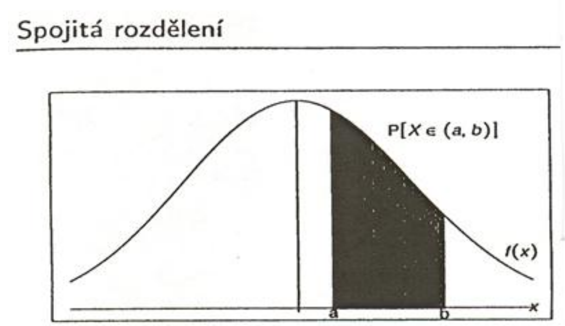
- Zatím jsme měli pouze diskrétní proměnné (tedy v oboru hodnot jsou mezery)
- Zajímá nás případ, kdy je oborem hodnot interval
- Nemůžeme mít vyjádřenou pst pro jednotlivé hodnoty z oboru hodnot (je jich nekonečno)-> zavedeme *hustotu* – spojitá funkce pro jednotlivé hodnoty z intervalu *f(x)*

Spojité rozdělení - F(x)

- Distribuční funkce je taky spojitá!
- Vztah mezi hustotou a distribuční funkcí

$$F(x) = P(X \leq x) = \int_{-\infty}^{x} f(t)dt$$

- Pro rozsahy používáme P(a $< X \leq b) =$ F(b)-F(a)
- Distribuční funkci budeme používat při výpočtech



Spojité rozdělení – EX a Var X

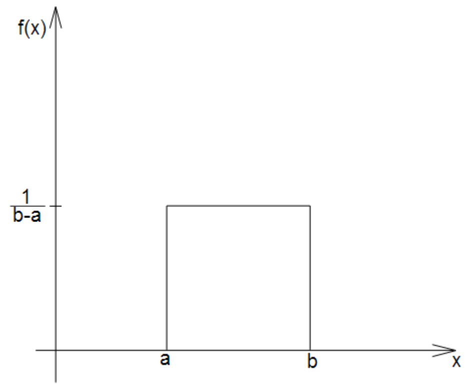
- Střední hodnota i rozptyl mají stejný význam, jen se počítají jinak
- $EX = \int_{-\infty}^{\infty} xf(x)dx$
- Var $X = \int_{-\infty}^{\infty} (x - EX)^2 f(x)dx$
- Naštěstí to už za nás pro vybrané rozdělení spočítal☺

Rovnoměrné rozdělení

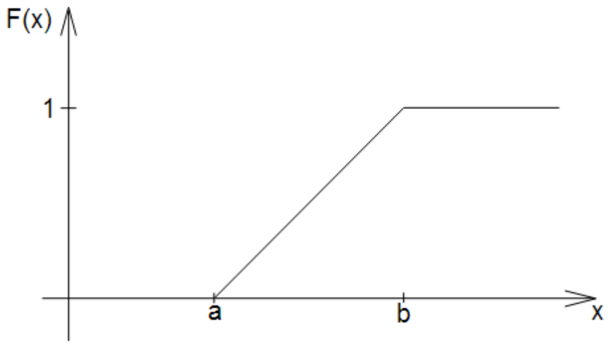
- Každé číslo z intervalu a,b je stejně pravděpodobné
- X=(a,b)
- Parametry
  - a: dolní rozsah intervalu
  - b: horní rozsah intervalu
- f(x)=$\frac{1}{b-a}$ pro a<x<b (hustota)
- F(x)=$\frac{x-a}{b-a}$ pro a<x<b (distribuční funkce)
- EX=$\frac{b+a}{2}$
  Var X=$\frac{(b-a)^2}{12}$
- Značí se X∼Ro(a,b)

Rovnoměrné rozdělení - grafy

- Hustota
  - Každá z hodnot je stejně pravděpodobná

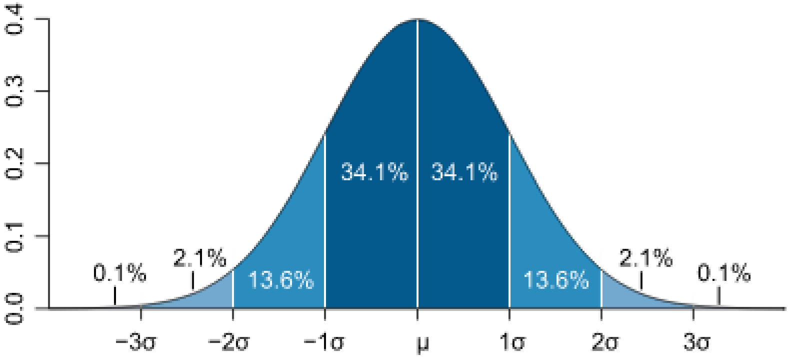


- Distribuční funkce



Normální (Gaussovo) rozdělení

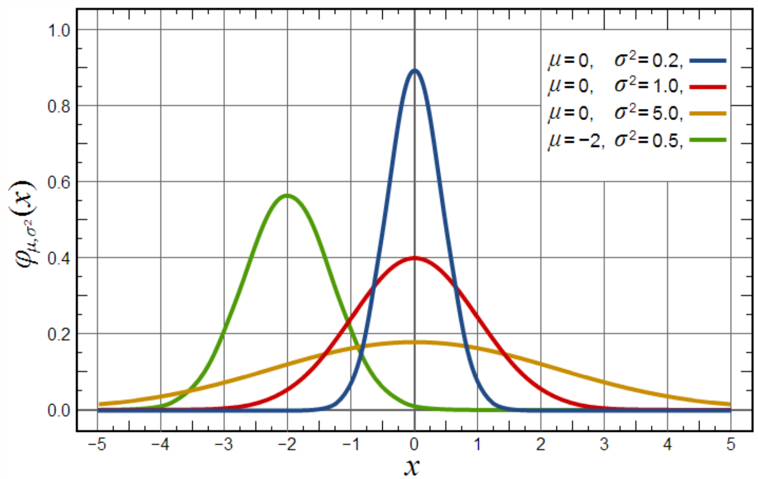
- Velmi důležité rozdělení – působí-li hodně faktorů zároveň, má veličina náhodné rozdělení
- X=$(-\infty, \infty)$
- Parametry:
  - μ: střední hodnota
  - σ^2: rozptyl



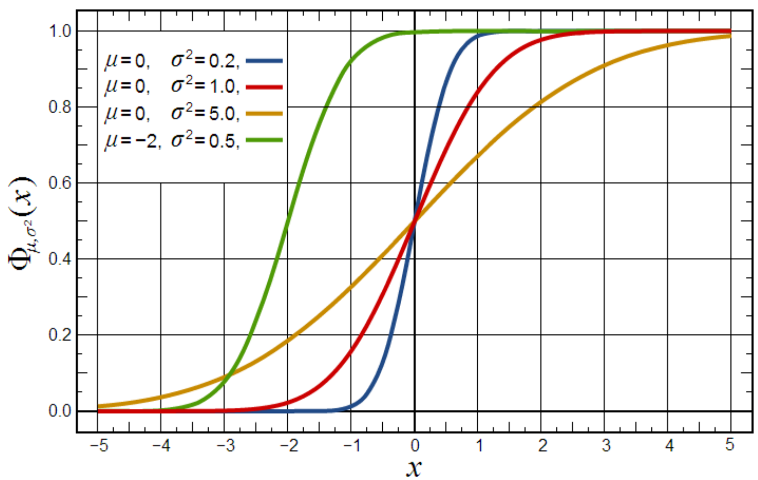
- f(x)=$\frac{1}{\sigma\sqrt{2\pi}} e^{\frac{(x-\mu)^2}{2\sigma^2}}$
- F(x) nemůže být vyjádřena funkcí (jedině numericky)
- EX=μ
  Var X=σ
- Značí se X~N(μ, σ^2)

Normální rozdělení -

- Hustota



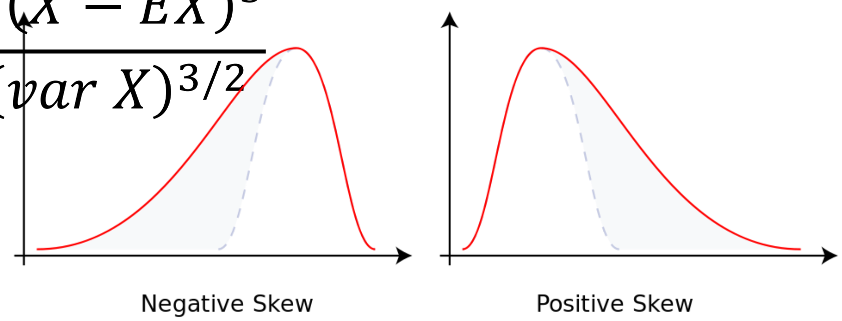
- Distribuční funkce
  - Analyticky sice nelze vyjádřit, ale graf můžeme zobrazit



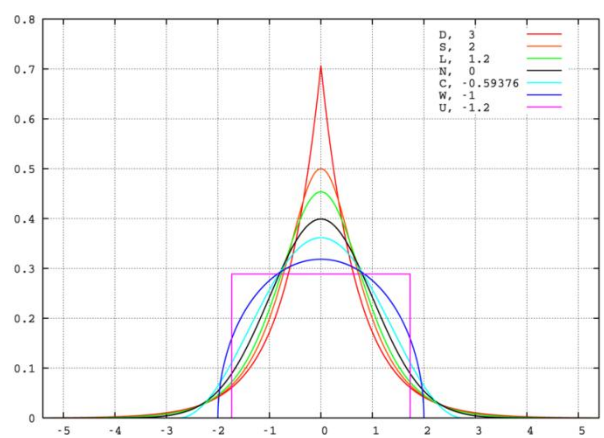
Šikmost a špičatost

- Pro popis rozdělení se používá kromě střední hodnoty a průměru ještě šikmost a špičatost
- Koeficient šikmosti vypovídá o symetrii rozdělení

$$\gamma_1 = \frac{E(X - EX)^3}{(var\ X)^{3/2}}$$



- Koeficient špičatosti vypovídá o velikosti „ramen“
  $\gamma_1 = \frac{E(X-EX)^4}{(var\ X)^2}$-3
- Existují i výběrové varianty těchto koeficientů



N(0,1) rozdělení

- Náhodnou veličinu značíme Z
- Jedná se o normální rozdělení s parametry $\mu=0$, $\sigma^2=1$
- Máme tabulku pro hodnoty distribuční funkce tohoto rozdělení
- Převádí se na něj všechna normální rozdělení (a hodnoty se poté najdou v tabulkách pro N(0,1) rozdělení)
- Značíme $P(Z \leq z)=\phi(z)$

Tabulky N(0,1) – směr z->p

- Počítáme p-hodnoty pro zadané z-hodnoty
- Máme jen pro z-hodnoty 0-4
  - Platí P(Z>4)=1 (proto nepotřebujeme vyšší rozsah)
  - Platí $\phi(z)=1-\phi(-z)$ (proto nepotřebujeme tabulky pro záporné hodnoty)
    - Př: z=-0.5
    - $\phi(-0.5)=1-\phi(0.5)=1-0.6915=0.3085$
- Ukázka tabulky:

z	$\phi(z)$
0.0	0.500
0.02	0.508
0.04	0.516
0.06	0.5239

Tabulky N(0,1) – směr p->z

- Počítáme p-hodnoty pro zadané z-hodnoty
- Pro p-hodnotu mezi 0.5 a 1 najdeme v tabulce (ale díváme se do druhého sloupce)
- Pro p-hodnotu mezi 0 a 0.5: najdeme v tabulce hodnotu pro „1-p“ a k výsledné z-hodnotě přidáme mínus
  - Př: $0.4=P(Z \leq z)$
    V tabulce najdeme z-hodnotu pro $p=0.6$
    -> 0.25 a přidáme mínus
    $z=-0.25$

Mccallova transformace

- Mccalova transformace je jedna z metod, které se používá k transformaci rozdělení s vysokou šikmostí na normální rozdělení
- Pokud transformujeme z binomického rozdělení, používáme i korekci na spojitost
- Hodím 5 krát mincí, můj hrubý skór bude součet orlů. Toto budu opakovat 1000 krát.

HS	četnost	Rel. četnost	Kum. rel. četnost	Korekce na spoj.	Z-skór	T-skór
0	150	0,15	0,15	0,075	-1,44	36
1	400	0,4	0,55	0,35	-0,385	46
2	200	0,2	0,75	0,65	0,3853	54
3	120	0,12	0,87	0,81	0,8779	59
4	80	0,08	0,95	0,91	1,3408	63
5	50	0,05	1	0,975	1,96	70

Mccallova transformace-výpočet

- Vycházíme ze sloupců HS a četnost
- Označíme si:
  - Četnost – Č
  - Relativní četnost – RČ
  - Kumulovaná relativní četnost – KRČ
  - Korekce na spojitost – KS
  - Z-skór- Z
  - T-skór-T
- Pokud budeme mluvit o i-tém řádku, označíme příslušné políčko dolním indexem i
- Platí
  - $R\check{C}_i = \frac{\check{C}_i}{n}$
  - $KR\check{C}_i = KR\check{C}_{i-1} + R\check{C}_i$

    $$KR\check{C}_1 = R\check{C}_1$$
  - $KS_i = KR\check{C}_{i-1} + 0.5R\check{C}_i$

    $$KS_1 = 0.5R\check{C}_1$$
  - Z – spočítáme přes inverzní hledání v tabulce
  - $T_i = Z_i \cdot 10 + 50$ a zaokrouhlíme
  - Pokud je KS_i=1, nepočítáme Z_i a za T_i dosadíme maximální hodnotu (tedy 100)
  - Pokud je KS_i=0, nepočítáme Z_i a za T_i dosadíme minimální hodnotu (tedy 0)

Důležitá vlastnost N(μ, σ^2)

- Má-li náhodná veličina X$\sim$N(μ, σ^2) a Y=aX+b, potom Y$\sim$N($\mathrm{a} + \mathrm{b}$μ, $b^2\sigma^2$), kde b≠0
  - Máme tedy předpis, jak na sebe rozdělení převádět
  - Konkrétně pokud X$\sim$N(μ, σ^2), potom Z=$\frac{X-\mu}{\sigma}$ má rozdělení N(0,1)
  - Vyskytne-li se nám někde výraz P($\frac{X-\mu}{\sigma}$ ≤z), přepíšeme si to na P($\frac{X-\mu}{\sigma}$ ≤z)=P(Z≤z)=ϕ(z) a najdeme v tabulkách

Příklad (výpočet psti)

- Počet bodů v testu má normální rozdělení N(50,144)
  - Jaká je pravděpodobnost, že náhodně vybraný člověk má více než 74 bodů?
- Řešení:
  - X$\sim$N$(50, 12^2)$ značí počet bodů v testu
    P(X>70) je pst, že náhodně vybraný člověk má více než 70 bodů
  - P(X>70)=1-P(X≤70)=1-P($\frac{X-50}{12} \leq \frac{74-50}{12}$)=1-P(Z ≤ 2)=1-ϕ(2)=1-0.9772=0.023

Příklad

- Počet bodů v testu má normální rozdělení N(50,144)
  - Jaká je pravděpodobnost, že náhodně vybraný člověk má mezi 30 a 45 body?
- Řešení:
  - $X \sim N(50,\ 12^2)$ značí počet bodů v testu
    P(30<X<45) je pst, že náhodně vybraný člověk má mezi 30 a 45 body
  - $P(30 \le X \le 45) = P(\frac{30-50}{12} \le \frac{X-50}{12} \le \frac{45-50}{12}) =$
    $= P(-1.66 \le Z \le -0.42) = P(-0.42) - P(-1.66) =$
    $= (1-\phi(0.42)) - (1-\phi(1.66)) = \phi(1.66) - \phi(0.42) = 0.9515 - 0.6554 =$
    $= 0.30$

Příklad (výpočet z-hodnoty)

- Výška člověka pochází z normálního rozdělení N(160,324). Jak je vysoký člověk, který je větší než 70% populace?
- Řešení:
  - X$\sim$N$(160,\ 18^2)$ značí počet bodů v testu
    0.7=P(X≤v)=P($\frac{X-160}{18}$≤$\frac{V-160}{18}$)= P(Z ≤$\frac{V-160}{18}$)
  - 0.7= P(Z ≤$\frac{V-160}{18}$) (a hledáme v tabulkách podle postupu p->z)
  - $\frac{V-160}{18}$=0.52
    V=0.52*18+160
    V=169.36

Problémy při inverzním výpočtu

- Pozor, pokud budeme mít rovnici v jiném tvaru, než $p=P(X\leq v)$, nemůžeme hledat v tabulkách.
- Př: $0.4=1-P(X\leq v)$ – nemůžeme hledat (nevíme, pro jakou hodnotu). Upravíme si rovnici do tvaru $0.6=P(X\leq v)$, a až poté budeme hledat v tabulkách

Přehled fint

- $P(Z>z)=1-P(Z\leq z)=1-\phi(z)$
  $P(Z\geq z)=1-P(Z<z)=1-\phi(z)$
  Aspoň na jedné straně musí být rovnost
- $P(X<a \text{ nebo } X>b)=1-P(a\leq X\leq b)$
  dá se to představit na číselné ose
- $\phi(-a)=1-\phi(a)$
  toto použijeme vždy, když budeme chtít najít zápornou hodnotu v tabulce
- $P(a<Z\leq b)=P(Z\leq b)-P(Z\leq a)=\phi(b)-\phi(a)$
  obvykle bývá a záporné, potom
  $P(-a<Z\leq b)=\phi(b)-\phi(-a)= \phi(b)-(1-\phi(a))=\phi(b)+\phi(a)-1$
- $P(-a<Z\leq a)=2\phi(a)-1$