Cvičení ze statistiky - 5

Filip Děchtěrenko

Minule bylo..

- Začali jsme pravděpodobnost
  - Klasická a statistická definice pravděpodobnosti
  - Náhodný jev
  - Doplněk, průnik, sjednocení
  - Podmíněná pravděpodobnost
  - Bayesova věta

Náhodná proměnná

- Popisování náhodných jevů pomocí písmen A,B,C se špatně kvantifikuje (mám-li minci, tak můžu popsat jev, že padla panna písmenem P, orel písmenem O)
- Zavedeme náhodnou proměnnou, která bude značit výsledek náhodného pokusu
- Značíme $P(X=k)=p$
  kde k je hodnota náhodné proměnné a p je pst
- !!!!!Pozor, X není neznámá!!!!!

Vlastnosti náhodné proměnné

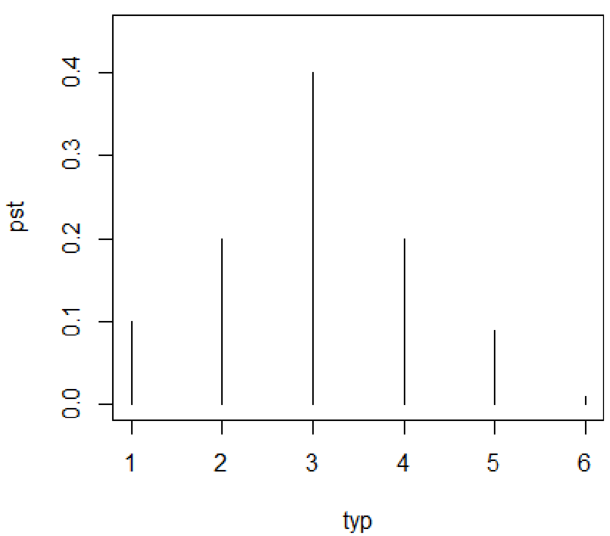
- Náhodná proměnná nabývá různých hodnot, všem možným hodnotám říkáme obor hodnot
- Máme-li mezi hodnotami mezery (tj. existují-li hodnoty, kterých náhodná proměnná nemůže nabývat), mluvíme o *diskrétní náhodné proměnné*
- Mohou-li hodnoty nabývat libovolné hodnoty z intervalu, mluvíme o *spojité náhodné proměnné*

Diskrétní náhodná proměnná

- Obor hodnot je konečný (spočetný), tedy ho můžeme zapsat jako množinu $\{x_1, x_2, \dots, x_n\}$
- Pro každé číslo z oboru hodnot máme definovanou pravděpodobnost jeho výskytu
$$P(X = x_1) = p_1, P(X = x_2) = p_2, \dots, P(X = x_n) = p_n$$
- Př: V továrně vyrábějí 6 typů výrobků. Pravděpodobnost, že další vyrobený výrobek bude typu 1-6 je:

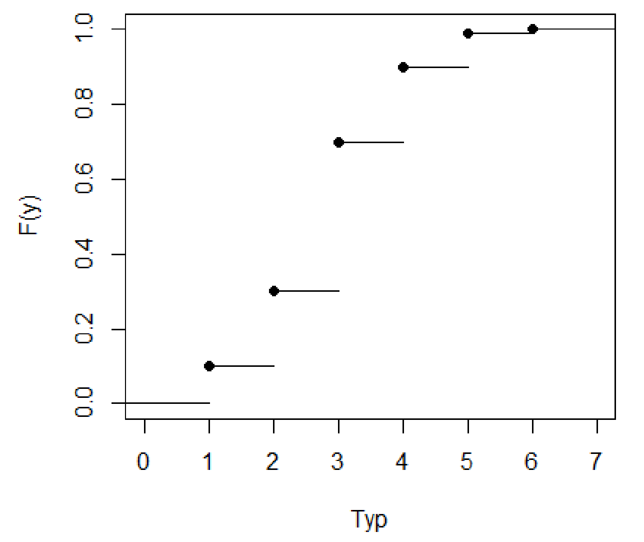
Typ	1	2	3	4	5	6
Pst	0.1	0.2	0.4	0.2	0.09	0.01

- Obor hodnot je tedy {1,2,3,4,5,6} a pravděpodobnosti jsou
  - P(X=1)=0.1
  - P(X=2)=0.2
  - P(X=3)=0.4
  - P(X=4)=0.2
  - P(X=5)=0.09
  - P(X=6)=0.01
- Musí platit, že součet dílčích pstí je 1



Distribuční funkce

- Jeden z prostředků popisu náhodné veličiny
- $F(x) = P(X \leq x)$
- Označuje, jaká je pravděpodobnost, že náhodná veličina je menší nebo rovna než zadané x
- Vlastnosti distribuční funkce:
  - Hodnoty mezi 0 a 1 ($0 \leq F(x) \leq 1$)
  - Funkce je neklesající
  - Je zprava spojitá (a u spojité je celá spojitá)
- Zobrazuje se grafem distribuční funkce
- Př: Nakreslete graf F pro hod kostkou



Sdružená pravděpodobnost

- Někdy se nám vyskytují proměnné spolu
- Př: s nějakou pravděpodobností mě může bolet zub, bolest zubu se vyskytuje s vyšší pstí, pokud mám kaz

	Zub bolí (X=1)	Zub nebolí (X=2)
Mám kaz(Y=1)	0.4	0.1
Nemám kaz(Y=2)	0.2	0.3

- Např.: P(X=1,Y=2)=0.2 značí pst, s jakou mě bolí zub, ale přitom nemám kaz
- Můžu se ptát, s jakou pstí mě budou bolet zuby nehledě na to, zda mám kaz - P(X=1)?
- Sečtu všechny dílčí psti u X=1, tedy:
$$P(X=1)=P(X=1,Y=1)+P(X=1,Y=2)=0.4+0.2=0.6$$
- Obecně $P(X=i)=\sum_j P(X=i,Y=j)$
P(X=i) říkáme *marginální pravděpodobnost*
- Náhodné proměnné jsou nezávislé, pokud platí $P(X=i,Y=j)=P(X=i)\cdot P(X=j)$

Sdružená pravděpodobnost příklad

- Mějme následující pravděpodobnosti:

	Y=1	Y=2	Y=3
X=1	1/8	1/8	1/32
X=2	1/32	1/8	0
X=3	0	1/16	1/2

- Kolik je $P(X=2,Y=1)$, $P(X=1)$,$P(X=3,Y>2)$? Jsou nezávislé?
- $P(X=2,Y=3)=1/32$
  $P(X=1)=3/32$
  $P(X=3,Y<3)=1/16$
  Nejsou nezávislé, protože $P(X=3,Y=2)\neq P(X=3)\cdot P(Y=2)$

Střední hodnota

- Typicky nás zajímá *střední hodnota* náhodné proměnné (také se někdy nazývá *očekávaná hodnota* nebo *populační průměr*)
- Jedná o číslo, které průměrně dostaneme při jednom pokusu
- $E(X) = \mu = \sum_i P(X = x_i) \cdot x_i$
- Vypovídá nám dobře o tom, co od náhodné proměnné můžeme čekat

Střední hodnota v příkladu

- $E(X) = \sum_i P(X = x_i) \cdot x_i$
- Továrna:

Typ	1	2	3	4	5	6
Pst	0.1	0.2	0.4	0.2	0.09	0.01

  - Jaká je očekávaná hodnota u výrobku v továrně?
  - $E(X) = 0.1 \cdot 1 + 0.2 \cdot 2 + 0.4 \cdot 3 + 0.2 \cdot 4 + 0.09 \cdot 5 + 0.01 \cdot$ 6=3
- Hod mincí:
  - Jaká je střední hodnota u mince (označím-li pannu jako 0 a orla jako 1)?
  - $E(X) = 0.5 \cdot 0 + 0.5 \cdot 1$
  - Na označení nezáleží, mohlo by to být klidně i panna 100 a orel 150. Střední hodnota se vztahuje zadaným číslům a ty mohou být kterákoli

Rozptyl

- Označuje míru rozptýlenosti okolo středu, někdy se taky nazývá populační rozptyl
- $var\ X = \sum (x_i - \mu)^2 P(X = x_i)$
- Alternativně: $var\ X = E(X - EX)^2 = EX^2 - (EX)^2$
- EX je obyčejná střední hodnota
  $EX^2 = \sum_i P(X = x_i) \cdot {x_i}^2$

Rozptyl v příkladu

- $var\ X = \sigma^2 = EX^2 - (EX)^2$
- Továrna:

Typ	1	2	3	4	5	6
Typ^2	1	4	9	16	25	36
Pst	0.1	0.2	0.4	0.2	0.09	0.01

  – Jaký je očekávaný populační rozptyl?

  – $var\ X = (0.1 \cdot 1 + 0.2 \cdot 4 + 0.4 \cdot 9 + 0.2 \cdot 16 + 0.09 \cdot 25 + 0.01 \cdot 36) - 3^2 = 10 - 9 = 1$
- Hod mincí:
  – Jaký je očekávaný rozptyl mince?

  – $var\ X = 0.5 \cdot 0 + 0.5 \cdot 1 - 0.5^2 = 0.5 - 0.25 = 0.25$

Funkce náhodné proměnné

- Samotná náhodná proměnná je pro formalizování světa málo, chtěli bychom s ní dělat základní aritmetické operace (něčím vydělit, přičíst konstantu,…)
- Př: na kostce padne 1 až 6, pokud chceme, aby nám vracela čísla -5 až 0, obrátím znaménko a přičtu 1, tedy $Y = 1 - X$
- Př: Hodím dvěma kostkami, chci jejich součet, tedy $Y = X_1 + X_2$
- Mění se pouze hodnoty y_i (původně x_i), pravděpodobnost se nemění! ($P(Y = y_i) = P(X = x_i)$)
- Vyvstává otázka, jak po těchto operacích vypadá E(Y) a var Y

Vlastnosti EY a var Y

- $E(c) = c$, kde c je konstanta
- $E(b \cdot X) = b \cdot E(X)$
- $E(a + X) = a + E(X)$
- $E(X_1 + X_2) = E(X_1) + E(X_2)$
- $var\ (c) = 0$, kde c je konstanta
- $var\ (bX) = b^2 \cdot var(X)$
- $var\ (a + X) = var(X)$
- Jsou-li nezávislé, pak $var(X + Y) = varX + varY$
  Nejsou-li nezávislé $var(X + Y) = varX + varY - cov(X, Y)$
  kde $cov(X, Y) = \sum_i \sum_j (x_i - EX)(y_i - EY)P(X = x_i . Y = y_i)$
- => máme-li spočítané EX a var X, nemusíme počítat EY a var Y z definice, ale stačí použít výše uvedené vztahy

Odvození var X

- Pomocí vztahů pro EX můžeme odvodit vzorec pro var X
- $var(X) = E(X - EX)^2 =$
  $E(X^2 - 2X(EX) + (EX)^2) =$
  $E(X^2) - 2(EX)(EX) + (EX)^2 =$
  $E(X^2) - (EX)^2$
- (spíš jen pro ilustraci)

Z rozdělení

- Velmi užitečná transformace je na Z rozdělení (měli jsme Z hodnotu
- $Z = \frac{X-\mu}{\sigma}$
- Potom platí, že E(Z)=0 a var(Z)=1
- Budeme to hodně využívat dál

Pravděpodobnostní modely

- Chceme popsat, jak vznikají data ve světě
- Známe-li vlastnosti modelu, můžeme předpovídat, jaká data dostaneme
- Př: mějme pravděpodobnostní model, ve kterém platí, že $P(X=1)=1$ (Tedy vždy padne 1), pak jsme schopní říct, kolik jedniček budeme mít po deseti pokusech

Hypergeometrický model

- Trochu jiný model, než ostatní
- Pro případy **bez vracení**
- Máme nějakou množinu prvků, kterou můžeme rozdělit na dvě skupiny (A a B). Vybereme z ní n prvků a zajímá nás, jaká je pst, že ve výběru je právě k prvků z A
- Parametry:
  - N: celkový počet prvků
  - N_A, N_B: počet prvků ve skupině A a B
  - n: výběr ze souboru
  - k: počet prvků ze skupiny A

Hypergeometrické rozdělní

- $P(X=k) = \frac{\binom{N_A}{k} \cdot \binom{N_B}{n-k}}{\binom{N}{n}}$
- Kde $\binom{N}{k}$ je kombinační číslo
$$\binom{N}{k} = \frac{N!}{k!\,(N-k)!}$$
- $n! = n \cdot (n-1) \cdot (n-2) \cdot \dots \cdot 2 \cdot 1$
tedy 4!=4·3·2·1=12
- $\binom{N}{k}$ značí počet všech k-tic z N prvků
- Nezáleží na pořadí (dvojice (A,B) je totéž jako (B,A))
- Jde vlastně o počet pozitivních případů, ku všem případům

Příklad

- Ve třídě je 20 dětí, z toho 8 kluků a 12 holek.
  - Náhodně vyberu 4 děti, jaká je pravděpodobnost, že právě dvě jsou holky?

$$P(X=2)=\frac{\binom{12}{2}\cdot\binom{8}{4-2}}{\binom{20}{4}}$$

  - Náhodně vyberu 4 děti, jaká je pravděpodobnost, že aspoň dvě jsou holky?

$$P(X\geq 2)=P(X=2)+P(X=3)+P(X=4)=\frac{\binom{12}{2}\cdot\binom{8}{4-2}}{\binom{20}{4}}+\ldots$$

Vlastnosti kombinačních čísel

- Pro kombinační číslo platí
  - $1 = \binom{n}{0} = \binom{n}{n}$
  - $\binom{n}{1}$=1
  - $\binom{n}{k}$=$\binom{n}{n-k}$

Alternativní rozdělení

- Základní diskrétní rozdělení
- X={0,1}
  tedy mám dvě možnosti, co mohou nastat
- Parametry:
  - π: pst, že nastane jev 0 (tedy 1- π, že nastane jev 1)
- EX=π
  Var X= π(1-π)
- Značí se X~Alt(π)
- Př: hod mincí (π=0.5)

Binomické rozdělení

- Součet n alternativních rozdělení
- $Y = X_1 + X_2 + \cdots X_n$
- Y={0,1,…,n}
- Parametry:
  - π: pst, že u každého z n jevů padne 0 (tedy 1- π, že padne 1)
  - n: počet alternativních rozdělení
- $P(X = k) = \binom{n}{k} \pi^k (1 - \pi)^{n-k}$
- EX=nπ
  Var X= nπ(1- π)
- Značí se X~Bi(n,π)
- Př: hod 10 mincemi, jaká je pst, že padne právě 5x panna

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