

Cvičení ze statistiky - 4

Filip Děchtěrenko

Minule bylo..

- Dokončili jsme deskriptivní statistiku
- Tyhle termíny by měly být známé:
 - Korelace
 - Regrese
 - Garbage in, Garbage out
 - Vícenásobná regrese

Pravděpodobnost

- U popisné statistiky jsme pracovali s „vyrobenými daty“
- Studium pravděpodobnosti nám řekne něco o tom, jak jsou data vyrobeny
- Každá data pochází z nějakého pravděpodobnostního modelu
- [Quincux](#) – data dole je náš vzorek, pyramida je náš model

Základní pojmy

- Elementární jev – nejmenší možná událost, co může nastat (na kostce nemůže padnout nic menšího, než jedno číslo), značíme ω
- Náhodný jev – je komplexní událost, která může nastat (kombinace elementárních jevů), značíme A, B, \dots
- Universum – Množina všech elementárních jevů, značíme U
- Pravděpodobnost náhodného jevu značíme **$P(A)$**

Náhodný jev

- Jde obecně o množinu elementárních jevů:
- Př:
 - U: hod kostkou, tedy
 - ω_1 – padne 1
 - ω_2 – padne 2
 - ...
 - ω_6 – padne 6
 - A: padne sudé číslo
 $A = \{\omega_2, \omega_4, \omega_6\}$
- S množinami můžeme provádět průnik, sjednocení, doplněk (za chvíli)

Dva pohledy na pravděpodobnost

1. Klasická definice pravděpodobnosti:

Jde o poměr žadáných výsledků ku všem možným výsledkům

$$P(A) = \frac{\#A}{\#U}$$

Počet všech **žadáných** výsledků

Počet všech **možných** výsledků

– Příklad: na kostce mohou padnout čísla 1, ..., 6 (6 možností), pravděpodobnost, že padne 4 (1 možnost), je:

$$P(\text{padne } 4) = \frac{1}{6}$$

Dva pohledy na pravděpodobnost (2)

2. Statistická definice pravděpodobnosti:
realizují náhodný pokus n -krát.

Pravděpodobnost jevu A je podíl žádaných výsledků ku počtu všech výsledků, za předpokladu, že n je hodně velké

$$P(A) = \lim_{n \rightarrow \infty} \frac{\#A}{n}$$

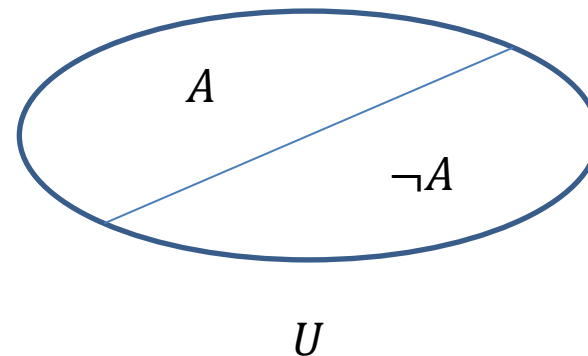
- Pokud jde n do nekonečna, vychází tato pravděpodobnost stejně jako klasická
- Příklad: Házení kostkou (na tabuli)

Podmíněná pravděpodobnost

- Závisí-li výskyt jednoho jevu na druhém, mluvíme o *závislosti*
- Pravděpodobnost toho, zda si vezmu deštník se bude měnit podle toho, zda včera pršelo nebo ne
- Pravděpodobnost výskytu dědičné choroby závisí na tom, zda ji měli moji rodiče
- Značíme $P(A | B)$
- Ve světě je všechno závislé na všem (ale můžeme to vzhledem k velikosti vlivu zanedbat)

Základní vlastnosti a značení

- Pravděpodobnost nabývá hodnot mezi 0 a 1 (odpovídá to 0% a 100%)
- Pokud platí, že $P(A)=1$, tak A nazýváme *jevem jistým*
 - A : půjdu nebo nepůjdu do školy
 $P(A)=1$
- Pokud platí, že $P(A)=0$, tak A nazýváme *jevem nemožným*
 - A : Napíšu písemku na 1 a zároveň ji nenapíši na 1
- Opak jevu A značíme \bar{A} , $\sim A$ nebo $\neg A$
 - Platí: $P(\neg A)=1-P(A)$
 - Jde o doplněk množiny A
 - A : bude pršet 24.12.2012
 $P(A)=0.2$
 $\neg A$: nebude pršet 24.12.2012
 $P(\neg A)=0.8$



Operace s náhodnými jevy - sjednocení

- Sjednocení (nebo)

- Stačí, aby nastal jeden z jevů

- $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

- Jsou-li nezávislé:

$$P(A \cup B) = P(A) + P(B)$$

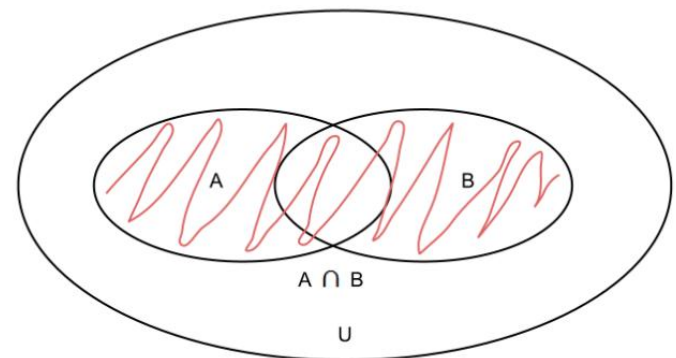
- A: padne šestka - $P(A)=1/6$

- B: padne pětka - $P(B)=1/6$

pětka a šestka nemůže padnou zároveň->jsou nezávislé

$$P(A \cup B) = 1/6 + 1/6 = \frac{1}{3}$$

- Co když nejsou nezávislé?



Operace s náhodnými jevy - průnik

- Průnik

- Oba jevy mají nastat zároveň

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

- $P(A|B)$ nazýváme *podmíněná pravděpodobnost*

- Jsou-li nezávislé

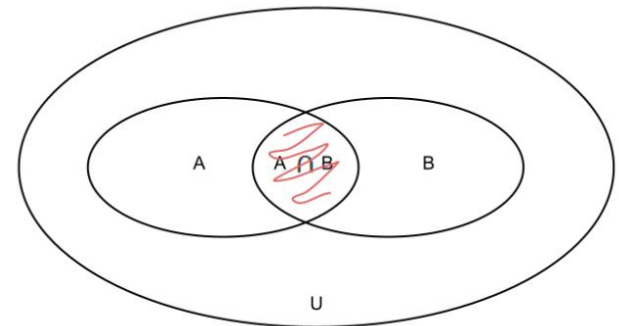
$$P(A|B) = P(A)$$

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

- A: bude pršet - $P(A)=0.8$

- B: vezmu si deštník – $P(B)=0.7$

- $P(A \cap B) = 0.8 \cdot 0.7 = 0.56$



Odbočka – alespoň, nejvýše,...

- Některé slovní konstrukce mohou napovídat o vztazích, které platí, či které nás budou zajímat

Slovní konstrukce	Matematický symbol
Mějme <i>alespoň</i> k kusů	$X \geq k$
Mějme <i>více než</i> k kusů	$X > k$
Mějme nejvýše k kusů	$X \leq k$
Mějme <i>méně než</i> k kusů	$X < k$
Mějme právě k kusů	$X = k$

- Příklady na procvičení jsou k nalezení negací výroků jsou [zde](#)

Příklady

1. Jaká je pravděpodobnost, že na kostce padne sudé číslo?
2. Jaká je pravděpodobnost, že na kostce padne sudé číslo, padla-li mi 3?
3. Hodím dvěma kostkami, jaká je pravděpodobnost, že na první padne 6, pokud na druhé padla 2?
4. Jaká je pravděpodobnost, že když hodím dvěma kostkami, že bude součet ok alespoň 10?
5. Hodíme dvěma kostkami, jaká je pravděpodobnost, že součet bude alespoň 10, víme-li, že padla-li na první kostce šestka?
6. Do obchodu chodili lidé v následujícím pořadí:
M,D,M,Ž,M,D,Ž,Ž,D,D,Ž,M,Ž,D,Ž (M-Muž, Ž-žena, D-dítě)
Jaká je pst, že přijde muž? Jaká je pst, že přijde žena, pokud předtím přišlo dítě?
7. Šance politiků A,B,C na vítězství jsou 0.4;0.3;0.2. Politik A musel kvůli korupční aféře odstoupit, jaká jsou nyní šance na zvolení politiků B,C?

Výsledky

1. A: padne sudé
 $P(A)=1/3$
2. A: padne sudé
B: padlo liché
 $P(A|B)=P(A)=1/3$
3. A: na první padne 6
B: na druhé padla 2
 $P(A|B)=P(A)=1/6$
4. Na kostkách může padnou:
1 1; 1 2; ...; 1 6; 2 1; ... 6 6 (celkem 36 možností)
Jen v případech 4 6; 5 5; 5 6; 6 5; 6 6 je součet alespoň 10 (5 možností), tedy $P(A)=5/36$
5. Pst je $\frac{1}{2}$
6. $P(M)=4/15$, $P(\check{Z}|D)=3/5$
7. Musí platit, že $P(B)+P(C)=1$ (nemám jinou možnost). Bez A jsou B a C v poměru 3:2, tedy musí platit, že $P(B)/P(C)=3/2$, tedy $P(B)=1.5P(C)$
Platí $P(B)+P(C)=1$
 $1.5P(C)+P(C)=1$
 $2.5P(C)=1$
 $P(C)=0.4 \rightarrow P(B)=0.6$

Bayesova věta

- Někdy může být těžké určit, $P(A | B)$, zatímco $P(B | A)$ je lehké
- Bayesova věta nám je dává do vztahu:

$$\text{Aposteriorní pst.} \rightarrow P(A_i | B) = \frac{P(B | A_i) \cdot P(A_i)}{P(B)} \leftarrow \text{Apriorní pst.}$$

- $P(B)$ je úplná pravděpodobnost - součet přes všechny možné alternativy A_j . Tedy

$$P(B) = \sum P(B | A_j) \cdot P(A_j)$$

Test HIV

- Máme test, který rozpozná HIV s pstí 95%, občas se ale splete a hlásí pozitivní výskyt i u lidí, kteří HIV nemají (s pstí 10%). V populaci má HIV 1%. Jaká je šance, že když do ordinace přijde člověk, u něhož test ukáže pozitivní výskyt, tak má opravdu HIV

- R : člověk má rakovinu

T : test je pozitivní ($\neg T$: test je negativní)

$$P(R) = 0.01$$

$$P(T|R) = 0.95; P(\neg T|R) = 0.05$$

$$P(T|\neg R) = 0.1; P(\neg T|\neg R) = 0.9$$

- $$P(R|T) = \frac{P(T|R) \cdot P(R)}{P(T)} = \frac{P(T|R) \cdot P(R)}{P(T|R)P(R) + P(T|\neg R)P(\neg R)} = \frac{0.95 \cdot 0.01}{0.95 \cdot 0.01 + 0.1 \cdot 0.99} = 0.088$$

Vysčítání přes všechny možnosti (jsou jen dvě, buď mám nebo nemám rakovinu)

- Tedy člověk má HIV s pstí <10%
- Baysův vzorec obrací kauzální a diagnostický vztah
- $P(\neg T|R)$ se nazývá *false negative*, $P(T|\neg R)$ se nazývá *false positive*
- Pozor: $P(T|R)$ a $P(T|\neg R)$ jsou úplně jiné pravděpodobnosti! Napadá vás případ, kde tyto psti jsou naprosto opačné?

Příklady – Bayesova věta

1. Máme tři šuplíky s ponožkami, v jednom máme červený pár, v druhém jednu ponožku červenou a druhou modrou, ve třetím jsou obě modré. Náhodně vyberu šuplík a vytáhnu modrou ponožku. Jaká je šance, že i druhá ponožka bude modrá?
2. Monty hall problém: Mám tři garáže. V dvou z nich je koza, ve třetí nové auto. Vyberu si jednu garáž, moderátor otevře jedny ze zbývajících dveří (s kozou). Poté vyberu ze zbývajících dvou dveří (buď zůstanu u původních nebo změním) a ty budou otevřeny. Je jedno, jestli změním výběr nebo zůstanu u původních?

Výsledky - Bayes

1. Pst, že druhá ponožka bude modrá=pst, že jsem vybral třetí šuplík

M: vytáhl jsem modrou

Š1: vybral jsem první šuplík; $P(\text{Š1})=1/3$

Š2: vybral jsem druhý šuplík; $P(\text{Š2})=1/3$

Š3: vybral jsem třetí šuplík; $P(\text{Š3})=1/3$

$$P(\text{Š3}|M) = \frac{P(M|\text{Š3})P(\text{Š3})}{P(M|\text{Š1})P(\text{Š1}) + P(M|\text{Š2})P(\text{Š2}) + P(M|\text{Š3})P(\text{Š3})}$$
$$= \frac{1 \cdot \frac{1}{3}}{0 \cdot \frac{1}{3} + 0.5 \cdot \frac{1}{3} + 1 \cdot \frac{1}{3}} = \frac{2}{3}$$

Monty hall problém

2. Lepší je změnit!

Formálně (trochu těžší):

Búno v prvním kroku jsem zvolil první dveře, a moderátor otevřel druhé dveře

A: Auto je v první garáži – $P(A)=1/3$

B: Auto je v druhé garáži – $P(B)=1/3$

C: Auto je ve třetí garáži – $P(C)=1/3$

O: moderátor otevřel dveře garáže

Otázka tedy je, zda $P(A|O)=P(C|O)$

$$P(A|O) = \frac{P(O|A) \cdot P(A)}{P(O|A) \cdot P(A) + P(O|B) \cdot P(B) + P(O|C) \cdot P(C)} = \frac{0.5 \cdot \frac{1}{3}}{0.5 \cdot \frac{1}{3} + 0 \cdot \frac{1}{3} + 1 \cdot \frac{1}{3}} = \frac{1}{3}$$

$$P(C|O) = \frac{P(O|C) \cdot P(C)}{P(O|A) \cdot P(A) + P(O|B) \cdot P(B) + P(O|C) \cdot P(C)} = \frac{1 \cdot \frac{1}{3}}{0.5 \cdot \frac{1}{3} + 0 \cdot \frac{1}{3} + 1 \cdot \frac{1}{3}} = \frac{2}{3}$$

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 p st
- !!!!!Pozor, X není neznámá!!!!