Pravděpodobnostní charakteristiky diagnostických testů, Bayesův vzorec

Prof.RND.Jana Zvárová, DrSc.

Motivace

- V medicíně má mnoho problémů pravěpodobnostní charakter
  - prognóza
  - diagnoza
  - účinnost léčby
- Počet pravděpodobnosti je základem induktivní statistiky
  - zobecnění směrem od výběru k populaci – nejistota; hladina významnosti, p-hodnoty, intervaly spolehlivosti

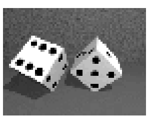
Náhodný pokus, náhodný jev

- Náhodný pokus: výsledek není jednoznačně určen podmínkami, předpokládáme opakovatelnost pokusu, jednotlivá opakování se neovlivňují
- Náhodný jev: tvrzení o výsledku pokusu, lze určit jeho pravdivost

  náhodné jevy A,B,C,D, …

  (Př.A...padnutí šestky, B…narození chlapce)

  negace ¬A



Relativní četnost, pravděpodobnost

- předpokládáme opakování pokusu, sledujeme výsledky:
  A, ¬A, A, ¬A, ¬A, A, A, A, ¬A, A
  jev nastal m krát z *n* pokusů
- Relativní četnost výskytu jevu A: m / n
- Pravděpodobnost jevu A…číslo P(A), které je mírou častosti výskytu A

$$P(A) = \lim_{n \to \infty} \frac{m}{n}$$

Příklad

- Sledujeme náhodný jev " narození chlapce „ v závislosti na rostoucím počtu novorozenců.

ABSOLUTNÍ ČETNOST *m* … počet narozených chlapců
RELATIVNÍ ČETNOST *m/n* … počet narozených chlapců k celkovému počtu novorozenců (často se udává v %)

Počet novorozenců *n*	Absolutní četnost *m*	Relativní četnost *m/n*
1	1	**1**
2	1	**0,50**
3	2	**0,66**
4	2	**0,50**
5	2	**0,40**
6	3	**0,50**
…	…	**…**

Základní vlastnosti

- pravděpodobnost jistého jevu je rovna 1
- pravděpodobnost nemožného jevu je 0
- pro libovolný A platí $0 \leq P(A) \leq 1$
- lze-li A rozložit na několik vzájemně se vylučujících (disjunktních) jevů $A_1, \dots, A_k$, pak $P(A) = P(A_1) + \dots + P(A_k)$
- je-li A částí B, pak $P(A) \leq P(B)$

Pravidla pro počítání

- většinou sledujeme nikoli jeden jev , ale více jevů a zajímají nás jejich vzájemné vztahy
- C=(A,B) … A a B nastanou současně
- D=(A nebo B) … nastane alespoň jeden z jevů A a B
- P(A nebo B) = P(A) + P(B) - P(A,B)

Jevy neslučitelné, opačné

- A a B jsou <u>neslučitelné</u>, když nemohou nastat oba současně, neboli P(A,B)=0
- P(A nebo B) = P(A) + P(B) … pravidlo o sčítání pravděpodobností
- obecněji: nechť $A_1, A_2, ..., A_k$ vzájemně neslučitelné, jev D=(A_1nebo A_2 nebo A_k) $P(D)= P(A_1) + … + P(A_k)=\Sigma P(A_i)$
- <u>opačný</u> (doplňkový) jev k jevu A (značíme ¬A) nastává právě tehdy, když A nenastává
- P(¬A) = 1- P(A)

Příklad

A … narození chlapce, P(A)=0,51
¬A... narození dívky, P(¬A) = 1-P(A)=0,49

Příklad: hod kostkou

Mějme 3 vzájemně neslučitelné jevy: A … padne 1, B … padne 3, C … padne 5
D …padne liché číslo, D=(A nebo B nebo C)
P(D)=P(A)+P(B)+P(C) = 1/6+1/6+1/6=0,5

Podmíněná pravděpodobnost

- pravd. nějakého jevu často závisí na tom, zda nastal jev jiný; nastal-li B může se změnit P(A)
- podmíněná pravděpodobnost jevu A za předpokladu, že nastal jev B

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

Nezávislost jevů

- Jevy A a B nezávislé, když výskyt jednoho neovlivňuje výskyt druhého

$$P(A|B) = P(A) \quad P(B \mid A) = P(B)$$

- pravidlo o násobení pravděpodobností

$$P(A,B) = P(A)P(B)$$

- obecněji: $A_1, A_2, \ldots, A_k$ nezávislé, $C=(A_1, A_2, \ldots, A_k)$

$$P(C) = P(A_1)P(A_2)\ldots P(A_k)$$

Příklad

A … zvýšený cholesterol, B …kouření

P(A) = 37/140=0,2643
P(B) = 98/140=0,7000
P(A,B) = 31/140 = 0,2214
P(A|B) = 0,2214 / 0,7000 = 0,3163
P(A|B)^{1} P(A) … A a B nejsou nezávislé

Příklad: hod kostkou

A…v 1.hodu 6, B…ve 2.hodu 6
P(A,B)=P(A)P(B)=(1/6)(1/6)=1/36=0,0278

Pravidlo o úplné pravděpodobnosti

- jevy B_i (i=1, 2, …,k) vzájemně neslučitelné a jeden z nich musí nastat

$$P(B_1 nebo B_2 nebo....nebo B_k) = \sum_{i=1}^{k} P(B_i) = 1$$

- A=(A,B_1) nebo (A,B_2) nebo … nebo (A,B_k)

$$P(A) = \sum_{i=1}^{k} P(A, B_i) = \sum_{i=1}^{k} P(A \mid B_i) P(B_i)$$

Příklad

A … úraz, zajímá nás P(A)

3 skupiny osob rozdělěné dle věku:
B_1…dítě, B_2… osoba v reprod.věku, B_3… osoba v postreprod.věku, B_i ... vzájemně neslučitelné, 1 musí nastat
$P(B_1)+P(B_2)+P(B_3)=0,25+0,60+0,15=1$

navíc známe podmíněné pravděpodobnosti:
$P(A| B_1)=0,2$; $P(A| B_2)=0,1$; $P(A| B_3)=0,4$

$P(A) = P(A| B_1) P(B_1) + P(A| B_2) P(B_2) + P(A| B_3) P(B_3) =$
$= 0,20*0,25 + 0,10*0,60 + 0,40*0,15 = 0,17$

Bayesův vzorec

- známe apriorní pravděpodobnosti $P(B_i)$ i=1,...,k
- známe podm. pravděpodobnosti $P(A|B_i)$ i=1,...,k
- zajímá nás aposteriorní pravděp. $P(B_j|A)$

$$P(B_j \mid A) = \frac{P(A \mid B_j) P(B_j)}{\sum_{i=1}^{k} P(A \mid B_i) P(B_i)}$$

Příklad

A …osoba je kuřák, zajímá nás P(A)

B_1…osoba s chron. bronchitidou, B_2… osoba bez chron.bronchitidy, $P(B_1)=0{,}40$, $P(B_2)=0{,}60$

navíc známe podmíněné pravděpodobnosti: $P(A|\,B_1)=0{,}75$; $P(A|\,B_2)=0{,}50$

$$P(B_1\,|\,A)=\frac{P(A\,|\,B_1)P(B_1)}{P(A\,|\,B_1)P(B_1)+P(A\,|\,B_2)P(B_2)}$$
$$=\frac{0{,}75*0{,}40}{0{,}75*0{,}40+0{,}50*0{,}60}=0{,}50$$

BAYESOVSKÝ PŘÍSTUP

SKRÍNINGOVÝ TEST T	NEMOC D + -	*CELKEM*
+ -	*a* *b* *c* *d*	*a + b* *c + d*
CELKEM	*a + c* *b + d*	*n*

SENSITIVITA a SPECIFICITA

SENSITIVITA (SE) je pravděpodobnost P (T^+/D^+) pozitivního výsledku testu u nemocné osoby

SE = a / (a + c)

SPECIFICITA (SP) je proevděpodobnost P(T^-/D^-) negativního výsledku testu u osoby bez nemoci

SP = d / (b + d)

NESPRÁVNÁ NEGATIVITA A NESPRÁVNÁ POZITIVITA

NESPRÁVNÁ NEGATIVITA (FN) je pravděpodobnost $P(T^-/D^+)$ *negativního* výsledku testu u nemocných

FN = c / (a + c)

NESPRÁVNÁ POZITIVITA (FP) je pravděpodobnost $P(T^+/D^-)$ of *pozitivního* výsledku testu u osob bez nemoci

FP = b / (b + d)

Hodnocení diagnostického či skríningového testu pro detekci nemoci

ALE: v klinické praxi nevíme, zda je nemoc přítomna či nikoli; známe jen výsledek testu a na jeho základě chceme predikovat přítomnost choroby ... P(D+|T+)
musíme na data nahlížet „ve směru“ výsledků testu ® prediktivní hodnoty

PREDIKTIVNÍ HODNOTY

PREDIKTIVNÍ HODNOTA POZITIVNÍHO TESTU je pravděpodobnost $P(D^+/T^+)$ výskytu nemoci v případě pozitivního výsledku testu

PV^+ = a / (a + b)

PREDIKTIVNÍ HODNOTA NEGATIVNÍHO TESTU je prevděpodobnost $P(D^-/T^-)$, že se nemoc nevyskytne v případě negativního výsledku testu

PV^- = d / (c + d)

Prediktivní hodnota pozitivního testu pomocí Bayesova vzorce

$$P(D+|T+) = \frac{P(T+|D+)P(D+)}{P(T+|D+)P(D+) + P(T+|D-)P(D-)}$$

$$= \frac{SE * P(D+)}{SE * P(D+) + (1-SP) * (1-P(D+)}$$

P(D+) ... apriorní předtestová pravděpodobnost D
P(D+|T+) ... aposteriorní potestová pravděpodobnost D

POZOR: pro SE=0,95, SP=0,95, P(D+)=0,01
dostaneme PV+=0,16
při skríningu obecné populace bude nevyhnutelně mnoho lidí nesprávně pozitivních

VZTAH MEZI SENZITIVITOU (SE), SPECIFICITOU (SP), PREVALENCÍ (P (D+)) A PREDIKTIVNÍMI HODNOTAMI (PV^+, PV^-) VYPLYVAJÍCÍ Z BAYESOVA VZORCE

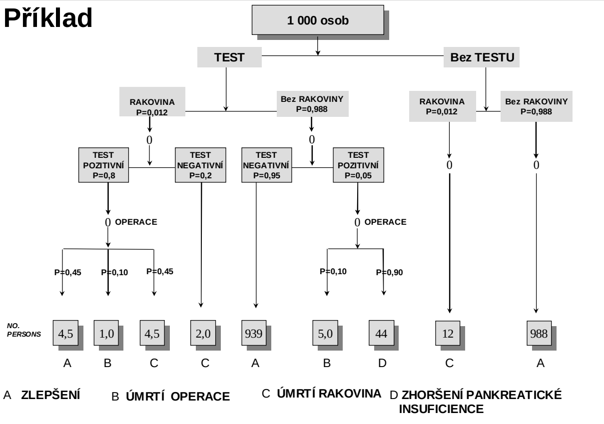
PV^+ =(SE. P (D +)) / (SE.P (D+) + (1 - SP).(1-P (D+)))

PV^- =(SP. (1-P (D +)) / (SP.(1-P (D+) + (1 - SE).P (D+))

ROC křivka

- řada diagnostických testů je kvantitativních
- jak stanovit dělící bod (cut-off point)?
- cíl: najít dělící bod tak, abychom dosáhli rovnováhy mezi FP a FN závěry (váhy nesprávných rozhodnutí)
- ROC křivka: spočteme SE a SP pro různé dělící body

Příklad



Šance

Řekneme, že šance (odds) závodního koně na první místo v dostihovém závodě (jev *A*) je 1 ku 4, znamená to, že kůň závod vyhraje s pravděpodobností

$$P(A) = 1/5 = 0{,}20$$

Abychom vyjádření pomocí šance převedli na vyjádření pomocí pravděpodobnosti, sečteme vlastně čísla 1 + 4 = 5 a dostaneme tak jmenovatel zlomku pro vyjádření pravděpodobnosti výhry, tj. 1/5.

$$O(A) = \frac{P(A)}{P(\neg A)} = \frac{P(A)}{1 - P(A)}$$

Pro libovolný náhodný jev *A* tedy platí: *šance* O(A) výskytu jevu *A* je

$$P(A) = \frac{O(A)}{1 + O(A)}.$$

Podíl šancí (Odds ratio)

Podíl šancí (odds ratio) OR udává podíl šanci, že se vyskytne nějaký jev *A* za určité podmínky (jev *B*), k šanci, že se jev *A* vyskytne, když podmínka neplatí (jev ¬B) . Podíl šancí se tedy vypočte jako

$$OR = \frac{O(A|B)}{O(A|\neg B)},$$

přičemž

$$O(A|B) = \frac{P(A|B)}{P(\neg A|B)} \text{ a } O(A|\neg B) = \frac{P(A|\neg B)}{P(\neg A|\neg B)}.$$

Věrohodnostní poměr

Věrohodnostní poměr (likelihood ratio) LR

udává podíl pravděpodobnosti, že se vyskytne nějaký jev *A* za určité podmínky (jev *B*), k pravděpodobnosti, že se jev *A* vyskytne, když podmínka neplatí (jev ¬ B), tedy

$$LR = \frac{P(A|B)}{P(A|\neg B)}.$$

Věrohodnostní poměr - příklad

Má-li pacient náhlou ztrátu paměti (jev *A*), chceme znát věrohodnostní poměr výskytu jevu *A* v případě, že má mozkový nádor (jev *B*), tj. podíl pravděpodobnosti, s jakou ztráta paměti vzniká při nádoru mozku, k pravděpodobnosti, s jakou vzniká v ostatních případech (jev ¬ B). Věrohodnostní poměr je tedy podíl podmíněných pravděpodobností

$$LR = \frac{P(A|B)}{P(A|\neg B)}.$$

Příklad

Ve statistické studii o rakovině plic bylo zjištěno, že šance na výskyt rakoviny plic (jev *A*) u kuřáků (jev *B*) je 5 ku 4 (5/4) a šance na výskyt rakoviny u nekuřáků (jev ¬ B) je 1 ku 8 (1/8). Potom podíl šancí je

$$\frac{\frac{5}{4}}{\frac{1}{8}} = \frac{40}{4} = 10,$$

což znamená, že šance dostat rakovinu plic je 10x větší u kuřáků než u nekuřáků.

Věrohodnostní poměr

Věrohodnostní poměr užíváme i při hodnocení skríningových a diagnostických testů a ve forenzní genetice. Například věrohodnostní poměr pozitivního skríningového testu je dán jako

$$P(T^+|D^+)/P(T^+|D^-)$$

Podobně věrohodnostní poměr negativního testu spočteme jako

$$P(T^-|D^+)/P(T^-|D^-)$$