



INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ

Rychlokurz forenzní DNA statistiky

21.10.2011

Anastassiya Žídková
anastazie.d@gmail.com



INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ

Program dnešního kurzu

- Úvod
- První část
 - Základní zákony pravděpodobnosti
- **Druhá část**
 - **Bayesova věta**
 - **Zásady při interpretaci výsledků**
 - **Nejčastější chyby při interpretaci výsledků**



INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ

Různé přístupy při interpretaci výsledků

- Frekventistický
 - Testuje se jedná hypotéza
 - Random Match Probability, Power of Exclusion
- Logický
 - Testuje se dvě protichůdné hypotézy
 - LR, Bayesova věta
- Plně Bayesovský
 - Může být testován libovolný počet hypotéz
 - Obecné znění Bayesovy věty pro i hypotéz



INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ

Druhá část

Bayesova věta



Thomas Bayes (1701 – 1761)



INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ

Výhody Bayesovy věty

- Je logická
- Je univerzální, používá se v různých odvětvích (epidemiologie, evoluce, finančnictví, filtrování spamu, odhalování plagiátorství)
- Definuje roli vědce
- Aktualizovat nejistotu tvrzení ve světle nových důkazů
- Umožňuje kombinovat forenzní vědu s požadavky soudu



INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ

Nevýhody Bayesovy věty

- Přiznává subjektivitu ve statistické analýze
- Někdy jde proti selskému rozumu
- V základních kurzech statistiky se neučí



INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ

Šance (Odds)

Šance jevu A – $O(A)$ je pravděpodobnost toho, že nastane jev A vydělena pravděpodobností toho, že jev A nenastane.

$$O(A) = \frac{P(A)}{P(\bar{A})}$$



INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ

Oblast hodnot šance a pravděpodobnosti

- Oblast hodnot šance
 $[0, \infty)$, pokud jsou šance menší než 1, tak je pravděpodobnější jmenovatel
- Oblast hodnot pravděpodobnosti
 $[0,1]$



INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ

Šance

Příklad: Jaká je šance, že při hodu kostkou padne číslo menší než 5 (jev A)?





INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ

Šance

Řešení: šance, že padne číslo menší než 5 je 4/6.

$$P(A) = 4/6 \quad P(\bar{A}) = 2/6$$

$$O(A) = \frac{P(A)}{P(\bar{A})} = \frac{4/6}{2/6} = 2$$

Je dvakrát větší šance, že padne číslo menší než 5, než že se tak nestane.



INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ

Šance

Příklad: Frekvence alely 8 v lokusu TPOX v české populaci je 0,5493. Zjistěte, jaká je šance, že vzorek z místa činu bude mít v lokusu TPOX alelu 8 (jev A)?



INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ

Šance

Řešení: $P(A)$ – šance, že vzorek bude mít alespoň jednu alelu 8 v lokusu TPOX.

$$P(A) = 1 - (1 - p_8)^2 = 1 - (1 - 0,5493)^2 = 0,7969$$

Frekvence zbývajících alel

$$O(A) = \frac{P(A)}{P(\bar{A})} = \frac{0,7969}{0,2031} = 3,92$$

Je 3,9 krát větší šance, že vzorek bude mít v lokusu TPOX alespoň jednu alelu 8, než že ji mít nebude.

Odvození pravděpodobnosti ze šance

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

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

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

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



INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ

Odvození pravděpodobnosti ze šance

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



INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ

Odvození pravděpodobnosti ze šance

Příklad: Vsadili jste celý svůj majetek na to, že při hodu kostkou padne jednička i přesto, že známý statistik Vám tvrdí, že šance padnutí jedničky je $1/5$ ($O(A)$). Jaká je pravděpodobnost, že při hodu kostkou padne kýžená jednička (jev A)?





INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ

Odvození pravděpodobnosti ze šance

$$\text{Řešení: } P(A) = \frac{O(A)}{1+O(A)} = \frac{\frac{1}{5}}{1+\frac{1}{5}} = \frac{\frac{1}{5}}{\frac{6}{5}} = \frac{1}{6} \approx 16,67\%$$



INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ

Odvození pravděpodobnosti ze šance

Příklad: Šance, že najdete velmi vzácný genotyp na hypotetickém ostrově je $1/500$. Jaká je frekvence tohoto genotypu?





INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ

Odvození pravděpodobnosti ze šance

$$\text{Řešení: } P(A) = \frac{O(A)}{1+O(A)} = \frac{\frac{1}{500}}{1+\frac{1}{500}} \approx 0,1996\%$$



INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ

Odvození Bayesovy věty

Budeme používat následující symboly:

H_p – hypotéza obžaloby (prosecutor's hypothesis)

H_d – hypotéza obhajoby (defence hypothesis)

E – vědecké důkazy

I – nevědecké důkazy



INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ

Odvození Bayesovy věty

Pomocí třetího zákona pravděpodobnosti

$$P(H_p, E | I) = P(H_p | E, I) \times P(E | I)$$

Potom

Tak lze zapsat průnik obou jevů

$$P(H_p | E, I) = \frac{P(H_p, E | I)}{P(E | I)} = \frac{P(E | H_p, I) \times P(H_p | I)}{P(E | I)}$$

Tento výraz rozepíšeme
taky pomocí třetího
zákonu pravděpodobnosti



INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ

Odvození Bayesovy věty

Totéž uděláme i pro pravděpodobnost hypotézy obhajoby

$$P(H_d, E | I) = P(H_d | E, I) \times P(E | I)$$

Potom

$$P(H_d | E) = \frac{P(H_d, E | I)}{P(E | I)} = \frac{P(E | H_d, I) \times P(H_d | I)}{P(E | I)}$$



INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ

Odvození Bayesovy věty

Uděláme poměr obou rovnic

$$\frac{P(H_p|E,I)}{P(H_d|E,I)} = \frac{\frac{P(E|H_p,I) \times P(H_p|I)}{P(E|I)}}{\frac{P(E|H_d,I) \times P(H_d|I)}{P(E|I)}} = \frac{P(E|H_p,I) \times P(H_p|I)}{P(E|H_d,I) \times P(H_d|I)}$$



INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ

Odvození Bayesovy věty

$$\frac{P(H_p|E,I)}{P(H_d|E,I)} = \frac{P(E|H_p,I)}{P(E|H_d,I)} \times \frac{P(H_p|I)}{P(H_d|I)}$$

Aposteriorní šance

Věrohodnostní poměr

Apriorní šance



INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ

Bayesova věta

Příklad: Máte v krabici $\frac{1}{3}$ kuliček, které mají na sobě písmeno H (jev H) a $\frac{2}{3}$ písmeno T (jev T). $\frac{1}{4}$ kuliček s písmenem H je černých a $\frac{3}{4}$ kuliček s písmenem H jsou bílé. $\frac{3}{4}$ kuliček s písmenem T je černých a $\frac{1}{4}$ kuliček s písmenem T jsou bílé. Jaká je pravděpodobnost, že pokud vytáhnete bílou kuličku, bude mít na sobě písmeno H?



INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ

Bayesova věta

Řešení: $P(H)=1/3$, $P(T)=2/3$, $P(B|H)=1/4$, $P(B|T)=3/4$,
 $P(W|H)=3/4$, $P(W|T)=1/4$

$$\frac{P(H|W,I)}{P(T|W,I)} = \frac{P(W|H,I)}{P(W|T,I)} \times \frac{P(H|I)}{P(T|I)} = \frac{3/4}{1/4} \times \frac{1/3}{2/3} = \frac{3}{2} = O(H)$$

Poté vypočítáme $P(H|W,I)$ ze vzorce šance

$$P(H|W,I) = \frac{O(H)}{1 + O(H)} = \frac{3/2}{1 + 3/2} = 3/5$$



INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ

Bayesova věta

Příklad: Opět máme hypotetické město, kde 25% obyvatelstva tvoří Španělé a 75% Afro-Američani. Krevní stopa pachatele nalezena na místě činu má v lokusu vWA genotyp 8, 11 (jev A). Žádní očití svědci nejsou, ale vyšetřovatel je přesvědčen, že pachatel je Afro-Američan, protože jich je ve městě více, než Španělů (jev B). Frekvence genotypu 8,11 v lokusu vWA je 0,182 ($A|B$) v populaci Španělů a 0,013 v populaci Afro-Američanů ($A|\bar{B}$). Jak ovlivní tyto frekvence pohled vyšetřovatele?



INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ

Bayesova věta

Řešení: $P(B) = 0,25$, $P(\bar{B}) = 0,75$ $P(A|B,I) = 0,182$

$P(A|\bar{B},I) = 0,013$

$$\frac{P(B|A,I)}{P(\bar{B}|A,I)} = \frac{P(A|B,I)}{P(A|\bar{B},I)} \times \frac{P(B|I)}{P(\bar{B}|I)} = \frac{0,182}{0,013} \times \frac{1/4}{3/4} \approx 4,67 = O$$

Poté vypočítáme $P(B|A,I)$ ze vzorce šance

$$P(B|A,I) = \frac{O}{1+O} = \frac{4,67}{1+4,67} = 0,823$$

Pravděpodobnost toho, že pachatel byl Španěl je 82%



INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ

Apriorní šance

- To je otázka pro soudce před tím, než nechá udělat znalecké posudky
- Je založena na analýze nevědecké informace
 - Kriminalistické případy: očití svědci, výslech oběti atd.
 - Paternitní případy: zda je údajný otec manžel matky dítěte, mají společnou domácnost, atd.
- Je téměř vždy subjektivní, ovšem využívající veškerou dostupnou relevantní informaci



INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ

Věrohodnostní poměr

- Otázka pro vědce
- Zjišťuje, zda jsou důkazy pravdivé za předpokladu tvrzení
- Aktualizuje apriorní šanci
- Sumarizuje veškerou informaci získanou znaleckým vyšetřením



INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ

Aposteriorní šance

- Otázka pro soudce po prozkoumání všech důkazů
- Zjišťuje, jaká je pravděpodobnost, že tvrzení je pravdivé za předpokladu všech důkazů



INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ

Otázky pro soud

- Jaká je pravděpodobnost toho, že hypotéza obžaloby je pravdivá za předpokladu zjištěných důkazů?

$$P(H_p | E, I)$$

- Jaká je pravděpodobnost toho, že hypotéza obhajoby je pravdivá za předpokladu zjištěných důkazů?

$$P(H_d | E, I)$$



INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ

Otázky pro forezního genetika

- Jaká je pravděpodobnost genotypizačních výsledků, pokud platí hypotéza obžaloby?

$$P(E | H_p, I)$$

- Jaká je pravděpodobnost genotypizačních výsledků, pokud platí hypotéza obhajoby?

$$P(E | H_d, I)$$



INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ

Zásady při interpretaci výsledků



INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ

První zásada

Pro zjištění nejistoty jakéhokoliv tvrzení je nezbytné uvažovat o nejméně jednom alternativním tvrzení.

- Nejčastěji uvažujeme o hypotéze obhajoby a obžaloby



INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ

Druhá zásada

Vědecká interpretace je založena na otázce typu „Jaká je pravděpodobnost výsledků za podmínky tvrzení?“



INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ

Třetí zásada

Vědecká interpretace je podmíněna nejen alternativními tvrzeními, ale i souborem doplňující nevědecké informace, která by měla být brána v úvahu.

- Příkladem takové informace je populace ze které pochází vinný.



INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ

Interpretace výsledků

Znalecký posudek, musí obsahovat kromě pravděpodobnosti spočítané pomocí Bayesovy věty i LR z důvodu již zmíněné subjektivity apriorní pravděpodobnosti.



INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ

- Na základě všech znaleckých posudků a dalších důkazů musí soudce stanovit prahovou hodnotu pro legální důsledky
- Soudce také zvažuje celospolečenské důsledky falešné positivity a falešné negativity



INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ

Nejčastější chyby při interpretaci výsledků



INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ

Chyba obžaloby (Transposed conditional - Záměna příčiny a důsledku)

- Věrohodnostní poměr zjišťuje pravděpodobnost výsledků za předpokladu hypotézy a ne naopak

$$P(E|H,I) \neq P(H|E,I) !!!$$

$$P(\text{zvíře je slon} | \text{zvíře má 4 nohy}) \neq P(\text{zvíře má 4 nohy} | \text{zvíře je slon})$$



INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ

Chyba obžaloby (Transposed conditional - Záměna příčiny a důsledku)

- **Správný závěr:** Výsledky DNA analýzy jsou 1000 krát pravděpodobnější, pokud obžalovaný zanechal stopu, než pokud ji zanechal jiný člověk.
- **Nesprávný závěr:** Je 1000 krát větší pravděpodobnost, že obžalovaný zanechal stopu, než pokud ji zanechal jiný člověk.



INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ

Chyba obhajoby - Příklad

- Pravděpodobnost náhodné shody DNA profilu je 1 z 100 000. Trestní čin byl spáchán ve městě s 1 000 000 obyvatel
- **Správný závěr:** Lze očekávat, že 10 lidí ve městě budou mít daný DNA profil.
- **Nesprávný závěr:** Obžalovaný (ten kdo má daný profil) má pravděpodobnost 1 z 10 být vinný.



INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ

Chyba obhajoby

- Chyba obhajoby spočívá v přidělování stejné pravděpodobnosti všem 10 lidem, u nichž je předpokládán daný DNA profil. Počet lidí ve městě s daným profilem může být jakékoliv číslo od 0 do 1 000 000. Očekávaná čísla nejsou skutečná čísla.



INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ

Chyba jedinečnosti - Příklad

- Pravděpodobnost náhodné shody DNA profilu je 1 z 1 000 000. Trestný čin byl spáchán ve městě s 1 000 000 obyvatel
- **Správný závěr:** Lze očekávat, že 1 člověk ve městě bude mít daný profil.
- **Nesprávný závěr:** Obžalovaný (ten kdo má daný profil) je vinný.



INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ

Chyba jedinečnosti

- Chyba jedinečnosti je v přehlédnutí faktu, že počet lidí s daným profilem ve městě může být jakékoliv číslo od 0 do 1 000 000. Očekávaná čísla nejsou skutečná čísla.



INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ

Význam frekvencí

- Co znamená výraz “frekvence shodného DNA profilu je 1 z 57 miliard“?
- Je to očekávaná pravděpodobnost, která byla získána násobením frekvencí jednotlivých alel a odkazuje na nekonečně velkou populaci s náhodným křížením. Nemá nic společného s velikostí celosvětové populace.



INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ

Použitá literatura

- Interpreting DNA evidence, Ian Evett and Bruce Weir, 1998
- Probability for dummies, Deborah Rumsey, 2006
- Introduction to statistics for forensic scientists, David Lucy, 2006
- Buckleton, Forensic DNA Evidence Interpretation, 2005



INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ

Děkuji za pozornost

Anastassiya Žídková
anastazie.d@gmail.com
<http://dnaint.martinzidek.com/>