Opravdu nemocný?

Daniela Hrbáčová, 18.9. 2019

Univerzita Karlova v Praze

Matematicko-fyzikální fakulta

Ukázky aplikací matematiky NMAG166

Situace

Pan G, občan České republiky, v poslední době často pociťuje žízeň bez zjevné příčiny, ba i hojně chodí na toaletu. Proto se rozhodl zajít k lékaři, jelikož cukrovka se vyskytuje nezřídka v jeho rodině. V ordinaci jej doktor dle obvyklých postupů (z ekonomických důvodů) neposlal na glukózový toleranční test a sestra pouze vyšetřila vzorek pánovy moči ponořením detekčního proužku. Diagnostický proužek pro analýzu glukózy v moči nezměnil barvu, což zřejmě znamená, že v pacientově moči není žádná glukóza. Tedy sestra sdělila pacientovi, že cukrovku nemá. Může pacient s klidem v duši odejít, že to byl planý poplach a cukrovkou tedy netrpí?

Analýza problému

Jak ze života víme, nic není stoprocentní. Tedy i testy mají své určité parametry – zejména se udává senzitivita a specificita každého testu. **Senzitivitou** chápeme pravděpodobnost pozitivního výsledku testu u osoby trpící danou nemocí (tedy, že test správně určí nemocnou osobu), kdežto **specificita** je její opak, pravděpodobnost negativního výsledku testu u osoby netrpící danou nemocí. [4] U použitého detekčního proužku výrobce udává senzitivitu 22 % a specificitu 99 %. [5] Dále se může uvádět pravděpodobnost výskytu falešně pozitivního, či negativního výsledku, to znamená, že test by dotyčnému vyšel pozitivně a nemocí by reálně netrpěl, případně by byl negativní výsledek testu a pacient by přesto byl nemocný. Tyto parametry se ale většinou neuvádějí. Dále ještě hraje důležitou roli jedna pravděpodobnost – pravděpodobnost výskytu dané choroby v testované populaci.

Teorie pravděpodobnosti

Klasická definice pravděpodobnosti: nechť elementární jevy jsou všechny stejně pravděpodobné, pak pravděpodobnost jevu A je podíl počtu příznivých jevů ku počtu všech elementárních jevů. [2] Tato definice zde není příliš vhodná, jelikož pravděpodobnost, že obyvatel České republiky trpí cukrovkou je asi 8 %. [1]

Zadefinuji v tomto případě **Kolmogorovu** pravděpodobnost podle [2].

Buď Ω množina elementárních jevů ω, řekneme, že systém podmnožin A množiny Ω je σ-algebra, jestliže pro něj platí následující:

(i) $\mathscr{A}$ je neprázdná,

(ii) $A_1, A_2, \cdots \in A \Rightarrow \bigcup_{i=1}^{\infty} A_i \in \mathscr{A}$,

(iii) $A \in \mathscr{A} \Rightarrow A^c \in \mathscr{A}$.

Dále definujme podle [2]:

****Pravděpodobnost*** *je reálná funkce $P(A)$: $A \in \mathscr{A} \rightarrow [0,1]$, pro kterou platí:*

(i) $P(\Omega) = 1$, $P(\emptyset) = 0$,

(ii) $P(A) \geq 0$ pro každé $A \in \mathscr{A}$,

(iii) jsou-li $A_1, A_2, \cdots$ disjunktní množiny z $\mathscr{A}$, pak $P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i)$.

****Podmíněná pravděpodobnost*** *jevu A za podmínky, že nastal jev B, pokud jev B nastane s pravděpodobností větší než 0 definujeme jako $P(A|B) = \frac{P(A \cap B)}{P(B)}$.* [3]

****Úplný systém jevů*** *je systém jevů $B_1, B_2, \ldots$ takový, že platí:*

(i) $\bigcup_i B_i = \Omega$,

(ii) $B_i \cap B_j = \emptyset$ pro $i \neq j$.

A na závěr doslova pár vět, které vychází z [2].

(Věta o úplné pravděpodobnosti)
Nechť $B_1, \ldots, B_n$ je úplný systém jevů a pro každé i platí $P(B_i) > 0$, pak pro každý jev A platí $P(A) = \sum_{i=1}^{n} P(A|B_i) \cdot P(B_i)$.

Důkaz:

A můžeme zapsat jako $A \cap \Omega$ a Ω je sjednocení B_i pro i od 1 do n, tedy $A \cap (B_1 \cup \cdots \cup B_n) = (A \cap B_1) \cup \cdots \cup (A \cap B_n)$. Jevy $A \cap B_i$ jsou disjunktní. Tedy $P(A) = \sum_{i=1}^{n} P(A \cap B_i) = \sum_{i=1}^{n} P(A|B_i) \cdot P(B_i)$.

(Bayesova věta)
Nechť $B_1, \ldots, B_n$ je úplný systém jevů; $P(B_i) > 0$ pro $i = 1, \ldots, n$; A je jev a $P(A) > 0$, pak platí $P(B_k|A) = \frac{P(A|B_k) \cdot P(B_k)}{\sum_{i=1}^{n} P(A|B_i) \cdot P(B_i)} = \frac{P(A|B_k) \cdot P(B_k)}{P(A)}$.

Důkaz:

$$P(B_k|A) = \frac{P(B_k \cap A)}{P(A)} = \frac{P(A|B_k) \cdot P(B_k)}{\sum_{i=1}^{n} P(A|B_i) \cdot P(B_i)}.$$

Výpočet

Shrňme si informace, které o testování pana G máme.

Specificita testu – 99 %, senzitivita – 22 %. [5] Pravděpodobnost, že náhodný člověk trpí cukrovkou (před testem) – P(C) = 8 %. [1]

Senzitivitu z definice můžeme vyjádřit jako P(+|C), kde P(+) je pravděpodobnost, že pacientovi vyšel pozitivně test a P(C) je pravděpodobnost, že pacient trpí cukrovkou. Specificitu jako $P(-|\bar{C})$, kde $P(-)$ je pravděpodobnost, že test vyšel negativně a $P(\bar{C})$ pravděpodobnost, že pacient netrpí cukrovkou.

Jelikož C a $\bar{C}$ je úplný systém jevů nastávajících s nenulovou pravděpodobností, můžeme ze zadaných údajů získat P(C|T) použitím Bayesovy věty. Tedy:

$$P(C/+) = \frac{P(+|C)\cdot P(C)}{P(+|C)\cdot P(C)+P(+|\bar{C})\cdot P(\bar{C})} = \frac{0{,}22\cdot 0{,}08}{0{,}22\cdot 0{,}08+0{,}01\cdot 0{,}92} \cong 0{,}657$$

Nás ovšem zajímá jiná podmíněná pravděpodobnost – P(C|$\bar{T}$), protože zkoumané osobě test na glukózu v moči vyšel negativně, ale má hypotézu, že je nemocná. Tuto pravděpodobnost můžeme také zjistit následujícím výpočtem:

$$P(C/-) = \frac{P(-|C)\cdot P(C)}{P(-|C)\cdot P(C)+P(-|\bar{C})\cdot P(\bar{C})} = \frac{0{,}78\cdot 0{,}08}{0{,}78\cdot 0{,}08+0{,}99\cdot 0{,}92} \cong 0{,}064$$

Tedy u neznámého člověka náhodně vybraného z populace je pravděpodobnost, že trpí cukrovkou, pokud mu vyšel test moči negativně, zhruba stejná jako byla před testem, asi 6,4 %. Jenže z lékařského hlediska zde musíme přihlédnout k tomu, že detekování glukózy v moči pomocí proužku může dopadnout pozitivně i z jiných důvodů, než že má pacient cukrovku. Například proto, že má nízký renální práh (ledviny nezadržují dostatečné množství glukózy, a tak i když je množství glukózy v krvi přijatelné, tak v moči je jí nadbytek). A naopak může vyjít negativně z důvodu vysokého renálního prahu – v krvi je množství glukózy nadbytečné, ale ledviny vyloučí do moči jen zanedbatelné množství. [5] A navíc u pana G víme, že se v jeho rodině vyskytuje často tato nemoc, tak pravděpodobnost, že se vyskytne i u něj je podstatně větší. Což výrazně ovlivní výsledek výpočtu. Kdyby P(C_g) byla pravděpodobnost, že pan G má cukrovku a odhadli bychom ji například padesáti procenty, pak by P(C_g|—) vzrostla z hodnoty pro průměrného člověka – 6,4 % na asi 44 % (což je podstatně vyšší hodnota blížící se polovině). Navíc si ještě musíme uvědomit, že P(—|C) = 1 – P(+|C) = 0,78, tedy cukrovkáři pravděpodobněji nevyjde test, než vyjde.

Další příklady

Příklad 1

Firma testuje namátkou zaměstnance po příchodu do práce na přítomnost omamných látek – drog. Paní M byla náhodně vybrána a otestována, test jí vyšel pozitivně. Má problém. Jaká je pravděpodobnost, že je skutečně uživatelkou drog?

U testu na přítomnost drog máme senzitivitu i specificitu 99 %. A pravděpodobnost, že náhodně vybraný člověk z populace je uživatelem drog odhadněme 0,5 %. [6]

Označme si P(D) jako pravděpodobnost, že náhodný člověk je uživatelem drog a P(+) jako v předchozím případě, že test vyšel pozitivně. Pak podmíněnou pravděpodobnost, že osobě vyjde test na přítomnost omamných látek pozitivně a osoba je skutečně uživatelem drog vypočteme jako:

$$P(D/+) = \frac{P(+|D)\cdot P(D)}{P(+|D)\cdot P(D)+P(+|\bar{D})\cdot P(\bar{D})} = \frac{0{,}99\cdot 0{,}005}{0{,}99\cdot 0{,}005+0{,}01\cdot 0{,}995} \cong 0{,}33$$

Tedy pokud paní M vyšel test pozitivně, pak s pravděpodobností pouhých 33 % skutečně užívá drogy, kdežto s pravděpodobností 67 % omamné látky přítomné nebyly. Je to docela překvapivý výsledek testu s 99% „správností“. V realitě naštěstí poté, co vyjde test na drogy pozitivně, je člověk odvezen do nemocnice a podstupuje kvalifikovanější vyšetření z krve a až na základě výsledku této analýzy se soudí.

Paní M se tedy moc nemusí znepokojovat snad falešnou pozitivitou výsledku a při druhém vyšetření pravděpodobně lékaři odhalí, že paní není uživatelkou drog a šlo o falešný poplach.

Příklad 2

Ženy nad 45 let by měly chodit co 2 roky na preventivní vyšetření prsu mamografickým screeningem z důvodu včasného zachycení rakoviny prsu. Slečna A jedno takové vyšetření právě podstoupila ve věku 47 let, vyšel jí pozitivní výsledek a chce od lékaře vědět, zda tedy opravdu nádor v prsu má.

V případě vyšetření prsu je senzitivita 80 % a specificita celých 90 %. [5] Pravděpodobnost výskytu karcinomu prsu u 47leté ženy je asi 0,1 %. [7]

Chápejme P(R) pravděpodobnost, že náhodně vybraná žena ve věku 47 let má rakovinu prsu a P(+), že mamograf potvrdil nález nádoru. Pak opět podmíněnou pravděpodobnost, že prs paní A obsahuje karcinom za podmínky, že screening potvrdil nález, vypočteme z Bayesovy věty následovně:

$$P(R/+) = \frac{P(+|R)\cdot P(R)}{P(+|R)\cdot P(R)+P(+|\bar{R})\cdot P(\bar{R})} = \frac{0{,}8\cdot 0{,}001}{0{,}8\cdot 0{,}001+0{,}1\cdot 0{,}999} \cong 0{,}008$$

Tedy slečně A hrozí rakovina prsu jen s pravděpodobností 0,8 %. Naopak kdyby mamograf vyšel negativně, pak bychom obdobným výpočtem došli k závěru, že pravděpodobnost, že trpí rakovinou a screening vyšel negativně je 0,022 %.

Důsledek

V situaci pana G by bylo více než vhodné být podroben glukózovému tolerančnímu testu z krve podle zlatého standardu. Byť by to bylo finančně náročnější, ale na druhou stranu by se pacient dozvěděl s největší pravděpodobností správný výsledek, jestli je, či není nemocen. Zlatý standard spočívá v tom, že je test proveden dvakrát za sebou, a tedy pravděpodobnost, že je výsledek testu pravdivý, se rapidně blíží 100 %. I u vyšetření přítomnosti glukózy v moči při druhém provedení testu se stejným (negativním) výsledkem se pravděpodobnost, že je testovaný skutečně zdravý vyšplhá k jedničce a nemocný by subjekt byl jen s minimální doplňkovou pravděpodobností.

Jako závěr bych z tohoto vyvodila to, že pokud by mě testovali na nějakou relativně výjimečně se vyskytující chorobu, tolik bych se neznepokojovala výsledkem prvního testování a pro jistotu nebo upřesnění bych se nechala otestovat ještě jednou.

Zdroje

[1] HATLAPATKOVÁ, Lenka. Počet diabetiků v České republice roste, ročně zhruba o deset tisíc. *Novinky* [online]. [cit. 2019-09-18]. Dostupné z: https://www.novinky.cz/vase-zpravy/clanek/pocet-diabetiku-v-ceske-republice-roste-rocne-zhruba-o-deset-tisic-40193043

[2] REXOVÁ, Páťa. *Bayesova věta a lékařská diagnostika* [online]. [cit. 2019-09-18]. Dostupné z: https://mks.mff.cuni.cz/library/BayesovaVetaPR/BayesovaVetaPR.pdf

[3] *ELEMENTÁRNÍ POČET PRAVDĚPODOBNOSTI* [online]. [cit. 2019-09-18]. Dostupné z: https://homel.vsb.cz/~dom033/predmety/statistika/ucebni_text/4pravdepodobnost.pdf Vysoká škola báňská - Technická univerzita Ostrava.

[4] TVRDÍK, Josef a Hana TOMÁŠKOVÁ. *Lékařská biofyzika, výpočetní technika I* [online]. [cit. 2019-09-18]. Dostupné z: http://www1.osu.cz/~tvrdik/wp-content/uploads/lekbio_statistika3.pdf. Ostravská univerzita.

[5] Příklady na využití hodnocení diagnostických testů. *KRAJSKÁ HYGIENICKÁ STANICE KRÁLOVÉHRADECKÉHO KRAJE SE SÍDLEM V HRADCI KRÁLOVÉ* [online]. [cit. 2019-09-18]. Dostupné z: http://www.khshk.cz/e-learning/kurs1b/kapitola_32__pklady_na_vyuit_hodnocen_diagnostickch_test.html

[6] Bayesova věta. In: *Wikipedia: the free encyclopedia* [online]. San Francisco (CA): Wikimedia Foundation, 2001- [cit. 2019-09-18]. Dostupné z: https://www.wikiskripta.eu/w/Bayesova_věta

[7] Výskyt zhoubných nádorů prsu v ČR. *Linkos* [online]. [cit. 2019-09-26]. Dostupné z: https://www.linkos.cz/pacient-a-rodina/onkologicke-diagnozy/nadory-prsu-c50/vyskyt-zhoubnych-nadoru-prsu-v-cr/

Obrázek

In: *U lékaře* [online]. 2008 [cit. 2019-09-18]. Dostupné z: https://www.ulekare.cz/clanek/cukrovka-diabetes-mellitus-927