

# The Prosecutor's Fallacy

1. november 2022

Tožilčeva zmota se pogosto pojavlja v epidemiologiji, vendar jo pogosto ne prepoznajo, deloma zato, ker preiskovalci nimajo močne intuicije o tem, kaj zmota sploh pomeni.

Tožilčeva zmota je dobro znana statistična zmota, ki izhaja iz napačnega razumevanja pogojnih verjetnosti in vprašanj večkratnega testiranja. Tu se osredotočamo na vidik pogojne verjetnosti tožilčeve zmote, pri kateri se domneva, da je verjetnost  $A$  glede na  $B$  enaka verjetnosti  $B$  glede na  $A$ :

$$\mathbb{P}(A \mid B) = \mathbb{P}(B \mid A).$$

Je posledica napačnega razumevanja Bayesovega izreka.

Klasičen primer tožilčeve zmote se pojavi, ko tožilec trdi, da

1. če bi bil obtoženec kriv, bi bila verjetnost dokazov (naprimer ujemanje DNK) velika;
2. zato mora biti, glede na dokaze, ki so pri roki, verjetnost obtoženčeve krivde velika.

Zmota nastane pri uporabi dokazov in statistike nasploh na sodiščih.

Zmota tožilca se pogosto pojavi v epidemiologiji in javnem zdravju, na primer tveganje za Downov sindrom narašča s starostjo matere, pri čemer je tveganje 17 - krat večje pri otrocih, rojenih ženskam, starim 40 let ali več, v primerjavi z otroki, rojenim ženskam, mlajšim od 30 let. Tako lahko domnevamo, da je večina žensk, ki rodi otroka z Downovim sindromom, starejših od 40 let; toda raziskava Albermana in Berryja ugotavljata, da se 51% vseh otrok z Downovim sindromom rodi ženskam mlajšim od 30 let, predvsem zato, ker te rojevajo več.

Zmota se tudi pogosto pojavi pri razlagi analitičnih rezultatov v epidemiologiji. Na primer, spletna stran Slate.com je objavil izjavo, da je imel eden od sedmih otrok s fetanlnim alkoholnim sindromom(FAS) mamo, ki je v prvem tromesečju nosečnosti spila eno do osem pijač na teden; to pomeni:

$$\mathbb{P}(\text{mama je spila 1-8 pijač na teden} \mid FAS) = \frac{1}{7}.$$

Toda ta količina nam malo pove o tveganju za FAS glede na določeno raven pitja matere ali

$$\mathbb{P}(FAS \mid \text{mama je spila 1-8 pijač na teden}).$$

Zmota se pojavi tudi v epidemiološki literaturi pri zamenjavi občutljivosti (verjetnost pozitivnega testa glede na resnični status bolezni) in pozitivne napovedne vrednosti (verjetnost resničnega statusa bolezni glede na pozitiven test).

## 1 Vizualizacija Tožilčeve zmote

Slika 1 prikazuje populacijo 100 moških, starih 50 let, ki se ne zdravijo zaradi hipertenzije in imajo skupni holesterol 235 mg/dl in krvni tlak 120 mmHg. Pričakuje se, da bo 9% teh moških (predstavljeno z devetimi črtastimi kvadrati) čez 10 let imelo miokardni infarkt(MI). Ena četrtnina moških je kadilcev (predstavljeno s 25 sivimi kvadrati); približno 16% naj bi bilemo MI v 10 letih; zato bo  $\frac{4}{25}$  kadilcev imelo MI (predstavljeno s črtastimi in sivimi kvadrati).

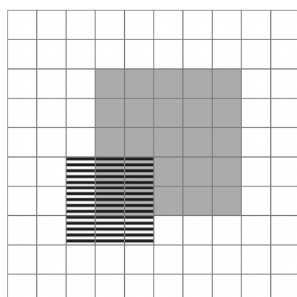
Število kadilcev, ki imajo MI, je očitno enako številu ljudi, ki imajo MI in so kadilci. Toda ali je delež kadilcev, ki imajo MI, enak deležu ljudi z MI, ki so kadilci? Iz slike 1 je odgovor očitno ne; verjetnost, da ste kadilec glede na to, da ste imeli MI, torej:

$$\mathbb{P}(\text{MI} \mid \text{kadilec}) \neq \mathbb{P}(\text{kadilec} \mid \text{MI}), \quad \text{ker}$$

$$\mathbb{P}(\text{MI} \mid \text{kadilec}) = \mathbb{P}(\text{črtasto} \mid \text{sivo}) = \frac{4}{25} = 0,16 \quad \text{in}$$

$$\mathbb{P}(\text{kadilec} \mid \text{MI}) = \mathbb{P}(\text{sivo} \mid \text{črtasto}) = \frac{4}{9} = 0,44.$$

Vizualno opazimo, da delež sivih kvadratov, ki se prekrivajo s črtastimi kvadrati, ni enak deležu črtastih kvadratov, ki se prekrivajo s sivimi kvadrati.



Slika 1: Populacija, predstavljena s 100 kvadrati, z 9 črtastimi, 25 sivimi in 4 črtastimi in sivimi

Slika 2 prikazuje izjeme pri tožilčevi zmoti, predstavljeni na sliki 1. Sedaj imamo 16 črtastih kvadratov, 16 sivih in 4 pikčaste in sive kvadrate, ker je splošna razširjenost črtastih in sivih kvadratov enaka, je

$$\mathbb{P}(\text{črtasto} \mid \text{sivo}) = \mathbb{P}(\text{sivo} \mid \text{črtasto}).$$

Tudi

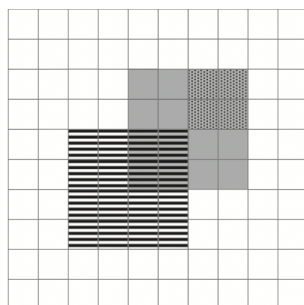
$$\mathbb{P}(\text{črtasto} \mid \text{pikčasto}) = \mathbb{P}(\text{pikčasto} \mid \text{črtasto}),$$

saj sta obe verjetnosti enaki 0. Oboje (podobno velike populacija in populacije brez prekrivanja) je ozka izjema zmote.

Po drugi strani pa so črtkani kvadrati na sliki 2 v celoti zajeti v sivih kvadratih; tako je

$$\mathbb{P}(\text{sivo} \mid \text{črtkano}) = 1, \quad \text{medtem ko je} \quad \mathbb{P}(\text{črtkano} \mid \text{sivo}) = 0,25.$$

Tožilčeva zmota velja, kadar je ena skupina podmnožica druge.



Slika 2: Populacija, predstavljena s 100 kvadrati, z 16 črtastimi, 16 sivimi in 4 pikčastimi in sivimi