

# STATISTIKA V KAZENSKEM PRAVU

## 1. BAYESOVA STATISTIKA

Bayesova statistika je statistična veja, ki nam s pomočjo matematičnih pristopov omogoča uporabo verjetnosti pri reševanju statističnih problemov. V svoje modele vključuje pogojno verjetnost, katero izračunamo z uporabo Bayesovega pravila.

Zlasti Bayesovo sklepanje razlaga verjetnost kot merilo verjetnosti ali zaupanja, ki ga lahko ima posameznik glede nastanka določenega dogodka. O nekem dogodku lahko že imamo predhodno prepričanje oziroma apriorno prepričanje, ki pa se lahko spremeni, ko se pojavijo novi dokazi. Bayesova statistika nam daje matematične modele za vključevanje naših apriornih prepričanj in dokazov za ustvarjenje novih prepričanj oziroma za pridobitev aposteriornega prepričanja, ki se lahko uporabi za kasnejše odločitve.

**1.1. Bayesovo pravilo.** Bayesovo sklepanje temelji na Bayesovim pravilom, ki izraža verjetnost nekega dogodka z verjetnostjo dveh dogodkov in obrnejnje pogojne verjetnosti. Pogojna verjetnost predstavlja verjetnost dogodka, glede na drug dogodek.

S pomočjo formule za pogojno verjetnost dobimo Bayesovo pravilo:

$$(1) \quad P(H|E) = \frac{P(E|H) \times P(H)}{P(E)},$$

verjetnost dogodka E lahko še razpišemo in dobimo:

$$(2) \quad P(H|E) = \frac{P(E|H) \times P(H)}{P(E|H)P(H) + P(E|\neg H)P(\neg H)}.$$

Obstaja še ena formulacija Bayesovega pravila, ki olajša izračune in je pogosto uporabljena pri Bayesovi analizi DNK dokazov:

$$(3) \quad \frac{P(H|E)}{P(\neg H|E)} = \frac{P(E|H)}{P(E|\neg H)} \times \frac{P(H)}{P(\neg H)}.$$

## 2. BAYESOVA ANALIZA

V kazenskih zadevah želimo vedeti, ali je obtoženec kriv ali ne. Če imamo torej na voljo dokaz  $E$ , nas zanima pogojna verjetnost  $P(kriv|E)$ , pri čemer nam je lahko v pomoč Bayesovo pravilo. To v teoriji drži, čeprav je v praksi izračun verjetnostne krivde lahko preveč zapleten. Ampak z Bayesovim pravilom lahko ocenimo verjetnosti vmesnih trditev oziroma dokazov, ki so ključnega pomena za ugotavljanje obtoženčeve krivde. Najbolj pogosta uporaba Bayesovega pravila je pri ugotavljanju, ali je obtoženec vir sledi DNK-ja s kraja zločina.

## 2.1. Poenostavljena Bayesova analiza. Naj bo:

$S$  ... trditev, da je obtoženec vir sledi DNK s kraja zločina;

$M$  ... trditev, da se obtoženec DNK ujema z DNK-jem s kraja zločina;

$f$  ... funkcija pogostosti ujemanja DNK z DNK-jem s kraja zločina.

Zelimo vedeti, kakšna je verjetnost  $S$  glede na  $M$ , tj.  $P(S|M)$ .

Bayesovo pravilo lahko uporabimo na naslednji način:

$$\frac{P(S|M)}{P(\neg S|M)} = \frac{P(M|S)}{P(M|\neg S)} \times \frac{P(S)}{P(\neg S)}.$$

Verjetnosti  $P(S)$  in  $P(\neg S)$  je težko oceniti, ker ne vemo kakšna je množica osumljencev. Smiselno bi bilo, da zanju upoštevamo interval predhodnih verjetnosti in ocenimo njihov vpliv na verjetnost trditve  $S$  in njene negacije. Nato moramo določiti vrednost  $P(M|S)$ , ki je običajno enaka ena - če bi obtoženec dejansko pustil sledove, bi laboratorijske analize pokazale ujemanje (to imenujemo lažno negativni rezultat); to je sicer poenostavitev, saj se lahko zgodi, da analize ne pokažejo ujemanja, čeprav je obtoženec pustil sledi. Potrebujemo še verjetnost  $P(M|\neg S)$  (verjetnost, da se bo našlo ujemanje, če obtoženec ni vir sledi na kraju zločina). To je običajno enakovredno pogostosti ujemanja DNK-ja z DNK-jem s kraja zločina (tj.  $f$ ); tudi to je poenostavitev, saj se lahko zgodi, da obtoženec nima enakega DNK profila, vendar so laboratorijske analize pokazale, da ga ima (to imenujemo lažno pozitivni rezultat). Sledi:

$$\frac{P(S|M)}{P(\neg S|M)} = \frac{1}{f} \times \frac{P(S)}{P(\neg S)}.$$

Ker poenostavljena Bayesova analiza ne upošteva možnosti lažno pozitivne in negativne laboratorijske analize si pogledajmo še izpopolnjeno Bayesovo analizo.

## 2.2. Izpopolnjena Bayesova analiza. Za upoštevanje možnosti laboratorijskih napak, bomo namesto $M$ uvedli spremenljivko $M_p$ .

$M_p$  ... poročano ujemanje laboratorijske analize;

$M_t$  ... trditev, da obstaja dejansko ujemanje v DNK-ju;

$\neg M_t$  ... teditev, da obstaja neujemanje v DNK-ju.

Sledi:

$$P(M_p|\neg S) = P(M_p|M_t)P(M_t|\neg S) + P(M_p|\neg M_t)P(\neg M_t|\neg S).$$

Sedaj je  $P(M_t|\neg S)$  enako  $f$  in zato  $P(\neg M_t|\neg S)$  enako  $1-f$ .  $P(M_p|\neg M_t)$  opisuje verjetnost lažno pozitivnih rezultatov laboratorija (oznaka  $FP$ ) in  $P(M_p|M_t)$  verjetnost resničnih pozitivnih rezultatov laboratorija (oznaka  $FN$ ). Sledi:

$$P(M|\neg S) = [(1 - FN) \times f] + [FP \times (1 - f)].$$

Formula pokaže, da za parvilno oceno verjetnosti  $P(M_p|\neg S)$  potrebujemo statistično oceno pogostosti profila DNK in stopnje napak laboratorijskih analiz, ki pa so redko na voljo.

Druga poenostavitev je, da predpostavimo, da je  $P(M_p|S) = 1$ , pri čemer ni upoštevana možnost lažnega negativnega rezultata. Kot zgoraj, imamo:

$$P(M_p|S) = P(M_p|M_t)P(M_t|S) + P(M_p|\neg M_t)P(\neg M_t|S).$$

Če je  $P(M_p|S) = 1$ , je  $P(\neg M_p|S) = 0$  in  $P(M_p|S) = 1 - FN$ . Potem sledi, da je:

$$\frac{P(M_p|S)}{P(M_p|\neg S)} = \frac{1 - FN}{[(1 - FN) \times f] + [FN \times (1 - f)]}.$$

### 3. DRUGI PRISTOPI

Da bi ocenila moč Bayesove analize dokazov DNK, jo bom primerjala z nekaterimi drugimi pristopi kot so frekvence in edinstvenost ter naključno ujemanje in razmerja verjetnosti.

**3.1. Frekvence in edinstvenost.** Predlagano je bilo s strani mnogih avtorjev, da je bolj naraven način za obravnavo verjetnosti uporaba naravnih frekvenc (angl. natural frequencies). Recimo, da je pogostost profila DNK  $f$  1 proti 10 milijonov, predpostavimo tudi, da ima obtoženec enak DNK in da je začetna populacija osumljencev 100 milijonov ljudi. Zanima nas kakšna je verjetnost, da je obtoženec vir DNK-ja s kraja zločina.

Naj bo:

$f$  ... pogostost profila DNK;

$m$  ... velikost populacije osumljencev;

$n$  ... število ljudi, ki imajo ustrezen DNK profil.

Po metodah, ki temeljijo na naravnih frekvencah, je potrebno izračunati, koliko ljudi z zadevnim profilom DNK je v populaciji osumljencev, tako da se  $f$  pomnoži z  $m$ . Če je posameznikov s takim profilom  $n > 1$ , je verjetnost, da je obtoženi vir,  $\frac{1}{n}$ . V primeru, ko je  $n < 1$  in so frekvence še posebej majhne, je bolje raziskovati ali je profil DNK edinstven ali ne - ali poleg obtoženca obstajajo še drugi posamezniki z enakim profilom DNK. Temu pravimo metoda edinstvenosti (angl. uniqueness method).

S formulo binomske porazdelitve lahko izračunamo verjetnost, da se bo dogodek  $X$ , na primer profil DNK, pojavil  $k$  - krat v  $s$  - kratnem številu ponovitev, pri čemer ima dogodek  $X$  frekvenco  $4f$ . Želimo vedeti, kolikšna je verjetnost, da ima točno en posameznik ustrezen DNK profil, ob pogoju da ga ima vsaj en posameznik, torej:

$$P(n = 1 | n \geq 1) = \frac{P(n = 1 \cap n \geq 1)}{P(n \geq 1)} = \frac{m \times f \times (1 - f)^{m-1}}{1 - (1 - f)^m}.$$

**3.2. Naključno ujemanje in razmerja verjetnosti.** Metoda verjetnost naključnega ujemanja (angl. random match probability) izraža možnost, da bi imel naključni posameznik, ki ni povezan z obdolžencem, ustrezen DNK profil. Ta verjetnost je enaka pogostosti profila DNK. Težava tega pristopa je, da verjetnost naključnega ujemanja lahko predstavljena oziroma razumevana narobe.

Pogosto se to verjetnost interpretira na slednji način:

- (1) če je verjetnost naključnega ujemanja na primer 1 proti 100 milijonom, potem je verjetnost, da ima profil DNK drug posameznik in ne obdolženec 1 proti 100 milijonom;
- (2) ker je to zelo majhna verjetnost, mora biti tudi verjetnost, da je sled DNK pustil nekdo drug na kraju zločina in ne obdolženec, zelo majhna;
- (3) zato mora biti verjetnost, da je vir sledi DNK s kraja zločina obtoženec zelo velika, ampak znaša 1 proti 100 milijonov.

Takšno sklepanje je napačno in je znano kot tožilčeva zmota(angl. Prosecutor's fallacy). Sestavlja jo enačba

$$(4) \quad 1 - f = P(S|M).$$

Zmota se pojavi v koraku (2), ko je zamenjano  $P(M|\neg S)$  s  $P(\neg S|M)$  in predpostavljeno, da sta obe verjetnosti enaki  $f$ .

Namesto verjetnosti naključnega ujemanja forenzični strokovnjaki pogosto pričajo o razmerju verjetnosti(angl. likelihood ratio) dokazov DNK, in sicer kot:

$$P(M|S) = P(M|\neg S).$$

Bayesovo pravilo vključuje razmerje verjetnosti in predhodno oziroma pariorno verjetnost, torej je celovitejši način predstavitve dokazov DNK. Težavo ocenjevanja apriorne verjetnosti pri Bayesovem pravilu, bi lahko odpravili z osredotočanjem le na verjetnost. Metoda razmerja verjetnosti je koristna v državah kot sta na primer Velika Britanija in Združene države Amerike, kjer se lahko Bayesovo pravilo šteje kot poseg v pravico porote do previdnosti: poroti naj ne bi bilo potrebno govoriti, kako naj razmišlja in presoja dokaze; Bayesovo pravilo pa je natanko metoda za tehtanje dokazov.

#### 4. ZMOTA

Ker večina ljudi pri razmišljanju o verjetnosti dela osnovne napake, obstaja mnogo zmot, ki izhajajo iz osnovnega razumevanja pravil teorije verjetnosti. Številne od teh zmot so zlasti posledica napačnega razumevanja pogojne verjetnosti. Bolj znana primera takih zmot sta Tožilčeva zmota(angl. Prosecutor's fallacy) in Zmota obrambnega odvetnika(angl. Defense attorney's fallacy). Na kratko je Tožilčeva zmota domneva, da je  $P(H|E)$  enaka  $P(E|H)$ , kjer je  $H$  predpostavka, da se najdejo dokazi o obtožencu,  $E$  pa predpostavka, da je obtoženec nedolžen. Zmota obrambnega odvetnika pa je, da ne upoštevamo velike spremembne verjetnosti v korist obtoženčeve krivde.

Manj znana zmota je tudi Zasliševalčeva zmota(angl. Interrogator's fallacy), ki razloži zakaj priznanja, pridobljena med zaslišanjem, ponujajo dvomljive dokaze v prid krivdi oziroma dokazi med sodbo ne morejo zmanjšati verjetnosti krivde.

**4.1. Tožilčeva zmota.** Tožilčeva zmota(angl. Prosecutor's fallacy) se pogosto pojavlja v kazenskem pravu, vendar jo pogosto neprepoznajo, 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. Napaka temelji na predpostavki, da je  $P(H|E) = P(E|H)$ , pri čemer  $H$  predstavlja primer, da se najdejo dokazi o obtožencu,  $E$  pa primer, da je obtoženec nedolžen. Vendar ta enakost ne drži: čeprav je  $P(H|E)$  običajno zelo majhen, je lahko  $P(E|H)$  še vedno veliko večji.

Za lažjo predstavo si oglejmo primer. Na primer, da ima storilec zločina enako krvno skupino kot obtoženec in da ima 10% prebivalstva enako krvno skupino. Potem je lahko na podlagi tega verjetnost, da je obtoženec kriv, 90 odstotna. Vendar je ta sklep skoraj pravilen le, če je bil obtoženec izbran kot glavni osumljenec na podlagi trdnih dokazov, ki so bili odkriti pred krvnim testom in z njim niso povezani, saj je lahko ujemanje krvi popolno naključje. V nasprotnem primeru je predstavljena utemeljitev napačna, saj ne upošteva predhodne verjetnosti, da gre za naključno nedolžno osebo. Denimo, da v mestu, kjer se je zgodil zločin, živi 1000 ljudi. To pomeni, da tam živi 100 ljudi, ki imajo krvno skupino storilca, zato je verjetnost, da je obtoženec kriv - na podlagi dejstva, da se njegova krvna skupina ujema s krvno skupino morilca - le 1%, kar je veliko manj kot 90%.

## 5. ZMOTA OBRAMBNEGA ODVETNIKA

Zmota zagovornika oziroma zmota obrambnega odvetnika (angl. Defense attorney's fallacy) se pojavi, ko se poroča o tem, koliko ljudi z določeno značilnostjo, se pojavi v določeni populaciji. Predpostavlja se, da je storilec del neke poljubno velike populacije in da ni na voljo drugih informacij, torej je za vse enako verjetno, da so storilci. Na podlagi teh predpostavk lahko sklepamo, da obstaja majhna verjetnost, da je osumljenec storilec kaznivega dejanja.