RAZMERJE VERJETNOSTI

NEŽA KRŽAN

1. Koncept verjetnosti

Definicija 1.1. Naj bo H nek dogodek in $\bar{H}$ negacija oziroma komplement dogodka H. Dogodka H in $\bar{H}$ sta znana kot komplementarna dogodka.

Pogosto se opravlja primerjava verjetnosti dokazov na podlagi dvah konkurenčnih predlogov, in sicer predloga tožilca in predloga obrambe.
$H_p \dots$ trditev, ki jo predlaga tožilstvo;
$H_d \dots$ trditev, ki jo predlaga obramba;
Hipoteze se lahko dopolnjujejo na enak način kot dogodki - ena in samo ena je lahko resnična, med seboj se izključujejo. Ni nujno, da so izbrane tako, da zajemajo vse možne razlage dokazov. Dve hipotezi lahko označujeta komplementarne dogodke(npr. resnično kriv in resnično nedolžen), vendar pa se lahko zgodi, da se označena dogodka ne dopolnjujeta.

Koncept verjetnosti je pomemben pri ocenjevanju dokazov, saj se le ti ocenjujejo glede na njihov vpliv na verjetnost določene domneve o interesni osebi(v nadaljevanju PoI) (preden pride do sojenja) ali obdolžencu(medtem, ko sojenje poteka). Zanima nas vpliv dokazov na verjetnost krivde(H_p) in nedolžnosti(H_d) osumljenca. Gre za dopolnjujoča se dogodka in razmerje verjetnosti teh dveh dogodkov,

$$\frac{P(H_p)}{P(H_d)}, \tag{1}$$

je verjetnost proti nedolžnosti ali verjetnost za krivdo. Ob upoštevanju dodatnih informacij E(oz. dokazov), je razmerje

$$\frac{P(H_p|E)}{P(H_d|E)}, \tag{2}$$

verjetnost v prid krivdi ob upoštevanju informacij E.

Občasno se zgodi, da predloga tožilstva in obrambe nista komplementarna in v takih primerih ni mogoče določiti $P(H_p)$ ali $P(H_d)$, ampak samo vpliv statistike, znane kot razmerje verjetnosti(angl. Likelihood Ratio; oznaka LR).

2. Razmerje verjetnosti(angl. Likelihood Ratio)

2.1. **Opredelitev.** V Bayesovi formuli nadomestimo H z $\bar{H}$ in enakovredna različica Bayesovega izreka je

$$P(\bar{H}|E) = \frac{P(E|\bar{H})P(\bar{H})}{P(E)}, \tag{3}$$

kjer $P(E) \neq 0$.

Če prvo enačbo delimo z drugo dobimo verjetnostno obliko Bayesovega izreka

$$\frac{P(H|E)}{P(\bar{H}|E)} = \frac{P(E|H)}{P(E|\bar{H})} \times \frac{P(H)}{P(\bar{H})}. \tag{4}$$

Leva stran je verjetnost dogodka H ob pogoju, da se je zgodil dogodek E. Pogojna verjetnost na desni strani dogodka, H in $\bar{H}$, sta v števcu in imenovalcu različna, medtem ko je dogodek E, katerega verjetnost nas zanima, enak. Na koncu pa imamo verjetnost v korist dogodka H brez kakršnih koli informacij o E.

Definicija 2.1. Razmerje

$$\frac{P(E|H)}{P(E|\bar{H})} \tag{5}$$

se imenuje razmerje verjetnosti(angl. Likelihood Ratio).

Oglejmo si dogodka E in H, ter njuni dopolnitvi. Razmerje verjetnosti je tu razmerje verjetnosti E, ko je H resničen in verjetnosti E, ko je H neresničen. Da bi upoštevali učinek E na verjetnost H, tj. da bi

$$\frac{P(H)}{P(\bar{H})}$$

spremenili v

$$\frac{P(H|E)}{P(\bar{H}|E)},$$

prvo pomnožimo z razmerjem verjetnosti. Verjetnost

$$\frac{P(H)}{P(\bar{H})}$$

je znana kot predhodna verjetnost v korist H, verjetnost

$$\frac{P(H|E)}{P(\bar{H}|E)}$$

pa je znana kot posteriorna verjetnost v korist H. Razlika med $P(E|H)$ in $P(H|E)$ je bistvena.Pri preučevanju vpliva E na H je treba upoštevati tako verjetnost E, ko je H resničen in ko je H neresničen. Pogosta napaka(zmota prenesene pogojne verjetnosti) je, da dogodek E, ki je malo verjeten, če je $\bar{H}$ resničen, pomeni dokaz v prid H. Da bi bilo tako, je treba dodatno zagotoviti, da E ni tako malo verjeten, če je H resničen. Razmerje verjetnosti je potem večje od 1 in pozitivna verjetnost je večja od predhodne verjetnosti.

2.2. **Razmerje verjetnosti v kazenskem pravu.** Obravnavajmo obliko Bayesovega izreka o verjetnosti v forenzičnem kontekstu ocenjevanja vrednosti nekaterih dokazov. Naj bo:
$H_p \ldots$ interesna oseba(PoI) oz. obtoženec je resnično kriv - nadomestimo H;
$H_d \ldots$ interesna oseba(PoI) je resnično nedolžen - nadomestimo $\bar{H}$;
$Ev \ldots$ obravnavani dokaz - nadomestimo dogodek E;

Oblika Bayesovega izreka nato omogoča, da se prehdodne verjetnosti(tj, pred predstavitvijo Ev) v korist krivde posodobijo v posteriorne verjetnosti ob upoštevanju Ev, na naslednji način:

$$\frac{P(H_p|Ev)}{P(H_d|Ev)} = \frac{P(Ev|H_p)}{P(Ev|H_d)} \times \frac{P(H_p)}{P(H_d)}.$$

Ob upoštevanju informacij o ozadju I, dobimo zapis

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

Pri vrednotenju dokazov Ev sta potrebni dve verjetnosti - verjetnost dokazov, če je PoI kriv in glede na informacije o ozadju, ter verjetnost dokazov, če je PoI nedolžen in glede na informacije o ozadju. Informacije o ozadju so včasih znane kot okvir okoliščin ali pogojne informacije.

Da lahko ocenimo oziroma določimo vrednost dokaza potrebujemo razmerje verjetnosti(angl. Likelihood ratio).

Definicija 2.2. Naj bosta H_p in H_d dve konkurenčni hipotezi ter I informacije o ozadju. Vrednost V dokaza Ev je podana z

$$V = \frac{P(Ev|H_p,I)}{P(Ev|H_d,I)},$$

razmerje verjetnosti, ki pretvori predhodne verjetnosti

$$\frac{P(H_p|I)}{P(H_d|I)}$$

v posteriorne verjetnosti

$$\frac{P(H_p|Ev,I)}{P(H_d|Ev,I)}.$$

Tabela 1. Kvalitativna lestvica za poročanje o vrednosti V podpore dokazov za H_p proti H_d(Vir: ENFSI, 2015).

1	$< V \leq$	2	brez podpore
2	$< V \leq$	10	šibka podpora prvi hipotezi
10	$< V \leq$	100	zmerna podpora prvi hipotezi
100	$< V \leq$	1000	srednje močna podpora prvi hipotezi
1000	$< V \leq$	10000	močna podpora prvi hipotezi
10000	$< V \leq$	1000000	zelo močna podpora prvi hipotezi
1000000	$< V$		izjemno močna podpora prvi hipotezi

2.3. **Utemeljitev uporabe razmerja verjetnosti.** Verjetnostna oblika Bayesovega izreka predstavlja prepričljiv intuitivni argument za uporabo razmerja verjetnosti kot merila vrednosti dokazov. Obstaja tudi matematični argument, ki upravičuje njegovo uporabo.

Želimo izmeriti vrednost V dokazov E v prid krivdi H_p. Pri tem bo obstajala odvisnost od osnovnih informacij I, vendar ta ne bo izrecno navedena. Predpostavlja se, da je ta vrednost V odvisna samo od verjetnosti E ob pogoju, da je PoI kriv(H_p), in od verjetnosti E ob pogoju, da je PoI nedolžen(H_d). Naj bo $x = P(E|H_p)$ in $y = P(E|H_d)$. Zgornja predpostavka pravi, da je $V = f(x, y)$ za neko funkcijo f. Vzemimo še en dokaz T, ki je neodvisen od E in H_p (in s tem H_d) in je tak, da je $P(T) = \theta$. Nato

$$P(E, T|H_p) = \tag{6}$$
$$= P(E|H_p)P(T|H_p) \tag{7}$$
$$= P(E|H_p)P(T) \tag{8}$$
$$= \theta x, \tag{9}$$

pri čemer iz (6) v (7) upoštevamo neodvisnost dokaza E in T ter iz (7) v (8) vrstico neodvisnost dokaza T in H_p. podobno

$$P(E, T|H_d) = \theta y.$$

Vrednost kombiniranih dokazov (E, T) je enaka vrednosti E, saj je bil T predpostavljen kot nepomemben. Vrednost (E, T) je $f(\theta x, \theta y)$, vrednost $E = V = f(x, y)$. Tako je $f(\theta x, \theta y) = f(x, y)$ za vsako θ v intervalu $[0, 1]$ možnih vrednosti $P(T)$. Razmerje med x in y v funkciji f ima lahko eno od štirih oblik, odvisno od štirih matematičnih operatorjev $+$, $\times$, - in $/$. Če pogledamo $\frac{x}{y}$ sledi

$$f(x, y) = f(\frac{x}{y})f(\theta x, \theta y) = f(\frac{\theta x}{\theta y}) = f(\frac{x}{y}).$$

To je enako $f(x, y)$ za vsako θ v intervalu $[0, 1]$. Iz tega sledi, da je f funkcija $\frac{x}{y}$ in torej, da je V funkcija

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

kar je razmerje verjetnosti.

Verjetnost hipoteze H na podlagi nekega dokaza E je verjetnost, da najdemo E, če je H resnična. Za alternativno hipotezo je LR razmerje obeh verjetnosti. LR nam pove, katera hipoteza je bolje podprta z dokazi. Kadar sta hipotezi medsebojno izključujoči in izčrpni, nam LR pove več. V tem primeru, če je verjetnost H večja od verjetnosti alternative, lahko sklepamo tudi, da se verjetnost H zaradi najdbe E poveča, medtem ko se verjetnost alternative zmanjša. Če je le mogoče, je treba upoštevati verjetnosti za vse razumne alternativne hipoteze (tako da je nabor hipotez izčrpen). Če se obravnavajo samo nekatere hipoteze, je treba pojasniti, da so predstavljene samo LR za pare teh hipotez. V primerih, ko je treba združiti več hipotez in/ali več dokazov, se lahko LR bolje uporablja v povezavi z drugimi metodami.
Kadar je treba količinsko ovrednotiti skupni učinek več dokazov, ki vključujejo različne povezane hipoteze (kot so hipoteze o ravni vira, ravni dejavnosti in ravni kaznivega dejanja), poenostavljene rešitve, ki neupravičeno predpostavljajo neodvisnost, niso ustrezne. Grafični prikazi dokazov so lahko v veliko pomoč pri modeliranju odvisnosti. Obstaja interaktivna programska oprema za izvajanje izračunov na grafičnih modelih(Bayesovih mrežah), ki uporabnikom omogoča, da raziščejo vplive različnih predpostavk. Čeprav je takšne metode težko uvesti neposredno na sodišču, so koristne za sintezo dokazov v kateri koli fazi preiskave pred sojenjem.

Ocena vrednosti razmerja verjetnosti je lahko podvržena številnim virom negotovosti, vključno s kakovostjo podatkov, pridobljenih z analizami, ki jih opravijo forenzični znanstveniki, izbiro kontrolnega vzorca in najdenih predmetov, ki jih lahko vzamejo različni preiskovalci ali analizirajo različni analitiki ali laboratoriji. Ocena znanstvenih dokazov na sodišču pogosto zahteva kombinacijo podatkov o pojavu ciljnih značilnosti skupaj z osebnim poznavanjem okoliščin iz določenega primera. Jasno je, da ima vsaka ocena verjetnosti, ki se nanaša na določen primer, tudi če jo obravnavamo v obliki frekvence, sestavino, ki temelji na osebnem znanju. Drugi viri negotovosti vključujejo pridobivanje predhodnih verjetnosti, pogojenih z razpoložljivim znanjem, ali celo izvajanje numeričnih postopkov za razreševanje računskih težav. Poročilo o vrednosti razmerja verjetnosti vključuje merilo njegove natančnosti, na primer z navedbo številčnega razpona vrednosti za verjetnost dokazov na podlagi konkurenčnih predlogov in s tem številčnega razpona vrednosti za razmerje verjetnosti. Vendar sta vrednost dokaza in moč posameznikovega prepričanja o vrednosti različna pojma in se ne smeta združevati v intervalu ali povzročiti spremembe vrednosti dokaza, kot se to na primer zgodi z navedbo spodnje meje neke poljubno izbrane ravni. V praksi je za kriminalistično preiskavo na voljo en niz podatkov o ozadju, ki so značilni za člane določene relevantne populacije, en niz kontrolnih podatkov in en niz izterjanih podatkov. Zato je za vrednotenje dokazov z določenim statističnim modelom na voljo ena sama vrednost V za povezano razmerje verjetnosti. Ponovno je treba upati, da so vsi različni kontrolni vzorci in pridobljeni podatki dovolj reprezentativni za populacije, iz katerih so bili izbrani, tako da se bodo razmerja verjetnosti po vrednosti le malo razlikovala.

2.4. **Bayesov faktor(angl. Bayes' Factor) in razmerje verjetnosti.** V forenziki se ta dva pojma, kljub pogostejši uporabi Bayesovega faktorja(BF), pogosto obravnavata kot sinonima. Bayesov faktor je glavni element Bayesove metodologije za primerjavo konkurenčnih predlogov. Opredeljen je kot sprememba, ki jo povzročijo novi dokazi (podatki) v verjetnosti pri prehodu od predhodne k posteriorni porazdelitvi v korist enega predloga k drugemu. Da se pokazati, da je razmerje verjetnosti poseben primer Bayesovega faktorja, kadar so konkurenčne hipoteze parametrizirane z enim samim parametrom (tj. preprosta hipoteza). Vendar pa lahko pride do primerov, ko se primerjajo sestavljene hipoteze. V takem primeru je Bayesov faktor razmerje dveh mejnih verjetnosti pri konkurenčnih hipotezah in se zdi, da ni več odvisen samo od podatkov.