

# Statistika v kazenskem pravu

## Dolga predstavitev diplomske naloge

Neža Kržan

Mentor: izred. prof. dr. Jaka Smrekar  
Fakulteta za matematiko in fiziko

Ljubljana, 5.maj 2023

# Statistika v kazenskem pravu

- Preverjamo **teorije** in **hipoteze**.
- Preučujemo razmerja med dvema ali več spremenljivkami.

**Odvisne spremenljivke** - empirični dogodki, ki jih želi raziskovalec pojasniti.

**Neodvisne spremenljivke** - dejavniki, za katere raziskovalec meni, da bi lahko vplivali na odvisne spremenljivke.

- 1 časovno zaporedje;
- 2 obstajati mora empirična povezava med odvisno in neodvisno spremenljivko;
- 3 razmerje med neodvisno in odvisno spremenljivko nepristransko;

# Težave

Določitev odvisnih in neodvisnih spremenljivk za modeliranje.

V proces določanja spremenljivk pogosto posežejo odvetniki, ki se sklicujejo na pravne zakone in načela. To lahko postane sporno, saj lahko takšni posegi ovirajo statistične znanstvenike pri izračunu verjetnostnega vpliva spremenljivk.

# Uporaba statistike pri pravnem postopku

Pred pričanjem na sodišču moramo vedeti

- 1 na kaj točno se podatki nanašajo,
- 2 kako so bili zbrani,
- 3 kakšen del podatkov manjka ali je neuporaben.

⇒ **Določimo ustrezen postopek analize podatkov.**

Potrebujemo

- 1 osnovne informacije odvetnika in drugih strokovnjakov,
- 2 določitev ustrezne populacije,
- 3 določitev parametrov, ki nas zanimajo,
- 4 določitev statističnega postopka.

⇒ **Oblikujemo ustrezne primerjalne skupine.**

# Raziskovalni proces

Raziskovalni proces v kazenskem pravosodju je običajno namenjen preučevanju problemov kriminala. Proces se izvaja po naslednjih točkah.

- 1 **Identifikacija problema.**
- 2 **Zasnova raziskave.**
- 3 **Analiza podatkov.**

# Vrednotenje dokazov

## Dokazni standard

je pravno vprašanje. Gre za abstraktno normo, ki je (podobno kot obstoj določenih predpostavk za določeno kaznivo dejanje) opredeljena s pravnim pravilom.

## **METODA DOKAZNE VREDNOSTI MODEL VERJETNOSTI HIPOTEZE**

V kazenskem pravu lahko zelo hitro pride do posebnih, edinstvenih predpostavk oziroma hipotez, ki pa predstavljajo težave pri vrednotenju oziroma merjenju v statističnih modelih.

# Koncept verjetnosti

Opravlja se primerjava verjetnosti dokazov na podlagi dveh konkurenčnih predlogov:

$H_p$  . . . trditev, ki jo predlaga tožilstvo;

$H_d$  . . . trditev, ki jo predlaga obramba.

Koncept verjetnosti je

- ključen pri ocenjevanju dokazov;
- omogoča objektivno oceno vpliva dokaza na verjetnost določene domneve o interesni osebi ali obdolžencu.

# Presoja dokazov

- Za presojo se uporabljajo različne metode in tehnike, ki temeljijo na statistični verjetnosti,
- metode omogočajo oceno, kako verjetno je, da so dokazi resnični in zanesljivi.

## KONTEKST DOKAZA

- Upoštevamo druge dokaze in okoliščine primera.
- Izvedemo bolj objektivno presojo dokazov.
- Vpliv na določeno domnevo oziroma hipotezo o interesni osebi ali obdolžencu.
- Upoštevamo verjetnost napake - verjetnost, da so dokazi napačni ali zavajajoči.



# Koncept verjetnosti

Verjetnost proti nedolžnosti ali verjetnost za krivdo

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

Verjetnost v prid krivdi ob upoštevanju informacij  $E$

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

Če imamo na voljo dokaz  $E$ , nas zanima pogojna verjetnost

$$P(kriv|E),$$

pri čemer nam je lahko v pomoč Bayesovo pravilo.

# Oprelitev

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.

Hipoteza: Oseba je »kriva«.

- 1 začnemo z nekim predhodnim prepričanjem o hipotezi;
- 2 ga posodabljam, ko se dokazi ponovno pojavijo.

# Bayesovo pravilo

Bayesovo sklepanje temelji na Bayesovem pravilu, ki izraža verjetnost nekega dogodka z verjetnostjo dveh dogodkov in obrnejnje pogojne verjetnosti.

## Definicija

*Bayesovo pravilo:*

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

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

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

# Bayesovo posodabljenje

Logična trditev, kako se sčasoma posodablajo apriorne oziroma predhodne verjetnosti dokazov glede na novo zbrane dokaze oziroma prepričanja.

## Trditev (Bayesovo posodabljanje)

*Če se dogodek  $E$  zgodi ob času  $t_1 > t_0$ , potem je  $P_1(H) = P_0(H|E)$ .*

- Predhodna oz. apriorna verjetnost =  $P_0(H)$  (ob času  $t_0$  dogodku  $H$  dodelimo verjetnost);
- Ko se zgodi dogodek  $E$  ob času  $t_1$ , ki vpliva na naša prepričanja o dogodku  $H$ , Bayesovo posodabljanje pravi, da je potrebno apriorno verjetnost dogodka  $H$  v času  $t_1$  enačiti z pogojno verjetnostjo dogodka  $H$  glede na dogodek  $E$  v času  $t_0$ .

# Bayesova teorija v kazenskem pravu

Bayesovo sklepanje razlaga verjetnost kot merilo verjetnosti ali zaupanja, ki ga lahko ima posameznik glede nastanka določenega dogodka.

- O dogodku lahko že imamo predhodno prepričanje oziroma apriorno prepričanje.
- Ta se lahko spremeni, ko se pojavijo novi dokazi.

Bayesova statistika nam daje matematične modele za vključevanje naših apriornih prepričanj in dokazov za ustvarjanje novih prepričanj.

# Bayesova teorija v kazenskem pravu

Ocena, kako informacije vplivajo na tožilčevo domnevo o obdolžencu oziroma storilcu kaznivega dejanja.

Postopek posodabljanja verjetnosti tožilčeve hipoteze na podlagi predhodnih oziroma apriornih verjetnosti.

verjetnost hipoteze pred upoštevanjem določenega dokaza (dokazov)



verjetnost hipoteze po upoštevanju določenega dokaza (dokazov)

# Predhodna verjetnost in določitev posteriorne verjetnosti

Recimo, da je statistični znanstvenik naprošen, da opravi analizo profila DNK krvi, najdene na kraju kaznivega dejanja, in rezultat primerja s profilom DNK obdolženca.

Odločitev porotnikov bo delno odvisna od njihove ocene dveh interesnih hipotez

$H_1$  ... vir krvi je obtoženec,

$H_2$  ... vir krvi je druga oseba.

Porotniki bodo morda želeli, da jim dokončno povemo, katera hipoteza je resnična, ali da jim navedemo verjetnosti vira. Za oceno verjetnosti vira mora statistični znanstvenik upoštevati tudi druge dokaze v kazenskem primeru.

# Predhodna verjetnost in določitev posteriorne verjetnosti

Obtoženec in kri s kraja zločina imata skupen niz t.i. genetskih označevalcev - najdemo pri eni osebi na 1 milijon prebivalcev.

→ **pogojna verjetnost ugotovitve rezultatov pri dveh hipotezah o medsebojni povezanosti**

- skupni genetski označevalci skoraj zagotovo najdeni v primeru  $H_1$  (vir je bil obtoženec);
- 1 možnost na milijon, da bodo najdeni v primeru  $H_2$  (vir je bil nekdo drug);

→ **predložimo razmerje verjetnosti**



# Težave z določitvijo predhodne verjetnosti

Različne metode za določitev in izračun predhodnih verjetnosti lahko dajejo različne rezultate.

Ali naj analitiki poskušajo določiti predhodne verjetnosti in če ja, kako naj jih določijo?

- **Nevtralno stanje** - analitiki predpostavi enake predhodne verjetnosti za vse hipoteze v primeru;
- Analitik naj uporabi svoje strokovno znanje za izračun apriorne verjetnosti na podlagi razpoložljivih podatkov in brez nepotrebnega vplivanja odvetnikov ali drugih udeležencev postopka;

# Poenostavljena Bayesova analiza

Naj bo:

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

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

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

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

Bayesovo pravilo lahko uporabimo na naslednji način:

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

# Poenostavljena Bayesova analiza

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

$P(M|S) = 1$  - **lažno negativni rezultat;**

$P(M|\neg S) = f$  - **lažno pozitivni rezultat;**

# Izpopolnjena Bayesova analiza I

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;

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

$f$  ... pogostost ujemnja DNK-ja z DNK-jem s kraja zločina;

$P(M_t | \neg S) = f$ ;

$P(\neg M_t | \neg S) = 1 - f$ ;

$P(M_p | \neg M_t)$  ... opisuje verjetnost lažno pozitivnih rezultatov laboratorija (oznaka  $FP$ );

$P(M_p | M_t)$  ... opisuje verjetnost resničnih pozitivnih rezultatov laboratorija (oznaka  $FN$ ).

Sledi:

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

# Izpopolnjena Bayesova analiza

- Relativno majhne stopnje napak lahko bistveno zmanjšajo dokazno vrednost DNK dokazov, saj močno zmanjšajo razmerje verjetnosti.
- Vpliv stopnje laboratorijskih napak kaže, da ne glede na to, kako nizka se izkaže pogostost profila, bo ta relativno nepomembna, če pogostosti ne spremlja ocena stopnje laboratorijskih napak.

Spremenljivost pogostosti profila lahko v Bayesovem okviru upoštevamo na dva načina:

- 1 s spremembo predhodne verjetnosti;
- 2 s spremembo pogostosti profila.

# Frekvence

- frekvence se običajno nanašajo na pojavljanje dokazov za posamezen primer;
- sklepanje o krivdi je lahko podprto s verjetnostnim sklepanjem z uporabo absolutnih ali relativnih frekvenc;

## Relativne frekvence

- navajajo (predpostavljajo) obstoj referenčnega vzorca, na podlagi katerega se lahko oceni pogostost zadevnega dogodka;
- lahko podpre vmesno sklepanje o moči dokazov;
- rutinsko vključene v znanstvene dokaze.

# Frekvence - obtoženec vir DNK-ja s kraja zločina

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.

**Kakšna je verjetnost, da je obtoženec vir DNK-ja s kraja zločina?**

$f$  ... pogostost profila DNK;

$m$  ... velikost populacije osumljencev;

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

Potrebno je izračunati, koliko ljudi z zadevnim profilom DNK je v populaciji osumljencev, tako da se  $f$  pomnoži z  $m$ .

# Frekvence - obtoženec vir DNK-ja s kraja zločina

- ①  $n \geq 1 \Rightarrow P(\text{obtoženi je vir}) = \frac{1}{n}$
- ②  $n < 1 \Rightarrow$  frekvence še posebej majhne

Verjetnost, da ima točno en posameznik ustrezen DNK profil, ob pogoju da ga ima vsaj en posameznik

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



# Primerjava z Bayesovo metodo

- **metoda edinstvenosti** - ni mogoče enostavno upoštevati številnih zapletov (npr. vpliv stopnje laboratorijskih napak);
- Bayesovo pravilo pa omogoča upoštevane različnih dejavnikov  
→ natančni izračuni in bolj zanesljivi rezultati;
- **frekvenčni modeli** - za dokaze, ki jih je mogoče uporabiti za prikaz, kako verjetno je, da bi se naključno izbrana oseba ujemala z vzorcem;

Izbira metode je odvisna od specifičnih zahtev problema.

# Metoda verjetnosti naključnega ujemanja

Izraža možnost, da bi imel naključni posameznik, ki ni povezan z obdolžencem, ustrezeni DNK profil. Ta verjetnost je enaka pogostosti profila DNK.

Pogosto se to verjetnost interpretira kot:

- 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.

# Metoda verjetnosti naključnega ujemanja

## ⇒ TOŽILČEVA ZMOTA

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

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 dokazov DNK:

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

# Opredelitev

## Definicija

*Razmerje*

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

*se imenuje **razmerje verjetnosti**.*

$\frac{P(H)}{P(\bar{H})}$  ... predhodna verjetnost v korist H;

$\frac{P(H|E)}{P(\bar{H}|E)}$  ... posteriorna verjetnost v korist H;

Torej iz Bayesovega izreka neposredno izhaja, da če je razmerje verjetnosti

- ➊ večje od 1, potem dokaz povečuje verjetnost krivde,
- ➋ če pa je manjše od 1, zmanjšuje verjetnost krivde.

# Razmerje verjetnosti v kazenskem pravu

$H_p$  ... interesna oseba(Pol) oz. obtoženec je resnično kriv - nadomestimo  $H$ ;

$H_d$  ... interesna oseba(Pol) je resnično nedolžen - nadomestimo  $\bar{H}$ ;

$Ev$  ... obravnavani dokaz - nadomestimo dogodek  $E$ ;

Bayesovov izreka omogoča, da se predhodne verjetnosti v korist krivde posodobijo v posteriorne verjetnosti

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

Ob upoštevanju informacij  $I$

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

# Razmerje verjetnosti v kazenskem pravu

## Definicija

*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*

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

# Utemeljitev uporabe razmerja verjetnosti

- 1 katera hipoteza je bolje podprta z dokazi;
- 2 v primerih, ko je treba združiti več hipotez in/ali več dokazov;
- 3 za količinsko ovrednotenje skupnega učinka več dokazov, ki vključujejo različne povezane hipoteze;

## Negotovosti

- kakovost podatkov, pridobljenih z analizami;
- izbira kontrolnega vzorca;
- osebno poznavanje okoliščin iz določenega primera;
- pridobivanje predhodnih verjetnosti, pogojenih z razpoložljivim znanjem;
- numerični postopki za razreševanje računskih težav;

⇒ **poročilo vključuje merilo njegove natančnosti**

# Bayesov faktor in razmerje verjetnosti

Bayesov faktor je glavni element Bayesove metodologije za primerjavo konkurenčnih predlogov.

- sprememba, ki jo povzročijo novi dokazi (podatki) v verjetnosti pri prehodu od predhodne k posteriorni porazdelitvi v korist enega predloga k drugemu;
- razmerje verjetnosti poseben primer Bayesovega faktorja;
- razmerje dveh mejnih verjetnosti pri konkurenčnih hipotezah;



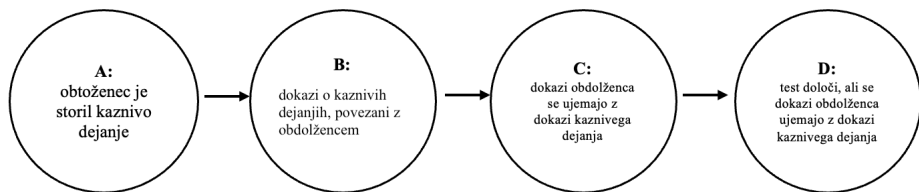
# Zmote v kazenskem pravu

Številne zmote so zlasti posledica napačnega razumevanja pogojne verjetnosti.

- **Tožilčeva zmota (angl. Prosecutor's fallacy),**
- **Zmota obrambnega odvetnika (angl. Defense attorney's fallacy) in**
- **Zasliševalčeva zmota (angl. Interrogator's fallacy).**

# Zmote v kazenskem pravu

Če upoštevamo vzročno verigo dokazov, predstavljeno na sliki, lahko razvrstitev zmot posplošimo na večino vrst dokazov.

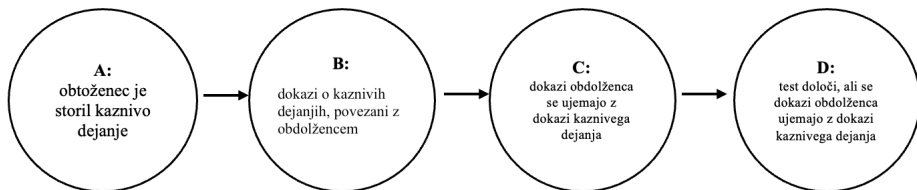


Slika: Vzorčna veriga dokazov

# Zmote v kazenskem pravu

## Tožilčeva zmota

enačimo  $P(C|\neg B)$  s  $P(\neg A|C)$ .

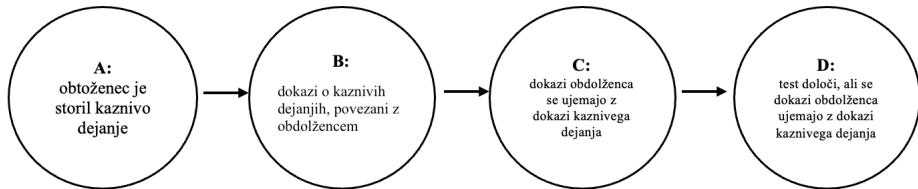


Slika: Vzorčna veriga dokazov

# Zmote v kazenskem pravu

## Napaka verjetnosti

$P(C|\neg B)$  enačimo z verjetnostjo (imenujmo jo  $q$ ), da ima vsaj en nedolžen član populacije ustreza dokazom.

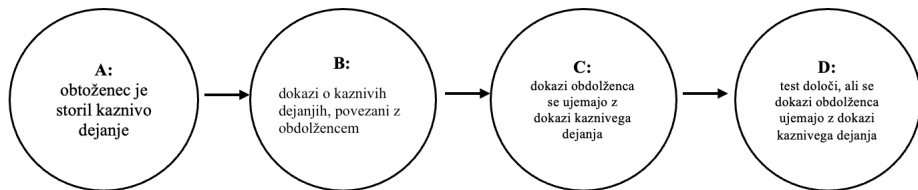


Slika: Vzorčna veriga dokazov

# Zmote v kazenskem pravu

## Zanemarjanje predhodnih verjetnosti

neupoštevanje predhodnih vrednosti, kot sta  $P(A)$  in  $P(B)$ .

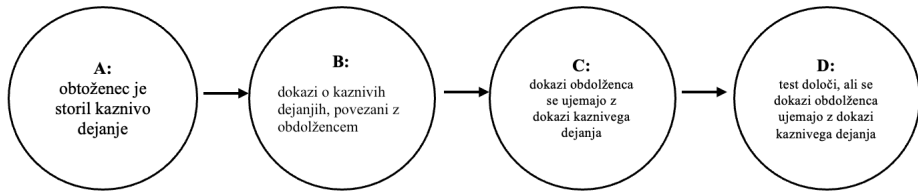


Slika: Vzorčna veriga dokazov

# Zmote v kazenskem pravu

## Napaka pri številčnem preračunavanju

gre za zamenjavo vrednosti  $P(C|\neg B)$  s pričakovanim številom drugih oseb, ki bi jih bilo treba testirati, preden bi našli ujemanje.

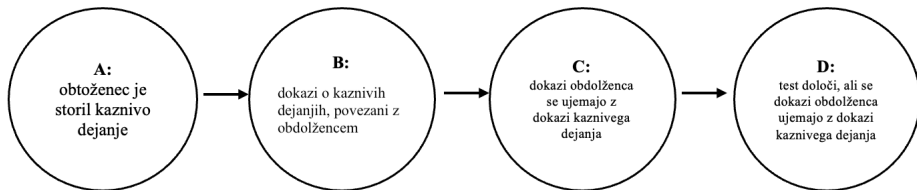


Slika: Vzorčna veriga dokazov

# Zmote v kazenskem pravu

## Pričakovane vrednosti, ki pomenijo edinstvenost

če je velikost populacije približno enaka  $1/P(\neg B|C)$ , potem mora biti obdolženec edini primerek. Binomski izrek pokaže, da obstaja več kot 25% verjetnost, da bosta v populaciji, katere velikost je  $1/P(\neg B|C)$ , vsaj dva ujemanja.

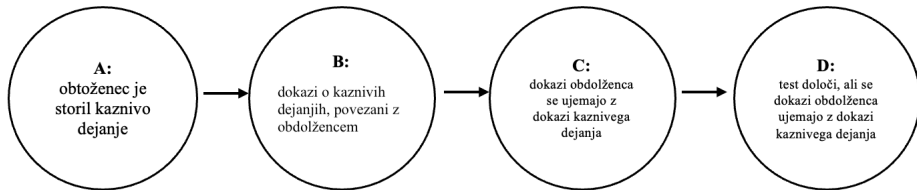


**Slika:** Vzorčna veriga dokazov

# Zmote v kazenskem pravu

## Zmota obrambnega odvetnika

dokaz  $C$  štejemo za nepomembnega, ker visoka predhodna verjetnost  $P(\neg A)$  še vedno povzroči visoko verjetnost  $P(\neg B|C)$ .



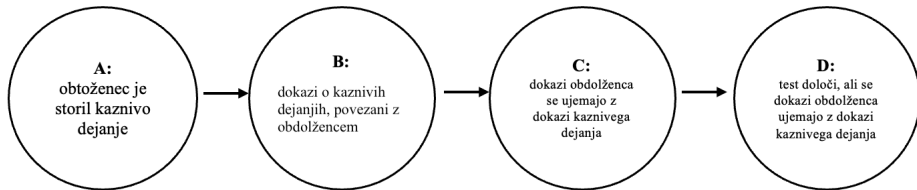
Slika: Vzorčna veriga dokazov



# Zmote v kazenskem pravu

## Napaka baze podatkov obrambnega odvetnika

verjetnost  $P(\neg B|C)$  temelji na drugačni populaciji, kot jo določa  $P(B)$  ali  $P(A)$ .

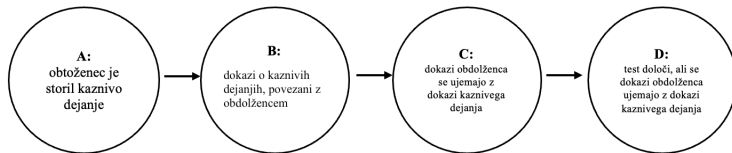


**Slika:** Vzorčna veriga dokazov

# Zmote v kazenskem pravu

## Zasliševalčeva zmota

v tem primeru je dokaz neposredno priznanje krivde. Če to ni potrjeno, to pomeni, da uporabljamo  $P(D|A)$  za informiranje  $P(A|D)$ . Napaka je, da ne upoštevamo  $P(D|\neg A)$ . Če je  $P(D|A) \leq P(D|\neg A)$ , potem dokaz nima vrednosti.



Slika: Vzorčna veriga dokazov

# Zmote v kazenskem pravu

## Zmota odvisnih dokazov

dva ali več dokazov, ki so odvisni, obravnavamo, kot da bi bili neodvisni.

Poseben primer te zmote je *logično odvisna dokazna zmota*.

## Napaka konjunkcije

preiskovalec ne upošteva dejstva, da je dokaz sestavljen iz več kot enega negotovega dogodka, in mu posledično pripiše večjo verjetnost, kot bi jo moral.

# Izogib zmotam z uporabo razmerja verjetnosti

## Prednost uporabe razmerja verjetnosti

odpravlja ugovor Bayesovemu izreku, in sicer upoštevanje predhodne verjetnosti za hipotezo, kot je »kriv«.

⇒ težave pri razumevanju razmerja verjetnosti kot pri razumevanju Bayesove teorije

# Izogibanje zmotam z uporabo Bayesovih omrežij

Bayesova omrežja pomagajo določiti ustrezne verjetnostne formule, ne da bi prikazali njihovo polno algebrsko obliko, in omogočajo skoraj popolno avtomatizacijo potrebnih verjetnostnih izračunov.

- 1 med konkurenčnimi hipotezami izberemo najverjetnejšo;
- 2 izbira mora biti podprta z znanstveno utemeljeno argumentacijo;
- 3 primerna so za analizo dogodka;
- 4 primerno za napovedovanje verjetnosti, da je k dogodku prispeval katerikoli od več možnih znanih vzrokov;

Prednosti Bayesovih mrež se najbolj izrazito pokažejo na zapletenih področjih z več spremenljivkami.

# Izogibanje zmotam z uporabo Bayesovih omrežij

- bistveno izboljšajo vrednotenje verjetnostnih razmerij, ki se uporabljajo za ocenjevanje znanstvenih dokazov;
- omogočajo kompleksnejše verjetnostne analize;
- če se za izdelavo ne uporablja dosleden okvir, lahko Bayesovo omrežje kaže različne rezultate;

Prikaz Bayesovega omrežja se mora ujemati z intuitivnim pripisovanjem vzročno-posledičnih povezav med končno hipotezo, podhipotezo in dokazi primera.