

UNIVERZA V LJUBLJANI
FAKULTETA ZA MATEMATIKO IN FIZIKO

Matematika – 1. stopnja

Neža Kržan

STATISTIKA V KAZENSKEM PRAVU

Delo diplomskega seminarja

Mentor: izr. prof. dr. Jaka Smrekar

Ljubljana, 2023

Kazalo

1	Uvod	7
2	Statistika v kazenskem pravu	7
3	Uporaba statistike pri pravnem postopku	8
4	Raziskovalni proces	9
5	Vrednotenje dokazov	10
6	Koncept verjetnosti	12
7	Bayesova statistika	13
7.1	Opredelitev	13
7.2	Bayesovo pravilo	13
7.3	Bayesovo posodabljanje	14
7.4	Primer - Taksi podjetja	14
7.5	Bayesova teorija v kazenskem pravu	15
7.6	Predhodna verjetnost in določitev posteriorne verjetnosti	16
8	Predstavitev sodnega procesa - medicinska sestra Lucia De Berk	18
8.1	Podatki in uporabljena metoda	18
8.2	Pristop z Bayesovo teorijo	20
9	Bayesova analiza	21
9.1	Poenostavljena Bayesova analiza	21
9.2	Izpopolnjena Bayesova analiza	21
10	Frekvence	23
11	Metoda verjetnosti naključnega ujemanja	25
12	Razmerje verjetnosti	26
12.1	Opredelitev	26
12.2	Razmerje verjetnosti v kazenskem pravu	27

Statistika v kazenskem pravu

POVZETEK

V diplomski nalogi se bom osredotočila na statistiko v kazenskem pravu in na zmote, ki se pojavljajo pri uporabi le te, zaradi pomanjkanja znanja statistike pri odvetnikih, sodnikih in poroti. Osredotočila se bom na uporabo Bayesove statistike v kazenskih postopkih oziroma na izračune, ki izhajajo iz Bayesove statistike in jo primerjala z drugimi metodami. V nadaljevanju bom opisala in razložila dve najpogostejši zmoti, prva je Tožilčeva zmota, ki je dobro znan statistični problem, druga večja, ki pa izhaja iz prve, pa je Zmota obrambnega odvetnika. Ker je uporaba statistike in verjetnostnega računa čedalje pogostejša v sodnih postopkih, bom na koncu pregledala resničen primer sodbe medicinski sestri Lucii de Berk.

Criminal justice statistics

ABSTRACT

...

Math. Subj. Class. (2020): ..., ...

Ključne besede: ..., ...

Keywords: ..., ...

1 Uvod

2 Statistika v kazenskem pravu

Statistične metode so temelj kazenskega pravosodja in kriminologije. Omogočajo oblikovanje in širjenje znanja o kriminaliteti in kazenskopravnem sistemu. Raziskave, ki preverjajo teorije ali preučujejo pojave v kazenskem pravosodju ter so objavljene v znanstvenih revijah in knjigah, so podlaga za večino tega, kar vemo o kaznivih dejanjih in sistemu, ki je bil oblikovan za njihovo obravnavanje. Večina teh raziskav ne bi bila mogoča brez statističnih podatkov.

Raziskave na področju kazenskega pravosodja in kriminologije so različne po naravi in namenu. Velik del raziskav vključuje preverjanje teorije in hipotez.

Definicija 2.1. *Teorija* je niz predlaganih in preverljivih razlag realnosti, ki jih povezujejo logika in dokazi.

Definicija 2.2. *Hipoteza* je posamezna trditev, izpeljana iz teorije, ki mora biti resnična, da bi teorija veljala za veljavno.

Teorije so predlagane razlage določenih dogodkov. Hipoteze so majhni deli teorij, ki morajo biti resnični, da bi celotna teorija držala. Teorijo si lahko predstavljamo kot verigo, hipoteze pa kot člene, ki sestavljajo to verigo.

Statistični znanstveniki na področju kazenskega pravosodja in kriminologije si običajno prizadevajo preučiti razmerja med dvema ali več spremenljivkami. Opazovani ali empirični pojavi pa sprožajo vprašanja. Vzemimo za primer umor. Empirični pojavi so umori in stopnje umorov na mesta kaznivega dejanja. Vendar pa statistični znanstveniki običajno želijo več kot le zapisati empirične ugotovitve - želijo vedeti, zakaj so stvari takšne, kot so. Poskušajo ugotoviti dejavnike, ki so prisotni, zato morajo določiti odvisne in neodvisne spremenljivke.

Definicija 2.3. *Odkvisna spremenljivka* je pojav, ki ga želi statistični znanstvenik preučiti, razložiti ali napovedati.

Definicija 2.4. *Neodvisna spremenljivka* je dejavnik ali značilnost, s katero se poskuša pojasniti ali napovedati odkvisno spremenljivko.

Odkvisne spremenljivke so empirični dogodki, ki jih želi statistični znanstvenik pojasniti. Statistični znanstveniki skušajo opredeliti spremenljivke, ki pomagajo napovedati ali pojasniti dogodke. Neodvisne spremenljivke so dejavniki, za katere statistični znanstvenik meni, da bi lahko vplivali na odkvisne spremenljivke. Glede na naravo raziskovalne študije določijo neodvisne in odkvisne spremenljivke.

Pomembno je razumeti, da neodvisno in odkvisno nista sinonima za vzrok in posledico. Določene neodvisne spremenljivke so lahko povezane z določenimi odkvisnimi spremenljivkami, vendar to še zdaleč ni dokončen dokaz, da so prve vzrok drugih. Za dokazovanje vzročnosti morajo statistični znanstveniki dokazati, da njihove študije izpolnjujejo tri merila. Prvo je časovno zaporedje, kar pomeni, da se mora neodvisna spremenljivka pojaviti pred odkvisno spremenljivko. Druga zahteva glede vzročnosti

je, da obstaja empirična povezava med neodvisno in odvisno spremenljivko. Zadnja zahteva je, da je razmerje med neodvisno spremenljivko in odvisno spremenljivko nepristransko. Nepristranskost je v kriminologiji in kazenskoopravnih raziskavah pogosto najtežje dokazati, saj je človeško vedenje zapleteno in ima vsako dejanje, ki ga oseba stori, več vzrokov. Razmejitev teh vzročnih dejavnikov je lahko težavna ali nemogoča. Razlog, zakaj je nepristranskost problem, je, da lahko obstaja tretja spremenljivka, ki pojasnjuje odvisno spremenljivko enako dobro ali celo bolje kot neodvisne spremenljivke. Ta tretja spremenljivka lahko delno ali v celoti pojasni razmerje med neodvisno in odvisno spremenljivko. Nenamerna izključitev ene ali več pomembnih spremenljivk lahko privede do napačnih zaključkov, saj lahko statistični znanstvenik zmotno verjame, da neodvisna spremenljivka močno napoveduje odvisno spremenljivko, medtem ko je v resnici razmerje dejansko delno ali v celoti posledica posrednih dejavnikov. Drug izraz za to težavo je pristranskost izpuščenih spremenljivk. Kadar je pristranskost izpuščenih spremenljivk prisotna v razmerju neodvisne spremenljivke - odvisne spremenljivke, vendar se ne prepozna, lahko pridemo do napačnega sklepa o zločinu.

Torej vse vrste spremenljivk so med seboj povezane, vendar je pomembno, da na podlagi statističnih povezav ne sklepamo prehitro o vzročnih posledicah, kar pa bi lahko bila težava.

Statistični znanstveniki se že na začetku sodnega procesa soočajo s prvimi težavami - določitvijo odvisnih in neodvisnih spremenljivk za modeliranje. V proces določanja spremenljivk pa pogosto v preveliki meri posegajo odvetniki, ki se sklicujejo na pravne zakone in načela. To lahko postane sporno, saj lahko takšni pretirani posegi ovirajo statistične znanstvenike pri izračunu verjetnostnega vpliva spremenljivk. Odvetniki sicer imajo pomembno vlogo pri zagovarjanju strank v sodnih procesih, vendar je njihovo znanje o statistiki in verjetnostnih izračunih pomanjkljivo. Po mojem mnenju zato odvetniki lahko napačno opredelijo odvisne in neodvisne spremenljivke ter s tem vplivajo na kakovost modeliranja, kar lahko privede do napačnih zaključkov in nepravilnih odločitev v sodnih procesih. Zagotovo določitev spremenljivk ne sme biti naloga le odvetnikov ali le statističnih znanstvenikov, ampak menim, da je sodelovanje med statistični znanstveniki in odvetniki pomembno, saj vsak prispeva svoj del poznavanja teorije, ki je v ozadju. S tem se lahko zagotovi pravilno opredelitev spremenljivk in pravilne verjetnostne izračune, ki bodo prispevali k pravičnim odločitvam v sodnih postopkih.

3 Uporaba statistike pri pravnem postopku

Sodišča ne potrebujejo statističnega strokovnega znanja le za rezultat statističnega postopka, temveč tudi za zagotovitev, da je metodologija primerna za podatke in da analiza razreši pravni problem. Pri skoraj vseh uporabah podatkov se sodni proces zanaša na razlago statističnih znanstvenikov, ki ocenijo zanesljivost podatkovne baze in pravilno razlagajo rezultate statistične analize. Pred pričanjem na sodišču moramo vedeti, na kaj točno se podatki nanašajo, kako so bili zbrani in kakšen del manjka ali je neuporaben, da se lahko odločimo za ustrezen postopek analize podatkov. Potrebujemo osnovne informacije odvetnika in drugih strokovnjakov, da lahko

oblikujemo ustrezne primerjalne skupine. Ta postopek vključuje določitev ustrezne populacije (populacij), ki jo (jih) je treba preučiti, parametrov, ki nas zanimajo, in statističnega postopka, ki ga je treba uporabiti. Statistični znanstveniki ne morejo določiti, katere vrednosti parametra so pravno pomembne, ker je parameter pravno določen.

Statistične informacije, ki jih dobi sodnik, so filtrirane prek odvetnikov. Odvetnik avtorju statistične analize postavlja vprašanja z namenom razlage statistične analize poroti, sodniku in drugim v sodni dvorani. Vprašanja so s strani odvetnikov seveda premišljeno postavljena, zato se lahko zgodi, da do temeljite razlage analize ne pride, ker statistični znanstvenik ne dobi primernih vprašanj. Kasneje so v delu obrazložene zmote, ki nastanejo zaradi pomanjkanja znanja verjetnosti pri sodnikih, poroti in odvetnikih, ampak mogoče pa nekatere izmed njih nastanejo tudi zaradi nepopolne razlage statistične analize. Do neke točke statistični znanstvenik sicer sam predstavi analizo, potem pa mora biti tudi on previden z razlago, zaradi porote, kajti poroti se predvidoma ne narekuje kako naj si razlaga dokaze, kar pa ponavadi preučujemo s statistično analizo. Mogoče bi moralo biti določeno kaj vse mora statistični znanstvenik predstaviti in razložiti, da bi se lahko izognili zmotam. Uporaba statističnih analiz na sodiščih prinaša tudi nove težave za statistiko. Temelj dobre statistične analize je dober pregled predpostavk, na katerih morajo temeljiti naše metode in upoštevanje pravil prava. Pomembne so tudi baze podatkov in velikosti vzorcev, kar lahko predstavlja težavo. Potrebno je kombiniranje različnih postopkov za pridobivanje informacij iz podatkov in razlago rezultatov, kar sem zasledila, da statistični znanstveniki velikokrat izkoriščajo, se ne poglobijo dovolj in predstavljena analiza postane nepopolna.

4 Raziskovalni proces

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

1. Identifikacija problema.

Na tej stopnji je treba navesti, zakaj je raziskava potrebna oziroma kakšen problem je potrebno razrešiti. Problem je treba jasno opredeliti in opisati. Navesti je potrebno hipotezo(e), dokaz(e) in spremenljivko(e), ki jih preučujejo.

Koncepti so posebne vrste spremenljivk (npr. socialno-ekonomski status), ki jih ni mogoče neposredno opazovati, vendar jih želijo izmeriti. Meritve morajo biti veljavne in zanesljive. Veljavnost je stopnja, do katere merilo natančno meri spremenljivko in njen osnovni koncept. Po mojem mnenju bi se na tem mestu že lahko vprašali oziroma soočili s problemom ali je izbrana mera jasen kazalnik zadevnega koncepta. Zanesljivost je stopnja, do katere je spremenljivka dosleden in zanesljiv kazalnik koncepta. Spremenljivka je namenjena merjenju teh opazovanj ali konceptov. Običajno ima več kot eno možno vrednost. Merjenje spremenljivke mora biti jasno opredeljeno oziroma mora imeti operativno opredelitev. Hipoteza je izražena v obliki odnosa med spremenljivkami. V smislu reševanja problemov hipoteza opisuje način, na katerega je mogoče rešiti problem. V raziskavi neodvisna spremenljivka

(X) povzroča učinek ali vpliv na odvisno spremenljivko (Y). Odvisna spremenljivka (Y) se lahko spremeni zaradi prisotnosti neodvisne spremenljivke. Hipoteza je torej napoved. Pričakujemo, da bo neodvisna spremenljivka povzročila učinek na odvisno spremenljivko.

2. Zasnova raziskave.

Več elementov raziskovalnega procesa se nanaša na postopek statistične analize. Vsi imajo ključno vlogo v logiki statistike. Zasnova raziskave nam pomaga ugotoviti, ali je metoda učinkovita, če bi jo poskušali izvajati na drugih mestih in v drugih časih. To opazimo z t.i. klasičnim eksperimentom. Cilj raziskave je dokazati, ali je imela spremenljivka želeni učinek ali ne. Da bi to ugotovili, poskušamo z raziskovalno zasnovo izolirati učinek ukrepa na problem. Klasični eksperiment vključuje razvrstitev udeležencev v eksperimentalno in kontrolno skupino. Ključni element postopka je naključna izbira, ki zagotavlja, da sta skupini primerljivi v vseh pomembnih vidikih. V statističnem smislu naključna dodelitev pomeni, da ima vsak član ciljne populacije enake možnosti, da bo izbran v eksperimentalno skupino. Drugi element je verjetnostno vzorčenje. Statistični znanstvenik običajno ne more preučiti vseh elementov populacije. Večina raziskav se izvaja z izbiro vzorca iz populacije. Zbiranje podatkov vključuje opredelitev in izbiro virov podatkov. Vir podatkov je lahko raziskava, uradne evidence ali uradni statistični podatki.

3. Analiza podatkov.

Po zbiranju podatkov se začne analiza, ki vključuje pravilno izbiro in uporabo statističnih metod. Toda interpretacija in predstavitev rezultatov raziskave sta prav tako ključna vidika celotnega procesa. Kot je že bilo omenjeno je pomembno, da statistični znanstveniki upoštevajo ciljno občinstvo, ki jim bodo raziskovalni rezultati namenjeni. Pri neustrezni predstavitvi lahko pride do napačnega razumevanja rezultatov in posledično do napačnih sklepov. Vendar pa se pri predstavitvi rezultatov pojavijo največje težave, ko gre za področje kazenskega pravosodja. Statistični znanstveniki morajo biti pozorni na posledice svojih rezultatov, zato svoje ugotovitve predstavijo na način, ki izraža resnico, hkrati pa je razumljiv in nepristranski. Poleg tega je treba upoštevati, da so posledice raziskav lahko dolgoročne in se lahko pojavijo šele v prihodnosti.

5 Vrednotenje dokazov

Ena od najbolj obravnavanih in kontroverznih tem med pravniki je vloga verjetnosti pri ocenjevanju pravnih dokazov, pridobljenih v sodnem procesu ugotavljanja dejstev in preiskave, ki je značilen za sodišča. Na mednarodnih kazenskih sodiščih je vodilno načelo prosta ocena dokazov, kar pomeni, da sodniki niso dolžni spoštovati pravil, kako ocenjevati dokaze, in zato lahko izberejo pristop, ki naj bi bil najprimernejši za oceno.

Postopek ugotavljanja dejstev zahteva oceno vseh dokazov, ki jih predložijo sodišču, da se določi, ali je obdolženec kriv ali ne. Uvedeta se pojma dokazni standard in vrednotenje dokazov.

Dokazni standard je pravno vprašanje. Gre za gre za abstraktno normo, ki je opredeljena s pravnim pravilom; podobno kot obstoj določenih predpostavk za določeno kaznivo dejanje.

Vrednotenje dokazov pa je pravni problem. Gre za odločitev, kako se dokazi, v določenem sodnem procesu oziroma kazenskem primeru, nanašajo na abstraktno normo.

Matematični pristopi za ocenjevanje dokazov določajo odstotke za dokazni standard, ki so manjši od popolne gotovosti, zato se predložene informacije (ali njihovo pomanjkanje) pretvorijo v številčne vrednosti (običajno približno 90-95-odstotna stopnja verjetnosti), ki se nato primerjajo z zahtevanim dokaznim standardom. Matematična tradicionalna verjetnost je opredeljena kot povezava med verjetnostjo nastanka dogodka s trenutno številčno vrednostjo. Po tem je možnost nastanka dogodka bolj verjetna, če spada med večino opazovanih dogodkov. Nasprotno pa možnost dogodka manj verjetna, če spada v manjšino opazovanih dogodkov. Verjetnost dogodka je torej razmerje med številom primerov, v katerih se je dogodek zgodil in število vseh relevantnih primerov, tj.

$$P(\text{dogodek}) = \frac{\text{število primerov, v katerih se je dogodek zgodil}}{\text{število relevantnih primerov}}$$

Najpogostejši uporabljeni metodi za ocenjevanje dokazov sta metoda dokazne vrednosti in model verjetnosti hipoteze.

Metoda dokazne vrednosti temelji na vrednosti, ki jo ima dokaz za dokazno temo, njen namen pa je ugotoviti, ali med dokazom in zadevno dokazno temo obstaja nalključna povezava, pri čemer je dokazna tema glavna obtožba v kazenskem primeru. S to metodo dokazujemo določen omejen nabor dokazov, njen cilj pa je oceniti verjetnost, da dokazi dokazujejo hipotezo.

Z modelom verjetnosti hipoteze pa ocenjujemo verjetnost hipoteze glede na dokaze. Cilj je ugotoviti, kako verjetno je, da je hipoteza, za katero dokazi zagotavljajo določeno stopnjo podpore, resnična. Glavna razlika z metodo dokazne vrednosti je, da predpostavlja, da obstaja začetna verjetnost za hipotezo pred obravnavo dokazov, t.i. predhodna verjetnost.

V kazenskem pravu pa lahko zelo hitro pride do posebnih, edinstvenih predpostavk oziroma hipotez, ki pa predstavljajo težave pri vrednotenju oziroma merjenju v statističnih modelih. Zato so dokazi razdeljeni na tri različne, vendar delno se prekrivajoče kategorije:

1. dokaz je usmerjen na pojav ali neobstoj dogodka, dejanja ali vrste ravnanja, na katerem temelji sodni spor;
2. dokaz je usmerjen na identiteto posameznika, odgovornega za določeno dejanje ali niz dejanj;
3. dokaz je usmerjen v namero ali kakšen drug element odgovornosti, kot je znanje ali provokacija.

Pomen, ustreznost in nevarnosti dokaza so močno odvisni od tega, ali naj bi takšen dokaz vplival na dogodek, identiteto ali miselnost.

6 Koncept verjetnosti

Definicija 6.1. 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 dveh 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 določena dogodka ne dopolnjujeta.

Koncept verjetnosti je ključen pri ocenjevanju dokazov, saj omogoča objektivno oceno njihovega vpliva na verjetnost določene domneve o interesni osebi (v nadaljevanju PoI) ali obdolžencu. Pri presoji dokazov se uporablja različne metode in tehnike, ki temeljijo na statistični verjetnosti. Te metode omogočajo oceno, kako verjetno je, da so dokazi resnični in zanesljivi. Pri presoji dokazov se najprej analizira njihova verjetnost, pri čemer je pomembno, da se upošteva tudi kontekst. Dokazi se namreč ne presojujejo izolirano, ampak v kontekstu celotnega primera. To pomeni, da se pri presoji dokazov upošteva tudi druge dokaze in okoliščine primera. Na ta način se lahko izvede bolj objektivna presoja dokazov in ugotovi, kako vplivajo na določeno domnevo o interesni osebi ali obdolžencu.

V skladu s konceptom verjetnosti se pri presoji dokazov upošteva tudi verjetnost napake. Ta se nanaša na verjetnost, da so dokazi napačni ali zavajajoči.

V splošnem nas zanima vpliv dokazov na verjetnost krivde(H_p) in nedolžnosti(H_d) osumljenca. Gre za dopolnjujoča se dogodka in razmerje verjetnosti teh dveh dogodkov,

$$\frac{P(H_p)}{P(H_d)}, \quad (6.1)$$

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

$$\frac{P(H_p|E)}{P(H_d|E)}, \quad (6.2)$$

verjetnost v prid krivdi ob upoštevanju dokazov E .

Ali je obtoženec kriv glede na znan doka E , je glavna stvar, ki nas pri sojenju zanima. 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.

7 Bayesova statistika

7.1 Opredelitev

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.

Bayesova analiza je standardna metoda za posodabljanje verjetnosti po opazovanju več dokazov, zato je zelo primerna za obravnavo in vrednotenje dokazov. Vsakdo, ki mora presoditi o hipotezi, kot je »krivda« (vključno s preiskovalci pred sojenjem, sodniki, porotami), neformalno začne z nekim predhodnim prepričanjem o hipotezi in ga posodablja, ko se dokazi ponovno pojavijo. Včasih lahko obstajajo celo objektivni podatki, na katerih temelji predhodna verjetnost. Pri uporabi Bayesovega sklepanja morajo statistični znanstveniki utemeljiti predhodne predpostavke, kadar je to mogoče, na primer z uporabo zunanjih podatkov; v nasprotnem primeru morajo uporabiti razpon vrednosti predpostavk in analizo občutljivosti, da preverijo zanesljivost rezultata glede na te vrednosti.

7.2 Bayesovo pravilo

Bayesovo sklepanje temelji na Bayesovem pravilu, ki izraža verjetnost nekega dogodka z verjetnostjo dveh dogodkov in obrnjene pogojne verjetnosti. Pogojna verjetnost predstavlja verjetnost dogodka, glede na drug dogodek.

Definicija 7.1. Pogojna verjetnost dogodka H , glede na dogodek E , je

$$P(H|E) = \frac{P(H \cap E)}{P(E)}, \quad (7.1)$$

ob predpostavki, da je $P(E) > 0$.

Formula (7.1) pove, da je verjetnost dogodka H ob pogoju, da se je zgodil dogodek E , enaka razmerju verjetnosti, da se zgodita oba dogodka in verjetnosti, da se je zgodil dogodek E . Potem pogojno verjetnost uporabimo še v števcu formule (7.1) in dobimo Bayesovo pravilo.

Izrek 7.2. (*Bayesovo pravilo*)

$$P(H|E) = \frac{P(E|H) \times P(H)}{P(E)}. \quad (7.2)$$

Verjetnost dogodka E lahko še razpišemo

$$P(H|E) = \frac{P(E|H) \times P(H)}{P(E|H)P(H) + P(E|\neg H)P(\neg H)}. \quad (7.3)$$

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

$$\frac{P(H|E)}{P(\neg H|E)} = \frac{P(E|H)}{P(E|\neg H)} \times \frac{P(H)}{P(\neg H)}. \quad (7.4)$$

7.3 Bayesovo posodabljanje

Bayesovo pravilo se razlikuje od Bayesovega posodabljanja. Prvo je matematični izrek, drugo pa logična trditev, kako se sčasoma posodablja apriorne oziroma predhodne verjetnosti dokazov glede na novo zbrane dokaze oziroma prepričanja.

Trditev 7.3. (*Bayesovo posodabljanje*) Če se dogodek E zgodi ob času $t_1 > t_0$, potem je $P_1(H) = P_0(H|E)$.

Ob času t_0 dogodku H dodelimo verjetnost $P_0(H)$; to se imenuje predhodna verjetnost oziroma apriorna 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 .

Recimo, da je dogodek H neka hipoteza oziroma prepričanje o zločinu in dogodek E dokazi, zbrani za ta zločin. Pri Bayesovem posodabljanju je videti, kot da je dokaz E nesporno resničen. Z drugimi besedami, predpostavka je, da moramo imeti po zbiranju dokazov E stopnjo zaupanja v E enako 1, torej če so dokazi zborni v času t_1 , je $P_1(E) = 1$.

7.4 Primer - Taksi podjetja

Za lažje razumevanje Bayesovega pravila sem pripravila spodnji primer.

Obstajata dve taksi podjetji, Zeleni Taksi in Modri taksi, katerih vozila so pobarvana zeleno oziroma modro. Podjetje Zeleni taksi pokriva 85 odstotkov trga, podjetje Modri taksi pa preostanek. Predpostavimo še, da v okolici ni drugih taksi podjetij. V meglenem dnevu taksi trči v mimoidočega pešca in ga poškoduje, vendar odpelje s kraja nesreče. Priča nesreče poroča, da je bilo vozilo modre barve. Priča ima prav le v 80 odstotkih primerov, kar pomeni, da je njegova zanesljivost enaka 0,8. Kolikšna je verjetnost, da je bil taksi, ki je povzročil nesrečo, modre barve glede na poročilo priče?

Vpeljimo oznake:

Z ... hipoteza, da je bil taksi zelen,

M ... hipoteza, da je bil taksi moder,

W_m ... dokaz, t.j. poročanje priče, da je bil taksi moder.

Problem je določiti pogojno verjetnost hipoteze M , ob pogoju, da je dokaz W_m resničen, torej $P(M|W_m)$.

Za uporabo Bayesovega pravila potrebujemo tri elemente: verjetnost dokazov glede na hipotezo, verjetnost dokazov in verjetnost hipoteze. Podjetje Zeleni taksi pokriva 85 odstotkov trga, zato je verjetnost $P(Z) = 0,85$. Po definiciji verjetnosti, je potem $P(M) = 1 - P(Z) = 0,15$. Vemo tudi, da ima priča v 80 odstotkih primerov prav, torej $P(W_m|M) = 0,8$ in $P(W_m|Z) = 0,2$. Izračunajmo še verjetnost dokaza $P(W_m)$:

$$P(W_m) = P(W_m|B)P(M) + P(W_m|Z)P(Z) = 0,8 \times 0,15 + 0,2 \times 0,85 = 0,29.$$

Sedaj lahko uporabimo Bayesovo pravilo:

$$P(M|W_m) = \frac{P(W_m|M)P(M)}{P(W_m|M)P(M) + P(W_m|Z)P(Z)} = \frac{0,8 \times 0,15}{0,29} = \frac{12}{29} \approx 0,41.$$

Verjetnost, da je bil taksi glede na pričanje dejansko modre barve, je precej majhna, tudi če ima priča v 80 odstotkih primerov prav. Razlog za to je, da je verjetnost M , ne glede na dokaze, majhna ($P(M) = 0,15$). Spodnja tabela kaže, da lahko s spreminjanjem verjetnosti M dobimo različne pogojne verjetnosti $P(M|W_m)$, pri čemer je zanesljivost priče nespremenjena:

$P(M)$	$P(Z)$	$P(W_m M)$	$P(M W_m)$
0,15	0,85	0,8	0,41
0,25	0,75	0,8	0,57
0,35	0,65	0,8	0,68
0,45	0,55	0,8	0,76
0,50	0,50	0,8	0,80
0,55	0,45	0,8	0,83
0,65	0,35	0,8	0,88
0,75	0,25	0,8	0,92
0,85	0,15	0,8	0,95

Tabela 1: Rezultati pogojne verjetnosti $P(M|W_m)$ pri spreminjanje verjetnosti M .

7.5 Bayesova teorija v kazenskem pravu

Bayesova teorija razlaga verjetnost kot merilo verjetnosti ali zaupanja, ki ga lahko ima posameznik glede nastanka določenega dogodka. Daje nam matematične modele za vključevanje naših apriornih prepričanj in dokazov, ki jih že lahko že imamo o

nekem dogodku, za ustvarjanje novih prepričanj oziroma za pridobitev posteriornih prepričanj, ki se lahko uporabijo za kasnejše odločitve, ko se pojavijo novi dokazi.

Sodniki ali porotniki, ki ugotavljajo sklep sodnega procesa, imajo na voljo vrsto dokazov. Njihova naloga je ocena, kako te informacije oziroma dokazi vplivajo na tožilčevo domnevo o obdolžencu oziroma storilcu kaznivega dejanja. Pri Bayesovi teoriji moramo upoštevati vsak dokaz posebej, kar je še posebej pomembno pri začetku sodnega procesa.

Gre za postopek posodabljanja verjetnosti tožilčeve hipoteze na podlagi predhodnih oziroma apriornih verjetnosti. Pretvorimo predhodno verjetnost, tj. verjetnost hipoteze pred upoštevanjem določenega dokaza (dokazov), v posteriorno verjetnost, tj. verjetnost hipoteze po upoštevanju določenega dokaza (dokazov).

Pomembno je tudi, da razlikujemo med dokazi pri Bayesovi teoriji in dokazi na sodišču. V Bayesovi teoriji je vsaka informacija dokaz, če je pomembna za verjetnost hipoteze. V kazenskem postopku pa je dokaz informacija, ki je bila predložena sodišču za podporo določeni izjavi na sodišču in je sprejeta kot pravno dopustna, kar pomeni, da je ta informacija znana sodniku, ki presoja tožilčevo domnevo o obdolžencu in je pomembna za verjetnost hipoteze.

7.6 Predhodna verjetnost in določitev posteriorne verjetnosti

Definicija 7.4. Predhodna verjetnost, ki je uporabljena v vsaki posodobitvi verjetnosti s pomočjo Bayesove teorije, je začetna verjetnost hipoteze oziroma tožilčeve domneve o obdolžencu oziroma storilcu kaznivega dejanja.

Razjasniti je potrebno, da ko odvetniki govorijo o predhodni verjetnosti, pogosto mislijo na verjetnost začetne hipoteze oziroma tožilčeve domneve o obdolžencu, preden so bili predloženi dokazi, kar je v skladu z opredelitvijo predhodne verjetnosti v Bayesovi teoriji. Po končni posodobitvi dobimo verjetnost hipoteze glede na vse dokaze, predložene na sojenju.

Recimo, da je statistični znanstvenik naprošen opraviti analizo profila DNK krvi, najdene na kraju kaznivega dejanja, in rezultat primerjati s profilom DNK obdolženca. O krivdi ali nedolžnosti obtoženca bo odločala porota. Odločitev porotnikov bo delno odvisna od njihove ocene dveh interesnih hipotez

$H_1 \dots$ vir krvi je obtoženec,

$H_2 \dots$ vir krvi je druga oseba.

Porotniki bodo morda želeli, da jim statistični znanstvenik dokončno pove, katera hipoteza je resnična, ali da jim navede verjetnosti vira. Za oceno verjetnosti vira mora upoštevati tudi druge dokaze v kazenskem primeru oziroma kontekst celotnega primera.

Recimo, da je statistični znanstvenik ugotovil, da imata obtoženec in kri s kraja zločina skupen niz t.i. genetskih označevalcev, ki jih najdemo pri eni osebi na 1 milijon prebivalcev v zadevni populaciji. Ne da bi upošteval druge dokaze, lahko statistični znanstvenik poda izjavo o pogojni verjetnosti ugotovitve teh rezultatov

pri dveh hipotezah o medsebojni povezanosti. Lahko na primer tudi izjavi, da so skupni genetski označevalci skoraj zagotovo najdeni v primeru H_1 (vir je bil obtoženec), vendar imajo le 1 možnost na milijon, da bodo najdeni v primeru H_2 (vir je bil nekdo drug). Na podlagi te ocene lahko statistični znanstvenik poroti predloži razmerje verjetnosti - na primer, da so rezultati profila DNK 1 milijonkrat bolj verjetni, če je bil vir krvi obtoženec in ne neka druga oseba. Vendar razmerje verjetnosti ni isto kot verjetnost vira.

Edini skladen način, kako na podlagi forenzičnih dokazov sklepati o verjetnosti virov, je uporaba Bayesovega pravila, ki zahteva, da začnemo s pripisom predhodnih verjetnosti za hipoteze, ki nas zanimajo, pri čemer bo Bayesov pristop bo deloval le, če bo statistični znanstvenik lahko začel s predhodno oziroma apriorno verjetnostjo.

Določitev predhodnih oziroma apriornih verjetnosti je resen problem pri Bayesovemu pristopu v kazenskih postopkih. Različne metode za določitev in izračun teh verjetnosti lahko dajejo rezultate, ki se med seboj precej razlikujejo, kar pa je problematično, ker celotna Bayesova teorija temelji ravno na teh začetnih izračunih. Bistveno vprašanje, ki se postavlja, je, ali naj statistični znanstvenik sploh poskušajo določiti predhodne verjetnosti in če ja, kako naj jih določijo. Nekateri strokovnjaki predlagajo, da bi morali predpostaviti enake predhodne verjetnosti za vse hipoteze v primeru, kar se imenuje nevtrarno stanje. To pomeni, da se statistični znanstvenik ne opredeli za nobeno od hipotez, preden zbere kakršne koli dokaze ali informacije, ampak predpostavlja enako verjetnost za vse hipoteze. Torej predpostavljajo, da sta predhodni verjetnosti H_1 in H_2 enaki, nato pa ju v skladu z Bayesovim pravilom pomnožijo z razmerjem verjetnosti, da določijo posteriorno verjetnost. To bi se lahko izkazalo za praktičen pristop, saj se lahko statistični znanstvenik s tem izogne vplivu lastnih in odvetniških predsodkov ter mnenj, ki bi lahko vplivali na predpostavke o verjetnosti. Na ta način se lahko zagotovi objektivnost analize, saj ne poskušamo prikazati ene hipoteze bolj verjetne od druge. Kljub temu pa mislim, da moramo biti do tega pristopa nekoliko kritični, saj je predpostavljanje enake verjetnosti za vse možnosti problematično - v realnosti se različne hipoteze razlikujejo po svoji verjetnosti.

Mnenje in poročila statističnih znanstvenikov naj bi bila ključni vir informacij, ki lahko prispevajo k objektivnemu in strokovnemu vrednotenju dokazov in verjetnosti v sodnih postopkih. Zato sem mnenja, da je potrebno zagotoviti njihovo neodvisnost in strokovnost ter jih zaščititi pred morebitnim poseganjem odvetnikov ali drugih udeležencev sodnega postopka v njihov proces dela, tako kot morajo to zagotoviti oni. Po mojem mnenju je pri določanju apriorne verjetnosti hipotez in dokazov zelo pomembno, da smo natančni in korektni, kajti vsi naslednji verjetnostni računi in statistične analize temeljijo ravno na tem. Zato naj statistični znanstvenik uporabi svoje strokovno znanje za izračun apriorne verjetnosti na podlagi razpoložljivih podatkov in brez nepotrebnega vplivanja odvetnikov ali drugih udeležencev postopka. Poleg tega sem ugotovila, da za statistiko v kazenskem pravu obstajajo pravila in smernice kako upoštevati zakonodajo, pravila in postopke sodnega procesa, ki jih po mojem mnenju dosledno upoštevajo, torej načeloma zagotovijo ustrezne predhodne verjetnosti. Tako bi dobili dovolj objektivno oceno verjetnosti krivde ali nedolžnosti obdolženca in zagotovili pravično sodno odločitev. Glavni očitke temu pristopu v okviru sodnega procesa bi lahko bil, da lahko statistični znanstveniki presežejo svoje

znanstveno znanje in si prisvojijo vlogo tistega, ki ugotavlja dejstva.

Ko na nek način le določijo predhodne oziroma apriorne verjetnosti tožilčeve hipoteze in dokazov torej sledi posodabljanje le teh. Tekom sodnega procesa se v realnosti vedno pojavljajo nove domneve o obtožencu in skoraj vedno najdejo nove dokaze s kraja zločina. Smiselno je, da vse to v postopku izračuna tudi upoštevamo. Zasledila sem, da se tekom sodnega procesa marsikateri dokaz najprej prizna in je znan sodniku, ki presoja tožilčevo domnevo o obdolžencu, torej ga upoštevajo v svojih izračunih za posodobitve predhodnih verjetnosti hipotez. Potem pa dokaz iz sodnega procesa umaknejo, ampak presnetilo me je, da dokaz največkrat ni umaknjen iz verjetnostnih računov. Mnenja sem, da bi morali statistični znanstvenik, ko se določen dokaz iz sodnega procesa, zaradi tehtnega razloga, umakne, posodobiti vse račune za nazaj in nato nadaljevati posodabljanje verjetnosti. Tako bi dobili primeren izračun posteriornih verjetnosti, na katerih bi potem lahko temeljil zaključek sodnega procesa.

8 Predstavitev sodnega procesa - medicinska sestra Lucia De Berk

Marca 2003 je bila medicinska sestra Lucia de Berk (v nadaljevanju osumljenka) v Haagu na Nizozemskem obsojena na dosmrtni zapor zaradi obtožb uboja in domnevnega uboja več bolnikov v dveh bolnišnicah, v katerih je delala v bližnji preteklosti (bolnišnici JZK in RKZ). V bolnišnici JKZ se je med Lucijinimi izmenami pojavilo nenavadno veliko smrti bolnikov, kar je sprožilo vprašanja glede njene morebitne vpletenosti v te dogodke. Razmišljalo se je, ali bi lahko bila Lucijina prisotnost pri toliko smrtih zgolj naključje.

Na zahtevo državnega tožilca, je statistično analizo podal statistik Henk Elffers. V grobem je bila njegova ugotovitev naslednja: ob predpostavkah, da

1. je verjetnost, da osumljenka doživi smrt bolnika med izmeno, enaka verjetnosti za katero koli drugo medicinsko sestro in
2. da so pojavi smrt bolnika neodvisni za različne delovne izmene,

potem je verjetnost, da je osumljenka doživela toliko smrti bolnikov, kot jih je dejansko doživela, manjša od 1 proti 342 milijonom. Henk Elffers je mnenja, da je verjetnost naključja tako majhna, da standardna statistična metodologija zavrne ničelno hipotezo o naključju. Vendar je poudaril, da to ne pomeni nujno, da je osumljenka kriva. Prvi problem, ki sem ga zasledila je, da Henk Elffers in sodišče nista upoštevala subjektivnega elementa pri izbiri verjetnostnega modela in s tem možnosti za obstoj več modelov z različnimi napovedmi ali celo različnimi odgovori na različna vprašanja.

8.1 Podatki in uporabljena metoda

Henk Elffers je temeljil na podatkih o izmenah osmuljenke in drugih medicinskih sester ter smrt bolnika, ki so se zgodili v teh izmenah, da bi utemeljil svoj model v celoti.

Ime bolnišnice (in številka izmene)	JKZ	RKZ-41	RKZ-42
Število vseh izmen	1029	336	339
Število izmen osumljenke	142	1	58
Število vseh incidentov	8	5	14
Število smrti v času izmene osumljenke	8	1	5

Tabela 2: Podatki o izmenah in smrtih bolnikov, ki jih je uporabil Henk Elffers (za določeno obdobje).

Tekom sodbe je bilo ugotovljeno, da je osumljenka v RKZ-41 dejansko opravila tri izmene in ne le ene, zato v kasnejših izračunih uporabimo to pravilno število.

Najprej bi se lahko odločili za izdelavo modela na podlagi epidemioloških podatkov o verjetnosti smrti med različnimi vrstami izmen; tako bi lahko izračunali verjetnost, da bi bila osumljenka naključno prisotna pri toliko smrtih, kolikor jim je bila dejansko priča. Težava tega pristopa pa je, da nimamo potrebnih podatkov. Zato je Henk Elffers poskušal sestaviti model, kjer uporabimo samo zgoraj navedene podatke. Predpostavil je, da

1. obstaja verjetnost p za nastanek smrti bolnika med določeno izmeno (torej p ni odvisen od tega, ali gre za dnevno ali nočno izmeno),
2. dogodki smrt bolnika se pojavljajo neodvisno drug od drugega.

Izračunajmo pogojno verjetnost dogodka, da se (recimo v JKZ) vse smrti zgodijo med izmenami osumljenke, ob upoštevanju skupnega števila smrti bolnikov in skupnega števila izmen v preučevanem obdobju. Naj bo

$n \dots$ skupno število izmen,

$r \dots$ število izmen osumljenke,

potem je pogojna verjetnost, da je bila osumljenka priča x smrtim, če se je zgodilo k smrti bolnikov, naslednja

$$\frac{\binom{r}{x} p^x (1-p)^{r-x} \binom{n-r}{k-x} p^{k-x} (1-p)^{n-r-k+x}}{\binom{n}{k} p^k (1-p)^{n-k}} = \frac{\binom{r}{x} \binom{n-r}{k-x}}{\binom{n}{k}}. \quad (8.1)$$

Ta porazdelitev je znana kot hipergeometrijska porazdelitev. S to formulo lahko enostavno izračunamo (pogojno) verjetnost, da je bila osumljenka priča vsaj takšnemu številu smrti, kot jih je dejansko doživela.

Po Elffersovem mnenju naj izračun ne bi bil povsem pošten do osumljenke. Bolj smiselno naj bi bilo, da se ne bi izračunavala verjetnost, da je bila osumljenka priča toliko smrtim, temveč verjetnost, da je bila neka medicinska sestra priča toliko smrtim bolnikov. V JKZ je za izmene skrbelo 27 medicinskih sester, zato je Henk Elffers domnevno, da bi dobil zgornjo mejo te verjetnosti, svoj rezultat pomnoži s 27, kar je imenovano naknadni popravek. Popravek naj bi bil poteben le za bolnišnico JKZ; v RKZ to ni potrebno, saj je bila osumljenka že opredeljen kot osumljenka na podlagi podatkov za bolnišnico JKZ. Prej omenjeni končni podatek, 1 proti 342 milijonom, je izračunal tako, da je pomnožil rezultate za vse tri bolnišnice.

Najbolj izpostavljena težava pri metodi, ki jo je uporabil Elfferson, je dvojna uporaba podatkov o JZK - uporabi jih za identifikacijo osumljenca in sum, da je prišlo do kaznivega dejanja, nato pa še pri izračuni verjetnosti. Gre za problem, ko se na podlagi določenih podatkov določi hipotezo, nato pa ta isti podatek uporabimo za preverjanje te hipoteze. Menim, da se je Henk Elfferson zavedal problema, saj je uvedel naknadni popravek glede tega.

8.2 Pristop z Bayesovo teorijo

Bayesov pristop bi rešil težave ponovne uporabe podatkov.

Naj bo

H_d ... Lucia de Berk je nedolžna;

H_p ... Lucia de Berk je kriva;

E ... razpoložljivi dokazi.

Z uporabo Bayesovega pravila dobimo

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

oziroma

$$\text{posteriorna verjetnost} = \text{razmerje verjetnosti} \times \text{predhodna verjetnost}$$

Verjetnost $P(H_d|E)$ je verjetnost hipoteze H_d po ovrednotenju dokazov E . Ob vsakih na novo pridobljenih dokazih lahko posodobimo verjetnosti na naslednji način. Naj bodo E' novi dokazi in

$$\text{stare posteriorne verjetnosti} = \frac{P(E|H_p)}{P(E|H_d)} \times \frac{P(H_p)}{P(H_d)},$$

potem, po pridobitvi novih dokazov, izračunamo nove posteriorne verjetnosti

$$\begin{aligned} \frac{P(H_p|E, E')}{P(H_d|E, E')} &= \frac{P(E' \cap E|H_p)}{P(E' \cap E|H_d)} \times \frac{P(H_p)}{P(H_d)} \\ &= \frac{P(E'|H_p, E)}{P(E'|H_d, E)} \times \text{stare posteriorne verjetnosti}. \end{aligned}$$

Menim, da imamo tudi tukaj podobne težave kot sem jih že opisala zgoraj - določitev verjetnosti hipotez $P(H_p)$ in $P(H_d)$, za katere dokaze je mogoče izračunati razmerje verjetnosti ter kako izračunamo verjetnost

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

saj ne vemo kako so različni dokazi med seboj povezani. Z določitvijo predhodnih verjetnosti lahko pridemo do zelo različnih rezultatov, ampak nimamo pa več težav z ponovno uporabo podatkov in naknadnimi popravki.

9 Bayesova analiza

Najbolj pogosta uporaba Bayesovega pravila je pri ugotavljanju, ali je obtoženec vir sledi DNK-ja s kraja zločina. V ta namen sem opisala poenostavljeno in izpopolnjeno Bayesovo analizo, v primeru, ko je dokaz DNK sled.

9.1 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.

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

Bayesovo pravilo lahko uporabimo na naslednji način:

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

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, kar 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 $P(M|\neg S)$, tj. 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, kar 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 pozitivnega in negativnega rezultata laboratorijske analize, sem si pogledala še izpopolnjeno Bayesovo analizo.

9.2 Izpopolnjena Bayesova analiza

Da upoštevam možnosti laboratorijskih napak, namesto M uvedem 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.

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

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

in zato

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

$P(M_p|\neg M_t)$ opisuje verjetnost lažno pozitivnih rezultatov laboratorijske analize (oznaka FP) in $P(M_p|M_t)$ verjetnost resničnih pozitivnih rezultatov laboratorijske analize (oznaka FN).

Sledi:

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

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

S predpostavko $P(M_p|S) = 1$ dobimo poenostavitev, kjer ni upoštevana možnost lažnega negativnega rezultata laboratorijske analize.

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:

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

Za predstavo, kako stopnje napak vplivajo na razmerje verjetnosti, predpostavim, da je pogostost profila DNK 1 proti milijardi. Predpostavim tudi, da sta stopnji lažno pozitivnih in negativnih laboratorijskih rezultatov (tj. FP in FN) enaki 0,01. Če je razmerje verjetnosti enako $\frac{1}{f}$, je milijarda. S formulo pa dobimo:

$$\frac{P(M_p|S)}{P(M_p|\neg S)} = \frac{1 - 0,01}{[(1 - 0,01) \times 0,000000001] + [0,01 \times (1 - 0,000000001)]} \approx 99.$$

Relativno majhne stopnje napak lahko bistveno zmanjšajo dokazno vrednost DNK dokazov, saj močno zmanjšajo razmerje verjetnosti; v našem primeru smo iz milijarde prišli na približno 100.

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. Bayesovo pravilo nam omogoča, da ta vidik upoštevamo.

Ker profili DNK predstavlja del naše genetske zasnove in imajo ljudje, ki so v sorodu, večjo verjetnost, da bodo imeli enak profil DNK, kot ljudje, ki niso v sorodu, morajo forenzični strokovnjaki svoje izjave vedno opredeliti z navedbo, da njihove ocene pogostosti veljajo za populacijo nepovezanih posameznikov. To spremenljivost pogostosti profila lahko v Bayesovem okviru upoštevamo na dva načina: s spremembo predhodne verjetnosti in s spremembo pogostosti profila. Izvedli bi lahko tudi različne izračune: enega za populacijo nepovezanih posameznikov in drugega za populacijo sorodnih posameznikov.

10 Frekvence

Da lahko ocenim moč Bayesove analize dokazov DNK, jo v nadaljevanju primerjam z nekaterimi drugimi pristopi. Ker je bilo predlagano s strani mnogih avtorjev, da je bolj naraven način za obravnavo verjetnosti uporaba naravnih frekvenc, sledi opis tega pristopa.

Definicija 10.1. Frekvenca f je posamezno število diskretnih statističnih enot iste vrednosti. Če je diskretnih podatkov zelo veliko ali če so podatki zvezni, jih združujemo v frekvenčne razrede.

Definicija 10.2. Absolutna frekvenca je število vrednosti statistične spremenljivke v k -tem razredu. Označimo jo z f_k za k -ti razred.

Definicija 10.3. Relativna frekvenca pa je delež absolutne frekvence f_k glede na celoto. Označimo jo z f'_k . Če je N število enot v populaciji ali morda vzorcu, je

$$f'_k = \frac{f_k}{N}.$$

V forenzičnem kontekstu se navedene frekvence običajno nanašajo na pojavljanje dokazov za posamezen primer, medtem ko se frekvence za vrednotenje dokazov običajno opisujejo kot osnovne stopnje. Sklepanje o krivdi je lahko podprto s statistično analizo ustreznih podatkov in verjetnostnim sklepanjem z uporabo absolutnih ali relativnih frekvenc, pri čemer je verjetnost, da bi določene podatke (dokaze) pridobili zgolj po naključju, izjemno majhna.

Relativne frekvence vedno navajajo ali predpostavljajo, da obstaja nek referenčni vzorec, na podlagi katerega se lahko oceni pogostost zadevnega dogodka. Nadaljna predpostavka je, da je ta primerjava poučna in pomembna za obravnavano zadevo. V okviru kazenskega postopka relativna frekvenca lahko podpre vmesno sklepanje o moči dokazov, ki se nanašajo na sporna dejstva, kar vodi do končnega sklepa, ali je obtoženec nedolžen ali kriv. Relativne frekvence so rutinsko vključene v znanstvene dokaze, ki se predložijo v kazenskih postopkih.

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 osumljencem, tako da se f pomnoži z m .

Če je posameznikov s takim profilom $n \geq 1$, 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.

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 f . Želimo vedeti, kolikšna je verjetnost, da ima točno en posameznik ustrezen DNK profil, ob pogoju da ga ima vsaj en posameznik, torej:

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

m	f	$P(n = 1 n \geq 1)$	$P(S M)[\text{Bayes}]$
10 milijonov	1 v 100 milojonov	0,9	0,9
100 milijonov	1 v 100 milojonov	0,62	0,5
1 milijarda	1 v 100 milojonov	0	0,09
10 milijonov	1 v 1 milijardi	1	0,99
100 milijonov	1 v 1 milijardi	0,9	0,9
1 milijarda	1 v 1 milijardi	0,62	0,5
10 milijonov	1 v 10 milijardah	1	0,999
100 milijonov	1 v 10 milijardah	1	0,99
1 milijarda	1 v 10 milijardah	0,9	0,9

Tabela 3: Primerjava rezultatov, dobljenih z frekvenčno metodo in Bayesovim pravilom.

Tabela vsebuje primerjavo rezultatov z rezultati, ki jih daje Bayesovo pravilo. Vidimo lahko, da so razlike zelo majhne, neke pa jih celo ni. Obe metodi, se lahko uporablja za reševanje določenih problemov, vendar se razlikujeta v načinu delovanja. Tako pri metodi edinstvenosti ni mogoče enostavno upoštevati številnih zapletov, kot je vpliv stopnje laboratorijskih napak. Bayesovo pravilo pa namreč omogoča upoštevanje različnih dejavnikov in zapletov, ki vplivajo na verjetnost določenega dogodka, s tem pa omogoča tudi bolj natančne izračune in posledično bolj zanesljive rezultate. Kljub temu pa ima tudi Bayesovo pravilo svoje omejitve, saj je odvisno od predpostavk in podatkov, ki jih uporabimo za izračun verjetnosti. V vsakem primeru je izbira metode odvisna od specifičnih zahtev problema in od tega, katera metoda bo najbolj primerna za njegovo reševanje.

Res je, da so frekvenčni modeli pomembni v primerih, kjer so na voljo DNK ali druge vrste dokazov, in jih je mogoče uporabiti za prikaz, kako verjetno je, da bi se naključno izbrana oseba ujemala z vzorcem ali v primerih z velikim številom žrtev za izbiro statistično ustreznih vzorcev celotne populacije žrtev. Kljub temu pa imajo ti modeli notranje pomanjkljivosti, ki jih ni mogoče zanemariti.

Bistvena teoretična pomanjkljivost frekvenčnih modelov je, da zahtevajo statistične dokaze, ki sodišču niso na voljo. Sodišča ne morejo številčno ovrednotiti nekega dejanskega dokaza, saj se ne zavedajo možnih anomalij pri uporabi drugih sredstev in storitev. Na primer, če priča trdi, da je videla osumljenca na kraju zločina, sodišče ne more oceniti verjetnosti, da je opazovanje določene priče v skladu s tem, kar se je dejansko zgodilo. To je zato, ker sodišče ponavadi nima na vseh voljo informacij o tem, koliko drugih ljudi bi lahko bilo na kraju zločina ob istem času. Poleg tega frekvenčni modeli temeljijo na predpostavki, da visoka vrednost verjetnosti, ki opisuje razmerje med obstoječimi dokazi in primerom, pomeni, da je vrednost tega dokaza visoka. Vendar to ni nujno vedno res. Merjenje skladnosti med predloženimi dokazi in tistim, kar se je resnično zgodilo, temelji na predpostavki, da obstajata reprezentativna populacija in skladen rezultat. V kazenskem primeru pa ti pogoji niso izpolnjeni, kar pomeni, da je uporaba frekvenčnih modelov lahko omejena.

11 Metoda verjetnosti naključnega ujemanja

Metoda verjetnost naključnega ujemanja izraža možnost, da bi imel naključni posameznik, ki ni povezan z obdolžencem, ustrezeni 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 milijonov, potem je verjetnost, da ima profil DNK drug posameznik in ne obdolženec 1 proti 100 milijonov;
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.

Sestavlja jo enačba

$$1 - f = P(S|M). \quad (11.1)$$

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, in sicer kot:

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

12 Razmerje verjetnosti

Občasno se zgodi, da predloga tožilstva in obrambe nista komplementarna in v takih primerih ni mogoče določiti $P(H_p)$ ali $P(H_d)$ (poglavje 1), ampak samo vpliv statistike, znane kot razmerje verjetnosti.

12.1 Opredelitev

V Bayesovi formuli (7.2) nadomestim H z \bar{H} in enakovredna različica Bayesovega izreka je

$$P(\bar{H}|E) = \frac{P(E|\bar{H})P(\bar{H})}{P(E)}, \quad (12.1)$$

kjer $P(E) \neq 0$.

Če prvo enačbo delim z drugo, dobim verjetnostno obliko Bayesovega izreka

$$\frac{P(H|E)}{P(\bar{H}|E)} = \frac{P(E|H)}{P(E|\bar{H})} \times \frac{P(H)}{P(\bar{H})}. \quad (12.2)$$

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 kakršnihkoli informacij o dogodku E .

Definicija 12.1. Razmerje

$$\frac{P(E|H)}{P(E|\bar{H})} \quad (12.3)$$

se imenuje razmerje verjetnosti.

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 proučevanju vpliva E na H je treba upoštevati tako verjetnost E , ko je H resničen in ko je H neresničen. Pogosta napaka, tj. zmotna 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. Torej iz Bayesovega izreka neposredno izhaja, da če je razmerje verjetnosti večje od 1, potem dokaz povečuje verjetnost krivde (pri čemer višje vrednosti pomenijo večjo verjetnost krivde), če pa je manjše od 1, zmanjšuje verjetnost krivde (in bolj ko se približuje ničli, manjša je verjetnost krivde).

12.2 Razmerje verjetnosti v kazenskem pravu

Če sedaj pogledamo obliko Bayesovega izreka o verjetnosti v forenzičnem kontekstu, tj. ocenjevanje vrednosti nekaterih dokazov.

Naj bo:

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

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

$Ev \dots$ obravnavani dokaz - nadomestimo dogodek E ;

Oblika Bayesovega izreka nato omogoča, da se predhodne verjetnosti, tj. pred predstavitvijo obravnavanih dokazov 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.

Definicija 12.2. 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

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

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

Tabela 4: Kvalitativna lestvica za poročanje o vrednosti V podpore dokazov za H_p proti H_d .

Pri uporabi takšnih tabel moramo biti previdni, če so hipoteze prelagane na podlagi podatkov; v tem primeru postanejo smiselne le če podamo predhodne verjetnosti za obravnavane hipoteze.

Slovar strokovnih izrazov