math.e

Hrvatski matematički elektronički časopis

Primjena Bayesove formule i algoritamskog pristupa Bayesovoj formuli na situacijama iz svakodnevnog života

algoritamski pristup Bayesovoj formuli **Bayesova formula** **potpun sustav događaja** **primjena u svakodnevnom životu**

Dragana Jankov Maširević, docentica na Odjelu za matematiku, Sveučilište u Osijeku, Trg Ljudevita Gaja 6, HR-31000 Osijek djankov@mathos.hr i Matea Klarić, student na Odjelu za matematiku, Sveučilište u Osijeku, Trg Ljudevita Gaja 6, HR-31000 Osijek mklaric@mathos.hr

Sažetak

U ovom članku je opisana Bayesova formula kao i algoritamski pristup korištenja Bayesove formule, koji je pogodan zbog jednostavnosti računanja uz upotrebu tablice (umjesto direktnog uvrštavanja u formulu). Također, dani su primjeri problema iz svakodnevnoga života čije je rješenje ilustrirano primjenom oba navedena pristupa.

1 Uvod

Engleski statističar i filozof Thomas Bayes (1702.–1762.) dokazao je čuveni teorem teorije vjerojatnosti koji po njemu nazivamo *Bayesov teorem* ili češće *Bayesova formula*. Kako bi uveli Bayesovu formulu potrebno je napraviti dekompoziciju prostora elementarnih događaja Ω na skup nepraznih, disjunktnih događaja $\mathcal{H} = \{H_i : i \in I \subseteq \mathbb{N}\} \subseteq \mathcal{F}$(pri čemu je $\mathcal{F}$ uobičajena oznaka za σ–algebru na Ω, a njezine elemente nazivamo *događaji*) takvih da je $\Omega = \bigcup_{i \in I} H_i$. Familiju $\mathcal{H}$ tada nazivamo *potpun sustav događaja*, a događaje H_i, $i \in I \subseteq \mathbb{N}$*hipoteze*. Sada Bayesovu formulu možemo interpretirati kao vjerojatnost hipoteze ako znamo da se dogodio određeni događaj [p. 64]-[9]. Ova formula ima široku primjenu, pogotovo u medicinskoj dijagnostici (vidi npr. [p. 23]-[6], [p. 184]-[8] , [p. 65]-[9] , [str. 43]-[10]) gdje može biti korisna za određivanje uzroka bolesti, o čemu ćemo reći nešto više u zadnjem dijelu.

2 Što je Bayesova formula i zašto je koristan algoritamski pristup?

Prije nego što je dokazao svoj poznati teorem, Bayes se bavio pitanjem uvjetne vjerojatnosti, odnosno vjerojatnosti da se dogodio događaj A ako nam je poznato da se realizirao događaj B, čiju je definiciju zapisao u sljedećem obliku [str. 51]-[4]

$$P(A|B) = \frac{P(A \cap B)}{P(B)}\,. \tag{1}$$

Takvu vjerojatnost još označavamo s $P_B(A)$ i često je nazivamo *uvjetna vjerojatnost od A uz uvjet B*.

Uz uvjetnu vjerojatnost povezujemo i *formulu potpune vjerojatnosti* koja tvrdi da vjerojatnost nekog proizvoljnog događaja $A \in \mathcal{F}$ možemo računati tako da odaberemo odgovarajuću familiju $\mathcal{H}$ koja čini *potpun sustav događaja*, a zatim vjerojatnost danog događaja A računamo kao (vidi [str. 36, Teorem 1.1]-[3])

$$P(A) = \sum_{i \in I} P(H_i)P(A|H_i). \tag{2}$$

Poznavanjem uvjetne vjerojatnosti (1) i formule potpune vjerojatnosti (2) možemo računati vjerojatnosti realizacije hipoteza iz potpunog sustava događaja uz informaciju o tome koji se događaj dogodio nakon izvođenja pokusa, odnosno

$$P(H_i|A) = \frac{P(H_i)P(A|H_i)}{P(A)}, \qquad i \in I \subseteq \mathbb{N}, \tag{3}$$

pri čemu vjerojatnost događaja A računamo uz pomoć formule (2). Kao što smo već spomenuli u uvodnom poglavlju, prethodnu formulu nazivamo *Bayesova formula*.

Račun koji uključuje Bayesovu formulu ponekad je dovoljno kompliciran za stvaranje mogućnosti pogreške ili nepravilne zamjene danih vrijednosti vjerojatnosti, stoga ćemo u nastavku opisati te koristiti i *algoritamski pristup* za određivanje vjerojatnosti koji je često jednostavniji, a pristupačan je i osobama koje ne poznaju zakone vjerojatnosti. U literaturi ga često možemo pronaći pod nazivom *intuitivni pristup Bayesovoj formuli*. Ovakav je algoritam prikladan jer se ostvaruje uz pomoć tablice, bez direktnog uvrštavanja u formulu. Naime, algoritam se temelji na ideji da na osnovi poznatih (zadanih) vjerojatnosti konstruiramo tablicu s odgovarajućim frekvencijama za svaku pojedinu mogućnost posebno. Ipak, treba uočiti da se ovakav tzv. *Bayesovski pristup* razlikuje od *frekvencionističkog pristupa*. Naime, statistička teorija se ponekad dijeli upravo na ta dva pristupa [1, 2, 12] i oni se razlikuju s obzirom na način na koji interpretiraju pojam vjerojatnosti: dok je frekvencionistička statistika poznata i kao *klasična* te vjerojatnost promatra kao limes relativnih frekvencija, zagovornici Bayesovske statistike smatraju da parametri statističkih modela moraju i sami biti modelirani kao slučajni. Kod Bayesovskog pristupa prije prikupljanja podataka imamo polaznu ideju o razdiobi nepoznatog parametra pri čemu se stupanj uvjerenja mijenja dolaskom novih informacija.

U literaturi često možemo naići na uputu da prilikom korištenja algoritamskog pristupa krećemo od pretpostavke da je skup jedinki koji promatramo *velik*, što se uobičajeno odnosi na broj veći ili jednak $10\,000$ [7, 8, 10, 13]. No, važno je napomenuti da rješenje ne ovisi o odabranom broju, što nije slučaj kod frekvencionističkog pristupa gdje promatrani pokus zaista moramo ponoviti veliki broj puta kako bi dobili što točniju vjerojatnost (sjetimo se najjednostavnijeg pokusa bacanja simetričnog novčića [str. 7, Primjer 1.8]-[3]). Stabilnost ovakvog algoritma ne ovisi o broju jedinki koje promatramo, već taj broj biramo u ovisnosti o tome koliko točnih znamenaka želimo dobiti prilikom aproksimacije rješenja: ako želimo dobiti n točnih znamenaka za polazni skup ćemo uzeti broj 10^{n+1}, što ćemo i ilustrirati na Primjeru 1. Dakle, čitatelj treba uočiti da odabrani broj jedinki nije apsolutan, nego u potpunosti ovisi o podacima s kojima raspolažemo.

3 Primjena Bayesove formule i algoritamskog pristupa

Na primjerima koji slijede ilustrirat ćemo primjenu Bayesove formule, a zatim i algoritamskog pristupa Bayesovoj formuli. Kada čitatelj usporedi rješenja dobivena ovim dvama pristupima uočit će da su ona jednaka, pri čemu aproksimacija rješenja algoritamskim pristupom ovisi, kako smo već istaknuli, o izboru broja jedinki skupa kojeg promatramo.

Primjer 1. *Ana će se uskoro početi pripremati za državnu maturu i potrebne su joj instrukcije iz matematike. U oglasima je pronašla veliki broj kontakt telefona i jedina dodatna informacija je da instrukcije drže osobe koje su diplomirale.*

- *[a)] Ana se pita kolika je vjerojatnost da osoba koju slučajno odabere iz oglasnika i nazove radi dogovora bude žena. Iz radoznalosti je otišla na stranice Državnog zavoda za statistiku (vidi* [14]*) i tamo pronašla podatke o diplomiranim studentima u 2014. godini u kojima piše da je te godine u Hrvatskoj diplomirao* $33\,741$ *student od kojih* 59.9% *čine žene, a ostalo muškarci. Također, na istim je stranicama pronašla zanimljiv podatak da od ukupnog broja diplomiranih muških studenata u 2014. godini njih* $6\,157$ *(odnosno* 45.5%*) spada u dobnu skupinu od najviše* 24 *godine, dok u istu skupinu spada* 51.7% *žena.*
- *[b)] Ana je u međuvremenu od urednika oglasnika, koji joj je prijatelj, saznala da su sve osobe iz oglasnika mlađe od* 27 *godina, što znači da su studenti koji su diplomirali 2014. godine tada imali najviše* 24 *godine. Kolika je sada vjerojatnost da će se Ani na slučajno odabrani broj telefona javiti osoba ženskog spola?*

Rješenje Bayesovom formulom. Potpun sustav događaja u ovom primjeru čine skupovi

Ž = $\{$ slučajno odabrana osoba je žena$\}$,
Ž^c = $\{$slučajno odabrana osoba je muškarac$\}$,

a od interesa su nam i događaji

G = $\{$slučajno odabrana osoba ima najviše 24 godine $\}$,
G^c = $\{$slučajno odabrana osoba ima više od 24 godine$\}$.

- [a)] Prema podacima koje je Ana pronašla na stranicama Državnog zavoda za statistiku znamo da 59.9% studenata koji su diplomirali 2014. godine čine žene, dakle pri slučajnom odabiru osobe iz oglasnika vjerojatnost da to bude žena je $P(\text{Ž}) = 0.599$.
- [b)] Kako je $P(\text{Ž}) = 0.599$, lako dobivamo i vjerojatnost suprotnog događaja $P(\text{Ž}^c) = 1 - 0.599 = 0.401$ Nadalje, s obzirom na dodatnu informaciju da od ukupnog broja diplomiranih ženskih studenata u 2014. godini 51.7% spada u dobnu skupinu od najviše 24 godine možemo zaključiti da je $P(G|\text{Ž}) = 0.517$, dok je prema podacima koje imamo za diplomirane studente muškog spola $P(G|\text{Ž}^c) = 0.455$. Kako familija $\{\text{Ž},\ \text{Ž}^c\}$ čini potpun sustav događaja, primjenom Bayesove formule (3) dobivamo

$$P(\text{Ž}|G) = \frac{P(\text{Ž}) \cdot P(G|\text{Ž})}{P(\text{Ž}) \cdot P(G|\text{Ž}) + P(\text{Ž}^c) \cdot P(G|\text{Ž}^c)}$$

$$= \frac{0.599 \cdot 0.517}{0.599 \cdot 0.517 + 0.401 \cdot 0.455}$$

$$= 0.62926 \approx 0.629.$$

Kada smo u obzir uzeli uvjet da su sve osobe iz oglasnika mlađe od 27 godina, vjerojatnost da će se Ani na telefon javiti osoba ženskog spola se povećala, jer je veći broj žena nego muškaraca koji su u 2014. godini diplomirali s najviše 24 godine.

Rješenje algoritamskim pristupom. Kao što smo naglasili u prethodnom poglavlju, najprije je potrebno odabrati određeni broj koji predstavlja broj jedinki promatranog skupa o čemu će ovisiti i točnost aproksimacije. U svrhu ilustracije tog pravila, u nastavku ćemo promotriti slučaj *jedne točne znamenke*, odnosno dobiveno rješenje će se u jednoj znamenci podudarati s rješenjem dobivenim Bayesovom formulom, te slučaj *tri točne znamenke*.

Jedna točna znamenka: Pretpostavimo najprije da je broj oglasa u kojima se nalaze informacije o instrukcijama iz matematike za državnu maturu 100. Zatim, kako je poznato (prema statističkim podacima koje je Ana pronašla te na osnovu kojih donosi svoje zaključke) da se među studentima koji su diplomirali 2014. godine nalazi 59.9% žena, slijedi da njih ima $0.599 \times 100 = 59.9$, odnosno 59, dok je u istoj populaciji 41 muškarac.

Kako od ukupnog broja diplomiranih ženskih studenata u 2014. godini u dobnoj skupini od najviše 24 godine pripada 51.7% žena, onda je takvih žena u našoj populaciji $0.517 \times 59 = 30.5$, odnosno 30. Dakle, žena koje su 2014. godine diplomirale s više od 24 godine ima $59 - 30 = 29$. Na isti način možemo zaključiti da je broj muškaraca koji su 2014. godine diplomirali s najviše 24 godine jednak 19, dok je onih koji su tada imali više od 24 godine bilo 22.

Radi preglednijeg zapisa, dobivene podatke možemo zapisati u obliku sljedeće tablice:

	G (dob ≤ 24)	G^c (dob> 24)	Ukupno
Ž (žene)	30	29	59
Ž^c (muškarci)	19	22	41
Ukupno	49	51	100

Sada iz tablice možemo iščitati odgovor na naše pitanje, odnosno možemo pronaći vjerojatnost da broj telefona koji je Ana na slučajan način odabrala iz oglasnika pripada ženskoj osobi koja je 2014. godine diplomirala s najviše 24 godine. Naime, među 49 osoba koje su 2014. godine diplomirale s najviše 24 godine, nalazi se 30 žena, pa je vjerojatnost koju tražimo jednaka $30 \div 49 = 0.612245 \approx 0.612$, što se u prvoj decimali podudara s vjerojatnosti koju smo prethodno izračunali primjenom Bayesove formule.

Tri točne znamenke: Ukoliko pretpostavimo da je broj oglasa u kojima se nalaze informacije o instrukcijama iz matematike za državnu maturu $10\,000$, na isti način, kao u prethodno opisanom slučaju, dolazimo do sljedeće tablice:

	G (dob ≤ 24)	G^c (dob> 24)	Ukupno
Ž (žene)	$3\,097$	$2\,893$	$5\,990$
Ž^c (muškarci)	$1\,825$	$2\,185$	$4\,010$
Ukupno	$4\,922$	$5\,078$	$10\,000$

Sada je vjerojatnost da broj telefona koji je Ana na slučajan način odabrala iz oglasnika pripada ženskoj osobi koja je 2014. godine diplomirala s najviše 24 godine jednaka $3\,097 \div 4\,922 = 0.629$ što je upravo vjerojatnost $P($ Ž$|G)$ koju smo prethodno izračunali primjenom Bayesove formule.

$\triangle$

Napomena 2. *U prethodnom primjeru familiju koja predstavlja potpun sustav događaja činila su dva skupa, jer smo imali podjelu na žene i muškarce. Kao što smo vidjeli u uvodnom dijelu, potpun sustav događaja može činiti i prebrojiva familija događaja. U sljedećem primjeru ilustrirat ćemo Bayesovu formulu i algoritamski pristup u slučaju kada imamo tri hipoteze. S obzirom da ćemo rješenje dobiveno uvrštavanjem u Bayesovu formulu zaokružiti na tri decimale, to ćemo u algoritamskom pristupu odabrati da je broj jedinki u promatranom skupu* $10^{3+1} = 10\,000$ *te će iz istog razloga taj broj u Primjeru 5 iznositi* $100\,000$*. No, u završnom dijelu rada vidjet ćemo, kroz slučaj u medicini, da je izbor promatranog skupa jedinki potpuno arbitraran. Također, napominjemo da u prethodnom primjeru nismo zaokruživali brojeve na standardni način [str. 12]-[5], jer je uobičajeno kod zaokruživanja decimalnog broja osoba uzeti najveći cjelobrojni dio.*

Primjer 3. *Kako bi se što bolje pripremila za polaganje mature iz matematike, Ana je preko interneta naručila zbirku riješenih zadataka koji su bili na maturi prethodnih godina, ne gledajući pri tome koji je izdavač. Od prijatelja koji su već polagali maturu čula je da su takvu zbirku objavile tri izdavačke kuće te da prva izdavačka kuća ima najbolju reklamu na internetu i ona štampa 80% zbirki iz kojih se budući studenti pripremaju, druga izdavačka kuća štampa 15% zbirki, dok je treća izdavačka kuća tek počela s radom i ona štampa 5% takvih zbirki. Prijatelji su također rekli Ani da ponekad naručene zbirke stignu na kućnu adresu nepotpune, odnosno u njima se nalaze samo tekstovi zadataka, a ne i rješenja. Naime, zbirke koje štampa prva izdavačka kuća štampaju se bez rješenja u 4% slučajeva, u drugoj izdavačkoj kući je taj postotak 6%, dok posljednja izdavačka kuća štampa 9% zbirki koje ne sadrže rješenje.*

a) Ana se zabrinula jer nije pogledala izdavača kada je naručivala i pita se kolika je vjerojatnost da je zbirku naručila od prvog izdavača?

b) Kada je zbirka stigla na kućnu adresu Ana je bila u školi, ali je njezin brat odmah otvorio paket i vidio da je zbirka stigla bez rješenja! Odmah je poslao poruku Ani. Sada kada je sigurna da zbirka nije stigla u obliku u kom se nadala, Ana se ponovno pita kolika je vjerojatnost da se radi o prvom izdavaču. Razlikuje li se ta vjerojatnost od one dobivene u prethodnom slučaju?

Rješenje Bayesovom formulom.
Možemo uočiti da potpun sustav događaja čine sljedeća tri skupa:

H_1 = $\{$ zbirka dolazi od prvog izdavača $\}$,
H_2 = $\{$zbirka dolazi od drugog izdavača $\}$,
H_3 = $\{$zbirka dolazi od trećeg izdavača$\}$.

Također, zanimaju nas i događaji

N = $\{$zbirka je nepotpuna (stigla je Ani bez rješenja)$\}$
N^c = $\{$zbirka je potpuna (stigla je s rješenjima)$\}$.

a) Ana je zbirku naručila slučajnim odabirom (nije gledala tko je izdavač) te je vjerojatnost da je ona izdana od strane prvog izdavača jednaka 0.8, jer prvi izdavač štampa 80% zbirki iz kojih budući studenti vježbaju.

b) Kada je Ana saznala da je zbirka došla bez rješenja, koristeći tu dodatnu informaciju krenula je opet izračunati vjerojatnost da ona dolazi od prvog izdavača, odnosno vjerojatnost $P(H_1|N)$, za što su joj trebali i sljedeći podaci koje je najprije zapisala iz podataka koje zna:

$$\begin{aligned}
P(H_1) &= 0.80 \qquad \text{(jer prvi izdavač štampa } 80\% \text{ zbirki)},\\
P(H_2) &= 0.15 \qquad \text{(jer drugi izdavač štampa } 15\% \text{ zbirki)},\\
P(H_3) &= 0.05 \qquad \text{(jer treći izdavač štampa } 5\% \text{ zbirki)},
\end{aligned}$$

$$\begin{aligned}
P(N|H_1) &= 0.04 \quad \text{(jer } 4\% \text{ zbirki bez rješenja dolazi od prvog izdavača)},\\
P(N|H_2) &= 0.06 \quad \text{(jer } 6\% \text{ zbirki bez rješenja dolazi od drugog izdavača)},\\
P(N|H_3) &= 0.09 \quad \text{(jer } 9\% \text{ zbirki bez rješenja dolazi od trećeg izdavača)}.
\end{aligned}$$

Uvrštavanjem navedenih vrijednosti u Bayesovu formulu (3) dobivamo

$$P(H_1|N) = \frac{0.80 \cdot 0.04}{0.80 \cdot 0.04 + 0.15 \cdot 0.06 + 0.05 \cdot 0.09} = 0.703.$$

Dakle, uz dani uvjet da je zbirka došla bez rješenja, vjerojatnost da ona dolazi od prvog izdavača se smanjila jer prvi izdavač štampa najmanji postotak nepotpunih zbirki, odnosno onih koje ne sadrže rješenja.

Rješenje algoritamskim pristupom. Kao u prethodnom primjeru, koristeći ovaj pristup, pronaći ćemo vjerojatnost $P(H_1|N)$ uz pomoć tablice. Pretpostavimo da imamo uzorak od $10\,000$ zbirki koje su puštene u prodaju. Prema danim podacima imamo $8\,000$ zbirki koje su izdane od strane prvog izdavača među kojima je 320 nepotpunih, odnosno štampanih bez rješenja. Dakle, imamo $7\,680$ potpunih zbirki od prvog izdavača. Analognim postupkom dobivamo i preostale vrijednosti koje su zapisane u sljedećoj tablici:

	N (nepotpune zbirke)	N^c (potpune zbirke)	ukupno
H_1	320	7 680	8 000
H_2	90	1 410	1 500
H_3	45	455	500
ukupno	455	9 545	10 000

Zanima nas vjerojatnost da je zbirku izdao prvi izdavač pri čemu znamo da je ona Ani stigla nepotpuna, odnosno ne sadrži potrebna rješenja. S obzirom da znamo da je zbirka nepotpuna, zanima nas prvi stupac prethodne tablice gdje možemo vidjeti da od ukupno 455 nepotpunih zbirki njih 320 dolazi od prvog izdavača, pa je tražena vjerojatnost jednaka $320 \div 455 \approx 0.703$. Možemo uočiti da smo isti rezultat dobili rješavanjem zadatka uz pomoć Bayesove formule.

$\triangle$

Napomena 4. *Prethodni primjer predstavlja ilustraciju Bayesove formule i algoritamskog pristupa u slučaju kada imamo tri hipoteze. Na osnovu toga, nije teško uočiti kako bi takve pristupe primjenili u slučaju četiri ili više hipoteza.*

4 Primjena u medicini

Kao što smo naveli u uvodnom dijelu, Bayesova formula se često primjenjuje u medicinskoj dijagnostici. Na primjer [str. 65]-[9], pretpostavimo da se na nekoj klinici liječe bolesnici od kojih svaki može imati jednu od bolesti $H_1, \ldots, H_n$ te da je kod slučajno odabranog bolesnika nakon pregleda ustanovljen skup od A simptoma. Uz pretpostavku da se na klinici bilježe statistički podaci o broju bolesnika koji su imali bolest H_i, $i = 1, \ldots, n$ možemo izračunati pripadne vjerojatnosti $P(H_i)$ i $P(A|H_i)$, $i = 1, \ldots, n$ što su, redom, relativna frekvencija bolesti H_i i relativna frekvencija utvrđivanja skupa simptoma A među svim bolesnicima koji su bolovali od bolesti H_i (relativna frekvencija je približno jednaka odgovarajućoj vjerojatnosti te k njoj teži s povećanjem uzorka koji promatramo). Ovakvi podaci nam omogućavaju primjenu Bayesove formule (ili algoritamskog pristupa) koja će nam dati odgovor na pitanje kolika je vjerojatnost da osoba ima bolest H_i ako je kod nje ustanovljen skup simptoma A. Primjenu ovih pristupa u medicinskoj dijagnostici ilustrirat ćemo sljedećim primjerom [11].

Primjer 5. *Ani je teško vježbati zadatke za državnu maturu jer ima simptome alergije. Već je išla na razna ispitivanja, ali svi testovi su bili negativni. Ostala je samo mogućnost da ima vrlo rijedak oblik alergije koji se nasumično pojavljuje kod jedne osobe u populaciji od $10\,000$ ljudi. Ako se Ana odluči testirati, uz pretpostavku da test daje točan rezultat u 99% slučajeva, i ako rezultat testa bude pozitivan, kolika je zaista vjerojatnost da Ana ima pomenutu alergiju?*

Rješenje Bayesovom formulom. Ako s A označimo događaj da osoba ima navedeni rijedak oblik alergije, a s B događaj da je rezultat testa pozitivan, možemo zaključiti da je $P(B|A) = 0.99$ i $P(A) = 0.0001$. Sada, primjenom formule (2), uz dani potpun sustav događaja $\{A,\ A^c\}$ dobivamo

$$P(B) = P(B|A)\,P(A) + P(B|A^c)\,P(A^c) = 0.99 \cdot 0.0001 + 0.01 \cdot 0.9999 = 0.01\,.$$

Primjenom Bayesove formule imamo da je

$$P(A|B) = \frac{P(A)\,P(B|A)}{P(B)} = 0.0099,$$

odnosno vjerojatnost da osoba ima alergiju, uz uvjet da će test biti pozitivan je manja od 1%, a razlog ovakvog zaključka je činjenica da je navedeni oblik alergije jako rijedak.

Rješenje algoritamskim pristupom. Vodeći se algoritamskim pristupom, s obzirom da nam je već poznato da u populaciji od $10\,000$ ljudi jedna (slučajna) osoba ima spomenutu rijetku alergiju, pretpostavimo da promatramo populaciju od $100\,000$ osoba. Jasno je da će tada 10 ljudi imati navedenu alergiju od čega će u 9 slučajeva test biti pozitivan, dok će u 999 slučajeva test biti lažno pozitivan. Zapisivanjem svih podataka u tablicu dobivamo

	B (test je pozitivan)	B^c (test je negativan)	ukupno
A	9	1	10
A^c	999	98 991	99 990
ukupno	1 008	98 992	100 000

Zanima nas vjerojatnost da u slučaju pozitivnog testa (što se dešava u $1\,008$ slučajeva) osoba zaista ima alergiju (takvih je osoba 9) te je tražena vjerojatnost jednaka $9 \div 1\,008 \approx 0.0089$, što je ponovno manje od 1% kao što smo već zaključili primjenom prethodnog pristupa.
Ukoliko čitatelj pretpostavi da promatrana populacija broji $1\,000\,000$ osoba, tražena vjerojatnost će iznositi približno 0.0098 što nas vodi do istog zaključka, no s rješenjem dobivenim Bayesovom formulom podudara se samo u prve tri znamenke.

$\triangle$

5 Zaključak

Kao što smo vidjeli u danim primjerima, Bayesova formula i algoritamski pristup Bayesovoj formuli se mogu primijeniti u raznim situacijama iz svakodnevnog života i dati nam odgovore na pitanja vezana uz vjerojatnosti neke od hipoteza koja nam je od interesa, a uz poznavanje dodatne informacije koja predstavlja vjerojatnost događaja koji se realizirao nakon izvođenja pokusa. Ipak, prilikom primjene ovih metoda u praksi moramo biti pažljivi [p. 65]-[9] jer su ponekad pored zadanih podataka koji su nam potrebni kako bi došli do određenih zaključaka, kao u Primjeru 3 kada ispitujemo vjerojatnost u medicinske svrhe, potrebne i neke dodatne netrivijalne pretpostavke. U spomenute pretpostavke ubrajamo, primjerice, veću vjerojatnost obolijevanja od određene bolesti temeljenu na nasljednom faktoru te razne čimbenike okruženja osoba koje promatramo, kao što su zagađenje zraka ili vode, i slično.

Bibliografija

[1] G. J. Babu, *Bayesian and frequentist approaches*, Online Proceedings of the Astronomical Data Analysis Conference (ADA VII), 2012.

[2] B. Basrak, *Aktuarska matematika II, 2. dio* (predavanja). PDF.
https://web.math.pmf.unizg.hr/∼bbasrak/pdf_files/AM2slides2dio.pdf

[3] M. Benšić, N. Šuvak, *Uvod u vjerojatnost i statistiku*, Grafika d.o.o., Osijek, 2014.

[4] F. M. Brückler, *Povijest Matematike II*, Zagreb, 2009.

[5] F. M. Brückler, I. Pažanin, *Matematika*1 za kemičare; Kako prevoditi s jezika kemije na jezik matematike i obrnuto?}, Zagreb, 2012.
http://prelog.chem.pmf.hr/∼fmbruckler/main1.pdf

[6] B. Draščić Ban, T. Poganj, *Primijenjena matematika*, Pomorski fakultet, Sveučilište u Rijeci (autorizirana predavanja i vježbe), Rijeka, 2009.

[7] M. Klarić, *Formula potpune vjerojatnosti i Bayesova formula*, završni rad, Sveučilište J.J. Strossmayera, Odjel za matematiku, Osijek, 2016.

[8] G. F. Luger, *Artificial intelligence, Structures and Strategies for Complex Problem Solving*, Addison–Wesley, Harlow, 2005.

[9] N. Sarapa, *Vjerojatnost i statistika*, Školska knjiga, Zagreb, 1993.

[10] J. V. Stone, *Bayes' Rule, A Tutorial Introduction to Bayesian Analysis*, Sebtel Press, Zagreb, 1993.

[11] F. E. Su, *Medical Tests and Bayes' Theorem*, Math Fun Facts
Dostupno: http://www.math.hmc.edu/funfacts

[12] T. Topić, *Bayesova statistika i procjena vrijednosti ulaganja*, diplomski rad, Sveučilište u Zagrebu, Prirodoslovno–matematički fakultet, Matematički odsjek, Zagreb, 2016.

[13] M. F. Triola, *Bayes' Theorem.* PDF.
http://faculty.washington.edu/tamre/BayesTheorem.pdf

[14] http://www.dzs.hr/

ISSN 1334-6083
© 2009 **HMD**