

Drugačiji pogled na statistiku

Damjan Krstajić

Moje prvobitno reagovanje kad čujem neku novu naučnu teoriju je nepoverenje, želja da nađem njene mane i po mogućstvu je odbacim. Stoga najveće misaone muke mi predstavljaju one teorije koje nisam u stanju ni da prihvatim ni da odbacim. E takav je Bajesov pogled na verovatnoću i statistiku.

Tomas Bajes (Thomas Bayes) je engleski teolog koji je u 18. veku došao do važne jednačine o inverznoj verovatnoći koja je po njemu i nazvana. On ju je zapisao u svojim papirima, a njegov prijatelj posthumno objavio sa naglaskom da je Bajes autor. Ono što je Pitagorina teorema u geometriji, to je Bajesova formula u teoriji verovatnoće. Kasnije su se drugi matematičari nadovezali i toliko dopunili svojim nalazima da sumnjam da bi Bajes danas, da može, uopšte prepoznao sebe u svemu onome što nosi njegovo ime.

Bajesova statistika je zahvaljujući pojavi računara doživela procvat krajem prošlog veka. Drugačijeg je pristupa od onih koji se mogu naći u našim udžbenicima i zbog toga zaslužuje dodatnu pažnju. Razvila se toliko da u svetskoj statistici danas razlikujemo Bajesovce i frekvencioniste, zastupnike standardne statistike. Rasprava među njima obiluje obostrano jakim argumentima i traži se matematičar koji će da je reši.

Bez uvođenja jednačina, dakle pojednostavljeno, razlika između ta dva pristupa bi se mogla svesti na sledeće:

**STANDARDNA
(FREKVENCIONISTIČKA)
STATISTIKA**

PODACI → VEROVATNOĆA

**BAJESOVA
STATISTIKA**

PRETHODNA VEROVATNOĆA + PODACI → NAKNADNA VEROVATNOĆA

(PRIOR PROBABILITY + DATA → POSTERIOR PROBABILITY)

U standardnoj statistici verovatnoća se računa samo na osnovu dobijenih podataka, dok kod Bajesove statistike imamo neku verovatnoću u hipotezu pre početka eksperimenta, dobijamo rezultate i na osnovu jednog i drugog, uz pomoć jednačina izračunamo novu verovatnoću. Međutim, na osnovu čega baziramo našu prethodnu verovatnoću? Zar to nije subjektivno, odnosno nenaučno? Bajesovci smatraju da je to dopustivo kad na osnovu nekih ranijih saznanja imamo neki stepen verovanja u hipotezu, dok frekvencionisti kažu da se tako uvodi subjektivnost u nauku.

Britanski matematičar Harold Džefris (Harold Jeffreys) je 1939. godine objavio fundamentalnu knjigu o teoriji verovatnoća u kojoj je detaljno opisao i obrazložio Bajesov pristup. Iako je do tada dao veliki doprinos geofizici i astronomiji, njegova knjiga je bila ignorisana od strane tadašnjih vodećih svetskih statističara. Bajesovci su većim delom 20. veka bili tretirani kao heretici. Pretpostavljam da je osnovni razlog tome što njihov pristup, gde imamo nešto apstraktno kao prethodna verovatnoća, nije bio praktičan.

Pojava moćnijih računara i veća količina informacija je promenila situaciju. Kompjuteri su omogućili da se simulira na hiljade mogućih prethodnih verovatnoća i da ispitamo kako se to odražava na naknadne verovatnoće. Takođe, na osnovu dostupnih podataka u nekim slučajevima možemo objektivno uzeti u obzir prethodne verovatnoće.

Pre desetak godina američka Agencija za lekove i hranu (FDA), inače poznata po veoma strogim statističkim principima, je prihvatila upotrebu Bajesove statistike u trajalima za medicinske instrumente. Ovo odobrenje može nekima, kao meni, da puno govori o njenoj validnosti. Međutim, jedna osoba je uspela da pokaže vrline i mane Bajesovog pristupa.

Nejt Silver (Nate Silver) je američki statističar koji je stekao slavu zahvaljujući svojim predikcionim modelima za američke izbore. On lično ne ispituje javno mnjenje, već koristi rezultate drugih agencija koja su javno dostupna. Na osnovu demografskih podataka o svakoj državi, izlaznosti i rezultatima na prošlim izborima, on je napravio kompleksan skup prethodnih verovatnoća koji zajedno sa rezultatima javnog mnjenja od raznih agencija daje predviđanje o verovatnoći pobede kandidata. Njegov model je dinamičan i kako se kampanja odvija i pojavljuju noviji podaci javnog mnjenja tako on izbacuje sveža predviđanja. Silver je 2008. godine sa Bajesovim pristupom tačno predvideo rezultate američkih predsedničkih izbora u 49 od 50 država, a 2012. u svih 50. Međutim 2016. godine je ozbiljno omanuo.

Silver je sjajan statističar i samo sa javno dostupnim podacima uspevaio je da bude bolji od drugih moćnijih igrača, ali nije imun na promene u očekivanjima, tj prethodnim verovatnoćama, koja su pre važila. I tu je suština kritike Bajesovog pristupa. Na osnovu čega možemo znati da su prethodne verovatnoće tačne? Čak i kad znamo da su pre važile, ne mora da znači da će i sada.

Standardna statistika je osmišljena da se primenjuje u izolovanom okruženju, dok sa Bajesovom vi možete da kombinujete zaključke drugih sa novim podacima i da dobijete nove zaključke koji su objektivni u onoj meri u kojoj su oni na koje se oslanjate. S obzirom na eksploziju podataka i sve moćnije računare, utisak mi je da je budućnost statistike u nauci Bajesova. Sa druge strane, kako u nauci slabo ko proverava objavljene rezultate, podozrevam da opasnost od zloupotrebe dolazi više sa Bajesove nego sa frekvencionističke strane.

Moj stav prema oba pravca u interpretaciji statistike je neutralan. Oba su pokazala praktičnu vrednost, ali ne bih umeo tačno reći kad koji da se primeni. U slučajevima kad imamo pouzdane prethodne verovatnoće onda Bajesova statistika, a u ostalim frekvencionistički pristup. Međutim, ono što nekome može izgledati pouzdano, meni može biti nesigurno i šta ćemo onda?

Reference koje podržavaju činjenice spomenute u članku

1. Bajesov pogled na verovatnoću i statistiku

https://en.wikipedia.org/wiki/Bayesian_probability

https://en.wikipedia.org/wiki/Bayesian_statistics

2. Tomas Bajes (Thomas Bayes)

https://en.wikipedia.org/wiki/Thomas_Bayes

3. Bajesova formula

https://en.wikipedia.org/wiki/Bayes%27_theorem

4. Različite interpretacije verovatnoća

https://en.wikipedia.org/wiki/Probability_interpretations

5. Harold Džefris (Harold Jeffreys)

https://en.wikipedia.org/wiki/Harold_Jeffreys

6. Knjiga o teoriji verovatnoće Harolda Džefrisa

<https://www.amazon.com/Theory-Probability-Classic-Physical-Sciences/dp/0198503687>

7. FDA uputstvo za upotrebu Bajesove statistike u kliničkim trajalima za medicinske instrumente

<https://www.fda.gov/MedicalDevices/ucm071072.htm>

8. Nejt Silver (Nate Silver)

https://en.wikipedia.org/wiki/Nate_Silver

9. Bradley Efron lepo opisuje legitimnost oba pristupa (Bajesovog i frekvencionističkog) u nauci

<https://www.tandfonline.com/doi/abs/10.1198/016214505000000033>