Sąlyginės tikimybės ir hipotezių tikrinimas

Ketvirtoji užduotis skirta sąlyginių tikimybių skaičiavimo ir hipotezių tikrinimo uždaviniams. Prieš pradedant spręsti uždavinius, reikėtų susipažinti su paskaitose nagrinėtais pavyzdžiais apie sąlygines tikimybes ir pilnosios tikimybės formulės taikymą. Papildomai aptarsime vieną pavyzdį.

Pilnosios tikimybės ir Bajeso formulės. Tarkime, kad slaptas pranešimas užšifruotas raidėmis a, b, c ir žinoma, kad paprastai pusę šifruoto teksto sudaro raidės a, o raidė b sutinkama dvigubai dažniau nei c. Be to, kol pasiekia adresatą, vidutiniškai 10% raidžių b bei 5% raidžių c iškraipomos ir virsta raidėmis a. Kokia tikimybė, kad tryliktas adresato gauto šifruoto teksto simbolis bus raidė a ?
Atsakymas būtų aiškus, jeigu žinotume koks buvo tryliktas šifro simbolis. Tačiau kaip išspręsti šį uždavinį to nežinant? Atsakymą padės rasti *pilnosios tikimybės formulė*:

$$P(A) = \sum_{i\in I} P(H_i)P(A|H_i)\,, \tag{1}$$

čia H_i, $(i \in I)$ - baigtinė arba skaiti poromis nesuderinamų įvykių šeima, tenkinanti sąlygą

$$P(\bigcup_{i\in I} H_i) = 1.$$

Pilnosios tikimybės formulė teigia, kad *apriorinę* įvykio A tikimybę galima rasti, žinant *aposteriorines* (sąlygines) A tikimybes, esant sąlygoms H_i, ir tų sąlygų susidarymo tikimybes. Esant toms pačioms prielaidoms kaip ir pilnosios tikimybės formulėje, galime rasti ir hipotezių aposteriorines tikimybes $P(H_j|A)$:

$$P(H_j|A) = \frac{P(H_j)P(A|H_j)}{\sum\limits_{i\in I} P(H_i)P(A|H_i)}\,. \tag{2}$$

(2) lygybė vadinama Bajeso hipotezių tikrinimo formule. Ja galime remtis tokioje sprendimų priėmimo situacijoje. Tarkime, žinome, jog įvyko vienas įvykis iš poromis nesuderinamų įvykių šeimos H_i $(i \in I)$ (teisinga viena iš kelių hipotezių) Kuris iš įvykių įvyko - nežinome, tačiau turime "netiesioginę" informaciją: įvyko įvykis A. Tarkime, reikia nuspręsti, kuria hipoteze H_i vadovautis, priimant sprendimą apie tolimesnius veiksmus. Mažiausia tikimybė suklysti bus tada, jei savo sprendimą grįsime ta hipoteze, kuriai $P(H_i|A)$ yra didžiausia.

Pavyzdys. Išspręsime suformuluotą uždavinį apie iškraipytą šifrą. Kadangi viskas priklauso nuo to koks buvo neiškraipyto šifro tryliktas simbolis, tai atitinkamai ir parinksime hipotezes H_1, H_2, H_3 :

H_1={tryliktas siunčiamo šifro simbolis buvo raidė a}, $\quad P(H_1) = \frac{1}{2}$;
H_2={tryliktas siunčiamo šifro simbolis buvo raidė b}, $\quad P(H_2) = \frac{1}{3}$;
H_3={tryliktas siunčiamo šifro simbolis buvo raidė c}, $\quad P(H_3) = \frac{1}{6}$;

Tegul A={tryliktas gauto šifro simbolis yra raidė a}. Tuomet, pagal uždavinio sąlygas,

$$P(A|H_1) = 1\,, \quad P(A|H_2) = 0,1\,, \quad P(A|H_3) = 0,05\,.$$

Pritaikę pilnosios tikimybės formulę, gauname

$$\begin{aligned} P(A) &= P(H_1)P(A|H_1) + P(H_2)P(A|H_2) + P(H_3)P(A|H_3) \\ &= \frac{1}{2}\cdot 1 + \frac{1}{3}\cdot\frac{1}{10} + \frac{1}{6}\cdot\frac{1}{20} = \frac{13}{24}\,. \end{aligned}$$

Galime formuluoti ir kitą, dažnai žymiai aktualesnį, klausimą. Tarkime, kad vienaip ar kitaip adresatas sugebėjo perskaityti tą nelemtą tryliktąjį gauto šifro simbolį - tai buvo raidė a. Kokia tikimybė, kad ji nėra iškraipyta? Kitaip sakant, mus dominanti tikimybė yra $P(H_1|A)$. Ją rasime, pasinaudoję Bajeso formule. Pastebėsime, kad trupmenos vardiklis (2) lygybėje, pagal pilnosios tikimybės formulę, yra lygus $P(A)$. Todėl

$$P(H_1|A) = \frac{P(H_1)P(A|H_1)}{P(A)} = \frac{\frac{1}{2}\cdot 1}{\frac{13}{24}} = \frac{12}{13}\,.$$