

Анализа података у биологији

Марија Цупарић

Трећи час



Нека уопштења:

Пребројив простор: Врше се независна понављања експеримента до прве појаве успеха, где је вероватноћа успеха p

- ★ Ако се новчић баца до прве појаве писма, онда је skup елементарних исхода

$$\Omega = \{n, gn, ggn, ggg, \dots\} \cup \{\omega_0\}, \text{ где је } \omega_0 = ggg\dots$$

- ★ Вероватноћа да писмо падне пре четвртог бацања је

$$P(A) = \frac{1}{2} + \frac{1}{2} \frac{1}{2} + \frac{1}{2} \frac{1}{2} \frac{1}{2} = \frac{7}{8}$$

Непребројив простор: Ω је неки ограничен подскуп у $\mathbb{R}^n$

- ★ $P(A) = \frac{m(A)}{m(\Omega)}$, где је $m(\cdot)$ мера скупа (дужина, површина, запремина,...)

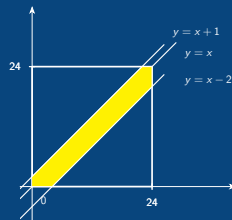
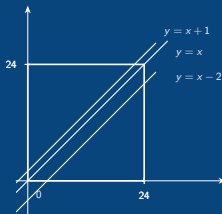
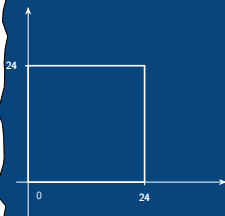
Пример

- (1) Нека су C и D тачке на дужи AB . Вероватноћа избора тачке са дужи CD ако се случајно бира тачка са дужи AB је $\frac{|CD|}{|AB|}$.
- (2) Нека је S круг полупречника 1, а S_1 круг полупречника $\frac{1}{2}$ са истим центром као S . Вероватноћа избора тачке из S_1 ако се случајно бира тачка из S је $\frac{\frac{1}{4}\pi}{\pi} = \frac{1}{4}$.



Пример. Два брода, независно један од другог и са истим вероватноћама стижу у луку у било ком тренутку у току 24 часа. Ако лука може да опслужи само један брод и ако се први брод задржава у луци 1 час, а други 2 часа, одредити вероватноћу да ће брод који стигне други, морати да чека.

x, y - времена долазака посматраних бродова



$$\Omega = \{(x, y) | x, y \in [0, 24]\}$$

$$A = \{(x, y) | x, y \in [0, 24], x < y < x + 1, y < x < y + 2\}$$

$$\Rightarrow P(A) = 1 - \frac{m(\bar{A})}{m(\Omega)} = 1 - \frac{\frac{22^2}{2} + \frac{23^2}{2}}{24^2} = 0.12$$

Уведимо сада формално (аксиоматски) појам вероватноће.

Дефиниција: Нека је Ω простор елементарних исхода. Функција P дефинисана на подскуповима од Ω је вероватноћа ако има следећа својства

- (A1) $P(\Omega) = 1$ (нормираност);
- (A2) $P(A) \geq 0$ за сваки догађај A над скупом Ω (ненегативност);
- (A3) Ако су $A_1, A_2, A_3, \dots$ догађаји, такви да за различите i и j важи $A_i \cap A_j = \emptyset$, онда важи једнакост

$$P\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} P(A_n) \quad (\sigma\text{-адитивност}).$$

Заправо, вероватноћа је било која функција која задовољава три наведене особине, а статистички, класични и бајесовски приступ нам дају смислен начин да до те функције и дођемо.



Пример. Брокер мисли да је вероватноћа да ће цена одређених акција расти током дана 0.3, а да је вероватноћа да ће падати 0.1. Одредити вероватноћу да ће доћи до промене цене, односно да ће цена или падати или расти.

A_1 — цена акција расте

A_2 — цена акција пада

$\Rightarrow$ дисјунктни догађаји

$$\Rightarrow P(A_1 \cup A_2) = P(A_1) + P(A_2) = 0.3 + 0.1 = 0.4.$$

На основу ова три својства можемо добити и нека друга својства вероватноће која су корисна.

а) Вероватноћа немогућег догађаја је 0, тј. $P(\emptyset) = 0$.

б) Ако су $A_1, A_2, \dots, A_n$, дисјунктни догађаји, онда важи једнакост:

$$P\left(\bigcup_{j=1}^n A_j\right) = \sum_{j=1}^n P(A_j) \text{ (коначна адитивност).}$$

в) Вероватноћа супротног догађаја је $P(\bar{A}) = 1 - P(A)$.

г) Ако је $A \subset B$, онда је $P(A) \leq P(B)$.

д) $P(A_1 \cup A_2 \cup A_3 \cup \dots) \leq P(A_1) + P(A_2) + P(A_3) + \dots$

ђ) $P(A \cup B) = P(A) + P(B) - P(AB)$.

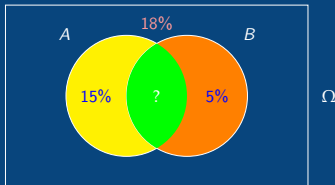


Пример. Еколози проучавају популацију риба у једном језеру како би проценили утицај загађења на здравље јединки. Посматрају се два догађаја: присуство повишене концентрације тешких метала у ткиву риба и појава физиолошких оштећења. Истраживање је показало да 15% риба има повишену концентрацију тешких метала у ткиву. Такође, 5% риба показује физиолошка оштећења. Даља анализа узорка показује да 18% риба има бар једну од ове две карактеристике — или повишену концентрацију тешких метала или физиолошка оштећења. Одредити вероватноћу да случајно изабрана риба из популације има и повишену концентрацију тешких метала и физиолошка оштећења.

A - догађај да риба има повишену концентрацију тешких метала, $P(A) = 0.15$

B - догађај да риба има физиолошка оштећења, $P(B) = 0.05$

$P(A \cup B) = 0.18$



$$\implies P(AB) = P(A) + P(B) - P(A \cup B) = 0.15 + 0.05 - 0.18 = 0.02.$$

Претпоставимо да имамо простор елементарних исхода Ω и неки одређени догађај A чија је вероватноћа $P(A)$.

Пример. Колика је вероватноћа да риба има физиолошко оштећење?

Затим добијамо додатну информацију да се догађај B реализовао. Желимо да утврдимо утицај, ако постоји, ове нове информације на $P(A)$.

Пример. Колика је вероватноћа да риба има физиолошко оштећење, ако се зна да има концентрацију тешких метала у ткиву?

Дефиниција: Нека су A и B догађаји дефинисани на истом Ω такви да је $P(B) \neq 0$. Условна вероватноћа од A при услову B је дата са

$$P(A|B) = \frac{P(AB)}{P(B)}.$$

Захтев да $P(B) \neq 0$ је неопходан, јер не смемо делити са нулом, али није рестриктиван јер нас у суштини не интересују догађаји који се реализују са вероватноћом 0.

Пример. $P(A|B) = \frac{P(AB)}{P(B)} = \frac{0.02}{0.05} = \frac{2}{5}.$

Из претходне дефиниције следи правило множења вероватноћа:

$$P(AB) = P(B)P(A|B).$$

Пример У Америци око 46% људи има крвну групу 0, а у регистрима је 4% оних који имају 0 грешком забележено као А. Колика је вероватноћа да случајно изабрани Американац стварно има 0, али су му забележили А?

О - има 0 крвну групу

А - забележено му је А

Дато нам је $P(O) = 0.46$ и $P(A|O) = 0.04$.

$$P(OA) = P(O)P(A|O) = 0.46 \cdot 0.04 = 0.018 \approx 2\%.$$



Размотримо догађаје:

- ★ A - у мензи се сервира пилетина за ручак, $P(A) = 0.7$
- ★ B - свемирска летелица безбедно слети

Поставља се питање да ако знамо да је свемирска летелица безбедно слетела, колика је сада вероватноћа да се послужи пилетина у мензи?

$\Rightarrow P(A|B) = 0.7$ (као и раније)

Дефиниција: Нека су A и B догађаји такви да $P(B) \neq 0$. За догађаје A и B такве да је $P(A|B) = P(A)$ се каже да су независни догађаји.

Два догађаја су независна ако реализација једног од њих нема утицаја на вероватноћу реализације другог догађаја.

Ако су A и B независни догађаји, тада $P(AB) = P(A)P(B)$.

Пример Бацају се плава и црвена коцкица. Дати су догађаји: A – добијени су исти бројеви, B – на црвеној је број 2 или 3.

$$P(A) = \frac{6}{36}, \quad P(B) = \frac{12}{36}, \quad P(AB) = \frac{2}{36}, \quad P(A | B) = \frac{\frac{2}{36}}{\frac{12}{36}} = \frac{2}{12}$$

$P(A | B) = P(A) \rightarrow A$ и B су независни.

Пример У Америци око 46% становника има 0 крвну групу, приближно 39% има негативан Rh фактор. Колика је вероватноћа да случајно изабрана особа има 0- крвну групу? Испитивања у генетици показују да крвна група појединца је независна од Rh фактора.

A – особа има крвну групу 0

B – особа има негативан Rh фактор

$$P(A) = 0.46, \quad P(B) = 0.39$$

$$P(AB) = P(A) P(B) = 0.46 \cdot 0.39 = 0.179$$

Дакле, око 17.9% популације има 0- крвну групу.

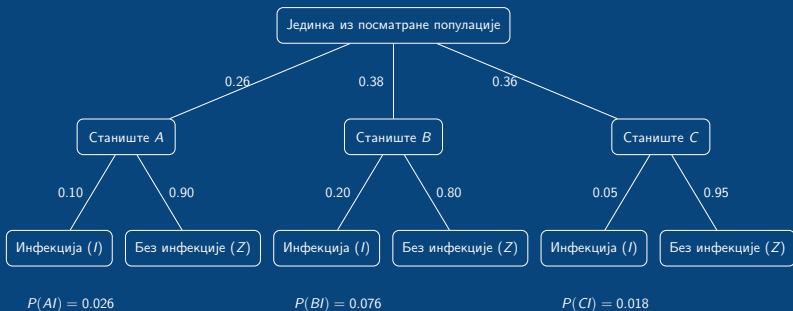
Претпоставимо да имамо n дисјунктних догађаја $A_1, \dots, A_n$ и $\bigcup_{i=1}^n A_i = \Omega$ (они чине такозвани потпун систем догађаја) и да желимо да одредимо вероватноћу неког догађаја B . Тада можемо користити формулу потпуне вероватноће

$$P(B) = P(BA_1 \cup BA_2 \cup \dots \cup BA_n) = P(A_1)P(B|A_1) + \dots + P(A_n)P(B|A_n).$$

Догађаји A_i се називају хипотезе.



Пример У истраживању једне врсте водоземаца, јединке се прате на три различита типа станишта: станиште A у 26% случајева, станиште B у 38% случајева и станиште C у преосталом делу случајева. Познато је да 10% јединки са станишта A , 20% јединки са станишта B и 5% јединки са станишта C показује знаке паразитске инфекције. Ако се једна јединка случајно бира из укупно посматране популације, одредити вероватноћу да та јединка има паразитску инфекцију.



$$P(I) = 0.026 + 0.076 + 0.018 = 0.12$$

Из претходне формуле и формуле условне вероватноће, директно следи Бајесова формула

$$P(A_i|B) = \frac{P(A_iB)}{P(B)} = \frac{P(A_iB)}{P(A_1)P(B|A_1) + \dots + P(A_n)P(B|A_n)}.$$

Дакле, ако знамо колика је вероватноћа догађаја B , можемо да одредимо и вероватноће појединачних хипотеза.



Пример У биолошким истраживањима узорци се обрађују у три различите лабораторијске јединице. Прва и најпоузданија јединица, A , даје неисправно обрађен узорак у само 1% случајева, док преостале две јединице, B и C , имају стопу грешке 5% и 7% редом. Јединица A обрађује узорке у 50% случајева, јединица B у 30% случајева, а јединица C у 20% случајева. Утврђено је да је један случајно изабрани узорак неисправно обрађен. Одредити вероватноћу да је тај узорак обрађен у јединици A .

A_1 - догађај да је узорак обрађен у јединици A , $P(A_1) = 0.5$

A_2 - догађај да је узорак обрађен у јединици B , $P(A_2) = 0.3$

A_3 - догађај да је узорак обрађен у јединици C , $P(A_3) = 0.2$

B - догађај да је узорак неисправно обрађен

$$P(B|A_1) = 0.01, \quad P(B|A_2) = 0.05, \quad P(B|A_3) = 0.07$$

Треба одредити $P(A_1|B)$. Тада

$$\begin{aligned} P(A_1|B) &= \frac{P(A_1)P(B|A_1)}{P(A_1)P(B|A_1) + P(A_2)P(B|A_2) + P(A_3)P(B|A_3)} \\ &= \frac{0.01 \cdot 0.5}{0.01 \cdot 0.5 + 0.05 \cdot 0.3 + 0.07 \cdot 0.2} = 0.1471. \end{aligned}$$

Дакле, ако знамо да је узорак неисправно обрађен, вероватноћа да потиче из јединице A износи око 15%.

За одређивање (класичне) вероватноће потребан је укупан број исхода и број начина реализације догађаја. У неким ситуацијама скуп свих исхода је велик да бисмо могли да испишемо све могућности, па тада прибегавамо методама пребројавања.

принцип множења: ако се експеримент може поделити у етапе, онда је број исхода једнак производу броја исхода у свакој етапи

Пример Студент треба да изабере три изборна предмета. Први бира од три попућене природне науке, други од четири друштвене науке, а трећи од пет спортова. На колико начина он то може да уради? $3 \cdot 4 \cdot 5 = 60$.

пермутације: низови објеката у одређеном редоследу; $n!$

Пример На колико начина се 8 спринтера може поставити на стартну линију?
 $8 \cdot 7 \cdot \dots \cdot 1 = 8! = 40320$

★ ако се објакти понављају: $\frac{n!}{n_1! \cdot n_2! \cdot \dots \cdot n_k!}$, где је $n_1 + n_2 + \dots + n_k = n$

комбинације: скупови објеката без одређеног редоследа; број комбинација r објеката изабраних од n различитих објеката је $\binom{n}{r} = \frac{n!}{r!(n-r)!}$

Пример На колико начина можемо изабрати 3 волонтера од 5 пријављених?
 $\binom{5}{3} = \frac{5!}{3! \cdot 2!} = \frac{5 \cdot 4}{2} = 10$

Пример У чекаоници ординације чекају 4 пацијента, од којих је један хитан случај, али тај податак није познат доктору.

- а) На колико начина пацијанти могу бити прегледани?
- б) У колико случајева ће хитан случај бити први прегледан?
- в) Ако се насумично бира пацијент, колика је вероватноћа да је први прегледан хитни случај?

а) $4 \cdot 3 \cdot 2 \cdot 1 = 4! = 24$

б) $1 \cdot 3 \cdot 2 \cdot 1 = 3! = 6$

в) $\frac{6}{24} = \frac{1}{4}$

