Vilniaus universitetas
Matematikos ir informatikos fakultetas

Informacijos teorija

(Paskaitų konspektas, I dalis)

A. Mačiulis

2013

Turinys

1 Pagrindinės tikimybių teorijos ir informacijos teorijos sąvokos **3**

1.1 Tikimybės. Atsitiktiniai dydžiai ir jų skirstiniai 3

1.2 Įvykių sistemos 13

1.3 Informacija ir entropija 16

1.4 Atsitiktinių dydžių entropijos 29

Literatūra **35**

1 Pagrindinės tikimybių teorijos ir informacijos teorijos sąvokos

1.1 Tikimybės. Atsitiktiniai dydžiai ir jų skirstiniai

1.1.1 apibrėžimas. *Tegu Ω yra baigtinė arba skaiti aibė, $\mathcal{P}(\Omega)$ visų jos poaibių sistema, $P(\omega)$, $\omega \in \Omega$, neneigiami skaičiai, tenkinantys sąlygą*

$$\sum_{\omega \in \Omega} P(\omega) = 1.$$

Tikimybe vadinsime funkciją $P : \mathcal{P}(\Omega) \to [0, 1]$, kiekvienam $A \subset \Omega$ apibrėžiamą lygybe

$$P(A) = \sum_{\omega \in A} P(\omega).$$

Porą (Ω, P) vadinsime diskrečiąja tikimybine erdve, o Ω - elementariųjų įvykių aibe.

Diskrečioji tikimybinė erdvė vadinama baigtine, kai jos elementariųjų įvykių aibės elementų skaičius $|\Omega|$ yra baigtinis.

Jei eksperimentas yra nusakomas tikimybine erdve (Ω, P) , tai aibės Ω elementai ω dar vadinami jo *elementariosiomis baigtimis*. Tada bet kuri to eksperimento baigtis A, sudaryta iš elementariųjų baigčių, vadinama *atsitiktiniu įvykiu* tikimybinėje erdvėje (Ω, P). Kitaip sakant, atsitiktiniu įvykiu laikysime bet kurį aibės Ω poaibį. Kaip įprasta, būtinąjį įvykį žymėsime Ω, negalimąjį, t.y. neturintį palankių elementariųjų baigčių, žymėsime tuščios aibės simboliu $\emptyset$, įvykiui A priešingą įvykį - $\overline{A}$.

1.1.1 pavyzdys. (*Klasikinis tikimybės apibrėžimas.*) Tai yra baigtinė tikimybinė erdvė, kurioje visi elementarieji įvykiai vienodai galimi. Taigi, jei

$$\Omega = \{\omega_1, \omega_2, \ldots, \omega_n\},$$

tai

$$P(\omega_1) = P(\omega_2) = \cdots = P(\omega_n) = \frac{1}{n}.$$

Todėl pagal apibrėžimą įvykio $A = \{\omega_{i_1}, \omega_{i_2}, \ldots, \omega_{i_k}\} \subset \Omega$ tikimybė bus

$$P(A) = \sum_{j=1}^{k} P(\omega_{i_j}) = \frac{k}{n}.$$

Kitaip sakant, įvykio tikimybė yra lygi jam palankių elementariųjų baigčių skaičiaus ir visų elementariųjų baigčių skaičiaus santykiui.

1.1.2 pavyzdys. Metamos trys simetriškos monetos. Raskime įvykio $A =$\{atsivertė bent vienas herbas\} tikimybę. Šiuo atveju galime sudaryti tokią elementariųjų įvykių aibę

$$\Omega = \{\omega_0, \omega_1, \omega_2, \omega_3\},$$

čia $\omega_i =$\{atsivertė i herbų\}. Tada $A = \{\omega_1, \omega_2, \omega_3\}$.
Atrodytų, kad $P(A) = \frac{3}{4}$. Tačiau patyrę monetų mėtytojai pastebės, kad taip nėra. Iš tikrųjų ne visi įvykiai ω_i yra vienodai galimi. Todėl klasikinis tikimybės apibrėžimas čia netinka, o eksperimento sąlygas atitinkančios elementariųjų įvykių tikimybės yra

$$P(\omega_0) = P(\omega_3) = \frac{1}{8}, \quad P(\omega_1) = P(\omega_2) = \frac{3}{8}.$$

Taigi

$$P(A) = P(\omega_1) + P(\omega_2) + P(\omega_3) = \frac{7}{8}.$$

1.1.2 apibrėžimas. *Įvykiai A ir B vadinami nesuderinamais, kai $P(A \cap B) = 0$. Jei $A \cap B = \emptyset$, tai A ir B vadinami nesutaikomais.*

Pastebėsime, kad bet kurie nesutaikomi įvykiai yra ir nesuderinami. Diskrečiojoje tikimybinėje erdvėje, neturinčioje nulinės tikimybės elementariųjų įvykių, šios dvi sąvokos ekvivalenčios.

1.1.3 pavyzdys. Dėžėje yra k baltų ir m juodų rutulių. Atsitiktinai be grąžinimo traukiame du rutulius. Nagrinėsime įvykius $A =$\{pirmasis rutulys baltas\} ir $B =$\{antrasis rutulys baltas\}. Tegu

$$\Omega = \{\omega_{bb}, \omega_{bj}, \omega_{jb}, \omega_{jj}\},$$

čia elementariųjų įvykių ω indeksai žymi ištraukto rutulio spalvą. Pavyzdžiui, $\omega_{bj} =$ \{pirmasis rutulys baltas, o antras - juodas\}. Tada

$$A = \{\omega_{bb}, \omega_{bj}\}, \quad B = \{\omega_{bb}, \omega_{jb}\}, \quad A \cap B = \{\omega_{bb}\} \neq \emptyset.$$

Matome, kad įvykiai A ir B yra sutaikomi. Rasime jų tikimybes. Kadangi

$$P(\omega_{bb}) = \frac{k(k-1)}{(k+m)(k+m-1)}, \quad P(\omega_{jj}) = \frac{m(m-1)}{(k+m)(k+m-1)},$$

$$P(\omega_{bj}) = P(\omega_{jb}) = \frac{k \cdot m}{(k+m)(k+m-1)}\,,$$

tai

$$P(A) = P(\omega_{bb}) + P(\omega_{bj}) = P(\omega_{bb}) + P(\omega_{jb}) = P(B) = \frac{k}{k+m}\,.$$

Tačiau įvykiai A ir B ne visada bus suderinami, nes

$$P(A \cap B) = P(\omega_{bb}) = \frac{k(k-1)}{(k+m)(k+m-1)} = 0\,,$$

kai $k \leq 1$.

Tegu $A,\ B,\ B_1,\ A_1, A_2, \ldots$ yra atsitiktiniai įvykiai diskrečiojoje tikimybinėje erdvėje (Ω, P). Priminsime pagrindines tikimybės savybes, išplaukiančias iš jos apibrėžimo.

1. $P(\emptyset) = 0, \quad P(\Omega) = 1.$

2. Jei $A \subset B$, tai $P(A) \leq P(B)$.

3. Jei A ir B nesuderinami ir $A_1 \subset A,\ B_1 \subset B$, tai įvykiai A_1 ir B_1 taip pat bus nesuderinami.

4. Jeigu įvykiai $A_1, A_2, \ldots$ poromis nesuderinami, t.y. $P(A_i \cap A_j) = 0$ visiems natūraliesiems $i \neq j$, tai
$$P(\bigcup_i A_i) = \sum_i P(A_i)\,.$$

5. Jei $A \subset B$, tai $P(B \setminus A) = P(B) - P(A)\,.$

6. $P(\overline{A}) = 1 - P(A)\,.$

7. $P(A \cup B) = P(A) + P(B) - P(A \cap B)\,.$

Tikimybių teorijoje vartojama ir abstrakti tikimybinės erdvės sąvoka. Bendruoju atveju Ω gali būti bet kuri netuščia aibė. Ką tuomet laikyti atsitiktiniais įvykiais? Kai Ω turi be galo daug skirtingų elementų, tai begalinės jų sąjungos bei sankirtos gali būti tokie Ω poaibiai, kad, įvedant tikimybės sąvoką, kils dideli matematiniai sunkumai. Todėl atsitiktinių įvykių aibe laikoma tik tam tikra, pakankamai "turtinga" Ω paibių sistema $\mathcal{F}$, turinti tokias savybes:

i) $\Omega \in \mathcal{F}$,

ii) $A \in \mathcal{F} \Rightarrow \overline{A} \in \mathcal{F}$,

iii) $A_1, A_2, \ldots \in \mathcal{F} \Rightarrow A_1 \cup A_2 \cup \ldots \in \mathcal{F}$.

Paibių sistema $\mathcal{F}$ vadinama atsitiktinių įvykių σ (sigma) algebra. Tada tikimybe vadinama funkcija $P : \mathcal{F} \to [0, 1]$, jei

i) $P(\Omega) = 1$;

ii) jei $A_i \in \mathcal{F}$ ir $A_i \cap A_j = \emptyset$ visiems natūraliesiems $i \neq j$, tai $P(A_1 \cup A_2 \cup \ldots) = P(A_1) + P(A_2) + \ldots$.

Taip apibrėžta tikimybė tenkina visas anksčiau suformuluotas savybes ir neprieštarauja 1.1.1 apibrėžimui. Mes neakcentuosime σ algebros vaidmens. Kalbėdami apie atsitiktinius įvykius, turėsime omenyje, kad jie priklauso tam tikrai σ algebrai. Pastebėsime, kad $\mathcal{P}(\Omega)$ yra σ algebra.

Sąlyginė tikimybė. Dažnai galimybė įvykti vienam įvykiui priklauso nuo to, ar įvyksta kitas įvykis. Tarkime norime rasti įvykio A tikimybę, žinodami, kad įvyko įvykis B. Tokia tikimybė vadinama įvykio A *sąlygine* tikimybe ir žymima $P(A|B)$ (skaitoma "tikimybė, kad įvyks A su sąlyga, kad įvyko B" arba "įvyks A , jeigu įvyko B"). Jei $P(B) > 0$, tai

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

Prisiminę 1.1.3 pavyzdžio eksperimentą, kai $k > 0$, nesunkiai rasime, kad ištraukus baltą rutulį, tikimybė vėl ištraukti baltą bus

$$P(B|A) = \frac{k-1}{k+m-1}.$$

Kaip matome, šiuo atveju $P(B|A) \neq P(B)$. Tai rodo, kad įvykio B tikimybė priklauso nuo to, ar įvyko įvykis A. Tokie įvykiai vadinami *priklausomais.*

Įvykių nepriklausomumas - tai viena iš svarbesniųjų tikimybių teorijos sąvokų. Pateiksime griežtesnį jos apibrėžimą. Iš sąlyginės tikimybės apibrėžimo išplaukia vadinamoji *tikimybių daugybos teorema* :

$$P(A \cap B) = P(B)P(A|B) = P(A)P(B|A). \tag{1.1}$$

Kai A ir B yra nepriklausomi įvykiai, turėsime $P(A|B) = P(A)$ ir $P(B|A) = P(B)$. Todėl, atsižvelgę į (1.1), gauname tokį įvykių nepriklausomumo apibrėžimą.

1.1.3 apibrėžimas. *Įvykius A ir B vadinsime nepriklausomais, jeigu*

$$P(A \cap B) = P(A)P(B).$$

Didesnio įvykių skaičiaus nepriklausomumas apibrėžiamas sudėtingiau. Pavyzdžiui, įvykiai A, B ir C vadinami nepriklausomais, jei teisingos visos keturios lygybės

$$\begin{aligned} P(A \cap B) &= P(A)P(B), \quad P(A \cap C) = P(A)P(C), \\ P(B \cap C) &= P(B)P(C), \quad P(A \cap B \cap C) = P(A)P(B)P(C). \end{aligned}$$

Nereikia šios sąvokos painioti su nesutaikomumu. Nepriklausomi įvykiai nebūtinai nesutaikomi. Dažnai įvykių nepriklausomumas pastebimas intuityviai. Tačiau intuicija gali ir suklaidinti.

1.1.4 pavyzdys. Metame lošimo kauliuką. Nagrinėsime įvykius

A={atsivertė lyginis skaičius akučių} ,

B={atsivertė ne mažiau kaip 4 akutės} ,

C={atsivertė daugiau kaip 4 akutės} .

Ar šie įvykiai priklausomi?

Apskaičiuosime įvykių tikimybes. Nesunku suprasti, kad

$$A = \{2, 4, 6\}, \quad B = \{4, 5, 6\}, \quad C = \{5, 6\}$$

$$A \cap B = \{4, 6\}, \quad A \cap C = \{6\}.$$

Todėl

$$P(A)P(C) = \frac{3}{6} \cdot \frac{2}{6} = \frac{1}{6} = P(A \cap C).$$

Taigi įvykiai A ir C nepriklausomi. Tačiau

$$P(A)P(B) = \frac{3}{6} \cdot \frac{3}{6} \neq \frac{2}{6} = P(A \cap B).$$

Todėl gauname, kad įvykiai A ir B , o tuo pačiu ir visi trys įvykiai A, B ir C , yra priklausomi.

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*:

$$P(A) = \sum_{i \in I} P(H_i) P(A|H_i)\,, \tag{1.2}$$

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 $P(H_j|A)$:

$$P(H_j|A) = \frac{P(H_j)P(A|H_j)}{\sum\limits_{i \in I} P(H_i)P(A|H_i)}\,. \tag{1.3}$$

(1.3) 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.

1.1.5 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 (1.3) lygybėje, pagal pilnosios tikimybės formulę, yra lygus $P(A)$. Todėl

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

Bernulio eksperimentai. Bernulio eksperimentų schema nusakoma taip: eksperimentą atlikus vieną kartą, jo sėkmės tikimybė lygi p. Atliekame n nepriklausomų eksperimentų. Sėkmių skaičių pažymėkime S_n. Kokia tikimybė, kad eksperimentas pavyks k kartų, t.y. $S_n = k$? Atsakymas į šį klausimą toks:

$$P(S_n = k) = \binom{n}{k} p^k (1-p)^{n-k}\,, \quad k = 0, 1, \ldots, n. \tag{1.4}$$

Bernulio schema yra vienodų ir nepriklausomų statistinių eksperimentų matematinis modelis. Ją naudojant skaičiuojamos tikimybės, susijusios su nepriklausomų vienodų bandymų seka, kai kiekviename bandyme galimos tik dvi baigtys.

1.1.6 pavyzdys. Informacija perduodama triukšmingu kanalu, kuris vidutiniškai iškraipo 1% visų siunčiamų bitų. Kokia tikimybė, kad baite bus ne daugiau dviejų iškraipytų bitų? Šiuo atveju sėkmė - gauti iškraipytą bitą. Pagal sąlygą tokios "sėkmės" tikimybė kiekvienu atveju yra $p = 0,01$. Mums reikalinga tikimybė, kad "sėkmių" skaičius po 8 bandymų būtų ne didesnis už 2. Pasinaudoję (1.4) lygybe, gausime

$$\begin{aligned} P(S_8 \le 2) &= P(S_8 = 0) + P(S_8 = 1) + P(S_8 = 2) \\ &= \binom{8}{0} 0,01^0 0,99^8 + \binom{8}{1} 0,01^1 0,99^7 + \binom{8}{2} 0,01^2 0,99^6 \approx 0,999946 \end{aligned}$$

Atsitiktiniai dydžiai. Apibrėždami atsitiktinius įvykius, kalbėjome apie eksperimentus ir jų elementariąsias baigtis. Praktiškai beveik visada susiduriame su skaitiniais stebimojo dydžio matavimais, t.y. su kokio nors atsitiktinio dydžio reikšmėmis.

1.1.4 apibrėžimas. *Tarkime, kad* (Ω, P) *yra diskrečioji tikimybinė erdvė. Atsitiktiniu dydžiu šioje erdvėje vadinama realioji funkcija* $X : \Omega \to \mathbb{R}$.

Taigi atsitiktinis dydis nusako taisyklę, pagal kurią kiekvienam elementariajam įvykiui priskiriama skaitinė reikšmė. Diskrečiosios tikimybinės erdvės atsitiktinių dydžių reikšmių aibė yra baigtinė arba skaiti. Tokie atsitiktiniai dydžiai X vadinami *diskrečiaisiais* ir dažniausiai nusakomi *reikšmių skirstiniu*, nurodant galimas reikšmes x_i ir jų tikimybes

$$p_i = P(X = x_i) = \sum_{\omega \in \Omega \,:\, X(\omega) = x_i} P(\omega)\,, \quad i = 1, 2, \ldots\,.$$

Pavyzdžiui sėkmių skaičius S_n, atlikus n Bernulio eksperimentų, yra diskretus atsitiktinis dydis, kurio reikšmių aibė $\{0, 1, 2, \ldots, n\}$, o tikimybės nusakomos (1.4) formule. Tokį skirstinį turintis atsitiktinis dydis vadinamas *binominiu* ir žymimas $S_n \sim \mathcal{B}(n, p)$.
Kai elementariųjų įvykių aibė Ω nėra skaiti, atsitiktinio dydžio reikšmių aibė gali būti labai "gausi" ir net nesunumeruojama. Pavyzdžiui, matuojant kliento aptarnavimo laiką, priklausomai nuo matavimo vienetų, rezultatas gali būti bet kuris tam tikro intervalo taškas. Šiuo atveju prasminga kalbėti ne apie pavienių reikšmių tikimybes, bet apie reikšmių priklausymo nurodytam intervalui tikimybę.

1.1.5 apibrėžimas. *Atsitiktinis dydis* X*, kurio patekimo į intervalą* $[a, b]$ *tikimybė skaičiuojama pagal formulę*

$$P(a \le X \le b) = \int_a^b p(x)\,dx\,, \quad p(x) \ge 0\,,$$

vadinamas absoliučiai tolydžiuoju dydžiu, o funkcija $p(x)$ *vadinama jo tankiu.*

Pastebėsime, kad bet kokiam absoliučiai tolydžiam atsitiktiniam dydžiui X ir realiajam skaičiui a "taškinė" tikimybė $P(X = a)$ lygi 0, o tankio funkcija $p(x)$ tenkina sąlygą

$$\int_{-\infty}^{\infty} p(x)\,dx = 1\,.$$

Kita vertus, pasirodo, kad bet kuri neneigiama funkcija funkcija $p(x)$, tenkinanti pastarąją lygybę, gali būti laikoma kažkokio atsitiktinio dydžio tankiu.

Dažnai atsitiktiniai dydžiai (tiek diskretieji, tiek tolydieji) nusakomi specialia - *pasiskirstymo funkcija.* Atsitiktinio dydžio X pasiskirstymo funkcija $F(x)$ yra

$$F(x) = P(X < x)\,, \quad x \in \mathbb{R}\,.$$

Absoliučiai tolydžiojo atsitiktinio dydžio pasiskirstymo funkciją ir tankį sieja lygybės:

$$\begin{aligned} F'(x) &= p(x)\,, \\ F(x) &= \int\limits_{-\infty}^{x} p(u)\,du\,. \end{aligned}$$

Kai nagrinėjami keli atsitiktiniai dydžiai, pasidaro svarbi jų tarpusavio priklausomybė. Natūralu pavadinti atsitiktinius dydžius X_1 ir X_2 nepriklausomais, kai su bet kuriais realiųjų skaičių poaibiais B_1, B_2 įvykiai $\{\omega : X_1(\omega) \in B_1\}$ ir $\{\omega : X_2(\omega) \in B_2\}$ yra nepriklausomi, kitaip sakant, jei

$$P(X_1 \in B_1,\, X_2 \in B_2) = P(X_1 \in B_1)P(X_2 \in B_2)\,.$$

Diskrečiųjų atsitiktinių dydžių atveju pakanka pareikalauti, kad visiems realiesiems x, y būtų tenkinama lygybė

$$P(X_1 = x,\, X_2 = y) = P(X_1 = x)P(X_2 = y)\,.$$

Priminsime kai kurias atsitiktinių dydžių skaitines charakteristikas. Pradėsime nuo vidurkio, nusakančio vidutinę atsitiktinio dydžio reikšmę. Diskrečiojo atsitiktinio dydžio skirstinį patogiausia užrašyti lentele

X	x_1	x_2	x_3	$\ldots$
P	p_1	p_2	p_3	$\ldots$

Žinoma, $p_1 + p_2 + \cdots = 1\,, \quad p_i \geq 0$. Taip nusakyto atsitiktinio dydžio *vidurkiu* vadinama suma

$$\mathbf{E}X = x_1p_1 + x_2p_2 + x_3p_3 + \ldots\,.$$

Jeigu X turi tankį $p(x)$, tai vidurkis apibrėžiamas kaip integralas

$$\mathbf{E}X = \int_{-\infty}^{\infty} xp(x)\,dx\,.$$

Galima apibrėžti ir atsitiktinio dydžio funkcijos vidurkį. Dikrečiojo ir tolydžiojo dydžių atvejais funkcijos vidurkis atitinkamai yra

$$\mathbf{E}f(X) = f(x_1)p_1 + f(x_2)p_2 + f(x_3)p_3 + \dots\,, \quad \mathbf{E}f(X) = \int_{-\infty}^{\infty} f(x)p(x)\,dx\,.$$

Suformuluosime pagrindines vidurkių savybes. Tegu $X, X_1, X_2, \dots, X_n$ yra atsitiktiniai dydžiai, turintys baigtinius vidurkius.

1. Su bet kokiomis konstantomis $c_1, c_2, \dots, c_n$ teisinga lygybė

$$\mathbf{E}\big(\sum_{i=1}^{n} c_i X_i\big) = \sum_{i=1}^{n} c_i \mathbf{E}X_i\,.$$

2. Jei $X_1 \le X_2$, tai $\mathbf{E}X_1 \le \mathbf{E}X_2$.
3. $|\mathbf{E}X| \le \mathbf{E}|X|$.
4. Jei X_1, X_2 yra nepriklausomi, tai

$$\mathbf{E}(X_1 \cdot X_2) = \mathbf{E}(X_1) \cdot \mathbf{E}(X_2)\,.$$

Kaip jau buvo minėta, vidurkis parodo vidutinę atsitiktinio dydžio X reikšmę. Jo sklaidą apie vidurkį aprašo *dispersija*

$$\mathbf{D}X = \mathbf{E}(X - \mathbf{E}X)^2\,.$$

Kvadratinė šaknis iš dispersijos vadinama *standartiniu nuokrypiu* $\sigma(X) = \sqrt{\mathbf{D}X}$.
Paminėsime keletą dispersijos savybių, laikydami, kad atsitiktiniai dydžiai
$X, X_1, X_2, \dots, X_n$ turi baigtines dispersijas.

1. $\mathbf{D}X \ge 0\,.$
2. $\mathbf{D}X = \mathbf{E}X^2 - (\mathbf{E}X)^2\,.$
3. $\mathbf{D}(X_1 + X_2) = \mathbf{D}X_1 + \mathbf{D}X_2 + 2cov(X_1, X_2)\,,$ čia

$$cov(X_1, X_2) = \mathbf{E}(X_1 - \mathbf{E}X_1)(X_2 - \mathbf{E}X_2) = \mathbf{E}X_1X_2 - \mathbf{E}X_1\mathbf{E}X_2$$

yra atsitiktinių dydžių X_1 ir X_2 *kovariacija*.

4. Jei $X_1, X_2, \ldots, X_n$ yra nepriklausomi atsitiktiniai dydžiai, tai su bet kokiomis konstantomis $c_1, c_2, \ldots, c_n$ teisinga lygybė

$$\mathbf{D}\Big(\sum_{i=1}^{n} c_i X_i\Big) = \sum_{i=1}^{n} c_i^2 \mathbf{D} X_i \,.$$

1.1.7 pavyzdys. Rasime binominio skirstinio, nusakyto (1.4) tikimybėmis, vidurkį ir dispersiją. Tegul $X_i = 1$, jei i-tasis Bernulio eksperimentas buvo sėkmingas, ir $X_i = 0$ - priešingu atveju. Tada sėkmių skaičius po n eksperimentų bus n nepriklausomų, vienodai pasiskirsčiusių atsitiktinių dydžių suma

$$S_n = X_1 + X_2 + \cdots + X_n \,.$$

Be to, visiems $i = 1, 2, \ldots, n$ atsitiktinio dydžio X_i skirstinys yra

X_i	0	1
P	$1-p$	p

Todėl

$$\begin{aligned} \mathbf{E}X_i &= \mathbf{E}X_i^2 = p \,, \\ \mathbf{D}X_i &= \mathbf{E}X_i^2 - (\mathbf{E}X_i)^2 = p(1-p) \,. \end{aligned}$$

Dabar jau nesunkiai randame nepriklausomų atsitiktinių dydžių X_i sumos vidurkį ir dispersiją

$$\begin{aligned} \mathbf{E}S_n &= \sum_{i=1}^{n} \mathbf{E}X_i = np \,, \\ \mathbf{D}S_n &= \sum_{i=1}^{n} \mathbf{D}X_i = np(1-p) \,. \end{aligned}$$

1.2 Įvykių sistemos

Tarkime $\mathcal{A} = \{A_i : i \in I\}$ yra diskrečiosios tikimybinės erdvės (Ω, P) įvykių šeima, I - baigtinė arba skaiti indeksų aibė.

1.2.1 apibrėžimas. *Jei* $P(A_i \cap A_j) = 0$ *visiems* $i, j \in I$, $i \neq j$ *ir*

$$P(\bigcup_{i \in I} A_i) = 1\,,$$

tai $\mathcal{A}$ *vadinsime tikimybinės erdvės* (Ω, P) *įvykių sistema.*

Pastebėsime, kad poromis nesuderinamiems įvykiams A_i

$$P(\bigcup_{i \in I} A_i) = \sum_{i \in I} P(A_i)\,.$$

Todėl antrojoje apibrėžimo sąlygoje sąjungos tikimybę pakeitę atitinkamų tikimybių suma, gautume ekvivalentišką įvykių sistemos apibrėžimą. Taip pat aišku, kad jei A_i poromis nesutaikomi ir

$$\bigcup_{i \in I} A_i = \Omega\,,$$

tai $\mathcal{A}$ - įvykių sistema. Atvirkščias teiginys teisingas tik, kai tikimybinėje erdvėje nėra nulinės tikimybės elementariųjų įvykių, t.y. $P(\omega) > 0$ visiems $\omega \in \Omega$.

1.2.2 apibrėžimas. *Tarkime* $\mathcal{A} = \{A_i : i \in I\}$ *ir* $\mathcal{B} = \{B_j : j \in J\}$ *yra tikimybinės erdvės* (Ω, P) *įvykių sistemos.* $\mathcal{A}$ *yra sistemos* $\mathcal{B}$ *apvalkalas, jei kiekvienam* $j \in J$ *galima rasti tokį* $i \in I$ *, kad* $P(A_i \cap B_j) = P(B_j)$*. Tokiu atveju sakysime, kad sistema* $\mathcal{B}$ *yra tikslesnė už sistemą* $\mathcal{A}$*.*

Kitaip sakant, kiekvienai tikslesnės sistemos aibei visada atsiras ją dengianti "grubesnės" sistemos aibė.

1.2.1 teorema. *Jei* $\mathcal{B}$ *yra tikslesnė už sistemą* $\mathcal{A}$*, tai visiems* $i \in I,\ j \in J$

$$P(A_i \cap B_j) = P(B_j) \quad arba \quad P(A_i \cap B_j) = 0.$$

Įrodymas. Kai $P(B_j) = 0$, šis teiginys akivaizdus. Tarkime, kad $P(B_j) \neq 0$ ir $P(A_i \cap B_j) \neq P(B_j)$. Lieka įsitikinti, kad tokiu atveju būtinai $P(A_i \cap B_j) = 0$. Pagal apvalkalo apibrėžimą, aibėje I galima rasti tokį $i_0 \neq i$, kad

$$P(A_{i_0} \cap B_j) = P(B_j).$$

Vadinasi

$$B_j = A_{i_0} \cap B_j \cup N\,, \quad P(N) = 0.$$

Taigi

$$P(A_i \cap B_j) = P(A_i \cap (A_{i_0} \cap B_j) \cup (A_i \cap N)) \leq P(A_i \cap A_{i_0}) + P(N) = 0.$$

□

1.2.3 apibrėžimas. *Tarkime $\mathcal{A} = \{A_i : i \in I\}$ ir $\mathcal{B} = \{B_j : j \in J\}$ yra tikimybinės erdvės (Ω, P) įvykių sistemos. Jų jungtinė sistema $\mathcal{A} \wedge \mathcal{B}$ nusakoma lygybe*

$$\mathcal{A} \wedge \mathcal{B} = \{A_i \cap B_j : (i,j) \in I \times J\}.$$

Akivaizdu, kad $\mathcal{A} \wedge \mathcal{B}$ tikslesnė ir už $\mathcal{A}$ ir už $\mathcal{B}$. Bet pasirodo, kad ji yra pati "grubiausia" iš visų tikslesnių už $\mathcal{A}$ ir $\mathcal{B}$. Teisingas toks teiginys.

1.2.2 teorema. *Jei sistema $\mathcal{C}$ yra tikslesnė už $\mathcal{A}$ ir $\mathcal{B}$, tai ji tikslesnė ir už $\mathcal{A} \wedge \mathcal{B}$.*

Įrodymas. Iš tikrųjų, iš teiginio prielaidos išplaukia, kad kiekvienam $C \in \mathcal{C}$ galima rasti tokius A_i ir B_j kad

$$P(A_i \cap C) = P(B_j \cap C) = P(C).$$

Todėl

$$B_j \cap C = C \setminus N\,, \quad N \subset C\,, \quad P(N) = 0.$$

Dabar gauname

$$\begin{aligned} P(A_i \cap B_j \cap C) &= P(A_i \cap (C \setminus N)) = P((A_i \cap C) \setminus (A_i \cap N)) \\ &= P(A_i \cap C) - P(A_i \cap N) = P(C). \end{aligned}$$

Pastaroji lygybė ir įrodo, kad $\mathcal{C}$ yra tikslesnė už $\mathcal{A} \wedge \mathcal{B}$. □

1.2.4 apibrėžimas. *Tikimybinės erdvės (Ω, P) įvykių sistemos $\mathcal{A} = \{A_i : i \in I\}$ ir $\mathcal{B} = \{B_j : j \in J\}$ vadinamos nepriklausomomis, jei*

$$P(A_i \cap B_j) = P(A_i)P(B_j)$$

visiems $(i,j) \in I \times J$.

1.2.1 pavyzdys. Tarkime X ir Y diskretieji atsitiktiniai dydžiai erdvėje (Ω, P), o $\{x_i \in \mathbb{R} : i \in I\}$ ir $\{y_j \in \mathbb{R} : j \in J\}$ - jų galimų reikšmių aibės. Apibrėžkime įvykius $A_i = \{X = x_i\}$ ir $B_j = \{Y = y_j\}$. Nesunku pastebėti, kad $\mathcal{A} = \{A_i : i \in I\}$ ir $\mathcal{B} = \{B_j : j \in J\}$ yra tikimybinės erdvės (Ω, P) įvykių sistemos. Jos bus nepriklausomos tada ir tik tada, kai nepriklausomi yra jas generuojantys atsitiktiniai dydžiai X ir Y.

1.3 Informacija ir entropija

Tarkime A yra tikimybinės erdvės (Ω, P) atsitiktinis įvykis, kurio tikimybė $P(A) = p$. Kiek informacijos gauname, įvykus šiam įvykiui ? Nekalbėsime apie gautos informacijos prasmę ar naudą. Mūsų tikslas - apibrėžti kiekybinį informacijos matą, priklausantį nuo įvykio tikimybės p. Įykio A prigimtis, skaičiuojant *informacijos kiekį* $I(A)$, nėra svarbi. Todėl informacijos kiekį apibrėšime kaip kintamojo p funkciją ir dažnai rašysime $I(A) = I(p)$. Jai kelsime tokius reikalavimus :

1. Informacija turi būti apibrėžta ir neneigiama, t.y. $I(p) \geq 0$, visiems $p \in (0,\ 1]$.

2. Nežymiai pakitus įvykio tikimybei, informacijos kiekis taip pat turėtų pasikeisti nedaug. Kitaip sakant funkcija $I(p)$ turi būti tolydi.

3. Funkcija $I(p)$ turi griežtai monotoniškai mažėjanti, t.y. kuo įvykio tikimybė mažesnė, tuo didesnį informacijos kiekį jam įvykus gauname. Jei pastarasis reikalavimas pasirodė keistas, panagrinėkite du atsitiktinius įvykius: A_1 ={ateinančių metų liepos septintą dieną Vilniuje snigs} ir A_2 ={ateinančių metų liepos septintoji Vilniuje bus saulėta}. Kurio įvykio tikimybė didesnė ir, kuriam įvykus, daugiau sužinotumėte apie Lietuvos klimato pokyčius?!

4. Įvykus dviem nepriklausomiems įvykiams, gautos informacijos kiekis turėtų būti lygus jų informacijų sumai. Prisiminę, kad dviems nepriklausomiems įvykiams A ir B tikimybė įvykti kartu yra $P(A \cap B) = P(A)P(B)$, turėsime tokį reikalavimą informacijos kiekio funkcijai: $I(p \cdot q) = I(p) + I(q)$ visiems $p, q \in (0,\ 1]$.

Pasirodo šie, iš pirmo žvilgsnio, paprasti reikalavimai vienareikšmiškai nusako juos tenkinančią funkciją.

1.3.1 teorema. *Funkcija $I(p)$ tenkina 1-4 sąlygas tada ir tik tada, kai egzistuoja $b > 1$, jog*

$$I(p) = \log_b \frac{1}{p}\,.$$

Įrodymas. Logaritminė funkcija aišku tenkina minėtus reikalavimus. Belieka įsitikinti, kad 1-4 sąlygas tenkinanti funkcija $I(p)$ yra būtinai logaritminė.

Tegul m ir n bet kokie natūralieji skaičiai. Iš 4 sąlygos išplaukia, kad

$$I(p^n) = I(p \cdot p^{n-1}) = I(p) + I(p^{n-1}) = I(p) + I(p) + I(p^{n-2}) = \cdots = nI(p)\,.$$

Todėl

$$I(p) = I((p^{1/m})^m) = mI(p^{1/m})$$

ir

$$I(p^{1/m}) = \frac{1}{m} I(p)\,.$$

Taigi, visiems teigiamiems racionaliesiems skaičiams $\frac{n}{m}$

$$I(p^{n/m}) = I((p^{1/m})^n) = \frac{n}{m} I(p)\,.$$

Dėl funkcijos $I(p)$ tolydumo iš čia išplaukia, kad

$$I(p^a) = aI(p)$$

visiems realiesiems $a \geq 0$. Todėl visiems $p \in (0,\ 1]$

$$I(p) = I\left(\left(\frac{1}{\mathrm{e}}\right)^{-\ln p}\right) = -I\left(\frac{1}{\mathrm{e}}\right) \ln p\,. \tag{1.5}$$

Kadangi funkcija $I(p)$ yra griežtai mažėjanti ir neneigiama, tai $I\left(\frac{1}{\mathrm{e}}\right) > 0$ ir galima rasti $b > 1$, kad

$$I\left(\frac{1}{\mathrm{e}}\right) = \frac{1}{\ln b}\,.$$

Iš čia ir (1.5) galutinai gauname

$$I(p) = -\frac{\ln p}{\ln b} = \log_b \frac{1}{p}\,.$$

□

Dabar jau galime apibrėžti įvykio informaciją. Logaritmo pagrindo pasirinkimas apsprendžia tik jos matavimo vienetus. Laikysime, kad informacijos vienetą gauname, įvykus įvykiui, kurio tikimybė $\frac{1}{2}$. Tada $b = 2$.

1.3.1 apibrėžimas. *Informacijos kiekiu, gaunamu įvykus įvykiui A, kurio tikimybė $p > 0$, vadinsime dydį*

$$I(A) = I(p) = \log_2 \frac{1}{p},$$

o jo matavimo vienetus - bitais.

Kartais naudojami ir kiti informacijos kiekio vienetai.

Logaritmo pagrindas (b)	*Informacijos kiekio vienetas*
2	bitas
3	tritas
e	natas
10	hartlis

Sąryšiai tarp šių matavimo vienetų nusakomi lygybėmis

$$1\ bitas\ = \log_3 2\ trito\ = \ln 2\ nato\ = \lg 2\ hartlio\,.$$

Beje, čia bitas ne atsitiktinai sutampa su dvejetainio skaičiaus skaitmens pavadinimu.

1.3.1 pavyzdys. Metame simetrišką monetą. Eksperimento baigčių S={atsivertė skaičius} ir H={atsivertė skaičius} tikimybės yra $P(S) = P(H) = 0,5$. Todėl bet kurios baigties atveju gaunamas $I(S) = I(H) = \log_2 2 = 1$ bitas informacijos. Jei moneta metama n kartų, tai bet kurią eksperimento baigtį galime nusakyti n dvejetainių skaitmenų, pavyzdžiui

$$\underbrace{011101110\ldots01101}_{n}$$

Čia 0 ir 1 žymi įvykius S ir H. Tokios baigties tikimybė yra 2^{-n}, o gaunamas informacijos kiekis

$$I\left(\frac{1}{2^n}\right) = \log_2 2^n = n\,,$$

t.y. lygiai tiek, kiek bitų užima informacija apie eksperimento rezultatą.

Pastebėsime, kad būtino įvykio informacija $I(\Omega) = 0$. Kitaip sakant, kai įvyksta tai "kas ir turėjo įvykti", mes nieko naujo nesužinome. Tačiau, jei įvyktų tai "kas įvykti negali", turėtume "labai daug" naujos informacijos. Toks pastebėjimas pateisina informacijos kiekio apibrėžimo papildymą nulinės tikimybės įvykiams. Taigi, jei $P(A) = 0$, tai

$$I(A) = \lim_{p \to 0} I(p) = \infty \, .$$

Galima apibrėžti ir dviejų įvykių sąlyginę informaciją.

1.3.2 apibrėžimas. *Tegul A ir B yra tikimybinės erdvės (Ω, P) atsitiktiniai įvykiai ir $P(B) > 0$. Įvykio A su sąlyga B informacija vadinsime dydį*

$$I(A|B) = \log_2 \frac{1}{P(A|B)} = -\log_2 \frac{P(A \cap B)}{P(B)} \, .$$

Pastebėsime, kad $I(A|B) = I(A)$ tada ir tik tada, kai įvykiai A ir B yra nepriklausomi. Aptarėme pavienio atsitiktinio įvykio informacijos kiekio sąvoką. Dabar pabandysime nusakyti įvykių sistemos informaciją. Pradėsime nuo pavyzdžio.

1.3.2 pavyzdys. Tegul galimos bandymo baigtys yra $A_1, A_2, \ldots A_n$, o jų tikimybės atitinkamai $p_1, p_2, \ldots p_n$. Kiek informacijos gausime atlikę tokį bandymą? Norėdami atsakyti į šį klausimą, galime samprotauti taip. Atlikus N tokių nepriklausomų bandymų, baigtis A_i pasikartos apytiksliai $N \cdot p_i$ kartų ir kiekvieną kartą gaunamos informacijos kiekis bus

$$I(A_i) = \log_2 \frac{1}{p_i} \, .$$

Taigi po visos bandymų serijos sukauptas informacijos kiekis bus

$$I_N \approx \sum_{i=1}^{n} N p_i \log_2 \frac{1}{p_i} \, .$$

Todėl vidutinis informacijos kiekis, gaunamas atlikus vieną bandymą, yra

$$\frac{I_N}{N} \approx \sum_{i=1}^{n} p_i \log_2 \frac{1}{p_i} \, .$$

Pastebėsime, kad dešinėje šios apytikslės lygybės pusėje esantis reiškinys yra lygus vidutinei įvykių $A_1, A_2, \ldots A_n$ informacijai.

1.3.3 apibrėžimas. *Tegul* $\mathcal{A} = \{A_1, A_2, \ldots, A_n\}$ *yra tikimybinės erdvės* (Ω, P) *įvykių su tikimybėmis* $p_i = P(A_i)$ *sistema. Įvykių sistemos* $\mathcal{A}$ *entropija vadinsime dydį*

$$H(\mathcal{A}) = H(p_1, p_2, \ldots, p_n) = \sum_{i=1}^{n} p_i \log_2 \frac{1}{p_i}\,.$$

Čia ir analogiškose sumose toliau visi dėmenys $p_i \log_2 \frac{1}{p_i}$ arba $p_i \log_2 p_i$ yra lygūs 0 , kai $p_i = 0$. Skaitytojams, kuriems toks susitarimas atrodo įtartinas, priminsime, kad

$$\lim_{p\to 0} p \log_2 \frac{1}{p} = \lim_{p\to 0} p \log_2 p = 0\,.$$

Aiškinantis įvairius sąryšius, kartais patogu interpretuoti entropiją kaip dydį, reiškiantį neapibrėžtumą, kurį jaučiame, nežinodami kuris iš sistemos $\mathcal{A}$ įvykių įvyks. Panagrinėkime dviejų įvykių su tikimybėmis p ir $1-p$ sistemą. Jos entropiją žymėsime $h(p) = H(p, 1-p)$. Taigi binarinės entropijos funkcija

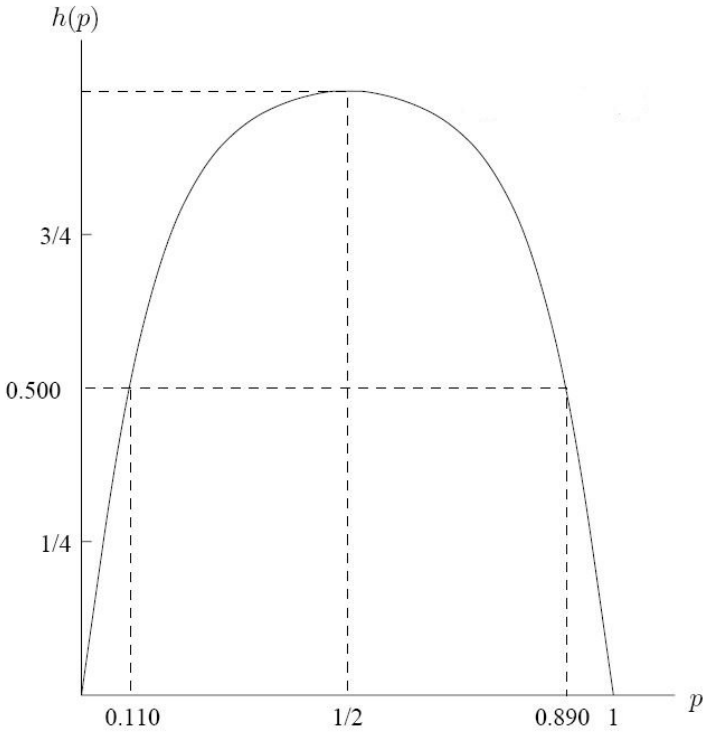
$$h(p) = p \log_2 \frac{1}{p} + (1-p) \log_2 \frac{1}{1-p}\,. \tag{1.6}$$

Šios funkcijos grafikas pavaizduotas 1.1 paveiksle. Matome, kad didžiausia entropijos reikšmė lygi $h\left(\frac{1}{2}\right) = 1$. Tai ir rodo, kad sunkiausiai prognozuojama dviejų įvykių sistema yra ta, kurioje abu įvykiai yra vienodai galimi, t.y. $p = \frac{1}{2}$. Ir priešingai - kai $p = 0$ arba $p = 1$, jokio neapibrėžtumo nėra, nes iš anksto aišku kuris iš dviejų įvykių įvyks. Tad nenuostabu, kad šiuo atveju entropija lygi $h(0) = h(1) = 0$.

Įvykus kokiam nors įvykiui B, sistemos $\mathcal{A}$ entropija gali pasikeisti. Pavyzdžiui, analizuodami prekybos centro duomenų bazės įrašus, pagal pirkėjo krepšelį bandome nuspėti pirkėjo amžių. Jei jis pirko tik duonos, tai aišku, nelabai ką tegalime pasakyti apie jo amžių. Bet, kai tarp jo pirkinių atrandame alų ir skrudintą duoną, neapibrėžtumas pirkėjo amžiaus atžvilgiu ženkliai sumažėja. Likusio neapibrėžtumo laipsnį nusako sąlyginė entropija

$$H(\mathcal{A}|B) = \sum_{i=1}^{n} P(A_i|B) I(A_i|B) = \sum_{i=1}^{n} P(A_i|B) \log_2 \frac{1}{P(A_i|B)}\,. \tag{1.7}$$

Tokių entropijų vidutinė reikšmė leidžia įvertinti neapibrėžtumą sistemos $\mathcal{A}$ atžvilgiu, kuris lieka, gavus informaciją apie kitą tos pačios tikimybinės erdvės įvykių sistemą $\mathcal{B} = \{B_1, B_2, \ldots, B_m\}$.



1.1 pav. Binarinės entropijos funkcija

1.3.4 apibrėžimas. *Dydį*

$$H(\mathcal{A}|\mathcal{B}) = \sum_{j=1}^{m} P(B_j) H(\mathcal{A}|B_j)$$

vadinsime sąlygine $\mathcal{A}$ *entropija* $\mathcal{B}$ *atžvilgiu, o entropijos pokytį*

$$I(\mathcal{A}, \mathcal{B}) = H(\mathcal{A}) - H(\mathcal{A}|\mathcal{B})$$

vadinsime sistemų $\mathcal{A}$ *ir* $\mathcal{B}$ *tarpusavio informacija.*

Sistemų $\mathcal{A}$ ir $\mathcal{B}$ jungtinės sistemos $\mathcal{A} \wedge \mathcal{B}$ entropiją žymėsime $H(\mathcal{A}, \mathcal{B})$. Prisiminę 1.2.3 apibrėžimą, turėsime

$$H(\mathcal{A}, \mathcal{B}) = H(\mathcal{A} \wedge \mathcal{B}) = \sum_{i=1}^{n} \sum_{j=1}^{m} P(A_i \cap B_j) \log_2 \frac{1}{P(A_i \cap B_j)} \,. \tag{1.8}$$

Šis apibrėžimas akivaizdžiai apibendrinamas ir didesniam įvykių sistemų skaičiui $k \geq 2$

$$H(\mathcal{A}_1, \mathcal{A}_2, \ldots, \mathcal{A}_k) = H(\mathcal{A}_1 \wedge \mathcal{A}_2 \wedge \cdots \wedge \mathcal{A}_k) \,.$$

Nagrinėjant entropijų savybes ir sąryšius, mums pravers vienas pagalbinis teiginys. Vektorių $(x_1, x_2, \ldots, x_n)$, sudarytą iš neneigiamų realiųjų skaičių, vadinsime tikimybiniu vektoriumi (kitaip: diskrečiuoju skirstiniu), jei

$$\sum_{i=1}^{n} x_i = 1\,.$$

1.3.2 lema (Gibbs'o nelygybė)**.** *Tegul $b > 1$. Tada bet kokiems tikimybiniams vektoriams $(x_1, x_2, \ldots, x_n)$ ir $(y_1, y_2, \ldots, y_n)$ teisinga nelygybė*

$$\sum_{i=1}^{n} x_i \log_b \left(\frac{y_i}{x_i}\right) \leq 0\,. \tag{1.9}$$

Nelygybė virsta lygybe tada ir tik tada, kai vektoriai $(x_1, x_2, \ldots, x_n)$ ir $(y_1, y_2, \ldots, y_n)$ sutampa.

Įrodymas. Kaip jau buvo minėta, kai $x_i = 0$, tai i - tasis nagrinėjamos sumos dėmuo taip pat lygus 0. Be to, pastebėsime, kad lemos teiginys yra teisingas, jei kuriam nors i, $x_i > 0$ ir $y_i = 0$, nes tada $x_i \log_b \left(\frac{y_i}{x_i}\right) = -\infty$. Todėl, nesiaurindami bendrumo, galime nagrinėti tik tokius vektorius, kuriems $y_i > 0$, jeigu $x_i > 0$. Taigi mums lieka įrodyti lemos teiginį, nelygybę (1.9) užrašius šitaip:

$$\sum_{i=1}^{n}{}^{*} x_i \log_b \left(\frac{y_i}{x_i}\right) \leq 0\,.$$

Čia simbolis * prie sumos ženklo reiškia, kad sumuojama tik pagal tas i reikšmes, kurioms $x_i > 0$. Vadinasi, ir visi y_i šioje sumoje yra teigiami. Padauginę pastarosios nelygybės abi puses iš teigiamo skaičiaus $\ln b$, pereisime prie natūraliųjų logaritmų

$$\sum_{i=1}^{n}{}^{*} x_i \ln \left(\frac{y_i}{x_i}\right) \leq 0\,. \tag{1.10}$$

Kadangi funkcija $y = \ln x$ yra iškila aukštyn, o jos grafiko liestinė taške $x = 1$ yra tiesė $y = x - 1$, tai

$$\ln x \leq x - 1$$

visiems $x > 0$. Be to, nelygybė virsta lygybe tik, kai $x = 1$. Iš čia išplaukia

$$\sum_{i=1}^{n}{}^{*} x_i \ln \left(\frac{y_i}{x_i}\right) \leq \sum_{i=1}^{n}{}^{*} x_i \left(\frac{y_i}{x_i} - 1\right) = \sum_{i=1}^{n}{}^{*} y_i - \sum_{i=1}^{n}{}^{*} x_i \leq 0\,.$$

Pastebėsime, kad abi pastarosios nelygybės, turi virsti lygybėmis, jei (1.10) suma lygi 0. Tačiau taip gali atsitikti tik, kai $x_i = y_i$ visiems $i = 1, 2, \ldots, n$. Lema įrodyta. □

Aptarsime pagrindines entropijos savybes. Pirmiausiai išsiaiškinsime kokios sistemos turi didžiausias ir kokios mažiausias entropijas.

1.3.3 teorema. *Įvykių sistemos* $\mathcal{A} = \{A_1, A_2, \ldots, A_n\}$ *entropija tenkina nelygybes*

$$0 \leq H(\mathcal{A}) \leq \log_2 n\,. \tag{1.11}$$

Mažiausią reikšmę $H(\mathcal{A}) = 0$ *ji įgyja tada ir tik tada, kai sistema sudaryta iš įvykių, kurių tikimybės yra* 0 *arba* 1. *Didžiausią entropiją* $H(\mathcal{A}) = \log_2 n$ *turės tos ir tik tos sistemos, kuriose visi įvykiai yra vienodai galimi, t.y.* $P(A_i) = \frac{1}{n}$ *visiems* $i = 1, 2, \ldots, n$.

Įrodymas. Pagal apibrėžimą entropija yra neneigiamų dėmenų suma

$$H(\mathcal{A}) = \sum_{i=1}^{n} P(A_i) \log_2 \frac{1}{P(A_i)}\,.$$

Todėl aišku, kad $H(\mathcal{A}) \geq 0$. Be to, tokia suma lygi 0 tada ir tik tada, kai visi dėmenys lygūs 0. Vadinasi, $P(A_i) = 0$ arba $P(A_i) = 1$ visiems $i = 1, 2, \ldots, n$. Iš tikrųjų tik viena iš šių tikimybių bus lygi 1, nes sistemą sudarančių įvykių tikimybių suma visada yra 1.

Antrąją nelygybę įrodysime skirtumui $H(\mathcal{A}) - \log_2 n$ pritaikę 1.3.2 lemą. Gausime

$$\begin{aligned} H(\mathcal{A}) - \log_2 n &= \sum_{i=1}^{n} P(A_i) \log_2 \frac{1}{P(A_i)} - \sum_{i=1}^{n} P(A_i) \log_2 n \\ &= \sum_{i=1}^{n} P(A_i) \log_2 \left(\frac{1/n}{P(A_i)} \right) \leq 0\,. \end{aligned}$$

Pastaroji nelygybė išplaukia iš (1.9) nelygybės, pasirinkus $x_i = P(A_i)$ ir $y_i = 1/n$. Iš čia pagal 1.3.2 lemą gauname ir paskutinį teoremos teiginį apie didžiausią entropiją. □

Pastaba. Įvykių sistemos entropija nepasikeistų iš sistemos pašalinus nulinės tikimybės įvykius (jei tokių yra). Todėl (1.11) nelygybėje n galima pakeisti teigiamą tikimybę turinčių sistemos $\mathcal{A}$ įvykių skaičiumi.

Atrodo, kad kuo sudėtingesnė ir daugiau įvykių turi sistema, tuo didesnė jos entropija. Tačiau tiesioginės priklausomybės tarp įvykių skaičiaus sistemoje ir jos entropijos dydžio, aišku, nėra.

1.3.3 pavyzdys. Tegul $\mathcal{A} = \{A_1, A_2, A_3, A_4\}$, $\mathcal{B} = \{B_1, B_2, B_3, B_4, B_5\}$ ir

$$\begin{aligned} P(A_1) &= P(A_2) = P(A_3) = P(A_4) = \frac{1}{4}; \\ P(B_1) &= P(B_2) = P(B_3) = P(B_4) = \frac{1}{16}, \quad P(B_5) = \frac{3}{4}. \end{aligned}$$

Apskaičiuojame sistemų $\mathcal{A}$ ir $\mathcal{B}$ entropijas

$$\begin{aligned} H(\mathcal{A}) &= 4 \cdot \frac{1}{4} \cdot \log_2 4; \\ H(\mathcal{B}) &= 4 \cdot \frac{1}{16} \cdot \log_2 16 + \frac{3}{4} \cdot \log_2 \frac{4}{3} \approx 1,3113. \end{aligned}$$

Kaip matome, $H(\mathcal{A}) > H(\mathcal{B})$. Gautąją nelygybę galime interpretuoti taip: numatyti, kuris iš įvykių įvyks, sistemoje $\mathcal{A}$ yra sunkiau, nei sistemoje $\mathcal{B}$.

Kai kurioms įvykių sistemų klasėms jų entropijų santykis visada nusakomas vienareikšmiškai. Pavyzdžiui, galime palyginti sistemos ir jos apvalkalo entropijas.

1.3.4 teorema. *Jei įvykių sistema* $\mathcal{B} = \{B_1, B_2, \ldots, B_m\}$ *yra tikslesnė už sistemą* $\mathcal{A} = \{A_1, A_2, \ldots, A_n\}$, *tai*

$$H(\mathcal{A}) \leq H(\mathcal{B}).$$

Įrodymas. Kadangi $\mathcal{B}$ yra tikslesnė už $\mathcal{A}$, tai pagal 1.2.1 teoremą

$$P(A_i) = \sum_{j=1}^{m} P(A_i \cap B_j) = \sum_{j \in J_i} P(B_j), \quad i = 1, 2, \ldots, n.$$

Čia J_i - poromis nesikertantys aibės $\{1, 2, \ldots, m\}$ poaibiai, tenkinantys sąlygą

$$\bigcup_{i=1}^{n} J_i = \{1, 2, \ldots, m\}.$$

Todėl

$$\begin{aligned} H(\mathcal{A}) &= -\sum_{i=1}^{n} P(A_i) \log_2 P(A_i) = -\sum_{i=1}^{n} \left(\sum_{j \in J_i} P(B_j) \right) \log_2 \left(\sum_{j \in J_i} P(B_j) \right) \\ &\leq -\sum_{i=1}^{n} \sum_{j \in J_i} P(B_j) \log_2 P(B_j) = -\sum_{j=1}^{m} P(B_j) \log_2 P(B_j) = H(\mathcal{B}) \end{aligned}$$

□

Prisiminkime sąlyginės entropijos $H(\mathcal{A}|B)$ apibrėžimą (1.7). Klausimas: ar, įvykus kokiam nors įvykiui B, sistemos $\mathcal{A}$ entropija sumažėja, kitaip sakant, ar galima tvirtinti, kad visada $H(\mathcal{A}|B) \leq H(\mathcal{A})$? Neigiamą atsakymą į šį klausimą pagrindžia toks pavyzdys.

1.3.4 pavyzdys. Tegul X ir Y yra atsitiktiniai dydžiai, įgyjantys reikšmes 0 ir 1, o jų bendrasis dvimatis skirstinys nusakytas lentele

$X\backslash Y$	0	1	$P(X=i)$
0	0,25	0,25	0,5
1	0	0,5	0,5
$P(Y=j)$	0,25	0,75	1

Nagrinėsime dvi įvykių sistemas

$$\mathcal{A} = \{\ \{X=0\},\ \{X=1\}\ \} \quad \text{ir} \quad \mathcal{B} = \{\ \{Y=0\},\ \{Y=1\}\ \}\,.$$

Kaip įprasta tokiais atvejais, sistemų $\mathcal{A}$ ir $\mathcal{B}$ entropijas žymėsime tiesiog $H(X)$ ir $H(Y)$. Prisiminę binarinės entropijos funkcijos $h(p)$ apibrėžimą (1.6), turėsime

$$\begin{aligned} H(X) &= h(0,5) = 1\,, \\ H(Y) &= h(0,25) \approx 0,811\,. \end{aligned}$$

Apskaičiuojame sąlygines tikimybes

$$\begin{aligned} P(Y=0|X=0) &= \frac{P(Y=0,\ X=0)}{P(X=0)} = \frac{0,25}{0,5} = 0,5\,, \\ P(Y=1|X=0) &= 1 - P(Y=0|X=0) = 0,5\,. \end{aligned}$$

Analogiškai gauname, kad

$$P(Y=0|X=1) = 0 \quad \text{ir} \quad P(Y=1|X=1) = 1\,.$$

Todėl, pagal (1.7) formulę, sąlyginės Y entropijos, kai žinomos atsitiktinio dydžio X reikšmės, bus

$$\begin{aligned} H(Y|X=0) &= h(0,5) = 1\,, \\ H(Y|X=1) &= h(0) = 0\,. \end{aligned}$$

Vidutinė šių entropijų reikšmė pagal 1.3.4 apibrėžimą yra

$$\begin{aligned} H(Y|X) &= P(X=0)\cdot H(Y|X=0) + P(X=1)\cdot H(Y|X=1) \\ &= \frac{1}{2}\cdot 1 + \frac{1}{2}\cdot 0 = \frac{1}{2}\,. \end{aligned}$$

Matome, kad $H(Y|X=1) < H(Y) < H(Y|X=0)$, tačiau $H(Y|X) < H(Y)$.

Šiame pavyzdyje vidutinė sąlyginė entropija $H(Y|X)$ nusako likusį (sumažėjusį) neapibrėžtumą Y atžvilgiu po to, kai gauta informacija apie atsitiktinį dydį X. Pasirodo toks sumažėjimas nėra atsitiktinis.

1.3.5 teorema. *Visoms įvykių sistemoms $\mathcal{A} = \{A_1, A_2, \ldots, A_n\}$ ir $\mathcal{B} = \{B_1, B_2, \ldots, B_m\}$ jų tarpusavio informacija yra neneigiama:*

$$I(\mathcal{A}, \mathcal{B}) \geq 0\,.$$

Be to, $I(\mathcal{A}, \mathcal{B}) = 0$ tada ir tik tada, kai sistemos $\mathcal{A}$ ir $\mathcal{B}$ yra nepriklausomos.

Įrodymas. Kadangi įvykiai su nulinėmis tikimybėmis neturi įtakos $I(\mathcal{A}, \mathcal{B})$ reikšmei, tai tarsime, kad $P(A_i)\cdot P(B_j) > 0$ visiems $i = 1, 2, \ldots, n;\quad j = 1, 2, \ldots, m\,.$
Pagal apibrėžimą

$$\begin{aligned} H(\mathcal{A}|\mathcal{B}) &= \sum_{j=1}^{m} P(B_j) H(\mathcal{A}|B_j) = \sum_{j=1}^{m}\sum_{i=1}^{n} P(B_j)P(A_i|B_j)\log_2 \frac{1}{P(A_i|B_j)} \\ &= \sum_{j=1}^{m}\sum_{i=1}^{n} P(A_i \cap B_j)\log_2\left(\frac{P(B_j)}{P(A_i\cap B_j)}\right)\,. \end{aligned} \tag{1.12}$$

Iš įvykių sistemos apibrėžimo išplaukia, kad visiems $i = 1, 2, \ldots, n$

$$P(A_i) = \sum_{j=1}^{m} P(A_i \cap B_j)\,.$$

Todėl

$$H(\mathcal{A}) = \sum_{i=1}^{n}\sum_{j=1}^{m} P(A_i\cap B_j)\log_2\frac{1}{P(A_i)}\,.$$

Prisiminę 1.3.4 apibrėžimą ir įstatę gautas išraiškas, turėsime

$$-I(\mathcal{A}, \mathcal{B}) = H(\mathcal{A}|\mathcal{B}) - H(\mathcal{A}) = \sum_{i=1}^{n}\sum_{j=1}^{m} P(A_i\cap B_j)\log_2\left(\frac{P(A_i)P(B_j)}{P(A_i\cap B_j)}\right) \leq 0\,. \tag{1.13}$$

Pastaroji nelygybė išplaukia iš 1.3.2 lemos, pritaikius ją tikimybiniams vektoriams $(x_1, x_2, \ldots, x_{mn})$ ir $(y_1, y_2, \ldots, y_{mn})$, kurių komponentės yra

$$x_{(i-1)m+j} = P(A_i \cap B_j) \quad \text{ir} \quad y_{(i-1)m+j} = P(A_i)P(B_j)\,, \quad i = 1, 2, \ldots, n\,, \quad j = 1, 2, \ldots, m\,.$$

Pagal tą pačią lemą gauname, kad (1.13) nelygybė virsta lygybe tada ir tik tada, kai visiems i ir j

$$P(A_i \cap B_j) = P(A_i)P(B_j)\,,$$

kitaip sakant, kai sistemos $\mathcal{A}$ ir $\mathcal{B}$ yra nepriklausomos. Teorema įrodyta. □

Iš ką tik įrodytos teoremos išplaukia, kad bet kokioms įvykių sistemoms $\mathcal{A}$ ir $\mathcal{B}$

$$H(\mathcal{A}|\mathcal{B}) \leq H(\mathcal{A})\,,$$

o šių entropijų lygybė yra sistemų $\mathcal{A}$ ir $\mathcal{B}$ nepriklausomumo kriterijus.
Jungtinės sistemos entropijos $H(\mathcal{A}, \mathcal{B})$ reikšmė priklauso ne tik nuo ją sudarančių sistemų, bet ir nuo jų priklausomumo laipsnio.

1.3.6 teorema. *Įvykių sistemoms $\mathcal{A}$ ir $\mathcal{B}$ teisingi sąryšiai*

$$H(\mathcal{A}, \mathcal{B}) = H(\mathcal{B}) + H(\mathcal{A}|\mathcal{B})\,, \tag{1.14}$$

$$H(\mathcal{A}, \mathcal{B}) = H(\mathcal{A}) + H(\mathcal{B}) - I(\mathcal{A}, \mathcal{B})\,. \tag{1.15}$$

Įrodymas. Pasiremsime kai kuriais 1.3.5 teoremos įrodymo tarpiniais rezultatais. Kadangi visiems $j = 1, 2, \ldots, m$

$$P(B_j) = \sum_{i=1}^{n} P(A_i \cap B_j)\,,$$

tai lygybę (1.12) galime parašyti taip

$$\begin{aligned} H(\mathcal{A}|\mathcal{B}) &= \sum_{j=1}^{m}\sum_{i=1}^{n} P(A_i \cap B_j) \log_2 \left(\frac{1}{P(A_i \cap B_j)}\right) - \sum_{j=1}^{m} \left(\sum_{i=1}^{n} P(A_i \cap B_j)\right) \log_2 \left(\frac{1}{P(B_j)}\right) \\ &= H(\mathcal{A}, \mathcal{B}) - H(\mathcal{B})\,. \end{aligned}$$

Iš čia gauname (1.14) lygybę.

Pagal tarpusavio informacijos apibrėžimą 1.3.4

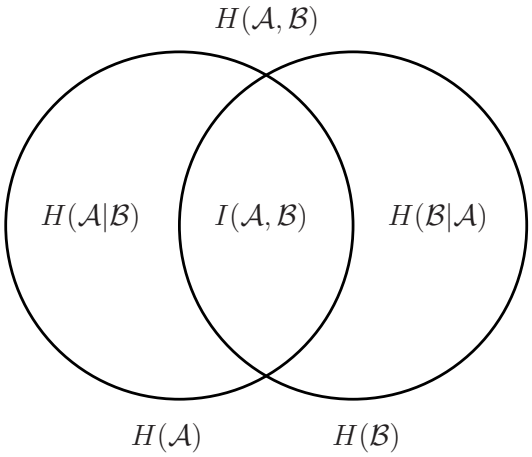
$$H(\mathcal{A}|\mathcal{B}) = H(\mathcal{A}) - I(\mathcal{A}, \mathcal{B})\,.$$

Įstatę šią sąlyginės entropijos išraišką į (1.14), gauname (1.15) lygybę ir tuo pačiu baigiame teoremos įrodymą. □

Pastebėsime, kad $H(\mathcal{A},\mathcal{B}) = H(\mathcal{B},\mathcal{A})$. Todėl iš ką tik įrodytos (1.15) lygybės išplaukia, kad simetriška yra ir tarpusavio informacija, t.y.

$$I(\mathcal{A},\mathcal{B}) = I(\mathcal{B},\mathcal{A}).$$

Dviejų sistemų entropiją, sąlyginių entropijų bei tarpusavio informacjos sąryšius patogu vaizduoti 1.2 paveiksle pateikiama diagrama.



1.2 pav. Entropija ir informacija

Iš 1.3.5 teoremos ir (1.15) lygybės gauname tokį jungtinės sistemos entropijos įvertį

$$H(\mathcal{A},\mathcal{B}) \leq H(\mathcal{A}) + H(\mathcal{B}).$$

Pritaikius indukciją, pastarąją nelygybę nesudėtinga apibendrinti didesniam įvykių sistemų skaičiui.

1.3.7 teorema. *Jei $\mathcal{A}_1, \mathcal{A}_2, \ldots, \mathcal{A}_k$ yra tos pačios diskrečiosios tikimybinės erdvės įvykių sistemos, tai*

$$H(\mathcal{A}_1, \mathcal{A}_2, \ldots, \mathcal{A}_k) \leq H(\mathcal{A}_1) + H(\mathcal{A}_2) + \cdots + H(\mathcal{A}_k).$$

Ši nelygybė virsta lygybe tada ir tik tada, kai $\mathcal{A}_1, \mathcal{A}_2, \ldots, \mathcal{A}_k$ yra nepriklausomos sistemos.

1.4 Atsitiktinių dydžių entropijos

Su diskrečiojo atsitiktinio dydžio entropijos sąvoka mes jau buvome susidūrę, nagrinėdami 1.3.4 pavyzdį. Tad visiškai suprantamas bus toks apibrėžimas.

1.4.1 apibrėžimas. *Diskrečiojo atsitiktinio dydžio X, įgyjančio reikšmes $x_1, x_2, \dots$, entropija $H(X)$ yra lygi įvykių sistemos, sudarytos iš įvykių $A_i = \{X = x_i\}$, $i = 1, 2, \dots$, entropijai, t.y.,*

$$H(X) = \sum_i P(A_i) \log_2 \frac{1}{P(A_i)} .$$

Todėl diskrečiojo atsitiktinio dydžio entropijos interpretacija bei savybės niekuo nesiskiria nuo 1.3 skyrelyje aptartos įvykių sistemos entropijos. Pateiksime keletą iš galimų, su entropijos sąvoka susijusių, uždavinių.

1.4.1 pavyzdys. Tegul T yra koks nors tekstas. Pavyzdžiui,

$$T = abrakadabra$$

Kaip matome, tekste sutinkamos penkios skirtingos raidės. Tad, norėdami parašyti dvejetainį T kodą, kiekvienai raidei turėsime skirti po tris bitus. Pavyzdžiui,

$$a \mapsto 000,\ b \mapsto 001,\ d \mapsto 010,\ k \mapsto 011,\ r \mapsto 100 . \tag{1.16}$$

Vadinasi, viso teksto kodo ilgis bus 33 bitai. Ar įmanoma sukonstruoti trumpesnį vienareikšmiškai dekoduojamą kodą?

Galime manyti, kad tekstas T yra sudarytas iš 11 atsitiktinio dydžio X realizacijų. Kitaip sakant, atsitiktinis dydis X reiškia atsitiktinai pasirinktą T raidę. Pagal raidžių dažnius tekste sudarome atsitiktinio dydžio X skirstinį

X	a	b	d	k	r
P	$\frac{5}{11}$	$\frac{2}{11}$	$\frac{1}{11}$	$\frac{1}{11}$	$\frac{2}{11}$

Apskaičiavę X entropiją (kartais ji dar vadinama teksto T entropija), sužinosime vidutinį informacijos kiekį, tenkantį vienai teksto raidei. Taigi

$$H(X) = \frac{5}{11} \log_2 \frac{11}{5} + 2 \cdot \frac{2}{11} \log_2 \frac{11}{2} + 2 \cdot \frac{1}{11} \log_2 11 \approx 2,0404 .$$

Tai yra vadinamasis *Šenono rėžis* vidutiniam vienos raidės kodo bitų skaičiui. Kitaip sakant, nėra tokio vienareikšmiškai dekoduojamo kodo, kuriuo pakeitus (1.16) kodą, tekstą T atvaizduotume trumpesne nei $11 \cdot H(X) \approx 22,44$ bitų seka. Tačiau ne visiems tekstams Šenono rėžis yra pasiekiamas. Tie skaitytojai, kurie šiek tiek žino duomenų kompresijos algoritmus, supras kaip konstruojamas tekstui T geriausią rezultatą duodantis kodas (beje ne vienintelis):

$$a \mapsto 0,\ b \mapsto 100,\ d \mapsto 110,\ k \mapsto 111,\ r \mapsto 101\,.$$

Jis tekstą T užkoduoja 23 bitų seka.

- Jei norite įsitikinti, kad Šenono rėžis yra pasiekiamas, pabandykite užkoduoti tekstą $T' = abrakadabrakaada$.

1.4.2 pavyzdys. Atliekamas tyrimas, siekiant nustatyti kokie faktoriai įtakoja galimybę susirgti tam tikra liga. Ligos požymį žymėsime kintamuoju Y, įgyjančiu reikšmes S ir N (serga, neserga). Nagrinėjami trys faktoriai:

X_1 – paciento lytis, galimos reikšmės $\{M, V\}$;

X_2 – rūkymas, galimos reikšmės $\{taip, ne\}$;

X_3 – kraujospūdis, galimos reikšmės $\{mažas, normalus, didelis\}$.

100 pacientų tyrimo duomenys pateikti 1.1 lentelėje. Kuris faktorius labiausiai įtakoja polinkį susirgti? Norėdami atsakyti į šį klausimą, turime išsiaiškinti, kaip pasikeičia Y entropija, vieno ar kito požymio atžvilgiu. Kitaip sakant, turime palyginti tris tarpusavio informacijas

$$I(Y, X_i) = H(Y) - H(Y|X_i)\,, \quad i = 1, 2, 3.$$

Pirmiausiai rasime $H(Y)$. Atsitiktinio dydžio Y skirstinys yra

Y	N	S
P	$0,44$	$0,56$

Vadinasi,

$$H(Y) = h(0,44) \approx 0,989587521$$

Paciento lytis (X_1)	*Rūkymas* (X_2)	*Kraujospūdis* (X_3)	*Ligos požymis* (Y)	Pacientų skaičius
M	*ne*	*mažas*	N	5
M	*ne*	*mažas*	S	2
M	*ne*	*normalus*	N	10
M	*ne*	*didelis*	S	6
M	*taip*	*mažas*	N	4
M	*taip*	*mažas*	S	2
M	*taip*	*normalus*	N	8
M	*taip*	*didelis*	N	1
M	*taip*	*didelis*	S	8
V	*ne*	*normalus*	N	8
V	*ne*	*didelis*	S	10
V	*ne*	*mažas*	N	2
V	*taip*	*mažas*	S	7
V	*taip*	*normalus*	N	6
V	*taip*	*normalus*	S	5
V	*taip*	*didelis*	S	16

1.1 lentelė. Ligą įtakojantys faktoriai

Skaičiuosime $H(Y|X_1)$. Pasinaudoję (1.7) formule, pagal 1.1 lentelės duomenis gausime

$$\begin{aligned} H(Y|X_1=M) &= P(Y=N|X_1=M)\log_2\frac{1}{P(Y=N|X_1=M)} \\ &+ P(Y=S|X_1=M)\log_2\frac{1}{P(Y=S|X_1=M)} \\ &= h\left(\frac{28}{46}\right)\approx 0,965636133 \end{aligned}$$

Analogiškai

$$H(Y|X_1=V)=h\left(\frac{16}{54}\right)\approx 0,876716289$$

Dabar, prisiminę 1.3.4 apibrėžimą, randame $H(Y|X_1)$

$$\begin{aligned} H(Y|X_1) &= P(X_1=M)H(Y|X_1=M)+P(X_1=V)H(Y|X_1=V) \\ &= 0,46\cdot h\left(\frac{28}{46}\right)+0,54\cdot h\left(\frac{16}{54}\right)\approx 0,917619417 \end{aligned}$$

Analogiškai skaičiuojame ir likusias entropijas

$$H(Y|X_2 = ne) = h\left(\frac{25}{43}\right) \approx 0,980798365;$$
$$H(Y|X_2 = taip) = h\left(\frac{19}{57}\right) \approx 0,918295834;$$
$$H(Y|X_2) = 0,43 \cdot h\left(\frac{25}{43}\right) + 0,57 \cdot h\left(\frac{19}{57}\right) \approx 0,945171922;$$
$$H(Y|X_3 = mažas) = h\left(\frac{11}{22}\right) = 1;$$
$$H(Y|X_3 = normalus) = h\left(\frac{32}{37}\right) \approx 0,571354974;$$
$$H(Y|X_3 = didelis) = h\left(\frac{1}{41}\right) \approx 0,165427034;$$
$$H(Y|X_3) = 0,22 \cdot 1 + 0,37 \cdot h\left(\frac{32}{37}\right) + 0,41 \cdot h\left(\frac{1}{41}\right) \approx 0,499226424;$$

Dabar jau galime rasti ieškomąsias tarpusavio informacijas

$$I(Y, X_1) \approx 0,071968104\,,$$
$$I(Y, X_2) \approx 0,044415599\,,$$
$$I(Y, X_3) \approx 0,490361097\,.$$

Kaip matome, diagnozuojant ligą, labiausiai informatyvus yra kraujospūdžio dydis. Tuo tarpu paciento lytis ir jo įprotis rūkyti galimybę susirgti įtakoja nežymiai.

Pastaba. 1.1 lentelėje pateikti duomenys yra fiktyvūs ir skirti tik aptariamų sąvokų iliustracijai. Todėl suformuluotos išvados jokiu būdu nereiškia, kad rūkymas nekenkia sveikatai!

Tolydžiųjų atsitiktinių dydžių entropija apibrėžiama kitaip. Skiriasi ir jos interpretacija.

1.4.2 apibrėžimas. *Tolydžiojo atsitiktinio dydžio X su tankio funkcija $p_X(x)$ entropija $H(X)$ yra lygi*

$$H(X) = -\int_{-\infty}^{\infty} p_X(x) \log_2 p_X(x)\, dx\,.$$

Iš karto reikia pažymėti, kad toldžiojo atsitiktinio dydžio entropijos negalima interpretuoti kaip vidutinės informacijos, nes šiuo atveju $P(X = x) = 0$, nepriklausomai nuo tankio

funkcijos $p_X(x)$ reikšmės. Be to, toldžiojo atsitiktinio dydžio entropija gali būti ir neigiama. Panagrinėkime tokį pavyzdį.

1.4.3 pavyzdys. Tegul X yra normalusis (kitaip: Gauso) atsitiktinis dydis su vidurkiu a ir dispersija σ^2, $\sigma > 0$. Tokio atsitiktinio dydžio tankio funkcija yra

$$p_X(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{(x-a)^2}{2\sigma^2}\right\}.$$

Rasime jo entropiją. Pagal apibrėžimą

$$\begin{aligned} H(X) &= \int_{-\infty}^{\infty} p_X(x) \left(\frac{(x-a)^2}{2\sigma^2}(\log_2 \mathrm{e}) + \log_2(\sqrt{2\pi}\sigma)\right) dx \\ &= \frac{\log_2 \mathrm{e}}{2\sigma^2} \int_{-\infty}^{\infty} (x-a)^2 p_X(x)\, dx + \log_2(\sqrt{2\pi}\sigma) \int_{-\infty}^{\infty} p_X(x)\, dx \\ &= \frac{\log_2 \mathrm{e}}{2\sigma^2} \cdot \sigma^2 + \log_2(\sqrt{2\pi}\sigma) \cdot 1 = \frac{1}{2}\log_2(2\mathrm{e}\pi\sigma^2). \end{aligned}$$

Matome, kad normaliojo atsitiktinio dydžio entropija priklauso tik nuo jo dispersijos σ^2 ir yra neigiama, kai $\sigma^2 < \frac{1}{2\mathrm{e}\pi}$.

Įrodysime dar vieną įdomią normaliojo atsitiktinio dydžio savybę. Tuo tikslu prisiminkime (1.9) nelygybę. Panašiai įrodoma ir jos "tolydžioji" versija

$$\int_{-\infty}^{\infty} p_X(x) \log_b\left(\frac{p_Y(x)}{p_X(x)}\right) dx \le 0. \tag{1.17}$$

Čia $b > 1$, o $p_X(x)$ ir $p_Y(x)$ - bet kokios tankio funkcijos. Pasinaudoję šia nelygybe, įrodysime tokį teiginį.

1.4.1 teorema. *Jei absoliučiai tolydus atsitiktinis dydis X turi baigtinę dispersiją $\mathbf{D}X = \sigma^2$, $\sigma > 0$, tai jo entropija*

$$H(X) \le \frac{1}{2}\log_2(2\mathrm{e}\pi\sigma^2).$$

Įrodymas. Atsitiktinio dydžio X vidurkį pažymėsime $\mathbf{E}X = a$. Tegul Y yra normalusis atsitiktinis dydis, turintis tokius pat vidurkį ir dispersiją, kaip ir atsitiktinis dydis X. Tada jo tankio funkcijos logaritmas

$$\log_2 p_Y(x) = -\frac{(x-a)^2}{2\sigma^2}(\log_2 \mathrm{e}) - \log_2(\sqrt{2\pi}\sigma). \tag{1.18}$$

Atsitiktiniams dydžiams X ir Y pritaikysime (1.17) nelygybę. Pasirinkę $b = 2$, ją galime parašyti taip

$$-\int_{-\infty}^{\infty} p_X(x) \log_2 p_X(x)\, dx \le -\int_{-\infty}^{\infty} p_X(x) \log_2 p_Y(x)\, dx\,.$$

Kairėje pastarosios nelygybės pusėje esantis reiškinys, pagal apibrėžimą, yra atsitiktinio dydžio X entropija $H(X)$. Vadinasi,

$$H(X) \le -\int_{-\infty}^{\infty} p_X(x) \log_2 p_Y(x)\, dx\,.$$

Įrašę $\log_2 p_Y(x)$ išraišką (1.18), gausime

$$\begin{aligned} H(X) &\le \int_{-\infty}^{\infty} p_X(x) \left(\frac{(x-a)^2}{2\sigma^2} (\log_2 \mathrm{e}) + \log_2(\sqrt{2\pi}\sigma) \right) dx \\ &= \frac{\log_2 \mathrm{e}}{2\sigma^2} \int_{-\infty}^{\infty} (x-a)^2 p_X(x)\, dx + \log_2(\sqrt{2\pi}\sigma) \int_{-\infty}^{\infty} p_X(x)\, dx \\ &= \frac{\log_2 \mathrm{e}}{2\sigma^2} \cdot \mathbf{D}X + \log_2(\sqrt{2\pi}\sigma) \cdot 1 = \frac{1}{2} \log_2(2\mathrm{e}\pi\sigma^2)\,. \end{aligned}$$

□

Iš 1.4.1 teoremos išplaukia, kad tarp visų absoliučiai tolydžių atsitiktinių dydžių, turinčių baigtinę dispersiją σ^2, didžiausią entropiją $\frac{1}{2}\log_2(2\mathrm{e}\pi\sigma^2)$ turi normalusis atsitiktinis dydis.

Literatūra

[1] V.Čekanavičius, G.Murauskas. *Statistika ir jos taikymai, I.* Vilnius: TEV, 2000.

[2] D.Hankerson, G.Harris, P.Johnson. *Introduction to Information Theory and Data Compression.* Chapman&Hall CRC, 2003.

[3] J.Kubilius. *Tikimybių teorija ir matematinė statistika.* Vilnius: Mokslas , 1980. Antrasis patais. ir papild. leid. Vilnius: VU, 1996.

[4] David J.C.MacKay. *Information Theory, Inference and Learning Algorithms.* Cambridge University Press, 2003.
Tinklalapis: http://www.inference.phy.cam.ac.uk/itprnn/book.pdf

[5] J.L.Massey. *Applied Digital Information Theory, I.* Lecture Notes, ETH, Zurich, 1998.
Tinklalapis: http://www.isiweb.ee.ethz.ch/archive/massey_scr/adit1.pdf

[6] S.Roman. *Coding and Information Theory.* Springer, 1992

[7] V.Stakėnas. *Informacijos kodavimas.* Vilnius: VU, 1996.

[8] V.Stakėnas. *Kodai ir šifrai.* Vilnius: TEV, 2007.