12 Paskaita. Ribinės teoremos

12.1 Markovo ir Čebyševo nelygybės

Teorema. Markovo nelygybė. *Tarkime, kad Y yra atsitiktinis dydis, turintis vidurkį. Tada bet kokiam $\delta > 0$ teisinga nelygybė*

$$P(|Y| \geqslant \delta) \leqslant \frac{\mathbf{E}|Y|}{\delta}. \tag{1}$$

Atitinkamai

$$P(|Y| < \delta) > 1 - \frac{\mathbf{E}|Y|}{\delta}.$$

Teorema. Čebyševo nelygybė. *Tarkime, kad X yra atsitiktinis dydis, turintis dispersiją. Tada bet kokiam $\varepsilon > 0$ teisinga tikslesnė nelygybė*

$$P(|X - \mathbf{E}X| \geqslant \varepsilon) \leqslant \frac{\mathbf{D}X}{\varepsilon^2}. \tag{2}$$

Atitinkamai

$$P(|X - \mathbf{E}X| < \varepsilon) > 1 - \frac{\mathbf{D}X}{\varepsilon^2}.$$

Pavyzdys. Atsitiktinio dydžio X skirstinys duotas lentele

X	1	2	3	4	5	6
P	0,05	0,1	0,25	0,3	0,2	0,1

Įvertinkite $\mathbf{P}(|X - EX| < 2)$ remdamiesi Čebyševo nelygybe. Raskite tikslią šios tikimybės reikšmę.

Sprendimas. Apskaičiuokime atsitiktinio dydžio X vidurkį ir dispersiją:
$\mathbf{E}X = \sum_{n=1}^{6} x_i p_i = 3,8,$
$\mathbf{E}X^2 = \sum_{n=1}^{6} x_i^2 p_i = 16,1,$
$\mathbf{D}X = \mathbf{E}X^2 - (\mathbf{E}X)^2 = 1,66.$
$\mathbf{P}(|X - EX| < 2) > 1 - \frac{\mathbf{D}X}{2^2} = 1 - \frac{1,66}{4} = 0,585.$
Taigi $\mathbf{P}(|X - EX| < 2) > 0,585.$
Dabar apskaičiuosime tikslią šios tikimybės reikšmę.
$\mathbf{P}(|X - EX| < 2) = \mathbf{P}(|X - 3,8| < 2) = \mathbf{P}(1,8 < X < 5,8|) = \mathbf{P}(X = 2) + \mathbf{P}(X = 3) + \mathbf{P}(X = 4) + \mathbf{P}(X = 5) = 0,1 + 0,25 + 0,3 + 0,2 = 0,85.$

12.2 Didžiųjų skaičių dėsnis

Teorema. *Tegu $X_1, X_2, \ldots$ yra nepriklausomi atsitiktiniai dydžiai, turintys tą patį vidurkį $\mathbf{E}X_j = a$ ir tą pačią dispersiją. Tada su kiekvienu $\varepsilon > 0$*

$$P\left(\left|\frac{X_1 + X_2 + \ldots + X_n}{n} - a\right| \geqslant \varepsilon\right) \to 0, n \to \infty.$$

Įrodymas. Teorema tvirtina, kad su didelėmis n reikšmėmis tikimybė, kad aritmetinis atsitiktinių dydžių reikšmių vidurkis skirsis nuo a daugiau kaip dydžiu ε, yra maža. Pažymėkime $\mathbf{D}X_j = \sigma^2$ ir $X = \dfrac{X_1 + X_2 + \ldots + X_n}{n}$. Aišku, kad

$$\mathbf{E}X = \frac{\mathbf{E}X_1 + \mathbf{E}X_2 + \ldots + \mathbf{E}X_n}{n} = \frac{na}{n} = a,$$
$$\mathbf{D}X = \frac{\mathbf{D}X_1 + \mathbf{D}X_2 + \ldots + \mathbf{D}X_n}{n^2} = \frac{n\sigma^2}{n^2} = \frac{\sigma^2}{n},$$

tai įstatę į Čebysevo nelygybę (2), gausime

$$P\left(\left|\frac{X_1 + X_2 + \ldots + X_n}{n} - a\right| \geqslant \varepsilon\right) \leqslant \frac{\mathbf{D}X}{\varepsilon^2} = \frac{\sigma^2}{n\varepsilon^2} \to 0, n \to \infty.$$

Didžiųjų skaičių dėsnis matematine forma išreiškia tai, ką mes intuityviai suprantame. Pavyzdžiui, gerai žinome, kad metus simetrišką monetą daug kartų (N kartų), herbas, tikriausiai, pasirodys maždaug $N/2$ kartų.

Jei turime Bernulio schemą, m – sėkmių skaičius atlikus n bandymų,

$$\mathbf{P}\left(\left|\frac{m}{n} - p\right| < \varepsilon\right) \to 1, n \to \infty.$$
$$\mathbf{P}\left(\left|\frac{m}{n} - p\right| < \varepsilon\right) > 1 - \frac{pq}{n\varepsilon^2}.$$

Pavyzdys. Pasėta 400 sėklų. Kiekviena jų išdygsta su tikimybe lygia 0,85. Įvertinkite tikimybę, kad išdygusių sėklų skaičius bus nuo 300 iki 380.

Sprendimas. Turime Bernulio schemą, X_j yra Bernulio atsitiktiniai dydžiai tikimybe $p = 0,85$. Pažymėkime $X = X_1 + X_2 + \ldots + X_{400}$.
$EX = np = 400 \cdot 0,85 = 340$.
$\mathbf{P}\left(300 < X < 380\right) = \mathbf{P}\left(|X - 340| < 40\right) = \mathbf{P}\left(\left|\dfrac{X}{400} - 0,85\right| < \dfrac{40}{400}\right)$
$\mathbf{P}\left(\left|\dfrac{m}{n} - p\right| < 0,1\right) > 1 - \dfrac{0,85 \cdot 0,15}{400 \cdot 0,1^2} \approx 0,97.$

Pavyzdys. Moneta mėtoma 1000 kartų. Įvertinkite tikimybę, kad herbo pasirodymų santykinio dažnio nukrypimas nuo $p = \dfrac{1}{2}$ bus mažesnis už 0,1.
Sprendimas. Turime Bernulio schemą.

$$\mathbf{P}\left(\left|\frac{m}{n} - p\right| < 0,1\right) > 1 - \frac{pq}{n \cdot \varepsilon^2} = 1 - \frac{\frac{1}{2} \cdot \frac{1}{2}}{1000 \cdot 0,01} = \frac{39}{40}.$$

12.3 Centrinė ribinė teorema

Tegu atsitiktiniai dydžiai $X_1, X_2, \ldots$ yra nepriklausomi ir turi tą pačią pasiskirstymo funkciją. Tada jų vidurkiai bei dispersijos (jeigu jos, žinoma, egzistuoja) bus tie patys. Pažymėkime $\mathbf{E}X_j = a, \mathbf{D}X_j = \sigma^2$. Dydžiams teisingas didžjųjų skaičių dėsnis, t.y. dydis

$$Y_n = \frac{X_1 + X_2 + \ldots + X_n}{n} - a = \frac{X_1 + X_2 + \ldots + X_n - na}{n},$$

kai n didėja, darosi vis labiau panašus į išsigimusį atsitiktinį dydį Y, įgyjantį reikšmę $0, P(Y = 0) = 1$. Dydžių Y_n vidurkiai lygūs nuliui, o dispersijos mažėja: $\mathbf{E}Y_n = 0, \mathbf{D}Y_n = \sigma^2/n \to 0, n \to \infty$. O dabar kiek pakeiskime šį dydį ir apibrėžkime

$$Z_n = \frac{X_1 + X_2 + \ldots + X_n - na}{\sigma\sqrt{n}}.$$

Aišku, kad $\mathbf{E}Z_n = 0$. Apskaičiuokime dispersiją pasinaudodami dydžių nepriklausomumu:

$$\mathbf{D}Z_n = \mathbf{D}\left[\frac{X_1 + X_2 + \ldots + X_n}{\sigma\sqrt{n}} - \frac{na}{\sigma\sqrt{n}}\right] = \mathbf{D}\left[\frac{X_1 + X_2 + \ldots + X_n}{\sigma\sqrt{n}}\right]$$

$$= \mathbf{D}\left[\frac{X_1}{\sigma\sqrt{n}}\right] + \ldots + \mathbf{D}\left[\frac{X_n}{\sigma\sqrt{n}}\right] = \frac{\sigma^2}{\sigma^2 n} + \ldots + \frac{\sigma^2}{\sigma^2 n} = 1.$$

Taigi visi dydžiai Z_n turi tą patį vidurkį ir tą pačią dispersiją. Ką gi apie juos galima pasakyti, kai $n \to \infty$? Ir čia pasirodo senas pažįstamas! Kai n didėja, dydžiai Z_n darosi vis panašesni į standartinį normalųjį dydį! Tokia yra vienos iš pačių svarbiausių tikimybių teoremų – centrinės ribinės teoremos – esmė.

Centrinė ribinė teorema. *Tegu $X_1, X_2, \ldots$ yra nepriklausomi atsitiktiniai dydžiai, turintys tą pačią pasiskirstymo funkciją, vidurkį ir dispersiją $\mathbf{E}X_j = a, \mathbf{D}X_j = \sigma^2$. Tada kiekvienam x, kai $n \to \infty$, teisingas teiginys*

$$P\left(\frac{X_1 + X_2 + \ldots + X_n - na}{\sigma\sqrt{n}} < x\right) \to \Phi(x), \tag{3}$$

čia

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-x^2/2} dx,$$

yra standartinio normaliojo dydžio $X \sim \mathcal{N}(0, 1)$ pasiskirstymo funkcija.

Taigi centrinė ribinė teorema teigia, kad nepriklausomai nuo individualių dydžių X_j savybių, dydis Z_n, kai n didelis, yra panašus į standartinį normalųjį dydį.

Muavro-Laplaso teorema, kurią naudojome Bernulio schemai tyrinėti, yra atskiras centrinės ribinės teoremos atvejis, kai X_j turi Bernulio skirstinį. Iš tikrųjų, jeigu atsitiktinis dydis $X_j = 1$, kai j-ajame Bernulio schemos bandyme įvyko sėkmė, ir $X_j = 0$, kai nesėkmė, tai dydžiai X_j yra nepriklausomi ir $a = \mathbf{E}X_j = p; \sigma^2 = \mathbf{D}X_j = pq$.

Pažymėkime $S_n = X_1 + X_2 + \ldots + X_n$. Bernulio schemos atveju S_n reiškia sėkmių skaičių atlikus n bandymų. Taigi (3) galime užrašyti taip:

$$P\left(\frac{S_n - np}{\sqrt{npq}} < x\right) \to \Phi(x),$$

arba

$$P\left(\frac{S_n - \mathbf{E}S_n}{\sqrt{\mathbf{D}S_n}} < x\right) \to \Phi(x).$$

Centrinę ribinę teoremą galima interpretuoti kaip didžiųjų skaičių dėsnio patikslinimą. Iš tikrųjų, didžiųjų skaičių dėsnis reiškia, kad

$$\frac{S_n}{n} = \frac{X_1 + X_2 + \ldots + X_n}{n} \approx a.$$

Savo ruožtu centrinę ribinę teoremą galime suvokti kaip tvirtinimą

$$Z_n = \frac{X_1 + X_2 + \ldots + X_n - na}{\sigma\sqrt{n}} = \frac{S_n/n - a}{\sigma/\sqrt{n}} \approx X, X \sim \mathcal{N}(0,1).$$

Tada

$$\frac{S_n}{n} \approx a + \frac{\sigma}{\sqrt{n}} \cdot X \sim \mathcal{N}(a, \sigma^2/n).$$

Taigi centrinė ribinė teorema teigia, kad su didelėmis n reikšmėmis aritmetinis vidurkis S_n/n yra panašus į normalųjį dydį su maža dispersija.
Jei turime Bernulio atsitiktinių dydžių sumą su tikimybėmis p, tai

$$\mathbf{P}\left(\left|\frac{m}{n} - p\right| < \varepsilon\right) \approx 2\Phi\left(\varepsilon\sqrt{\frac{n}{pq}}\right) - 1.$$

Pavyzdys. Šaudoma į apvalų taikinį, suskirstytą 4 koncentriniais apskritimais. Atsižvelgiant į pataikymą, šaulys gauna taškus, kurių skaičius apibrėžtas skirstiniu:

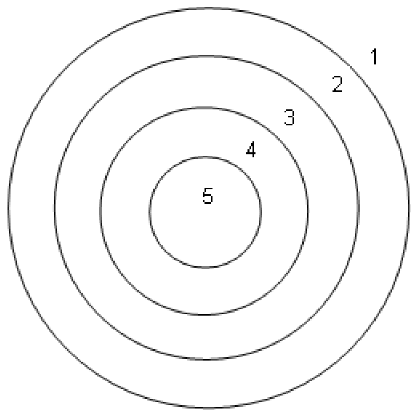
X	1	2	3	4	5
$\mathbf{P}_X$	0,3	0,25	0,2	0,15	0,1

Apskaičiuokite tikimybę, kad po 100 šūvių šaulys surinks nuo 240 taškų iki 280 taškų.
Sprendimas. Pažymėkime X_i – i-ju šūviu surinktų taškų skaičių.

$\mathbf{E}X_i = 1 \cdot 0,3 + 2 \cdot 0,25 + 3 \cdot 0,2 + 4 \cdot 0,15 + 5 \cdot 0,1 = 2,5,$

$\mathbf{E}X_i^2 = 8, \mathbf{D}X_i = 1,75, S_{100} = X_1 + X_2 + \ldots + X_{100},$

$\mathbf{E}S_{100} = 100 \cdot 2,5 = 250, \quad \mathbf{D}S_{100} = 100 \cdot 1,75 = 175.$



1 pav.: Koncentriniai apskritimai

S_{100} – po 100 šūviu surinktų taškų skaičius. Pagal centrinę ribinę teoremą

$\mathbf{P}\left(\frac{S_{100}-\mathbf{E}S_{100}}{\sqrt{\mathbf{D}S_{100}}}<x\right)\approx\Phi(x).$

$\mathbf{P}\left(240\leqslant S_n<280\right)\approx\Phi\left(\frac{280-250}{\sqrt{175}}\right)-\Phi\left(\frac{240-250}{\sqrt{175}}\right)\approx\Phi(2,27)-\Phi(-0,76)=$ $\Phi(2,27)+\Phi(0,76)-1\approx 0,9893+0,7764-1\approx 0,7657.$

Pavyzdys. Monte Carlo metodas.

Didžiųjų skaičių dėsnis – vienas iš pagrindinių instrumentų, kuriais naudojamės, kai norime gauti žinių iš sukauptų duomenų. Tarkime, mums rūpi ryšio tarp mūsų kompiuterio ir kokio nors serverio nustatymo laikas. Šis laikas priklauso nuo įvairiausių faktorių, taigi yra atsitiktinis dydis. Pažymėkime jį T. Kokia tikimybė, kad ryšiui nustatyti prireiks mažiau kaip, pavyzdžiui, 5 milisekundžių, t. y. kam lygi tikimybė $p=P(T<5)=F_T(5)$? Dydžio pasiskirstymo funkcijos nežinome, tačiau galime atlikti daug bandymų ir gauti daug duomenų. Pažymėkime X_1 atsitiktinį dydį, įgyjantį reikšmes 0 ir 1:

$$X_1=\begin{cases}0, & \text{pirmajame bandyme ryšiui užmegzti prireikė mažiau nei 5 ms,}\\ 1, & \text{pirmajame bandyme ryšys užmegztas greičiau nei per 5 ms.}\end{cases}$$

Su kitais bandymais susiekime analogiškai apibrėžtus atsitiktinius dydžius $X_2, X_3, \ldots$. Galime padaryti prielaidą, kad dydžiai nepriklausomi. Be to,

$$P(X_j=1)=P(T<5)=p,\quad P(X_j=0)=1-p, \mathbf{E}X_j=p.$$

Iš didžiųjų skaičių dėsnio išplaukia, kad su didelėmis n reikšmėmis

$$\frac{X_1+X_2+\ldots+X_n}{n}\approx p.$$

Taigi naudodamiesi didžiųjų skaičių dėsniu įvertinome nežinomą su stebimu atsitiktiniu dydžiu susijusio įvykio tikimybę.