1.11 Teoreme limită

Teoremele limită constituie rezultatele teoretice cele mai importante ale teoriei probabilităţilor. Dintre acestea, cel mai importante sunt:

- legile numerelor mari (dau condiţii în care media aritmetică a unui şir de variabile aleatoare converge într-un anumit sens la o anumită constantă - valoarea medie a mediei aritmetice respective);
- teoreme limită centrale (dau condiţii în care suma unui număr de variabile aleatoare, corespunzătpr normată, converge la o variabilă aleatoare normală).

Vom demonstra mai întâi două inegalităţi:

Propoziţia 1.11.1 (Inegalitatea lui Markov) *Dacă X este o variabilă aleatoare ce ia numai valori ne-negative (adică $X \geq 0$), atunci pentru orice $a > 0$ avem:*

$$P(X \geq a) \leq \frac{M(X)}{a}.$$

Demonstraţie. Vom face demonstraţia în cazul unei variabile aleatoare X continue, având densitatea $f(x)$.

Dacă $M(X) = +\infty$, inegalitatea este evident adevărată, şi deci putem presupune că media $M(X)$ este finită. Avem

$$\begin{aligned} M(X) = & \int_{-\infty}^{+\infty} x f(x)\, dx \overset{X \geq 0}{=} \int_{0}^{+\infty} x f(x)\, dx \\ \overset{a>0}{\geq} & \int_{a}^{+\infty} x f(x)\, dx \geq \int_{a}^{+\infty} a f(x)\, dx \\ = & \; aP(X \geq a), \end{aligned}$$

de unde rezultă (deoarece $a > 0$)

$$P(X \geq a) \leq \frac{M(X)}{a}.$$

■

Propoziţia 1.11.2 (Inegalitatea lui Cebâşev) *Dacă X este o variabilă aleatoare cu medie $M(X) = \mu$ şi dispersie $D^2(X) = \sigma^2$ finite, atunci pentru orice $a > 0$ avem:*

$$P(|X - \mu| \geq a) \leq \frac{\sigma^2}{a^2}.$$

Demonstraţie. Considerăm variabila aleatoare ne-negativă $Y = (X-\mu)^2$. Din inegalitatea lui Markov, obţinem:

$$\begin{aligned} P(|X-\mu| \geq a) &= P\left((X-\mu)^2 \geq a^2\right) = P\left(Y \geq a^2\right) \\ &\leq \frac{M(Y)}{a^2} = \frac{M\left((X-\mu)^2\right)}{a^2} = \frac{D^2(X)}{a^2} \\ &= \frac{\sigma^2}{a^2}. \end{aligned}$$

■

Observaţia 1.11.3 *Inegalităţile anterioare sunt importante deoarece ele dau margini superioare pentru anumite probabilităţi, atunci când numai media/dispersia sunt cunoscute, ca]n exemplul următor:*

Exemplul 1.11.4 *Dacă X este o variabilă aleatoare cu medie $M(X) = 2$ şi dispersie $D^2(X) = 9$, atunci din inegalitatea lui Cebâşev (pentru $a = 5$, spre exemplu) avem:*

$$P(|X-2| > 5) \leq \frac{D^2(X)}{5^2} = \frac{9}{25} = 0.36.$$

Dacă în plus $X \geq 0$, atunci din inegalitatea lui Markov (pentru $a = 7$, spre exemplu) avem:

$$P(X > 7) \leq \frac{M(X)}{7} = \frac{2}{7} \approx 0.285\ldots$$

Observaţia 1.11.5 *Din ingalitatea lui Cebâşev rezultă că dacă dispersia unei variabile aleatoare X este $D^2(X) = 0$, atunci*

$$P(|X-\mu| > a) = 0,$$

oricare ar fi $a > 0$, şi deci

$$P(|X-\mu| > 0) = P\left(\bigcap_{n \geq 1} |X-\mu| > \frac{1}{n}\right) = \lim_{n\to\infty} P\left(|X-\mu| > \frac{1}{n}\right) = 0,$$

adică

$$P(X = \mu) = 1,$$

ceea ce arată că singurele variabile aleatoare având dispersie 0 sunt cele constante.

Teorema 1.11.6 (Legea slabă a numerelor mari) *Fie $X_1, X_2, \ldots$ un şir de variabile aleatoare independente şi identic distribuite, având medie $M(X_i) = \mu$ – finită. Atunci oricare ar fi $\varepsilon > 0$ avem*

$$P\left(\left|\frac{X_1 + \ldots + X_n}{n} - \mu\right| > \varepsilon\right) \underset{n\to\infty}{\to} 0$$

(adică şirul $\frac{X_1+\ldots+X_n}{n}$ converge în probabilitate către μ).

Demonstraţie. Vom face demonstraţia în ipoteza suplimentară că există dispersia $D^2(X_i) = \sigma^2 < +\infty$.

Pentru un $n \geq 1$ arbitrar fixat, considerând variabila aleatoare

$$Y = \frac{X_1 + \ldots + X_n}{n},$$

avem

$$M(Y) = M\left(\frac{X_1 + \ldots + X_n}{n}\right) = \frac{1}{n}(M(X_1) + \ldots + M(X_n)) = \frac{n\mu}{n} = \mu$$

şi

$$\begin{aligned} D^2(Y) &= D^2\left(\frac{X_1 + \ldots + X_n}{n}\right) = \frac{1}{n^2} D^2(X_1 + \ldots + X_n) \\ &\underset{\text{independente}}{\overset{X_1, X_2, \ldots}{=}} \frac{1}{n^2}\left(D^2(X_1) + \ldots + D^2(X_n)\right) \\ &= \frac{n\sigma^2}{n^2} \\ &= \frac{\sigma^2}{n}. \end{aligned}$$

Din inegalitatea Cebâşev avem

$$P\left(\left|\frac{X_1 + \ldots + X_n}{n} - \mu\right| > \varepsilon\right) = P(|Y - \mu| > \varepsilon) \leq \frac{D^2(Y)}{\varepsilon^2} = \frac{\sigma^2/n}{\varepsilon^2} = \frac{\sigma^2}{n\varepsilon^2} \underset{n \to \infty}{\to} 0.$$

■

Teorema limită centrală este un rezultat fundamental, cu numeroase aplicaţii practice, ce afirmă că suma unui număr mare de variabile aleatoare independente este apoximativ o variabilă aleatoare normal distribuită, fapt ce explică de ce frcvenţele relative ale multor populaţii statistice au o formă de clopot, adică sunt aproximativ normal distribuite.

Reamintim că o variabilă aleatoare normală $\mathcal{N}(\mu, \sigma^2)$ cu medie μ şi dispersie σ^2 este o variabilă aleatoare continuă având densitatea

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}.$$

Teorema 1.11.7 (Teorema limită centrală) *Fie $X_1, X_2, \ldots$ un şir de variabile aleatoare idenpendente şi identic distribuite, cu medie $M(X_i) = \mu$ şi dispersie $D^2(X_i) = \sigma^2$ finite. Atunci distribuţia variabilei aleatoare $\frac{X_1 + \ldots + X_n - n\mu}{\sigma\sqrt{n}}$ tinde (pentru $n \to \infty$) la distribuţia normală standard $\mathcal{N}(0,1)$, adică:*

$$P\left(\frac{X_1 + \ldots + X_n - n\mu}{\sigma\sqrt{n}} < a\right) \underset{n \to \infty}{\to} \int_{-\infty}^{a} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx = \Phi(x).$$

Pentru demonstraţie vom folosi următoarea:

Lema 1.11.8 *Fie $Z_1, Z_2, \ldots$ un şir de variabile aleatoare având funcţiile de distribuţie F_{Z_n} şi funcţiile generatoare de moment φ_{Z_n}, şi fie Z o variabilă aleatoare cu funcţia de distribuţie F_Z şi funcţia generatoare de moment φ_Z. Atunci*

$$\varphi_{Z_n} \underset{n\to\infty}{\to} \varphi_Z \Longrightarrow F_{Z_n} \underset{n\to\infty}{\to} F_Z,$$

pentru orice valoare $t \in \mathbb{R}$ pentru care funcţia de distribuţie F_Z este continuă în acest punct.

Folosind faptul că funcţia generatoare de moment a unei variabile aleatoare standard $Z \in \mathcal{N}(0,1)$ este $\varphi_Z(t) = e^{t^2/2}$, în particular avem următoarea:

Lema 1.11.9 *Fie $Z_1, Z_2, \ldots$ un şir de variabile aleatoare având funcţiile de distribuţie F_{Z_n} şi funcţiile generatoare de moment φ_{Z_n}. Dacă $\varphi_{Z_n} \underset{n\to\infty}{\to} e^{t^2/2}$, atunci*

$$F_{Z_n} \underset{n\to\infty}{\to} \Phi_Z = \int_{-\infty}^{a} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx,$$

oricare ar fi $t \in \mathbb{R}$.

Demonstraţia teoremei limită centrale. Vom face demonstraţia în ipoteza suplimentară că funcţiile generatoare de moment $\varphi_{X_n}(t) \overset{\text{not}}{=} \varphi_X(t)$ a variabilelor aleatoare X_n există şi sunt finite oricare ar fi $t \in \mathbb{R}$.

Considerăm mai întâi cazul în care $\mu = 0$ şi $\sigma^2 = 1$.

Avem în acest caz

$$\varphi_{\frac{X_i}{\sqrt{n}}}(t) = M\left(e^{\frac{tX_i}{\sqrt{n}}}\right) = \varphi_X\left(\frac{t}{\sqrt{n}}\right), \qquad 1 \le i \le n,$$

şi cum variabilele aleatoare $X_1, \ldots, X_n$ sunt independente, avem

$$\begin{aligned}
\varphi_{\frac{X_1+\ldots+X_n}{\sqrt{n}}}(t) &= M\left(e^{t\frac{X_1+\ldots+X_n}{\sqrt{n}}}\right) = M\left(\prod_{i=1}^{n} e^{t\frac{X_i}{\sqrt{n}}}\right) \\
&= \prod_{i=1}^{n} M\left(e^{t\frac{X_i}{\sqrt{n}}}\right) = \prod_{i=1}^{n} \varphi_X\left(\frac{t}{\sqrt{n}}\right) \\
&= \left(\varphi_X\left(\frac{t}{\sqrt{n}}\right)\right)^n.
\end{aligned}$$

Dezvoltând în serie Taylor funcţia $\varphi_X(t)$ în jurul punctului $t = 0$, avem:

$$\varphi_X(t) = \varphi_X(0) + \frac{\varphi_X'(0)}{1!} t + \frac{\varphi_X''(0)}{2!} t + R(t),$$

unde $\lim_{t\to\infty} \frac{R(t)}{t^2} = 0$.

Avem deci:

$$\begin{aligned}
\varphi_X(0) &= 1+\frac{M(X)}{1!}t+\frac{M\left(X^2\right)}{2!}t^2+R(t)= \\
&= 1+\frac{\mu}{1!}t+\frac{\sigma^2+\mu^2}{2!}t^2+R(t) \\
&= 1+\frac{0}{1!}t+\frac{1+0^2}{2!}t^2+R(t) \\
&= 1+\frac{t^2}{2}+R(t),
\end{aligned}$$

şi deci

$$\left(\varphi_X\left(\frac{t}{\sqrt{n}}\right)\right)^n=\left(1+\frac{t^2}{2n}+R\left(\frac{t}{\sqrt{n}}\right)\right)^n,$$

de unde obţinem

$$\begin{aligned}
\lim_{n\to\infty}\varphi_{\frac{X_1+\ldots+X_n}{\sqrt{n}}}(t) &= \lim_{n\to\infty}\left(\varphi_X\left(\frac{t}{\sqrt{n}}\right)\right)^n \\
&= \lim_{n\to\infty}\left(1+\frac{t^2}{2n}+R\left(\frac{t}{\sqrt{n}}\right)\right)^n \\
&= e^{\lim_{n\to\infty} n\left(\frac{t^2}{2n}+R\left(\frac{t}{\sqrt{n}}\right)\right)} \\
&= e^{\frac{t^2}{2}+\lim_{n\to\infty} t^2\frac{R\left(\frac{t}{\sqrt{n}}\right)}{\left(\frac{t}{\sqrt{n}}\right)^2}} \\
&= e^{\frac{t^2}{2}}.
\end{aligned}$$

Conform lemei anterioare rezultă că

$$F_{\frac{X_1+\ldots+X_n}{\sqrt{n}}}(t)\to\Phi_Z=\int_{-\infty}^{a}\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}dx,$$

oricare ar fi $t\in\mathbb{R}$, încheiând demonstraţia în cazul particular $\mu=0$ şi $\sigma^2=1$.

Pentru cazul general, considerăm variabilele aleatoare $\tilde{X}_i=\frac{X_i-\mu}{\sigma}$, $i\geq 1$, şi cum variabilele aleatoare $X_1,X_2,\ldots$ sunt independente şi identic distribuite, cu medie $M(X_i)=\mu$ şi dispersie $D^2(X_i)=\sigma^2$, rezultă că variabilele aleatoare $\tilde{X}_1,\tilde{X}_2,\ldots$ sunt de asemenea independente şi identic distribuite, cu medie $M\left(\tilde{X}_i\right)=0$ şi dispersie $D^2\left(\tilde{X}_i\right)=0$.

Coform cazului anterior demonstrat rezultă că

$$F_{\frac{\tilde{X}_1+\ldots+\tilde{X}_n}{\sqrt{n}}}(t)\to\Phi_Z=\int_{-\infty}^{a}\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}dx,\qquad t\in\mathbb{R},$$

sau echivalent

$$F_{\frac{X_1+\ldots+X_n-n\mu}{\sigma\sqrt{n}}}(t)\to\Phi_Z=\int_{-\infty}^{a}\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}dx,\qquad t\in\mathbb{R},$$

încheiând demonstraţia. ■

Observaţia 1.11.10 *Există o variantă mai generală a legii numerelor mari (legea tare a numerelor mari). Ea afirmă următoarea:*

Teorema 1.11.11 (Legea tare a numerelor mari) *Dacă $X_1, X_2, \ldots$ sunt variabile aleatoare independente şi identic distribuite, cu medie $M(X_i) = \mu$ finită, atunci*

$$P\left(\lim_{n\to\infty} \frac{X_1 + \ldots X_n}{n} = \mu\right) = 1.$$

În particular, considerâm un eveniment E ce poate apare în urma efectuării unui experiment cu probabilitate $P(E) = p$, şi considerăm n repetiţii ale acestui experiment. Definind şirul de variabile aleatoare

$$X_i = \begin{cases} 1, & \text{dacă ev. } E \text{ a apărut în experimentul } i \\ 0, & \text{în caz contrar} \end{cases}, \qquad i \geq 1,$$

din legea tare a numerelor mari obţinem:

$$\begin{aligned} 1 &= P\left(\lim_{n\to\infty} \frac{X_1 + \ldots + X_n}{n} = M(X_1)\right) \\ &= P\left(\lim_{n\to\infty} \frac{\#\text{ apariţii a lui } E}{\#\text{ total de încercări}} = p\right) \\ &= P\left(\lim_{n\to\infty} \text{frecvenţa relativă a lui } E = p\right), \end{aligned}$$

adică frecvenţa relativă de apariţie a evenimentului E tinde când $n \to \infty$ aproape sigur către probabilitatea p de apariţie a evenimentului E.

1.11.1 Exerciţii

1. O fabrică produce în medie 50 produse pe săptămână.

   (a) Ce se poate spune despre probabilitatea $P(X > 75)$?

   (b) Dacă dispersia numărului de piese produse este $\sigma^2 = 25$, ce se poate spune despre probabilitatea $P(40 < X < 60)$?

   Indicaţie: se folosesc inegalităţile Markov şi Cebâşev.

2. Se consideră o variabilă aleatoare X uniform distribuită pe intervalul $(0, 10)$.

   (a) Să se calculeze media μ a variabilei aleatoare X

   (b) Să se calculeze direct probabilitatea $P(|X - \mu| > 4)$

   (c) Să se estimeze probabilitatea $P(|X - \mu| > 4)$ folosind inegalitatea lui Cebâşev.

3. Considerăm $X_1, \ldots, X_{20}$ variabile aleatoare independente Poisson cu parametrul $\lambda = 1$.

(a) Să se utilizeze inegalitatea Markov pentru a estima probabilitatea $P \sum_{i=1}^{20} X_i > 15$.

(b) Să se utilizeze Teoreme limită centrală pentru a estima probabilitatea $P\left(\sum_{i=1}^{20} X_i > 15\right)$.

4. Din experienţă, un profesor ştie că nota obţinută la examen de un student este o variabilă aleatoare cu media 75 (din 100 puncte posibile).

   (a) Ce se poate spune despre probabilitatea $P(X > 85)$?

   (b) Dacă dispersia notelor obţinute este $\sigma^2 = 25$, ce se poate spune despre probabilitatea $P(65 < X < 85)$?

5. Considerăm o variabilă aleatoare X având media $\mu = 20$ şi dispersia $\sigma^2 = 20$. Ce se poate spune despre probabilitatea $P(0 \leq X \leq 40)$?

6. Numărul de candidaţi X ce se înscriu la un concurs este o variabilă aleatoare Poisson cu parametrul $\lambda = 100$. S-a hotărât ca daca se înscriu mai mult de 120 de candidaţi, să se organizeze două examene, altfel să se organizeze un singur examen. Care este probabilitatea să fie nevoie să se organizeze două examene?

   Indicaţie: Pentru calculul probabilităţii $P(X > 120)$ se va folosi aproximarea variabilei aleatoare Poisson prin variabila aleatoare binomială cu parametrii n şi p, astfel încât $np = \lambda = 100$ (spre exemplu $n = 200$ şi $p = \frac{1}{2}$), şi deci $X = X_1 + \ldots + X_{200}$, unde $X_1, \ldots, X_{200}$ sunt varibile aleatoare Bernoulli cu parametrul $p = \frac{1}{2}$. Se foloseşte acum Teorema limită centrală pentru aproximarea probabilităţii cerute. Răspuns: $0.0228\ldots$

7. Se aruncă 10 zaruri. Să se estimeze probabilitatea ca suma punctelor obţinute să fie între 30 şi 40.

   Indicaţie: se foloseşte Teorema limită centrală. Răspuns: $0.65\ldots$

8. Fie $X_1, \ldots, X_{10}$ variabile aleatoare independente şi uniform distribuite pe intervalul $(0, 1)$. Să se aproximeze probabilitatea $P\left(\sum_{i=1}^{10} X_i > 6\right)$.

   Indicaţie: se foloseşte Teorema limită centrală. Răspuns: $0.16\ldots$

9. Se aruncă de 100 de ori un zar. Să se estimeze probabilitatea $P\left(\prod_{i=1}^{100} X_i \leq a^{100}\right)$, unde $a > 0$ este un număr arbitrar fixat (caz particular $a = 2$).

   Indicaţie: probabilitatea dată se poate scrie

$$P\left(\prod_{i=1}^{100} X_i \leq a^{100}\right) = P\left(\sum_{i=1}^{100} \ln X_i \leq 100 \ln a\right),$$

   şi se foloseşte Teorema limită centrală.