Universitatea Tehnică din Cluj-Napoca
Controlul Avansat al Proceselor

Matematici Avansate
Decembrie 2023

LEGEA NUMERELOR MARI

CEOCA OVIDIU

Cuprins

1. Introducere

Legea numerelor mari este o teoremă fundamentală din teoria probabilităților care descrie rezultatele unui experiment repetat de mai multe ori (numărul repetiților tinde spre ∞). Conform acestei legi, rezultatul mediu obținut se apropie tot mai mult de valoarea așteptată (o constantă), cu cât experimentul se repetă de mai multe ori. Aceasta se explică prin faptul că abaterile întâmplătoare într-un sens sau altul se compensează reciproc.

Această lege a fost formulată de **Jakob Bernoulli** și a pus bazele teoriei probabilităților ca știință.

1.1. **Exemplu.** Să presupunem următorul experiment: aruncați un zar comun. Acum să luăm în considerare evenimentul în care obținem numărul 1. După cum știm, probabilitatea ca numărul 1 să apară este 1/6 (zarul are 6 fețe, una dintre ele este una).

Ce ne spune legea numerelor mari? Ne spune că pe măsură ce creștem numărul de repetări ale experimentului nostru (facem mai multe aruncări ale zarului), frecvența cu care evenimentul se va repeta (obține 1) se va apropia de o constantă, care va avea o valoare egală. valoarea la probabilitatea sa (1/6 sau 16,66%).

Posibil, în primele 10 sau 20 de lansări, frecvența cu care obținem 1 nu va fi 16%, ci un alt procent precum 5% sau 30%. Dar pe măsură ce facem din ce în ce mai multe tonuri (să zicem 10.000), frecvența la care apare 1 va fi foarte aproape de 16,66%.

1.2. **Interpretarea greșită a legii.** Mulți oameni interpretează greșit legea numerelor mari, crezând că un eveniment va tinde să depășească pe altul. Astfel, de exemplu, ei cred că, deoarece probabilitatea ca numărul 1 să se arunce pe aruncarea unui zar ar trebui să fie aproape de 1/6, atunci când numărul 1 nu apare la primele 2 sau 5 aruncări, este foarte probabil că va apărea În următorul. Acest lucru nu este adevărat, deoarece legea numerelor mari se aplică doar pentru multe repetări, așa că putem petrece toată ziua aruncând un zar și nu ajungem la frecvența 1/6.

Lansarea unui zar este un eveniment independent și, prin urmare, atunci când apare un anumit număr, acest rezultat nu afectează următoarea aruncare. Numai după mii de repetări vom putea verifica dacă legea numerelor mari există și că frecvența relativă de obținere a unui număr (în exemplul nostru 1) va fi 1/6.

Interpretarea greșită a teoriei poate determina oamenii (în special jucătorii de noroc) să piardă bani și timp.

2. Inegalitatea Cebîșev

Înainte de a discuta de Legea numerelor mari, trebuie prima dată amintită o inegalitate foarte importantă ce poartă denumirea de **Inegalitate Cebîșev**.

2.1. **Inegalitatea Cebîșev.** Acest tip de inegalitate este valabilă atât pentru variabile discrete, cât și pentru variabile continue, însă pentru simplificare se va studia doar inegalitatea pentru valori discrete.

Teorema 2.1 (Inegalitatea Cebîșev)**.** *Fie X o variabilă aleatoare discretă, iar valoarea sa estimată fiind $\mu = E(X)$, și fie ϵ orice număr pozitiv cu proprietatea că $\epsilon > 0$. În acest caz se poate afirma faptul că*

$$P(|X-\mu| \geq \epsilon) \leq \frac{V(X)}{\epsilon^2} \tag{2.1}$$

sau, echivalent,

$$P(|X-\mu| < \epsilon) \geq 1 - \frac{V(X)}{\epsilon^2} \tag{2.2}$$

În caz particular, $\mu = E(X)$, $V(X) = \sigma^2 < \infty$ și $\epsilon = k\sigma$, atunci prima inegalitate a lui Cebîșev (2.1) devine

$$P(|X-\mu| \geq k\sigma) \leq \frac{1}{k^2} \tag{2.3}$$

sau, echivalent,

$$P(|X-\mu| < k\sigma) \geq 1 - \frac{1}{k^2} \tag{2.4}$$

pentru orice $k > 0$

Inegalitățile lui Markov și a lui Cebîșev au loc pentru orice tip de variablă aleatoare (**v.a.**). În schimb, estimările furnizate de aceste sunt destul de grosiere. Inegalitatea lui Cebîșev este foarte utilă pentru a demonstra convergența în probabilitatea unui șir de v.a.

Proof. Fie $m(x)$ funcția de distribuție a lui X. Atunci, probabilitatea ca X să difere de μ cu cel puțin ϵ este dată de relația

$$P(|X-\mu| \geq \epsilon) = \sum_{|x-\mu|\geq\epsilon} m(x) \tag{2.5}$$

Se cunoaște faptul că

$$V(X) = \sum_{x} (x-\mu)^2 m(x) \tag{2.6}$$

iar acest lucru este evident cel puțin la fel de mare ca

$$\sum_{|x-\mu|\geq\epsilon} (x-\mu)^2 m(x) \tag{2.7}$$

deoarece toate sumele sunt pozitive, iar intervalul sumei a fost restricționat în cea de-a doua sumă. Cu toate acestea, această ultimă sumă este cel puțin

$$\begin{aligned} &\sum_{|x-\mu|\geq\epsilon} \epsilon^2 m(x) \\ &= \epsilon^2 \sum_{|x-\mu|\geq\epsilon} m(x) \\ &= \epsilon^2 P(|X-\mu| \geq \epsilon) \end{aligned} \tag{2.8}$$

Astfel, dacă X este orice variabilă aleatoare discretă și ϵ orice număr pozitiv, se poate afirma faptul că

$$P(|X-\mu| \geq \epsilon) \leq \frac{V(X)}{\epsilon^2} \tag{2.9}$$

□

3. Legea numerelor mari

3.1. **Legea numerelor mari.** În următoarea secțiune, se vor folosi următoarele notații. Dacă $(X_k)_{k\in\mathbb{N}^*}$ este un șir de v.a., atunci

$$S_n = \sum_{k=1}^{n} X_k \tag{3.1}$$

și

$$\overline{X}_n = \frac{S_n}{n} \tag{3.2}$$

unde $n \in \mathbb{N}^*$

Definitia 3.1. Dacă un șir $(X_k)_{k\in\mathbb{N}^*}$ de v.a. satisface

$$\frac{S_n}{n} - E\left(\frac{S_n}{n}\right) \xrightarrow{\mathbb{P}} 0 \tag{3.3}$$

sau, echivalent,

$$\overline{X}_n - E(\overline{X}_n) \xrightarrow{\mathbb{P}} 0 \tag{3.4}$$

pentru orice $n \to \infty$. Sau, echivalent pentru orice $\epsilon > 0$,

$$\lim_{n\to\infty} \mathbb{P}\left(\left|\frac{\sum_{k=1}^{n} X_k}{n} - E\left(\frac{\sum_{k=1}^{n} X_k}{n}\right)\right| > \epsilon\right) = 0 \tag{3.5}$$

Atunci spunem că șirul dat satisface **legea slabă a numerelor mari (LSNM)**.

Teorema 3.2. *Fie $(X_k)_{k\in\mathbb{N}^*}$ un șir de v.a. integrabile. Atunci, șirul dat satisface **LSNM**, adică*

$$Z_n = \frac{S_n}{n} - E\left(\frac{S_n}{n}\right) \xrightarrow{\mathbb{P}} 0 \tag{3.6}$$

pentru $n \to \infty$ dacă și numai dacă

$$\lim_{n\to\infty} E\left(\frac{Z_n^2}{1+Z_n^2}\right) = 0 \tag{3.7}$$

Teorema 3.3. *Fie $(X_k)_{k\in\mathbb{N}^*}$ un șir de v.a. de pătrat integrabil astfel încât*

$$\lim_{n\to\infty} V\left(\frac{S_n}{n}\right) = 0 \tag{3.8}$$

*Atunci șirul $(X_k)_{k\in\mathbb{N}^*}$ satisface **LSNM**.*

Proof. Deoarece $E(Z_n)$=0, obținem, folosind

$$V(X - E(X)) = E[(X - E(X)) - E(X - E(X))]^2 = V(X) \tag{3.9}$$

unde v.a. $X - E(X)$ are media zero deci dispersia v.a. centrate coincide cu dispersia v.a., ecuația următoare

$$0 \le E\left(\frac{Z_n^2}{1+Z_n^2}\right) \le E(Z_n^2) = D^2(Z_n) = D^2\left(\frac{S_n}{n}\right) \tag{3.10}$$

□

În cazul în care v.a. sunt independente două câte două. obținem, ținând cont de faptul că daca v.a. $(X_i)_{i=\overline{1,n}}$ admit dispersie, atunci, dacă v.a. sunt necorelate două câte două, atunci dispersia sumei este suma dispersiilor,

$$(X_i)_{i=\overline{1,n}} necorelate \Longrightarrow D^2\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} D^2(X_i) \tag{3.11}$$

următoarea teoremă.

Teorema 3.4. *Fie $(X_k)_{k\in\mathbb{N}^*}$ un șir de v.a. de pătrat integrabil și independente două câte două astfel încât*

$$\lim_{n\to\infty}\frac{1}{n^2}\sum_{k=1}^{n}D^2(X_k)=0 \tag{3.12}$$

*Atunci, șirul $(X_k)_{k\in\mathbb{N}^*}$ satisface **LSNM**.*

Definitia 3.5. Fie $(X_k)_{k\in\mathbb{N}^*}$ un șir de v.a. de tip i.i.d.[1] cu

$$X_k \sim Bernoulli(p)=\mathfrak{B}(1,p) \tag{3.13}$$

unde $k\in\mathbb{N}^*$. Putem observa X_k ca v.a. care desemnează numărul de apariții ale unui eveniment A (numit **Succes**) la încercarea k, cu probabilitatea p de apariție a **Succesului**

$$X_k:\begin{pmatrix}0 & 1\\ q & p\end{pmatrix} \tag{3.14}$$

V.a.

$$f_n=\sum_{k=1}^{n}X_k \tag{3.15}$$

se numește **frecvența absolută de apariție a Succesului** în cele n probe și are drept valori, numărul de apariții ale *Succesului* în cele n observații.

Prin urmare, f_n urmează o distribuție de tip binomial cu $f_n\sim\mathfrak{B}(n,p)$.

V.a.

$$\frac{f_n}{n} \tag{3.16}$$

se numește **frecvența relativă de apariție a Succesului** sau echivalent

$$\frac{f_n}{n}\xrightarrow{\mathbb{P}}p \tag{3.17}$$

pentru $n\to\infty$

adică șirul frecvențelor relative de apariție a *Succesului* converge în probabilitate la p (care este probabilitatea, teoretică, de apariție a Succesului la o singură încercare).

Cu alte cuvinte, dacă f_n este frecvența absolută de apariție a unui eveniment A în n probe independente și p este probabilitatea de apariție a lui A (indiferent de probă), atunci frecvența relativă $\frac{f_n}{n}$ de apariție a evenimentului A în cele n probe tinde în probabilitate la p (3.17).

[1]Dacă v.a. $(X_i)_{i=\overline{1,n}}$ sunt independente și identic distribuite atunci se va scrie prescurtat **i.i.d**

Conform exemplului precedent, se obține că limita 3.17 se poate scrie sub forma:

Teorema 3.6 (Teorema lui Bernoulli)**.** *Dacă* $(A_k)_{k\in\mathbb{N}^*}$ *este un șir de evenimente independente astfel încât* $\mathbb{P}(A_k)=p$ *pentru orice* $k\in\mathbb{N}^*$ *atunci are loc:*

$$\frac{1}{n}(A_1+A_2+A_3+\cdots+A_n)\xrightarrow{\mathbb{P}} p, \quad pentru \quad n\to\infty \tag{3.18}$$

Se observă că dacă șirul $(X_k)_{k\in\mathbb{N}^*}$ satisface *LSNM*, adică

$$\frac{S_n}{n}-E\left(\frac{S_n}{n}\right)\xrightarrow{\mathbb{P}} 0, \quad pentru \quad n\to\infty \tag{3.19}$$

și dacă, în plus, are loc

$$E\left(\frac{S_n}{n}\right)\to\mu, \quad pentru \quad n\to\infty \tag{3.20}$$

atunci are loc și

$$\frac{S_n}{n}\xrightarrow{\mathbb{P}}\mu, \quad pentru \quad n\to\infty \tag{3.21}$$

Într-adevăr, avem următoarea convergență a șirurilor numerice (văzute ca v.a. constante)

$$E\left(\frac{S_n}{n}\right)\xrightarrow{F}\mu, \quad pentru \quad n\to\infty \tag{3.22}$$

În toate rezultatele de mai sus s-a cerut ca v.a. să fie pătrat integrabil. Se poate arăta că șirul $(X_k)_{k\in\mathbb{N}^*}$ satisface **LSNM** presupunând doar integrabilitatea v.a.. Demonstrația se face prin trunchierea v.a., luând, mai întâi, $\tilde{X}_k=X_{k\{|X_k|\le k\}}$

Teorema 3.7. *Fie* $(X_k)_{k\in\mathbb{N}^*}$ *un șir de v.a. de tip i.i.d. (independența este două câte două) astfel încât*

$$E(X_k)=\mu<\infty, \quad pentru \quad orice \quad k\in\mathbb{N}^* \tag{3.23}$$

Atunci șirul $(X_k)_{k\in\mathbb{N}^*}$ *satisface LSNM, mai precis se obține*

$$\frac{S_n}{n}\xrightarrow{\mathbb{P}}\mu, \quad pentru \quad n\to\infty \tag{3.24}$$

În continuare este important ca relația 3.3 să aibă loc aproape sigur.

Definitia 3.8. Dacă un șir $(X_k)_{k\in\mathbb{N}^*}$ de v.a. satisface

$$\frac{S_n}{n}-E\left(\frac{S_n}{n}\right)\xrightarrow{a.s.} 0, \quad pentru \quad n\to\infty \tag{3.25}$$

sau, echivalent,

$$\overline{X}_n-E(\overline{X}_n)\xrightarrow{a.s.} 0, \quad pentru \quad n\to\infty \tag{3.26}$$

sau echivalent,

$$\mathbb{P}\left(\left\{\omega : \lim_{n\to\infty}\left(\frac{\sum\limits_{k=1}^{n} X_k(\omega)}{n} - E\left(\frac{\sum\limits_{k=1}^{n} X_k}{n}\right)\right) = 0\right\}\right) = 1 \tag{3.27}$$

Atunci spunem că șirul dat satisface **legea tare a numerelor mari (LTNM)**.

4. Exerciții rezolvate

4.1. **Exercițiul 1.** Care este probabilitatea ca la 1000 de aruncări ale unei monede, frecvența absolută de apariție a stemei să fie între 400 și 650?

4.1.1. *Rezolvare.* Să definim v.a. independente X_k care iau drept valori, numărul de apariții ale stemei la aruncarea k a monedei, deci

$$X_k : \begin{pmatrix} 0 & 1 \\ \frac{1}{2} & \frac{1}{2} \end{pmatrix}, \quad k \in \mathbb{N}^* \tag{4.1}$$

Atunci v.a.

$$S_{10^3} = \sum_{k=1}^{10^3} X_k \tag{4.2}$$

are drept valori numărul de apariții ale stemei la aruncarea de 1000 de ori a unei monede, adică frecvența absolută de apariție a stemei la $n = 10^3$ aruncări. Deoarece $X_k \sim \mathfrak{B}(1,p)$, $k \in \mathbb{N}^*$, obținem $S_{10^3} \sim \mathfrak{B}(10^3, p)$. Pentru calculul mediei și dispersiei putem folosi folosi formulele:

$$E(X) = n\cdot \quad și \quad D^2(X) = n \cdot pq \tag{4.3}$$

Dar se poate calcula și direct cu ajutorul

$$E(S_{10^3}) = \sum_{k=1}^{10^3} E(X_k) = 10^3 \cdot E(X_1) = 10^3 \cdot \frac{1}{2} \tag{4.4}$$

iar dispersia se poate obține folosind independența și dispersia v.a. X_k:

$$\begin{aligned} D^2(S_{10^3}^2) &= D^2\left(\sum_{k=1}^{10^3} X_k\right) \\ &= \sum_{k=1}^{10^3} D^2(X_k) \\ &= 10^3 \cdot D^2(X_1) \\ &= 10^3 \cdot \frac{1}{2} \cdot \frac{1}{2} \\ &= 250 \end{aligned} \tag{4.5}$$

Dacă luam $\epsilon = 10^2$ în inegalitatea lui Cebîșev obținem

$$\begin{aligned}\mathbb{P}(400 < S_{10^3} < 650) &> \mathbb{P}(400 < S_{10^3} < 600) \\ &= \mathbb{P}(|S_{10^3} - 500| < 10^2) \\ &\geq 1 - \frac{250}{10^4} \\ &= \frac{39}{40}\end{aligned} \tag{4.6}$$

References

[1] Lucian Maticiuc (2022) *Teoria Probabilităților -Teorie și Aplicații-*, Iași.
[2] Andrei-George Oprina, Emil Simion (Februarie 2017) *Elemente de Teoria Probabilităților*, București.
[3] Anamaria Popescu (2016) *Teoria Probabilităților Caiet de Seminar*, Petroșani.
[4] Grinstead, Charles Miller, and James Laurie Snell (4 Iulie 2006) *Grinstead and Snell's introduction to probability.*
[5] Radu Trîmbițaș (Decembrie 2012) *Legea numerelor mari și legi limită*, UBB.