Probabilităţi şi Statistică - Curs 7

30 martie 2026

Table of contents

Teorema limită centrală

Theorem 1.1

(Teorema limită centrală, Lindeberg-Lévy) Fie $(X_n)_{n\geqslant 1}$ un şir de variabile aleatoare independente şi identic distribuite cu media μ şi dispersia σ^2. Atunci

$$\frac{\frac{1}{n}\sum_{i=1}^{n} X_i - \mu}{\frac{\sigma}{\sqrt{n}}} \to N(0,1) \text{ sau}$$

$$\lim_{n\to\infty} P\left(\frac{\sum_{i=1}^{n} X_i - n\mu}{\sigma\sqrt{n}} \leqslant a\right) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{a} \exp(-t^2/2)dt, \forall a \in \mathbb{R}.$$

Teorema limită centrală

- Teorema limită centrală permite estimarea unor probabilităţi asociate sumelor de variabile (independente şi identic distribuite).
- Pe de altă parte, teorema explică de ce atât de multe procese (din ştiinţele sociale, biologie, psihologie etc) urmează o lege normală.
- În esenţa ei teorema limită centrală spune că, pentru eşantioane suficient de mari ($n \geqslant 30$), variabila

$$\frac{\sum_{i=1}^{n} X_i - n\mu}{\sigma\sqrt{n}}$$

  urmează o lege normală standard, $N(0,1)$.
- Teorema limită centrală are loc chiar şi pentru variabile dependente, dacă au corelaţia foarte scăzută.

Aproximarea normală a distribuţiei binomiale

- Fie X_n un şir de variabile Bernoulli(p) independente.
- $X = \sum_{i=1}^{n} X_i$ are o distribuţie binomială, $B(n, p)$.
- Folosind teorema limită centrală obţinem teorema de Moivre-Laplace care spune că, pentru n suficient de mare, variabila

$$Y = \frac{X - \mathbb{E}[X]}{\sqrt{Var[X]}} = \frac{X - np}{\sqrt{np(1-p)}}$$

este o variabilă normală standard $(N(0,1))$.
- Aproximarea este bună pentru $np(1-p) \geqslant 10$.

Aproximarea normală a distribuţiei binomiale

- Un alt mod de a vedea acest rezultat este următorul: când k este aproape de np

$$\binom{n}{k} p^k (1-p)^{n-k} \sim \frac{\exp -\frac{(k-np)^2}{2np(1-p)}}{\sqrt{2\pi np(1-p)}}.$$

- Considerăm următorul exemplu: fie X numărul de apariţii ale stemei în 40 de aruncări ale unei monede.
- Cât este $P(X = 20)$?

$$P(X = 20) = P(19.5 \leqslant X \leqslant 20.5) =$$

$$= P\left(\frac{19.5-20}{\sqrt{10}} \leqslant \frac{X-20}{\sqrt{10}} \leqslant \frac{20.5-20}{\sqrt{10}}\right) = P\left(-0.16 \leqslant \frac{X-20}{\sqrt{10}} \leqslant 0.16\right)$$

$$\sim \Phi(0.16) - \Phi(-0.16) = 0.1272,$$

unde $\Phi(\cdot)$ este funcţia de repartiţie a variabilei $N(0,1)$.

Corecţia continuă

- Corecţia continuă este o ajustare care se face ori de câte ori o distribuţie discretă este aproximată printr-una continuă.
- $P(X = 10) = P(9.5 \leqslant X \leqslant 10.5)$, $P(X > 15) = P(X \geqslant 15.5)$, $P(X < 13) = P(X \leqslant 12.5)$.

Generarea de numere aleatoare uniform distribuite

- Când vorbim despre numere aleatoare ne gândim cel mai adesea la numere aleatoare uniform distribuite.
- Există două tipuri de variabile aleatoare uniforme: discretă şi continuă.
- De exemplu, pentru a alege un număr întreg aleator uniform distribuit între 1 şi n (câteodată între 0 şi $(n-1)$) trebuie să generăm o valoare a unei variabile aleatoare discrete uniforme U_n.
- Pe de altă parte, dacă dorim să alegem un număr aleator uniform din intervalul $[0,1]$ trebuie să generăm o valoare a unei variabile continue uniforme $U_{[0,1]}$.
- În general, *a simula o anumită variabilă aleatoare înseamnă a genera valori care urmează acea distrubuţie*.

Generarea de numere aleatoare

- Aproape orice limbaj de programare conţine câte un generator de numere aleatoare unifome (discrete şi continue); noi vom utiliza generatoarele din R care acoperă şi alte distribuţii în afară de cele uniforme.
- Trecem în revistă comenzile R pentru distribuţiile discrete sau continue uzuale.
- Funcţiile care încep cu p, q, d şi r returnează funcţie de repartiţie (sau CDF - cumulative distribution function), inversa CDF, function de densitate de probabilitate (PDF), respectiv o valoare a unei variable aleatoare având distribuţia specificată.
- Pentru a genera doar valori discrete uniforme se poate utiliza funcţia $sample()$.

Generarea de numere aleatoare

Distribuţia	Comenzi			
Binomial	$pbinom()$	$qbinom()$	$dbinom()$	$rbinom()$
Geometric	$pgeom()$	$qgeom()$	$dgeom()$	$rgeom()$
Poisson	$ppois()$	$qpois()$	$dpois()$	$rpois()$
Uniform	$punif()$	$qunif()$	$dunif()$	$runif()$
Exponential	$pexp()$	$qexp()$	$dexp()$	$rexp()$
Normal	$pnorm()$	$qnorm()$	$dnorm()$	$rnorm()$
Student	$pt()$	$qt()$	$dt()$	$rt()$
Gamma	$pgamma()$	$qgamma()$	$dgamma()$	$rgamma()$

- Detalii se pot afla folosind $help(nume)$ în R sau Rstudio.

Generarea de numere aleatoare

- Pentru a simula o variabilă aleatoare discretă este nevoie de repartiţia ei.

$$X : \begin{pmatrix} x_1 & x_2 & \dots & x_k & \dots \\ p_1 & p_2 & \dots & p_k & \dots \end{pmatrix}$$

- X se poate simula astfel: generăm o valoare aleatoare uniformă (continuă) U şi returnăm x_i dacă $\sum_{j=1}^{i-1} p_j \leqslant U < \sum_{j=1}^{i} p_j$.

Aplicaţii ale LNM şi TLC

Exemplul 1. (*LNM - Problema acului lui Buffon*) Problema (formulată în 1733 şi rezolvată în 1777 de naturalistul şi matematicianul francez Comte de Buffon) cere să se determine probabilitatea ca un ac de lungime l să intersecteze o dreaptă când este aruncat pe o suprafaţă plană pe care sunt desenate (o infinitate de) drepte paralele aflate la distanţa $2d$.

Soluţie. Vom presupune că lungimea acului este mai mică decât distanţa dintre drepte (cea mai uşoară variantă de analizat); există două variabile care determină poziţia relativă a acului faţă de cea mai apropiată dreaptă: unghiul, x, pe care îl face acul cu direcţia liniilor şi distanţa de la mijlocul acului la această dreaptă, y. Acul va intersecta cea mai apropiată linie dacă şi numai $y \leqslant l/2 \sin x$, oricare ar fi $x \in [0, \pi]$.

Aplicaţii ale LNM şi TLC

Toate situaţiile posibile sunt complet descrise de perechile $(x, y) \in [0, \pi] \times [0, d]$, iar cazurile favorabile sunt perechile care dau puncte aflate sub graficul funcţiei $f : [0, \pi] \to \mathbb{R}$, $f(x) = l/2 \sin x$.
Astfel, probabilitatea este

$$\frac{\int_0^\pi f(x)\, dx}{\pi \cdot d} = \frac{1}{\pi d} \int_0^\pi \frac{l}{2} \sin x\, dx = \frac{l}{2\pi d} \left[-\cos x\right]_0^\pi = \frac{l}{\pi d}.$$

Pentru $l = d = 1$, adică atunci când acul are lungimea egală cu jumătate din distanţa dintre drepte, probabilitatea este $1/\pi$.

Aplicaţii ale LNM şi TLC

Introducem experienţa aleatoare care constă în a arunca acul şi notăm cu X variabila Bernoulli care are valoarea 1 dacă şi numai dacă acul intersectează o dreaptă; probabilitatea de succes şi media variabilei X au valoarea $1/\pi$.

Dacă repetăm în mod independent experimentul de n ori vom obţine un eşantion de dimensiune n, $(X_i)_{i=\overline{1,n}}$. Datorită Legii Numerelor Mari $\overline{x}_n \to 1/\pi$, astfel, pentru valori mari ale lui n,

$$\overline{x}_n = \frac{\text{numărul de succese}}{n} \approx \frac{1}{\pi}.$$

Acest tip de relaţie a fost utilizat pentru a obţine aproximări experimentale ale lui π. Mai mulţi "aruncători" de ace au efectuat deja acest experiment.

Aplicaţii ale LNM şi TLC

Exemplul 2. (*Verificarea LNM*) Considerăm o repartiţie X cu media μ şi dispersia σ^2 şi un şir de n variabile aleatoare independente şi identic distribuite cu X, X_i, $i = \overline{1,n}$. Legea numerelor mari spune că (într-un anumit sens, probabilistic) media de selecţie converge la medie:

$$\overline{x}_n \to \mu \text{ as } n \to \infty$$

Verificăm această lege utilizând o distribuţie Poisson cu diverşi parametri λ (pentru o astfel de distribuţie $\mu = \lambda$).

λ	2	3	4	6	8	12	15
$\overline{x}_n$	1.955	2.977	4.003	6.027	8.018	12.093	14.925

Se observă că mediile de selecţie obţinute ($n = 5000$) sunt foarte apropiate de mediile corespunzătoare. (Eşantioanele au fost obţinute folosind $rpois(n, \lambda)$.)

Aplicaţii ale LNM şi TLC

Dacă repetăm testul anterior cu distribuţia Gamma pentru diverse valori ale parametrilor (α, λ) (media este $\mu = \alpha/\lambda$) obţinem

α	2	2	3	4	6	6	6	12
λ	1.5	2	2	3	5	4	8	4
$\overline{x}_n$	1.361	1.009	1.489	1.345	1.204	1.501	0.752	2.973
μ	1.333	1.000	1.500	1.333	1.200	1.500	0.750	3.000

Mediile de selecţie obţinute ($n = 5000$) sunt foarte apropiate de mediile corespunzătoare. (Eşantioanele au fost obţinute folosind $rgamma(n, \alpha, \lambda)$.)

Aplicaţii ale LNM şi TLC

Examplul 3. (*TLC - de Moivre-Laplace*) Mărimea ideală a anului I la un colegiu particular este de 150 studenţi. Conducerea colegiului, ştiind din experienţă, că, în medie, doar 30% din elevii care trec examenul de admitere vor urma cursurile, aprobă 450 de locuri pentru admitere. Calculaţi probabilitatea ca cel puţin 151 de elevi admişi să rămână şi să urmeze cursurile colegiului.

Soluţie. Fie X numărul de elevi admişi care urmează cursurile colegiului; vom presupune că fiecare elev admis va urma independent cursurile colegiului. Atunci $X : B(450, 0.3)$ şi

$$P(X > 150) = P(X \geqslant 150.5) = P\left(\frac{X - np}{\sqrt{np(1-p)}} \geqslant \frac{150.5 - np}{\sqrt{np(1-p)}}\right) =$$

$$= P\left(\frac{X - 135}{\sqrt{81}} \geqslant \frac{15.5}{\sqrt{81}}\right) \approx P(Z \geqslant 1.722)$$

unde $Z : N(0,1)$. Astfel $P(X \geqslant 150) \approx 1 - pnorm(1.722) = 0.0425$.

Aplicaţii ale LNM şi TLC

Examplul 4. (*TLC*) O populaţie de muncitori are media greutăţii 167 şi deviaţia standard 27. Dacă se alege un eşantion de 36 muncitori, care este probabilitatea ca media de selecţie să fie cuprinsă între 163 şi 170?

Soluţie. Fie $\overline{x}_n$ media de selecţie, din TLC, $\dfrac{\overline{x}_n - \mu}{\sigma/\sqrt{n}}$ urmează cu aproximaţie o distribuţie normală standard, astfel

$$P(163 \leqslant \overline{x}_n \leqslant 170) = P\left(\frac{163-167}{4.5} \leqslant \frac{\overline{x}_n - 167}{4.5} \leqslant \frac{170-167}{4.5}\right) =$$

$$= P\left(-0.888 \leqslant \frac{\overline{x}_n - 167}{4.5} \leqslant 0.888\right) \approx P(-0.888 \leqslant Z \leqslant 0.888) =$$

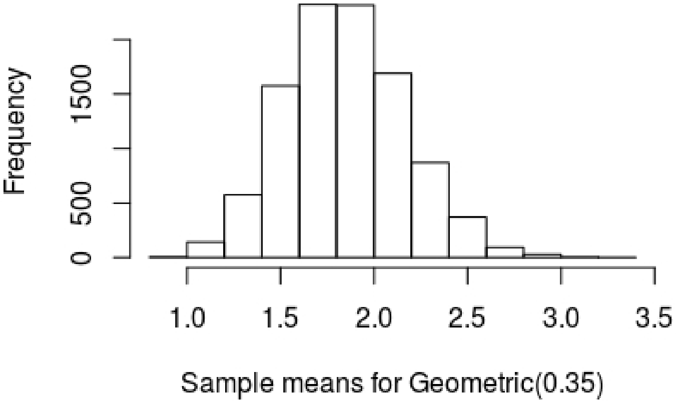
$$= pnorm(0.888) - pnorm(-0.888) = 2 \cdot pnorm(0.888) - 1 = 0.625$$

Aplicaţii ale LNM şi TLC

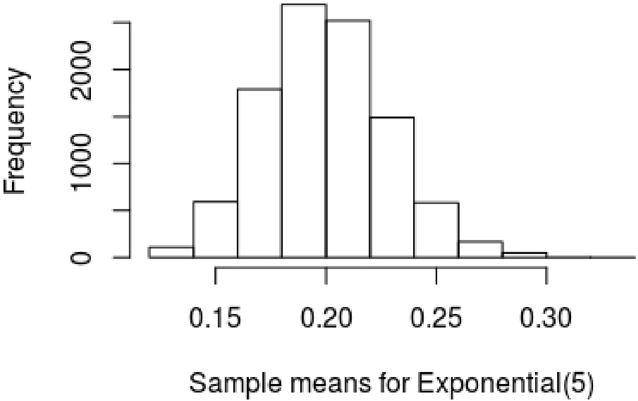
Exemplul 5. (*Verificarea TLC*) Considerăm o distribuţie de probabilitate, X, cu media μ şi dispersia σ^2 şi un şir format din n variabile aleatoare independente şi identic distribuite X_i, $i = \overline{1, n}$. Conform TLC, pentru valori mari ale lui n, media de selecţie, $\overline{x}_n$, are o distribuţie normală, $N(\mu, \sigma^2/n)$.

Dorim să verificăm această afirmaţie şi considerăm N astfel de medii de selecţie şi construim o histogramă. Exemplele de mai jos folosesc distribuţia geometrică $G(0.35)$ şi distribuţia exponenţială $Exp(5)$ ($n = 50$, $N = 10000$).

Aplicaţii ale LNM şi TLC



Applications of LLN and CLT



Aplicaţii ale LNM şi TLC

Exemplul 6. (*Verificarea TLC*) Considerăm o distribuţie de probabilitate, X, cu media μ şi dispersia σ^2 şi un şir format din n variabile aleatoare independente şi identic distribuite X_i, $i = \overline{1,n}$. Acest şir poate fi văzut ca un eşantion aleator simplu; dacă $\overline{x}_n$ este media de selecţie, TLC spune că

$$\lim_{n\to\infty} P\left[\frac{\overline{x}_n - \mu}{\sigma/\sqrt{n}} \leqslant z\right] = P(Z \leqslant z),$$

unde $Z : N(0,1)$. De obicei, pentru valori mari ale lui n putem face următoarea aproximare

$$P_n(z) = P\left[\frac{\overline{x}_n - \mu}{\sigma/\sqrt{n}} \leqslant z\right] \approx P(Z \leqslant z).$$

O methodă pentru a verifica cât de precisă este această aproximare: alegem independent N eşantioane (şiruri) $\left(X_i^k\right)_{i=\overline{1,n}}^{k=\overline{1,N}}$ şi calculăm

$$P^N = \frac{\left|\{k : \overline{x}_n^k \leqslant z\sigma/\sqrt{n} + \mu\}\right|}{N}.$$

Aplicaţii ale LNM şi TLC

Altfel spus, P^N este numărul de eşantioane (dintre cele N) care satisfac inegalitatea $\dfrac{\overline{x}_n - \mu}{\sigma/\sqrt{n}} \leqslant z$ supra numărul total de eşantioane. Această statistică ar trebui să aproximeze $P[Z \leqslant z]$. Pentru distribuţia exponenţială cu $\lambda = 2$, $n = 50$ şi $N = 2000$ rezultatele se găsesc mai jos (un singur eşantion de dimensiune n poate fi obţinut cu $rexp(n, \lambda)$).

z	-1.5	-1.0	-0.5	0	0.5	1.0	1.5
$P^N(z)$	0.055	0.154	0.313	0.509	0.723	0.831	0.931
$Rel.err$	16%	2.5%	1.6%	1.8%	4.6%	1.8%	0.2%
$pnorm(z)$	0.066	0.158	0.308	0.5	0.691	0.847	0.933

Eroarea relativă este egală cu $\dfrac{|P(Z \leqslant z) - P^N(z)|}{P(Z \leqslant z)}$.

Aplicaţii ale LNM şi TLC

Pentru a calcula $P^N(z)$ am folosit următorul algoritm:

$\mu \leftarrow 1/\lambda$; // why?
$\sigma \leftarrow 1/\lambda$; // why?
$c \leftarrow z * \sigma/\sqrt{n} + \mu$;
$j \leftarrow 1$;
for($i = \overline{1, N}$)
  if($mean(rexp(n, \lambda)) \leqslant c$)
    j++;
return j/N;

Sfârşit

Bibliography I

- Baron, M., *Probability and Statistics for Computer Scientist*, Chapman & Hall/CRC Press, 2013 sau ediţia electronică https://ww2.ii.uj.edu.pl/~z1099839/naukowe/RP/rps-michael-byron.pdf
- Johnosn, J. L., *Probability and Statistics for Computer Science*, Wiley Interscience, 2008.
- Lipschutz, S., *Theory and Problems of Probability*, Schaum's Outline Series, McGraw-Hill, 1965.
- Ross, S. M., *A First Course in Probability* , Prentice Hall, 5th edition, 1998.
- Shao, J., *Mathematical Statistics*, Springer Verlag, 1998.

Bibliography II

Stone, C. J., *A Course in Probability and Statistics*, Duxbury Press, 1996.