

Proiect la Statistică

Grupele 311, 312, 322

- I. **Inegalitatea Berry-Essen** este un rezultat celebru cu ajutorul căruia putem determina acuratețea aproximării pe care o realizează TLC. Cerând în plus față de condițiile din TLC ca $E|X_i|^3 < \infty$, avem că:

$$\sup_x |P(Z_n \leq x) - \Phi(x)| \leq \frac{33}{4} \frac{E|X_1 - \mu|^3}{\sqrt{n}\sigma^3},$$

unde X_1, X_2, \dots, X_n i.i.d., $\Phi(x)$ este funcția de repartiție a normalei standard, $\mu = E(X_1)$

$$\sigma = \sqrt{\text{Var}(X_1)} \text{ și } Z_n = \frac{\sqrt{n}(\bar{X}_n - \mu)}{\sigma}.$$

Cerinte:

- 1) Calculați $P(Z_n \leq x)$ pentru repartițiile:
Binomială, Geometrică, Poisson, Uniformă pe caz discret și respectiv continuu, Exponențială, Gamma și respectiv Beta.
- 2) Reprezentați grafic funcțiile obținute la 1).
- 3) Folosind funcția *optimize* aproximați pentru repartițiile de mai sus:
$$\sup_x |P(Z_n \leq x) - \Phi(x)|$$
- 4) Construiți câte o funcție în R care să calculeze $E(X)$ și respectiv $\text{Var}(X)$, unde tipul repartiției v.a. X este transmis fie printr-o denumire, fie prin funcția de masă/funcția densitate de probabilitate, în cazul discret și respectiv continuu.
- 5) Construiți o funcție în R care să calculeze $E|X_1 - \mu|^3$
- 6) Construiți un dataframe în R care să conțină marginile date de inegalitatea Berry-Essen pentru repartițiile: Binomială, Geometrică, Poisson, Uniformă pe caz discret și respectiv continuu, Exponențială, Gamma și respectiv Beta, pentru un volum al eșantionului $n \in \{30, 100, 1000\}$. Alegerea valorilor parametrilor repartițiilor ilustrate vă revine vouă.
- 7) Ilustrați grafic pentru $n \in \{30, 100, 1000\}$ pentru un interval pe care îl considerați relevant evoluția diferenței $P(Z_n \leq x) - \Phi(x)$ pentru fiecare din repartițiile de mai sus.
- 8) Construiți o funcție în R care să calculeze marginea dată de inegalitatea Berry-Essen pentru o v.a. pentru care se cunoaște funcția de masă/funcția densitate de probabilitate.

- II. Folosiți **metoda respingerii** pentru a genera observații din densitatea de probabilitate definită prin $f(x) \propto \exp(-x^2/2) \{\sin(6x)^2 + 3\cos(x)^2 \sin(4x)^2 + 1\}$ parcurgând pașii următori:

(OBS: Notăția “ \propto ” înseamnă că $f(x)$ este proporțional cu expresia din dreapta)

- a) Reprezentați grafic $f(x)$ și arătați că aceasta este mărginită de $Mg(x)$ unde $g(x)$ este densitatea de probabilitate a repartiției normale standard. Determinați o valoare potrivită pentru constanta M , chiar dacă nu este optimă.
(**Indiciu:** Folosiți funcția *optimise* din **R**)
- b) Generați 25000 de observații din densitatea de mai sus folosind metoda respingerii.
- c) Deduceți, pornind de la rata de acceptare a acestui algoritm, o aproximare a *constantei de normalizare* a lui $f(x)$, apoi comparați histograma valorilor generate cu reprezentarea grafică a lui $f(x)$ normalizată.

III. Folosind funcția ***check.convergence*** din pachetul **R *ConvergenceConcepts*** verificați dacă pentru următoarele exemple sunt verificate convergența în lege (în distribuție), în probabilitate și respectiv convergența aproape sigură. Interpretați și comentați rezultatele obținute.

- 1) Fie X_1, X_2, \dots, X_n v.a. i.i.d $X_i \sim \text{Beta}(\frac{1}{n}, \frac{1}{n})$ și $X \sim \text{Binomial}(1, \frac{1}{2})$. Verificați dacă $X_n \xrightarrow{D} X$. Dar în cazul în care $X_i \sim \text{Beta}(\frac{a}{n}, \frac{b}{n})$, cu $a > 0, b > 0$?
- 2) Fie X_1, X_2, \dots, X_n v.a. i.i.d uniform distribuite pe mulțimea de valori $\{\frac{1}{n}, \frac{2}{n}, \dots, 1\}$ și $X \sim \text{Unif}(0, 1)$. Verificați dacă $X_n \xrightarrow{D} X$. Dar $X_n \xrightarrow{P} X$?
- 3) Fie X_1, X_2, \dots, X_n v.a. i.i.d. Notăm cu m și respectiv M infimumul și respectiv supremumul mulțimii valorilor pe care le poate lua X_1 .
(i.e. $P(m \leq X \leq M) = 1$, $P(X_1 < a) > 0$ și $P(X_1 > b) > 0$ pentru orice $a > m$ și respectiv $b < M$).
Verificați că $\min\{X_1, X_2, \dots, X_n\} \xrightarrow{a.s.} m$ și că $\max\{X_1, X_2, \dots, X_n\} \xrightarrow{a.s.} M$.

Notă: Pentru utilizarea pachetului ***ConvergenceConcepts*** folosiți documentația primită la laborator.

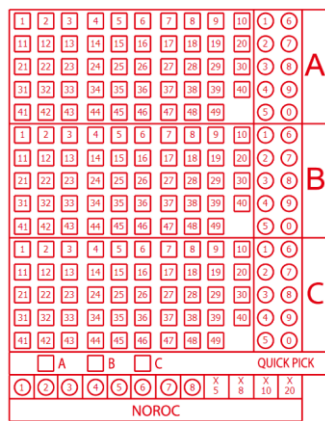
IV. O problemă de LOTO

Se consideră date următoarele informații despre un bilet la LOTO 6/49:

- Un *bilet simplu*, pentru care se pot alege 6 numere din 49 disponibile costa 7 lei
- În fiecare săptămână există două extrageri
- În urma extragerii, dacă exact 3 numere din cele alese se regăsesc printre cele extrase se primește suma de 30 lei, pentru exact 4 numere fondul de câștig este 363 350 lei, pentru exact 5 numere fondul de câștig este 390 000 lei iar pentru exact 6 numere fondul de câștig este 1 090 000 lei. Cu excepția primei situații, în care se primește o sumă fixă, fondul de câștiguri se împarte la numărul de câștigători din categoria respectivă.
- Un bilet asociat unei *variante combinate* constă în alegerea între 7 și 48 de numere (echivalent cu combinații de n luate câte 6 *variante simple*, $6 < n < 49$). Prețul asociat este calculat înmulțind 7 lei cu câte variante simple sunt echivalente cu acea variantă compusă.

Cerinte:

- 1) Folosindu-vă de simulare (10^6 valori), estimați următoarele cantități, iar acolo unde este posibil, comparați cu rezultatul teoretic:
 - a) Probabilitatea de a câștiga din prima încercare, cu un bilet simplu, la fiecare dintre categoriile posibile de câștig.
 - b) Probabilitatea de a câștiga după k încercări (k fiind un parametru), cu un bilet simplu, la fiecare dintre categoriile posibile de câștig.
 - c) Probabilitatea de a câștiga de r ori din k încercări (r și k fiind parametri), cu un bilet simplu, la fiecare dintre categoriile posibile de câștig.
 - d) Probabilitatea de a câștiga de r ori după k eșecuri (r și k fiind parametri), cu un bilet simplu, la fiecare dintre categoriile posibile de câștig.
 - e) Probabilitatea de a câștiga, în medie, cel puțin o dată pe an, dacă se joacă săptămânal câte un bilet simplu, timp de 30 ani.
 - f) Probabilitatea de a juca săptămânal, timp de un an, câte un bilet simplu și a nu câștiga niciodată
 - g) Probabilitatea de a juca săptămânal, timp de un an, câte un bilet simplu și de a câștiga, cumulat, o sumă mai mare decât costul total al biletelor jucate.
 - h) Probabilitatea de a câștiga din nou, la orice categorie, în săptămâna următoare, cu un bilet simplu, presupunând că în săptămâna curentă jucătorul a câștigat premiul cel mare.
 - i) Care este cea mai probabilă combinație de categorii de câștiguri realizate pentru un jucător care cumpără câte un bilet simplu pentru fiecare extragere, timp de un an? (ex. câștigă de 12 ori câte 3 numere, de 2 ori 4 numere, o dată 5 numere și niciodată cele 6 numere)
- 2) Folosindu-vă de simulare și/sau rezultate teoretice răspundeți justificat la următoarele întrebări:
 - a) Care este câștigul mediu anual estimat al Loteriei, știind că există 2 extrageri săptămânale?
 - b) Care este numărul minim de bilete simple care trebuie cumpărate într-un an pentru ca Loteria să nu fie în pierdere?
 - c) Un jucător are un buget lunar de 70 lei. Care ar fi strategia optimă de joc (bilete simple jucate în săptămâni diferite, un singur bilet cu o variantă complexă jucată o singură dată etc.) care să-i aducă cele mai mari șanse de câștig? Estimați/Calculați aceste șanse!
 - d) Care este numărul mediu de câștigători de la fiecare categorie de la o extragere? Puteți identifica o repartiție asociată?
- 3) Loteria dorește să introducă un nou tip de bilet, *biletul "stânga-dreapta"*. Pentru acest tip de bilet se consideră câștigătoare situația în care fie au fost alese exact numerele câștigătoare, fie vecinii săi imediați din stânga sau din dreapta, la nivel de rând, conform aranjării numerelor pe biletul LOTO. De exemplu, dacă este extras 10, dar jucătorul a ales 9 atunci va fi considerat număr câștigător, însă nu se întâmplă același lucru în cazul lui 11 (pentru că se află pe alt rând).



- Estimați cât ar trebui să fie prețul unui asemenea bilet pentru ca jucătorii să fie interesați să-l cumpere.
 - Estimați cât ar trebui să fie prețul minimum al unui asemenea bilet pentru ca Loteria să fie interesată să-l comercializeze.
 - Pornind de la a) și b) construiți o animație(folosind, de exemplu, pachetul animation <https://www.jstatsoft.org/article/view/v053i01>) care să arate cum evoluează interesul cumpărătorului și respectiv al Loteriei față de a cumpăra/vinde asemenea bilete în funcție de modificarea prețului acestuia.
 - Construiți o strategie de a alege numerele cu cele mai mari șanse de câștig cu un asemenea bilet și evaluați care sunt aceste șanse.
 - În ce condiții un asemenea bilet ar aduce pentru un jucător regulat un câștig mediu anual estimat ca fiind superior celui unui bilet simplu?
- 4) Loteria pune la dispoziție istoricul extragerilor anterioare din ultimii 20 de ani. În ce manieră puteți folosi aceste informații pentru a crește șansele de câștig cu un bilet simplu? Justificați răspunsul și ilustrați-l cu grafice relevante.

Precizări importante

- Fiecare echipa va trimite prin liderul echipei **un singur** e-mail la adresa simona.cojocsa@fmi.unibuc.ro ce va conține o arhivă cu 2 componente: codul R(comentat!) și documentația proiectului, în format .docx sau .pdf.
- Documentația proiectului trebuie să conțină, pe prima pagină, numele membrilor echipei, liderul echipei și grupa din care face parte fiecare.
Pentru fiecare exercițiu în parte documentația trebuie să conțină:
 - **Calculule matematice solicitate**(pot fi tehnoredactate sau scrise de mână și scanate, însă dacă apelați la ultima variantă vă rog să vă asigurați că ați scris **citeț**, fără ștersături sau zone scanate deficitar)
 - **Codul R comentat**
 - **Graficele realizate**(acolo unde au fost cerute)
 - **Comentariile și concluziile voastre**
- Pentru obținerea punctajului maxim trebuie să rezolvați toate exercițiile propuse. Dacă odată cu rezolvarea lor, identificați și rezolvați unele cerințe suplimentare care sunt **relevante** pentru subiectul respectiv puteți obține un bonus de până la 10 p, fără ca nota finală de laborator să poată depăși 50 p.
- Termenul de trimitere a proiectului este **4 februarie 2024 ora 22:00**.