

## Lista de proiecte la Probabilități și Statistică, CTI, an II, grupele 251-254

### Precizări importante

1. Proiectele se realizează în echipă de 2-4 persoane. Fiecare echipa va desemna un leader care va fi precizat în documentație.
2. Fiecare echipa alege unul din cele 2 proiecte pe care vrea să le realizeze.
3. Leaderul echipei va trimite pe adresa [simona.cojocea@fmi.unibuc.ro](mailto:simona.cojocea@fmi.unibuc.ro) până la data de **18 iunie 2023 ora 22:00** o singură arhivă care va conține fișierele sursă ale proiectului împreună cu documentația.
4. Documentația este **obligatorie** și lipsa ei atrage necorectarea proiectului.
5. Documentația trebuie să conțină :
  - numele membrilor echipei
  - descrierea problemei
  - aspecte teoretice folosite în rezolvarea problemei care depășesc nivelul cursului
  - precizări privind pachete software folosite și surse de inspirație
  - comentarea codului și a soluției prezentate
  - identificarea unor eventuale dificultăți în realizarea cerințelor
  - concluzii
6. Pentru **proiectul 2**: găsiți informații relevante despre pachetul Shiny și exemple de aplicații aici: <https://shiny.rstudio.com/>
7. Dacă la oricare din cele 2 proiecte se realizează cerințe suplimentare față de cele date, cerințe care să fie relevante, se poate obține un bonus de **5p**, fără însă ca nota finală asociată laboratorului să poată depăși **30 p**.

**Proiectul 1: Construirea unui pachet R pentru lucru cu variabile aleatoare 20 p.**

Folosind documentul suport(*rpackage\_instructions.pdf*) și orice alte surse de documentare considerați potrivite construiți un pachet R care să permită lucru cu variabile aleatoare . Pentru a primi punctaj maxim, pachetul trebuie să implementeze cel puțin 8 din următoarele cerințe, cu precizarea că cerința 9 este **obligatorie**:

- 1) Verificarea dacă o funcție introdusă de utilizator este funcție de masă/densitate de probabilitate. Arătați în mod explicit cum diferențiați între cazul funcției de masă și respectiv al densității de probabilitate.
- 2) Fiind dată o funcție  $f$  , introdusă de utilizator, determinarea unei *constante de normalizare*  $k$ (cu precizarea dacă această constantă duce la o funcție de masă sau la o densitate de probabilitate). În cazul în care o asemenea constantă nu există, afișarea unui mesaj corespunzător către utilizator.
- 3) Reprezentarea grafică a densității de probabilitate/funcției de masă și a funcției de repartiție pentru diferite valori ale parametrilor repartiției. În cazul în care funcția de repartiție nu este dată într-o formă *explicită*(ex. repartiția normală) se acceptă reprezentarea grafică a unei aproximări a acesteia.
- 4) Calculul mediei, dispersiei și a momentelor inițiale și centrate până la ordinul 4(dacă există). Atunci când unul dintre momente nu există, se va afișa un mesaj corespunzător către utilizator.
- 5) Calculul mediei și dispersiei unei variabile aleatoare  $g(X)$ , unde  $X$  are o repartiție continuă cunoscută iar  $g$  este o funcție continuă precizată de utilizator.
- 6) Crearea unei funcții **P** care permite calculul diferitelor tipuri de probabilități asociate unei variabile aleatoare(ex. probabilități condiționate). Exemplificați cu cel puțin 4 timpuri de astfel de situații.
- 7) Calculul covarianței și coeficientului de corelație pentru două variabile aleatoare (**Atenție:** Trebuie să folosiți *funcția de masă/densitatea comună* a celor două variabile aleatoare!)
- 8) Pornind de la funcția de masă/densitatea comună a două variabile aleatoare, construirea funcțiilor de masă/densităților marginale și a funcțiilor de masă/densităților condiționate.
- 9) Se dau seturile următoare de valori:
  - a) 7 4 2 11 2 1 2 1 6 6 0 1 3 9 7 0 1 14 0 5 1 5 2 4 3 1 0 0 26 1
  - b) -1.91 -0.97 4.59 2.19 -0.86 -0.74 -0.60 -1.29 0.93 1.42 2.14 -2.01 2.60 1.45 2.60 -3.32 -3.62 3.09 2.91 3.60 -0.83 -0.27 1.82 -1.38 -1.76 1.43 -0.59 -1.34 2.07 1.02
  - c) 0.90 8.91 0.06 1.85 1.61 6.50 0.26 0.04 0.62 1.01 3.42 1.45 3.44 0.46 0.55 0.09 2.22 0.65 0.61 6.45 0.27 4.81 2.27 0.34 4.51 0.42 3.71 2.59 0.42 11.18
  - d) 4.83 4.37 5.57 4.22 5.96 5.11 5.52 4.81 5.19 4.19 4.73 5.92 5.63 4.53 4.67 4.84 5.25 5.06 5.98 5.25 4.60 4.11 4.32 5.09 5.25 5.10 4.36 5.40 5.33 4.65
  - e) 11 11 10 10 10 6 5 9 11 10 14 8 11 6 13 9 14 16 14 10 7 7 11 12 9 5 12 15 9 12

Pentru fiecare dintre seturile de valori de mai sus efectuați:

9.1. Faceți histograma valorilor. Calculați mediana, media și deviația standard și ilustrați pe desen aceste valori.

9.2. Găsiți o manieră de a identifica din ce repartiție au fost generate valorile de mai sus, cu identificarea parametrilor și justificarea alegerii făcute.

9.3 Presupuneți că valorile au fost extrase dintr-o repartiție normală de parametri necunoscuți. Folosind metoda verosimilității maxime și respectiv metoda momentelor estimați acești parametri în baza celor 5 eșantioane. Comparați estimările obținute prin cele două metode și comentați rezultatele.

9.4 Identificați în care din cazurile de mai sus este verosimil ca valorile să fie extrase dintr-o repartiție normală și justificați răspunsul. Argumentați și în caz contrar.

10) Construiți o funcție care răspunde la întrebările de la 9) pentru un set oarecare de valori, de lungime variabilă, care urmează să fie introdus de utilizator(de la tastatură/preluat dintr-un fișier).

## Proiectul 2: Crearea unei aplicații pentru lucrul cu variabile aleatoare folosind Shiny 20p

Folosind pachetul R *Shiny* construiți o aplicație web care să permită lucrul cu variabile aleatoare(discrete și continue). Pentru obținerea punctajului maxim este necesar să implementați **cel puțin 6 din următoarele cerințe**, dintre care cerința 9) este **obligatorie**:

- 1) Crearea unui meniu din care poate fi aleasă o repartiție de variabile aleatoare(trebuie să aveți **cel puțin 8 repartiții** disponibile, dintre care *cel puțin 3 trebuie fie continue!*), cu particularizarea parametrilor. Utilizatorul va vizualiza o scurtă descriere a respectivei repartiții(în stilul Wikipedia), cu reprezentarea grafică a densității de probabilitate/funției de masă și a funcției de repartiție, afișarea mediei, dispersiei și a altor elemente ce caracterizează respectiva repartiție.
- 2) Crearea unei opțiuni în meniul principal care permite utilizatorului să-și introducă propriile repartiții de variabile aleatoare, care vor putea fi accesate ulterior în aceeași manieră ca cele preexistente.
- 3) *Lucru cu evenimente*: se dau două evenimente A și B despre care se precizează dacă sunt independente, incompatibile sau dacă nu se știe nimic despre ele. Pornind de la un set de informații despre niște probabilități legate de ele să se determine toate celelalte probabilități(ex. Știu  $P(A), P(B), P(A \cap B)$  și determin  $P(A \cup B)$  folosind formula lui Poincare,  $P(A|B)$  și  $P(B|A)$  din formula probabilității condiționate)
- 4) *Generalizarea lucrului cu evenimente*: implementarea formulei probabilității totale, inegalității lui Boole, calculul unei probabilități condiționate în care sunt implicate mai mult de 3 evenimente(ex.  $P(A|C \cup B)$ ).
- 5) Afișarea unei v.a. discrete. În cazul în care numărul său de valori este foarte mare să existe posibilitatea de a alege prima valoare care se dorește a fi vizualizată(ex. X ia valori de la 1 la 100, iar eu vreau să vizualizez v.a. începând cu poziția 53).
- 6) Crearea unei opțiuni într-un meniu care determină o transformare a unei variabile aleatoare  $g(X)$ , unde X este o variabilă aleatoare discretă, iar g este o funcție furnizată de utilizator.
- 7) Pornind de la o v.a. continuă X a cărei repartiție e aleasă de utilizator din Galerie și de la o funcție g introdusă de utilizator, calculul și afișarea media și dispersia v.a.  $g(X)$ .
- 8) Crearea unui meniu care să permită utilizatorului să introducă elemente ale repartiției comune a două v.a. discrete, urmând ca restul elementelor să fie calculate și afișate la apăsarea unui buton *Completează*. Pentru aceste două v.a. să se determine: repartițiile marginale, media, dispersia, covarianța și coeficientul de corelație.
- 9) Se dau seturile următoare de valori:
  - f) 7 4 2 11 2 1 2 1 6 6 0 1 3 9 7 0 1 14 0 5 1 5 2 4 3 1 0 0 26 1
  - g) -1.91 -0.97 4.59 2.19 -0.86 -0.74 -0.60 -1.29 0.93 1.42 2.14 -2.01 2.60 1.45 2.60 -3.32 -3.62 3.09 2.91 3.60 -0.83 -0.27 1.82 -1.38 -1.76 1.43 -0.59 -1.34 2.07 1.02
  - h) 0.90 8.91 0.06 1.85 1.61 6.50 0.26 0.04 0.62 1.01 3.42 1.45 3.44 0.46 0.55 0.09 2.22 0.65 0.61 6.45 0.27 4.81 2.27 0.34 4.51 0.42 3.71 2.59 0.42 11.18
  - i) 4.83 4.37 5.57 4.22 5.96 5.11 5.52 4.81 5.19 4.19 4.73 5.92 5.63 4.53 4.67 4.84 5.25 5.06 5.98 5.25 4.60 4.11 4.32 5.09 5.25 5.10 4.36 5.40 5.33 4.65
  - j) 11 11 10 10 10 6 5 9 11 10 14 8 11 6 13 9 14 16 14 10 7 7 11 12 9 5 12 15 9 12

Pentru fiecare dintre seturile de valori de mai sus efectuați:

- 9.1. Faceți histograma valorilor. Calculați mediana, media și deviația standard și ilustrați pe desen aceste valori.
- 9.2. Găsiți o manieră de a identifica din ce repartiție au fost generate valorile de mai sus, cu identificarea parametrilor și justificarea alegerii făcute.
- 9.3 Presupuneți că valorile au fost extrase dintr-o repartiție normală de parametri necunoscuți. Folosind metoda verosimilității maxime și respectiv metoda momentelor estimați acești parametri în baza celor 5 eșantioane. Comparați estimările obținute prin cele două metode și comentați rezultatele.
- 9.4 Identificați în care din cazurile de mai sus este verosimil ca valorile să fie extrase dintr-o repartiție normală și justificați răspunsul. Argumentați și în caz contrar.
- 10) Construiți o funcție care răspunde la întrebările de la 9) pentru un set oarecare de valori, de lungime variabilă, care urmează să fie introduse de utilizator(de la tastatură/preluate dintr-un fișier).