

Lista de proiecte la Probabilități și Statistică, Informatică, an II, seria 23

Proiectul 1: Construirea unui pachet R pentru lucru cu variabile aleatoare continue 40 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 continue. Pentru a primi punctaj maxim, pachetul trebuie să implementeze cel puțin 8 din următoarele cerințe(oricare 8!):

- 1) Fiind dată o funcție f , introdusă de utilizator, determinarea unei *constante de normalizare* k . În cazul în care o asemenea constantă nu există, afișarea unui mesaj corespunzător către utilizator.
- 2) Verificarea dacă o funcție introdusă de utilizator este densitate de probabilitate.
- 3) Crearea unui obiect de tip variabilă aleatoare continuă pornind de la o densitate de probabilitate introdusă de utilizator.
- 4) Reprezentarea grafică a densității 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.
- 5) 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.
- 6) 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.
- 7) Crearea unei funcții **P** care permite calculul diferitelor tipuri de probabilități asociate unei variabile aleatoare continue(similar funcției **P** din pachetul *discreteRV*)
- 8) Afișarea unei “*fișe de sinteză*” care să conțină informații de bază despre respectiva repartiție(cu precizarea sursei informației!). Relevant aici ar fi să precizați pentru ce e folosită în mod uzual acea repartiție, semnificația parametrilor, media, dispersia etc.
- 9) Generarea a n valori(unde n este precizat de utilizator!) dintr-o repartiție de variabile aleatoare continue(vezi materialul suport pentru partea de *simulare*).
- 10) Calculul covarianței și coeficientului de corelație pentru două variabile aleatoare continue(**Atentie**:Trebuie să folosiți *densitatea comună* a celor două variabile aleatoare!)
- 11) Pornind de la densitatea comună a două variabile aleatoare continue, construirea densităților marginale și a densităților condiționate.
- 12) Construirea sumei și diferenței a două variabile aleatoare continue independente(folosiți formula de *convoluție*)

Proiectul 2: Crearea unei aplicații pentru lucrul cu variabile aleatoare folosind Shiny 35p

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 8 din următoarele cerințe:

- 1) Crearea unui meniu din care poate fi aleasă o repartiție de variabile aleatoare(trebuie să aveți *cel puțin 15 repartiții* disponibile!), cu particularizarea parametrilor. Utilizatorul va vizualiza o scurtă descriere a respectivei repartiții(în stilul Wikipedia), cu reprezentarea grafică a densității de probabilitate/funcției 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) Fiind dată o variabilă aleatoare X cu repartiția dată dintre cele disponibile în Galerie să se permită utilizatorului să se calculeze diferite probabilități(asemănător comportamentului funcției **P** din pachetul *discreteRV*), atât pentru variabile discrete cât și pentru cele continue.
- 7) 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.
- 8) 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. $g(X)$.
- 9) 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.
- 10) Crearea unui meniu care să permită utilizatorului ca, pornind de la repartiția comună a două v.a. discrete X și Y(furnizată de utilizator) să fie construită și afișată repartiția comună a alte două v.a. Z și T, unde $Z=g(X,Y)$ iar $T=h(X,Y)$, iar g și h sunt furnizate de utilizator. (ex. $Z=\max(X,Y)$, iar $T=\min(X,Y)$).
- 11) Crearea unui meniu care să permită utilizatorului introducerea unor valori numerice, sau importarea unui fișier care să conțină aceste valori. Pentru acest set de date să se determine mediana, cuartilele și să se afișeze histograma și diagrama boxplot.
- 12) Crearea unei opțiuni în meniu care să permită operații cu v.a. discrete independente(sumă, diferență, produs, raport, etc.).

Proiectul 3: Câteva aplicații cu variabile aleatoare în R Studio 30 p

1. Fie două variabile aleatoare discrete X și Y cu repartițiile:

$$X : \begin{pmatrix} x_1 & x_2 & \dots & x_n \\ p_1 & p_2 & \dots & p_n \end{pmatrix} \text{ și respectiv } Y : \begin{pmatrix} y_1 & y_2 & \dots & y_m \\ q_1 & q_2 & \dots & q_m \end{pmatrix}$$

- a) Construiți o funcție **frepcomgen** care primește ca parametri m și n și care generează un tabel cu repartiția comună a v.a. X și Y incompletă, dar într-o formă în care poate fi completată ulterior.

Observație: Se cere la a) să generați valorile lui X, valorile lui Y și suficient de multe valori pentru p_i , q_j și respectiv π_{ij} astfel încât să poată fi determinată repartiția comună a celor două v.a.

Nota: În construirea algoritmului puteți începe de la cazul particular m=3 și n=2. Dacă reușiți să oferiți soluția doar pentru acest caz particular, dar nu și pentru cazul general veți primi punctaj parțial.

- b) Construiți o funcție **fcompleprecom** care completează repartiția comună generată la punctul anterior (pentru cazul particular sau pentru cazul general).

Nota: În cazul în care nu știți să rezolvați punctul a) puteți construi o funcție care să determine repartiția comună pornind de la un exemplu discutat la seminar.

- c) Având la dispoziție repartiția comună a v.a. X și Y de la punctul b) calculați:

1) $\text{Cov}(2X+3, 4Y-11)$

2) $P(0 < X < 8/Y > 3)$

3) $P(X > 2, Y < 7)$

- d) Pentru exemplul obținut la punctul b) construiți două funcții **fverind** și respectiv **fvernecor** cu ajutorul cărora să verificați dacă variabilele X și Y sunt:

1) independente

2) necorelate

2. Construiți o funcție în R care să preia ca date de intrare densitățile de probabilitate a două variabile aleatoare continue independente și opțiunea utilizatorului (un număr de la 1 la 4) care să returneze următoarele rezultate:

-opțiunea 1- suma v.a. (folosiți formula de convoluție)

-opțiunea 2-diferența v.a. (folosiți formula de convoluție)

-opțiunea 3-media și dispersia celor 2 v.a.

-opțiunea 4-reprezentarea grafică, în același reper, a celor 2 densități, cu culori diferite

3. Construiți o funcție în R care să se comporte ca un *generator de numere aleatoare* (se generează n valori, unde n este dat de utilizator) având următoarele specificații:

- 1) Pentru prima valoare (x_1) se citește timpul sistemului, se ia numărul format din minute și secunde (t_1) și se calculează modulo 12.

(ex. Dacă ora sistemului este 12:15:23 atunci $t_1=1523$)

Dacă $t_1 \bmod 12=0$ atunci x_1 se generează folosind funcția **rnorm** din R cu parametri dați de numărul minutelor și respectiv numărul secundelor.

Dacă $t_1 \bmod 12=1$ atunci x_1 se generează folosind funcția **rpois** din R cu parametru dat de număr reprezentând minutele și se adună la el un număr y_1 generat cu funcția **runif** din R (cu parametrii 0 și 1).

Dacă $t_1 \bmod 12=5$ atunci x_1 se generează folosind funcția **rexp** din R cu parametru dat de numărul reprezentat de ora sistemului.

Dacă $t_1 \bmod 12=7$ atunci x_1 se generează folosind funcția **rbinom** din R cu parametrii dați de ora sistemului și $1/nr_minute$ și se adună la el un număr y_1 generat cu funcția **runif** din R(cu parametrii 0 și 5) .

Dacă $t_1 \bmod 12=8$ atunci x_1 se generează folosind funcția **runif** din R(de parametri 5 și 12).

Dacă $t_1 \bmod 12=9$ atunci x_1 se generează folosind funcția **rgamma** din R și se scade din el un număr y_1 generat cu funcția **rhyper** din R.

În celelalte cazuri se reia procesul și se citește din nou ora sistemului. Dacă procesul a fost reluat de 3 ori și nu s-a intrat pe unul dintre cazurile de mai sus atunci x_1 se generează folosind funcția **rnorm** din R cu parametrii 0 și 1.

- 2) Pentru valorile x_n cu $n>1$ se folosește următoarea formulă de recurență:
 $x_n=a* x_{n-1}+b$ unde a este o valoare generată cu funcția **rexp** din R de parametru 2, iar b este o valoare generată cu funcția **rnorm** din R de parametri 5 și 1

Funcția returnează un vector cu valorile generate și realizează histograma lor.

4. Folosind setul de date **X** efectuați operații de statistică descriptivă pentru variabilele din acest set de date(medie, varianța, quartile, boxplot, interpretări).

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 3 proiecte pe care vrea să le realizeze.
3. Leaderul echipei va trimite pe adresa simona.cojocsa@fmi.unibuc.ro până la data de **1 februarie 2021 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. Pentru **proiectul 1**: dacă alegeți implementarea cerinței legate de generarea de numere aleatoare dintr-o repartiție dată, solicitați materiale suplimentare!
8. Pentru **proiectul 3**: acolo unde anumite detalii nu au fost furnizate(de ex. parametrii unei repartiții) rămân la alegerea voastră. Setul de date X va fi furnizat fiecărei echipe în parte, odată ce șefii de grupă îmi trimit listele cu echipele formate.
9. Dacă la oricare din cele 3 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 **50 p**.