Proiect

Grupele 301 și 321

Notă privind redactarea proiectului: Raportul trebuie scris prin intermediul pachetului `Rmarkdown` din `R`. Informații introductive despre modul de folosire a acestui pachet pot fi găsite aici sau aici iar pentru mai multe detalii se poate consulta cartea R Markdown: The definitive guide. Toate simulările, figurile și codurile folosite trebuie incluse în raport. Se va folosi doar limbajul *R*. Raportul trimis trebuie să conțină pe lângă fișierul generat (*HTML*, *Microsoft Word* sau LaTeX) și fișierul `.Rmd` care conține codul sursă.

1 Problema 1

Un segment de lungime 1 este rupt în trei bucăți. Presupunând că punctele de ruptură sunt date de două variabile aleatoare X și Y repartizate pe $[0, 1]$, scopul acestui exercțiu este să determinăm, în funcție de procedura de alegere a punctelor de ruptură, care este probabilitatea de formare a unui triunghi cu lungimile celor trei segmente obținute.

Procedura 1: Presupunem că punctele de ruptură sunt date de două variabile aleatoare X și Y, independente și repartizate uniform pe intervalul $[0, 1]$.

1. Fie a, b și c lungimile celor trei segmente obținute (luate de la stanga la dreapta). Arătați că lungimile celor trei segmente pot forma un triunghi dacă și numai dacă fiecare dintre cele trei lungimi este mai mică sau egală cu $\frac{1}{2}$. Traduceți această condiție in funcție de v.a. X și Y.
2. Într-o primă etapă dorim să aproximăm lungimile medii ale celor trei segmente obținute. Pentru aceasta, simulăm $N = 5000$ de realizări independente ale cuplului (X, Y). Care sunt valoriile lungimilor medii ? Ce teoremă limită justifică acest rezultat ?
3. Dorim de asemenea să răspundem într-o manieră mai riguroasă la intrebarea problemei:

a. Cuplul de puncte (X, Y) de ruptură poate fi văzut ca un punct în pătratul unitate $[0, 1]^2$. Plecând de la 5000 de realizări independente ale cuplului (X, Y), reprezentați grafic punctele (X_i, Y_i) din interiorul pătratului $[0, 1]^2$ care determină cele trei segmente cu ajutorul cărora putem forma un triunghi, cu albastru, și pe celelalte cu roșu.

b. Plecând de la $N = 5000$ de realizări independente ale cuplului (X, Y), estimați probabilitatea căutată.

c. Găsiți această probabilitate teoretic și comparați cu rezultatul găsit la punctul anterior.

4.[1] Presupunând că punctele de ruptură sunt date de procedura 1, ce puteți spune despre probabilitatea de formare a unui triunghi obtuzunghic ? Justificați atât teoretic cât și prin simulare.

Ne întrebăm acum ce se întâmplă cu probabilitatea de formare a unui triunghi cu ajutorul celor trei segmente determinate de punctele de ruptură dacă adoptăm următoarele două proceduri:

Procedura 2: Alegem pentru început un punct de ruptură X repartizat uniform $\mathcal{U}([0, 1])$ și dintre cele două segmente formate îl alegem pe cel mai lung pe care alegem un al doilea punct, Y, repartizat uniform pe acest segment.

[1] Această întrebare este mai dificilă și nerezolvarea ei nu scade punctajul proiectului pentru grupele formate din cel mult doi studenți.

5. Reconsiderați subpunctele b. și c. de la punctul 3.

6. Pentru a ilustra grafic regiunea determinată de perechile care verifică problema (formează un triunghi) putem considera variabila aleatoare Z repartizată uniform pe intervalul $[0,1]$ și care să fie independentă de X și să exprimăm lungimile celor trei segmente în funcție de aceasta. Plecând de la 5000 de realizări independente ale cuplului (X,Z), reprezentați grafic, cu albastru, punctele (X_i,Z_i) din interiorul pătratului $[0,1]$ care determină cele trei segmente cu ajutorul cărora putem forma un triunghi și pe celelalte cu roșu.

Procedura 3: Alegem pentru început un punct de ruptură X repartizat uniform $\mathcal{U}([0,1])$ care va împărții segmentul $[0,1]$ în două subsegmente. Aruncăm, de manieră independentă, o monedă echilibrată și alegem în funcție de rezultatul aruncării, cap sau pajură, segementul din stânga (cap) sau cel din dreapta (pajură). Pe segmentul selecționat alegem un al doilea punct de ruptură, Y, repartizat uniform pe acest segment.

7. Reconsiderați punctele 5. și 6. de la *Procedura 2*.

2 Problema 2

Fie $\mathbf{X}=(X_1,X_2)$ un vector aleator din $\mathbb{R}^2$ cu densitate f_1 proporțională cu

$$\tilde{f}_1(x_1,x_2)=\exp\left[-\frac{1}{2}\left(\frac{x_1^2}{4}+x_2^2\right)\right]\mathbf{1}_{\{|x_2|\leq 1\}},$$

și $\mathbf{Y}=(Y_1,Y_2)$ un vector aleator din $\mathbb{R}^2$ cu densitate f_2 proporțională cu

$$\tilde{f}_2(y_1,y_2)=\left[\cos^2(y_1)+0.5\sin^2(3y_2)\cos^4(y_1)\right]\exp\left[-\frac{1}{2}\left(\frac{y_1^2}{4}+y_2^2\right)\right].$$

Obiectivul acestei probleme este de a construi un algoritm (prin metoda respingerii) care să permită simularea de observații din repartițiile f_1 și respectiv f_2.

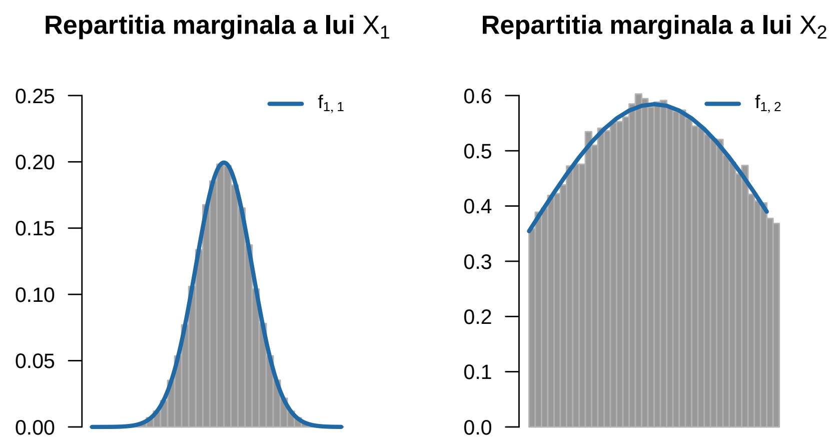
1. Arătați că pentru a genera o observație din repartiția f putem aplica algoritmul de acceptare-respingere funcției $\tilde{f}$.

2. Determinați constantele M_1, M_2 și o densitate g astfel încât pentru $\mathbf{x}=(x_1,x_2)\in\mathbb{R}^2$

$$\tilde{f}_1(\mathbf{x})\leq M_1 g(\mathbf{x})\quad \text{și}\quad \tilde{f}_2(\mathbf{x})\leq M_2 g(\mathbf{x}).$$

3. Deduceți o metodă de simulare din repartițiile de densitate f_1 și f_2. Construiți câte o funcție în `R` care să implementeze această metodă pentru fiecare densitate în parte și generați câte un eșantion de talie $n=100000$ din fiecare densitate bidimensională. Puteți optimiza procedura, vectoriza codul ?

4. Comparați densitățile marginale empirice rezultate din eșantionul generat anterior pentru repartiția de densitate f_1 cu densitățile marginale teoretice ale lui f_1 prin trasarea unei histograme peste care suprapunem densitatea teoritică (a se vedea figura de mai jos).



5. Comparați repartiția empirică cu cea teoretică în cazul f_1 printr-o reprezentare 2D/3D cu ajutorul pachetului `plot3D` (a se vedea funcțiile `hist3D`, `persp3d` și `image2D`). (a se vedea figura de mai jos)

Indicație: Pentru trasarea repartiției empirice (a histogramei 3d) aveți nevoie să construiți o funcție care să returneze repartiția empirică a eșantionului 2d pe un grid dat (o partiție 2d). Pentru a crea diviziunea bidimensională (echidistantă) puteți folosi funcția `cut` (pe fiecare dimensiune) iar pentru a determina câte puncte cad în fiecare subregiune dreptunghiulară puteți folosi funcția `table`. Țineți cont și de normalizarea histogramei !