Probleme de antrenament pentru lucrarea III

1. Un lanţ Markov are matricea de tranziţie:

$$Q = \begin{bmatrix} 0.1 & 0.2 & 0.4 & 0.3 \\ 0 & 0.7 & 0.2 & 0.1 \\ 0 & 0.3 & 0.3 & 0.4 \\ 0.4 & 0 & 0.3 & 0.3 \end{bmatrix}$$

a) Desenaţi graful de tranziţie şi explicaţi de ce lanţul este ireductibil şi aperiodic.

b) Să se calculeze $P(X_4 = 3|X_0 = 2, X_1 = 2, X_2 = 3, X_3 = 1)$;

c) Care este probabilitatea ca lanţul să evolueze pe traiectoria de noduri $(4, 1, 3, 2)$? ştiind că distribuţia iniţială de probabilitate, π_0, este distribuţia uniformă pe mulţimea nodurilor.

d) Matricea Q are vectorul propriu $v = (0.5, 0.5, 0.5, 0.5)$ corespunzător valorii proprii 1, iar Q^T are vectorul propriu $w = (0.22, 0.660.52, 0.48)^T$ corespunzător aceleiaşi valori. Care din cei 2 vectori conduce la distribuţia de echilibru a lanţului Markov? Determinaţi coordonatele distribuţiei de echilibru, π, şi explicaţi ce reprezinta a două corrdonată, π_2.

2. După evaluarea anuală angajaţii unei firme au fost clasificaţi astfel: background redus, `R`, background mediu, `M`, şi ridicat, de top, `T`. Estimările probabilitătilor de trecere dintr-o clasă în alta, într-un an, sunt date în tabelul:

Q=

	R	M	T
R	0.65	0.28	0.07
M	0.15	0.67	0.18
T	0.12	0.36	0.52

a) Dacă un angajat are 3 ani consecutiv nivelul redus, care este probabilitatea ca în anul al patrulea să aibă nivel `M`?

b) Dacă tu ai acum nivelul `M`, care este probabilitatea ca peste trei ani să ai nivel `T`, fără ca în aceşti ani să treci prin `R`?

c) Ce semnificaţie are în contextul problemei fiecare coordonată $\pi(1), \pi(2), \pi(3)$ a distribuţiei de echilibru a acestui lanţ Markov?

3. Un lanţ Markov, ce are matricea de tranziţie:

$$Q = \begin{bmatrix} 0.85 & 0.05 & 0.05 & 0 & 0.05 \\ 0.2 & 0.6 & 0.1 & 0 & 0.1 \\ 0.15 & 0.05 & 0.5 & 0.25 & 0.05 \\ 0.3 & 0.05 & 0.05 & 0 & 0.6 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

modelează comportamentul unui navigator pe WEB, pe site-ul unui shop online. Cele 5 noduri ale grafului de tranziţie codifică una din acţiunile navigatorului pe site:

1- Dă click-uri pe diverse produse;

2 Citeşte fişierele ce descriu produsele sau conţin review-uri;

3 Cumpără un produs (completează formularul online cu adresa de expediere şi detaliile cardului);

4 Dă click pentru confirmarea comenzii;

5 Părăseşte site-ul.

a) Ce fel de lanţ este acest lanţ Markov? Este el ireductibil şi aperiodic sau este un lanţ absorbant? Justificaţi răspunsul!

b) Scrieţi pseudocodul care determină distribuţia de probabilitate π_n. Ce reprezintă coordonatele vectorului probabilist π_6?

c) Observăm că numărul mediu de vizite ale nodului 4 înainte de a ieşi de pe site este de fapt numărul mediu de produse cumpărate de pe acest site. Scieţi etapele de calcul, care conduc la aflarea acestui număr mediu de produse cumpărate, ştiind că navigarea a inceput din nodul 1.

4. Un lanţ Markov pe un spaţiu cu 4 stări are matricea de tranziţie: $Q = \begin{bmatrix} 0 & 0 & 1 & 0 \\ 0.3 & 0 & 0 & 0.7 \\ 0 & 0.4 & 0.1 & 0.5 \\ 0.2 & 0.4 & 0.4 & 0 \end{bmatrix}$

Să se deseneze graful de tranziţie. Dacă distribuţia iniţială de probabilitate este distribuţia uniformă să se calculeze probabilitatea ca un agent ce navighează de la un nod la altul al grafului conform matricii de tranziţie să parcurgă următoarea traiectorie: $4, 3, 3, 2, 1, 3, 4$.

Să se determine apoi probabilitatea ca la doilea pas în graf agentul să fie în nodul 4.

Transpusa matricii Q are vectorul propriu $v = [0.13793, 0.24631, 0.29557, 0.3202]^T$, corespunzător valorii proprii 1. Ce reprezintă vectorul v pentru lanţul Markov şi ce interpretare au coordonatele sale?

5. La o celulă a unui sistem de comunicaţie wirelles apelurile sosesc conform unui proces Poisson de rată 4 apeluri pe minut. Să se calculeze:

• probabilitatea ca în primele 5.5 minute să nu sosească nici un apel;

• probabilitatea ca în primele 180 secunde să se înregistreze 5 apeluri, iar în primele 300 de secunde 8 apeluri (atenţie, la a doua întrebare se cere probabilitatea unui singur eveniment $A \cap B$)).

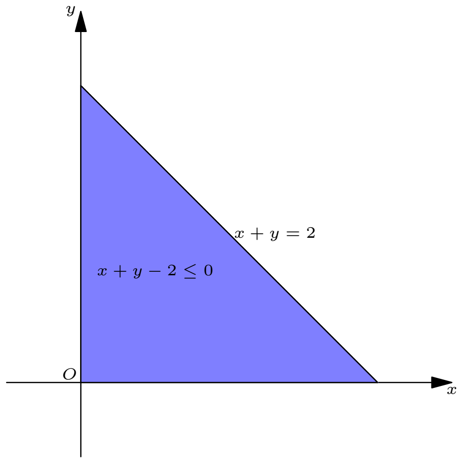
• Care este lungimea medie (în secunde) a intervalului de timp între două apeluri?

Indicaţie: vezi problemele rezolvate din tema relativ la procesul Poisson.

6. Durata de viaţa a unor chip-uri VLSI fabricate de producătorul XX este normal distribuită de medie $m = 5\cdot 10^6$ ore şi abatere standard $\sigma = 5\cdot 10^5$ore. Dell cere ca cel puţin 95% dintr-un lot de chip-uri să aibă durata de viaţa mai mare de $4 \cdot 10^4$ore. Îndeplinesc chip-urile produse de XX condiţia de acceptare a Dell, pentru a încheia contractul de furnizare chip-uri?

7. Variabilele aleatoare X, Y sunt independente şi uniform distribuite, respectiv pe intervalele $[-1, 2]$, $[3, 5]$. Ce distribuţie de probabilitate are vectorul aleator (X, Y)? Justificaţi răspunsul! Scrieţi pseudocodul de simulare a vectorului (X, Y). Care este probabilitatea ca simulând vectorul (X, Y) să fie returnat un punct din pătratul $[0, 1] \times [3, 4]$?

8. Un vector aleator (X, Y) este uniform distribuit pe domeniul $D = \{(x, y) \,|\, x + y - 2 \leq 0, x \in [0, 2], y \in [0, 2]\}$.



• Determinaţi densitatea de probabilitate, $f_{X,Y}$, a vectorului aleator (X, Y) şi probabilitatea $P((X, Y) \in D$.

• Daţi pseudocodul de simulare a vectorului aleator (X, Y).

Indicaţie: se aplică definiţia unui vector aleator uniform distribuit pe un domeniu mărginit, D, şi anume densitatea lui de probabilitate este egală cu $1/aria(D)$, pentru $(x, y) \in D$ şi 0 în rest.

Pentru a scrie pseudocodul de simulare prin metoda respingerii, trebuie să formulăm condiţia ca un punct uniform distribuit pe cel mai mic pătrat ce-l include pe D, $[0, 2] \times [0, 2]$, să nu aparţină lui D:

```
do{
...
}while( x+y-2>0);
return (x,y);
```

9. Generatorul Mesenne–Twister atunci când generează consecutiv n numere uniform distribuite, acestea sunt practic observaţii asupra a n variabile aleatoare i.i.d. uniform distribuite. $U_1, U_2, \ldots, U_n$ ~Unif[0,1). Aflaţi funcţia de repartiţie a variabilei $Y = \max(U_1, U_2, \ldots, U_n)$.

Indicaţie: Funcţia de repartiţie a lui Y este:

$$F_Y(x) = P(Y \leq x) = P(\max(U_1, U_2, \ldots, U_n) \leq x) = P(U_1 \leq x, U_2 \leq x, \ldots, U_n \leq x)$$

Ultima egalitate este adevărată, pentru ca daca maximul este mai mic sau egal cu x, atunci si $(U_1 \le x)$ si $(U_2 \le x)$, $\ldots, (U_n \le x)$. Folosiţi apoi faptul că variabilele $U_1, U_2, \ldots, U_n$ sunt independente.

10. Tensiunea receptată într-un sistem de comunicaţie este suma a două variabile aleatoare $V = X + Y$, unde X ~Bernoulli(p) este mesajul aleator, iar $Y \sim N(0,1)$ este zgomotul. Presupunând că X şi Y sunt independente să se determine funcţia de repartiţie a lui V, condiţionată de $X = b$, $b \in \{0,1\}$

Indicaţie: $P(V \le x | X = b) = P(b + Y \le x) = P(Y \le x - b)$.

11. Vectorul aleator (X,Y) este uniform distribuit pe un dreptunghi $[-1,3] \times [-2,5]$.

a) Să se scrie expresia analitică a densităţii sale de probabilitate, şi să se calculeze $P((X,Y) \in D)$, unde D este domeniul triunghiular ale cărui vârfuri sunt $O(0,0)$ $A(3,0)$, $B(0,5)$.

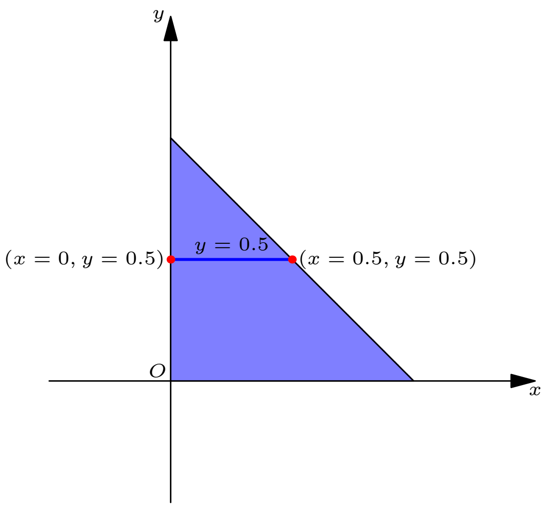
b) Scrieţi pseudocodul de generare a $n = 1000$ puncte în triunghiul ABC.

12. Un vector aleator (X,Y) are densitatea de probabilitate:

$$f_{X,Y}(x,y) = \begin{cases} 6x & x \ge 0, y \ge 0, x + y \le 1 \\ 0 & \text{în rest} \end{cases}$$

a) Să se determine densitatea marginală f_Y.

b) Să se calculeze densitatea condiţionată a variabilei $(X|Y = 0.5)$ şi valoarea ei medie.



a)

$$f_Y(y) = \begin{cases} \int_{-\infty}^{\infty} f_{X,Y}(x,y)dx = \int_0^{1-y} 6x dx & y \in [0,1] \\ 0 & \text{în rest} \end{cases}$$

Explicaţie pentru limitele de integrare: pentru fiecare y fixat în $[0,1]$, x merge de la 0 până "întâlneşte" dreapta $x + y = 1$, adică până $x = 1 - y$. b)

$$f(x|y = 0.5) = \begin{cases} \dfrac{f_{X,Y}(x, 0.5)}{f_Y(0.5)} & \text{pt. } x \in [0, 0.5] \\ 0 & \text{în rest} \end{cases}$$

În lungul dreptei $y = 0.5$, x-ul merge de la 0 la 0.5 (vezi figura).

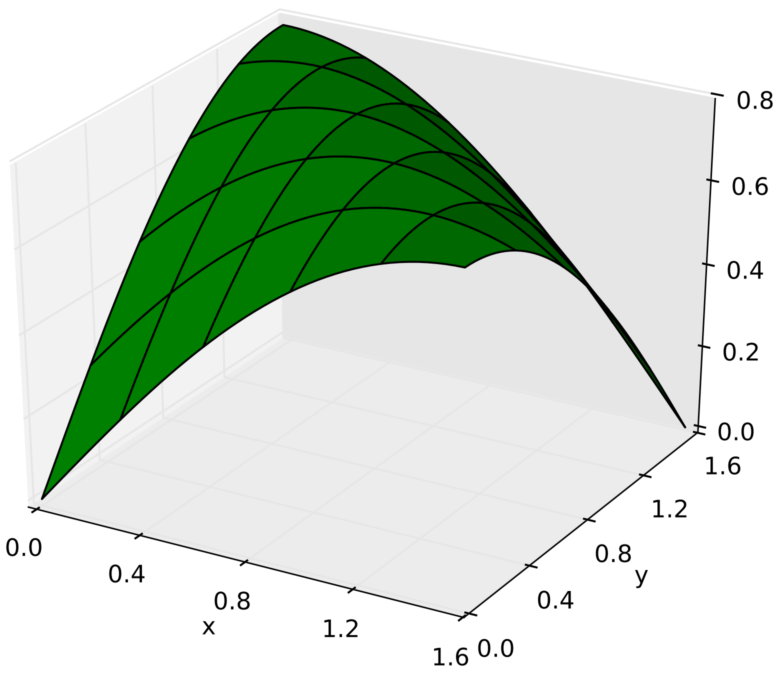
Valoarea medie a variabilei condiţionate, $(X|Y = 0.5)$ se calculează ca pntru orice variabilă aleatoare continuă şi anume:

$$M(X|Y = 0.5) = \int_{-\infty}^{\infty} x f(x|0.5) dx = \int_{0}^{0.5} x f(x|0.5) dx$$

13. Vectorul aleator (X, Y) are densitatea de probabilitate:

$$f(x, y) = \begin{cases} \frac{1}{a}\sin(x + y) & x \in [0, \pi/2], y \in [0, \pi/2] \\ 0 & \text{în rest,} \end{cases}$$

unde $a = 2 - \sqrt{2}/2$. Graficul densităţii este ilustrat în figura de mai jos.



a) Calculaţi $P(X \geq \pi/4, Y \leq \pi/4)$ (indicaţie $\int_a^b (\sin(x + y)dx = -\cos(x + y)|_a^b = -\cos(a + y) - cos(b + y)$, etc;

Să se determine densitatea marginală f_X şi densitatea distribuţiei condiţionate, $f(y|x = \pi/4)$.

14. O variabilă aleatoare $X = -2 + 1.2Z$, unde $Z \sim N(0, 1)$. Ce distribuţie de probabilitate are X? calculaţi $P(-1 < X < 1.5)$.

15. Să se determine în funcţie de repartiţia Φ a distribuţiei normale standard $N(0,1)$, funcţia de repartiţie a variabilei aleatoare $Y = 5X$, unde $X \sim N(1, \sigma = 0.7)$.

Indic:

$$F_Y(x) = F_{5X}(x) = P(5X \leq x) = P(X \leq x/5) = \cdots$$

16. Timpul cât îi ia unui calculator să se conecteze la un server este aleator şi are distribuţia de probabilitate, $T \sim N(m = 3.3sec, \sigma = 0.66sec)$.

a) Un calculator se zice că are conexiune rapidă dacă se conectează în mai puţin de 2.5 sec. Care este procentul de calculatoare ce fac parte din această categorie?

b) Care este timpul minim de conexiune pentru cele mai lente 5% calculatoare?

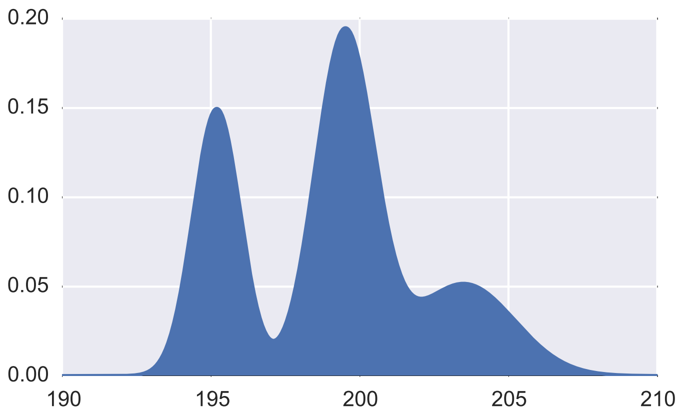
Indicaţie:

a) $P(T < 2.5) =???$; b) $P(T > t) = 0.05$

17. O clasă de imagini color folosite ca date de antrenament pentru un algoritm de recunoaştere a formelor are distribuţia de probabilitate a culorilor de bază R, G, B, o mixtura gaussiana 1D:

$$f = 0.3f_R + 0.48f_G + 0.22f_B$$

unde f_R, f_G, f_B, sunt respectiv densităţile distribuţiilor normale: $N(m = 195.2, \sigma = 0.8)$, $N(199.5, 1.0)$, $N(203.5, 1.7)$ (media pt fiecare reprezintă numarul mediu de pixeli coloraţi cu acea culoare).



Să se calculeze probabilitatea ca o imagine să aibă mai mult de 200 de pixeli din această mixtură.

Indicaţie: Notăm cu Im variabila aleatoare ce are ca distribuţie a culorilor, mixtura f. Din defniţia mixturii (vezi cursul aferent) avem că funcţia de repartiţie a mixturii este:

$$F_{Im}(x) = 0.3F_R(x) + 0.48F_G(x) + 0.22F_B(x)$$

Atunci probabilitatea cerută este:

$$P(Im > 200) = 1-P(Im \leq 200) = 1-F_{Im}(200) = 1-(0.3F_R(200)+0.48F_G(200)+0.22F_B(200))$$

Fiecare funcţie de repartiţie F_R, F_G, F_B fiind reparţia unei distribuţii normale se calculează cu ajutorul repartiţie normale standard $N(0,1)$, Φ.

18. Două variabile aleatoare X, Y au o dependenţa liniară $Y = -3X + 1$. a) Cât este coeficientul de corelaţie al celor două variabile aleatoare? b) Calculaţi covarianţa $cov(X,Y)$ ştiind că $\sigma^2(X) = 0.64$ şi $\sigma^2(Y) = 1.21$.

19. Un semnal S normal distribuit, $S \sim N(1, \sigma = 1.2)$, este transmis printr-un canal de comunicaţie şi este receptat ca o variabilă aleatoare $R = S + Z$, unde $Z \sim N(0, 0.25)$ este zgomotul independent de S, adăugat semnalului prin canal. Calculaţi media şi dispersia semnalului receptat, R şi $cov(S,R)$.

Indicaţie: În general $cov(X, Y + Z) = cov(X,Y) + cov(X,Z)$, pt că:

$$\begin{aligned} cov(X, Y+Z) &= M(X(Y+Z)) - M(X)M(Y+Z) = \\ &= M(XY) + M(XZ) - M(X)M(Y) - M(X)M(Z) = \\ &= M(XY) - M(X)M(Y) + M(XZ) - M(X)M(Z) = cov(X,Y) + cov(X,Z) \end{aligned}$$