Probleme de antrenament pentru lucrarea I

1. Numererele maşinilor dintr-un judeţ conţin 5 cifre. Poliţia generează cele 5 cifre independent şi uniform (cu aceeaşi probabilitate).

a) Calculaţi probabilitatea de a primi un număr cu 5 cifre distincte. (nu va gândiţi că numărul a fost atribuit deja).

b) Calculaţi probbailitatea ca numărul să contină două cifre egale.

c) Care este probabilitatea ca toate cele 5 cifre să fie egale?

2. Câte numere pare de 4 cifre pot fi formate din cifrele $0, 1, 2, 5, 6, 9$, dacă fiecare cifră apare o singură dată într-un număr?

Indicaţie: Exprimăm mulţimea numerelor ce trebuie formate după regula precizată, ca reuniune a două submulţimi disjuncte:

$$A = \{\overline{xyz0}, x \neq 0\}, \quad B = \{\overline{xyzu}, x \neq 0, u \neq 0\}$$

şi numărăm câte elemente are fiecare submulţime.

3. Un calculator are trei hard–diskuri numerotate $0, 1, 2$. Când bootează el alege la întâmplare un harddisk pentru a stoca fişierele temporare. Care este probabilitatea să fie ales hardul 0 sau 2?

4. Este adevărat sau fals că $P(A \backslash B) = P(A) - P(B)$? Pentru a răpunde corect exprimaţi A ca reuniune de $A \setminus B$ şi $A \cap B$ (desenaţi diagrame Venn pentru A şi B si discutaţi în funcţie de poziţia lui B faţă de A).

5. Dacă A şi B sunt evenimente independente şi $P(A) = 0.3, P(B) = 0.42$, calculaţi $P(A \cup B)$. Scrieţi formula de condiţionare iterată a evenimentelor $A_1, A_2, \ldots, A_n$ şi demonstraţi-o pentru $n = 2$ şi $n = 3$.

Dacă o propozitie, $f = w_1 w_2 w_3 w_4$, este succesiunea a 4 cuvinte, scrieţi folosind formula conditionarii iterate care este probabilitatea ca într-un corpus (baza de texte) să existe această propoziţie.

Dacă $P(A) = 0.3$, $P(B) = 0.3$ şi $P_B(A) = 0.3$ deduceţi care din afirmaţiile de mai jos sunt adevărate:

a) A şi B sunt mutual exclusive;
b) $P(A \cap B) = 0$;
c) A şi B sunt independente;
d) toate proprietăţile de la a)–c);
e) nici una dintre a)–d)

6. Într-o parcare circulară există 15 locuri şi numerotate $1, 2, \ldots, 15$. Când ajungi în parcare găseşti 5 locuri libere. Care este probabilitatea ca acestea să fie unul după altul

7. **Background pt problemă:** În transmiterea informaţiei prin reţele digitale de comunicaţie pachetele de informaţie trec printr-un număr de routere în loc să fie transmise direct de la "expeditor" la "destinatar". În limbaj de reţele se spune că fiecare router cauzează un *hop*. Numărul de *hop*-uri este astfel nr de routere prin care trece un pachet. Fiecare *hop* induce o întârziere în transmitere.

Şi acum problema: Într-o reţea de comunicaţie fiecare pachet traveseazā două subreţele. În prima subreţea pachetul poate avea 1, 2, 3, 4, 5 *hop*-uri, egal probabile, iar în a doua 4 hop-uri egal probabile. Cu alte cuvinte ruta unui pachet este aleatoare, nu fixă. Numărul de hop-uri dintr-o sub-reţea este independent de cel din cea de-a doua. Fie X variabila aleatoare care dă numărul total de hop-uri prin care traversează un pachet de la sursă la destinaţie.

a) Să se determine distribuţia de probabilitate a variabilei X, adică să se calculeze probabilităţile $P(X = k)$, unde k ia valori într-o multime pe care o deduceţi voi pe bază de raţionament.

b) Dacă fiecare hop cauzează o întârziere de 1 milisecundă, care este întârzierea medie a unui pachet cauzată de hopurile din cele două subreţele?

Indicaţie: X=Y+Z, unde Y ia valorile 1,2,3,4,5, cu 1/5 si Z ia valorile 1,2,3,4 cu 1/4.

8. Să se calculeze probabilitatea ca într-un corpus (bază de texte digitale) să existe succesiunea de cuvinte *students attended courses*, ştiind că în notaţia $w_1 =$*students*, $w_2 =$*attended*, $w_3 =$*courses*, au fost estimate din texte următoarele probabilităţi: $P(w_1) = 0.35, P(w_2|w_1) = 0.2, P(w_3|w_1w_2) = 0.15$

9. Probabilitatea de a primi adresa IP de la ISP-ul (Internet Service Provider) la care ai abonament este în orice moment 0.85. Care este probabilitatea ca primele 3 încercări să se soldeze cu eşec? Să se calculeze probabilitatea de a primi IP-ul la prima încercare.

10. La bordul unui avion sunt instalate 4 calculatoare de control ce funcţionează în paralel şi independent. În fiecare etapă critică a zborului aplicaţia inteligentă de pe fiecare calculator determină acţiunea ce se impune. Probabilitatea ca aplicaţia de pe un calculator să afişeze o decizie nepotrivită este de 10^{-4}. Dacă X este variabila aleatoare ce indică numărul de calculatoare ale căror aplicaţii iau decizie nepotrivită într-o situaţie critică, să se calculeze $P(X = 1)$ (explicaţi cum se exprimă evenimentul $(X = 1)$ şi apoi calculaţi probabilitatea lui).

Indicaţie: Notăm cu C_i evenimentul ca sistemul i să ia decizie nepotrivită, $i = \overline{1, 4}$. Exprimaţi evenimentul $(X = 1)$ în funcţie de evenimente compuse din C_i şi opusele lor.

11. O variabilă aleatoare discretă X are distribuţia de probabilitate:

$$P_X(x) = \begin{cases} \dfrac{|x|}{c} & \text{pt } x = -2, -1, 0, 1, 2 \\ 0 & \text{înrest} \end{cases}$$

Să se scrie distribuţia lui X sub formă de tabel, să se determine constanta c şi să se calculeze, $M(X)$ şi distribuţia de probabilitate a variabilei aleatoare $Y = (X - M(X))^2$.

12. Un digipass folosit în online-banking are un generator de cifre uniform distribuite pe mulţimea $\{0, 1, 2, 3, 4, 5, 6, 7, 8, 9\}$. Prin apăsarea succesivă a unui buton, se afişează cifre generate în mod independent. Dacă butonul este tastat de 4 ori consecutiv, care este probabilitatea să se genereze cifrele 1,3,2, 7, în orice ordine?

13. Studenţilor din anul întâi li se oferă 3 şanse de a promova examenul de probabilităţi. Fie variabila aleatoare X ce asociază fiecărui student promovat numărul (1, 2 sau 3) al încercării în care acesta a promovat examenul. Distribuţia de probabilitate a variabilei X dedusă din anii precedenţi este definită prin $p_X(k) = \dfrac{0.4^{k-1}0.6}{0.936}$, $k = 1, 2, 3$. Care este procentul de studenţi ce au promovat în cel puţin două încercări?

14. Inginerii software schimbă compania la care lucrează după aproximativ 2 ani, pentru a ocupa in altă paerte o poziţie superioara.

O companie este interesată de relaţia (dacă ea există) între experienţa în muncă a angajaţilor săi şi poziţia lor profesională în companie. Fie X v.a. ce dă numărul de locuri de muncă (inclusiv prezentul loc) pe care l-au avut angajaţii companiei. Presupunem că X ia valorile $1, 2, 3, 4$. Variabila Y, luând valorile $1, 2, 3, 4$, indică poziţia profesional–ierarhică în companie (1 codifică poziţie joasă, 4 codifică poziţia superioară). Dacă este selectat un angajat la întâmplare există 16 rezultate posibile pentru vectorul (X, Y). Presupunem că distribuţia de probabilitate a acestuia este:

			Y		
		1	2	3	4
	1	0.03	0.05	0.1	0.12
X	2	0.05	0.06	0.08	0.07
	3	0.07	0.06	0.06	0.02
	4	0.07	0.09	0.05	0.02

a) Să se calculeze probabilitatea ca un angajat selectat aleator să fie la al treielea loc de muncă.

b) Stiind că un angajat are poziţia ierarhică 2, adică $Y = 2$, calculaţi probabilitatea ca el să fie la al 3 lea loc de muncă.

15. Pentru a dezvolta o aplicaţie de tip *troubleshooting detection* manipulaţi trei variabile aleatoare discrete, X, Y, Z ce ia fiecare, respectiv m, n, q valori.

a) Ce dimensiune are tabloul ce stochează distribuţia de probabilitate a vectorului aleator (X, Y, Z)?

b) Dacă se dau ca date de intrare probabilităţile $p_{ijk} = P(X = i, Y = j, Z = k)$, deduceţi cum calculaţi $\pi_k = P(Z = k)$.

c) Descrieţi algoritmul de calcul al probabilităţii condiţionate $P(Z = k|X = i)$.

16. Fie X variabila aleatoare ce ia valoarea 1 dacă evenimentul A se produce şi 0 în caz contrar. Care este distribuţia de probabilitate a lui X? Calculaţi valoarea medie şi dispersia, ştiind că $P(A) = 0.48$.

17. Variabilele aleatoare X şi Y iau fiecare valorile $1, 2, 3$, iar distribuţia de probabilitate a vectorului aleator (X, Y) este dată în matricea:

$$P = (p_{ij}) = \begin{pmatrix} 0.125 & 0 & 0.125 \\ 0 & 0.5 & 0 \\ 0.125 & 0 & 0.125 \end{pmatrix}$$

Să se determine distribuţiile marginale $p_X(i) = P(X = i)$, $p_Y(j) = P(Y = j)$, $i = \overline{1,3}$ şi distribuţia variabilei $(Y|X = 3)$.

Să se calculeze $P(X < Y)$ şi să se deducă dacă X şi Y sunt independente.

18. Un calculator on-line are 4 linii de comunicaţie de intrare cu caracteristicile din tabelul de mai jos:

Linia	Procentul din trafic pe fiecare linie	Proporţia de mesaje fără eroare pe fiecare linie
1	0.4	0.998
2	0.3	0.999
3	0.1	0.997
4	0.2	0.992

Care este probabilitatea ca un mesaj selectat la întâmplare să fi sosit fără eroare?

Indicaţie: Fie L_i evenimentul ”mesaj intrat pe linia i”, $i = \overline{1,4}$ şi F evenimentul ”mesaj intrat fără eroare”. Evident că L_1, L_2, L_3, L_4 sunt evenimente mutual exclusive două câte două. Pentru calculul probabilităţii $P(F)$ se foloseşte... formula probabilităţii totale.

19. Ştiind că variabillel aleatoare X şi Y iau valorile $1, 2, 3$, fiecare şi că distribuţia de probabilitate a lui Y este dată de $p_Y = [1/6, 2/3, 1/6]$, iar distribuţiile de probabilitate ale variabilelor $(X|Y = j)$, $j = 1, 2, 3$, sunt date respectiv de coloanele matricii următoare

$$(\pi_{ij}) = \begin{pmatrix} 0.25 & 0.25 & 0.25 \\ 0.5 & 0.5 & 0.5 \\ 0.25 & 0.25 & 0.25 \end{pmatrix}$$

să se deducă distribuţia de probabilitate a vectorului aleator (X, Y).

Două probleme de interviu la Google:

20. In a country in which people only want boys, every family continues to have children until they have a boy. If they have a girl, they have another child. If they have a boy, they stop. What is the proportion of boys to girls in the country?

21. If the probability of observing a car in 30 minutes on a highway is 0.95, what is the probability of observing a car in 10 minutes (assuming constant default probability)?

22. • Ce contorizează o variabilă aleatoare ce are distribuţia Poisson? Precizaţi distribuţia de probabilitate a unei v.a. Poisson $X \sim$Poiss(λ) precum (valorile pe care le ia şi probabilităţile cu care ia fiecare valoare). Ce relaţie există între $M(X)$, $\sigma(X)$?.

• La un nod internet, pachetele de informaţie sosesc cu o rată de 100 pachete pe minut.

Care este probabilitatea ca în nodul respectiv să nu sosească nici un pachet în 5 secunde?

23. Un fişier al unei baze de date conţine 1 milion de înregistrări ce ocupă un disc de stocare cu o densitate de 10/bloc. O actualizare săptămânală modifică 3% din fişier şi modificările sunt distribuite uniform în mulţimea înregistrărilor. Sistemul trebuie să rescrie fiecare bloc odată cu modificarea uneia din înregistrările ce le stochează.

a) Care este probabilitatea ca sistemul să rescrie blocul i în timpul unei actualizări?

b) Care este numărul mediu de blocuri rescrise?

Rezolvare: a) Fie $1 \leq j \leq 10$ şi definim variabila aleatoare Y_{ij} ce ia valoarea 1 dacă înregistrarea j din blocul i este modificată şi valoarea 0 în caz contrar. Notâm cu $X_i = \sum_{j=1}^{10} Y_{ij}$ v.a. ce dă numărul de înregistrări modificate în blocul i. Din modul de definire Y_{ij} este o v.a. Bernoulli de parametru $p = 0.03$ (deoarece probabilitatea succesului, adică a unei modificări este de $3\% = 0.03$). Prin urmare $X_i \sim \text{Bin}(n = 10, p = 0.03)$. Sistemul trebuie să rescrie blocul i dacă $X_i > 0$.

$$P(X_i > 0) = 1 - P(X_i = 0) = 1 - C_{10}^0 (0.03)^0 (1 - 0.03)^{10} = 0.2626$$

Observăm că rezultatul este independent de eticheta i a blocului. Deci fiecare bloc este rescris cu o probabilitate de 0.2626.

b) Fişierul conţine 1000000 nregistrări cu o densitate de 10/bloc. Rezultă deci că există 100 000 blocuri. Considerăm v.a. Bernoulli B_i, cu parametrul $p = 0.2626$, $B_i = 1$ dacă blocul i este rescris şi 0 în caz contrar. Variabila aleatoare $Z = \sum_{i=1}^{100000} B_i$ are distribuţia binomială $\text{Bin}(n = 100000, p = 0.2626)$. Deci numărul mediu de blocuri rescrise este:

$$M(Z) = np = 100000 \cdot 0.2626 = 26260$$

24. Fie S mulţimea permutărilor şirului de caractere "NASOL". Dacă pe această mulţime avem distribuţia uniformă, care este probabilitatea ca un algoritm ce selectează la întâmplare o permutare să returneze "LASON"?

25. Ce este un experiment Bernoulli? Ce contorizează o variabilă aleatoare $X \sim \text{Bin}(n, p)$ asociată unui experiment Bernoulli cu n încercări? Dar $Y \sim \text{Geom}(p)$?

Dacă într-o încercare Bernoulli codificăm cu 1 succesul şi cu 0 eşecul şi $X_1, X_2, \ldots, X_n$ sunt n încercări, în care probabilitatea eşecului este $q = 0.1$, calculaţi $P(X_1, X_2, \ldots, X_n) = (b_1, b_2, \ldots, b_n)$, unde $b_i \in \{0, 1\}$, $i = \overline{1, n}$.

26. Fie X, Y două variabile aletoare Bernoulli, independente ce iau valorile 1 şi 0 cu probabilităţile 2/3, respectiv, 1/3. Să se determine distribuţia de probabilitate sumei modulo 2, $X+Y$.

27. Protocolul TCP/IP se foloseşte pentru a transmite pachete de informaţie pe canale internet. Receptorul TCP verifică pachetele pentru a identifica eventualele erori de transmitere. Dacă se detectează o eroare într-un pachet se cere retransmiterea pachetului respectiv. Ştiind că probabilitatea erorii pe pachet este de 0.01, şi că transmiterea eronată a unui pachet este independentă de calitatea transmiterii celorlalte pachete, să se calculeze (şi să se precizeze variabila aleatoare implicată în calculele ce le faceţi, mulţimea de valori şi distribuţia ei):

(i) Probabilitatea ca dintr-o succesiune de transmiteri al treielea pachet e primul care necesită retransmitere.

(ii) Probabilitatea să fie necesară o retransmitere de pachet într-o succesiune de 10 pachete transmise.

(iii) Numărul mediu de retransmiteri într-o succesiune de 1000 de pachete.

28. Setul de date de intrare pentru un program conţine 60% date de tipul I şi 40% date de tipul II. Intrările pot produce mesaje de atenţionare în procent de 25% (cele de tipul I), respectiv 15% (cele de tipul II). Dacă după rulare este afişat un mesaj de atenţionare, care este probabilitatea ca el să fie cauzat de datele de intrare de tipul I?