Tema 1

Soluții

Exercițiul 1

Zece cartonașe pe care sunt scrise cifrele $0, 1, 2, 3, 4, 5, 6, 7, 8$ și 9 sunt puse într-un bol. Cinci cartonașe sunt extrase la întâmplare și sunt așezate în rând în ordinea extragerii. Care este probabilitatea ca numărul de cinci cifre extras să fie divizibil cu 495 ?

Să presupunem că numerele extrase (în ordinea extragerii) sunt a, b, c, d și e. Atunci numărul format cu cifrele extase este

$$N = 10000a + 1000b + 100c + 10d + e.$$

Este ușor de observat că numărul cazurilor posibile în acest experiment este $10 \times 9 \times 8 \times 7 \times 6 = 30240$. Pentru a determina numărul cazurilor favorabile să observăm că $495 = 5 \times 9 \times 11$ deci N trebuie să fie divizibil cu 5, cu 9 și cu 11. Cum

$$10000a + 1000b + 100c + 10d$$

este întotdeauna multiplu de 5, găsim că N este divizibil cu 5 dacă și numai dacă $e = 0$ sau $e = 5$. Mai mult cum N se poate scrie sub forma

$$N = (9999a + 999b + 99c + 9d) + (a + b + c + d + e)$$

deducem că el este divizibil cu 9 dacă și numai dacă suma cifrelor lui este multiplu de 9 ($a + b + c + d + e$ este divizibil cu 9). De asemenea dacă scriem

$$N = (9999a + 1001b + 99c + 11d) + (a - b + c - d + e)$$

atunci prima paranteză este multiplu de 11 iar pentru ca N să fie divizibil cu 11 este necesar ca $a-b+c-d+e$ să fie divizibil cu 11.

Cum cifrele sunt distincte avem că

$$\begin{aligned} a + b + c + d + e &\geq 0 + 1 + 2 + 3 + 4 = 10, \\ a + b + c + d + e &\leq 9 + 8 + 7 + 6 + 5 = 35 \end{aligned}$$

prin urmare $a+b+c+d+e$ este divizibil cu 9 dacă și numai dacă $a+b+c+d+e \in \{18, 27\}$. Vom considera cele două cazuri separat.

1. Fie $a + b + c + d + e = 18$. Deoarece $a - b + c - d + e = 18 - 2b - 2d$ deducem că $a - b + c - d + e$ este par iar

$$|a - b + c - d + e| < 18$$

astfel $a-b+c-d+e = 0$ deoarece 0 este singurul multiplu de 11 par mai mic ca 18. Găsim că $a+c+e = b+d = 9$.

Dacă $e = 0$ atunci $a + c = b + d = 9$ deci perechile (a, c) și (b, d) pot fi extrase din cele 8 perechi $(1, 8), (2, 7), \dots, (8, 1)$. O dată ce am ales (a, c) mai rămân 6 perechi pentru (b, d) conducând astfel la un total de $8 \cdot 6 = 48$ de numere.

Dacă $e = 5$ atunci $a + c = 4$ și $b + d = 9$ prin urmare (a, c) poate fi una din perechile $(0, 4), (4, 0), (1, 3), (3, 1)$. Dacă $(a, c) \in \{(0, 4), (4, 0)\}$ atunci $(b, d) \in \{(1, 8), (2, 7), (3, 6), (6, 3), (7, 2), (8, 1)\}$ iar dacă $(a, c) \in \{(1, 3), (3, 1)\}$ atunci $(b, d) \in \{(0, 9), (2, 7), (7, 2), (9, 0)\}$ de unde obținem un total de $2 \cdot 6 + 2 \cdot 4 = 20$ de numere. Numărul de cazuri favorabile pentru scenariul 1 este 68.

2. Fie $a + b + c + d + e = 27$. În acest caz $a - b + c - d + e$ este impar iar din

$$27 > a - b + c - d + e > 27 - 2 \cdot 9 - 2 \cdot 8 = -7$$

găsim că $a - b + c - d + e = 11$ (singurul multiplu de 11 impar în mulțimea de valori posibile). Deoarece $11 = a - b + c - d + e = 27 - 2b - 2d$ deducem că $b + d = 8$ și respectiv $a + c + e = 19$. În acest caz nu putem avea $e = 0$ pentru că asta ar implica $a + c = 19$ dar $a + c < 9 + 8 = 17$ ceea ce conduce la $e = 5$. Avem $a + c = 14$ de unde $(a, c) \in \{(8, 6), (6, 8)\}$ iar $(b, d) \in \{(1, 7), (7, 1)\}$. Am obținut $2 \cdot 2 = 4$ numere favorabile în acest caz.

Combinând cele două cazuri avem $68 + 4 = 72$ cazuri favorabile iar probabilitatea căutată este $\frac{72}{30240} \approx 0.0023$.

Exercițiul 2

Se dorește verificarea fiabilității unui test de pentru depistarea nivelului de alcool al automobiliștilor. În urma studiilor statistice pe un număr mare de automobiliști, s-a observat că în general 0.5% dintre aceștia depășesc nivelul de alcool autorizat. Niciun test nu este fiabil 100%. Probabilitatea ca testul considerat să fie pozitiv atunci când doza de alcool autorizată este depășită precum și probabilitatea ca testul să fie negativ atunci când doza autorizată nu este depășită sunt egale cu $p = 0.99$.

1. Care este probabilitatea ca un automobilist care a fost testat pozitiv să fi depășit în realitate nivelul de alcool autorizat ?
2. Cât devine valoarea parametrului p pentru ca această probabilitate să fie de 95% ?
3. Un polițist afirmă că testul este mai fiabil sambăta seara (atunci când tinerii ies din cluburi). Știind că proporția de automobiliști care au băut prea mult în acest context este de 30%, determinați dacă polițistul are dreptate.

Considerăm evenimentele următoare:

- $A = \{\text{testul considerat este pozitiv}\}$
- $B = \{\text{automobilistul a depășit nivelul de alcool autorizat}\}$

1. Din ipoteză știm că $\mathbb{P}(B) = 0.005$, $\mathbb{P}(A|B) = \mathbb{P}(A^c|B^c) = 0.99$. Vrem să găsim probabilitatea $\mathbb{P}(B|A)$. Avem

$$\begin{aligned}
\mathbb{P}(B|A) &= \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)} = \frac{\mathbb{P}(A|B)\mathbb{P}(B)}{\mathbb{P}(B)} = \frac{\mathbb{P}(A|B)\mathbb{P}(B)}{\mathbb{P}(A|B)\mathbb{P}(B) + \mathbb{P}(A|B^c)\mathbb{P}(B^c)} \\
&= \frac{\mathbb{P}(A|B)\mathbb{P}(B)}{\mathbb{P}(A|B)\mathbb{P}(B) + [1 - \mathbb{P}(A^c|B^c)]\,(1 - \mathbb{P}(B))} = \frac{0.99 \times 0.005}{0.99 \times 0.005 + 0.01 \times 0.995} \approx 0.332.
\end{aligned}$$

2. Căutăm p așa încât $\mathbb{P}(B|A) = 0.95$. Am văzut că $\mathbb{P}(B|A) = \frac{p\mathbb{P}(B)}{p\mathbb{P}(B) + (1-p)(1-\mathbb{P}(B))}$ de unde

$$p = \frac{(1-\mathbb{P}(B))\mathbb{P}(B|A)}{(1-\mathbb{P}(B))\mathbb{P}(B|A)+(1-\mathbb{P}(B|A))\mathbb{P}(B)} = \frac{0.995\times 0.95}{0.995\times 0.95 + 0.05\times 0.005} \approx 0.99973.$$

3. Știm că $\mathbb{P}(B) = 0.3$, prin urmare $\mathbb{P}(A) = \mathbb{P}(A|B)\mathbb{P}(B) + \mathbb{P}(A|B^c)\mathbb{P}(B^c) = 0.99\times 0.3 + 0.01\times 0.7 \approx 0.304$ și

$$\mathbb{P}(B|A) = \frac{\mathbb{P}(A|B)\mathbb{P}(B)}{\mathbb{P}(A|B)\mathbb{P}(B)+\mathbb{P}(A|B^c)\mathbb{P}(B^c)} \approx 0.9769,$$

de unde tragem concluzia că testul este mult mai fiabil în această situație.

Exercițiul 3

Efectuăm aruncări succesive a două zaruri echilibrate și suntem interesați în găsirea probabilității evenimentului ca suma 5 (a fețelor celor două zaruri) să apară inaintea sumei 7. Pentru aceasta presupunem că aruncările sunt *independente*.

1. Calculați pentru început probabilitatea evenimentului E_n: *în primele* $n-1$ *aruncări nu a apărut nici suma* 5 *și nici suma* 7 *iar în a* n*-a aruncare a apărut suma* 5. Concluzionați.
2. Aceeași întrebare, dar înlocuind 5 cu 2.

1. Considerăm evenimentele următoare:
   - $A_i = \{$suma celor două zaruri la cea de- a i-a aruncare este 5$\}$
   - $B_i = \{$suma celor două zaruri la cea de- a i-a aruncare este 7$\}$

   Evenimentul E_n se scrie

$$E_n = (A_1^c \cap B_1^c) \cap (A_2^c \cap B_2^c) \cap \cdots \cap (A_{n-1}^c \cap B_{n-1}^c) \cap A_n.$$

   Aplicand independenţa avem că

$$\begin{aligned}
\mathbb{P}(E_n) &= \mathbb{P}((A_1^c \cap B_1^c) \cap (A_2^c \cap B_2^c) \cap \cdots \cap (A_{n-1}^c \cap B_{n-1}^c) \cap A_n) \\
&\overset{indep.}{=} \mathbb{P}(A_1^c \cap B_1^c) \times \mathbb{P}(A_2^c \cap B_2^c) \times \cdots \times \mathbb{P}(A_{n-1}^c \cap B_{n-1}^c) \times \mathbb{P}(A_n) \\
&= \mathbb{P}(A_1^c \cap B_1^c)^{n-1}\mathbb{P}(A_n).
\end{aligned}$$

   Observăm că spaţiul stărilor la cea de-a n-a lansare este $\Omega = \{(i,j)|1 \leq i,j \leq 6\}$ şi probabilitatea ca suma celor două zaruri să fie 5 este $\mathbb{P}(A_n) = \frac{4}{36}$, deoarece cazurile favorabile sunt $\{(1,4),(2,3),(4,1),(3,2)\}$. Obţinem de asemenea că probabilitatea ca suma să nu fie nici 5 şi nici 7 la prima lansare este $\mathbb{P}(A_1^c \cap B_1^c) = \frac{26}{36}$, deoarece situaţiile in care suma este 7 sunt $\{(1,6),(2,3),(3,4),(4,3),(5,2),(6,1)\}$.

   In concluzie, probabilitatea evenimentului

$$A = \{\text{suma 5 (a feţelor celor două zaruri) să apară inaintea sumei 7}\}$$

   este

$$\begin{aligned}\mathbb{P}(A) &= \mathbb{P}\left(\bigcup_{n=1}^{\infty} E_n\right) = \sum_{n=1}^{\infty}\mathbb{P}(E_n) = \sum_{n=1}^{\infty}\left(\frac{26}{36}\right)^{n-1}\frac{4}{36} \\ &= \frac{1}{9}\sum_{n=0}^{\infty}\left(\frac{13}{18}\right)^{n} = \frac{1}{9}\frac{1}{1-\frac{13}{18}} = \frac{2}{5}.\end{aligned}$$

2. Fie F_n evenimentul ce corespunde la: *in primele* $n-1$ *aruncări nu a apărut nici suma* 2 *şi nici suma* 7 *iar in a* n*-a aruncare a apărut suma* 2 şi C_i evenimentul ce corespunde la suma celor două zaruri la cea de-a i-a aruncare este 2. Avem

$$F_n = (C_1^c \cap B_1^c) \cap (C_2^c \cap B_2^c) \cap \cdots \cap (C_{n-1}^c \cap B_{n-1}^c) \cap C_n$$

şi probabilitatea lui $\mathbb{P}(F_n)$ este

$$\begin{aligned}\mathbb{P}(F_n) &= \mathbb{P}((C_1^c \cap B_1^c) \cap (C_2^c \cap B_2^c) \cap \cdots \cap (C_{n-1}^c \cap B_{n-1}^c) \cap C_n) \\ &\overset{indep.}{=} \mathbb{P}(C_1^c \cap B_1^c) \times \mathbb{P}(C_2^c \cap B_2^c) \times \cdots \times \mathbb{P}(C_{n-1}^c \cap B_{n-1}^c) \times \mathbb{P}(C_n) \\ &= \mathbb{P}(C_1^c \cap B_1^c)^{n-1}\mathbb{P}(C_n).\end{aligned}$$

Avem $\mathbb{P}(C_n) = \frac{1}{36}$ (deoarece doar $(1,1)$ ne dă suma 2) şi $\mathbb{P}(C_1^c \cap B_1^c) = \frac{36-7}{36} = \frac{29}{36}$. Prin urmare probabilitatea evenimentului căutat, pe care il notă cu B, este

$$\begin{aligned}\mathbb{P}(B) &= \mathbb{P}\left(\bigcup_{n=1}^{\infty} F_n\right) = \sum_{n=1}^{\infty}\mathbb{P}(F_n) = \sum_{n=1}^{\infty}\left(\frac{29}{36}\right)^{n-1}\frac{1}{36} \\ &= \frac{1}{36}\sum_{n=0}^{\infty}\left(\frac{29}{36}\right)^{n} = \frac{1}{36}\frac{1}{1-\frac{29}{36}} = \frac{1}{7}.\end{aligned}$$

Exercițiul 4

Fie Y o variabilă aleatoare (v.a.) și m mediana ei, i.e. $\mathbb{P}(Y \leq m) \geq \frac{1}{2}$ și $\mathbb{P}(Y \geq m) \geq \frac{1}{2}$. Arătați că pentru orice numere reale a și b așa încât $m \leq a \leq b$ sau $m \geq a \geq b$ avem

$$\mathbb{E}|Y-a| \leq \mathbb{E}|Y-b|.$$

Putem presupune că $\mathbb{E}[|Y|] < \infty$ deoarece în caz contrar am avea $\infty = \mathbb{E}[|Y-a|] \leq \mathbb{E}[|Y-b|] = \infty$.

Să considerăm cazul în care $m \leq a \leq b$. Avem

$$\begin{aligned}\mathbb{E}[|Y-b|] - \mathbb{E}[|Y-a|] &= \mathbb{E}[(b-Y)\mathbf{1}_{\{Y\leq b\}}] + \mathbb{E}[(Y-b)\mathbf{1}_{\{Y>b\}}] - \mathbb{E}[(a-Y)\mathbf{1}_{\{Y\leq a\}}] - \mathbb{E}[(Y-a)\mathbf{1}_{\{Y>a\}}] \\ &= (b-a)\mathbb{E}[\mathbf{1}_{\{Y\leq a\}}] + \mathbb{E}[(a+b-2Y)\mathbf{1}_{\{a<Y\leq b\}}] + (a-b)\mathbb{E}[\mathbf{1}_{\{Y\geq b\}}] \\ &= 2\mathbb{E}[(b-Y)\mathbf{1}_{\{a<Y\leq b\}}] + (b-a)\left[\mathbb{E}[\mathbf{1}_{\{Y\leq a\}}] - \mathbb{E}[\mathbf{1}_{\{Y>a\}}]\right] \\ &\geq (b-a)\left[2\mathbb{P}(Y\leq a) - 1\right] \geq 0\end{aligned}$$

deoarece $\mathbb{P}(Y \leq a) \geq \mathbb{P}(Y \leq m) \geq \frac{1}{2}$. Dacă $m \geq a \geq b$ atunci $-m \leq -a \leq -b$ și $-m$ este mediana lui $-Y$ de unde avem concluzia.

Exercițiul 5

a) Fie X o variabilă repartizată exponențial (de parametru α). Arătași că are loc următoarea relație (proprietatea lipsei de memorie):

$$\mathbb{P}(X > s+t|X > s) = \mathbb{P}(X > t) \tag{1}$$

b) Fie X o variabilă aleatoare care verifică relația (1). Arătați că X este repartizată exponențial.

a) Dacă $X \sim Exp(\alpha)$ atunci densitatea sa este $f(x) = \alpha e^{-\alpha x}\mathbf{1}_{(0,\infty)}(x)$ şi are funcţia de repartiţie $F(x) = (1-e^{-\alpha x})\mathbf{1}_{(0,\infty)}(x)$. Observăm că $\mathbb{P}(X > t) = 1 - \mathbb{P}(X \leq t) = 1 - F(t) = e^{-\alpha t}$, deci

$$\mathbb{P}(X > t+s|X > s) = \frac{\mathbb{P}(X > t+s, X > s)}{\mathbb{P}(X > s)} = \frac{\mathbb{P}(X > t+s)}{\mathbb{P}(X > s)} = \frac{e^{-\alpha(t+s)}}{e^{-\alpha s t}} = e^{-\alpha t} = \mathbb{P}(X > t).$$

Interpretare: Să presupunem că X reprezintă durata de viaţă a unei componente electrice. Ştiind că X a funcţionat deja un timp s ($X > s$), probabilitatea ca X să funcţioneze un timp t suplimentar ($X > t+s$) este egală cu probabilitatea ca X să funcţioneze un timp t. Faptul că a funcţionat deja un timp s nu influenţează probabilitatea să funcţioneze un timp t in plus.

b) Din relaţia

$$\mathbb{P}(X > s+t|X > s) = \frac{\mathbb{P}(X > s+t, X > s)}{\mathbb{P}(X > s)} = \frac{\mathbb{P}(X > s+t)}{\mathbb{P}(X > s)} = \mathbb{P}(X > t)$$

obţinem $\mathbb{P}(X > s+t) = \mathbb{P}(X > s)\mathbb{P}(X > t)$ de unde notand cu $h(t) = \mathbb{P}(X > t)$ avem $h(s+t) = h(s)h(t)$ pentru $s > 0,\ t > 0$.

Pentru a verifica că v.a. X este distribuită exponenţial trebuie să rezolvăm ecuaţia funcţională Cauchy. Observăm pentru inceput că dacă $s = t$ atunci $h(2s) = h^2(s)$ şi prin inducţie avem $h(ks) = h^k(s)$ pentru $k \in \mathbb{N}$. Luand $s = \frac{1}{2}$ avem $h(1) = h^2(\frac{1}{2})$ de unde $h(\frac{1}{2}) = h^{\frac{1}{2}}(1)$ şi pentru $s = \frac{1}{k}$ rezultă că $h(\frac{1}{k}) = h^{\frac{1}{k}}(1)$ (prin aceeaşi argumentare). Combinand rezultatele avem $h(\frac{m}{n}) = h(m\frac{1}{n}) = h^m(\frac{1}{n}) = h^{\frac{m}{n}}(1)$. Prin urmare $h(q) = h^q(1)$ pentru orice $q \in \mathbb{Q}_+$. Dacă $r \in \mathbb{R}_+ - \mathbb{Q}_+$ există un şir $(q_n)_n \subset \mathbb{Q}_+$ aşa incat $q_n \downarrow r$ şi folosind continuitatea la dreapta avem $h(q_n) \downarrow h(r)$ deci $h(r) = a^r$, unde $a = h(1)$. In final am găsit că $h(t) = e^{-t\log\frac{1}{h(1)}}$.

Exercițiul 6

Arătați că:

a) Dacă X este o variabilă aleatoare cu valori în $\mathbb{N}$ atunci

$$\mathbb{E}[X] = \sum_{n\geq 1}\mathbb{P}(X \geq n).$$

b) Dacă X este o variabilă aleatoare cu valori pozitive atunci

$$\mathbb{E}[X] = \int_0^{+\infty}\mathbb{P}(X \geq x)\,dx.$$

a) Observăm că:

$$\mathbb{P}(X > n) = \mathbb{P}(X = n+1) + \mathbb{P}(X = n+2) + \cdots = \sum_{k=n+1}^{\infty} \mathbb{P}(X = k)$$

și

$$\begin{aligned}\sum_{n=0}^{\infty}\sum_{k=n+1}^{\infty} \mathbb{P}(X = k) &= (\mathbb{P}(X = 1) + \mathbb{P}(X = 2) + \dots) + (\mathbb{P}(X = 2) + \mathbb{P}(X = 3) + \dots)\\ &+ (\mathbb{P}(X = 3) + \mathbb{P}(X = 4) + \dots) + \dots\\ &= 1\cdot\mathbb{P}(X = 1) + 2\cdot\mathbb{P}(X = 2) + 3\cdot\mathbb{P}(X = 3) + \dots\\ &= \sum_{k=1}^{\infty} k\mathbb{P}(X = k) = \sum_{k=0}^{\infty} k\mathbb{P}(X = k) = \mathbb{E}[X].\end{aligned}$$

O altă idee ar fi să luăm $X = \sum_{i=0}^{\infty} \mathbf{1}_{\{X>i\}}$ și să inversăm $\sum$ cu $\mathbb{E}$ (de ce putem ?).

b) Avem

$$\int_0^{+\infty} \mathbb{P}(X \geq x)\,dx = \int_0^{+\infty} \mathbb{E}[\mathbf{1}_{\{X\geq x\}}]\,dx \overset{\text{Tonelli-Fubini}}{=} \mathbb{E}\left[\int_0^{\infty} \mathbf{1}_{\{X\geq x\}}\,dx\right] = \mathbb{E}\left[\int_0^{X} dx\right] = \mathbb{E}[X].$$

Exercițiul 7

Un tehnician dintr-un laborator de biologie face două măsurători considerate independente și repartizate normal de medie 0 și de varianță 1. Calculați corelația dintre valoarea cea mai mică și cea mai mare a celor două măsurători.

Fie X și Y cele două măsurători, iar conform ipotezei acestea sunt variabile aleatoare independente și $\mathcal{N}(0,1)$ repartizate. Fie de asemenea $M = \max(X, Y)$ și $L = \min(X, Y)$, cea mai mare și respectiv cea mai mică dintre valori. Cum pentru $x, y \in \mathbb{R}$ avem

$$\max(x, y) + \min(x, y) = x + y \quad \text{și} \quad \max(x, y) - \min(x, y) = |x - y|$$

deducem că

$$\begin{aligned}\mathbb{E}[M] + \mathbb{E}[L] &= \mathbb{E}[M + L] = \mathbb{E}[X + Y] = \mathbb{E}[X] + \mathbb{E}[Y] = 0\\ \mathbb{E}[M] - \mathbb{E}[L] &= \mathbb{E}[M - L] = \mathbb{E}[|X - Y|].\end{aligned}$$

Pentru a calcula $\mathbb{E}[|X - Y|]$ să observăm că $X - Y \sim \mathcal{N}(0, 2)$ (deoarece X și Y sunt independente) prin urmare notând $X - Y = Z\sqrt{2}$, cu $Z \sim \mathcal{N}(0,1)$, obținem

$$\begin{aligned}\mathbb{E}[|X - Y|] &= \sqrt{2}\mathbb{E}[|Z|] = \sqrt{2}\int_{-\infty}^{\infty} |z|\frac{1}{\sqrt{2\pi}}e^{-\frac{z^2}{2}}\,dz\\ &= 2\sqrt{2}\int_0^{\infty} z\frac{1}{\sqrt{2\pi}}e^{-\frac{z^2}{2}}\,dz = 2\sqrt{2}\frac{1}{\sqrt{2\pi}} = \frac{2}{\sqrt{\pi}}.\end{aligned}$$

Astfel deducem că $\mathbb{E}[M] = \frac{1}{\sqrt{\pi}}$ și $\mathbb{E}[L] = -\frac{1}{\sqrt{\pi}}$ iar

$$Cov(M, L) = \mathbb{E}[ML] - \mathbb{E}[M]\mathbb{E}[L] = \mathbb{E}[XY] + \frac{1}{\pi} = \frac{1}{\pi}$$

deoarece $\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y] = 0$. Cum corelația dintre M și L este definită prin

$$\rho(M, L) = \frac{Cov(M, L)}{\sqrt{Var(M)}\sqrt{Var(L)}},$$

trebuie să mai calculăm $Var(M)$ și $Var(L)$. Cum $M + L = X + Y$, luând varianța avem

$$Var(M + L) = Var(X + Y) = Var(X) + Var(Y) = 2$$

de unde

$$Var(M) + Var(L) + 2Cov(M, L) = 2$$

deci $Var(M) + Var(L) = 2\left(1 - \frac{1}{\pi}\right)$. Cum $\max(X, Y) = -\min(-X, -Y)$ și cum, din simetria repartiției normale standard, $(-X, -Y)$ este repartizat la fel ca (X, Y) deducem că

$$Var(M) = Var(\max(X, Y)) = Var(-\min(-X, -Y)) = Var(\min(-X, -Y)) = Var(\min(X, Y)) = Var(L)$$

prin urmare $Var(M) = Var(L) = 1 - \frac{1}{\pi}$ de unde

$$\rho(M, L) = \frac{Cov(M, L)}{\sqrt{Var(M)}\sqrt{Var(L)}} = \frac{\frac{1}{\pi}}{1 - \frac{1}{\pi}} = \frac{1}{\pi - 1}.$$

Exercițiul 8

Un proces Bernoulli de parametru p este un șir de variabile aleatoare independente $(X_n)_{n\geq 1}$ cu $X_n \in \{0, 1\}$ și $\mathbb{P}(X_n = 1) = p$.

a) Arătațic că v.a. $S_n = X_1 + X_2 + \cdots + X_n$ este repartizată $\mathcal{B}(n, p)$ și calculați media și varianța acesteia.

b) Fie L cel mai mare număr natural pentru care $X_1 = X_2 = \cdots = X_L$ și M cel mai mare număr natural așa încât $X_{L+1} = X_{L+2} = \cdots = X_{L+M}$. Găsiți distribuțiile v.a. L și M.

c) Arătați că $\mathbb{E}[L] \geq \mathbb{E}[M]$, $\mathbb{V}[L] \geq \mathbb{V}[M] \geq 2$ și calculați $Cov[L, M]$.

d) Calculați $\lim_{k\to\infty} \mathbb{P}(M = n \,|\, L = k)$.

a) Este definiția binomialei. Avem $\mathbb{E}[S_n] = np$ și $\mathbb{V}[S_n] = np(1 - p)$.

b) Avem că $\{L = n\} = \{X_1 = \cdots = X_n = 1, X_{n+1} = 0\} \cup \{X_1 = \cdots = X_n = 0, X_{n+1} = 1\}$ de unde $\mathbb{P}(L = n) = p^n q + pq^n$, $n \geq 1$, $q = 1 - p$. Rezultă că

$$\mathbb{E}[L] = \sum_{n\geq 1} n\mathbb{P}(L=n) = \sum_{n\geq 1} n(p^n q + pq^n) = 2 + \frac{(p-q)^2}{pq}$$
$$\mathbb{V}[L] = \sum_{n\geq 1} n^2\mathbb{P}(L=n) - \mathbb{E}[L]^2 = 2 + \frac{(1+pq)(p-q)^2}{p^2q^2}$$

Pentru a găsi legea lui M să ne uităm la cuplul (L, M) și să observăm că evenimentul $\{L = n, M = m\}$ este dat de $\{X_1 = \cdots = X_n = 1, X_{n+1} = \cdots = X_{n+m} = 0, X_{n+m+1} = 1\} \cup \{X_1 = \cdots = X_n = 0, X_{n+1} = \cdots = X_{n+m} = 1, X_{n+m+1} = 0\}$ de unde

$$\begin{aligned}\mathbb{P}(L=n, M=m) = \mathbb{P}(X_1 = \cdots = X_n = 1, X_{n+1} = \cdots = X_{n+m} = 0, X_{n+m+1} = 1) +\\ \mathbb{P}(X_1 = \cdots = X_n = 0, X_{n+1} = \cdots = X_{n+m} = 1, X_{n+m+1} = 0) = p^n q^m p + p^m q^n q\end{aligned}$$

și prin urmare legea lui M este

$$\mathbb{P}(M=m) = \sum_{n\geq 1} \mathbb{P}(L=n, M=m) = q^{m-1}p^2 + p^{m-1}q^2.$$

Obținem asfel că $\mathbb{E}[M] = 2$ (independent de p) și că

$$\mathbb{V}[M] = 2 + \frac{2(p-q)^2}{pq}.$$

c) Este evident că $\mathbb{E}[L] = 2 + \frac{(p-q)^2}{pq} \geq 2 = \mathbb{E}[M]$ și că $\mathbb{V}[L] = 2 + \frac{(1+pq)(p-q)^2}{p^2q^2} \geq 2 + \frac{2(p-q)^2}{pq} = \mathbb{V}[M] \geq 2$. Se poate calcula ușor că

$$\mathbb{E}[LM] = \sum_{n,m\geq 1} nm\mathbb{P}(L=n, M=m) = \frac{1}{pq}$$

de unde rezultă că $Cov[L, M] = -\frac{(p-q)^2}{pq}$.

d) Observăm că

$$\lim_{k\to\infty} \mathbb{P}(M=n \,|\, L=k) = \lim_{k\to\infty} \frac{\mathbb{P}(L=k, M=n)}{\mathbb{P}(L=k)} = \lim_{k\to\infty} \frac{p^{k+1}q^n + q^{k+1}p^n}{p^k q + q^k p}$$

și studiind comportamentul raportului $\frac{p}{q}$ (dacă este > 1 sau nu după cum e p) deducem că

$$\lim_{k\to\infty} \mathbb{P}(M=n \,|\, L=k) = \begin{cases} p^{n-1}q, & \text{dacă } p < \frac{1}{2} \\ q^{n-1}p, & \text{dacă } p > \frac{1}{2} \\ 2^{-n}, & \text{dacă } p = \frac{1}{2} \end{cases}$$

Exercițiul 9

a) Fie X o variabilă aleatoare de medie 0 și varianță $\sigma^2 < \infty$. Arătați că penru orice $a > 0$ are loc inegalitatea

$$\mathbb{P}(X \geq a) \leq \frac{\sigma^2}{\sigma^2 + a^2}$$

b) Un grup de 200 de persoane, din care jumătate sunt bărbați, este divizat într-o 100 de perechi de câte 2 persoane. Dați o margine superioară pentru probabilitatea ca cel mult 30 dintre acestea să fie perechi mixte.

a) Fie $b > 0$ și să observăm că evenimentul $\{X \geq a\}$ este echivalent cu evenimentul $\{X + b \geq a + b\}$ de unde

$$\mathbb{P}(X \geq a) = \mathbb{P}(X + b \geq a + b) \leq \mathbb{P}\left((X + b)^2 \geq (a + b)^2\right)$$

deoarece dacă $\omega \in \{X + b \geq a + b\}$ atunci $\omega \in \left\{(X + b)^2 \geq (a + b)^2\right\}$. Aplicând inegalitatea lui Markov pentru $Y = (X + b)^2 \geq 0$ avem că

$$\mathbb{P}(X \geq a) \leq \mathbb{P}\left((X + b)^2 \geq (a + b)^2\right) \leq \frac{\mathbb{E}\left[(X + b)^2\right]}{(a + b)^2} = \frac{\sigma^2 + b^2}{(a + b)^2}, \quad \forall b > 0,$$

unde în ultima egalitate am folosit faptul că variabila aleatoare X este de medie 0. Dacă luăm valoarea lui b care minimizează membrul drept, aceasta este $b = \frac{\sigma^2}{a}$, obținem rezultatul cerut și anume

$$\mathbb{P}(X \geq a) \leq \frac{\sigma^2}{a^2 + \sigma^2}.$$

b) Observăm că dacă X este o variabilă aleatoare de medie μ și varianță σ^2 atunci inegalitatea de la punctul a) aplicată variabilelor $X - \mu$ și respectiv $\mu - X$ implică, pentru $a > 0$

$$\begin{aligned}\mathbb{P}(X \geq \mu + a) &\leq \frac{\sigma^2}{a^2 + \sigma^2}\\ \mathbb{P}(X \leq \mu - a) &\leq \frac{\sigma^2}{a^2 + \sigma^2}.\end{aligned}$$

Dacă numerotăm bărbații de la 1 la 100 și definim variabilele aleatoare X_i, de tip Bernoulli, care iau valoarea 1 dacă bărbatul i este împerechiat cu o femeie și 0 altfel, atunci numărul de perechi mixte este

$$S = \sum_{i=1}^{100} X_i.$$

Deoarece bărbatul i are șanse egale să fie împerecheat cu oricare din cele 199 de persoane rămase, din care 100 sunt femei, avem că

$$\mathbb{E}[X_i] = \mathbb{P}(X_i = 1) = \frac{100}{199}$$

iar pentru $i \neq j$ avem

$$\mathbb{E}[X_i X_j] = \mathbb{P}(X_i = 1, X_j = 1) = \mathbb{P}(X_i = 1)\mathbb{P}(X_j = 1 | X_i = 1) = \frac{100}{199} \times \frac{99}{197}.$$

Prin urmare găsim că

$$\mathbb{E}[S] = \sum_{i=1}^{100} \mathbb{E}[X_i] = 100 \times \frac{100}{199} \approx 50.25$$

iar

$$Var(S) = \sum_{i=1}^{100} Var(X_i) + 2\sum_{i<j} Cov(X_i, X_j) = \sum_{i=1}^{100} Var(X_i) + 2\sum_{i<j} \left(\mathbb{E}[X_i X_j] - \mathbb{E}[X_i]\mathbb{E}[X_j]\right),$$

de unde

$$Var(S) = 100 \times \frac{100}{199}\frac{99}{199} + 2\binom{100}{2}\left[\frac{100}{199} \times \frac{99}{197} - \left(\frac{100}{199}\right)^2\right] \approx 25.126.$$

Aplicând inegalitatea de mai sus deducem că

$$\mathbb{P}(S \leq 30) = \mathbb{P}(S \leq \underbrace{50.25 - 20.25}_{\mu - a}) \leq \frac{25.126}{25.126 + 20.25^2} \approx 0.058.$$