Alexei Leahu, Ion Pârţachi

PROBABILITĂŢI ŞI STATISTICĂ
(Prin exemple şi probleme propuse)
Partea I. PROBABILITĂŢI

Chişinău

CUPRINS

3.1. Parametri de poziţie: valoarea medie, moda, mediana, cuantile

3.2. Disperia (varianţa), abaterea standard, covarianţa, coeficientul de corelaţie, regresie liniară

3.3. Momente ale variabilei aleatoare (iniţiale, centrale), asimetria, aplatizarea

4. Modele (distribuţii, densităţi de distribuţie) probabiliste uzuale, inegalităţi, Legea Numerelor Mari, Teorema Limită Centrală pentru variabile aleatoare independente identic distribuite

4.1. Distribuţii probabiliste uzuale in caz discret (Uniformă, Bernoulli, Binomială, Geometrică, Poisson, Multinomială, Hypergeometrică, Hypergeometrică Multivariată)

4.2. Distribuţii probabiliste uzuale in caz (absolut) continuu (Uniformă, Exponenţială, Normală, Hi-pătrat (χ^2), T-Student)

4.3. Inegalitatea Chebyshev, Legea Numerelor Mari (în formele Cebyshev, Bernoulli, Hincin), Teorema Limită Centrală

5. Elemente de Teoria Informaţiei

5.1. Obiectul de studiu al Teoriei Informaţiei

5.2. Entropia ca masură a nedeterminării/cantităţii de informaţie

5.3. Proprietăţile entropiei

5.4. Transmiterea informaţiei. Codificarea. Teoreme de codificare

BIBLIOGRAFIE

Prefaţă

Utilizarea modelelor şi metodelor statistico-probabiliste la analiza şi interpretarea datelor a devenit o practică uzuală în toate domeniile ştiinţifice, fie că este vorba de prelucrarea datelor cu caracter economic sau datelor experimentale în tehnică sau de analiză şi interpretarea datelor legate de elaborarea unor noi tehnologii informaţionale, etc.

Disciplina matematică **Probabilităţi şi Statistică** cuprinde elementele de bază ale **Teoriei Probabilităţilor, Statisticiii Descriptive şi Matematice**, dar şi exemple de **Aplicaţii** ale lor. **Teoria Probabilităţilor** este cea care:

1) Oferă noţiunile care stau la baza modelelor ce descriu comportamentul probabbilist al fenomenelor aleatoare (eveniment aleator, probabilitate, variabilă aleatoare (v.a.) şi funcţia ei de distribuţie (f.d.), valoare medie, dispersie şi alte caracteristici numerice ale v.a., etc.);

2) Pune la dispoziţie formule şi rezultate teoretice fructificate în metodele de calcul ale probabilităţilor (formulele adunării, înmulţirii, probabilităţii totale, lui Bayes, dar şi formulele de calcul bazate pe funcţia de distribuţie, distribuţia sau densitatea de distribuţie a v.a.);

3) Serveşte, graţie proprietăţilor caracteristice f.d., drept sursă inepuizabilă de modele probabiliste, scoţánd în evidenţă modelele probabiliste uzuale sau clasice (distribuţiile Uniformă, Bernoulli, Binomială, Geometrică, Hypergeometrică, Exponenţială, Normală, Hi-patrat, Student, etc.);

4) Fundamentează matematic Principiul Regularitării Statistice prin prisma Legii Numerelor Mari, care, împreună cu Teorema Limită Centrală, conferă Statisticii un suport matematic solid.

5) Aflându-se la baza Teoriei Informaţiei, permite aplicarea metodelor cantitative aspura conceptelor legate de informaţie, inclusiv ín contextul cercetării sistemelor de prelucrare şi transmitere a informaţiei.

Statistica Matematică este cea care, prin prisma Teoriei Probabilităţilor, oferă un cadru riguros pentru analiza şi interpretarea datelor statistice. Aceasta permite ca rezultatele prelucrării primare a datelor realizate cu ajutorul Statisticiii Descriptive să fie aprofundate prin a cerceta calitatea, exactitatea estimatorilor utilizaţi, dar şi prin lansarea, verificarea şi validarea ipotezelor statistice privind legitatile probabiliste care guvernează sau generează datele statistice colectatele. Cu alte cuvinte, în baza unor observaţii parţiale ale unui fenomen aleator putem, fie şi cu o anumită aproximare şi grad de incredere, anticipa comportamentul probabilist al acestuia.

1. Calculul Probabilităţilor

1.1. Obiectul de studiu al Teoriei Probabilităţilor şi locul ei in Statistica Matematica, probabilitate frecvenţială, probabilitate subiectivă

Noţiunea de *statistică* se naşte odată cu apariţia şi dezvoltarea relaţiilor economce, până în secolul XIX aceasta fiind tratată ca ştiinţă politică. Cuvântul în cauză provine de la latinescul *status*, care înseamnă stare.

Începând cu secolul XIX, *statistica* prinde conturul ştiinţei care, actualmente, *are drept obiect de studiu metodele, procedeele de colectare, organizare, prelucrare, analiză şi interpretare a datelor ce vizează rezultatele observărilor facute asupra fenomenelor sau experimentelor aleatoare.* Statistica modernă, mai ales acea parte a ei care se numeşte *Statistica matematica* , bazându-se esenţial pe realizarile ştiinţelor matematice, foloseşte din plin *Teoria probabilităţilor.*

Apariţia Teoriei probabilităţilor ca ramură a matematicii ce studiază modele matematice ale fenomenelor aleatoare datează din secolul XVII şi este legată de numele marilor matematicieni Blaise Pascal (1623-1662), Pierre Fermat (1601-1665), Christian Huygens (1629-1695) şi Jacob Bernoulli (1654-1705).

Pentru a înţelege mai bine ce studiază Teoria probabilităţilor ca parte a Statisticii matematice, avem nevoie de cateva explicatii suplimentare, ţinând cont că mulţitudinea de fenomene care se întâlnesc în lumea înconjurătoare se împarte în două clase: *fenomene deterministe* si *fenomene indeterministe* sau *aleatoare.*

Astfel, spunem ca fenomenul este *determinist* daca observatorul poate anticipa cu certitudine evoluţia sau derularea acestuia. In calitate de exemplu putem lua fenomenul atracţiei universale. Observaţiile facute asupra acestui fenomen i-au permis marelui matematician şi fizician englez Isaac Newton (1642-1727) sa formuleze legea atracţiei universale:

$$F = k\frac{m_1 m_2}{r^2}\ .$$

Acesta este un exemplu tipic de model matematic a unui fenomen (in cazul dat) determinist. Dealtfel, *a modela matematic (spre a fi cercetat) un fenomen, proces, experiment, eveniment sau obiect oarecare înseamnă a-l descrie cu ajutorul noţiunilor şi formulelor matematice, adică a-l descrie în limbajul* *matematic.* Unul şi acelaşi model matematic poate descrie două fenomene diferite în esenţă. De exemplu, formula de mai sus poate servi in calitate de model matematic şi pentru fenomenul atracţiei a două particule elementare (legea lui Coulomb).

Spunem despre un fenomen că este *indeterminist (aleator)* , dacă observatorul fenomenului *nu* poate anticipa cu certitudine evoluţia lui. Din punct de vedere al observatorului, observaţiile făcute asupra unui fenomen sau măsurătorile corespunzătoare echivalează cu o experimentare legată de fenomenul dat. Or, prin *experiment* vom înţelege observarea unui fenomen dat.

Acum putem spune că *Teoria probabilităţilor* şi *Statistica matematică* studiază modele matematice ale experimentelor (fenomenelor) aleatoare.

Aici se impun cateva precizări. Experimentele aleatoare (indeterministe) se împart, la rândul lor, în două subclase: **(a)** *experimente aleatoare care posedă proprietatea regularităţii (stabilităţii) statistice* şi **(b)** *experimente aleatoare care nu posedă proprietatea regularităţii statistice.*

Definiţia 1. Vom spune că un experiment aleator $\mathcal{E}$ posedă *proprietatea regularităţii (stabilităţii) statistice* daca acesta verifică urmatoarele proprietăţi:

1) *poate fi reprodus ori de câte ori dorim practic în aceleaşi condiţii;*

2) *pentru orice eveniment A asociat lui $\mathcal{E}$ frecvenţa lui relativă în n probe*

$$f_n(A) = \frac{\text{numărul de probe în care } s-a \text{ produs } A}{\text{numărul total de probe}} = \frac{n(A)}{n}$$

oscilează în jurul unui număr notat cu $P(A)$, $P(A) \in [0,1]$, $f_n(A)$ devenind, odată cu creşterea lui n, "tot mai aproape şi mai aproape de $P(A)$";

3) *pentru două serii diferite, respectiv de n şi m probe, atunci când n şi m sunt foarte mari, avem că $f_n(A) \approx f_m(A)$.*

În concluzie, stabilitatea statistică a frecvenţelor relative conferă verosimilitate ipotezei, conform căreia pentru orice eveniment A, posibil ca rezultat observabil al unui experiment aleator $\mathcal{E}$, putem defini numărul $P(A)$ cu ajutorul căruia măsurăm gradul (şansele) de realizare a lui A într-un număr foarte mare de probe. Astfel, în Teoria probabilităţilor devine postulat afirmaţia, conform căreia pentru orice eveniment A asociat unui experiment aleator $\mathcal{E}$ există (obiectiv) un număr $P(A)$ numit probabilitate a lui A. Proprietatea firească a acestui număr rezidă în faptul că odată cu creşterea numărului n de probe (experimente) "independente" frecvenţa relativă $f_n(A)$ se apropie tot mai mult de $P(A)$. Numărul $P(A)$ se numeşte *probabilitate statistică (sau frecvenţială) a evenimentului* A. Aceasta din urmă afirmatie/proprietate este conforma cu Legea Numerelor Mari, care in Teoria probabilităţilor reprezintă suportul matematic al proprietăţii regularităţii/stabilităţii statistice, ceea ce inseamnă că aceasta teorie se pliază perfect pe aplicaţiile ei la rezolvarea problemelor ce apar in cadru studierii modelelor matematice ale experimentelor aleatoare ce fac parte din subclasa (a).

Exemplul 1. Considerăm în calitate de experiment aleator $\mathcal{E}$ aruncarea unei monede simetrice (cu centrul de greutate nedeplasat) sau, cum se mai spune, "perfecte". Fie A evenimentul ce constă în apariţia stemei. Observăm, astfel, că.

$$f_{1000}\,(A) \approx \frac{1}{2} = P(A),\; f_{2000}(A) \approx \frac{1}{2} = P(A).$$

Prin urmare, putem afirma ca probabilitatea (statistică) a apariţiei stemei la aruncarea monedei o singura dată este egală cu $1/2$, ceea ce inseamnă, ca ăruncând moneda de un număr suficient de mare de ori, stema va apare în aproximativ 50% de cazuri .

Putem aduce şi alte exemple de fenomene aleatoare: rezultatele aruncării unui zar, greutatea unui bob ge grâu ales la întâmplare, numărul de bacterii într-o picătură de apă, durata vieţii unui calculator produs de întreprinderea dată, numărul de apeluri telefonice înregistrate la o staţie telefonică pe durata unei zile, etc., etc. Enumerarea lor poate continua la nesfârşit, însă ele toate vor avea acelaşi caracter, fiind însoţite de astfel de noţiuni imprecise (deocamdată) ca aruncare "onestă", moneda "perfectă", "probe independente", etc.

Interesant este faptul ca frecventa relativa, prin urmare şi probabilitatea statistică (frecvenţială), posedă proprietăţile puse in anul 1934 la baza Teoriei axiomatice a probabilităţilor, teorie lansată de către ilustrul matematician rus A. N. Kolmogorov (1903-1987). Iata aceste proprietăţi care pot fi verificate cu uşurinţă pe cale experimentală în baza următorului exemplu.

Exemplul 2. Considerăm experimentul aleator aleator $\mathcal{E}$ ce constă în efectuarea un sondaj printre potentialii săi clienţi de către o companie producatoare de 3 tipuri de băuturi răcoritoare A, B şi C, pentru ca în funcţie de rezultatele sondajului sa se facă ajustări, privind cotele de producţie a fiecarui tip de bautura. Exprimarea preferintei presupune alegerea unui singur raspuns: A, B sau C. Considerând că in eşantion au fost incluşi un număr n, suficient de mare. de respondenţi, atunci compania va putea identifica cotele de productie, acestea fiind direct proporţional cu frecventele relative $f_n(A)$, $f_n(B)$ şi $f_n(C)$. Se poate observa cu usurinţă că:

1. $0 \leq f_n(A) \leq 1$ (rezultă din definitia frecvenţei relative);

2. $f_n(A \text{ } sau \text{ } B \text{ } sau \text{ } C) = 1$ (deoarece raspunsurile A *sau* B *sau* C se exclud mutual, doua cate doua , iar alte raspunsuri sunt excluse, cu alte cuvinte înregistrarea ca raspuns A *sau* B sau C, privit ca eveniment aleator, este eveniment *sigur*);

3. $f_n(A \text{ } sau \text{ } B) = f_n(A) + f_n(B)$, $f_n(A \text{ } sau \text{ } C) = f_n(A) + f_n(C)$, $f_n(B \text{ } sau \text{ } C) = f_n(B) + f_n(C)$.

Remarca1. Probabilitatea statistică nu poate fi aplicată întotdeauna, deoarece nu orice experiment poate fi repetat în condiţii identice ori de câte ori dorim.

Drept exemplu putem lua rezultatul pe care-l va obtine un student *concret* la examenul de Probabilităţi şi Statistică matematică. Într-adevăr, observăm ca acest examen (experiment aleator), raportat la un student concret, nu poate fi *poate fi reprodus ori de câte ori dorim practic în aceleaşi condiţii.* Daca, însă, vom modifica experimentul prin a ne referi in fiecare proba (repetare de examen) la un student "ales la întâmplare", atunci acest experiment va satisface rigorile 1-3 ale regularităţii statistice, apelând, de exemplu, la istoricul rezultatelor de la acest examen din anii anteriori. Or, experimentele aleatoare care posedă proprietatea regularităţii statistice ţin de fenomenele de masă. Pentru studiul experimentelor care nu posedă această proprietate, putem folosi noţiunea de probabilitate subiectivă.

Definiţia 2. Prin *probabilitate subiectivă* vom înţelege acea regulă P conform căreia o persoană dată îi asociază fiecărui eveniment A un număr $P(A) \in [0, 1]$, numit *probabilitatea* evenimentului A.

Astfel, orice student concret îşi poate, bineînţeles, evalua probabilitatea subiectivă că va promova din prima încercare examenul de Probabilităţi şi Statistică Matematică, această probabilitate variind în functie de student.

La fel, putem vorbi despre probabilitatea subiectiva, evaluata, să zicem, de un expert NASA, că până în anul 2025 va avea loc prima expediţie a omului pe Marte.

Pentru studiul fenomenelor aleatoare indeterministe, în afară de probabilitate *subiectivă* şi probabilitate *frecvenţială*, există şi noţiunile de probabilitate *clasică*, probabilitate *geometrică*, probabilitate *discretă* şi probabilitate definită în sens *axiomatic*. Toate aceste noţiuni au ca scop definirea unei modalităţi de măsurare a şanselor (gradelor) de realizare, gradelor de certitudine in producerea evenimentelor aleatoare date, *definiţia axiomatică a probabilităţii*, fiind, într-un anumit sens, *acoperitoare pentru toate noţiunile de probabilitate* enumerate mai sus.

Remarca 2. În ceea ce priveşte fenomenele aleatoare întâlnite în domeniile ştiinţelor inginereşti acestea au, de regulă, un caracter de masă. Însă, dacă ne vom referi la domeniile economic şi al afacerilor, acestea fiind vulnerabile la factorii subiectivi, trebuie sa fim foarte atenţi dacă fenomenele aleatoare cu caracter economic, puse de noi în discuţie, satisfac rigorilor impuse asupra experimentelor aleatoare pentru ca acestea sa fie tratate/clasificate ca fiind experimente ce posedă proprietatea regularităţii statistice sau nu.

Precizăm, ca *Teoria probabilităţilor si Statistica Matematica,* expusă de noi in continuare, *vizeaza doar modele matematice ale experimentelor aleatoare ce posedă proprietatea regularităţii statistice.* Cu alte cuvinte vom vorbi doar despre *probabilităţi obiective.*

Probleme propuse

1. La prima lecţie practică, privind calculul probabilităţilor, studenţilor le-a fost propusă următoarea întrebare: "Cu ce este egală probabilitatea că, ieşind la plimbare pe bulevardul Ştefan cel Mare din Chişinău, întâmplător, vă va ieşi în cale un Dinozaur?" Unul din studenţi, familiarizat cu definiţia clasică a probabilităţii încă din liceu, a răspuns că probabilitatea în cauză este egala cu 1/2, deoarece..."este posibil unul din două cazuri cu şanse egale de realizare: ori vom întâlni un Dinosaur, ori Nu". Răspunsul acesta, fiind unul ridicol, se impune întrebarea: ce metodă de evaluare a probabilităţii se potriveşte în acest caz? Pe cale experimentală sau subiectivă? Cu ce este egala probabilitatea respectivă? Argumentaţi.

2. Biblioteca Naţională "Vasile Alecsandri" din Chişinău a înregistrat pe parcursul anului 2017 cititori nou veniţi care, privind nivelul lor de studii, s-au repartizat conform următorului tabel:

Nivelul de studii	**Frecvenţa**
Superioare	720
Studenţi/superioare incomplete	1106
Medii (liceu, şcoala medie)	502
Medii incomplete	52
Fără studii	0
Total	2380

Evaluaţi probabilitatea că un cititor luat la întâmplare din lista celor inregistraţi la Biblioteca Naţională "V. Alecsandri" va fi unul cu studii.

Evaluaţi probabilităţile respective pentru fiecare nivel de studii şi arătaţi că suma lor este egala cu 1.

3. Multe instituţii de învăţământ preuniversitar oferă elevilor acces la internet. Astfel, conform Anuarului Statistic din SUA publicat in 1997, în anul 1996 acces la internet aveau 21 733 din 51 745 şcoli primare, 7286 din 14 012 şcoli gimnaziale si 10 682 din 17 229 licee.

Cu ce este egală probabilitatea ca:

a) Alegând la întâmplare pentru a vizita *una din şcolile primare*, aceasta va fi conectată la Internet;

b) Alegând la întâmplare pentru a vizita *una din şcolile gimnaziale*, aceasta va fi conectată la Internet;

c) Alegând la întâmplare pentru a vizita *unul din licee*, acesta va fi conectată la Internet;

d) Alegând la întâmplare pentru a vizita *una din instituţiile de învăţământ menţionate mai sus*, aceasta va fi conectată la Internet.

4. O companie producătoare de pastă de dinţi a iniţiat un studiu privind gradul de preferintă a cumpărătorului în funcţie de unul din cele 5 tipuri de ambalaje propuse de designeri. Drept rezultat, au fost inregistrate următoarele rezultate: 5 cumpărători au preferat ambalajul nr. 1, 15- ambalajul nr. $2, 30$- ambalajul nr. 3, 40- ambalajul nr.4 şi 10- ambalajul nr. 5.

Sunt oare aceste date suficiente pentru a depista ambalajul cu şansele cele mai mari de a fi preferat de catre cumpărători? Argumentaţi.

5. O companie specializată in construirea blocurilor de locuit, ce consta din apartamente cu 1, 2 si 3 camere, a estimat că probabilităţile subiective, privind cererea din partea unui client, în funcţie de tipul apartamentului solicitat, sunt urmatoarele: 0.30 pentru un apartament cu o cameră, 0.40 pentru un apartament cu 2 camere şi 0.25 pentru un apartament cu 3 camere.

a) Sunt oare valide aceste estimari ale probabilităţilor mentionate mai sus? De ce DA sau de ce NU?

b) Ce intervenţie asupra valorilor probabilităţilor trebuie aplicate ca acestea sa devină valide?

6. Specificaţi care din exemplele descrise mai jos se referă la experimentele aleatoare ce poseda proprietatea Regularităţii Statistice si care nu:

a) Rezultatul unui meci concret in cadrul campionatului mondial la fotbal;

b) Rezultatul jocului cu o singura variantă la Lotosport "6 din 49";

c) Eşecul sau succesul unui businessman la sfârşitul primului an de activitate în cadrul unei afaceri cu caracter economic;

d) Mărimea încălţămintei solicitate de un cumpărător ales la întâmplare dintre cei care au vizitat o secţie de încălţăminte a unui Mall;

e) Numărul de apeluri telefonice înregistrate în 24 de ore la un centru de urgenţă medicală.

1.2. Noţiuni şi rezultate auxiliare din Combinatorică

Analiza Combinatorie sau Combinatorica este acea disciplină matematică în care sunt studiate metodele de numărare a tuturor combinărilor ce pot fi alcatuite din elementele unei mulţimi finite. Aceasta se bazează esenţial pe doua principii: *Principiul adunării* şi *Principiul înmulţirii* expuse mai jos.

Fie A şi B sunt două mulţimi finite, atunci distingem două situaţii, după cum cele două mulţimi pot fi disjuncte sau nu. Se verifică cu uşurinţă că are loc

Principiul adunării (caz disjunct) Dacă A şi B sunt mulţimi finite şi disjuncte, adică $A \cap B = \emptyset$, $card(A) = n$ şi $card(B) = m$, atunci

$$card(A \cup B) = n + m.$$

Corolar. Daca $A_1, A_2, ..., A_k$ sunt mulţimi finite disjuncte două câte două, atunci

$$card(\bigcup_{i=1}^{k} A_i) = \sum_{i=1}^{k} card\ A_i$$

Principiul adunării (caz general). Dacă A şi B sunt mulţimi finite, $A, B \subseteq \Omega$, $card(A) = n$, $card(B) = m$ şi $card(A \cap B) = k$, atunci $card(A \cup B) = n + m - k$.

Demonstraţie (opţional). Folosind proprietăţile operaţiilor asupra mulţimilor,deducem că oricare ar fi mulţimile $A, B \subseteq \Omega$ avem:

$$A = (A \cap B) \cup (A \cap \overline{B}),\ B = (A \cap B) \cup (\overline{A} \cap B),$$
$$A \cup B = (A \cap B) \cup (A \cap \overline{B}) \cup (\overline{A} \cap B).$$

Conform principiului adunării în cazul disjunct obţinem:

$$cardA = card(A \cap B) + card(A \cap \overline{B}),$$

$$card(B) = card(A \cap B) + card(\overline{A} \cap B),$$

$$card(A \cup B) = card\ (A \cap B) + card\ (A \cap \overline{B}) + card\ (\overline{A} \cap B) + card\ (A \cap B)$$

$$-card(A \cap B) = card(A) + card(B) - card(A \cap B) = n + m - k.\square$$

Remarcă. Folosind inducţia matematică putem deduce Principiul adunării pentu un număr arbitrar k de mulţimi finite $A_1, A_2, ..., A_k$:

$$\begin{aligned} card \bigcup_{i=1}^{k} A_i &= \sum_{i=1}^{k} card\ A_i - \sum_{1 \le i < j \le k} card\ (A_i \cap A_j) + ... \\ &+ (-1)^{m-1} \cdot \sum_{1 \le i_1 < i_2 < ... i_m \le k} card\ (A_{i_1} \cap A_{i_2} \cap ... A_{i_m}) + .. \\ &+ (-1)^{k-1} \cdot card\ (A_1 \cap A_2 \cap ... A_k) \end{aligned}$$

Exemplul 1. Considerăm o grupă de studenţi despre care ştim că fiecare student posedă cel puţin una din limbi străine: 20 de studenţi cunosc limba engleză, 15 limba franceză, 10 limba germană, 5 limbile engleză şi franceză, 5 limbile franceză şi germană,4 limbile engleză şi germană şi 1 student limbile engleză, franceză şi germană. Câţi studenţi sunt în grupă ?

Notând prin E, F şi G mulţimile de studenţi care posedă, respectiv, limba engleză, franceză, germană şi ţinând cont de datele problemei, deducem: $cardE = 20$, $cardF = 15$, $cardG = 10$, $card(E \cap F) = 5$, $card(E \cap G) = 4$, $card(F \cap G) = 5$, $card(E \cap F \cap G) = 1$ şi atunci

$$card(E \cup F \cup G) = card(E) + card(F) + card(G) - card(E \cap F) -$$
$$card(E \cap G) - card(F \cap G) + card(E \cap F \cap G) = 32.$$

Principiul înmulţirii *(în limbajul produsului cartezian).* *Dacă A şi B sunt două mulţimi finite astfel încât $card(A) = n$ şi $card(B) = m$, atunci*

$$card(A \times B) = n \cdot m.$$

Demonstraţie (optional). Este evident ca dacă

$$A = \{a_1, a_2, ..., a_n\}, B = \{b_1, b_2, ..., b_m\},$$

atunci mulţimile $A\times\ B$ şi $\{\ (i,j)\ \big| i = \overline{1,n},\ j = \overline{1,m}\}$ au acelaşi număr de elemente. Daca în sistemul cartezian de coordonate xOy vom plasa valorile $i = \overline{1,n}$ pe axa Ox iar valorile $j = \overline{1,m}$ pe axa Oy, atunci elementului (i,j) îi corespunde punctul (i,j) din planul xOy, având, astfel, în planul xOy o reţea de $n \cdot m$ puncte. □

Pentru orice număr k de mulţimi finite, aplicând metoda inducţiei matematice, putem demonstra că are loc formula:

$$card(A_1 \times A_2 \times \ldots \times A_n) = \prod_{i=1}^{n} cardA_i.$$

Demonstraţia ei se bazează esential pe faptul ca are loc egalitatea $A_1 \times A_2 \times \ldots \times A_i = (A_1 \times A_2 \times \ldots \times A_{i-1}) \times A_i$, $i = \overline{2,k},$.

Principiul înmulţirii în limbajul produsului cartezian poate fi reformulat *în limbajul acţiunilor.*

Principiul înmulţirii *(în limbajul acţiunilor).* *Dacă o acţiune poate fi realizată în k etape succesive astfel încât etapa i poate fi realizată în n_i modalităţi, $i = \overline{1,k}$, atunci această acţiune poate fi realizată în $n_1 \cdot n_2 \cdot \ldots \cdot n_k$ modalităţi.*

Exemplul 2. Presupunem ca un safeu poate fi deschis cunoscând un cod de forma

$$i_1 \quad i_2 \quad i_3 \quad i_4 \quad i_5 \quad i_6\ ,$$

unde $i_k = \overline{0,9}$, $k = \overline{1,6}$. Câte coduri diferite de acest gen există?

Mulţimea Ω a tuturor codurilor de acest gen coincide cu produsul cartezian a mulţimii $\{0, 1, 2, ..., 9\}$ de 6 ori cu ea însăşi, adică

$$\Omega = \left\{(i_1, i_2, ..., i_6)\ \mid i_k \in \{1, 2, ..., 9\},\ k = \overline{1,6}\right\}.$$

Acesta are, conform Principiului înmulţirii în limbajul produsului cartezian,10^6 elemente (coduri).

Dacă avem informaţia ca acest cod este format din cifre diferite atunci vom observa ca Principiul înmulţirii *în limbajul acţiunilor* se aplică mai usor decât Principiului înmulţirii în limbajul produsului cartezian. Într-adevăr, a forma un cod din 6 cifre diferite este echivalent cu a efectua o acţiune în 6 etape succesive, astfel încât prima etapă poate fi realizată în 10 modalităţi, cea de a doua în 9 modalităţi,etc., ultima (a şasea) în $10-(6-1)=5$ modalităţi. Conform *Principiului înmulţirii în limbajul acţiunilor*, numărul tuturor codurilor este egal cu $10\cdot 9\cdot 8\cdot 7\cdot 6\cdot 5 = A_{10}^6$. Mai observam că raspunsul din exemplul nostru se exprima prin *formula aranjamentelor*.

Exemplul 3. Presupunem ca avem o multime Ω formată din n elemente. Atunci numarul tuturor submulţimilor pe care le putem forma din ea este egal cu

$$card\{A \ \mid A\subseteq\Omega\} = 2^n.$$

Într-adevar, pentru a forma o submulţime a lui Ω e ca şi cum ai realiza o acţiune in n etape succesive: la etapa cu nr. i decidem ce facem cu elementul cu nr.i, îl includem sau nu in mulţime, adică fiecare etapa, $i=\overline{1,n}$, poate fi realizata in 2 modalităţi. Apelând la Principiul înmulţirii obţinem formula de mai sus.

Definiţia 1. Fie A o mulţime formată din n elemente diferite, $A=\{a_1, a_2, \cdots, a_n\}$, atunci vom numi *aranjament din n elemente luate câte k* orice mulţime ordonată de forma $(a_{i_1}, a_{i_2}, \cdots, a_{i_k})$ cu proprietatea că $i_1\neq i_2\neq \cdots\neq i_k$, $a_{i_j}\in A$, $i_j=\overline{1,n}$, $j=\overline{1,k}$. Evident, noţiunea are sens pentru $k=\overline{1,n}$. Mulţimea tuturor aranjamentelor de n elemente luate câte k se notează cu $\mathcal{A}_n^k$, adica

$$\mathcal{A}_n^k = \left\{(a_{i_1}, a_{i_2}, \cdots, a_{i_k}) \ \mid i_1\neq i_2\neq\cdots\neq i_k,\ a_{i_j}\in A, i_j=\overline{1,n},\ j=\overline{1,k}\right\}.$$

Cardinalul acestei mulţimi se notează cu A_n^k şi este numărul tuturor aranjamentelor din n elemente luate câte k.

Conform principiului înmulţirii în limbajul acţiunilor, a construi un *aranjament din n elemente luate câte k* este echivalent cu a realiza o acţiune în k etape succesive, astfel încât prima etapă poate fi realizată în n modalităţi, cea de a doua în $n-1$ modalităţi, etc., ultima (etapa nr. k) în $n-(k-1)=n-k+1$ modalităţi. Or, numărul tuturor aranjamentelor din n elemente luate câte k este egal cu

$$A_n^k = n\cdot(n-1)\cdot(n-2)\cdot\cdots\cdot(n-k+1) = \frac{n!}{(n-k)!}$$

Prin definiţie, atunci când $k = n$, aranjamentul se numeşte *permutare de n elemente*. Deci mulţimea tutror permutărilor de n elemente notată prin $\mathbf{P}_n$ coincide cu $\mathcal{A}_n^n$, ceea ce înseamnă ca *numărul tutror permutărilor de n elemente* P_n este egal cu A_n^k, adică $P_n = n!$.

Definiţia 2. Orice submulţime de forma $\{a_{i_1}, a_{i_2}, \cdots, a_{i_k}\}$, $i_1 \neq i_2 \neq \cdots \neq i_k$, $a_{i_j} \in A$, $i_j = \overline{1,n}$, $j = \overline{1,k}$, se numeşte combinare din n elemente luate câte k. Evident noţiunea are sens pentru $k = \overline{1,n}$. Mulţimea tuturor combinărilor de n elemente luate câte k elemente o vom nota prin

$$\mathcal{C}_n^k = \left\{ \{a_{i_1}, a_{i_2}, \cdots, a_{i_k}\} \mid i_1 \neq i_2 \neq \cdots \neq i_k,\ a_{i_j} \in A, i_j = \overline{1,n}, j = \overline{1,k} \right\}.$$

Cardinalul acestei mulţimi îl vom nota cu $\mathbb{C}_n^k$.

Observăm ca dintr-o *combinare din n elemente luate câte k* putem forma $k!$ *aranjamente din n elemente luate câte k*. Or, a forma un aranjament din n elemente luate câte k este echivalent cu a realiza o acţiune în două etape succesive:

1. alegem o combinare din n elemente luate câte k, etapă pentru care avem $\mathbb{C}_n^k$ modalităţi de a o efectua;

2. din această combinare, formăm un aranjament din n elemente luate câte k, etapă care se poate realiza în $k!$ modalităţi.

Rezultă că $A_n^k = k! \cdot \mathbb{C}_n^k$, adică

$$\mathbb{C}_n^k = \frac{A_n^k}{k!} = \frac{n!}{k!(n-k)!}.$$

Exemplul 3. Considerăm că avem o mulţime de n elemente astfel încât n_1 elemente sunt de tipul 1, n_2 elemente sunt de tipul 2, $\cdots$, n_k elemente sunt de tipul k, $n_1 + n_2 + ... + n_k = n$. Alegem la întâmplare, unul câte unul, toate elementele mulţimii şi le aranjăm în ordinea extragerii lor. Să se calculeze cardinalul mulţimii Ω a tuturor rezultatelor posibile ce corespunzd acestui experiment.

Pentru a alcătui un rezultat posibil ca element a mulţimii Ω ce corespunzde acestui experiment este suficient să realizăm o acţiune în k etape succesive.

Etapa 1: din n locuri disponibile pentru a aranja elementele extrase, alegem n_1 locuri pe care vom plasa elementele de tipul1. Această acţiune o putem realiza în $\mathbb{C}_n^{n_1}$ modalităţi;

Etapa 2: din cele $n-n_1$ locuri, disponibile dupa etapa 1 , alegem n_2 locuri pe care vom plasa elementele de tipul 2. Această acţiune o putem realiza în $\mathbb{C}_{n-n_1}^{n_2}$ modalităţi, etc.,

Etapa k: din cele $n-n_1-n_2-\cdots-n_{k-1}=n_k$ locuri, disponibile dupa etapa k, alegem n_k locuri pe care vom plasa elementele de tipul k. Această acţiune o putem realiza în $\mathbb{C}_{n-n_1-n_2-\cdots-n_{k-1}}^{n_k}=\mathbb{C}_{n_k}^{n_k}$ modalităţi.

Conform principiului înmulţirii, avem :

$$card(\Omega)=\mathbb{C}_n^{n_1}\cdot\mathbb{C}_{n-n_1}^{n_2}\cdot\cdots\cdot\mathbb{C}_{n-n_1-n_2-\cdots-n_{k-1}}^{n_k}=\frac{n!}{n_1!n_2!\cdots n_k!}.$$

Formula obţinută este, de fapt, formula de calcul pentru $P(n_1,n_2,...,n_k)$, numărul permutărilor a n elemente, din care n_1 elemente sunt de tipul 1, n_2 elemente sunt de tipul 2, $\cdots$, n_k elemente sunt de tipul k, $n_1+n_2+...+n_k=n$:

$$P(n_1,n_2,...,n_k)=\frac{n!}{n_1!n_2!\cdots n_k!}.$$

Ultima mai poarta denumirea de **formula permutarilor cu repetare** .

Exemplul 4. Presupunem că avem la dispoziţie 10 cartonaşe marcate cu litere astfel: M, M, A, A, A, T, T, I, E, C. Un copil se joacă, extăgând la întâmplare câte un cartonaş şi aranjându-l în ordinea extragerii. Care este probabilitatea să obţinem cuvântul MATEMATICA?

Întrucât considerăm cartonaşele marcate la fel ca fiind de acelaşi tip, rezultă că avem 2 cartonaşe de tip M, 3 cartonaşe de tip A, 2 cartonaşe de tip T, 1 cartonaş de tip I, 1 cartonaş de tip E şi 1 cartonaş de tip C. Folosind formula dedusă mai sus, cardinalul spaţiului de evenimente elementare asociat experimentului este

$$card(\Omega)=\frac{10!}{3!2!1!1!1!}$$

Rezultă că probabilitatea obţinerii cuvântului MATEMATICA este egală cu

$$\frac{1}{card(\Omega)}=\frac{3!2!1!1!1!}{10!}=\frac{1}{2\cdot5\cdot6\cdot7\cdot8\cdot9\cdot10}=3.306\,9\times10^{-6}$$

Exemplul 5. Presupunem că dispunem de n cutii şi r bile identice. Plasăm bilele, una câte una, la întâmplare, în una din cutii. Să se calculeze cardinalul multimii tuturor variantelor de plasare a bilelor în cutii.

Soluţie. În cele ce urmează vom reprezenta n cutii prin intermediul a $n+1$ bare verticale, iar r bilel prin intermediul a r asteriscuri. De exemplu, situaţia când 5 bile identice, fiind plasate in 3 cutii astfel încât in prima cutie nimeresc 0 bile, in cutia a doua 2 bile şi în cutia a treia 3 bile poate fi reprezentată astfel:

$$|| ** | *** |;$$

iar situaţia când toate bilele nimeresc in prima cutie poate fi reprezentată astfel:

$$| ***** |||\,.$$

Or, pentru o astfel de reprezentare schematică avem nevoie de $n+1$ locuri pentru bare (pereţii cutiilor) şi r locuri pentru asteriscuri (bile). Din exemplele aduse vedem că orice repartizare concretă a r bile identice în n cutii este univoc determinata de pozitia a $n-1$ bare (pereţi) interioriare si a r asteriscuri (bile) pe cele $r+n-1$ locuri interioare, cele două bare (pereţi) exterioare rămânând de fiecare dată fixe. Drept consecinţă alegerea a $n-1$ locuri pentru bare (sau r locuri pentru asteriscuri) din totalul de $n+r-1$ locuri, poate fi făcută în $\mathbb{C}_{n+r-1}^{n-1} = \mathbb{C}_{n+r-1}^{r}$ modalităţi, $\mathbb{C}_{n+r-1}^{r}$, find cunoscut ca numărul **combinarilor din n elemente luate câte r cu repetare.**

Exemplul 6. Intr-o cafenea sunt expuse 7 tipuri de îngheţată. O familie formată din 4 persoane (părinţii cu doi copii) a comandat la întâmplare 4 îngheţate. În câte modalităţi poate fi realizată această comanda, care este privită ca o multime neordonată formată din 4 îngheţate?

Soluţie. Observăm ca problema noastra se reduce la problema calculării numărului de *combinari din* $n=7$ *elemente luate câte* $r=4$ *cu repetare,* tipul îngheţatei fiind considerat, convenţional, cutie, iar cele 4 ingehţate-bile identice. Prin urmare, comanda în cauză poate fi realizată in $\mathbb{C}_{7+4-1}^{4} = 210$ modalităţi.

Probleme propuse.

1. Un om de afaceri are de ales unul din traseele turistice pe care le propun 2 firme turistice. Câte variante de alegere are acesta dacă:

a) Prima firmă propune 10 trasee turistice iar cea de a doua 20 de trasee, astfel încât in ofertele ambelor firme nu se regăseşte niciun traseu comun;

b) Prima firmă propune 10 trasee turistice iar cea de a doua 20 de trasee, astfel încât 5 trasee se regăsesc in ofertele ambelor firme.

2. Producătorul de autoturisme Dacia Piteşti produce mai multe versiuni de model noul Dacia Sandero Stepway care depind de unul din 2 tipuri de alimentare combustibil (benzină/motorină), 2 variante de volum de motor, 2 tipuri de transmisie (manuală sau automată), 3 pachete dotari interioare şi 4 culori vopsea pentru exterior. Cate versiuni autoturism in total trebuie să solicite de la producator un dealer autorizat care vinde acest model de autoturism?

3. Într-un autoturism cu 5 locuri, inclusiv locul şoferului, se aşează, la întâmplare, 5 persoane, din care numai 2 persoane posedă permis de conducere, fiecare persoana ocupând un singur loc. Cu ce este egal numarul total de ocupare a celor 5 locuri? Câte variante de ocupare a locurilor favorizează evnimentul că la volanul autoturismului va nimeri o persoană care posedă permis de conducere.

4. Dintr-o populaţie statistică, reprezentată de toţi studenţii unei universităţi, a fost efectuat un sondaj bazat pe un eşantion de volum 500, ce vizeaza atitudinea studenţilor faţă de fumat. Câte eşantioane diferite de volum 500 sunt posibile în total daca numărul total al studentilor este egal cu 11000, prin eşantion inţelegând orice submultime ordonată formata din 500 de studenţi aleşi la întâmplare din aceasta populaţie statistică? Câte submulţimi diferite de acest fel sunt posibile daca:

a) Eşantionarea/selecţia este făcută fără repetare, adică studentul, odată fiind ales, este exclus din lista tuturor studentilor din care se face selectarea;

b) Eşantionarea/selecţia este făcută cu repetare, adică studentul ales poate apare în eşantion în mod repetat, fiind lăsat de fiecare dată în listă.

Care este numărul total de eşantioane diferite de volum 500, posibile în acest sondaj daca în eşantion sunt incluşi *nu* studenţii aleşi, ci raspunsurile lor, *DA* sau *NU*, în funcţie de este sau nu fumător studentul chestionat? În acest caz, depinde oare acest număr de tipul de eşantionare (cu sau fara repetare)? Dar de numărul N de studenti care studiază la universitate, unde, bineînţeles, $N \geq 500$, in cazul selecţiei fără repetare?

5. Registrul Auto Român (RAR) a ales, in calitate de tip de numerotare a autoturismelor, o secvenţă formată din acronimul judeţului sau a mun. Bucureşti, funcţie de viza de reşedinţă sau adresa persoanei fizice sau juridice, ce deţine autoturismul, urmata de 2 cifre şi 3 litere (fără diacritice) al alfabetului latin. Care este numărul maxim de autoturisme care pot fi numerotate astfel intr-un judet sau mun Bucureşti? Dar numarul maxim de autoturisme care pot fi numerotate astfel in România, ştiind că România este împărţită în 41 de judeţe.

6. La pizzeria Andy's au fost aduse pentru ziua curenta, n componente alimentare pentru producerea Pizzei, turtele pentru Pizza fiind produse aparte la pizerie. Câte tipuri de Pizza pot fi produse daca fiecare tip admite, cel putin, una din aceste componente sau, cel mult, toate componentele aduse. Dar câte tipuri de Pizza pot fi produse pentru $n = 6$ daca numarul de componente folosite:

a) trebuie sa fie egal cu 3;

b) poate fi, cel putin, egal cu 1 şi, cel mult, egal cu 5.

7. Un reprezentant pentru vânzări din SUA urmează să întreprindă vizite de lucru in oraşele Omaha, Dallas, Wichita şi Oklahoma Sity, între care exista legătură directă cu avionul. Presupunem că ordinea vizitării lor este aleasă la întâmplare. Câte rezultate posibile corespund acestui experiment aleator?

8. În condiţiile problemei anterioare reprezentantul de vânzări are sarcina sa aleagă la întâmplare si să viziteze doar trei din oraşele enumerate, nu contează în ce ordine. Câte rezultate posibile corespund acestui experiment aleator?

10. Un grup de 8 participanţii la o Conferinţă Internaţională, printre care Ionescu şi Petrescu, au fost cazaţi şi repartizaţi de către organizatori, la întâmplare, intr-un hotel particular ce are numai 4 camere: una cu 3 locuri, două cu 2 locuri si una cu un singur loc. Câte rezultate posibile sunt în acest experiment aleator? Dar câte dintre aceste rezultate favorizează faptul ca Ionescu şi Petrescu vor nimeri in camera de 3 locuri?

11. Secţia Literatură Tehnica a Bibliotecii Naţionale "Vasile Alecsandri" din Chisinău are cărţi ce ţin de domeniile Matematică, Fizică, Chimie, etc., în total 16 domenii ale Ştiinţei. Experimentul rezidă în alegerea la întâmplare a 4 comenzi de carte pentru a vedea cum se repartizează acestea pe domenii. Câte rezultate posibile sunt în acest experiment aleator? Câte din aceste rezultate favorizează faptul ca toate comenzile se referă la domenii de Ştiinţa diferite? Dar câte din aceste rezultate favorizează faptul ca toate comenzile se referă la unul si acelaşi domeniu de Ştiinţa diferite?

1.3. Spaţii de evenimente elementare, evenimente aleatoare şi operaţii asupra lor, definiţia axiomatică a probabilităţii

A modela matematic un experiment aleator înseamnă, de fapt, a descrie matematic (**a**) *mulţimea de rezultate posibile* în acest experiment, (**b**) *evenimentelor aleatoare* asociate acestui experiment şi (**c**) *probabilitatea*, adică aplicaţia sau regula conform careia fiecărui eveniment aleator îi punem în corespondenţă un număr ce caracterizeaza şansele (gradul) lui de realizare. Vom începe cu dezideratul (**a**).

Spaţii de evenimente elementare

Chiar daca intr-un experiment aleator nu putem anticipa cu certitudine care rezultat anume se va produce, este firesc să presupunem ca multimea tuturor rezultatelor posibile poate fi descrisă cu exactitate. Aducem, drept confirmare, câteva exemple.

Exemplul 1. Experimentul $\mathcal{E}$ constă în aruncarea unei monede o singură dată. Mulţimea de rezultate posibile este dată de

$$\Omega = \{S, B\} = \{0, 1\} = \{\omega_1, \omega_2\},$$

unde prin $S, 0$ sau ω_1 se subînţelege "stema" iar prin $B, 1$ sau ω_2 se subînţelege "banul".

Exemplul 2. Experimentul $\mathcal{E}$ constă în aruncarea unei monede de trei ori succesiv. Mulţimea de rezultate posibile este dată de

$$\Omega = \{SSS, SSB, SBS, SBB, BSS, BSB, BBS, BBB\}.$$

Exemplul 3. Experimentul $\mathcal{E}$ constă în aruncarea zarului o singură dată. Mulţimea de rezultate posibile este dată de

$$\Omega = \{1, 2, 3, 4, 5, 6\} = \{i \mid i = \overline{1, 6}\} = \{\omega_i \mid i = \overline{1, 6}\}.$$

Exemplul 4. Experimentul $\mathcal{E}$ constă în aruncarea unei monede până la prima apariţie a stemei. Mulţimea de rezultate posibile este dată de

$$\Omega = \{S, BS, BBS, BBBS, \ldots, \underbrace{BB..B}_{n-1\ ori}S, ...\}$$

care este o *mulţime infinită numărabilă*, unde prin $\underbrace{BB..B}_{n-1\ ori}S$ se subînţelege că prima apariţie a "stemei" este precedata de apariţia "banului" de $n-1$ ori succesiv, $n \geq 1$.

Exemplul 5. Experimentul $\mathcal{E}$ constă în alegerea unui punct la întâmplare pe segmentul $[0, 1]$. Mulţimea de rezultate posibile este dată de *mulţimea infinită nenumărabilă*

$$\Omega = \{x \mid x \in [0, 1]\},$$

unde x este coordonata punctului ales.

Exemplul 6. Experimentul $\mathcal{E}$ constă în înregistrarea greutăţii unui student ales la întâmplare de la una din universităţi. Mulţimea de rezultate posibile este dată de

$$\Omega = \{x \mid x > 0\}.$$

Remarca 1. Deoarece multimile finite sau infinite, cel mult, numărabile *se mai numesc mulţimi discrete,* spunem că exemplele 1-4 se referă la cazul *discret*, în caz contrar, cum sunt exemplele 5-6, spunem că se referă la cazul *continuu*. Cu toate acestea exemplele invocate ne conduc la următoarea definiţie valabilă în *caz general*.

Definiţia 1. Vom numi *spaţiu de evenimente elementare* orice mulţime nevidă $\Omega \neq \emptyset$, ale cărei elemente reprezintă (descriu) toate rezultatele posibile într-un experiment aleator. Fiecare element al mulţimii Ω (care corespunde unui singur rezultat posibil) se numeşte *eveniment elementar.*

Remarca 2. Definiţia nu exclude situaţia, când unuia şi aceluiaşi experiment aleator îi putem asocia, în dependenţă de scopul urmărit, spaţii de evenimente elementare diferite. Astfel, în exemplul 2 dacă ne interesează numai numărul de steme apărute la aruncarea monedei de trei ori, atunci în calitate de spaţiu de evenimente elementare putem lua

$$\Omega = \{0, 1, 2, 3\}.$$

Acum suntem pregătiţi sa răspundem dezideratului (**b**).

Evenimente aleatoare şi operaţii asupra lor

Pentru a introduce noţiunea matematica cu ajutorul căreia să descriem evenimentele aleatoare vom pleca de la exemplul legat de aruncarea zarului o singură dată pentru care spaţiul corespunzător de evenimente elementare este $\Omega = \{1, 2, 3, 4, 5, 6\}$. Considerăm următoarele evenimente aleatoare pe care le vom nota, aici şi în continuare, cu primele majuscule ale alfabetului latin: $A = \{$*numărul de puncte va fi par*$\}$; $B = \{$*numărul de puncte va fi impar*$\}$; $C = \{$*numărul de puncte va fi mai mic sau egal cu* $4\}$; $D = \{$*numărul de puncte va fi mai mic decât* $13\}$; $E = \{$*numărul de puncte va fi mai mare decât* $13\}$; $F = \{$*numărul de puncte va fi divizibil cu* $3\}$. Observam că putem stabili următoarea corespondenţă biunivocă: $A \longleftrightarrow \{2, 4, 6\}$; $B \longleftrightarrow \{1, 3, 5\}$; $C \longleftrightarrow \{1, 2, 3, 4\}$; $D \longleftrightarrow \{1, 2, 3, 4, 5, 6\} = \Omega$; $E \longleftrightarrow \emptyset$; $F \longleftrightarrow \{3, 6\}$.

În concluzie, dacă experimentului aleator îi corespunde un spaţiu discret de evenimente elementare Ω, atunci evenimentele aleatoare pot fi descrise ca fiind submulţimi ale lui Ω. Aşadar, având un spaţiu de evenimente elementare Ω, exemplul nostru arata ca putem, cel puţin in caz discret, formula

Definiţia 2. Vom numi *eveniment aleator* (**în caz discret**) orice submulţime $A \subseteq \Omega$. În particular, Ω se va numi *evenimentul sigur*, iar $\emptyset$ se va numi *eveniment imposibil*. Spunem că evenimentul elementar $\omega \in \Omega$ favorizează apariţia evenimentului aleator A dacă şi numai dacă $\omega \in A$.

Astfel, in exemplul de mai sus, evenimentul $D = \Omega$ şi se numeşte *eveniment sigur*, iar evenimentul E - *eveniment imposibil*.

Or, operaţiile care pot fi aplicate asupra mulţimilor pot fi aplicate şi asupra evenimentelor aleatoare.

Definiţia 3. *Sumă (reuniune)* a două evenimente aleatoare $A, B \subseteq \Omega$ vom numi evenimentul

$$C = A \cup B = \{\omega \in \Omega \mid \omega \in A \textit{ sau } \omega \in B\},$$

adică C se produce dacă *sau* se produce A, *sau* se produce B, *sau* se produc A *şi* B concomitent . Cu alte cuvinte, *suma* evenimentelor A şi B se produce atunci şi numai atunci când se produce *cel puţin unul* din aceste două evenimente.

În exemplul nostru: $A \cup F = \{2, 3, 4, 6\}$, $A \cup B = \Omega$.

Definiţia 4. *Produs (intersecţie)* a două evenimente aleatoare $A, B \subseteq \Omega$ vom numi evenimentul

$$C = A \cap B = \{\omega \in \Omega \mid \omega \in A \textit{ şi } \omega \in B\},$$

adică C se produce dacă se produce *şi* A *şi* B. Cu alte cuvinte evenimentul *produs al* evenimentelor A şi B are loc atunci şi numai atunci când aceste două evenimente se produc concomitent. Produsul a două evenimente A şi B se mai notează $A \cdot B$ sau AB.

În exemplul nostru: $A \cap F = \{6\}$, $A \cap B = \emptyset$.

Definiţia 5. Pentru orice $A \subseteq \Omega$ vom numi *eveniment "non-A"* sau *eveniment opus evenimentului A* complementara acestei submulţimi în raport cu Ω, adică evenimentul notat cu $\bar{A}$ sau A^c, unde

$$A^c = \{\omega \in \Omega \mid \omega \notin A\}$$

Or, evenimentul $\bar{A}$ se produce atunci şi numai atunci când *nu* se produce evenimentul A.

În exemplul nostru: $\overline{A} = B$, $\overline{B} = A$, $\overline{D} = E$, $\overline{E} = D$.

Definiţia 6. Dacă avem două evenimente aleatoare $A, B \subseteq \Omega$ spunem că evenimentul A *implică* evenimentul B dacă şi numai dacă $A \subseteq B$. Altfel spus, A *implică* B atunci şi numai atunci când din faptul că s-a produs evenimentul A rezultă că s-a produs şi evenimentul B. Dacă $A \subseteq B$ şi $B \subseteq A$, atunci spunem că $A = B$ (A este *echivalent* cu B).

Cum operaţiile asupa evenimentelor aleatoare sunt, de fapt, operaţii asupra mulţimilor din Ω putem formula

Propoziţia 1. *Operaţiile asupra evenimentelor aleatoare au următoarele proprietăţi*: $A \cup A = A$; $A \cup \Omega = \Omega$; $A \cap A = A$; $A \cap \Omega = \Omega$; $A \cup \emptyset = A$; $A \cap \emptyset = \emptyset$; $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$; $\overline{A \cup B} = \overline{A} \cap \overline{B}$; $\overline{A \cap B} = \overline{A} \cup \overline{B}$.

Remarca 3. Ultimele două proprietăţi se numesc formulele de *dualitate* ale lui De Morgan. Prin analogie, operaţiile de *sumă* şi *produs* pot fi extinse asupra evenimentelor $A_i \subseteq \Omega$, $i = 1,2,\ldots$,adică putem opera cu evenimentele : $\bigcup\limits_{n=1}^{k} A_n$, $\bigcup\limits_{n=1}^{\infty} A_n$, $\bigcap\limits_{n=1}^{k} A_n$, $\bigcap\limits_{n=1}^{\infty} A_n$, unde $k \geq 2$.

Definiţia 7. Dacă $A, B \subseteq \Omega$ sunt astfel încât $A \cap B = \emptyset$, adică A, B nu se pot produce concomitent, atunci spunem că A şi B sunt *evenimente incompatibile* (sau *disjuncte*).

Se verifică cu uşurinţă că în caz discret familia $\mathcal{F} = \{A \mid A$ *este eveniment aleator legat de experimentul aleator căruia îi corespunde spaţiul de evenimente elementare* $\Omega\}$ coincide cu familia tuturor submultimilor spaţiului de evenimente elementare Ω, adica cu $\mathcal{F} = \{A \mid A \subseteq \Omega\}$ şi aceasta verifică următoarele axiome (proprietăţi):

1. Dacă $A \subseteq \mathcal{F}$, atunci $\overline{A} \in \mathcal{F}$;

2. Dacă $A_1, A_2, ..., A_n, ... \in \mathcal{F}$, atunci $\bigcup\limits_{i=1}^{\infty} A_i \in \mathcal{F}$.

Definiţia 8. Familia $\mathcal{F}$ de submulţimi ale lui Ω care verifică axiomele de mai sus se numeşte câmp (borelian) de evenimente. Elementele lui $\mathcal{F}$ se numesc evenimente aleatoare, iar spaţiul de evenimente elementare Ω înzestrat cu un câmp de evenimente $\mathcal{F}$ se numeşte *spaţiu măsurabil* şi se notează prin $(\Omega, \mathcal{F})$.

Această din urmă definitie a noţiunii de eveniment aleator este acoperitoare pentru ambele cazuri, discret si continuu, adica este valabilă în caz general. Mai mult, are loc următoarea

Propoziţie 2. *Dacă* $(\Omega, \mathcal{F})$ este un *spaţiu măsurabil atunci au loc următoarele proprietăţi:*

a) Evenimentele sigur şi imposibil fac parte si ele din câmpul de evenimente aleatoare, adică $\Omega \in \mathcal{F}$, $\varnothing \in \mathcal{F}$;

b) Dacă $A_1, A_2, ..., A_n, ... \in \mathcal{F}$ *, atunci şi* $\bigcap\limits_{i=1}^{\infty} A_i \in \mathcal{F}$.

Cu alte cuvinte, dacă familia $\mathcal{F}$ de submulţimi ale lui Ω este un câmp (borelian) de evenimente, atunci garantat că evenimente aleatoare vor fi mulţimile Ω şi $\varnothing$ cât si complementarele tuturor submulţimilor din Ω care fac parte din $\mathcal{F}$, dar si intersectia (produsul) tuturor submultimilor A_1,A_2,...,A_n,... din Ω ce fac parte din $\mathcal{F}$, nu numai reuniunea (suma) lor. Lucrul acesta este suficient pentru a modela matematic, toate situatiile care prezintă interes din punct de vedere practic in legatura cu evenimentele aleatoare pentru care dorim sa definim noţiunea de probabilitate.

Acum, pentru a întregi modelarea matematica a experimentelor aleatoare, avem la dispozitie toate datele pentru a defini noţiunea generală de probabilitate, răspunzând, astfel, la dezideratul (**c**).

Pentru a da o definitie cat mai generală, aşa cum arată ilustrul matematician rus A. N. Kolmogorov, părintele teoriei moderne a probabilităţilor, se impune o abordare axiomatică. De notat ca abordarea axiomatica la construirea unei teorii nu este straină, de exemplu, Mecanicii Teoretice sau chiar şi stiinţelor economice. Astfel, Teoria Consumului se bazeaza pe un set de supoziţii comportamentale ale consumatorilor, supoziţii considerate ca fiind adevăruri axiomatice. În cazul nostru, deoarece Probabilitatea este privită ca masură a gradului de realizare a oricarui eveniment aleator asociat unui experiment aleator, este firesc sa o privim ca pe o funcţie(aplicaţie) definită pe multimea/familia tuturor evenimentelor aleatore legate de acest experiment, adica pe multimea $\mathcal{F}$ definită mai sus.

Aşadar, fie $(\Omega, \mathcal{F})$ un spaţiu masurabil, aunci putem formula

Definiţia axiomatică a probabilităţii. Vom numi *probabilitate* orice aplicaţie (funcţie) P:$\mathcal{F} \to \mathbb{R}$ dacă aceasta satisface

Axioma 1. Pentru orice eveniment $A \in \mathcal{F}$ probabilitatea lui $P(A) \geq 0$;

Axioma 2. Probabilitaea evenimentului sigur $P(\Omega) = 1$;

Axioma 3. Probabilitatea producerii a cel puţin unuia din evenimentele A_1, A_2, ..., A_n,... din $\mathcal{F}$, incompatibile (disjuncte) două câte două, coincide cu suma probabilităţilor acestor evenimente, adica $P(A_1 \cup A_2 \cup ... \cup A_n \cup ...) = P(A_1) + P(A_2) + ... P(A_n) + ...$, pentru $\forall$ $\{A_n\}_{n\geq 1} \in \mathcal{F}$, $A_i \cap A_j = \varnothing$, $i \neq j$, i,$j \geq 1$.

În concluzie, probabilitatea P fiind o aplicaţie (funcţie) definită pe familia $\mathcal{F}$ a tuturor evenimentelor aleatoare asociate experimentului dat, este firesc sa numim probabilitate a evenimentului A valoarea $P(A)$ a acestei funcţii.

Remarca 4. Dacă confruntăm proprietăţile frecvenţelor relative cu Definiţia Axiomatică a Probabilităţii, deducem că aceasta din urmă este acoperitoare, de fapt, pentru proprietăţile Probabilităţii Frecvenţiale (Statistice).

Definiţia 9. Vom numi *câmp de probabilitate* sau *spaţiu de probabilitate* orice triplet de forma $(\Omega,\mathcal{F},P)$, unde $(\Omega,\mathcal{F})$ este un spaţiu măsurabil asociat unui experiment aleator iar P este o probabilitate definită pe $\mathcal{F}$.

Notiunea de *câmp de probabilitate* poate servi in calitate de model matematic (model probabilist) care descrie comportamentul probabilist al evenimentelor aleatoare asociate experimentului aleator în cauză. Drept confirmare putem aduce următoarele exemple.

Exemplul 7. Considerăm experimentul aleator $\mathcal{E}$ ce constă în aruncarea unui zar "perfect" o singura data. Atunci spatiul de evenimente elementare corespunzător $\Omega = \{1, 2, 3, 4, 5, 6\}$ descrie toate rezultatele posibile iar familia $\mathcal{F}$ =$\{A \mid A \subseteq \Omega\}$ ale tuturor submulţimilor din Ω descrie toate evenimentele aleatoare asociate acestui experiment. Observam că aplicaţia P:$\mathcal{F} \rightarrow \mathbb{R}$ conform formulei

$$P(A) = \frac{Card\ A}{Card\ \Omega} = \frac{1}{6} Card\ A,$$

este o probabilitate în sensul definiţiei de mai sus. Prin urmare tripletul $(\Omega,\ \mathcal{F},\ P)$ este un câmp de probabilitate ce descrie comportamentul probabilist al oricarui eveniment aleator asociat experimentului $\mathcal{E}$. De exemplu probabilitatea evenimentului $A = \{$*la aruncarea unui zar "perfect" o singură dată va apare un număr par de puncte*$\}$ este egala cu $P(A) = \frac{1}{6} Card$ $\{2, 4, 6\} = \frac{3}{6} = \frac{1}{2}$.

Remarca 5. Finalmente, verificarea faptului ca modelul probabilist $(\Omega,$ $\mathcal{F},\ P)$ descrie adecvat experimentul aleator $\mathcal{E}$ se poate face pe cale experimentală. Conform proprietăţii regularităţii statistice, dacă , odată cu creşterea lui n, pentru orice $A \in \mathcal{F}$, frecvenţile relative $f_n(A) \simeq P(A)$, atunci ipoteza ca modelul in cauza este adecvat experimentului $\mathcal{E}$ devine credibilă, dar validarea modelului presupune aplicarea unor metode avansate ce ţin de Statistica Matematica.

Exemplul 8. Considerăm următorul joc de noroc: Ion si Petru aruncă pe rând o monedă "perfecta", jocul terminandu-se deîndata ce unul din jucători va înregistra aparitia stemei. Primul arunca moneda Ion. Ne intereseaza, de exemplu, cine are şanse mai mari de a câştiga jocul. Conform Exemplului 3, spatiul de evenimente elementare $\Omega = \{S,\ BS,\ BBS,$

$BBBS, \ldots, \underbrace{BB..B}_{n-1\ ori}S, ...\} = \{\omega_1, \omega_2, ..., \omega_n, ...\}$. Se observă cu uşurinţă ca orice eveniment aleator asociat acestui joc este submultime a lui Ω. Prin urmare $\mathcal{F} = \{A \mid A \subseteq \Omega\}$. De exemplu evenimentele $\{jocul\ va\ fi\ câstigat\ de\ Ion\} = \{\omega_1, \omega_3, ..., \omega_{2k+1}, ...\}$, $\{jocul\ va\ fi\ câstigat\ de\ Petru\} = \{\omega_2, \omega_4, ..., \omega_{2k}, ...\}$. În continuare, asociem fiecărui eveniment elementar probabilitatea $P\{\ \omega_n\} = \frac{1}{2^n}$, $n = 1, 2, ...$ iar fiecărui eveniment $A \in \mathcal{F}$ - probabilitatea calculată după formula $P(A) = \sum\limits_{n:\omega_n \in A} P\{\ \omega_n\}$. Evident, $P(A) \geq 0$. Axioma 1 a probabilităţii este, astfel, îndeplinită. Este îndeplinită şi Axioma 2 deoarece

$$P(\Omega) = \sum_{n \geq 1} \frac{1}{2^n} = \frac{1}{2}(1 + \frac{1}{2} + \frac{1}{2^2} + ... + \frac{1}{2^n} + ..) = \frac{1}{2} \cdot \frac{1}{1 - \frac{1}{2}} = 1.$$

Cum pentru $\forall\ \{A_n\}_{n \geq 1} \in \mathcal{F}$, $A_i \cap\ A_j = \varnothing$, $i \neq j$ avem ca

$$P(A_1 \cup A_2 \cup ... \cup A_n \cup ...) = \sum_{n:\omega_n \in A_1 \cup A_2 \cup ... \cup A_n \cup ...} P(\ \omega_n) =$$

$$\sum_{n:\omega_n \in A_1} P(\ \omega_n) + \sum_{n:\omega_n \in A_2} P(\ \omega_n) + ... = P(A_1) + P(A_2) + ...$$

rezulta ca si Axioma 3 este valabilă. Prin urmare funcţia P:$\mathcal{F} \rightarrow \mathbb{R}$ definită mai sus este o probabilitate, iar $(\Omega, \mathcal{F}, P)$ un model de probabilitate pentru jocul de noroc descris anterior. Astfel,

$$P\{jocul\ va\ fi\ câstigat\ de\ Ion\} = \ P\{\omega_1, \omega_3, ..., \omega_{2k+1}, ...\} =$$

$$\frac{1}{2} + \frac{1}{2^3} + \frac{1}{2^5} + ... = \frac{1}{2} \cdot \frac{1}{1 - \frac{1}{2^2}} = \frac{2}{3}.$$

În mod similar aflăm că $P\{jocul\ va\ fi\ câstigat\ de\ Petru\} = \frac{1}{3}$. Cu alte cuvinte, Ion are de 2 ori mai multe şanse de a câştiga jocul decât Petru.

Exemplul 9. Considerăm experimentul aleator ce constă în fixarea duratei x de viaţă, masurată în ore, a memoriei hard a unui laptop marca HP. Evident, spatiul de evenimente elementare corespunzător este mulţimea $\Omega = \{x \mid x \geq 0\}$. Luăm în calitate de $\mathcal{F}$ orice familie de submulţimi ale lui Ω daca aceasta este un câmp borelian de evenimente, adică $\mathcal{F}$ contine toate intervalele din Ω de forma $(a, b]$, $[a, b)$, (a, b), reuniunile, intersectiile si complementarele unor astfel de intervale. Aceasta ne permite sa afirmăm ca aplicatia P:$\mathcal{F} \to \mathbb{R}$ dupa formula $P(A) = \int\limits_A 10^{-6} e^{-10^{-6}x} dx$, privită ca funcţie de multimi din $\mathcal{F}$, este o probabilitate. Intr-adevăr, $P(A) \geq 0$, deoarece pentru orice $x \in A \subseteq \Omega$ funcţia de sub integrală este nenegativă, iar

$$P(\Omega) = \int\limits_{\Omega} 10^{-6} e^{-10^{-6}x} dx = \int\limits_0^{+\infty} 10^{-6} e^{-10^{-6}x} dx = (1 - e^{-10^{-6}x}) \mid_0^{+\infty} = 1,$$

şi din proprietăţile integralei rezultă că

$$P(A_1 \cup A_2 \cup ... \cup A_n \cup ...) = \int\limits_{A_1 \cup A_2 \cup ... \cup A_n \cup ...} 10^{-6} e^{-10^{-6}x} dx =$$

$$\int\limits_{A_1} 10^{-6} e^{-10^{-6}x} dx + \int\limits_{A_2} 10^{-6} e^{-10^{-6}x} dx + ... + \int\limits_{A_n} 10^{-6} e^{-10^{-6}x} dx + ... =$$

$$P(A_1) + P(A_2) + ... P(A_n) + ...,$$

pentru $\forall \{A_n\}_{n \geq 1} \in \mathcal{F}, A_i \cap A_j = \varnothing, i \neq j, i, j \geq 1$.Aceasta inseamnă ca P verifica Axiomele 1-3 din definiia probabilităţii. Or, $(\Omega$,$\mathcal{F}$,$P)$ poate fi privit ca un model matematic ce descrie comportamentul probabilist al duratei vieţii memoriei hard a unui laptop marca HP. Astfel, probabilitattea că acesta va avea o durată de viaţă mai mare decât T este egala cu

$$\int\limits_{(T,+\infty)} 10^{-6} e^{-10^{-6}x} dx = \int\limits_T^{+\infty} 10^{-6} e^{-10^{-6}x} dx = (1 - e^{-10^{-6}x}) \mid_T^{+\infty} = e^{-10^{-6}T}.$$

Astfel, pentru $T = 43800$ ore, adica aproximativ 5 ani, aceasta probabilitate este egala cu $e^{-10^{-6} \times 43800} = 0.957\,15$, ceea ce denotă o fiabilitate sporită pentru memoria hard a unui calculator marca HP. În ce măsură modelul matematic descris mai sus corespunde realităţii aceasta esta o problemă ce ţine de Statistica Matematică.

Probleme propuse.

1.Pentru fiecare dintre experimentele aleatoare de mai jos descrieţi spaţiul de evenimente elementare corespunzător.

a. La încheerea activităţii zilnice, contabilitatea centrului comercial UNIC duce evidenţa numarului de tranzacţii in numerar ai primilor 100 de clienţi.

b. Biroul Naţional de meteorologie inregistrează zilnic datele temperaturii pentru fiecare localitate. Interes prezintă, de exemplu, perechea de date ce constă din temperatura minimă si temperatura maximă exprimată Celsius.

c. Compania Rompetrol asigură toate staţiile sale de alimentare cu combustibil inclusiv cu benzină fara plumb. Interes prezinta cantitatea totală de benzină fără plumb solicitată de toate staţiile de alimentare Rompetrol din Republica Moldova.

d. Managerul de birou al unei companii specializate in servicii de copiere contorizează numărul de copii pe care le produce o maşină de copiat inainte de a suferi un blocaj de hârtie. Interes prezintă anume acest număr de copii.

2. Pentru fiecare din exemplele aduse in problema anterioară sa se specifice tipul de mulţime a rezultatelor posibile (multime finită, infinită numărabilă-caz discret, infinită nenumărabilă-caz continuu). Argumentaţi răspunsurile.

3. Pentru fiecare din exemplele **a-d,** aduse in problema **1**, sa se descrie multimea (famiia) $\mathcal{F}$ de evenimente aleatoare (câmpul borelian de evenimente) pe care-l putem asocia (în mod firesc) experimentului corespunzător.

4. O firmă din Chişinău specializată în vânzarea autoturismelor noi japoneze are următoarea echipă de comercianţi:

Prenume	Experienţă de vânzare	Vârsta	Studiile	Căsătorit
Ion	4	34	liceale	*Da*
Doina	12	31	liceale	*Nu*
Mihai	21	56	superioare	*Da*
Gheorghe	9	42	liceale	*Da*
Elena	3	24	superioare	*Nu*
Mircea	7	29	liceale	*Nu*
Raluca	12	44	superioare	*Da*
Nicolae	2	25	liceale	*Nu*

Un client venit sa discute procurarea unui autoturism nou alege la intâmplare un comerciant. Alegând metoda potrivita de asociere a probabilitătilor, calculaţi probabilităţile următoarelor evenimente:

a. Va fi aleasă o persoana de sex feminin; **b.** Va fi ales un barbat cu vârsta sub 40 de ani; **c.** Va fi aleasă o persoană cu o vechime de activitate în domeniul vânzărilor de cel puţin 10 ani; **d.** Va fi aleasă o persoană cu studii superioare şi căsătorită; **e.** Va fi aleasă o persoană de sex feminin , cu studii liceale şi cu experienţă de activitate în domeniul vânzărilor de cel puţin 5 ani; **f.** Va fi aleasă o persoană cu o vechime de activitate în domeniul vânzărilor de cel puţin 2 ani şi vârsta de cel puţin 21 de ani; **g.** Va fi aleasă o persoană cu vârsta sub 60 de ani; **h.** Va fi aleasă o persoană cu vârsta su 24 de ani.

5. Pentru fiecare din următoarele exemple de triplete de forma $(\Omega, \mathcal{F}, P)$ determinaţi dacă acesta reprezintă un model probabilist (câmp de probabilitate).

a) Spatiul de evenimente elementare $\Omega = \{1, 2, 3, 4, 5, 6, 7, 8\}$, câmpul de evenimente aleatoare $\mathcal{F} = \{A : A \subseteq \Omega\}$, aplicaţia (funcţia) P definită pe $\mathcal{F}$ este dată de formula $P(A) = \sum\limits_{k\in\Omega : k\in A} \frac{k}{36}$, $\forall A \in \mathcal{F}$; b) $\Omega = [0, +\infty)$, $\mathcal{F} = \{A : A$ este un subinterval a lui Ω sau submultime a lui Ω ca reuniune, intersectie sau complementara unor astfel de intervale$\}$, $P(A) = \int\limits_A e^{-x} dx$, $\forall A \in \mathcal{F}$; c) $\Omega = \{1,\ 2,\ ...,\ n,\ ...\}$, $\mathcal{F} = \{A : A \subseteq \Omega\}$, $P(A) = \sum\limits_{k\in\Omega : k\in A} \frac{k^2}{10^5}$, $\forall A \in \mathcal{F}$; d) $\Omega = (0, 1)$, $\mathcal{F} = \{A : A$ este un subinterval a lui Ω sau submultime a lui Ω ca reuniune, intersectie sau complementara unor astfel de intervale$\}$, $P(A) = \int\limits_A 12x(1-x)^2 dx$, $\forall A \in \mathcal{F}$.

1.4. Proprietăţile probabilităţii drept consecintă din definiţia axiomatică a probabilităţii

Fie $(\Omega,\mathcal{F},P)$ un spatiu de probabilitate. Din definiţia axiomatică a probabilităţii P, privită ca funcţie de submulţimi (evenimente) ale spaţiului de evenimente elementare, rezulta un set de proprietăţi generale ale probabilităţii ce simplifica procesul de calcul ale unor probabilităţi ale unor evenimente aleatoare ce se exprima prin alte evenimente aleatoare ale caror probailitati sunt cunoscute sau se identifică mai uşor. Le vom centraliza in următoarea

Propoziţie (*Proprietăţile probabilităţii*). *Orice probabilitate P definită pe campul de eveneminte aleatoare $\mathcal{F}$ posedă următoarele proprietăţi:*

a) $0 \le P(A) \le 1$ pentru orice eveniment A din $\mathcal{F}$;

b) $P(\bar{A}) = 1 - P(A)$ sau în formă echivalentă, $P(A) = 1 - P(\bar{A})$, pentru orice eveniment A din $\mathcal{F}$;

c) Probabilitatea evenimentului imposibil este egala cu zero, cu alte cuvinte $P(\varnothing) = 0$;

d) **(*Formula adunării probabilităţilor*)**. *Dacă A şi B sunt două evenimente aleatoare legate de unul şi acelaşi câmp de evenimente $\mathcal{F}$*, atunci

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

sau, în caz general, dacă $A_1, A_2, ..., A_n$ sunt n evenimente aleatoare legate de unul şi acelaşi câmp de evenimente F, atunci

$$\begin{aligned} P(\overset{n}{\underset{i=1}{\cup}} A_i) &= \sum_{i=1}^{n} P(A_i) - \sum_{1 \leq i < j \leq n} P\ (A_i \cap A_j) + ... \\ &+(-1)^{m-1} \cdot \sum_{1 \leq i_1 < i_2 < ... i_m \leq n} P\ (A_{i_1} \cap A_{i_2} \cap ... A_{i_m}) + ... \\ &+(-1)^{n-1} \cdot P\ (A_1 \cap A_2 \cap ... A_n) \end{aligned}$$

e) Dacă A şi B sunt evenimente din $\mathcal{F}$ şi A implică B, adică $A \subseteq B$, atunci $P(A) \leq P(B)$ şi $P(B \setminus A) = P(B) - P(A)$;

d) Dacă A şi B sunt evenimente din $\mathcal{F}$, atunci

$$P(AB) \leq P(A) \leq P(A \cup B) \leq P(A) + P(B),$$

$$P(AB) \leq P(B) \leq P(A \cup B) \leq P(A) + P(B).$$

Următoarea teoremă vine sa confirme faptul noţiunea de Probabilitate Clasica studiată la nivel de liceu este un caz particular al noţiunii de Probabilitate Axiomatică.

Teoremă (*Probabilitate Clasică*). *Daca spatiul de evenimente elementare Ω corespunzător unui experiment aleator $\mathcal{E}$ constă din n rezultate posibile ce au aceleaşi şanse de realizare (adică sunt echiprobabile) iar A este un eveniment aleator ce contine k elemente, atunci probabilitatea evenimentului dat ce se calculează după formula probabilităţii clasice*

$$P(A) = \frac{card\ A}{card\ \Omega} = \frac{k}{n}$$

este o probabilitate şi în sens axiomatic.

Demonstraţie. Conform condiţiilor teoremei, spatiul corespunzător de probabilitate poate fi descris ca fiind tripletul $(\Omega,\ \mathcal{F},\ P)$, unde $\Omega = \{\omega_1,$ $\omega_2,\ ....\ ,\omega_n\ \}$, câmpul de evenimente fiiind $\mathcal{F} = \{A \mid A \subseteq \Omega\}$. Atunci putem arăta că aplicaţia $P : \mathcal{F} \longrightarrow \mathbb{R}$, unde $P(A)$ se calculează după formula clasică a probabilităţii satisface axiomelor 1-3 ale probabilităţii. Într-adevăr, valabilitatea Axiomei 1 este evidentă. Deoarece evenimente elementare sunt echiprobabile, adică $P\{\omega_1\} = ... = P\{\omega_n\} = 1/n$, rezultă că

$$P(\Omega) = 1 = P(\bigcup_{i=1}^{n}\{\omega_i\}) = P(\omega_1) + P(\omega_2) + ...P(\omega_n) = n \cdot \frac{1}{n} = 1,$$

ceea ce confirma valabilitatea Axiomei 2.

Pentru a arăta că este valabilă Axioma 3, deoarece $card\ \mathcal{F} = 2^n$, ne vom referi la orice multime finita de evenimente $A_1,\ A_2,\ ...,\ A_k$ din $\mathcal{F}$, disjuncte două câte două. Cum

$$\bigcup_{j=1}^{k} A_j = \left\{\omega \in \Omega\ \middle|\omega \in \bigcup_{j=1}^{k} A_j\right\} = \bigcup_{j=1}^{k}\ \{\omega \in \Omega\ |\omega \in A_j\},$$

$$A_i \cap A_j = \emptyset,\ \forall i \neq j, i, j = \overline{1, k.}$$

conform Principiului Adunării (caz disjunct), avem egaliatatea $Card(\bigcup_{j=1}^{k} A_j) = \sum_{j=1}^{k} CardA_j$. Prin urmare

$$P(\bigcup_{j=1}^{k} A_j\}) = \frac{Card(\bigcup_{j=1}^{k} A_j)}{Card\ \Omega} = \frac{\sum_{j=1}^{k} CardA_j}{Card\Omega} = \sum_{j=1}^{k} P(A_j). \qquad \square$$

Remarcă. Teorema de mai sus arată că definiţia Axiomatică a Probabilităţii este acoperitoare pentru Definiţia Clasică a Probabilităţii. Echiprobabilitatea evenimentelor elementare poate fi postulată, prin urmare aceasta din urma poate fi utilizată atunci când intr-o problemă de calcul a probabilităţilor avem de a face cu experimente de genul aruncarii unei monede "perfecte"sau unui zar "perfect", de alegere "la întâmplare" a unui element dintr-o multime finită de elemente, etc. Validarea acestor presupuneri se poate verifica doar pe cale experimentală, comparând rezultatele obtinute pe cale teoretica cu cele obtinute pe cale experimentală.

Exemplul 1. Considerăm aruncarea unui zar perfect de n ori. Cu ce este egală probabilitatea că faţa 6 va apare cel puţin o dată?

Soluţie. Spaţiul de evenimente elementare poate fi descris ca fiind multimea de seturi ordonate $\Omega = \{(i_1, i_2, ..., i_n)\} \mid i_j = \overline{1,6},\ j = \overline{1,n}\}$. Observăm că multimea Ω coincide cu produsul cartezian al multimii $\{1,2,3,4,5,6\}$ cu ea însăşi de n ori. Prin urmare $Card\,\Omega = 6^n$, iar din faptul ca zarul e perfect conchidem toate evenimentele elementare sunt echiprobabile, având probabilitatea $\frac{1}{6^n}$. Pentru a calcula probabilitatea evenimentului $A = \{$*la aruncarea unui zar perfect de n ori faţa* 6 *va pare cel puţin o dată*$\}$ observăm că este mai uşor sa aplicăm proprietaea *b)* a probabilităţii. Intr-adevăr, evenimentul opus $\overline{A}$ poate fi descris ca fiind $\overline{A} = \{(i_1, i_2, ..., i_n)\} \mid i_j = \overline{1,5},\ j = \overline{1,n}\}$, prin urmare $Card\,(\overline{A}) = 5^n$. Folosind definiţia clasică a probabilităţii, găsim că $P(\overline{A}) = \frac{5^n}{6^n} = (\frac{5}{6})^n$. Deci $P(A) = 1 - P(\overline{A}) = 1 - (\frac{5}{6})^n$.

Exemplul 2. La examenul de "Probabilităţi, Statistică Matematică" sunt propuse 24 de bilete de examinare din care 20 de bilete "norocoase" şi 4−"nenorocoase". La examen s-au prezentat 24 de studenţi, fiecare extrăgând la întâmplare, fără repetare, cate un bilet. Care student are, in ordinea extragerii, probabilitatea mai mare de a extrage un bilet "norocos", primul, al doilea,..., sau ultimul?

Soluţie. Presupunem ca biletele "norocoase" sunt numerotate de la 1 până la 20 iar cele "nenorocoase " de la 21 până la 24. Atunci spaţiul de evenimente elementare Ω şi evenimentele $A_k = \{$*studentul cu numărul de ordine k va extrage un bilet "norocos"*$\}$ se descriu, respectiv, ca fiind

$$\Omega = \{(i_1, i_2, ..., i_{24})\} \mid i_1 \neq i_2 \neq ... \neq i_{24},\ i_j = \overline{1,24}, j = \ \overline{1,24}\},$$

$$A_k = \{(i_1, i_2, ..., i_k, ..., i_{24})\} \in \Omega \mid \ \ i_k = \overline{1,20}, \}, k = \overline{1,n}.$$

Observăm ca Ω conţine toate permutările posibile ale numerelor de la 1 la 24, prin urmare $Card\,\Omega = 24!$. La fel, A_k contine toate permutările din Ω, dar care au pe locul k, $k = \overline{1,24}$ un numar "norocos" de bilet de examinare. cu alte cuvinte ne putem imagina că evenimentele elementare din Ω care favorizează evenimentiul aleator A_k pot fi obţinute drept rezultat al unei acţiuni realizate în 2 etape succesive, astfel încât la prima etapă alegem şi fixăm pe locul k un număr de bilet de la 1 până la 20, iar la etapa a doua punem pe locurile ramase orice rezultat a unei permutari din celelalte 23 numere de bilete rămase. Prima etapa poate fi realizată in 20 de modalităţi, iar cea de a doua în 23! modalităţi. In concluzie, din Principiul Înmulţirii, deducem că numărul de evenimente elementare care favorizează extragerea unui bilet "norocos" pentru studentul nr. k, $k=\overline{1,24}$, este egal cu $Card$ $(A_k) = 20 \times 23!$. Or, indiferent de numarul k,

$$P(A_k) = \frac{20 \times 23!}{24!} = \frac{20}{24} = \frac{5}{6}.$$

Morala: nu contează în ce ordine te prezinţi la examen, conteaza...bagajul de cunoştinţe.

Exemplul 3. *Selecţie aleatoare.* În statistica, atunci când cercetarea vizează o *populaţie statistică*, privită ca o multime de elemente omogene in raport cu o anumita caracteristică/proprietate, studiul se bazează, de fapt, pe cercetarea unei submultimi (ordonate) de n elemente a populaţiei statistice, numite *eşanion de volum n.* Pentru ca rezultatele cercetarii facute asupra eşantionului să redea, cu aproximarea dorită, fidel, proprietăţile aceleaşi caracteristici în întreaga populaţie, se cere ca eşantionul în cauză sa fie *reprezentativ.* Aceasta inseamnă, de fapt, că *selecţia este aleatoare* , adică fiecare element inclus în populaţia statistică este ales la întâmplare, cu alte cuvinte, are aceleaşi şanse de a fi selectat şi fiecare eşantion, adica orice submultime de elemente ordonate în ordinea selectării lor, are aceeaşi probabilitate de a fi extrasă din populaţia statistică. În presupunerea ca populatia statistică Ω consta din N elemente diferite vom arata că fiecare eşantion de volum n are aceeaşi probabilitate de a fi selectat indiferent de tipul selectiei *fără repetare* sau *cu repetare.* Într-adevăr:

Cazul selecţiei aleatoare fără repetare. Aceasta inseamnă ca elementul este extras la întâmplare şi nu este returnat în populaţia $\Omega = \{\omega_1, \omega_2, ..., \omega_N\}$, fiecare eşantion reprezentând o submulţime de forma $(\omega_{i_1}, \omega_{i_2}, ..., \omega_{i_n})$: $\omega_{i_1} \neq \omega_{i_2} \neq ... \neq \omega_{i_n}$, $\omega_{i_k} \in \Omega$, $i_k=\overline{1,N}$, $k=\overline{1,n}$, $n \leq N$. Este aplicabilă definiţia clasică a probabilităţii pentru care spaţiul de evenimente elementare coincide cu mulţimea $\mathcal{A}_N^n = \{(\omega_{i_1}, \omega_{i_2}, ..., \omega_{i_n}) : \omega_{i_1} \neq \omega_{i_2} \neq ... \neq \omega_{i_n}, \omega_{i_k} \in \Omega,$ $i_k=\overline{1,N}$, $k=\overline{1,n}\}$. Cum

$$Card\mathcal{A}_N^n = A_N^n = N \cdot (N-1) \cdot ... \cdot (N-n+1),$$

rezultă că fiecare rezultat posibil in cadrul unei atare eşantionări are aceeaşi probabilitate egala cu $1/A_N^n$. Dealtfel, se poate arăta, folosind metodele analizei combinatorii, ca pentru acest tip de selecţie aleatoare orice element al populaţiei poate nimeri în eşantion cu una şi aceeaşi probabilitate egală cu n/N (*Demonstraţi!*), ceea ce corespunde criteriului de reprezentativitate a eşantionului.

Cazul selecţiei aleatoare cu repetare. Aceasta inseamnă ca elementul este extras la întâmplare şi după ce este înregistrat, acesta este returnat în populaţia $\Omega = \{\omega_1, \omega_2, ..., \omega_N\}$, fiecare eşantion reprezentând o submulţime de forma $(\omega_{i_1}, \omega_{i_2}, ..., \omega_{i_n}) : \omega_{i_k} \in \Omega$, $i_k = \overline{1, N}$, $k = \overline{1, n}$. $n \geqslant 1$ Spaţiul de evenimente elementare coincide cu multimea $\Omega \times \Omega \times ... \times \Omega = \Omega^n = \{(\omega_{i_1}, \omega_{i_2}, ..., \omega_{i_n}) : \omega_{i_k} \in \Omega, i_k = \overline{1, N}, k = \overline{1, n}\}$. Conform Principiului Înmulţirii $Card$ $\Omega^n = N^n$,prin urmare fiecare rezultat posibil in cadrul unei atare eşantionări are aceeaşi probabilitate egala cu $1/N^n$. Cât priveşte probabilitatea de a nimeri în eşantion pentru orice element al populaţiei, reeşind din caracterul selecţiei, aceasta este egala cu $1/n$ (*Demonstraţi!*).

Exemplul 4. Presupunem că o familie posedă două autoturisme, starea lor fiind de aşa natură încât probabilitatea că primul autoturism va fi funcţional, este egala cu 0.8, iar pentru al doilea, aceeaşi probabilitate, este egală cu 0.4. În plus, probabilitatea că ambele autoturisme vor fi, concomitent, funcţionale, este egală cu 0.3.

a) Descrieţi evenimentele aleatoare despre care este vorba în problemă şi asociaţi-le probabilităţile corespunzătoare;

b) Cu ce este egală probabilitatea că va fi funcţional cel puţin un autoturism?

c) Dar probabilitaea ca niciun autoturism nu va fi funcţional?

Soluţie. a) Conform condiţiilor descrise în problemă putem introduce evenimentele $A = \{$*pimul autoturism va fi funcţional*$\}$, $A = \{$*al doilea autoturism va fi funcţional*$\}$. Atunci evenimentul $A \cap B = \{$*ambele autoturisme vor fi, concomitent, funcţionale*$\}$. Probabilităţile corespunzătoare vor fi: $P(A) = 0.8$, $P(B) = 0.4$, $P(A \cap B) = 0.3$;

b) Aici ne intereseaza probabilitatea evenimentului $A \cup B = \{$*cel putin unul din autoturisme va fi funcţional*$\}$. Aplicând Formula Adunării Probabilităţilor găsim că

$$P(A \cup B) = P(A) + P(B) - P(A \cap B) = 0.8 + 0.4 - 0.3 = 0.9;$$

c) Observăm că evenimentul $\overline{A \cup B} = \{$ *niciun autoturism nu va funcţiona*$\}$, prin urmare probabilitatea lui

$$P(\overline{A \cup B}) = 1 - P(A \cup B) = 1 - 0.9 = 0.1.$$

Exemplul 5. Un studiu privind gradul de angajare în câmpul muncii are la baza un sondaj aleator bazat pe un eşantion în care au fost incluse

1309 de persoane. Rezultatele au fost centralizate în următorul tablou:

	Neangajaţi în câmpul muncii	*Angajaţi în câmpul muncii*	*Total pe linie*
Bărbaţi	206	412	618
Femei	386	305	691
Total pe coloană	592	717	1309

Folosind aceste date, aflaţi probabilitatea că, *alegând la întâmplare* o persoană inclusă în eşantion, aceasta va fi *sau* o femeie *sau* o persoana neangajată în câmpul muncii?

Soluţie. Spatiul de evenimente elementare consta din 1309 evenimente elementare. Introducem următoarele evenimente aleatoare: $A = \{$ *persoana aleasă la întâmplare din eşantion va fi o femeie*$\}$ şi $B = \{$ *persoana aleasă la întâmplare din eşantion va fi una neangajată în câmpul muncii*$\}$. Atunci, deoarece alegerea este la întâmplare, conform definiţiei clasice, probabilitatea evenimentului căutat coincide cu $P(A \cup B)$. Dar $P(A) = \frac{691}{1309}$, $P(B) = \frac{592}{1309}$ iar $P(A \cap B) = \frac{386}{1309}$. Prin urmare $P(A \cup B) = P(A) + P(B) - PA \cap B) = \frac{691}{1309} + \frac{592}{1309} - \frac{386}{1309} = \frac{897}{1309} = 0.685\,26$.

Probleme propuse.

Tema: *Proprietătile probabilităţii, probabilitate clasică.*

1. Intr-o grupă de studenţi, din care face parte si studentul Pacală, fiecare student este *sau* de sex feminin *sau* are părul blond *sau* indrăgeste disciplina Matematica. In grupă sunt 20 de studente din care 12 au părul blond şi doar una din studentele cu părul blond au indrăgit Matematica. Numarul total de studenţi/studente cu părul blond este egal cu 24, din care doar 12 indrăgesc Matematica. Numarul total de studenţi/studente care indrăgesc Matematica este egal cu 17, din care 6 sunt studente. Cu ce este egala probabilitatea ca, alegând la intamplare un student din aceasta grupa, acesta va chiar Pacală?

2. Intr-un microbuz cu 17 locuri, inclusiv locul şoferului, au urcat 17 persoane, din care 4 persoane posedă permis de conducere al unui vehucul de acest tip. Cu ce este egală probabilitatea ca microbuzul va putea pleca, daca ştim ca fiecare persoană ocupă, la intamplare, unul din aceste 17 locuri?

3. O grupă de studenţi enumara 35 de studenti. Dintr-aceştea, 20 de studenţi s-au inscris în clubul sportiv al universităţii, 10 studenti s-au inscris în cercul de dansatori al universităţii, iar 10 studenţi nu s-au inscris la nicio activitate. Din aceasta grupa este ales la intamplare un student. Calculati probabilitatea ca:

a) acesta va fi unul inscris la ambele activitati;
b) acesta va fi unul inscris numai in clubul sportiv ;
c) acesta va fi unul inscris numai la cercul de dans.

4. Dintr-o sută de studenţi, 28 de studenţi cunosc limba engleza, 30-germana, 42-franceza, 8-engleza si germana, 10-engleza si franceza, 5-germana si franceza, 2-toate trei limbi. Este ales la intamplare un student. Cu ce este egală probabilitatea ca acesta nu cunoaste nicuna din aceste trei limbi.

5. Presupunem ca un zar "perfect" este aruncat o singura data. Calculati probabilitaţile urmatoarelor evenimente: $A = \{$*numarul de puncte aparute va fi egal cu* $6\}; B = \{$*numarul de puncte aparute va fi multiplu lui* $3\}; C = \{$*numarul de puncte va fi par si ,totodata, mai mare decat* $2\}$.

6. Considerăm aruncarea unui zar "perfect" de două ori succesiv. Calculati probabilitaţile urmatoarelor evenimente: $A = \{$*la ambele aruncari va apare acelasi numar de puncte*$\}$; $B = \{$*numarul de puncte aparute la prima aruncare va fi mai mare decat numarul de puncte apărute la aruncarea a doua*$\}$: $C = \{$*suma punctelor aparute la ambele aruncari va fi pară*$\}$; $D = \{$*produsul punctelor apărute la ambele aruncări va fi egală cu* $6\}$.

7.Se alege la intamplare un număr natural format din 5 cifre. Calculati probabilităţile urmatoarelor evenimente: $A = \{$*numărul citit de la stanga la dreapta sau invers, va ramâne neschimbat, ca de exemplu 13531*$\}$ $B = \{$*numărul va fi multiplu lui* $5\}; C = \{$*numărul va fi format numai din cifre pare*$\}$.

8. Considerăm o mulţime formata din primele 10 litere ale alfabetului latin. Cate alfabete formate din 3 litere putem alcătui din aceasta mulţime de litere. Cu ce este egala probabilitatea că un alfabet de acest fel ales la intâmplare va conţine litera A?

9. Dintr-un lot de 10 calculatoare, din care 3 calculatoare sunt cu defecte, sunt alese la întamplare 3. Calculaţi probabilitaţile urmatoarelor evenimente: $A = \{$*dintre cele 3 calculatoare alese, cel putin unul, va avea defecte*$\}$ $B = \{$*toate calculatoarele alese vor avea defecte*$\}; C = \{$*dintre cele* 3 *calculatoare alese exact* 2 *vor avea defecte*$\}$.

10. Un cub, ale cărui feţe sunt vopsite, a fost tăiat într-o mie de cubuleţe de aceeaşi dimensiune. Cu ce este egală probabilitatea că un cubuleţ extras la întâmplare va avea exact două feţe vopsite?

11. Un grup de 8 persoane ocupa fiecare, la intâmplare, unul din cele 8 scaune aşezate in jurul unei mese rotunde. Cu ce este probabilitatea că 2 persoane anume vor numeri alături.

12. Un grup de 8 persoane ocupă fiecare, la intamplare, unul din cele

8 scaune aşezate intr-un rand de 8 locuri. Cu ce este probabilitatea că 2 persoane anume vor numeri alături.

13. Pe 5 cartonaşe sunt scrise cifrele de la 1 pana la 5. Se aleg la întamplare, fara intoarcere, unul dupa altul, 3 cartoanaşe, acestea fiind puse alaturi de la stanga la dreapta in ordinea extragerii. Calculati probabilităţile urmatoarelor evenimente: $A = \{$*va apare numărul* 123$\}$, $B = \{$nu va apare cifra 3$\}$, $C = \{$va apare un numar par$\}$.

14. Numerele $1, 2, ..., 9$ sunt scrise in ordine aleatoare. Calculati probabilităţile urmatoarelor evenimente: $A = \{$*numerele vor apare in ordinea lor crescătoare*$\}$ $B = \{$*numerele* 1 *şi* 2 *vor nimeri alături in ordine crescătoare*$\}$; $C = \{$*pe locuri pare vor nimeri numere pare*$\}$.

15. Consideram un alfabet format din literele $a, b, c, d, m.$ Alegem la intamplare, succesiv, *cu intoarcere*, 4 litere, scriindu-le in ordinea extragerii lor. Cu ce este egală probabilitatea că vom obţine cuvântul *mama*?

16. Considerăm aruncarea unui zar "perfect" de 10 ori succesiv. Calculaţi probabilitaţile următoarelor evenimente: $A = \{$*la nici o aruncare nu va apare faţa* 6 $\}$ $B = \{$*la cel puţin o aruncare va apare faţa* 6 $\}$; $C = \{$*exact la* 3 *aruncari va apare faţa* 6$\}$.

17. Intr-un lift al unei case cu 7 nivele, la nivelul de jos, au urcat 6 pasageri. Stiind ca fiecare pasager poate iesi, la intamplare, la oricare din cele 6 nivele, calculaţi probabilităţile următoarelor evenimente: $A = \{$*toţi pasagerii vor iesi la nivele diferite*$\}$ $B = \{$*toţi pasagerii vor iesi la acelaşi nivel*$\}$; $C = \{$*la nivelele* 4, 5 *si* 6 *vor ieşi câte* 2 *pasageri*$\}$.

18. Un copil se joacă cu 11 cartonaşe pe care sunt imprimate literele *I, N, F, O, R, M, A, T, I, C, A,* aranjându-le la întamplare unul langa altul. Cu ce este egala probabilitatea că acesta va obţine, astfel, cuvântul *INFORMATICA*.

19. Considerăm aruncarea unui zar "perfect" de 6 ori succesiv. Calculaţi probabilitaţile următoarelor evenimente: $A = \{$*de trei ori va apare faţa* 1, *de doua ori faţa* 3 *si o data faţa* 6$\}$; $B = \{$*vor apare feţe diferite*$\}$; $C = \{$*de* 3 *ori va apare aceeaşi faţă*$\}$.

20. Care este probabilitatea că, jucând cu o singură variantă la LOTO-SPORT "5 din 35", nu vom fi în pierdere, adica vom caştiga ceva?

21. Într-un tren cu 3 vagoane se urcă la întâmplare 7 persoane. Care este probabilitatea că în primul vagon vor urca 4 persoane?

22. *Problema cavalerului DeMere:* De câte ori trebuie să aruncăm un zar "perfect" pentru că probabilitatea apariţiei feţei 6, cel puţin o dată, să fie mai mare decât 1/2?

23. O grupă este formată din 23 de studenţi. Calculaţi probabilitaţile următoarelor evenimente: $A = \{$*toţi studenţii vor avea zile de naştere diferite*$\}$; $B = \{$*se vor găsi, cel puţin doi studenţi care au aceeaşi zi de naştere*$\}$.**Nota.** Excludem cazul când in grupă sunt studenţi gemeni.

24. Să se arate că probabilitatea de a obţine în urma aruncării a 4 zaruri, cel puţin o singură dată faţa 1, este mai mare decât probabilitatea de a obţine după 24 de aruncări a unei perechi de zaruri cel puţin o singură dată două feţe 1. (Răspunsul explică **paradoxul cavalerului de Mere**, care considera aceste probabilităţi egale, fapt ce nu corespunde observărilor empirice).

25. $2n$ echipe de fotbal, printre care echipele Dacia si Zimbru, au fost impărtite, prin tragere la sorţi, in 2 subgrupe a cate n echipe. Deduceţi formulele de calcul pentru probabilităţile următoarelor evenimente: $A = \{$*Dacia si Zimbru vor nimeri in grupe diferite*$\}$, $B = \{$*Dacia si Zimbru vor nimeri in aceeaşi grupă*$\}$. Calculaţi aceste probabilităţi pentru $n = 20$.

26. O urnă conţine m bile albe şi n bile negre. Din această urnă făcându-se extracţii cu întoarcere, să se determine formula de calcul pentru:

(a) Probabilitatea ca primele k bile extrase să fie negre.

(b) Probabilitatea ca prima bilă albă să apară la a k-a extracţie.

(c) Probabilitatea ca printre primele k bile extrase vor fi i bile albe.

Calculati aceste probabilitati pentru $m = 5$, $n = 4$.

27. Primul rand al unei săli de Cinema are $2n$ locuri. n bărbaţi si n femei ocupa la intâmplare, fiecare, cate un loc. Deduceţi formulele de calcul pentru probabilităţile următoarelor evenimente: $A = \{$niciun barbat nu va nimeri alături de barbat$\}$, $B = \{$toti bărbatii vor nimeri alături$\}$.Calculaţi aceste probabilităţi pentru $n = 10$.

28. La un turneu de tenis s-au inscris 40 de sportivi. Prin tragere la sorţi aceştia au fost imparţiţi in 4 subgrupe a cate 10 sportivi. Cu ce este egală probabilitatea ca 4 din cei mai puternici tenismeni vor nimeri in grupe diferite.

29. O firmă producătoare de calculatoare acceptă achiziţionarea procesoarelor pentru calculatoare de la o alta firma numai dacă în urma verificarii a 5% din procesoare alese la întâmplare dintr-un lot propus spre a fi cumparat, niciunul nu va avea defecte. Presupunem că într-un lot de 1000 de procesoare, propuse spre achiziţie, se afla cinci procesoare defecte. Cu ce este egala probabilitatea că în urma controlului, lotul va fi acceptat spre a fi cumpărat?

1.5. Probabilităţi clasice, discrete si geometrice drept cazuri particulare ale definiţiei axiomatice a probabilităţii

După cum am vazut, din Teorema demonstrată în paragraful anterior, probabilitatea clasică devine un caz particular al Definiţiei Axiomatice. Folosind limbajul acesteia din urma putem formula o definiţie matematică mai stricta pentru prima şi anume

Definiţia probabilităţii clasice. *Vom spune ca avem de a face cu un câmp de probabilitate clasică* $(\Omega, \mathcal{F}, P)$ *dacă*

a) Spaţiul de evenimente elementare Ω *conţine un număr finit de evenimente elementare;*

b) Familia de evenimente aleatoare $\mathcal{F}$ *este reprezentata de toate submultimile posibile ale lui* Ω*;*

c) probabilitatea este o aplicatie P *definita pe* $\mathcal{F}$ *cu valori in multimea numerelor reale calculate conform formulei:*

$$P(A) = \frac{card\ A}{card\ \Omega},\ \forall A \in \mathcal{F}.$$

Remarca 1. Definiţia formulată astfel atrage dupa sine, în mod automat, faptul ca toate evenimentele elementare sunt echiprobabile. Intr-adevar, pentru orice eveniment elementar $\omega \in \Omega$ găsim că

$$P\{\omega\} = \frac{card\ \{\omega\}}{card\ \Omega} = \frac{1}{card\ \Omega}.$$

Însă aplicabilitatea probabilităţii clasice este limitată de condiţia a), dar şi de faptul că, chiar dacă această condiţie este valabilă, nu toate evenimentele elementare sunt echiprobablie. Urmatoarele exemple confirmă această afirmaţie.

Exemplul 1. *(Spatiul de evenimente elementare este finit, dar evenimentele elementare nu sunt echiprobabile).* Considerăm aruncarea unui zar cu centrul de greutate deformat astfel încât probabilităţile aparitiei feţelor lui se raportează ca $1:2:3:4:5:6$. Definiţia clasica nu este aplicabilă, fapt ce se poate verifica experimental cu orice zar imperfect.

Exemplul 2. *(Spatiul de evenimente elementare este infinit, dar numărabil).* Considerăm jocul de noroc descris in Exemplul 8 din p.1.3, cu singura precizare ca moneda aruncată are probabilitatea apariţiei stemei egală cu $p, 0 < p < 1$. Chiar dacă moneda era perfectă (vezi exemplul invocat), probabilitatea clasică nu este aplicabilă din cauza neîndeplinirii condiţiei a) din definiţie.

Exemplul 3. *(Evenimentele elementare sunt echiprobabile, dar spaţiul de evenimente elementare este infinit nenumarabil).* Considerăm experimentul imaginar ce constă in aruncarea la întâmplare a unui punct pe segmentul [0.1]. Sintagma "la întâmplare" ne sugerează că toate evenimentele elementare din spaţiului de evenimente elementare $\Omega = [0.1]$ sunt echiprobabile, dar probabilitatea clasică nu poate fi aplicată din acelaş motiv ca şi în exemplul anterior.

Totuşi, cum arată câmpul de probabilitate în fiecare din aceste exemple? Deoarece spaţiile de evenimente elementare din exemplele 1 şi 2 sunt mulţimi finite sau cel mult numărabile la ele se poate aplica

Definiţia probabilităţii discrete. Vom spune ca avem de a face cu o *probabilitate discretă* P dacă

a) Spaţiul de evenimente elementare Ω *reprezinta o multime finită sau infinită, cel mult, numerabila, adica* Ω=$\{\omega_1, \omega_2, ..., \omega_n\}$ *sau* Ω=$\{\omega_1, \omega_2, ..., \omega_n, ...\}$ *;*

b) Câmpul de evenimente aleatoare $\mathcal{F}$ *este reprezentat de toate submulţimile posibile ale lui* Ω*;*

c) P *este o aplicaţie definită pe* $\mathcal{F}$ *cu valori in mulţimea numerelor reale calculate conform formulei:*

$P(A)$=*suma probabilităţilor pentru fiecare eveniment elementar ce favorizează evenimentul* $A=\sum\limits_{n:\omega_n\in A} P\{\ \omega_n\}$ *, unde* P *verifică următoarele* 2 *axiome:*

A1. $P\{\omega_i\} \geq 0$, *pentru orice* $i \geq 1$;

A2. $P(\Omega) = 1$;

$P(A)$, reprezentând un numar, se numeste *probabilitatea evenimentului* A, iar tripletul (Ω,$\mathcal{F}$, P) - *câmp de probabilitate discretă.*

Remarca 2. Se poate arăta că orice *câmp de probabilitate axiomatică* aplicat la cazul discret este, de fapt un *câmp de probabilitate discretă* şi viceversa, probabilitatea discretă fiind, întucâtva, mai uşor de manipulat în calcule.

Exemplul 1 (*Continuare*). Spaţiul de evenimente elementare fiind $\Omega = \{1, 2, ..., 6\}$, câmpul de evenimente $\mathcal{F}$ *este reprezentat de toate submultimile posibile ale lui* Ω. Aceasta inseamna valabilitatea condiţiilor *a)* şi *b).* Pe deoparte, din faptul că $P\{1\} : P\{2\} : ...: P\{6\}$=$1 : 2 : 3 : 4 : 5 : 6$ rezultă că $P\{2\} = 2P\{1\}$, $P\{3\} = 3P\{1\}$, ..., $P\{6\} = 6P\{1\}$. Pe de alta parte $1 = P(\Omega) = P(\{1\} \cup \{2\} \cup ... \cup \{6\}) = P\{1\}+P\{2\}+ ... +P\{6\}$. Prin urmare $P\{1\}+P\{2\}+ ... +P\{6\} = P\{1\}+2P\{1\}+ ... +6P\{1\} = 21P\{1\} = 1$, adica $P\{1\} = \frac{1}{21}$, $P\{2\} = \frac{2}{21}$, ... ,$P\{6\} = \frac{6}{21}$. Axiomele 1,2 ale probabilitatii discrete fiind întrunite, rezultă, de exemplu ca $P(A) =$

$P\{$*la o singură aruncare a zarului va apare faţa pară*$\}=P\{2,4,6\}=\frac{2+4+6}{21}=\frac{12}{21}$. Vedem că, spre deosebire de cazul zarului simetric, $P(A)>\frac{1}{2}$.

Exemplul 2 (*Continuare*). Deoarece experimentul constă în aruncarea unei monede până la prima apariţie a stemei, putem considera rezultat elementar numărul total de aruncări efectuate, adică $\Omega=\{1,\ 2,\ ...,\ n,\ ...\}$. Atunci câmpul de evenimente $\mathcal{F}$ *va fi reprezentat de toate submultimile posibile ale lui* Ω. Nu prezintă greutate să se verifice experimental cu orice monedă ca probabilităţile $P\{k\}=P\{$*până la prima apariţie a stemei vor fi efectuate k aruncări*$\}=p(1-p)^{k-1}$, $k=1,\ 2,\ ...$, deîndată ce probabilitatea apariţiei stemei la o singură aruncarea monedei se ştie ca este egala cu p, $0<p<1$. Cum $P\{k\}\geq 0$, iar $P(\Omega)=P\{1,\ 2,\ ...,\ n,\ ...\}=P\{1\}+P\{2\}+ ...\ P\{n\}+...=p+p(1-p)^1+...+p(1-p)^{n-1}+...=p\cdot\frac{1}{1-(1-p)}=1$, rezultă ca putem aplica Definiţia probabilităţii discrete. Astfel, dacă în jocul de noroc descris in exemplul 8, p.1.3, moneda nu este perfecta, adică are probabilitatea apariţiei stemei egală cu p, $0<p<1$, $p\neq\frac{1}{2}$, atunci, să zicem, probabilitatea $P(A)=P\{jocul\ va\ fi\ câstigat\ de\ Ion\}=P\{1\}+P\{3\}+...+P\{2k-1\}+...= =p+p(1-p)^2+...+p(1-p)^{2k-2}+...=p\cdot\frac{1}{1-(1-p)^2}$. De exemplu, pentru $p=1/3$, avem că $P(A)=\frac{1}{3}\cdot\frac{1}{1-(1-\frac{1}{3})^2}=\frac{3}{5}$, adica probabilitatea $P(A)$ e mai mică decât atunci când moneda era simetrică, dar oricum mai mare decât probabilitatea că jocul va fi câştigat de către rivalulul lui Ion, Petru.

Cât priveşte experimentele aleatoare similare celui din Exemplul 3, acestea pot fi modelate matematic apelând la

Definiţia probablitaţii geometrice. Vom spune ca avem de a face cu o *probabilitate geometrică* P dacă

a) Spaţiul de evenimente elementare Ω *reprezintă o multime infinită nenumărabilă din* $\mathbb{R}^n$ *pentru care* $mes\Omega<+\infty$*, unde mes reprezinta lungimea in* $\mathbb{R}^1$*, aria in* $\mathbb{R}^2$ *sau volumul in* $\mathbb{R}^n$ *pentru* $n\geq 3$;

b) Câmpul de evenimente aleatoare $\mathcal{F}$ *este reprezentat de toate submulţimile măsurabile* A *ale lui* Ω*, adica pentru care* $mesA$ *poate fi definită*;

c) P *este o aplicaţie definită pe* $\mathcal{F}$ *cu valori in mulţimea numerelor reale calculate conform formulei*:

$$P(A)=\frac{mesA}{mes\Omega}.$$

Tripletul $(\Omega,\mathcal{F},P)$ poartă denumirea, în acest caz, de *spaţiu de probabilitate geometrică.*

Astfel, in exemplul invocat, aplicând definitia probabilităţii geometrice aflăm că probabilitatea că un punct aruncat la intamplare pe $[0,1]$ va nimeri in punctul x este egală cu $P\{x\} = mes\{x\}/mes([0,1]) = 0/1 = 0$, pentru orice x din $[0,1]$. Daca ne intereseaza, de exemplu, probabilitatea ca un punct aruncat la intamplare pe $[0,1]$ va nimeri in prima jumătate a acestui interval este egală cu $P([0,0.5]) = mes([0,0.5])/mes([0,1]) = 0.5/1 = 0.5$. Dealtfel, observăm că $P([0,0.5] = P([0,0.5)) = P([0.5,1])$. In genere, probabilitatea ca un punct aruncat la intamplare pe $[0,1]$ va nimeri intr-un interval (a,b) din $[0,1]$ coincide cu lungimea acestui interval.

Remarca 3. Exemplul analizat arată ca proprietatea *c)* a probabilităţii (vezi Proprietaţile Probabilităţii din p.1.4.), conform careia, dacă evenimentul $A = \varnothing$, $P(A) = 0$, arată că reciproca acestei proprietăţi nu are loc, adică există exemple de evenimente aleatoare $A \neq \varnothing$, dar care au $P(A) = 0$. Cu alte cuvinte, afirmaţia ca probabilitatea unui eveniment este egala cu zero nu atrage dupa sine afirmatia că acest eveniment este imposibil. Un alt exemplu, probabilitatea ca un fir de cablu electric, situat intre doi stalpi, se va rupe, in timpul unei furtuni, exact la mijloc este egală cu 0, dar evenimentul in cauză nu este imposibil. Aceste exemple fac deosebirea dintre notiunea teoretică (ideală) a probabilităţii şi cea empirică, cum ar fi probabilitatea frecvenţială.

Probleme propuse.

Tema: *Probabilităţi discrete.*

1. Considerăm aruncarea o singură dată a unui tetraedru regulat, ale cărui feţe sunt numerotate cu numerele de la 1 până la 4, iar centrul său de greutate este deplasat astfel încât probabilităţile apariţiei fiecărei feţe se raportează ca $P\{1\}$:$P\{2\}$: ... :$P\{4\}$ = 1:2: ... :4. Calculaţi probabilităţile următoarelor evenimente: $A_k = \{$*va apare faţa* $k\}$, $k = \overline{1,4}$; $B =$*va apare o fata pară*$\}$; $C = \{$va apare o faţă numerotată cu un numar prim$\}$.

2. Consideram aruncarea o singură dată a unui zar al cărui centru de greutate este deplasat astfel încât, probabilitatile apariţiei fiecărei dintre feţele $k = \overline{1,5}$ coincid intre ele, iar probabilitatea aparitiei fetei 6 coincide cu suma probabilitatilor anterioare. Aflati probabilitatile aparitiei pentru fiecare fata in parte, dar si probabilitatile evenimentelor $B = \{$*va apare un număr par de puncte*$\}$; $C = \{$va apare un numar prim de puncte$\}$.

3. Consideram aruncarea o singură dată a unui tetraedru regulat, ale cărui feţe sunt numerotate cu numerele de la 1 până la 4, iar centrul sau de greutate este deplasat astfel încât probabilitatile $P\{k\}$ ale apariţiei fiecărei feţe k, $k = \overline{1,4}$, sunt legate intre ele astfel: $P\{1\} : P\{2\} : P\{3\} = 1 : 2 : 3$, iar $P\{4\} = P\{1\} + P\{2\} + P\{3\}$. Calculaţi probabilităţile $P\{k\}$, $k = \overline{1,4}$, dar si probabilitatile următoarelor evenimente: $B = \{va\ apare\ \ o\ față\ pară\}$; $C = \{va\ apare\ o\ față\ numerotată\ cu\ un\ număr\ prim\}$.

4. Presupunem că alegem la întâmplare câte o literă din cuvintele *mama* si *vama*.Descrieti spaţiul de evenimente elementare si calculaţi probabilitatea că literele extrase vor fi aceleaşi.

5. Doi jucători, *Ion* si *Petru*, practica următorul joc de noroc: primul arunca moneda *Ion*; daca apare "stema", acesta este declarat castigator; daca nu, arunca *Petru*; daca apare "stema" , acesta este declarat castigator; daca nu, din nou arunca moneda *Ion*; etc., etc., jocul se termina atunci cand unul din jucatori inregistreaza , primul, aparitia stemei. Pentru fiecare jucator aparte, aflati probabilitatea ca acesta va castiga jocul, stiind ca moneda este deformata astfel, incat "stema" apare cu probabilitatea $p, 0 < p < 1$? Exista oare vre-o valoare a lui $p, 0 < p < 1$ astfel incat Ion si Petru sa aibă şanse egale de castigare a jocului?

Indicatie: Sa se considere că probabilitatea că jocul se va termina la aruncarea k, $k = 1, 2, ...,$ este egala cu $p(1-p)^{k-1}$.

6. Trei jucători, *Ion*, *Petru* si *Mihai*, practică următorul joc de noroc: primul aruncă moneda *Ion*; dacâ apare "stema", acesta este declarat castigător; dacă nu, aruncă *Petru*; dacă apare "stema" , acesta este declarat castigator; dacă nu, aruncă moneda *Mihai*; dacă apare "stema" , acesta este declarat castigător; dacă nu, atunci din nou aruncă moneda *Ion*; etc., etc., jocul se termină atunci, cand unul din jucători inregistrează , primul, apariţia stemei. Pentru fiecare jucător aparte, aflaţi probabilitatea ca acesta va caştiga jocul, , stiind ca moneda este deformata astfel, incat "stema" apare cu probabilitatea $p, 0 < p < 1$.

Indicatie: Sa se considere că probabilitatea că jocul se va termina la aruncarea k, $k = 1, 2, ...,$ este egala cu $p(1-p)^{k-1}$.

Tema: *Probabilitate geometrică.*

7. *Problema întâlnirii.* Doua persoane si-au fixat o intalnire intre orele 12.00 şi 13.00 cu condiţia că primul sosit la locul întâlnirii îl aşteaptă pe al doilea cel mult 15 minute şi daca acesta din urma nu soseşte între timp, primul părăseşte locul întâlnirii . Considerând ca fiecare dintre persoane soseşte la locul întâlnirii in mod intâmplător, sa se calculeze probabilitatea ca întâlnirea nu va avea loc.

8. Presupunem ca autobusele de ruta dată circulă la intervale fixe de 30 de minute. Un potenţial pasager soseşte la una din staţiile acestei rute într-un moment întâmplător. Cu ce este egala probabilitatea ca acesta va aştepta, cel mult, 5 minute pana la sosirea următorului autobus? Dar probabilitatea ca acesta nu va fi nevoit sa aştepte deloc?

9. Un baston de lungimea L este rupt la întâmplare in trei segmente. Cu ce este egală probabilitatea ca din aceste segmente de bason se poate construi un triunghi.

1.6. Probabilitate condiţionată. Formula înmulţirii probabilităţilor

Fie A şi B doua evenimente aleatoare legate de acelaşi câmp de probabilitate $(\Omega, \mathcal{F}, P)$, unde $P(B) > 0$. Presupunem că acest spaţiu de probabilitate modelează comportamentul probabilist al unui experiment aleator $\mathcal{E}$. Presupunerea că $P(B) > 0$ înseamnă că într-un număr n suficient de mare de repetări a acestui experiment frecvenţa relativă a evenimentului B este mai mare ca zero, adică $f_n(B) > 0$. Întrebarea firească este, în ce masurā faptul (informaţia) că s-a produs evenimentul B influienţează probabilitatea evenimentului A ? Atunci putem vorbi despre probabilitatea evenimentului A condiţionată de evenimentul B notată $P(A/B)$ sau despre frecvenţa relativă $f_{n(B)}(A)$ a evenimentului A in $n(B)$ probe $\mathcal{E}$ în care s-a produs evenimentul B, unde, conform Principiului Regularităţii Statistice, $f_{n(B)}(A) \simeq P(A/B)$ pentru $n(B)$ suficient de mare. Dar, din acelaşi Principiu avem ca,

$$P(A/B) \simeq f_{n(B)}(A) = \frac{n(A \cap B)}{n(B)} = \frac{\frac{n(A \cap B)}{n}}{\frac{n(B)}{n}} \simeq \frac{P(A \cap B)}{P(B)}.$$

Aceasta justifică următoarea

Definiţie. Se numeşte probabilitate a evenimentului A condiţionată de evenimentul B, $P(B) > 0$, mărimea notată cu $P(A/B)$ şi calculată după formula

$$P(A/B) = \frac{P(A \cap B)}{P(B)}.$$

Exemplul 1. O companie specializată în comercializarea calculatoarelor, în urma unei cercetări statistice, că 67% din toţi clienţii lor cumpără calculatoare, iar 45% preferă tablete. a) Cu ce este egală probabilitatea că un client ales la întâmplare va fi unul care va cumpără şi un calculator şi o tabletă? b) Dar probabilitatea că acest client va cumpăra apă o tabletă, dacă ştim că acesta este unul care a cumpărat un calculator?

Soluţie. Deoarece, procentual vorbind, din numarul total de 100% cumpărători, 67% din ei cumpără calculatoare, iar 45% - apa tablete, rezultă că tot ce excede 100% din suma 67%+45%, adică 12% din ei, vor fi cei care preferă să cumpere ambele tipuri de apă minerală (vezi Principiul Adunării in caz general). Introducem evenimentele $A = \{$*un cumpărător ales la întâmplare va fi unul care va cumpăra un calculator*$\}$, $B = \{$*un cumpărător ales la întâmplare va fi unul care va cumpăra o tabletă*$\}$. Atunci, folosind formula probabilităţii clasice, găsim :

a) $P\{$*un client ales la întâmplare va fi unul care va cumpără şi calculator şi tabletă*$\} = P(A \cap B) = 12/100 = 0.12$.

b) $P(B) = 0.45$; $P(A/B) = \frac{P(A\cap B)}{P(B)} = \frac{0.12}{0.45} = 0.266\,67$.

Din definiţie, mai exact din formula de calcul a probabilităţii condiţionate, rezultă *formula înmulţirii probabilităţilor* pentru două evenimente aleatoare:

$$P(AB) = P(A)P(B/A), \text{ dacă } P(A) > 0$$

$$\text{sau } P(AB) = P(B)P(A/B), \text{ dacă } P(B) > 0.$$

Exemplul 2. Se ştie că orice întreprindere producătoare, spre exemplu, de calculatoare, supune produsele sale controlului calitătii inainte de a fi comercializate. Presupunem că, prin metode statistico-matematice specifice controlului calităţii se ştie că ponderea calculatoarelor cu defecte ascunse este de 2%, iar atinci când un calculator defect este supus controlului acesta poate fi admis (din greşeală) spre comercializare cu probabilitatea 0.05. Cu ce este egală probabilitatea ca un calculator ales la întâmplare va fi unul cu defecte şi admis, totodată, spre comercializare.

Soluţie. Întroducem evenimentele $D = \{$*un calculator ales la întâmplare va fi unul cu defecte*$\}$ şi $C = \{$*un calculator ales la întâmplare va fi admis spre comercializare*$\}$. Conform condiţiilor din problemă $P(D) = 0.02$, $P(C/D) = 0.05$. Deoarece $P(D) > 0$, din *formula înmulţirii probabilităţilor* avem că probabilitatea că un calculator ales la întâmplare va fi unul cu defecte şi admis, totodată, spre comercializare coincide cu

$$P(CD) = P(D)P(C/D) = 0.02 \times 0.05 = 0.001.$$

Formula înmulţirii probabilităţilor pentru două evenimente aleatoare este doar un caz particular a unei formule mai generale. Are loc urmatoarea

Teoremă (*Formula înmultirii probabilitătilor în caz general*). *Dacă* $(\Omega, \mathcal{F}, P)$ *este un câmp de probabilitate şi* A_1 , A_2 , ... , A_n sunt evenimente *aleatoare legate de acest câmp de evenimente cu proprietatea* $P(A_1A_2 \ldots A_{n-1}) > 0$, *atunci*

$$P(A_1A_2...A_n) = P(A_1)P(A_2\ /\ A_1)P(A_3\ /\ A_1A_2)...P(A_n\ /\ A_1A_2...A_{n-1}).$$

Exemplul 3. Cu ce este egală probabilitatea că, jucând Poker şi efectuând patru extrageri succesive (fără repetare) dintr-un butuc de cărţi bine amestecate, vom extrage patru aşi?

Soluţie. Considerăm evenimentele $A_k = \{$*la extragerea cu numărul de ordine k va apare un as*$\}$, $k = \overline{1,4}$. Aplicând Teorema anterioară, gasim că probabilitatea în cauză este egală cu

$$P\left(\bigcap_{k=1}^{4} A_k\right) = P(A_1A_2...A_4) =$$

$$= P(A_1)P(A_2\ /\ A_1)P(A_3\ /\ A_1A_2)P(A_4\ /\ A_1A_2...A_3) =$$

$$= \left(\frac{4}{52}\right)\left(\frac{3}{51}\right)\left(\frac{2}{50}\right)\left(\frac{1}{49}\right) = \frac{1}{270\,725} = 3.693\,8 \times 10^{-6}.$$

Probleme propuse.

1. Studiul statistic al vânzărilor zilnice ale unui magazin specializat în vânzarea produselor electronice audio arata că fiecare din următoarele evenimente $A = \{$ *un client va cumpara cel puţin un CD-player*$\}$, $B = \{$ *un client va cumpara cel puţin un set de căşti audio*$\}$ şi $C = \{$ *un client va cumpara cel puţin un CD- player şi un set de căşti audio*$\}$ sunt egale, respectiv, cu $P(A) = 0.6$, $P(B) = 0.75$ şi $P(C) = 0.5$. Aflaţi valorile următoarelor probabilităţi $P(A \cup B)$, $P(A/B)$, $P(B/A)$, $P(\overline{A} \cap \overline{B})$, $P(\overline{A} \cup \overline{B})$, $P(\overline{A}/\overline{B})$, $P(\overline{B}/\overline{A})$.

2. În legătură cu aruncarea unui zar "perfect" de trei ori succesiv considerăm evenimentele $A = \{$ *de fiecare dată va apare o faţă diferită*$\}$, $B = \{$ *cel putin o dată va apare faţa* $6\}$. Calculaţi probabilităţile $P(A/B)$, $P(B/A)$.

3. Consideram aruncarea unei monede "perfecte" sau pana cand apare "stema" sau pana cand "banul " apare de 3 ori succesiv. Cu ce este egala probabilitaea ca "banul" va apare de 3 ori succesiv daca se stie ca la prima aruncare a aparut "banul".

4. Considerăm câmpul de probabilitate $(\Omega, \mathcal{F}, P)$ şi evenimentul aleator $B \in \mathcal{F}$ cu $P(B) > 0$. Arataţi ca tripletul $(B,\ B \cap \mathcal{F},\ P_B)$ reprezinta un nou câmp de probabilitate (un nou model probabilist), unde $P_B(A) = P(A/B) = \frac{P(A \cap B)}{P(B)}$ pentru orice eveniment $A \in B \cap \mathcal{F}$.

Tema: *Formula înmulţirii probabilităţilor.*

5. Presupunem ca un PC consta din N blocuri, calculatorul ieşind din funcţiune deîndată ce iese din funcţiune unul din blocuri, ieşirea simultană din funcţiune a două sau mai multe blocuri fiind exclusă. Depanatorul de calculatoare verifică, luând la intâmplare, unul după altul câte un bloc, pană când va depista blocul defectat. Cu ce este egală probabilitatea că depanatorul va depista blocul defectat la incercarea cu numărul de ordine k, $k = 1$, 2, ..., N.

6. Un lot de 100 de calculatoare este supus controlului calităţii, selectând la întâmplare 5 calculatoare. Dacă se depistează că , cel puţin, unul din aceste calculatoare este defect, atunci întreg lotul este respins. Cu ce este egala probabilitaea ca lotul de calculatoare supus controlului va fi respins, daca se stie ca 5% de calculatoare din lot sunt cu defecte?

1.7. Independenţa evenimentelor aleatoare, formula lui Poisson

În situaţia când două evenimente aleatoare A şi B, legate de acelaşi câmp de probabilitate $(\Omega, \mathcal{F}, P)$, unde $P(B) > 0$, au proprietatea că $P(A/B) = P(A)$ este firesc să spunem că evenimentul A nu depinde de B sau pe scurt A şi B sunt independente. Vom arăta că definiţia de mai jos este acoperitoare şi pentru toate situaţiile de acest tip.

Definţia independenţei (*cazul a două evenimente*). Vom spune că *două evenimente aleatoare* A şi B, legate de acelaşi câmp de probabilitate $(\Omega, \mathcal{F}, P)$, *sunt independente* dacă $P(AB) = P(A)P(B)$. În caz contrar vom spune că *evenimentele* A şi B *sunt dependente.*

Într-adevăr, de aici rezultă imediat că dacă A şi B sunt independente, unde $P(B) > 0$, atunci $P(A \ / \ B) = \frac{P(A \cap B)}{P(B)} = \frac{P(A)P(B)}{P(B)} = P(A)$, ceea ce confirmă aşteptările noastre.

Remarca 1. Definiţia independenţei a două evenimente aleatoare, cuprinde chiar si cazurile când, cel puţin, unul din evenimentele A şi B are probabilitatea egală cu zero. De exemplu, fie $P(B) = 0$. Atunci, din Axioma 1 a probabilităţii si din proprietatea $P(AB) \leq P(B) = 0$, rezultă că $P(AB) = 0$. Dar aceasta inseamna că $P(AB) = P(A)P(B)$, adică A şi B sunt independente.

Mai mult, are loc următoarea

Propoziţia 1. Dacă evenimente aleatoare A şi B, legate de acelaşi câmp de probabilitate $(\Omega, \mathcal{F}, P)$ sunt independente, atunci independente vor fi si fiecare din perechile de evenimente: $(\overline{A}, B)$, $(A, \overline{B})$, $(\overline{A}, \overline{B})$.

Exemplul 1. Angajaţii Corporaţiei Excelsior sunt distribuiţi în funcţie de gen şi tipul angajării lor în felul următor:

	Tipul	**de**	**angajare**	
Sexul	**Vânzâri**	**Conducere**	**Producţie**	**Total**
Masculin	825	675	750	2250
Feminin	1675	825	250	2750
Total	2500	1500	1000	5000

Pentru a stimula loialitatea faţă de companie, compania alege la întâmplare un angajat, asigurându-i lunar o scurtă vacanţă plus cheltuielile aferente.

a) Este oare evenimentul $F = \{va\ fi\ aleasă\ o\ femeie\}$ independent de evenimentul $C = \{va\ fi\ ales\ o\ persoana\ din\ conducerea\ corporaţiei\}$?

b) Dar acelaşi eveniment F, este oare independent de evenimentul $D = \{va\ fi\ ales\ o\ persoana\ din\ sectorul\ de\ Producţie\}$?

Soluţie. Din tabel deducem că

$$P(F) = \frac{2750}{5000} = 0.55, P(C) = \frac{1500}{5000} = 0.3, P(D) = \frac{1000}{5000} = 0.2,$$

$$P(F \cap C) = \frac{825}{5000} = 0.165, P(F \cap D) = \frac{250}{5000} = 0.05.$$

a) Cum $P(F \cap C)$=0.165=$P(F)P(C)$=0.55 $\times$ 0.3 rezultă că evenimentele F şi C sunt *independente*;

b) Dar $P(F \cap D)$=0.05 $\neq$ $P(F)P(C)$=0.55 $\times$ 0.02=0.11. În concluzie, evenimentele F şi D sunt *dependente.*

Exemplul 2. La ora de Statistică Matematica s-au prezentat 20 de studenţi din care 8 sunt fumători, 12 poartă ochelari iar 6 din cei care poartă ochelari sunt şi fumători. Este scos la tablă la intâmplare unul din studenţi. Aflaţi dacă evenimentele $A = \{studentul\ scos\ la\ tabla\ este\ unul\ fumător\}$ şi $B = \{studentul\ scos\ la\ tabla\ este\ unul\ ochelarist\}$ sunt independente sau dependente.

Soluţie. Din condiţiile problemei, folosind definiţia probabilităţii clasice, aflăm că $P(AB) = \frac{3}{10} \neq P(A)P(B) = \frac{8}{20} \cdot \frac{12}{20}$, de unde deducem că evenimentele A şi B sunt dependente. Această concluzie, trebuie, însa, tratată cu prudenţa, aceasta fiind valabilă doar pentru acest exemplu concret, dar în care eşantionul nu este reprezentativ, deoarece numarul de 20 de studenţi nu este suficient de mare pentru a lansa o cercetare statistica mai amplă.

Aceasta în pofida faptului că pare a fi plauzibilă ipoteza, conform căreia fumatul ar influienţa negativ acuitatea vederii. O atare cercetare, pentru un numar n de studenţi (persoane) suficient de mare, dacă ar da diferente considerabile între frecvenţa relativă $f_n(B)$ şi frecvenţa relativă condiţionată $f_n(B/A)$=$f_n(A \cap B)/f_n(A)$, atunci ipoteza în cauză ar fi confirmată statistic.

Noţiunea de independenţă a evenimentelor aleatoare poate fi extinsă (generalizată) asupra mai mult de două evenimente.

Definiţia independenţei evenimentelor (*în totalitate*).Vom spune ca evenimentele aleatoare A_1 , A_2 , ... , A_n, legate de acelaşi câmp de probabilitate (Ω, $\mathcal{F}$, P), sunt *independente (in totalitate)* daca $P(A_{i_1}A_{i_2}...A_{i_k}) = P(A_{i_1})P(A_{i_2})P(A_{i_3})...P(A_{i_k}\ )$ pentru orice set de indici diferiţi $\{i_1,\ i_2\ ,...,i_k\}$ din multimea de indici $\{1,2,\ ...,n\}$, $k = 2,\ 3,\ ...,\ n$.

Exerciţiu. Folosind definiţia de mai sus, prin negarea ei, formulaţi, desfăşurat, **Definiţia dependenţei evenimentelor** (*în totalitate*).

Remarca 2. Atunci când evenimentele A_1 , A_2 , ... , A_n sunt independente (în totalitate) Formula Înmulţirii Probabilităţilor se simplifică si are forma

$$P(A_1A_2...A_n) = P(A_1)P(A_2)P(A_3)...P(A_n\ ).$$

Remarca 3. Independenţa evenimentelor A_1 , A_2 , ... , A_n *două câte două,* mai exact, faptul că $P(A_{i_1}A_{i_2}) = P(A_{i_1})P(A_{i_2})$ pentru orice set de indici diferiţi $\{i_1,\ i_2\}$ din multimea de indici $\{1,2,\ ...,n\}$ nu garantează independenţa lor *în totalitate*. Contraexemplul de mai jos confirmă această afirmaţie.

Contraexemplu (*Trei evenimente independente două câte două, dar nu şi independente în totalitate*). Considerăm aruncarea o singură dată a unui tetraedru regulat, cu centru de greutate simetric, feţele căruia sunt colorate astfel: faţa 1-albastru, faţa 2-galben, faţa 3-roşu, faţa 4-albastru, galben, roşu.

Arătăm ca evenimentele A=$\{$*va apare culoarea albastră*$\}$, G=$\{$*va apare culoarea galbenă*$\}$, R=$\{$*va apare culoarea roşie*$\}$ sunt independente 2 cate 2, dar nu si in totalitate. Într-adevăr,

$$P(A) = P(G) = P(R) = \frac{1}{2}, P(AG) = \frac{1}{4} = P(A)P(G) = \frac{1}{2} \cdot \frac{1}{2},$$

$$P(AR) = \frac{1}{4} = P(A)P(R) = \frac{1}{2} \cdot \frac{1}{2}, P(GR) = \frac{1}{4} = P(G)P(R) = \frac{1}{2} \cdot \frac{1}{2},$$

dar

$$P(AGR)=\frac{1}{4}\neq P(A)P(G)P(R)=\frac{1}{2}\cdot\frac{1}{2}\cdot\frac{1}{2}. \qquad \square$$

Propoziţia 2. (*Formula lui Poisson*). Daca evenimentele A_k sunt independente (in totalitate) si probabilităţile $P(A_k)= p_k$, k=1,2,. . . ,n sunt cunoscute, atunci probabilitatea

$$P\{se\ va\ produce\ cel\ puţin\ unul\ din\ evenimentele A_k\}=P\left(\underset{k=1}{\overset{n}{\cup}}A_k\right)=$$

$$1-[(1-P(A_1))(1-P(A_2))...(1-P(A_n)]=1-[(1-p_1)(1-p_2)\dots(1-p_n)].$$

Exemplul 3. Un aparat constă din trei elemente care în timpul funcţionării lui se pot deteriora, independent unul de altul. Notăm prin $A_i=\{elementul\ i\ se\ va\ deteriora\}$, $i=1,2,3$. Să se calculeze probabilitatea evenimentului $A=\{se\ va\ deteriora\ un\ singur\ element\}$, $B=\{se\ va\ deteriora\ ,\ cel\ puţin,\ un\ element\}$, dacă se ştiu probabilităţile: $p_1=P(A_1)=0.13$, $p_2=P(A_2)=0.06$, $p_3=P(A_3)=0.12$.

Soluţie. Vom exprima evenimentul aleator A prin intermediul evenimentelor A_1, A_2 şi A_3. Evenimentul A se va produce, atunci si numai atunci cand, se va deteriora primul element iar al doilea – nu şi al treilea – nu, **sau** se va deteriora al doilea element, iar primul - nu şi al treilea – nu, **sau** se va deteriora al treilea element, iar primul – nu şi al doilea – nu. Prin urmare, conform definiţiilor operaţiior asupra evenimentelor aleatoare, avem:

$P(A)$=$P\left(A_1\cdot\overline{A_2}\cdot\overline{A_3}\cup\overline{A_1}\cdot A_2\cdot\overline{A_3}\cup\overline{A_1}\cdot\overline{A_2}\cdot A_3\right)$.

Calculăm probabilitatea evenimentului A folosind succesiv: formula de adunare a probabilităţilor pentru evenimente incompatibile (disjuncte) doua câte două, formula înmulţirii probabilităţilor evenimentelor independente (in totalitate) şi formula de calcul al probabilităţii evenimentului opus.

$P(A)=0.13\times(1-0.06)\times(1-0.12)$+$(1-0.13)\times0.06\times(1-0.12)$+$(1-0.13)\times(1-0.06)\times0.12$=0.251 61.

La calcularea valorii $P(B)=P\left(\underset{k=1}{\overset{3}{\cup}}A_k\right)$ se aplica Formula lui Poisson. Deci, $P(B)$=$1-(1-0.13)\times(1-0.06)\times(1-0.12)$=0.280 34. $\square$

Probleme propuse.

Tema: *Independenţa evenimentelor aleatoare.*

1. Cum aţi explica pe înţelesul unei persoane inteligente, dar care nu cunoaşte Teoria probabilităţilor, care este deosebirea dintre independenta a două evenimente aleatoare şi proprietatea lor de a fi incompatibile/disjuncte?

Pot fi oare independente si incompatibile concomitent două evenimente aleatoare A şi B, legate de acelaşi câmp de probabilitate $(\Omega, \mathcal{F}, P)$?

2. Sunt oare independente două evenimente aleatoare A şi B, legate de acelaşi câmp de probabilitate $(\Omega, \mathcal{F}, P)$ dacă $P(A) > 0$, $P(B) > 0$,$AB = \varnothing$?

3. Arătaţi că dacă un eveniment aleator A nu depinde de el însuşi, atunci probabilitatea acestuia este egala cu 0 sau cu 1.

4. Juriul unui concurs consta din 3 persoane care iau decizie corecta independent unul de altul. Prima si a doua persoana iau decizie corecta cu una si aceeasi probabilitate $p, 0 < p < 1$, iar cea de a treia, pentru a lua decizie, arunca o moneda "perfecta". Decizia finala se ia cu majoritate de voturi. Cu ce este egala probabilitatea ca Juriul va lua o decizie corecta.

5. Presupunem ca un PC marca DELL produs in China este de caliate superioara cu probabilitatea 0.7, iar acelasi calculator produs in Honkong este de calitate superioara cu probabilitatea 0.8. Sunt luate la intamplare 3 PC-uri produse in China si 4 PC-uri produse in Honkong. Cu ce este egala probabilitatea ca toate calculatoare vor fi de calitate superioara.

6. Contraexemplu (*Trei evenimente pentru care probabilitatea produsului lor coincide cu produsul probabilităţilor lor, dar care nu sunt independente in totalitate*). Considerăm aruncarea unui zar "perfect" de două ori succesiv. Arătaţi că evenimentele $A = \{$*numărul punctelor apărute la prima aruncare nu va întrece* $3\}$, $B = \{$*numărul punctelor apărute la prima aruncare nu va fi mai mic de* 3 *şi, totodată, nu mai mare de* $5\}$ şi $C = \{$*suma punctelor apărute va fi egală cu* $9\}$. Arătaţi ca $P(ABC) = P(A)P(B)P(C)$, dar A, B şi C nu sunt independente două câte două, cu alte cuvinte, *nu sunt independente în totalitate.*

Tema: *Formula lui Poisson.*

7. Presupunemm că 3% din producţia de procesoare pentru telefoanele mobile **iPhone 6s**, produse de firma asociata, au defecte. Controlului sunt supuse 20 de procesoare luate la întâmplare. Cu ce este egala probabilitaea că printre ele se va depista, cel puţin, un procesor cu defecte?

8. Care este numarul minim de numere aleatoare din multimea de numere $\{1, 2, ..., 9\}$, care te trebuie generate pe calculator, pentru a fi siguri cu probabilitatea nu mai mica decât 0.9, ca printre ele se va întâlni, cel putin un număr par?

9. Presupunem cunoscut faptul că intr-un experiment aleator $\mathcal{E}$ probabilitatea apariţiei, cel puţin o dată, a evenimentului A in patru probe independente $\mathcal{E}$ este egală cu $1/2$. Cu ce este egală probabilitatea evenimentului

A dacă aceasta este aceeaşi in fiecare probă $\mathcal{E}$.

1.8. Formulele probabilităţii totale si a lui Bayes

Noţiunile introduse permit calcularea probabilităţilor unor evenimente mai complicate, cunoscând unele probabilităţi, inclusiv condiţionate, mai uşor de identificat. Este vorba de următoarea

Teoremă *(Formulele probabilităţii totale şi a lui Bayes).* *Dacă* A *şi* H_1, H_2, ..., H_n,... *sunt evenimente aleatoare legate de acelaşi câmp de probabilitate* $(\Omega, \mathcal{F}, P)$ *şi satisfac condiţiile* :

a) *evenimentul* A *implica producerea a cel putin unuia din evenimentele* H_1, H_2, ..., H_n,..., adică $A \subseteq H_1 \cup H_2 \cup ...$;

b) *evenimentele* H_1, H_2, ..., H_n, ... *sunt incompatibile două câte două, adică* $H_i \cap H_j = \varnothing$, $\forall i \neq j$, $i, j \geq 1$;

c) $P(H_i) > 0$, *atunci au loc*

formula probabilităţii totale (FPT)

$$P(A) = \sum_{k \geq 1} P(A/H_k)P(H_k)$$

şi ***formula lui Bayes***

$$P(H_j/A) = \frac{P(H_j)P(A/H_j)}{\sum\limits_{k \geq 1} P(A/H_k)P(H_k)}, \textit{ pentru orice } j \geq 1.$$

Remarcă. Probabilitatea $P(A)$ se numeşte probabilitate *apriori* deoarece aceasta este calculată înainte de a se efectua experimentul aleator corespunzător; graţie condiţiei $a)$ evenimentele H_1, H_2, ..., H_n, ... se mai numesc *ipoteze,* iar probabilitatile $P(H_j/A)$, $j \geq 1$ se numesc probabilităţi *aposteriori* deoarece acestea sunt probabilităţi ale ipotezelor, probabilităţi recalculate in condiţia că, drept rezultat, în urma experimentului s-a produs evenimentul A.

Exemplul 1 (*Problema „Monty-Hall”)***.** Această problemă a fost popularizată în Statele Unite de o emisiune de divertisment a canalului CBS din 1963.

Concurentul este pus în faţa a trei uşi. În spatele a două uşi este câte o capră iar în spatele unei uşi este un autoturism. Scopul jocului este desigur câştigarea maşinii. Jucătorul va câştiga ce se află în spatele uşii alese.

Participantul trebuie să aleagă o uşă. Apoi, prezentatorul îi deschide o altă uşă, din celelalte două rămase, în spatele căreia este o capră. Apoi îl întreabă dacă vrea sau nu să îşi schimbe alegerea iniţială în favoarea uşii a treia. Această întrebare generează, de fapt, problema calcularii a doua probabilitati corespunzătoare respectării a două strategii de joc.

Strategia 1: Participantul *nu-şi schimbă* alegerea iniţială.

Strategia 2: Participantul *îşi schimbă* alegerea iniţială.

Care dintre strategii conduce la o probabilitate mai mare de a câştiga maşina?

Soluţie. Deoarece participantul alege, la început, la întâmplare una din cele trei uşi, rezulta că nu conteaza în spatele cărei uşi se află autoturismul. Aşa că presupunerea că maşina se află în spatele uşii numărul 1 nu afectează rezolvarea problemei. Considerăm evenimentul $A = \{$*participantul va câştiga autoturismul*$\}$. Observăm că acest eveniment implică valabilitatea a cel puţin uneia din ipoteze H_i=$\{$*participantul va alege iniţial uşa cu numărul* $i\}$, $i = 1$, 2, 3.

În aceste notaţii, evenimentele A, H_i, verifica condiţiile aplicării FPT. Deci, $P(A)$=$P(A/H_1)P(H_1)$+$P(A/H_2)P(H_2)$+$P(A/H_3)P(H_3)$, unde $P(H_i) = 1/3$, $i = 1$, 2, 3, indiferent de strategia folosită. În cazul folosirii Strategiei 1 avem: $P(A/H_1) = 1$, iar $P(A/H_2)$=$P(A/H_3) = 0$, prin urmare $P(A)$=$1 \cdot \frac{1}{3}$+$0 \cdot \frac{1}{3}$+$0 \cdot \frac{1}{3}$. În cazul folosirii Strategiei 2 avem: $P(A/H_1) = 0$, iar $P(A/H_2)$=$P(A/H_3) = 1$, prin urmare $P(A)$=$0 \cdot \frac{1}{3}$+$1 \cdot \frac{1}{3}$+$1 \cdot \frac{1}{3} = \frac{2}{3}$. În concluzie, aplicarea Strategiei 2 este mai eficientă deoarece măreşte şansele câştigării autoturismului de două ori faţă de cazul aplicării Strategiei 1.

Exemplul 2 *(Problema lui Lewis Carrol)*. Într-o cutie se află o bilă, despre culoarea căreia se ştie că este albă sau neagră cu una şi aceeaşi probabilitate. Întroducem în această cutie o bilă albă, după care extragem la intâmplare o bilă.

a) Cu ce este egală probabilitatea că bila extrasă va fi de culoare albă.

b) Cu ce este egala probabilitatea ca bila initiala este de culoare alba, dacă bila extrasă este albă.

Soluţie. Întroducem evenimentul A=$\{$*bila extrasă va fi albă*$\}$ şi ipotezele H_1=$\{$*bila iniţial aflată în cutie va fi albă*$\}$, H_2=$\{$*bila iniţial aflată în cutie va fi neagră*$\}$. Din condiţiile problemei deducem ca: *a)* $A \subseteq H_1 \cup H_2$; *b)* $H_1 \cap H_2$=$\varnothing$; *c)* $P(H_1)$=$P(H_2)$=$\frac{1}{2}$, ceea ce înseamnă că sunt valabile condiţiile în care putem aplica formulele probabilităţii totale şi a lui Bayes. Prin urmare, deoarece $P(A/H_1) = 1$, $P(A/H_2) = \frac{1}{2}$,

a) $P(A){=}P(A/H_1)P(H_1) + P(A/H_2)P(H_2){=}1 \cdot \frac{1}{2} + \frac{1}{2} \cdot \frac{1}{2}{=}\frac{3}{4}$;
b) $P(H_1/A){=}\frac{P(H_1)P(A//H_1)}{P(A)}{=}\frac{\frac{1}{2}\cdot 1}{\frac{3}{4}}{=}\frac{2}{3}$.

În concluzie, întroducerea artificiala a unei bile suplimentare de culoare albă mareşte probablitatea evenimentului A de la 1/2 la 2/3.

Probleme propuse

1. Să se rezolve *Problema „Monty-Hall"* din exemplul 1 anterior în presupunerea că alegerea Strategiei 1 sau 2 este una *randomizată* cu probabilitatea p, $0 \leq p \leq 1$. Mai exact, dupa ce participantul a făcut alegerea iniţială, acesta aruncă o monedă deformată, astfel încât Stema apare cu probabilitatea p iar Banul cu probabilitatea $1 - p$. Dacă apare Stema, atunci participantul optează pentru Strategia 1, în caz contrar optează pentru Strategia 2. Să se arate că în cazul alegerii randomizate a strategiei probabilitatea $P(A) = (4 - p)/6$, prin urmare şansele maxime de câştigare a autoturismului corespund valorii $p = 0$, ceeea ce corespunde folosirii, numai şi numai, a Strategiei 2.

2. Un lot de PC-uri, din care 10% sunt cu defecte, este supus controlului calitatii. Schema controlului este de aşa natura, incat defectul (daca acesta exista) este depistat cu probabilitatea 0.95, iar probabilitatea ca un calculator fara defecte va fi declarat defect este egala cu 0.03. Cu ce este egala probabilitatea ca un calculator ales la intamplare din lot va fi declarat defect? Cu ce este egala probabilitatea ca PC-ul ales la intamplare intr-adevar este defect daca se stie ca acest PC a fost, in urma controlului, declarat a fi defect?

3. Consideram ca la un magazin de calculatoare au fost aduse un lot de PC-uri marca HP, din care 30% sunt produse in China, 20%-in Singapore si 50%-in Honkong. Cu ce este egala probabilitatea ca un PC cumparat la intamplare are defecte ascunse daca astfel de defecte au 20% de calculatoare produse in China, 10%-cele produse in Singapore si 5%-cele produse in Honkong? Cu ce este egala probabilitatea ca PC-ul cumparat la intamplare este produs in China , daca se stie ca acesta s-a dovedit a avea defecte ascunse?

4. Intr-o cutie sunt 20 de mingi de tenis, din care 15 sunt noi noute, iar 5 sunt folosite la joc. Pentru primul joc sunt alese la intamplare doua mingi, dupa care sunt puse la loc in cutie. Pentru jocul urmator sunt alese, la fel, doua mingi. Cu ce este egala probabilitatea ca ambele mingi alese pentru cel de al doilea joc vor fi noi noute? Cu ce este egala probabilitatea ca pentru primul joc au fost extrase 2 mingi noi noute daca se stie ca mingiile extrase pentru cel de al doilea joc s-au dovedit a fi noi noute?

5. Avem doua cutii, astfel incat in prima cutie se afla 6 bile albe si 4 bile negre, iar intr-a doua cutie se afla 3 bile albe si 2 bile negre. Din prima cutie este extrasa la intamplare o bila si pusa intr-a doua cutie. dupa care dintr-a doua cutie este extrasa la intamplare o bila. Cu ce este egala probabilitatea ca aceasta va fi de culoare alba? Cu ce este egala probabilitatea ca din prima cutia a fost extrasa o bila alba daca se stie ca din cutia a doua a fost extrasa o bila alba?

6. Presupunem exact una din 10 000 000 de monede perfecte are imprimata Stema pe ambele parti ale ei. Cu ce este egala probabilitatea ca a fost aleasa la întâmplare moneda cu ambele fete marcate cu Stemă daca se stie ca in urma aruncarii ei de 10 ori succesiv a aparut Stema?

7. Se stie ca mesajele scurte (SMS-urile) transmise prin intermediul telefoniei mobile sunt codificate cu ajutorul cifrelor/semnalelor 0 sau 1. Presupunem ca transmiterea semnalelor este supusa bruiajelor, astfel incat sunt deformate $2/5$ semnale 0 si $1/3$ semnale 1. Presupunem ca ponderea semnalului 0 in mesajul transmis este egala cu $5/8$ iar ponderea semnalului 1 este egala cu $3/8$. Cu ce este egala probabilitatea receptionarii corecte a primului semnal din mesaj daca se stie ca a fost receptionat: a) semnalul 0; b) semnalul 1.

8. Presupunem ca avem un lot de 5 PC-uri despre care se stie, doar, ca este echiprobabila orice ipoteza H_k despre numarul k de PC-uri defecte in acest lot, $k = 0, 1, 2, ..., 5$. Care ipoteza are probabilitatea cea mai mare daca se stie ca, alegand la intamplare un PC, acesta s-a dovedit a fi cu defecte?

9. Sa se determine probabilitatea ca intr-un lot de 1000 de calculatoare nu exista niciunul cu defecte, daca se stie ca 100 de calculatoare din acest lot, supuse controlului, s-au dovedit a fi fara defecte, presupunand ca sunt valabile, cu una si aceeasi probabilitate, oricare din ipotezele $H_k = \{$*numarul de calculatoare defecte, printre cele* 1000 *de calculatoare din lot, este egal cu* k $\}$, $k = 0, 1, 2, 3, 4, 5$.

10. Intr-o cutie sunt 7 bile albe si 3 bile negre. Sunt extrase la intamplare, fara intoarcere, doua bile, din care, cel putin, una s-a dovedit a fi de culoare neagra. Cu ce este egală probabilitatea ca cealaltă bilă extrasă este albă.

11. O reprezentanţă pentru vânzări de automobile cunoaşte, din experienţa anterioară că 10% din cei care vizitează showroom-ul şi discută cu un vânzător vor cumpăra în cele din urmă o maşină. Pentru a mări şansele de succes, reprezentanţa oferă o cină gratuită cu un agent de vânzări pentru toţi oamenii care sunt de acord să asculte o prezentare completă a vânzărilor. Aceasta ştiind că unii vor face totul pentru o cină gratuită, chiar dacă nu intenţionează să cumpere o maşină, iar unii ar prefera să nu petreacă o cină cu un vânzător de maşini. Pentru a testa eficacitatea acestui stimulent de promovare a vânzărilor proiectul este derulat pentru 6 luni, constatând că 40% dintre persoanele care au cumpărat maşini au avut o cină gratuită. În plus, 10% din persoanele care nu au cumpărat maşini au avut o cină gratuită. Conducerea este interesată în a afla dacă:

a) Oamenii care acceptă cina au o probabilitate mai mare de a cumpăra o maşină noua?

b) Care este probabilitatea ca o persoană care nu acceptă o cină gratuită să achiziţioneze o maşină?

12. Pe baza unei examinări a înregistrărilor anterioare ale soldurilor conturilor unei societăţi, un auditor constată că 15% au conţinut erori. Din aceste solduri în eroare, 60% au fost considerate ca fiind valori neobişnuite bazate pe istoricul activităţii societăţii. Din totalul tuturor soldurilor contului, 20% prezentau valori neobişnuite. Dacă cifra pentru un anumit sold pare neobişnuită pe această bază, cu ce este egală probabilitatea că acesta conţine erori?

13. Datele statistice arată ca 5% din persoanele de sex masculin suferă de daltonism pe cand doar 0.025% din persoanele de sex feminin suferă de această tulburare de vedere cromatică. Dintr-o societate formată din acelaşi număr de bărbaţi şi de femei a fost aleasă la întâmplare o persoană. Cu ce ste egală probabilitatea că aceasta persoană este una de sex masculin daca aceasta s-a dovedit a fi una afectată de daltonism.

2. Variabile aleatoare

2.1. Introducere

În legătură cu modelarea matematica a experimentelor aleatoare (ce posedă proprietatea regularităţii statistice) trebuie sa ţinem cont că, în realitate interes prezintă, deseori, nu rezultatele posibile (evenimentele elementare) ca atare ci nişte mărimi numerice ce depind de acestea. De exemplu, intr-un sondaj ce vizează cercetarea veniturilor salariale ale angajaţilor, sa zicem, din sfera bugetară, interes prezintă venitul salarial al unui angajat inclus în eşantion, nu persoana în cauză. Cu alte cuvinte vom avea de a face cu o marime (variabilă) ce depinde de evenimentele elementare, doar salariul variază de la angajat la angajat. Noţiunea matematica corespunzătoare este cea de *variabilă aleatoare* care in Statistică mai este cunoscuta şi sub denumirea de *caracteristică sau variabilă statistică.* Dar în legatură cu fiecare eveniment elementar putem asocia una, două sau mai multe caracteristici statistice. Astfel, in exemplul nostru, fiecărui angajat din sfera bugetara îi putem asocia astfel de caracteristici ca marimea salariului, nivelul de studii, vârsta, etc. Drept consecinţă, se impune introducerea noţiunilor de variabile aleatoare unidimensionale, bidimensionale (multidimensionale).

2.2. Variabilă aleatoare (unidimensională), funcţia ei de distribuţie (repartiţie)

Pentru început vom introduce noţiunea de variabilă aleatoare unidimensională.

Definiţia 1. Fie $(\Omega, \mathcal{F}, P)$ un câmp de probabilitate, atunci vom numi *variabilă aleatoare (v.a.) definită pe acest câmp* orice aplicaţie (regulă, funcţie) $X : \Omega \longrightarrow \mathbb{R}$ care verifică condiţia

$$\{\omega \in \Omega : X(\omega) \leqslant x\} \in \mathcal{F} \textit{ pentru orice } x \in \mathbb{R}. \quad (1)$$

Remarca 1. Dacă suntem in cazul discret, adică, în cazul când spatiul de evenimente elementare Ω este o multime finită sau, cel mult, numărabilă, atunci câmpul (familia) de evenimente aleatoare $\mathcal{F}$ coincide cu familia tuturor submulţimilor din Ω. Prin urmare, in acest caz, putem numi variabila aleatoare orice aplicaţie $X : \Omega \longrightarrow \mathbb{R}$, deoarece in caz discret condiţia ca $\{\omega \in \Omega : X(\omega) \leqslant x\} \in \mathcal{F}$ este valabilă automat pentru orice $x \in \mathbb{R}$.

Evenimentul care figurează în condiţia (1) se notează, pe scurt, astfel: $\{\omega : X(\omega) \leqslant x\}$, sau $\{X \leqslant x\}$. Mărimea $X(\omega)$ se numeşte valoare a variabilei aleatoare X. Din condiţia (1) rezultă că pentru orice $x \in \mathbb{R}$ putem determina probabilitatea evenimentului aleator $\{X \leqslant x\}$.

În calitate de exemple de v.a. întâlnite in practică putem lua: suma de puncte apărute la aruncarea unui zar de două ori, durata funcţionării unui dispozitiv electronic, numărul de particule alfa emise de o substanţă radioactivă într-o unitate de timp, cantitatea anuală de precipitaţii atmosferice într-o anumită regiune, numărul de apeluri telefonice înregistrate pe parcursul a 24 de ore la o statie de ajutor medical, numărul de accidente auto înregistrate pe parcursul unui anumit interval de timp, numărul despăgubirilor sau valoarea totală a despăgubirilor platite de o Companie de Asigurări, etc., etc.

Este important sa facem deosebire dintre v.a. X şi multimea ei de valori posibile $\mathcal{X}$. De exemplu, daca X reprezinta multimea de puncte apărute la aruncarea unui zar o singura dată, atunci $X \in \mathcal{X}$ =$\{1, 2, ..., 6\} \subseteq \mathbb{R}$.

Definiţia matematică a noţiunii de v.a. este suficient de flexibilă. Aceasta se vede din

Proprietăţile variabilei aleatoare. *a) Dacă X este o variabilă aleatoare, atunci pentru orice a, $b \in \mathbb{R}$, $a < b$ sunt evenimente aleatoare şi prin urmare sunt definite probabilităţile lor pentru* $\{X > a\}$,$\{X \geq a\}$,$\{X = a\}$, $\{a < X \leq b\}$, $\{a < X < b\}$, etc.;

b) Fie $(\Omega, \mathcal{F}, P)$ un câmp de probabilitate, $a \in \mathbb{R}$, iar $X : \Omega \longrightarrow \mathbb{R}$ şi $Y : \Omega \longrightarrow \mathbb{R}$ sunt variabile aleatoare. Atunci sunt variabile aleatoare şi funcţiile: 1) aX; 2) X^k, k= 1, 2, ...; 3) $X - Y$; 4) $X + Y$; 5) $X \cdot Y$; 6) $1/Y$ dacă $Y(\omega) \neq 0$, $\forall \omega \in \Omega$; 7) X / Y dacă $Y(\omega) \neq 0$, $\forall \omega \in \Omega$; 8) $X \pm a$.

c) În genere, daca avem un şir finit de v.a. definite pe unul si acelasi câmp de probabilitate, atunci v.a. va fi si orice functie de aceste variabile, in caz că aceasta este functie continua. Astfel suma de v.a., diferenta lor, produsul lor, minimumul sau maximumul de aceste variabile, etc., vor fi v.a.

Drept alternativă la noţiunile de v.a. X şi de câmp de probabilitate $(\Omega, \mathcal{F}, P)$ pe care este definită aceasta, privite, în ansamblu, ca model matematic ce descrie comportamentul probabilist al v.a. X, se foloseşte pe larg, inclusiv în Statistica Matematică, noţiunea de funcţie de distribuţie (repartiţie) a v.a. X.

Definiţia 2. *Fie* $(\Omega, \mathcal{F}, P)$ un câmp de probabilitate şi $X : \Omega \longrightarrow \mathbb{R}$ o variabilă aleatoare definită pe el. Atunci funcţia $F : \mathbb{R} \longrightarrow \mathbb{R}$ definită prin relaţia

$$F(x) = P(X \leq x), \text{ pentru orice } x \in \mathbb{R}, \quad (2)$$

se numeşte *funcţie de distribuţie (f.d.) sau de repartiţie a v.a. X*, numită şi *funcţie cumulativă de distribuţie (f.c.d.).*

Remarca 2. Funcţia cumulativă de distribuţie este, de fapt, imaginea teoretică a funcţiei empirice din Statistica Descriptivă, mai exact a frecvenţelor relative cumulate crescător.

Exemplul 1. În legatură directă cu aruncarea unei monede "perfecte" o singură dată vom descrie f.d. a numarului X de "steme" apărute şi vom trasa graficul ei. Observăm ca valorile posibile ale lui X sunt valori din $\mathcal{X} \in \{0, 1\}$, unde $P(X \in \mathcal{X}) = 1$. Atunci f.d. a v.a. X

$$F(X) = P(X \leq x) = \begin{cases} P(\varnothing), & \text{pentru orice } x < 0, \\ P\{\textit{stema nu va apare}\}, & \text{pentru orice } 0 \leq x < 1, \\ P\{\Omega\}, & \text{pentru orice } 1 \leq x, \end{cases}$$

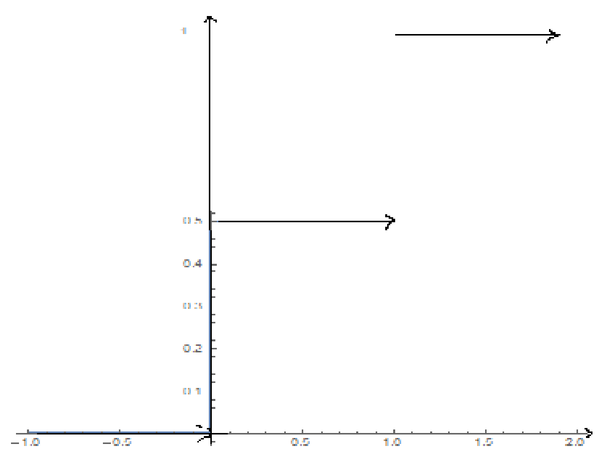
$$\begin{cases} 0, & \text{pentru orice } x < 0, \\ \frac{1}{2}, & \text{pentru orice } 0 \leq x < 1, \\ 1, & \text{pentru orice } 1 \leq x. \end{cases}$$

Folosind funcţia indicator

$$I_A(x) = \begin{cases} 1, \ \textit{pentru } x \in A, \\ 0, \ \textit{pentru } x \notin A, \end{cases}$$

observăm că funcţia de distribuţie $F(x)$ poate fi scrisă intr-o formă mai compacta: $F(x)=\frac{1}{2}I_{[0,1)}(x)+I_{[1,+\infty)}(x)$.

Graficul acestei functii este următorul



Teoremă. ***(Proprietăţile caracteristice ale f.d.).*** *Dacă $F(x)$ este o f.d. a unei v.a., atunci au loc următoarele proprietăţi:*

1^0. *F(x) este monoton nedescrescătoare, adică,* $F(x_1) \leq F(x_2)$ deîndată ce $x_1 \leq x_2$;

2^0. *F(x) este continuă la dreapta pentru orice $x \in \mathbb{R}$, adică, pentru orice şir monoton descrescător de valori x_n care tinde la x, atunci când n tinde* *la* $+\infty$, *şirul corespunzător de valori* $F(x_n)$ *are drept limită valoarea* $F(x)$, *fapt ce se noteaza, pe scurt,* $F(x+0) = F(x)$;

3^0. $F(-\infty) = 0$ *şi* $F(+\infty) = 1$.

Remarca 3. Importanta Teoremei anterioare rezidă in faptul ca proprietăţile **1**^{0}-**3**^{0} sunt caracteristice *numai şi numai* funcţiilor de distribuţie in sensul că, are loc şi reciproca acestei teoreme, conform căreia: pentru orice funcţie $F : \mathbb{R} \longrightarrow \mathbb{R}$, ce posedă aceste proprietăţi, putem construi (neunivoc) un câmp de probabilitate $(\Omega, \mathcal{F}, P)$ şi o v.a. X definită pe el, astfel încât funcţia ei de distribuţie va coincide cu F. In concluzie, orice funcţie $F(x)$ poate fi privita ca un model matematic ce descrie o *Legitate* (distribuţie, repartiţie) ce guvernează comportamentul probabilist al unei v.a., deîndata ce a aceasta funcţie posedă proprietăţile **1**^{0}-**3**^{0}. Mai mult, în postură de model probabilist, cunoscând f.d. $F(x)$, putem calcula probabilitătile cele mai importante din punct de vedere practic, probabilităţi legate de v.a. corespunzălegate X. Aceasta se vede din următoarea

Propoziţie (***Formule de calcul ale probabilitătilor pe baza f.d.***)

Fie X *o v.a. cu f.d.* $F(x)$*. Atunci pentru orice* $a < b$, $a, b \in \mathbb{R}$, *au loc urmatoarele formule de calcul ale probabilităţilor:*

a) $P(a < X \leq b) = F(b) - F(a)$;

b) $P(X > a) = 1 - F(a)$;

c) $P(X = a) = F(a) - F(a-0)$, *prin* $F(a-0)$ *fiind notată limita la stânga a funcţiei* F *în punctul* a;

d) $P(a \leq X \leq b) = F(b) - F(a-0)$;

e) $P(a < X < b) = F(b-0) - F(a)$;

f) $P(a \leq X < b) = F(b-0) - F(a-0)$.

Remarca 4. Din formula *c)* rezultă că, atunci când f.r. este continuă, $P(X = a) = 0$, deorece în acest caz şi $F(a-0) = F(a)$.

Remarca 5. În continuare vom spune că *graficele care au forma celui din exemplul anterior* sunt *grafice de formă scarată.* Acestea ne permit sa restabilim cu uşirinţă repartiţia v.a., mai exact, mulţimea de valori posibile ale v.a. corespunzătoare, ca fiind punctele de salt ale graficului, iar probabilităţile corespunzătoare coincid cu înălţimile treptelor în cauză.

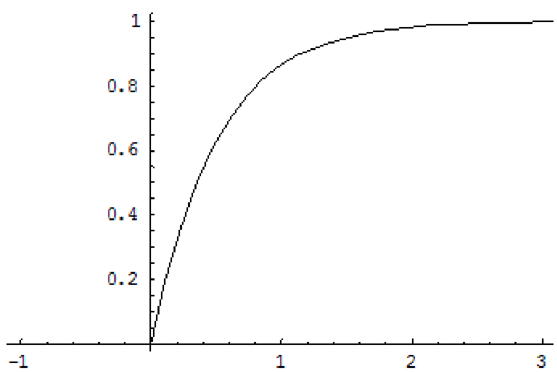
Exemplul 2. Olaf Motors, Inc. este un dealer auto într-un mic oraş sudic din SUA. Pe baza unei analize a acestuia istoriei vânzărilor sale , managerii ştiu că în fiecare zi numărul de maşini Toyota Prius vândute poate varia de la 0 la 5. Cum poate funcţia de distribuţie $F(x)$ a numărului X de maşini vândute într-o zi să fie utilizată pentru planificarea stocului zilnic daca se stie că

$$F(x) = 0.15 I_{[0,1)}(x) + 0.45 I_{[1,2)}(x) + 0.65 I_{[2,3)}(x) + 0.85 I_{[3,4)}(x) +$$

$$0.95 I_{[4,5)}(x) + I_{[5,+\infty)}(x).$$

Soluţie. Variabila aleatoare $X \in \mathcal{X} = \{0, 1, 2, 3, 4, 5\}$. Deoarece $P(X \leq 4) = F(4) = 0.95$, rezultă că dacă Olaf Motors va tine zilnic în stoc 4 autoturisme Prius, atunci clientii vor fi pe deplin satisfacuţi în 95% de cazuri, pe cand daca acest dealer va ţine zilnic în stoc 2 autoturisme, atunci în 35% de cazuri vor exista clienţi cu solicitări de cumpărare nesatisfacute, deoarece $P(X > 2) = 1 - F(2) = 1 - 0.65 = 0.35$.

Exemplul 3. Considerăm funcţia $F(x) = (1 - e^{-2x}) \cdot I_{[0,+\infty)}(x)$. Observăm că $F(x)$ are toate proprietăţile caracteristice f.d., ceea ce se vede şi din graficul ei de mai jos ca este o functie monoton crescatoare, continuă si pentru care $F(-\infty) = 0$, $F(+\infty) = 1$. Aceasta se vede şi din graficul ei dat mai jos



Prin urmare putem considera ca $F(x)$ descrie comportamentul probabilist al unei v.a. X, valorile căreia sunt concentrate pe intervalul $[0, +\infty)$, deoarece:

$$P(X \leq 0) = F(0) = 1 - e^{-2 \cdot 0} = 1 - 1 = 0$$

iar

$$P(X > 0) = 1 - F(0) = 1 - (1 - e^{-2 \cdot 0}) = 1 - 0 = 1.$$

Probleme propuse.

1. Considerând funcţia de distribuţie $F(x)$ din exemplul anterior ca fiind funcţia de distribuţie a duratei vieţii X (măsurată în ani) a unui corp de iluminat pe bază de neon, aflaţi valorile următoarelor probabilităţi: a) $P(X \leq 1)$; b) $P(X > 1)$; c) $P(X = 1)$; d) $P(X > 0)$. Arătaţi că în acest caz $P(a < X \leq b)$=$P(a \leq X \leq b)$=$P(a \leq X < b)$=$F(b) - F(a)$=$e^{-2a} - e^{-2b}$, pentru orice $a \leq b$, $a, b \in \mathbb{R}$.

2. Care din urmatoarele exemple reprezinta funcţii de distribuţie, adică sunt modele probabiliste ale unor variabile aleatoare?

a) $F(x) = (1-p)I_{[0,1)}(x) + I_{[1,+\infty)}(x), p \in (0,1)$;
b) $F(x) = \frac{1}{4}I_{[0,1)}(x) + \frac{1}{8}I_{[1,2)}(x) + I_{[2,+\infty)}(x)$;
c) $F\ (x)$=$\frac{1}{4}I_{[0,1)}(x)$+$\frac{1}{2}I_{[1,2)}(x)$+$I_{[2,+\infty)}(x)$;
d) $F(x)$=$xI_{[0,1]}(x)$+$I_{(1,+\infty)}(x)$; e) $F(x)$=$2xI_{[0,1]}(x)$+$I_{(1,+\infty)}(x)$;
f) Trasaţi graficul fiecarei funcţii.

2.3. Variabile aleatoare de tip discret, distribuţii (repartiţii)

În Teoria Probabilităţilor sunt studiate 3 tipuri de v.a.: discrete, (absolut) continue si singulare, interesante din punct de vedere ale aplicaţiilor fiind doar primele două. Vom începe cu variabilele aleatoare discrete (v.a.d.).

Fie X o v.a. definită pe câmpul probabilist $(\Omega, \mathcal{F}, P)$.

Definiţia 1. Variabila X se numeşte variabilă aleatoare de tip discret dacă mulţimea valorilor posibile ale acesteia este finită sau infinită, cel mult numerabilă, mai exact, dacă putem identifica o mulţime de forma $\mathcal{X}$=$\{x_1$,x_2, ..., $x_n\}$ sau $\mathcal{X}$=$\{x_1$,x_2, ..., x_n,...$\}$ cu proprietatea $P(X \in \mathcal{X}) = 1$, unde x_1<x_2< ...< x_n<... .

Drept exemple de v.a. de tip discret putem lua numarul de steme apărute la aruncarea unei monede de n ori, numărul de puncte apărute la aruncarea unui zar o singură data, numărul de apeluri telefonice înregistrate la Urgenţa Medicală pe parcursul a 24 de ore, numărul de erori descoperite în urma compilării unui soft, numărul de cumpărători, dintr-un număr total de n cumpărători, care au cumpărat un anume produs comercial, etc.,etc.

În paragraful anterior am vazut că orice v.a. X definită pe câmpul probabilist $(\Omega, \mathcal{F}, P)$ poate fi modelată matematic într-o manieră echivalentă, cu ajutorul funţiei ei de distributie $F(X)$ =$P(X \leq X)$. După cum vom vedea în cazul v.a. de tip discret, mai există încă o alternativă, echivalentă cu f.d.

Definiţia 2. Vom numi *distribuţie (repartiţie) probabilistă* (simplu, *dis-*

tribuţie) a v.a. X setul de perechi ordonate $(x_i, p_i)_{i\geq 1}$ sau tabloul de forma

$$X : \begin{pmatrix} x_1 & x_2 & ... & x_n & ... \\ p_1 & p_2 & ... & p_n & ... \end{pmatrix}, \text{ unde } p_i = P(X = x_i) \geq 0,\ i \geq 1,\ \sum_{i\geq 1} p_i = 1.$$

Următoarea afirmaţie arată că, în caz discret, f.d. şi distribuţia v.a. X sunt două forme echivalente de modelare matematica (probabilistă) a ei.

Propoziţie. *F.d. şi distribuţia unei v.a. X de tip discret sunt legate între ele conform următoarelor formule:*

a) $F(x) = \sum\limits_{i:x_i\leq x} p_i$;

b) mulţimea de valori posibile ale v.a. X coincide cu mulţimea $\mathcal{X}$=$\{x_1$, x_2, ..., x_n, ...$\}$=$\{x \in \mathbb{R} : F(x) - F(x-0) > 0\}$ iar probabiltăţile

p_i= $P(X = x_i) = F(x_i) - F(x_i - 0)$, $i \geq 1$.

Remarcă. Din această propoziţie rezultă că f.d. şi distribuţia unei v.a. de tip discret sunt două forme echivalente de modelare probabilistă ce descriu legea care guvernează comportamentul probabilist al v.a.. Mai mult, o v.a. de tip discret este definită, dacă se cunoaşte legea ei de repartiţie: sau sub formă de funcţie de repartiţie, sau sub formă de repartiţie. Mai observăm că formula de legătură dintre distribuţia şi f.d. a v.a X (vezi p.a) din propozidia de mai sus) mai poate fi scrisă şi în forma

$$F(x) = \sum_{i\geq 1} p_i I_{[x_i,+\infty)}(x).$$

Exemplul 1. Revista *The American Almanac of Jobs and Salaries,* a comunicat, în urma unui studiu statistic pe anii $1994-95$, că din numărul total de absolvenţ licenciaţi la specializarea Contabilitate care şi-au găsit un loc de muncă 25% s-au angajat în sectorul public. Cu ce este egală probabilitatea că alegând la întamplare 15 astfel de absolvenţi vom descoperi că cel puţin 3 dintre ei vor fi angajaţi din sectorul public.

Soluţie. Spatiul de evenimente elementare poate fi descris ca fiind Ω=$\{(a_1, a_2, ..., a_{15}) : a_k \in \{0{,}1\},\ k = \overline{1,15}\}$, unde

$$a_k = \begin{cases} 1, \textit{ daca absolventul nr. } k \textit{ va fi unul angajat din sectorul public}, \\ \qquad\qquad\qquad 0, \textit{ în caz contrar.} \end{cases}$$

Câmpul de evenimente aleatoare $\mathcal{F}$=$\{A : A \subseteq \Omega\}$, iar pentru fiecare eveniment elementar $\{(a_1, a_2, ..., a_{15})\}$, ştiind ca evenimentele aleatoare $\{a_k\}$

sunt independente si probabilitătile lor sunt egale cu

$$P\{a_k\} = (0.25)^{a_k}(1-0.25)^{1-a_k} = (0.25)^{a_k}(0.75)^{1-a_k},\ k=\overline{1,15}\ ,$$

putem determina probabilitatea lui

$$P\{(a_1,a_2,...,a_{15})\} = P\{a_1\}P\{a_2\}...P\{a_1\} = \prod_{k=1}^{15}(0.25)^{a_k}(0.75)^{1-a_k} =$$

$$= (0.25)^{\sum\limits_{k=1}^{15} a_k}(0.75)^{15-\sum\limits_{k=1}^{15} a_k}.$$

Deci pentru orice eveniment $A \in \mathcal{F}$ putem calcula probabilitatea lui dupa formula:

$$P(A) = \sum_{(a_1,a_2,...,a_{15})\in\Omega\ :\ (a_1,a_2,...,a_{15})\in A} P\{(a_1,a_2,...,a_{15})\}.$$

Notăm prin X numărul absolvenţilor angajaţi în sectorul public din cei 15 absolvenţi angajaţi în câmpul muncii. Observăm ca X poate fi privită ca fiind v.a definită pe câmpul de probabilitate $(\Omega,\ \mathcal{F},\ P)$ mai sus, unde $X \in \{0,$ $1,\ ...,\ 15\}$ cu probabilitatea 1, $\{X{=}k\}{=}\{(a_1,\ a_2,\ ...,\ a_{15}) \in \Omega :\ X(a_1,\ a_2,\ ...,$ $a_{15}) = k\}{=}\{(a_1,\ a_2,\ ...,\ a_{15}) \in \Omega :\ a_1+\ a_2+\ ...+\ a_{15} = k\},\ k = \overline{1,15}$. Prin urmare putem spune ca

$$P\{X=k\} = \sum_{(a_1,a_2,...,a_{15})\in\Omega\ :\ (a_1,a_2,...,a_{15})\in\{X=k\}} P\{(a_1,a_2,...,a_{15})\} =$$

$$= \sum_{(a_1,a_2,...,a_{15})\in\Omega\ :\ a_1+a_2+...+a_{15}=k} P\{(a_1,a_2,...,a_{15})\} =$$

$$\sum_{(a_1,a_2,...,a_{15})\in\Omega\ :\ a_1+a_2+...+a_{15}=k} (0.25)^{\sum\limits_{k=1}^{15} a_k}(0.75)^{15-\sum\limits_{k=1}^{15} a_k} =$$

$$(0.25)^k(0.75)^{15-k} \cdot \sum_{(a_1,a_2,...,a_{15})\in\Omega\ :\ a_1+a_2+...+a_{15}=k} 1 =$$

$$card(\{(a_1,a_2,...,a_{15})\in\Omega :\ a_1+a_2+...+a_{15}=k\})\cdot(0.25)^k(0.75)^{15-k} =$$

$$= \mathbb{C}_n^k(0.25)^k(0.75)^{15-k},\ k=\overline{1,15}.$$

V.a. X are distribuţia dată de formula $P\{X=k\}$=$\mathbb{C}_n^k(0.25)^k(0.75)^{15-k}$, $k=\overline{1,15}$. Atunci evenimentul A=\{ *cel puţin 3 din 15 absolvenţi vor fi angajaţi din sectorul public*\}= $\{X \geq 3\}$, iar $P(A)=1-P(\overline{A})=1-P(X<3)$. Dar,

$$P(X<3)=P(\{X=0\}\cup\{X=1\}\cup\{X=2\})=$$

$$P(\{X=0\}\cup\{X=1\}\cup\{X=2\})=$$

$$P\{X=0\}+P\{X=1\}+P\{X=2\}=\sum_{k=0}^{3}\mathbb{C}_n^k(0.25)^k(0.75)^{15-k}=0.236088.$$

Or, $P(X\leq 3)=1-0.236088=0.763\,91$.

Modelul (distribuţia) probabilistă dată de formula

$$P\{X=k\}=\mathbb{C}_n^k(0.25)^k(0.75)^{15-k}, k=\overline{0,15}$$

reprezintă un caz particular al distribuţiei numită *binomială*.

Probleme propuse.

1. O unitate medicala specializată trateaza 10 pacienţi afectaţi de o infectie bacterială letală. Acestora li se aplica un antibiotic eficienţa căruia este confirmată experimental, ca fiind egala cu 95% din cazuri favorabile raportat la numarul total de cazuri. Dacă tratamentul nu este eficient, atunci pacientul decedează.

a) Identificaţi v.a. ce descrie numarul de pacienti care au supravieţuit infectiei în urma aplicarii antibioticului. Descrieti multimea de valori posibile a acestei variabile. Ce reprezintă câmpul de evenimente aleator corespunzător acestei v.a.

b) Prin analogie cu Exemplul 1 din p.2.3., asociaţi repartiţia (modelul) ce descrie comportamentul probabilist a v.a. întroduse.

c) Calculaţi probabilitatea ca vor supravieţui toţi cei 10 pacienţi supuşi tratamentului.

d) Dar cu ce este egală probabilitatea ca vor deceda, cel mult, 2 pacienţi.

e) În cazul decedării, să zicem, a 5 pacienţi, Ministerul Ocrotirii Sănătăţii declanşează o investigaţie privind corectitudinea practicilor unităţii medicale. Ce argumente de ordin probabilist pot fi aduse în favoarea acestei investigaţii guvernamentale.

f) Aflaţi f.d. a v.a. cercetate şi trasaţi graficul ei.

2. Magazinul "Torrent Computers" specializat în vânzarea calculatoarelor a stabilit, în baza activiţii sale de lunga durată, că numărul X de calculatoare vândute zilnic este guvernată de distribuţia

$$X : \left(\begin{array}{ccccccc} 0 & 1 & 2 & 3 & 4 & 5 & 6 \\ 0.05 & 0.1 & 0.2 & 0.2 & 0.2 & 0.15 & 0.1 \end{array} \right).$$

Calculaţi următoarele probabilităţi: a) $P(3 \leq X \leq 6)$; b) $P(X > 3)$; c) $P(X \leq 4)$; d) $P(2 < X \leq 5)$.

3. V.a. X reprezintă suma punctelor aparute la aruncarea unui zar "perfect" de doua ori succesiv. Aflaţi distribuţia probabilistă a v.a. X şi calculaţi probabilitatea ca X va lua valori din intervalul $[-1, 4)$.

4. V.a. X reprezintă produsul punctelor aparute la aruncarea unui zar "perfect" de doua ori succesiv. Aflaţi distribuţia probabilistă a v.a. X şi calculaţi probabilitatea ca X va lua valori din intervalul $[-1, 0)$.

5. V.a. X reprezintă numarul minim din cele doua numere de puncte aparute la aruncarea unui zar "perfect" de doua ori succesiv.Aflaţi distribuţia probabilistă a v.a. X şi calculaţi probabilitatea ca X va lua valori din intervalul $[-1, 5)$.

6. V.a. X reprezintă numarul maxim din cele doua numere de puncte aparute la aruncarea unui zar "perfect" de doua ori succesiv. Aflaţi distribuţia probabilistă a v.a. X şi calculaţi probabilitatea ca X va lua valori din intervalul $[-1, 5)$.

7. Monetăria statului utilizează o maşină de ştanţat monede. La fiecare ştanţare se produc 10 monede. Numărul de ordine a ştanţării de la care maşina se defectează şi incepe sa produca monede rebutate poate fi privit ca o v.a. X a carei distribuţie probabilistă are forma $P(X = k){=}q(1 - p)^{k-1} I_{\{1,2,3,\ldots\}}(k)$, unde $p \in (0, 1)$.

a) Exista vre-o restrictie la alegerea constantei q? Daca da, precizaţi care anume? **b)** Este oare v.a. X de tip discret sau de tip (absolut) continuu? De ce? **c)** Dacă se ştie ca probabilitatea defectării maşinii chiar de la prima ştanţare este este egală cu 0.5, concretizaţi cum arată ditribuţia probabbilistă. Calculaţi probabilitatea că maşina se va defecta la ştanţarea a 10-a. **d)** Deduceţi formula de calcul pentru valorile f.d. $F(x){=}P(X \leq x)$ şi calculaţi cu ajutorul ei probabilitatea că maşina va efectua fără defecte cel puţin 10 ştanţări. **e)** Calculaţi probablitatea că maşina va efectua fără defecte cel puţin 20 ştanţări dacă se ştie că aceasta a efectuat fără defecte cel puţin 10 ştanţări.

2.4. Variabile aleatoare de tip (absolut) continue, densităţi de distribuţie (repartiţie)

Pentru început vom relua Exemplul 3 din p.1.5 în care experimentul aleator considerat rezidă în aruncarea /alegerea la întâmplare a unui punct din intervalul $[0, 1]$. Din punct de vedere matematic acest experiment se descrie, cum am vazut, de campul de probabilitate geometrică $(\Omega,\ \mathcal{F},\ P)$, unde

Ω=$[0, 1]$, $\mathcal{F} = \{A \subseteq \Omega :\ \textit{există}\ mesA\}$ iar $P(A) = \frac{mesA}{mes\Omega}$. Daca vom nota prin X coordonata unui punct aruncat la întâmplare pe intervalul $[0, 1]$, atunci X devine v.a. definită pe câmpul de probabilitate $(\Omega,\ \mathcal{F},\ P)$, unde $X(\omega) = \omega$, pentru orice $\omega \in \Omega$. Mai mult ca atât, dat fiind specificul campului de probabilitatye geometrică funcţia ei de distribuţie

$$F(x) = P(X \le x) = \frac{mes(\Omega \cap (-\infty, x))}{mes\Omega} = \begin{cases} 0,\ \textit{dacă}\ x < 0, \\ x,\ \textit{dacă}\ x \in [0, 1], \\ 1,\ \textit{dacă}\ x > 1. \end{cases}$$

Observam că f.d. $F(x)$ este nu numai continuă, dar si derivabilă aproape pentru orice $x \in \mathbb{R}$, mai exact, cu excepţia punctelor $x \in \{0, 1\}$ şi derivata ei

$$f(x) = \frac{dF(x)}{dx} = \begin{cases} 0,\ \textit{dacă}\ x \notin [0,1], \\ 1,\ \textit{dacă}\ x \in [0, 1], \end{cases} \ge 0,\ \textit{pentru orice}\ x \in \mathbb{R}.$$

În plus,

$$F(x) = \int_{-\infty}^{x} f(u)du.$$

Faptul ca v.a. X din exemplul nostru ia valori dintr-o multime *infinită nenumărabilă* (de putere continuum) cu f.d. continuă şi pentru care exista o functie nenegativa $f : \mathbb{R} \longrightarrow [0, +\infty)$ cu proprietatea de mai sus face ca toate v.a. cu proprietăţi similare să scoată în evidenţă un tip aparte de v.a. marcat de următoarea

Definiţie. Vom spune că v.a. X *este o v.a. de tip (absolut) continuă* dacă pentru funcţia ei de distribuţie $F(x) = P(X \le x)$ exsta o funcţie $f(x) \ge 0$, $\forall x \in \mathbb{R}$, astfel încât

$$F(x) = \int_{-\infty}^{x} f(u)du.$$

Funcţia $f(x)$ cu aceste două proprietăţi se numeşte *densitate de distribuţie (repartiţie) a probabilităţilor,* pe scurt, *(d.d).*
Remarca 1. F.d., dar si d.d. analizate în exemplul de mai sus poartă denumirea de *distribuţie uniformă pe* $[0, 1]$, iar experimentul (imaginar) descris mai sus poate fi simulat (imitat) pe calculator, folosind orice limbaj de programare evoluat (C++, Java, etc.) in baza funcţiei program *random float*, accesarea căreia conduce la generarea unei valori a unei v.a. X uniform distribuite pe acest interval. Dealtfel aceasta funcţie program, împreună cu varianta ei discretă stau la baza algoritmilor de simulare (generare) a valorilor v.a. X cu distributia probabilista arbitrară. Cu alte cuvinte, cu ajutorul calculatorului putem simula orice fenomen aleator, deîndată ce este cunoscută legitatea probabilistă sub forma ei de f.d., distribuţie sau d.d..

Drept alte exemple de v.a. de tip (absolut) continue putem lua durata vieţii unui dispozitiv electronic, durata dintre doua apeluri telefonice succesive inregistrate la un post telefonic, durata dintre doua despagubiri succesive acordate de o companie de asigurări, eroarea de masurare cu ajutorul unui indstrument de măsurare, înălţimea unui barbat (femeie) ales (aleasă) la întâmplare dintr-o populaţie a unei ţări anume, etc., toate acestea fiind exemple de v.a. ale caror valori posibile formează o mulţime infinită nenumărabilă.

Remarca 2. Cuvântul *"absolut"* din Definiţie nu este luat întâmplător în paranteze, deoarece in Teoria probabiltăţilor se regăsesc exemple de v.a. ale căror f.d. este continuă, dar care nu satisfac condiţiilor suplimentare impuse asupra f.d. pentru ca v.a. corespunzătoare să fie *v.a. de tip (absolut) continuă.* Este vorba de *v.a. de tip singular* care in acest curs nu vor fi abordate, datorită faptului ca aceste nu au o aplicatie practica evidentă. Folosirea cuvântului *"densitate"* , la fel, nu apare întâmplător în denumirea de *densitate de repartiţie a probabilităţilor*, deoarece acesta ne aminteşte de sensul fizic al densităţii. Intr-adevar, pentru orice v.a. X de tip absolut continuă cu f.d. $F(x)$, are d.d. $f(x)$ ce poseda proprietatea că pentru o valoare $\triangle x > 0$, oricât de mică ar fi ea,

$$P(x < X \leq x + \triangle x) = F(x + \triangle x) - F(x) = \int_{-\infty}^{x+\triangle x} f(u)du - \int_{-\infty}^{x} f(u)du =$$

$$\int_{x}^{x+\triangle x} f(u)du \simeq f(x)\triangle x.$$

În plus, toate v.a. X de tip (absolut) continuu, dat fiind faptul că acestea au functii de distribuţie continue, au proprietatea că pentru orice numar real a

$$P(X = a) = F(a) - F(a - 0) = F(a) - F(a) = 0.$$

Din acelaşi motiv au loc următoarele formule de calcul ale probabilităţilor în baza f.d. sau a d.d.

$$P(a < X \leq b) = P(a \leq X \leq b) = P(a < X < b) = P(a \leq X < b) =$$

$$= F(b) - F(a) = \int_a^b f(u)du,$$

iar

$$P(X > a) = P(X \geq a) = 1 - F(a) = \int_a^{+\infty} f(u)du.$$

Similar cu cazul discret, si in cazul (absolut) continuu d.d. $f(x)$ este o alternativă la f.d. $F(x)$ a v.a. privita ca model matematic ce descrie comportamentul probabilist al v.a. X. Concluzia aceasta se poate trage din urmatoarea

Propoziţie *(Legatura dintre d.d. şi f.d., proprietăţile d.d).* *Dacă* X este o *v.a. de tip (absolut) continuă , atunci f.d.* $F(x) = P(X \leq x)$ *şi d.d. corespunzătoare* $f(x)$ *au proprietăţile*

1^o. Cunoscând f.d. $F(x) = P(X \leq x)$ a v.a. X putem restabili d.d. corspunzătoare $F(x)$, conform relaţiei

$$f(x) = \frac{dF(x)}{dx};$$

2^o. Cunoscând d.d. $f(x)$ a v.a. X putem restabili f.d. corspunzătoare $F(x)$, conform relaţiei

$$F(x) = \int_{-\infty}^{x} f(u)du;$$

3^o.Orice funcţie $f(x)$ nenegativă, integrabilă Riemann pe $\mathbb{R}$ şi pentru care

$$\int_{-\infty}^{+\infty} f(x)dx = 1$$

reprezintă o d.d. probabilistă a unei v.a. X de tip (absolut) continuă.

Remarca 3. Multimea $\mathcal{X}$=$\{x \in \mathbb{R}:$ *d.d.* $f(x)$ *a v.a.* X *are proprietatea că* $f(x) > 0\}$ arată că valorile posibile ale v.a. X sunt concentrate pe $\mathcal{X}$. Într-adevar:

$$1 = \int_{-\infty}^{+\infty} f(x)dx = \int_{\mathcal{X}} f(x)dx + \int_{\mathbb{R}\setminus\mathcal{X}} f(x)dx = \int_{\mathcal{X}} f(x)dx + \int_{\mathbb{R}\setminus\mathcal{X}} 0\ dx = \int_{\mathcal{X}} f(x)dx.$$

Exemplul 1. Multe modele legate de cercetarea unor fenomene aleatoare ce apar în domeniile tehnicii**,** asigurărilor, demografiei, finanţelor, etc., vizeaza v.a. de tip absolut continuu ale caror comportament probabilist este descris prin intermediul densitaţii de distribuţie de forma

$$f(x) = \begin{cases} 0, \ \textit{dacă}\ x < 0, \\ \lambda e^{-\lambda x}\ , \textit{dacă}\ x \geq 0. \end{cases} = \lambda e^{-\lambda x} \cdot I_{[0,+\infty)}(x)$$

unde parametrul $\lambda > 0$.

Cum $f(x) \geq 0$ şi

$$\int_{-\infty}^{+\infty} f(x)dx = \int_{-\infty}^{0} 0dx + \int_{0}^{+\infty} \lambda e^{-\lambda x}dx = (1 - e^{-\lambda x})\ |_0^{+\infty} = (1-0) - (1-1) = 1,$$

din proprietatea $\mathbf{3^o}$ formulată în propozitia anterioară, conchidem că $f(x)$ reprezintă un model matematic a unei v.a. Echivalentul acestui model este f.d. corespunzătoare d.d. $f(x)$, adică

$$F(x) = \int_{-\infty}^{x} f(u)du = \begin{cases} 0, \ \textit{dacă}\ x < 0, \\ 1 - e^{-\lambda x}\ , \textit{dacă}\ x \geq 0. \end{cases} = (1 - e^{-\lambda x}) \cdot I_{[0,+\infty)}(x).$$

Acest model poartă denumirea de *distribuţie probabilistă exponenţială* cu parametrul $\lambda > 0$ (a se vedea exemplele 9, din p.1.3 şi 3, din p.2.2.).

Probleme propuse.

1. Care din urmatoarele funcţii pot fi considerate ca fiind d.d. (modele probabiliste) ale unor v.a.:

a) $f(x) = e^{-0.25x} I_{[0,+\infty)}(x)$; **b)** $f(x) = 0.25e^{-0.25x} I_{[0,+\infty)}(x)$; **c)** $f(x) = x$, pentru orice $x \in \mathbb{R}$;

d) $f(x)=2xI_{[0,1]}(x)$. În acele cazuri în care $f(x)$ este d.d. trasaţi graficul ei, determinaţii f.d. corespunzătoare şi trasaţi graficul ei.

2. Pentru fiecare din următoarele exemple să se determine valoarea constantei k astfel încât funcţia $f(x)$ să fie o d.d., dupa care determinaţi f.d. $F(x)$ corespunzătoare d.d. $f(x)$ şi trasaţi graficele lor:

a) $f(x)=k(x-1)I_{[1,3]}(x)$; b) $f(x)=kxI_{[0,3]}(x)$; c) $f(x)=kx^3I_{[0,2]}(x)$; d) $f(x)=k(\sin x)I_{[0,\pi/2]}(x)$; e) $f(x)=k(\cos x)I_{[-\pi/2,\pi/2]}(x)$; f) $f(x)=k\sqrt{x}I_{[0,4]}(x)$.

3. Pentru fiecare din următoarele exemple de f.d. $F(x)$ să se determine d.d. $f(x)$ corespunzătoare: a) $F(x)=[1-e^{-\lambda x}(1+x)]I_{[0,+\infty)}(x)$; b) $F(x)=\frac{1}{2\pi}\int\limits_{-\infty}^{x} e^{-\frac{t^2}{2}}dt$; c) $F(x)=\frac{(x+2)}{4}I_{[-2,2]}(x)$.

4. Pentru estimarea cantităţii zilnice Q de combustibil, solicitate unei firme de aprovizionare cu combustibil auto, firma în cauză foloseşte formula $Q=100-10p+E$, unde $p \in [0,8]$. E reprezdntând eroarea de estimare, aceasta este o v.a. dată de d.d. $f(x)=I_{[-20,20]}(x)$.Cantitatea Q solicitata se masoară în mii de galoane, iar preţul p este exprimat în dolari per galon.

a) Cu ce este egală probabilitatea că va fi solicitată o cantitate de combustibil mai mare decât 70000 de galoane, dacă preţul este egal cu 3$ galonul? Dar pentru preţul de 4$ per galon? **b)** Dacă ştim costul mediu variabil de furnizare a benzinei fără plumb, că este dat de formula $C(Q)=\sqrt{Q}/2$, ce v.a. poate fi utilizată pentru a reprezenta profitul zilnic în funcţie de preţul variabil p ? **c)** Cu ce este egală probabilitatea că profitul zilnic va fi unul pozitiv dacă pretţul de vânzare este stabilit să fie 4$ per galon/ Dar daca preţul este egal cu 5$? Dar 3$?

2.5. Variabile aleatoare mixate discrete-continue

Multitudinea de modele probabiliste ale v.a. des întâlnite in cercetarea fenomenelor aleatore din lumea înconjurătoare cuprinde, înafară de v.a. de tip discret şi de tip (absolut) continue, v.a. distribuţiile probabiliste ale cărora întrunesc proprietăţi ale ambelor tipuri de v.a. Este vorba de variabilele aleatoare mixate discrete-continue.

Exemplu. Considerăm v.a. X ce reprezintă durata vieţii, masurată în mii de ore, a memoriei hard a unui laptop marca HP. În Exemplul 9 din p.1.3. distributia probabilista propusa corespundea unei v.a. aleatoare de tip (absolut) continue cu f.d. $F(x)$=$(1 - e^{-10^{-6}x})I_{[0,+\infty)}(x)$. Prin urmare probabilitatea evenimentului A =$\{$*unui laptop nou nouţ îi va pica memoria hard chiar la prima lui pornire*$\}$ = $\{X = 0\}$ este egală cu zero (a se vedea Remarca 2 din paragraful anterior). In realitate, însă, aceasta probabilitate chiar dacă este mică este, totuşi, nenulă. Urmatorul exemplu de f.d. a duratei vieţii memoriei hard a unui laptop marca HP ia în calcul o astfel de posibilitate. Sa zicem, că

$$F(x) = 0.05 \cdot I_{[0+\infty)}(x) \ + 0.95 \cdot (1 - e^{-10^{-6}x})I_{(0,+\infty)}(x).$$

Proprietăţile caracteristice f.d. fiind respectate, rezulta ca funcţia $F(x)$ poate fi luată în calitate de model matematic ce descrie mai adecvat comportamentul probabilist al v.a. X în cazul că serviciul cu controlul calităţii constată că în 5% din astfel de calculatoare hardul a picat chiar de la prima pornire. Într-adevăr, $P(X = 0) = 0.05$, iar probabilitattea că memoria hard va avea o durată de viaţă mai mare decât T este egala cu $P(X > T) = 1 - P(X \leq T) = 1-[0.05 \cdot I_{[0+\infty)}(x) + 0.95 \cdot (1-e^{-10^{-6}x})I_{(0,+\infty)}(t)]$. Astfel, pentru $T = 0$, $P(X > 0) = 1 - 0.05 = 0.95$, iar pentru $T = 43800$ ore, adica aproximativ 5 ani, $P(X > 43800) = 1 - P(X \leq 43800)$. Dar

$$P(X \leq 43800) = P(X = 0) + P(0 < X \leq 43800) = 0.05 + (F(43800) - F(0)) =$$

$$0.05 + [0.05 + 0.95(1 - e^{-10^{-6} \cdot 43800}) - 0.05] =$$

$$0.05 + 0.95 \cdot (1 - e^{-10^{-6} \cdot 43800}) = 9.0712 \times 10^{-2}.$$

Prin urmare, $P(X > 43800) = 1 - 9.0712 \times 10^{-2} = 0.90929$.

Comparând probabilitatea $P(X > 43800) = 0.95715$, că durata vieţii memoriei hardului va fi mai mare de 5 ani, în cazul în care nu este luat în calcul faptul că serviciul cu controlul calităţii constată că în 5% din astfel de calculatoare hardul a picat chiar de la prima pornire, cu aceeaşi probabilitate în caz ca informaţia în cauză este surprinsă in model, probabilitate care este egală cu 0.90929, deducem ca în ultimul caz şansele de supravieţuire a unui astfel de calculator mai mult de 5 ani s-au diminuat cu peste 4%. $\square$

Funcţia de distribuţie anlizată în exemplul de mai sus se încadrează în următoarea

Definiţie. Vom numi v.a. *mixată discretă-continuă* orice v.a. X dacă functia ei de distribuţie $F(x)$ are forma

$$F(x) = pF_d(x) + (1 - p)F_{ac}(x),$$

unde $0 \leq p \leq 1$, $F_d(x)$ reprezintând o f.d. de tip discret iar $F_{ac}(x)$ o f.d. de tip (absolut) continuă.

Astfel, în exemplul de mai sus avem că $p = 0.05$,

$$F_d(x) = I_{[0+\infty)}(x) = \begin{cases} 0, \; \textit{dacă } x < 0, \\ 1, \; \textit{dacă } x \geqslant 0, \end{cases}$$

iar

$$F_{ac}(x) = (1 - e^{-10^{-6}x}) I_{(0,+\infty)}(x) = \begin{cases} 0, \; \textit{dacă } x \leqslant 0, \\ (1 - e^{-10^{-6}x}), \quad \textit{dacă } x > 0. \end{cases}$$

Remarcă. Apelând la forma specifică de scriere a f.d. in cazurele discret şi absolut continuu, deducem ca

$$F_d(x) = \sum_{i \geq 1} p_i I_{[x_i \leq +\infty]}(x),$$

şi corespunde unei v.a. discrete X_1 date de distribuţia

$$X_1 : \begin{pmatrix} x_1 & x_2 & ... & x_n & ... \\ p_1 & p_2 & ... & p_n & ... \end{pmatrix}, \text{ unde } p_i = P(X = x_i) \geq 0, \; i \geq 1, \; \sum_{i \geq 1} p_i = 1,$$

iar

$$F_{ac}(x) = \int_{-\infty}^{x} f(u) du$$

şi corespunde unei v.a. (absolut) continue X_2 cu d.d. $f(x)$. În acest sens, v.a. X dată de f.d.

$$F(x) = p \cdot \sum_{i \geq 1} p_i I_{[x_i \leq +\infty]}(x) + (1 - p) \cdot \int_{-\infty}^{x} f(u) du$$

prezintă un amestec (mixaj) de proprietăţi ale v.a. discrete, dar şi a v.a. (absolut) continue.

Probleme propuse.

1. Dacă presupunem că autobusul de ruta dată circulă la intervale de timp egale cu T_0, $T_0 > 0$, atunci in calitate de model matematic ce descrie timpul de aşteptare X a unui pasager sosirea în staţie a acestui autobus, putem lua v.a. X de tip (absolut) continue cu d.d. $f(x)=\frac{1}{T_0}I_{[0,T_0]}(x)$. Cu alte cuvinte f.d. a timpului X de aşteptare este dată de

$$F(x)=P(X\leq x)=\int_{-\infty}^{x}\frac{1}{T_0}I_{[0,T_0]}(t)\ dt,$$

adică v.a. T este *uniform distribuită* pe intervalul închis $[0,T_0]$. Conform acestui model, avem că probabilităţile $P(X=0)$=$P(X=T_0)$=0. În realitate, însă, se poate întâmpla (cei drept cu o probabilitate foarte mică) că pasagerul sa prindă autobusul chiar odată cu sosirea sa în statie atunci durata lui de aşteptare X=0 sau, dimpotriva, ca pasagerul sa sosească chiar în momentul când autobusul a inchis uşile si pleacă, atunci acesta va trebui sa aştepte autobusul o durată de timp X=T_0. Să zicem că aceste probabilităti sunt egale, respectiv cu $P(X=0)$=$P(X=T_0)$=0.025, în rest comportamentul probabilist al v.a. X, fiind descris de modelul descris mai sus.

a) Arăţi că, în realitate, tumpul de aşteptare este o v.a. X^* dată de f.d.

$$F^*(x)=0.025[I_{[0,+\infty)}(x)+I_{[T_0,+\infty)}(x)]+0.95\cdot\int_{-\infty}^{x}\frac{1}{T_0}I_{[0,T_0]}(t)dt;$$

b) Trasaţi grafic, funcţiile $F(x)$ şi $F^*(x)$;

c) Calculaţi si comparaţi perechile de probabilităţi $P(0<X\leq T_0)$, $P(0<X^*\leq T_0)$ şi $P(0<X\leq T_0/2)$, $P(0<X^*\leq T_0/2)$.

2.6. Variabilă aleatoare multidimensională (vectorială), funcţia ei de distribuţie, funcţii de distribuţie marginale

În paragrafele anterioare din acest capitol am introdus conceptul de variabila aleatoare, ca fiind o singură aplicaţie/funcţie cu valori numerice definită pe mulţimea evenimentelor elementare ce pot fi observate într-un experiment aleator, dar problemele practice vizează şi cazuri când interes prezintă comportamentul probabilistic a două sau chiar mai multe astfel de (v.a.) aplicaţii/funcţii legate de unul şi acelaşi experiment aleator (câmp de probabilitate). De exemplu, din punctul de vedere al unei cercetări statistice, interes prezintă comportamentul probabilistic a unei perechi de variabile aleatoare (X_1,X_2), unde X_1 reprezinta nivelul de studii iar X_2 venitul anual al unui IT-ist, angajat în câmpul muncii, ales la întâmplare din Republica

Moldova. Dealtfel, orice cercetare bazată pe un sondaj statistic, ce implica un eşantion (x_1, x_2, ..., x_n) de observaţii făcute asupra unei v.a. X, pentru a aplica Statistica Matematică, se admite punctul de vedere al matematicianului, conform căruia (x_1, x_2, ..., x_n) pot fi privite ca fiind n v.a. independente, identic distribuite ca şi v.a. X şi aceasta, spre deosebire de punctul de vedere al statisticianului, care a înregistrat (colectat) nişte valori numerice concrete x_1, x_2, ..., x_n. Toate aceste exemple conduc la notiunea de v.a. multidimensionale sau vectoriale.

Definiţia 1. Vom numi *variabilă aleatoare* $n-$*dimensionala (vectorială)* orice vector X=(X_1, X_2, ..., X_n), unde X_1, X_2, ..., X_n sunt v.a. definite pe unul şi acelaşi câmp de probabilitate $(\Omega, \mathcal{F}, P)$.

Următoarea Teoremă ne arată că putem formula o definiţie echivalentă a noţinii de v.a. multidimensională.

Teorema 1. *Vectorul* X = (X_1, X_2, ..., X_n) *este o v.a. n-dimensională definită pe câmpul de probabilitate* (Ω, $\mathcal{F}$, P) *dacă şi numai dacă mulţimea*

$$\{\omega \in \Omega : X_1(\omega) \leq x_1, X_2(\omega) \leq x_2, ..., X_n(\omega) \leq x_n\} \in \mathcal{F}, \forall (x_1, x_2, ..., x_n) \in \mathbb{R}^n.$$

Remarca 1. Aceasta teoremă garantează, de fapt, că pentru orice v.a. X = (X_1, X_2, ..., X_n) mulţimile invocate sunt evenimente aleatoare şi prin urmare putem calcula probabilităţile lor, care pe scurt pot fi scrise ca fiind

$$P(X \leq x) = P(X_1 \leq x_1, X_2 \leq x_2, ..., X_n \leq x_n), \text{ unde } x = (x_1, x_2, ..., x_n) \in \mathbb{R}^n.$$

Mai mult, drept consecinţa, se poate arăta că evenimente aleatoare sunt si mulţimile, de exemplu, de forma

$$(a_1 < X_1 \leq b_1, a_2 < X_2 \leq b_2, ..., a_n < X_n \leq b_n),$$

$$(a_1 < X_1, a_2 < X_2, ..., a_n < X_n),$$

pentru orice (a_1, a_2, ..., a_n),(b_1, b_2, ..., b_n) $\in \mathbb{R}^n$, $a_i \leq b_i$, $i = \overline{1,n}$, etc. Dar anume astfel de mulţimi şi prezintă interes din punct de vedere practic. Mai observăm că noţiunea de *v.a. multidimensională* generalizează noţiunea de v.a. studiată anterior şi care poate fi privită ca *v.a. unidimensională*.

Ca şi pentru v.a. unidimensionale, drept alternativă la noţiunile de vector aleator X = (X_1, X_2, ..., X_n) şi de câmp de probabilitate $(\Omega, \mathcal{F}, P)$ pe care este definit acesta, în ansamblu, privite ca model matematic ce descrie comportamentul probabilist al v.a. X, se foloseşte pe larg, inclusiv în Statistica

Matematică, noţiunea de funcţie de distribuţie a unei v.a. multidimensionale X.

Definiţia 2. *Fie* $(\Omega, \mathcal{F}, P)$ un câmp de probabilitate şi $X = (X_1, X_2, ..., X_n) : \Omega \longrightarrow \mathbb{R}^n$ o variabilă aleatoare n-dimensională definită pe el. Atunci funcţia $F : \mathbb{R}^n \longrightarrow \mathbb{R}$ definită prin relaţia

$$F(x) = P(X \leq x) = F(x_1, x_2, ..., x_n) =$$

$$= P(X_1 \leq x_1, X_2 \leq x_2, ..., X_n \leq x_n),\ \forall\ x = \ (x_1, x_2, ..., x_n) \in \mathbb{R}^n$$

se numeşte *funcţie de distribuţie (f.d.) a v.a. n- dimensionale* X sau *funcţie de distribuţie (în ansamblu) a v.a.* X_1, X_2, ..., X_n.

Exemplul 1. În legătură cu experimentul aleator ce constă în aruncarea unei perechi de zaruri "perfecte" considerăm v.a. X şi Y , ce coincid, respectiv, cu indicatorul evenimentului A=\{*suma punctelor apărute va fi pară*\} şi B=\{*produsul punctelor apărute va fi par*\}. Să se determine funcţia de distribuţie a v.a. bidimensionale (X, Y).

Soluţie. Deoarece mulţimea de valori posibile al v.a. X sau ale v.a. Y este mulţimea $\{0, 1\}$, rezultă că v.a. bidimensională (X, Y) ia valori din mulţimea $\{(0, 0),\ (0{,}1),\ (1, 0),\ (1, 1)\}$. Folosind definiţia clasică aflăm că $P(X = 0, Y = 0)$=$P(\overline{A}\ \overline{B}) = 0$, $P(X = 0, Y = 1)$=$P(\overline{A}\ B) = 1/2$, $P(X = 1, Y = 0)$=$P(A\overline{B}) = 1/4$, $P(X = 1, Y = 1)$=$P(AB) = 1/4$. F.d. $F(x, y)$ a v.a. $(X,\ Y)$ este o funcţie definită pentru orice $(x,\ y) \in \mathbb{R}^2$. Dar, ţinând cont de probabilităţile anterioare, deducem că $P(X \leq x,\ Y \leq y)$=0, pentru orice

$$(x, y) \in (-\infty, 0) \times (-\infty, +\infty) \cup [0, 1) \times (-\infty, 1) \cup [1, +\infty) \times (-\infty, +0),$$

$P(X \leq x, Y \leq y) = 1/2$, pentru orice $(x, y) \in [0, 1) \times [1, +\infty)$, $P(X \leq x$, $Y \leq y)$=1/4, pentru orice $(x, y) \in [1, +\infty) \times [0, 1)$, $P(X \leq x,\ Y \leq y)$=1, pentru orice $(x, y) \in [1, +\infty) \times [1, +\infty)$. Cu alte cuvinte, am definit f.d. $F(x, y)$ pe intreg spaţiul $\mathbb{R}^2$.

Teorema 2 ***(Proprietăţile caracteristice f.d. ale v.a. multidimensionale*** *). O funcţie de n- variabile* $F(x_1,\ x_2,\ ...,\ x_n)$ *definită* pe $\mathbb{R}^n$ *cu valori în* $\mathbb{R}$ *poate fi considerată f.d. a unui vector aleator n-dimensional dacă şi numai dacă aceasta posedă urmatoarele proprietăţi*

1^0. $\sum\limits_{\varepsilon_1=0}^{1} \sum\limits_{\varepsilon_2=0}^{1} ... \sum\limits_{\varepsilon_n=0}^{1} (-1)^{\varepsilon_1+\varepsilon_2+...+\varepsilon_n} F(\varepsilon_1 a_1 + (1-\varepsilon_1) b_1,\ \varepsilon_2 a_2 + (1-\varepsilon_2) b_2,$ $...,\ \varepsilon_n a_n + (1-\varepsilon_n) b_n) \geq 0$ *deîndată ce* $(a_1,\ a_2,\ ...,\ a_n)$, $(b_1,\ b_2,\ ...,\ b_n) \in \mathbb{R}^n$, $a_1 \leq b_1,\ a_2 \leq b_2,\ ...,\ a_n \leq b_n$;

2^0. *$F(x_1,\ x_2,\ ...,\ x_n)$ este continuă la dreapta pentru fiecare variabilă x_i în parte, adică,*

$$\lim_{a_k \downarrow x_i} F(x_1, x_2, ..., a_k, ..., x_n) = F(x_1, x_2, ..., x_i, ..., x_n)$$

pentru orice şir monoton descrescator de valori a_k care tinde la x_i, atunci când k tinde la $+\infty$, fapt ce se noteaza, pe scurt,

$$F(x_1, x_2, ..., x_i + 0, ..., x_n) = F(x_1, x_2, ..., x_i, ..., x_n);$$

3^0. $\lim\limits_{x_i \longrightarrow -\infty} F(x_1, x_2, ..., x_n) = F(x_1, x_2, ..., -\infty, ..., x_n) = 0$ *pentru* $\forall i = \overline{1, n}$ *şi* $F(+\infty, +\infty, ..., +\infty) = \lim\limits_{x_i \longrightarrow +\infty, i=\overline{1,n}} F(x_1, x_2, ..., x_n) = 1.$

Remarca 2. Ca şi în cazul f.d. ale v.a. unidimensionale, pentru f.d. ale v.a. multidimensionale este valabila Remarca 3, p.2.2., ceea ce arata că orice functie de n variabile care satisface proprietaţile $\mathbf{1}^0 - \mathbf{3}^0$ de mai sus poate fi considerată un model probabilist, adică f.d. a unei v.a. multidimensionale. Mai mult, are loc următoarea

Propoziţie. *Dacă v.a.* $X = (X_1,\ X_2,\ ...,\ X_n)$ este dată de f.d. $F(x_1,\ x_2,\ ...,\ x_n)$, atunci:

a)

$$P(a_1 < X_1 \le b_1, a_2 < X_2 \le b_2, ..., a_n < X_n \le b_n) =$$

$$\sum_{\varepsilon_1=0}^{1}\sum_{\varepsilon_2=0}^{1}...\sum_{\varepsilon_n=0}^{1}(-1)^{\varepsilon_1+\varepsilon_2+...+\varepsilon_n}F(\varepsilon_1 a_1+(1-\varepsilon_1)b_1, \varepsilon_2 a_2+(1-\varepsilon_2)b_2, ..., \varepsilon_n a_n+(1-\varepsilon_n)b_n),$$

$(a_1,\ a_2,\ ...,\ a_n),\ \in \mathbb{R}^n,\ a_1 \le b_1,\ a_2 \le b_2,\ ...,\ a_n \le b_n;$

b) f.d. $F_i(x)$ a fiecărei v.a. X_i în parte poate fi refăcută din f.d. în ansamblu a v.a. $(X_1,\ X_2,\ ...,\ X_n)$ după formula

$$F_i(x) = F(+\infty, +\infty, ..., x, ..., +\infty).$$

Consecinţă. *Dacă v.a. $X = (X_1,\ X_2)$ este dată de f.d. $F(x_1,\ x_2)$, atunci:*

$$P(a_1 < X_1 \le b_1, a_2 < X_2 \le b_2) = F(b_1, b_2) + F(a_1, a_2) - F(a_1, b_2) - F(b_1, a_2),$$

deîndată ce $a_1 \le b_1,\ a_2 \le b_2$.

Definiţia 3. Funcţiile de distribuţie determinate conform p.b) din propozitia anterioară se numesc *f.d. marginale ale v.a.* X_i, $i = \overline{1, n}$.

Exemplul 2. Considerăm următoarea funcţie de doua variabile

$$F(x,y) = \begin{cases} 0, \; \textit{dacă} \; \min(x,y) < 0, \\ \min(x,y), \; \textit{dacă} \; 0 \leq \min(x,y) < 1, \\ 1, \; \textit{dacă} \; 1 \leq \min(x,y). \end{cases}$$

Să se arate ca $F(x,y)$ este o f.d. a unei v.a. bidimensionale (X,Y) şi să se afle f.d. a fiecărei v.a.

Soluţie. Proprietăţile $\mathbf{1}^0 - \mathbf{3}^0$ caracteristice f.d. de mai multe variabile se verifică cu uşurinţa, ţinând cont şi de faptul că $\min(x,y)$ este o funcţie continuă ca funcţie de două variabile x,y, deoarece $\min(x,y) = (x+y-|x-y|)/2$. Prin urmare $F(x,y)$ poate fi privita ca fiind o f.d. a unei v.a. (X,Y). Atunci, conform punctului b) din propozitia de mai sus distribuţiile (*marginale)* a fiecarei v.a. în parte sunt egale, respectiv, cu

$$F_X(x) = F(x,+\infty) = \begin{cases} 0, \; \textit{dacă} \; x < 0, \\ x, \; \textit{dacă} \; 0 \leq x \leq 1, \\ 1, \; \textit{dacă} \; 1 < x, \end{cases}$$

$$F_Y(y) = F(+\infty,) = \begin{cases} 0, \; \textit{dacă} \; y < 0, \\ y, \; \textit{dacă} \; 0 \leq y \leq 1, \\ 1, \; \textit{dacă} \; 1 < y. \end{cases}$$

Concluzie: ambele v.a. X şi Y sunt (a se vedea remarca 1 din p.2.4.) *uniform, identic distribuite pe segmentul* $[0,1]$.

2.7. Tipurile de variabile aleatoare multidimensionale (bidimensionale), distribuţii, densităţi de distribuţie, independenţa v.a.

Pentru simplitate, fără a afecta cazul general, vorbind despe variabile aleatore vectoriale de dimensiunea n, vom considera cazul $n=2$, adică vom considera cazul v.a. bidimensionale. Aşadar, fie (X,Y) o v.a. bidimensională definită pe câmpul de probabilitate $(\Omega,\mathcal{F},P)$ sau, ceea ce este echivalent, dată (guvernată) de funcţia de distribuţie $F(x.y)$, notat pe scurt $(X,Y) \sim (\Omega,\mathcal{F},P)$ sau $(X,Y) \sim f.d. \; F(x,y)$. Ca şi în cazul v.a. unidimensionale, din punct de vedere al aplicaţiilor lor practice, cel ma des sunt întâlnite v.a. multidimensionale de tip discret şi de tip (absolut) continue, inclusiv in varianta lor mixată discret-continue.

Dacă o v.a. este de tip discret se poate afla folosind

Definiţia 1. Vom spune ca v.a. (X,Y) *este o v.a. de tip discret* dacă fiecare din v.a. (unidimensionale) X, Y este v.a de tip discret.

Din definiţie rezultă că v.a. (X,Y) este o v.a. de tip discret dacă mulţimile de valori posibile $\mathcal{X}$, $\mathcal{Y}$ ale v.a. X,Y au formele $\mathcal{X}=\{x_1,x_2, ..., x_n\}$ sau $\mathcal{X}=\{x_1,x_2, ..., x_n,...\}$ şi respectiv $\mathcal{Y}=\{y_1,y_2, ..., y_n\}$ sau $\mathcal{Y}=\{y_1,y_2, ..., y_n,...\}$, cu proprietatea $P((X,Y) \in \mathcal{X} \times \mathcal{Y}) = 1$, unde $x_1<x_2< ...< x_n<...$ iar $y_1<y_2< ...< y_n<...$, $\mathcal{X} \times \mathcal{Y}$ fiind produsul cartezian ale mulţimilor $\mathcal{X}$, $\mathcal{Y}$.

Definiţia 2. Vom numi *distribuţie probabilista a v.a. bidimensionale de tip discret* (X,Y) orice set de forma $\{(x_i,y_j), p_{ij}\}_{i,j\geq 1}$ sau orice tabel de forma

$X \setminus Y$	y_1	y_2	...	y_j	...
x_1	p_{11}	p_{12}	...	p_{1j}	...
x_2	p_{21}	p_{22}	...	p_{2j}	...
...	...	...	...	...	...
x_i	p_{i1}	p_{i2}	...	p_{ij}	...
...	...	...	...	....	...

unde $p_{ij} = P(X = x_i, Y = y_j) \geq 0$, $\sum\limits_{i,j\geq 1} p_{ij} = 1$.

Tabelul din exemplul analizat mai sus este, din cate vedem, un exemplu de repartiţie probabilista, iar scrierea f.d. $F(x,y)$ arată că, ştiind distribuţia v.a. (X,Y) putem scrie f.d., folosind formula

$$F(x,y) = \sum_{i,j:x_i\leq x, y_j\leq y} p_{ij}.$$

Mai mult, ştiind distribuţia probabilista a v.a. (X,Y), putem afla distribuţia probabilistă a fiecrei dintre v.a. X, Y. Astfel,

$$P(X = x_i) = P(\{X = x_i\} \cap \Omega) = P(\{X = x_i\} \cap \bigcup_{j\geq 1}\{Y = y_j\}) =$$

$$P(\bigcup_{j\geq 1}(\{X = x_i\} \cap \{Y = y_j\})) = P(\bigcup_{j\geq 1}\{X = x_i, Y = y_j\}) =$$

$$= \sum_{j\geq 1} P(X = x_i, Y = y_j) = \sum_{j\geq 1} p_{ij} = p_{i\cdot}, \text{ pentru orice } i \geq 1.$$

Analogic, $P(Y = y_j) = \sum\limits_{i\geq 1} p_{ij} = p_{\cdot j}$, pentru orice $j \geq 1$. Aceste două distribuţii poartă denumirea de *distribuţii marginale*, prin analogie cu noţiunea de f.d. marginală.

Exemplul 1 ***(Continuare).*** În condiţiile Exemplului 1 din paragraful anterior sa se determine distribuţia v.a. (X, Y), unde X, Y reprezintă indicatorii parităţii sumei si respectiv produsului punctelor apărute. Să afle distribuţia fiecărei v.a. X, Y în parte.

Soluţie. Aşa cum am văzut, v.a.

$$(X, Y) \in \mathcal{X} \times \mathcal{Y} = \{0, 1\} \times \{0, 1\} = \{(0, 0)(0, 1), (1, 0), (1, 1)\}$$

şi $P((X, Y) \in \mathcal{X} \times \mathcal{Y})$=1. iar valorile p_{ij}=$P(X = i, Y = j)$, i,j =0, 1, valori care au stat la baza aflării f.d. $F(x, y)$, reprezintă distribuţia centralizată in urmatorul tabel

$$\begin{array}{ccc} X \setminus Y & 0 & 1 \\ 0 & 0 & 0.5 \\ 1 & 0.25 & 0.25 \end{array},$$

după care scrierea f.d. $F(x, y)$ a v.a.(X, Y) devine simplă

$$F(x, y) = P(X \le x, Y \le y) =$$
$$\begin{cases} 0, & \text{dacă } (x, y) \in (-\infty, 0) \times (-\infty, +\infty)\cup \\ & \cup\ [0, 1) \times (-\infty, 1) \cup [1, +\infty) \times (-\infty, +0), \\ 0.5, & \text{dacă } (x, y) \in [0, 1) \times [1, +\infty), \\ 0.25, & \text{dacă } (x, y) \in [1, +\infty) \times [0, 1), \\ 1, & \text{dacă } (x, y) \in [1, +\infty) \times [1, +\infty). \end{cases}$$

Acest exemplu sugerează, că şi în cazul v.a. multidimensionale de tip discret, ca şi în cazul v.a. unidimensionale de tip discret, o alternativă la f.d., in calitate de model probabilist, poate servi noţiunea de distribuţie probabilistă.

Având distribuţia (în ansamblu) a v.a. X şi Y scrisă mai sus şi folosind formulele pentru aflarea distribuţiilor marginale, gasim că indicatorul X a parităţii sumei punctelor apărute la aruncarea a două zaruri perfecte are distribuţia

$$X : \begin{pmatrix} 0 & 1 \\ 0.5 & 0.5 \end{pmatrix},$$

iar indicatorul Y a parităţii produsului punctelor apărute la aruncarea a două zaruri perfecte are distribuţia următoare

$$Y : \begin{pmatrix} 0 & 1 \\ 0.25 & 0.75 \end{pmatrix},$$

fapt ce coincide si cu rezultatele calculelor directe.

Un alt experiment aleator (fie şi imaginar) care conduce la identificarea unui nou tip de v.a. este descris în

Exemplul 2. Considerăm aruncarea (sau alegerea) unui punct la întâmplare în patratul de latură 1, adică în domeniul $[0,1]\times[0,1]\subset\mathbb{R}^2$.

Notăm prin (X,Y) coordonatele acestui punct. Date fiind condiţiile experimentului, acest vector este un v.a. bidimensional definit pe câmpul de probabilitate geometrica $(\Omega,\mathcal{F},P)$, unde Ω=$[0,1]\times[0,1]$, $\mathcal{F}=\{A\subseteq\Omega:$ *există* $mesA\}$ iar $P(A)$=$\frac{mesA}{mes\Omega}$=$mesA$, pentru orice eveniment $A\in\mathcal{F}$, deoarece $mes\Omega$=$mes([0,1]\times[0,1]$=1. Observăm că aplicaţia (X,Y) este o aplicaţie definită pe Ω=$[0,1]\times[0,1]$ cu valori in (X,Y) conform regulii

$$(X,Y):(\omega_1,\omega_2)\longrightarrow(X(\omega_1,\omega_2),Y(\omega_1,\omega_2))=(\omega_1,\omega_2)$$

pentru orice $(\omega_1,\omega_2)\in[0,1]\times[0,1]$. Evenimentul aleator

$$A=\{(\omega_1,\omega_2)\in[0,1]\times[0,1]:X(\omega_1,\omega_2)\le x,Y(\omega_1,\omega_2)\le y\}=$$
$$\{(\omega_1,\omega_2)\in[0,1]\times[0,1]:X(\omega_1,\omega_2)\le x,Y(\omega_1,\omega_2)\le y\}=$$
$$\{(\omega_1,\omega_2)\in[0,1]\times[0,1]:\omega_1\le x,\omega_2\le y\},$$

iar

$$mesA=mes([0,1]\times[0,1]\cap(-\infty,x]\times(-\infty,y]).$$

Putem, aşadar, determina f.d. $F(x,y)$ a v.a. bidimensionale (X,Y):

$$F(x,y)=P(X\le x,Y\le y)=\frac{mesA}{mes\Omega}=\begin{cases}0, & dac\breve{a}\ X<0\ sau\ Y<0,\\ x, & dac\breve{a}\ 0\le x\le 1,\ y>1,\\ y, & dac\breve{a}\ \ x>1,\ 0\le y\le 1,\\ xy, & dac\breve{a}\ 0\le x\le 1,\ 0\le y\le 1,\\ 1, & dac\breve{a}\ x>1,\ y>1.\end{cases}$$

Pentru noi acest exemplu este remarcabil prin faptul ca funcţia de distribuţie obţinută are proprietatea ca există o funcţie $f(x,y)\ge 0$, pentriu orice $(x,y)\in\mathbb{R}^2$, integrabilă, astfel încât

$$F(x,y)=\int_{-\infty}^{x}\int_{-\infty}^{y}f(x,y)dudv,$$

unde

$$f(x,y)=I_{[0,1]\times[0,1]}(x)=\begin{cases}0, dac\breve{a}\ (x,y)\notin[0,1]\times[0,1],\\ 1, dac\breve{a}\ (x,y)\in[0,1]\times[0,1].\end{cases}$$

Această proprietate este în consens cu

Definiţia 3. Vom spune ca v.a. bidimensionala (X, Y) *este de tip (absolut) continue* daca funcţia ei de distribuţie are proprietatea că există o funcţie $f(x, y) \geq 0$, pentriu orice $(x, y) \in \mathbb{R}^2$, integrabilă, astfel încât

$$F(x,y) = \int\limits_{-\infty}^{x}\int\limits_{-\infty}^{y} f(u,v)dudv.$$

Funcţia $f(x, y)$ cu aceste proprietăţi se numeşte *densitate de distribuţie (d.d.) a v.a.* (X, Y).

Propoziţia 1. *Dacă* (X, Y) *este o v.a de tip (absolut) continue, atunci f.d.* $F(x, y)$ *şi d.d* $f(x, y)$ *a acestei variabile au următoarele proprietăţi:*

1. $f(x, y) \geq 0$, *pentru orice* $(x, y) \in \mathbb{R}^2$;

2. $F(x,y) = \int\limits_{-\infty}^{x}\int\limits_{-\infty}^{y} f(u,v)dudv$, *pentru orice* $(x, y) \in \mathbb{R}^2$;

3. Cu excepţia unei mulţimi A de puncte din $\mathbb{R}^2$, aria (măsura) căreia $mesA = 0$, există $\frac{\partial^2 F(x,y)}{\partial x \partial y}$ şi

$$\frac{\partial^2 F(x,y)}{\partial x \partial y} = f(x,y);$$

4. $\int\limits_{-\infty}^{+\infty}\int\limits_{-\infty}^{+\infty} f(x,y)dxdy = 1;$

5. $P(a_1 < X_1 \leq b_1, a_2 < X_2 \leq b_2) = \int\limits_{a_1}^{b_1}\int\limits_{a_2}^{b_2} f(x,y)dxdy$, *pentru orice* (a_1, b_1), $(a_2, b_2) \in \mathbb{R}^2$, $a_1 \leq b_1$, $a_2 \leq b_2$);

6. d.d. $f_1(x)$ a v.a. X şi d.d. $f_2(y)$ a v.a. Y pot fi aflate, stiind, d.d. $f(x, y)$ a v.a. (X, Y) din formulele respective,

$$f_1(x) = \int\limits_{-\infty}^{+\infty} f(x,y)dy, \;\; f_2(y) = \int\limits_{-\infty}^{+\infty} f(x,y)dx.$$

Definiţia 4. *D.d.* $f_1(x)$, $f_2(y)$ determinate prin formulele de la p.6. al Propoziţiei anterioare se numesc *d.d. marginale.*

Remarcă. Din pp. 2-3 a propoziţiei 1 rezultă ca f.d. $F(x, y)$ şi d.d. $f(x, y)$ ale v.a. bidimensionale, privite *ca modele matematice ce descriu comportamentul probabilist al v.a.* (X, Y), *sunt echivalente* in sens că, ştiind f.d.

putem restabili d.d. şi viceversa. Proprietatea 2 arată că în caz (absolut) continuu f.d. $F(x,y)$ este continuă ca funcţie de doua variabile. Reciproca nu este valabila, adică se poate aduce un exemplu de f.d. $F(x,y)$ continuă ca funcţie de doua variabile, dar nu si (absolut) continue ca f.d..

Exemplul 3. Intr-adevăr, f.d.

$$F(x,y)=\begin{cases} 0,\ dac\breve{a}\ \min(x,y)<0,\\ \min(x,y),\ dac\breve{a}\ 0\le\min(x,y)<1,\\ 1,\ dac\breve{a}\ 1\le\min(x,y),\end{cases}$$

adusă drept exemplu în paragraful precedent, este continuă, dar nu exista nicio funcţie cu proprietăţile $1-4$ ale d.d. prezentate in propozitia anterioară. Dealtfel, acesta este un exemplu tipic de f.d. a unei *v.a.* (X,Y) *de tip singular*, acest tip de v.a. nefiind abordat in cursul nostru. Mai mult, observăm ca f.d. marginale $F_1(x)$, $F_2(y)$ ce corespund acesteia reprezintă *f.d. ale unor v.a. uniform distribuite*, mai exact

$$F_1(x)=\begin{cases} 0,\ dac\breve{a}\ x<0,\\ x,\ dac\breve{a}\ x\in[0,1],\\ 1,\ dac\breve{a}\ x>1,\end{cases},\ F_2(y)=\begin{cases} 0,\ dac\breve{a}\ y<0,\\ y,\ dac\breve{a}\ y\in[0,1],\\ 1,\ dac\breve{a}\ y>1.\end{cases}$$

Surprinzător, dar aceleaşi f.d. marginale corespund si f.d. $F(x,y)$ din Exemplul 2 de mai sus. Mai mult, în Exemplul 2, f.d. $F(x,y)=F_1(x)\cdot F_2(y)$, pe când în exemplul nostru f.d. $F(x,y)\neq F_1(x)\cdot F_2(y)$. Explicaţia vine din

Definiţia 5. Vom spune ca *v.a.* X_1, X_2, ..., X_n definite pe unul şi acelaşi câmp de probabilitate *sunt independente* dacă f.d. $F(x_1,\ x_2,\ ...,\ x_n)$ a vectorului aleator $(X_1,\ X_2,\ ...,\ X_n)$ are proprietatea că

$$F(x_1,x_2,...,x_n)=F_1(x_1)\cdot F_2(x_2)\cdot ...\cdot F_n(x_n),$$

unde $F_1(x_1),F_2(x_2)$, ..., $F_n(x_n)$ reprezintă f.d. marginale ce corespund v.a. X_1, X_2, ..., X_n. În caz contrar vom spune că X_1, X_2, ..., X_n *sunt dependente.*

Or, explicaţia promisă rezidă în faptul ca v.a. X,Y ce reprezinta coordonatele unui punct aruncat la întâmplare în patratul $[0,1]\times[0,1]$ sunt, conform definiţiei 5, v.a. independente. În schimb v.a. X,Y ce au f.d. invocată in exemplul din paragraful anterior sunt dependente. De remarcat că aplicarea directă a definiţiei 5 este greoaie, deaceea, putem aplica variantele ei echivalente ce ţin cont de specificul tipului de v.a. şi ce rezultă din

Propoziţia 2. *V.a.* (X, Y) *sunt independente atunci şi numai atunci când*

a) în caz discret: distribuţia lor în ansamblu $\{P(X = x_i, Y = y_j)\}_{i,j\geq 1}$ *are proprietatea că* $P(X = x_i, Y = y_j) = P(X = x) \cdot P(Y = y_j)$ *pentru orice* $i, j \geq 1$;

b) în caz (absolut) continuu: densitatea ei de distribuţie $f(x, y)$ *şi densităţile marginale* $f_1(x)$, $f_2(y)$ *au proprietatea că* $f(x, y) = f_1(x) \cdot f_2(y)$ *pentru orice* $x, y \in \mathbb{R}$.

Exemplul 1. *(Continuare). Pentru a verifica dacă v.a.* X, Y analizate în acest exemplu *sunt sau nu independente* putem aplica afirmatia a) din Propozitia 2. Constatăm cu uşurinţă, de exemplu, ca $P(X = 0, Y = 0) = 0 \neq P(X = 0)P(Y = 0) = 0.5 \cdot 0.25$, ceea ce este suficient să afirmăm că *v.a.* X *şi* Y *sunt dependente.*

Exemplul 3. Compania Ocean Fish are deschise in Republica Moldova două intreprinderi de procesare a peştelui despre care se ştie ca proportiile de procesare a acestora, raportate la intreaga capacitate de procesare zilnica a companiei, reprezintă o v.a. bidimensionala (X, Y) cu d.d. $f(x, y) = (x + y)I_{[0,1]}(x)I_{[0,1]}(y)$. a) Sa se calculeze $P(X \leq 0.5, Y \leq 0.5)$. b) Să se afle d.d. a proporţiei capacităţii de procesare zilnică la care operează fiecare întreprindere în parte. c) Să se calculeze $P(X \leq 0.5)$, $P(Y \leq 0.5)$. d) Sunt oare proporţiile X şi Y independente?

Soluţie. a) Din definiţia f.d. a v.a. de tip (absolut) continuu deducem ca

$$P(X \leq 0.5, Y \leq 0.5) = F(0.5, 0.5) = \int_{-\infty}^{0.5}\int_{-\infty}^{0.5} f(x, y)dxdy =$$

$$\int_{-\infty}^{0.5}\int_{-\infty}^{0.5} (x + y)I_{[0,1]}(x)I_{[0,1]}(y)dxdy =$$

$$\int_0^{0.5}\left(\int_0^{0.5}(x + y)dx\right)dy = \int_0^{0.5}\left[\left(\frac{x^2}{2} + xy\right)\Big|_0^{0.5}\right]dy = \int_0^{0.5}\left(\frac{1}{8} + \frac{y}{2}\right)dy = \frac{2}{16} = 0.125.$$

b) Folosind formula 6 din Propoziţia 1 aflam d.d. marginală $f_1(x)$ a v.a. X:

$$f_1(x) = \int_{-\infty}^{+\infty} f(x, y)dy = \int_0^1 (x + y)I_{[0,1]}(x)dy = \left(x + \frac{1}{2}\right)I_{[0,1]}(x).$$

Din considerente de simetrie, observăm că d.d. marginală $f_2(y)$ a v.a. Y este dată de egalitatea $f_2(y) = \left(y + \frac{1}{2}\right) I_{[0,1]}(y)$. Prin urmare, răspunsul la p. c) este dat de egalităţile

$$P(X \leq 0.5) = \int_0^{0.5} (x + \frac{1}{2})dx = P(Y \leq 0.5) = \int_0^{0.5} (y + \frac{1}{2})dy = \frac{1}{8} + \frac{1}{4} = 0.375.$$

d) Apelând la Definiţia 3 a independenţei v.a. multidimensionale, deducem că v.a. X şi Y sunt *dependente* deoarece

$$\begin{aligned} F(0.5, 0.5) &= P(X \leq 0.5, Y \leq 0.5) = 0.125 \neq F_1(0.5)F_2(0.5) = \\ P(X &\leq 0.5)P(Y \leq 0.5) = 0.375^2 = 0.14063. \end{aligned}$$

Probleme propuse.

1. Care din următoarele tabele reprezintă o distribuţie probabilistă a unui vector aleator?

$$1)\ \begin{array}{cccc} X\backslash Y & -1 & 0 & 1 \\ 0 & 1/8 & 0 & 1/8 \\ 1 & 1/4 & 1/8 & 1/4 \end{array},\ 2)\ \begin{array}{cccc} X\backslash Y & -1 & 0 & 1 \\ 0 & 1/8 & 0 & 1/4 \\ 1 & 1/4 & 1/8 & 1/4 \end{array},3)\ \begin{array}{cccc} X\backslash Y & -1 & 0 & 1 \\ 0 & 1/8 & 0 & 1/8 \\ 1 & 1/4 & 1/2 & 1/4 \end{array}.$$

În caz că avem de a face cu o distribuţie probabilista aflaţi: a) $P(X < 1, -1 \leq Y < 1)$; b) distribuţia fiecărei v.a. în parte; c) daca v.a. X,Y sunt sau nu independente.

2. O pepinieră funcţionează în baza unei echipe de 7 angajaţi din care 3 raspund de sectorul vânzări iar ceilalti 4 sunt responsabili de sectorul grădinărit. Evident, numarul, relativ mic, de angajaţi poate crea dificultăţ legate de absenteism. Numarul de persoane absente în ziua dată din sectorul vânzări, dar şi a acelor absente din sectorul gradinărit reprezintă un vector aleator (X, Y) distribuit conform cu tabelul urmator:

$X\diagdown Y$	0	1	2	3	4
0	0.75	0.025	0.01	0.01	0.03
1	0.06	0.03	0.01	0.01	0.003
2	0.025	0.01	0.005	0.005	0.002
3	0.005	0.004	0.003	0.002	0.001

.

a) Calculaţi probabilitatea că în ziua dată vor absenta mai mult de doi angajaţi (indiferent de profilul lor); b) Aflaţi distribuţia absentarii angajaţilor din sectorul gradinărit. Calculaţi probabilitatea că în ziua dată vor absenta mai mult de doi angajaţi din sectorul grădinărit. c) Sunt oare v.a. X,Y independente?

3. Considerăm doua v.a. bidimensionale (X, Y) şi (U, V) independente. Sunt oare v.a. Y şi U independente? Argumentaţi de ce sau de ce nu.

4. F.d. $F(x, y)$ a v.a. bidimensionale (X, Y) este dată de formula

$$F(x, y) = (1 - e^{-x/10} - e^{-y/2} + e^{-(x+5y)/10})I_{[0,+\infty)}(x)I_{[0,+\infty)}(y).$$

Aflaţi: a) d.d. a v.a. (X, Y), folosind formula 3 din Propozitia 1 anterioară; b) d.d. marginală a v.a. X; c) f.d. a v.a. X; d) dacă v.a. X şi Y sunt independente?

3. Caracteristici numerice ale variabilelor aleatoare

3.1. Parametri de poziţie: valoarea medie, moda, mediana, cuantile

În directă legătură cu cercetarea comportamentului probabilist al unei v.a. (caracteristici statistice) X nu întotdeauna prezintă interes modelul matematic exhaustiv, fie sub formă de camp de probabilitate $(\Omega, \mathcal{F}, P)$ pe care este definită X, fie sub forma de f.d. $F(x)$ a acestei v.a., fie sub forma de distributie, daca X este o v.a. de tip discret, fie sub forma de d.d. $f(x)$, dacă X este o v.a. de tip (absolut) continue, ci doar unele caracteristici numerice sumare, numite şi parametri. Printre ei se află şi parametrii numiţi *parametri de pozitie sau parametri ai tendinţei centrale* şi care reprezinta nişte valori numerice sumare, de referinţă cu care putem compara valorile (posibile) individuale ale v.a X . Sa luăm un exemplu concret de experiment aleator care ne va conduce la noţiunea de valoare medie, pe scurt, medie.

Exemplul 1. Este vorba de o loterie în care sunt emise şi vândute, să zicem, 10000 de bilete şi în care un participant cu un singur bilet de loterie poate câstiga una din sumele băneşti x_1, x_2, ..., x_k, unde $x_1 < x_2 < ... < x_k$. Evident, suma totală acordată biletelor câştigătoare nu poate fi mai mare decât costul tuturor biletelor. Dacă vom considera "câştig" şi valoarea x_0 ce coincide cu preţul unui bilet luat cu semnul minus, chiar si atunci când acesta este necâştigător, atunci câştigul X a unui jucator, ce a cumpărat un singur bilet de loterie, este o v.a. de tip discret ale cărei valori posibile fac parte din mulţimea $\mathcal{X}$=$\{x_0$, x_1, x_2, ..., $x_k\}$. Paticipând de n ori la această loterie, cumpărând de fiecare dată un singur bilet de loterie şi câştigând de n_i ori valoarea x_i , i=0, 1, ..., k, $n_0 + n_1 + ... + n_k$=n, putem calcula media câştigului după formula de calcul, cunoscută în Statistică sub denumirea de medie de selecţie:

$$\overline{x} = \frac{1}{n}\sum_{i=0}^{k} n_i \cdot x_i = \sum_{i=0}^{k} \frac{n_i}{n} \cdot x_i = \sum_{i=0}^{k} f_n(X = x_i) \cdot x_i,$$

unde $f_n(X = x_i)$ este frecvenţa relativă a incasării câştigului x_i jucând de n ori, de fiecare dată cu un singur bilet de loterie. Dar, conform Principiului Regularităţii Statistice, frecvenţa relativă $f_n(X = x_i)$ se apropie, odată cu creşterea lui n, tot mai mult şi mai mult de probabilitatea (teoretică) $P(X = x_i)$, $i=\overline{1,k}$, ceea ce înseamnă că pentru n suficient de mare:

$$\overline{x} = \sum_{i=0}^{k} f_n(X = x_i) \cdot x_i \simeq \sum_{i=0}^{k} P(X = x_i) \cdot x_i.$$

Dar, $\{P(X = x_i)\}_{i=\overline{1,n}}$ reprezintă distribuţia probabilistă a v.a. X. Prin urmare, dacă am cunoaşte, apriori, această distribuţie, am putea evalua o medie a câştigului nostru, pentru a decide, de exempu, dacă merită sau nu să participăm la această loterie. Considerentele de mai sus reprezintă unul din motivele pentru care putem da

Definiţia 1. Vom numi *valoare medie a v.a.* X numărul $\mathbb{E}X$ calculat dupa formula

$$\mathbb{E}X = \begin{cases} \sum\limits_{i\geq 1} x_i P(X = x_i), \text{ în caz discret,} \\ \int\limits_{-\infty}^{+\infty} x f(x) dx, \text{ în caz (absolut) continuu,} \end{cases}$$

considerând că *valoarea medie există*, dacă în caz discret

$$\sum_{i\geq 1} |x_i|\, P(X = x_i) < +\infty,$$

iar în caz (absolut) continuu

$$\int_{-\infty}^{+\infty} |x|\, f(x) dx < +\infty,$$

cu alte cuvinte, *daca suma sau integrala Riemann respectivă converge absolut.*

Remarca 1. Din Definiţia 1 deducem că, atunci când v.a. (in caz discret) ia valori dintr-o mulţime finită de valori sau (in caz (absolut) continuu) d.d. $f(x)$ este nenulă pe un numar finit de intervale de lungimi finite din $\mathbb{R}$, valoarea medie există întotdeauna.

Exemplul 1 *(Continuare).* Dacă se ştie că v.a. X are distribuţia $\{(x_i,p_i)\}_{i=\overline{1,k}}$, $p_i = P(X = x_i) \geq 0, \sum\limits_{i\geq 1}^{k} p_i = 1$, atunci valoarea medie exista întotdeauna şi este egală cu

$$\mathbb{E}X = \sum_{i=1}^{k} x_i p_i \ .$$

Pe de altă parte, convergenţa absolută, nu este o condiţie de prisos impusă pentru a garanta existenta valorii medii. Aceasta o demonstrează

Exemplul 2 (*O v.a. de tip discret care nu posedă valoare medie*). Considerăm v.a. X dată de distribuţia $P(X = (-2)^{n+1}) = \frac{1}{2^n}$, $n = 1,\ 2, \ldots$.

Observăm că seria (suma)

$$\sum_{n=1}^{+\infty} x_n P(X = x_n) = \sum_{i=1}^{+\infty} (-2)^{n+1} \frac{1}{2^n} = \sum_{i=1}^{+\infty} (-2)$$

asociată cu valoarea medie $\mathbb{E}X$, evident, nu converge absolut, prin urmare nu exista valoarea medie a acestei v.a.

Exemplul 3. Considerăm v.a. X ce corespunde coordonatei unui punct ales la întâmplare din intervalul $[0, 1]$, adică v.a. dată de d.d. $f(x) = I_{[0,1]}(x)$. Atunci valoarea ei medie $\mathbb{E}X$ există şi

$$\mathbb{E}X = \int_{-\infty}^{+\infty} x f(x) dx = \int_{-\infty}^{+\infty} x I_{[0,1]}(x) dx = \int_{-\infty}^{0} 0 dx + \int_{0}^{1} x dx + \int_{1}^{+\infty} 0 dx = \frac{1}{2}.$$

Concluzie: Coordonata unui punct ales la întâmplare din intervalul $[0, 1]$ va coincide, in medie, cu mijlocul acestui interval.

Propoziţia 1. ***(Proprietăţile valorii medii).*** *Valoarea medie a unei v.a. posedă următoarele proprietăţi:*

a) Dacă v.a. X este nenegativă cu probabilitatea 1 *şi exista valoarea ei medie $\mathbb{E}X$, atunci $\mathbb{E}X \geq 0$ şi $\mathbb{E}X = 0$* dacă şi numai dacă $P(X = 0)$=1;

b) Dacă valoarea medie a v.a. X există, atunci există valoarea medie a v.a. $aX + b$ şi $\mathbb{E}(aX + b) = a\mathbb{E}X + b$, pentru orice $a, b \in \mathbb{R}$;

c) Dacă valoarea medie a v.a. X există, atunci există valoarea medie a v.a. $|X|$ şi $|\mathbb{E}X| \leq \mathbb{E}\,|X|$;

d) Dacă valoarea medie a v.a. X şi Y există, atunci există şi valoarea medie a v.a. $X + Y$ şi $\mathbb{E}(X + Y) = \mathbb{E}X + \mathbb{E}Y$;

e) Dacă v.a. X şi Y sunt independente şi există valorile lor medii $\mathbb{E}X$, $\mathbb{E}Y$, atunci exista şi valoarea medie a v.a. XY şi $\mathbb{E}(X \cdot Y)$=$\mathbb{E}X \cdot \mathbb{E}Y$.

Exemplul 4. Considerăm două v.a. independente X şi Y pentru care $\mathbb{E}X = 1$, $\mathbb{E}Y = 2$. Să se calculeze valoarea medie a v.a. $3(X+1)Y+5$.

Soluţie. Folosind proprietăţile valorii medii găsim că

$$\mathbb{E}[3(X+1)Y+5] = \mathbb{E}(3XY+3Y)+5] = 3\mathbb{E}XY + 3\mathbb{E}Y + 5 =$$

$$3\mathbb{E}X\mathbb{E}Y + 3\mathbb{E}Y + 5 = 3\cdot 1\cdot 2 + 3\cdot 2 + 5 = 15.$$

Propoziţia 2. ***(Formula de transport).*** *a) Dacă $g : \mathbb{R} \longrightarrow \mathbb{R}$ este o funcţie reală de o singură variabilă şi X o v.a., astfel încât $g(X)$ este o v.a pentru care exista $\mathbb{E}g(X)$, atunci are loc următoarea formulă de calcul*

$$\mathbb{E}g(X) = \begin{cases} \sum\limits_{i\geq 1} g(\ x_i)P(X = x_i), \text{ în caz discret,} \\ \int\limits_{-\infty}^{+\infty} g(x)f(x)dx, \text{ în caz (absolut) continuu;} \end{cases}$$

b) Dacă $g : \mathbb{R}^2 \longrightarrow \mathbb{R}$ este o funcţie reală de două variabile şi (X,Y) o v.a. bidimensională, astfel încât $g(X,Y)$ este o v.a. pentru care exista $\mathbb{E}g(X,Y)$, atunci are loc următoarea formulă de calcul

$$\mathbb{E}g(X,Y) = \begin{cases} \sum\limits_{i\geq 1}\sum\limits_{j\geq 1} g(\ x_i, y_j)P(X = x_i, Y = y_j), \text{ în caz discret,} \\ \int\limits_{-\infty}^{+\infty}\int\limits_{-\infty}^{+\infty} g(x,y)f(x,y)dxdy, \text{ în caz (absolut) continuu.} \end{cases}$$

Din această propoziţie, drept caz particular al proprietăţii b), rezultă următoarea

Consecinţă. *Daca v.a. bidimensionalaă este data de distribuţia ei (in caz discret) sau densitatea ei de distribuţie (in caz absolut continuu), atunci putem calcula valorile medii* $\mathbb{E}X$ şi $\mathbb{E}Y$ conform formulelor:

$$\mathbb{E}X = \begin{cases} \sum\limits_{i\geq 1}\sum\limits_{j\geq 1} x_i P(X = x_i, Y = y_j), \text{ în caz discret,} \\ \int\limits_{-\infty}^{+\infty}\int\limits_{-\infty}^{+\infty} xf(x,y)dxdy, \text{ în caz (absolut) continuu;} \end{cases}$$

$$\mathbb{E}Y = \begin{cases} \sum\limits_{i\geq 1}\sum\limits_{j\geq 1} y_j P(X = x_i, Y = y_j), \text{ în caz discret,} \\ \int\limits_{-\infty}^{+\infty}\int\limits_{-\infty}^{+\infty} yf(x,y)dxdy, \text{ în caz (absolut) continuu;} \end{cases}$$

Remarca 2. Sensul şi utilitatea formulei de transport rezidă în faptul că, atunci când avem de a face cu o v.a. privită ca o funcţie $g(X)$ de v.a. X, indiferent de dimensionalitatea v.a. X, pentru calcularea valorii medii a v.a. compuse $g(X)$ nu este obligatorie cunoaşterea distribuţiei sau densităţii ei de distribuţie.

În plus, dacă v.a. X este de tip discret, atunci şi v.a. $g(X)$ va fi de tip discret, dar dacă v.a. X este de tip (absolut) continuu, atunci v.a. $g(X)$ poate fi sau de tip discret sau de tip (absolut) continuu sau una de tip mixat discret-continuu, aceasta în dependenţă de specificul funcţiei $g(\cdot)$.

Exemplul 5. Considerăm o v.a. X dată de distribuţia $P(X=0)=1-p$, $P(x=1)=p$, $0<p<1$. Atunci, folosind formula de transport pentru $g(x)=X^r$, $r\geq 1$, gasim că $\mathbb{E}X^r=0^r\cdot(1-p)+1^r\cdot p=p^r$. Dealtfel, anticipând, valorile de tipul $\mathbb{E}X^r$, dacă acestea există, se *numesc momente de ordinul* r, valoarea medie $\mathbb{E}X$ fiind un caz particular pentru $r=1$.

Exemlpul 6. Considerăm v.a. bidimensională (X,Y) analizată în exemplul 2, p. 2.7., iar pentru $g(x,y)=\min(x,y)$, v.a. $Z=\min(X,Y)$. Observăm că v.a. (X,Y) are d.d. $f(x,y)=I_{[0,1]\times[0,1]}(x)$, prin urmare, folosind formula de transport, găsim că

$$\mathbb{E}Z=\mathbb{E}\min(X,Y)=\int\limits_{-\infty}^{+\infty}\int\limits_{-\infty}^{+\infty}\min(x,y)f(x,y)dxdy=$$

$$\int\limits_{-\infty}^{+\infty}\int\limits_{-\infty}^{+\infty}\min(x,y)I_{[0,1]\times[0,1]}(x)dxdy=\int\limits_{0}^{1}\int\limits_{0}^{1}\min(x,y)dxdy=1/3.$$

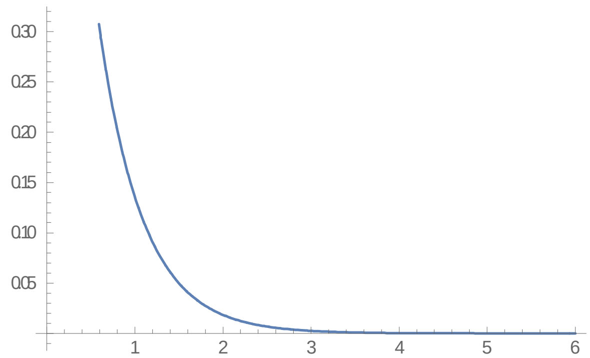
Un alt parametru de poziţie utilizat este *moda,* acesta nefiind un parametru care face parte din categoria parametrilor de tip momente.

Definiţia 2. Vom numi *modă a v.a. X numărul*

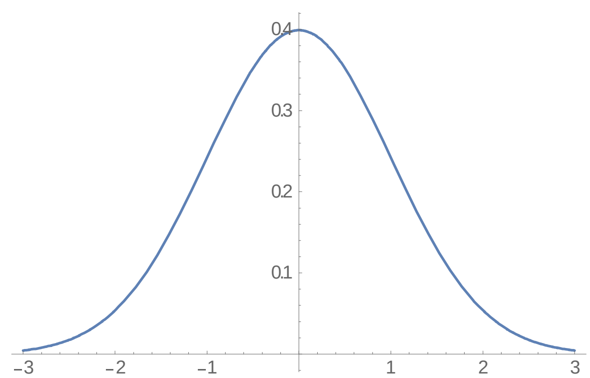
$$\operatorname{mod}(X)=\begin{cases} x_{i_0}\colon \max\limits_{i\geq 1}P(X=x_i)=P(X=x_{i_0}), \text{ în caz discret,}\\ x_0\colon\max\limits_{x\in\mathbb{R}}f(x)=f(x_0), \text{ în caz (absolut) continuu .}\end{cases}$$

dacă acesta există, iar în cazul când $\operatorname{mod}(X)$ există şi este unic, atunci spunem ca *distribuţia* $\{P(X=x_i)\}_{i\geq 1}$ *sau d.d.* $f(x)$ *a v.a.* X *este unimodală*, altfel, dacă $\operatorname{mod}(X)$ există şi nu este unic, atunci spunem ca *distribuţia sau d.d. a v.a. X este multimodală.*

Exemplul 7. (***Exemplu când moda nu există***). Considerăm v.a. X dată de d.d. $f(x) = 2e^{-2x} \cdot I_{[0,+\infty)}(x)$. Graficul d.d. $f(x)$ de mai jos arată în mod evident că $f(x)$ nu posedă o valoare a argumentului x pentru care d.d să ia valoare maximală.



Exemplul 8. (***Exemplu de d.d. unimodală***). Considerăm v.a. X dată de d.d. $f(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$, pentru orice $x \in \mathbb{R}$. Graficul d.d. $f(x)$ de mai jos arată în mod evident că $f(x)$ posedă o singură valoare a argumentului x, $x = 0$ pentru care d.d să ia valoare maximală. Prin urmare $\text{mod}(X) = 0$.



Exemplul 9. (***Exemplu de distribuţie multimodală***). Considerăm v.a. X dată de distribuţia $P(X = 0) = P(x = 1) = 1/2$. Conform definiţiei, $\text{mod}(X) \in \{0, 1\}$, cu alte cuvinte *distribuţia* în cauză *este bimodală.*

Un al parametru de pozitie care nu face parte din categorii parametrilor de tip momente de ordinul r este prezentat în

Definiţia 3. Vom numi *mediană a v.a. X numărul a* determinat din condiţia că $P(X \leq a) \geq 1/2$ şi totodată $P(X \geq a) \geq 1/2$. Mediana se notează $Med(X)$.

Remarca 3. Spre deosebire de *modă* şi chiar spre deosebire de *valoarea medie,* mediana există întotdeauna şi, în caz discret, $Med(X)$ coincide cu acea valoare reală a pentru care distribuţia $\{P(X = x_i)\}_{i\geq 1}$ a v.a. X are proprietatea că

$$\sum_{i:x_i\leq a} P(X = x_i) \geq 1/2 \text{ şi } \sum_{i:x_i\geq a} P(X = x_i) \geq 1/2,$$

iar în caz (absolut) continuu, cu valoarea a pentru care d.d. $f(x)$ a v.a. X are proprietatea

$$\int_{-\infty}^{a} f(x)dx = 1/2, \int_{a}^{+\infty} f(x)dx = 1/2.$$

Mai mult, în caz (absolut) continuu soluţia este unică, dacă f.d.

$$F(x) = \int_{-\infty}^{x} f(u)du$$

este continuă şi strict monoton crescătoare. Această proprietate arată că valorile posibile ale v.a. X, poziţionate la stânga şi la dreapta de mediana ei, au aceeaş masă (pondere) probabilistă. Dealtfel, folosind interpretarea geometrică a integralei si faptul că în caz (absolut) continuu d.d. $f(x)$ a v.a. X posedă proprietatea

$$\int_{-\infty}^{+\infty} f(x)dx = 1,$$

rezulta ca aria figurii cuprinse între graficul d.d. $f(x)$ şi axa Ox de la -∞ până la $Med(X)$ coincide cu aria figurii cuprinse între graficul d.d. $f(x)$ şi axa Ox de la $Med(X)$ până la $+\infty$, ambele fiind egale cu 1/2. A se vedea, in acest sens, v.a. X din exemplul 9 de mai sus, exemplu pentru care $Med(X) = Mod(X) = 0$, graficul d.d $f(x) = \frac{1}{\sqrt{2\pi}}e^{-x^2/2}$ confirmând interpretarea noastră.

Exemplul 10 (*O v.a. X pentru care mediana nu este unică*). Considerăm v.a. X ce coincide cu numărul de puncte apărute la aruncarea unui zar "perfect" o singură dată. Atunci distribuţia v.a. X fiind $P(X =$

$i) = 1/6$, $i = \overline{1,6}$, rezultă că $Med(X)$ nu este unică deoarece inegalitile

$$\sum_{i:i\leq a}\frac{1}{6} \geq 1/2 \text{ şi } \sum_{i:i\geq a}\frac{1}{6} \geq 1/2$$

au loc pentru orice $3 \leq a \leq 4$.

Mediana reprezintă, la rândul ei, un caz particular a unei noţiuni mai generale numite *quantilă.*

Definiţia 4. Vom numi *quantilă de ordinul* α, pe scurt, $\alpha-$*quantilă a v.a.* X (sau $\alpha \cdot 100\%$ *procentilă a distributiei/d.d. a v.a.* X) acea valoare b pentru care $P(X \leq b) \geq \alpha$ şi totodată $P(X \geq b) \geq 1-\alpha$, unde $0 < \alpha < 1$.

Or, vedem că *mediana coincide* cu $0.5-$*quantila v.a.* X (50% *procentila distribuţiei/d.d. a v.a.* X). Ca şi mediana, $\alpha-$*quantila* întotdeauna există, dar nu pentru orice v.a. este unică. Excepţie fac v.a. de tip absolut continuu.

Probleme propuse.

1. Consideră în calitate de v.a. X numărul de steme apărute la aruncarea unei monede "imperfecte" pentru care probabilitatea apariţiei stemei este egala cu $p, 0 < p < 1$. Aflaţi valoarea ei medie.

2. Fie X o v.a. dată de distribuţia $P(X = k) = p(1-p)^{k-1}$, $k = 1, 2, 3, ...$, unde $0 < p < 1$. Dealtfel, această distribuţie descrie comportamentul probabilist al numarului de aruncari a unei monede "imperfecte" (pentru care probabilitatea apariţiei stemei este egala cu $p, 0 < p < 1$) până la prima apariţie a stemei. Aflaţi valoarea medie a v.a. X.

3. La Las Vegas, roata de ruletă are un 0 şi un 00 şi apoi numerele de la 1 până la 36 marcate pe sloturi egale; roata este rotită şi o bilă se opreşte la întâmplare într-unul din sloturi. Când un jucător pariază 1 dolar pe un anumit număr, el primeşte 36 de dolari dacă bila se opreşte la acest număr, având, astfel, un câştig net de 35 de dolari; in caz contrar, el îşi pierde pariul în dolari. Aflaţi valoarea medie pentru câştigul său.

4. Íntr-o a doua versiune de ruletă la Las Vegas, un jucător pariază pe roşu sau negru. Jumătate din numerele de la 1 la 36 sunt roşii, iar jumătate sunt negre. Dacă un jucător pariază un dolar pe negru şi dacă bila se opreşte pe un număr negru, atunci acesta îşi primeşte dolarul înapoi şi încă un dolar. Dacă bila se opreşte pe un număr roşu sau pe 0 sau 00, el îşi pierde dolarul. Găsiţi valoarea medie a câştigului jucătorului în această versiune a jocului.

5. În legătură cu aruncarea unui zar "perfect" de două ori succesiv considerăm v.a. bi-dimensională (X, Y), unde X reprezinta suma punctelor apărute

iar Y diferenţa lor (numărul de puncte apărute la primul zar minus numărul de puncte apărute la al doilea). Arătaţi că $\mathbb{E}XY = \mathbb{E}X\mathbb{E}Y$ şi că X, Y sunt dependente. Cu alte cuvinte, reciproca proprietăţii *e)* din Propoziţia 1 nu este valabilă.

6. O companie de asigurare de viaţă oferă unui bărbat în vârstă de 50 de ani o valoare nominală de 1000 USD, pentru o poliţă de asigurare de viaţă pe un an pentru o primă de 14 dolari. Tabelele de mortalitate standart indică faptul că probabilitatea ca un bărbat din această categorie de vârstă să moară în cursul anului este egală cu 0.006. Care este valoarea aşteptată (medie) a câştigului pentru fiecare poliţă vândută de către această companie de asigurări de viaţă?

7. Un mare producător auto efectuează sondaje trimestriale privind satisfacţia clienţilor care au achiziţionat automobile noi în ultimii trei ani. Proporţia celor care nu au dat raspuns într-un trimestru dat este privita, din punct de vedere matematic, ca fiind un rezultat al unei v.a. X cu densitatea de distributie $f(x) = 3x^2 I_{[0,1]}(x)$. Cu ce este egala, în medie, proporţia celor care nu au dat răspuns într-un trimestru dat? Dar mediana acestei proporţii, cu ce este egalaă?

8. Profitul unei firme este o v.a. de forma $\Pi(X) = pq(X) - rX$, unde X e o v.a. ale carei valori posibile reprezintă cantitatea zilnică de marfă (exprimata in kg), sub formă de produs agricol perisabil, furnizată firmei spre procesare, unde $p = 5$ reprezintă preţul produsului procesat per kg, $r = 2$ este preţul de cumparare per kg de către firmă a mărfii, iar funcţia $q(x) = x^{0.9}$ arată care este cantitatea de produs, rezultat în urma procesării unei cantităţi x de marfă. În presupunerea că X este o v.a. a carui comportament probabilist este guvernat de densitatea de distribuţie $f(x) = \frac{1+2x}{110} I_{[0,10]}(x)$, aflaţi valoarea medie a profitului zilnic.

9. O firmă producătoare de calculatoare vinde calculatoare cu o garanţie de 3 ani, ce constă în înlocuirea calculatorului cu unul nou dacă în acest rastimp iese din funcţiune memoria hard. Presupunem că durata vieţii memoriei hard, exprimată în ani, este o v.a. X cu densitatea de distribuţie $f(x) = 0.005e^{-0.005x} I_{[0,+\infty)}(x)$. *a)* Cu ce este egală probabilitatea că memoria hard a unui astfel de calculator va avea o dura de viaţa de cel mult 3 ani; *b)* Calculaţi durata medie a funţionalitătii memoriei hard pentru un astfel de calculator; *c)* Calculaţi *mediana* v.a. X; d) Este oare v.a. X *unimodală* sau *multimodală*?

10. Presupunem ca durata vieţii unui bec electric de tip Super Light este o v.a. X dată de densitatea de distribuţie $f(x) = \lambda^2 x e^{-\lambda x} I_{[0,+\infty)}(x)$ cu $\lambda =$

0.05. Aflaţi durata medie a vieţii a acestui tip de bec electric.

11. Dealerul unei firme de vânzare a autoturismelor este specializat în vânzarea a două mărci de maşini de teren: Range Rover si Dacia Duster. Numărul de autoturisme vândute săptămânal, corespunzător acestor marci, reprezintă o realizare a unei v.a. bidimensionale (X, Y) guvernată de următoarea distribuţie probabilista în ansamblu:

X / Y	0	1	2	3	4
0	0.20	0.15	0.075	0.05	0.03
1	0.10	0.075	0.04	0.03	0.02
2	0.05	0.03	0.02	0.01	0.01
3	0.04	0.03	0.02	0.01	0.01

La salariul sau de 100 USD săptămânal dealerul mai incasează prime a câte 200 USD pentru fiecare maşină Range Rover vândută şi 100 USD pentru fiecare maşină Dacia Duster vândută.

a) Sunt oare v.a. X şi Y independente?

b) Aflati valoarea medie a comisionului săptămânal incasat de dealer pentru vânzarea autoturismelor. Dar valoarea totală a plaţii incasate de dealer saptămânal pentru activitatea sa de vânzător auto cu ce este egală?

c) Aflati valoarea medie a comisionului săptămânal incasat de dealer pentru vânzarea autoturismelor Range Rover. Dar valoarea medie a comisionului săptămânal incasat de dealer pentru vânzarea autoturismelor Dacia Duster cu ce este egală?

3.2. Dispersia (varianţa), abaterea standard, covarianţa, coeficientul de corelaţie, regresie liniară

Vom începe cu un exemplu care demonstrează necesitatea întroducerii unui parametru sau a unei caracteristici numerice a v.a. X pentru a masura gradul de împrăştiere a valorilor ei individuale în jurul unui parametru de poziţie cum ar fi, de exemplu, valoarea medie a v.a. X.

Exemplul 1. Din câte se ştie, nu există aparate de măsurare absolut exacte, rezultatele masurărilor fiind influienţate de o mulţime de factori (temperatura, umiditatea mediului înconjurător, materialul din care este produs aparatul, cine este producătorul, etc.). Or, modelele matematice corespunzătoare sunt de natură probabilistă. Considerăm, de pildă, că v.a. X, ce reprezinta rezultatul unei masurări executate cu un aparat de măsurare, este dată de distribuţia

$$X : \begin{pmatrix} a-\varepsilon & a & a+\varepsilon \\ (1-p)/2 & p & (1-p)/2 \end{pmatrix}, 0 < p < 1,$$

a fiind valoarea exactă a mărimii măsurate iar $\varepsilon > 0$, eroarea de măsurare.

Valoarea ei medie a v.a. X este egală cu

$$\mathbb{E}X = (a-\varepsilon) \cdot \frac{1-p}{2} + a \cdot p + (a+\varepsilon) \cdot \frac{1-p}{2} = a.$$

Cu alte cuvinte, un astfel de aparat ne arată "în medie" valoarea exactă a mărimii măsurate, dar această constatare nu ne ajută cu nimic la caracterizarea calităţii aparatului nostru de măsurat. Noţiunea următoare vine să compenseze această lacună.

Definiţia 1. Vom numi *dispersie* sau *varianţă a v.a.* X numărul $\mathbb{D}X = \mathbb{E}(X - \mathbb{E}X)^2$, dacă această valoare medie există, bineînţeles.

Exemplul 1 (*Continuare*). În exemplul nostru pentru a calcula dispersia v.a. X putem folosi Formula de transport din p.3.1 şi faptul că $\mathbb{D}X = \mathbb{E}g(X)$, unde $g(x) = (x - \mathbb{E}X)^2 = (x-a)^2$. Prin urmare

$$\mathbb{D}X = (a-\varepsilon-a)^2 \cdot \frac{1-p}{2} + (a-a)^2 \cdot p + (a+\varepsilon-a)^2 \cdot \frac{1-p}{2} = \varepsilon^2 \cdot (1-p).$$

Distribuţia v.a. X arată că instrumentul corespunzător de masurare este cu atât mai bun cu cât valoarea erorii de măsurare ε este mai mică iar probabilitatea $p = P(X = a)$ este mai aproape de 1, dar aceasta are loc atunci si numai atunci când valoarea dispersiei este mai aproape de 0. Cu alte cuvinte, cu cât dispersia este mai mică, cu atât este mai mic gradul de împrăştiere a valorilor individuale a v.a. X faţă de valoarea ei medie $\mathbb{E}X = a$.

Pentru a simplifica, uneori, calculul dispersiei este utilă

Propoziţia 1. *Dispersia v.a. X poate fi calculată după formula*

$$\mathbb{D}X = \mathbb{E}X^2 - (\mathbb{E}X)^2.$$

Exemplul 2. Calculaţi dispersia v.a. X date de d.d. $f(x) = I_{[0,1]}(x)$, adică X fiind uniform distribuită pe $[0,1]$.

Soluţie. Deoarece,

$$\mathbb{E}X = \int_0^1 x dx = 1/2, \mathbb{E}X^2 = \int_0^1 x^2 dx = 1/3,$$

rezultă $\mathbb{D}X = \mathbb{E}X^2 - (\mathbb{E}X)^2 = 1/3 - (1/2)^2 = 1/12$.

Remarca 1.Din definiţia dispersiei v.a. X desprindem, că dispersia nu are aceeaşi unitate de masură ca şi v.a. X sau valoarea ei medie. Astfel, în exemplul 1, daca X se masoară în cm valoarea ei medie se masoară, la fel, în cm, dar dispersia ei se masoară în cm^2. Notiunea introdusă în definiţia următoare păstrează calităţile dispersiei ca masură a gradului de împrăştiere, dar este, în schimb, exprimată în aceeaşi unitate de măsură ca şi v.a. X.

Definiţia 2. Vom numi *abatere standard a v.a.* X numărul $\sigma = \sqrt{\mathbb{D}X}$.

La calcularea dispersiei, prin urmare si a abaterii standard, sunt utile proprietăţile dispersiei centralizate in

Propoziţia 2. *Dispersia posedă următoarele proprietăţi:*

a) $\mathbb{D}X \geq 0$, *iar* $\mathbb{D}X = 0$ *atunci şi numai atunci, când* $P(X = \mathbb{E}X) = 1$;

b) Dacă există dispersia $\mathbb{D}X$ *a v.a., atunci pentru orice* $a,b \in \mathbb{R}$ *există dispersia v.a.* $aX \pm b$ *şi* $\mathbb{D}(aX \pm b) = a^2\mathbb{D}X$;

c) Dacă există dispersiile $\mathbb{D}X$ şi $\mathbb{D}Y$ *ale v.a.* X *şi* Y, *atunci există dispersia v.a.* $X \pm Y$ *şi*

$$\mathbb{D}(X \pm Y) = \mathbb{D}X + \mathbb{D}Y \pm 2 \cdot \mathbb{E}(X - \mathbb{E}X)(Y - \mathbb{E}Y);$$

d) Dacă v.a. X *şi* Y *sunt independente şi există dispersiile lor* $\mathbb{D}X$ şi $\mathbb{D}Y$, *atunci există dispersia v.a.* $X \pm Y$ *şi*

$$\mathbb{D}(X \pm Y) = \mathbb{D}X + \mathbb{D}Y.$$

Definiţia 3. Vom numi *covarianţă a două v.a.* X *şi* Y *numărul* $Cov(X,Y) = \mathbb{E}(X - \mathbb{E}X)(Y - \mathbb{E}Y)$. Dacă această valoare medie există, atunci, din Formula de Transport rezultă ca, ea poate fi calculată dupa formula

$$Cov(X,Y) == \begin{cases} \sum\limits_{i\geq 1}\sum\limits_{j\geq 1}(x_i - \mathbb{E}X)(y_j - \mathbb{E}Y)P(X = x_i, Y = y_j), \textit{ in caz} \\ \textit{discret, } \{P(X = x_i, Y = y_j)\}_{i,j\geq 1} \textit{ fiind distrib. v.a. } (X,Y); \\ \int\limits_{-\infty}^{+\infty}\int\limits_{-\infty}^{+\infty}(x_i - \mathbb{E}X)(y_j - \mathbb{E}Y)f(x,y)dxdu, \textit{ in caz (absolut)} \\ \qquad \textit{continuu, } f(x,y) \textit{ fiind d.d. a v.a. } (X,Y). \end{cases}$$

Din definiţia covarianţei, dar şi din pp. c)-d) ale Propoziţiei 2, deducem următoarea

Consecinţă. *a) Dacă covarianţa v.a.* X *şi* Y *există, atunci aceasta poate fi calculată după formula*

$$Cov(X,Y) = \mathbb{E}XY - \mathbb{E}X\mathbb{E}Y;$$

b) Dacă v.a. X şi Y sunt independente atunci $Cov(X,Y)=0$.

Remarca 2. Reciproca afirmaţiei *b)* din Consecinţă nu are loc, aşa cum arată şi următorul

Contraexemplu (*Două v.a. de covarinţă nula, dar dependente*). Considerăm v.a. X,Y independente cu valorile medii $\mathbb{E}X{=}\mathbb{E}Y{=}0$ şi pentru care există $\mathbb{E}X^2$.

Atunci v.a. $Z=XY$, fiind dependentă de v.a. X, va avea $\mathbb{E}Z=\mathbb{E}XY=\mathbb{E}X\mathbb{E}Y=0$, dar aceste două v.a. X şi Z au, totuşi,

$$Cov(X,Z)=\mathbb{E}(X-\mathbb{E}X)(Z-\mathbb{E}Z)=\mathbb{E}XZ=\mathbb{E}X^2Y=\mathbb{E}X^2\mathbb{E}Y=\mathbb{E}X^2{\cdot}0=0.$$

Concluzie. Consecinţa formulată mai sus, împreună cu contraexemplul nostru, arată, de fapt, că *orice două v.a. X,Y care au $Cov(X,Y)\neq 0$ sunt dependente.*

Propoziţia 3 (*Inegalitatea Cauchy-Buniakovschi*). *Dacă există dispersiile $\mathbb{D}X$ şi $\mathbb{D}Y$ ale v.a. X şi Y, atunci există $Cov(X,Y)$ şi are loc inegalitatea*

$$|Cov(X,Y)|\leq\sqrt{\mathbb{D}X\mathbb{D}Y}.$$

Exemplul 1, p.2.6 (*Continuare*). Aşa cum am văzut, v.a. (X,Y), unde X reprezintă indicatorul parităţii sumei punctelor apărute iar Y - indicatorul parităţii produsului punctelor apărute la aruncarea unui zar "perfect" de două ori succesiv, are distribuţia (în ansamblu)

$X\setminus Y$	0	1
0	0	1/2
1	1/4	1/4

cu distribuţiile marginale

$$X:\begin{pmatrix}0 & 1\\ 1/2 & 1/2\end{pmatrix}, Y:\begin{pmatrix}0 & 1\\ 1/4 & 3/4\end{pmatrix}.$$

Atunci,

$$\mathbb{E}X=1/2,\mathbb{D}X=1/4,\sigma_X=1/2,\mathbb{E}Y=3/4,\mathbb{D}Y=3/16,\sigma_Y=\sqrt{3}/4,$$

iar

$$\begin{aligned}\mathbb{E}XY &= 0\cdot 0\cdot P(X=0,Y=0)+0\cdot 1\cdot P(X=0,Y=1)+\\ +1\cdot 0\cdot P(X &= 1,Y=0)+1\cdot 1\cdot P(X=1,Y=1)=1/4\end{aligned}$$

$$Cov(X,Y) = \mathbb{E}XY - \mathbb{E}X\mathbb{E}Y = 1/4 - (1/2)(3/4) = -1/8.$$

Inegalitatea Cauchy-Buniakovschi se verifică cu uşurinţă:

$$|Cov(X,Y)| = 1/8 = 0.125 \leq \sigma_X\sigma_Y = \sqrt{3}/8 = 0.216\,51.$$

În altă ordine de idei, la sfârşitul p.2.7, am arătat, reeşind din definiţia independenţei (dependenţei) v.a., ca v.a. X, Y din exemplul de mai sus sunt dependente, dar acelaşi lucru rezultă şi din faptul că $Cov(X,Y) \neq 0$. Într-adevar, presupunem contrariul, că X, Y sunt independente. Atunci, drept consecinţă, $Cov(X,Y) = 0$. Dar in exemplul nostru $Cov(X,Y) = -1/8$. Contradicţie, ce arata ca nenulitatea covarianţei a două v.a. poate fi considerată o condiţie suficientă, dar nu şi necesară (vezi contraexemplul de mai sus) ca acestea să fie dependente. Doar atat, nu putem, însă, spune nimic despre gradul lor de dependenţă sau asociere. În acest scop serveşte noţiunea din

Definiţia 4. Vom numi *coeficient de corelaţie a două v.a.* X *şi* Y numărul $\rho(X,Y) = \frac{Cov(X,Y)}{\sqrt{\mathbb{D}X\mathbb{D}Y}} = \frac{Cov(X,Y)}{\sigma_X\sigma_Y}$.

Aşa cum arată Inegalitatea Cauchy-Buniakovschi, coeficientul de corelaţie există, deîndată ce exista dispersiile $\mathbb{D}X$, $\mathbb{D}Y$. Mai mult, din aceeşi inegalitate şi consecinţa de mai sus rezultă

Propoziţia 4. *Coeficientul de corelaţie* $\rho(X,Y)$ *a două v.a.* X *şi* Y *are proprietăţile:*

a) Dacă v.a. X *şi* Y *sunt independente, atunci* $\rho(X,Y) = 0$;

b) $-1 \leq \rho(X,Y) \leq 1$.

În exemplul de mai sus $\rho(X,Y) = \frac{Cov(X,Y)}{\sqrt{\mathbb{D}X\mathbb{D}Y}} = \frac{-1/8}{\sqrt{3}/8} = -0.577\,35$. Despre ce fel de dependenţă şi în ce grad putem vorbi în baza valorii acestea? Pentru a putea raspunde la această întrebare luăm, pentru comparaţie, cazul dependenţei liniare: $Y = aX + b$, care se mai numeşte *regresie liniară*, unde $a, b \in \mathbb{R}$, $a \neq 0$. Cum $\mathbb{D}Y = a^2\ \mathbb{D}X$,

$$\begin{aligned} Cov(X,Y) &= \mathbb{E}XY - \mathbb{E}X\mathbb{E}Y = \mathbb{E}X(aX+b) - \\ -\mathbb{E}X(a\mathbb{E}X+b) &= a[\mathbb{E}X^2 - (\mathbb{E}X)^2] = a\mathbb{D}X, \end{aligned}$$

rezultă că

$$\rho(X,Y) = \frac{Cov(X,Y)}{\sqrt{\mathbb{D}X\mathbb{D}Y}} = \frac{a\mathbb{D}X}{\sqrt{a^2(\mathbb{D}X)^2}} = \frac{a}{|a|} = \begin{cases} 1,\ dacă\ a > 0, \\ -1,\ dacă\ a < 0. \end{cases}$$

Aşadar, dacă v.a. Y depinde liniar de v.a. X, atunci $|\rho(X,Y)| = 1$. Important e că are loc şi reciproca acestei afirmaţii. Ma exact are loc

Propoziţia 5. *Coeficientul de corelaţie* $|\rho(X,Y)| = 1$, *atunci şi numai atunci când există* $a, b \in \mathbb{R}$, $a \neq 0$, *astfel încât probabilitatea* $P(Y = aX+b) = 1$.

Remarca 3. Dacă $\rho(X,Y) = 1$, atunci *regresia liniară este pozitivă*, deoarece în acest caz coeficientul $a > 0$, ceea ce atrage dupa sine *creşterea lui* Y odată cu *cresterea lui* X şi invers, dacă $\rho(X,Y) = -1$, atunci *regresia liniară este negativă*, deoarece în acest caz coeficientul $a < 0$, ceea ce atrage dupa sine *descreşterea lui* Y odată cu *cresterea lui* X.

În concluzie, dacă revenim la exemplul nostru de mai sus, faptul ca $\rho(X,Y) = -0.577\,35$ arată, că regresia dintre Y şi X este *neliniară.*

Are loc următoarea

Teoremă *(Predicţia liniară cea mai bună pentru valoarea posibilă a v.a.* Y*). Dacă v.a.* X şi Y *posedă, cel puţin, momentele iniţiale de ordinul* 2, *atunci există două valori* $a, b \in \mathbb{R}^2$, $a = cov(X,Y)/\mathbb{D}X$, $b = \mathbb{E}Y - a\mathbb{E}X$, *pentru care valoarea medie a patratului distanţei dintre* Y *şi* $\widehat{Y} = aX + b$, *fiind egală cu* $\mathbb{E}(Y - \widehat{Y})^2 = \mathbb{E}[Y - (aX+b)]^2$, *ia valoarea cea mai mică.*

Exemplu 3. Preţul X în *Euro per litru* al combustibilului şi al cantităţii Y de combustibil exprimat in mii de tone vândute de catre o companie de carburanţi intr-o zi pot fi modelate, din punct de vedere matematic ca fiind o v.a. (X,Y) guvernată de o legitate probabilistă dată de densitatea de distribuţie

$$f(x,y) = 2xe^{-xy} I_{[0.5,1]}(x) I_{(0,+\infty)}(y).$$

a) Determinaţi predicţia liniară cea mai buna pentru cantitatea Y în funcţie de preţul X;

b) Cu ce va fi egală cantitatea medie de combustibil vândut daca preţul combustibilului este dat de valoarea $x = 0.95$ *Euro*?

c) Care va fi suma medie în euro a incasării zilnice pentru combustibilul vândut?

Soluţie. Calculăm, mai întâi, valorile necesare pentru a determina coeficienţii a, b care definesc predicţia liniară cea mai buna.

$$\mathbb{E}X = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} x f(x,y)\,dxdy = \int_{0.5}^{1}\int_{0}^{+\infty} x \cdot 2xe^{-xy}\,dxdy =$$

$$\int_{0.5}^{1} 2x^2(\frac{-1}{x}e^{-xy}\mid_0^{+\infty})dx = \int_{0.5}^{1} 2xdx = x^2\mid_{0.5}^{1}= 1 - 0.25 = 0.75.$$

$$\mathbb{E}X^2 = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} x^2 f(x,y)dxdy = \int_{0.5}^{1}\int_{0}^{+\infty} x^2 \cdot 2xe^{-xy}dxdy = 0.583333.$$

$$\mathbb{D}X = \mathbb{E}X^2 - (\mathbb{E}X)^2 = 0.583333 - 0.75^2 \simeq 0.021.$$

$$\mathbb{E}Y = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} yf(x,y)dxdy = \int_{0.5}^{1}\int_{0}^{+\infty} y \cdot 2xe^{-xy}dxdy =$$

$$\int_{0}^{+\infty} 2y\left(\int_{0.5}^{1} xe^{-xy}dx\right)dy = \int_{0}^{+\infty} 2y(1 - e^{-xy})\mid_{x=0.5}^{1} dy \simeq 1.39$$

$$\mathbb{E}XY = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} xyf(x,y)dxdy = \int_{0.5}^{1}\int_{0}^{+\infty} xy \cdot 2xe^{-xy}dxdy = 1$$

$$Cov(X,Y) = \mathbb{E}XY - \mathbb{E}X\mathbb{E}Y \simeq 1 - 0.75 \cdot 1.39 = -0.0425$$

Prin urmare:

a) Din Teorema anterioară deducem ca $a{=}cov(X,Y)/\mathbb{D}X{=}{-}0.0425/0.021{=}$ -2.0238, iar $b{=}\mathbb{E}Y - a\mathbb{E}X \simeq 1.39 - ((-0.0425)/0.021) \times 0.75{=}2.9079$. Prin urmare cea mai buna predictie liniară pentru v.a. Y în funcţie de X este v.a. $\widehat{Y} = aX + b = -2.0238X + 2.9079$.

b) Atunci când preţul $X = 0.95$ $Euro/litru$ predicţia cantităţii zilnice de combustibil vândut $\widehat{Y} = -2.0238 \times 0.95 + 2.9079 = 0.98529$ *mii tone* de combustibil;

c) Suma medie în euro a incasării zilnice pentru combustibilul vândut conform predicţiei va fi egala cu

$$\mathbb{E}(X \times (1000)^2\widehat{Y}) = (1000)^2\mathbb{E}X\widehat{Y} = (1000)^2\mathbb{E}(2.9079X - 2.0238X^2) =$$

$$(1000)^2 \times (2.9079\mathbb{E}X - 2.0238\mathbb{E}X^2) =$$

$$(1000)^2 \times (2.9079 \times 0.75 - 2.0238 \times 0.583333) = 1.0004 \times 10^6 Euro$$

Probleme propuse.

1. În condiţiile problemelor $1-4$, $6-10$ din lista de Probleme propuse in paragraful anterior, calculaţi şi dispersia v.a. abordate la calcularea valorii medii.

2. Un joc de noroc este considerat ca fiind unul "echitabil" sau "corect" daca valoarea medie a sumei câştigate de către jucător este nulă. Sa se verifice dacă următorul joc este "echitabil" si să se calculeze dispersia sumei câştigate. Jocul constă în următoarele: jucătorul aruncă o perche de zaruri "perfecte", plătind suma de N unităţi monetare (u.m.). Dacă suma punctelor apărute este egală cu 7 sau 11, atunci jucătorului i se returnează suma egală cu N u.m., plus o suma în valoare de $2N$ u.m., în caz contrar jucătorul pierde suma introdusă.

3. Fie X un număr ales la întâmplare diin mulţimea de numere $\{-1, 0, 1\}$. Calculaţi valorea medie, dispersia şi abaterea standart a v.a. X.

4. Considerăm v.a. X pentru care valorea medie $\mathbb{E}X = 100$ şi abaterea standart $\sigma = 15$. Aflaţi cu ce sunt egale valorile: a) $\mathbb{E}X^{\ 2}$; b) $\mathbb{E}(3X + 10)$; c) $\mathbb{E}(-X)$; d) $\mathbb{D}X$; e) $\mathbb{D}(3X + 10)$; f) $\mathbb{D}(-3X - 10)$; g) $\mathbb{D}(3X - 10)$.

5. Un sondaj statistic organizat de către guvern interoga populaţia privind proiectul de construire a unei staţii electrice atomice, avînd la bază un eşantion ce includea 2400 de persoane alese la întâmplare. Drept rezultat au fost înregistrate 40% de raspunsuri favorabile, restul fiind nefavorabile. Aflaţi dispersia si abaterea standart pentru numarul S_{2400} ce coincide cu numărul persoanelor incluse în eşantion care s-au exprimat in favoarea construirii unei staţii atomice.

5. Este dată o v.a. X cu valorea medie $\mathbb{E}X = \mu$ şi abaterea standart $\sqrt{\mathbb{D}X} = \sigma$. Construim o nouă v.a. $Y = (X - \mu)/\sigma$ numită *normare* sau *standartizare* a v.a. X. Arătaţi că $\mathbb{E}Y = 0$ iar $\mathbb{D}Y = 1$.

6. Considerăm două v.a. X, Y independente, identic distribuite cu d.d. $f(x) = I_{[0,1]}(x)$. Introducem v.a. $Z = X + Y$ şi $W = X - Y$. Calculaţi coeficienţii de corelaţie: a) $\rho(X, Y)$; b) $\rho(X, Z)$; c) $\rho(Y, W)$; d) $\rho(Z, W)$.

7. Preţul estimativ X exprimat in USD şi cantitatea totala Y de pixuri de o anumită marcă realizată într-o reţea de magazine de specialitate, având ca unitate de măsura 10000 de pixuri, pe durata unei anumite perioade de vânzare, reprezintă o v.a. bidimensională (X, Y) cu d.d. $f(x, y) = 10xe^{-ps} I_{[0.20,20]}(x) I_{[0,+\infty]}(y)$.

a) Determinaţi cea mai buna predicţie liniară $\widehat{Y}$ pentru cantitatea Y în funcţie de preţul X;

b) Cu ce va fi egală cantitatea medie de pixuri vândute daca preţul unui pix este dat de valoarea $x = 0.12$ USD?

c) Cu ce va fi egală valoarea medie în USD a incasării totale pentru vanzările de pixuri pentru perioada dată, i.e. valoarea $\mathbb{E}(10000X\widehat{Y})$ USD.

3.3. Momente ale variabilei aleatoare (iniţiale, centrale), asimetria, boltirea (aplatizarea)

Atunci când sunt puse în discuţie astfel de aspecte ca tendinţa centrală, gradul de îmrăştiere a unei v.a. X , forma distribuţiei sau d.d. a acestei v.a., dar si în cadrul aplicării unor proceduri statistice, sunt utile nu numai valoarea medie si dispersia, dar si valorile medii ale v.a. X^r sau $(X-\mathbb{E}X)^r$, calculate şi pentru valori ale lui r diferite de 1 sau de 2. Toate aceste valori medii poartă denumirea de *momente,* cu ele întâlnindu-ne în exemplul 5, p.3.1. După cum vom vedea, deosebim *momente iniţiale* şi *momente centrale* conform cu

Definiţia 1. Vom numi *moment iniţial de ordinul $r \geq 0$ a v.a. X* numărul $\alpha_r = \mathbb{E}X^r$, dacă această valoare medie există, bineînţeles.

Definiţia 2. Vom numi *moment central de ordinul $r \geq 0$ a v.a. X* numărul $\mu_r = \mathbb{E}(X-\mathbb{E}X)^r$, dacă această valoare medie există, bineînţeles.

Observăm ca $\alpha_0 = \mu_0 = 1$, iar $\alpha_1 = \mathbb{E}X$, $\mu_1 = 0$, $\mu_2 = \mathbb{D}X$. În plus, atunci când α_r, μ_r există, aplicarea Formulei de transport conduce la următoarele formule de calcul direct, care în cazul discret, adică în cazul când v.a. este dată de distribuţia $\{P(X = x_i)\}_{i\geq 1}$, sunt formulele

$$\alpha_r = \sum_{i\geq 1} x_i^r P(X = x_i),\ \mu_r = \sum_{i\geq 1}(x_i - \mathbb{E}X)^r P(X = x_i),$$

iar în cazul (absolut) continuu, adică în cazul când v.a. este dată de d.d. $f(x)$, sunt formulele

$$\alpha_r = \int_{-\infty}^{+\infty} x^r f(x)dx,\ \mu_r = \int_{-\infty}^{+\infty}(x-\mathbb{E}X)^r f(x)dx.$$

Înloc de formula directă, la calcularea momentelor centrală, devine utilă formula ce exprimă legătura dintre momentele iniţiale si cele centrale din

Propoziţia 1 (*Momentele centrale ca funcţie de momentele iniţiale*). *Dacă pentru valoarea $r \in \{1, 2, ..., n, ...\}$ există momentul central μ_r de ordinul r al v.a. X , atunci*

$$\mu_r = \sum_{i=0}^{r} (-1)^i \mathbb{C}_r^i \alpha_i \alpha_{r-i}.$$

Privitor la existenţa momentelor este utilă şi

Propoziţia 2. *Dacă, pentru $r > 0$, există momentul iniţial $\mathbb{E}X^r$, respectiv, momentul central $\mathbb{E}(X - EX)^r$, atunci pentru orice $s \in [0, r]$ există momentul iniţial $\mathbb{E}X^s$, respectiv, momentul central $\mathbb{E}(X - EX)^s$.*

Exemplul 1. Presupunem că pentru v.a. X există momentul central de ordinul $r = 4$. Atunci, din propoziţiile 1,2 rezultă că exista si momentele iniţiale şi centrale de ordinele $1, 2, 3$ şi

$$\begin{aligned}
\mu_1 &= \alpha_1, \\
\mu_2 &= \alpha_2 - \alpha_1^2, \\
\mu_3 &= \alpha_3 - 3\alpha_1\alpha_2 + 2\alpha_1^3, \\
\mu_4 &= \alpha_4 - 4\alpha_1\alpha_3 + 6\alpha_1^2\alpha_2 - 3\alpha_1^4.
\end{aligned}$$

Aceste din urmă formule sunt cele care se aplică cel mai des în Statistică, mai ales în Statistica Descriptivă, pentru a calcula, în baza datelor statistice cei doi parametri ce descriu forma distributiei sau d.d. a v.a. (caracteristicii statistice) X: coeficientul de Asimetrie si coeficientul fe Aplatizare (Boltire).

Definiţia 3. Vom numi *coeficient de asimetrie a distribuţiei/d.d. a v.a.* X numărul $As(X) = \mu_3/\sigma^3$, unde μ_3 este momentul central de ordinul 3, iar σ abaterea standard a v.a. X, dacă μ_3 există, bineînţeles. Vom spune că *distribuţia/d.d. v.a. X este: simetrica* dacă $As(X) = 0$ sau $(\mu_3 = 0)$, *positiv asimetrică* dacă $As(X) > 0$ sau $(\mu_3 > 0)$, *negativ asimetrică* dacă $As(X) < 0$ sau $(\mu_3 < 0)$.

Exemplul 2 (*O v.a. cu d.d. simetrică*). Astfel, v.a. X (standard normal distribuită) cu d.d. $f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$ este simetrică. A se vedea gfaficul acestei d.d. în exemplul 8, p.3.1. Mai mult, functiiile $\frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$, $\frac{1}{\sqrt{2\pi}} x^3 e^{-\frac{x^2}{2}}$ fiind funcţii impare pe domeniul $(-\infty, +\infty)$, rezultă că

$$\mathbb{E}X = \int_{-\infty}^{+\infty} x \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx = 0, \mu_3 = \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}} x^3 e^{-\frac{x^2}{2}} dx = 0,$$

ceea ce înseamnă ca $As(X) = 0$. Apropo, în acest exemplu $\mathbb{E}X = Med(X) = Mod(X) = 0$.

Pentru caracterizarea formei distribuţii/d.d. a unei v.a., înafară de coeficientul de asimetrie, se mai foloseşte si coeficientul de Boltire (Aplatizare).

Definiţia 4. Vom numi *coeficient de boltire (aplatizare) a distribuţiei/d.d. a v.a.* X numărul $Ap(X) = \mu_4/\sigma^4$, unde μ_4 este momentul central de ordinul 4, iar σ abaterea standard a v.a. X, dacă μ_4 există, bineînţeles.

Exemplul 2 (*Continuare*). Pentru v.a. X (*standard normal distribuită*), folosind formula integrării prin părţi, găsim că momentele centrale de ordinele 2 şi 4 este sunt egale, respectiv, cu

$$\mu_2 = \mathbb{D}X = \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}} x^2 e^{-\frac{x^2}{2}} dx = 1,\ \mu_4 = \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}} x^4 e^{-\frac{x^2}{2}} dx = 3.$$

Prin urmare abaterea standard a v.a. X (acum putem spune, a v.a. *standard normal distribuită cu media* 0 *şi dispersia* 1) este egală cu 1, iar coeficientul de aplatizare $Ap(X) = \mu_4/\sigma^4 = 3$.

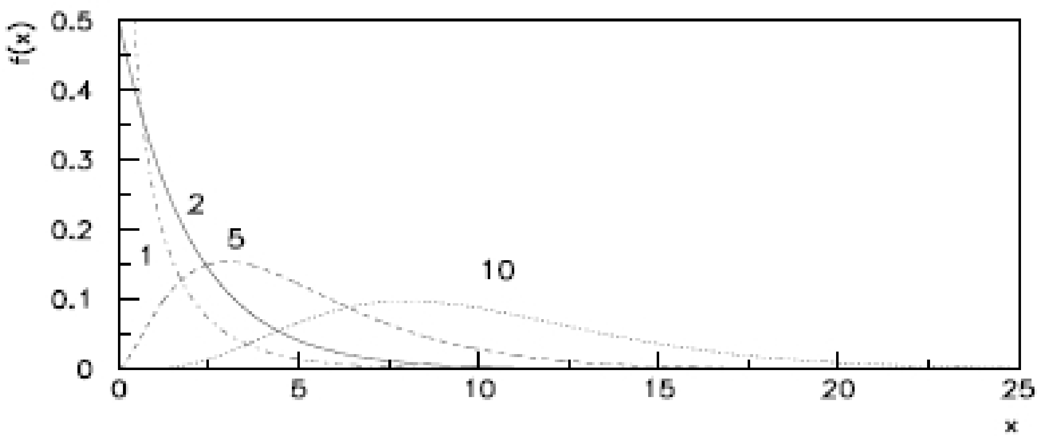
Remarca 1. Comparând definiţia coeficientului de boltirere si faptul, ca pentru distribuţia normală cu media 0 şi dispersia 1, acesta este egal cu 3,deacum înainte gradul de boltire a distribuţiei/d.d. se va compara cu gradul de boltire a d.d. a unei v.a. standard normal distribuite, spunând ca *boltirea unei d.d. e negativă*, dacă $\mu_4/\sigma^4 < 3$, adică *boltirea ei e mai mică (graficul ei este mai plat) decât boltirea d.d. a v.a. standard normal distribuite cu media* 0 *şi dispersia* 1,ori că *boltirea unei d.d. e pozitivă*, dacă $\mu_4/\sigma^4 > 3$, adică *boltirea ei e mai mare (graficul ei este mai ascuţit) decât boltirea d.d. a v.a. standard normal distribuite cu media* 0 *şi dispersia* 1.

Exemplul 3 (*V.a. cu d.d. pozitiv asimetrică în diverse variante de boltire (aplatizare)*. Considerăm v.a. X cu d.d.

$$f(x) = \frac{x^{(n-2)/2} exp(-x/2)}{2^{\frac{n}{2}} \Gamma(\frac{n}{2})} \cdot I_{[0,+\infty)}(x),$$

unde $\Gamma(u) = \int_0^{+\infty} x^{u-1} e^{-x} dx$ este funcţia gamma.

Modelul acesta, adică d.d., coincide cu d.d. a v.a. $X_1^2 + X_2^2 + ... + X_n^2$, unde $X_1, X_2, ..., X_n$ sunt variabile aleatoare, independente, identic distribuite (v.a.i.i.d.) normal cu media 0 şi dispersia 1, d.d. ce poartă denumirea de *distribuţie Hi-patrat cu n grade de libertate* . Conform cărţii Forbes C., Evans M., et alls, *Statistical Distributions*, Ed. John Willey&Sons, USA, 2010, $\mathbb{E}X = n$, $\mathbb{D}X = 2n$, $Med(X) \simeq n - \frac{2}{3}$, $Mod(X) = n - 2$, pentru $n \geq 2$, coeficientul de asimerie $As(X)$=$\frac{2^{3/2}}{\sqrt{n}}$, coeficientul de aplatizare $Ap(X)$ =$3 + 12/n$. Iata cum arata pe unul si acelaşi grafic d.d. Hi-patrat pentru valorile parametrului ν=1, 2, 5, 10.



Or, din grafice se vede cum arată *d.d. pozitiv asimetrică.* Cum $0 < As(X)$=$\frac{2^{3/2}}{\sqrt{n}} \underset{n\to\infty}{\longrightarrow} 0$, $Ap(X)$ =$3+12/n \underset{n\to\infty}{\longrightarrow} 3$, se vede cum graficul d.d. tinde să devină simetric, gradul de aplatizare apropiindu-se de cel al d.d. normale. Cu alte cuvinte, odată cu creşterea numărului gradelor de libertate ν d.d. Hi-patrat se aproape de d.d. normală, lucru folosit pe larg in Statistica Matematică.

In opoziţie cu *asimetria pozitiva*, *devine clar cum arata grafic o distribuţie/d.d. negativ asimetrică.*

Remarca 2. Exemplele aduse mai scot in evidenţă urmatoarele situatii:

a) daca distribuţia/d.d. este simetrică ($As(X) = 0$), *atunci* $Mod(X) = Med(X) = \mathbb{E}X$;

b) daca distribuţia/d.d. este pozitiv asimetrică ($As(X) > 0$),
atunci $Mod(X) < Med(X) < \mathbb{E}X$;

c) daca distribuţia/d.d. este negativ asimetrică ($As(X) < 0$),
atunci $\mathbb{E}X < Med(X) < Mod(X)$.

Probleme propuse.

1. Pentru urmatoarele v.a. date de distribuţia probabilistă sau d.d. determinaţi dacă exista valoarea medie, dispersia. În plus, aflaţi modul, mediana şi stabiliţi care distribuţie este simetrică:

a) $P(X = x) = \mathbb{C}_4^x (0.2)^x (0.8)^{4-x} I_{\{0,1,2,3,4\}}(x)$;

b) $f(x) = 3x^2 I_{[0,1]}(x)$;

c) $f(x) = 2x^{-3} I_{[1,+\infty)}(x)$;

d) $f(x) = [\pi(1 + x^2)]^{-1} I_{(-\infty,+\infty)}(x)$.

2. Primele trei momente iniţiale ale v.a. X sunt egale cu $\alpha_1 = 0.5$, $\alpha_{12} = 0.5$, $\alpha_3 = 0.75$.

a) Determinaţi primele trei momente centrale μ_i, $i = \overline{1,3}$, ale v.a X;

b) Stabiliţi dacă distributia/d.d. este antisimetrica. De ce sau de ce nu?

4. Modele (distribuţii, d.d.)) probabiliste uzuale, inegalităţi, Legea Numerelor Mari, Teorema Limită Centrală

4.1. Distribuţii probabiliste uzuale in caz discret (Uniformă, Bernoulli, Binomială, Geometrică, Poisson, Multinomială, Hypergeometrică)

Cum distribuţia în caz discret sau d.d. a v.a. în caz (absolut) continuu reprezintă, de fapt, adevărate modele matematice ale unor variabile aleatoare ce prezintă interes din punct de vedere practic noi vom scoate în evidenţă câteva dintre cele mai răspândite modele probabiliste.

Distribuţia Uniformă (caz discret).

Definiţia 1. Vom spune că *v.a. X este distribuită uniform pe mulţimea de valori posibile* $\{0,1, 2, ..., N\}$, pe scurt $X \sim U\{0,1, 2, ..., N\}$, dacă probabilităţile

$$P(X = k) = \frac{1}{N + 1},\ k = 0, 1, ..., N,$$

numărul N având rol de parametru, unde N este un număr întreg *nenegativ.*

Se arată ca dacă $X \sim U\{0,1, 2, ..., N\}$, atunci $\mathbb{E}X$=$N/2$, $\mathbb{D}X$=$(N^2 - 1)/12$, μ_3=0.

Analogic, vom spune că $X \sim U\{1, 2, ..., N\}$, dacă $P(X = k) = \frac{1}{N}$, $k = 1$, ..., N sau $X \sim U\{a_1, a_2, ..., a_N\}$, $P(X = a_k) = \frac{1}{N}$, $k = 1, ..., N$, unde a_1, a_2, ..., $a_N \in \mathbb{R}$, $a_1 < a_2 < ... < a_N$, iar parametrul N este un număr întreg *pozitiv.*

Remarca 1. Din punct de vedere matematic distribuţia uniformă modelează, de exemplu, alegerea la întâmplare a unui element din multimea de N elemente diferite. Experimentele aleatoare de acest gen pot fi simulate şi pe calculator deoarece orice limbaj de programare evoluat (C^{++}, Java, etc.) are in biblioteca sa de functii program funcţia *random*, corespunzătoare cazului discret, accesarea căreia are drept rezultat generarea unui număr ales la întâmplare din mulţimea $\{0,1, 2, ..., N\}$. Acest model este utilizat şi în Sondajele Statistice, atunci când se urmăreşte ca valoarea aleasă şi inclusă în eşantion dintr-o Populaţie Statistică să fie aleasă aleator.

Exemplul 1. Presupunem ca un PC consta din N blocuri, calculatorul ieşind din funcţiune deîndată ce iese din funcţiune unul din blocuri, ieşirea simultană din funcţiune a două sau mai multe calculatoare fiind exclusă. Depanatorul de calculatoare verifică, luând la intâmplare, unul după altul câte un bloc, pană când va depista blocul defectat. Cu ce este egală probabilitatea că depanatorul va depista blocul defectat la incercarea cu numărul de ordine k, $k = 1, 2, ..., N$.

Soluţie. Întroducem evenimentele aleatoare $A_k = \{$*depanatorul va depista blocul defectat la incercarea cu numărul de ordine* $k\}$, $k = 1, 2, ..., N$. Considerăm v.a. $X =$*numărul de ordine al încercării la care va fi depistat blocul defect*, unde $X \in \{1, 2, ..., N\}$. Atunci,

$P(X{=}1){=}P(A_1){=}\frac{1}{N}$,

$P(X{=}2){=}P(\overline{A_1}\ A_2){=}P(\overline{A_1})P(\overline{A_1}\ /A_2){=}\frac{N-1}{N}\frac{1}{N-1}{=}\frac{1}{N}$,

...,

$P(X{=}N){=}P(\overline{A_1}\,\overline{A_2}...\overline{A_{N-1}}A_{N)}){=}P(\overline{A_1})P(\overline{A_2}\ /\ \overline{A_1})P(\overline{A_3}\ /\ \overline{A_1A_2})...P(A_N$ $/\ \overline{A_1A_2...A_{N-1}}) = \frac{N-1}{N}\frac{N-2}{N-1}\frac{N-3}{N-2}...\frac{1}{N-(N-1)}{=}\frac{1}{N}$.

În concluzie, v.a. $X \sim U\{1, 2, ..., N\}$.

Distribuţia Bernoulli.

Definiţia 2. Vom spune că *v.a.* X *este distribuită Bernoulli cu parametrul* p, $0 \le p \le 1$, pe scurt $X \sim Bernoulli(p)$, dacă *mulţimea ei de valori posibile* este $\{0,1\}$, *iar probabilităţile*

$$P(X = 0) = 1 - p,\ P(X = 1) = p.$$

Caracteristicele numerice de bază ale acestei v.a. sunt $\mathbb{E}X{=}p$, $\mathbb{D}X{=}p(1-p)$, $\mu_3{=}2p^3 - 3p^2 + p$.

Remarca 2. Din punct de vedere matematic distribuţia Bernoulli modelează comportamentul probabilist al numărului de "succese" într-o singură probă Bernoulli, aceasta din urmă noţiune fiind introdusă în

Definiţia 3. Vom numi *probă Bernoulli* orice experiment aleator ce posedă proprietăţile:

1. Spaţiul de evenimente elementare $\Omega = \{"succes", "insucces"\}$;

2. Probabilitatea $p = P\{"succes"\}$, $0 \leq p \leq 1$, nu variază de la probă la probă;

3. Rezultatele unei probe nu influienţează rezultatele celorlalte probe (independenţa probelor).

Drept exemplu de probă Bernoulli poate servi aruncarea unei monede o singură dată, *"succes"* considerând $\{aparitia\ stemei\}$. În genere, orice experiment aleator poate fi considerat probă Bernoulli deîndată ce acesta poate fi repetat de fiecare dată independent de celelalte repetări, iar "succes" fiind considerat orice eveniment A asociat experimentului, unde $p{=}P\{"succes"\}{=}P(A)$. Observăm, astfel, că numărul X de "succese" într-o singură probă Bernoulli are distribuţia Bernoulli.Dealtfel, obsevăm că pentru $p = \frac{1}{2}$ distribuţia $Bernoulli(p)$ coincide cu distribuţia $U\{01\}$.

Exemplul 2. Pe raftul unui magazin de calculatoare sunt expuse spre vânzare 100 de laptopuri de acelaşi model, din care 3 laptopuri au defecte a scunse, dar care nimeresc sub incidenţa garanţiei. Un client cumpără, alegând la întâmplare, unul din laptopuri. Codificam rezultatul alegerii cumpărătorului prin X, unde $X = 0$ dacă laptopul va fi returnat în urma depistării unui defect ascuns şi $X = 1$ în caz contrar. Atunci v.a. $X \sim Bernoulli(p)$, unde parametrul $p{=}P\{laptopul\ cumpărat\ nu\ va\ avea\ defecte\ ascunse\}{=}0.97$.

Distribuţia Binomială.

Definiţia 4. Vom spune că *v.a. X este distribuită Binomial cu parametrii n, p,* $n \in \{1, 2, ...\}$, $0 \leq p \leq 1$, pe scurt $X \sim Bi(n;p)$, dacă *mulţimea ei de valori posibile* este $\{0,1, 2, ..., n\}$, *iar probabilităţile*

$$P_n(k) = P(X = k) = \mathbb{C}_n^k\, p^k(1-p)^{n-k},\ k = 0, 1, 2, ..., n.$$

Caracteristicele numerice de bază ale acestei v.a. sunt $\mathbb{E}X{=}np$, $\mathbb{D}X{=}np(1-p)$, $\mu_3{=}np(1-p)(1-2p)$.

Pentru $n = 1$ avem că $Bi(1; p) \equiv Bernoulli(p)$.

Remarca 3. Din punct de vedere matematic distribuţia binomială modelează comportamentul probabilist al numarului total de "succese" în n probe Bernoulli deoarece este valabilă

Propoziţia 1. *Numărul total de "succese" în n probe Bernoulli cu probabilitatea "succesului" p în fiecare probă este o v.a.* $X \sim Bi(n; p)$.

Exemplul 3. În baza înregistrărilor de la casele de marcaj a centrului comercial Mall-Dova s-a constatat ca doar 30% din cumpărători îşi achită cumpărăturile cu cardul de credit, restul sub formă de cash. Cu ce este egală probabilitatea că din următorii 5 cumparatori 3 îşi vor achita cumpărăturile cu cardul de credit.

Soluţie. Modul de plata al fiecarui cumpărător poate fi interpretat ca rezultat al unei probe Bernoulli: plata cu cardul="succes", plata cash=

"insucces", unde probabilitatea $p = P\{"succes"\} = 30/100 = 0.3$.Prin urmare numarul platitorilor cu cardul din numărul total de $n = 5$ cumpărători este o v.a. $X \sim Bi(5; 0.3)$. Deci $P\{$*din următorii* 5 *cumparatori* 3 *îşi vor achita cumpărăturile cu cardul de credit*$\}$=$P(X = 3)$=$\mathbb{C}_5^3(0.3)^3(1 - 0.3)^{5-3}$ =0.1323.

Distribuţia Geometrică.

Definiţia 5. Vom spune că *v.a. X este distribuită Geometric cu parametrul p*, $0 < p < 1$, pe scurt $X \sim Geom(p)$, dacă *mulţimea ei de valori posibile este* $\{0, 1, 2, ..., n, ...\}$, *iar probabilităţile*

$$p_k = P(X = k) = p(1 - p)^k,\ k = 0, 1, 2, ...$$

sau dacă *mulţimea ei de valori posibile este* $\{1, 2, ..., n, ...\}$, *iar probabilităţile*

$$P(X = k) = p(1 - p)^{k-1},\ k = 1, 2, ...,$$

ultima fiind numită şi *distribuţie geometrică trunchiată în zero.*

Dealtfel, dacă v.a. $X \sim Geom(p)$ trunchiată în zero, atunci v.a. $X - 1 \sim Geom(p)$.

Caracteristicele numerice de bază ale v.a. $X \sim Geom(p)$ trunchiată în zero sunt $\mathbb{E}X$=$1/p$, $\mathbb{D}X$=$(1 - p)/p^2$, μ_3=$(1 - p)(2 - p)/p^3$.

Distribuţia geometrică (trunchiată în zero) este unica v.a. de tip discret ce posedă proprietatea remarcabilă enunţată în

Propoziţia 2 (*Proprietatea "lipsei memoriei" a distribuţiei geometrice*). *V.a. X de tip discret, pentru care mulţimea de valori posibile este* *multimea numerelor naturale, este distribuită geometric trunchiată în zero, adică*

$$P(X=k)=p(1-p)^{k-1},\ k=1,2,...,$$

dacă şi numai daca aceasta posedă proprietatea "lipsei memoriei":

$$P(X=n+k\ /\ X>n)=P(X=k\ ),\ \textit{pentru orice}\ n,k\geq 1.$$

Consecinţă. *Dacă v.a.* $X\sim Geom(p)$ *trunchiată în zero, atunci*

$$P(X>n+k\ /\ X>n)=P(X>k\ ),\ \textit{pentru orice}\ n,k\geq 1.$$

Remarca 4. Din punct de vedere matematic distribuţia Geometrică apare în contextul experimentului aleator ce constă în repetarea unei probe Bernoulli cu probabilitatea "succesului" p până la prima apariţie a "succesului" aşa cum se vede din

Propoziţia 3. ***Numărul total de probe Bernoulli****, cu probabilitatea "succesului" p în fiecare probă, efectuate până la prima apariţie a "succesului" este o v.a.* $X\sim Geom(p)$ *trunchiată în zero, iar* ***numărul total de "insuccese"*** *înregistrate în acelaşi experiment este o v.a.* $Y\sim Geom(p)$ *, unde* $0<p<1$.

Exemplul 4 (*Continuare la exemplul 3*). În condiţiile enunţate anterior, ne interesează cu ce este egală probabilitatea că de la începutul următoarei zile de lucru a centrului comercial până la prima apariţie a unui cumpărător plătitor cu cardul de credit va fi inregistrat un număr mai mare decât 10 cumparători ce-şi vor face cumpărături? Dar probabilitatea că, ştiind că primii, cel puţin, 10 cumpători şi-au achitat cumpăturile cash, după aceasta, pâna la prima apariţie a unui cumpărător plătitor cu cardul de credit, cel puţin, încă 10 cumpărători îşi vor achita cumpărăturile cash?

Soluţie. De această dată, conform Propoziţiei 3, numărul de cumpărători înregistraţi până la apariţia primului cumparător platitor cu cardul este o v.a. $X\sim Geom(0.3)$ trunchiată în zero, iar raspunsul la prima întrebare se reduce la calcularea

$$P(X>10)=P(X=11)+P(X=12)+...=0.3(0.7)^{10}+0.3(0.7)^{11}+...=$$

$$0.3\cdot 0.7^{10}[1+0.7+0.7^2+...]=0.3\cdot 0.7^{10}\cdot\frac{1}{1-0.7}=0.02\,824\,8.$$

Răspunsul la cea de a doua întrebare vizează calcularea probabilităţii condiţionate $P(X>10+10\ /\ X>10)$=$P(X>20\ /\ X>10)$. Dar, conform consecinţei de mai sus $P(X>20\ /\ X>10)$=$P(X>10)$=0.02 824 8.

Distribuţia Poisson.

Definiţia 6. Vom spune că *v.a.* X *este distribuită Poisson cu parametrul* λ, $\lambda > 0$, pe scurt $X \sim Poisson(\lambda)$, dacă *mulţimea ei de valori posibile este* $\{0,1, 2, ..., n, ...\}$, *iar probabilităţile*

$$p_k = P(X = k) = \frac{\lambda^k}{k!}e^{-\lambda},\ k = 0, 1, 2, ...$$

Caracteristicele numerice de bază ale v.a. $X \sim Poisson(\lambda)$ $\mathbb{E}X$=$\mathbb{D}X$=μ_3=λ. Dacă λ este număr întreg, atunci p_k îşi atinge valoarea maximă pentru $k_0 = \lambda$ şi $k_0 = \lambda + 1$. Dacă λ este fracţionar, atunci $Mod(X)$=$[\lambda] + 1$.

Un context aparte în care apare distribuţia Poisson este legat de noţiunea dată de

Definiţia 7. Vo numi *flux de evenimente* un şir de evenimente aleatoare, care se produc în momente aleatoare de timp. Un flux de evenimente se numeşte *flux Poisson* dacă el are proprietăţile:

a) este staţionar, adică probabilitatea că într-un anume interval de timp se vor realiza exact k evenimente depinde numai de numărul k şi de lungimea (durata) intervalului de timp şi nu depinde de începutul lui;

b) probabilitatea realizării a k evenimente într-un anume interval de timp nu depinde de numărul de evenimente care s-au realizat înainte de începerea acestui interval;

c) realizarea a două sau mai multe evenimente într-un interval mic de timp are, practic, probabilitate nulă.

Numărul mediu de evenimente dintr-un flux Poisson care se realizează într-o unitate de timp se numeşte *intensitate a fluxului.* Vom nota intensitatea fluxului cu λ. Atunci are loc

Propoziţia 4. *Numărul total de evenimente produse pe un interval de timp de durată* t *într-un flux Poisson de intensitate* λ, $\lambda > 0$, *este o v.a.* $X(t)$ *distribuită Poisson cu parametrul*

$$p_k(t) = P(X(t) = k) = \frac{(\lambda t)^k}{k!}e^{-\lambda t},\ k = 0, 1, 2, ...$$

Remarca 5. Mulţimea de v.a. $\{X(t), t \geq 0\}$, unde $X(t)$ reprezintă *numărul total de evenimente produse pe un interval de timp de durată* t *într-un flux Poisson de intensitate* λ, $\lambda > 0$ este un proces aleator ce numeşte *proces Poisson cu parametrul* λ.

Drept exemple de v.a. $X(t)$ ce apar în contextul fluxului Poisson putem, cu o anumită doză de exactitate, considera:

1. Numărului de cutremure de pământ care au loc într-o regiune seismică într-un interval de timp;
2. Numărului de accidente rutiere produse într-un oraş, într-un interval de timp;
3. Numărului de decese printre asiguraţii unei companii de asigurare într-un interval de timp;
4. Numărului de particule (alfa) emise de o substanţă radioactivă intr-un anumit interval de timp;
5. Numărului de automobile care vin la o staţie de alimentare cu benzină într-un interval de timp;
6. Numărului de clienţi care se adresează la un oficiu poştal într-o zi;
7. Numărului de apeluri la un post telefonic într-un interval de timp. etc., etc.

Remarca 6. Din punct de vedere matematic distribuţia Poisson mai modelează, de exemplu, comportamentul probabilistic ale unor astfel de v.a. ca:

1. Numărului de erori de programare comise de un programator intr-un soft de o anumită lungime;
2. Numărului de bacterii descoperite intr-o picătură de apă;
3. Numărului de erori de tipar care se conţin pe o pagină (sau un grup de pagini) dintr-o carte;
4. Numărului de 3 gemeni noi născuţi în decurs de un an într-o ţară anume;
5. Numărului de oameni dintr-o anumită ţară care au depăşit vârsta de 100 de ani, etc.

Printre exemplele enumerate sunt şi exemple ce nimeresc sub incidenţa distribuţiei Binomiale, dar cate poate fi aproximată cu distribuţia Poisson conform

Propoziţia 5 (*Teorema Limită Poisson*). *Dacă v.a.* $X \sim Bi(n;p)$ *şi* $np \underset{n\to+\infty}{\longrightarrow} \lambda$, $\lambda > 0$, *deîndată ce* $p \longrightarrow 0$, *atunci probablităţile*

$$P_n(k) = P(X=k) = \mathbb{C}_n^k\, p^k(1-p)^{n-k} \underset{n\to+\infty}{\longrightarrow} \frac{(\lambda)^k}{k!}e^{-\lambda},\ k = 0, 1, 2, \ldots .$$

Exemplul 5. Conform normelor tehnologice la 1000 de cozonaci cu stafide sunt prevăzute 10000 de stafide. Cu ce este egală probabilitatea ca într-un cozonac ales la întâmplare nu vom descoperi nicio stafidă? Dar care este numărul cel mai probabil de stafide care vor nimeri în cozonacul ales?

Soluţie. Dacă sunt respectate normele tehnologice, atunci pentru fiecare din cele 10000 srafide probabilitatea de a nimeri în cozonacul ales ("succes") este egală cu $p = 1/1000$, astfel, avem de a face cu $n = 10000$ de probe Bernoulli cu probabilitatea succesului 1/1000 în fiecare probă. prin urmare Numărul stafidelor numerite în cozonacul ales este o v.a. $X \sim Bi(10000; 1/1000)$. Cum n este suficient de mare, iar p suficient de mic, putem aplica propozitia 5, considerând că $\lambda = np = 10000 \cdot 1/1000 = 10$. Prin urmare probabilitatea, de exemplu, că într-un cozonac ales la întâmplare nu vom descoperi nicio stafidă este egală cu

$$P_{10000}(0) = P(X = 0) = \mathbb{C}^0_{10000}(1/1000)^0(1 - 1/1000)^{10000-0} =$$

$$= (999/1000)^{10000} \simeq \frac{\lambda^k}{k!}e^{-\lambda} = \frac{10^0}{0!}e^{-10} = 4.54 \times 10^{-5} \simeq 0.$$

Or, dacă în urma unui control de rutină, se constată că într-un cozonac ales la întâmplare nu a fost descoperită nicio stafidă, atunci ipoteza cea mai verosimilă e că... au fost încălcate normele tehnologice anunţate mai sus. În plus, deoarece λ=10, acesta fiind un numar intreg, rezultă că pentru $k_0 = \lambda$ şi $k_0 = \lambda + 1$, probabilităţile $P(X = 10)$=$P(X = 11)$ au valoarea cea mai mai mare, egală cu $\frac{10^{10}}{10!}e^{-10} = 1.251\,1 \times 10^{-11}$.

Distribuţia Multinomială.

Este vorba despre distribuţia unei v.a. $n-$dimensionale, $n \geq 2$, dar care pentru $n = 2$ coincide cu distribuţia $Bi(n; , p)$.

Definiţia 8. Vom spune că v.a. X=(X_1, X_2, ..., X_n) este *multinomial distribuită cu parametrii m şi* p_1, p_2, ..., $p_n > 0$, p_1+ p_2+ ...+ p_n=1, pe scurt $X \sim Multi(m; p_1, p_2, ..., p_n)$, daca mulţimea de valori posibile a vectorului X este $\mathcal{X} = \{(m_1, m_2, ..., m_n) : m_i = \overline{0, m},\ i = \overline{1, n},\ m_1 + m_2 + ... + m_n = m\}$, iar

$$p(m_1, m_2, ..., m_n) = P(X_1 = m_1, X_2 = m_n, ..., X_n = m_n) =$$

$$= \frac{m!}{m_1!m_2!...m_n!}p_1^{m_1}p_2^{m_2}...p_n^{m_n}$$

pentru orice $m_i = \overline{0, m}$, $i = \overline{1, n}$, $m_1 + m_2 + ... + m_n = m$.

Caracteristicile numerice marginale pentru fiecare v.a. X_i, $i=\overline{1,n}$, sunt $\mathbb{E}X_i=mp_i$, $\mathbb{D}X_i=mp_i(1-p_i)$, $\mu_{3,i}=mp_i(1-p_i)(1-2p_i)$, analogia cu cazul distribuţiei binomiale fiind concludentă.

Remarca 7. Este util să ştim că fiecare v.a. în parte $X_i \sim Bi(m;p_i)$, $i=\overline{1,n}$. Mai mult, pentru orice subset de indici $1\leq i_1<i_2<...<i_k$, $1\leq k\leq n$ vectorul aleator $(X_{i_1}, X_{i_2}, ..., X_{i_k}) \sim Multi(m;p_{i_1}, p_{i_2}, ..., p_{i_k})$. Din punct de vedere matematic distribuţia multinomiala descrie toate situaţiile conforme cu

Propoziţia 6. *Fie X_i numarul de înregistrări a evenimentului A_i $i=\overline{1,n}$, în urma a m repetări independente a experimentului aleator $\mathcal{E}$ în care se poate produce de fiecare dată doar unul din evenimentele A_i, unde $\overset{n}{\underset{i=1}{\cup}} A_i=\Omega$, $A_i\cap A_j=\varnothing$, pentru orice $i\neq j$, $i,j=\overline{1,n}$, iar probabilităţile $p_i=P(A_i)>0$, $p_1+p_2+\ldots+p_n=1$ sunt cunoscute. Atunci $X=(X_1,\ X_2,\ ...,\ X_n)$ este o v.a. n-dimensionala şi $X\sim Multi(m;p_1,\ p_2,\ ...,\ p_n)$.*

Exemplul 6. Presupunem ca în finala campionatului mondial la şah au ajuns doi şahişti, din care primul este campionul mondial *en titre*, despre care se mai ştie că, în baza partidelor de şah jucate anterior între aceştia, 20% din partide le-a câştigat primul, 20% al doilea şahist, iar restul partidelor s-au terminat la egalitate. Meciul pentru titlul de campion consta din 12 partide, care, in caz de, de acumulare unui număr egal de puncte la sfârşit campionul en titre îşi păstrează titlul. Cu ce ste egală probabilitatea că primul jucător va câştiga 4 partide, al doilea 5, resul partidelor terminându-se la egalitate? Dar probabilitatea că cel de al doilea jucător va deveni campion mondial?

Soluţie. Putem considera fiecare partidă de şah un experiment aleator $\mathcal{E}$ *în care se poate produce de fiecare dată doar unul din evenimentele* $A_1=\{$*partida dată va fi câştigată de primul jucător*$\}$, $A_2=\{$*partida dată va fi câştigată de al doilea jucător*$\}$,$A_1=\{$*partida dată se va termina la egalitate*$\}$, *unde* $\overset{3}{\underset{i=1}{\cup}} A_i=\Omega$, $A_i\cap A_j=\varnothing$, *pentru orice* $i\neq j$, $i,j=\overline{1,3}$, *iar probabilităţile* $p_1=P(A_1)=p_2=P(A_2)=0.2$, $p_2=P(A_2)=0.6$. Atunci, fiind aplicabilă propozitia 6, deducem: probabilitatea că primul jucător va câştiga 4 partide, al doilea 5, resul partidelor terminându-se la egalitate este egală cu

$$p(4,5,3)=\frac{12!}{4!5!3!}0.2^4 0.2^4 0.6^3=0.015328.$$

Ambii jucători, având aceeaşi putere de joc, pentru a-şi păstra titlul de campion mondial, primul jucător este avantajat, totuşi, de un meci terminat la egalitate. Dar probabilitatea ca meciul se va termina la egalitate este egală cu

$p(6,6,0)+p(5,5,2)+p(4,4,4)+p(3,3,6)+p(2,2,8)+p(1,1,10)+p(0,0,12) =$

$$\frac{12!}{6!6!0!}0.2^6 0.2^6 0.6^0+\frac{12!}{5!5!2!}0.2^5 0.2^5 0.6^2+\frac{12!}{4!4!4!}0.2^4 0.2^4 0.6^4+\frac{12!}{3!3!6!}0.2^3 0.2^3 0.6^6+$$
$$\frac{12!}{2!2!8!}0.2^2 0.2^2 0.6^8+\frac{12!}{1!1!10!}0.2^1 0.2^1 0.6^{10}+\frac{12!}{0!0!12!}0.2^0 0.2^0 0.6^{12}=0.18121.$$

Or, probabilitatea ca cel de al doilea jucator va deveni noul campion mondial este egala cu $(1-0.18121)/2 = 0.409\,40$, fapt ce i-a determinat pe organizatorii campionatelor mondiale la şah să renunţe la condiţia enunţată în problemă, înlocuind-o cu condiţia că in caz că cele 12 partide nu vor desemna un câştigător net, atunci meciul continuă pâna la prima partidă câştigată de unul din jucători, acesta şi fiind declarat învingător.

Distribuţia Hipergeometrică.

Definiţia 9. Vom spune că v.a. X este *hipergeometric distribuită cu parametrii* N, M, n, pe scurt $X \sim Hypergeom(N,M,n)$, unde (N,M,n): N=1, 2,...; M=0, 1,..., N; n=1, 2, ..., N, daca mulţimea de valori posibile a v.a. X este $\mathcal{X} = \{k : \min[0, n-(N-M)] \le k \le \min(n, M)\}$, iar

$$P_{N,M,n}(k) = P(X=k) = \frac{\mathbb{C}_M^k \mathbb{C}_{N-M}^{n-k}}{\mathbb{C}_N^n},$$

$\min[0, n-(N-M)] \le k \le \min(n, M)$.

Caracteristicile numerice de bază ale v.a. X sunt: $\mathbb{E}X = n\frac{M}{N}$, $\mathbb{D}X = n\frac{M}{N} \cdot \frac{N-M}{N} \cdot \frac{N-n}{N-1}$, $\mu_3 = n\frac{M}{N} \cdot \frac{NM}{N} \cdot \frac{N-2M}{N} \cdot \frac{N-n}{N-1} \cdot \frac{N-2n}{N-2}$.

Remarca 8. Din punct de vedere matematic distribuţia hipergeometrică descrie extragerea la întãmplare, fără întoarcere/repetare a n elemente dintr-o mulţime de N elemente, din care M elemente sunt de tipul 1, restul $N-M$ elemente fiind de tipul 2. Atunci, folosind definiţia clasică a probabilităţii, se arată că numărul de elemente de tipul 1 care se vor regăsi printre cele n elemente extrase este o v.a. $X \sim Hypergeom(N,M,n)$, $(N,M,n) \in \{(N,M,n)$: N=1, 2,...; M=0, 1,..., N; n=1, 2, ..., $M\}$. Menţionăm că modul de extragere, fară întoarcere sau cu întoarcere, este principial. Astfel, dacă extragerea succesivă a n bile este cu întoarcere, atunci, folosind definiţia clasică a probabilităţii, se demonstrează că numărul de elemente de tipul 1 care se vor regăsi printre cele n elemente extrase este o v.a. $X \sim Bi(n;p)$, unde $p = P\{\text{"}succes\text{"}\} = P\{elementul\ extras\ la\ extragerea\ i\ va\ fi\ unul\ de\ tipul\ 1\} = M/N$, pentru orice $i=\overline{1,n}$. Aceasta este esential, de exemplu, la modelarea matematica a eşantionării aleatoare în funcţie de tipul de selectare/extragere (cu întoarcere sau fără întoarcere) a elementelor din populaţia statistică.

Exemplul 7 (*Loto 6 din 49*). În România se bucură de popularitate printre amatorii de jocuri de noroc Loteria Naţională 6 din 49. Interes prezintă probabilitatea că, jucând cu o singură variantă, vom incasa câştigul maxim, adică vom ghici $X = 6$ numere câştigătoare. Deoarece ne aflăm în situaţia aplicării distribuţiei $X \sim Hypergeom(N, M, n)$, unde N=49, M=6, n=6, deducem ca

$$P(X = 6) = P_{49,6,6}(6) = \frac{\mathbb{C}_6^{6k}\mathbb{C}_{49-6}^{6-6}}{\mathbb{C}_{49}^6} = \frac{1}{\mathbb{C}_{49}^6} = \frac{1}{13\,983\,816}.$$

Distribuţia Hipergeometrică Multivariată.

Distribuţia Hipergeometrică Multivariată generalizează distribuţia Hipergeometrică definită anterior.

Definiţia 10. Vom spune că v.a. $X = (X_1, X_2, ..., X_k)$ este *hipergeometric multivariată distribuită cu parametrii* $(N, M_1, M_2, ..., M_k, n)$, pe scurt $X \sim Hypergeom(N, M_1, M_2, ..., M_k, n)$, unde $N \in \{1, 2, ...\}$, $M_i \in \{0, 1, ..., N\}$, $i=\overline{1,k}$, $\sum_{i=1}^{k} M_i=N$, $n \in \{1, 2, ..., N\}$, daca mulţimea de valori posibile a v.a. X este $\mathcal{X} = \{(x_1, x_2, ..., x_k) : x_i \in \{0, 1, ..., N\}, i=\overline{1,k}, \sum_{i=1}^{k} x_i=N\}$, iar

$$P_{N,M_1,M_2,...,M_k,n}(x_1, x_2, ..., x_k) =$$

$$= P(X_1 = x_1, X_2 = x_2, ..., X_k = x_k) = \frac{\prod_{i=1}^{k}\mathbb{C}_{M_i}^{x_i}}{\mathbb{C}_N^n}.$$

Caracteristicile numerice marginale pentru fiecare v.a. X_i, $i=\overline{1,k}$, sunt $\mathbb{E}X=n\frac{M_i}{N}$, $\mathbb{D}X=n\frac{M_i}{N} \cdot \frac{N-M_i}{N} \cdot \frac{N-n}{N-1}$, $\mu_3=n\frac{M_i}{N} \cdot \frac{NM_i}{N} \cdot \frac{N-2M_i}{N} \cdot \frac{N-n}{N-1} \cdot \frac{N-2n}{N-2}$.

Remarca 9. Este util să ştim că fiecare v.a. în parte $X_i \sim Hypergeom(N, M_i, n)$, $i=\overline{1,k}$. Mai mult, pentru orice subset de indici $1 \leq i_1<i_2<...<i_j$, $1 \leq j \leq k$ vectorul aleator $(X_{i_1}, X_{i_2}, ..., X_{i_j}) \sim Hypergeom(N, M_{i_1}, M_{i_2}, ..., M_{i_j}, n)$.

Remarca 10. Din punct de vedere matematic distribuţia hipergeometrică Multivariată descrie extragerea la întãmplare, *fără întoarcere/repetare* a n elemente dintr-o mulţime de N elemente, din care M_i elemente sunt de tipul i, $i{=}\overline{1,k}$, $\sum\limits_{i=1}^{k} M_i{=}N$. Atunci, folosind definiţia clasică a probabilităţii, se arată că numărul de elemente x_i de tipul i care se vor regăsi printre cele n elemente extrase, $i{=}\overline{1,k}$, $\sum\limits_{i=1}^{k} x_i{=}N$ este o v.a. k-dimensionala $X \sim Hypergeom(N,\, M_1,\, M_2,\, ...,\, M_k,\, n)$.

Exemplul 8. Dintr-o cutie în care se află 30 de mingi de tenis, din care 5 de culoare roşie, 10 de culoare galbenă şi 15 de culoare albastră au fost extrase la întâmplare, fără întoarcere 9 mingi. Calculaţi probabilitatea că printre mingile extrase se vor afla câte 3 mingi de fiecare culoare.

Soluţie. În conformitate cu ultima Remarcă această probabilitate este egală cu

$$P_{30,5,10,15:,9}(3,3,3) = \frac{\mathbb{C}_5^3 \cdot \mathbb{C}_{10}^3 \cdot \mathbb{C}_{15}^3}{\mathbb{C}_{30}^9} \simeq 0.0381627$$

Probleme propuse.

Acolo unde e cazul, în problemele urmatoare sa se foloseasca, pentru calcule numerice, facilităţile site-ului *wolframalpha.com*.

1. Presupunem că probabilitatea statistică ca un copil nou născut va fi băiat este egala cu 0.51. Se cere: 1) să se determine distribuţia v.a. X care reprezintă numărul de băieţi printre 10000 de copii noi născuţi; 2) să se calculeze probabilitatea ca printre 10000 de copii noi născuţi numărul băieţilor va fi cuprims între 4500 şi 5500.

2. Un ţintaş trage in tinta pana cand va nimeri in ea, probabilitatea de a nimeri fiind egala cu 0.1. Se cere: 1) să se determine distribuţia v.a. X care reprezintă numărul total de trageri in ţinta; 2) să se calculeze probabilitatea ca v.a. X va depasi valoarea 15 .

3. Consideram in calitate de v.a. X numarul de bile rosii inregistrate printre 10 bile extrase la intamplare, una cate una, cu intoarcere, dintr-o cutie cu 2 bile albe si 3 bile rosii .Se cere: 1) să se determine distribuţia v.a. X ; 2) să se calculeze probabilitatea ca v.a. X va depasi valoarea 1 .

4. Consideram in calitate de v.a. X numarul de bile rosii inregistrate printre 10 bile extrase la intamplare, una cate una, fara intoarcere, dintr-o cutie cu 200 bile albe si 300 bile rosii .Se cere: 1) să se determine repartiţia v.a. X ; 2) să se calculeze probabilitatea ca v.a. X va depasi valoarea 1 .

5. Presupunem ca avem o sală de laborator in care avem 30 de calculatoare, din care 5 sunt nefuncţionale, în care fiecare student, al unei grupe de 20 de studenţi, ocupă la înâmplare câte un calculator. Considerăm in calitate de v.a. X numarul studentilor care vor ocupa calculatoare nefuncţionale. Se cere: 1) să se determine distribuţia v.a. X ; 2) să se calculeze probabilitatea ca v.a. X va varia inte valorile 1 si 4.

6. Considerăm aruncarea unui zar "perfect" de 10000 de ori succesiv, in calitate de v.a. X luând numarul de câte ori apare faţa 6. Se cere: 1) să se determine distribuţia v.a. X ; 2) să se calculeze probabilitatea ca v.a. X va varia intre valoarile 100 si 300.

7. Considerăm in calitate de v.a. X numărul de aruncări nereuşite ale unui zar "perfect" pana la prima apariţie a unui număr de puncte multiplu lui 3. Se cere: 1) să se determine distribuţia v.a. X ; 2) să se calculeze probabilitatea ca v.a. X va varia inte valorile 100 si 300.

8. Considerăm în calitate de v.a. X numărul de peştişori roşii înregistraţi printre cei 100 de peştişori scoşi la intâmplare, fara întoarcere, dintr-un acvariu enorm cu 1000 de peştişori, din care doar 10 sunt de culoare roşie. Se cere: 1) să se determine distribuţia v.a. X ; 2) să se calculeze probabilitatea ca v.a. X va depasi valoarea 2.

9. Consideram in calitate de v.a. X numarul de PC-uri de acelasi tip supuse unui control, unul cate unul, controlul terminându-se in momentul depistării unui calculator defect. Ştiind ca un PC ales la intâmplare va fi defect este egala cu $p = 0.05$, se cere: 1) să se determine distribuţia v.a. X ; 2) să se calculeze probabilitatea că v.a. X va depaşi valoarea 1.

10. Considerăm in calitate de v.a. X numarul de numere ghicite, jucând cu o singura variantă la "LOTOSPORT 5 din 35". Scrieţi distribuţia v.a. X.Care este probabilitatea că, jucând cu o singură variantă la LOTOSPORT "5 din 35", nu vom fi în pierdere, adica vom caştiga ceva?

11. Considerăm in calitate de v.a. X numărul format in felul următor: dintr-o cutie cu 9 bile numerotate de la 1 la 9 sunt extrase la intâmplare, succesiv, cu întoarcere, 2 bile, formând astfel un numar din doua cifre, prima cifra find numarul primei bile, iar cea de a doua cifra, fiind numarului celei de a doua bile extrase. 1) Determinati repartitia v.a. X; 2) Calculati probabilitatea că numarul X nu va intrece valoarea 50.

12. Un dispozitiv electronic constă din 1000 de elemente, astfel incât fiecare element poate ieşi din funcţiune, independent unul de celelalte, pe o durată de timp T, cu una si aceeaşi probabilitate $p = 5 \cdot 10^{-4}$. Notam prin X numarul total de elemente ieşite din funcţiune pe durata de timp T. 1)

Aflati repartitia v.a. X; 2) Calculati probabilitatea ca pe durata T va iesi din functiune, cel putin, un element.

13. Consideram in calitate de v.a. X numărul format in felul următor: dintr-o cutie cu 9 bile numerotate de la 1 la 9 sunt extrase la intâmplare, succesiv, fară întoarcere, 2 bile, formând astfel un număr din două cifre, prima cifră find numărul primei bile, iar cea de a doua cifră, fiind numărului celei de a doua bile extrase. 1) Determinaţi distribuţia v.a. X; 2) Calculati probabilitatea ca numărul X nu va intrece valoarea 90.

14. Considerăm in calitate de v.a. X numărul total de aruncări a două zaruri "perfecte " până când, prima data, *suma* punctelor apărute va fi un număr multiplu lui 3. 1) Aflat distribuţia v.a. X; 2) Calculaţi probabilitatea că valoarea lui X va depăşi 1000.

15. Considerăm in calitate de v.a. X numarul total de "succese" la aruncarea a doua zaruri "perfecte " de 5000 de ori, considerând "succes" la fiecare aruncare faptul ca *produsul* punctelor apărute va fi un numar multiplu lui 3. 1) Aflat distribuţia v.a. X; 2) Calculati probabilitatea că valoarea lui X va depăşi 1000.

16. Considerăm in calitate de v.a. X numărul total de aruncări a două zaruri "perfecte " până când, prima dată, *minimul* punctelor apărute va fi un număr multiplu lui 3. 1) Aflaţi distribuţia v.a. X; 2) Calculati probabilitatea că valoarea lui X va depăşi 1000.

17. Considerăm in calitate de v.a. X numărul total de "succese" la aruncarea a două zaruri "perfecte " de 5000 de ori, considerand "succes" la fiecare aruncare faptul că *maximul* punctelor apărute va fi un număr multiplu lui 3. 1) Aflat distribuţia v.a. X; 2) Calculati probabilitatea că valoarea lui X va depăsi 1000.

18. Considerăm în calitate de v.a. X numărul total de elevi care au fost depistaţi ca fiind infectaţi TBC, dintr-un număr total de 1000 de elevi supuşi controluli medical , dacă se ştie că de această infecţie este afectată 0.1% din populaţie. 1) Aflat repartitia v.a. X; 2) Calculaţi probabilitatea că valoarea lui X va depaşi 100.

19. Presupunem ca un an va fi secetos, independent de anii precedenţi, cu probabilitatea 0.05. Notam prin X numărul de ani nesecetoşi inregistraţi pana la înregistrarea pentru prima dată a unui an secetos. 1) Aflat distribuţia v.a. X; 2) Calculati probabilitatea ca valoarea lui Xva depasi 100.

20. O companie de asigurări are incheiate 10000 de poliţe de asigurare anuală de tip RCA. Considerăm in calitate de v.a. X numărul total de despăgubiri pe care le va plăti compania pe parcursul unui an. Ştiind că

probabilitatea unei avarii auto este egală cu $p = 0.001$: 1) Aflaţi distribuţia v.a. X; 2) Calculaţi probabilitatea că valoarea lui X va depăşi 100.

21. Numărul X de particule alfa emise de un gram de substanţă radioactivă într-o secundă este o v.a. cu distribuţia Poisson cu parametrul a, unde a este numărul mediu de particule alfa emise într-o secundă. 1) Să se determine distribuţia v.a. X. 2) Să se calculeze probabilităţile evenimentelor: $A = \{$*într-o secundă vor fi emise nu mai mult de două particule alfa*$\}$ şi $B = \{$*într-o secundă vor fi emise cinci particule alfa*$\}$, $C = \{$*într-o secundă vor fi emise mai mult de zece particule alfa*$\}$. Care este numărul de particule alfa care corespunde celei mai mari probabilităţi? Să se considere că $a = 0,25$.

22. Numărul X de apeluri telefonice înregistrate la Urgenţa Medicală în 24 de ore este o v.a. cu distribuţia Poisson cu parametrul a, unde a este numărul mediu de apeluri telefonice înregistrate în 24 de ore. 1) Să se determine distribuţia v.a. X. 2) Să se calculeze probabilităţile evenimentelor: $A = \{$*în* 24 *de ore fi inregistrate cel mult* 100 *de apeluri telefonice*$\}$ şi $B = \{$*în* 24 *de ore vor fi inregistrate exact* 100 *de apeluri telefonice*$\}$, $C = \{$*în* 24 *de ore vor fi inregistrate mai mult de* 1000 *de apeluri telefonice*$\}$. Aflati numărul de apeluri telefonice înregistrate in 24 de ore, care corespunde celei mai mari probabilităţi? Să se considere că $a = 100$.

23. Numărul X de pene de curent înregistrate la un Centru de Calcul pe parcursul unui an este o v.a. cu distribuţia Poisson cu parametrul a, unde a este numărul mediu de pene de curent electric înregistrate pe parcursul uni an. 1) Să se determine distribuţia v.a. X. 2) Să se calculeze probabilităţile evenimentelor: $A = \{$*într-un an vor fi inregistrate cel mult 10 pene de curent*$\}$ şi $B = \{$*într-un an vor fi inregistrate exact 10 pene de curent*$\}$, $C = \{$*într-un an vor fi inregistrate mai mult de 100 pene de curent*$\}$. Aflaţi numărul de pene de curent înregistrate într-un an, care corespunde celei mai mari probabilităţi? Să se considere că $a = 10$.

24. Numărul X de erori de programare comise de un programator la elaborarea unui soft de proporţii este o v.a. cu distribuţia Poisson cu parametrul a, unde a este numărul mediu de erori de programare comise de programator la elaborarea unui astfel de soft. 1) Să se determine distribuţia v.a. X. 2) Să se calculeze probabilităţile evenimentelor: $A = \{$ *vor fi comise cel mult 10 erori de programare*$\}$ şi $B = \{$*vor fi comise exact* 10 *erori de programare*$\}$, $C = \{$*numarul de erori* de programare comise *va varia intre* 100 *si* 500$\}$. Aflati numărul de erori de programare comise la elaborarea unui soft de proportii, care corespunde celei mai mari probabilităţi? Să se considere că $a = 50$.

25. Numărul X de bacterii descoperite într-o picatură de apă luată dintr-un lac este o v.a. cu distribuţia Poisson cu parametrul a, unde a este numărul mediu de bacterii descoperite intr-o picatură de apa. 1) Să se determine distribuţia v.a. X. 2) Să se calculeze probabilităţile evenimentelor: $A = \{$ *vor fi descoperite cel mult 10 bacterii*$\}$ şi $B = \{$*vor fi descoperite exact* 10 *bacterii*$\}$, $C = \{$*numarul de bacterii va varia intre* 500 *si* 5000$\}$. Aflaţi numărul de erori de programare comise la elaborarea unui soft de proporţii, care corespunde celei mai mari probabilităţi? Să se considere că $a = 1000$.

26. Numărul de accidente aviatice inregistrate anual la o anumită companie aviatică este o v.a. cu distribuţia Poisson cu parametrul a , unde a este numărul mediu de accidente aviatice înregistrate anual.1) Să se determine repartiţia v.a. X. 2) Să se calculeze probabilităţile evenimentelor: $A = \{$ *intr-un an vor fi înregistrate cel mult* 5 *accidente*$\}$ şi $B = \{$*într-un an vor fi înregistrate exact* 5 *accidente*$\}$, $C = \{$*numarul de accidente înregistrate într-un an va varia între* 10 *si* 20 *accidente*$\}$. Aflati numărul de acciente înregistrate intr-un an, care corespunde celei mai mari probabilităţi? Să se considere că $a = 0.1$.

27. Numărul de stafide descoperite într-un cozonac cu stafide ales la intamplare dintr-o partida de astfel de cozonaci este o v.a. cu repartitia Poisson cu parametrul a , unde a este numărul mediu de stafide care pot fi descoperite într-un astfel de cozonac.1) Să se determine distribuţia v.a. X. 2) Să se calculeze probabilităţile evenimentelor: $A = \{$ *vor fi descoperite cel mult* 10 *stafide*$\}$ şi $B = \{$*vor fi descoperite exact* 10 *stafide*$\}$, $C = \{$*numărul de bacterii va varia intre* 500 *si* 1000$\}$. Aflati numărul de stafide descoperite într-un cozonac cu stafide ales la întamplare, care corespunde celei mai mari probabilităţi? Să se considere că $a = 10$.

28. Numărul de grindine înregistrate per m^2 în timpul unei ploi cu grindină este o v.a. cu distribuţia Poisson cu parametrul a , unde a este numărul mediu de grindine înregistrate pe m^2. 1) Să se determine repartiţia v.a. X. 2) Să se calculeze probabilităţile evenimentelor: $A = \{$ *pe o suprafata de* 1 m^2 *vor fi inregistrate cel mult* 50 *de grindine*$\}$ şi $B = \{$*vor fi descoperite exact* 50 *de grindine*$\}$, $C = \{$*numarul de bacterii va varia intre* 100 *si* 200$\}$. Aflaţi numărul de stafide descoperite pe o suprafaţă de $1m^2$ aleasă la întâmplare, număr care corespunde celei mai mari probabilităţi? Să se considere că $a = 30$.

29. Numărul de accidente auto înregistrate in or. Chişinău in 24 de ore este o v.a. cu distribuţia Poisson cu parametrul a, unde a este numărul mediu de accidente auto in acest interval de timp. 1) Să se determine distribuţia

v.a. X. 2) Să se calculeze probabilităţile evenimentelor: $A = \{ X \leq 5\}$ şi $B = \{ X = 5\}$, $C = \{ 5 \leq X \leq 10\}$. Aflati numărul de accidente auto, care corespunde celei mai mari probabilităţi? Să se considere că $a = 0.5$.

30. Numărul anual de decese printre asiguraţii unei companii de asigurare este o v.a. cu distribuţia Poisson cu parametrul a, unde a este numarul mediu de decese anual. 1) Să se determine distribuţia v.a. ξ. 2) Să se calculeze probabilităţile evenimentelor: $A = \{ X \leq 10\}$ şi $B = \{ X = 10\}$, $C = \{ 10 \leq X \leq 20\}$. Aflati numărul anual de decese printre asiguratii unei companii de asigurare , care corespunde celei mai mari probabilităţi? Să se considere că $a = 3$.

4.2. Distribuţii probabiliste uzuale in caz (absolut) continuu (Uniformă, Exponenţială, Normală, Hi-pătrat (χ^2), T-Student)

Distribuţia Uniformă (caz (absolut) continuu).

Definiţia 1. Vom spune că v.a. X *este distribuită uniform pe segmentul* $[a, b]$, pe scurt $X \sim U[a, b]$, $a,b \in \mathbb{R}$, $a < b$, dacă densitatea ei de distribuţie $f(x) = \frac{1}{b-a} \cdot I_{[a,b]}(x)$.

Caracteristicile numerice de bază ale v.a. X sunt: $\mathbb{E}X=\frac{a+b}{2}$, $\mathbb{D}X=\frac{(b-a)^2}{12}$, $\mu_3=0$.

Remarca 1. Din punct de vedere matematic distribuţia uniformă modelează, de exemplu, alegerea la întâmplare a unui punct de pe segmentul $[a, b]$. Experimentele aleatoare de acest gen pot fi simulate şi pe calculator deoarece orice limbaj de programare evoluat (C^{++}, Java, etc.) are in biblioteca sa de functii program funcţia Random, corespunzătoare cazului absolut , accesarea căreia are drept rezultat generarea unui număr ales la întâmplare din multimea $[0, 1]$. Dat fiind faptul că in memoria calculatorului numerele iraţionale, de exemplu, sunt, din cauza lungimii limitate a cuvântului de calculator (32 sau 64 biţi), aproximate de numere raţionale, generatorul corespunzător poartă denumirea de *Generator de Numere Pseudoaleatoare.* Acest generator in ambele sale ipostaze (discretă şi (absolut0 continue) se afla la baza simularii pe calculator a oricărei legităţi probabiliste. De exempu, având la bază generatorul de valori ale v.a. $X \sim U[0, 1]$, putem simula valori ale v.a. $Y \sim U[a, b]$ şi viceversa. Algoritmul se desprinde din

Propoziţia 1. *Daca v.a.* $X \sim U[0, 1]$ *, atunci v.a.* $Y=(b - a)X + a \sim U[a, b]$ *pentru orice* $a,b \in R$, $a < b$. *Dimpotriva, dacă v.a.* $Y \sim U[a, b]$ *, atunci v.a.* $X = (Y - a)/(b - a) \sim U[0, 1]$.

Are loc şi

Propoziţia 2. *Daca v.a.* $X \sim U[0,b]$ *, atunci şi v.a.* $Y{=}b-X \sim U[0,b]$.

Exemplul 1. Un troleibus soseşte in staţie din 5 în 5 minute. Care este probabilitatea că un pasager, care vine in staţie într-un moment aleator de timp, va aştepta troleibuzul cel mult 2 minute?

Soluţie. Din faptul că momentul de sosire în staţie este unul aleator, rezultă ca acest moment X este, de fapt, rezultatul alegerii la întâmplare a unui punct de pe un segment de lungime 5, cu alte cuvinte, $X \sim U[0,5]$. Timpul de aşteptare a următorului troleibus coincide cu v.a. $Y{=}b-X$, prin urmare, conform propoziţiei 2, $Y \sim U[0,5]$, ceea ce înseamnă că probabilitatea că un pasager, care vine in staţie într-un moment aleator de timp, va aştepta troleibuzul cel mult 2 minute coincide cu

$$P(0 \leq Y \leq 2) = \int_0^2 \frac{1}{5} dx = \frac{2}{5} = 0.4.$$

Distribuţia exponenţială.

Definiţia 2. Vom spune că v.a. X *este distribuită exponenţial* cu parametrul λ, $\lambda > 0$, pe scurt $X \sim \exp\{\lambda\}$, dacă mulţimea ei de valori posibile coincide cu $[0,+\infty)$, iar $f(x){=}\lambda e^{-\lambda x} \cdot I_{[0,+\infty)}(x)$.

Caracteristicile numerice de bază ale v.a. X sunt: $\mathbb{E}X{=}\frac{1}{\lambda}$, $\mathbb{D}X{=}\frac{1}{\lambda^2}$, $\mu_3{=}\frac{2}{\lambda^3}$. Ţinând cont de legătura dintre f.d. si d.d. a v.a. X, găsim ca f.r. a unei v.a. $X \sim \exp\{\lambda\}$ este egală cu $F(x){=}(1-e^{-\lambda x}) \cdot I_{[0,+\infty)}(x)$.

Această distribuţie apare în directă legatură cu fluxurile (procesele Poisson), legatură evidenţiată în

Propoziţia 3. *Durata intervalului de timp dintre două momente succesive în care se produc două evenimente într-un flux Poisson cu intensitatea* $\lambda > 0$ *este o v.a.* $X \sim \exp\{\lambda\}$.

Prin urmare, în toate exemplele de flux Poisson enumărate în paragraful anterior putem invoca şi distribuţia exponenţială.

Remarca 2. Dacă în caz discret singura distribuţie care posedă proprietatea "lipsei memoriei" este distribuţia geometrică, apoi în caz (absolut) continuu singura v.a. ce posedă aceasta proprietate este v.a. $X \sim \exp\{\lambda\}$. Într-adevăr, are loc

Propoziţia 4 (*Proprietatea "lipsei memoriei" a distribuţiei exponenţiale*). *V.a. X de tip (absolut) continuu, pentru care mulţimea de valori posibile coincide $[0,+\infty)$,este o v.a. distribuită exponenţial, adică*

$$F(x) = (1 - e^{-\lambda x}) \cdot I_{[0,+\infty)}(x),$$

dacă şi numai daca aceasta posedă proprietatea "lipsei memoriei":

$$P(X \leq x + h \ / \ X > x) = P(X \leq h\ ),\ \textit{pentru orice}\ h \geq 0.$$

Consecinţă. *Dacă v.a. $X \sim \exp\{\lambda\}$, atunci*

$$P(X > x + h \ / \ X > x) = P(X > h\ ),\ \textit{pentru orice}\ h \geq 0.$$

Proprietatea "lipsei memoriei" numită si "proprietate markoviană" face ca această distribuţie sa fie folosită pe larg acolo unde apare noţiunea de durată (a vieţii): în Demogragie, Teoria Fiabilităţii, Teoria Aşteptării, Actuariat, etc. A se vedea, în acest sens, exemplele 9 din p.1.3 şi 2 din p.2.2.

Exemplul 2. Timpul de asteptare la o staţie de alimentare cu carburanţi este o v.a. X cu o durată medie de aşteptare egală cu T_0. Să se afle probabilităţile următoarelor evenimente: A=$\{\frac{1}{2}T_0 < X \leq \frac{3}{2}T_0\}$, B=$\{X > 2T_0\}$.

Soluţie. Cum $X \sim \exp\{\lambda\}$ iar $\mathbb{E}X$=$\frac{1}{\lambda} = T_0$, rezultă că $X \sim \exp\{\frac{1}{T_0}\}$. Prin urmare

$$P(A) = (1 - \exp\{-\frac{1}{T_0} \cdot \frac{3}{2}T_0\}) - (1 - \exp\{-\frac{1}{T_0} \cdot \frac{1}{2}T_0\}) =$$

$$\exp\{-\frac{1}{2}\} - \exp\{-\frac{3}{2}\} = 0.383\,4,$$

$$P(B) = 1 - (1 - \exp\{-\frac{1}{T_0} \cdot 2T_0\}) = \exp\{-2\} = 0.135\,34.$$

Distribuţia normală.

Definiţia 3. Vom spune că v.a. X *este distribuită normal* cu parametrul m şi σ, $m \in \mathbb{R}$, $\sigma > 0$, pe scurt $X \sim N(m;\sigma)$, dacă mulţimea ei de valori posibile coincide cu $(-\infty,+\infty)$ iar $f(x)$=$\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{(x-m)^2}{2\sigma^2}}$. Atunci când m=0, σ=1 vom spune că X *este standard normal distribuită sau Gauss distribuită*, pe scurt, $X \sim N(0;1)$ (vezi exemplul 8, p.3.1.)

Caracteristicile numerice de bază ale v.a. X sunt: $\mathbb{E}X=m$, $\mathbb{D}X=\sigma^2$, $\mu_3=0$. Putem, astfel, vorbi că $X \sim N(m;\sigma)$ *este o v.a. normal distribuită cu media* m *şi abaterea standard* σ.

F.d. a v.a. $X \sim N(m;\sigma)$ este funcţia $F(x) = \frac{1}{\sigma\sqrt{2\pi}} \int\limits_{-\infty}^{x} e^{-\frac{(u-m)^2}{2\sigma^2}} du$ iar în cazul când X este standard normal distribuită f.d. este $\Phi(x)=\frac{1}{\sqrt{2\pi}} \int\limits_{-\infty}^{x} e^{-\frac{u^2}{2}} du$ şi d.d. $\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$. De remarcat faptul, ca valorile f.d. $\Phi(x)$, deci si $F(x)$ nu pot fi determinate decât prin metode numerice aproximative deoarece integrala $\frac{1}{\sqrt{2\pi}} \int\limits_{-\infty}^{x} e^{-\frac{u^2}{2}} du$ nu se calculează exact, dar în cadrul aplicaţiilor concrete putem utiliza Tabelul valorilor f.d. $\Phi(x)$ (Vezi Anexa 1) sau calculatorul online pentru distribuţia normala cu media si abaterea standard cunoscute, ce poate fi accesat la adresa, de exemplu,

http : //onlinestatbook.com/2/calculators/normal_dist.html .

Important e că Tabelul valorilor f.d. $\Phi(x)$ poate fi utilizat şi la calcularea valorilor f.d. $F(x)$, conform formulei $F(x)=\Phi(\frac{x-m}{\sigma})$, daca $F(x)$ corespunde distributiei $N(m;\sigma)$. Aceasta din urmă rezulta din

Propoziţia 5. *Dacă v.a.* $X \sim N(0;1)$, *atunci v.a.* $Y = \sigma X + m \sim N(m;\sigma)$. *Invers, dacă v.a.* $Y \sim N(m;\sigma)$, *atunci v.a.* $X = \frac{Y-m}{\sigma} \sim N(0;1)$.

Transformarea v.a. $Y \sim N(m;\sigma)$ în v.a. $\frac{Y-m}{\sigma}$ se numeşte *standardizare sau normare.*

Propoziţia 6. *F.d.* $\Phi(x)$ *a v.a.*$X \sim N(0;1)$ *posedă următoarele proprietăţi: a)* $\Phi(-x) + \Phi(x)=1$ pentru orice $x \in \mathbb{R}$*; b)* $P(-3 \leq X \leq 3)=\Phi(3) - \Phi(-3)=0.9973$; c) dacă $Y \sim N(m;\sigma)$, atunci $P(\alpha \leq Y \leq \beta)=P(\frac{\alpha-m}{\sigma} \leq \frac{Y-m}{\sigma} \leq \frac{\beta-m}{\sigma})=\Phi(\frac{\beta-m}{\sigma}) - \Phi(\frac{\alpha-m}{\sigma})$ pentru orice $\alpha, \beta \in \mathbb{R}$, $\alpha \leq \beta$.

Proprietatea a) arată că Tabelul valorilor f.d. $\Phi(x)$ poate fi definit doar pentru valorile argumentului $x \geq 0$, iar proprietatea b) justifică de ce Tabelul din Anexa 1 este dat doar pentru valorile $-3.99 < x < 3.99$. Mai mult, din proprietatea c) deducem următoarea

Consecinţă (*Regula* "3σ"). Pentru orice v.a.$Y \sim N(m;\sigma)$,indiferent de valoarea ei medie m şi abaterea ei standard σ, probabilitatea

$$P(m - 3\sigma \leq Y \leq m + 3\sigma) = 0.9973.$$

Cu alte cuvinte, în peste 99% din cazuri, valorile v.a. $Y \sim N(m;\sigma)$ se vor situa în jurul valorii ei medii la o distanţă de 3σ.

O alta proprietate utilă ne-o oferă

Propoziţia 7. *Dacă* X_1, X_2, ..., X_n *sunt v.a.i.i.d. Normal cu valoarea medie* $\mathbb{E}X_k = m$ *şi dispersia* $\mathbb{D}X_k = \sigma^2$, *atunci v.a.* $Y = X_1 + X_2 + \ldots + X_n \sim N(nm; \sqrt{n}\sigma)$.

Remarca 3. Exemple de v.a. distribuite normal sunt: cantitatea anuală de precipitaţii atmosferice dintr-o anumita regiune, eroarea care se obţine în urma măsurarii unei mărimi cu un aparat de măsurat, înălţimea unui bărbat luat la intamplare, etc. Mai mult, în anumite condiţii, sume de v.a. într-un număr suficient de mare pot fi considerate, cu o anumită doză de aproximare, ca fiind v.a. normal distribuite, ceea ce simplifică esenţial analiza statistica a datelor.

Exemplul 3. În cadrul planificarii producţiei de costume pentru bărbaţi destinate comercializării în Republica Moldova, Fabrica de Confecţii "Ionel" din Chişinău urmează să determine cota parte de costume produse pentru bărbaţii ce au înălţimea ce variază între 164 şi 170 cm., ştiind doar faptul că înălţimea unui bărbat ales la întâmplare din regiunea dată este o v.a. X normal distribuită cu media m=170cm şi abaterea standard $\sigma = 5cm$. Evident că această cotă exprimată în procente este egală cu $P(164 \leq X \leq 170) \cdot 100\%$. Dar

$$P(164 \leq X \leq 170) = \Phi(\frac{170-170}{5}) - \Phi(\frac{164-170}{5}) =$$

$$\Phi(0) - \Phi(-1.2) = 0.5 - 0.11505 = 0.38495.$$

Prin urmare, cota parte în cauză este egală cu aproximativ 38.5%.

Distribuţia Hi-pătrat (χ^2).

Definiţia 4. Vom spune că variabila v.a.$\chi^2(n)$ *este repartizată hi-pătrat cu n grade de libertate* daca aceasta are aceeaşi d.r. ca şi v.a. $X_1^2 + X_2^2 + \ldots + X_n^2$, unde X_1, X_2, ..., X_n sunt variabile aleatoare independente, identic standart normal distribuite, sau ceea ce este echivalent, dacă aceasta are d.d.

$$f(x) = \frac{x^{(n-2)/2} exp(-x/2)}{2^{\frac{n}{2}}\Gamma(\frac{n}{2})} \cdot I_{[0,+\infty)}(x),$$

$\Gamma(u) = \int\limits_0^{+\infty} x^{u-1}e^{-x}dx$ fiind $\Gamma(u) = \int\limits_0^{+\infty} x^{u-1}e^{-x}dx$. Distribuţia hi-patrat este tabelata si in functie de n si x , şi astfel putem afla valoarea probabilitatii

$P(\chi^2(n) \le x)$. Aceasta din urma poate fi calculată şi online la adresa

http : //onlinestatbook.com/2/calculators/normal_dist.html .

Din exemplul 3, p. 3.3. putem vedea cum variază grafic forma d.r. a v.a.$\chi^2(n)$ in functie de valoarea parametrului n.

Caracteristicile numerice de bază ale v.a. $\chi^2(n)$ sunt: $\mathbb{E}\chi^2(n)=n$, $\mathbb{D}\chi^2(n)=2n$, $\mu_3=8n$.

Remarca 4. Această distribuţie este utilizată in calitate de model matematic pentru diverse proceduri de estimare de interval sau de verificare a ipotezelor în Statistica Matematică.

Exemplul 4. Considerăm doua v.a. independente $X \sim \chi^2(15)$ şi $Y \sim \chi^2(5)$. Să se determine valorile medie si dispersia v.a. $U = X + Y$, dar şi valorile $u_{0.05}$ şi $u_{0.95}$ pentru care $P(U \leqslant u_{0.05}) = 0.05$, $P(U \leqslant u_{0.95}) = 0.95$.

Soluţie. Din definiţia distribuţiei $\chi^2(n)$ deducem că $X \sim X_1^2+X_2^2+...+X_{15}^2$ iar $Y \sim X_{16}^2+X_{17}^2+...+X_{20}^2$, unde X_1, X_2, ..., X_{15}, X_{16}, X_{17}, ..., X_{20} sunt v.a.i.i.d. standard normal distribuite. Prin urmare $U{=}X{+}Y \sim X_1^2+X_2^2+...$ $+X_{15}^2+X_{16}^2+X_{17}^2+...+X_{20}^2 \sim \chi^2(20)$. Or, $\mathbb{E}(X+Y) = \mathbb{E}\chi^2(20)=20$, $\mathbb{D}(X+Y) = \mathbb{D}\chi^2(20)=40$.

Cum $U \sim \chi^2(20)$, din tabelul prezentat in Anexa 2, vedem că $P(U \leqslant u_\alpha) = P(\chi^2(20) \le Q(\alpha, 20) = \alpha$, $0 < \alpha < 1$. Cu alte cuvinte $u_{0.05} = 10.851$ iar $u_{0.95} = 31.41$.

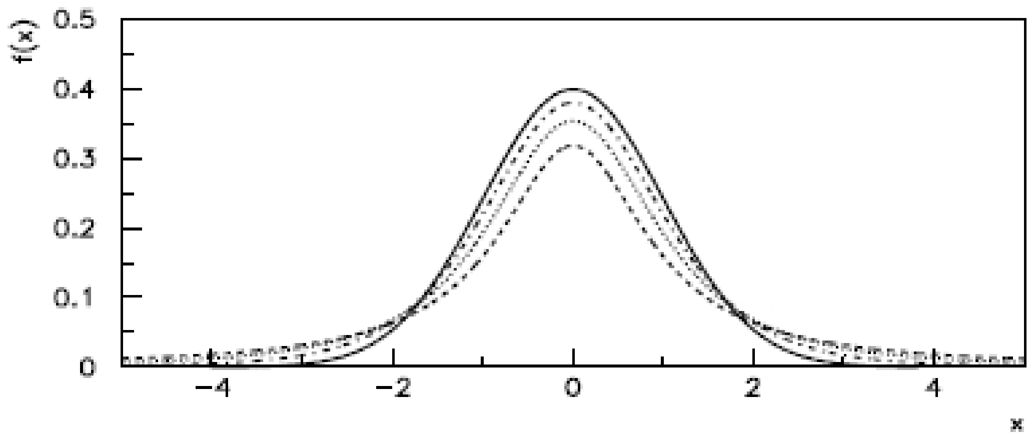
Distribuţia T-Student.

Definiţia 5. Vom spune că variabila v.a. $T(n)$ *este Student repartizată cu n grade de libertate,* pe scurt, $T(n) \sim Student(n)$, daca aceasta are d.d. ce coincide cu d.d. a v.a. $\frac{X}{\sqrt{\chi^2(n)/n}}$, unde X, $\chi^2(n)$ sunt variabile aleatoare independente respectiv, standard normal distriuită şi hi-pătrat distribuită cu n grade de libertate.

Caracteristicile numerice de bază ale v.a. $T(n)$ sunt: $\mathbb{E}T(n) = 0$, $\mathbb{D}T(n) = n/(n-2)$ pentru $n > 2$, $\mu_3=0$

La fel, aceasta distribuţie se foloseste pe larg in Statistica Matematica la construirea unor estimatori de interval, dar şi la verificarea ipotezelor statistice.

Pentru calculul probabilităţilor de forma $P(T(n) \le x)$ se folosesc tabele care iau in calcul doar cazurile cand numarul de grade de libertate nu intrece 120, deoarece odata cu cresterea lui n aceasta distribuţie se apropie de cea normala standart, ceea ce se vede şi din graficele densităţii de distribuţie Student in functie de n (graficul cu valoarea maximală cea mai mică pentru n=1, apoi $n = 2, 5, \infty$, ultimul fiind trasat cu linie ingroşată) :



Probleme propuse.

1. Înălţimea unui bărbat este o v.a. cu distribuţia normală. Presupunem că această distribuţia are parametrii $m = 175$ cm şi $\sigma = 6$ cm. Să se formeze programul de confecţionate a costumelor bărbăteşti pentru o fabrică de confecţii care se referă la asigurarea cu costume a bărbaţilor, înălţimile cărora aparţin intervalelor: $[150, 155)$, $[155, 160)$, $[160, 165)$, $[165, 170)$, $[170, 175)$, $[175, 180)$, $[180, 185)$, $[185, 190)$, $[190, 195)$, $[195, 200]$.

2. Presupunem că o convorbire telefonică durează în medie 5 minute şi este o v.a. de distribuţie exponenţială. 1) Să se scrie f.dr. şi să se traseze graficul ei. 2) Dacă vă apropriaţi de o cabină telefonică imediat după ce o persoană a întrat în ea atunci care este probabilitatea că o să aşteptaţi nu mai mult de 2 minute?

3. Un autobus circulă regulat cu intervalul 30 minute. 1) Să se scrie d.d. a v.a.c. care reprezintă durata aşteptării autobusului de către un pasager care soseste in staţie într-un moment aleator de timp. 2) Să se traseze graficul d.d. 3) Să se determine f.d. şi să se traseze graficul ei. 4) Care este probabilitatea că, sosind in staţie, pasagerul va aştepta autobusul nu mai mult de 10 minute, unde numărul n coincide cu ultima cifra a numarului de ordine a problemei.

4. Cantitatea anuală de precipitaţii atmosferice are distribuţia normală. Presupunem că anual, cantitatea de precipitaţii într-o anumita regiune este

o v.a. aleatoare de repartiţie normală de parametrii $m = 500mm$ şi $\sigma = 150mm$. Care este probabilitatea că în anul viitor cantitatea de precipitaţii va fi cuprinsă între 400 si 500 milimetri. Considerând că un an este secetos când cantitatea de precipitaţii nu depăşeşte 300 mm, aflati probabilitatea că doi din viitorii zece ani vor fi secetoşi?

5.Fie X o v.a. care urmează o distribuţie normală cu media $m = -2$ şi abaterea standard $\sigma = 3$.

a) Calculaţi $P(-2.5 < X < 1.5)$;

b) Calculaţi $P(-7 < X < -3.2)$;

c) Calculaţi $P(X > 1.9)$;

d) Identificaţi valoarea $x_{0.25}$ astfel încât $P(X \leqslant x_{0.25}) = 0,25$.

6. Diametrul mingilor de Ping-Pong (notat cu X) fabricate într-o fabrică mare este de aşteptat să fie normal distribuite cu o medie de $3,3cm$ şi o abatere standard de $0,6cm$.

Care este probabilitatea ca o minge de Ping-Pong aleasă la întâmplare va avea un diametru

a) Mai mic de $3,25cm$?

b) Între $3.25cm$ şi $3.35cm$?

c) Să se determine cele două valori ale extremităţilor intervalului cu centrul în valoarea medie pentru care 60% din diametrele mingilor de Ping-Pong vor nimeri în acesta?

Dacă sunt selectate multe eşantioane aleatoare de 25 de mingi Ping-Pong

(d) Ce distribuţie va urma media eşantionul?

(e) Ce proporţie din eşantioane vor avea o medie mai mică de $3,25cm$?

(f) Ce proporţie din eşantioane vor avea o medie între $3.25cm$ şi $3.35cm$?

(g) Să se determine cele două valori ale extremităţilor intervalului cu centrul în valoarea medie a *mediei* diametrelor mingilor incluse în eşantion pentru care 60% din *mediile* diametrelor mingilor incluse în eşantion vor nimeri în acest interval?

(h) Ce este mai probabil: să apară o minge luată aparte cu diametrul între $3,25$ şi $3,35cm$ sau un eşantion cu o medie intre $3,25$ şi $3,35cm$? Explicaţi.

4.3. Inegalitatea Chebyshev, Legea Numerelor Mari (în formele Cebyshev, Bernoulli, Hincin), Teorema Limită Centrală

Modelele matematice ce-şi gasesc aplicare în Statistica Matematică vizează, de fapt, v. a. n-dimensionale $X= (X_1, X_2, ..., X_n)$ sau funcţii de forma $g(X_1,$

X_2, ..., X_n) privite ca v.a., interes prezentând comportamentul lor probabilist atunci când $n \longrightarrow \infty$. În acest scop sunt utile inegalităţi de tip Inegalitatea Chebyshev, care scot în evidenţa faptul că, în anumite condiţii, cunoaşterea doar a unor caracteristici sumare ale v.a. X, cum ar fi valoarea medie, momentul central de ordinul 2 sau dispersia, sunt suficiente pentru a evalua probabilităţi legate de această v.a.

Propziţia 1 (*Inegalitatea Markov*). *Dacă pentru v.a. X există $\mathbb{E}|X|$, atunci are loc inegalitatea*

$$P(|X| \geq \varepsilon) \leq \frac{\mathbb{E}|X|}{\varepsilon}, \forall \varepsilon > 0.$$

Consecinţa 1. *Dacă v.a. $X \geq 0$ şi există $\mathbb{E}X$, atunci are loc inegalitatea*

$$P(X \geq \varepsilon) \leq \frac{\mathbb{E}X}{\varepsilon}, \forall \varepsilon > 0.$$

Consecinţa 2 (*Inegalitatea Chebyshev*). *Dacă pentru v.a. X există momentul iniţial EX^2, atunci are loc inegalitatea Chebyshev în formele*

$$P(|X| \geq \varepsilon) \leq \frac{\mathbb{E}X^2}{\varepsilon^2}, \forall \varepsilon > 0.$$

şi

$$P(|X - \mathbb{E}X| \geq \varepsilon) \leq \frac{\mathbb{D}X}{\varepsilon^2}, \forall \varepsilon > 0.$$

Consecinţa 2 (*Regula "$k\sigma$"*). *Dacă pentru v.a. X există momentul iniţial $\mathbb{E}X^2$, atunci pentru k=2, 3, 4*

$$P(\mathbb{E}X - k\sigma < X < \mathbb{E}X + k\sigma) > 1 - \frac{1}{k^2}, .$$

Această consecinţă arată, de exemplu, pentru $k = 3$, că în cel puţin $(1 - \frac{1}{3^2})100\% \simeq 89\%$ din cazuri, valorile v.a. X se vor situa în intervalul $(\mathbb{E}X - 3\sigma, \mathbb{E}X + 3\sigma)$. Aşa cum arată regula 3σ, formulată drept consecinţă din propoziţia 6 pentru distribuţia normală cu media m şi abaterea standard σ, aceeaşi regulă dedusă din inegalitatea Chebyshev este mai puţin exacta decât prima, dar aceasta din urmă având avantajul de a fi aplicabilă pentru toate v.a. pentru care există media si dispersia.

Exemplul 1. În presupunerea că v.a. X are valoarea medie $\mathbb{E}X = 1$ şi abaterea standard $\sigma = 0.2$ putem evalua probabilităţile (graniţele lor de jos) evenimentelor

$$A = \{0.5 < X < 1.5\}, B = \{0.75 < X < 1.35\}, C = \{X < 2\}.$$

Într-adevăr, $A = \{1 - 0.5 < X < 1 + 0.5\} = \{|X - \mathbb{E}X| < 0.5\}$. Dar

$$P(\overline{A}) = P(|X - \mathbb{E}X| \geq 0.5) \leq \frac{\sigma^2}{0.5^2} = \frac{0.2^2}{0.5^2} = 0.16.$$

Prin urmare $P(A) = 1 - P(\overline{A}) \geq 1 - 0.16 = 0.84$. Analogic, deoarece $B \supseteq \{0.75 < X < 1.25\} = \{|X - \mathbb{E}X| < 0.25\}$, avem

$$P(B) = 1 - P(\overline{B}) \geq 1 - P(|X - \mathbb{E}X| < 0.25 \geq \frac{0.2^2}{0.25^2} = 0.64.$$

Cum $C \supseteq \{0 < X < 2\} = \{|X - \mathbb{E}X| < 1\}$, avem că

$$P(C) = 1 - P(\overline{C}) \geq 1 - P(|X - \mathbb{E}X| < 1) \geq 1 - \frac{0.2^2}{1^2} = 1 - 0.04 = 0.96.$$

Exemplul 2. Viteza vântului în $km/oră$ în orice punct al Pământului este o v.a. X. Să se evalueze graniţa de sus a probabilităţii evenimentului $\{X \geq 80\ km/ora\}$ în fiecare din următoarele cazuri: *a)* în urma unor măsurări facute mai mulţi ani la rând s-a stabilit că $\mathbb{E}X = 16\ km/oră$; *b)* în urma unor măsurări facute suplimentar s-a stabilit că abaterea standard a v.a. X este egală cu $\sigma = 4\ km/oră$.

Soluţie. *a)* Deoarece X este o v.a. nenegativa, putem aplica Inegaliattea Markov:

$$P\{X \geq 80 km/ora\} \leq \frac{16\ km/ora}{80\ km/ora} = 0.2.$$

b) În acest caz, înainte să aplicăm Inegalitatea Cebyshev, observăm că

$$\{X \geq 80\} = \{X - \mathbb{E}X \geq 80 - \mathbb{E}X\} = \{X - \mathbb{E}X \geq 64\} \subseteq \{|X - \mathbb{E}X| \geq 64\}.$$

Prin urmare

$$P\{X \geq 80\} \leq P\{|X - \mathbb{E}X| \geq 64\} \leq \frac{\mathbb{D}X}{64^2} = \frac{4^2}{64^2} \simeq 0.004.$$

Remarca 1. Exemplul 2 este sugesitiv, în sensul că arată foarte clar efectul cunoaşterii suplimentare a dispersiei v.a. X asupra mariiri exactităţii estimării probabilităţii necunoscute.

Definiţia 1. Vom spune că şirul de v.a. X_1, X_2, ..., X_n, ... *este guvernat de Legea Numerelor Mari (LNM)* dacă

$$\lim_{n\longrightarrow\infty} P\left(\left|\frac{1}{n}\sum_{k=1}^{n} X_k - \frac{1}{n}\sum_{k=1}^{n}\mathbb{E}X_k\right| < \varepsilon\right) = 1,\ \forall\varepsilon > 0.$$

Teorema 1 (*LNM în forma Cebyshev*). *Dacă X_1, X_2, ..., X_n, ... este un şir de v.a. necorelate două câte două pentru care există dispersia fiecărei v.a. şi o constantă c astfel încât dispersiile $\mathbb{D}X_k \leq c$, $k = 0{,}1{,}2$, ..., atunci şirul acesta este guvernat de Legea Numerelor Mari.*

Consecinţa 1. *Dacă X_1, X_2, ..., X_n, ... este un şir de v.a. independente pentru care există dispersia fiecărei v.a. şi o constantă c astfel încât dispersiile $\mathbb{D}X_k \leq c$, $k = 0{,}1{,}2$, ..., atunci şirul acesta este guvernat de Legea Numerelor Mari.*

Consecinţa 2 (*LNM în forma Bernoulli*). *Dacă X_1, X_2, ..., X_n, ... este un şir de v.a. i.i.d. Bernoulli cu parametrul p, $0 < p < 1$, atunci şirul acesta este guvernat de Legea Numerelor Mari, care în acest caz ia forma*

$$\lim_{n\longrightarrow\infty} P\left(\left|\frac{1}{n}\sum_{k=1}^{n} X_k - p\right| < \varepsilon\right) = 1,\ \forall\varepsilon > 0.$$

Remarca 2. LNM în forma Bernoulli serveşte în calitate de fundament matematic pentru Principiul Regularităţii (Stabilităţii) Statistice. Întradevăr, cum orice experiment aleator $\mathcal{E}$ ce poate fi repetat independent unul de altul în aceleaşi condiţii, pentru fiecare eveniment aleator A de probabilitate nenulă $p = P(A)$, poate fi transformat în probă Bernoulli cu probabilitaea "succesului" p, rezultă că v.a.

$$X_k = \begin{cases} 1, & \textit{dacă în proba cu nr. } k \textit{ se produce } A, \\ 0, & \textit{în caz contrar,} \end{cases} \quad k =, 1, 2, ...$$

sunt guvernate de LNM în forma Bernoulli. Dar frecvenţa relativă

$$f_n(A) = \frac{1}{n}\sum_{k=1}^{n} X_k.$$

Prin urmare pentru astfel de experimente aleatoare LNM ia forma

$$\lim_{n\longrightarrow\infty} P\left(|f_n(A) - P(A)| < \varepsilon\right) = 1,\ \forall\varepsilon > 0,$$

care devine, din punct de vedere matematic, o expresie riguroasă a Principiul Regularităţii Statistice.

De remarcat ca LNM expusă mai sus în diverse forme este, totuşi, condiţionată de existenţa valorii medii şi a dispersiei, dar în anumite condiţii existenţa dispersiei nu este obligatorie, aşa cum se vede din

Teorema 2 (*LNM în forma Hincin*). *Dacă X_1, X_2, ..., X_n, ... este un şir de v.a.i.i.d. şi pentru care există valoarea medie $\mathbb{E}X_k = m$, $k=1, 2, ...$, atunci şirul acesta este guvernat de Legea Numerelor Mari, care în acest caz ia forma*

$$\lim_{n\longrightarrow\infty} P\left(\left|\frac{1}{n}\sum_{k=1}^{n} X_k - m\right| < \varepsilon\right) = 1,\ \forall\varepsilon > 0.$$

Exemplul 3. În Statistică, atunci când avem de a face cu valorile (X_1, X_2, ..., X_n) unei v.a. X, valori obţinute în urma unor observaţii independente asupra acestei v.a., media aritmetica $\overline{X}=\frac{1}{n}\sum\limits_{k=1}^{n} X_k$ serveşte în calitate de estimator pentru valoarea medie (teoretică) $\mathbb{E}X = m$, m find necunoscut. LNM în forma Hincin garantează că, odată cu cresterea lui n, aceast aproximare devine tot mai bună pentru m. Din păcate, însă, LNM nu ne spune nimic despre viteza cu care acest estimator se apropie de valoarea estimată m. Aceasta lacună vine să o înlăture

Teorema 3 (*Teorema Limită Centrală pentru v.a.i.i.d.*). *Dacă X_1, X_2, ..., X_n, ... este un şir de v.a.i.i.d. şi pentru care există valoarea medie $\mathbb{E}X_k = m$ şi dispersia $\mathbb{D}X_k = \sigma^2$, $k=1, 2, ...$, atunci*

$$\lim_{n\longrightarrow\infty} P\left(\frac{\sum\limits_{k=1}^{n} X_k - nm}{\sigma\sqrt{n}} \leq x\right) = \Phi(x) = \frac{1}{\sqrt{2\pi}}\int\limits_{-\infty}^{x} e^{-\frac{u^2}{2}}du,\ \forall x \in \mathbb{R}.$$

Consecinţa 1. (*Teorema Limită Centrală în forma Moivre-Laplace*). *Dacă X_1, X_2, ..., X_n, ... este un şir de v.a. i.i.d. Bernoulli cu parametrul*

p, $0 < p < 1$, atunci

$$\lim_{n \longrightarrow \infty} P\left(\frac{\sum\limits_{k=1}^{n} X_k - np}{\sqrt{np(1-p)}} \leq x \right) = \Phi(x) = \frac{1}{\sqrt{2\pi}} \int\limits_{-\infty}^{x} e^{-\frac{u^2}{2}} du, \forall x \in \mathbb{R}.$$

Următoarea afirmaţie vine sa ilustreze modul în care $\frac{1}{n}\sum\limits_{k=1}^{n} X_k$ converge către valoarea medie.

Consecinţa 2. *Dacă X_1, X_2, ..., X_n, ... este un şir de v.a.i.i.d. şi pentru care există valoarea medie $\mathbb{E}X_k = m$ şi dispersia $\mathbb{D}X_k = \sigma^2$, k=1, 2, ..., atunci:*

a) *pentru $\varepsilon > 0$ dat şi n suficient de mare probabilitatea*

$$P\left(\left| \frac{1}{n} \sum_{k=1}^{n} X_k - m \right| \leq \varepsilon \right) \simeq 2\Phi(\frac{\varepsilon\sqrt{n}}{\sigma}) - 1.$$

b) *pentru orice $\varepsilon > 0$ şi $0 < \alpha < 1$ probabilitatea*

$$P\left(\left| \frac{1}{n} \sum_{k=1}^{n} X_k - m \right| \leq \varepsilon \right) \geq 1 - \alpha$$

deîndată ce $n \geq n_0 = \left[\left(\frac{\sigma x_{1-\alpha/2}}{\varepsilon} \right)^2 \right] + 1$, unde $[a]$ este partea întreagă a numărului a, iar $x_{1-\alpha/2}$ este $1 - \alpha/2$ - quantilă a v.a. standard distribuită, adică

$$x_{1-\alpha/2} : \Phi(x_{1-\alpha/2}) = 1 - \alpha/2.$$

Remarca 2. Importanţa Teoremi Limită Centrală rezidă, mai ales, în faptul că aceasta arată că, în anumite condiţii, o sumă $S_n = \sum\limits_{k=1}^{n} X_k$ de v.a.independente, luate într-un număr suficient de mare, indiferent care este distribuţia lor, se comportă din punct de vedere probabilist ca şi cum aceasta ar fi o v.a. normal distribuită. Mai exact, avem

Consecinţa 3. *Dacă X_1, X_2, ..., X_n, ... este un şir de v.a.i.i.d. şi pentru care există valoarea medie $\mathbb{E}X_k = m$ şi dispersia $\mathbb{D}X_k = \sigma^2$, k=1, 2, ..., atunci pentru n* suficient de mare

$$P\left(\sum_{k=1}^{n} X_k \leq x \right) \approx \Phi(\frac{x - nm}{\sigma\sqrt{n}}),$$

sau cu alte cuvinte $S_n = \sum\limits_{k=1}^{n} X_k \simeq N(nm; \sigma\sqrt{n})$.

Exemplul 4. O Companie de asigurări are asigurate 10000 de automobile. Probabilitatea unei avarieri, ce intră sub incidenţa poliţei de asigurare, fiind egală cu p=0.006. Fiecare proprietar de automobil asigurat plăteşte o poliţă de asigurare anuală de 50 de euro, primind, în caz de avariere o sumă de 4000 de euro. Să se calculeze probabilităţile următoarelor evenimente: A=\{*la sfârşit de an Compania fa fi în pierdere*\}, B_m=\{*la sfârşit de an Compania va avea un profit de, cel puţin,* m euro\}, m=160000, 240000, 320000.

Soluţie. Considerăm v.a.

$$X_k = \begin{cases} 1, \textit{dacă automobilul cu nr. } k \textit{ va fi avariat}, \\ 0, \textit{în caz contrar,} \end{cases}$$

$k = 1, 2, ..., 10000$. Atunci evenimentul

$$A \equiv \{4000 \cdot \sum_{k=1}^{10000} X_k \geq 500000\} = \{\sum_{k=1}^{10000} X_k \geq 125\}.$$

Deoarece v.a. $X_k \sim Bernoulli(0.006)$, k=1, 2, ..., 10000, n=10000 fiind suficient de mare, putem aplica Teorema limită Centrală în forma Moivre-Laplace şi Consecinţa 3:

$$P(A) = P\left(\sum_{k=1}^{10000} X_k \geq 125\}\right) \approx 1 - \Phi\left(\frac{125 - 10000 \cdot 0.006}{\sqrt{10000 \cdot 0.006 \cdot (1 - 0.006)}}\right) =$$

$$= 1 - \Phi(8.416\,8) = 1 - 1 = 0.$$

Analogic,

$$B_{160000} \equiv \{4000 \cdot \sum_{k=1}^{10000} X_k \leq 500000 - 160000\} = \{\sum_{k=1}^{10000} X_k \leq 85\},$$

prin urmare

$$P(B_{160000}) = P\{\sum_{k=1}^{10000} X_k \leq 85\} \approx \Phi\left(\frac{85 - 10000 \cdot 0.006}{\sqrt{10000 \cdot 0.006 \cdot (1 - 0.006)}}\right) =$$

$$= \Phi\,(3.\,237\,2) = 0.9987.$$

Cititorul poate verifica desinestător că $P(B_{240000}) \approx 0.7413$ şi $P(B_{320000}) \approx 0.1587$.

Probleme propuse.

1. Preţul de vânzare al unui anumit produs agricol la Piaţa Centrală variază de la zi, fiind o v.a. X cu valooare medie egala cu 2 lei/kg. Un cumpărător afirmă că probablitatea că în ziua următoare preţul la acest produs va fi mai mare sau egal cu 10 lei/kg este mai mică sau egală cu 0.2. Este oare această afirmaţie fondată?

2. Producatorul de automobile Mitsubishi Colt garantează, conform cărţii tehnice a acestui autoturism, un consum mediu de benzina în afara oraşului egal cu 5 $l/100km$ cu o abatere standart egala cu 0.25 $l/100km$. Să se estimeze valoarea probabilităţii că proprietarul unui atare autoturism va înregistra un consum real de benzina în afara oraşului cuprins între 4.5 si 5.5 $l/100km$.

3. Presupunem ca durata vieţii măsurată în sute de ore a memoriei hard produs pentru un anumit tip de laptop este o v.a. măsurată în sute de ore şi definită de d.d. $f(x) = \lambda e^{-\lambda x} I_{[0,+\infty)}(x)$. Considerăm că v.a. X_1, X_2, ..., X_n sunt v.a.i.i.d. cu d.d. $f(x)$. Pentru v.a. $\overline{X}=\frac{1}{n}\sum\limits_{k=1}^{n} X_k$, luând n=100 si λ=1/10, calculaţi cu aproximare probabilitatea $P(\overline{X} \geq 15)$.

5. Elemente de Teoria Informaţiei

5.1. Obiectul de studiu al Teoriei Informaţiei

Teoria Informaţei (TI) studiază modele matematice (preponderant statisticoprobabiliste) ale fenomenelor legate de recepţionarea, stocarea şi transmiterea informaţiei. Părintele TI este Claude Shannon, anul 1948 fiind considerat anul apariţiei acestei teorii.

Noţiunea de informaţie este o noţiune primară, adică imposibil de încadrat într-o definiţie strictă, la fel ca noţiunea de punct in geometrie sau noţiunea de mulţime din Teoria Mulţimilor. Încercând, totuşi, să o explicăm, am putea spune ca informaţia pentru un sistem oarecare, biologic sau tehnic, este un mesaj despre evenimente care au sau au avut sau vor avea loc atât în exteriorul sistemului cât şi în interiorul lui.

Menţionăm că între informaţie şi nedeterminare există o legătură strânsă. O informaţie este informaţie, în sensul adevărat al cuvântului , dacă şi numai dacă ea înlătură o anumită nedeterminare.

Într-un experiment se obţine o informaţie numai atunci când nu cunoaştem rezultatul său înainte de efectuarea experimentului, adică numai atunci când rezultatul înlătură nedeterminarea care exista la început.Mai mult, cu cât nedeterminarea de la începutul experimentului este mai mare, cu atât mai mare va fi cantitatea de informaţie pe care o obţinem atunci când aflăm rezultatul experimentului.

5.2. Entropia ca masură a nedeterminării/cantităţii de informaţie

În cazul unui eveniment aleator A asociat experimentului aleator $\mathcal{E}$ şi pentru care probabilitaea sa $P(A) > 0$, Claude Shannon defineste cantitatea de informaţie ce se conţine în mesajul "s-a produs evenimentul A" ca fiind numărul $I(A) = -log2P(A)$. Cea mai mica unitate a cantităţii de infornmaţie se numeşte bit, unde $I(A) = 1bit$ atunci când $P(A) = 1/2$.

Evident, $I(A)$ poate fi interpretată şi ca, măsura nedeterminării legate de producerea evenimentului A. această nedeterminare fiind cu atât mai mare cu cât probabilitatea $P(A)$ va fi mai mica. În extremis, gradul de nedeterminare este egal cu 0 dacă $P(A) = 1$.

Ne interesează măsura nedeterminării înlăturate sau cantitatea de informaţie furnizată de realizarea unui experiment aleator $\mathcal{E}$. Pentru aceasta presupunem că în acest experiment observatorul este interest (înregistrează) producerea unuia din evenimentele A_1,A_2,. . . , An care alcătuiesc un grup complet de evenimente, adică acestea sunt disjuncte doua cate doua si suma lor coincide cu evenimentul sigur. Realizarea efectivă a experimentului E, înregistrarea unui anumit eveniment dintre evenimentele A_1 , A_2 ,. . . , An , ca rezultat a experimentului, înlătură nedeterminarea pe care am avut-o la început.

Exemplu. Fie 2 experimente aleatoare:

$\mathcal{E}_1$-aruncarea unei monede "perfecte" şi grupul complet de evenimente A_1=\{*apariţia stemei*\}=\{S\}, A_2=\{*apariţia banului*\}=\{B\}

$\mathcal{E}_2$-aruncarea unei monede deformate, astfel încât $P\{S\} = 0.9, P\{B\} = 0.1$.

Avem, astfel, repartiţiile corespunzătoare:

$$\mathcal{E}_1 : \begin{pmatrix} S & B \\ 05 & 08 \end{pmatrix}, \mathcal{E}_2 : \begin{pmatrix} S & B \\ 09 & 01 \end{pmatrix}.$$

Evident, $\mathcal{E}_1$ conţine mai multă nedeterminare decât $\mathcal{E}_2$, fapt care poate fi confirmat folosind noţiunea de entropie definite în continuare.

Fie un experiment aleator $\mathcal{E}$ în care putem observa, în urma efectuării lui, unul din evenimentele A_1,A_2,$\ldots$, An care alcătuiesc un grup complet de evenimente. Presupunem că $p_i = P(A_i) > 0, i = 1, \ldots, n$.

Cantitatea de informaţie I furnizată în urma efectuării experimentului $\mathcal{E}$ depinde de evenimentul A_i care se va produce, prin urmare aceasta este o v.a. data de repartiţia ,

$$I : \begin{pmatrix} I(A_1) & I(A_2) & \ldots & I(A_n) \\ p_1 & p_2 & \ldots & p_n \end{pmatrix}, p_i > 0, i = 1, \ldots, n, (\mathbf{1})$$

Anume valoarea medie $\mathbb{E}I$ a acestei v.a. poate fi luată în calitate de măsurător al cantităţii de informaţie / gradului de nedeterminare furnizat de experimentul $\mathcal{E}$. Mai exact, putem da următoarea

Definiţie. Vom numi *entropie* sau *măsură a nedeterminării* sau *cantitate de informaţie a experimentului* $\mathcal{E}$ numărul

$$H(p_1, p_2, \ldots, p_n) = -\sum_{i=1}^{n} p_i \cdot log_2 p_i,$$

considerând că $p \cdot log_2 p = 0$ dacă $p = 0$.

Revenind la exemplul de mai sus, constatăm ca entropiile lor sunt egale cu $H(\mathcal{E}_1) = 1bit$, $H(\mathcal{E}_2) = -(0.9log_2 0.9 + 0.1log_2 0.1) < H(E_1)$ deoarece funcţia $H(p) = -(plog2p + (1 - p)log2(1 - p)$ are valoare maximală ,egala cu 1, pentru $p = 0.5$. Cantitatea de informaţie I furnizată în urma efectuării experimentului $\mathcal{E}$ depinde de evenimentul A_i care se va produce, prin urmare aceasta este o v.a. data de repartiţia , $p_i > 0, i = 1, \ldots, n$,

5.3. Proprietăţile entropiei

Entropia posedă un şir de proprietăţi, cele mai importante fiind prezentate în următoarele teoreme.

Teorema 1. *Entropia* $H(p_1, p_2, \ldots, p_n)$ *ce corespunde experimentului* $\mathcal{E}$ *cu repartiţia (1) posedă următoarele proprietăţi :*

a) $H(p_1, p_2, \ldots, p_n) \geq 0$;

b) $H(p_1, p_2, \ldots, p_n) = 0$ *dacă şi numai dacă există un indice io astfel încât* $p_{i_o} = 1$;

c) $H(p1, p2, \ldots, pn) < H(1/n, 1/n, \ldots 1/n)$ *pentru orice*

$(p_1, p_2, \dots, p_n), p_i \geq 0, i = 1, \dots, n,$;

d) $\quad H(p_1, p_2, \dots, p_n, 0) = H(p_1, p_2, \dots, p_n).$

Având două experimente $\mathcal{E}_1$ şi $\mathcal{E}_2$ independente putem define entropia experimentului $\mathcal{E} = (\mathcal{E}_1, \mathcal{E}_2)$ ce constă în efectuarea lor simultană. Notăm entropia experimentului cu $H(\mathcal{E})$. Are loc

Teorema 2.$H(\mathcal{E}) = H(\mathcal{E}_1, \mathcal{E}_2) = H(\mathcal{E}_1) + H(\mathcal{E}_2)$.

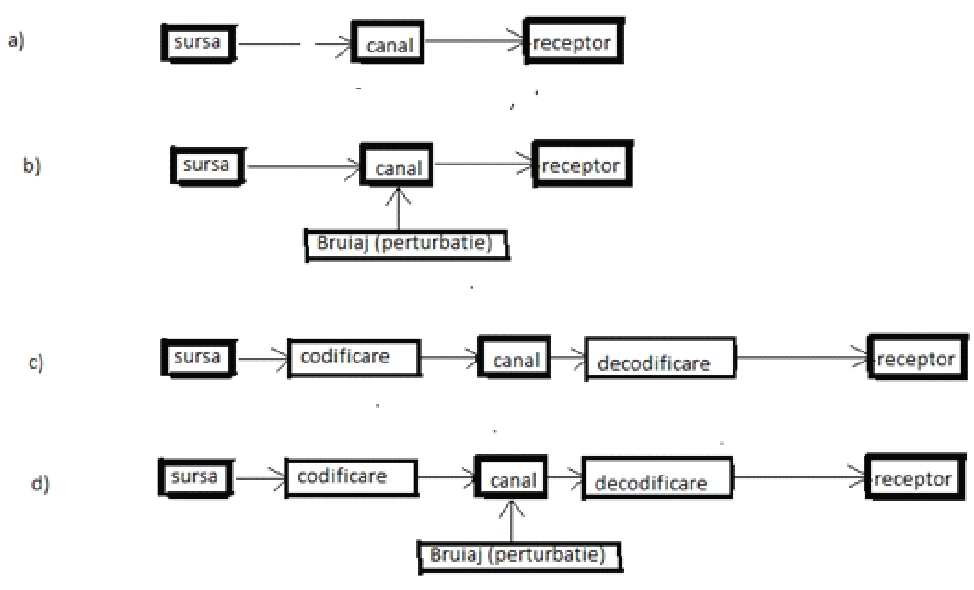
Exemplu. Consideram in calitate de experiment aleator $\mathcal{E} = (\mathcal{E}_1, \mathcal{E}_2)$ aruncarea simultană a două monede perfecte, $\mathcal{E}_1$ fiind aruncarea primei monede, iar $\mathcal{E}_2$ aruncarea celei de a doua monede. Evident, $H(\mathcal{E}_1) = H(\mathcal{E}_2) = 1bit$, dar cantitatea de informaţie $I(\mathcal{E})$ furnizată de rezultatul efectuării experimentului $\mathcal{E}$ este o v.a. cu repartiţia

$$I : \begin{pmatrix} I(SS) & I(SB) & I(BS) & I(BB) \\ 14 & 14 & 14 & 14 \end{pmatrix}.$$

Prin urmare $H(\mathcal{E}) = H(\mathcal{E}_1, \mathcal{E}_2) = H(\mathcal{E}_1) + H(\mathcal{E}_2)$.

5.4. Transmiterea informaţiei. Codificarea. Teoreme de codificare

Orice system de transmitere a informaţiei se încadrează în una din următoarele scheme generale:



Pentru a înţelege rolul noţiunii de entropie în Teoria Informaţiei ne vom referi în continuare, drept exemplu, la schema c) de transmitere a infpormaţiei cu codificare/decodificare, dar fără bruiaj. Dar pentru început vom vedea ce reprezintă codificarea/decodificarea în sistemele de comunicaţie de orice natură, pornind de la următorul

Exemplu. Consideram trimiterea unui mesaj prin sistemul SMS (system de mesaje scurte) al unui telefon mobil. Pentru aceasta folosim, initial, in calitate de semnale caracterele tastaturii QWERTY, numarul lor total fiind notat cu N. Cu alte cuvinte aceste caractere formeaza o mulţime $X = \{x_1, x_2, \ldots, x_N\}$ de semnale. În momentul trimiterii lui (apasarii butonului "send") fiecare semnal/character x din X este codificat digital, adică cu ajutorul unei mulţimi de semnale simple $Y = \{y_1, y_2, \ldots, y_Q\} = \{0, 1\}$.

Definiţie. Vom numi codificare (codare) asocierea la un anumit system de semnale care poartă o informaţie (mesaj) a unor succesi-uni/şiruri de semnale a unui alt system de semnale. Primele se numesc semnale initiale, cele de a doua se numesc semnale simple.

Este firesc ca multimea Y de semnale simple să conţină mai puţine semnale decât multimea X cea ce se şi întâmplă în exemplul nostru, $Q < N$.

În Teoria Informaţiei se folosesc, în funcţie de rezultatele scontate, diferiţi algoritmi de codare/decodare.

Motivele care ne conduc la necesitatea codarii informatiei sunt multiple. Iata câteva dintre ele:

a) Natura sistemului concret de transmitere a informaţiei (telegraf, telefon mobil, sistemul Morse, etc.) implică un anumit mod de codare;

b) Prezenţa bruiajelor (perturbaţiilor) canalelor de transmitere a informaţiei implică utilizarea codării pentru a diminua efectul nociv asupra corectitudinii informaţiei recepţionate;

c) La fel, caracterul intim, secret al informaţiei implică codarea ei.

Observatie. Motive de tipul *c)* pot avea drept consecintă faptul că, uneori, $cardY \geq cardX$.

Fie $X = \{x_1, x_2, \ldots, x_N\}$ mulţimea de semnale iniţiale, iar $Y = \{y_1, y_2, \ldots, y_Q\}$ multimea de semnale simple. În procesul alcătuirii/trimiterii unui mesaj (unei informaţii), folosind semnalele din X, fiecare xi din X are o anumita probabilitate (frecventa relativă) $p(x_i)$ cu care se intalneste in acest mesaj,

. Să luăm, drept exemplu, mesajul "*mâine_va_ploua*". Acesta are lungimea 14, incluzând simbolul "*pauză*"- "_", iar semnalele "*m*", "*a*" şi "*b*", luate ca exemplu, au frecventa relativă egală, respectiv, cu $p(m) = 1/14, p(a) = 1/7, p(b) = 0$. Prin urmare, transmiterii unui mesaj $\mathcal{E}$ i se poate asocia cantitatea de informaţie $H(\mathcal{E})$. Cum în procesul codării (criptarii) fiecarui semnal xi i se asociază un şir de semnale simple din Y, şirul având lungimea ni , $i = 1, 2, \ldots, N$, atunci putem calcula lungimea medie L a unui şir de semnale simple folosite în procesul de codare:

Acum putem formula, drept examplu, două rezultate teoretice ce vizează procesul de codificare a semnalelor transmise prin intermediul unui system/canal de transmitere a informatiei fără bruiaj, rezultate care arata care sunt limitele şi posibilităţile codificari.

Propoziţie. Pentru ca să fie posibilă codificarea prin intermediul şirurilor de semnale simple de lungimea n_i , $i = 1, 2, \ldots, N$, care se fie ataşate semnalelor iniţiale $x_1, x_2, \ldots, x_N$, este necesar şi suficient să fie îndeplinită inegalitatea

$$\sum_{i=1}^{N} Q^{-n_i} < 1.$$

Teorema codificării. Ca să putem efectua o codificare, folosind Q semnale simple, pentru a transmite o cantitate de informaţie $H(\mathcal{E})$, este necesar ca lungimea medie L a şirurilor de codificare, ataşate semnalelor iniţiale din multimea X , purtătoare de informaţie, să nu fie inferioară numărului $H(\mathcal{E}) log_2 Q$, adică $L \geq H(\mathcal{E}) log_2 Q$.

Exemplu (*continuare*). În directă legatură cu trimiterea mesajului "*mâine_va_ploua*", respectiv, experimentul $\mathcal{E}$, cantitatea de informaţie I va fi o v.a. cu repartitia

$$I : \begin{matrix} I(\hat{a}) & I(a) & I(e) & I(0) & I(u) & I(i) & I(m) & I(n) & \ldots & I(v) & I(\_) \\ 114 & 214 & 114 & 114 & 114 & 114 & 114 & 114 & \ldots & 114 & 214 \end{matrix}$$

Prin urmare $H(\mathcal{E}) = 2[-(2/14) \cdot log_2(2/14)] + 10[-(1/14) \cdot log_2(1/14)] \approx 3.5216$.

În cazul codificării digitale $Y = \{0, 1\}, Q = 2$. Prin urmare ca acest mesaj sa poata fi codificat digital este necesar ca lungimea medie a sirurilor de codificare sa fie mai mare sau egală cu $H(\mathcal{E}) log_2 Q = 3.5216$.

BIBLIOGRAFIE

[1]. Kolmogorov A.N., *Noţiunile de bază ale teoriei probabilităţilor*, M., Ed. Nauka, 1974 (l. rusă).

[2]. Feller W., *An Introduction to Probability Theory and its Applications*, Volume I, 3rd edition, N.-Y., London, John Wiley & Sons Inc., 1968.

[3]. Feller W., *An Introduction to Probability Theory and its Applications*, Volume II, 2nd edition, N.-Y., London, John Wiley & Sons Inc.,, 1971.

[4]. Iosifescu M., Mihoc Gh., Theodorescu R., *Teoria probabilităţilor si statistica matematică,* Ed. Tehnică, Bucureşti, 1966.

[5]. Jay L. Devore, *Probability&Statistics (for Enginering and the Sciences)*, 8th edition, BROOKS ⁄COLE, Chicago, 2009

[6]. Mittelhammer R. C., *Mathematical statistics for economics and bussiness*, Ed. Springer-Verlag, N.-Y. Inc., 1996.

[7]. Silviu Guiaşu, Radu Theodorescu, *Matematica si Informatia,* Edit. Stiintifica, Bucuresti, 1967.

ANEXA 1

Tabelul de valori a distribuţiei normale standard:

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{u^2}{2}} du$$

x	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
-3.9	.00005	.00005	.00004	.00004	.00004	.00004	.00004	.00004	.00003	.00003
-3.8	.00007	.00007	.00007	.00006	.00006	.00006	.00006	.00005	.00005	.00005
-3.7	.00011	.00010	.00010	.00010	.00009	.00009	.00008	.00008	.00008	.00008
-3.6	.00016	.00015	.00015	.00014	.00014	.00013	.00013	.00012	.00012	.00011
-3.5	.00023	.00022	.00022	.00021	.00020	.00019	.00019	.00018	.00017	.00017
-3.4	.00034	.00032	.00031	.00030	.00029	.00028	.00027	.00026	.00025	.00024
-3.3	.00048	.00047	.00045	.00043	.00042	.00040	.00039	.00038	.00036	.00035
-3.2	.00069	.00066	.00064	.00062	.00060	.00058	.00056	.00054	.00052	.00050
-3.1	.00097	.00094	.00090	.00087	.00084	.00082	.00079	.00076	.00074	.00071
-3.0	.00135	.00131	.00126	.00122	.00118	.00114	.00111	.00107	.00104	.00100
-2.9	.00187	.00181	.00175	.00169	.00164	.00159	.00154	.00149	.00144	.00139
-2.8	.00256	.00248	.00240	.00233	.00226	.00219	.00212	.00205	.00199	.00193
-2.7	.00347	.00336	.00326	.00317	.00307	.00298	.00289	.00280	.00272	.00264
-2.6	.00466	.00453	.00440	.00427	.00415	.00402	.00391	.00379	.00368	.00357
-2.5	.00621	.00604	.00587	.00570	.00554	.00539	.00523	.00508	.00494	.00480
-2.4	.00820	.00798	.00776	.00755	.00734	.00714	.00695	.00676	.00657	.00639
-2.3	.01072	.01044	.01017	.00990	.00964	.00939	.00914	.00889	.00866	.00842
-2.2	.01390	.01355	.01321	.01287	.01255	.01222	.01191	.01160	.01130	.01101
-2.1	.01786	.01743	.01700	.01659	.01618	.01578	.01539	.01500	.01463	.01426
-2.0	.02275	.02222	.02169	.02118	.02068	.02018	.01970	.01923	.01876	.01831
-1.9	.02872	.02807	.02743	.02680	.02619	.02559	.02500	.02442	.02385	.02330
-1.8	.03593	.03515	.03438	.03362	.03288	.03216	.03144	.03074	.03005	.02938
-1.7	.04457	.04363	.04272	.04182	.04093	.04006	.03920	.03836	.03754	.03673
-1.6	.05480	.05370	.05262	.05155	.05050	.04947	.04846	.04746	.04648	.04551
-1.5	.06681	.06552	.06426	.06301	.06178	.06057	.05938	.05821	.05705	.05592
-1.4	.08076	.07927	.07780	.07636	.07493	.07353	.07215	.07078	.06944	.06811
-1.3	.09680	.09510	.09342	.09176	.09012	.08851	.08691	.08534	.08379	.08226
-1.2	.11507	.11314	.11123	.10935	.10749	.10565	.10383	.10204	.10027	.09853
-1.1	.13567	.13350	.13136	.12924	.12714	.12507	.12302	.12100	.11900	.11702
-1.0	.15866	.15625	.15386	.15151	.14917	.14686	.14457	.14231	.14007	.13786
-0.9	.18406	.18141	.17879	.17619	.17361	.17106	.16853	.16602	.16354	.16109
-0.8	.21186	.20897	.20611	.20327	.20045	.19766	.19489	.19215	.18943	.18673
-0.7	.24196	.23885	.23576	.23270	.22965	.22663	.22363	.22065	.21770	.21476

-0.6	.27425	.27093	.26763	.26435	.26109	.25785	.25463	.25143	.24825	.24510
-0.5	.30854	.30503	.30153	.29806	.29460	.29116	.28774	.28434	.28096	.27760
-0.4	.34458	.34090	.33724	.33360	.32997	.32636	.32276	.31918	.31561	.31207
-0.3	.38209	.37828	.37448	.37070	.36693	.36317	.35942	.35569	.35197	.34827
-0.2	.42074	.41683	.41294	.40905	.40517	.40129	.39743	.39358	.38974	.38591
-0.1	.46017	.45620	.45224	.44828	.44433	.44038	.43644	.43251	.42858	.42465
-0.0	.50000	.49601	.49202	.48803	.48405	.48006	.47608	.47210	.46812	.46414
0.0	.50000	.50399	.50798	.51197	.51595	.51994	.52392	.52790	.53188	.53586
0.1	.53983	.54380	.54776	.55172	.55567	.55962	.56356	.56749	.57142	.57535
0.2	.57926	.58317	.58706	.59095	.59483	.59871	.60257	.60642	.61026	.61409
0.3	.61791	.62172	.62552	.62930	.63307	.63683	.64058	.64431	.64803	.65173
0.4	.65542	.65910	.66276	.66640	.67003	.67364	.67724	.68082	.68439	.68793
0.5	.69146	.69497	.69847	.70194	.70540	.70884	.71226	.71566	.71904	.72240
0.6	.72575	.72907	.73237	.73565	.73891	.74215	.74537	.74857	.75175	.75490
0.7	.75804	.76115	.76424	.76730	.77035	.77337	.77637	.77935	.78230	.78524
0.8	.78814	.79103	.79389	.79673	.79955	.80234	.80511	.80785	.81057	.81327
0.9	.81594	.81859	.82121	.82381	.82639	.82894	.83147	.83398	.83646	.83891
1.0	.84134	.84375	.84614	.84849	.85083	.85314	.85543	.85769	.85993	.86214
1.1	.86433	.86650	.86864	.87076	.87286	.87493	.87698	.87900	.88100	.88298
1.2	.88493	.88686	.88877	.89065	.89251	.89435	.89617	.89796	.89973	.90147
1.3	.90320	.90490	.90658	.90824	.90988	.91149	.91309	.91466	.91621	.91774
1.4	.91924	.92073	.92220	.92364	.92507	.92647	.92785	.92922	.93056	.93189
1.5	.93319	.93448	.93574	.93699	.93822	.93943	.94062	.94179	.94295	.94408
1.6	.94520	.94630	.94738	.94845	.94950	.95053	.95154	.95254	.95352	.95449
1.7	.95543	.95637	.95728	.95818	.95907	.95994	.96080	.96164	.96246	.96327
1.8	.96407	.96485	.96562	.96638	.96712	.96784	.96856	.96926	.96995	.97062
1.9	.97128	.97193	.97257	.97320	.97381	.97441	.97500	.97558	.97615	.97670
2.0	.97725	.97778	.97831	.97882	.97932	.97982	.98030	.98077	.98124	.98169
2.1	.98214	.98257	.98300	.98341	.98382	.98422	.98461	.98500	.98537	.98574
2.2	.98610	.98645	.98679	.98713	.98745	.98778	.98809	.98840	.98870	.98899
2.3	.98928	.98956	.98983	.99010	.99036	.99061	.99086	.99111	.99134	.99158
2.4	.99180	.99202	.99224	.99245	.99266	.99286	.99305	.99324	.99343	.99361
2.5	.99379	.99396	.99413	.99430	.99446	.99461	.99477	.99492	.99506	.99520
2.6	.99534	.99547	.99560	.99573	.99585	.99598	.99609	.99621	.99632	.99643
2.7	.99653	.99664	.99674	.99683	.99693	.99702	.99711	.99720	.99728	.99736
2.8	.99744	.99752	.99760	.99767	.99774	.99781	.99788	.99795	.99801	.99807
2.9	.99813	.99819	.99825	.99831	.99836	.99841	.99846	.99851	.99856	.99861
3.0	.99865	.99869	.99874	.99878	.99882	.99886	.99889	.99893	.99896	.99900

3.1	.99903	.99906	.99910	.99913	.99916	.99918	.99921	.99924	.99926	.99929
3.2	.99931	.99934	.99936	.99938	.99940	.99942	.99944	.99946	.99948	.99950
3.3	.99952	.99953	.99955	.99957	.99958	.99960	.99961	.99962	.99964	.99965
3.4	.99966	.99968	.99969	.99970	.99971	.99972	.99973	.99974	.99975	.99976
3.5	.99977	.99978	.99978	.99979	.99980	.99981	.99981	.99982	.99983	.99983
3.6	.99984	.99985	.99985	.99986	.99986	.99987	.99987	.99988	.99988	.99989
3.7	.99989	.99990	.99990	.99990	.99991	.99991	.99992	.99992	.99992	.99992
3.8	.99993	.99993	.99993	.99994	.99994	.99994	.99994	.99995	.99995	.99995
3.9	.99995	.99995	.99996	.99996	.99996	.99996	.99996	.99996	.99997	.99997

ANEXA 2

Tabel pentru $\alpha-$cuantilele uzuale $Q(\alpha)$, $\alpha \in (0,1)$ ale v.a. $\chi^2(\nu)$ distribuite Hi-pătrat cu ν grade de libertate:
$P(\chi^2(\nu) \leq Q(\alpha, \nu)) = \alpha$, unde pentru $\nu \geq 40$ quantilele $Q(\alpha, \nu) \approx Q(\alpha, 40)$

ν	$Q(.005)$	$Q(.01)$	$Q(.025)$	$Q(.05)$	$Q(.1)$	$Q(.9)$	$Q(.95)$	$Q(.975)$	$Q(.99)$	$Q(.995)$
1	0.000	0.000	0.001	0.004	0.016	2.706	3.841	5.024	6.635	7.879
2	0.010	0.020	0.051	0.103	0.211	4.605	5.991	7.378	9.210	10.597
3	0.072	0.115	0.216	0.352	0.584	6.251	7.815	9.348	11.345	12.838
4	0.207	0.297	0.484	0.711	1.064	7.779	9.488	11.143	13.277	14.860
5	0.412	0.554	0.831	1.145	1.610	9.236	11.070	12.833	15.086	16.750
6	0.676	0.872	1.237	1.635	2.204	10.645	12.592	14.449	16.812	18.548
7	0.989	1.239	1.690	2.167	2.833	12.017	14.067	16.013	18.475	20.278
8	1.344	1.646	2.180	2.733	3.490	13.362	15.507	17.535	20.090	21.955
9	1.735	2.088	2.700	3.325	4.168	14.684	16.919	19.023	21.666	23.589
10	2.156	2.558	3.247	3.940	4.865	15.987	18.307	20.483	23.209	25.188
11	2.603	3.053	3.816	4.575	5.578	17.275	19.675	21.920	24.725	26.757
12	3.074	3.571	4.404	5.226	6.304	18.549	21.026	23.337	26.217	28.300
13	3.565	4.107	5.009	5.892	7.042	19.812	22.362	24.736	27.688	29.819
14	4.075	4.660	5.629	6.571	7.790	21.064	23.685	26.119	29.141	31.319
15	4.601	5.229	6.262	7.261	8.547	22.307	24.996	27.488	30.578	32.801
16	5.142	5.812	6.908	7.962	9.312	23.542	26.296	28.845	32.000	34.267
17	5.697	6.408	7.564	8.672	10.085	24.769	27.587	30.191	33.409	35.718
18	6.265	7.015	8.231	9.390	10.865	25.989	28.869	31.526	34.805	37.156
19	6.844	7.633	8.907	10.117	11.651	27.204	30.143	32.852	36.191	38.582
20	7.434	8.260	9.591	10.851	12.443	28.412	31.410	34.170	37.566	39.997
21	8.034	8.897	10.283	11.591	13.240	29.615	32.671	35.479	38.932	41.401
22	8.643	9.542	10.982	12.338	14.041	30.813	33.924	36.781	40.290	42.796
23	9.260	10.196	11.689	13.091	14.848	32.007	35.172	38.076	41.638	44.181
24	9.886	10.856	12.401	13.848	15.659	33.196	36.415	39.364	42.980	45.559
25	10.520	11.524	13.120	14.611	16.473	34.382	37.653	40.647	44.314	46.928
26	11.160	12.198	13.844	15.379	17.292	35.563	38.885	41.923	45.642	48.290
27	11.808	12.879	14.573	16.151	18.114	36.741	40.113	43.195	46.963	49.645
28	12.461	13.565	15.308	16.928	18.939	37.916	41.337	44.461	48.278	50.994
29	13.121	14.256	16.047	17.708	19.768	39.087	42.557	45.722	49.588	52.336
30	13.787	14.953	16.791	18.493	20.599	40.256	43.773	46.979	50.892	53.672
31	14.458	15.655	17.539	19.281	21.434	41.422	44.985	48.232	52.192	55.003
32	15.134	16.362	18.291	20.072	22.271	42.585	46.194	49.480	53.486	56.328
33	15.815	17.074	19.047	20.867	23.110	43.745	47.400	50.725	54.775	57.648
34	16.501	17.789	19.806	21.664	23.952	44.903	48.602	51.966	56.061	58.964
35	17.192	18.509	20.569	22.465	24.797	46.059	49.802	53.204	57.342	60.275
36	17.887	19.233	21.336	23.269	25.643	47.212	50.998	54.437	58.619	61.581
37	18.586	19.960	22.106	24.075	26.492	48.364	52.192	55.668	59.893	62.885
38	19.289	20.691	22.878	24.884	27.343	49.513	53.384	56.896	61.163	64.183
39	19.996	21.426	23.654	25.695	28.196	50.660	54.572	58.120	62.429	65.477
40	20.707	22.164	24.433	26.509	29.051	51.805	55.759	59.342	63.691	66.767