Curs 1

Cristian Niculescu

1 Introducere

Bibliografie:
Jeremy Orloff, Jonathan Bloom, *Introduction to Probability and Statistics*, MIT OpenCourseWare, 2014
Cristian Niculescu, *Probabilităţi şi statistică*, Editura Universităţii din Bucureşti, 2015

1.1 Probabilităţi vs. statistică

Probabilităţile şi statistica sunt strâns legate deoarece toate afirmaţiile statistice sunt la bază afirmaţii despre probabilităţi. În ciuda acestui fapt, cele 2 domenii se simt uneori ca subiecte foarte diferite. Probabilităţile sunt independente logic; sunt câteva reguli şi toate răspunsurile rezultă logic din aceste reguli, cu toate că pot fi calcule complicate. În statistică aplicăm probabilităţile pentru a trage concluzii din date.

Exemplu probabilistic

Avem o monedă corectă (probabilităţi egale pentru avers sau revers). O aruncăm de 100 de ori. Care este probabilitatea a cel puţin 60 de aversuri? Există un singur răspuns (aproximativ 0.02844397) şi vom învăţa cum să-l calculăm.

Exemplu statistic

Avem o monedă de provenienţă necunoscută. Pentru a investiga dacă este corectă o aruncăm de 100 de ori şi numărăm aversurile. Să spunem că numărăm 60 de aversuri. Sarcina noastră ca statisticieni este să tragem o concluzie (inferenţă) din aceste date. Sunt multe moduri de a proceda, atât în termeni de forma pe care o ia concluzia, cât şi în ce priveşte calculele probabilităţilor folosite pentru a justifica concluzia. De fapt, statisticieni diferiţi pot trage concluzii diferite.

Observăm că în primul exemplu procesul aleator este deplin cunoscut (probabilitatea aversului = 0.5). Obiectivul este să găsim probabilitatea unui anume rezultat (cel puţin 60 de aversuri) provenind din procesul aleator. În al 2-lea exemplu, rezultatul este cunoscut (60 de aversuri) şi obiectivul este să clarificăm procesul aleator necunoscut (probabilitatea aversului).

1.2 Interpretări frecvenţioniste vs. Bayesiene

Există 2 şcoli proeminente şi uneori conflictuale de statistică: Bayesiană şi frecvenţionistă. Abordările lor sunt bazate pe interpretări diferite ale sensului probabilităţii.
Frecvenţioniştii spun că probabilitatea măsoară frecvenţa diferitelor rezultate ale unui experiment. De exemplu, spunând că o monedă corectă are o probabilitate de 50% a aversurilor înseamnă că, dacă o vom arunca de multe ori, atunci ne aşteptăm ca aproximativ jumătate din aruncări să fie aversuri.
Bayesienii spun că probabilitatea este un concept abstract care măsoară o stare de cunoaştere sau un grad de încredere într-o propoziţie dată. În practică, Bayesienii nu atribuie o singură valoare pentru probabilitatea aversului unei monede. Mai degrabă ei consideră un domeniu de valori, fiecare cu propria probabilitate de a fi adevărată.
Abordarea frecvenţionistă a fost mult timp dominantă în domenii ca biologia, medicina, sănătatea publică şi ştiinţele sociale. Abordarea Bayesiană s-a bucurat de o renaştere în era computerelor performante şi datelor mari. Este utilă în special când se încorporează date noi într-un model statistic existent, de exemplu, când se antrenează un sistem de recunoaştere a vorbirii sau a feţei. Astăzi, statisticienii creează instrumente puternice folosind ambele abordări în moduri complementare.

1.3 Aplicaţii, modele didactice şi simulare

Probabilităţile şi statistica sunt larg folosite în ştiinţele naturii, inginerie, medicină, ştiinţele sociale, economie şi informatică. Lista aplicaţiilor este mare: teste ale unui tratament medical contra altuia (sau a unui placebo), măsuri ale legăturii genetice, căutarea particulelor elementare, învăţarea automată pentru vedere sau vorbire, strategii şi probabilităţi de jocuri de noroc, meteorologie, prognoză economică, epidemiologie, marketing, căutare pe net...
Date fiind atât de multe aplicaţii captivante, poate vă miraţi de ce vom consuma atât de mult timp gândindu-ne la modele didactice ca monede şi zaruri. Înţelegându-le pe acestea cu desăvârşire, vom dezvolta un bun simţ pentru fondul simplu interior multor probleme complexe din lumea reală. De fapt, modesta monedă este un model realist pentru orice situaţii cu 2 rezultate posibile: succes sau eşec al unui tratament, motor de avion, pariu, sau chiar al unui curs.

Uneori o problemă este atât de complicată încât cel mai bun mod de înţelegere a ei este prin simulare pe computer. Aici folosim software pentru a face experimente *virtuale* de multe ori pentru a estima probabilităţi. La acest curs vom folosi R pentru simulare, calcul şi vizualizare.

2 Numărare şi mulţimi

2.1 Scopurile învăţării

1. Să ştie definiţiile şi notaţiile pentru mulţimi, intersecţie, reuniune, complementară.
2. Să poată să vizualizeze operaţiile cu mulţimi folosind diagrame Venn.
3. Să înţeleagă cum este folosită numărarea la calculul probabilităţilor.
4. Să poată folosi regula produsului, principiul includerii şi al excluderii, permutările, aranjamentele şi combinările pentru a calcula numărul elementelor unei mulţimi.

2.2 Numărare

2.2.1 Întrebări motivante

Exemplul 1. O monedă este *corectă* dacă, aruncând-o, obţinem avers sau revers cu aceeaşi probabilitate. Aruncăm o monedă de 3 ori. Care este probabilitatea să obţinem exact un avers?
Răspuns: Cu 3 aruncări, putem lista uşor cele 8 cazuri posibile (avers = H, revers = T):

$$\{HHH, HHT, HTH, HTT, THH, THT, TTH, TTT\}.$$

3 din aceste cazuri au exact 1 avers:

$$\{HTT, THT, TTH\}.$$

Deoarece toate cazurile sunt egal probabile, avem

$$P(1 \text{ avers în 3 aruncări}) = \frac{\text{numărul de cazuri cu 1 avers}}{\text{numărul total de cazuri}} = \frac{3}{8}.$$

Gândiţi: Ar fi practică listarea cazurilor la 10 aruncări?

Un pachet de 52 de cărţi de joc are 13 ranguri (2,3,...,9,10,J,Q,K,A) şi 4 culori ($\clubsuit, \diamondsuit, \heartsuit, \spadesuit$). O mână de poker constă în 5 cărţi. *O pereche* este o mână care constă în 2 cărţi având un rang şi 3 cărţi având alte 3 ranguri, de

exemplu, $\{3\clubsuit, 3\diamondsuit, 5\heartsuit, 8\spadesuit, 10\spadesuit\}$.
Testaţi-vă intuiţia: probabilitatea unei perechi este:
(a) sub 5%
(b) între 5% şi 10%
(c) între 10% şi 20%
(d) între 20% şi 40%
(e) peste 40%.
În acest moment putem doar să ghicim această probabilitate. Unul din scopurile noastre este să învăţăm cum s-o calculăm exact. Pentru a începe, observăm că, deoarece fiecare mulţime de 5 cărţi este egal probabilă, putem calcula probabilitatea unei perechi astfel

$$P(\text{o pereche}) = \frac{\text{numărul de "o pereche"}}{\text{numărul total de mâini}}.$$

Deci, pentru a afla probabilitatea exactă, avem nevoie să calculăm numărul de elemente din fiecare din aceste mulţimi. Şi trebuie s-o facem iscusit, deoarece sunt prea multe elemente pentru a le lista pe toate.

Am observat deja de câteva ori că toate cazurile posibile erau egal probabile şi am folosit asta pentru a afla o probabilitate numărând. Afirmăm aceasta riguros în următorul principiu.
Principiu: Presupunem că sunt n cazuri posibile pentru un experiment şi fiecare caz este egal probabil. Dacă sunt k cazuri favorabile, atunci probabilitatea unui caz favorabil este k/n.
Vă puteţi gândi la un scenariu în care cazurile posibile nu sunt egal probabile?
Iată un scenariu: la un examen, puteţi lua orice notă de la 1 la 10. Sunt 10 cazuri. Este probabilitatea de a lua sub 5 egală cu 4/10?

2.2.2 Mulţimi şi notaţii

Definiţii
O mulţime S este o colecţie de elemente.
Element: Scriem $x \in S$ pentru faptul că elementul x este în mulţimea S.
Submulţime: Spunem că mulţimea A este submulţime a lui S dacă toate elementele lui A sunt în S. Scriem $A \subset S$.
Complementară: Complementara lui A faţă de S este mulţimea elementelor lui S care nu sunt în A. Scriem A^c, $\complement A$, $\overline{A}$ sau $S \setminus A$.
Reuniune: Reuniunea lui A şi B este mulţimea tuturor elementelor din A **sau** B (sau din ambele). Scriem $A \cup B$.
Intersecţie: Intersecţia lui A şi B este mulţimea tuturor elementelor din A **şi**

B. Scriem $A \cap B$.
Mulţimea vidă: Mulţimea vidă este mulţimea care nu are niciun element. O notăm cu $\emptyset$.
Disjuncte: A şi B sunt **disjuncte** dacă n-au niciun element comun. Adică $A \cap B = \emptyset$.
Diferenţă: Diferenţa lui A şi B este mulţimea elementelor din A care nu sunt în B. Scriem $A \setminus B$ sau $A - B$.
Exemplul 2. Plecăm cu o mulţime de 10 animale

$$S = \{\text{antilopă, albină, pisică, câine, elefant, broască, ţânţar, hienă, iguană, jaguar}\}.$$

Considerăm 2 submulţimi:

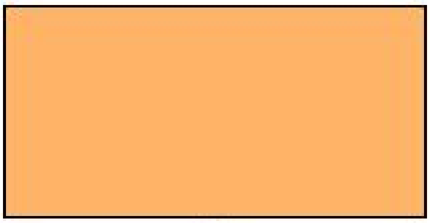
$$M = \{\text{antilopă, pisică, câine, elefant, hienă, jaguar}\}$$
$$W = \{\text{antilopă, albină, elefant, broască, ţânţar, hienă, iguană, jaguar}\}.$$

Scopul nostru aici este să cercetăm diferite operaţii cu mulţimi.
Intersecţie: $M \cap W = \{\text{antilopă, elefent, hienă, jaguar}\}$.
Reuniune:
$M \cup W = \{\text{antilopă, albină, pisică, câine, elefant, broască, ţânţar, hienă, iguană, jaguar}\}$.
Complementară: $M^c = \{\text{albină, broască, ţânţar, iguană}\}$.
Diferenţă: $M \setminus W = \{\text{pisică, câine}\}$.
Adesea sunt multe moduri de a obţine aceeaşi mulţime, de exemplu, $M^c = S \setminus M$, $M \setminus W = M \cap \complement W$.

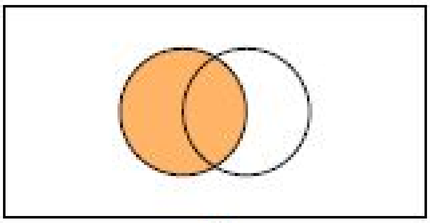
Relaţiile dintre reuniune, intersecţie şi complementară sunt date de legile lui DeMorgan:

$$(A \cup B)^c = A^c \cap B^c$$
$$(A \cap B)^c = A^c \cup B^c.$$

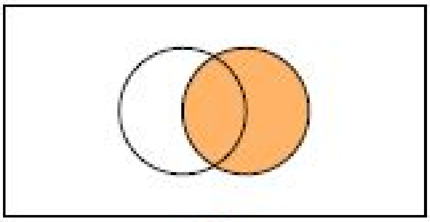
Diagrame Venn-Euler
Diagramele Venn-Euler dau un mod uşor de a vizualiza operaţiile cu mulţimi. În toate figurile, S este regiunea din interiorul dreptunghiului mare, L este regiunea din interiorul cercului din stânga şi R este regiunea din interiorul cercului din dreapta. Regiunea colorată arată mulţimea scrisă sub fiecare figură.



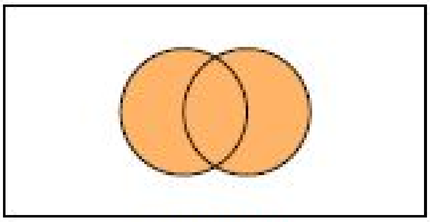
S



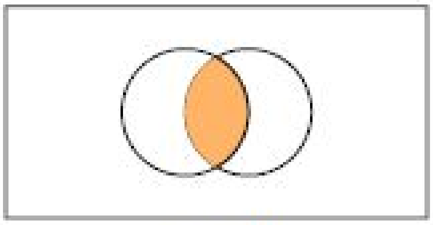
L



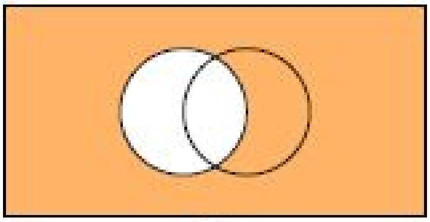
R



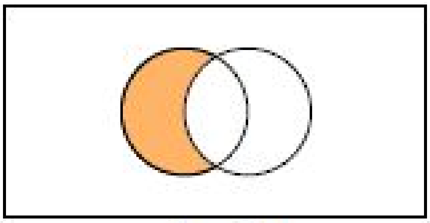
$L \cup R$



$L \cap R$

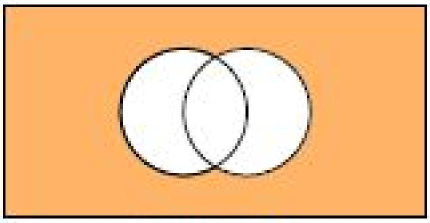


L^c

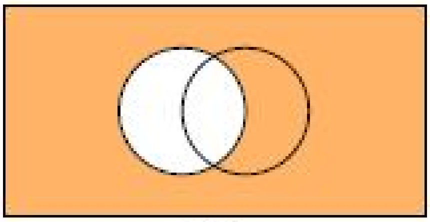


$L - R$

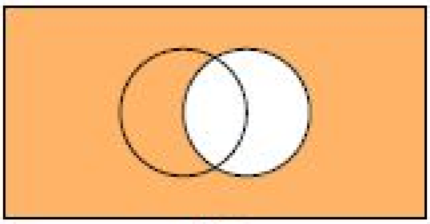
Demonstraţia legilor lui DeMorgan



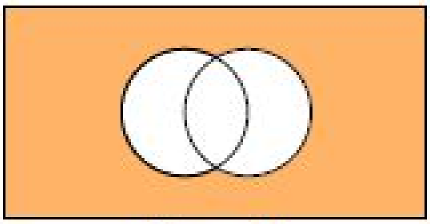
$(L \cup R)^c$



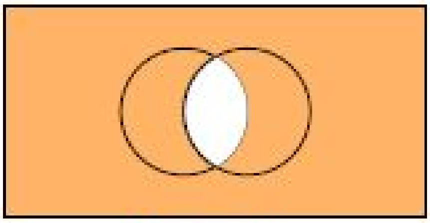
L^c



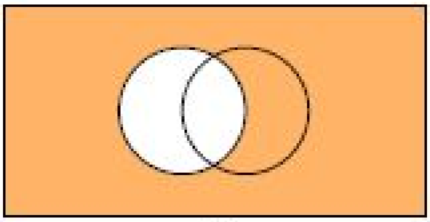
R^c



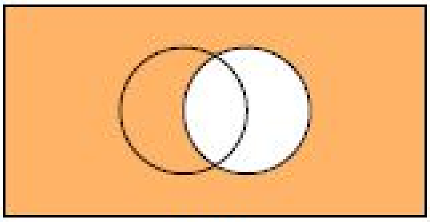
$L^c \cap R^c$



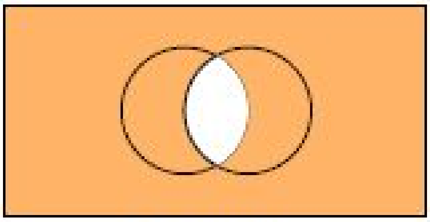
$(L \cap R)^c$



L^c



R^c



$L^c \cup R^c$

Exemplul 3. Verificaţi legile lui DeMorgan pentru submulţimile $A = \{1, 2, 3\}$ şi $B = \{3, 4\}$ ale mulţimii $S = \{1, 2, 3, 4, 5\}$.
Răspuns: Pentru fiecare lege, calculăm ambii membri ai relaţiei şi arătăm că sunt egali.
1. $(A \cup B)^c = A^c \cap B^c$:
Membrul stâng: $A \cup B = \{1, 2, 3, 4\} \Rightarrow (A \cup B)^c = \{5\}$.
Membrul drept: $A^c = \{4, 5\}$, $B^c = \{1, 2, 5\} \Rightarrow A^c \cap B^c = \{5\}$.
Cei 2 membri sunt egali, q.e.d.
2. $(A \cap B)^c = A^c \cup B^c$:
Membrul stâng: $A \cap B = \{3\} \Rightarrow (A \cap B)^c = \{1, 2, 4, 5\}$.
Membrul drept: $A^c = \{4, 5\}$, $B^c = \{1, 2, 5\} \Rightarrow A^c \cup B^c = \{1, 2, 4, 5\}$.
Cei 2 membri sunt egali, q.e.d.

Produsul cartezian

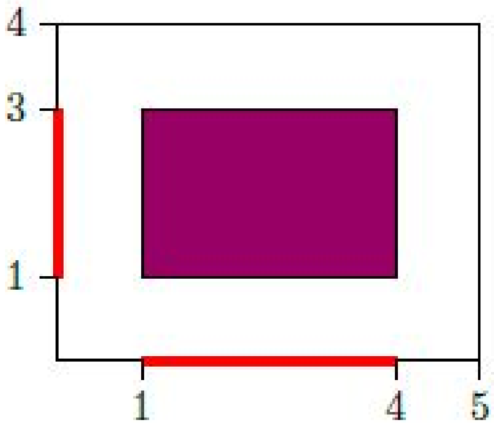
Produsul cartezian al mulţimilor S şi T este mulţimea perechilor ordonate:

$$S \times T = \{(s, t) | s \in S, t \in T\}.$$

În cuvinte, membrul drept se citeşte "mulţimea perechilor ordonate (s, t) cu proprietatea că s este în S şi t este în T".
Următoarele diagrame arată 2 exemple de produs cartezian.

×	1	2	3	4
1	(1,1)	(1,2)	(1,3)	(1,4)
2	(2,1)	(2,2)	(2,3)	(2,4)
3	(3,1)	(3,2)	(3,3)	(3,4)

$\{1, 2, 3\} \times \{1, 2, 3, 4\}$



$[1, 4] \times [1, 3] \subset [0, 5] \times [0, 4]$

Ultima figură ilustrează de asemenea faptul că dacă $A \subset S$ şi $B \subset T$, atunci $A \times B \subset S \times T$.

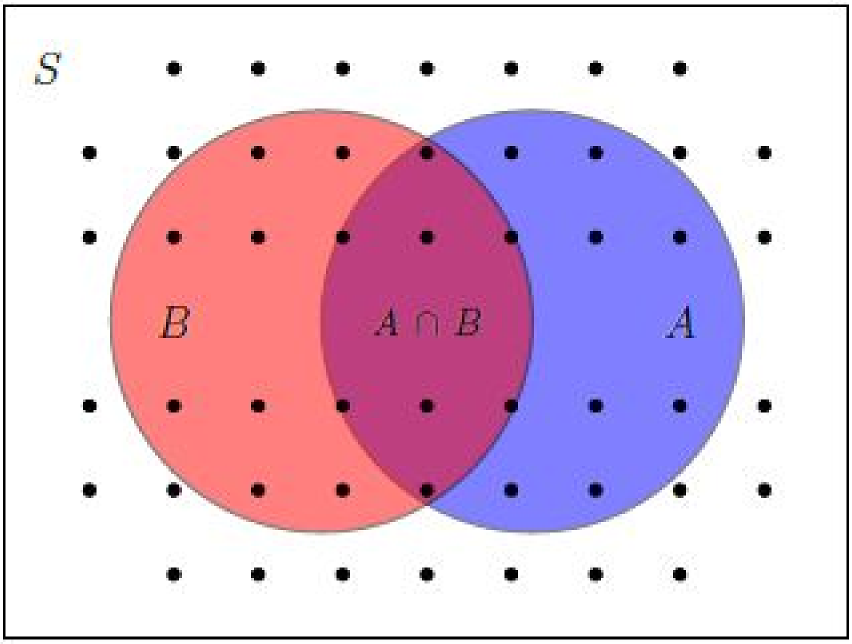
2.2.3 Numărare

Dacă S este finită, folosim $|S|$ sau $\#S$ pentru a nota numărul elementelor lui S (cardinalul lui S).
2 principii de numărare utile sunt *principiul includerii şi excluderii* şi *regula produsului*.

Principiul includerii şi excluderii
Principiul includerii şi excluderii spune

$$|A \cup B| = |A| + |B| - |A \cap B|.$$

Putem ilustra asta printr-o diagramă Venn-Euler. S este mulţimea tuturor punctelor, A este mulţimea punctelor din cercul albastru, iar B este mulţimea punctelor din cercul roşu.



$|A|$ este numărul de puncte din A şi analog pentru celelalte mulţimi. Figura arată că $|A| + |B|$ numără *de 2 ori* $|A \cap B|$, de aceea $|A \cap B|$ este scăzut în formula includerii şi excluderii.

Exemplul 4. Într-o trupă de cântăreţi şi chitarişti sunt 7 cântăreţi, 4 chitarişti şi 2 ambele. Cât de mare este trupa?

Răspuns: Fie S mulţimea cântăreţilor şi G mulţimea chitariştilor. Principiul includerii şi excluderii spune că

$$\text{mărimea trupei} = |S \cup G| = |S| + |G| - |S \cap G| = 7 + 4 - 2 = 9.$$

Regula produsului

Regula produsului spune:

Dacă există n moduri de a face acţiunea 1 şi apoi m moduri de a face acţiunea 2, atunci există $n \cdot m$ moduri de a face acţiunea 1 urmată de acţiunea 2.

O numim de asemenea regula multiplicării.

Exemplul 5. Dacă avem 3 cămăşi şi 4 pantaloni, atunci putem face $3 \cdot 4 = 12$ costumaţii.

Gândiţi: Un punct extrem de important este că regula produsului are loc chiar dacă modurile de a face acţiunea 2 depind de acţiunea 1, atât timp cât *numărul* de moduri de a face acţiunea 2 este independent de acţiunea 1. Pentru a ilustra asta:

Exemplul 6. Există 5 competitori în finala de 100m la olimpiadă. În câte moduri pot fi date medaliile de aur, argint şi bronz? (Presupunem că

egalităţile sunt excluse.)
Răspuns: Există 5 moduri de a da medalia de aur. Odată ce aceasta este dată, există 4 moduri de a da medalia de argint şi apoi 3 moduri de a da medalia de bronz. Răspuns $5 \cdot 4 \cdot 3 = 60$ de moduri.
Observăm că alegerea medaliatului cu aur afectează câştigătorul argintului, dar numărul posibililor medaliaţi cu argint este totdeauna 4.

2.2.4 Permutări, aranjamente şi combinări

Permutări şi aranjamente
O permutare a unei mulţimi este o ordonare particulară a elementelor sale. De exemplu, mulţimea $\{a, b, c\}$ are 6 permutări: $abc, acb, bac, bca, cab, cba$. Am aflat numărul permutărilor listându-le pe toate. Puteam de asemenea să aflăm numărul permutărilor folosind regula produsului. Adică, există 3 moduri de a alege primul element, apoi 2 moduri pentru al 2-lea şi 1 pentru al 3-lea. Asta dă un total de $3 \cdot 2 \cdot 1 = 6$ permutări.
În general, regula produsului ne spune că numărul de permutări ale unei mulţimi de k elemente este

$$k! = k \cdot (k-1) \cdot ... \cdot 3 \cdot 2 \cdot 1.$$

Vorbim de asemenea de aranjamentele a k elemente dintr-o mulţime cu n elemente. Arătăm ce înseamnă asta cu un exemplu.
Exemplul 7. Listaţi toate aranjamentele de 3 elemente ale mulţimii $\{a, b, c, d\}$.
Răspuns: Este o listă mai lungă,

abc	acb	bac	bca	cab	cba
abd	adb	bad	bda	dab	dba
acd	adc	cad	cda	dac	dca
bcd	bdc	cbd	cdb	dbc	dcb

Observăm că abc şi acb se numără ca aranjamente distincte. Adică, pentru aranjamente ordinea contează.
Există 24 de aranjamente. Observăm că regula produsului ne-ar fi spus că sunt $4 \cdot 3 \cdot 2 = 24$ permutări fără a ne mai deranja să le listăm.

Combinări
În contrast cu aranjamentele, la combinări ordinea nu contează: aranjamentele sunt liste, iar combinările sunt mulţimi. Arătăm ce vrem să spunem printr-un exemplu.
Exemplul 8. Listaţi toate combinările de 3 elemente din mulţimea $\{a, b, c, d\}$.
Răspuns: O astfel de combinare este o colecţie de 3 elemente fără a consi-

dera ordinea. Astfel, *abc* şi *cab* reprezintă ambele aceeaşi combinare. Putem lista toate combinările listând toate submulţimile de exact 3 elemente.

$$\{a, b, c\}\ \{a, b, d\}\ \{a, c, d\}\ \{b, c, d\}.$$

Există doar 4 combinări, în contrast cu cele 24 de aranjamente din exemplul anterior. Factorul 6 apare deoarece fiecare combinare de 3 elemente poate fi scrisă în 6 ordini diferite.

Formule

Folosim următoarele notaţii.

A_n^k = numărul aranjamentelor (listelor) de k elemente distincte dintr-o mulţime cu n elemente.

$C_n^k = \binom{n}{k}$ = numărul combinărilor (submulţimilor) de k elemente dintr-o mulţime cu n elemente.

Subliniem că prin numărul de combinări de k elemente înţelegem numărul de submulţimi de k elemente.

Acestea au următoarele notaţii şi formule:

$$\text{Aranjamente: } A_n^k = \frac{n!}{(n-k)!} = n(n-1)...(n-k+1).$$

$$\text{Combinări: } C_n^k = \frac{n!}{k!(n-k)!} = \frac{A_n^k}{k!}.$$

Notaţia A_n^k se citeşte "aranjamente de n luate câte k". Notaţia C_n^k se citeşte "combinări de n luate câte k". Formula pentru A_n^k rezultă din regula produsului. Ea implică de asemenea formula pentru C_n^k, deoarece o submulţime de k elemente poate fi ordonată în $k!$ moduri.

Putem ilustra relaţia dintre aranjamente şi combinări aliniind rezultatele din cele 2 exemple anterioare.

abc	*acb*	*bac*	*bca*	*cab*	*cba*	$\{a, b, c\}$
abd	*adb*	*bad*	*bda*	*dab*	*dba*	$\{a, b, d\}$
acd	*adc*	*cad*	*cda*	*dac*	*dca*	$\{a, c, d\}$
bcd	*bdc*	*cbd*	*cdb*	*dbc*	*dcb*	$\{b, c, d\}$

Aranjamente: A_4^3 Combinări: C_4^3

Observăm că fiecare linie din lista de aranjamente constă din toate cele 3! permutări ale mulţimii corespunzătoare din lista de combinări.

Exemple

Exemplul 9. Calculaţi:

(i) Numărul de moduri de a alege 2 din 4 obiecte (ordinea nu contează).
(ii) Numărul de moduri de a lista 2 din 4 obiecte.
(iii) Numărul de moduri de a alege 3 din 10 obiecte.
Răspuns: (i) $C_4^2 = \frac{4!}{2! \cdot 2!} = 6$.
(ii) $A_4^2 = \frac{4!}{2!} = 12$.
(iii) $C_{10}^3 = \frac{10!}{3! \cdot 7!} = \frac{10 \cdot 9 \cdot 8}{3 \cdot 2 \cdot 1} = 120$.
Exemplul 10. (i) Calculaţi numărul de moduri de a obţine 3 aversuri într-o secvenţă de 10 aruncări ale unei monede.
(ii) Dacă moneda este corectă, care este probabilitatea a exact 3 aversuri din 10 aruncări?
Răspuns: (i) Se cere numărul de secvenţe de 10 aruncări (aversuri sau reversuri) cu exact 3 aversuri. Adică, trebuie să alegem exact 3 din 10 aruncări să fie aversuri. Asta este aceeaşi întrebare ca ultima din exemplul precedent.

$$C_{10}^3 = \frac{10!}{3! \cdot 7!} = \frac{10 \cdot 9 \cdot 8}{3 \cdot 2 \cdot 1} = 120.$$

(ii) Fiecare aruncare are 2 cazuri posibile (avers sau revers). Astfel, regula produsului spune că există $2^{10} = 1024$ secvenţe de 10 aruncări. Deoarece moneda este corectă, fiecare secvenţă este egal probabilă. Deci probabilitatea a 3 aversuri este

$$\frac{120}{1024} = 0.1171875.$$