

Curs 5

Cristian Niculescu

1 Demonstrații

1.1 Introducere

Dăm mai mult material matematic formal pentru a sublinia că atunci când facem matematică ar trebui să avem grijă să specificăm complet ipotezele noastre și să dăm argumente deductive clare pentru a demonstra afirmațiile noastre.

1.2 Cu probabilitate mare histograma de densitate seamănă cu graficul funcției densitate de probabilitate

Am afirmat că o consecință a legii numerelor mari este că, atunci când numărul datelor crește, histograma de densitate a datelor are o probabilitate crescătoare de a se potrivi cu graficul pdf sau pmf. Aceasta este o bună regulă a degetului mare, dar este mai degrabă imprecisă. Putem face afirmații mai precise.

Presupunem că avem un experiment care produce date conform variabilei aleatoare X și presupunem că generăm n date independente din X . Le notăm

$$x_1, x_2, \dots, x_n.$$

Printr-un coș înțelegem o mulțime de valori, de exemplu $(b_k, b_{k+1}]$. Pentru a face o histogramă de densitate împărțim domeniul de valori al lui X în m coșuri și calculăm fracția din date din fiecare coș.

Fie p_k probabilitatea ca o dată aleatoare să fie în al k -lea coș. Aceasta este probabilitatea pentru o variabilă aleatoare (Bernoulli) indicator $B_{k,j}$, care este 1 dacă a j -a dată este în coș și 0 altfel.

Propoziția 1. Fie $\bar{p}_k = (\text{numărul datelor din coșul } k)/n$. Când numărul n al datelor devine mare, probabilitatea că \bar{p}_k este aproape de p_k se apropie de 1. Altfel spus, fiind dat un număr pozitiv oricât de mic, să-l numim a ,

probabilitatea $P(|\bar{p}_k - p_k| < a)$ depinde de n și, când $n \rightarrow \infty$, această probabilitate tinde la 1.

Demonstrație. Fie \bar{B}_k media lui $B_{k,j}$. Deoarece $E(B_{k,j}) = p_k$, legea numerelor mari spune exact că

$$P(|\bar{B}_k - p_k| < a) \rightarrow 1 \text{ când } n \rightarrow \infty.$$

Dar, deoarece $B_{k,j}$ sunt variabile indicator, media lor este exact \bar{p}_k . Înlocuirea lui \bar{B}_k cu \bar{p}_k în relația de mai sus dă

$$P(|\bar{p}_k - p_k| < a) \rightarrow 1 \text{ când } n \rightarrow \infty.$$

Aceasta este exact ce s-a cerut în propoziția 1.

Propoziția 2. Aceeași propoziție este valabilă simultan pentru un număr finit de coșuri. Adică, pentru coșurile de la 1 la m avem

$$P(|\bar{B}_1 - p_1| < a, |\bar{B}_2 - p_2| < a, \dots, |\bar{B}_m - p_m| < a) \rightarrow 1 \text{ când } n \rightarrow \infty.$$

Demonstrație. Mai întâi scriem următoarea regulă de probabilitate, care este o consecință a principiului includerii și excluderii: dacă 2 evenimente A și B au $P(A) = 1 - \alpha_1$ și $P(B) = 1 - \alpha_2$, atunci $P(A \cap B) \geq 1 - (\alpha_1 + \alpha_2)$. Acum, propoziția 1 spune că $\forall \alpha > 0$, de la un n suficient de mare avem $P(|\bar{B}_k - p_k| < a) > 1 - \alpha/m$ separat pentru fiecare coș. Din regula de probabilitate, probabilitatea intersecției tuturor acestor evenimente este cel puțin $1 - \alpha$. Deoarece putem lua $\alpha > 0$ oricât de mic vrem, lăsând n să tindă la ∞ , la limită obținem probabilitatea 1 așa cum am afirmat.

Propoziția 3. Dacă f este o densitate de probabilitate continuă cu domeniul de valori $[a, b]$, atunci, luând suficiente date și având lățimea coșurilor suficient de mică, putem garanta că, cu mare probabilitate, histograma de densitate este cât de aproape vrem de graficul lui f .

Demonstrație. (Schită.) Presupunem că lățimea coșului din jurul lui x este Δx . Dacă Δx este suficient de mic, atunci probabilitatea că o dată este în coș este aproximativ $f(x)\Delta x$. Propoziția 1 garantează că, dacă n este suficient de mare, atunci, cu probabilitate mare, (numărul datelor din coș)/ n este de asemenea aproximativ $f(x)\Delta x$. Deoarece aceasta este aria coșului în histogramă, vedem că înălțimea coșului va fi aproximativ $f(x)$. Adică, cu mare probabilitate, înălțimea histogramei peste orice punct x este aproape de $f(x)$. Aceasta este ceea ce a afirmat propoziția 3.

Observație. Dacă domeniul de valori este nemărginit sau densitatea tinde la ∞ în vreun punct, trebuie să fim mai atenți. Există rezultate și pentru aceste cazuri.

1.3 Inegalitatea lui Cebâșev

O demonstrație a LoLN rezultă din următoarea inegalitate cheie.

Inegalitatea lui Cebâșev. Presupunem că Y este o variabilă aleatoare cu media μ și dispersia σ^2 . Atunci, $\forall a > 0$ avem

$$P(|Y - \mu| \geq a) \leq \frac{Var(Y)}{a^2}.$$

În cuvinte, inegalitatea lui Cebâșev spune că probabilitatea ca Y să difere de medie cu cel puțin a este mărginită superior de $Var(Y)/a^2$. Practic, cu cât mai mică este dispersia lui Y , cu atât mai mică este probabilitatea că Y este departe de media lui.

Demonstrația legii numerelor mari. Deoarece $Var(\bar{X}_n) = Var(X)/n$, $\lim_{n \rightarrow \infty} Var(\bar{X}_n) = 0$. Astfel, din inegalitatea lui Cebâșev pentru $Y = \bar{X}_n$ și a fixat $\implies \lim_{n \rightarrow \infty} P(|\bar{X}_n - \mu| \geq a) = 0 \implies \lim_{n \rightarrow \infty} P(|\bar{X}_n - \mu| < a) = 1$, ceea ce este LoLN.

Demonstrația inegalității lui Cebâșev. Demonstrația este în esență aceeași pentru Y discretă și Y continuă. Presupunem Y continuă și $\mu = 0$, deoarece înlocuind Y cu $Y - \mu$, dispersia nu se schimbă. Astfel,

$$\begin{aligned} P(|Y| \geq a) &= \int_{-\infty}^{-a} f(y)dy + \int_a^{\infty} f(y)dy \leq \int_{-\infty}^{-a} \frac{y^2}{a^2} f(y)dy + \int_a^{\infty} \frac{y^2}{a^2} f(y)dy \\ &\leq \int_{-\infty}^{\infty} \frac{y^2}{a^2} f(y)dy = \frac{Var(Y)}{a^2}. \end{aligned}$$

Prima inegalitate utilizează faptul că $y^2/a^2 \geq 1$ pe intervalele de integrare. A 2-a inegalitate rezultă deoarece includerea intervalului $[-a, a]$ nu micșorează integrala, deoarece integrantul este nenegativ.

1.4 Nevoia de dispersie

Am trecut cu vederea un fapt tehnic. Peste tot am presupus că repartițiile aveau dispersie. De exemplu, demonstrația legii numerelor mari a folosit dispersia prin intermediul inegalității lui Cebâșev. Dar sunt repartiții care nu au dispersie deoarece suma sau integrala pentru dispersie nu converge. Pentru astfel de repartiții legea numerelor mari poate fi falsă.

2 Repartiții comune, independență

2.1 Scopurile învățării

1. Să înțeleagă ce înseamnă o pmf, pdf și cdf **comună** a 2 variabile aleatoare.
2. Să poată calcula probabilități și marginale dintr-o pmf sau pdf comună.
3. Să poată testa dacă 2 variabile aleatoare sunt independente.

2.2 Introducere

În știință și viața reală, suntem adesea interesați de 2 (sau mai multe) variabile aleatoare în același timp. De exemplu, putem măsura înălțimea și greutatea girafelor, sau IQ-ul și greutatea la naștere a copiilor, sau frecvența exercițiilor și rata bolilor de inimă la adulți, sau nivelul poluării aerului și rata bolilor respiratorii în orașe, sau numărul de prieteni pe Facebook și vârsta membrilor Facebook.

Gândiți: Ce relație ne-am aștepta să fie în fiecare din cele 5 exemple de mai sus? De ce?

În astfel de situații variabilele aleatoare au o **repartiție comună** care ne permite să calculăm probabilități ale evenimentelor implicând ambele variabile și să înțelegem relația dintre variabile. Cel mai simplu este când variabilele sunt **independente**. Când nu sunt, folosim **covarianța** și **corelația** ca măsuri ale naturii dependenței dintre ele.

2.3 Repartiția comună

2.3.1 Cazul discret

Presupunem că X și Y sunt 2 variabile aleatoare discrete și că X ia valorile din $\{x_1, x_2, \dots, x_n\}$, iar Y ia valorile din $\{y_1, y_2, \dots, y_m\}$. Perechea ordonată (X, Y) ia valorile din produsul cartezian $\{(x_1, y_1), (x_1, y_2), \dots, (x_n, y_m)\}$. **Funcția masă de probabilitate comună** (pmf comună) a lui X și Y este funcția

$$p(x_i, y_j) = P(X = x_i, Y = y_j), i = \overline{1, n}, j = \overline{1, m}.$$

O organizăm într-un [tabel de probabilitate comună](#):

$X \backslash Y$	y_1	y_2	\dots	y_j	\dots	y_m
x_1	$p(x_1, y_1)$	$p(x_1, y_2)$	\dots	$p(x_1, y_j)$	\dots	$p(x_1, y_m)$
x_2	$p(x_2, y_1)$	$p(x_2, y_2)$	\dots	$p(x_2, y_j)$	\dots	$p(x_2, y_m)$
\dots	\dots	\dots	\dots	\dots	\dots	\dots
\dots	\dots	\dots	\dots	\dots	\dots	\dots
x_i	$p(x_i, y_1)$	$p(x_i, y_2)$	\dots	$p(x_i, y_j)$	\dots	$p(x_i, y_m)$
\dots	\dots	\dots	\dots	\dots	\dots	\dots
x_n	$p(x_n, y_1)$	$p(x_n, y_2)$	\dots	$p(x_n, y_j)$	\dots	$p(x_n, y_m)$

Exemplul 1. Aruncăm 2 zaruri corecte. Fie X numărul de pe primul zar și Y numărul de pe cel de-al 2-lea. Atunci atât X cât și Y iau valori de la 1 la 6, iar pmf comună este $p(i, j) = 1/36, \forall i, j = \overline{1, 6}$. Iată tabelul de probabilitate comună:

$X \backslash Y$	1	2	3	4	5	6
1	1/36	1/36	1/36	1/36	1/36	1/36
2	1/36	1/36	1/36	1/36	1/36	1/36
3	1/36	1/36	1/36	1/36	1/36	1/36
4	1/36	1/36	1/36	1/36	1/36	1/36
5	1/36	1/36	1/36	1/36	1/36	1/36
6	1/36	1/36	1/36	1/36	1/36	1/36

Exemplul 2. Aruncăm 2 zaruri corecte. Fie X numărul de pe primul zar și T totalul de pe cele 2 zaruri. Iată tabelul de probabilitate comună:

$X \backslash T$	2	3	4	5	6	7	8	9	10	11	12
1	1/36	1/36	1/36	1/36	1/36	1/36	0	0	0	0	0
2	0	1/36	1/36	1/36	1/36	1/36	1/36	0	0	0	0
3	0	0	1/36	1/36	1/36	1/36	1/36	1/36	0	0	0
4	0	0	0	1/36	1/36	1/36	1/36	1/36	1/36	0	0
5	0	0	0	0	1/36	1/36	1/36	1/36	1/36	1/36	0
6	0	0	0	0	0	1/36	1/36	1/36	1/36	1/36	1/36

O funcție masă de probabilitate comună trebuie să satisfacă 2 proprietăți:

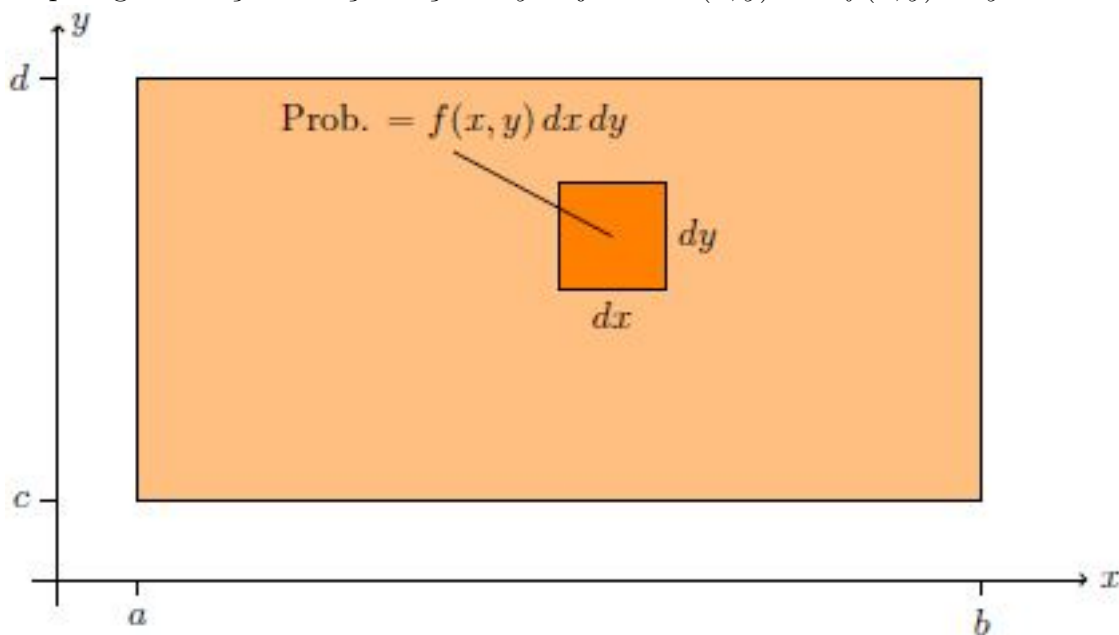
1. $0 \leq p(x_i, y_j) \leq 1, \forall i = \overline{1, n}, j = \overline{1, m}$.
2. Probabilitatea totală este 1. Putem exprima aceasta cu o **sumă dublă**:

$$\sum_{i=1}^n \sum_{j=1}^m p(x_i, y_j) = 1.$$

2.3.2 Cazul continuu

Cazul continuu este în esență la fel: doar înlocuim mulțimile discrete de valori cu intervale, funcția masă de probabilitate comună cu o **funcție densitate de probabilitate comună** și sumele cu integrale.

Dacă X ia valori în $[a, b]$ și Y ia valori în $[c, d]$, atunci perechea (X, Y) ia valori în produsul cartezian $[a, b] \times [c, d]$. **Funcția densitate de probabilitate comună** (pdf comună) a lui X și Y este o funcție $f(x, y)$ dând densitatea de probabilitate în (x, y) . Adică, probabilitatea că (X, Y) este într-un mic dreptunghi de lățime dx și înălțime dy în jurul lui (x, y) este $f(x, y)dx dy$.



O funcție densitate comună de probabilitate trebuie să satisfacă 2 proprietăți:

1. $f(x, y) \geq 0, \forall x, y$.
2. Probabilitatea totală este 1. Exprimăm aceasta cu o **integrală dublă**:

$$\int_c^d \int_a^b f(x, y) dx dy = 1.$$

Observație: ca la pdf a unei singure variabile aleatoare, pdf comună $f(x, y)$ poate lua valori > 1 ; este o densitate de probabilitate, **nu** o probabilitate.

- Ar trebui să puteți calcula integrale duble pe dreptunghiuri.
- Pe o regiune nedreptunghiulară, când $f(x, y) = c$ este constantă, integrala dublă este egală cu $c \cdot (\text{aria regiunii})$.

2.3.3 Evenimente

Variabilele aleatoare sunt utile pentru a descrie evenimente. Reamintim că un eveniment este o mulțime de rezultate și că variabilele aleatoare atribuie numere rezultatelor. De exemplu, evenimentul " $X > 1$ " este mulțimea tuturor rezultatelor pentru care X este > 1 . Aceste concepte se extind la perechi de variabile aleatoare și rezultate comune.

Exemplul 3. În exemplul 1, descrieți evenimentul $B = "Y - X \geq 2"$ și aflați-i probabilitatea.

Răspuns. Putem descrie B ca o mulțime de perechi (x, y) :

$$B = \{(1, 3), (1, 4), (1, 5), (1, 6), (2, 4), (2, 5), (2, 6), (3, 5), (3, 6), (4, 6)\}.$$

Putem să-l descriem și vizual:

$X \backslash Y$	1	2	3	4	5	6
1	1/36	1/36	1/36	1/36	1/36	1/36
2	1/36	1/36	1/36	1/36	1/36	1/36
3	1/36	1/36	1/36	1/36	1/36	1/36
4	1/36	1/36	1/36	1/36	1/36	1/36
5	1/36	1/36	1/36	1/36	1/36	1/36
6	1/36	1/36	1/36	1/36	1/36	1/36

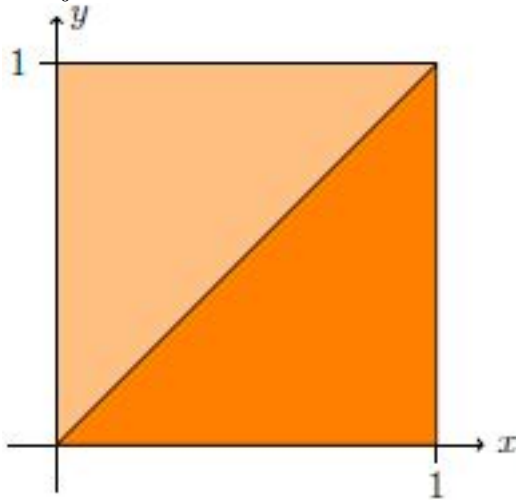
Evenimentul B constă din rezultatele corespunzătoare dreptunghiurilor portocalii.

Probabilitatea lui B este suma probabilităților din dreptunghiurile portocalii, astfel, $P(B) = 10/36 = 5/18$.

Exemplul 4. Presupunem că atât X cât și Y iau valori în $[0, 1]$ cu densitatea uniformă $f(x, y) = 1$. Vizualizați evenimentul " $X > Y$ " și aflați-i probabilitatea.

Răspuns. În comun, X și Y iau valori în pătratul unitate. Evenimentul

" $X > Y$ " corespunde triunghiului mai închis din dreapta-jos din figura de mai jos.

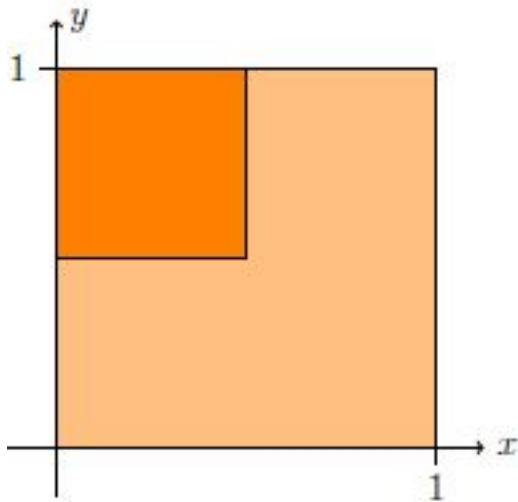


Evenimentul " $X > Y$ " în pătratul unitate.

Deoarece densitatea este constantă, probabilitatea este chiar fracția din aria totală ocupată de eveniment. În acest caz este 0.5.

Exemplul 5. Presupunem că atât X cât și Y iau valori în $[0,1]$ cu densitatea $f(x, y) = 4xy$. Arătați că f este o pdf comună validă, vizualizați evenimentul $A = "X < 0.5$ și $Y > 0.5"$ și aflați-i probabilitatea.

Răspuns. În comun, X și Y iau valori în pătratul unitate.



Evenimentul A în pătratul unitate.

Pentru a arăta că f este o pdf validă trebuie să verificăm că este nenegativă

(ceea ce este clar) și că probabilitatea totală este 1.

$$\begin{aligned}\text{Probabilitatea totală} &= \int_0^1 \int_0^1 4xy dx dy = \int_0^1 2x^2 y \Big|_{x=0}^{x=1} dy = \int_0^1 2y dy \\ &= y^2 \Big|_0^1 = 1, \text{q.e.d.}\end{aligned}$$

Evenimentul A este chiar pătratul din stânga sus. Deoarece densitatea nu este constantă, trebuie să calculăm o integrală pentru a afla probabilitatea.

$$P(A) = \int_0^{0.5} \int_{0.5}^1 4xy dy dx = \int_0^{0.5} 2xy^2 \Big|_{y=0.5}^{y=1} dx = \int_0^{0.5} \frac{3x}{2} dx = \frac{3x^2}{4} \Big|_0^{0.5} = \frac{3}{16}.$$

2.3.4 Funcția de repartiție comună

Presupunem că X și Y sunt variabile aleatoare repartizate comun. Folosim notația " $X \leq x, Y \leq y$ " pentru evenimentul " $X \leq x$ și $Y \leq y$ ". [Funcția de repartiție comună](#) (cdf comună) este definită astfel

$$F(x, y) = P(X \leq x, Y \leq y)$$

Cazul continuu: Dacă X și Y sunt variabile aleatoare continue cu densitatea comună $f(x, y)$ pe domeniul de valori $[a, b] \times [c, d]$, atunci cdf comună este dată de integrala dublă

$$F(x, y) = \int_c^y \int_a^x f(u, v) du dv.$$

Pentru a recupera pdf comună, derivăm cdf comună. Deoarece sunt 2 variabile, avem nevoie să folosim derivate parțiale:

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

Cazul discret: Dacă X și Y sunt variabile aleatoare discrete cu pmf comună $p(x_i, y_j)$, atunci cdf comună este dată de suma dublă

$$F(x, y) = \sum_{x_i \leq x} \sum_{y_j \leq y} p(x_i, y_j).$$

2.3.5 Proprietățile cdf comună

Cdf comună $F(x, y)$ a lui X și Y trebuie să satisfacă proprietățile:

1. F este crescătoare: i.e. dacă x sau y cresc, atunci F trebuie să rămână constantă sau să crească.

2. $F(x, y) = 0$ la stânga jos a domeniului de valori comun.

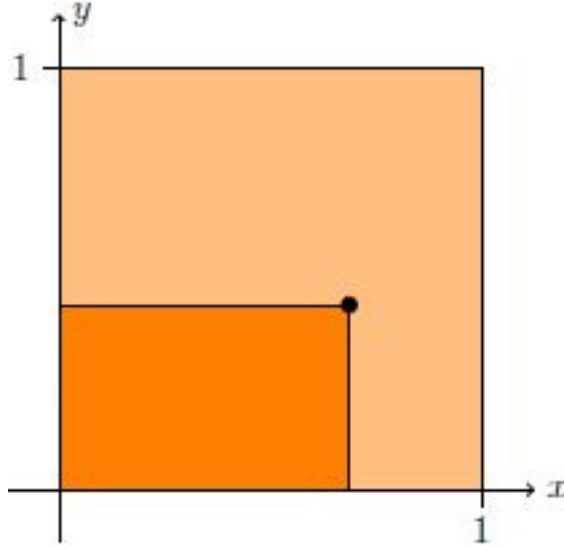
$$\lim_{(x,y) \rightarrow (-\infty, -\infty)} F(x, y) = 0.$$

3. $F(x, y) = 1$ la dreapta sus a domeniului de valori comun.

$$\lim_{(x,y) \rightarrow (\infty, \infty)} F(x, y) = 1.$$

Exemplul 6. Aflați cdf comună pentru variabilele aleatoare din exemplul 5.

Răspuns. Evenimentul " $X \leq x$ și $Y \leq y$ " este un dreptunghi în pătratul unitate.



Pentru a afla cdf F , calculăm o integrală dublă:

$$F(x, y) = \int_0^y \int_0^x 4uv \, du \, dv = \int_0^y 2u^2 v \Big|_{u=0}^{u=x} dv = \int_0^y 2x^2 v \, dv = x^2 v^2 \Big|_{v=0}^{v=y} = x^2 y^2.$$

Exemplul 7. În exemplul 1, calculați $F(3.5, 4)$.

Răspuns. Redesenăm tabelul de probabilitate comună. Figura este similară cu cea din exemplul precedent.

$F(3.5, 4)$ este probabilitatea evenimentului " $X \leq 3.5$ și $Y \leq 4$ ". Putem vizualiza acest eveniment ca dreptunghiurile colorate din tabel:

$X \backslash Y$	1	2	3	4	5	6
1	1/36	1/36	1/36	1/36	1/36	1/36
2	1/36	1/36	1/36	1/36	1/36	1/36
3	1/36	1/36	1/36	1/36	1/36	1/36
4	1/36	1/36	1/36	1/36	1/36	1/36
5	1/36	1/36	1/36	1/36	1/36	1/36
6	1/36	1/36	1/36	1/36	1/36	1/36

Evenimentul " $X \leq 3.5$ și $Y \leq 4$ ".

Adunând probabilitățile din dreptunghiurile colorate obținem

$$F(3.5, 4) = 12/36 = 1/3.$$

Observație. O diferență nefericită între vizualizările continuă și discretă este că pentru variabile continue valoarea crește când mergem în sus în direcția verticală în timp ce contrarul este adevărat pentru cazul discret.

2.3.6 Repartiții marginale

Când X și Y sunt variabile aleatoare repartizate comun, putem considera doar una din ele, să zicem X . În acest caz avem nevoie să aflăm pmf (sau pdf sau cdf) a lui X fără Y . Aceasta este numită o **pmf** (sau pdf sau cdf) **marginală**. Următorul exemplu ilustrează modul de a calcula aceasta și motivul pentru termenul "marginal".

2.3.7 Pmf marginală

Exemplul 8. În exemplul 2 am aruncat 2 zaruri corecte și fie X valoarea primului zar și T totalul ambelor zaruri. Calculați pmf-urile marginale ale lui X și T .

Răspuns. În tabel fiecare linie reprezintă o singură valoare a lui X . Astfel, evenimentul " $X = 3$ " este linia colorată a tabelului. Pentru a afla $P(X = 3)$ trebuie să adunăm probabilitățile din această linie. Punem suma în **marginea dreaptă** a tabelului. Analog, $P(T = 5)$ este chiar suma coloanei cu $T = 5$. Punem suma în **marginea de jos** a tabelului.

$X \backslash T$	2	3	4	5	6	7	8	9	10	11	12	$p(x_i)$
1	1/36	1/36	1/36	1/36	1/36	1/36	0	0	0	0	0	1/6
2	0	1/36	1/36	1/36	1/36	1/36	1/36	0	0	0	0	1/6
3	0	0	1/36	1/36	1/36	1/36	1/36	1/36	0	0	0	1/6
4	0	0	0	1/36	1/36	1/36	1/36	1/36	1/36	0	0	1/6
5	0	0	0	0	1/36	1/36	1/36	1/36	1/36	1/36	0	1/6
6	0	0	0	0	0	1/36	1/36	1/36	1/36	1/36	1/36	1/6
$p(t_j)$	1/36	2/36	3/36	4/36	5/36	6/36	5/36	4/36	3/36	2/36	1/36	1

Calculul probabilităților marginale $P(X = 3) = 1/6$ și $P(T = 5) = 4/36 = 1/9$.

Observație. În acest caz știam deja pmf-urile lui X și T . Rezultatele de aici coincid cu cele anterioare.

Pmf-urile marginale se obțin din pmf comună prin adunare:

$$p_X(x_i) = \sum_j p(x_i, y_j), \quad p_Y(y_j) = \sum_i p(x_i, y_j).$$

Termenul "marginal" se referă la faptul că valorile sunt scrise pe marginea tabelului.

2.3.8 Pdf marginală

Pentru densitatea comună continuă $f(x, y)$ cu domeniul $[a, b] \times [c, d]$, pdf-urile marginale sunt:

$$f_X(x) = \int_c^d f(x, y) dy, \quad f_Y(y) = \int_a^b f(x, y) dx.$$

Comparând aceasta cu pmf-urile marginale de mai sus, ca de obicei sumele sunt înlocuite de integrale.

Spunem că pentru a obține marginala pentru X , **integrăm după y** pdf comună și viceversa.

Exemplul 9. Presupunem că (X, Y) ia valori în pătratul $[0, 1] \times [1, 2]$ cu pdf comună $f(x, y) = \frac{8}{3}x^3y$. Aflați pdf-urile marginale $f_X(x)$ și $f_Y(y)$.

Răspuns. Pentru a afla $f_X(x)$ integrăm după y și pentru a afla $f_Y(y)$ integrăm după x .

$$f_X(x) = \int_1^2 \frac{8}{3}x^3y dy = \frac{4}{3}x^3y^2 \Big|_{y=1}^{y=2} = \frac{16}{3}x^3 - \frac{4}{3}x^3 = 4x^3.$$

$$f_Y(y) = \int_0^1 \frac{8}{3} x^3 y dx = \frac{2}{3} x^4 y \Big|_{x=0}^{x=1} = \frac{2}{3} y.$$

Exemplul 10. Presupunem că (X, Y) ia valori în pătratul unitate $[0, 1] \times [0, 1]$ cu pdf comună $f(x, y) = \frac{3}{2} (x^2 + y^2)$. Aflați pdf marginală $f_X(x)$ și utilizați-o pentru a afla $P(X < 0.5)$.

Răspuns.

$$f_X(x) = \int_0^1 \frac{3}{2} (x^2 + y^2) dy = \left(\frac{3}{2} x^2 y + \frac{y^3}{2} \right) \Big|_{y=0}^{y=1} = \frac{3}{2} x^2 + \frac{1}{2}.$$

$$P(X < 0.5) = \int_0^{0.5} f_X(x) dx = \int_0^{0.5} \left(\frac{3}{2} x^2 + \frac{1}{2} \right) dx = \left(\frac{1}{2} x^3 + \frac{1}{2} x \right) \Big|_0^{0.5} = \frac{1}{16} + \frac{1}{4} = \frac{5}{16}.$$

2.3.9 Cdf marginală

Aflarea cdf marginală din cdf comună se face astfel: dacă X și Y iau comun valori pe $[a, b] \times [c, d]$, atunci

$$F_X(x) = F(x, d), \quad F_Y(y) = F(b, y).$$

Dacă $d = \infty$ (caz în care $[c, d]$ se consideră deschis în d), atunci $F_X(x) = \lim_{y \rightarrow \infty} F(x, y)$. Analog pentru $F_Y(y)$.

Exemplul 11. Cdf comună în ultimul exemplu a fost $F(x, y) = \frac{1}{2} (x^3 y + x y^3)$ pe $[0, 1] \times [0, 1]$. Aflați cdf-urile marginale și utilizați $F_X(x)$ pentru a calcula $P(X < 0.5)$.

Răspuns. Avem $F_X(x) = F(x, 1) = \frac{1}{2} (x^3 + x)$ și $F_Y(y) = F(1, y) = \frac{1}{2} (y + y^3)$. Deci $P(X < 0.5) = F_X(0.5) = \frac{1}{2} (0.5^3 + 0.5) = \frac{5}{16}$, exact la fel ca la exemplul 10.

2.3.10 Vizualizare 3D

Am vizualizat $P(a < x < b)$ ca aria de sub pdf $f(x)$ pe intervalul $[a, b]$. Deoarece domeniul de valori al lui (X, Y) este deja o regiune bidimensională în plan, graficul lui $f(x, y)$ este o suprafață peste acea regiune. Putem atunci vizualiza probabilitatea ca [volumul](#) de sub suprafață.

2.4 Independență

Reamintim că evenimentele A și B sunt independente \iff

$$P(A \cap B) = P(A)P(B).$$

Variabilele aleatoare X și Y definesc evenimente ca " $X \leq 2$ " sau " $Y > 5$ ". Astfel, X și Y sunt independente dacă **orice** eveniment definit de X este independent de **orice** eveniment definit de Y . Iată definiția care garantează aceasta:

Definiție. Variabilele aleatoare repartizate comun X și Y sunt **independente** \iff cdf comună a lor este produsul cdf-urilor marginale:

$$F(x, y) = F_X(x)F_Y(y).$$

Pentru variabile discrete aceasta este echivalent cu pmf comună să fie produsul pmf-urilor marginale:

$$p(x_i, y_j) = p_X(x_i)p_Y(y_j).$$

Pentru variabile continue aceasta este echivalent cu pdf comună să fie produsul pdf-urilor marginale:

$$f(x, y) = f_X(x)f_Y(y).$$

Exemplul 12. Pentru **variabile discrete** independența înseamnă că probabilitatea din orice celulă trebuie să fie produsul dintre probabilitățile marginale de pe linia și coloana ei. În primul tabel de mai jos aceasta este adevărat: fiecare probabilitate marginală este $1/6$ și fiecare celulă are $1/36$, adică produsul probabilităților marginale. De aceea X și Y sunt independente. În al 2-lea tabel de mai jos, există probabilități ale celulelor care nu sunt produsul probabilităților marginale. De exemplu, nicio probabilitate marginală nu este 0, deci nicio probabilitate 0 dintr-o celulă nu poate fi produsul marginalelor. De aceea X și T nu sunt independente.

$X \setminus Y$	1	2	3	4	5	6	$p(x_i)$
1	1/36	1/36	1/36	1/36	1/36	1/36	1/6
2	1/36	1/36	1/36	1/36	1/36	1/36	1/6
3	1/36	1/36	1/36	1/36	1/36	1/36	1/6
4	1/36	1/36	1/36	1/36	1/36	1/36	1/6
5	1/36	1/36	1/36	1/36	1/36	1/36	1/6
6	1/36	1/36	1/36	1/36	1/36	1/36	1/6
$p(y_j)$	1/6	1/6	1/6	1/6	1/6	1/6	1

$X \setminus T$	2	3	4	5	6	7	8	9	10	11	12	$p(x_i)$
1	1/36	1/36	1/36	1/36	1/36	1/36	0	0	0	0	0	1/6
2	0	1/36	1/36	1/36	1/36	1/36	1/36	0	0	0	0	1/6
3	0	0	1/36	1/36	1/36	1/36	1/36	1/36	0	0	0	1/6
4	0	0	0	1/36	1/36	1/36	1/36	1/36	1/36	0	0	1/6
5	0	0	0	0	1/36	1/36	1/36	1/36	1/36	1/36	0	1/6
6	0	0	0	0	0	1/36	1/36	1/36	1/36	1/36	1/36	1/6
$p(y_j)$	1/36	2/36	3/36	4/36	5/36	6/36	5/36	4/36	3/36	2/36	1/36	1

Exemplul 13 Pentru **variabile continue** independența înseamnă că putem factoriza pdf sau cdf comună ca produsul unei funcții de x cu o funcție de y .

(i) Presupunând că X are domeniul $[0, 1/2]$, Y are domeniul $[0, 1]$ și $f(x, y) = 96x^2y^3$, atunci X și Y sunt independente. Densitățile marginale sunt $f_X(x) = 24x^2$ și $f_Y(y) = 4y^3$.

(ii) Dacă $f(x, y) = 1.5(x^2 + y^2)$ pe pătratul unitate, atunci X și Y nu sunt independente deoarece nu există niciun mod de a factoriza $f(x, y)$ într-un produs $f_X(x)f_Y(y)$.

(iii) Dacă $F(x, y) = \frac{1}{2}(x^3y + xy^3)$ pe pătratul unitate, atunci X și Y nu sunt independente deoarece cdf nu se factorizează într-un produs $F_X(x)F_Y(y)$.

3 Covarianță și corelație

3.1 Scopurile învățării

1. Să înțeleagă covarianța și corelația.
2. Să poată calcula covarianța și corelația a 2 variabile aleatoare.

3.2 Covarianța

Covarianța este o măsură a cât de mult 2 variabile aleatoare variază împreună. De exemplu, înălțimea și greutatea girafelor au covarianța pozitivă deoarece, când una este mare, cealaltă tinde de asemenea să fie mare.

Definiție. Presupunem că X și Y sunt variabile aleatoare cu mediile μ_X și μ_Y . **Covarianța** lui X și Y este

$$Cov(X, Y) = E((X - \mu_X)(Y - \mu_Y)).$$

3.2.1 Proprietățile covarianței

1. $Cov(aX + b, cY + d) = acCov(X, Y), \forall a, b, c, d$ constante.
2. $Cov(X_1 + X_2, Y) = Cov(X_1, Y) + Cov(X_2, Y)$.
3. $Cov(X, X) = Var(X)$.
4. $Cov(X, Y) = E(XY) - \mu_X\mu_Y$.
5. $Var(X + Y) = Var(X) + Var(Y) + 2Cov(X, Y), \forall X, Y$.
6. Dacă X și Y sunt independente, atunci $Cov(X, Y) = 0$.

Avertizare. Reciproca este falsă: covarianța 0 nu implică totdeauna independență.

Observații. 1. Proprietatea 4 este ca proprietatea similară pentru dispersie.

Într-adevăr, dacă $X = Y$, este exact acea proprietate:

$$Var(X) = E(X^2) - \mu_X^2.$$

Proprietatea 6 reduce formula din proprietatea 5 la formula anterioară

$$Var(X + Y) = Var(X) + Var(Y) \text{ când } X \text{ și } Y \text{ sunt independente.}$$

2. $Cov(X, Y) = Cov(Y, X), \forall X, Y$.

3.2.2 Sume și integrale pentru calculul covarianței

Deoarece covarianța este definită ca o medie, o calculăm ca de obicei ca o sumă sau integrală.

Cazul discret. Dacă X și Y au pmf comună $p(x_i, y_j)$ atunci

$$Cov(X, Y) = \sum_{i=1}^n \sum_{j=1}^m p(x_i, y_j) (x_i - \mu_X)(y_j - \mu_Y) = \left(\sum_{i=1}^n \sum_{j=1}^m p(x_i, y_j) x_i y_j \right) - \mu_X \mu_Y.$$

Cazul continuu. Dacă X și Y au pdf comună $f(x, y)$ pe domeniul $[a, b] \times [c, d]$, atunci

$$Cov(X, Y) = \int_c^d \int_a^b (x - \mu_X)(y - \mu_Y) f(x, y) dx dy = \left(\int_c^d \int_a^b xy f(x, y) dx dy \right) - \mu_X \mu_Y.$$

3.2.3 Exemple

Exemplul 1. Aruncăm o monedă corectă de 3 ori. Fie X numărul de aversuri în primele 2 aruncări și Y numărul de aversuri din ultimele 2 aruncări (astfel încât este o suprapunere peste aruncarea din mijloc). Calculați $Cov(X, Y)$.

Răspuns. Metoda I (folosind tabelul probabilităților comune și definiția covarianței).

Cu 3 aruncări sunt doar 8 rezultate

$\{HHH, HHT, HTH, HTT, THH, THT, TTH, TTT\}$, deci putem face direct tabelul de probabilitate comună.

$X \backslash Y$	0	1	2	$p(x_i)$
0	1/8	1/8	0	1/4
1	1/8	2/8	1/8	1/2
2	0	1/8	1/8	1/4
$p(y_j)$	1/4	1/2	1/4	1

Din marginale calculăm $E(X) = 1 = E(Y)$. Acum folosim definiția:

$$Cov(X, Y) = E((X - \mu_X)(Y - \mu_Y)) = \sum_{i,j} p(x_i, y_j)(x_i - 1)(y_j - 1).$$

Scriem suma eliminând termenii care sunt 0, i.e. toți termenii unde $x_i = 1$ sau $y_j = 1$ sau probabilitatea este 0.

$$Cov(X, Y) = \frac{1}{8}(0 - 1)(0 - 1) + \frac{1}{8}(2 - 1)(2 - 1) = \frac{1}{8} + \frac{1}{8} = \frac{1}{4}.$$

Metoda II (folosind tabelul probabilităților comune și proprietatea 4).

Ca la metoda I se calculează tabelul probabilităților comune și mediile lui X și Y . Din tabel calculăm

$$E(XY) = \sum_{i,j} p(x_i, y_j)x_i y_j = 1 \cdot \frac{2}{8} + 2 \cdot \frac{1}{8} + 2 \cdot \frac{1}{8} + 4 \cdot \frac{1}{8} = \frac{1}{4} + \frac{1}{4} + \frac{1}{4} + \frac{1}{2} = \frac{5}{4}.$$

Din proprietatea 4,

$$Cov(X, Y) = E(XY) - \mu_X \mu_Y = \frac{5}{4} - 1 = \frac{1}{4}.$$

Metoda III (folosind proprietățile covarianței). Ca de obicei, fie X_i rezultatul celei de-a i -a aruncări, deci $X_i \sim \text{Bernoulli}(0.5)$. Avem

$$X = X_1 + X_2 \quad \text{și} \quad Y = X_2 + X_3.$$

Știm că $E(X_i) = 1/2$ și $Var(X_i) = 1/4, \forall i = 1, 2, 3$. De aceea, folosind proprietatea 2 a covarianței și observația 2, avem

$$Cov(X, Y) = Cov(X_1 + X_2, X_2 + X_3) = Cov(X_1, X_2) + Cov(X_1, X_3) + Cov(X_2, X_2) + Cov(X_2, X_3).$$

Deoarece aruncările diferite sunt independente, din proprietatea 6,

$$Cov(X_1, X_2) = Cov(X_1, X_3) = Cov(X_2, X_3) = 0.$$

Deci, din proprietatea 3,

$$Cov(X, Y) = Cov(X_2, X_2) = Var(X_2) = \frac{1}{4}.$$

Exemplul 2. (Covarianță 0 nu implică independența.) Fie $X \sim \begin{pmatrix} -2 & -1 & 0 & 1 & 2 \\ 1/5 & 1/5 & 1/5 & 1/5 & 1/5 \end{pmatrix}$.

Fie $Y = X^2$. Arătați că avem $Cov(X, Y) = 0$, dar X și Y nu sunt independente.

Răspuns. Facem un tabel de probabilitate comună

$Y \backslash X$	-2	-1	0	1	2	$p(y_j)$
0	0	0	1/5	0	0	1/5
1	0	1/5	0	1/5	0	2/5
4	1/5	0	0	0	1/5	2/5
$p(x_i)$	1/5	1/5	1/5	1/5	1/5	1

Folosind marginalele calculăm $E(X) = 0$ și $E(Y) = 2$.

Calculăm covarianța folosind proprietatea 4:

$$Cov(X, Y) = \sum_{i,j} p(x_i, y_j) x_i y_j - \mu_X \mu_Y = \frac{1}{5}(-8 - 1 + 1 + 8) - 0 \cdot 2 = 0.$$

Arătăm că X și Y nu sunt independente. Pentru a face asta, trebuie să găsim un loc unde regula produsului eșuează, i.e. unde $p(x_i, y_j) \neq p(x_i)p(y_j)$:

$$P(X = -2, Y = 0) = 0 \text{ dar } P(X = -2)P(Y = 0) = (1/5)(1/5) = 1/25.$$

Deoarece acestea nu sunt egale, X și Y nu sunt independente.

Discuție. Acest exemplu arată $Cov(X, Y) = 0$ nu implică X și Y sunt independente. De fapt, X și X^2 sunt atât de dependente cât pot fi variabilele aleatoare: dacă știm valoarea lui X , atunci știm valoarea lui X^2 cu 100% certitudine.

$Cov(X, Y)$ măsoară relația liniară dintre X și Y . În exemplul de mai sus X și X^2 au o relație pătratică complet pierdută de $Cov(X, Y)$.

3.2.4 Demonstrațiile proprietăților covarianței

Proprietățile 1 și 2 rezultă din proprietățile similare pentru medie.

3.

$$\text{Cov}(X, X) = E((X - \mu_X)(X - \mu_X)) = E((X - \mu_X)^2) = \text{Var}(X).$$

4.

$$\begin{aligned}\text{Cov}(X, Y) &= E((X - \mu_X)(Y - \mu_Y)) \\ &= E(XY - \mu_X Y - \mu_Y X + \mu_X \mu_Y) \\ &= E(XY) - \mu_X E(Y) - \mu_Y E(X) + \mu_X \mu_Y \\ &= E(XY) - \mu_X \mu_Y - \mu_X \mu_Y + \mu_X \mu_Y \\ &= E(XY) - \mu_X \mu_Y.\end{aligned}$$

5. Folosind proprietățile 3 și 2 și observația 2 obținem

$$\begin{aligned}\text{Var}(X + Y) &= \text{Cov}(X + Y, X + Y) = \text{Cov}(X, X) + 2\text{Cov}(X, Y) + \text{Cov}(Y, Y) \\ &= \text{Var}(X) + \text{Var}(Y) + 2\text{Cov}(X, Y).\end{aligned}$$

6. Reamintim că $E(X - \mu_X) = 0$. Dacă X și Y sunt independente, atunci $f(x, y) = f_X(x)f_Y(y)$. De aceea

$$\begin{aligned}\text{Cov}(X, Y) &= \int \int (x - \mu_X)(y - \mu_Y) f_X(x) f_Y(y) dx dy \\ &= \int (x - \mu_X) f_X(x) dx \int (y - \mu_Y) f_Y(y) dy \\ &= E(X - \mu_X) E(Y - \mu_Y) \\ &= 0.\end{aligned}$$

3.3 Corelația

Unitățile de măsură ale covarianței $\text{Cov}(X, Y)$ sunt ”unitățile lui X ori unitățile lui Y ”. Aceasta face grea compararea covarianțelor: dacă schimbăm scalele, atunci se schimbă și covarianța. Corelația este un mod de a elimina scala din covarianță.

Definiție. Coeficientul de corelație între X și Y este definit de

$$\text{Cor}(X, Y) = \rho = \frac{\text{Cov}(X, Y)}{\sigma_X \sigma_Y}.$$

3.3.1 Proprietăți ale corelației

1. ρ este covarianța standardizărilor lui X și Y .

2. ρ este fără dimensiune.

3. $-1 \leq \rho \leq 1$. Mai mult,

$\rho = 1 \iff Y = aX + b$ cu $a > 0$,

$\rho = -1 \iff Y = aX + b$ cu $a < 0$.

Proprietatea 3 arată că ρ măsoară relația **liniară** între variabile. Când corelația este pozitivă, atunci, dacă X este mare, Y va tinde să fie mare de asemenea. Când corelația este negativă, atunci, dacă X este mare, Y va tinde să fie mic.

Exemplul 2 de mai sus arată că relațiile de ordin mai mare pot fi ratate complet de corelație.

3.3.2 Demonstrația proprietății 3 a corelației

$$0 \leq \text{Var} \left(\frac{X}{\sigma_X} - \frac{Y}{\sigma_Y} \right) = \text{Var} \left(\frac{X}{\sigma_X} \right) + \text{Var} \left(\frac{Y}{\sigma_Y} \right) - 2\text{Cov} \left(\frac{X}{\sigma_X}, \frac{Y}{\sigma_Y} \right) = 2 - 2\rho \implies \rho \leq 1.$$

$$\text{Analog } 0 \leq \text{Var} \left(\frac{X}{\sigma_X} + \frac{Y}{\sigma_Y} \right) \implies \rho \geq -1.$$

$$\rho = 1 \iff \text{Var} \left(\frac{X}{\sigma_X} - \frac{Y}{\sigma_Y} \right) = 0 \iff \frac{X}{\sigma_X} - \frac{Y}{\sigma_Y} = c \iff Y = aX + b \text{ cu } a = \frac{\sigma_Y}{\sigma_X} > 0 \text{ și } b = -c\sigma_Y.$$

$$\text{Analog } \rho = -1 \iff Y = aX + b \text{ cu } a = -\frac{\sigma_Y}{\sigma_X} < 0.$$

Exemplu. Continuăm exemplul 1. Pentru a calcula corelația împărțim covarianța la deviațiile standard. În exemplul 1 am aflat $\text{Cov}(X, Y) = 1/4$ și $\text{Var}(X) = 2\text{Var}(X_j) = 1/2$. Deci, $\sigma_X = 1/\sqrt{2}$. Analog $\sigma_Y = 1/\sqrt{2}$. Astfel,

$$\text{Cor}(X, Y) = \frac{\text{Cov}(X, Y)}{\sigma_X \sigma_Y} = \frac{1/4}{1/2} = \frac{1}{2}.$$

Vedem o corelație pozitivă, ceea ce înseamnă că X mai mare tinde să meargă cu Y mai mare și X mai mic cu Y mai mic. În exemplul 1 aceasta se întâmplă deoarece aruncarea 2 este inclusă atât în X cât și în Y , deci contribuie la mărimea ambelor.

3.3.3 Repartiții normale bivariate

Repartiția normală bivariată are densitatea

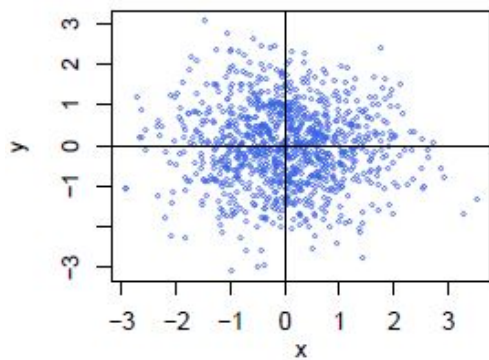
$$f(x, y) = \frac{e^{\frac{-1}{2(1-\rho^2)} \left[\frac{(x-\mu_X)^2}{\sigma_X^2} + \frac{(y-\mu_Y)^2}{\sigma_Y^2} - \frac{2\rho(x-\mu_X)(y-\mu_Y)}{\sigma_X \sigma_Y} \right]}}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho^2}}.$$

Pentru această repartiție, repartițiile marginale pentru X și Y sunt normale și corelația între X și Y este ρ .

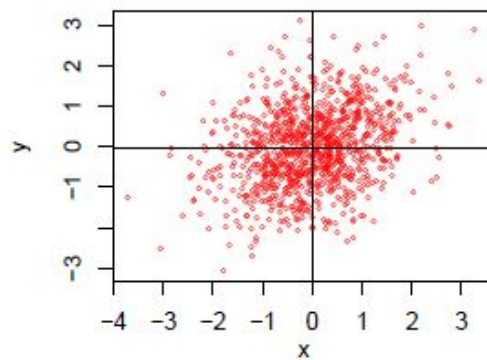
În figurile de mai jos a fost utilizat R pentru a simula repartiția pentru diferite valori ale lui ρ . Individual X și Y sunt normale standard, i.e. $\mu_X = \mu_Y = 0$ și $\sigma_X = \sigma_Y = 1$. Figurile arată reprezentările rezultatelor.

Aceste reprezentări arată o proprietate importantă a corelației. Împărțim datele în cadrane desenând o linie orizontală la media datelor y și una verticală la media datelor x . O corelație pozitivă corespunde tendinței datelor de a se afla în cadranele 1 și 3. O corelație negativă corespunde tendinței datelor de a se afla în cadranele 2 și 4. Putem vedea datele adunându-se de-a lungul unei linii când ρ devine apropiat de ± 1 .

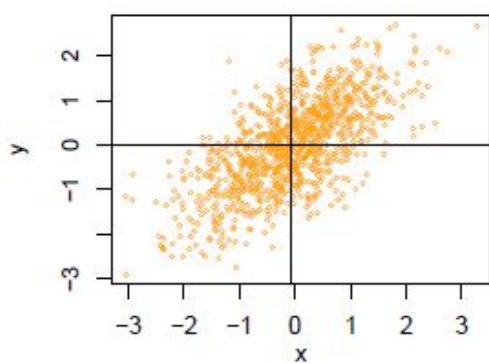
$\rho=0.00$



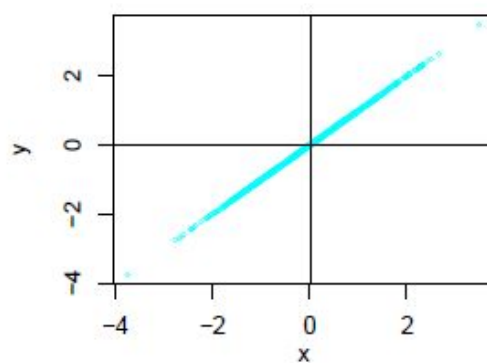
$\rho=0.30$



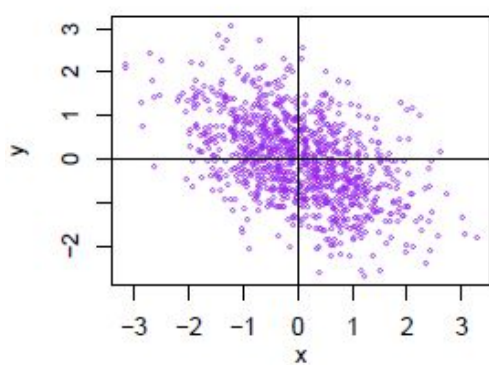
$\rho=0.70$



$\rho=1.00$



$\rho=-0.50$



$\rho=-0.90$

