Seminar 11

Partea1: Covarianţă şi coeficient de corelaţie.

11.1 Teorie: covarianţă şi coeficientul de corelaţie

Covarianţa unui vector (X, Y) se defineşte ca fiind

$$cov(X,Y) := M[(X - M(X))(Y - M(Y)]$$

Covarianţa are următoarele proprietăţi:

- $cov(X,Y) = M(XY) - M(X)M(Y)$
- $cov(X,X) = \sigma^2(X)$;
- $cov(X,a) = 0$
- $cov(X,Y) = cov(Y,X)$;
- $cov(aX,Y) = a\,cov(X,Y)$, $a \in \mathbb{R}$;
- $cov(X+Y,Z) = cov(X,Z) + cov(Y,Z)$;
- $\sigma^2(X+Y) = \sigma^2(X) + \sigma^2(Y) + 2cov(X,Y)$;
- Dacă în plus X şi Y sunt variabile aleatoare independente, atunci
$$\sigma^2(X+Y) = \sigma^2(X) + \sigma^2(Y).$$
- X, Y sunt independente, atunci $cov(X,Y) = 0$ (**Reciproca nu este adevărată**!)

itemize **Coeficientul de corelaţie** a două variabile aleatoare X şi Y, de abateri standard nenule σ_X, σ_Y, este

$$\rho(X,Y) = \frac{cov(X,Y)}{\sigma_X \sigma_Y} = \frac{cov(X,Y)}{\sigma(X)\sigma(Y)} \in [-1, 1]$$

- Dacă între variabilele aleatoare X şi Y există o relaţie liniară de forma
$$Y = aX + b,\ a, b \in \mathbb{R},\ a \neq 0,$$
atunci
$$\rho(X,Y) = \begin{cases} -1, & \text{dacă } a < 0, \\ 1, & \text{dacă } a > 0. \end{cases}$$

- Reciproc, dacă $|\rho(X,Y)| = 1$, atunci între ele există o relaţie liniară,

$$Y = aX + b, \quad a \neq 0$$

Legătura dintre două variabile X şi Y poate fi determinată folosind coeficientul de corelaţie astfel:

- $\rho(X,Y) = 0$, atunci X şi Y sunt **necorelate**;
- $\rho(X,Y)$ este apropiat de zero, atunci X şi Y sunt **slab corelate** (intensitatea legăturii dintre ele este redusă);
- $\rho(X,Y) = 1$, atunci $Y = aX + b, \quad a > 0$, X şi Y sunt **pozitiv corelate**;
- $\rho(X,Y) = -1$, atunci $Y = aX + b, \quad a < 0$, X şi Y sunt **negativ corelate**;
- $|\rho(X,Y)|$ are o valoare apropiată de 1, **relaţia dintre variabilele aleatoare este "aproape liniară"**, adică valorile (x,y) ale vectorului aleator (X,Y) sunt uşor dispersate în jurul unei drepte de ecuaţie $y = ax + b$.

Matricea de covarianţă a vectorului aleator $\mathbf{X} = (X_1, X_2, \ldots, X_n)^T$ este matricea notată cu Σ, ale cărei elemente sunt $\sigma_{ij} = \text{cov}(X_i, X_j)$, $i, j \in \{1, \ldots, n\}$. **Observaţii:**

- $\sigma_{ii} = cov(X_i, X_i) = \sigma^2(X_i)$
- Σ este simetrică şi semipozitiv definită
- $\Sigma = M(\mathbf{YY}^T)$, unde $\mathbf{Y} = \mathbf{X} - \mathbf{m} = (X_1 - m_1, X_2 - m_2, \ldots, X_n - m_n)^T$ iar $M(\mathbf{YY^T})$ notează matricea mediilor elementelor matricii $\mathbf{YY^T}$.

11.2 Probleme rezolvate: covarianţă şi coeficientul de corelaţie

1. Fie X o variabilă cu valorile $-2, -1, 0, 1, 2$, fiecare valoare fiind luată cu probabilitate $p = \frac{1}{5}$. Se consideră $Y = X^2$. Să se arate că $cov(X,Y) = 0$, dar variabilele X şi Y nu sunt independente.

Rezolvare: Distribuţia de probabilitate a vectorului (X,Y) este:

		Y			$\sum p_i$
		0	1	4	
	-2	0	0	$\frac{1}{5}$	$\frac{1}{5}$
X	-1	0	$\frac{1}{5}$	0	$\frac{1}{5}$
	0	$\frac{1}{5}$	0	0	$\frac{1}{5}$
	1	0	$\frac{1}{5}$	0	$\frac{1}{5}$
	2	0	0	$\frac{1}{5}$	$\frac{1}{5}$
$\sum q_i$		$\frac{1}{5}$	$\frac{2}{5}$	$\frac{2}{5}$	

Avem că $M(X) = 0, M(Y) = 2$. Din tabel se observă că X şi Y nu sunt independente: de exemplu, $P(X = -2, Y = 0) = 0 \neq P(X = -2) \cdot P(Y = 0) = \frac{1}{25}$.
Pentru a calcula covarianta vectorului (X, Y) vom folosi formula

$$cov(X, Y) = M(XY) - M(X)M(Y)$$

Variabila XY are distribuţia $XY = \begin{pmatrix} -8 & -4 & -2 & -1 & 0 & 1 & 2 & 4 & 8 \\ \frac{1}{5} & 0 & 0 & \frac{1}{5} & 0 & \frac{1}{5} & 0 & 0 & \frac{1}{5} \end{pmatrix}$, iar $M(XY) = \frac{1}{5}(-8 - 1 + 1 + 8) = 0$. Deci, $cov(X, Y) = 0$.
Media variabilei XY se poate determina folosind LOTUS, fără a calcula deci distribuţia variabilei XY. Avem $M(XY) = \sum xyP(X = x, Y = y) = (-2) \cdot 4 \cdot \frac{1}{5} + (-1) \cdot 1 \cdot \frac{1}{5} + 1 \cdot 1\frac{1}{5} + 2 \cdot 4 \cdot \frac{1}{5} = 0$.

2. Fie X o variabilă aleatoare ce are media $M(X) = 3$ şi dispersia $\sigma^2(X) = 1$, iar $Y = -2X + 5$. Să se calculeze covarianţa şi coeficientul de corelaţie pentru variabilele X, Y. Să se determine matricea de covarianţă asociată vectorului (X, Y).
Rezolvare: Deoarece între X şi Y există o relaţie liniară de forma $Y = aX + b$ cu $a < 0$, coeficientul de corelaţie este

$$\rho(X, Y) = -1.$$

Dar cum

$$\rho(X, Y) = \frac{cov(X, Y)}{\sigma(X)\,\sigma(Y)},$$

calculând $\sigma^2(Y) = \sigma^2(-2X + 5) = 4\sigma^2(X) = 4$, rezultă că $-1 = \dfrac{cov(X, Y)}{2}$, deci $cov(X, Y) = -2$.
Matricea de covarianţă asociată vectorului (X, Y) are forma $\Sigma = \begin{pmatrix} \sigma^2(X) & cov(X, Y) \\ cov(X, Y) & \sigma^2(Y) \end{pmatrix}$.
Obţinem $\Sigma = \begin{pmatrix} 1 & -2 \\ -2 & 4 \end{pmatrix}$.

3. Fie X şi Y două variabile aleatoare independente, $X, Y \sim N(0, 1)$. Se consideră $Z = 1 + X + XY^2$ şi $W = 1 + X$. Să se calculeze $cov(Z, W)$.

Rezolvare: Folosind proprietăţile şi definiţia covarianţei , putem arăta că:
$cov(Z, W) = cov(1 + X + XY^2, 1 + X) = cov(X + XY^2, X) = cov(X, X) + cov(XY^2, X) = \sigma^2(X) + M(X^2Y^2) - M(XY^2)M(X)$.
Variabilele X şi Y sunt independente, deci $M(X^2Y^2) = M(X^2)M(Y^2)$.
Vom obţine: $cov(Z, W) = \sigma^2(X) + M(X^2)M(Y^2) - M(X)^2M(Y^2) = 1 + 1 - 0 = 2$.

4. Se consideră vectorul aleator continuu (X, Y) cu densitatea de probabilitate

$$f(x, y) = \begin{cases} 2 & \text{dacă } 0 \leq y \leq x \leq 1 \\ 0 & \text{dacă altfel} \end{cases}$$

Să se determine $cov(X, Y)$.
Rezolvare: Vom aplica formula $cov(X, Y) = M(XY) - M(X)M(Y)$.
Pentru a determina $M(X)$ trebuie sa determinăm densitatea marginală $f_X(x)$ (analog pt $M(Y)$). Avem, pentru $x \in (0, 1)$,

$$f_X(x) = \int_{-\infty}^{\infty} f(x, y)dy = \int_0^x 2dy = 2x$$

Deci, $f_X(x) = \begin{cases} 2x & \text{dacă } 0 \leq x \leq 1 \\ 0 & \text{dacă altfel} \end{cases}$ Atunci $M(X) = \int_0^1 x f_X(x)dx = \frac{2}{3}$. Analog, pentru $y \in (0, 1)$ avem

$$f_Y(y) = \int_{-\infty}^{\infty} f(x, y)dx = \int_y^1 2dx = 2(1 - y)$$

Deci, $f_Y(y) = \begin{cases} 2(1-y) & \text{dacă } 0 \leq y \leq 1 \\ 0 & \text{dacă altfel} \end{cases}$ Atunci $M(Y) = \int_0^1 y f_Y(y)dy = \frac{1}{3}$.

In plus, $M(XY) = \int\int_{\mathbb{R}^2} xyf(x, y)dxdy = \int_0^1\int_0^x 2xydy\, dx = \int_0^1 x^3 dx = \frac{1}{4}$.
In concluzie, avem $cov(X, Y) = \frac{1}{36}$.

11.3 Probleme propuse: covarianţă, coeficientul de corelaţie

5. Se consideră vectorul aleator discret (X, Y) cu densitatea de probabilitate

$$P(x, y) = \begin{cases} \frac{1}{21} & \text{dacă } x = 0, 1, 2, 3, 4, 5, \quad y = 0, 1, \ldots, x \\ 0 & \text{dacă altfel} \end{cases}$$

Să se determine $cov(X, Y)$.

6. Se consideră vectorul aleator continuu (X, Y) cu densitatea de probabilitate

$$f(x, y) = \begin{cases} \frac{x+y}{3} & \text{dacă } 0 \leq x \leq 1, 0 \leq y \leq 2 \\ 0 & \text{dacă altfel} \end{cases}$$

Să se determine $cov(X, Y)$.

7. Daca matricea de covarianta a vectorului (X, Y) este $\Sigma = \begin{pmatrix} 4 & -4 \\ -4 & 25 \end{pmatrix}$, sa se calculeze $\rho(X, Y)$ şi $\sigma^2(X + 2Y)$.

8. Fie variabilele aleatoare X şi Y înre care există relaţia $Y = X - 2$. Să se calculeze covarianţa şi coeficientul de corelaţie pentru variabilele X, Y, şiind că $\sigma^2(X) = 0.01$. Să se determine matricea de covarianţă asociată vectorului (X, Y).

9. Dacă X şi Y sunt două variabile aleatore astfel încât $\rho(X,Y)=0$, atunci sunt X şi Y independente? Dar necorelate?

10. Se consideră vectorul aleator continuu (X,Y) cu densitatea de probabilitate

$$f(x,y)=\begin{cases} 4xy & \text{dacă } 0\le x\le 1, 0\le y\le 1 \\ 0 & \text{dacă altfel} \end{cases}$$

Să se determine $cov(X,Y)$.