Joint distributions. Covariance and correlation

Probability theory

Anton Afanasev

Higher School of Economics

DSBA 221

October 7, 2023

Seminar Overview

Quiz

1. Consider the following p.m.f. of a random variable X:

x	1789	1790	1791	1792	1793
$\mathsf{P}_X(x)$	$1/5$	$1/4$	$1/10$	$1/4$	$1/5$

   Find the following **WITHOUT CALCULATOR**:
   - $\mathsf{E}(X)$,
   - $\mathsf{V}(X)$.

2. X is a random variable with $\mathsf{E}\left(X^2\right) = 3.6$ and $\mathsf{P}(X = 2) = 0.6$ and $\mathsf{P}(X = 3) = 0.1$. X takes just one other value between 0 and 3. Find the variance of X.

Problem 1

A random variable X has a binomial distribution with mean 10 and variance 6. Find $\mathsf{P}(X = 4)$.

Problem 2

A box contains two gold balls and three silver balls. You are allowed to choose successively balls from box at random. You win 1 dollar each time you draw a gold ball and lose 1 dollar each time you draw a silver ball. After a draw, the ball is not replaced. Show that, if you draw until you are ahead by 1 dollar or until there are no more gold balls, this is a favorable game.

Problem 2

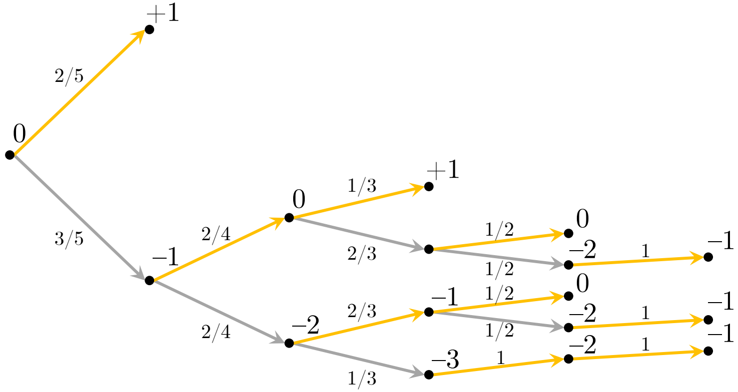

Figure: Probability tree of the Problem 2.

Joint distribution function

Definition

Joint distribution function $F_{X_1,\dots,X_n}$ of random variables $X_1, \dots, X_n$:

$$\forall x_i \in \mathbb{R} : \quad F_{X_1,\dots,X_n}(x_1, \dots, x_n) = \mathsf{P}(X_1 \le x_1, \dots, X_n \le x_n).$$

1. $F_{X_1,\dots,X_n}(x_1, \dots, x_n) \in [0, 1]$,
2. Monotonically non-decreasing in each variable (others are fixed),
3. Right-continuous in each variable (others are fixed),
4. $\lim\limits_{\substack{x_1 \to +\infty \\ \dots \\ x_n \to +\infty}} F_{X_1,\dots,X_n}(x_1, \dots, x_n) = 1$,
5. $\lim\limits_{\substack{x_1 \to -\infty \\ \dots \\ x_n \to -\infty}} F_{X_1,\dots,X_n}(x_1, \dots, x_n) = 0$,
6. $\forall a_i, b_i \in \mathbb{R} : \quad \mathsf{P}(a_1 < X_1 \le b_1, \dots, a_n < X_n \le b_n) \ge 0.$

Probabilities via c.d.f.

- One variable X with c.d.f. F_X:

$$\mathsf{P}(x_1 < X \leq x_2) = F_X(x_2) - F_X(x_1).$$

- Two variables X, Y with c.d.f. $F_{X,Y} = F$:

$$\mathsf{P}(x_1 < X \leq x_2,\ y) = F(x_2, y) - F(x_1, y),$$
$$\mathsf{P}(x,\ y_1 < Y \leq y_2) = F(x, y_2) - F(x, y_1),$$
$$\Downarrow$$

$$\mathsf{P}(x_1 < X \leq x_2, y_1 < Y \leq y_2) = F(x_2, y_2) - F(x_1, y_2) - F(y_1, x_2) + F(x_1, y_1).$$

Joint and marginal mass functions

- Joint p.m.f. of $X_1, \dots, X_n$:

$$\mathsf{P}_{X_1,\dots,X_n}(x_1, \dots, x_n) = \mathsf{P}(X_1 = x_1, \dots, X_n = x_n).$$

- Marginal p.m.f.-s of $X_1, \dots, X_n$:

$$\mathsf{P}_{X_i}(\alpha) = \sum_{\substack{x_j \in X_j \\ j \neq i}} \mathsf{P}_{X_1,\dots,X_n}(x_1, \dots, x_{i-1}, \alpha, x_{i+1}, \dots, x_n).$$

- Marginal p.m.f.-s of X, Y:

$$\mathsf{P}_X(x) = \sum_{y \in Y} \mathsf{P}_{X,Y}(x, y),$$

$$\mathsf{P}_Y(y) = \sum_{x \in X} \mathsf{P}_{X,Y}(x, y).$$

Marginal distribution function

- Marginal c.d.f.-s of X, Y:

$$F_X(x) = \lim_{y\to\infty} F_{X,Y}(x,y) = F_{X,Y}(x,\infty),$$
$$F_Y(y) = \lim_{x\to\infty} F_{X,Y}(x,y) = F_{X,Y}(\infty,y).$$

- Marginal c.d.f.-s of $X, \ldots, X_n$:

$$F_{X_i}(\alpha) = \lim_{\substack{x_j\to\infty \\ j\neq i}} F_{X_1,\ldots,X_n}(x_1,\ldots,x_{i-1},\alpha,x_{i+1},\ldots,x_n).$$

Problem 3

Consider two random variables with the following joint distribution:

$X \setminus Y$	1	2
3	$1/4$	$1/4$
5	$1/6$	$1/3$

1. Find the marginal distributions of X and of Y.
2. Are X and Y independent?
3. Find $\mathsf{E}(X+2Y)$, $\mathsf{E}(XY)$, $\mathsf{V}(X+Y)$.
4. Suppose that random variables U and V have same distributions as X and Y, but are independent. Find the joint distribution of U and V.

Independence

Definition

Random variables X and Y are independent iff:

$$\forall x, y \in \mathbb{R} : \quad F_{X,Y}(x, y) = F_X(x) \cdot F_Y(y).$$

- Inherently, discrete X and Y are independent iff:

$$\forall x, y \in \mathbb{R} : \quad \mathsf{P}_{X,Y}(x, y) = \mathsf{P}_X(x) \cdot \mathsf{P}_Y(y).$$

Definition

Random variables $X_1, \ldots, X_n$ are collectively (mutually) independent iff:

$$\forall x_i \in \mathbb{R} : \quad F_{X_1,\ldots,X_n}(x_1, \ldots, x_n) = F_{X_1}(x_1) \cdot \ldots \cdot F_{X_n}(x_n).$$

Covariance and correlation

- Covariance of X and Y:

$$\mathsf{Cov}(X,Y) = \mathsf{E}(X - \mathsf{E}(X))(Y - \mathsf{E}(Y)) = \mathsf{E}\,(XY) - \mathsf{E}\,(X) \cdot \mathsf{E}\,(Y)\,.$$

- Correlation coefficient of X and Y:

$$\rho(X,Y) = \mathsf{Corr}(X,Y) = \frac{\mathsf{Cov}(X,Y)}{\sigma(X) \cdot \sigma(Y)}.$$

- $\rho(X,Y)$ is a cosine-like measure between "vectors" X and Y:

$$-1 \leq \rho(X,Y) \leq 1,$$

with Cauchy–Schwarz inequality:

$$\mathsf{Cov}(X,Y)^2 \leq \mathsf{V}(X) \cdot \mathsf{V}(Y),$$

and equality is achieved iff X and Y are linearly dependent.

Covariance and correlation

- If $\text{Cov}(X, Y) = 0$ then X and Y are called uncorrelated.
- Some identities:

$$
\begin{aligned}
\text{Cov}(X, Y) &= \text{Cov}(Y, X), \\
\text{Cov}(X, X) &= \text{V}(X), \\
\text{Cov}(a, X) &= 0, \\
\text{Cov}(aX + b, cY + d) &= ac\,\text{Cov}(X, Y), \\
\text{V}(X \pm Y) &= \text{V}(X) + \text{V}(Y) \pm 2\text{Cov}(X, Y),
\end{aligned}
$$

  where a, b, c and d are constants.
- $\text{Cov}(X, Y)$ is bilinear:

$$\text{Cov}(X + Y, V + W) = \text{Cov}(X, V) + \text{Cov}(X, W) + \text{Cov}(Y, V) + \text{Cov}(Y, W).$$

Problem 4

Suppose that X and Y have the following joint probability mass function:

$X \setminus Y$	1	2	3
1	0.25	0.25	0
2	0	0.25	0.25

What is the correlation coefficient?

Problem 5

The probability distribution of a random variable X is:

$X = x$	-1	0	1
$\mathsf{P}(X = x)$	a	b	a

What is the correlation coefficient between X and X^2?

Uncorrelatedness and independence

- $\mathsf{Corr}(X, Y)$ is a measure of linear dependence between X and Y.
- Uncorrelatedness and independence are not equivalent:

$$\text{Independence} \Rightarrow \text{Uncorrelatedness},$$
$$\text{Uncorrelatedness} \not\Rightarrow \text{Independence}.$$

- If random variables are dependent non-linearly, they could have $\mathsf{Corr}(X, Y) \neq 0$.

That's all Folks