Probabilities and Statistics - Lecture 5

Olariu E. Florentin

March 25, 2024

Table of contents

Covariance

Definition 1

Let X and Y two discrete random variables which have expectations.

(i) *The* covariance *of X and Y (if exists) is defined as*

$$cov[X, Y] = \mathbb{E}\left[(X - \mathbb{E}[X]) \cdot (Y - \mathbb{E}[Y])\right] =$$

$$= \sum_{i,j} (x_i - \mathbb{E}[X]) \cdot (y_j - \mathbb{E}[Y]) \cdot P\{X = x_i \cap Y = y_j\}.$$

(ii) *The* correlation coefficient *(or just the* correlation*) of X and Y (non-degenerate variables) is*

$$\rho(X, Y) = \frac{cov[X, Y]}{StDev[X] \cdot StDev[Y]}.$$

Joint probability distribution and covariance - example

Example. We have two boxes: B_1 containing 2 white balls, 2 black balls, and 3 red balls, and B_2 containing 3 white balls, 2 black balls, and 1 red ball. From B_1 we withdraw a ball which we put in B_2, then we withdraw another ball from the second box. Let X be the number of white withdrawn balls, and Y be the number of black withdrawn balls.

(a) Determine the joint probability distribution of X and Y.

(b) Determine the distribution and the expectation of XY.

(c) Determine the covariance and the correlation betwen X and Y.

Solution: We observe first that X and Y are related like this: $X + Y \leqslant 2$. Let A_i be the event "the i-th withdrawn ball is white", B_i be the event "the i-th withdrawn ball is black", and C_i be the event "the i-th withdrawn ball is red" $(i = \overline{1,2})$.

Joint probability distribution and covariance - example

		X			
		0	1	2	
Y	0	6/49	11/49	8/49	25/49
	1	?	?	0	?
	2	?	0	0	?
		?	?	8/49	

Covariance of two random variables

Proposition 1

Let X and Y two random variables which have expectations and covariance. Then

(i) $cov[X, Y] = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y]$.

(ii) $Var[X + Y] = Var[X] + 2cov[X, Y] + Var[Y]$.

(iii) $-1 \leqslant \rho[X, Y] = \rho[Y, X] \leqslant 1$ *and* $\rho[X, X] = 1$ *(i. e.,* $cov[X, X] = Var[X]$*).*

(iv) *(exercise)* $\rho[aX + b, Y] = \rho[X, Y]$, *if* $a \in \mathbb{R}^*$, $b \in \mathbb{R}$.

(v) $cov[aX + bY + c, Z] = a \cdot cov[X, Z] + b \cdot cov[Y, Z]$, *for* $a, b, c \in \mathbb{R}$.

(vi) *(exercise)* $cov\left[\sum_{i=1}^{n} X_i, \sum_{j=1}^{m} Y_j\right] = \sum_{i=1}^{n}\sum_{j=1}^{m} cov[X_i, Y_j]$.

Covariance of two random variables

proof: For (i)

$$cov[X, Y] = \mathbb{E}\left[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])\right] =$$

$$= \mathbb{E}\left[XY - \mathbb{E}[Y]X - \mathbb{E}[X]Y + \mathbb{E}[X]\mathbb{E}[Y]\right] =$$

$$= \mathbb{E}[XY] - 2\mathbb{E}[X]\mathbb{E}[Y] + \mathbb{E}[X]\mathbb{E}[Y] = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y].$$

For (ii)

$$Var[X + Y] = \mathbb{E}\left[(X + Y)^2\right] - \mathbb{E}^2[X + Y] =$$

$$= \mathbb{E}\left[X^2 + 2XY + Y^2\right] - \left(\mathbb{E}^2[X] + 2\mathbb{E}[X]\mathbb{E}[Y] + \mathbb{E}^2[Y]\right) =$$

$$= \left(\mathbb{E}\left[X^2\right] - \mathbb{E}^2[X]\right) + 2\left(\mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y]\right) + \left(\mathbb{E}\left[Y^2\right] - \mathbb{E}^2[Y]\right) =$$

$$= Var[X] + 2cov[X, Y] + Var[Y].$$

Covariance of two random variables

Now for (iii), since $0 \leqslant Var[tX + Y] = t^2 Var[X] + 2t \cdot cov[X, Y] + Var[Y]$, for every $t \in \mathbb{R}$, we must have:

$$\Delta = 4cov^2[X, Y] - 4Var[X]Var[Y] \leqslant 0 \Leftrightarrow$$

$$\Leftrightarrow |cov[X, Y]| \leqslant StDev[X] \cdot StDev[Y].$$

Then, $cov[X, X] = \frac{1}{2}(Var[2X] - 2Var[X]) = Var[X]$.

For (v):

$$cov[aX + bY + c, Z] = \mathbb{E}[aXZ + bYZ + cZ] - \mathbb{E}[aX + bY + c] \cdot \mathbb{E}[Z] =$$

$$= a\mathbb{E}[XZ] + b\mathbb{E}[YZ] + c\mathbb{E}[Z] - (a\mathbb{E}[X] + b\mathbb{E}[Y] + c) \cdot \mathbb{E}[Z]. \blacksquare$$

Covariance of two random variables - example

Example. Let X_1, Y_1 and X_2, Y_2 be two pairs of random variables with the following joint probability distributions:

		X_1 = 1	X_1 = 2	
Y_1	2	1/4	1/4	1/2
	4	1/4	1/4	1/2
		1/2	1/2	

		X_2 = 1	X_2 = 2	
Y_2	2	1/2	0	1/2
	4	0	1/2	1/2
		1/2	1/2	

Prove that $\rho[X_1, Y_1] \neq \rho[X_2, Y_2]$ and $cov[X_1, Y_1] \neq cov[X_2, Y_2]$.

Covariance of two random variables - example

Solution: X_1 and X_2 (Y_1 and Y_2) have the same distribution, $StDev[X_1] = StDev[X_2]$ and $StDev[Y_1] = StDev[Y_2]$, therefore it will be sufficient to show that: $cov[X_1, Y_1] \neq cov[X_2, Y_2]$.

$$\mathbb{E}[X_1] = \mathbb{E}[X_2] = 1 \cdot \frac{1}{2} + 2 \cdot \frac{1}{2} = \frac{3}{2}, \mathbb{E}[Y_1] = \mathbb{E}[Y_2] = 2 \cdot \frac{1}{2} + 4 \cdot \frac{1}{2} = 3.$$

We compute the expectation of $X_1 Y_1$ and the covariance of X_1 and Y_1:

$$X_1 Y_1 : \begin{pmatrix} 2 & 4 & 8 \\ \frac{1}{4} & \frac{1}{2} & \frac{1}{4} \end{pmatrix} \Rightarrow \mathbb{E}[X_1 Y_1] = 2 \cdot \frac{1}{4} + 4 \cdot \frac{1}{2} + 8 \cdot \frac{1}{4} = \frac{9}{2},$$

$$cov[X_1, Y_1] = \mathbb{E}[X_1 Y_1] - \mathbb{E}[X_1]\mathbb{E}[Y_1] = \frac{9}{2} - \frac{9}{2} = 0.$$

Covariance of two random variables - example

We now compute the expectation of $X_2 Y_2$ and the covariance of X_2 and Y_2

$$X_2 Y_2 : \begin{pmatrix} 2 & 8 \\ \frac{1}{2} & \frac{1}{2} \end{pmatrix} \Rightarrow \mathbb{E}[X_2 Y_2] = 2 \cdot \frac{1}{2} + 8 \cdot \frac{1}{2} = 5,$$

therefore

$$cov[X_2, Y_2] = \mathbb{E}[X_2 Y_2] - \mathbb{E}[X_2]\mathbb{E}[Y_2] = 5 - \frac{9}{2} = \frac{1}{2}. \clubsuit$$

Independent random variables

Definition 2

Two random variables X and Y are independent *if, for every $A \subseteq X(\Omega)$ and $B \subseteq Y(\Omega)$, we have*

$$P\{(X \in A) \cap (Y \in B)\} = P\{X \in A\} \cdot P\{Y \in B\}.$$

Since $P\{X = x_i \cap Y = y_j\} = P\{X = x_i\} \cdot P\{Y = y_j\} = p_i \cdot q_j$, it follows that the joint distribution can be computed like this: $r_{ij} = p_i q_j$.

Theorem 2.1

Let X and Y be two independent discrete random variables, then:

(i) $\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y]$.

(ii) $Var[X + Y] = Var[X] + Var[Y]$.

(iii) $cov[X, Y] = 0$.

Independent random variables

proof: We consider only the case when both variables are finite. For (i):

$$\mathbb{E}[XY] = \sum_z zP\{XY = z\} = \sum_z z \cdot \left(\sum_{z = x_i y_j} P\{X = x_i \cap Y = y_j\} \right) =$$

$$= \sum_{i,j} x_i y_j P\{X = x_i \cap Y = y_j\} = \sum_{i,j} x_i y_j P\{X = x_i\} P\{Y = y_j\} =$$

$$= \left(\sum_i x_i P\{X = x_i\} \right) \cdot \left(\sum_j y_j P\{Y = y_j\} \right) = \mathbb{E}[X]\mathbb{E}[Y].$$

Independent random variables

From the last relation we get (iii): $cov[X,Y] = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y] = 0$.

For (ii) we use Proposition 1 or proceed directly:

$$Var[X+Y] = \mathbb{E}\left[(X+Y)^2\right] - (\mathbb{E}[X+Y])^2 =$$

$$= \mathbb{E}\left[X^2 + 2XY + Y^2\right] - (\mathbb{E}[X] + \mathbb{E}[Y])^2 =$$

$$= \mathbb{E}\left[X^2\right] + 2\mathbb{E}[XY] + \mathbb{E}\left[Y^2\right] - \mathbb{E}^2[X] - 2\mathbb{E}[X]\mathbb{E}[Y] - \mathbb{E}^2[Y] =$$

$$= \mathbb{E}\left[X^2\right] - \mathbb{E}^2[X] + \mathbb{E}\left[Y^2\right] - \mathbb{E}^2[Y] = Var[X] + Var[Y]. \blacksquare$$

The converse of this theorem is not necessarily true.

Independent random variables - example

Example. We roll two dice. Determine the expectation of the product and the variance of sum.

Solution: Let X_1 and X_2 the values on the dice. These two variables are independent (and have the same distribution), hence

$$\mathbb{E}[X_1 X_2] = \mathbb{E}[X_1]\mathbb{E}[X_2] = \frac{49}{4} \text{ and}$$

$$Var[X_1 + X_2] = Var[X_1] + Var[X_2].\clubsuit$$

Introduction

- The theoretical results from the following section help to compute lower and upper bounds for probabilities which are related to a random variable (when this variable has expectation and variance).

Markov inequality

Theorem 2.1

(Markov inequality.) Let $X \geqslant 0$ be a discrete random variable with $\mathbb{E}[X] = \mu$. Then

$$P\{X \geqslant t\} \leqslant \frac{\mu}{t}, \forall t > 0.$$

proof: When X is a discrete random variable having the following distribution

$$X : \begin{pmatrix} x_1 & x_2 & \dots & x_n & \dots \\ p_1 & p_2 & \dots & p_n & \dots \end{pmatrix},$$

where $x_1 < x_2 < \dots < x_n < \dots$. Let us suppose that $t \in (x_{k-1}, x_k]$ ($x_0 = -\infty$), then

$$\mu = \mathbb{E}[X] = \sum_i p_i x_i \geqslant \sum_{i \geqslant k} p_i x_i \geqslant t \sum_{i \geqslant k} p_i = t \cdot P\{X \geqslant t\}.$$

∎

Markov inequality

Proposition 2

The Markov inequality becomes equality if and only if

$$P\{X = 0\} + P\{X = t\} = 1.$$

proof: If we transform the sequence of inequalities from the above proof in equalities we get

$$p_i x_i = 0, \forall i < k \text{ and } p_i x_i = p_i t, \forall i \geqslant k.$$

As the distribution contains only pozitive probabilities (i. e., $p_i > 0$, $\forall i$), we must have that

$$X \text{ has two values, } x_1 = 0 \text{ and } x_2 = t$$

$$\text{or has only one value, } X \equiv t.$$

■

Markov inequality - example

Example. Let $X \geqslant 0$ be a random variable with $\mathbb{E}[X] = 1$. Find upper bounds for the following probabilities.

$$P\{X \geqslant 2\}, P\{X \geqslant 4\} \text{ and } P\{X \geqslant 2^k\}.$$

Solution: Using Markov inequality

$$P\{X \geqslant 2\} \leqslant \frac{\mathbb{E}[X]}{2} = \frac{1}{2}, P\{X \geqslant 4\} \leqslant \frac{\mathbb{E}[X]}{4} = \frac{1}{4},$$

$$P\{X \geqslant 2^k\} \leqslant \frac{\mathbb{E}[X]}{2^k} = \frac{1}{2^k} \clubsuit$$

Observation. The probability that a random variable $X \geqslant 0$ (with finite expectation), has values greater than a given $t > 0$, becomes smaller as t increases.

Chebyshev inequality

Theorem 2.2

(Chebyshev inequality.) *Let X be a discrete random variable with $\mathbb{E}[X] = \mu$ and $Var[X] = \sigma^2$. Then*

$$P\left\{|X-\mu| \geqslant t\right\} \leqslant \frac{\sigma^2}{t^2}, \forall t > 0.$$

proof:
Consider the variable $Y = (X-\mu)^2$ for which $\mathbb{E}[Y] = Var[X]$, then using Markov inequality,

$$P\left\{|X-\mu| \geqslant t\right\} = P\left\{(X-\mu)^2 \geqslant t^2\right\} \leqslant \frac{\mathbb{E}[Y]}{t^2} = \frac{\sigma^2}{t^2}, \forall t > 0.$$

∎

Chebyshev inequality

- A possible interpretation of this inequality: if a variable has small variance, then the probability that this variable takes values far away from its expectation is small.
- The following Chebyshev inequality consequence says that the probability that a variable takes values at least k standard deviations far from its expectation is at most $\frac{1}{k^2}$.
- In this way we can say that the standard deviation measures the spreading of the values of a variable around its expectation.

Chebyshev inequality

Corollary 2.1

Let X be a variable with $\mathbb{E}[X] = \mu$ and $Var[X] = \sigma^2 > 0$.

$$P\{|X - \mu| \geqslant k\sigma\} \leqslant \frac{1}{k^2}, \forall k > 0.$$

Proposition 3

The Chebyshev inequality becomes equality if and only if

$$P\{X = \mu - t\} + P\{X = \mu\} + P\{X = \mu + t\} = 1.$$

proof: Chebyshev inequality is based on Markov inequality, hence the equality holds if and only if $P\{Y = 0\} + P\{Y = t^2\} = 1$.

$$P\{Y = 0\} = P\{(X - \mu)^2 = 0\} = P\{X = \mu\}, \text{ and}$$

$$P\{Y = t^2\} = P\{(X - \mu)^2 = t^2 = P\{|X - \mu| = t\} =$$

$$= P\{X - \mu = t\} + P\{X - \mu = -t\}.$$

Chebyshev inequality - example

Example. Let X be a random variable with $\mathbb{E}[X] = 1$ and $Var[X] = 4$. Find (lower or upper) bounds for the following probabilities

$$P\{X \geqslant 3\}, P\{|X-1| \geqslant 6\} \text{ and } P\{X \leqslant -9\}.$$

Solution: Using Chebyshev inequality

$$P\{X \geqslant 3\} = P\{X-1 \geqslant 2\} \leqslant P\{|X-1| \geqslant 2\} \leqslant \frac{Var[X]}{2^2} = 1,$$

$$P\{|X-1| \geqslant 6\} \leqslant \frac{4}{36} = \frac{1}{9},$$

$$P\{X \leqslant -9\} = P\{X-1 \leqslant -10\} \leqslant P\{|X-1| \geqslant 10\} \leqslant \frac{1}{25} \clubsuit$$

Chernoff bounds

Theorem 3.1

Let $(X_i)_{1\leqslant i\leqslant n}$ *independent variables, each being Bernoulli distributed with a parameter* p_i*, respectively. If* $X = \sum_{i=1}^{n} X_i$, $\mu = \mathbb{E}[X]$, *then*

$$P\{X > (1+\delta)\mu\} < \left[\frac{e^{\delta}}{(1+\delta)^{1+\delta}}\right]^{\mu} \text{ (upper tail), } \forall \delta > 0 \text{ and}$$

$$P\{X < (1-\delta)\mu\} < \left[\frac{e^{-\delta}}{(1-\delta)^{1-\delta}}\right]^{\mu} \text{ (lower tail), } \forall \delta \in [0,1).$$

Chernoff bounds

- The first inequality says that the sum of Bernoulli independent variables exponentially decays as we move to the right of its expectation:

$$\lim_{\delta \to +\infty} \left[\frac{e^{\delta}}{(1+\delta)^{1+\delta}} \right]^{\mu} = 0.$$

- Both inequalities has simpler forms as we will see below.

Chernoff bounds

Corollary 3.1

Let $(X_i)_{1 \leqslant i \leqslant n}$ *be independent variables, having Bernoulli distribution with parameter* p_i. *If* $X = \sum_{i=1}^{n} X_i$, $\mu = \mathbb{E}[X]$, *then*

$$P\{X > (1 \pm \delta)\mu\} < \exp\left(\frac{-\delta^2\mu}{2+\delta}\right), \forall \delta \geqslant 0.$$

Chernoff bounds - an application

Application. We flip a coin n times; let X_i be a variable equal with 1 if we get the head at the i-th flip and 0 otherwise. $X = \sum_{i=1} X_i$ is the number of heads from all flips.

$$\mathbb{E}[X_i] = p_i = \frac{1}{2}, Var[X_i] = p_i(1 - p_i) = \frac{1}{4},$$

$$\mu = \mathbb{E}[X] = \sum_{i=1}^{n} \mathbb{E}[X_i] = \frac{n}{2} \text{ and } \sigma^2 = Var[X] = \sum_{i=1}^{n} Var[X_i] = \frac{n}{4}.$$

(This is a different way to compute the characteristics of a binomial variable.) Using Chernoff bounds we get

Chernoff bounds - an application

$$P\{X > \mu + \lambda\} = P\left\{X > \left(1 + \frac{\lambda}{\mu}\right)\mu\right\} < exp\left(\frac{-\lambda^2}{\lambda + 2\mu}\right) =$$
$$= \exp\left(\frac{-\lambda^2}{\lambda + n}\right).$$

We compare this result with those from Markov and Chebyshev inequalities:

$$P\{X \geqslant \mu + \lambda\} \leqslant \frac{\mu}{\mu + \lambda} = \frac{n}{n + 2\lambda} \text{ (Markov)},$$

$$P\{X \geqslant \mu + \lambda\} \leqslant P\{|X - \mu| \geqslant \lambda\} \leqslant \frac{\sigma^2}{\lambda^2} = \frac{n}{4\lambda^2} \text{ (Chebyshev)}.$$

Note that Markov inequality is weaker than Chebyshev inequality which is weaker than Chernoff bounds. On the other hand Markov inequality (and Chebyshev's) don't need the independence of the n variables.

Hoeffding bounds

Theorem 4.1

Let $X_1, X_2, \dots, X_n$ be independent bounded random variables: $a_i \leqslant X_i \leqslant b_i$, $a_i \neq b_i \in \mathbb{R}$, $i = \overline{1, n}$ and $X = \sum_{i=1}^{n} X_i$. Then

$$P\{X - \mathbb{E}[X] \geqslant \delta\} \leqslant \exp\left(-\frac{2\delta^2}{\sum_{i=1}^{n}(b_i - a_i)^2}\right), \forall \delta \geqslant 0.$$

Corollary 4.1

In the above conditions we have

$$P\{|X - \mathbb{E}[X]| \geqslant \delta\} \leqslant 2\exp\left(-\frac{2\delta^2}{\sum_{i=1}^{n}(b_i - a_i)^2}\right), \forall \delta \geqslant 0.$$

Exercises for seminar

- Joint Probability Distributions: I.2, II.2, II.5, II.7, II.13, II.14
- Markov and Chebyshev's Inequalities: III.2, III.6, III.7, III.8, III.10
- Reserve: II.6, II.7, III.1, III.9

End

Exercises - Covariance

I.1. Show that (X, Y and Z are random variables)

(a) $cov[aX + bY + c, Z] = a \cdot cov[X, Z] + b \cdot cov[Y, Z], \forall a, b, c \in \mathbb{R}$.

(b) $cov\left[\sum_{i=1}^{n} X_i, \sum_{j=1}^{m} Y_j\right] = \sum_{i=1}^{n}\sum_{j=1}^{m} cov[X_i, Y_j]$, for every random variables $(X_i)_{1\leqslant i\leqslant n}$ and $(Y_i)_{1\leqslant j\leqslant m}$ (use induction).

(c) $Var\left[\sum_{i=1}^{n} X_i\right] = \sum_{i=1}^{n} Var[X_i] + 2\sum_{i<j} cov[X_i, X_j]$, for every random variables $X_1, X_2, \ldots, X_n$.

I.2. Let X be the random variable from below and $Y = X^2$. Show that X and Y are not independent but $cov[X, Y] = 0$.

$$X : \begin{pmatrix} -1 & 0 & 1 \\ 0.25 & 0.5 & 0.25 \end{pmatrix}$$

Exercises - Joint Distributions and Independent Variables.

II.1. Suppose that X and Y have the following joint probability distribution

		Y		
		-3	2	4
X	1	0.2	?	0.2
	3	0.3	0.05	0.05

(a) Determine the distributions of X and Y.

(b) Compute $cov[X, Y]$ and $\rho[X, Y]$.

(c) Are X and Y independent?

Exercises - Joint Distributions and Independent Variables.

II.2. A coin is flipped three times. Let X equals 1 if we get tail at the first flip and 0 otherwise, and Y a variable which equals the number of tails in all three flips. Find

(a) the JPD of the two variables.

(b) the distributions of X and Y and their covariance.

II.3. Let X be a random variable and $Y = X^2$. We know that

$$\begin{pmatrix} -2 & -1 & 1 & 2 \\ \frac{1}{4} & \frac{1}{8} & \frac{1}{4} & \frac{3}{8} \end{pmatrix}.$$

Find

(a) the distribution of Y and the JPD of X and Y.

(b) the covariance and the correlation of X and Y.

Exercises - Joint Distributions and Independent Variables.

II.4. Let X and Y two random independent variables such that

$$X : \begin{pmatrix} 1 & -1 \\ 0.6 & 0.4 \end{pmatrix}, \; Y : \begin{pmatrix} -1 & 0 & 1 \\ 0.2 & 0.5 & 0.3 \end{pmatrix}.$$

(a) Determine their JPD and covariance.

(b) Determine the distribution and the expectation of $X + Y$.

II.5. In a box we have three red and five black balls. We withdraw a ball from the box and we replace it by one of opposite colour. Then we withdraw another ball from the box. Let X be the number of red and Y the number of black withdrawn balls.

(a) Determine the JPD of variables X and Y.

(b) Are X and Y independent?

Exercises - Joint Distributions and Independent Variables.

II.6. In a box we have four withe balls (two labeled with 1 and two labeled with 2) and three black balls (two labeled with 1 and one labeled 2). We with draw, one by one without replacement, two balls. Let X be the number of white balls and Y the number of balls labeled with 2.

(a) Find the JPD of X and Y.

(b) Variables X and Y are independent?

II.7. A coin is tossed three times. Let X be the number of tails obtained at the first two tossings and Y be the number of tails at the last tossing. Determine

(a) the distribution of X and Y.

(b) the JPD of X and Y. (Are X and Y independent?)

(c) the distribution of $X + Y$.

Exercises - Joint Distributions and Independent Variables.

II.8. X and Y have the following JPD

		X			
		-1	1	2	3
Y	-1	0	$1/36$	$1/6$	$1/12$
	0	$1/18$	0	$1/18$	0
	1	0	$1/36$	$1/6$	$1/12$
	2	$1/12$	0	$1/12$	?

(a) Compute $P(X \geqslant 2 \text{ and } Y \leqslant 0)$.

(b) Are X and Y independent?

(c) Determine the distribution of $X + Y$.

Exercises - Joint Distributions and Independent Variables.

II.9. Two players P_1 and P_2 compete in a match of tennis. The winner is the first player to win two sets in a best-of-three. P_1 independently wins a set with probability $1/3$. Let X be the number of the sets played by P_1 up to the end of the match and by Y be the number of sets P_2 wins in this match. Determine

(a) the joint probability distribution of X and Y;

(b) the covariance of the two variables; X and Y are independent variables?

II.10. We have a biased coin: the probability of heads in a ny given toss is $1/3$. We toss the coin three times. Let X be the number of tail occurrences and Y be the maximum number of head occurrences in a row. Find

(a) the joint probability distribution of X and Y;

(b) the covariance of the two variables; X and Y are independent variables?

Exercises - Joint Distributions and Independent Variables.

II.11. A die is rolled three times. X is a variable that denotes how many times we get an even number, and Y denotes how many times we get an prime number. Determine

- (a) the joint probability distribution of X and Y;
- (b) the covariance of the two variables; are X and Y independent?

II.12. A box contains 5 white and 4 red balls. We withdraw at random a ball from the box and we replaced it with one of opposite color. Then, we withdraw another ball from the box. Let X be the number of white balls and Y be the number of red balls obtained. Determine

- (a) the joint probability distribution of X and Y; what is the functional relation between the two random variables?
- (b) the covariance of the two variables; are X and Y independent?

Exercises - Joint Distributions and Independent Variables.

II.13. A box contains 3 black and 5 green balls. We withdraw a ball from the box and, if we get a black one we return the ball in the box together with a green one, otherwise we replace it with two black balls. Then, we withdraw another ball from the box. Let X be the number of black balls and Y be the number of green balls obtained. Find

(a) the joint probability distribution of X and Y

(b) the covariance of X and Y; are X and Y independent?

II.14. We have two boxes: B_1 contains 2 white and 2 black balls, and B_2 contains 1 white and 2 black balls. We roll a die and, if we get a multiple of 3 we withdraw a ball from B_1, otherwise we withdraw a ball from B_2. Let X be the the number of white balls remaining in B_1 and Y be the number of black balls remaining in B_2.

(a) Find the joint probability distribution of X and Y.

(b) Find the covariance of X and Y; are X and Y independent?

Exercises - Markov and Chebyshev's Inequalities

III.1. A random variable $X \geqslant 0$ has its expectation and variance both equal with 20. Using Markov and/or Chebyshev's inequalities find lower and/or upper bounds for $P\{X \geqslant 40\}$ and $P\{-60 \leqslant X \leqslant 100\}$.

III.2. Let $X \geqslant 0$ be random variable with $\mathbb{E}[X] = Var[X] = 1$. Using Markov and/or Chebyshev's inequalities find lower and/or upper bounds for

$$P\{X \geqslant 2\}, P\{|X - 1| \geqslant 2\}, P\{X \leqslant -3\}.$$

III.3. Let $X \geqslant 0$ be random variable with $\mathbb{E}[X] = Var[X] = 2$. Using Markov and/or Chebyshev's inequalities find lower and/or upper bounds for $P\{X \geqslant 8\}$ and $P\{|X - 2| \geqslant 8\}$.

III.4. Let $X \geqslant 0$ be random variable with $\mathbb{E}[X] = 2$ and $Var[X] = 1$. Using Markov and/or Chebyshev's inequalities find lower and/or upper bounds for $P\{X \geqslant 6\}$ and $P\{|X - 1| \geqslant 5\}$.

Exercises - Markov and Chebyshev's Inequalities

III.5. A random variable $X \geqslant 0$ has $\mathbb{E}[X] = 2$ and $Var[X] = 3$. Using Markov and/or Chebyshev's inequalities what can you say about $P\{X \geqslant 8\}$ and $P\{|X - 2| \geqslant 4\}$?

III.6. The probability to get the head on a biased coin is 0.3. We toss this coin 300 times. Find an upper bound for the probability that we get the head at least 100 times.

III.7. The probability to get the head on a biased coin is 0.2. We toss this coin n times. Find an upper bound for the probability that we get the head at least 50% times.

III.8. Two biased coins have the probabilities of the tail occurrence 0.25 and 0.8, respectively. The coins are 25 times flipped. Using Markov and Chebyshev inequalities find upper bounds for the probability that we get two tails at least 10 times.

Exercises - Markov and Chebyshev's Inequalities

III.9. We roll two dice 36 times. Using Markov and Chebyshev inequalities find upper bounds for the probability that we get two a product which is a prime number at least 10 times.

III.10. Two biased coins are flipped 32 times. The probability of getting a tail is 1/3 for the first coin and 3/4 for the second. Using Markov and Chebyshev inequalities find upper bounds for the probability that we get tail on both coins at least 12 times.

III.11. We toss a fair coin n times. Let X be the number of tails we get. Find upper bounds for

(a) $P\{|X - n/2| > \sqrt{n}\}$ and $P\{X > n/2 + \sqrt{n}\}$;

(b) $P\{|X - n/2| > 5\sqrt{n}\}$ and $P\{X > n/2 + 5\sqrt{n}\}$.

III.12. Let X be a Poisson distributed variable with cu parameter λ. Estimate the probability that X deviates from $\mathbb{E}[X]$ with at least $2\sqrt{\lambda}$.

Exercises - Markov and Chebyshev's Inequalities

III.13*. (Borel-Cantelli lemma) Let $(A_n)_{n \geqslant 1}$ be a sequence of random events such that $\sum_{n \geqslant 1} P(A_n) < +\infty$. Prove that the probability of occurrence of at most k of these events is at least

$$1 - \frac{\sum_{n \geqslant 1} P(A_n)}{k}.$$

(Hint: Use Markov's inequality for a variable which numbers the occurring events.)

Annex

proof: (for Theorem 3.1) Let $t > 0$; by Markov inequality we have

$$P\{X > (1+\delta)\mu\} = P\{e^{tX} > e^{t(1+\delta)\mu}\} < \frac{\mathbb{E}\left[e^{tX}\right]}{e^{t(1+\delta)\mu}}.$$

Using the independence we have

$$\mathbb{E}\left[e^{tX}\right] = \mathbb{E}\left[\exp\left(t\sum_{i=1}^{n} X_i\right)\right] = \mathbb{E}\left[\prod_{i=1}^{n} e^{tX_i}\right] =$$

$$= \prod_{i=1}^{n} \mathbb{E}\left[e^{tX_i}\right] = \prod_{i=1}^{n} \left[p_i e^t + (1-p_i)\right],$$

Thus

$$P\{X > (1+\delta)\mu\} < \frac{\prod_{i=1}^{n}\left[1 + p_i(e^t - 1)\right]}{e^{t(1+\delta)\mu}}.$$

Annex

Using the inequality $x+1<e^x, \forall x\in\mathbb{R}$, we get

$$P\{X>(1+\delta)\mu\}<\frac{\prod_{i=1}^{n}\left[\exp\left(p_i(e^t-1)\right)\right]}{e^{t(1+\delta)\mu}}=\frac{\exp\left(\sum_{i=1}^{n}p_i(e^t-1)\right)}{e^{t(1+\delta)\mu}},$$

hence

$$P\{X>(1+\delta)\mu\}<\frac{\exp\mu(e^t-1)}{\exp t\mu(1+\delta)}, \forall t>0.$$

We look for the minimum of $f:\mathbb{R}_+^*\to\mathbb{R}$, $f(t)=(e^t-1)\mu-t\mu(1+\delta)$:

$$f'(t)=\mu(e^t-1-\delta), f'(t)=0\Leftrightarrow t=\ln(1+\delta),$$

Annex

f is nod-increasing on $(0, \ln(1+\delta)]$ and increasing on $[\ln(1+\delta), +\infty)$.
We get

$$P\{X > (1+\delta)\mu\} < \left[\frac{e^{\delta}}{(1+\delta)^{1+\delta}}\right]^{\mu}.$$

A similar proof for the second inequality from the theorem. ■

Annex

proof: (for Corollary 3.1) The upper bound from Theorem 3.1 is

$$\left[\frac{e^{\delta}}{(1+\delta)^{1+\delta}}\right]^{\mu} = \left(\exp\left[\delta - (1+\delta)\ln(1+\delta)\right]\right)^{\mu}.$$

It can be shown that $ln(1+\delta) > \dfrac{2\delta}{1+\delta}, \forall\delta > 0$, therefore

$$\delta - (1+\delta)\ln(1+\delta) \leqslant \frac{-\delta^2\mu}{2+\delta}.$$

∎

Annex

proof: (for Theorem 4.1) We prove only one of the inequalities:

$$P\{X - \mathbb{E}[X] \geqslant \delta\} = P\{e^{tX} \geqslant e^{t(\delta + \mathbb{E}[X])}\} \overset{(Markov)}{\leqslant} \frac{\mathbb{E}\left[e^{tX}\right]}{e^{t(\delta+\mathbb{E}[X])}} \leqslant$$

$$\overset{(indep.)}{\leqslant} \exp\left[-t(\mathbb{E}[X] + \delta)\right] \cdot \prod_{i=1}^{n} \mathbb{E}\left[e^{tX_i}\right] = e^{-t\delta} \cdot \prod_{i=1}^{n} \mathbb{E}\left[e^{tX_i}\right] e^{-t\mathbb{E}[X_i]}.$$

We analyze the function $f : \mathbb{R}_+ \to \mathbb{R}_+$, $f(t) = e^t$:

$f'(t) = f''(t) = e^t > 0, \forall t \in \mathbb{R}_+$, hence f is concave on $\mathbb{R}_+$, i. e.,

$$f\left[\lambda t_1 + (1-\lambda)t_2\right] \leqslant \lambda f(t_1) + (1-\lambda) f(t_2), \forall t_1, t_2 \in \mathbb{R}_+, \lambda \in [0,1].$$

Annex

As $X_i \in [a_i, b_i]$, we have $X_i = \lambda a_i + (1-\lambda) b_i$, where $\lambda = \dfrac{X_i - a_i}{b_i - a_i} \in [0,1]$.
Thus,

$$e^{tX_i} = exp[t\lambda a_i + t(1-\lambda)b_i] \leqslant \lambda e^{ta_i} + (1-\lambda)e^{tb_i} \text{ therefore}$$

$$\mathbb{E}\left[e^{tX_i}\right] \leqslant \mathbb{E}\left[\lambda e^{ta_i} + (1-\lambda)e^{tb_i}\right] = e^{ta_i}\mathbb{E}\left[\frac{X_i - a_i}{b_i - a_i}\right] + e^{tb_i}\mathbb{E}\left[\frac{b_i - X_i}{b_i - a_i}\right],$$

$$\mathbb{E}\left[e^{t(X_i - \mathbb{E}[X_i])}\right] \leqslant e^{-t\mathbb{E}[X_i]}\left(e^{ta_i} \cdot \frac{\mathbb{E}[X_i] - a_i}{b_i - a_i} + e^{tb_i} \cdot \frac{b_i - \mathbb{E}[X_i]}{b_i - a_i}\right).$$

We will show that this expression is smaller or equal with

$$\exp\left[\frac{t^2(b_i - a_i)^2}{8}\right].$$

Annex

Let $\theta = t(b_i - a_i)$ and $\alpha = \dfrac{\mathbb{E}[X_i] - a_i}{b_i - a_i}$, and consider the function $g : \mathbb{R}_+ \to \mathbb{R}$ (by applying the logarithm on the above expression):

$$g(\theta) = -\theta\alpha + \ln\left(1 - \alpha + \alpha e^{\theta}\right) - \frac{\theta^2}{8}$$

and show that $g(\theta) \leqslant 0$, $\forall \theta \in \mathbb{R}_+$.

$$g'(\theta) = -\alpha - \frac{\theta}{4} + \frac{\alpha e^{\theta}}{(1-\alpha) + \alpha e^{\theta}} = -\alpha - \frac{\theta}{4} + \frac{\alpha}{(1-\alpha)e^{-\theta} + \alpha},$$

$$g''(\theta) = -\frac{1}{4} + \frac{\alpha(1-\alpha)e^{-\theta}}{\left[(1-\alpha)e^{-\theta} + \alpha\right]^2} =$$

$$= -\frac{1}{4} + \frac{\alpha}{(1-\alpha)e^{-\theta} + \alpha} \cdot \frac{(1-\alpha)e^{-\theta}}{(1-\alpha)e^{-\theta} + \alpha} \leqslant$$

Annex

$$\leqslant -\frac{1}{4} + \frac{1}{4}\left(\frac{\alpha}{(1-\alpha)e^{-\theta}+\alpha} + \frac{(1-\alpha)e^{-\theta}}{(1-\alpha)e^{-\theta}+\alpha}\right)^2 = 0$$

From Taylor's formula it exists a $\theta_0 \in [0, \theta]$, such that

$$g(\theta) = g(0) + g'(0)\frac{\theta}{1!} + g''(\theta_0)\frac{\theta_0^2}{2!} \leqslant 0.$$

Hence

$$\mathbb{E}\left[e^{t(X_i - \mathbb{E}[X_i])}\right] \leqslant \exp\left[\frac{t^2(b_i - a_i)^2}{8}\right], \forall i$$

and

$$P\{X - \mathbb{E}[X] \geqslant \delta\} \leqslant \exp\left[\frac{t^2}{8}\sum_{i=1}^{n}(b_i - a_i)^2 - t\delta\right], \forall t \geqslant 0.$$

Annex

The above exponent is a quadratic function in t, and it reach its miminum in $t = \frac{4\delta}{\sum_{i=1}^{n}(b_i - a_i)^2}$:

$$\exp\left[\frac{t^2}{8}\sum_{i=1}^{n}(b_i - a_i)^2 - t\delta\right] \geqslant \exp\left[-\frac{2\delta^2}{\sum_{i=1}^{n}(b_i - a_i)^2}\right].$$

Bibliography

- Bertsekas, D. P., J. N. Tsitsiklis, *Introduction to Probability*, Athena Scietific, 2002.
- Gordon, H., *Discrete Probability*, Springer Verlag, New York, 1997.
- Hoeffding, W., *Probability inequalities for sums of bounded random variables*, J. of the Amer. Statistical Assoc. vol. 58, issue 301, pp. 13-30, 1963.
- Lipschutz, S., *Theory and Problems of Probability*, Scahaum's Outline Series, McGraw-Hill, 1965.
- Ross, S. M., *A First Course in Probability* , Prentice Hall, 5th edition, 1998.
- Stone, C. J., *A Course in Probability and Statistics*, Duxbury Press, 1996.