Probability Theory Notes

January 4, 2022

Contents

Chapter 1

Числові характеристики випадкових величин

1.1 Попередні зауваження

Розглянемо дискретну випадкову величину ξ

$$\xi(\omega) = \sum_{i=1}^{n} x_i \mathbb{1}_{A_i}(\omega)$$

$$\{A_1, \dots, A_n\} \text{ - повна група подій}$$

Проведено n незалежних випробувань в кожному з яких спостерігається

$$\xi_n(\omega) = \sum_{i=1}^{m} x_i \cdot \mathbb{1}^n_{A_i}(\omega)$$

Розглянемо

$$\hat{\xi} = \frac{\xi_1 + \dots + \xi_n(\omega)}{n} = \frac{1}{n} \sum_{i=1}^{n} \sum_{j=1}^{m} x_j \cdot \mathbb{1}^i_{A_j}(\omega) =$$

$$= \sum_{i=1}^{m} x_j \cdot \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}^i_{A_j}(\omega)$$

$$\frac{1}{n} \sum_{i=1}^{n} \mathbb{1}^i_{A_j}(\omega) \text{ - частота появи } A_j \text{ в } n \text{ випробуваннях } \to_{n\to\infty} P(A_j).$$

$$\hat{\xi} = \frac{\xi_1 + \dots + \xi_n}{n} \to_{n\to\infty} \sum_{j=1}^{m} x_j \cdot P(A_j).$$

Припустимо $\Omega[0,1];\ \mathcal{F} = \mathcal{B}([0,1])$, P міра Лебега, $P((a,b]) = b - a$
для дискретної ймовірності:

$$S_1 = x_1 \cdot P(A_1) = x_1 \cdot |A_1|$$
$$S_2 = x_2 \cdot P(A_2) = x_2 \cdot |A_2|$$
$$S \sim \sum_{j=1}^{m} x_j P(A_j) \text{ - площа}$$

для неперевного випадку:

$$\hat{\xi} \sim \int\limits_{\Omega} \xi(\omega) P(d\omega).$$

1.2 Definition and examples of expected value

Нехай $(\omega, \mathcal{F}, P)$ - ймовірністний простір. ξ - випадкова величина на цьому просторі.

Definition 1. *Математичним сподіванням випадкової величини ξ називається число*

$$M\xi = \int\limits_{\Omega} \xi(\omega) P(d\omega).$$

$$\textit{(expectation)} \quad E\xi = \int\limits_{\Omega} \xi(\omega) P(d\omega).$$

ξ індукує міру P_ξ на $\mathbb{R}$:

$$P_\xi((a, b]) = F_\xi(b) - F_\xi(a).$$

Заміна $\xi(\omega) = x$ приводить до інтеграла Лебега-Стілтьєса:

$$M\xi = \int\limits_{\mathbb{R}} x P_\xi(dx).$$

Звідси маємо інтеграл Стілтьєса:

$$M\xi = \int\limits_{\mathbb{R}} x dF_\xi(x).$$

Для дискретної випадкової величини ξ:

$$E\xi = \sum_{i=1}^{\infty} x_i \cdot P(\xi = x_i) \tag{1.1}$$

Якщо ξ має щільність $f_\xi(x)$:

$$E\xi = \int\limits_{\mathbb{R}} x f_x(x) dx \tag{1.2}$$

Remark. *It is considered that expectation exists if series (1.1) or integral (1.2) is absolutely convergent.*

Example. *If $A \in \mathcal{F}$ then $\xi(\omega) = \mathbb{1}_A(\omega)$*

$$E\xi = 0 \cdot P(\xi = 0) + 1 \cdot P(\xi = 1) = P(A).$$

Example.

$$P(\xi = i) = \frac{1}{i(i+1)}, i = 1, 2, \ldots.$$

$$\sum_{i=1}^{\infty} i \cdot P(\xi = i) = \sum_{i=1}^{\infty} \frac{1}{i+1} = +\infty \Rightarrow E\xi \textit{ does not exist}.$$

Example.

$$\xi \sim U(a,b);\ f_\xi(x) = \frac{1}{b-a}\mathbb{1}(x \in (a,b]).$$

$$E\xi = \frac{1}{b-a}\int\limits_a^b x dx = \frac{b^2-a^2}{2(b-a)} = \frac{a+b}{2}.$$

For uniform distribution the expectation is the middle of the segment.

Example.

$$\xi \sim C(0,1);\ f_\xi(x) = \frac{1}{\pi(a+x^2)}.$$

Whereas $\int\limits_{-\infty}^{\infty} \frac{xdx}{\pi(1+x^2)}$ *- divergent then* $E\xi$ *does not exist.*

Let g - Borel function. Then $g(\xi)$ - stochastic variable. For $Mg(\xi)$ have:

$$Eg(\xi) = \int\limits_\Omega g(\xi(\omega))P(d\omega) = \int\limits_\mathbb{R} g(x)P_\xi(dx) = \int\limits_\mathbb{R} g(x)dF_\xi(x).$$

For discrete stochastic variable:

$$Eg(\xi) = \sum_{i=1}^{+\infty} g(x_i)\cdot P(\xi = x_i).$$

For absolutely continuous:

$$Eg(\xi) = \int\limits_\mathbb{R} g(x)f_\xi(x)dx.$$

If $\xi = (\xi_1, \dots, \xi_n)$ with density $f_\xi(x_1, \dots, x_n),\ g: \mathbb{R}^n \to \mathbb{R}$ - Borel function.

$$Eg(\xi_1, \dots, \xi_n) = \int \cdots \int\limits_{\mathbb{R}^n} g(x_1, \dots, x_n)f_\xi(x_1, \dots, x_n)dx_1 \dots dx_n.$$

Theorem 1.2.1. *Properties of expectation*

1. $Ec = c,\ c = const$
2. $E(a\xi + b) = a\cdot E\xi + b,\ a, b = const$
3. $E(\xi_1 + \xi_2) = E\xi_1 + E\xi_2$
4. $E[\xi_1 \cdot \xi_2] = E\xi_1 \cdot E\xi_2$
   ξ_1, ξ_2 *are independent stochastic variables*
5. $\xi \geq 0 \Rightarrow M\xi \geq 0$
   $\xi \leq \eta \Rightarrow E\xi \leq E\eta$
6. $|E\xi| \leq E|\xi|$

Proof. 4. Let ξ_1, ξ_2 - absolutely continuous stochastic variables with densities

$$f_{\xi_1}(x), f_{\xi_2}(y).$$

$$E[\xi_1, \xi_2] = \iint\limits_{\mathbb{R}^2} x \cdot y \cdot f_{(\xi_1,\xi_2)}(x,y)dxdy =$$
$$= \iint\limits_{\mathbb{R}^2} x \cdot y \cdot f_{\xi_1}(x) \cdot f_{\xi_2}(y)dxdy =$$
$$= \int\limits_{\mathbb{R}} xf_{\xi_1}(x)dx \int\limits_{\mathbb{R}} yf_{\xi_2}(y)dy = E\xi_1 \cdot E\xi_2.$$

□

Remark. *For arbitrary number of stochastic variables:*

$$E(\xi_1 + \dots + \xi_n) = \sum_{i=1}^{n} E\xi_i.$$

$$E(\xi_1 \cdot \dots \cdot \xi_n) = \prod_{i=1}^{n} E\xi_i.$$

for $\xi_1, \dots, \xi_n$ that are independent together.

Example. *let $\xi \sim Bin(n,p)$; $\;E\xi-?$*

$$P(\xi = k) = C_n^k p^k (1-p)^{n-k},\; k = \overline{0,n}.$$

$$M\xi = \sum_{k=0}^{n} k \cdot C_n^k p^k (1-p)^{n-k}.$$

Using:

$$\xi = \sum_{i=1}^{n} \xi_i \text{ where } \xi_i \sim B(p) : .$$
$$P(\xi_i = 1) = p; P(\xi_i = 0) = 1 - pM\xi_i = 1 \cdot p + 0 \cdot (1-p) = p.$$

Then:

$$M\xi = \sum_{i=1}^{n} M\xi_i = n \cdot p.$$

1.3 Dispersion

Definition 2. *Dispersion of stochastic variable is called a number*

$$\mathcal{D}\xi = M(\xi - M\xi)^2.$$

Remark.

$$\mathcal{D}\xi = M(\xi^2 - 2M\xi \cdot \xi + (M\xi)^2) = M\xi^2 - 2M\xi \cdot M\xi + (M\xi)^2 = M\xi^2 - (M\xi)^2.$$

$$\mathcal{D}\xi = M\xi^2 - (M\xi)^2 \tag{1.3}$$

Definition 3. *Number $M\xi^2$ is called second momentum os stochastic variable ξ.*

Example.

.

$$\xi \sim B(p);\ M\xi = p;$$
$$M\xi^2 = 1 \cdot P(\xi = 1) + 0 \cdot P(\xi = 0) = p\mathcal{D}\xi = p - p^2 = p(1-p).$$

Example.

$$\xi \sim U(a,b);\ M\xi = \frac{a+b}{2}$$
$$M\xi^2 = \int\limits_a^b x^2 \frac{1}{b-a} dx = \frac{b^2 - a^2}{3(b-a)} = \frac{a^2 + ab + b^2}{3}$$
$$\mathcal{D}\xi = \frac{a^2 + ab + b^2}{3} - \frac{(a+b)^2}{4} = \frac{4(a^2+ab+b^2) - 3(a+b)^2}{12} = \frac{(b-a)^2}{12}$$

Theorem 1.3.1 (properties of dispersion). *1.*

$$\mathcal{D}\xi \geq 0$$
$$\mathcal{D}\xi = 0 \iff \xi = c = const$$

2.

$$\mathcal{D}(a\xi + b) = a^2 \cdot \mathcal{D}\xi$$

3. If ξ_1 and ξ_2 are independent, then

$$\mathcal{D}(\xi_1 + \xi_2) = \mathcal{D}\xi_1 + \mathcal{D}\xi_2$$

Proof. 1.

$$\mathcal{D}\xi = M(\xi - M\xi)^2$$
$$(\xi - M\xi)^2 \geq 0 \Rightarrow M(\xi - M\xi)^2 \geq 0$$
$$M(\xi - M\xi)^2 = 0 \iff (\xi - M\xi)^2 = 0 \iff \xi = M\xi = const$$

2.

$$\mathcal{D}(a\xi + b) = M((a\xi + b) - m(a\xi + b))^2 = M(a\xi + b - aM\xi - b)^2 = Ma^2(\xi - m\xi)^2 =$$
$$= a^2 \cdot M(\xi - M\xi)^2 = a^2 \cdot \mathcal{D}\xi.$$

3. Let ξ_1 and ξ_1 independent.

$$\mathcal{D}(\xi_1 + \xi_2) = M(\xi_1 + \xi_2 - M(\xi_1 + \xi_2))^2 = M((\xi_1 - M\xi_1) + (\xi_2 - M\xi_2))^2 =$$
$$= M((\xi_1 - M\xi_1)^2 + 2(\xi_1 - M\xi_1)(\xi_2 - M\xi_2) + (\xi_2 - M\xi_2)^2) =$$
$$= \mathcal{D}\xi_1 + 2 \cdot M[(\xi_1 - M\xi_1)(\xi_2 - M\xi_2)] + \mathcal{D}\xi_2.$$
$$\xi_1, \xi_2 \text{ independent } \Rightarrow \mathcal{D}\xi_1 + \mathcal{D}\xi_2 + 2M(\xi_1 - M\xi_1) \cdot M(\xi_2 - M\xi_2) = \mathcal{D}\xi_1 + \mathcal{D}\xi_2.$$

□

Example.

$$\xi \sim \textit{Bin}(n,p);\ M\xi = n\cdot p;\ \mathcal{D}\xi = ?$$
$$M\xi^2 = \sum_{k=0}^{n} k^2 \cdot C_n^k \cdot p^k \cdot (1-p)^{n-k}$$
$$\xi = \sum_{i=1}^{n} \xi_i,\ \xi_i \sim B(p),\ \xi_2,\dots,\xi_n \textit{ - independent}$$
$$\mathcal{D}\xi = \sum_{i=1}^{n} \mathcal{D}\xi_i = \sum_{i=1}^{n} p\cdot(1-p) = np(1-p).$$

Remark.

$$M\xi = \underset{a}{\textit{argmin}} M(\xi-a)^2.$$

$$M(\xi-a)^2 = M((\xi-M\xi)+(M\xi-a))^2 =$$
$$M(\xi-M\xi)^2 + 2(M\xi-a)M(\xi-M\xi) + (M\xi-a)^2 =$$
$$\mathcal{D}\xi + (M\xi-a)^2 \geq \mathcal{D}\xi$$
$$\textit{moreover } M(\xi-a)^2 = \mathcal{D}\xi \iff (M\xi-a)^2 = 0$$
$$\Rightarrow a = M\xi.$$

Example. *Numerical characteristics of the main probability distributions*

1. $\xi \sim B(p), M\xi = p, \mathcal{D}\xi = p(1-p)$
2. $\xi \sim \textit{Bin}(n,p), M\xi = np,\ \ \mathcal{D}\xi = np(1-p)$
3. $\xi \sim \textit{Poiss}(\lambda)$

$$M\xi = \sum_{k=0}^{\infty} k\cdot\frac{\lambda^k}{k!}e^{-\lambda} = e^{-\lambda}\cdot\sum_{k=1}^{\infty}\frac{\lambda^k}{(k-1)!} = \lambda e^{-\lambda}\cdot\sum_{k=0}^{\infty}\frac{\lambda^k}{k!} = \lambda e^{-\lambda}\cdot e^{\lambda} = \lambda$$
$$M\xi^2 = \sum_{k=0}^{\infty} k^2\cdot\frac{\lambda^k}{k!}e^{-\lambda} = e^{-\lambda}\sum_{k=1}^{\infty} k\frac{\lambda^k}{(k-1)!} =$$
$$e^{-\lambda}\sum_{k=1}^{\infty}((k-1)+1)\cdot\frac{\lambda^k}{(k-1)!} = e^{-\lambda}(\lambda^2\sum_{k=2}^{\infty}\frac{\lambda^{k-1}}{(k-2)!} + \lambda\sum_{k=1}^{\infty}\frac{\lambda^{k-1}}{(k-1)!}) =$$
$$= \lambda^2 e^{-\lambda}e^{\lambda} + \lambda\cdot e^{-\lambda}\cdot e^{\lambda} = \lambda^2+\lambda$$
$$\mathcal{D}\xi = \lambda^2+\lambda-\lambda^2 = \lambda$$

4. $\xi \sim \textit{Geom}(p) : p(\xi = k) = (1-p)^k,\ k = 0, 1, \dots$

$$M\xi = \sum_{k=0}^{\infty} k \cdot (1-p)^k \cdot p$$

$$\sum_{k=0}^{\infty} (1-p)^k = \frac{1}{p}$$

$$\sum_{k=0}^{\infty} k(1-p)^{k-1} = \frac{1}{p^2} \quad | \ p \cdot (1-p)$$

$$\sum_{k=0}^{\infty} k(1-p)^k \cdot p = \frac{1-p}{p}$$

$$M\xi^2 = \sum_{k=0}^{\infty} k^2(1-p)^k \cdot p$$

$$\sum_{k=0}^{\infty} k(1-p)^k = \frac{1-p}{p^2} = \frac{1}{p^2} - \frac{1}{p}$$

$$\sum_{k=0}^{\infty} k^2(1-p)^{k-1} = \frac{2}{p^3} - \frac{1}{p^2} \quad | \ (1-p) \cdot p$$

$$\sum_{k=0}^{\infty} k^2(1-p)^k \cdot p = \frac{2(1-p)}{p^2} - \frac{1-p}{p}$$

$$\mathcal{D} = \frac{2(1-p)}{p^2} - \frac{1-p}{p} - \left(\frac{1-p}{p}\right)^2 = \frac{2(1-p)}{p^2} - \frac{1-p}{p}\left(1 + \frac{1-p}{p}\right) =$$

$$= \frac{2(1-p)}{p^2} - \frac{1-p}{p^2} = \frac{1-p}{p^2};$$

$$\mathcal{D}\xi = \frac{1-p}{p^2}; \quad M\xi = \frac{1-p}{p}$$

5. $\xi \sim U(a,b); \quad M\xi = \frac{a+b}{2}; \quad \mathcal{D}\xi = \frac{(b-a)^2}{12}$

6. $\xi \sim \textit{Exp}(\lambda) : \ f_\xi(x) = \lambda e^{-\lambda x} \cdot \mathbb{1}(x \geq 0)$

$$M\xi = \int_0^{\infty} x\lambda e^{-\lambda x} dx = -\int_0^{\infty} x de^{-\lambda x} = -x \cdot e^{-\lambda x} \ |_0^{\infty} + \int_0^{\infty} e^{-\lambda x} dx = \frac{1}{\lambda}.$$

$$M\xi^2 = \int_0^{\infty} x^2\lambda e^{-\lambda x} dx = \int_0^{\infty} x^2 de^{-\lambda x} = \int_0^{\infty} 2xe^{-\lambda x} dx = \frac{2}{\lambda}\int_0^{\infty} x\lambda e^{-\lambda x} dx = \frac{2}{\lambda^2}$$

$$\mathcal{D}\xi = \frac{2}{\lambda^2} - \frac{1}{\lambda^2} = \frac{1}{\lambda^2} \quad M\xi = \frac{1}{\lambda}$$

7. $\xi \sim \textit{N}(a, \sigma^2);$

$$f_\xi(x) = \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{(x-a)^2}{2\sigma^2}}$$

$$M\xi = a; \quad \mathcal{D}\xi = \sigma^2$$

$$M\xi = \int\limits_{\mathbb{R}} x\frac{1}{\sqrt{2\pi}\sigma}e^{\frac{(x-a)^2}{2\sigma^2}}dx =$$

$$\frac{x-a}{\sigma} = z$$

$$dx = \sigma dz$$

$$= \frac{1}{\sqrt{2\pi}}\int\limits_{\mathbb{R}}(\sigma z + a)e^{-z^2/2}dz == \frac{\sigma}{\sqrt{2\pi}} \cdot \int\limits_{\mathbb{R}} ze^{-z^2/2}dz + \frac{a}{\sqrt{2\pi}}\int\limits_{\mathbb{R}} e^{-z^2/2}dz = a$$

$$\mathcal{D}\xi = M(\xi - M\xi)^2 = M(\xi - a)^2 = \int\limits_{\mathbb{R}}(x-a)^2\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(x-a)^2}{2\sigma^2}}dx =$$

$$\frac{x-a}{\sigma} = z$$

$$dx = \sigma dz$$

$$= \frac{1}{\sqrt{2\pi}}\int\limits_{\mathbb{R}}\sigma^2 z^2 \cdot e^{-z^2/2}dz = \frac{2\sigma^2}{\sqrt{2\pi}}\int\limits_0^\infty z^2 e^{-z^2/2}dz = -\frac{2\sigma^2}{\sqrt{2\pi}}\int\limits_0^\infty zde^{-z^2/2} =$$

$$= -\frac{2\sigma^2}{\sqrt{2\pi}}z \cdot e^{-z^2/2}\mid_0^\infty + \frac{2\sigma^2}{\sqrt{2\pi}}\int\limits_0^\infty e^{-z^2/2}dz = \sigma^2 \cdot \frac{2}{\sqrt{2\pi}}\int\limits_0^\infty e^{-z^2/2}dz =$$

$$= \sigma^2 \cdot \frac{1}{\sqrt{2\pi}}\int\limits_{-\infty}^{+\infty} e^{-z^2/2}dz = \sigma^2$$

Chapter 2

Covariance of random variables. Correlation coefficient.

2.1 Covariance of random variables

Consider $\xi = (\xi_1, \xi_2)$ - random vector.

Definition 4. *Covariation of stochastic variables ξ_1, ξ_2 is a number:*

$$\mathit{cov}(\xi_1, \xi_2) = M[(\xi_1 - M\xi_1) \cdot (\xi_2 - M - \xi_2)] \tag{2.1}$$

(assuming that $M\xi_i$ exist)

If ξ_1, ξ_2 are discrete random variables, then

$$\text{cov}(\xi_1, \xi_2) = \sum_{i=1}^{\infty} \sum_{j=1}^{\infty} (x_i - M\xi_1) \cdot (y_j - M\xi_2) \cdot P(\xi_1 = x_i, \xi_2 = y_j). \tag{2.2}$$

If ξ_1, ξ_2 have common distribution density $f_\xi(x, y)$, then

$$\text{cov}(\xi_1, \xi_2) = \int\limits_{\mathbb{R}^2}\int (x - M\xi_1)(y - M\xi_2) f_\xi(x, y) dx dy \tag{2.3}$$

From definition Rightarrow:

$$\text{cov}(\xi_1, \xi_2) = M[\xi_1 \cdot \xi_2 - \xi_1 \cdot M\xi_2 - \xi_2 \cdot M\xi_1 + M\xi_1 \cdot M\xi_2] =$$
$$M[\xi_1 \cdot \xi_2] - M\xi_2 \cdot M\xi_1 - M\xi_1 \cdot M\xi_2 + M\xi_1 \cdot M\xi_2 = M[\xi_1 \cdot \xi_2] - M\xi_1 \cdot M\xi_2.$$

$$\text{cov}(\xi_1, \xi_2) = M[\xi_1 \cdot \xi_2] - M\xi_1 \cdot M\xi_2 \tag{2.4}$$

Proposition 2.1.1. *If ξ_1, ξ_2 are independent, then*

$$\mathit{cov}(\xi_1, \xi_2) = 0.$$

It is said, that ξ_1, ξ_2 are uncorrelated.

Indeed, if ξ_1, ξ_2 are independent, then from the properties of expectation:

$$M[\xi_1 \cdot \xi_2] = M\xi_1 \cdot M\xi_2.$$

then

$$\mathit{cov}(\xi_1, \xi_2) = M\xi_1 \cdot M\xi_2 - M\xi_1 \cdot M\xi_2 = 0.$$

Inverse statement is not true: from uncorrelated does not Rightarrow independency.

Remark.

$$\mathcal{D}(\xi_1+\xi_2) = \mathcal{D}\xi_1 + \mathcal{D}\xi_2 + 2M(\xi_1 - M\xi_1)(\xi_2 - M\xi_2) = \mathcal{D}\xi_1 + \mathcal{D}\xi_2 + 2\,\text{cov}(\xi_1,\xi_2).$$

Theorem 2.1.2. *Main properties of variation:*

1. $$\text{cov}(\xi,\xi) = \mathcal{D}\xi$$

2. $$\text{cov}(a_1\xi_2 + b_1, a_2\xi_2 + b_2) = a_1 \cdot a_2\,\text{cov}(\xi_1,\xi_2)$$

3. $$|\,\text{cov}(\xi_1,\xi_2)| \le \sqrt{\mathcal{D}\xi_1 \cdot \mathcal{D}\xi_2}$$

4. *Equality* $|\,\text{cov}(\xi_1,\xi_2)| = \sqrt{\mathcal{D}\xi_1 \cdot \mathcal{D}\xi_2}$ *is true if and only if* ξ_1 *and* ξ_2 *and linearly dependent.*

$$\exists a, b = const : \xi_2 = a\xi_2 - b.$$

Proof. 1.

$$\text{cov}(\xi,\xi) = M[\xi\cdot\xi] - M\xi \cdot M\xi = M\xi^2 - (M\xi)^2 = \mathcal{D}\xi$$

2.

$$\begin{gathered}\text{cov}(a_1\xi_1 + b_1, a_2\xi_2 + b_2) = M(a_1\xi_2 + b_1 - (a_1 M\xi_1 + b_1)) \cdot (a_2\xi_2 + b_2 - (a_2 M\xi_2 + b_2)) = \\ = M(a_1(\xi_1 - M\xi_1) \cdot a_2(\xi_2 - M\xi_2)) = a_1 \cdot a_2 \cdot M((\xi_1 - M\xi_1)(\xi_2 - M\xi_2)) = \\ = a_1 \cdot a_2 \cdot \text{cov}(\xi_1,\xi_2)\end{gathered}$$

3. Consider stochastic variable:

$$\eta(x) = x\cdot\xi_1 - \xi_2, \quad x \in \mathbb{R}$$

$$\mathcal{D}\eta(x) = \mathcal{D}(x\xi_1 - \xi_2) = x^2\cdot\mathcal{D}\xi_1 + \mathcal{D}\xi_2 - 2x\cdot\text{cov}(\xi_1,\xi_2)$$

As $\mathcal{D}\eta(x) \ge 0 \ \ \forall x \in \mathbb{R}$, so discriminant in the right part is not positive.

$$\mathcal{D} = (2\,\text{cov}(\xi_1,\xi_2))^2 - 4disp\xi_1 \cdot \mathcal{D}\xi_2 \le 0$$

$$\Rightarrow |\,\text{cov}(\xi_1,\xi_2)| \le \sqrt{\mathcal{D}\xi_1 \cdot \mathcal{D}\xi_2}$$

4.

$$\begin{gathered}|\,\text{cov}(\xi_1,\xi_2)| = \sqrt{\mathcal{D}\xi_1 \cdot \mathcal{D}\xi_2} \iff \mathcal{D} = 0 \iff \text{ equation } \mathcal{D}\eta(x) = 0 \text{ has solution } a \iff \\ \iff \mathcal{D}\eta(a) = 0 \iff \eta(a) = b = const \iff a\xi_1 - \xi_2 = b \iff \xi_2 = a\xi_1 - b.\end{gathered}$$

□

Remark. *Covariation of stochastic variables shows how much their dependency is close to linear.*

2.2 Correlation coefficient

Definition 5. *Correlation coefficient of random variables* ξ_1, ξ_2 *is a number:*

$$\rho(\xi_1, \xi_2) = \frac{\text{cov}(\xi_1, \xi_2)}{\sqrt{\mathcal{D}\xi_1 \cdot \mathcal{D}\xi_2}}$$

considering that $\mathcal{D}\xi_i > 0$.

Theorem 2.2.1. *Properties of covariation coefficient:*

1. $\rho(\xi, \xi) = 1$
2. ξ_1 *and* ξ_2 *are independent and* $\mathcal{D}\xi_i > 0 \Rightarrow \rho(\xi_1, \xi_2) = 0$
3. $|\rho(\xi_1, \xi_2)| = 1 \Rightarrow \xi_1$ *and* ξ_2 *have linear dependency:*

$$\xi_2 = a\xi_1 - b$$

for any constants a, b.

4. $\rho(a_1\xi_1 + b_1, a_2\xi_2 + b_2) = \pm\rho(\xi_1, \xi_2) =$

$$= \begin{cases} \rho(\xi_1, \xi_2) & a_1 \cdot a_2 > 0 \\ -\rho(\xi_1, \xi_2) & a_1 \cdot a_2 < 0 \end{cases}$$

Proof. DO IT YOURSELF □

Example. *Let* ξ_1, ξ_2 *air temperature of some two consistent days of the year. Consider that:*

$$M\xi_1 = m_1,\ M\xi_2 = m_2;\ \ \sigma_1^2 = \mathcal{D}\xi_1,\ \sigma_2^2 = \mathcal{D}\xi_2;\ \ \rho(\xi_1, \xi_2) = \rho..$$

Consider linear prediction:

$$\widetilde{\xi_2} = a\xi_1 + b.$$

where a, b *are some constants. Find* a, b *from the condition of minimization of standard deviation* $\widetilde{\xi_2}$ *and* ξ_2*, otherwords:*

$$M(\widetilde{\xi_2} - \xi_2)^2 \to min$$

Calcualte $M(\widetilde{\xi_2} - \xi_2)^2$*:*

$$M(\widetilde{\xi_2} - \xi_2) = \mathcal{D}(\widetilde{\xi_2} - \xi_2) + (M(\widetilde{\xi_2} - \xi_2))^2 = \mathcal{D}\widetilde{\xi_2}\mathcal{D}\xi_2 - 2\,\text{cov}(\widetilde{\xi_2}, \xi_2) + (M\widetilde{\xi_2} - M\xi_2)^2 =$$
$$= a^2 \cdot \mathcal{D}\xi_1 + \mathcal{D}\xi_2 - 2a \cdot \text{cov}(\xi_1, \xi_2) + (aM\xi_1 + b - M\xi_2)^2 =$$
$$(a^2 \cdot \sigma_1^2 + \sigma_2^2 - 2a\rho\sigma_1 \cdot \sigma_2) + (am_1 + b - m_2)^2$$

$$(am_1 + b - m_2)^2 \geq 0$$

Consider $am_1 + b - m_2 = 0;\ \ b = m_2 - am_1$ *Minimize first part* $a^2 \cdot \sigma_1^2 + \sigma_2^2 - 2a\rho\sigma_1 \cdot \sigma_2$*:*

$$2a \cdot \sigma_1^2 - 2\rho\sigma_1\sigma_2 = 0$$
$$a = \rho\frac{\sigma_2}{\sigma_2}$$
$$2\sigma_1^2 > 0$$

So the best predition:

$$\widetilde{\xi_2} = \rho\frac{\sigma_2}{\sigma_1}\cdot\xi_1 + (m_2 - \rho\frac{\sigma_2}{\sigma_1}m_1) = \rho\frac{\sigma_2}{\sigma_1}(\xi_1 - m_1) + m_2; .$$

$$M(\widetilde{\xi_2} - \xi_2)^2 = \sigma_2^2(1-\rho^2)..$$

If $|\rho| = 1$ *then* $M(\widetilde{\xi_2} - \xi_2)^2 = 0 \Rightarrow$ *min prediction*

$$\widetilde{\xi_2} = \rho\frac{\sigma_2}{\sigma_1}(\xi_1 - m_1) + m_2.$$

is precise.

If $|\rho| = 0 \Rightarrow M(\widetilde{\xi_2} - \xi_2)^2 = \sigma^2$ *and* $\widetilde{\xi_2} = m_2$ *does not depend on* ξ_1.

2.3 Equation of full probability for expectation

Example. *Firstly, dices are rolled, then a coin is flipped times the points on dice. How to find expectation of tails number?*

Let ξ *- number of points of dice rolled.*

η *- number of tails within* ξ *flips.*

$$\eta = \sum_{i=1}^{\xi} \mathbb{1}(\textit{tail within i-th flip}).$$

- number of tails.

Number of additions is random.

(will continue soon...)

Definition 6. *Probability distribution:*

$$P(\xi = x_i \mid H), i \geq 1.$$

is conditional distribution of discrete random value ξ *within* H*, where* H *is random event,* $P(H) > 0$.

Definition 7. *Conditional expectation of random value* ξ *within* H *is*

$$M(\xi \mid H) = \sum_{i=1}^{\infty} x_i \cdot P(\xi = x_i \mid H).$$

Theorem 2.3.1 (formula of full probability)**.** *Let* ξ *random value;* $\{H_1, \dots, H_n\}$ *- full group of events. Then:*

$$M\xi = \sum_{i=1}^{n} P(H_i)\cdot M(\xi \mid H_i).$$

Proof.

$$\sum_{i=1}^{n} P(H_i) \cdot M(\xi \mid H_i) = \sum_{i=1}^{n} P(H_i) \cdot \sum_{j=1}^{m} x_j \cdot P(\xi = x_j \mid H_i) =$$
$$\sum_{i=1}^{n} \sum_{j=1}^{m} x_j \cdot P(H_i) \cdot \frac{P(\xi = x_i, H_i)}{P(H_i)} = \sum_{j=1}^{m} x_j \cdot \sum_{i=1}^{n} P(\xi = x_i, H_i) =$$
$$\sum_{j=1}^{n} x_j \cdot P(\xi = x_j) = M\xi$$

□

Example (continuation)**.**

$$\eta = \sum_{i=1}^{\xi} \mathbb{1}(\textit{tails in i-th attempt})$$

$$M\eta = \sum_{k=1}^{6} P(\xi = k) M[\eta \mid \xi = k] = \sum_{k=1}^{6} \frac{1}{6} \cdot M \sum_{i=1}^{k} \mathbb{1}(\textit{tail on i-th attempt}) =$$
$$= \sum_{k=1}^{6} \frac{1}{6} \cdot \sum_{i=1}^{k} P(\textit{tails on i-th attempt}) = \frac{1}{6} \sum_{k=1}^{6} k \cdot \frac{1}{2} = \frac{1}{6} \cdot \frac{1}{2} \cdot \frac{1+6}{2} \cdot 6 = \frac{7}{4} = 1.75.$$

Example. *Let* $\xi \sim$ *Poiss*(λ)*;* $x_1, x_2, \ldots$ *- independent equally distributed,* $x_i \sim$ *Exp*(α)*.*

$$\eta = \sum_{i=1}^{\xi} P(\xi = k) \cdot M[\eta \mid \xi = k] = \sum_{k=0}^{\infty} P(\xi = k) \cdot M \sum_{i=1}^{k} x_i = \sum_{k=0}^{\infty} P(\xi = k) \cdot \sum_{i=1}^{k} M x_i =$$
$$M x_i = M x_2 = \cdots = M x_k$$
$$= \sum_{k=0}^{\infty} P(\xi_k) \cdot k \cdot M x_1 = M\xi \cdot M x_1$$

$$M\eta = \frac{\lambda}{\alpha}.$$

2.4 Inequalities related to moments of random values

2.4.1 Chebyshev inequality

Let ξ - integral (невід'ємна) random value. Then:

$$\forall \varepsilon > 0 : P(\xi \geq \varepsilon) \leq \frac{M\xi}{\varepsilon}.$$

Proof.

$$\xi - \xi \cdot \mathbb{1}(\xi \geq \varepsilon) + \xi \cdot \mathbb{1}(\xi < \varepsilon) \geq \xi \cdot \mathbb{1}(\xi \geq \varepsilon) \geq \varepsilon \cdot \mathbb{1}(\xi \geq \varepsilon) \Rightarrow$$
$$\Rightarrow M\xi \geq M(\varepsilon \cdot \mathbb{1}(\xi \geq \varepsilon)) = \varepsilon \cdot P(\xi \geq \varepsilon) \Rightarrow$$
$$\Rightarrow P(\xi \geq \varepsilon) \leq \frac{M\xi}{\varepsilon}$$

□

Corollary 2.4.1. 1. *If ξ - arbitrary random value, then*

$$P(|\xi| \geq \varepsilon) \leq \frac{M|\xi|}{\varepsilon}.$$

2. $P(|\xi| \geq \varepsilon) = P(|\xi|^k \geq \varepsilon^k) \leq \frac{M|\xi|^k}{\varepsilon^k}$

3. $P(|\xi - M\xi| \geq \varepsilon) \leq \frac{\mathcal{D}\xi}{\varepsilon^2}$

   Indeed:

$$P(|\xi - M\xi| \geq \varepsilon) = P(|\xi - M\xi|^2 \geq \varepsilon^2) \leq \frac{M(\xi - M\xi)srJ}{\varepsilon^2} = \frac{\mathcal{D}\xi}{\varepsilon^2}$$

Example. 1. *Rule of "three sigm"*

Let ξ random value with expectation $M\xi$ and $\mathcal{D}\xi = \sigma^2$;

$\sigma = \sqrt{\mathcal{D}\xi}$ *- standard deviation of random value.*

$$P(|\xi - M\xi| > 3\sigma) \leq \frac{\mathcal{D}\xi}{9\sigma^2} = \frac{\sigma^2}{0\sigma^2} = \frac{1}{9} \Rightarrow$$
$$P(|\xi - M\xi| < 3\sigma) \geq 1 - \frac{1}{9}.$$

If $\xi_1, \xi_2 \ldots, \xi_N$ independent equally distributed random values, then at least 90% of observations will be in interval:

$$(m - 3\sigma, m + 3\sigma).$$

where $m = M\xi_1$.

$\xi_1, \xi_2, \ldots, \xi_N \sim N(0,1)$ *independent. Then $\approx 90\%$ will be in interval $(-3,3)$.*

2. *Let p - unknown part of population of a country support some resolution. For definition p is used social poll.*

   n persons are polled:

$$\sum_{i=1}^{n} \mathbb{1}(\textit{i-th person support the resolution}).$$

$\frac{S_n}{n}$ *- part of those, who support the resolution.*

$\frac{S_n}{n} \approx p$ *- within large n.*

The question is, how large must be n for the deviation $\frac{S_n}{n}$ to be quite small. For instance:

$$P\left(\left|\frac{S_n}{n} - p\right| \geq 0.1\right) \leq 0.05.$$

Notice that:

$$M\left(\frac{S_n}{n}\right) = \frac{1}{n} \cdot MS_n = \frac{1}{n} \cdot np = p..$$

$$P\left(\left|\frac{S_n}{n} - M\left(\frac{S_n}{n}\right)\right| \leq 0.1\right) \leq \frac{\mathcal{D}\left(\frac{S_n}{n}\right)}{(0.1)^2} = \frac{\mathcal{D}S_n}{n^2 \cdot (0.1)^2} = \frac{n \cdot p \cdot (1-p)}{n^2 \cdot (0.1)^2} = \frac{p(1-p)}{n(0.1)^2} \leq$$
$$\leq \frac{1}{4n(0.1)^2} \quad \textit{as} \quad \forall p \in (0,1) : p(1-p) \leq \frac{1}{4}.$$

Find n from the condition:

$$\frac{1}{4n(0.1)^2} \leq 0.05 \Rightarrow n \geq \frac{1}{4 \cdot 0.05 \cdot (0.1)^2}.$$

Chapter 3

Inequalities. The law of large numbers in the form of Chebyshev. Borel-Cantelli lemma

3.1 Cauchy-Bunyakovsky inequality

Let ξ, η - stochastic variables such that $M\xi^2 < \infty, M\eta^2 < \infty$. Then $M|\xi\eta| < \infty$ and

$$M|\xi \cdot \eta| \leq \sqrt{M\xi^2} \cdot \sqrt{M\eta^2} \tag{3.1}$$

Proof.

$$\widetilde{\xi} \equiv \frac{\xi}{\sqrt{M\xi^2}}, \ \widetilde{\eta} \equiv \frac{\eta}{\sqrt{M\eta^2}}.$$

Whereas

$$(|\widetilde{\xi}| - |\widetilde{\eta}|)^2 \geq 0$$

, then

$$2|\widetilde{\xi}| \cdot |\widetilde{\eta}| \leq \widetilde{\xi}^2 + |\widetilde{\eta}|^2$$

Take expectation:

$$2M[|\widetilde{\xi}| \cdot |\widetilde{\eta}|] \leq M\widetilde{\xi}^2 + M|\widetilde{\eta}|^2 = 2$$

$$\Rightarrow M[|\widetilde{\xi}| \cdot |\widetilde{\eta}|] \leq 1$$

$$\Rightarrow M|\xi| \cdot |\eta| \leq \sqrt{M\xi^2} \cdot \sqrt{M\eta^2}$$

□

3.2 Jensen's inequality

Let $g(x)$ convex downward Borel function (опукла донизу борелівська функція) and $M|\xi| < \infty$. Then

$$g(M\xi) \leq Mg(\xi).$$

Proof. If g is convex downward, then

$$\forall x_0 \in \mathbb{R}\ \exists \lambda = \lambda(x_0) : g(x) \geq g(x_0) + (x - x_0) \cdot \lambda.$$

Consider $x = \xi, x_0 = M\xi$. Got

$$g(\xi) \geq g(M\xi) + (\xi - M\xi) \cdot \lambda.$$

Apply expectation:

$$Mg(\xi) \geq Mg(M\xi) + \lambda \cdot M(\xi - M\xi)$$
$$Mg(M\xi) = const; \quad M(\xi - M\xi) = 0$$
$$\Rightarrow Mg(\xi) \geq g(M(\xi))$$

□

3.3 Lyapunov inequality

If $0 < s < t$, then

$$(M|\xi|^s)^{1/s} \leq \left(M|\xi|^t\right)^{1/t}.$$

Proof. □

Corollary 3.3.1.

$$M|\xi| \leq \left(M|\xi|^2\right)^{1/2} \leq \left(M|\xi|^3\right)^{1/3} \leq \cdots \leq (M|\xi|^n)^{1/n}.$$

3.4 Helder inequality

Let $1 < p < \infty$ and $1 < q < \infty$, $\frac{1}{p} + \frac{1}{q} = 1$.

If $M|\xi|^p < \infty$, $M|\eta|^q < \infty$, then

$$M|\xi\eta| < \infty \quad \text{and} \quad M|\xi \cdot \eta| \leq (M|\xi|^p)^{1/p} \cdot (M|\eta|^q)^{1/q}.$$

(at $p = q = 2$ we obtain Cauchy-Bunyakovsky inequality)

Proof. Let $\tilde{\xi} = \frac{\xi}{(M|\xi|^p)^{1/p}}$; $\tilde{\eta} = \frac{\eta}{(M|\eta|^q)^{1/q}}$. Function $\ln x$ is convex upward. That's why we have $\forall a, b > 0$ and $a + b = 1$ within $x, y > 0$:

$$\ln(ax + by) \geq a \ln x \cdot b \ln y = \ln x^a y^b$$

$$\Rightarrow ax + by \geq x^a y^b$$

Let $x = |\tilde{\xi}|^p$, $y = |\tilde{\eta}|^q$, $a = \frac{1}{p}$, $b = \frac{1}{q}$. Then got:

$$|\tilde{\xi} \cdot \tilde{\eta}| \leqslant \tfrac{1}{p}\left|\tilde{\xi}\right|^p + \tfrac{1}{q}\,|\tilde{\eta}|^q$$
$$M|\tilde{\xi} \cdot \tilde{\eta}| \leqslant \tfrac{1}{p}\underbrace{M|\tilde{\xi}|^p}_{=1} + \tfrac{1}{q}\underbrace{M|\tilde{\eta}|^q}_{=1} = 1$$
$$M|\xi \cdot \eta| \leqslant (M|\xi|^p)^{1/p}\,(M|\eta|^q)^{1/q}$$

□

3.5 Minkovkiy inequality

If $M|\xi|^p < \infty$, $M|\eta|^p < \infty$, $1 \le p < \infty$, then

$$M|\xi + \eta|^p < \infty$$

and

$$(M|\xi + \eta|^p)^{1/p} \le (M|\xi|^p)^{1/p} + (M|\eta|^p)^{1/p}.$$

Proof. □

3.6 The law of large numbers in the form of Chebyshev

Let $X_1, X_2, \dots$ - sequence of random values with finite expectation $m_i = MX_i$.

Definition 8. *The sequence $\{X_n\}_{n\ge1}$ satisfies the law of large numbers, if*

$$\forall\varepsilon > 0 : P\left(\left|\frac{1}{n}\sum_{i=1}^{n} X_i - \frac{1}{n}\sum_{i=1}^{n} m_i\right| \ge \varepsilon\right) \underset{n\to\infty}{\longrightarrow} 0.$$

Theorem 3.6.1 (The Law of large numbers in the form of Chebyshev)**.** *Let $\{x_n\}_{n\ge1}$ sequence of independent random values that have finite expectations $Mx_i = m_i$ and $\mathcal{D}x_I = \sigma_i^2$, moreover the dispersions are evenly limited:*

$$\forall i \;\; \sigma_i^2 \le C < \infty.$$

then $\{x_n\}_{n\ge1}$ satisfies the law of large numbers.

X-

Corollary 3.6.2. *For independent evenly distributed random values:*
if $\{x_n\}_{n\ge1}$ sequence with finite $m = Mx_1, \sigma^2 = \mathcal{D}x_1$ then

$$\forall\varepsilon > 0 \;\; P\left(\left|\frac{1}{n}\sum_{i=1}^{n} x_i - m\right| \ge \varepsilon\right) \underset{n\to\infty}{\longrightarrow} 0.$$

Remark. *For use of the law of large numbers the finiteness of expectation is enough (will be proven later).*

Example. • *Bernoulli theorem:*

Theorem 3.6.3 (Bernoulli)**.** *Let S_n be a number of «successes» in n unrelated repeated trials with probability of «success» p in each one. Then*

$$\forall\varepsilon > 0 : P\left(\left|\frac{S_n}{n} - p\right| \ge \varepsilon\right) \underset{n\to\infty}{\longrightarrow} 0$$

Indeed — $S_n = \sum\limits_{i=1}^{n} X_i$, where $X_i \sim B(p)$; X_i are independent;

$MX_i = p$; $\mathcal{D}X_i = p(1-p) \le \frac{1}{4}$.

- *Poisson theorem*

  Theorem 3.6.4 (Poisson). *Let S_n be a number of successes in n trials. In k-th trials the probability of success in p_k. Then:*

  $$\forall \varepsilon > 0 : P\left(\left|\frac{S_n}{n} - \frac{1}{n}\sum_{i=1}^{n} p_i\right| \geq \varepsilon\right) \underset{n\to\infty}{\longrightarrow} 0$$

  Indeed: $S_n = \sum\limits_{i=1}^{n} X_i$ *,*
  $X_i \sim B(p_i);\ MX_i = p_i;\ \mathcal{D}X_i = p_i(1-p_i) \leq \frac{1}{4}$

3.7 Borel-Cantelli

The next lemma is the main tool for analysis of properties with probability of 1.

Consider $\{A_n\}_{n\geq}$ - a sequence of random events from σ-algebra $\mathcal{F}$.

Call to mind next notation:

$$\overline{\lim_{n\to\infty}} A_n = \{A_n \text{ occurs for infinitely many } n\} \equiv \{A_n \text{ i.o. (infinitely often)}\} =$$

$$= \bigcap_{n=1}^{\infty} \bigcup_{k\geq n} A_k$$

$$\forall n\ \exists k \geq n : A_k \text{ occurred}$$

$$\underline{\lim_{n\to\infty}} A_n = \{ \text{ from some number all the } A_n \text{ events occur } \} = \bigcup_{n=1}^{\infty} \bigcap_{k\geq n} A_k$$

$$(\exists n\ \forall k \geq n A_k \text{ occurs })$$

Lemma 3.7.1 (Borel-Cantelli). *Got several cases:*

a. If $\sum\limits_{n=1}^{\infty} P(A_n) < \infty$*, then* $P(A_n$ *i.o.* $) = 0$*.*

b. If $\sum\limits_{n=1}^{\infty} P(A_n) = \infty$ *and* $A_1, A_2, \ldots$ *are independent, then* $P(A_n$ *i.o.* $) = 1$*.*

Proof. a. By definition:

$$P\left(\overline{\lim_{n\to\infty}} A_n\right) = P\left(\bigcap_{n=1}^{\infty} \bigcup_{k\geq n} A_k\right)$$

Create the following subsequence:

$$B_1 = \bigcup_{k=1}^{\infty} A_k \supset B_2 = \bigcup_{k=2}^{\infty} A_k \supset \ldots$$

The $P\left(\bigcap_{n=1}^{\infty} \bigcup_{k\geq n} A_k\right)$ is continuous from above:

$$P\left(\bigcap_{n=1}^{\infty} \bigcup_{k\geq n} A_k\right) = \lim_{n\to\infty} P\left(\bigcup_{k\geq n} A_k\right) \leq \lim_{n\to\infty} \sum_{k\geq n} P(A_k) = 0$$

as long as the series $\sum_{k=1}^{\infty} P(A_k)$ is convergent.

$\square$