# Випадкові вектори. Розподіл, числові характеристики.

## 1 Теоретичні відомості

Нехай $(\Omega, \mathcal{F}, P)$ – імовірнісний простір. Далі для зручності розглядаються вектори-стовпчики.

Випадковим вектором $\vec{\xi} : \Omega \to \mathbb{R}^d$ називається $(\mathcal{F}, \mathcal{B}(\mathbb{R}^d))$-вимірним відображення.

Еквівалентно, випадковим вектором називається вектор $\vec{\xi} = (\xi_1, \ldots, \xi_d)^T$, координати якого є випадковими величинами.

Дійсно, якщо перше означення має місце, то для довільного $[a, b) \in \mathcal{B}(\mathbb{R})$:

$$\begin{aligned}\{\omega \in \Omega \mid \xi_1(\omega) \in [a, b)\} &= \{\omega \in \Omega \mid \xi_1(\omega) \in [a, b)\} \cap \Omega = \\ &= \{\omega \in \Omega \mid \xi_1(\omega) \in [a, b)\} \cap \{\omega \in \Omega \mid \xi_j(\omega) \in \mathbb{R},\ j = \overline{2, d}\} = \\ &= \{\omega \in \Omega \mid \vec{\xi}(\omega) \in [a, b) \times \mathbb{R}^{d-1}\} \in \mathcal{F}.\end{aligned}$$

Останнє включення має місце, оскільки $[a, b) \times \mathbb{R}^{d-1} \in \mathcal{B}(\mathbb{R}^d)$.

Навпаки, нехай $\vec{\xi} = (\xi_1, \ldots, \xi_d)^T$ є вектором з випадкових величин. Тоді для $\prod_{k=1}^{d}[a_k, b_k) \in \mathbb{R}^d$:

$$\left\{\omega \in \Omega \,\middle|\, \vec{\xi}(\omega) \in \prod_{k=1}^{d}[a_k, b_k)\right\} = \bigcap_{k=1}^{r}\{\omega \in \Omega \mid \xi_k(\omega) \in [a_k, b_k)\} \in \mathcal{F},$$

Вибір множин обгрунтовується тим, що $\mathcal{B}(\mathbb{R}^d) = \sigma a(\mathcal{P}_d)$. Останнє включення має місце внаслідок того що $(\xi_j)$ є випадковими величинами.

Функцією розподілу випадкового вектора $\vec{\xi}$ називають функцію кількох змінних

$$F_{\vec{\xi}}(\vec{x}) = P(\xi_1 < x_1, \ldots, \xi_d < x_d),\ \vec{x} = (x_1, \ldots, x_d)^T \in \mathbb{R}^d.$$

Часто також розглядають міру Лебега-Стілтьєса $\lambda_F$ на $\mathcal{B}(\mathbb{R}^d)$, породжену функцією розподілу випадкового вектора $F_{\vec{\xi}}(\cdot)$:

$$\lambda_F\left(\prod_{k=1}^{d}[a_k, b_k)\right) := F_{\vec{\xi}}(\vec{b}) - F_{\vec{\xi}}(\vec{a}),\ \vec{a} = (a_1, \ldots, a_d)^T,\ \vec{b} = (b_1, \ldots, b_d)^T,\ \vec{a}, \vec{b} \in \mathbb{R}^d.$$

Якщо існує така вимірна функція $f : \mathbb{R}^d \to \mathbb{R}$, що

$$\lambda_F(B) = \int_B f(\vec{x})\lambda_d(d\vec{x}),\ B \in \mathcal{B}(\mathbb{R}^d), \tag{1}$$

то випадковий вектор $\vec{\xi}$ називають абсолютно неперервним, а функцію $f$ називають щільністю розподілу цього вектора. (Проведіть аналогію з теорією абсолютно неперервних випадкових величин)

Математичним сподіванням випадкового вектора $\vec{\xi}$ називають вектор з математичних сподівань координат:

$$E[\vec{\xi}] = (E[\xi_1], \ldots, E[\xi_d])^T.$$

Коваріацією випадкових величин $\xi$ та $\eta$ називається число

$$\mathrm{cov}(\xi, \eta) = E[(\xi - E[\xi])(\eta - E[\eta])] = E[\xi\eta] - E[\xi]E[\eta].$$

Коваріаційною матрицею випадкового вектора $\vec{\xi}$ називають матрицю, що складається з коваріацій пар координат цього вектора:

$$\mathrm{Cov}(\vec{\xi}) = E[(\vec{\xi} - E[\vec{\xi}])(\vec{\xi} - E[\vec{\xi}])^T] = (\mathrm{cov}(\xi_i, \xi_j))_{i,j=1}^d$$

Далі нагадаємо деякі відомості про незалежні випадкові величини, які можна нескладно узагальнити для випадкових векторів.

Випадкові величини $\xi$ та $\eta$ називають незалежними, якщо для довільних $A, B \in \mathcal{B}(\mathbb{R})$:

$$P(\xi \in A, \eta \in B) = P(\xi \in A)P(\eta \in B) \tag{2}$$

Для незалежності двох випадкових величин досить вимагати виконання (2) на півінтервалах вигляду $A = (-\infty, x)$, $B = (-\infty, y)$, для всіх $x, y \in \mathbb{R}$, тобто

$$P(\xi < x, \eta < y) = P(\xi < x)P(\eta < y).$$

Набір випадкових величин $\{\xi_j\}_{j=1}^n$ називають:

1. Попарно незалежними, якщо довільна пара величин $\xi_i, \xi_j$ є незалежною, $i \neq j$,
2. Незалежними в сукупності, якщо для довільної підмножини $J \subset \{1, \ldots, n\}$:

$$P(\cap_{j\in J}\{\xi_j \in B_j\}) = \prod_{j\in J} P(\xi_j \in B_j),\ \{B_j\}_{j\in J} \subset \mathcal{B}(\mathbb{R}).$$

Якщо $\xi$ та $\eta$ є незалежними, тоді вони є некорельованими, тобто $\mathrm{cov}(\xi, \eta) = 0$. Навпаки твердження, взагалі кажучи, невірне.

Якщо $\{\xi_1, \ldots, \xi_n\}$ та $\{\eta_1, \ldots, \eta_m\}$ є незалежними наборами, тоді їх перетворення $f(\xi_1, \ldots, \xi_n)$ та $g(\eta_1, \ldots, \eta_m)$ є незалежними.

Так, а тепер питання: чи можна ввести аналог LOTUS на перетворення від випадкових векторів? Відповідь – можна.

**Law of the unconscious statistician (LOTUS).** Нехай вимірна функція $g : \mathbb{R}^d \to \mathbb{R}$ є інтегровною за мірою $\lambda_F$ (див. (1)). Тоді $E[g(\vec{\xi})]$ існує, причому

$$E[g(\vec{\xi})] = \int_{-\infty}^{+\infty} g(\vec{x})\lambda_F(d\vec{x}).$$

# 2 Задачі

## 2.1 Задача 1

Математик проводить $n$ незалежних випробувань. Кожне з випробувань може мати один з $r$ результатів. Імовірності настання окремого результату у випробуваннях відомі: якщо $X_k$ є номером результату в $k$-му випробуванні, то

$$P(X_k = i) = p_i,\ k = \overline{1,n},\ i = \overline{1,r}$$

Тут $p_j > 0$ для всіх $j = \overline{1,r}$, та $\sum_j p_j = 1$.

Введемо вектор абсолютних частот: $\vec{\nu} = (\nu_1, \ldots, \nu_r)$, де $\nu_i$ – кількість випробувань, в яких отримали $i$-ий результат:

$$\nu_i = \sum_{k=1}^{n} 1\{X_k = i\}.$$

Задача полягає в наступному:

1. Знайти розподіл $\vec{\nu}$.
2. Знайти маргінальні розподіли $\vec{\nu}$. Чи є $\{\nu_j\}_{j=1}^{n}$ незалежними?
3. Знайти коваріаційну матрицю $\vec{\nu}$.

### Розв'язання

Нехай серед $n$ випробувань маємо $k_1$ з 1-им результатом, $k_2$ з 2-им результатом, ..., і $k_n$ випробувань з $r$-им результатом. Всі результати вичерпані, тому $\sum_{j=1}^{r} k_j = n$. Кількість способів отримати вищеописаний результат становить $C_n^{k_1} \cdot C_{n-k_2}^{k_2} \cdot \ldots \cdot C_{n-k_1-\ldots-k_{r-1}}^{k_r} = \frac{n!}{k_1!k_2!\ldots k_r!}$. Імовірність отримати конкретну реалізацію із заданою кількістю появ результатів є $p_1^{k_1} \cdot p_2^{k_2} \cdot \ldots \cdot p_r^{k_r}$ (пригадайте розподіл результатів в одному випробуванні та сукупну незалежність випробувань в схемі). Отже, для $\vec{k} = (k_1, \ldots, k_r) \in \mathbb{Z}_+^r$ таких, що $\sum_j k_j = n$ маємо

$$P(\vec{\nu} = \vec{k}) = P(\cap_{j=1}^{r}\{\nu_j = k_j\}) = \frac{n!}{k_1! \cdot \ldots \cdot k_r!} \cdot p_1^{k_1} \cdot \ldots \cdot p_r^{k_r}$$

Для отримання розподілу координати можна миттєво отримати без сумісного розподілу $\vec{\nu}$: щоб отримати $P(\nu_j = k)$, обираємо $k$ з $n$ випробувань, в яких отримали $j$-ий результат, далі множимо на імовірність отримати конкретну реалізацію, де є $k$ $j$-их результатів та $n-k$ 'невдач': відповідно $p_j(1-p_j)$. Отже $P(\nu_j = k) = C_n^k p_j^k (1-p_j)^{n-k}$ для $k = \overline{0,n}$.

Але ніхто не забороняє взяти 'проінтегрувати' за тими координатами, що не цікавлять:

$$P(\nu_1 = k) = P(\nu_1 = k, \nu_1 + \ldots + \nu_r = n) = \sum_{k_j \geq 0, j=\overline{2,n}:\, k+k_2+\ldots+k_r=n} P(\nu_1 = k, \nu_2 = k_2, \ldots, \nu_r = k_r) =$$

$$= \sum_{k_2,\ldots,k_r \geq 0:\, k_2+\ldots+k_r=n-k} \frac{n!}{k!k_2!\ldots k_r!} p_1^k p_2^{k_2} \ldots p_r^{k_r} = \frac{n!}{(n-k!)k!} p_1^k \sum_{k_2,\ldots,k_r \geq 0:\, k_2+\ldots+k_r=n-k} \frac{(n-k)!}{k_2!\ldots k_r!} p_2^{k_2} \ldots p_r^{k_r} =$$

$$= \left| \text{Поліноміальна теорема} \right| = C_n^k p_1^k (p_2 + \ldots + p_r)^{n-k} = C_n^k p_1^k (1 - p_1)^{n-k},\ k = \overline{0, n}.$$

Отже одновимірні маргінальні розподіли є біноміальними, тобто $\nu_j \sim Bin(n, p_j)$. Для біноміального розподілу відомо (якщо ні, то доведіть) що $E[\nu_j] = np_j$ та $Var[\nu_j] = np_j(1 - p_j)$.

Тепер поговоримо про незалежність координат. Якби координати були б попарно незалежними, тоді для всіх $k_i + k_j = n$, $k_j \geq 0$:

$$P(\nu_i = k_i, \nu_j = k_j) = P(\nu_i = k_i)P(\nu_j = k_j) = C_n^{k_i} C_n^{k_j} p_i^{k_i} p_j^{k_j} (1 - p_i)^{n-k_i} (1 - p_j)^{n-k_j}$$

Але двовимірний маргінальним розподіл (простими словами – сумісний розподіл двох координат) має вигляд

$$P(\nu_i = k_i, \nu_j = k_j) = \frac{n!}{k_i! k_j! (n - (k_i + k_j))!} p_i^{k_i} p_j^{k_j} (1 - (p_i + p_j))^{n-(k_i+k_j)}$$

Щоб отримати результат вище, скористайтеся довільними з міркувань вище для розподілів.

Тут неважко вже побачити, що сумісний розподіл не розпадається на добутки розподілів координат. Наприклад

$$P(\nu_i = 0, \nu_j = 0) = (1 - (p_i + p_j))^n \neq (1 - p_i)^n (1 - p_j)^n = P(\nu_i = 0)P(\nu_j = 0).$$

Дісйно, чому? Побачимо, що ліву та праву частини можна подати у вигляді $(1 - x - y)^n$ та $(1 - x - y + xy)^n$, $0 < x < 1$, $0 < y < 1 - x$. Вирази в дужках невід'ємні, отже можна позбутися степенів та розглядати

$$1 - x - y = 1 - x - y + xy \Leftrightarrow xy = 0 \Rightarrow \emptyset \text{ (згідно умов на } x,\ y)$$

Отже координати не є попарно незалежними. Значить, вони й не можуть бути незалежними в сукупності (яка вимагає незалежність довільної підмножини з координат). Тобто абсолютні частоти у векторі залежні.

Тепер знайдемо $\mathrm{Cov}(\vec{\nu}) = (\mathrm{cov}(\nu_i, \nu_j))_{i,j=1}^r$. Досить знайти елементи поза діагоналлю, оскільки по діагоналі стоять дисперсії координат (якщо не впевнені, то дивимося уважно на коваріацію випадкової величини на себе). Отже, для $i \neq j$ знаходимо $E[\nu_i \nu_j]$. Щоб знайти математичне сподівання добутку координат, можна:

1. Скористатися LOTUS для $\vec{\nu}$ з $g(\vec{\nu}) = \nu_i \cdot \nu_j$.
2. Скористатися LOTUS для $(\nu_i, \nu_j)$ з $g(\nu_i, \nu_j) = \nu_i \cdot \nu_j$.
3. Перейти до інших випадкових величин (до представлень), щоб легше рахувати інтеграл.

Далі скористаємося третім способом. Перші два теж допустимі, але з ними довше піде підрахунок (втім для звірки результату закликаю спробувати порахувати 'в лоб').

Отже, знаючи що $\nu_j = \sum_{k=1}^n 1\{X_k = j\}$:

$$E[\nu_i \nu_j] = E\left[\sum_{k_1=1}^n 1\{X_{k_1} = i\} \sum_{k_2=1}^n 1\{X_{k_2} = j\}\right] = \sum_{k_1=1}^n \sum_{k_2=1}^n E[1_{X_{k_1}=i, X_{k_2}=j}] =$$
$$= \sum_{k_1=1}^n \sum_{k_2=1}^n P(X_{k_1} = i, X_{k_2} = j) = \dots$$

Якщо $k_1 = k_2$, то імовірність береться від перетину несумісних подій (очевидно?), отже $P(X_{k_1} = i, X_{k_2} = j) = 0$.

Інакше, при $k_1 \neq k_2$ маємо з незалежності випадкових величин

$$P(X_{k_1} = i, X_{k_2} = j) = P(X_{k_1} = i) P(X_{k_2} = j) = p_i p_j.$$

Врахувавши попередні міркування, маємо

$$\dots = \sum_{k_1=1}^n \sum_{k_2=\overline{1,n}, k_2 \neq k_1} p_i p_j = n(n-1) p_i p_j.$$

Отримавши $E[\nu_i \nu_j]$, маємо коваріацію

$$\operatorname{cov}(\nu_i, \nu_j) = E[\nu_i \nu_j] - E[\nu_i] E[\nu_j] = n(n-1) p_i p_j - n p_i n p_j = -n p_i p_j,\ i \neq j.$$

## 2.2 Задача 2

Розглянемо абсолютно неперервни випадковий вектор $(\xi, \eta)$ що має щільність $f(x, y)$:

$$f(x,y) = 1_{(0,1)}(x)1_{(0,1)}(y)\frac{6}{5}x(x+2y).$$

Задача полягає в наступному:

1. Обчислити $P(\xi \in (1/4, 3/4), \eta < \xi)$.
2. Знайти маргінальні розподіли випадкового вектора.
3. Чи є координати вектора незалежними?
4. Обчислити $\mathrm{cov}(\xi, \eta)$.

### Розв'язання

В основному доведеться оперувати з сумісною щільністю розподілу $f(x, y)$.

Спочатку обчислимо $P(\xi \in (1/4, 3/4), \eta < \xi)$. Для цього

$$P(\xi \in (1/4,3/4), \eta < \xi) = \int\limits_{1/4}^{3/4}\int\limits_{-\infty}^{x} f(x,y)dydx = \int\limits_{1/4}^{3/4}\int\limits_{0}^{x} \frac{6}{5}x(x+2y)dydx =$$

$$= \int\limits_{1/4}^{3/4}\int\limits_{0}^{x} \frac{6}{5}x(x+2y)dydx = \int\limits_{1/4}^{3/4} \frac{6}{5}x(x^2+x^2)dx =$$

$$= \int\limits_{1/4}^{3/4} \frac{6}{5}2x^3dx = \frac{12}{5}\cdot\frac{x^4}{4}\bigg|_{1/4}^{3/4} = \frac{3}{5}\cdot\left(\frac{3^4-1^4}{4^4}\right).$$

Знайдемо маргінальні розподіли. Побачимо, що

$$P(\xi < u) = P(\xi < u, -\infty < \eta < +\infty) = \int\limits_{-\infty}^{u}\int\limits_{-\infty}^{+\infty} f(x,y)dydx \Rightarrow f_\xi(u) = \frac{d}{du}P(\xi < u) = \int\limits_{-\infty}^{+\infty} f(u,y)dy \text{ м.с.}$$

тобто принаймні треба вміти правильно інтегрувати. Отже,

$$f_\xi(u) = \int\limits_{-\infty}^{+\infty} f(u,y)dy = \int\limits_{-\infty}^{+\infty} 1_{(0,1)}(u)1_{(0,1)}(y)\frac{6}{5}u(u+2y)dy = 1_{(0,1)}(u)\frac{6}{5}u\int\limits_{0}^{1}(u+2y)dy = 1_{(0,1)}(u)\frac{6}{5}(u^2+u)$$

Аналогічно для $\eta$ отримаємо відповідну щільність розподілу

$$f_\eta(v) = \int\limits_{-\infty}^{+\infty} f(x,v)dx = 1_{(0,1)}(v)\frac{6}{5}(1/3+v)$$

Для перевірки незалежності координат достатньо переконатися, що сумісна щільність розподілу є добутком (розпадається) на добуток маргінальних щільностей. Неважко побачити, що

$$f_\xi(x)f_\eta(y) = 1_{(0,1)}(x)\frac{6}{5}(x^2+x)1_{(0,1)}(y)\frac{6}{5}(1/3+y) \neq 1_{(0,1)}(x)1_{(0,1)}(y)\frac{6}{5}x(x+2y) = f(x,y).$$

Тобто координати випадкового вектора залежні.

Залишається підрахувати коваріацію координат. Для цього досить знати $E[\xi\eta]$ та $E[\xi]$, $E[\eta]$, оскільки (переконайтеся)

$$\operatorname{cov}(\xi,\eta) = E[\xi\eta] - E[\xi]E[\eta].$$

Згідно LOTUS для $(\xi,\eta)$ та $g(x,y) = x\cdot y$:

$$E[\xi\eta] = \int\limits_{-\infty}^{+\infty}\int\limits_{-\infty}^{+\infty} xyf_{(\xi,\eta)}(x,y)dydx = \frac{6}{5}\int\limits_0^1\int\limits_0^1 xy\cdot x(x+2y)dydx = \frac{6}{5}\int\limits_0^1 (x^3/2+2x^2/3)dx = \frac{6}{5}\cdot\frac{25}{72} = \frac{5}{12}.$$

Можна по суті з сумісного розподілу вичепити теоретичні моменти координат, але ж ми вже маємо їх розподіли. Тому тут можна через одновимірні розподіли піти та отримати

$$E[\xi] = \frac{6}{5}\cdot\int\limits_0^1 u(u^2+u)du = \frac{6}{5}\cdot\left(\frac{1}{4}+\frac{1}{3}\right) = \frac{6}{5}\cdot\frac{7}{12} = \frac{7}{10}$$

$$E[\eta] = \frac{6}{5}\cdot\int\limits_0^1 v(1/3+v)dv = \frac{6}{5}\cdot\left(\frac{1}{3}\cdot\frac{1}{2}+\frac{1}{3}\right) = \frac{6}{5}\cdot\frac{1}{2} = \frac{3}{5}$$

Тепер маємо все для підрахунку коваріації:

$$\operatorname{cov}(\xi,\eta) = \frac{5}{12} - \frac{7}{10}\cdot\frac{3}{5} = -\frac{1}{300}$$

## 2.3 Задача 3

Розглянемо два незалежні випадкові величини $u_1$ та $u_2$ з рівномірним розподілом на $[0, 1]$. Розглянемо $S = u_1 + u_2$. Знайти розподіл $S$, обчислити $E[S]$, $Var[S]$.

### Розв'язання

Миттєво обчислимо числові характеристики розподілу суми. Математичне сподівання отримаємо з лінійності

$$E[S] = E[u_1 + u_2] = E[u_1] + E[u_2] = 1/2 + 1/2 = 1.$$

Для дисперсії скористаємося тим, що незалежні випадкові величини є некорельованими:

$$\begin{aligned} Var[S] &= \mathrm{cov}(S, S) = \mathrm{cov}(u_1 + u_2, u_1 + u_2) = \mathrm{cov}(u_1, u_1) + 2\mathrm{cov}(u_1, u_2) + \mathrm{cov}(u_2, u_2) = \\ &= Var[u_1] + Var[u_2] = 1/12 + 1/12 = 1/6. \end{aligned}$$

Тепер знайдемо розподіл суми через розподіл випадкового вектора $(u_1, u_2)$. Через $\lambda_F$ введемо міру Лебега-Стілтьєса, породжену сумісним розподілом $(u_1, u_2)$. Можна переконатися, що $d\lambda_F(x_1, x_2) = 1_{(0,1)^2}(x_1, x_2)dx_1dx_2$. Для $s \in \mathbb{R}$:

$$F_S(s) = P(S < s) = P(u_1 + u_2 < s) = P(u_2 < s - u_1) = \lambda_F(\Delta_s \cap \square) = \dots$$

Розберемося з областю інтегрування в останньому інтегралі. По суті це перетин 'трикутника' $\Delta_s = \{(x, y) \in \mathbb{R}^2 \mid y \leq s - x\}$ з квадратом $\square = (0, 1) \times (0, 1)$. Тоді

$$\Delta_s \cap \square = \begin{cases} \emptyset, & s \leq 0, \\ \{(x, y) \in \mathbb{R}^2 \mid 0 < x \leq s,\ 0 < y \leq s - x\}, & 0 < s \leq 1, \\ \square \setminus \{(x, y) \in \mathbb{R}^2 \mid s - 1 < x < 1, s - x < y < 1\}, & 1 < s \leq 2, \\ \square, & 2 < s. \end{cases}$$

Отже,

$$\dots = \begin{cases} 0, & s \leq 0, \\ \int\limits_0^s \int\limits_0^{s-x_1} dx_2dx_1 = \int\limits_0^s (s - x_1)dx_1 = s^2 - \dfrac{s^2}{2} = \dfrac{s^2}{2}, & 0 < s \leq 1, \\ 1 - \int\limits_{s-1}^1 \int\limits_{s-x_1}^1 dx_2dx_1 = 1 + \int\limits_{s-1}^1 ((s - 1) - x_1)dx_1 = 1 - \dfrac{(2 - s)^2}{2}, & 1 < s \leq 2, \\ 1, & 2 < s. \end{cases}$$

Тобто щільність розподілу суми $S$ має вигляд

$$f_S(s) = 1_{(0,1]}(s) \cdot s + 1_{(1,2)}(s) \cdot (2 - s) \quad \text{м.н. } \lambda_1$$

Тобто ми отримали трикутниковий розподіл.