

§8. Векторные случайные величины. Независимость случайных величин

Лекция 13

При определении действительной случайной величины мы интерпретировали ее как некоторую числовую характеристику исследуемого объекта. Однако на практике мы чаще сталкиваемся с одновременным наблюдением нескольких числовых характеристик – случайным вектором, распределение которого так же, как и в одномерном случае, порождается распределением на измеримом пространстве (Ω, \mathcal{A}) элементарных исходов статистического эксперимента. Чтобы провести аналогию с определением скалярной случайной величины, мы должны вспомнить строение борелевских множеств в \mathbb{R}^n . Роль интервалов здесь играют *прямоугольники* – подмножества \mathbb{R}^n вида $B = B_1 \times \dots \times B_n$, где каждое B_k есть открытый (a_k, b_k) , полуоткрытый $(a_k, b_k]$ и $[a_k, b_k)$ или замкнутый $[a_k, b_k]$ интервал на действительной прямой \mathbb{R} . Конечные объединения непересекающихся прямоугольников образуют булеву алгебру подмножеств \mathbb{R}^n , а наименьшая σ -алгебра \mathcal{B}^n , содержащая эту булеву алгебру, образует класс измеримых подмножеств \mathbb{R}^n или *событий*. Таким образом мы получаем измеримое пространство $(\mathbb{R}^n, \mathcal{B}^n)$.

Определение 8.1. *Векторной случайной величиной* или *случайным вектором* называется измеримое отображение $X^{(n)} = X^{(n)}(\omega) = (X_1(\omega), \dots, X_n(\omega))$ пространства элементарных исходов Ω , наделенного σ -алгеброй измеримых подмножеств \mathcal{A} , в n -мерное эвклидово пространство \mathbb{R}^n с борелевской σ -алгеброй \mathcal{B}^n . Для любого $B \in \mathcal{B}^n$ справедливо включение $X^{(n)-1}(B) = \{\omega : X^{(n)}(\omega) \in B\} \in \mathcal{A}$.

Теперь, по аналогии с одномерным случаем, зададим вероятность P_n на $(\mathbb{R}, \mathcal{B}^n)$, порожденную вероятностью P на (Ω, \mathcal{A}) , соотношением

$$P_n(B) = P(X^{(n)-1}(B)), \quad \forall B \in \mathcal{B}^n.$$

Как будет видно в дальнейшем, исходное вероятностное пространство (Ω, \mathcal{A}, P) играет более важную роль в характеристике распределения $X^{(n)}$, если $n > 1$. Мы будем изучать вероятностные модели, которые можно записать в виде интеграла Лебега

$$P(X^{(n)} \in B) = \int_B f(x_1, \dots, x_n) d\mu_1(x_1) \cdots d\mu_n(x_n)$$

от неотрицательной функции $f(x_1, \dots, x_n)$ по мере $d\mu = d\mu_1 \cdots d\mu_n$, где каждая σ -конечная мера μ_i , $i = 1, \dots, n$ на борелевской прямой $(\mathbb{R}, \mathcal{B})$ является или считающей мерой, или мерой Лебега. В таком случае вычисление вероятности событий $B \in \mathcal{B}^n$ сводится или к суммированию вероятностей отдельных точек в \mathbb{R}^n , или к вычислению кратных интегралов Римана. Функция f в данном случае выступает в роли n -мерной *функции плотности*. Естественно, можно ввести также понятие n -мерной *функции распределения*

$$F(x_1, \dots, x_n) = P(X_1 < x_1, \dots, X_n < x_n) = \int_{-\infty}^{x_1} \cdots \int_{-\infty}^{x_n} f(x_1, \dots, x_n) d\mu_1(x_1) \cdots d\mu_n(x_n),$$

однако при $n > 1$ с помощью этой функции можно выразить только вероятности „прямоугольников“ в \mathbb{R}^n , в то время как вероятность попадания случайного вектора в подмножества более сложной конфигурации (например, эллипсоиды) приходится вычислять с помощью интеграла от функции плотности. Как и в одномерном случае, n -мерная *функция распределения однозначно определяет распределение вероятностей на $(\mathbb{R}^n, \mathcal{B}^n)$* , то есть имеет место n -мерный аналог теоремы 4.1.

Из определения функции распределения вытекает, что в случае непрерывного распределения ($\mu = \mu_1 \times \cdots \times \mu_n$ – мера Лебега) функция плотности f выражается через функцию распределения посредством дифференцирования

$$f(x_1, \dots, x_n) = \frac{\partial^n F(x_1, \dots, x_n)}{\partial x_1 \cdots \partial x_n},$$

а в дискретном случае (μ – считающая мера, приписывающая единицу каждой точке \mathbb{R}^n с целочисленными координатами)

$$f(x_1, \dots, x_n) = P(X^{(n)} = x^{(n)}) = P(X_1 = x_1, \dots, X_n = x_n).$$

Я полагаю, вы сами сможете записать аналогичные связи между F и f в „смешанном“ дискретно-непрерывном случае, когда часть компонент случайного вектора имеет непрерывное распределение, а другая – дискретное.

Как вычислить совместное распределение отдельных компонент X_{i_1}, \dots, X_{i_k} случайного вектора $X^{(n)}$? Для этого достаточно в функции распределения $X^{(n)}$ устремить к $+\infty$ все переменные, отличные

от x_{i_1}, \dots, x_{i_k} , или, что то же, проинтегрировать функцию плотности по каждой из переменных, отличных от x_{i_1}, \dots, x_{i_k} , в пределах $\pm\infty$.

Заметим, что в теории вероятностей принято называть распределения каждой компоненты (случайной величины) $X_i, i = 1, \dots, n$, — *маргинальными* или *частными* распределениями.

Пример 8.1 (равномерное распределение на круге.) В часть плоскости \mathbb{R}^2 , ограниченную окружностью $x^2 + y^2 = r^2$, наугад бросается точка, так что ее координаты (x, y) представляют реализацию случайного вектора (X, Y) . Как и в случае с бросанием точки на отрезок прямой, термин „наугад” понимается в смысле зависимости вероятности попадания точки в некоторую, измеримую по Лебегу часть B круга только от площади B . Те же рассуждения, что и при выводе равномерного распределения на отрезке, приводят нас к равномерному распределению (X, Y) с функцией плотности (по мере Лебега $d\mu = dx dy$) $f(x, y)$, равной постоянной $1/\pi r^2$, если $x^2 + y^2 \leq 1$, и равной нулю вне этого круга.

Найдем функцию плотности $f^X(x)$ маргинального распределения X . Для этого мы должны проинтегрировать функцию $f(x, y)$ по переменной y в пределах $\pm\infty$ при каждом фиксированном значении $x \in \mathbb{R}$. Если x фиксировано, то $f(x, y)$ отлична от нуля и равна $1/\pi r^2$ только при значениях y , удовлетворяющих неравенству $-\sqrt{r^2 - x^2} \leq y \leq \sqrt{r^2 - x^2}$. Следовательно,

$$f^X(x) = \frac{1}{\pi r^2} \int_{-\sqrt{r^2 - x^2}}^{\sqrt{r^2 - x^2}} dy = \frac{2}{\pi r^2} \sqrt{r^2 - x^2},$$

если $|x| \leq r$, и $f^X(x) = 0$ в противном случае. Легко видеть, что маргинальное распределение второй компоненты Y случайного вектора имеет тот же вид. Таким образом, маргинальные распределения компонент отличны от равномерного и имеют четко выраженную моду, совпадающую с началом координат.

Маргинальные плотности компонент случайного вектора наиболее просто находятся в том случае, когда функция плотности $X^{(n)}$ распадается в произведение функций плотности отдельных компонент. Понятно, что хотя бы в дискретном случае это говорит о некоторой „независимости” компонент случайного вектора. Чтобы ввести строгое определение такой независимости, мы должны обратиться

ся к σ -подалгебрам алгебры \mathcal{A} , порожденным каждой компонентой X_i , $i = 1, \dots, n$, вектора $X^{(n)}$.

Пусть $X = X(\omega)$ – случайная величина на (Ω, \mathcal{A}) со значениями в измеримом пространстве $(\mathbb{R}, \mathcal{B})$. Рассмотрим класс $\mathcal{A}_X = \{X^{-1}(B), B \in \mathcal{B}\}$ всех прообразов элементов борелевского поля \mathcal{B} , полагая $X^{-1}(\mathbb{R}) = \Omega$. Имеет место

Предложение 8.1. *Класс \mathcal{A}_X подмножеств Ω является σ -алгеброй (подалгеброй \mathcal{A} .)*

Доказательство. Достаточно проверить аксиомы булевой σ -алгебры (см. определение 2.5).

(A1). По определению \mathcal{A}_X пространство элементарных исходов $\Omega = X^{-1}(\mathbb{R}) \in \mathcal{A}_X$.

(A2). Покажем, что дополнение $(X^{-1}(B))^c \in \mathcal{A}_X$, каково бы ни было $B \in \mathcal{B}$. Действительно, событие, противоположное $X^{-1}(B) = \{\omega : X(\omega) \in B\}$, означает, что $X(\omega)$ не принадлежит B , то есть $X(\omega) \in B^c$. Так как $B^c \in \mathcal{B}$, то $X^{-1}(B^c) = (X^{-1}(B))^c \in \mathcal{A}_X$.

(A3)_S. Рассуждения, аналогичные предыдущему пункту, показывают, что

$$\bigcup_1^\infty X^{-1}(B_i) = X^{-1}\left(\bigcup_1^\infty B_i\right) \in \mathcal{A}_X.$$

Легко понять, что данное утверждение справедливо не только для скалярных случайных величин, но и случайных векторов. Теперь мы в состоянии ввести одно из фундаментальнейших понятий теории вероятностей и математической статистики.

Определение 8.2. Случайные величины (случайные векторы) X_1, \dots, X_n , заданные на одном и том же измеримом пространстве (Ω, \mathcal{A}) , называются *независимыми в совокупности* или *совместно независимыми*, если независимы σ -подалгебры $\mathcal{A}_{X_1}, \dots, \mathcal{A}_{X_n}$ σ -алгебры \mathcal{A} , порожденные соответствующими случайными величинами.

Таким образом, в соответствии с определением 3.4 независимости σ -алгебр, для любых элементов (событий) B_1, \dots, B_n борелевского поля \mathcal{B} справедливо равенство

$$P(X_1 \in B_1, \dots, X_n \in B_n) = \prod_1^n P(X_i \in B_i), \quad (1)$$

то есть совместное распределение независимых случайных величин распадается в произведение их маргинальных распределений. Оказывается, для независимости случайных величин достаточно потребовать выполнения более слабого условия, состоящего в возможности представления совместной функции распределения X_1, \dots, X_n в виде произведения маргинальных функций распределения.

Предложение 8.2. (*критерий независимости случайных величин*). Случайные величины X_1, \dots, X_n независимы в совокупности тогда и только тогда, когда их совместная функция распределения (функция плотности) распадается в произведение маргинальных функций распределения (маргинальных функций плотности):

$$F(x_1, \dots, x_n) = \prod_1^n F^{X_i}(x_i), \quad f(x_1, \dots, x_n) = \prod_1^n f^{X_i}(x_i).$$

Доказательство. Необходимость условия очевидна, поскольку свойство мультипликативности совместной функции распределения есть частный случай равенства (1). Таким образом достаточно показать, что для двух случайных величин X и Y равенство $F(x, y) = F^X(x)F^Y(y)$ при любых $x, y \in \mathbb{R}$ влечет $P(X \in B_1, Y \in B_2) = P(X \in B_1)P(Y \in B_2)$ для любых $B_1, B_2 \in \mathcal{B}$; общий случай, касающийся независимости $n > 2$ случайных величин, рассматривается с привлечением метода математической индукции.

Перепишем условие независимости

$$\begin{aligned} P(X \in (-\infty, x))P(Y \in (-\infty, y)) = \\ P(\{X \in (-\infty, x)\} \cap \{Y \in (-\infty, y)\}) \end{aligned}$$

в виде

$$\begin{aligned} F^X(x) = P(X \in (-\infty, x)) = \\ \frac{P(\{X \in (-\infty, x)\} \cap \{Y \in (-\infty, y)\})}{F^Y(y)}. \end{aligned} \quad (2)$$

Поскольку функция распределения $F^X(x)$ однозначно определяет маргинальное распределение $P(X \in B_1)$ случайной величины X , (теорема 4.1), то равенство (2) влечет

$$P(X \in B_1) = \frac{P\{X \in B_1\} \cap \{Y \in (-\infty, y)\}}{F^Y(y)}$$

или, что то же,

$$F^Y(y) = P(Y \in (-\infty, y)) = \frac{P\{X \in B_1\} \cap \{Y \in (-\infty, y)\}}{P(X \in B_1)},$$

для любых $B_1 \in \mathcal{B}$. Используя снова теорему 4.1 об однозначном определении распределения вероятностей случайной величины Y посредством ее функции распределения $F^Y(y)$, получаем требуемое определение независимости:

$$P(X \in B_1, Y \in B_2) = P(X \in B_1)P(Y \in B_2),$$

каковы бы ни были $B_1, B_2 \in \mathcal{B}$.

Утверждение теоремы, касающееся функции плотности, следует немедленно из соотношения между функцией распределения и функцией плотности:

$$F(x, y) = \int_{-\infty}^x \int_{-\infty}^y f(u, v) d\mu_1(u) d\mu_2(v).$$

Следующее утверждение, относящееся к функциям от независимых случайных величин, позволяет вычислять моментные характеристики некоторых распределений значительно проще, чем это делалось в §6.

Предложение 8.3. *Если X_1, \dots, X_n независимы в совокупности, то*

1⁰. *независимы в совокупности случайные величины $Y_1 = g_1(X_1), \dots, Y_n = g_n(X_n)$, где $g_i, i = 1, \dots, n$ – измеримые функции;*

$$2^0. \mathbf{E} \prod_1^n X_i = \prod_1^n \mathbf{E} X_i;$$

$$3^0. \mathbf{D} \sum_1^n X_i = \sum_1^n \mathbf{D} X_i.$$

Доказательство. 1⁰. Поскольку σ -алгебры, порожденные случайными величинами Y_1, \dots, Y_n , являются подалгебрами соответствующих σ -алгебр, порожденных X_1, \dots, X_n , а последние независимы (см. определение 3.4), то данное утверждение следует непосредственно из определения 8.2 независимости случайных величин.

2⁰. Пусть $f(x_i)$ – функция плотности X_i по мере $\mu_i, i = 1, \dots, n$. Тогда, в силу предложения 8.2, совместная функция плотности

$$f(x_1, \dots, x_n) = \prod_1^n f_i(x_i),$$

так что

$$\mathbf{E} \prod_1^n X_i = \int_{\mathbb{R}} x_1 f(x_1) d\mu(x_1) \cdots \int_{\mathbb{R}} x_n f(x_n) d\mu(x_n) = \prod_1^n \mathbf{E} X_i.$$

3⁰. Используя только что доказанное утверждение (2) и свойство линейности математического ожидания, получаем

$$\begin{aligned} \mathbf{D} \sum_1^n X_i &= \mathbf{E} \left(\sum_1^n (X_i - \mathbf{E} X_i) \right)^2 = \\ &= \mathbf{E} \left[\sum_1^n (X_i - \mathbf{E} X_i)^2 + \sum_{i \neq j} (X_i - \mathbf{E} X_i)(X_j - \mathbf{E} X_j) \right] = \\ &= \sum_1^n \mathbf{E} (X_i - \mathbf{E} X_i)^2 + \sum_{i \neq j} \mathbf{E} (X_i - \mathbf{E} X_i) \cdot \mathbf{E} (X_j - \mathbf{E} X_j) = \\ &= \sum_1^n \mathbf{E} (X_i - \mathbf{E} X_i)^2 = \sum_1^n \mathbf{D} X_i. \end{aligned}$$

Как будет видно в дальнейшем, вывод ряда вероятностных моделей строится на стохастическом представлении наблюдаемой случайной величины в виде суммы независимых случайных величин: $X = X_1 + \dots + X_n$, и при этом распределение каждой X_i , $i = 1, \dots, n$ имеет достаточно простой вид, например, вычислить моменты X_i намного проще, чем моменты X . В таком случае формулы предложения 8.3 указывают прямой путь к вычислению моментов, а иногда и распределения, случайной величины X . В сущности, мы уже использовали технику таких представлений, когда выводили биномиальное распределение – распределение числа успехов в испытаниях Бернулли.

Пример 8.2 (о некоторых свойствах биномиального распределения). Результат каждого i -го испытания в схеме Бернулли можно регистрировать как значение индикаторной функции успеха, обозначая цифрой 1 успех, а цифрой 0 неудачу. Таким образом, с i -ым испытанием соотносится случайная величина X_i , принимающая значение 1 с вероятностью p и значение 0 с вероятностью $1 - p$. Последовательности из n независимых испытаний Бернулли ставится в соответствие случайный вектор $X^{(n)} = (X_1, \dots, X_n)$, состоящий из независимых, одинаково распределенных по закону $B(1, p)$ компонент (напомним, $B(1, p)$ есть частный случай биномиального распределения, которое мы назвали двухточечным распределением). В таких обозначениях случайная величина X , реализация которой равна числу успехов в n испытаниях (числу X_i , принявших значение 1), представима в виде $X = \sum_1^n X_i$, и в силу предложения 8.3

$$\mathbf{E}X = \sum_1^n \mathbf{E}X_i = n\mathbf{E}X_1, \quad \mathbf{D}X = \sum_1^n \mathbf{D}X_i = n\mathbf{D}X_1.$$

Имеем: $\mathbf{E}X_1 = 1 \cdot p + 0 \cdot (1 - p) = p$, $\mathbf{E}X_1^2 = \mathbf{E}X_1 = p$, $\mathbf{D}X_1 = p - p^2 = p(1 - p)$, откуда немедленно получаем известные нам и полученные в результате более сложных выкладок формулы моментов биномиального распределения: $\mathbf{E}X = np$, $\mathbf{D}X = np(1 - p)$.

Укажем еще на одно интересное применение стохастического представления биномиальной случайной величины X в виде суммы независимых случайных величин.

Предложение 8.4 (теорема сложения для биномиального распределения). Если X_1, \dots, X_m независимы в совокупности и $X_k \sim B(n_k, p)$, $k = 1, \dots, m$, то

$$X = \sum_1^m X_k \sim B(n, p),$$

где $n = n_1 + \dots + n_m$.

Доказательство. Каждое X_k есть сумма n_k независимых, одинаково распределенных по закону $B(1, p)$ случайных величин. Следовательно, X есть сумма n таких же величин, откуда $X \sim B(n, p)$.

Распределения, для которых справедливы *теоремы сложения*, составляют особый класс *устойчивых* законов распределений, и изучению свойств таких распределений посвящаются отдельные монографии. Вы, наверное, догадываетесь, что устойчивым является пуассоновское распределение, как предел биномиального. В дальнейшем мы покажем, что это в действительности так, разработав более совершенный математический аппарат доказательств теорем сложения. А сейчас мы докажем устойчивость нормального закона, получив предварительно общую формулу для распределения суммы независимых случайных величин.

Предложение 8.5 (*формула свертки распределений*). Пусть X_1 и X_2 независимы и имеют непрерывные распределения с функциями плотности $f_1(x)$ и, соответственно, $f_2(x)$ по мере Лебега $d\mu = dx$. Тогда функция плотности $f(x)$ распределения случайной величины $X = X_1 + X_2$ есть свертка функций f_1 и f_2 :

$$f(x) = \int_{-\infty}^{\infty} f_1(t)f_2(x-t)dt = \int_{-\infty}^{\infty} f_2(t)f_1(x-t)dt.$$

Доказательство. Совместная функция плотности $f(x_1, x_2)$ независимых случайных величин X_1 и X_2 равна (см. предложение 8.2) произведению их функций плотности: $f(x_1, x_2) = f_1(x_1)f_2(x_2)$. Используя известную нам формулу (см. формулы после определения 8.1)

$$P((X_1, X_2) \in B) = \int_B f_1(x_1)f_2(x_2)dx_1dx_2$$

для вычисления вероятностей попадания случайного вектора в любую измеримую область B на плоскости \mathbb{R}^2 , представим функцию распределения суммы случайных величин в виде

$$F(x) = P(X_1 + X_2 < x) = \int_{t+s < x} f_1(t)f_2(s)dt ds = \int_{-\infty}^{\infty} f_1(t)dt \int_{-\infty}^{x-t} f_2(s)ds.$$

Дифференцируя правую часть последнего равенства по x , получаем искомую первую формулу для плотности $f(x)$. Вторая формула справедлива в силу симметрии вхождения функций f_1 и f_2 в интегральное представление $F(x)$.

Непосредственное применение формулы свертки к функциям плотности нормального распределения показывает, что имеет место

Предложение 8.6 (теорема сложения для нормального распределения.) Если X_1, \dots, X_n независимы в совокупности и каждое $X_k \sim \mathcal{N}(\mu_k, \sigma_k^2)$, $k = 1, \dots, n$, то

$$X = \sum_1^n X_k \sim \mathcal{N}\left(\sum_1^n \mu_k, \sum_1^n \sigma_k^2\right).$$

Доказательство. Предложение достаточно доказать для случая $n = 2$, поскольку для произвольного числа слагаемых доказательство проводится методом индукции. При $n = 2$ формула свертки дает следующее выражение для функции плотности $f(x)$ случайной величины $X = X_1 + X_2$:

$$f(x) = \frac{1}{2\pi\sigma_1\sigma_2} \int_{-\infty}^{\infty} \exp\left\{-\frac{1}{2}\left(\frac{t-\mu_1}{\sigma_1}\right)^2 - \frac{1}{2}\left(\frac{x-t-\mu_2}{\sigma_2}\right)^2\right\} dt.$$

Приводя квадратическую форму под знаком экспоненты к виду

$$-\frac{1}{2}\left(\frac{t-a}{b}\right)^2 + h(a,b),$$

где a и b зависят от параметров μ_i и σ_i , $i = 1, 2$, и используя известную нам формулу

$$\frac{1}{\sqrt{2\pi b}} \int_{-\infty}^{\infty} \exp\left\{-\frac{1}{2}\left(\frac{t-a}{b}\right)^2\right\} dt = 1,$$

находим искомую функцию плотности

$$f(x) = \frac{1}{\sqrt{2\pi(\sigma_1^2 + \sigma_2^2)}} \exp\left\{-\frac{(x - \mu_1 - \mu_2)^2}{2(\sigma_1^2 + \sigma_2^2)}\right\}.$$