

6. Случайные величины и их распределения.

I. Дискретные случайные величины

Определение. Дискретной случайной величиной называется числовая функция ξ , заданная на Ω и принимающая конечное или счетное множество значений, $\xi(\omega) \in \{x_1, x_2, \dots, x_n, \dots\}$, такая, что множества $A_k = \{\omega: \xi(\omega) = x_k\}$ являются событиями. Набор вероятностей

$$p_k = P(A_k) = P\{\xi = x_k\}, \quad k = 1, 2, \dots$$

называется **распределением вероятностей случайной величины** ξ . Числа $x_1, x_2, \dots, x_n, \dots$ предполагаются различными, тогда события $\{A_k\}$ попарно несовместны и в сумме составляют достоверное событие Ω , следовательно, $\sum_k p_k = 1$ (это равенство называют **условием нормировки**).

Пример 1. Вероятность ошибки при передаче сообщения по каналу связи равна p . Для борьбы с ошибками сообщение передается 5 раз, оно считается принятым правильно, если число ошибочных экземпляров не превышает 2. Считая ошибки в канале независимыми, построить распределение случайной величины ξ , равной числу ошибочных копий; найти вероятность правильного приема сообщения.

Применяя схему Бернулли с $n = 5$ и заданным p , получаем

$$p_k = P\{\xi = k\} = C_5^k p^k (1 - p)^{5-k}, \quad k = 0, 1, 2, 3, 4, 5.$$

Вероятность правильного приема равна $p_0 + p_1 + p_2$, например, при $p = 0,3$ эта вероятность составляет 0,837.

Биномиальное распределение есть распределение целочисленной случайной величины ξ , равной числу успехов в схеме Бернулли (см. Параграф 4) с числом испытаний n и вероятностью успеха p :

$$p_k = P\{\xi = k\} = C_n^k p^k (1 - p)^{n-k}, \quad k = 0, 1, 2, \dots, n.$$

Геометрическое распределение (число попыток до первого успеха).

Предположим, в условиях схемы Бернулли испытания проводятся до наступления первого успеха и пусть ξ обозначает номер испытания, в котором наступил первый успех. Событие $\{\xi = k\}$ означает, что вначале было подряд $k - 1$ неуспехов, а затем успех; вследствие независимости испытаний вероятность такой последовательности равна произведению вероятностей, так что

$$p_k = P\{\xi = k\} = (1 - p)^{k-1} p, \quad k = 1, 2, 3, \dots$$

Это и есть геометрическое распределение.

Проверим, что для него выполнено условие нормировки, то есть сумма всех вероятностей равна 1. Обозначая для краткости $q = 1 - p$, запишем

$$\sum_{k=1}^{\infty} p_k = p \sum_{k=1}^{\infty} q^{k-1} = p \sum_{k=0}^{\infty} q^k = p \frac{1}{1 - q} = 1,$$

здесь использована формула для суммы геометрической прогрессии:

$$\sum_{k=0}^{\infty} q^k = \frac{1}{1-q}.$$

Распределение Пуассона. Теорема Пуассона в схеме Бернулли (Теорема 4.1.) привела к следующему дискретному распределению:

$$p_k = P\{\xi = k\} = \frac{\lambda^k}{k!} e^{-\lambda}, \quad k = 0, 1, 2, \dots$$

Условие нормировки для него легко проверяется:

$$\sum_{k=0}^{\infty} p_k = e^{-\lambda} \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} = 1,$$

поскольку сумма справа есть разложение экспоненты e^{λ} в ряд Тейлора.

Пример 2. Геометрическое распределение возникло как распределение числа испытаний до первого успеха. Рассмотрим теперь случайную величину η , равную числу испытаний до m -го успеха ($m \geq 1$).

Случайная величина η принимает значение k на такой последовательности длины k ($k \geq m$), состоящей из нулей и единиц, которая заканчивается единицей, а на предыдущих $k-1$ позициях расположены $(m-1)$ нулей и $(k-1) - (m-1) = (k-m)$ единиц. Вероятность получить такую последовательность в серии независимых испытаний равна $p^{m-1}(1-p)^{k-m}p$, а общее число таких последовательностей есть C_{k-1}^{m-1} , следовательно,

$$p_k = P\{\eta = k\} = C_{k-1}^{m-1} p^m (1-p)^{k-m}, \quad k = m, m+1, m+2, \dots$$

Проверим условие нормировки для этого распределения:

$$\sum_{k=m}^{\infty} p_k = p^m \sum_{k=m}^{\infty} C_{k-1}^{m-1} (1-p)^{k-m} = \frac{p^m}{(m-1)!} \sum_{k=m}^{\infty} \frac{(k-1)!}{(k-m)!} q^{k-m};$$

выполнив замену переменной $k-m = l$, получим

$$\begin{aligned} \sum_{k=m}^{\infty} p_k &= \frac{p^m}{(m-1)!} \sum_{k=m}^{\infty} \frac{(l+m-1)!}{l!} q^l = \\ &= \frac{p^m}{(m-1)!} \left[(m-1)! + m! q + \frac{(m+1)!}{2!} q^2 + \dots \right] = p^m \left[1 + mq + \frac{m(m+1)}{2!} q^2 + \dots \right] \end{aligned}$$

Сумма в скобках, как легко убедиться, представляет собой разложение в ряд Тейлора дроби

$$\frac{1}{(1-q)^m} = \frac{1}{p^m}$$

откуда и следует равенство единице всей суммы вероятностей.

Упражнение 1. Отрицательным биномиальным распределением называется распределение случайной величины, равной числу неудач, наступивших до m -го успеха в схеме Бернулли. Напишите формулу для распределения вероятностей этой случайной величины.

II. Непрерывные случайные величины

Определение. Пусть на множестве Ω задана числовая функция $\xi: \Omega \rightarrow \mathbb{R}$ (\mathbb{R} – множество вещественных чисел). Функция ξ называется случайной величиной, если для любого $x \in \mathbb{R}$ множество $\{\omega: \xi(\omega) \leq x\}$ является случайным событием.

Определение. Функция вещественной переменной $x \in \mathbb{R}$

$$F(x) = F_{\xi}(x) = P\{\xi \leq x\}$$

называется **функцией распределения** случайной величины ξ .

Теорема 1. Функция распределения обладает следующими свойствами:

1) Для $x_1 < x_2$ справедливо равенство

$$P\{x_1 < \xi \leq x_2\} = F(x_2) - F(x_1),$$

из которого также следует, что $F(x)$ неубывающая функция, $F(x_2) \geq F(x_1)$;

2) имеют место свойства

$$F(+\infty) = 1, \quad F(-\infty) = 0.$$

Доказательство. Если $x_1 < x_2$, по свойству аддитивности ($P(A) = P(B) + P(C)$, если $A = B + C$) событие $\{\xi \leq x_2\}$ можно представить как сумму двух несовместных событий, $\{\xi \leq x_2\} = \{\xi \leq x_1\} + \{x_1 < \xi \leq x_2\}$, откуда и следует первое утверждение. Из него получаем свойство, полезное при вычислениях вероятностей.

$$F(x_2) = F(x_1) + P\{x_1 < \xi \leq x_2\} \quad (1)$$

Второе утверждение следует из равенств $\{\xi \leq -\infty\} = \emptyset$, $\{\xi \leq \infty\} = \Omega$. ■

Определение. Плотностью распределения случайной величины ξ , принимающей значения на некотором интервале вещественной оси, называется функция $p_{\xi}(x)$, такая что для любых $a < b$

$$P\{a < \xi \leq b\} = \int_a^b p_{\xi}(x) dx.$$

Из этого определения и свойств интеграла Римана следует:

1). $F_{\xi}(x) = \int_{-\infty}^x p_{\xi}(x) dx$;

$$F_{\xi}(x) = \int_{-\infty}^x p_{\xi}(u) du = P\{-\infty < \xi \leq x\}$$

2). в точках непрерывности функции $p_{\xi}(x)$

$$p_{\xi}(x) = \frac{dF_{\xi}(x)}{dx}$$

3). для любого интервала $(a, b]$

$$F_{\xi}(b) - F_{\xi}(a) = P\{a < \xi \leq b\} = \int_a^b p_{\xi}(x) dx.$$

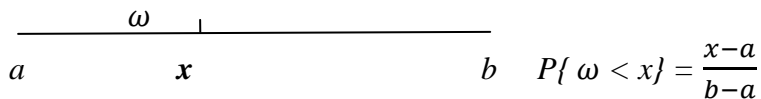
Если существует плотность, то говорят, что распределение $F_{\xi}(x)$ является абсолютно непрерывным. Плотность распределения является неотрицательной функцией, для которой выполнено **условие нормировки**:

$$\int_{-\infty}^{\infty} p_{\xi}(x) dx = 1;$$

оно получается из предыдущей формулы при $a = -\infty$, $b = \infty$.

Замечание. Функции распределения рассматриваются и для непрерывных и для дискретных случайных величин. Например, функция распределения $F_{\xi}(x)$ случайной величины из Примера 1. имеет вид ступенчатой функции, слева от 0 она равна 0, в каждой точке 0,1,2,3,4,5 имеет скачок величины p_k , так что $F_{\xi}(5) = 1$. Напишите формулу для функции распределения дискретной случайной величины в общем виде.

Рассмотрим теперь некоторые простые примеры абсолютно непрерывных распределений (далее в тексте абсолютно непрерывное распределение будем называть просто непрерывным).



$$P\{\omega < x\} = \frac{x-a}{b-a}$$

Равномерное распределение. Для равномерного распределения на интервале $[a, b]$ плотность принимает постоянное значение на этом интервале и равна нулю вне его. Если обозначить c это постоянное значение, то согласно условию нормировки, $\int_a^b c \cdot dx = 1$, откуда получаем значение константы; таким образом, равномерное распределение имеет плотность

$$p(x) = \begin{cases} \frac{1}{b-a}, & x \in [a, b] \\ 0, & x \notin [a, b], \end{cases}$$

соответствующая функция распределения

$$F(x) = \begin{cases} \frac{x-a}{b-a}, & x \in [a, b] \\ 0, & x < a \\ 1, & x > b \end{cases}.$$

Наиболее часто применяется равномерное распределение на интервале $[0, 1]$:

$$p(x) = \begin{cases} 1, & x \in [0, 1] \\ 0, & x \notin [0, 1], \end{cases} \quad F(x) = \begin{cases} x, & x \in [0, 1] \\ 0, & x < 0 \\ 1, & x > 1. \end{cases}$$

Упражнение. Если задана случайная величина ξ с равномерным распределением на $[0, 1]$, то как из нее получить случайную величину с равномерным распределением на $[a, b]$.

Экспоненциальное распределение применяется для описания систем, поведение которых можно охарактеризовать как отсутствие памяти (в вероятностном смысле).

Чтобы понять, как выглядит статистическое описание системы "без памяти", применим условную вероятность. Представим себе систему, где в случайные моменты времени ξ происходят некоторые события, а нас интересуют вероятности вида $P\{\xi > t\}$, где t обозначает время ожидания следующего события, а t – заданное число. То есть, мы хотим иметь возможность вычислять вероятности того, что следующего события придется ждать не меньше, чем t единиц времени (например, t секунд).

Предположим, что эта гипотетическая система ведет себя следующим образом: если мы уже ждали не менее s секунд, то вероятность того, что придется ждать еще t секунд, не зависит от этого s . Мы предполагаем, что для любого t событие $\{\xi > t\}$ является случайным событием, поэтому на языке условных вероятностей это свойство отсутствия памяти выглядит так:

$$P\{\xi > s + t / \xi > s\} = P\{\xi > t\}$$

но

$$\begin{aligned} P\{\xi > s + t / \xi > s\} &= \\ &= \frac{P(\{\xi > s + t\} \cap \{\xi > s\})}{P\{\xi > s\}} = \frac{P\{\xi > s + t\}}{P\{\xi > s\}} = P\{\xi > t\} \end{aligned}$$

следовательно,

$$P\{\xi > s + t\} = P\{\xi > s\}P\{\xi > t\}$$

Обозначим $P\{\xi > t\} = G(t)$, тогда для функции $G(t)$ мы получили уравнение

$$G(s + t) = G(s)G(t) \quad (2)$$

которое имеет единственное решение $G(t) = e^{-\lambda t}$, λ – некоторое положительное число. Если предположить, что $G(t)$ дифференцируема, то решение получить несложно.

Действительно, из функционального уравнения (2) получаем

$$\frac{G(t + s) - G(t)}{s} = G(t) \left(\frac{G(s) - 1}{s} \right);$$

очевидно, $G(0) = 1$, поэтому при переходе к пределу $s \rightarrow 0$ выражение справа в скобках стремится к производной $G'(0)$, ее обозначим $G'(0) = a$ и получаем дифференциальное уравнение для нашей функции,

$$\frac{dG(t)}{dt} = aG(t)$$

Оно имеет решение $\ln G = at + \text{const}$ ($\text{const} = 0$, так как $G(0) = 1$), тогда $G = e^{at}$, но при $t \rightarrow \infty$ функция G стремится к 0, а потому константа a должна быть отрицательной, $a = -\lambda$.

$$G(t) = P\{\xi > t\} = 1 - P\{\xi < t\} = 1 - F_{\xi}(x)$$

Конструкцию системы "без памяти" можно применять к моделированию таких процессов как радиоактивный распад атомных ядер, распад словарного состава языка и т.д. Для этого рассмотрим случайную величину ξ , принимающую неотрицательные значения, с функцией распределения

$$F(x) = 1 - e^{-\lambda x}$$

это - **экспоненциальное распределение**. Вероятность $G(x) = 1 - F(x) = P\{\xi > x\}$, соответствует свойству отсутствия памяти; если ξ имеет смысл времени ожидания, то вероятность того, что придется ждать еще время s при условии, что ждем уже t , от t не зависит:

$$P\{\xi > t + s / \xi > t\} = P\{\xi > s\}$$

Экспоненциальное распределение имеет плотность

$$p(x) = \lambda e^{-\lambda x}, x \geq 0.$$

Смысл параметра λ поясняет следующий пример.

Пример 3. Медианой распределения вероятностей назовем число m , которое удовлетворяет условию $F(m) = \frac{1}{2}$. Физики называют m периодом полураспада: пусть ξ – время жизни нестабильного атомного ядра; поскольку модель радиоактивного распада ядер обладает свойством отсутствия памяти, то за время m распадется половина имеющейся массы вещества. Предположим, экспериментально установлено, что период полураспада равен T лет, тогда для экспоненциального распределения из равенства $F(T) = \frac{1}{2}$ получаем

$$\lambda = \frac{\ln 2}{T}.$$

В лингвистике существует такая точка зрения: слова языка живут по тому же закону, что и ядра атомов радиоактивного вещества: вероятность случайно взятому слову прожить время x соответствует экспоненциальному распределению. Предположим, статистическими измерениями установлено – период полураспада базового словарного состава для некоторых древних языков составляет $T = 2000$ лет. Рассмотрим гипотезу, согласно которой венгерский и финский языки произошли от одного языка; если согласно статистическим оценкам, два языка имеют 21% общих слов, то сколько лет назад эти языки разделились?

Обозначим этот срок t_0 , тогда получаем уравнение: $P\{\xi > t_0\} = 1 - F(t_0) = 0,21$, откуда $e^{-\lambda t_0} = 0,21$; решая относительно t_0 и заменяя λ через период полураспада, находим

$$t_0 = -\frac{\ln(0,21)}{\lambda} = -\frac{\ln(0,21)}{\ln 2} T \approx 4500 \text{ лет.}$$

По другой оценке, эти два языка имеют 27% общих слов.

Пример 4. Рассмотрим типичную задачу из учебников теории вероятностей на вычисления, связанные с распределением случайной величины. Пусть задана плотность распределения

$$p(x) = c \cdot \cos(x), -\frac{\pi}{2} \leq x \leq \frac{\pi}{2}$$

Непрерывно постройте график этой плотности. Необходимо найти постоянную c , построить функцию распределения и вычислить вероятность $P\{0 \leq \xi \leq \frac{\pi}{4}\}$.

Подобная константа c называется **нормировочная постоянная**, она часто рассматривается в подобных задачах. Константу c находим из условия нормировки:

$$1 = \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} p(x) dx = c \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \cos(x) dx = c \cdot \left[\sin\left(\frac{\pi}{2}\right) - \sin\left(-\frac{\pi}{2}\right) \right] = 2c$$

Функция распределения тогда равна

$$F_{\xi}(x) = \int_{-\frac{\pi}{2}}^x p_{\xi}(x) dx$$

$$F(x) = \frac{1}{2} [\sin(x) + 1],$$

и с помощью свойства 3) функции распределения находим

$$P\left\{0 \leq \xi \leq \frac{\pi}{4}\right\} = \frac{1}{2} \sin\left(\frac{\pi}{4}\right) = \frac{\sqrt{2}}{4}.$$

Самостоятельно постройте графики плотности и функции распределения в этом примере.
Нарисовать картинку!!!