Вероятность и случайные величины

1. Случайность в теории вероятностей и статистике

Большинство явлений окружающего мира слишком сложны для того, чтобы их можно было описать простыми детерминированными законами. Например, предсказания атмосферных явлений не являются абсолютно точными, хотя законы, по которым существуют отдельные молекулы атмосферы, исследованы достаточно хорошо. Тем не менее, во взаимодействии молекул принимают участие слишком много различных факторов, что делает невозможным построение исчерпывающей прогнозирующей модели.

Более простым процессом является подбрасывание кубика, но и здесь точное предсказание результата не представляется возможным. Можно отойти от построения сложной физической модели подбрасывания и перейти к рассмотрению некоего «черного ящика». Этот ящик по неизвестным законам генерирует случайные события, соответствующие числам, выпадающим на кубике. В математике такие «черные ящики» называются *случайными величинами*, а генерируемые события — *реализациями случайной величины*. Набор реализаций случайной величины называется *выборкой* из нее.

Нельзя предугадать, какое событие произойдет в следующий момент наблюдения за «черным ящиком», однако, если вести наблюдения достаточно долго, начнут прослеживаться определенные закономерности. Например, каждое число на кубике будет выпадать примерно одинаковое количество раз. Именно такие закономерности являются объектом изучения теории вероятностей и математической статистики.

Если проводить эксперимент со случайной величиной бесконечно, то каждому событию можно будет поставить в соответствие его *вероятность* — долю испытаний, завершившихся наступлением события (определение нестрогое). Вероятность не может быть измерена на практике в силу данного определения.

Теория вероятностей изучает модели случайных величин и свойства этих моделей. **Статистика и анализ данных** пытаются по свойствам конечных выборок определить свойства случайной величины, чтобы понять, как она будет вести себя в будущем. Осуществить такой переход позволяет *закон больших чисел*: на большой выборке частота события хорошо приближает его вероятность (формулировка нестрогая).

2. Свойства вероятности

Основные свойства вероятности:

1) $0 \leqslant P(A) \leqslant 1$, то есть вероятность любого события лежит на отрезке от нуля до единицы.
2) $P(\varnothing) = 0$ — событие, вероятность которого равна нулю, называется *невозможным*.
3) $P(\bar{A}) + P(A) = 1$. Для события A всегда можно определить событие «не A», которое соответствует событию «A не произошло». Вероятности таких событий в сумме дают единицу.

Для пары событий возможны следующие отношения:

1) *Вложенность*: $A \subseteq B$. Примером такого отношения является стрельба из арбалета по мишени. Событие A — это попадание в «десятку», событие B — набор пяти или более очков. Тогда событие A вложено в событие B, причем мишень визуализирует это буквально (см. рис. 1).

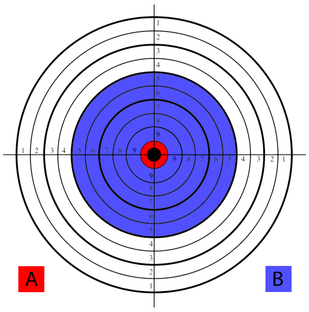

Рис. 1.

Вероятности таких событий связаны друг с другом неравенством:

$$A \subseteq B \quad \Rightarrow \quad P(A) \leqslant P(B).$$

2) *Произведение событий* AB и *сумма событий* $A+B$. Пусть событие A — это попадание в синюю область на мишени, событие B — в зеленую (см. рис. 2). Тогда событие AB — это пересечение этих областей, то есть произошло каждое из данных событий (см. рис. 3).

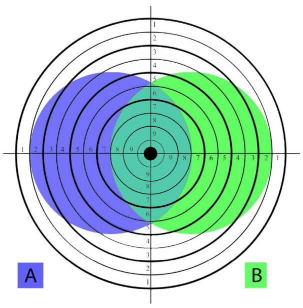

Рис. 2.

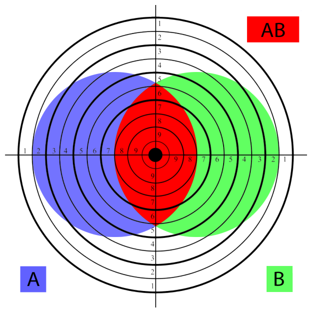

Рис. 3.

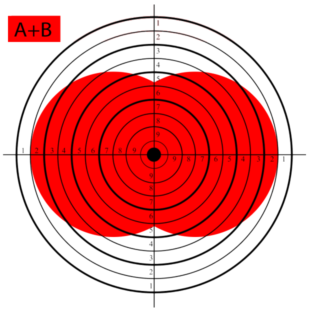

Рис. 4.

Сумма событий $A+B$ означает попадание в область, соответствующую объединению этих двух областей, то есть произошло хотя бы одно из данных событий (см. рис. 4).

Вероятности событий AB и $A+B$ связаны следующим образом:

$$P(A+B) = P(A) + P(B) - P(AB).$$

3) *Дополнение*: $B\backslash A$. Здесь происходит событие B, но не происходит событие A. Вероятность такого события задаётся следующим выражением:

$$P\big(B\backslash A\big) = P(B) - P(AB).$$

В случае, когда A полностью лежит в B, формула упрощается:

$$P\big(B\backslash A\big) = P(B) - P(A).$$

В примере из пункта 1 это соответствует попаданию в область от пяти до девяти очков (см. рис. 5).

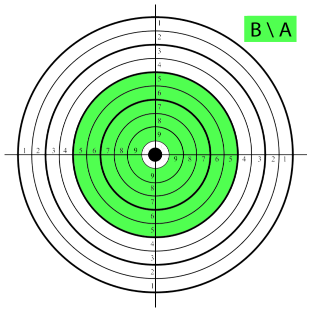

Рис. 5.

4) *Независимость событий*: $P(AB)$. Пусть событие A — это попадание в нижнюю половину мишени, а событие B — попадание в правую половину. Событием AB будет пересечение этих двух событий — попадание в нижнюю правую четверть (см. рис. 6).

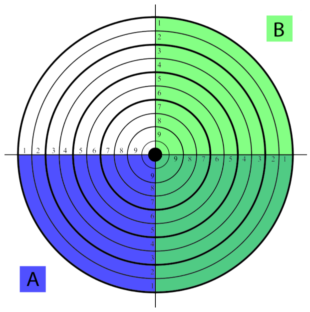

Рис. 6.

События A и B являются *независимыми*, если вероятность их пересечения равна произведению вероятностей компонент:

$$P(AB) = P(A)P(B).$$

В примере с мишенью, если можно считать, что арбалет хорошо настроен и не дует ветер, то:

$$P(A) = 0{,}5, \quad P(B) = 0{,}5, \quad P(AB) = 0{,}25,$$

что означает независимость двух рассматриваемых событий.

Пусть событие A — это попадание в «десятку», событие B — попадание в любое место мишени. Если известно, что событие B произошло, то вероятность события A повышается. *Условная вероятность* события A при условии, что произошло событие B, определяется следующим образом:

$$P\big(A|B\big) = \frac{P(AB)}{P(B)}.$$

Если попадание в мишень происходит в восьмидесяти процентах случаев, а в «десятку» — в пяти процентах, то:

$$P(AB) = P(A) = 0{,}05 \quad \Rightarrow \quad P\big(A|B\big) = \frac{0{,}05}{0{,}8} = 0{,}0625.$$

Формула полной вероятности имеет вид:

$$P(A) = P\big(A|B\big)P(B) + P\big(A|\bar{B}\big)P\big(\bar{B}\big).$$

Условные вероятности двух событий связаны друг с другом с помощью **формулы Байеса**:

$$P\big(A|B\big) = \frac{P(A)P(B|A)}{P(B)}.$$

3. Дискретные случайные величины

В испытании с подбрасыванием кубика возможны шесть исходов. Эти исходы можно пронумеровать $1, 2, 3, 4, 5, 6$ согласно значению, выпавшему на кубике. Данному множеству исходов ставится в соответствие вектор вероятности:

$$\{1, 2, 3, 4, 5, 6\} \quad \Longrightarrow \quad \begin{pmatrix} p_1 \\ p_2 \\ p_3 \\ p_4 \\ p_5 \\ p_6 \end{pmatrix},$$

причем:

$$p_i \geqslant 0,\ i = 1, \ldots, 6; \quad \sum_{i=1}^{6} p_i = 1.$$

Именно так устроена *дискретная случайная величина* X. Она принимает счетное множество значений $A = \{a_1, a_2, a_3, \ldots\}$ с вероятностями $p_1, p_2, p_3, \ldots$, где:

$$p_i \geqslant 0\ \forall i, \quad \sum_{i=1}^{\infty} p_i = 1.$$

Вероятность того, что $X = a_i$, равна p_i, тогда *функция вероятности* имеет вид:

$$P\big(X = a_i\big) = p_i.$$

Одной из наиболее часто встречающихся случайных величин является **дискретная случайная величина с двумя исходами**. Примером в данном случае является подбрасывание монеты. Можно обозначить выпадение решки за 1 («успех»), а выпадение орла — за 0 («неудача»). Пусть вероятность «успеха» равна p:

$$P(X = 1) = p, \quad P(X = 0) = 1 - p.$$

Именно так устроена *бернуллиевская случайная величина*:

$$X \sim \mathrm{Ber}(p).$$

Другим примером является **сумма независимых бинарных случайных величин**. Пусть вероятность попадания мяча в баскетбольное кольцо равна p, имеется n попыток, а число попаданий равно X. В силу независимости попыток:

$$P(X = n) = p^n.$$

В общем случае:

$$P(X = k) = C_n^k p^k (1 - p)^{n-k}, \ k = 0, 1, 2, \ldots, n.$$

Здесь используется *биномиальный коэффициент*:

$$C_n^k = \frac{n!}{k!(n-k)!}.$$

Такая случайная величина X называется *биномиальной*. Биномиальное распределение имеет два параметра: целочисленный n и $p \in [0, 1]$:

$$X \sim \mathrm{Bin}(n, p).$$

Еще одним классом дискретных случайных величин являются **счетчики**. В качестве иллюстрации можно рассмотреть текст романа Набокова «*Приглашение на казнь*». Из всех произведений Набокова составляется словарь, затем текст данного романа сверяют с этим словарем. Некоторые слова не встречаются в романе ни разу, например, слово «шахматист». Некоторые слова встречаются в романе редко, например, слово «аляповатость» (1 раз). Есть и слова, которые используются очень часто (например, «год» — 21 раз).

Пусть X — это число использований слова в тексте. Вероятность того, что X равно k, можно описать *распределением Пуассона*:

$$P(X = k) = \frac{\lambda^k \, \mathrm{e}^{-\lambda}}{k!}, \quad \lambda > 0, \quad k = 0, 1, 2, \ldots.$$

$$X \sim \mathrm{Pois}(\lambda).$$

Распределением Пуассона описывается, например, число автобусов, которые проезжают за час мимо автобусной остановки, или число радиоактивных распадов, которые улавливает счетчик Гейгера.

4. Непрерывные случайные величины

Непрерывные случайные величины нельзя задавать с помощью функции вероятности в силу того, что если множество A несчетное, то вероятность события нулевая:

$$|A| > \aleph_0 \quad \Rightarrow \quad P(X = a) = 0 \ \forall a \in A.$$

Первый способ определения таких величин — с помощью *функции распределения* (см. рис. 7):

$$F(x) = P(X \leqslant x).$$

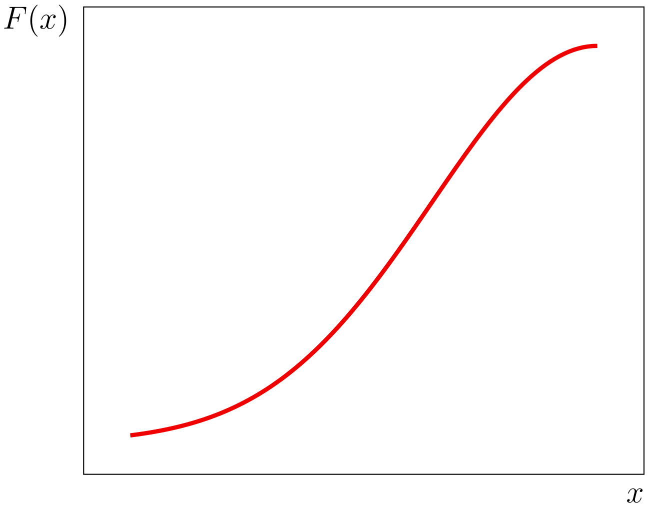

Рис. 7.

Функция распределения всегда принимает значение от 0 до 1 и не убывает по аргументу x.

Второй способ определения непрерывных случайных величин — с помощью *плотности распределения*:

$$f(x): \int_a^b f(x)\,dx = P(a \leqslant X \leqslant b).$$

Плотность связана с функцией распределения следующим образом:

$$F(x) = \int\limits_{-\infty}^{x} f(u)\,du.$$

Для непрерывной случайной величины верно равенство:

$$\int\limits_{-\infty}^{+\infty} f(u)\,du = P(-\infty \leqslant X \leqslant +\infty) = 1.$$

В отличие от функции распределения, плотность распределения на графике может принимать различный вид.

Примером непрерывной случайной величины является *равномерная случайная величина.* Пусть X — это время ожидания на светофоре до того, как можно будет перейти дорогу. Если на светофоре нет счетчика, то нельзя угадать, сколько именно придется ждать зеленого сигнала. Время ожидания может быть любым числом от 0 до, например, 30 секунд. Именно так устроено *равномерное распределение* — случайная величина на отрезке $[a, b]$ принимает любое значение с одинаковой вероятностью:

$$X \sim U(a, b).$$

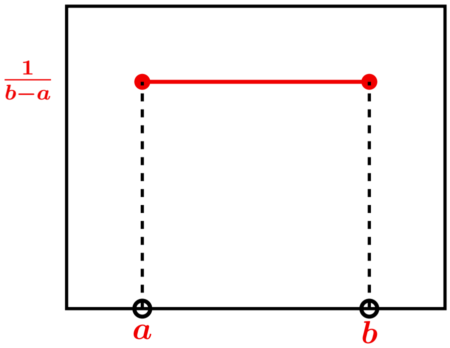

Рис. 8.

Плотность вероятности для равномерной случайной величины имеет вид (см. рис. 8):

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

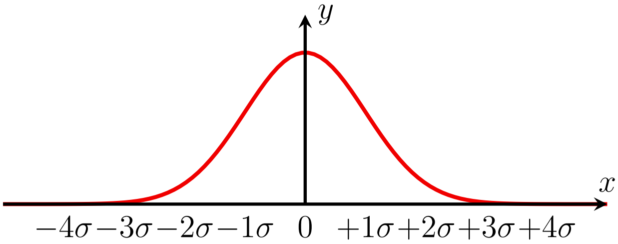

Рис. 9.

Другим примером непрерывной случайной величины является *нормальная случайная величина.* Человек из примера любит приходить на работу к 11 часам, но точное время его прихода варьируется — иногда он проснется раньше, иногда он опаздывает. Таким образом, точное время его прихода на работу X представляет собой результат взаимодействия большого количества слабо зависимых случайных факторов. Именно такие величины хорошо моделируются *нормальным (Гауссовым) распределением*:

$$X \sim N\left(\mu, \sigma^2\right).$$

В данном примере параметр μ отвечает за среднее время прихода, а параметр σ определяет разброс вокруг среднего. Функция плотности вероятности нормального распределения имеет вид (см. рис. 9):

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right).$$

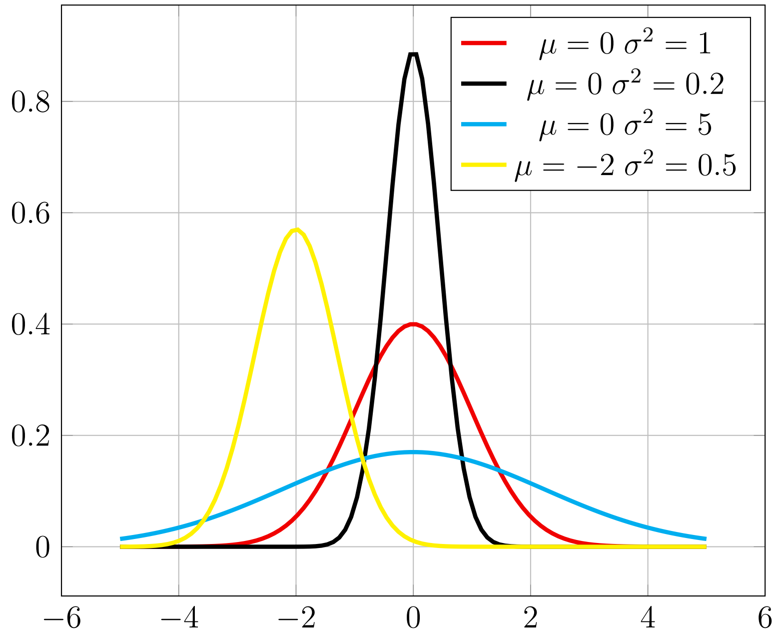

Рис. 10.

Варьируя значения параметров μ и σ, можно влиять на форму графика функции плотности вероятности нормального распределения (см. рис. 10).

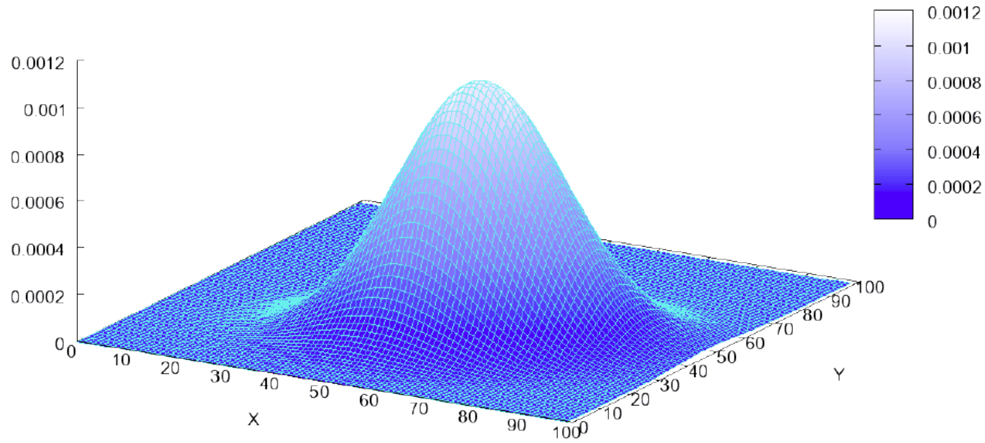

Рис. 11.

Нормальное распределение может быть *многомерным*. В этом случае случайная величина является не скалярной, а векторной:

$$X \sim N(\mu, \Sigma), \quad \mu \in \mathbb{R}^k, \Sigma \in \mathbb{R}^{k \times k},$$

причем матрица Σ является положительно определенной. Функция плотности вероятности в данном случае имеет вид (см. рис. 11):

$$f(x) = (2\pi)^{-\frac{k}{2}} \left|\Sigma\right|^{-\frac{1}{2}} \exp\left(-\frac{1}{2}(x-\mu)^T |\Sigma|^{-1} (x-\mu)\right).$$