# 2 КРАТКИЕ СВЕДЕНИЯ ИЗ ТЕОРИИ ВЕРОЯТНОСТИ И МАТЕМАТИЧЕСКОЙ СТАТИСТИКИ

## 2.1.Случайные величины

Всякая случайная величина $X$ обладает тем свойством, что при производстве в неизменных условиях $n$ наблюдений будут зарегистрированы неодинаковые значения $\chi_1$, $\chi_2$, …, $\chi_n$, среди которых, однако, могут встречаться и повторяющиеся. Различают *дискретные* и *непрерывные* случайные величины.

**Дискретная случайная величина** принимает лишь отдельные, изолированные одно от другого значения. Таким свойством обладают, прежде всего, параметры, называемые атрибутивными признаками (например – цвет предмета, сорт продукции, годное или бракованное изделие и т. д.). Отдельные значения подобных параметров определяются путем счета и поэтому значениями дискретной случайной величины могут быть только натуральные числа.

**Непрерывная случайная** величина принимает любые значения из некоторого свойственного ей числового интервала. Таким свойством обладают количественные признаки (механические свойства материала, фактические размеры продукции, производительность агрегата при обработке конкретного профилеразмера и т. п.) Отдельные значения таких параметров определяются путем измерений или расчетов и поэтому могут быть любыми действительными числами.

И для дискретной, и для непрерывной случайной величины характерны, по крайней мере, две закономерности:

1. Всякая случайная величина имеет некоторый ограниченный с обеих сторон интервал варьирования, величина и положение которого на числовой оси обусловлены ее физической природой. Случайность проявляется лишь в том, что значение, которое будет обнаружено в некоторый момент наблюдения за анализируемым параметром, зара-

нее не известно.

2. Каждое значение случайной величины проявляется внутри интервала ее варьирования с совершенно определенной вероятностью.

## 2.2.Закон распределения

Закон распределения – соотношение, устанавливающее связь между возможными значениями случайной величины и соответствующими им вероятностями. Часто также используют термин «Распределение вероятности».

Закон распределения представляют либо функцией распределения, либо плотностью распределения.

**Функция распределения** отображает вероятность события, заключающегося в том, что случайная величина (например $X$) примет значение меньше, чем произвольное действительное число $x$ (т. е. вероятность события $X < x$):

$$F(x) = P(X < x). \quad (2.1)$$

В виде функции распределения можно отобразить закон распределения как непрерывной, так и дискретной случайной величины (рис. 2.1).

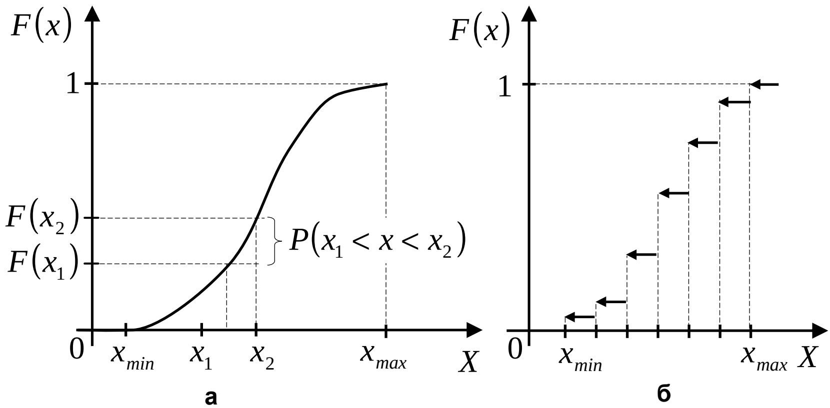


Рис. 2.1. Функции распределения непрерывной (а) и дискретной (б) случайных величин

Независимо от вида случайной величины функция распределения обладает следующими свойствами:

1. $F(x)$ есть неубывающая функция $x$ и если $x_1 < x_2$, то $F(x_1) < F(x_2)$. Разность двух ординат, соответствующих точкам $x_1$ и $x_2$, дает вероятность того, что значения случайной величины будут лежать в интервале между $x_1$ и $x_2$:
$$P(x_1 < X < x_2) = F(x_2) - F(x_1). \qquad (2.2)$$
2. Значения функции распределения при предельных значениях аргумента (т. е. соответствующей случайной величины) равны 0 и 1. Для генеральной совокупности это свойство записывают следующим образом:
$$F(x_{min}) = 0, F(x_{max}) = 1. \qquad (2.3)$$

Особенностью функции распределения дискретной случайной величины является то, что она есть разрывная ступенчатая функция, скачки которой происходят в точках, соответствующих возможным значениям случайной величины, и равны вероятностям (частотам) этих значений (рис. 2.1,б).

**Плотность распределения** (используют также термины *плотность распределения вероятности, плотность вероятности*) отображает вероятность события, состоящего в том, что произвольное значение $x$ случайной величины $X$ находится в некотором наперед заданном интервале $\{x_1, x_2\}$ (т. е. вероятность события $x_1 < x < x_2$):
$$f(x) = P(x_1 < X < x_2). \qquad (2.4)$$

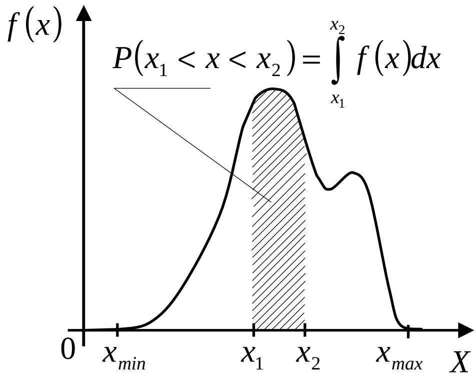

Рис. 2.2. Плотность распределения вероятности случайной величины

График плотности распределения некоторой случайной величины изображен на рис. 2.2.

Плотность распределения существует только для непрерывной случайной величины и если функция распределения данной случайной величины непрерывна и дифференцируема, то:

$$f(x) = F'(x). \tag{2.5}$$

По сравнению с функцией распределения описание распределения с помощью плотности вероятности более удобно и наглядно, так как позволяет отобразить его особенности на различных участках внутри интервала варьирования данной случайной величины.

Свойства плотности распределения:

Плотность распределения определяет случайную величину также полно, как и функция распределения и является неотрицательной функцией:

$$f(x) \geq 0. \tag{2.6}$$

Площадь, ограниченная числовой осью, кривой плотности распределения, а также прямыми $x = x_1$ и $x = x_2$ равна вероятности того, что случайная величина примет некоторое значение в рассматриваем интервале:

$$\int_{x_1}^{x_2} f(x)dx = P(x_1 < x < x_2). \tag{2.7}$$

Так как событие $x_{min} < x < x_{max}$ является достоверным, площадь под кривой плотности распределения равна 1:

$$\int_{x_{min}}^{x_{max}} f(x)dx = 1. \tag{2.8}$$

## 2.3.Нормальное распределение и его особенности

К настоящему времени обнаружен целый ряд законов распределения вероятности. Для дискретных случайных величин наиболее характерны распределение Пуассона и биноминальное. Для непрерывных случайных величин известны показательный и нормальный законы распределения, а также связанные с нормальным, - распределения Стьюдента, Фишера и хи-квадрат.

Нормальное распределение занимает особое положение, так как в соответствии с *центральной предельной теоремой* теории вероятности оно является предельным законом, к которому, при весьма часто встречающихся условиях, приближаются все другие распределения.

*Нормальное распределение* - это распределение *непрерывной* случайной величины, для которого характерна плотность распределения вида:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} exp\left[ -\frac{(x - M_x)^2}{2\sigma^2} \right], \qquad (2.9)$$

где $M_x$ – математическое ожидание (характеристика положения истинного значения случайной величины);

$\sigma$ - стандартное отклонение (характеристика вариации значений случайной величины).

Функция и плотность нормального распределения схематично изображены на рис. 2.3.

Нормальное распределение имеет следующие важные для практического использования свойства:

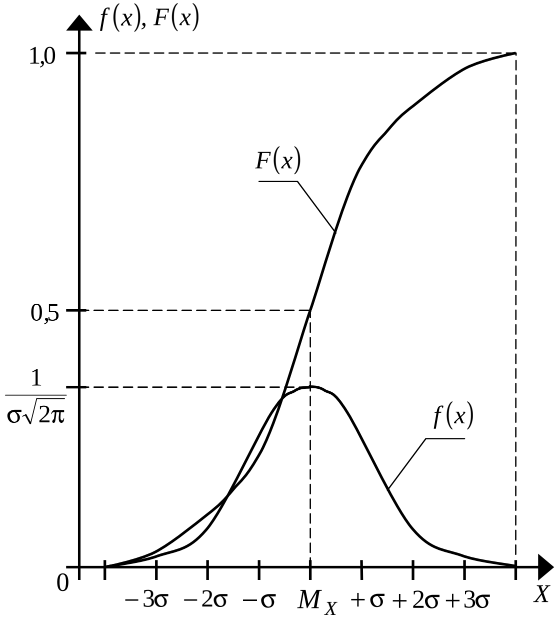


Рис. 2.3. Функция и плотность нормального распределения

1. Кривая плотности распределения симметрична относительно прямой $x = M_x$ и при этой абсциссе достигает максимума, который равен

$$f_{max} = \frac{1}{\sigma\sqrt{2\pi}} = \frac{0{,}3989}{\sigma}. \qquad (2.10)$$

2. Площадь под кривой, ограниченная ординатами $f(M_x - 3\sigma)$ и $f(M_x + 3\sigma)$ равна 0,9973 (т. е. практически 1). Это означает, что при нормальном распределении случайной величины вероятность проявления ее значений, отличающихся от математического ожидания более, чем на $3\sigma$, практически равна нулю.
3. Площадь под кривой, ограниченная ординатами $f(M_x - 2\sigma)$ и $f(M_x + 2\sigma)$равна 0,9544 (с достаточной для практики точностью 0,95). Это означает, что при нормальном распределении случайной величины вероятность проявления ее значений, отличающихся от математического ожидания более, чем на $2\sigma$, не превышает 0,05 (т. е. 5 %).
4. Площадь под кривой, ограниченная ординатами $f(M_x - \sigma)$ и $f(M_x + \sigma)$ равна 0,6826 (для практических целей можно принять 0,7). Это означает, что при нормальном распределении случайной величины вероятность проявления ее значений, отличающихся от математического ожидания более, чем на $\sigma$, равна 0,3 (т. е. 30 %).
5. Нормальное распределение обладает свойством линейности, которое формулируется следующим образом. Если *независимые* случайные величины $X_1$ и $X_2$ имеют нормальные распределения, то для произвольных чисел $\alpha$ и $\beta$ величина $Y = \alpha X_1 + \beta X_2$ также имеет нормальное распределение, причем из свойств математического ожидания и дисперсии следует

$$M_Y = \alpha M_{X1} + \beta M_{X2}; \qquad (2.11)$$

$$\sigma_Y = \sqrt{\alpha^2 \sigma_{X1}^2 + \beta \sigma_{X2}^2}. \qquad (2.12)$$

## 2.4.Стандартное нормальное распределение и функция Лапласа

Нормальное распределение зависит от двух параметров - математического ожидания $M_x$ и стандартного отклонения $\sigma$, что затрудняет ее представление в табличном (табулированном) виде. Поэтому было предложено использовать нормированную случайную величину:

$$Z = \frac{x - M_x}{\sigma},$$

для которой функция плотности нормального распределения (2.9) принимает вид:

$$f(x) = \varphi(z) = \frac{1}{\sqrt{2\pi}} exp\left(-\frac{z^2}{2}\right). \qquad (2.13)$$

График плотности стандартного нормального распределения приведены на рис. 2.5.

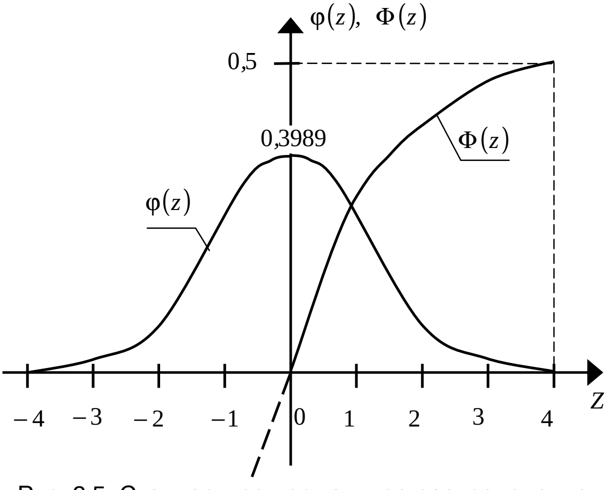


Рис. 2.5. Стандартное нормальное распределение: $\varphi(Z)$ – кривая Гаусса; $\Phi(Z)$ – Функция Лапласа

График плотности стандартного нормального распределения называют “*кривая вероятностей*”, “*кривая Гаусса* ”. Основным отличием кривой Гаусса от кривой ненормированного нормального распределения является то, что она фактически строится для $M(x)=0$ и $\sigma=1$ (рис. 2.5). Из указанной особенности следует ряд специфических свойств.

1. Плотность стандартного нормального распределения симметрична относительно оси ординат и имеет максимум, равный 0,3989.
2. Плотность стандартного нормального распределения является четной функцией:
$$\varphi(-z)=\varphi(z).$$
3. При $z=4$ плотность стандартного нормального распределения равна нулю
$$\varphi(4)=0{,}0001\approx 0.$$
Поэтому при ее табулировании указываются значения для нормированных значений случайной величины от 0 до 4.
4. Значение функции ненормированного нормального распределения равно значению функции стандартного нормального распределения.

Чтобы получить функцию стандартного нормального распределения необходимо выполнить интегрирование зависимости (2.13), однако результат не может быть выражен через элементарные функции. Поэтому используется функция Лапласа:

$$\Phi(z)=\frac{1}{\sqrt{2\pi}}\int_0^z e^{-\frac{u^2}{2}}du. \quad (2.14)$$

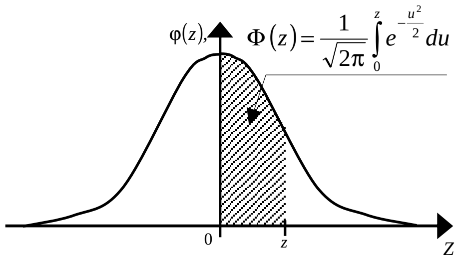

Рис.2.6. К определению Функции Лапласа

Смысл функции Лапласа иллюстрируется рис.2.6 - $\Phi(z)$ представляет собой значение вероятности, с которой случайная величина, обладающая стандартным нормальным распределением, принимает значение из интервала $\{0;\ z\}$. Основные свойства функции Лапласа:

1. Функция Лапласа является нечетной:
$$\Phi(-z) = -\Phi(z).$$

Ее график симметричен относительно начала координат (рис. 2.6) и $\Phi(0)=0$.

2. Функция Лапласа является монотонно возрастающей в пределах от $\Phi(-4) = -0,5$ до $\Phi(+4) = +0,5$.

С учетом указанных свойств переход от функции Лапласа к функции стандартного нормального распределения осуществляется следующим образом:

$$F(z,0,1)=0,5+\Phi(z).$$

## 2.5.Выборка и выборочные характеристики

Как уже отмечалось, случайная величина может принимать множество значений, которые, однако, соответствуют объективно существующему интервалу ее варьирования. Совокупность всех возможных значений случайной величины $\{\chi_j\}_N$ для всех возможных условий наблюдений называется генеральной совокупностью. Количество значений случайной величины в генеральной совокупности (объем генеральной совокупности) $N$ определяется характером случайной величины и для непрерывной величины $N \to \infty$.

По некоторым причинам, например в связи с ограниченностью продолжительности наблюдений, из генеральной совокупности удается обнаружить лишь $n < N$ значений случайной величины $\{x_i\}_n$. Совокупность ограниченного числа значений случайной величины, которые получены в результате наблюдения, называют выборкой, а количество $n$ значений в выборке - объемом выборки. Выборка является подмножеством генеральной совокупности:

$$\{x_i\}_n \in \{\chi_j\}_N .$$

Для решения прикладных задач разработаны и применяются специальные числовые характеристики распределения вероятности. Каждая такая характеристика имеет строго определенный смысл, который не изменяется от того, что рассматривает исследователь - генеральную совокупность или выборку из нее. Однако формулы для расчета числовых характеристик распределения в каждом из указанных случаев могут быть различными.

При использовании числовых характеристик случайной величины весьма важным является следующее обстоятельство. Изучая случайную величину, мы не можем охватить генеральную совокупность и поэтому, имеем в своем распоряжении лишь ограниченную выборку из нее. Таким образом, по выборке будут получены не истинные значения той или иной характеристики распределения изучаемой случайной величины, а лишь их приближенные значения (оценки).

В общем случае для оценивания некоторого параметра $\xi$ генеральной совокупности используется некоторая величина $\theta$, вычисляемая по результатам выборки. По смыслу $\xi$ и $\theta$ характеризуют одно и то же свойство случайной величины. Однако величина $\theta$ является *выборочной оценкой* действительного значения анализируемого параметра: $\theta \approx \xi$, в литературе ее часто называют *статистикой*. Статистика есть случайная величина, распределение которой отличается и от распределения случайной величины в генеральной совокупности, и от распределения оцениваемого параметра $\xi$. В связи с этим выборочные оценки должны отвечать ряду требований.

1. *Выборочная оценка должна быть **состоятельной***. Оценка $\theta$ является состоятельной, если с увеличением объема выборки $n$ вероятность ее значения стремится к вероятности значения оцениваемого параметра генеральной совокупности:
$$\lim_{n \to \infty} Prob\left\{\left|\theta - \xi\right| < \Delta\right\} = 1 \text{ или } \theta \xrightarrow{Prob} \xi .$$
2. *Выборочная оценка должна быть **несмещенной***. Оценка $\theta$ является несмещенной, если при любом объеме выборки ее математическое ожидание равно оцениваемому параметру:
$$M_{\theta} = \xi .$$
3. *Выборочная оценка должна быть **эффективной***. Из различных оценок одного и того же параметра генеральной совокупности наиболее эффективной является та, которая обладает наименьшей дисперсией:
$$\lim_{n \to \infty} D_{\theta} = 0 .$$

## 2.6.Контрольные вопросы

1. Как классифицируют параметры объекта исследования с позиций обработки и анализа числовой информации? Каков характер параметров объекта исследования?
2. Перечислите и кратко поясните наиболее распространенные задачи, решаемые обработкой и анализом числовой информации.
3. Назовите виды и закономерности случайной величины.
4. Функция и плотность распределения вероятности.
5. Нормальное распределение вероятности и его особенности.
6. Стандартное нормальное распределение вероятности и его особенности.
7. Функция Лапласа.
8. Генеральная совокупность и выборка.
9. Понятие о выборочных оценках характеристик случайной величины и требования к ним.