ЛЕКЦИЯ 4

Непрерывные случайные величины, их числовые характеристики. Равномерное распределение непрерывной случайной величины. Нормальное распределение. Понятие о центральной предельной теореме. Интегральная теорема Лапласа.

1.12. НЕПРЕРЫВНЫЕ СЛУЧАЙНЫЕ ВЕЛИЧИНЫ.

1.12.1. Геометрическая модель теории вероятностей.

Вероятность случайного события в классической модели находится в предположении, что множество исходов опыта конечно. На практике встречаются задачи, в которых множество исходов опыта бесконечно (континуально). Примером может служить число точек на отрезке прямой; число точек, принадлежащих области на плоскости, области в пространстве. Аналогичная ситуация возникает, когда в результате эксперимента фиксируется непрерывно меняющаяся величина, например, масса, температура, время. *Один из подходов* к описанию таких ситуаций – *геометрическая вероятность*.

Пусть внутри ограниченной области Ω на плоскости расположена некоторая область D. В область Ω наудачу бросается «случайная точка». Пусть вероятность попадания точки в любую часть области Ω будет пропорциональна площади этой части и не будет зависеть ни от ее расположения, ни от ее формы.

Тогда вероятность попадания точки в область D находится по формуле:

$$P(D) = \frac{S_D}{S_\Omega}, \tag{1.2}$$

где S_D – площадь области D, S_Ω – площадь области Ω.

При аналогичных предположениях можно утверждать следующее.

Если отрезок a составляет часть отрезка L и на отрезок L наудачу бросается «случайная точка», то вероятность попадания этой точки в отрезок a равна $P(a) = l_a / l_L$, где l_a – длина отрезка a, l_L – длина отрезка L.

Если внутри трехмерной области Ω расположена область G и в область Ω наудачу бросается «случайная точка», то вероятность попадания этой точки в область G равна $P(G) = V_G / V_\Omega$, где V_G – объем области G, V_Ω – объем области Ω.

Задача 1.5. Внутрь круга радиуса R наудачу брошена точка. Найти вероятность того, что точка окажется внутри вписанного в круг квадрата. Предполагается, что вероятность попадания точки в квадрат пропорциональна площади квадрата и не зависит от его расположения относительно круга.

Решение

Площадь круга $S_{круга} = \pi R^2$. Сторона вписанного в круг квадрата равна $R\sqrt{2}$. Тогда площадь квадрата $S_{кв} = 2R^2$. Вероятность попадания «случайной точки» в квадрат согласно формуле (1.2)

$$P = \frac{S_{кв}}{S_{круга}} = \frac{2R^2}{\pi R^2} = \frac{2}{\pi}.$$

Задача 1.6. Отрезок KB разделен точкой C в отношении 2:1. На отрезок наудачу брошена точка. Найти вероятность того, что эта точка окажется правее точки C.

Решение

Если длина отрезка CB равна одной единице длины, то длина всего отрезка равна трем единицам длины. Вероятность искомого события равна отношению длины отрезка CB к длине отрезка KB.

$$P = \frac{l_{CB}}{l_{KB}} = \frac{1}{3}.$$

Задача 1.7. В куб наудачу помещают точку. Найти вероятность того, что она попадет в пирамиду, вершина которой совпадает с центром верхнего основания куба, а основание пирамиды совпадает с основанием куба.

Решение

Сформулируем событие A = {точка, случайным образом размещенная в кубе, окажется внутри пирамиды}, тогда

$$p = P(A) = \frac{V_{\text{пир.}}}{V_{\text{куба}}} = \frac{\frac{1}{3} S_{осн.} h}{a^3} = \frac{\frac{1}{3} a^2 a}{a^3} = \frac{1}{3},$$

где обозначено: a – длина ребра куба; $S_{осн.}$ – площадь основания пирамиды; h – высота пирамиды.

1.12.2. Непрерывные случайные величины, их числовые характеристики.

Определение. Случайная величина X имеет *непрерывное распределение*, если её функция распределения: $F(x) = P(X < x)$ непрерывна на всей числовой оси.

Другими словами случайная величина X имеет *непрерывное распределение*, если она может принимать любые значения из некоторого конечного или бесконечного промежутка. Число возможных значений непрерывной случайной величины бесконечно. Можно дать и другое определение *непрерывной случайной величины*: случайная величина называется *непрерывной*, если математическое ожидание любой функции $g(X)$ можно записать в виде:

$$M(g(X)) = \int_{-\infty}^{+\infty} g(x) \cdot \varphi(x) dx. \qquad (1.41)$$

Под “любой” функцией $g(x)$ имеется ввиду такая, для которой интеграл (1.41) существует и сходится абсолютно.

Функция $\varphi(x)$ называется *плотностью распределения вероятностей* непрерывной случайной величины X и обладает следующими свойствами:

1. Вероятность попадания величины X в произвольный интервал A на оси $0x$ равна

$$P(X \in A) = \int_A \varphi(x) dx, \qquad (1.42)$$

т. е. интегралу по A от функции плотности.

Таким образом, функция плотности $\varphi(x)$ полностью характеризует распределение случайной величины X.

2. В частности, для интервала (x_1, x_2), получаем:

$$P(X \in (x_1,\ x_2)) = \int_{x_1}^{x_2} \varphi(x)dx\,. \tag{1.43}$$

3. Так как вероятность неотрицательна, то из (1.42) следует, что $\varphi(x) \geq 0$ для любого x.

4. Вероятность достоверного события равна 1, поэтому

$$\int_{-\infty}^{+\infty} \varphi(x)dx = P(X \in (-\infty,+\infty)) = 1\,. \tag{1.44}$$

Последнее равенство называется *условием нормировки* функции плотности.

График функции плотности распределения $\varphi(x)$ называется *кривой распределения* (рис. 1.10).

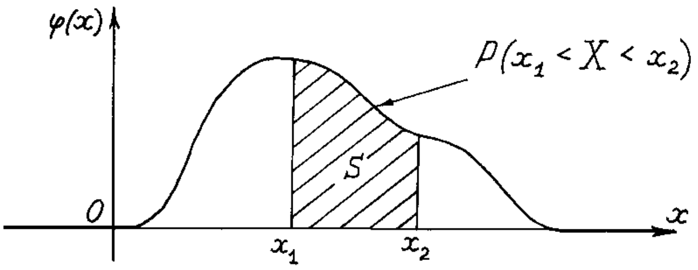

Рис. 1.10. График плотности распределения $\varphi(x)$ (кривая распределения).

Вероятность попадания случайной величины X в интервал (x_1, x_2) численно равна площади соответствующей криволинейной трапеции. Из условия нормировки следует, что площадь области, ограниченной сверху кривой распределения, а снизу – осью $0x$, равна 1.

Заметим, что вероятность попадания непрерывной случайной величины в любую отдельную точку равна нулю.

Функцией распределения случайной величины X является функция $F(x)$, равная вероятности события ($X < x$), т.е. вероятности того, что случайная величина X примет значение, меньшее значения аргумента x (см. п. 1.8, формула (1.17)).

Для непрерывной случайной величины функция распределения равна

$$F(x) = P(X < x) = \int_{-\infty}^{x} \varphi(t)dt \tag{1.45}$$

и обладает следующими свойствами:

1. $0 \le F(x) \le 1$ для всех x;

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

3. $F(x)$ – неубывающая функция на всей оси;

4. $F(x)$ – непрерывная функция, в точках непрерывности $\varphi(x)$ она имеет производную:

$$F'(x) = \varphi(x). \tag{1.46}$$

Вероятность попадания непрерывной случайной величины X в произвольный интервал (x_1, x_2) можно вычислить с помощью функции распределения следующим образом:

$$P(X \in (x_1, x_2)) = P(X < x_2) - P(X < x_1) = F(x_2) - F(x_1). \tag{1.47}$$

Поэтому функция распределения $F(x)$ так же, как и функция плотности распределения $\varphi(x)$, полностью характеризует распределение вероятностей случайной величины X и даже более удобна для расчетов вероятностей, так как не требует интегрирования. Для многих распределений, встречающихся в статистике, функции распределения табулированы.

В задачах статистики часто бывает нужно найти такое значение x по заданной вероятности $\mathcal{P}$, что

$$\mathcal{P} = P(X < x) = F(x). \tag{1.48}$$

Данное уравнение может иметь, вообще говоря, множество решений. Но для большинства распределений, встречающихся в статистике, функция плотности распределения $\varphi(x)$ строго положительна для всех X из некоторого интервала и равна нулю вне этого интервала. Поэтому внутри этого интервала функция $F(x)$ строго монотонно возрастает.

В этих случаях решение уравнения (1.43) существует и единственно для всех $\mathcal{P} \in (0;\ 1)$. Оно называется *квантилью распределения* и обозначается $x_{\mathcal{P}}$ (рис. 1.11). Обычно для квантилей распределения также составляют таблицу, которая представляет собой таблицу значений функции, обратной функции распределения.

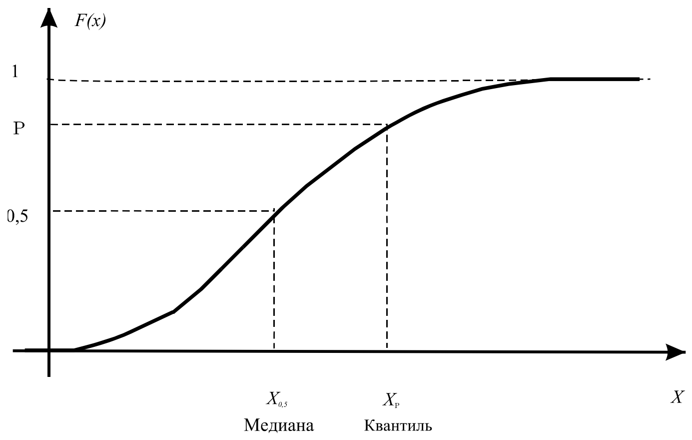

Рис. 1.11. График функции распределения, квантиль и медиана случайной величины X.

Некоторые квантили имеют специальное название. Так, *медианой* непрерывной случайной величины X называется действительное число m_X, удовлетворяющее условию:

$$P(X < m_X) = P(X > m_X) = 0{,}5, \tag{1.49}$$

т. е. решение уравнения $F(x) = 0{,}5$.

Математическое ожидание и дисперсия непрерывной случайной величины X находят по формулам, которые следуют из выражения (1.41) :

$$M(X) = \int_{-\infty}^{+\infty} x \cdot \varphi(x)dx. \tag{1.50}$$

$$D(X) = M(X - M(X))^2 = \int_{-\infty}^{+\infty} (x - M(X))^2 \cdot \varphi(x)dx$$

Дисперсию проще рассчитывать по следующей формуле:

$$D(X) = M(X^2) - (M(X))^2 = \int_{-\infty}^{+\infty} x^2 \varphi(x)dx - (M(X))^2. \qquad (1.51)$$

1.12.3. Равномерное распределение непрерывной случайной величины

К часто встречающимся непрерывным распределениям вероятностей относится *равномерное распределение*. Говорят, что случайная величина X распределена равномерно в интервале (a, b), если все ее возможные значения сосредоточены в этом интервале и плотность вероятности в нем постоянна:

$$\varphi(x) = \begin{cases} c, & x \in (a,b), \\ 0, & x \notin (a,b). \end{cases} \qquad (1.52)$$

Из условия нормировки находим, что $c = 1/(b-a)$. Тогда вероятность попадания случайной величины X в интервал (x_1, x_2), лежащий внутри интервала (a, b), равна, согласно формуле (1.39),

$$P(X \in (x_1, x_2)) = c(x_2 - x_1) = \frac{x_2 - x_1}{b - a}.$$

Математическое ожидание, дисперсия и среднее квадратическое отклонение случайной величины, имеющей равномерное распределение, равны, соответственно:

$$M(X) = \frac{a+b}{2}; \qquad D(X) = \frac{(b-a)^2}{12}; \qquad \sigma(X) = \frac{b-a}{2\sqrt{3}}.$$

1.12.4. Нормальное распределение.

К числу наиболее важных для приложений относится *нормальное распределение.*

Случайная величина U имеет *стандартное нормальное распределение*, если ее плотность задается формулой

$$\varphi(u) = \frac{1}{\sqrt{2\pi}} e^{-u^2/2}, \quad (-\infty < u < +\infty). \tag{1.53}$$

Кривая стандартного нормального распределения представлена на рис. 1.14. Это распределение симметрично. Его параметры: $M(U) = 0; \quad D(U) = M(U^2) = 1.$

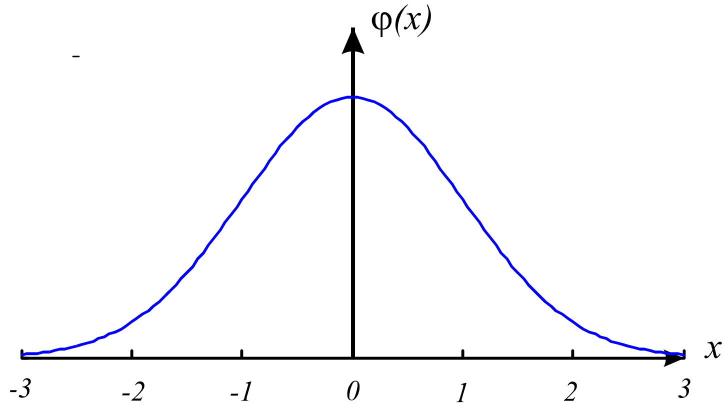

Рис. 1.14. Кривая стандартного нормального распределения.

Расчеты вероятностей для стандартного нормального распределения выполняются с помощью интеграла вероятностей:

$$\Phi(t) = \int_0^t \varphi(u)du = \frac{1}{\sqrt{2\pi}} \int_0^t e^{-u^2/2} du \quad (t > 0), \quad \Phi(-t) = -\Phi(t). \tag{1.54}$$

Интеграл вероятностей $\Phi(t)$ связан с функцией распределения $F(t) = \int_{-\infty}^{t} \varphi(u)du$ формулой $F(t) = 0{,}5 + \Phi(t)$. Поэтому вероятность попадания случайной величины, имеющей стандартное нормальное распределение, в интервал (t_1, t_2) рассчитывают по формуле:

$$P(t_1 < U < t_2) = F(t_2) - F(t_1) = \Phi(t_2) - \Phi(t_1), \tag{1.55}$$

а для симметричного интервала $(-t, t)$, пользуются формулой :

$$P(|U| < t) = 2P(0 < U < t) = 2\Phi(t).$$

Значения интеграла вероятностей представлены в таблице П1 приложения.

Случайная величина X имеет *нормальное распределение с параметрами* a, σ (будем обозначать $X \sim N(a; \sigma)$), если она может быть представлена в виде $X = a + \sigma U$, где U имеет стандартное нормальное распределение, $\sigma > 0$. Для такого распределения $M(X) = a$, $D(X) = \sigma^2$, $\sigma(X) = \sigma$, так что параметры распределения представляют собой математическое ожидание и среднее квадратическое отклонение величины X.

Функция плотности нормально распределенной случайной величины имеет вид

$$\phi(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-a)^2/(2\sigma^2)}, \qquad (-\infty < x < +\infty)\,. \tag{1.56}$$

Влияние параметров a и σ на график функции плотности видно из рис. 1.15. На этом рисунке приведены кривые нормального распределения при значениях $\sigma = 1$, $\sigma < 1$ и $\sigma > 1$. График симметричен относительно прямой $x = a$, так что значение a является центром распределения. При $\sigma = 1$ это та же кривая, что и на рис. 1.14, но смещенная вдоль оси абсцисс на a единиц. С уменьшением параметра σ кривая функции плотности сжимается вдоль оси абсцисс и вытягивается вдоль вертикальной прямой $x = a$, что приводит к увеличению вероятности попадания случайной величины X в любую фиксированную окрестность точки $x = a$. С увеличением параметра σ, наоборот, кривая функции плотности растягивается вдоль оси абсцисс и сжимается вдоль прямой $x = a$. Вероятности попадания случайной величины X в любую фиксированную окрестность точки $x = a$ уменьшается.

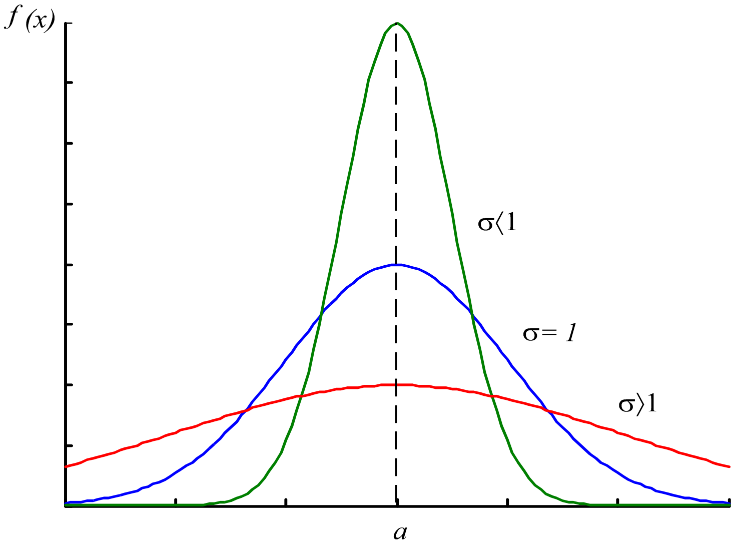

Рис. 1.15. Кривые нормального распределения.

Расчет вероятности попадания в произвольный интервал (x_1, x_2) для нормально распределенной величины X проводится с помощью интеграла вероятностей (1.54):

$$P(x_1 < X < x_2) = \Phi\left(\frac{x_2 - a}{\sigma}\right) - \Phi\left(\frac{x_1 - a}{\sigma}\right). \qquad (1.57)$$

В частности, для симметричного интеграла

$$P\left(\left|X - a\right| < t\sigma\right) = 2\Phi(t). \qquad (1.58)$$

Случайная величина с нормальным распределением обладает следующими важными свойствами:

1. Сумма двух величин X_1 и X_2, каждая из которых имеет нормальное распределение с произвольными параметрами, также имеет нормальное распределение.

2. При умножении случайной величины, имеющей нормальное распределение, на произвольное число (константу) получившаяся величина также имеет нормальное распределение.

Из этих свойств следует важный вывод. Если есть конечное число случайных величин $X_1, X_2, ..., X_n$, каждая из которых имеет нормальное распределение с произвольными параметрами, то их линейная комбинация

$$c_1X_1 + c_2X_2 + \ldots + c_nX_n$$

также имеет нормальное распределение при любых коэффициентах $c_1, c_2, \ldots, c_n$.

1.13. НЕПРЕРЫВНЫЕ ДВУМЕРНЫЕ СЛУЧАЙНЫЕ ВЕЛИЧИНЫ.

Часто приходится решать задачи, в которых рассматриваются события, описываемые не одной, а несколькими – в частности, двумя случайными величинами. Так если станок-автомат штампует цилиндрические валики, то диаметр валика D и его высота H образуют систему двух случайных величин

Двумерной случайной величиной называют систему из двух случайных величин (X, Y), для которой определена вероятность $P\,[(X < x), (Y < y)]$ совместного выполнения неравенств $X < x,\ Y < y$, где x и у –любые действительные числа.

Функция двух переменных

$$F(x, y) = P\,[(X < x), (Y < y)] \tag{1.59}$$

определенная для любых x и y, называется *функцией распределения* системы двух случайных величин (X, Y)

Будем рассматривать X и Y как декартовы координаты точки на плоскости. Точка $M\,(X, Y)$ может занимать то или иное положение на плоскости OXY. Тогда функция распределения даст вероятность того, что случайная точка $M\,(X, Y)$ попадает в область ω, изображенную на рис. 1.25.

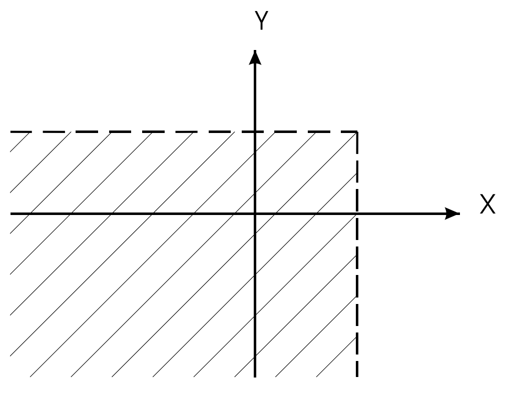

Рис. 1.25

Двумерная величина (*X*, *Y*) называется *непрерывной,* если существует такая непрерывная неотрицательная функция $\varphi(x, y)$, двух переменных, что вероятность того, что точка *M*(*X*, *Y*) содержится в некоторой области ω плоскости *OXY* равна двойному интегралу от функции $\varphi(x, y)$ по области ω:

$$P(M(X,Y) \in \omega) = \iint_{\omega} \varphi(x, y)dxdy. \tag{1.60}$$

Функция $\varphi(x, y)$ называется *плотностью совместного распределения вероятностей* системы двух величин *X* и *Y*. Отсюда, в частности, следует, что если область ω имеет вид, изображенный на рис. 1.25, то функцию распределения системы случайных величин можно записать следующим образом:

$$F(x, y) = P\,[(X < x), (Y < y)] = \iint_{\omega} \varphi(x, y)dxdy = \int_{-\infty}^{x} dx \int_{-\infty}^{y} \varphi(x, y)dy. \tag{1.61}$$

Равенство

$$\int_{-\infty}^{+\infty} dx \int_{-\infty}^{+\infty} \varphi(x, y)dy = 1 \tag{1.62}$$

является *условием нормировки* функции плотности совместного распределения вероятностей.

Непрерывные случайные величины *X* и *Y* называются *независимыми,* если

$$\varphi(x, y) = \varphi_1(x)\ \varphi_2(y)$$

где $\varphi_1(x)$ и $\varphi_2(y)$ – соответственно плотности распределения вероятностей случайных величин *X* и *Y*. В этом случае

$$F(x, y) = P\,[(X < x), (Y < y)] = \int_{-\infty}^{x} dx \int_{-\infty}^{y} \varphi_1(x)\varphi_2(y)dy =$$

$$= \int_{-\infty}^{x} \varphi_1(x)dx \int_{-\infty}^{y} \varphi_2(y)dy = F_1(x) \cdot F_2(y),$$

где $F_1(x)$ и $F_2(y)$ – соответственно функции распределения величин *X* и *Y* [см. формулу (1.45)].

Зная функцию распределения $F(x, y)$ двумерной случайной величины (X, Y), легко найти как функцию распределения, так и плотность распределения каждой из случайных величин X и Y в отдельности.

Действительно, пусть $F_1(x)$ – функция распределения случайной величины X. Тогда $F_1(x) = P(X < x)$. Так как в этом случае Y может принимать любое значение, то ясно, что

$$P(X < x) = P[(X < x), (-\infty < Y < +\infty)].$$

Следовательно, по формуле (1.61) имеем

$$F_1(x) = P(X < x) = P[(X < x), (-\infty < Y < +\infty)] = \int_{-\infty}^{x} dx \int_{-\infty}^{+\infty} \varphi(x, y) dy.$$

Дифференцируя последнее равенство по x, согласно правилу дифференцирования интеграла по переменной верхней границе получим

$$\varphi_1(x) = F_1'(x) = \int_{-\infty}^{+\infty} \varphi(x, y) dy. \qquad (1.63)$$

Аналогичным образом получаем

$$F_2(y) = P(Y < y) = \int_{-\infty}^{y} dy \int_{-\infty}^{+\infty} \varphi(x, y) dx.$$

и, следовательно,

$$\varphi_2(y) = \int_{-\infty}^{+\infty} \varphi(x, y) dx. \qquad (1.64)$$

Таким образом, чтобы получить плотность распределения одной из составляющих двумерной случайной величины, надо проинтегрировать в границах от $-\infty$ до $+\infty$ плотность распределения системы $\varphi(x, y)$ по переменной, соответствующей другой случайной величине.

Для системы случайных величин (X, Y) математическое ожидание любой функции $f(X, Y)$ вычисляется по формуле (аналог формулы (1.41)):

$$M[f(X,Y)] = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} f(x,y)\varphi(x,y)dxdy;. \qquad (1.65)$$

1.14. ПОНЯТИЕ О ЦЕНТРАЛЬНОЙ ПРЕДЕЛЬНОЙ ТЕОРЕМЕ. ИНТЕГРАЛЬНАЯ ТЕОРЕМА ЛАПЛАСА.

Часто приходится иметь дело с такими случайными величинами, которые являются суммами большого числа независимых случайных величин. Оказывается, что при некоторых весьма общих условиях закон распределения суммы этих величин чрезвычайно близок к нормальному закону распределения. Установление таких условий и составляет содержание *центральной предельной теоремы*, которая была доказана русским математиком А.М.Ляпуновым в 1900 году. Мы не будем формулировать полностью теорему Ляпунова, формулировка весьма сложна и требует значительных пояснений. Остановимся лишь на ряде частных случаях и общем смысле теоремы.

Если слагаемые суммы независимы, имеют произвольное, но одинаковое распределение, то теорема справедлива, т.е. распределение суммы достаточно большего числа слагаемых близко к нормальному. Причем, чем больше слагаемых в сумме, тем распределение ближе к нормальному (говорят, что оно имеет *асимптотически нормальное распределение*).

Теорема остается справедливой и в том случае, когда слагаемые суммы независимы, имеют произвольные и различные между собой распределения. Но в этом случае распределения слагаемых должны удовлетворять ряду условий. Смысл этих условий грубо можно описать так: вклад каждой случайной величины в рассеяние суммы должен быть мал (сильно не выделяться по сравнения с вкладом остальных слагаемых).

Частным случаем центральной предельной теоремы является *интегральная теорема Лапласа.*

Теорема. Пусть производится n независимых опытов, в каждом из которых вероятность наступления события A одна и та же и равна p ($0 < p < 1$). Пусть m – число появлений события A в n опытах. Тогда случайная величина $X = m$ имеет биномиальное распределение (см. п. 1.8). Но при достаточно больших n распределение случайной величины X близко к нормальному с параметрами $a = M(X) = np$, $\sigma = \sqrt{D(X)} = \sqrt{npq}$ ($q = 1 - p$).

Из интегральной теоремы Лапласа следует, что при достаточно больших n вероятность того, что случайная величина X, т.е. число наступлений события A в n опытах, принадлежит интервалу (x_1, x_2) можно приближенно вычислить по формуле (1.57):

$$P(x_1 < X < x_2) \approx \Phi\left(\frac{x_2 - np}{\sqrt{npq}}\right) - \Phi\left(\frac{x_1 - np}{\sqrt{npq}}\right), \tag{1.66}$$

где $\Phi(x)$ – интеграл вероятностей. Условие применения: n велико, а p и q – не очень малы, так что $npq \geq 20$.

Запишем еще два следствия из теоремы Лапласа.

1) Вероятность отклонения числа m наступления события A от математического ожидания np определяется по формуле

$$P(|m - np| \leq \delta) \approx 2\Phi\left(\frac{\delta}{\sqrt{npq}}\right) \tag{1.67}$$

2) Вероятность отклонения относительной частоты события $\frac{m}{n}$ от его вероятности p находится по формуле

$$P\left(\left|\frac{m}{n} - p\right| \leq \varepsilon\right) \approx 2\Phi\left(\frac{\varepsilon\sqrt{n}}{\sqrt{pq}}\right) \tag{1.68}$$