UDC 519.2

One Criteria of Consent of Normal Distribution Law

Serezha N. Sandryan

North Caucasus State Technical University, Russia
2 Kulakov Pr., Stavropol 355029
PhD (Physical and Mathematical Sciences), associate professor
E-mail: sandrun@yandex.ru

Abstract. According to the Central limit theorem, normal probability distribution law is most often found in random phenomena. In this work the linear criteria of consent is developed for the verification of statistical hypotheses of the normal distribution law of the statistical population.

Keywords: Normal probability distribution law; Central limit theorem; criteria of consent; statistical hypotheses; method of least squares; linear regression; correlation; statistical population; Laplace function.

Введение. Для многих реальных задач научного, производственного и экономического характера нормальный закон распределения вероятностей рассматривается как вероятностная модель для описания случайных величин. В самом деле хорошо известно, что при достаточных общих реальных ограничениях нормальный закон распределения согласно центральной предельной теоремы появляется как предельный закон для описания случайных величин [1].

Поэтому часто выбор функции распределения в практических задачах гипотетически принимается как нормальный закон распределения. Тогда вполне естественна постановка следующей задачи.

Постановка задачи.

По результатам экспериментальных данных

$$x_1, x_2, ..., x_n \tag{1}$$

установить, согласуются ли данные наблюдения с выдвинутой нулевой гипотезой:

$$H_0 : F(x) = N(x, a, \sigma)$$

где $\mathrm{F}(\mathrm{x})$ - функция распределения генеральной совокупности X;

$N(x, a, \sigma)$ - нормальный закон распределения с параметрами a и σ.

Один из наиболее широко употребимых критериев согласия является критерием хи-квадрат χ^2 или критерий χ^2 - Пирсона [2,3].

В работе автором предлагается линейная критерия согласия для принятия гипотезы о нормальности распределения генеральной совокупности:

$$H_0 : F(x) = N(x, a, \sigma)$$

$$H_a : F(x) \neq N(x, a, \sigma)$$

Известно, что случайная величина ξ имеет нормальный (Гауссовский) закон распределения с параметрами a и σ, если её функция распределения $F(x) = P(\xi < x)$ задается формулой [3]

$$F(x) = \Phi\left(\frac{x-a}{\sigma}\right), \tag{2}$$

где $\Phi(z) = \frac{1}{\sqrt{2\pi}} \int_{\infty}^{z} e^{-\frac{x^2}{2}} dx$ - функция Лапласа.

Функция Лапласа – табулированная функция[3].

Для проверки гипотезы о нормальном законе распределения выборки (1) воспользуемся законом больших чисел [1 ,4]:выборочная функция $F_n(x) = \frac{\nu(X < x)}{n}$ при больших объемах выборки $n(n \to \infty)$ равномерно близка к теоретической функции распределения

$$\forall x: \lim_{n \to \infty} F_n(x) = F(x).$$

Теперь в формуле (2) $F(x)$ заменяем на выборочную функцию $F_n(x)$ и рассмотрим уравнение

$$F_n(x) = \Phi(\frac{x-a}{\sigma}). \tag{3}$$

Учитывая монотонность функции Лапласа, применяя обратную функцию Лапласа Φ^{-1} обе части (3),тогда для любого x получаем

$$\Phi^{-1}(F_n(x)) = \frac{x-a}{\sigma}. \tag{4}$$

Между функциями Φ и Φ^{-1} можно установить табличную связь[3].

Введем обозначение $z(x) = \Phi^{-1}(F_n(x))$. При верности нулевой гипотезы H_0, согласно соотношению (4) функция $z(x)$ является линейной функцией от переменной x и наоборот. Таким образом, получается что, существование линейной зависимости между x и $z(x)$ равносильно верности нулевой гипотезы. В этом и заключена суть построения линейной критерии согласия нормального закона распределения. Далее, для установления этой линейной зависимости по выборочной функции распределения $F_n(x)$ построим обратную функцию $\Phi^{-1}(F_n(x))$ и находим последовательность точек $(x, z(x))$. Далее с помощью метода наименьшего квадрата[2] и коэффициента корреляции выявим линейную связь и составим уравнение регрессии $z(x)$ на x. Уравнение линейной регрессии составляется для нахождения оценки параметров нормального закона.

Пусть уравнение линейной регрессии имеет вид $z(x) = \alpha x + \beta$, тогда из этого уравнения и соотношения (4) заодно при верности нулевой гипотезы можно определить оценки для параметров нормального распределения по формулам :

$$\sigma = \frac{1}{\alpha}, a = -\frac{\beta}{\alpha} \quad . \tag{5}$$

Для выявления существования линейной зависимости между координатами точек $(x, z(x))$ для удобства вычисления и построения выборочной функции распределения перейдем от выборки (1) к вариационному ряду $x_{(1)}, x_{(2)}, \ldots, x_{(n)}$. Не усложняя вычисления, будем предполагать, что объем выборки и число вариантов совпадают.

Тогда $F_n(x)$ – кусочно-постоянная функция, которая в каждой из точек $x_{(i)}$ совершает скачок равный $\frac{1}{n}$ (рис. 1).

Правилом обратной функции построим обратную по отношению к $y = F_n(x)$ функцию $x = \Phi^{-1}(y)$ – обратная функция Лапласа. Для этого предварительно построим график функции $y = F_{n0}(x)$:

дополним график функции $y = F_n(x)$ вертикальными отрезками в точках разрыва $x_{(i)}$ до непрерывной линии $y = F_{n0}(x)$.

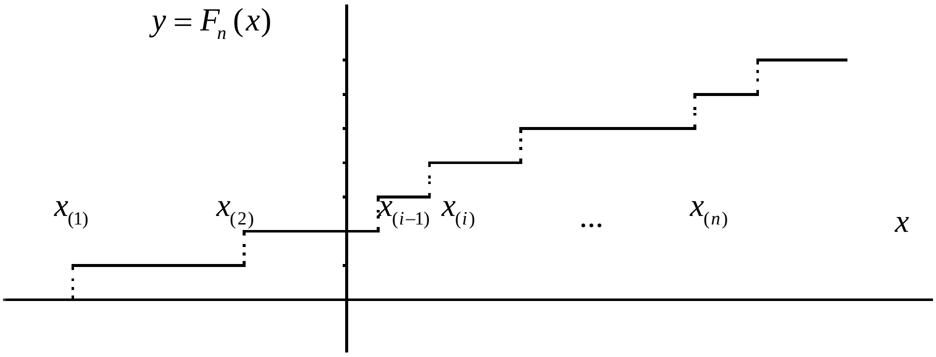

Рис. 1. график $F_n(x)$

Теперь, основываясь на непрерывной линии $y = F_{n0}(x)$ на интервале [0,1], построим однозначную обратную функции Лапласа $\Phi^{-1}(y)$ следующим образом.

$$\Phi^{-1}(y) = \begin{cases} F_{n0}^{-1}(y), & \text{в точках непрерывности } F_{n0}^{-1}(x), \\ F_{n0}^{-1}(y+0), & \text{в точках разрыва } x_{(i)}. \end{cases}$$

Таким образом, построенная функция соответствует обратной функции Лапласа $\Phi^{-1}(y)$ на интервале [0,1] и имеет следующий вид (рис. 2).

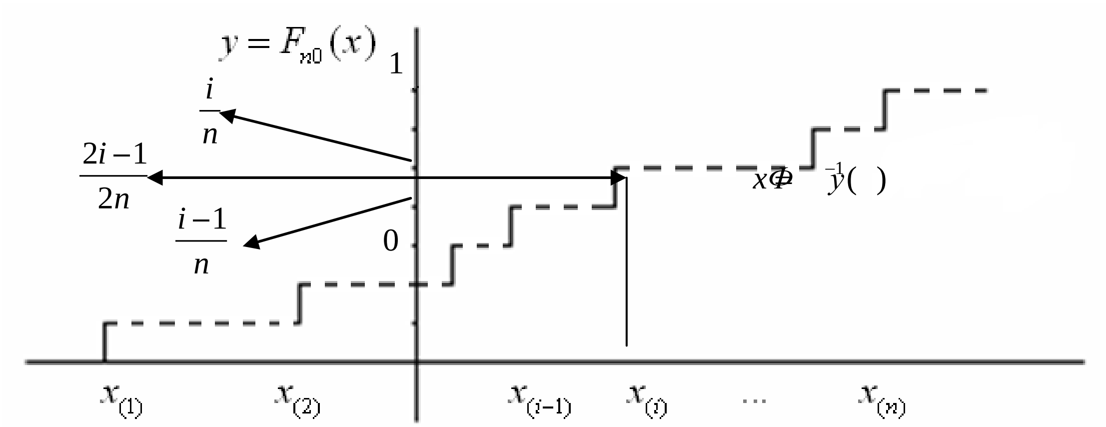

Рис. 2. график обратно функции Лапласа

Рассмотрим *i*-й интервал $\left[\frac{i-1}{n}, \frac{i}{n}\right]$ на оси ординат обратной функции Лапласа $\Phi^{-1}(y)$. Обозначим через y_i^c точкой середины *i*-го интервала, координата которого равна $y_i^c = \frac{2i-1}{2n}$. Тогда обратная функция Лапласа $\Phi^{-1}(y)$ точке y_i^c сопоставляет точку разрыва $x_{(i)}$ выборочной функции $F_n(x)$
(см . рис.2). Таким образом, получаем следующее функциональное равенство

$$\Phi^{-1}\left(\frac{2i-1}{2n}\right) = x_{(i)} \quad \text{для } i = 1, 2, \ldots, n. \tag{6}$$

Положим $z_i = \Phi^{-1}\left(\frac{2i-1}{2n}\right)$.

Теперь для принятия нулевой гипотезы H_0 рассмотрим последовательность точек $(z_i, x_{(i)})$ и для удобства расчета представим их в виде таблицы

z_i	z_1	z_2	$\cdots$	z_n
$x_{(i)}$	$x_{(1)}$	$x_{(2)}$	$\cdots$	$x_{(n)}$

. Для принятия линейной связи как показатель линейности выбираем коэффициент корреляции Пирсона [2]

$$r = \frac{n\sum_{i=1}^{n} x_{(i)} z_i - \sum_{i=1}^{n} x_{(i)} \sum_{i=1}^{n} z_i}{\left(n\sum_{i=1}^{n} x_{(i)}^2 - \left(\sum_{i=1}^{n} x_{(i)}\right)^2\right)\left(n\sum_{i=1}^{n} z_i^2 - \left(\sum_{i=1}^{n} z_i\right)^2\right)}, \; |r| \le 1.$$

Теперь построим статистическую гипотезу о наличии линейной связи между переменными z_i и $x_{(i)}$:

H_0: между переменными z_i и $x_{(i)}$ отсутствует линейная связь.

H_a: между переменными z_i и $x_{(i)}$ есть линейная связь.

Задается доверительная вероятность γ и проводится двусторонняя проверка $\alpha = \frac{1-\gamma}{2}$.

По таблице t-распределения Стьюдента[3] находим граничные точки $\pm t_{\alpha;n-2}$.

Согласно статистике

$$t = \sqrt{r^2(n-2)/(1-r^2)}$$

принимается или отвергается гипотеза на уровне значимости $(1-\gamma)$.

Если принимается нулевая гипотеза H_0, то связь между координатами z_i и $x_{(i)}$ задается уравнением линейной регрессии, z_i на $x_{(i)}$, которое имеет вид [2]

$$z = \alpha x + \beta,$$

где $$\alpha = \frac{n\sum_{i=1}^{n} x_{(i)} z_i - \sum_{i=1}^{n} x_{(i)} \sum_{i=1}^{n} z_i}{n\sum_{i=1}^{n} x_{(i)}^2 - \left(\sum_{i=1}^{n} x_{(i)}\right)^2}$$ и $$\beta = \frac{1}{n}\left(\sum_{i=1}^{n} z_i - \alpha\sum_{i=1}^{n} x_{(i)}\right).$$

Для упрощения вычисления последовательности

$$z_i = \Phi^{-1}\left(\frac{2i-1}{2n}\right), \quad i = 1, 2, \ldots, n. \tag{7}$$

установим одно важное свойство обратной функции Лапласа.

Лемма. Для любого значения x справедливо тождество

$$\Phi^{-1}(1-x) = -\Phi^{-1}(x).$$

Доказательство леммы в работе не приводится.

В силу утверждения леммы получаем:

$$z_i = \Phi^{-1}\left(\frac{2n-i}{2n}\right) = \Phi^{-1}\left(1-\frac{i}{2n}\right) = -\Phi^{-1}\left(\frac{i}{2n}\right). \qquad (8)$$

При вычислении значений числовой последовательности (7) на основе соотношения (8) объем вычислительной работы сокращается наполовину.

Для практического применения линейной критерии согласия для определенных значений объема выборки n целесообразно построить специальную таблицу. Далее продолжить работу по структуре статьи.

Примечания:

1. Гнеденко Б.В. Курс теории вероятностей. Учебник для университетов. 6-е изд. перераб. и доп. М.: Наука, 1988. 448 с.
2. Четыркин Е.М., Калихман И.Л. Вероятность и статистика. М.: финансы и статистика, 1982. 319 с., ил.
3. Вентцель Е.С. Теория вероятностей. Учебник для вузов. 7-е изд. стер. М.: Высш. шк., 2001. 575 с.: ил.
4. Феллер В. Введение в теорию вероятностей и её приложения. Т. 2. М.: Мир, 1984.

УДК 519.2

Одна критерия согласия нормального закона

Сережа Николаевич Сандрян

Северо-Кавказский Федеральный Университет, Россия
355029, г. Ставрополь., пр. Кулакова, 2.
Кандидат физико-математических наук, доцент
E-mail: sandrun@yandex.ru

Аннотация. Согласно центральной предельной теореме, нормальный закон распределения вероятностей наиболее часто встречается в случайных явлениях. В работе разработана линейная критерия согласия для проверки статистических гипотез нормального закона распределения генеральной совокупности.

Ключевые слова: Нормальный закон распределения; центральная предельная теорема; критерия согласия; статистическая гипотеза; метод наименьших квадратов; линейная регрессия; корреляция; генеральная совокупность; функция Лапласа.