УДК 621.317.08

*И. П. ЗАХАРОВ, канд. техн. наук, Н. В. ШТЕФАН, Г. Г. САФАРЯН, М. П. СЕРГИЕНКО*

# ИДЕНТИФИКАЦИЯ ФОРМЫ ЗАКОНА РАСПРЕДЕЛЕНИЯ ПОГРЕШНОСТЕЙ ПО ВЫБОРКАМ МАЛОГО ОБЪЕМА

При статистической обработке данных важной задачей является определение формы закона распределения погрешностей.

Для решения задачи определения вида закона распределения рекомендуются статистические методики, требующие построения гистограмм или полигонов распределения с последующей их проверкой по какому-либо из так называемых критериев согласия. Разработан целый ряд таких критериев, применение каждого из которых обусловлено определенными условиями [1, 2]. Одним из наиболее важных условий, ограничивающих применения того или иного критерия согласия, является объем выборки. На практике же получение выборок большого объема не всегда экономически обосновано, приводит к значительным затратам времени и средств, приводит к износу оборудования.

При малом объеме выборки для идентификации нормальных распределений применяются критерий W [2] и составной критерий [3].

В работе [4] проведено исследование применения составного критерия для распределений, отличных от нормального, и рассчитаны критические области параметра этого критерия для наиболее распространенных распределений. Показано, что эти области для различных распределений перекрываются и применение данного критерия может привести к ошибке.

Однако для малого числа наблюдений есть возможность построить закон распределения. Для этого применяются графические методы [5].

Порядок построения эмпирической интегральной функции состоит в следующем. Исходными данными располагаются в порядке неубывания $x_1 \le x_2 \le ... \le x_n$, тогда оценку $\hat{F}(x_i)$ находят по формуле

$$\hat{F}(x_i) = \frac{i}{n+1} (i = 1,2,...,n). \quad (1)$$

При проведении линии по нанесенным точкам (полученным экспериментальным путем) графика эмпирической функции распределения получают графическую оценку теоретической функции распределения $F(x)$ (рис. 1).

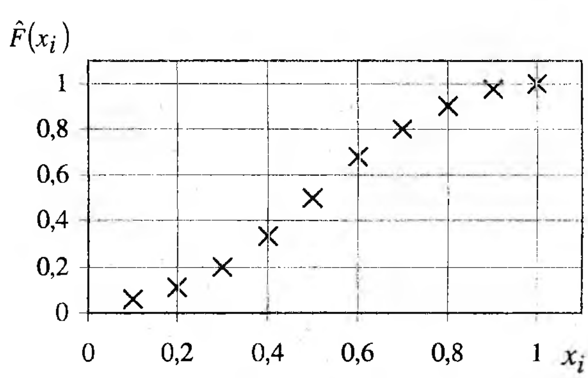


Рис. 1

Для идентификации формы распределения графическим методом применяются вероятностные сетки, которые представляют собой прямоугольную сетку, на которой масштаб выбран таким образом, чтобы график функции этого распределения, построенный на сетке, представлял прямую линию.

Следовательно, основанием для построения вероятностных сеток является целесообразное преобразование одной или обеих прямоугольных координат $\{x, F(x)\}$ графика функции распределения таким образом, чтобы взаимная зависимость обеих преобразованных переменных стала линейной.

Для выборки объемом $n$ значений случайной величины $X$ на вероятностную сетку для данного вида распределения наносят точки эмпирической функции распределения $\hat{F}(x)$. Затем по этим точкам проводят прямую так, чтобы нанесенные точки отклонялись от нее как можно меньше. Если нанесенные эмпирические точки мало отклоняются от проведенной прямой, то это свидетельствует о том, что опытные данные не противоречат тому виду распределения, для которого была построена сетка.

Построение вероятностных сеток производится для заданного распределения. Рассмотрим построение сеток для некоторых распределений. Вероятностная сетка для распределений нормального, равномерного, треугольного, Лапласа и арксинуса содержит по оси абсцисс равномерную шкалу, по оси ординат откладывают значения $y$, а соответствуют им значения $F(y)$, которые рассчитываются по формулам, приведенным в табл. 1 для конкретного распределения.

Таблица 1

| Распределение | Значение $F(y)$, соответствующее величине $y$ |
|---|---|
| Нормальное | $F(y)=\frac{1}{2\pi}\int\limits_{-\infty}^{y} e^{-\frac{x^2}{2}}dx$ |
| Лапласа | $F(y)=\begin{cases}\frac{1}{2}e^{y}, & y<0\\ 1-\frac{1}{2}e^{-y}, & y\ge 0\end{cases}$ |
| Треугольное | $F(y)=\begin{cases}\left(\frac{y}{\sqrt{12}}+\frac{1}{\sqrt{2}}\right)^2, & y<0\\ 1-\left(\frac{y}{\sqrt{12}}-\frac{1}{\sqrt{2}}\right)^2, & y\ge 0\end{cases}$ |
| Арксинуса | $F(y)=\frac{1}{2}+\frac{1}{\pi}\arcsin\frac{y}{\sqrt{2}}$ |
| Равномерное | $F(y)=\frac{y-\sqrt{3}}{2\sqrt{3}}$ |

Поскольку вид закона распределения априорно не известен, то приходится производить построение несколько раз на вероятностных сетках для различных распределений и выбрать из них то, которое при построении более линейно. Этот процесс, выполняемый вручную, весьма трудоемок. Кроме того, процесс принятия решений при использовании вероятностных сеток носит субъективный характер.

Нами предлагается следующий метод идентификации формы распределения. Данные, полученные в результате измерений, преобразовать таким образом, чтобы при построении экспериментальной функции распределения в прямоугольной сети координат с равномерными осями абсцисс и ординат она представляла прямую линию. Т.е. производить преобразование не осей, а самих данных. Очевидно, что преобразующей функцией является выражения интегральной функции распределения для заданного закона. Если преобразующая функция не соответствует экспериментальным данным, то экспериментальная функция распределения не преобразуется в прямую. Преобразуя экспериментальные данные различными функциями и аппроксимируя полученные данные прямыми линиями методом наименьших квадратов, определяем, для какой преобразующей функции сумма квадратов невязок окажется наименьшей. Основываясь на этом, можно предложить следующий критерий определения вида закона распределения:

1. Получение $n$ значений отсчета и расположение их в порядке возрастания.
2. Определения оценок центра распределения и рассеяния.

3. Нормирование полученных данных по формуле $x = \frac{x - \hat{m}}{\hat{\sigma}}$.

4. Преобразование полученных данных функциями из таблицы 1.

5. Аппроксимация полученных преобразованных данных прямыми линиями с помощью метода наименьших квадратов и определение квадратов невязок. Экспериментальное распределение соответствует той преобразующей функции, для которой сумма квадратов невязок окажется наименьшей.

Проиллюстрируем применение предложенного критерия на следующем примере. Возьмем выборки объемом $n = 10$ и $n = 15$, подчиняющиеся различным законам распределения, которые априорно известны: закону распределения арксинус, равномерному, треугольному, нормальному и распределению Лапласа. В каждой выборке произведем упорядочивание и нормирование членов, и для каждого члена выборки рассчитаем значение оценки $\hat{F}(x)$ по формуле (2). Произведем преобразование каждой из выборок функциями, соответствующими нормированным и центрированным интегральным функциям распределения вышеперечисленных законов .

Аппроксимируем преобразованные функции распределения прямыми вида $\hat{F}(x) = A_0 + A_1 x$ методом наименьших квадратов. Рассчитанные значения коэффициентов $A_0$, $A_1$ и невязок $S(\delta^2)$ приведены в табл. 2.

Таблица 2

| Преобразующая функция | Фактические распределения | Число наблюдений $n = 10$ | | | Число наблюдений $n = 15$ | | |
|---|---|---|---|---|---|---|---|
| | | Коэффициент $A_0$ | Коэффициент $A_1$ | Сумма квадратов невязок $S(\delta^2)$ | Коэффициент $A_0$ | Коэффициент $A_1$ | Сумма квадратов невязок $S(\delta^2)$ |
| Арксинус | Арксинус | **0,039** | **0,961** | **0,006** | **0,113** | **1,053** | **0,064** |
| | Равномерное | - | - | - | - | - | - |
| | Треугольное | - | - | - | - | - | - |
| | Нормальное | - | - | - | - | - | - |
| | Лапласа | - | - | - | - | - | - |
| Равномерное | Арксинус | -0,002 | 0,944 | 0,020 | 0,194 | 0,741 | 0,108 |
| | Равномерное | **0,032** | **1,081** | **0,004** | **-0,017** | **1,021** | **0,033** |
| | Треугольное | 0,164 | 0,335 | 0,024 | 0,012 | 0,941 | 0,035 |
| | Нормальное | -0,132 | 1,428 | 0,011 | -0,032 | 0,995 | 0,066 |
| | Лапласа | -0,084 | 1,027 | 0,006 | -0,15 | 1,309 | 0,072 |
| Треугольное | Арксинус | -0,068 | 1,052 | 0,013 | 0,096 | 0,764 | 0,089 |
| | Равномерное | 0,121 | 0,724 | 0,005 | 0,020 | 0,835 | 0,070 |
| | Треугольное | **-0,056** | **1,009** | **0,003** | **0,017** | **1,021** | **0,033** |
| | Нормальное | -0,061 | 0,877 | 0,005 | -0,022 | 0,985 | 0,044 |
| | Лапласа | -0,044 | 0,938 | 0,007 | -0,103 | 1,205 | 0,050 |
| Нормальное | Арксинус | | | | 0,228 | 0,706 | 0,120 |
| | Равномерное | | | | 0,032 | 0,812 | 0,072 |
| | Треугольное | | | | 0,056 | 0,852 | 0,050 |
| | Нормальное | | | | **0,024** | **0,0943** | **0,045** |
| | Лапласа | | | | -0,109 | 1,204 | 0,050 |
| Лапласа | Арксинус | -0,061 | 1,005 | 0,008 | 0,255 | 0,654 | 0,131 |
| | Равномерное | 0,155 | 0,663 | 0,008 | 0,057 | 0,749 | 0,090 |
| | Треугольное | 0,103 | 0,632 | 0,014 | 0,091 | 0,777 | 0,065 |
| | Нормальное | 0,027 | 0,751 | 0,010 | 0,048 | 0,843 | 0,061 |
| | Лапласа | **0,027** | **0,792** | **0,010** | **-0,017** | **1,021** | **0,033** |

Как видно из табл. 2, наименьшую $S(\delta^2)$ имеют те функции, для которых распределение и преобразующая функция совпадают. Кроме того, у них коэффициент $A_0$ близок к 0, а коэффициент $A_1$ близок к 1. Это можно считать вторичным признаком для идентификации распределения (кроме минимума квадратов невязок).

Графическое представление преобразований приведено на рис 2. На этом рисунке 1 – график эмпирической функции Лапласа, построенный по опытным данным, 2-6 – данные, имеющие распределения из табл. 2 (2 – Лапласа, 3 – равномерное, 4 – нормальное, 5 – арксинус, 6 – треугольное) и преобразованные функцией Лапласа.

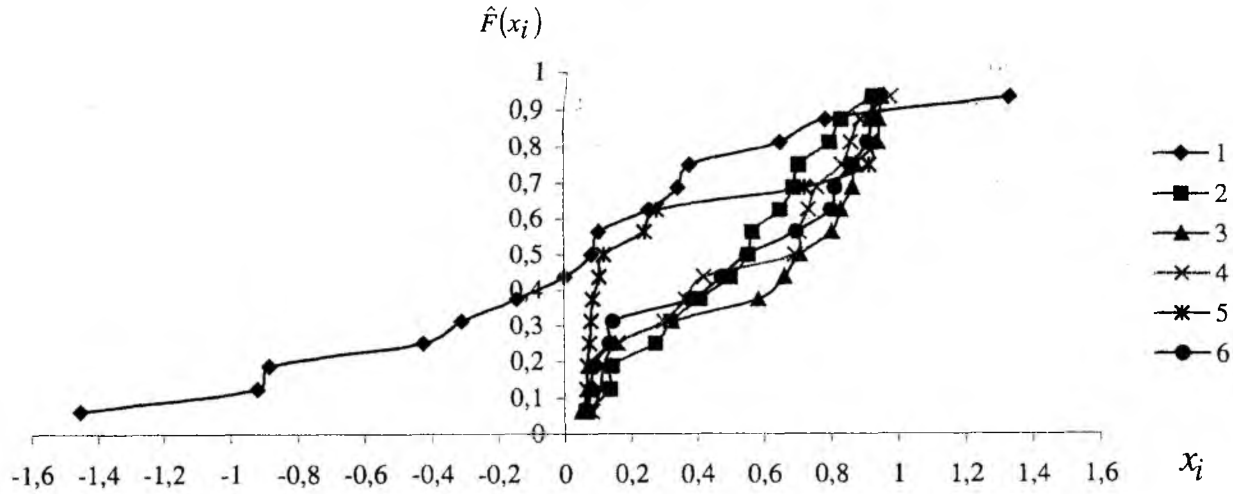


Рис. 2

Как видно из рис. 2, визуально различить преобразованные кривые довольно сложно, применение же математического аппарата позволяет это сделать. Это дает возможность не только исключить принятие субъективного решения, но и применять данный метод в интеллектуальных средствах измерения.

**Список литературы:** 1. *Закс Л.* Статистическое оценивание. М.: Статистика, 1976. 598 с. 2. *ГОСТ 11.006-74*. Правила проверки согласия опытного распределения с теоретическим. Введ 01.01.1976 г. М.: Изд-во стандартов, 1981. 31 с. 3. *Кудряшова Ж. Ф., Рабинович С.Г., Резник К.А.* Рекомендация по методам обработки результатов наблюдений при прямых измерениях // Методы обработки результатов наблюдений при измерениях. год вып. 134 (194). С. 5 – 113. (Тр. Метрологич. ин-тов СССР) 4. *Захаров И.П., Штефан Н.В.* Об эффективности применения составного критерия для оценки нормальности распределения // Тр. III Междунар. науч.- технич. конфер. «Метрологічне забезпечення в галузі електричних, магнітних та радіовимірювань» Харьков: ХГНИИМ. 2000. Т. 2. С. 164 – 166. (Метрологія в електрониці – 2000). 5. *Хан Г., Шапиро С.* Статистические модели в инженерных задачах. М.: Мир,1969. 396 с.

*Харьковский национальный*

*университет радиоэлектроники*

*Поступила в редколлегию 23.04.2002*