УДК 519.21

Л.О. Кириченко, А.В. Шкловец, Д.А. Поляков, Е.А. Боковиков

Харьковский национальный университет радиоэлектроники, Харьков

ОБ ОДНОМ МЕТОДЕ ИДЕНТИФИКАЦИИ ЗАКОНА ВЫБОРОЧНОГО РАСПРЕДЕЛЕНИЯ

В работе предложена новая характеристика непрерывной случайной величины – идентификатор распределения, определяемая как отношение квадрата математического ожидания размаха, к её дисперсии. Показано, что данная характеристика не зависит от линейного преобразования случайной величины. Численное моделирование случайных сигналов показало возможность определения закона распределения сигналов оцениванием данной характеристики.

Ключевые слова: *выборочное распределение, непрерывная случайная величина, численное моделирование.*

Постановка проблемы и анализ последних публикаций

При разработке интеллектуальных средств измерений, контроля и диагностики возникают задачи автоматического распознавания объекта исследований, представленного временным рядом наблюдений.

В случае, когда исходные данные носят случайный характер, распознавание заключается в идентификации вида распределения вероятностей этих данных, что позволяет выбрать оптимальные методы дальнейшей обработки информации.

© Л.О. Кириченко, А.В. Шкловец, Д.А. Поляков, Е.А. Боковиков

Общепринятым методом идентификации закона распределения случайной величины (СВ) является построение гистограммы, выбор подходящей функции плотности распределения и проверка по критерию согласия. При использовании этого метода необходимо либо участие человека для выдвижения гипотезы о законе распределения, либо автоматический перебор большого количества вариантов. Другим методом является анализ выборочных моментных характеристик и сравнение их с теоретическими. Эти методы и их компьютерная реализация подробно описаны в работах [1, 2].

Основываясь на некоторых идеях работы [3], авторами предложен способ применения описанного метода для непрерывных СВ с конечной дисперсией. В работе [4] был рассмотрен метод определения закона распределения непрерывной случайной величины (СВ), значения которой ограничены на некотором интервале, с помощью оценивания числовой характеристики данной СВ по её выборочным данным.

Целью данной работы является представление числовой характеристики непрерывной СВ, которая зависит лишь от закона распределения СВ, является идентификатором типа распределения и ее оценка позволяет определять закон распределения по выборочным данным.

Изложение основного материала

Теоретические замечания. Пусть задана выборка $x_1,\dots,x_n$ из непрерывной генеральной совокупности ξ. Упорядочив выборочные значения в порядке возрастания, получим вариационный ряд $x^{(1)} < \dots < x^{(n)}$. Под размахом выборки будем понимать статистику $R = x^{(n)} - x^{(1)}$. Размах выборки выражается через функцию распределения $F_\xi(x)$ генеральной совокупности ξ следующим образом [5]:

$$F_n(R) = n\int_{-\infty}^{\infty}\left(F(x+R)-F(x)\right)^{n-1}dF(x). \qquad (1)$$

Математическое ожидание размаха равно

$$M[R] = \int_{-\infty}^{\infty}\left(1-F^n(x)-(1-F(x))^n\right)dx\,. \qquad (2)$$

Идентификатор распределения. Пусть СВ ξ представлена выборкой длиной n из генеральной совокупности и СВ $R_\xi(n)$ определяет ее размах. Тогда числовой характеристикой закона распределения $F_\xi(x)$ является функция, названная идентификатором распределения

$$Id_\xi(n) = \frac{\left(M[R_\xi(n)]\right)^2}{D[\xi]}, \qquad (3)$$

где $M[R_\xi(n)]$ – математическое ожидание размаха, $D[\xi]$ – дисперсия СВ ξ. Идентификатор распределения определен для СВ с конечной дисперсией.

Докажем основное свойства идентификатора распределения.

Утверждение 1. Идентификатор распределения инвариантен относительно линейного преобразования случайной величины.

Доказательство. Пусть СВ ξ_0 с единичной дисперсией и нулевым математическим ожиданием имеет функцию распределения $F_{\xi_0}(x)$. Утверждение будет доказано, если показать, что идентификатор распределения для СВ $\xi = \sigma\xi_0 + m$ равен идентификатору распределения СВ ξ_0. Воспользовавшись формулой (2), получаем, что идентификатор распределения СВ ξ_0 равен:

$$Id_{\xi_0}(n) = \left(M[R_{\xi_0}(n)]\right)^2 \Big/ D[\xi] =$$

$$= \left(\int_{-\infty}^{\infty}\left(1-F_{\xi_0}^n(x)-\left(1-F_{\xi_0}(x)\right)^n\right)dx\right)^2.$$

Функция распределения СВ ξ имеет вид $F_\xi(x) = F_{\xi_0}\left(\frac{x-m}{\sigma}\right)$, а идентификатор распределения СВ ξ будет равен:

$$Id_\xi(n) = \left(M[R_\xi(n)]\right)^2 \Big/ \sigma^2 =$$

$$= \frac{\left(\int_{-\infty}^{\infty}\left(1-F_\xi^n(x)-\left(1-F_\xi(x)\right)^n\right)dx\right)^2}{\sigma^2} =$$

$$= \left(\int_{-\infty}^{\infty}\left(1-F_{\xi_0}^n\left(\frac{x-m}{\sigma}\right)-\left(1-F_{\xi_0}\left(\frac{x-m}{\sigma}\right)\right)^n\right)d\frac{x-m}{\sigma}\right)^2 =$$

$$= \left(\int_{-\infty}^{\infty}\left(1-F_{\xi_0}^n(y)-\left(1-F_{\xi_0}(y)\right)^n\right)dy\right)^2 = Id_{\xi_0}(n).$$

Утверждение доказано.

Рассмотрим основные законы распределений, такие как равномерное, нормальное, показательное и др. Из инвариантности равномерного распределения относительно линейного преобразования следует, что все равномерно распределенные СВ обладают одной и той же функцией $Id_\xi(n)$. Из инвариантности показательного распределения относительно линейного преобразования, где m=0, следует, что все показательно распределенные СВ независимо от значения λ имеют одинаковую функцию $Id_\xi(n)$. Также можно показать инвариантность Id от параметров для нормального, логистического, распределения арксинуса и других. Выведем формулу идентификатора распределения для равномерного распределения. Пусть СВ $\xi \sim R[a,b]$, тогда

$$F_\xi(x) = \frac{x-a}{b-a}, \text{ при } x\in[a,b], \; D[\xi] = \frac{(b-a)^2}{12}.$$

$$Id_\xi(n)=\left(\int_{-\infty}^{\infty}\left(1-F_\xi^n(x)-\left(1-F_\xi(x)\right)^n\right)dx\right)^2\Big/D[\xi]=$$

$$=2\left(\int_a^b\left(1-\left(\frac{x-a}{b-a}\right)^n-\left(1-\frac{x-a}{b-a}\right)^n\right)dx\right)^2\Big/(b-a)^2=$$

$$=12\left(b-a-\frac{\int_a^b(x-a)^n dx}{(b-a)^n}-\frac{\int_a^b(b-x)^n dx}{(b-a)^n}\right)^2\Big/(b-a)^2=$$

$$=\frac{12\left(b-a-\frac{(b-a)^{n+1}}{(b-a)^n(n+1)}-\frac{(b-a)^{n+1}}{(b-a)^n(n+1)}\right)^2}{(b-a)^2}=$$

$$=12\left((n-1)/(n+1)\right)^2.$$

Для приведенных в табл. 1 законов распределений идентификатор распределения не выражается через элементарные функции.

Таблица 1

Идентификаторы распределений

Тип распределения	$Id_\xi(n)$
Нормальное	$\int_{-\infty}^{\infty}\left(1-\Phi_0^n(x)-\left(1-\Phi_0(x)\right)^n\right)dx$
Экспоненциальное	$\int_0^{\infty}\left(1-\left(1-e^{-x}\right)^n-e^{-nx}\right)dx$
Логистическое	$\dfrac{3\int_{-\infty}^{\infty}\left(1-\frac{1}{(1+e^x)^n}-\left(1-\frac{1}{1+e^x}\right)^n\right)dx}{\pi^2}$
Арксинуса	$8\left(1-\left(\frac{2}{\pi}\right)^n\int_0^{\frac{\pi}{2}}\left(x^n\sin(2x)-(1-x)^n\sin(2x)\right)dx\right)$

Замечание: $\Phi_0(x)$ *– функция Лапласа.*

На рис. 1 приведено графическое изображение идентификаторов для логистического, экспоненциального, нормального, арксинуса, равномерного распределений (снизу вверх).

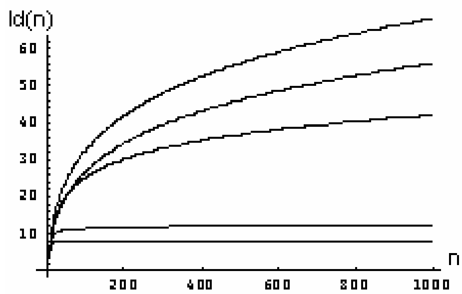

Рис. 1. Графики зависимостей $Id_\xi(n)$

Оценка идентификатора распределения

Оценкой, полученной по k выборкам, каждая объемом n, является величина

$$\hat{Id}_\xi(n)=\left(\frac{1}{k}\sum_{i=1}^{k}\left(X_{max}^i(n)-X_{min}^i(n)\right)\right)^2\Big/\left(\frac{1}{k}\sum_{i=1}^{k}S_i^2\right),$$

где X_{max}^i и X_{min}^i – максимальное и минимальное значения i-й выборки соответственно, S_i – среднеквадратическое отклонение i-й выборки:

$$S_i=\frac{1}{n-1}\sum_{j=1}^{n}\left(X_j^i-\frac{1}{n}\sum_{l=1}^{n}X_l^i\right)^2, i=\overline{1,k}.$$

Проведенные численно-аналитические исследования показали, что оценка $Id_\xi(n)$ является состоятельной

$$\hat{Id}_\xi(n)\xrightarrow[k\to\infty]{P}Id_\xi(n).$$

Распознавание по выборочным данным. Рассмотрим вопрос о распознавании закона распределения по данным, представленными одной выборкой. С помощью оценки идентификатора распределения можно решить следующую задачу: определить путем выбора из конечного числа законов распределения, какой закон распределения имеет генеральная совокупность, представленная выборочными значениями, либо определить, что такого закона среди рассматриваемых нет.

На рис. 2 приведены теоретические кривые зависимостей $Id_\xi(n)$ для нормального, логистического и распределения Лапласа, а так же выборочная кривая, полученная по одной выборке с нормальной генеральной совокупностью.

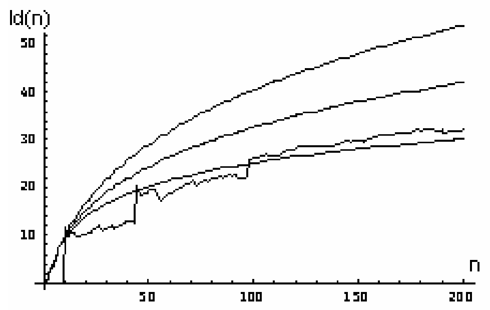

Рис. 2. Теоретические зависимости $Id_\xi(n)$ и выборочная кривая для нормального распределения

Для численного определения степени близости между теоретическими и выборочными кривыми $Id_\xi(n)$ в качестве критерия было выбрано среднее значение отклонений выборочного идентификатора от теоретического:

$$K=\frac{1}{n}\sum_{i=1}^{n}\left|Id_\xi(i)-\hat{Id}(i)\right|.$$

На рис. 3 представлена выборочная плотность распределения K-статистики для выборок объемом

$n = 50$ из равномерной и нормальной совокупностей. Левая гистограмма отображает отклонения равномерного выборочного идентификатора от равномерного теоретического, а правая – от теоретического нормального. Значение $K_{кр}$ соответствует уровню значимости $\alpha = 0,05$.

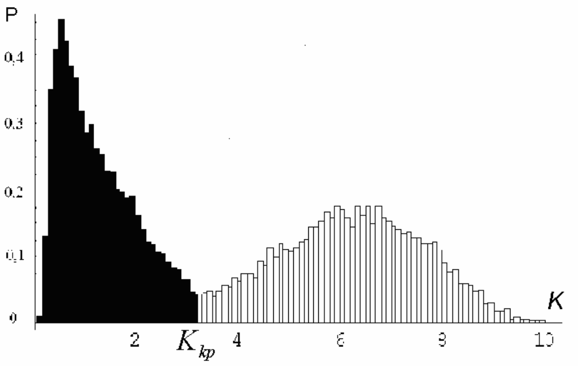

Рис. 3. Выборочная плотность распределения K-статистики

Использование данного метода позволяет определять закон распределения СВ при объеме выборки 50 элементов с вероятностью 90%.

В качестве еще одного критерия было предложено использовать среднюю скорость возрастания выборочного идентификатора. Для этого сравниваются приращения теоретических и практических идентификаторов по следующей формуле

$$\Delta = \frac{1}{n}\sum_{i=1}^{n}\left|\Delta^{i}_{практ} - \Delta^{i}_{теор}\right|,$$

где $\Delta^{i}_{теор} = \left|Id_{\xi}(i+1) - Id_{\xi}(i)\right|$, $\Delta^{i}_{практ} = \left|I\hat{d}(i+1) - I\hat{d}(i)\right|$.

Далее применяется та же схема идентификации распределения, что и предыдущем методе. Т.е. выборка будет иметь тот закон распределения, теоретическому идентификатору распределения которого будет соответствовать наименьшее значение Δ.

На рис. 4 представлена выборочная плотность распределения Δ-статистики для выборок объемом $n = 50$. Левая гистограмма отображает отклонения равномерного выборочного идентификатора от равномерного теоретического, центральная – от теоретического нормального, а правая – от теоретического Лапласа. Значение $\Delta_{кр}$ соответствует уровню значимости $\alpha = 0,05$. Данный критерий позволяет определять закон распределения СВ при объеме выборки 50 элементов с вероятностью около 90%.

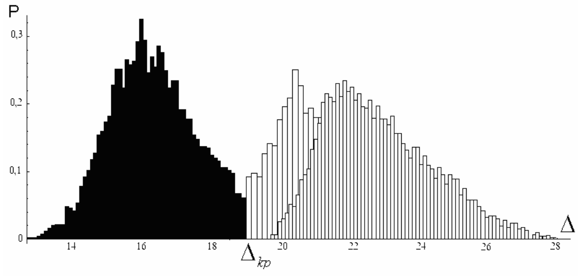

Рис. 4. Выборочная плотность распределения Δ-статистики

Вышеперечисленные методы распознавания с помощью идентификатора распределения апробированы для малых объемов выборочных данных. Если же выборочные данные представлены в большом объеме, то для уточнения полученных результатов предлагается разбивать выборку на k независимых выборок одного объёма и рассчитывать оценку Id не по одной выборке, а по нескольким.

Выводы

Таким образом, в работе предложена числовая функциональная характеристика закона распределения непрерывной случайной величины, получены аналитические выражения идентификатора распределения для распространенных распределений, предложен метод идентификации закона распределения непрерывной случайной величины по выборочным данным для больших и малых выборок.

Список литературы

1. *Афифи А., Эйзен С. Статистический анализ. Подход с использованием ЭВМ / Пер. с англ. И.С. Енюкова, И. Д. Новикова. – М.: Мир, 1982. – 486 с.*

2. *Тюрин Ю.Н., Макаров А.А. Статистический анализ данных на компьютере. – М: Инфра, 1997. – 528 с.*

3 *Кликушин Ю.Н. Фрактальная шкала для измерения распределений вероятности // Журнал радиоэлектроники. – М., 2000. – № 3 [Электронный ресурс]. – Режим доступа: http://jre.cplire.ru/koi/mar00/2/text.html.*

4 *Кириченко Л.О., Шкловец А.В. Об одном методе идентификации закона распределения случайной величины // Системи обробки інформації. – Х.: ХУ ПС, 2005. – Вип. 4 (44). – С. 197-201.*

5 *Справочник по теории вероятностей и математической статистике / В.С. Королюк, Н.И. Портенко, А.В. Скороход, А.Ф. Турбин. – М.: Наука, 1985. – 640 с.*

Поступила в редколлегию 8.07.2008

Рецензент: д-р техн. наук, проф. Е.П. Путятин, Харьковский национальный университет радиоэлектроники, Харьков.

ПРО ОДИН МЕТОД ІДЕНТИФІКАЦІЇ ЗАКОНУ ВИБІРКОВОГО РОЗПОДІЛУ

Л.О. Кіріченко, А.В. Шкловець, Д.А. Поляков, Є.А. Боковіков

У роботі запропонована нова характеристика безперервної випадкової величини – ідентифікатор розподілу, визначувана як відношення квадрата математичного очікування розмаху до її дисперсії. Показано, що дана характеристика не залежить від лінійного перетворення випадкової величини. Чисельне моделювання випадкових сигналів показало можливість визначення закону розподілу сигналів оцінюванням даної характеристики.

Ключові слова: *вибірковий розподіл, безперервна випадкова величина, чисельне моделювання.*

ABOUT ONE METHOD OF AUTHENTICATION OF LAW OF SAMPLING DISTRIBUTION

L.O. Kirichenko, A.V. Shklovets, D.A. Poland, J.A. Bokovikov

New description of continuous random quantity is in-process offered is a distributing identifier, determined as a relation of square of the expected value of scope, to its dispersion. It is rotined that this description does not depend on linear transformation of random quantity . The numerical simulation of casual signals rotined possibility of determination of law of distributing of signals the evaluation of this description.

Keywords: *sampling distribution,continuous random quantity , numerical simulation.*