УДК 658.562:519.23 **doi: 10.26906/SUNZ.2019.1.099**

Р. М. Трищ, А. Н. Денисенко, Е. Н. Черняк

Украинская инженерно–педагогическая академия, Харьков, Украина

МЕТОД ОПРЕДЕЛЕНИЯ ЗАКОНА РАСПРЕДЕЛЕНИЯ ПОКАЗАТЕЛЕЙ КАЧЕСТВА ИЗДЕЛИЙ КАК СЛУЧАЙНОЙ ВЕЛИЧИНЫ

Целью статьи является создание метода определения закона распределения случайной величины не требующего большого количества выборочных данных и обеспечивающего достаточную надежность расчета. В статье рассмотрены существующие подходы идентификации вида закона распределения показателей качества изделий. Предлагается метод идентификации закона распределения случайных величин показателей качества изделий с использованием теории порядковых статистик. Экспериментально доказано, что предложенный метод позволит определить закон распределения при малом количестве статистических данных. Математическая модель реализована с помощью доступного и простого компьютерного математического пакета Mapl 9 для определения, из какого закона из трех: нормального, равномерного или Симпсона произведена выборка. Для реализации метода предлагается методика идентификации закона распределения.

Ключевые слова: закон распределения; порядковые статистики; малая выборка; графоаналитический метод; плотность распределения; математическое ожидание; идентификация закона; показатель качества.

Введение

Определение вида закона распределения показателей качества изделий, как случайной величины, важный процесс при их оценивании, особенно при ограниченном количестве существующей статистической информации, так как от правильности идентификации закона распределения зависит достоверность решения практических задач статистическими методами. Идентификация вида закона распределения показателей качества изделий (закон распределения) по опытным данным занимает одно из центральных мест при обработке результатов экспериментов статистическими методами. Этому вопросу уделено много внимания в литературе [1–5].

Традиционный подход при решении задач сводится к расчету параметров эмпирического распределения, принятию их в качестве оценок параметров генеральной совокупности случайных величин с последующей проверкой сходимости эмпирического распределения с предполагаемым теоретическим по критериям χ^2 (Пирсона), λ (Колмогорова), ω^2 (Мизеса). Такой χ^2подход имеет следующие недостатки: зависимость методики обработки результатов эксперимента от предполагаемого теоретического распределения, большой объем вычислений, особенно при использовании χ^2 и ω^2 [6]. Некоторые новые критерии [7] не имеют удовлетворительного теоретического обоснования, а в ряде случаев, как это показано в работе [7], не обладают достаточной мощностью. Б. Е. Янковский [8] предложил информационный способ определения закона распределения, суть которого заключается в использовании энтропии эмпирического распределения случайной величины. Недостаток предложенного метода в том, что отсутствуют оценки допустимого расхождения между эмпирической и теоретической энтропиями, что не позволяет его эффективного применения на практике.

Цель работы – создание метода определения закона распределения случайной величины, не требующего большого количества выборочных данных и обеспечивающего достаточную надежность расчета.

Основная часть

Графоаналитический метод определения закона распределения. Функции распределения вида $F\left((x-\mu)/\sigma\right)$ относятся к весьма широкому классу распределений, включающему все распределения, подобные нормальному. Они характеризуются параметром расположения μ и мерой рассеивания σ, а также могут включать все распределения с одним параметром, например: экспоненциальное распределение с известным началом отсчета - ноль. Причем параметры μ и σ не обязательно должны быть математическим ожиданием и стандартным отклонением, как у нормального закона распределения.

Так как плотность распределения $f(x) = F'(x)$, то она имеет вид:

$$f(x)= g\left((x-\mu)/\sigma\right)/\sigma, \qquad (\sigma > 0).$$

Отсюда, стандартизированная случайная величина $Y = (x-\mu)/\sigma$ имеет плотность распределения g(y), которая не зависит от μ и σ. Так, например, для равномерного распределения $f(x) = 1/(b-a)$ при $a \le x \le b$. Плотность распределения случайной величины $Y = (x-\mu)/\sigma$ имеет вид: g(y)=1, при $0 \le y \le 1$.

Пусть имеется n независимых наблюдений x_1, x_2, …, x_n непрерывной случайной величины x с функцией распределения $F\left((x-\mu)/\sigma\right)$. Расположим наблюдения по возрастанию их величины и обозначим упорядоченные результаты: $x_{(1)}$, $x_{(2)}$, …, $x_{(n)}$.

Так как упорядоченные случайные величины X и Y (для выборки объема n) связаны соотношением

$$Y_{(i)} = \left(x_{(i)}-\mu\right)/\sigma \qquad (i=1, 2,…, n), \quad (1)$$

то моменты случайной величины зависят только от вида g(y), но не от μ и σ.

Тогда математическое ожидание i-ой порядковой статистики (нормированной) случайной величины $Y_{(i)}$ есть некоторое число, независящее от параметров, вычисляемое по формуле:

$$M(Y_{(i)}) = \frac{n!}{(i-1)!(n-i)!} \times \\ \times \int_{-\infty}^{\infty} y\left[F(y)\right]^{i-1}\left[1-F(y)\right]^{n-i} f(y)dy. \quad (2)$$

Отсюда следует, что для всех распределений вида $F\left((x-\mu)/\sigma\right)$ можно раз и навсегда вычислить математическое ожидание нормированной случайной величины $Y_{(i)}$ по формуле (2).

Из свойств математических ожиданий и соотношения (1) имеем:

$$M(X_{(i)}) = \mu + \sigma M(Y_{(i)}), \quad (3)$$

Таким образом, $M(X_{(i)})$ является линейной комбинацией параметров μ и σ с известными коэффициентами $M(Y_{(i)})$.

Отсюда, при заданном объеме выборки n можно на оси абсцисс нанести значения $M(Y_{(i)})$, а на оси ординат выборочные упорядоченные результаты (рис. 1).

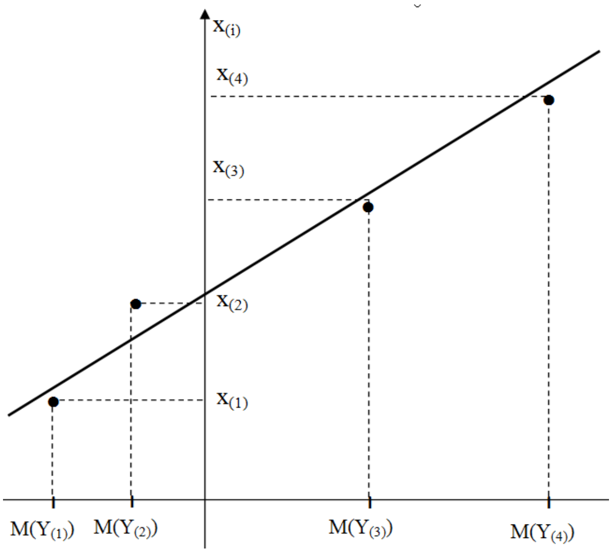

Рис. 1. Линейная зависимость порядковых статистик от нормированных математических ожиданий порядковых статистик

Принимая, что $M(X_{(i)})=x_{(i)}$, находим по способу наименьших квадратов ту прямую (рис. 1), которая аппроксимирует эти точки, т.е. такую прямую, чтобы сумма квадратов отклонений была минимальна.

$$L = \sum_{i=1}^{n} (x_{(i)} - \mu - \sigma M(Y_{(i)}))^2 \quad (4)$$

В этом случае параметры получат оценки:

$$\mu = \frac{A_2 S_1 - A_1 S_2}{n A_2 - A_1^2}; \quad (5)$$

$$\sigma = \frac{n S_2 - A_1 S_1}{n A_2 - A_1^2}, \quad (6)$$

где $A_1 = \sum_{i=1}^{n} M(Y_{(i)})$; $A_2 = \sum_{i=1}^{n} \left[M(Y_{(i)})\right]^2$;

$$S_1 = \sum_{i=1}^{n} x_{(i)}; \; S_2 = \sum_{i=1}^{n} x_{(i)} \cdot M(Y_{(i)}). \quad (7)$$

Итак, чтобы определить закон распределения по малой выборке начиная с объема n=10, нужно найти L по формуле (4) для различных законов, используя математические ожидания их порядковых статистик и формулы (5), (6) и (7). Для выборок объема от 3 и больше, необходимо применять оптимальные линейные оценки параметров распределений и сравнивать их по величине. Наименьшее из L будет соответствовать принадлежности выборки определенному распределению.

Методика идентификации закона распределения. Для проверки предлагаемой идеи, определения принадлежности выборочных значений из генеральной совокупности небольшого объема одному из возможных законов распределения случайных величин предлагается использовать метод имитационного моделирования. Для этого необходимо сгенерировать случайные величины, имеющие один из законов распределения с определенными параметрами. Например, сгенерируем случайные числа по нормальному закону распределения с параметрами μ=2 и σ=3 с определенным объемом, например, n=10:

$x_{(1)}$= -0,35304; $x_{(2)}$= 0,283042; $x_{(3)}$= 0,914675;
$x_{(4)}$= 0,977164; $x_{(5)}$= 1,635863; $x_{(6)}$= 2,192437; (8)
$x_{(7)}$= 3,432361; $x_{(8)}$= 3,542162; $x_{(9)}$= 4,636724;
$x_{(10)}$= 7,18877.

Пользуясь найденным $M(Y_{(i)})$ для n=10 имеем:
$M(Y_{(1)})$ = - 1,5387527308; $M(Y_{(2)})$ = - 1,0013570446;
$M(Y_{(3)})$ = - 0,6560591057; $M(Y_{(4)})$ = - 0,375764697;
$M(Y_{(5)})$ = - 0,1226677523; $M(Y_{(6)})$ = 0,1226677523;
$M(Y_{(7)})$ = 0,375764697; $M(Y_{(8)})$ = 0,6560591057;
$M(Y_{(9)})$ = 1,0013570446; $M(Y_{(10)})$ = 1,5387527308.

По формулам (6) и (5) имеем:

μ=2,445015800; σ=2,360192827.

Из (4) имеем: L=3,117140239.

Для равномерного закона математическое ожидание нормированной порядковой статистики:

$$M(X_{(i)}) = \frac{i}{n+1}$$

и для него по результатам выборки оценка:
μ= 1,535372994; σ=7,9600777587, L=3,994140401.

Для закона Симсона при объеме выборки n=10 были найдены математические ожидания нормированных порядковых статистик:
$M(Z_{(1)})$ = - 0,308895; $M(Z_{(2)})$ = - 0,213317;
$M(Z_{(3)})$ = - 0,141475; $M(Z_{(4)})$ = - 0,080991;
$M(Z_{(5)})$ = - 0,026373; $M(Z_{(6)})$ = -0,026373;
$M(Z_{(7)})$ = 0,080991; $M(Z_{(8)})$ = 0,141475;
$M(Y_{(9)})$ = 0,213317; $M(Z_{(10)})$ = 0,308895.

Фрагменты таблиц математических ожиданий порядковых статистик для рассматриваемых законов распределения для объемов n = 2-10 представлены в табл. 1-3.

Таблица 1 – **Математические ожидания порядковых статистик в выборке объемом n из совокупности, подчиняющейся нормальному распределению**

n	$Y_{(1)}$	$Y_{(2)}$	$Y_{(3)}$	$Y_{(4)}$	$Y_{(5)}$	$Y_{(6)}$	$Y_{(7)}$	$Y_{(8)}$	$Y_{(9)}$	$Y_{(10)}$
10	-1,538752	-1,001357	-0,656059	-0,375764	-0,122667	0,122667	0,375764	0,656059	1,001357	1,538752
9	-1,485013	-0,932297	-0,571970	-0,274525	0	0,274525	0,571970	0,932297	1,485013	
8	-1,423600	-0,852224	-0,472822	-0,152514	0,152514	0,472822	0,852224	1,423600		
7	-1,352178	-0,757374	-0,352707	0	0,352707	0,757374	1,352178			
6	-1,267206	-0,641755	-0,201547	0,201547	0,641755	1,267206				
5	-1,162964	-0,495019	0	0,495019	1,162964					
4	-1,029375	-0,297011	0,297011	1,029375						
3	-0,846284	0	0,846284							
2	-0,564189	0,564189								

Таблица 2 – **Математические ожидания порядковых статистик в выборке объемом n из совокупности, подчиняющейся распределению Симсона**

n	$Z_{(1)}$	$Z_{(2)}$	$Z_{(3)}$	$Z_{(4)}$	$Z_{(5)}$	$Z_{(6)}$	$Z_{(7)}$	$Z_{(8)}$	$Z_{(9)}$	$Z_{(10)}$
10	-0,308895	-0,213317	-0,141475	-0,080991	-0,02637	0,026373	0,080991	0,141475	0,213317	0,308895
9	-0,299337	-0,198949	-0,123330	-0,059144	0	0,059144	0,123330	0,198949	0,299337	
8	-0,288183	-0,182145	-0,101935	-0,032858	0,032858	0,101935	0,182145	0,288183		
7	-0,274928	-0,162092	-0,076031	0	0,076031	0,162092	0,274928			
6	-0,258809	-0,137503	-0,043446	0,043446	0,137503	0,258809				
5	-0,238591	-0,106151	0	0,106151	0,238591					
4	-0,212103	-0,063690	0,063690	0,212103						
3	-0,175000	0	0,175000							
2	-0,116667	0,116667								

Таблица 3 – **Математические ожидания порядковых статистик в выборке объемом n из совокупности, подчиняющейся равномерному распределению**

n	$X_{(1)}$	$X_{(2)}$	$X_{(3)}$	$X_{(4)}$	$X_{(5)}$	$X_{(6)}$	$X_{(7)}$	$X_{(8)}$	$X_{(9)}$	$X_{(10)}$
10	0,090909	0,181818	0,272727	0,363636	0,454545	0,545455	0,636364	0,727273	0,818182	0,909091
9	0,1	0,2	0,3	0,4	0,5	0,6	0,7	0,8	0,9	
8	0,111111	0,222222	0,333333	0,444444	0,555556	0,666667	0,777778	0,888889		
7	0,125	0,25	0,375	0,5	0,625	0,75	0,875			
6	0,142857	0,285714	0,428571	0,571429	0,714286	0,857143				
5	0,166667	0,333333	0,5	0,666667	0,833333					
4	0,2	0,4	0,6	0,8						
3	0,25	0,5	0,75							
2	0,333333	0,666667								

По результатам выборки (7) имеем для этого распределения по формулам (5) и (6):

$$\mu = 2{,}4450115800;$$

$$\sigma = 11{,}42630684;$$

$$L = 3{,}28560048.$$

Из проведенных расчетов видно, что из этих трех распределений близким является нормальное распределение, что действительно есть, так как нами генерировалась выборка из нормального закона.

Такой результат был получен нами и для генерированного равномерного закона и закона Симпсона при различных объемах выборки $n=10\div20$.

Итак, данный метод действительно может дать ответ, из какого закона распределения произведена малая выборка.

Нами составлена программа в математическом пакете Mapl 9, для определения, из какого закона из трех: нормального, равномерного или Симпсона произведена выборка при объемах n от 10 до 20.

Предложим методику идентификации закона распределения случайных величин по выборке небольшого объема, которая заключается в следующем:

1) берется выборка из партии контролируемых изделий, объемом n = 10-20 шт. и измеряются действительные значения контролируемого показателя качества;

2) значения действительных значений показателей качества выстраивается в вариационный ряд (в порядке возрастания);

3) определяется оценки параметров для различных законов распределения по формулам (5) и

(6) или по оптимальным линейным оценкам, если они получены для этих законов.

4) определяется сумма квадратов отклонений аппроксимирующей прямой эмпирических значений от аппроксимирующей прямой математических ожиданий порядковых статистик по формуле (4);

5) закон, которому принадлежит наименьшее из значений L является законом распределения эмпирических значений контролируемого показателя качества.

Выводы

Данный метод позволит определить по малому числу наблюдений выбор закона распределения случайной величины X. Составленная нами программа для трех распределений: нормального, равномерного и Симпсона для выборок объема $n=2 \div 20$ позволяет определить из какого закона взята выборка.

Список литературы

1. Трищ Р. М. Обобщённая точечная и интервальная оценки качества изготовления детали ДВС / Р. М. Трищ, Е. А. Слитюк. // Восточно-Европейский журнал передовых технологий. – 2006. – №1/2 (19). – С. 63–67.
2. Арпентьєв Б. М. Особливості систем управління якістю механозбирального виробництва / Б. М. Арпентьєв, Р. М. Тріщ. // Стандартизація. Сертифікація. Якість. – 2005. – №1. – С. 68 – 72.
3. Трищ Р. М. Оценка модели точности изготовления деталей на адекватность // Вісник НТУ „ХПІ”. Збірник наукових праць. Тематичний випуск: Нові рішення в сучасних технологіях. – Харків: НТУ „ХПІ”. – 2005. – № 57 – С. 61-65.
4. Трищ Р. М. Определение закона распределения точности механической обработки // Восточно-Европейский журнал передовых технологий. – Харьков. – 2005. – № 5/2 (17) – С. 46-48.
5. Трищ Р. М. Применение чувствительных характеристик для определения модели точности изготовления деталей // Восточно-Европейский журнал передовых технологий. – Харьков. – 2005. – № 6/2 (18) – С. 11-13.
6. Информационные методы в управлении качеством / В. Г. Григорович, С. В. Юдин, Н. О. Козлова, В. В. Шильдин. – Москва: РИА «Стандарты и качество», 2001. – 208 с.
7. Розно М. И. Статистический контроль партий продукции по альтернативному признаку при измененном допуске // Надежность и контроль качества. – 1992. - № 2. – С. 44-52.
8. Янковский Б. Е. Информационный способ определения вида закона распределения // Надежность и контроль качества. – 1971. - № 2. – С. 71-79.

Рецензент: д-р техн. наук, проф. Г. І. Канюк,
Українська інженерно-педагогічна академія, Харків
Received (Надійшла)
Accepted for publication (Прийнята до друку)

Метод визначення закон розподілу показників якості виробів як випадкової величини

Р. М. Тріщ, А. М. Денисенко, О. М. Черняк

Метою статті є створення методу визначення закону розподілу випадкової величини який не потребує великої кількості вибіркових даних і забезпечує достатню надійність розрахунку. У статті розглянуті існуючі підходи ідентифікації виду закону розподілу показників якості виробів. Пропонується метод ідентифікації закону розподілу випадкових величин показників якості виробів з використанням теорії порядкових статистик. Експериментально доведено, що запропонований метод дозволить визначити закон розподілу при малій кількості статистичних даних. Математична модель реалізована за допомогою доступного і простого комп'ютерного математичного пакета Mapl 9 для визначення, з якого закону з трьох: нормального, рівномірного або Сімпсона проведена вибірка. Для реалізації методу пропонується методика ідентифікації закону розподілу.

Ключові слова: закон розподілу; порядкові статистики; мала вибірка; графоаналітичний метод; щільність розподілу; математичне очікування; ідентифікація закону; показник якості.

Method for determining the law of distribution of the indicators of quality of products as random quantity

R. Trishch, A. Denysenko, O. Cherniak

The aim of the article is to provide a method for determining the law of the random variable distribution does not require a large amount of sample data and provide sufficient reliability calculation. The article discusses the existing approaches to identifying the type of distribution law for product quality indicators. A method is proposed for identifying the distribution law of random variables of product quality indicators using the theory of ordinal statistics. To determine the belonging of sample values from the general population of a small volume to one of the possible laws of the distribution of random variables, it is proposed to use a simulation method. It is proposed to generate random variables that have one of the distribution laws with certain parameters. Experimentally generated random numbers according to the normal law. From the calculations carried out in the article, it can be seen that from these three distributions the closest is the normal distribution, which really is, since we generated a sample from the normal law. Thus, it is proved that the proposed method will allow to determine the distribution law with a small amount of statistical data. The mathematical model is implemented using the accessible and simple computer math package Mapl 9 to determine from which law of the three: normal, uniform, or Simpson, the sample was taken. To implement the method, a method of identification of the distribution law is proposed.

Keywords: distribution law; ordinal statistics; small excerpt; graphoanalytical method; distribution density; expected value; law identification; quality indicator.