УДК 519.24

М.С.Токмачёв

ХАРАКТЕРИЗАЦИЯ ЛОГАРИФМИЧЕСКОГО РАСПРЕДЕЛЕНИЯ УСЛОВИЕМ ПОСТОЯНСТВА РЕГРЕССИИ

Институт электронных и информационных систем НовГУ

Conditions for logarithmic distributions characterization by quality of regress constancy of the fourth order statistics to the linear form are derived. Both statistics are deduced on representative sampling of n volume.

Характеризации вероятностных распределений некоторыми стохастическими свойствами статистик от независимых наблюдений представляют теоретическую ветвь развития математической статистики, они призваны систематизировать и вычленить классы распределений в соответствии с их структурными «глубинными» свойствами. Подобный подход к распределениям как совокупности характерных специфических обуславливающих их признаков зачастую является определяющим в конкретном исследовании, поскольку позволяет редуцировать исходные вероятностные задачи к более простому и осмысленному виду.

В предлагаемой работе рассматривается характеризация вероятностного однопараметрического (p) распределения с характеристической функцией (х. ф.)

$$f(t) = \frac{1}{\ln p} \ln(1 - qe^{it}), \tag{1}$$

где $p \in (0;1)$; $q = 1 - p$, известного в литературе как логарифмическое распределение [1].

Дискретная случайная величина (с.в.) X, имеющая логарифмическое распределение, принимает значения $x_m = m$, где $m = 1, 2, \ldots$, с соответствующими вероятностями

$$p_m = -\frac{(1-p)^m}{m \ln p}.$$

Число p — параметр распределения, интерпретируемый как вероятность ($0 < p < 1$).

Для указанного распределения известно, что

$$M(X) = -\frac{1-p}{p \ln p},\ D(X) = -\frac{1-p}{p^2 \ln p}\left(1 + \frac{1-p}{\ln p}\right). \tag{2}$$

Характеризационным условием является условие постоянства регрессии полиномиальной статистики Q на линейную форму Λ, где Q и Λ — некоторые функции независимых одинаково распределенных случайных величин, в частности, Q и Λ — статистики определенного вида, составленные по репрезентативной выборке $X_1, X_2, \ldots, X_n$ из генеральной совокупности X. Подобные характеризации известны для многих распределений: нормального, пуассоновского, гамма-распределения, биномиального, отрицательного биномиального, типа гиперболического косинуса, Лапласа и ряда других [2-10]. При этом статистика Q, используемая для характеризации, имеет, как правило, второй либо третий порядок, ибо повышение порядка рассматриваемой статистики влечет резкое увеличение сложности поиска положительно-определенных решений дифференциальных уравнений, соответствующих условию постоянства регрессии. В данной работе полиномиальная статистика Q имеет четвертый порядок, причем старшая степень переменных — третья. Коэффициенты Q взяты конкретные и зависящие лишь от n (количества с. в.). Эта фиксация взаимосвязи коэффициентов приводит, естественно, не к общему исследованию для статистик соответствующего порядка в зависимости от соотношения коэффициентов, а к характеризации приведенного выше распределения, осуществляемой с помощью условия нулевой регрессии статистик.

При этом единственный параметр распределения p формируется, исходя из соотношения первых моментов, μ и σ^2, а не коэффициентов статистик, Никаких ограничений на моменты, вызванных характеризационным условием, не предполагается. Тем самым несмотря на конкретные коэффициенты статистик Q и Λ имеет место характеризация логарифмического распределения с любым допустимым параметром.

Проведение такого исследования в общем виде (см. напр.,[4], [8]) для указанных статистик 4-го порядка — задача, эквивалентная решению соответствующего дифференциального уравнения 3-го порядка с произвольными коэффициентами, что весьма не тривиально, ибо, как известно, даже для многих дифференциальных уравнений 1-го порядка не всегда можно найти общее решение (а также и решения в характеристических функциях). Отметим, что подобные статистики использовались в [3] для характеризации геометрического и пуассоновского распределений.

Итак, пусть $X_1, X_2, \ldots, X_n$ — независимые, одинаково распределенные с.в., обладающие моментами до 3-го порядка включительно*. Введем две статистики:

$$Q = \frac{1}{n(n-1)} \sum_{k=1}^{n} \sum_{\substack{j=1 \\ j \neq k}}^{n} X_k^3 X_j - \frac{2}{n(n-1)} \sum_{k=1}^{n} \sum_{\substack{j=1 \\ j \neq k}}^{n} X_k^2 X_j^2 + \frac{1}{n(n-1)} \sum_{k=1}^{n} \sum_{\substack{j=1 \\ j \neq k}}^{n} X_k^2 X_j, \tag{3}$$

и

$$\Lambda = X_1 + X_2 + \ldots + X_n. \tag{4}$$

Пусть выполнено условие нулевой регрессии

$$E(Q \mid \Lambda) = E(Q) = 0, \tag{5}$$

где $E(Q \mid \Lambda)$ — условное математическое ожидание.

Теорема. Условие нулевой регрессии (5), для статистик вида (3), (4) является характеристическим свойством распределений с.в. $X_1, X_2, \ldots, X_n$ с х.ф. $f(t)$ вида (1), $E(X_k) = \mu$, $D(X_k) = \sigma^2$. При этом характеризация логарифмического распределения имеет место в случае справедливости соотношения

$$0 < \frac{\mu}{\mu^2 + \sigma^2} < 1,$$

а в случае справедливости соотношения

$$\frac{\mu}{\mu^2 + \sigma^2} \geq 1 \tag{6}$$

или

$$\frac{\mu}{\mu^2 + \sigma^2} \leq 0, \tag{7}$$

(что равносильно $\mu \leq 0$) условие (5) характеризует лишь вырожденное распределение.

Доказательство. Рассмотрим условие (5), которое, как известно [5], равносильно равенству

$$E(Qe^{it\Lambda}) = E(Q)E(e^{it\Lambda}) = 0. \tag{8}$$

Подставив в выражение $E(Qe^{it\Lambda})$ конкретные статистики Q и Λ вида (3), (4), вводя х.ф. $f(t)$ как $f(t) = E(e^{itX_k})$, $k = 1, 2, \ldots, n$, следовательно,

$$f^{(m)}(t) = i^m E(X_k^m e^{itX_k}), \tag{9}$$

и проделав соответствующие выкладки, приходим вместо (8) к дифференциальному уравнению в характеристических функциях

$$\frac{f''' f' f^{n-2}}{i^4} - 2\frac{(f'')^2 f^{n-2}}{i^4} + \frac{f'' f' f^{n-2}}{i^3} = 0.$$

*Полученное в итоге характеризации логарифмическое распределение обладает моментами любого порядка.

Поскольку нас интересует не общее решение, а лишь решения в характеристических функциях, и для любой х.ф. всегда $f(0) = 1$, то по непрерывности можем полагать, что в некоторой окрестности нуля $f(t) \neq 0$. Следовательно, поделив обе части последнего равенства на $f^{n-2}(t)$, получаем

$$f''' f' - 2(f'')^2 + if'' f' = 0. \qquad (10)$$

Ищем частное решение уравнения (10), удовлетворяющее начальным условиям

$$f(0) = 1,\ f'(0) = i\mu,\ f''(0) = -(\mu^2 + \sigma^2). \qquad (11)$$

Эти соотношения справедливы для любой х.ф. и следуют из формулы (9).

Понизим порядок (10) с помощью замены функции

$$z = f'. \qquad (12)$$

Тогда $f'' = z'$, $f''' = z''$, и уравнение принимает вид

$$z'' z - 2(z')^2 + iz' z = 0.$$

Поделим обе части равенства на z^2. При этом заметим, что равенство $z = 0$ приводит к $f = \mathrm{const}$, т.е. лишь к вырожденному распределению с.в. X_k, $k = 1, 2, \ldots n$. Таким образом, получаем

$$\frac{z''}{z} - 2\left(\frac{z'}{z}\right)^2 + i\frac{z'}{z} = 0$$

и далее осуществляем вновь замену переменной:

$$U = \ln z. \qquad (13)$$

Следовательно,

$$U' = \frac{z'}{z},\ U'' = \frac{z''}{z} - \left(\frac{z'}{z}\right)^2,$$

и наше уравнение сводится к уравнению

$$U'' - (U')^2 + iU' = 0, \qquad (14)$$

которое вновь допускает понижение порядка.

Замена

$$y = U' \qquad (15)$$

приводит (14) к уравнению Бернулли:

$$y' + iy - y^2 = 0. \qquad (16)$$

Для классического уравнения (16) используем соответствующую замену

$$y = Ve^{-it}, \qquad (17)$$

которая переводит (16) в дифференциальное уравнение с разделяющимися переменными

$$V' = V^2 e^{-it}. \qquad (18)$$

Интегрируя (18), получаем

$$V = \frac{1}{C_1 - ie^{-it}},$$

откуда и из (17) следует, что

$$y = \frac{e^{-it}}{C_1 - ie^{-it}}. \qquad (19)$$

При этом полагаем $C_1 \neq 0$, ибо в противном случае, возвращаясь к переменной f, немедленно получаем цепочку утверждений

$$y = i,\ U' = i,\ \frac{z'}{z} = i,\ \ln z = it + C_2,\ z = f' = e^{it+C_2},$$

и f — х.ф. лишь вырожденного распределения.

Итак, при $C_1 \neq 0$, исходя из (19), рассматриваем (см. (15))

$$y = U' = \frac{e^{-it}}{C_1 - ie^{-it}}. \qquad (20)$$

Проинтегрируем последнее уравнение:

$$dU = \frac{e^{-it}}{C_1 - ie^{-it}}dt,$$

$$U = \int \frac{e^{-it}dt}{C_1 - ie^{-it}} = -\ln(C_1 - ie^{-it}) + \ln C_2.$$

Таким образом,

$$U = \ln\frac{C_2}{C_1 - ie^{-it}},\ C_2 \neq 0. \qquad (21)$$

Далее, исходя из начальных условий, выразим постоянные C_1 и C_2 через моменты μ и σ^2.

Поскольку (см. (15), (13), (12))

$$y = U' = \frac{z'}{z} = \frac{f''}{f'},$$

то, согласно (11)

$$y(0) = \frac{f''(0)}{f'(0)} = -\frac{\mu^2 + \sigma^2}{i\mu}.$$

Используя (20), приходим к равенству

$$y(0) = \frac{1}{C_1 - i} = -\frac{\mu^2 + \sigma^2}{i\mu},$$

откуда следует

$$C_1 = i - \frac{i\mu}{\mu^2 + \sigma^2}. \qquad (22)$$

Из (13) и (21) получаем

$$U = \ln z = \ln\frac{C_2}{C_1 - ie^{-it}},$$

следовательно,

$$z = \frac{C_2}{C_1 - ie^{-it}}. \qquad (23)$$

По (12) и начальному условию (11) находим из (23)

$$z(0) = f'(0) = i\mu = \frac{C_2}{C_1 - i}.$$

Отсюда и из (22) выражаем C_2:

$$C_2 = i\mu(C_1 - i) = \frac{\mu^2}{\mu^2 + \sigma^2}. \qquad (24)$$

Напомним, что $C_2 \neq 0$, следовательно, согласно (24) и $\mu \neq 0$.

Возвращаясь к (23) и замене (12), интегрируем уравнение далее

$$z = f' = \frac{C_2}{C_1 - ie^{-it}},$$

$$f = C_2\int\frac{dt}{C_1 - ie^{-it}} = C_2\int\frac{d(it)}{C_1 i + e^{-it}}.$$

Произведем замену переменной интегрирования

$$C_1 i + e^{-it} = \omega,$$

тогда

$$f = -C_2\int\frac{d\omega}{\omega(\omega - C_1 i)} = -C_2\left(\frac{1}{C_1 i}\ln(\omega - C_1 i) - \frac{1}{C_1 i}\ln\omega + C_3\right).$$

Следовательно, окончательно получаем решение

$$f(t) = \frac{C_2}{C_1 i}\ln(1 + iC_1 e^{it}) - C_2 C_3, \qquad (25)$$

где C_3 — новая произвольная постоянная.

Рассмотрим два возможных случая: $C_3 = 0$ и $C_3 \neq 0$.

При $C_3 = 0$, полагая

$$iC_1 = -q,\ \frac{iC_1}{C_2} = \ln p,$$

приходим к х.ф. логарифмического распределения вида (1):

$$f(t) = \frac{1}{\ln p}\ln(1 - qe^{it}).$$

При этом, зная C_1 и C_2 (см. (22), (24)), необходимо обеспечить выполнение соответствующих ограничений: $0 < p < 1$, $p + q = 1$. Так как

$$q = -i\,C_1 = \frac{\mu^2 + \sigma^2 - \mu}{\mu^2 + \sigma^2} = 1 - \frac{\mu}{\mu^2 + \sigma^2},\ q = 1 - p,$$

то полагаем

$$p = \frac{\mu}{\mu^2 + \sigma^2} \tag{26}$$

и предполагаем, что $0 < p < 1$, т. е.

$$\mu > 0,\ \frac{\mu}{\mu^2 + \sigma^2} < 1. \tag{27}$$

Условие, что

$$\ln p = \frac{iC_1}{C_2} = \frac{\mu - \mu^2 - \sigma^2}{\mu^2}, \tag{28}$$

можно проверить, исходя из известной зависимости моментов μ и σ^2 от параметра p (см. (2)):

$$\mu = -\frac{1-p}{p\ln p},\ \sigma^2 = -\frac{1-p}{p^2\ln p}\left[1 + \frac{1-p}{\ln p}\right].$$

Тогда

$$\frac{\mu - \mu^2 - \sigma^2}{\mu^2} = \frac{1}{\mu} - 1 - \frac{\sigma^2}{\mu^2} = -\frac{p\ln p}{1-p} - 1 +$$

$$+\frac{1-p}{p^2\ln p}\left[1 + \frac{1-p}{\ln p}\right]\frac{p^2\ln^2 p}{(1-p)^2} = -\frac{p\ln p}{1-p} - 1 + \left[1 + \frac{1-p}{\ln p}\right]\frac{\ln p}{1-p} =$$

$$= \frac{1}{1-p}[-p\ln p - 1 + p + \ln p + 1 - p] = \frac{1}{1-p}(-p\ln p + \ln p) = \ln p.$$

Таким образом, при условиях на моменты (27) имеет место характеризация логарифмического распределения.

Далее рассмотрим решение (25) при $C_3 \neq 0$. Выразим значение C_3 через моменты, используя начальные условия (11):

$$f(0) = 1 = \frac{C_2}{C_1\, i}\ln(1 + iC_1) - C_2C_3,$$

$$C_3 = \frac{1}{C_2}\left[\frac{C_2}{C_1\, i}\ln(1 + iC_1) - 1\right] =$$

$$= \frac{\mu^2 + \sigma^2}{\mu - \mu^2 - \sigma^2}\ln\frac{\mu}{\mu^2 + \sigma^2} - \frac{\mu^2 + \sigma^2}{\mu^2}.$$

Подставив полученный коэффициент C_3 в (25), приходим к формуле

$$f(t) = \frac{\mu^2}{\mu - \mu^2 - \sigma^2}\ln\left(1 + \frac{\mu - \mu^2 - \sigma^2}{\mu^2 + \sigma^2}e^{it}\right) -$$

$$-\left(\frac{\mu^2}{\mu - \mu^2 - \sigma^2}\ln\frac{\mu}{\mu^2 + \sigma^2} - 1\right),$$

т. е.

$$f(t) = \frac{\mu^2}{\mu - \mu^2 - \sigma^2}\ln\left(\frac{\mu^2 + \sigma^2}{\mu} + \frac{\mu - \mu^2 - \sigma^2}{\mu}e^{it}\right) + 1.$$

Учитывая соотношения (26), (28), выразим $f(t)$ через параметр p:

$$f(t) = \frac{1}{\ln p}\ln\left(\frac{1}{p} + (1 - \frac{1}{p})e^{it}\right) + 1 = 1 + \frac{1}{\ln p}\ln\left(\frac{1 + (p-1)e^{it}}{p}\right).$$

Следовательно,

$$f(t) = 1 + \frac{1}{\ln p}(\ln(1 - qe^{it}) - \ln p) = \frac{1}{\ln p}\ln(1 - qe^{it}).$$

При $p \le 0$ или $p \ge 1$ функция $f(t)$ вида (1) не является характеристической. Однако в этих случаях характеристическая функция вырожденного закона, $f(t) = e^{i\mu t}$, удовлетворяет дифференциальному уравнению (10). Следовательно, при соотношениях (6) или (7) условие нулевой регрессии (5) характеризует лишь вырожденное распределение.

Теорема доказана.

Характеризация логарифмического распределения условием постоянства регрессии полиномиальной статистики на линейную статистику дополняет еще одним классическим распределением множество известных распределений, обладающих подобным свойством. Отметим, что сам порядок статистик, представляющих характеризации, указывает на некие внутренние свойства распределений, а это позволяет ввести новую систематизацию с установлением ранее не известных связей и зависимостей между конкретными распределениями.

1. Справочник по теории вероятностей и математической статистике / В.С.Королюк, Н.И.Портенко, А.В.Скороход, А.Ф.Турбин. М.: Наука, 1985. 640 с.
2. Lukacs E. // Ann. Math. Statist. 1942. V.13. №1. P.91.
3. Lukacs E. Characterization problems for discrete distributions // Proc. Internat. Symp. on classical and contaqions distributions. Montreal, 1963. P.65-74.
4. Клебанов Л.Б. // Теория вероятности и ее применение. 1979. Т. XXIV. №3. С.646-648.
5. Каган А.М., Линник Ю.В., Рао С.Р. Характеризационные задачи математической статистики. М.: Наука, 1972. 656 с.
6. Токмачёв М.С. // Деп. в ВИНИТИ №1575-В94 от 24.06.94. 10 с.
7. Токмачёв М.С. // Деп. в ВИНИТИ №1542-В94 от 21.06.94. 11 с.
8. Токмачёв М.С. // Вестник НовГУ. Сер.: Естеств. и техн. науки. 1995. № 1. С.139-141.
9. Токмачёв М.С. // Вестник НовГУ. Сер.: Естеств. и техн. науки. 1996. № 3. С.93-96.
10. Токмачёв М.С. // Деп. в ВИНИТИ №61-В97 от 08.01.97. 13 с.