В.В.Нешитой, *доктор технических наук*

ВЫРАВНИВАНИЕ СТАТИСТИЧЕСКИХ РАСПРЕДЕЛЕНИЙ ОБОБЩЕННЫМИ ПЛОТНОСТЯМИ

Проблема установления закона распределения случайной величины по выборочным данным к настоящему времени до конца не решена. К сожалению, основным методом ее решения все еще считается выдвижение гипотез о выравнивающем (аппроксимирующем) распределении и проверка каждой из них по критериям согласия. Такой метод не дает однозначного решения.

Для надежного установления наилучшего распределения необходимо использовать универсальные (обобщенные) распределения, включающие как частные случаи практически все известные распределения.

Три системы непрерывных распределений. Для наиболее точного описания статистических рядов распределения предлагается использовать три системы непрерывных распределений автора, заданные четырехпараметрическими плотностями [1]:

$$p(x) = Ne^{\gamma x}\left(1 - \alpha u e^{\beta x}\right)^{\frac{1}{u}-1} \tag{1}$$

$$p(t) = Nt^{\gamma-1}\left(1 - \alpha u t^{\beta}\right)^{\frac{1}{u}-1} \tag{2}$$

$$p(y) = \frac{N}{y}(\ln y)^{\gamma-1}\left[1 - \alpha u (\ln y)^{\beta}\right]^{\frac{1}{u}-1}, \tag{3}$$

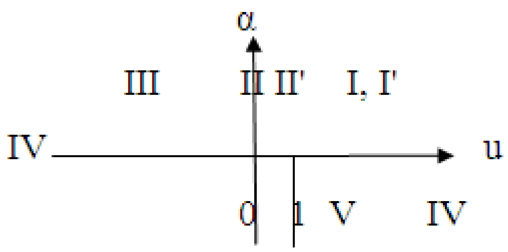
где N – нормирующий множитель; α, β, γ, u – параметры. Плотности (2) и (3) получаются из плотности (1) соответственно при $X = \ln T$, $X = \ln\ln Y$.

В зависимости от значений параметров u, α, а также знака параметров β, γ все распределения каждой из трех систем разделяются на типы (см. рис).

Введем в плотность (2) дополнительный параметр сдвига l и перепишем ее в виде

$$p(t) = M(t-l)^{\gamma-1}\left[1 - \alpha u (t-l)^{\beta}\right]^{\frac{1}{u}-1}. \tag{4}$$

На основании плотности (4) можно получить три дополнительные системы непрерывных распределений.



Классификация распределений (типы со штрихом – при $\beta, \gamma < 0$).

Первая дополнительная система в общем случае задается формулой (4) при $|\beta| = 1$, а в случае симметричных распределений – при $\beta=2, \gamma=1$.

Вторая дополнительная система непрерывных распределений получается из первой при $T = lnY$ и прежних значениях параметров β, γ, а третья – при $y=lnw$.

Три системы непрерывных распределений, заданные плотностями (1) – (3), можно расширить за счет включения в них распределений типов 1.1 и 2.1, которые относятся к дополнительным системам непрерывных распределений (с параметром $\beta=1$). Для примера рассмотрим первую и вторую системы непрерывных распределений.

В общем случае первая система будет включать три обобщенные плотности

$$\left.\begin{aligned} p(x) &= Ne^{k\beta x}\left(1-\alpha u e^{\beta x}\right)^{\frac{1}{u}-1} \\ p(t) &= N(t-l)^{k-1}\left[1-\alpha u(t-l)\right]^{\frac{1}{u}-1} \\ p(t) &= N\left[1-\alpha u(t-\bar{t})^2\right]^{\frac{1}{u}-1} \end{aligned}\right\}, \qquad (5)$$

где $k=\gamma/\beta$. Сюда входит как частный случай нормальный закон при $u \to 0$.

Распределения второй системы можно получить из первой как распределения функций случайных аргументов: $X=lnT$ – для первой плотности; $T=lnY$ – для двух других плотностей. Третья плотность включает логарифмически нормальный, или логнормальный закон при $u \to 0$.

Выбор системы непрерывных распределений для вычисления аппроксимирующей кривой. Наличие трех систем непрерывных распределений, а также знание свойств этих распределений освобождает исследователя от выдвижения многочисленных гипотез об аппроксимирующем распределении.

Каждая из трех систем непрерывных распределений предназначена для описания преимущественно своего класса статистических распределений.

Так, первая система непрерывных распределений в соответствии с ее характерными особенностями наряду с другими должна описывать статистические распределения таких случайных величин, последующие значения которых получаются из предыдущих путем их изменения (сдвига) на всех интервалах на постоянную величину *C* без изменения частот интервалов. Эти распределения содержат параметры сдвига – $\alpha, l, \bar{t}$, при изменении которых выравнивающая кривая перемещается по горизонтальной оси без изменения формы.

Кроме того, если случайная величина растет во времени по линейному закону, либо она задана на всей числовой оси, то ее распределение должно описываться первой системой непрерывных распределений.

Аналогично вторая система непрерывных распределений должна описывать статистические распределения таких *неотрицательных* случайных величин, последующие значения которых на всех интервалах получаются из предыдущих путем их умножения на всех интервалах на постоянную *C* без изменения частот интервалов (при этом ширина интервалов и их границы увеличиваются в *C* раз).

Кроме того, если случайная величина растет во времени по показательному закону, то ее распределение должно описываться второй системой непрерывных распределений.

Вычисление аппроксимирующих распределений по устойчивому методу. Метод оценивания параметров аппроксимирующих распределений является устойчивым, если он мало чувствителен к выбросам на концах статистического распределения. Такой метод, разработанный автором [3], излагается ниже.

Рассмотрим обобщенную плотность (1), которая задает первую основную систему непрерывных распределений. Введем два показателя – асимметрии *B* и островершинности *H*, которые зависят от двух параметров формы $k=\gamma/\beta$, *u*. По этим показателям устанавливается тип аппроксимирующего распределения и находятся оценки параметров *k, u*. Оценки двух других параметров рассчитываются по простым формулам. Заметим, что этот метод требует предварительного группирования статистических данных.

Для обобщенной плотности $p(x)$ показатели *B, H* равны

$$\left.\begin{aligned} B &= M\left[p(x)(x - M(x))\right] = f(k,u) \\ H &= S_3 / S_1^3 = f(k,u) \end{aligned}\right\}, \tag{6}$$

где

$$S_\gamma = M\left[p(x)\right]^\gamma = f(\beta,k,u).$$

Исследования показали, что величина *H* задана на интервале $\sqrt{2} < H < 2$, а величина *B* – на интервале $-1/4 < B < 1/4$.

Вычислим для разных типов распределений значения показателей *B, H* при различных значениях параметров *k, u*. Далее построим номограмму [1]. Она применима для трех основных систем непрерывных распределений, заданных плотностями (1) – (3). При этом плотности (2) и (3) должны быть приведены к форме плотности (1), т.е. представлены в виде *tp(t)=f(lnt), ylnyp(y)=f(lnlny).*

Показатели *B, H* однозначно определяют тип распределения, приведенного к форме плотности *p(x).* Более того, с помощью этих показателей могут быть найдены оценки параметров *k, u* непосредственно из номограммы. Далее находится оценка параметра β, которая для всех типов равна

$$\beta = \frac{S_1}{S_1^{(\cdot)}}. \tag{7}$$

Тогда $\gamma = k\beta$. Оценки параметра α для распределений II, II′ типов и произведения αu для остальных типов рассчитываются по формулам:

$$I - I':\ \alpha u = e^{\pm(\nu_1^{(z)} - \beta\nu_1)};\ II - II':\ \alpha = e^{\pm(\nu_1^{(z)} - \beta\nu_1)};\ III - V:\ \alpha u = -e^{\nu_1^{(z)} - \beta\nu_1}.$$

В зависимости от типа распределения величины $\nu_1^{(z)}$ и $S_1^{(z)}$ рассчитываются по формулам:

$$I - I': \nu_1^{(z)} = \pm\left[\psi(k) - \psi(k + 1/u)\right] \quad S_1^{(z)} = \frac{1}{2\sqrt{\pi}} \frac{2(k+1/u)-1}{2/u-1} \cdot \frac{g(k)g(1/u)}{g(k+1/u)}$$

$$II - II': \nu_1^{(z)} = \pm\psi(k), \quad S_1^{(z)} = g(k)/2\sqrt{\pi};$$

$$III - V: \nu_1^{(z)} = \psi(k) - \psi(1 - 1/u - k), \quad S_1^{(z)} = \frac{1}{2\sqrt{\pi}} \frac{g(k)g(1-1/u-k)}{g(1-1/u)}.$$

Величины $g(x) = \Gamma(x+1/2)/\Gamma(x)$ и $\psi(x) = d(\ln\Gamma(x))/dx$ могут быть вычислены по приближенным формулам:

$$g(x) \approx \frac{x(x+0{,}875)}{(x+0{,}5)\sqrt{x+1}};$$

$$\psi(x) \approx -\left(\frac{1}{x} + \frac{1}{x+1}\right) + \ln(x+2) - \frac{1}{2(x+2)} - \frac{1}{12(x+2)^\wedge 2}.$$

Для установления типа аппроксимирующего распределения и нахождения оценок параметров по устойчивому методу достаточно найти значения статистических показателей ν_1^*, S_1^*, B^*, H^* и приравнять их соответствующим теоретическим. Эти показатели для каждой системы непрерывных распределений вычисляются по-своему. Но номограмма применима ко всем трем системам непрерывных распределений.

Оценки статистических показателей в случае аппроксимирующих распределений, заданных плотностью $p(x)$ (см. формулу (1)), вычисляются по формулам:

$$\left.\begin{aligned} &\nu_1^* = \bar{x} = \sum_{i=1}^{n} x_i p_i h_i \\ &S_1^* = \sum_{i=1}^{n} p_i^2 h_i,\ S_3^* = \sum_{i=1}^{n} p_i^4 h_i \\ &B_1^* = \sum_{i=1}^{n} x_i p_i^2 h_i - \nu_1^* S_1^*;\ \mathrm{H}^* = \frac{S_3^*}{(S_1^*)^3} \end{aligned}\right\},$$

где $p_i = m_i/(Mh_i)$ – эмпирическая плотность распределения; m_i –

наблюденная частота случайной величины X в i-ом интервале $(i = 1,2..., n); M = \sum_{i=1}^{n} m_i$ – наблюденная частота во всех n интервалах (объем выборки); h_i – ширина i-го интервала; x_i – значение случайной величины X в середине i-го интервала.

В случае плотности $p(t)$ статистические показатели рассчитываются по формулам:

$$\left.\begin{aligned}
&\nu_1^* = \overline{\ln t} = \sum_{i=1}^{n} \ln t_i p_i h_i; \quad S_1^* = \sum_{i=1}^{n} t_i p_i^2 h_i \\
&S_3^* = \sum_{i=1}^{n} t_i^3 p_i^4 h_i; \; H^* = \frac{S_3^*}{(S_1^*)^3} \\
&B^* = \sum_{i=1}^{n} t_i \ln t_i p_i^2 h_i - \nu_1^* S_1^*
\end{aligned}\right\}.$$

После вычисления эмпирических показателей B^*, H^* их следует приравнять теоретическим, с помощью номограммы установить тип аппроксимирующего распределения и найти оценки параметров *k, u* (по точке с координатами *B, H*). При известных оценках параметров *k, u* по приведенным выше формулам легко вычисляются оценки параметров *β, α* или произведения *αu*. Далее остается вычислить нормирующий множитель, который в зависимости от типа распределения задается формулами (в случае распределений (1) – (3)):

Типы I, I′: $N = \dfrac{\beta(\alpha u)^k \Gamma(k+1/u)}{\Gamma(k)\Gamma(1/u)}$; Типы II, II′: $N = \dfrac{\beta\alpha^k}{\Gamma(k)}$;

Типы III–V: $N = \dfrac{\beta(-\alpha u)^k \Gamma(1-1/u)}{\Gamma(k)\Gamma(1-1/u-k)}$.

Нормирующий множитель вначале целесообразно прологарифмировать, вычислить lnN, а затем найти значение $N = e^{\ln N}$. Логарифм гамма-функции можно вычислить по приближенной формуле

$$\ln \Gamma(x) \approx 0{,}91894 - \ln[x(x+1)] + (x+1{,}5)\ln(x+2) - (x+2) + \frac{1}{12(x+2)}.$$

Подставив найденные оценки параметров в соответствующее распределение, можно рассчитать теоретические значения плотности при заданных значениях случайной величины.

---

1. Нешитой, В.В. Элементы теории обобщенных распределений: монография / В.В.Нешитой. – Мн.: РИВШ, 2009. – 203 с.

2. Нешитой, В.В. Универсальные законы рассеяния и старения публикаций / В.В.Нешитой // Веснік Бел. дзярж. ун-та культ. і маст. – 2007. – № 8. – С. 128–133.

3. Нешитой, В.В. Оценивание параметров обобщенных распределений / В.В.Нешитой; БелНИИНТИ. – Мн., 1983. – 96 с. – Деп. в БелНИИНТИ 11.03.84, № 857.