УДК 62–192:519.2

ОБ АППРОКСИМАЦИИ СТАТИСТИЧЕСКИХ РАСПРЕДЕЛЕНИЙ ОБОБЩЕННЫМИ ПЛОТНОСТЯМИ

В.В. НЕШИТОЙ

При аппроксимации статистических распределений, характеризующих, например, прочность или долговечность изделий, обычно используют известные непрерывные распределения, в том числе Вейбулла, нормальное или логарифмически нормальное и др. Далее с помощью критериев согласия выбирают наиболее подходящую аппроксимирующую кривую.

Подбор таких кривых можно значительно облегчить, если вместо отдельных распределений использовать обобщенные плотности, включающие как частные случаи широко известные непрерывные распределения.

Рассмотрим три вида распределений [1,2], заданных плотностями

$$p(t) = N t^{\gamma-1} (1 - \alpha u t^{\beta})^{1/u-1}; \tag{1}$$

$$p(x) = N \exp(\gamma x) (1 - \alpha u \exp(\beta x))^{1/u-1}; \tag{2}$$

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

где N – нормирующий множитель; α, β, γ, u – параметры.

В зависимости от значений u, α распределения каждого из трех видов можно разделить на несколько типов (рис. 1). Для распределений типов $I-V$ параметры β, γ положительны, для типов I', II' – отрицательны.

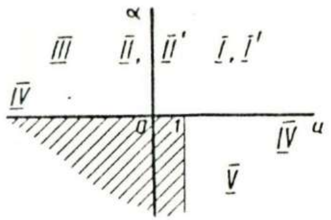

Рис. 1

Распределения, заданные плотностью вида (1), могут иметь правостороннюю или левостороннюю асимметрию, а также быть симметричными (при $\beta = 2$; $\gamma = 1$).

Среди распределений типов $III - V$, заданных плотностью вида (2), также имеются симметричные. Параметры их удовлетворяют условию $\gamma/\beta = k = (1 - 1/u)/2$. При $(1 - 1/u)/2 < k < 1 - 1/u$ эти распределения имеют правостороннюю асимметрию, при $0 < k < (1 - 1/u)/2$ – левостороннюю.

Распределения, заданные плотностью вида (3), имеют правостороннюю асимметрию.

Нормирующий множитель N в зависимости от типа распределения задается формулами:

$$N_{I,\,I'} = \beta\ (\alpha\, u)^k\, \Gamma(k + 1/u) / (\Gamma(k)\ \Gamma(1/u))\,; \tag{4}$$

$$N_{II,\,II'} = \beta \alpha^k / \Gamma(k)\,; \tag{5}$$

$$N_{III-V} = \beta\,(-\alpha\, u)^k\, \Gamma(1 - 1/u) / (\Gamma(k)\ \Gamma(1 - 1/u - k))\,, \tag{6}$$

где $\Gamma(\cdot)$ – гамма-функция.

Для симметричных распределений нормирующий множитель уменьшается вдвое, а отношение $k = 1/2$.

Из (4 –6) можно определить интервалы изменения параметров для различных типов распределений:

$0 < u < \infty$; $\alpha > 0$; $0 < k < \infty$ – для типов I, I';

$u \to 0$; $\alpha > 0$; $0 < k < \infty$ – для типов II, II';

$-\infty < u < 0$; $\alpha > 0$; $0 < k < 1 - 1/u$ – для типа III;

$u \to \mp\infty$; $\alpha \to \pm 0$; $0 < k < 1$ – для типа IV;

$1 < u < \infty$; $\alpha < 0$; $0 < k < 1 - 1/u$ – для типа V.

Можно показать, что как частный случай из (1) при $\beta = 1$ следует система кривых Пирсона [3].

Для видов распределений (1 – 3) автором разработан единый метод оценивания их параметров и установления типа аппроксимирующей кривой по показателям асимметрии B и островершинности H [2]. Суть его заключается в том, что указанные показатели являются функциями параметров формы u и k, оценки которых находятся по эмпирическим значениям B и H. В случае распределений, заданных (2),

$$B = M\,[p\,(x)\,(x - \nu_1)\,]\,;\quad H = S_3/S_1^3\,, \tag{7}$$

где ν – центр распределения (начальный момент первого порядка); M – символ математического ожидания; $S_1 = M\,[p\,(x)\,]\,;\ S_3 = M\,[p\,(x)\,]^3$.

Для нахождения оценок параметров распределений, заданных плотностями (1, 3), их надо привести к виду (2). При этом положим

$$x_1 = \ln t,\ p_1\,(x) = t\,p\,(t)\,;\ x_2 = \ln \ln y,\ p_2\,(x) = y\,p\,(y)\,\ln y, \tag{8}$$

для (1) и (3). В этом случае в формулы (7) вместо x, $p\,(x)$ необходимо подставить их значения из (8).

Оценки параметров β, α (или произведения $\alpha\,u$) находятся по эмпирическим величинам S_1^*, ν_1^*:

для всех типов

$$\beta = S_1^*\,/S_1'\,; \tag{9}$$

для типов I, I'

$$\alpha\,u = \exp\,(\pm\,(\nu_1' - \beta\,\nu^*))\,;$$

для типов II, II'

$$\alpha = \exp\,(\pm\,(\nu_1' - \beta\,\nu^*))\,;$$

для типов $III - V$

$$\alpha\,u = -\exp\,(\nu_1' - \beta\,\nu^*)\,. \tag{10}$$

В этих формулах и ниже при двойном знаке „плюс" и „минус" относятся к типам I, II и I', II', соответственно.

Величины S_1', ν_1' вычисляются в зависимости от типа кривой. Для распределений типов I, I'

$$S_1' = \left(\frac{\Gamma\,(k + 1/u)}{\Gamma\,(k)\,\Gamma\,(1/u)}\right)^2 \frac{\Gamma\,(2k)\,\Gamma\,(2/u - 1)}{\Gamma\,(2\,(k + 1/u) - 1)} - \frac{(2\,(k + 1/u) - 1)\,g\,(k)\,g\,(1/u)}{2\sqrt{\pi}\,(2/u - 1)\,g\,(k + 1/u)}\,;$$

$$\nu_1' = \pm\,(\Psi\,(k) - \Psi\,(k + 1/u))\,,$$

где $g\,(k) = \Gamma\,(k + 1/2)\,/\Gamma\,(k)$; $\Psi\,(k) = d\,\ln \Gamma\,(k)\,/dk$ – логарифмическая производная гамма-функции.

Для распределений типов *II, II'* $S_1' = \Gamma(2k)/[2^{2k}(\Gamma(k))^2] = g'(k)/2\sqrt{\pi}$; $\nu' = \pm\Psi(k)$.

Для распределений типов *III – V*

$$S_1' = \left(\frac{\Gamma(1-1/u)}{\Gamma(k)\ \Gamma(1-1/u-k)}\right)^2 \frac{\Gamma(2k)\,\Gamma(2(1-1/u-k))}{\Gamma(2(1-1/u))} =$$

$$= \frac{g(k)\ g(1-1/u-k)}{2\sqrt{\pi}\,g(1-1/u)}; \tag{11}$$

$$\nu' = \Psi(k) - \Psi(1-1/u-k).$$

Величины $\Psi(k)$, $\ln\Gamma(k)$ и, следовательно, $\Gamma(k)$ с любой степенью точности могут быть вычислены по формулам, которые следуют из ряда Стирлинга:

$$\Psi(k) \approx -\sum_{s=k}^{k+n}(1/s) + \ln(k+n) + \frac{1}{2(k+n)} - \frac{1}{12(k+n)^2} +$$

$$+ \frac{1}{120(k+n)^4} - \ldots;$$

$$\ln\Gamma(k) = \frac{1}{2}\ln 2\pi - \sum_{s=k}^{k+n}\ln s + \left(k+n+\frac{1}{2}\right)\ln(k+n) - (k+n) +$$

$$+ \frac{1}{12(k+n)} - \frac{1}{360(k+n)^3} + \ldots,$$

где n – целое. При $k+n \geqslant 5$ эти формулы дают погрешность, по абсолютной величине не превышающую значения $2{,}54 \cdot 10^{-7}$. Величина $g(k) = k(k+0{,}875)/((k+0{,}5)(k+1)^{3/2})$.

На рис. 2 представлены показатели приведенных к виду (2) обобщенных распределений с левосторонней асимметрией при $0 < B < 0{,}25$ и симметричных. Распределения с правосторонней асимметрией характеризуются значениями $-0{,}25 < B < 0$. Если при этом они относятся к типам *III – V*, то из рис. 2 вначале находятся оценки параметров k', u по $|B|$, а затем вычисляется величина $k = 1 - 1/u - k'$. В остальных случаях, включая распределения типов *I', II'*, оценки k, u находятся непосредственно из рис. 2, по $|B|$.

Рассмотрим аппроксимацию статистического распределения на примере выборки из $N = 160$ электрических ламп. Задано распределение срока службы ламп [4, с. 128] значениями T середин временны́х интервалов (при ширине каждого $\Delta t = 200$ ч) и частотой их появления m_i (табл. 1).

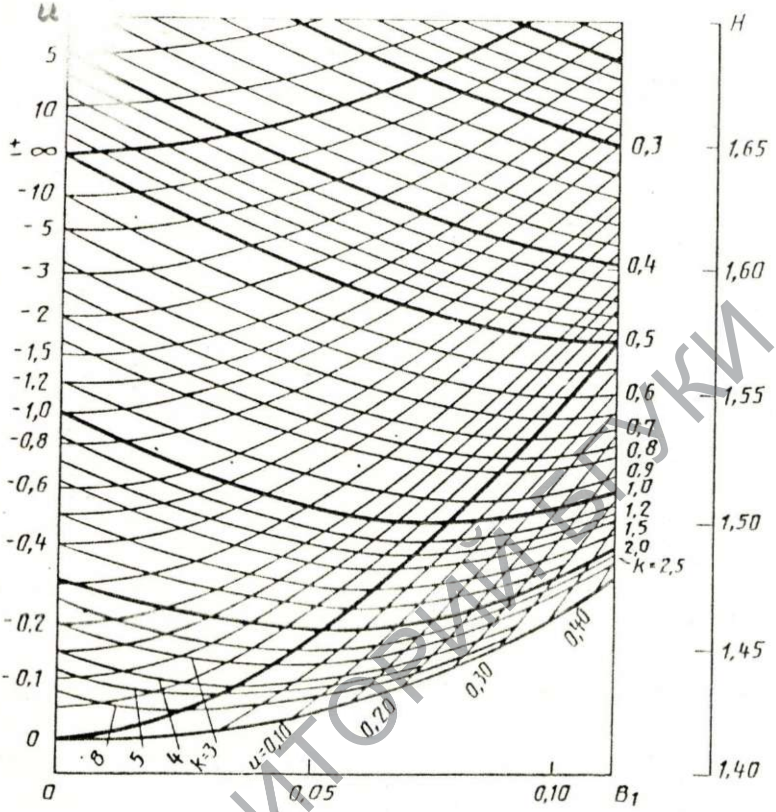

Рис. 2

В качестве аппроксимирующего возьмем обобщенное распределение, заданное плотностью вида (1). Примем ширину интервала Δt за единицу и внесем в таблицу новые значения t_i середин интервалов.

Вычисление статистических показателей ν^*, S_1^*, B^*, H^*, S_3^* произведем по формулам, которые являются эмпирическими аналогами (7), с учетом (8):

$$\nu^* = \sum_{i=1}^{n} \ln t_i (m_i/M)\ \Delta t;\quad S_1^* = \sum_{i=1}^{n} t_i (m_i/M)^2\ \Delta t;\quad S_3^* = \sum_{i=1}^{n} t_i^3\ (m_i/M)^4\ \Delta t;$$

$$H^* = S_3^*/(S_1^*)^3;\quad B^* = \sum_{i=1}^{n} t_i \ln t_i (m_i/M)^2\ \Delta t - \nu^* S_1^*;\quad M = \sum_{i=1}^{n} m_i.$$

В результате расчетов получим: $\nu^* = 1{,}882576$; $S_1^* = 0{,}698672$; $S_3^* = 0{,}502937$; $H^* = 1{,}474666$; $B^* = 0{,}035043$. Приравнивая значения показателей H^* и B^* статистического распределения соответствующим значе-

Таблица 1

T, ч	m_i	t_i	m_i/M	$p(t_i)$
100	1	0,5	0,0062	0,00040
300	3	1,5	0,0187	0,00871
500	2	2,5	0,0125	0,03370
700	10	3,5	0,0625	0,07358
900	10	4,5	0,1250	0,11570
1100	24	5,5	0,1500	0,14417
1300	25	6,5	0,1562	0,15006
1500	20	7,5	0,1250	0,13532
1700	18	8,5	0,1125	0,10887
1900	11	9,5	0,0687	0,08010
2100	8	10,5	0,0500	0,05506
2300	5	11,5	0,0312	0,03600
2500	5	12,5	0,0312	0,02271
2700	4	13,5	0,0250	0,01399
2900	2	14,5	0,0125	0,00850
3100	1	15,5	0,0062	0,00513
3300	1	16,5	0,0062	0,00309

ниям этих показателей теоретического распределения, с помощью номограммы для установления типа аппроксимирующей кривой и оценивания параметров k, u, которая изображена на рис. 2, находим $u \approx -0{,}2$, $k \approx 1{,}5$. Аппроксимирующее распределение относится к типу III, для которого $-\infty < u < 0$. Далее по формулам (6, 9 – 11) вычисляем $S' = 0{,}2738$; $\nu' = -1{,}3524$; $\beta = 2{,}555$, $a\,u = -0{,}0021$; $N = 1/349{,}4$.

Аппроксимирующее распределение типа III задается плотностью (1), которую целесообразно представить в виде

$$p(t) = \frac{\beta(-a\,u)^{k}\,\Gamma(1-1/u)}{\Gamma(k)\,\Gamma(1-1/u-k)}\;\frac{t^{k\beta-1}}{(1+(t/(-1/a\,u)^{1/\beta})^{\beta})^{1-1/u}}, \qquad (12)$$

откуда с учетом найденных оценок параметров

$$p(t) = t^{2{,}8325}/\{349{,}4\,[1+(t/11{,}170)^{2{,}555}]^{6}\}. \qquad (13)$$

В табл. 1 приведены значения плотности вероятностей, вычисленные по формуле (13), а также относительные частоты m_i/M. Несмотря на малый объем выборки, значения относительных частот и плотности вероятностей в серединах интервалов оказались близкими между собой.

Отметим, что величина S_1^* может служить показателем качества изделий. Действительно, чем больше срок службы ламп и чем большая доля их имеет такой срок службы, тем качество выше.

В табл. 2 приведены примеры аппроксимирующих распределений для различных рядов наблюдений, характеризующих прочность изделий. Статистические данные, взятые из [3, с. 131, 269, 291; 4, с. 199], получены на выборках достаточно больших объемов ($400 \leqslant N \leqslant 1000$). Из табл. 2 видно, что все аппроксимирующие распределения относятся к типу III.

Таким образом, распределения, характеризующие прочность и долговечность изделий (по крайней мере в рассмотренных примерах) могут быть описаны плотностью (12).

Таблица 2

Характеристика	Параметры распределений						N
	H^*	B^*	k	u	β	$(-\frac{1}{a u})^{1/\beta}$	
Хрупкость	1,5294	0,0513	0,8	−0,50	4,8686	9,8775	1/985,67
Модуль упругости	1,5217	0,0752	0,8	−0,15	4,6557	9,3828	1/226,54
Предел прочности при сжатии	1,5585	0,0709	0,6	−0,50	5,4368	7,4367	1/118,38
То же при растяжении	1,4951	0,0706	1,1	−0,05	3,0087	15,7120	1/83,85
Число циклов нагружения	1,5935	−0,0006	0,7	−2,50	3,7690	2,3835	1/4,98

В заключение отметим, что описанный метод оценивания параметров аппроксимирующих распределений требует исследования точности даваемых оценок. Это позволит установить его место в ряду других методов (моментов, наибольшего правдоподобия), а также обосновать использование в практических расчетах. Однако решение этих вопросов выходит за рамки настоящей статьи.

СПИСОК ИСПОЛЬЗОВАННОЙ ЛИТЕРАТУРЫ

1. Нешитой В.В. Построение системы непрерывных распределений/БелНИИНТИ. — Минск, 1979. — Деп. в БелНИИНТИ 30.07.80, № 174.
2. Нешитой В.В. Оценивание параметров обобщенных распределений/БелНИИНТИ. — Минск, 1983. — Деп. в БелНИИНТИ 11.03.84, № 857.
3. Митропольский А.К. Техника статистических вычислений. — М.: Наука, 1971.
4. Смирнов Н.В., Дунин-Барковский И.В. Курс теории вероятностей и математической статистики для технических приложений. — М.: Наука, 1969.

Статья поступила 08.08.86