УДК 519.213

КАК ВЫЧИСЛИТЬ ЗАКОН РАСПРЕДЕЛЕНИЯ СЛУЧАЙНОЙ ВЕЛИЧИНЫ?

Нешитой В.В.

Белорусский государственный университет культуры и искусств, г. Минск

В настоящее время подбор теоретического закона распределения для аппроксимации статистического ряда осуществляется путем выдвижения гипотез и проверки каждой из них по критериям согласия. Такой метод не дает однозначного решения даже при выдвижении нескольких гипотез подряд. Спрашивается, можно ли непосредственно по статистическому ряду вычислить закон распределения? Оказывается, можно, если предварительно решить несколько проблем:

– дать классификацию непрерывных случайных величин в зависимости от их свойств;

– для каждого класса случайных величин разработать свою систему непрерывных распределений, способную с высокой точностью аппроксимировать все многообразие статистических распределений своего класса;

– разработать метод вычисления закона распределения и оценок параметров по статистическому ряду, единый для всех систем непрерывных распределений. Метод должен быть устойчивым;

– довести этот метод до программной реализации.

Рассмотрим каждую из проблем.

Классификация случайных величин. Непрерывные случайные величины в зависимости от их свойств могут быть отнесены к трем основным классам.

К *первому классу* отнесем такие случайные величины, которые могут быть заданы на любом интервале оси абсцисс. Последующие их значения образуются из предыдущих путем прибавления некоторой величины С. Если на равных интервалах времени величина С постоянна, то среднее такой случайной величины растет во времени по линейному закону. Статистические распределения этих случайных величин будем описывать первой системой непрерывных распределений.

Ко *второму классу* отнесем такие случайные величины, которые заданы на положительной полуоси. Последующие их значения образуются из предыдущих путем умножения на некоторую величину С>0. Если величина С на равных интервалах времени постоянна, то среднее такой случайной величины растет во времени по показательному закону, а логарифм среднего – по линейному закону. Статистические распределения этих случайных величин будем описывать второй системой непрерывных распределений.

К *третьему классу* отнесем такие случайные величины, которые также заданы на положительной полуоси, но последующие значения образуются из предыдущих путем их возведения в некоторую степень С. Средние значения таких случайных величин могут расти во времени по двойному показательному закону. Статистические распределения этих случайных величин будем описывать третьей системой непрерывных распределений.

Из вышесказанного следует, что распределение и динамика случайных величин взаимосвязаны. По одному из указанных законов роста среднего случайной величины безошибочно определяется система непрерывных распределений. Если среднее случайной величины растет по другому закону, то ее распределение может быть описано одной из дополнительных систем непрерывных распределений.

Системы непрерывных распределений. Построим вначале вторую систему непрерывных распределений. Обратимся к уравнению прямой. Если не принимать во внимание усеченных распределений, то уравнение прямой позволяет записать три плотности и в результате их интегрирования – три функции распределения случайной величины T>0:

$$p(t)=\alpha(1-\alpha t/2);\ F(t)=1-(1-\alpha t/2)^2; \qquad (1)$$

$$p(t)=\alpha;\ F(t)=\alpha t=1-(1-\alpha t); \qquad (2)$$

$$p(t)=2\alpha t;\ F(t)=\alpha t^2=1-(1-\alpha t^2). \qquad (3)$$

Обобщим попарно функции распределения (1), (2) и (2), (3) путем введения новых параметров β, u (вместо показателей степени 1 и 2).

В первом и втором случаях соответственно получим

$$F(t)=1-(1-\alpha ut)^{1/u},\ F(t)=1-(1-\alpha t^{\beta}). \qquad (4)$$

Дальнейшее обобщение двухпараметрических функций распределения (4) дает трехпараметрическую функцию распределения

$$F(t)=1-(1-\alpha ut^{\beta})^{1/u}, \qquad (5)$$

из которой дифференцированием по t найдем плотность распределения

$$p(t)=\alpha\beta t^{\beta-1}(1-\alpha ut^{\beta})^{1/u-1}. \qquad (6)$$

Последняя плотность может быть еще более расширена за счет введения четвертого параметра k

$$p(t)=Nt^{k\beta-1}(1-\alpha ut^{\beta})^{1/u-1}. \qquad (7)$$

Обобщенная плотность (7) задает вторую систему непрерывных распределений. На ее базе при X=LnT находится обобщенная плотность для первой системы непрерывных распределений: p(x)=p(t)dt/dx, или

$$p(x)=Ne^{k\beta x}\left(1-\alpha ue^{\beta x}\right)^{1/u-1}. \qquad (8)$$

Найдем *дополнительные плотности*. Они следуют из формулы (7), в которую введен новый параметр *l*. Для первой системы непрерывных распределений дополнительные плотности имеют вид:

$$p(t) = N(t-l)^{k-1}\left[1-\alpha \mathrm{u}(\mathrm{t}-\mathrm{l})\right]^{1/u-1},\ p(t) = N\left[1-\alpha u(t-\bar{t})^2\right]^{1/u-1}.$$

В итоге каждая система распределений содержит три плотности.

Общий устойчивый метод вычисления закона распределения. Рассмотрим обобщенную плотность (8). Введем два показателя – асимметрии B и островершинности H, которые зависят от двух параметров формы k, u. По статистическим оценкам этих показателей однозначно устанавливается тип аппроксимирующего распределения и находятся оценки параметров k, u с помощью бинарной сетки (номограммы) [1,2], которая применима к трем основным системам непрерывных распределений, заданным первыми плотностями. При этом плотность p(t) приводится к форме плотности p(x), т.е. tp(t)=f(lnt). Для дополнительных систем автором построена другая номограмма, которая является продолжением первой. Оценки остальных двух параметров (α и β) вычисляются по специальным формулам. Для обобщенной плотности p(x) показатели B, H задаются формулами

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

где $S_r = M\left[p(x)\right]^r = f(\beta,\mathrm{k},\mathrm{u})$.

Величины B, H заданы на интервалах: –1/4<B<1/4; $\sqrt{2} < H < 2$.

Наличие теории обобщенных распределений [1] позволяет утверждать, что во многих случаях закон распределения заранее известен и только для вычисления оценок его параметров требуется обработка статистических данных. Решение множества теоретических задач также значительно упрощается. Отныне нет необходимости выдвигать различные гипотезы. В зависимости от свойств случайной величины выбирается система непрерывных распределений и по статистическому ряду вычисляется закон распределения.

Литература:

1. Нешитой, В.В. Элементы теории обобщенных распределений: монография / В.В.Нешитой. – Минск: РИВШ, 2009. – 204 с.

2. Нешитой, В.В. Математико-статистические методы анализа в библиотечно-информационной деятельности: учеб.-метод. пособие / В.В.Нешитой. – Минск: БГУКИ, 2009. – 203 с.