УДК 531.314.4: 311.2

Э. Г. ГЛАДКИЙ

ИСПОЛЬЗОВАНИЕ ВЕРОЯТНОСТНОГО РАСПРЕДЕЛЕНИЯ ЯКОБИ ДЛЯ АППРОКСИМАЦИИ ЭМПИРИЧЕСКИХ СТАТИСТИЧЕСКИХ РАСПРЕДЕЛЕНИЙ

Государственное предприятие «Конструкторское бюро «Южное»,
ул. Криворожская, 3, 49008, Днепр, Украина; e-mail: edgladky@gmail.com

Цель статьи продемонстрировать возможности использования вероятностного распределения Якоби для подгонки статистических совокупностей. Проанализированы универсальные системы распределений Пирсона и Джонсона, обобщенное лямбда-распределение и распределение Грама–Шарлье, которые нашли широкое распространение для подгонки статистических совокупностей. Указывается, что основной недостаток таких распределений состоит в том, что они не позволяют учесть реальные ограниченные области варьирования случайных величин. Рассмотрены теоретические вопросы построения одномерного вероятностного распределения Якоби, получаемого на основе разложения неизвестной функции плотности по системе ортогональных полиномов Якоби, имеющих вариацию на ограниченном отрезке. Сформулированы принципы оптимальности распределения Якоби при аппроксимации статистических данных и даны практические рекомендации его построения. В частности, наилучшие результаты подгонки получаются для распределения Якоби, построенного с использованием ультрасферических ортогональных полиномов Якоби. Определена область использования распределения Якоби, которая значительно шире, чем у распределения Грама–Шарлье. Приведены способы определения предельных точек распределения Якоби. На примерах продемонстрировано преимущество распределения Якоби при подгонке статических совокупностей в сравнении с используемыми на практике универсальными распределениями.

Мета статті продемонструвати можливості використання імовірнісного розподілу Якобі для підгонки статистичних сукупностей. Проаналізовано універсальні системи розподілів Пірсона та Джонсона, узагальнений лямбда розподіл і розподіл Грама–Шарл'є, які знайшли широке розповсюдження при підгонці статистичних сукупностей. Зазначається, що основний недолік таких розподілів полягає в тому, що вони не дозволяють урахувати реальні обмежені області варіювання випадкових величин. Розглянуто теоретичні питання побудови одновимірного імовірнісного розподілу Якобі, отриманого на основі розкладу невідомої функції щільності за системою ортогональних поліномів Якобі, які мають варіювання на обмеженому відрізку. Сформульовано принципи оптимальності розподілу Якобі при апроксимації статистичних даних і надано практичні рекомендації його побудови. Зокрема, найкращі результати підгонки виходять для розподілу Якобі, побудованого з використанням ультрасферичних ортогональних поліномів Якобі. Визначена область використання розподілу Якобі, яка є значно ширшою від розподілу Грама–Шарл'є. Наведено способи визначення граничних точок розподілу Якобі. На прикладах продемонстровано перевагу розподілу Якобі при підгонці статичних сукупностей в порівнянні з універсальними розподілами, що використовуються на практиці.

The paper purpose is to demonstrate the opportunities of the Jacobi probability distributions for fitting the statistical populations. The universal Pearson and Johnson systems of the distributions, the generalized lambda distribution and the Gram-Charlier distribution, which are widely used for fitting statistical populations, are analyzed. It is pointed out that the main disadvantage of these distributions is that they do not take into account real limited ranges of variations in the random variables. The paper considers the theoretical problems of the construction of the one-dimensional Jacobi probability distribution, based on the expansion the unknown density function in the term of the system of the orthogonal Jacobi polynomials with variations in a limited interval. The optimality principles of the Jacobi distribution are formulated to approximate the statistical data, and practical recommendations are given for its construction. In particular, the best fitting results are obtained for the Jacobi distribution constructed with the ultraspherical orthogonal Jacobi polynomials. The application of the Jacobi distribution is determined, which is significantly wider than the application of the Gram-Charlier distribution. Methods for determining the limited points of the Jacobi distribution are presented. Examples demonstrate the advantages of the Jacobi distribution for fitting the statistical populations in comparison with the universal distributions used in practice.

Ключевые слова*: подгонка статистических данных, распределение Якоби, распределение Грама–Шарлье, обобщенное лямбда-распределение, область изменения случайной величины.*

Анализ статистических данных – неотъемлемая часть практически любого научного исследования. При этом достаточно часто исследователь сталкивается с необходимостью подгонки статистических данных с использованием вероятностного распределения. Настоящей удачей являются случаи, когда при построении распределения для рассматриваемой выборки удается учесть механизм образования случайности. Однако в большинстве практических ситуаций приходится использовать некоторые эмпирические формы, в качестве которых широкое применение нашли семейства распределений Пирсона и Джонсона, обобщенное лямбда-распределение, распределение Грама–Шарлье. Дадим краткую характеристику каждой из указанных систем и распределений.

Анализ традиционных распределений, используемых в практике подгонки статистических данных. Распределения системы Пирсона вводятся при помощи дифференциального уравнения [4, 6]

$$\frac{df(x)}{dx} = \frac{(x-a)}{b_0 + b_1 x + b_2 x^2} f(x), \tag{1}$$

где $f(x)$ – плотность распределения непрерывной случайной величины (СВ) X; a, b_0, b_1, b_2 – параметры распределения. Путем решения (1) получаются семь основных типов кривых Пирсона [2, 6].

Система распределений Джонсона строится с использованием следующего преобразования нормированной нормальной СВ

$$\frac{y-\gamma}{\delta} = g(x;\mu,\lambda), \tag{2}$$

где y – нормированная нормальная СВ; x – СВ, для которой необходимо подобрать распределение; $g(\bullet)$ – некоторая монотонная функция; γ, μ – параметры сдвига; δ, λ – параметры масштаба. Джонсоном были предложены три типа функций, определяющих соответствующие семейства кривых: так называемые кривые S_L-, S_b-, S_U-распределений [2, 4, 8].

Распределения семейств (1) и (2) позволяют накрыть практически всю область возможных значений коэффициентов скоса и эксцесса, что дает возможность описывать как близкие к нормальной, так и отличные от нее статистические совокупности (смотри, например, [9]). В то же время, использование распределений семейств Пирсона и Джонсона связано с определенными трудностями. Прежде всего, для имеющихся статистических данных необходимо выбрать соответствующий тип кривой. Так, для выбора соответствующего распределения семейства Пирсона используется критерий

$$æ = \frac{r_3^2 (r_4 + 3)^2}{4(4r_4 - 3r_3^2)(2r_4 - 3r_3^2 - 6)},$$

где $r_3 = \frac{m_3}{m_2^{\frac{3}{2}}}$, $r_4 = \frac{m_4}{m_2^2}$ – статистические оценки приведенных моментов 3-го и 4-го порядка (m2, m3, m4 – оценки центральных моментов второго, третьего и четвертого порядков соответственно). Для выбора распределения системы Джонсона, соответствующего статистической совокупности, обычно используют каппа-диаграмму [8] или метод процентилей [8, 15]. Учитывая это, исследователь часто ищет распределение, которое в рамках одной аналитической формы позволяет описать статистические данные с различными значениями коэффициентов скоса и эксцесса. В этом смысле более удобными являются обобщенное лямбда-распределение и распределение Грама–Шарлье.

Обобщенное лямбда-распределение (ОЛР) определяется через квантильную функцию $Q(u)$ для $0 \le u \le 1$, представляющую собой функцию, обратную функции распределения СВ $F(x)$, так что $Q \equiv F^{(-)}$. На практике наибольшее распространение нашла квантильная функция вида [13, 14]

$$Q(u) = \lambda_1 + \frac{u^{\lambda_3} - (1-u)^{\lambda_4}}{\lambda_2}, \tag{3}$$

где λ_1 – параметр положения, λ_2 – параметр масштаба, λ_3, λ_4 – параметры формы ОЛР.

Распределение (3) обладает значительной гибкостью, однако область его использования ограничивается соотношением $1{,}8 + 1{,}7r_3^2 \le r_4 \le 9 + 1{,}7r_3^2$ [10, 11], которое учитывает возможность получения решения для параметров, а также ограничение на существование моментов третьего и четвертого порядков ОЛР.

Распределение Грама–Шарлье [4, 5] получается на основе разложения неизвестной функции плотности по системе ортогональных полиномов Чебышева–Эрмита (весовая функция полиномов представляет собой нормальную плотность). Для нормированной СВ функция плотности распределения Грама–Шарлье имеет вид

$$f_{GC}(x) = \left\{1 + \frac{\beta_1}{3!}(x^3 - 3x) + \frac{\beta_2}{4!}(x^4 - 6x^2 + 3)\right\}N(x), \tag{4}$$

где $\beta_1 = \frac{\mu_3}{\mu_2^{\frac{3}{2}}} = \rho_3$ – коэффициент асимметрии; $\beta_2 = \frac{\mu_4}{\mu_2^2} - 3 = \rho_4 - 3$ – коэффициент плосковершинности (μ_2, μ_3, μ_4 – центральные моменты второго, третьего и четвертого порядков соответственно); $N(x)$ – плотность стандартного нормального распределения. При подгонке статистических данных с использованием (4) вместо β_1 и β_2 берутся их статистические оценки.

Несомненным преимуществом распределения Грама–Шарлье является его потенциальная возможность в рамках одной аналитической формы описать любые отличные от нормальной статистические совокупности. Однако на практике оказывается, что использование (4) возможно только лишь для определенных сочетаний значений коэффициентов скоса и эксцесса, вследствие возникновения у кривой плотности полимодальности (несколько мод) и отрицательных «хвостов» (выходов в отрицательные области). На рис. 1 показана область использования распределения Грама–Шарлье. Из графика, в частности, следует, что распределение Грама–Шарлье имеет ограниченное использование в случае близких к нулю и отрицательных значений коэффициентов эксцесса.

Анализируя все указанные выше семейства и распределения, можно отметить один момент, который не нашел должного отражения при анализе статистических данных. Это учет реальных ограниченных областей изменения СВ. Анализ процессов любой физической природы показывает, что области изменения характеристик этих процессов, как правило, ограничены некоторыми пределами [4, 8]. В [4] прямо указывается на необходимость проводить подгонку статистических совокупностей кривыми плотностей только в интервалах существования СВ. Таким образом, для более корректного описания СВ следует использовать так называемые ограниченные вероятностные распределения, в которых СВ варьируются в пределах некоторой замкнутой области (в одномерном случае – отрезка варьирования).

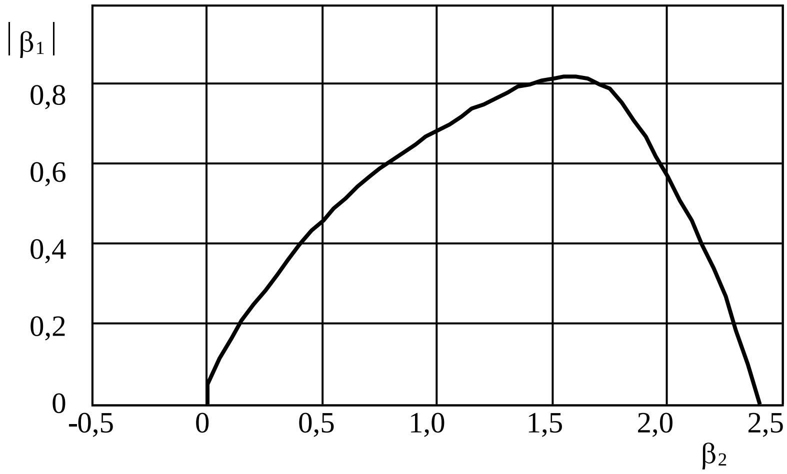

Рис. 1 – Область отсутствия отрицательных «хвостов» и полимодальности распределения Грама–Шарлье

Среди распределений семейств Пирсона и Джонсона есть четырехпараметричные распределения с ограниченной областью изменения варианты. Это распределение Джонсона S_b и распределение Пирсона типа 1. Недостатком этих распределений является тот факт, что их применение возможно только для определенных сочетаний коэффициентов скоса и эксцесса, а предельные точки этих распределений связаны с числовыми характеристиками и не могут быть выбраны из некоторых физических соображений. Последнее соображение в полной мере относится и к ОЛР. В этом несложно убедиться: если $\lambda_3 > 0$, $\lambda_4 > 0$ [11], распределение (3) имеет ограниченный интервал варьирования $[\lambda_1 - \lambda_2^{-1}, \lambda_1 + \lambda_2^{-1}]$.

Отметим одну важную особенность распределения Пирсона типа 1. Оно является весовой функцией ортогональных полиномов Якоби, с использованием которых может быть построено универсальное ограниченное вероятностное распределение, пригодное для использования в широком диапазоне значений коэффициентов скоса и эксцесса. Хотя в [9] и отмечено, что распределение, полученное на основе разложения по ортогональным полиномам Якоби, не нашло широкого применения, в то же время такое распределение, как будет показано ниже, может быть с успехом использовано для подгонки статистических совокупностей (цель настоящей статьи).

Теоретические основы построения распределения Якоби. Распределение Якоби было рассмотрено автором в работе [1] с точки зрения использования его в задачах параметрической надежности. Базируясь на этой работе, приведем теоретические основы построения вероятностного распределения Якоби.

Неизвестная плотность распределения $f(x)$ с областью варьирования $[a, b]$ может быть представлена в виде ряда разложения по ортогональным полиномам Якоби следующим образом:

$$f(x) \cong \sum_{k=0}^{\infty} c_k P_k(x;\alpha,\beta,a,b) h(x;\alpha,\beta,a,b)\,, \tag{5}$$

где $P_k(x;\alpha,\beta,a,b)$ – ортогональные полиномы Якоби с весовой функцией $h(x;\alpha,\beta,a,b)$; c_k – коэффициенты разложения, определяемые из условия ортогональности. Последовательность одномерных многочленов Якоби $P_k(x;\alpha,\beta,a,b)$ представляет собой полную систему функций, ортогональных на сегменте $[a, b]$ с весовой функцией

$$h(x;\alpha,\beta,a,b) = (x-a)^{\alpha}(b-x)^{\beta}\,, \tag{6}$$

где $\alpha > -1$ и $\beta > -1$. Весовая функция (6) непрерывна на сегменте ортогональности $[a, b]$, и ее свойства аналогичны распределению Пирсона типа 1.

Ортогональные полиномы $P_k(x;\alpha,\beta,a,b)$ удобно определять через ортогональные полиномы Якоби для отрезка варьирования [-1; 1], свойства которых хорошо исследованы [7]. Выражения для первых пяти ортогональных полиномов $P_k(x;\alpha,\beta,-1,1)$ ($k=\overline{0,4}$) приведены в таблице 1.

Таблица 1 – Выражения для ортогональных полиномов Якоби

k	Полиномы Якоби $P_k(x;\alpha,\beta,-1,1)$
0	1
1	$(\alpha+\beta+2)x+\beta-\alpha$
2	$(\alpha+\beta+3)(\alpha+\beta+4)x^2 +$ $+2(\alpha+\beta+3)(\beta-\alpha)x+(\beta-\alpha)^2 - (\alpha+\beta+4)$
3	$(\alpha+\beta+4)(\alpha+\beta+5)(\alpha+\beta+6)x^3 +$ $+3(\alpha+\beta+4)(\alpha+\beta+5)(\beta-\alpha)x^2 +$ $+3(\alpha+\beta+4)[(\beta-\alpha)^2 - (\alpha+\beta+6)]x +$ $+(\beta-\alpha)[(\alpha+\beta+4)(\alpha+\beta+5) - 4(\alpha+3)(\beta+3)]$
4	$(\alpha+\beta+5)(\alpha+\beta+6)(\alpha+\beta+7)(\alpha+\beta+8)x^4 +$ $+4(\alpha+\beta+5)(\alpha+\beta+6)(\alpha+\beta+7)(\beta-\alpha)x^3 +$ $+6(\alpha+\beta+5)(\alpha+\beta+6)[(\beta-\alpha)^2 - (\alpha+\beta+8)]x^2 +$ $+4(\beta-\alpha)(\alpha+\beta+5)[(\alpha+\beta+6)(\alpha+\beta+7) - 4(\alpha+4)(\beta+4)]x +$ $+(\alpha+\beta+5)(\alpha+\beta+6)[(\alpha+\beta+7)(\alpha+\beta+8) - 8(\alpha+4)(\beta+4)] +$ $+16(\alpha+4)(\beta+4)(\alpha+3)(\beta+3)$

Переход на произвольный отрезок варьирования $[a, b]$ может быть осуществлен при помощи замены переменных следующим образом

$$P_k(x;\alpha,\beta,a,b) = \left(\frac{b-a}{2}\right)^k P_k\left(\frac{2x-b-a}{b-a};\alpha,\beta,-1{,}1\right).$$

Проводя преобразования, выражение для $P_k(x;\alpha,\beta,a,b)$ можно записать так:

$$P_k(x;\alpha,\beta,a,b) = \sum_{i=0}^{k} \tilde{b}_k^{(i)} x^i \;, \tag{7}$$

где $\tilde{b}_k^{(i)} = \sum_{j=i}^{k} b_k^{(j)} (b-a)^{k-j} 2^{i-k} C_j^i (-b-a)^{j-i}$, $b_k^{(i)}$ – коэффициент полинома $P_k(x;\alpha,\beta,\text{-}1{,}1)$, стоящий при переменной в степени *i*.

Свойство ортогональности для полиномов Якоби

$$\int_a^b P_k(x;\alpha,\beta,a,b) P_m(x;\alpha,\beta,a,b) h(x;\alpha,\beta,a,b)dx = \begin{cases} 0, k \neq m \\ \|P(x;\alpha,\beta,a,b)\|^2 \end{cases}$$

устанавливается с помощью интегрирования по частям (смотри [7]). Здесь лишь приведем окончательное выражение для нормы

$$\|P_k(x;\alpha,\beta,a,b)\|^2 = k!(b-a)^{\alpha+\beta+2k+1} \frac{\Gamma(\alpha+k+1)\Gamma(\beta+k+1)}{(\alpha+\beta+2k+1)\Gamma(\alpha+\beta+k+1)} ,$$

которое обобщает традиционное выражение для нормы ортогональных полиномов Якоби $P_k(x;\alpha,\beta,-1{,}1)$.

Коэффициенты разложения c_k находятся из условия ортогональности (полагается, что ряд (5) равномерно сходится)

$$c_k = \frac{1}{\|P_k(x;\alpha,\beta,a,b)\|^2} \int_a^b P_k(x;\alpha,\beta,a,b) f(x)dx = \frac{1}{\|P_k(x;\alpha,\beta,a,b)\|^2} f_k(\overline{\nu}) , \tag{8}$$

где в формуле $f_k(\overline{\nu})$ есть результат замены в выражении соответствующего полинома $P_k(x;\alpha,\beta,a,b)$ (формула (7)) степеней x^i соответствующими начальными моментами ν_i.

Ряд (5) при некоторых условиях сходится к $f(x)$ и, следовательно, может служить для представления $f(x)$ с любой степенью точности. Однако сходимость или расходимость ряда (5) не имеет практического значения, важно лишь, чтобы плотность $f(x)$ могла быть с достаточной точностью представлена с помощью небольшого числа членов ряда. Усеченный до некоторого члена ряд (5) представляет собой распределение Якоби. По аналогии с распределением Грама–Шарлье ограничимся пятью членами (5), что приведет к необходимости использования в расчетах подгонки моментов до четвертого порядка включительно (при практическом использовании вместо моментов берутся их оценки, а, как известно, статистические оценки моментов более высоких порядков имеют большую погрешность). Выражение для плотности распределения Якоби можно переписать в виде

$$f_{Ja}(x) = c_0 h(x;\alpha,\beta,a,b)\sum_{i=0}^{4} b_{sum}^{(i)} x^i \;, \qquad (9)$$

где $c_0 = \dfrac{\Gamma(\alpha+\beta+2)}{(b-a)^{\alpha+\beta+1}\Gamma(\alpha+1)\Gamma(\beta+1)}$, $b_{sum}^{(i)} = \dfrac{1}{c_0}\sum_{j=0}^{4-i} \tilde{b}_{4-j}^{(i)} c_{4-j}$.

Несложно заметить, что произведение первых сомножителей в (9) представляет собой плотность распределения Пирсона типа 1.

Функция распределения Якоби находится путем интегрирования усеченного ряда (5) и имеет вид

$$F_{Ja}(x) = I_{\bar{x}}(\alpha+1,\beta+1) + \sum_{i=1}^{4} c_i P_{i-1}(x;\alpha+1,\beta+1,a,b)h(x;\alpha+1,\beta+1,a,b)\,,$$

где $I_t(p,q) = \dfrac{\Gamma(p+q)}{\Gamma(p)\Gamma(q)}\int_0^t x^{p-1}(1-x)^{q-1}dx$ – неполная бета функция; $\bar{x} = \dfrac{x-a}{b-a}$.

Построение распределения Якоби при подгонке статистических данных. Построение распределения Якоби при подгонке статических совокупностей состоит в определении параметров весовой функции α, β, a, b, а также коэффициентов разложения c_k ($k = \overline{0,4}$). Последние при имеющихся значениях α, β, a, b определяются по методу моментов (согласно (8)), исходя из оценок числовых характеристик СВ. Получаемая кривая распределения Якоби должна быть в некотором смысле оптимальной, поэтому неизвестные параметры α, β, a, b желательно выбирать исходя из некоторого критерия оптимальности. Будем полагать, что по имеющимся статистическим данным проведена операция группирования. В таком случае под оптимальной кривой Якоби будем понимать функцию плотности, которая обладает следующими свойствами:

- минимальное отклонение, в смысле некоторого критерия, аппроксимирующей кривой от гистограммы исходной статистической совокупности;
- отсутствие выходов кривой плотности в отрицательные области;
- одномодальность аппроксимирующей кривой.

Представляется целесообразным неизвестные параметры распределения Якоби α, β, a, b определять путем минимизации целевой функции взвешенной суммы квадратов отклонений аппроксимирующей кривой Якоби от гистограммы исходной статистической совокупности в точках, являющихся центрами интервалов группирования

$$G(\mathbf{U}) = \sum_{i=1}^{m} w_i (n_i - n \cdot f_{Ja}(x_i;\mathbf{U}))^2 \to \min\;, \qquad (10)$$

где $\mathbf{U} = \{\alpha, \beta, a, b\}$ – вектор варьируемых (управляющих) параметров процесса оптимизации (параметры распределения Якоби); w_i – некоторый весовой коэффициент; n_i – частота i-того разряда (количество наблюдений попавших в данный разряд); n – объем выборки; x_i – середина i-го разряда; $f_{Ja}(x_i;\mathbf{U})$ – значение функции плотности распределения Якоби в точке x_i; m – количество интервалов группирования.

Если в $G(\mathbf{U})$ используется весовой коэффициент $w_i = 1$, тогда (10) соответствует методу наименьших квадратов для сглаживания аппроксимирующей кривой. Если $w_i = (n \cdot f_{ja}(x_i;\mathbf{U}))^{-1}$, то (10) может рассматриваться как метод минимума критерия хи-квадрат. В итоге, подгонка распределения Якоби представляет собой комбинированный метод, состоящий из метода моментов и метода сглаживания.

Определение параметров α, β, a, b в соответствии с (10) не обеспечивает автоматического выполнения всех требований оптимальности кривой распределения Якоби, аппроксимирующего статистические данные. Для удовлетворения требований по отсутствию у кривой (9) отрицательных «хвостов» и полимодальности будем использовать обобщенную целевую функцию вида

$$G'(\mathbf{U}) = G(\mathbf{U}) + \delta(\mathbf{U}/\Omega)) \to \min , \tag{11}$$

где $\delta(\mathrm{U}/\Omega)$ – индикаторная функция; Ω – область значений вектора варьируемых параметров, для которой функция плотности распределения Якоби не имеет выходов в отрицательные области и имеет одну моду (область «0/1»). Индикаторная функция определяется следующим образом:

$$\delta(\mathbf{U}/\Omega) = \begin{cases} 0, \text{если } \mathbf{U} \in \Omega; \\ +\infty, \text{если } \mathbf{U} \notin \Omega. \end{cases}$$

Таким образом, задача поиска параметров распределения Якоби свелась к задаче условной оптимизации.

В процессе подгонки предварительно необходимо осуществить замену переменной

$$x' = \frac{x - \hat{X}_0}{c} , \tag{12}$$

где x – значение переменной исходной статистической совокупности; $\hat{X}_0$ – статистическая оценка моды; c – величина интервала группирования. Оценка $\hat{X}_0$ определяется следующим образом [6]:

$$\hat{X}_0 = \overline{X} - \frac{sr_3}{2}\frac{t+2}{t-2},$$

где $\overline{X}$ – выборочное среднее; s – выборочное среднее квадратическое отклонение; $t = \dfrac{6(r_4 - r_3^2 - 1)}{3r_3^2 - 2r_4 + 6}$.

Определение предельных точек отрезка варьирования распределения Якоби. Использование (11) на пространстве варьируемых параметров α, β, a', b' (a', b' – пределы, получающиеся в соответствии с (12)) позволяет определить предельные точки отрезка варьирования СВ исходя из критерия наилучшего качества аппроксимирующей кривой. В то же время при построении распределения Якоби определение неизвестных предельных точек a, b области варьирования СВ может рассматриваться как отдельная задача и в этом случае они не включаются в перечень варьируемых параметров процесса минимизации.

В ряде случаев пределы изменения СВ удается установить прямым исследованием физического процесса. Например, нулевое значение является естественным нижним пределом для ряда физических характеристик, например скорости ветра.

В простейшем случае в качестве пределов варьирования СВ для группированных данных можно принять следующие значения

$$\begin{aligned} a &= x_1 - c \\ b &= x_m + c \end{aligned}, \tag{13}$$

где x_1, x_m – соответственно середины первого и последнего интервалов группированных данных (m – количество интервалов группирования).

Заметим, что значения, получаемые с использованием (13), могут выступать в качестве предельных значений, ограничивающих область варьирования СВ в процессе минимизации целевой функции (11).

Границы области изменения СВ могут быть определены с использованием статистических методов, основная часть которых базируется на анализе крайних значений упорядоченной выборки. Например, для упорядоченного по возрастанию вариационного ряда наблюдений $X_{(1)} < X_{(2)} < \ldots < X_{(n)}$, в качестве предельных точек области варьирования СВ можно непосредственно принять значения $\hat{a} = X_{(1)}$, $\hat{b} = X_{(n)}$. Эти оценки будут смещенными и тем ближе к реальным значениям предельных точек, чем больше объем статистической совокупности. Существует целая группа методов нахождения предельных точек, основанная на исключении такого смещения. Простейший из них предложен в работе [3], где в качестве нижней точки трехпараметричного распределения Вейбулла предложено брать $\hat{a} = 0{,}9X_{(1)}$. Соответственно, для верхней точки распределения по аналогии можно принять $\hat{b} = 1{,}1X_{(n)}$. Теоретически такие коэффициенты, устраняющие смещение, ничем не обоснованны и представляют элемент чисто эвристического подхода.

Математически более корректно устранить смещение порядка n^{-1} для оценок предельных точек позволяют соотношения, получаемые согласно методу [12]

$$\hat{b} = nX_{(n)} - \frac{n-1}{n}\left[(n-1)X_{(n)} + X_{(n-1)}\right] = \frac{2n-1}{n}X_{(n)} - \frac{n-1}{n}X_{(n-1)},$$

$$\hat{a} = nX_{(1)} - \frac{n-1}{n}\left[(n-1)X_{(1)} + X_{(2)}\right] = \frac{2n-1}{n}X_{(1)} - \frac{n-1}{n}X_{(2)}.$$

Для оценки пределов варьирования СВ также могут быть использованы статистические методы, предложенные в [16] и основанные на предположении, что поведение СВ на «хвостах» описывается экстремальным распределением третьего типа [2, 3].

Необходимо отметить, что вышеуказанные способы определения предельных точек области варьирования СВ позволяют уменьшить размер массива управляющих параметров, а значит, упростить процесс нахождения минимума (13), однако при этом несколько ухудшается качество подгонки. Возможен также комбинированный способ, когда один из пределов варьирования СВ выбирается на основании анализа физического процесса или статистическими методами, а второй вводится в совокупность варьируемых параметров.

Область использования распределения Якоби. Поскольку распределение Якоби получено путем разложения неизвестной функции плотности с использованием ортогональных полиномов, оно, как и распределение Грама-Шарлье, склонно к образованию отрицательных «хвостов» и полимодальности. Это может приводить к возникновению отрицательных частот при подгонке статистических данных кривыми Якоби.

Определение областей отсутствия отрицательных «хвостов» и полимодальности для распределения Якоби связано с большими трудностями вычислительного характера, поскольку приходится одновременно варьировать значениями четырех параметров: показателями степеней α, β и пределами варьирования a, b. Исследование, проведенное для нормированной СВ ($m = 0$, $\sigma = 1$) в пределах интервалов значений параметров $-5 \le a \le -3$; $3 \le b \le 5$ и $0 \le \alpha = \beta \le 10$, позволило получить указанную область, которая показана на рис. 2.

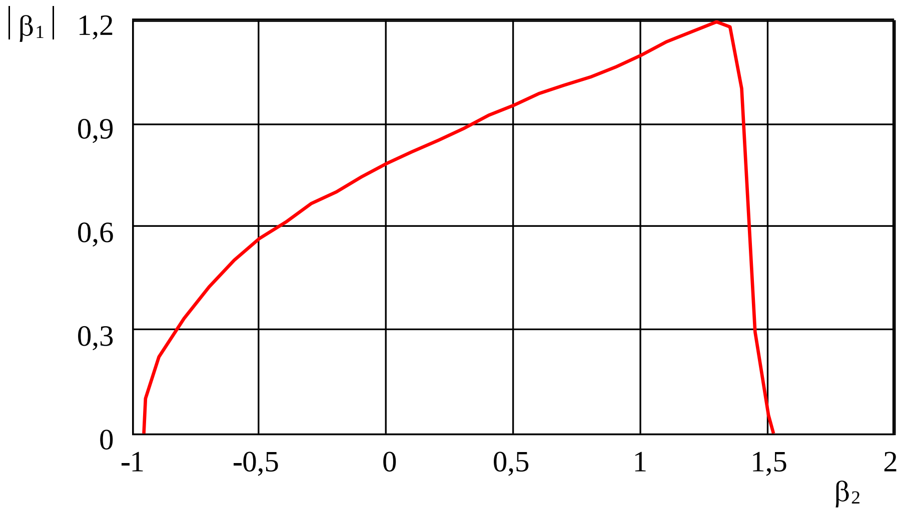

Рис. 2 – Область отсутствия отрицательных «хвостов» и полимодальности распределения Якоби

Как следует из рис. 2, область использования распределения Якоби даже с учетом вышеуказанных ограничений по параметрам является достаточно широкой. Распределение (9), в отличие от распределения Грама–Шарлье, можно использовать даже в области нулевых и отрицательных эксцессов.

Расширяя область варьирования граничных точек a и b, можно получить еще большую область отсутствия отрицательных «хвостов» и полимодальности распределения Якоби. В частности, при нулевом эксцессе, допустив варьирование по нижнему пределу в интервале $a \in [-5,-1]$, можно использовать распределение Якоби вплоть до значений $\beta_1 = 0{,}9$.

Необходимо также заметить, что для конкретных значений β_1, β_2, соответствующих области использования распределения Якоби (рис. 2), существуют определенные области значений α, β, a, b. Для их поиска в процессе подгонки и минимизации целевой функции (11) необходимо дополнительно исследовать области отсутствия отрицательных «хвостов» и полимодальности.

Примеры использования распределения Якоби для подгонки статистических совокупностей. С использованием распределения Якоби была проведена многочисленная аппроксимация статистических массивов. Для анализа качества подгонки статистических данных использовались как близкие к нормальной совокупности, так и статистические ряды, явно отличные от нормального случая. Результаты показали, что в обоих случаях получается хорошее согласие кривых Якоби с гистограммами статистических совокупностей. В процессе построения аппроксимирующих кривых Якоби оказалось, что минимальные значения критерия (11) реализуются при близких значениях показателей степеней $\alpha \approx \beta$ в весовой функции (6), для случая ультрасферических полиномов Якоби. Использование в (11) весового коэффициента $w_i = (n \cdot f_{Ja}(x_i))^{-1}$ давало возможность лучше описать статистические совокупности на «хвостах» – в областях, удаленных от математического ожидания.

Также было установлено, что для улучшения результатов подгонки желательно предварительно провести варьирование предельных точек a, b для $\alpha \approx \beta$ с целью выявления областей отсутствия отрицательных «хвостов» и полимодальности.

Для иллюстрации подгонки статистических данных с использованием распределения Якоби рассмотрим статистические ряды: длины бобов [4] и скорости ветра у поверхности Земли [6]. Числовые характеристики указанных рядов приведены в табл. 2.

Таблица 2 – Числовые характеристики статистических рядов

№ ряда	Физический параметр	$\overline{X}$	s	r_3	r_4	æ	$\hat{X}_0$	c
1	Длина бобов, мм	14,404	0,8998	-0,9106	4,8658	0,6072	14,7158	0,5
2	Скорость ветра на уровне Земли, м/с	4,58	2,89	1,5107	6,399	-57,5	2,375	2,0

Блоки управляющих параметров при минимизации целевой функции (11) для рассматриваемых статистических рядов формировались по-разному. Для ряда 1 оба предела области варьирования вводились в массив управляющих параметров и предельными значениями для них служили значения, полученные согласно (13). При подгонке статистического ряда 2 оба предела задавались жестко a = 0 м/с, b = 19 м/с. Указанные значения были получены путем предварительного исследования областей отсутствия отрицательных «хвостов» и полимодальности (рис. 3). На рис. 3 на пространстве значений α, β для различных отрезков варьирования СВ [a, b] показаны области «0/1», а также области «0/3», представляющие собой области отсутствия отрицательных «хвостов» и наличия у кривой плотности Якоби трех мод.

Минимизация (11) проводилась с использованием метода случайного поиска. В процессе минимизации использовалось несколько стартовых точек для поиска наилучшего варианта эмпирического распределения Якоби. В результате минимизации целевой функции $G'(\mathbf{U})$ получены параметры весовой функции распределения Якоби, приведенные в табл. 3.

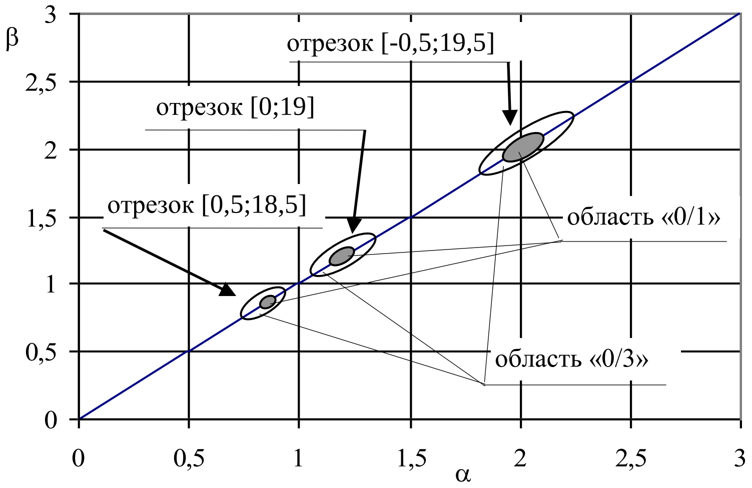

Рис. 3 – Области «0/1» и «0/3» распределения Якоби, аппроксимирующего ряд 2, для различных отрезков варьирования СВ

Таблица 3 – Значения параметров плотности распределения Якоби

№ ряда	α	β	a' (a)	b' (b)	c_0	$b_{sum}^{(i)}$				
						0	1	2	3	4
1	8,9019	8,9	-12,6753 (8,378)	5,9972 (17,714)	$1{,}0062\cdot10^{-18}$	4,3615	2,8991	0,7389	0,0836	$3{,}54\cdot10^{-3}$
2	1,35	1,35	-1,1875 (0)	8,3126 (19)	$2{,}5719\cdot10^{-03}$	5,7218	-3,7613	0,9407	-0,1051	$4{,}45\cdot10^{-3}$

Например, выражение для плотности распределения Якоби, аппроксимирующего ряд 1, имеет вид

$$f_{Ja}(x) = 1{,}0062\cdot10^{-18}\cdot\left(\frac{x-14{,}7158}{0{,}5}+12{,}6753\right)^{8{,}9019}\left(5{,}9972-\frac{x-14{,}7158}{0{,}5}\right)^{8{,}9}\times$$
$$\times\left[3{,}54\cdot10^{-3}\cdot\left(\frac{x-14{,}7158}{0{,}5}\right)^{4}+0{,}0836\left(\frac{x-14{,}7158}{0{,}5}\right)^{3}+0{,}7389\left(\frac{x-14{,}7158}{0{,}5}\right)^{2}+\right.$$
$$\left.+2{,}8991\left(\frac{x-14{,}7158}{0{,}5}\right)+4{,}3615\right].$$

Отметим, что для получения более высокого качества подгонки необходимо учитывать для параметров распределения Якоби как минимум четыре знака после запятой.

Результаты выравнивания указанных статистических рядов с использованием распределения Якоби приведены в табл. 4, 5. Для сравнения в таблицах также приведены выравнивающие частоты, полученные для соответствующих распределений семейств Пирсона и Джонсона (для ряда 1), распределения Грама–Шарлье и ОЛР. Часть результатов была заимствована из соответствующих источников. При построении обобщенного лямбда распределения (3) использовался метод моментов [10, 11]. Качество подгонки рассматриваемыми распределениями оценивалось с использованием критерия согласия хи-квадрат [5].

Для ряда 1 можно отметить следующее. Подгонка с использованием распределения Грама–Шарлье (4) приводит к появлению отрицательных частот. Для распределения Грама–Шарлье с четырьмя членами разложения (в (4) исключен член, содержащий коэффициент плосковершинности), хотя отрицательных частот и нет, качество подгонки вряд ли можно считать удовлетворительным.

Выравнивающие частоты, полученные с использованием ОЛР (значения параметров: $\lambda_1 = 14{,}830407$; $\lambda_2 = 0{,}032243$; $\lambda_3 = 0{,}023334$; $\lambda_4 = 0{,}009156$, и соответственно, пределы варьирования [8,833; 17,338]), показывают неплохое совпадение с исходным статистическим рядом. Результаты, полученные для ОЛР, с позиций критерия критерия хи-квадрат, оказываются лучше чем для Пирсона типа 4. Выравнивающий ряд, полученный для распределения Джонсона, параметры которого определялись по методу процентилей, показывает неплохое соответствие по отношению к исходному статическому ряду. Наилучшими же являются результаты выравнивания статистического ряда, полученные с использованием распределения Якоби (9).

Результаты подгонки ряда 2 (табл. 5 и рис. 4) показывают, что функция плотности гамма-распределения (Пирсона типа 3) начинается в точке $a = 0{,}754$ м/с, а не с нулевого значения, как следует из физических соображений. Последнее представляется весьма важным, поскольку скорости ветра, близкие к нулю, не редкость. Кривая плотности распределения Грама–Шарлье (4) полимодальна и выходит в отрицательную область, что является неприемлемым. Для ОЛР (параметры соответственно равны $\lambda_1 = 2{,}147012$; $\lambda_2 = 0{,}012065$; $\lambda_3 = 0{,}003922$; $\lambda_4 = 0{,}034404$) пределы варьирования СВ составляют [-0,793; 24,657]. Нижний предел имеет отрицательное значение, что физически невозможно (вероятность получить значения скорости ветра меньше 0 при этом составляет 0,0011). Распределение Якоби по критерию хи-квадрат, а также области варьирования СВ и в этом случае дает наилучший результат подгонки.

Заключение. В статье предложено использовать универсальное распределение Якоби, построенное с использованием разложения неизвестной функции плотности по системе ортогональных полиномов Якоби, для описания статистических совокупностей. Показаны широкие возможности использования распределения Якоби, которое позволяет:

– осуществить подгонку статистических данных на ограниченных отрезках варьирования, что естественно приближает математическую модель к реальному физическому процессу;

– одной аналитической формой охватить достаточно широкую область возможных значений коэффициентов скоса и эксцесса статистической совокупности (включая нулевые и отрицательные эксцессы);

– наилучшим образом (в рамках выбранного критерия оптимальности) описывать статистические данные за счет оптимального выбора параметров распределения.

Таблица 4 – Выравнивающие частоты для ряда 1

Длина бобов, мм	Наблюденные частоты	IV тип Пирсона[1)]	Грам-Шарлье (4)[1)]	Грам-Шарлье (4 члена)[1)]	Тип S_U Джонсона[1)]	Тип S_U Джонсона[2)]	ОЛР	Якоби
<9,25	–	1,9	1,7	0,9	2,6	2,1	3,5	1,8
9,5	1	2,6			2,7	2,4		
10,0	7	5,4			5,8	5,3	5,6	8,4
10,5	18	11,3	10,0	5,9	12,1	11,4	12,3	19,0
11,0	36	24,2	43,5	29,6	25,7	23,2	26,7	33,9
11,5	70	52,5	117,0	98,7	55,2	49,1	57,0	67,1
12,0	115	113,8	178,1	206,2	118,0	107,6	119,7	116,7
12,5	199	243,7	132,1	258,7	249,3	232,2	245,5	189,9
13,0	437	503,6	199,0	280,7	508,7	491,8	488,7	418,9
13,5	929	968,9	921,3	713,4	970,6	975,2	926,6	994,9
14,0	1787	1638,9	2082,6	1788,4	1642,5	1688,8	1605,9	1800,2
14,5	2294	2229,8	2506,4	2593,0	2240,6	2294,9	2310,9	2281,8
15,0	2082	2132	1833,0	2155,4	2130,3	2086,2	2230,8	1975,8
15,5	1129	1181,6	926,2	1012,7	1151,5	1087,5	1083,9	1111,7
16,0	275	299,3	370,4	241,7	290,1	315,3	266,2	363,0
16,5	55	28,5	116,6	25,6	32,2	56,6	46,9	54,0
17,0	6	1,4	13,7	12,8	2,0	9,4	7,9	2,9
>17,25	–	–	–15,2	16,3	0,1	1,0		
Всего	9440	9440	9440	9440	9440	9440	9440	9440
χ^2	–	96,2	–	>400	74,8	48,4	59,9	37,3

Примечания: [1)] результаты заимствованы из работы [4], где в качестве метода оценки параметров использовался метод моментов; [2)] результаты заимствованы из работы [15], где для определения неизвестных параметров распределения использовался метод процентилей.

Таблица 5 – Выравнивающие частоты для ряда 2

Скорость ветра, м/с	Частота реализации	Выравнивающие частоты			
		Грам-Шарлье	Гамма[1)] (Пирсона 3)	ОЛР	Якоби
1,50	24	18,4	26	24,2	24,6
3,50	34	40,6	33,7	34,6	33,8
5,50	23	29,1	20,4	20,4	23,4
7,50	10	3,4	10,6	10,6	11,1
9,50	5	0,7	5,2	5,2	3,9
11,5	2	4,5	2,4	2,5	1,5
13,5	1	2,7	1,1	1,1	1,0
15,5	0	0,6	0,5	0,5	0,9
17,5	1	0,1	0,2	0,2	0,6
Всего	100	100	100	99,5	100
χ^2	–	16,886	0,539	0,377	0,285

Примечания: [1)] результаты заимствованы из работы [6].

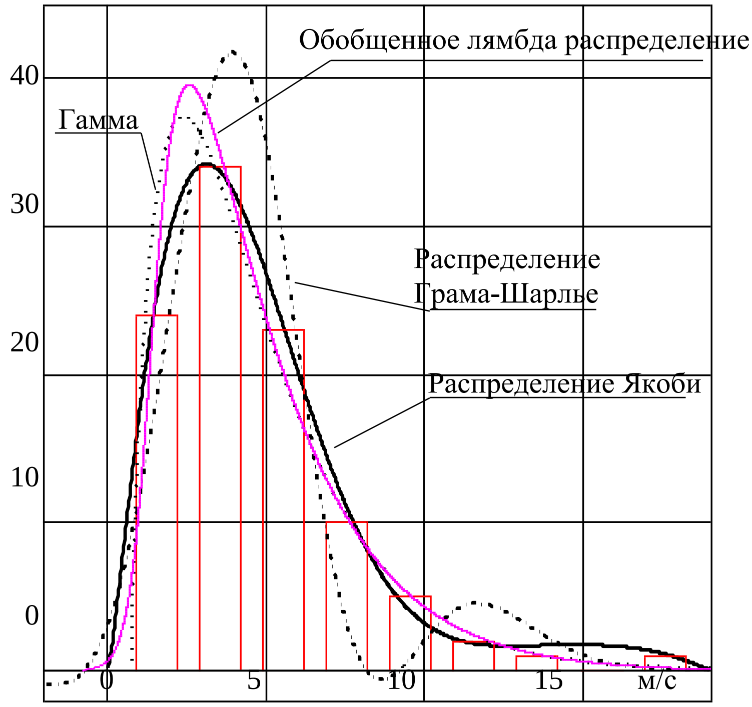

Рис. 4 – Аппроксимация статистического ряда скорости ветра.

Результаты подгонки статистических данных подтвердили лучшее качество аппроксимации распределением Якоби по сравнению с другими универсальными системами и распределениями.

1. *Гладкий Э. Г.* Использование ограниченного вероятностного распределения Якоби в задаче расчета параметрической надежности механических систем летательных аппаратов. Системне проектування та аналіз характеристик аерокосмічної техніки. Зб. наук. пр. Дн–ськ: ДДУ. 1998. Т. 1. С. 32 – 41.

2. *Губарев В. В.* Таблицы характеристик случайных величин и векторов. Новосибирск: Изд-во НЭИ, 1980. 280с.
3. *Капур К., Ламберсон Л.* Надежность и проектирование систем. М.: Мир, 1980. 604 с.
4. *Кендалл М. Дж., Стьюарт А.* Теория распределений М.: Наука, 1966. 588 с.
5. *Крамер Г.* Математические методы статистики. М.: ГИИЛ, 1975. 648 с.
6. *Митропольский А. К.* Техника статистических вычислений. М.: Наука. 1971. 576 с.
7. *Суетин П. К.* Классические ортогональные многочлены. М.: Наука, 1976. 328 с.
8. *Хан Г., Шапиро С.* Статистические модели в инженерных задачах. М.: Мир, 1969. 396 с.
9. *Jonson N. L., Kotz S., Balakrishnan N.* Continuous Univariate Distributions. Volume 1. N.Y.e.a. John Wiley and Sons, 1972. 769 p.
10. *Freimer M., Mudholkar G., Kollia G., Lin C.* (1988) A study of the generalized Tukey Lambda family, Communications in Statistics, Theory and Methods, 17(10). P. 3547 – 3567.
11. *Karian Z., Dudewicz E.* Fitting Statistical Distributions: The Generalized Lambda Distribution and Generalized Bootstrap Methods. CRC Press, Boca Raton, 2000.
12. *Quenouille M. H.* Notes on bias in estimation // Biometrika. 1956. Vol. 43. P. 353 – 364.
13. *Ramberg J., Schmeiser B.* An approximate method for generating asymmetric random variables. Communications of the ACM. 17(2). 1974. P. 78 – 82.
14. *Ramberg J. S., Tadikamalla P. R., Dudewicz E. J. and Mykytka E. F.* (1979) A probability distribution and its uses in fitting data. Technometrics, 21. P. 201 – 214.
15. *Slifker James F., Shapiro Samuel S.* The Johnson System: Selection and Parameter estimation/ Technometrics. 1980. Vol.22. No.2 May. P. 239 – 246.
16. *Smith R. L., Weissman I.* Maximum likelihood estimation of the lower tail of probability distribution. J. R. Statist. Soc. Section B. 1985. Vol. 47, № 2. P. 285 – 298.

Получено 28.02.2017,
в окончательном варианте 20.03.2017