МЕТОД ЯКОБИ В ЗАДАЧЕ СПЛАЙН-АППРОКСИМАЦИИ ПЛОТНОСТИ РАСПРЕДЕЛЕНИЯ

Безуглов Д.А., Миронович Д.В., Решетникова И.В.

Ростовский институт сервиса (филиал) Южно – Российского государственного
Университета экономики и сервиса

Во многих научно-прикладных задачах цифровой обработки радиотехнических сигналов характерной является задача получения оценки плотности распределения случайной величины.

Однако процесс статистической обработки измерений сопряжен с рядом трудностей. Во-первых, получение выборки всегда связано с ошибками и аномальными выбросами измеряемых величин. Во-вторых, существующие методы обработки предполагают наличие большой выборки, в идеале стремящейся к бесконечности. В практических же задачах размер выборки часто ограничен возможностями исследуемой системы. А в тех случаях, когда такие ограничения отсутствуют, необходимость обработки больших массивов результатов измерений также не желательна, поскольку ведет к увеличению вычислительных затрат и снижению оперативности обработки информации.

Целью настоящей статьи является разработка высокоэффективного алгоритма оценки неизвестной плотности распределения случайной величины по малой выборке на базе сглаживающих нормализованных кубических В-сплайнов с использованием итерационного метода Якоби, позволяющего существенно снизить объем вычислительных затрат.

Анализ существующих методов статистической обработки показывает, что в случае малой выборки построение эмпирической функции распределения случайной величины более обосновано, нежели плотности распределения этой случайной величины. Это объясняется тем, что функция распределения монотонно возрастает на всей области определения и поэтому менее критична к размерам выборки. Таким образом, при получении аналитического выражения для плотности распределения случайной величины логично воспользоваться известным соотношением, описывающим взаимосвязь функции распределения и плотности распределения случайной величины: $p(x) = \dfrac{dF(x)}{dx}$.

(1)

Однако, этот путь получения плотности распределения связан с невозможностью дифференцирования функции распределения частот - эмпирического аналога функции распределения, поскольку она имеет ступенчатый вид и, следовательно, в узлах разбиения вариационного ряда результатов измерений ее производная терпит разрыв. Для решения этой задачи по ряду причин целесообразно применить математический аппарат сглаживающих сплайнов. Таким образом, задачу получения эмпирической плотности распределения по малой выборке будем решать в следующей последовательности: по результатам измерений определим функцию накопления частот; полученную функцию сгладим нормализованным кубическим В-сплайном; продифференцировав этот сплайн в аналитическом виде получим выражение для плотности распределения случайной величины.

Пусть имеется набор измерений $x_1, x_2, ..x_N$. В результате их статистической обработки можно получить диаграмму накопления частот вида: $\hat{F}_N(x) = \sum_{i=1}^{\mu_N} \dfrac{1}{N_i}$, при $N \longrightarrow \infty$ $\hat{F}_N(x) \longrightarrow F(x)$, (2)

где: μ_N - число элементов выборки; N - число шагов разбиения вариационного ряда исходных данных.

Рассмотрим область $\phi=[a;b]$, на которой определена функция $\hat{F}_N(x)$. Введем на ϕ множество узлов с равным шагом разбиения h : $\Delta: x_0<x_1=a<x_2<...<x_N=b<x_{N+1}$. (3)

Получим аналитическое выражение для сплайн - аппроксимации функции накопления частот (2) в виде системы сглаживающих кубических нормализованных В-сплайнов дефекта 1:

$$F(x) = \sum_{i=0}^{N+1} b_i B_3^i(\chi_{i3}), \quad \chi_{i3} = \frac{1}{h}(x - x_{i+\frac{n+1}{2}}), \tag{4}$$

где: b_i - коэффициенты сплайна; $x_{i+\frac{n+1}{2}}$ - координаты средины носителя В-сплайна; n - степень сплайна.

$$B_3^i(\chi_{i3}) = \frac{1}{6}(2+\chi_{i3})^3 B_0^i + \left(\frac{2}{3} - \chi_{i3}^2 - \frac{\chi_{i3}^3}{2}\right)B_3^{i+1} + \left(\frac{2}{3} - \chi_{i3}^2 + \frac{\chi_{i3}^3}{2}\right)B_3^{i+2} + \frac{1}{6}(2-\chi_{i3})^3 B_0^{i+3}. \tag{5}$$

$$B_0^i = \begin{cases} 1 \text{ при } x \in [x_i; x_{i+1}] \\ 0 \text{ при } x \notin [x_i; x_{i+1}] \end{cases}. \tag{6}$$

Задачу сглаживания будем решать минимизируя функционал вида [1]:

$$J = \rho \int_a^b \left|F''(x)\right|^2 dx + \sum_{i=1}^{N} \left(\hat{F}(x_i) - F(x_i)\right)^2, \tag{7}$$

где ρ - коэффициент сглаживания; $\hat{F}(x_i)$ - значение сглаживаемой функции в узлах сетки; $F(x_i)$ - значение сплайна (4) в точке x_i; $|F''(x)|$ - модуль второй производной сплайна (4).

После несложных арифметических преобразований, группирования относительно коэффициентов сплайна и подстановки в (4) получим аналитическое выражение для сплайна в виде:

$$F(\chi)=\sum_{i=1}^{N-1}\left(\frac{1}{6}\chi^3(b_{i+2}-3b_{i+1}+3b_i-b_{i-1})+\frac{1}{2}\chi^2(b_{i+1}-2b_i+b_{i-1})+\frac{1}{2}\chi(b_{i+1}-b_{i-1})+\frac{1}{6}(b_{i+1}+4b_i+b_{i-1})\right). \tag{8}$$

Таким образом функционал (7) запишется в виде:

$$J=\rho\sum_{i=1}^{N-1}\int_{x_i}^{x_{i+1}}|F''(x)|^2\,dx+\sum_{i=1}^{N}\left[\hat{F}(x_i)-\frac{1}{6}(b_{i+1}+4b_i+b_{i-1})\right]^2. \tag{9}$$

Для нахождения коэффициентов сплайна (8) приносящего минимум функционалу (9) вычислим частные производные $\frac{\partial J}{\partial b_j}$ $(i=\overline{0;N+1})$ и приравняем их нулю. В результате этой операции получим систему из $N+2$ линейных уравнений матричная форма, которых имеет вида: $\frac{1}{6}$(***A*** ***B***)= ***Z***, (10)

где

$$\boldsymbol{A}=\begin{Vmatrix}1+12\rho & 4-18\rho & 1 & 6\rho & 0 & 0 & 0 & 0\\ 4-18\rho & 17+48\rho & 8-36\rho & 1 & 6\rho & 0 & 0 & 0\\ 1 & 8-36\rho & 18+84\rho & 8-54\rho & 1 & 6\rho & 0 & 0\\ 6\rho & 1 & 8-54\rho & 18+96\rho & 8-54\rho & 1 & 6\rho & 0\\ . & . & . & . & . & . & . & .\\ 0 & 6\rho & 1 & 8-54\rho & 18+96\rho & 8-54\rho & 1 & 6\rho\\ 0 & 0 & 6\rho & 1 & 8-54\rho & 18+84\rho & 8-36\rho & 1\\ 0 & 0 & 0 & 6\rho & 1 & 8-36\rho & 17+48\rho & 4-18\rho\\ 0 & 0 & 0 & 0 & 6\rho & 1 & 4-18\rho & 1+12\rho\end{Vmatrix} \tag{11}$$

$$\boldsymbol{B}=\begin{Vmatrix}b_0\\ b_1\\ b_2\\ \vdots\\ b_{N-1}\\ b_N\\ b_{N+1}\end{Vmatrix},\qquad \boldsymbol{Z}=\begin{Vmatrix}z_1\\ 4z_1+z_2\\ z_1+4z_2+z_3\\ \vdots\\ z_{N-1}+4z_N+z_{N+1}\\ z_{N-1}+4z_N\\ z_N\end{Vmatrix}.$$

Матрица коэффициентов системы (11) имеет семидиагональный вид и хорошо обусловлена.

Для сокращения вычислительных затрат применим для решения системы (11) итерационный метод Якоби [2]. При этом решение с точностью ε , где $\varepsilon>0$ - заданное число, запишется в следующем виде:

$$b_1^{n+1}=-\left(\frac{4-18\rho}{1+12\rho}\cdot b_2^n+\frac{b_3^n}{1+12\rho}+\frac{6\rho}{1+12\rho}\cdot b_4^n\right)+\frac{z_1}{1+12\rho}$$

$$b_2^{n+1}=-\left(\frac{4-18\rho}{17+48\rho}\cdot b_1^n+\frac{8-36\rho}{17+48\rho}\cdot b_3^n+\frac{b_4^n}{17+48\rho}+\frac{6\rho}{17+48\rho}\cdot b_5^n\right)+\frac{4\cdot z_1+z_2}{17+48\rho}$$

$$b_3^{n+1}=-\left(\frac{b_1^n}{18+84\rho}+\frac{8-36\rho}{18+84\rho}\cdot b_2^n+\frac{8-54\rho}{18+84\rho}\cdot b_4^n+\frac{b_5^n}{18+84\rho}+\frac{6\rho}{18+84\rho}\cdot b_6^n\right)+\frac{z_1+4\cdot z_2+z_3}{18+84\rho}$$

$$b_4^{n+1}=-\left(\frac{6\rho}{18+96\rho}\cdot b_1^n+\frac{b_2^n}{18+96\rho}+\frac{8-54\rho}{18+96\rho}\cdot b_3^n+\frac{8-54\rho}{18+96\rho}\cdot b_5^n+\frac{b_6^n}{18+96\rho}+\frac{6\rho}{18+96\rho}\cdot b_7^n\right)+\frac{z_2+4\cdot z_3+z_4}{18+96\rho}$$

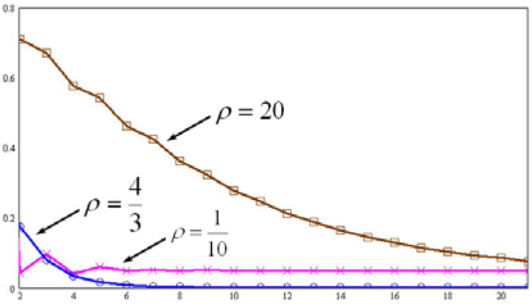
$$\ldots \tag{12}$$

$$b_{N-2}^{n+1}=-\left(\frac{6\rho}{18+96\rho}\cdot b_{N-5}^n+\frac{b_{N-4}^n}{18+96\rho}+\frac{8-54\rho}{18+96\rho}\cdot b_{N-3}^n+\frac{8-54\rho}{18+96\rho}\cdot b_{N-1}^n+\frac{b_N^n}{18+96\rho}+\frac{6\rho}{18+96\rho}\cdot b_{N+1}^n\right)+\frac{z_{N-1}+4\cdot z_N+z_{N+1}}{18+96\rho}$$

$$b_{N-1}^{n+1}=-\left(\frac{6\rho}{18+84\rho}\cdot b_{N-5}^n+\frac{b_{N-4}^n}{18+84\rho}+\frac{8-54\rho}{18+84\rho}\cdot b_{N-3}^n+\frac{8-36\rho}{18+84\rho}\cdot b_{N-2}^n+\frac{b_N^n}{18+84\rho}\right)+\frac{z_{N-1}+4\cdot z_N+z_{N+1}}{18+84\rho}$$

$$b_N^{n+1}=-\left(\frac{6\rho}{17+48\rho}\cdot b_{N-3}^n+\frac{b_{N-2}^n}{17+48\rho}+\frac{8-36\rho}{17+48\rho}\cdot b_{N-1}^n+\frac{4-18\rho}{17+48\rho}\cdot b_{N+1}^n\right)+\frac{z_{N-1}+4\cdot z_N}{17+48\rho}$$

$$b_{N+1}^{n+1} = -\left(\frac{6\rho}{1+12\rho}\cdot b_{N-2}^{n} + \frac{b_{N-1}^{n}}{1+12\rho} + \frac{4-18\rho}{1+12\rho}\cdot b_{N}^{n}\right) + \frac{z_N}{1+12\rho},$$



где $n = 0,1,2,\ldots,n$ - номер итерации.

Окончание итераций определяется либо заданием максимального числа итераций $n_{\max}$, либо условием

$$\max_{1\le i\le m}\left|b_i^{n+1} - b_i^n\right| < \varepsilon \qquad (13)$$

Рис.1. Невязка при заданных коэффициентах ρ.

На рис. 1 приведены результаты вычислительного эксперимента. Решалась система размерностью М=20. При этом предложенный метод позволяет существенно уменьшить объем вычислительных затрат и распараллелить вычислительный процесс, что важно при построении современных цифровых систем обработки сигналов. Это следует из того, что при решении системы (11) путем обращения матрицы ***A*** потребуется порядка $(N+2)^3 + (N+2)^2(N) \approx 2(N+2)^3$ операций. Как следует из рис 1. в силу диагонального вида метод Якоби в нашем случае хорошо сходится уже после 6-10 итераций. Таким образом, решение в соответствии с (12) потребует $13mN$ операций, где m – число итераций. При m=10, N>8 выигрыш в вычислительных затратах составит $\frac{(N+2)^2}{65}$ раз. При этом граница гарантированной устойчивости метода Якоби определится в нашем случае как $0 < \rho \le \frac{4}{3}$. При $\rho = 0$ матрица ***A*** плохо обусловлена, система решений не имеет.

В дальнейшем, дифференцируя в аналитическом виде (8), получим выражение для оценки плотности распределения случайной величины в виде:

$$p(x) = \sum_{i=1}^{N-1}\left(\frac{\chi^2}{2h}(b_{i+2} - 3b_{i+1} + 3b_i - b_{i-1}) + \frac{\chi}{h}(b_{i+1} - 2b_i + b_{i-1}) + \frac{1}{2h}(b_{i+1} - b_{i-1})\right). \qquad (14)$$

Для получения окончательного выражения необходимо соблюсти условие нормировки плотности распределения случайной величины. Поскольку функция $p(x)$ имеет значения отличные от нуля только на области *ф=[a;b]*, то условие нормировки можно записать в виде: $\int_a^b p(x)dx = 1$. (15)

Таким образом, задача нормировки результатов сглаживания сводится к определению нормирующего коэффициента для полученной оценки плотности распределения. Выражение для этого коэффициента, с учетом (14) и (15) запишется в виде:

$$\mathrm{K}\left(\frac{1}{2h}\sum_{i=1}^{N-1}\int_{x_i}^{x_{i+1}}\chi^2(b_{i+2} - 3b_{i+1} + 3b_i - b_{i-1}) + 2\chi\,(b_{i+1} - 2b_i + b_{i-1}) + (b_{i+1} - b_{i-1})\right)\mathrm{dx} = 1. \qquad (16)$$

где K - нормирующий коэффициент оценки плотности распределения.

В результате получим выражение, определяющее нормирующий коэффициент:

$$\mathrm{K} = \frac{6}{b_{N-2} + 4b_{N-1} + b_N - b_2 - 4b_1 - b_0}. \qquad (17)$$

С учетом (17) выражение (14) примет вид:

$$p(x) = \mathrm{K}\sum_{i=1}^{N-1}\left(\frac{\chi^2}{2h}(b_{i+2} - 3b_{i+1} + 3b_i - b_{i-1}) + \frac{\chi}{h}(b_{i+1} - 2b_i + b_{i-1}) + \frac{1}{2h}(b_{i+1} - b_{i-1})\right). \qquad (18)$$

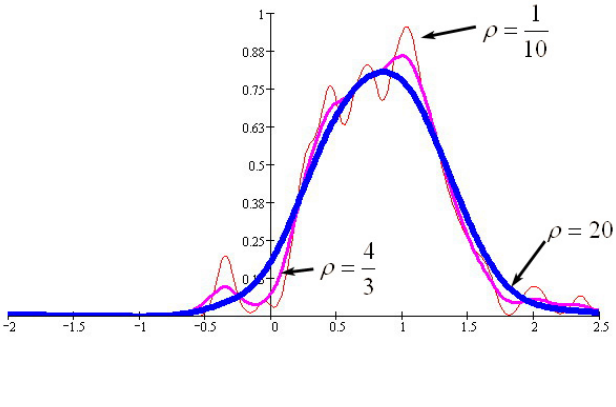


Рис. 2. Плотности распределения при заданных коэффициентах ρ.

Результаты исследования влияния коэффициента сглаживания ρ на вид получаемой плотности распределения приведена рис.2. Таким образом, увеличение коэффициента   в данном случае эквивалентно увеличению выборки.

Выводы. Таким образом, в результате проведенных аналитических выкладок получено выражение для сглаживающего кубического нормализованного В-сплайна одной переменной, на его основе разработан алгоритм для получения оценки плотности распределения по малой выборке результатов измерений с использованием итерационного метода Якоби, позволяющего существенно сократить вычислительные затраты и распараллелить вычисления. Полученные границы гарантированной сходимости метода позволяют обоснованно выбрать параметр ρ.

Литература

1. Безуглов Д.А., Поморцев П.М, Скляров А.В. Обработка результатов измерений на базе аппроксимации плотности распределения сглаживающими кубическими B- сплайнами. Измерительная техника, 2000, №9, с. 32-36;

2. Самарский А.А., Гулин А.В. Численные методы: Учебное пособие для вузов.- М.: Наука. Гл. ред. физ-мат. Лит., 1989.- 432 с.;

METHOD JACOBY IN A TASK OF SPLINE-APPROXIMATION OF DENSITY OF DISTRIBUTION

Bezuglov D., Mironovich D., Reschetnikova I.

The Rostov institute of service (branch) The South - Russian state university of economy and service

In many scientific - applied tasks of digital processing of radio engineering signals characteristic the task of reception of an estimation of density of distribution of casual size is.

The purpose of present clause is the development of highly effective algorithm of an estimation of unknown density of distribution of casual size on small sample on the basis of smoothing normalized cubic B - splines with use of an iterative method Jacoby, allowing it is essential to lower volume of computing expenses.

The analysis of existing methods of statistical processing shows, that in case of small sample the construction of empirical function of distribution of casual size is more proved, rather than density of distribution of this casual size. It is explained to that the function of distribution monotonously grows on all range of definition and consequently is less critical to the sizes of sample. However, this way of reception of density of distribution is connected to impossibility of differentiation of function of distribution of frequencies - empirical analogue function of distribution, as it has a step kind and, so, in units of splitting of a variational number of results of measurements its derivative bears break. Thus, a task of reception of empirical density of distribution on small sample we shall solve in the following sequence: by results of measurements we shall determine function of accumulation of frequencies; received function shall smooth normalized cubic B - spline; differentiating this spline in an analytical kind we shall receive expression for density of distribution of casual size.

The iterative method Jacoby is applied for the decision of the received system of the linear equations for reduction of computing expenses in job [2].

The system by dimension M=20 was solved. Thus the offered method allows essentially to reduce volume of computing expenses and to devide computing process, that is important at construction of modern digital systems of processing of signals. It follows that at the decision of system of the linear equations by the reference of a matrix it is required $(N+2)^3 + (N+2)^2(N) \approx 2(N+2)^3$ operations. As follows from the results, received in job, the method Jacoby in our case well converges already after 6-10 iterations. Thus, at n=10, N > 8 where n - number of iterations, the prize in computing expenses will make $\frac{(N+2)^2}{65}$ of time. Thus the border of guaranteed stability of a method Jacoby will be determined in our case as $0 < \rho \le \frac{4}{3}$. At $\rho = 0$ the matrix of system is badly caused, and she of the decisions has no.

Thus, as a result of the carried out analytical calculations the expression for smoothing cubic normalized In - сплайна one variable is received, on its basis the algorithm for reception of an estimation of density of distribution on small sample of results of measurements with use of an iterative method Jacoby is developed, allowing it is essential to reduce computing expenses and распараллелить of calculation. The received borders of guaranteed convergence of a method allow is proved to choose parameter ρ.

Literature

1. Bezuglov D.A., Pomorcev P.M, Sklyarov A.V. Processing of results of measurements on the basis of approximation of density of distribution smoothing cubic B - splines. Izmeratelnaya tehnika, 2000, №9, p. 32-36;

2. Samarskiy A.A., Gulin A.V. Numerical methods: Text book for institutes.- M.: NAUKA. Chief edition of physics – mathematical literature., 1989.- 432 p.