$\hat{\sigma}_\tau$ являются состоятельными. Тогда из свойств оценок Юла – Уокера [9] следует, что и построенные оценки $\hat{\sigma}^2\left(\hat{\sigma}_0,\ldots,\hat{\sigma}_p\right)$, $\hat{\theta}_i\left(\hat{\sigma}_0,\ldots,\hat{\sigma}_p\right), i=1,\ldots,p,$ также являются состоятельными. □

1. Хьюбер П. Робастность в статистике. М., 1984.
2. Хампель Ф., Рончетти Э., Рауссеу П., Штаэль В. Робастность в статистике. Подход на основе функций влияния. М., 1989.
3. Харин Ю. С. Оптимальность и робастность в статистическом прогнозировании. Минск, 2008.
4. Sen P. K. // IMS Lecture Notes. 1995. № 27.
5. Gomez G., Espinal A., Lagakos W. // Statistics in medicine. 2003. P. 409. № 22.
6. Park J. W., Genton M. G., Ghosh S. K. // The Canadian Journal of Statistics. 2007. Vol. 35. № 1. P. 151.
7. Андерсон Т. Статистический анализ временных рядов. М., 1976.
8. Андерсон Т. Введение в многомерный статистический анализ данных. М., 1963.
9. Харин Ю. С., Зуев Н. В., Жук Е. Е. Теория вероятностей, математическая и прикладная статистика. Минск, 2011.
10. Jagdish K. P., Campbell B. R. Handbook of the normal distribution. New York, 1982.
11. Фихтенгольц Г. М. Курс дифференциального и интегрального исчисления: в 3 т. М., 2003. Т. 1.

Поступила в редакцию 07.06.12.

Игорь Александрович Бодягин – аспирант кафедры математического моделирования и анализа данных. Научный руководитель – Ю. С. Харин.

Юрий Семенович Харин – член-корреспондент НАН Беларуси, доктор физико-математических наук, профессор, заведующий кафедрой математического моделирования и анализа данных.

УДК 519.234

Е. Г. КРАСНОГИР

ИНТЕГРАЛЬНОЕ ВЗВЕШЕННОЕ СРЕДНЕКВАДРАТИЧНОЕ РАССТОЯНИЕ В ЗАДАЧЕ НЕПАРАМЕТРИЧЕСКОГО ЯДЕРНОГО ОЦЕНИВАНИЯ ПЛОТНОСТИ ВЕРОЯТНОСТЕЙ ПО ВЫБОРКАМ ЗАВИСИМЫХ ОТСЧЕТОВ

In this paper we investigate the possibility of application of weighted mean integrated squared distance for bandwidth calculation in kernel density estimation.

Пусть X – стационарный марковский случайный процесс с неизвестной маргинальной плотностью распределения вероятностей $f(x)$, $x \in \mathbf{R}$, и пусть $X_1, \ldots, X_n$ – отсчеты со значениями в $\mathbf{R}$ этого процесса соответственно в моменты времени t, $t + \Delta$, $t + 2\Delta$, …, $t + (n-1)\Delta$ $(\Delta > 0)$. В качестве оценки функции $f(x)$ используется статистика

$$f_h(x) = \frac{1}{nh}\sum_{i=1}^{n} K\left(\frac{x - X_i}{h}\right), \tag{1}$$

где $h > 0$ – параметр размытости; $K = K(z) \in \mathbf{R}$ – ядерная функция (ядро), $z \in \mathbf{R}$ [1], имеющая свойства симметричной плотности распределения вероятностей [2], т. е.

$$\int_{-\infty}^{+\infty} K(z)dz = 1,\ \ K(z) \geq 0,\ \ K(z) = K(-z). \tag{2}$$

Функция $f_h(x)$ называется непараметрической ядерной оценкой для $f(x)$.

Форма ядерной функции слабо влияет на качество оценки (1) [2], что позволяет при необходимости выбирать наиболее удобный для вычислений вид ядерной функции.

Способы вычисления параметра размытости. С целью вычисления оптимального параметра размытости выбирается некий критерий или расстояние между истинной плотностью вероятностей и ее непараметрической оценкой, а затем данное расстояние минимизируется по h. Значение параметра размытости, при котором расстояние будет наименьшим, называется оптимальным параметром размытости.

Наиболее популярным критерием является $E\int_{-\infty}^{\infty}(f_h(x) - f(x))^2\,dx$ – интегральная среднеквадратичная ошибка [3] (E – оператор математического ожидания). Для достаточно больших n с использованием формулы Тейлора строится и минимизируется по h аппроксимация данной функции. Решением задачи минимизации для случая, когда $X_1, \ldots, X_n$ – независимые в совокупности одинаково распределенные наблюдения со значениями в **R** некоторой случайной величины X, является

$$h^* = B(K)\left(n\int_{-\infty}^{\infty} f''^2(y)dy\right)^{-1/5}, \tag{3}$$

где $B(K)$ – константа, значение которой зависит только от выбора ядра.

Очевидный недостаток выражения (3) – наличие в его правой части неизвестной функции $f''^2(y)$. В связи с этим возникает необходимость исследования других критериев. В частности, для распределений со строго возрастающей функцией распределения в [4] рассматривается критерий

$$E\int_0^1 (t - F_h(F^{-1}(t)))^2\,dt, \tag{4}$$

где $F_h(x) = \int_{-\infty}^{x} f_h(u)du$ – непараметрическая ядерная оценка функции распределения $F(x)$ случайной величины X; $F^{-1}(t)$ – функция, обратная $F(x)$. Используя формулу Тейлора, можно получить аппроксимацию критерия (4), минимизация которой для независимых наблюдений $X_1, \ldots, X_n$ приводит к асимптотически оптимальному значению

$$h^* = C(K)\left(n\int_{-\infty}^{\infty} f'^2(y)f(y)dy \Big/ \int_{-\infty}^{\infty} f^2(y)dy\right)^{-1/3}, \tag{5}$$

где $C(K)$ – зависящая только от выбора ядра константа. Преимущество данной формулы состоит в том, что в нее входит производная неизвестной плотности вероятностей $f(x)$ порядка меньше, чем в (3), что позволяет вычислять параметр размытости h^* (5) для более широкого набора плотностей вероятностей. Например, если $f(x)$ – плотность хи-квадрат распределения с m степенями свободы, то интеграл от f''^2 сходится при $m > 5$ (поэтому и h^* (3) вычисляется только при $m > 5$), а интеграл от $f'^2 f$ – при $m > 8/3$.

Заметим, однако, что в [4] исследуются свойства аппроксимации критерия (4) при $n \to \infty$, а для точного вычисления параметра размытости необходимо решать задачу минимизации самого критерия. Поэтому является целесообразным подробное исследование свойств функции (4), которое позволило бы гарантировать, что (4) может применяться для решения задачи вычисления оптимального параметра размытости.

В связи с этим рассмотрим критерий

$$Q(h) = E\int_{-\infty}^{\infty} (F(x) - F_h(x))^2 f(x)\,dx. \tag{6}$$

Выражение (6) будем называть интегральным взвешенным среднеквадратичным расстоянием для функции распределения. Заметим, что данный критерий получается из (4) путем замены переменных $x = F^{-1}(t)$ в интеграле.

Явное выражение для $Q(h)$. Пусть $f_j(t, s)$ – совместная плотность вероятностей отсчетов X_k и X_{k+j} стационарного случайного процесса X, $k = \overline{1, n-j}$, $j = \overline{1, n-k}$. Введем также обозначение

$$IK(x) = \int_{-\infty}^{x} K(t)\,dt. \tag{7}$$

Теорема 1. *Выражение для $Q(h)$ (6) может быть записано в виде*

$$Q(h) = \frac{1}{3} - 2\int_{-\infty}^{\infty} F(x)f(x)\left(\int_{-\infty}^{\infty} IK\left(\frac{x-y}{h}\right)f(y)dy\right)dx + \frac{1}{n}\int_{-\infty}^{\infty} f(x)\left(\int_{-\infty}^{\infty} IK^2\left(\frac{x-y}{h}\right)f(y)dy\right)dx +$$

$$+\frac{2}{n^2}\sum_{j=1}^{n-1}(n-j)\int_{-\infty}^{\infty} f(x)\left(\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} IK\left(\frac{x-y}{h}\right)IK\left(\frac{x-z}{h}\right)f_j(y,z)\,dydz\right)dx \tag{8}$$

или

$$Q(h)=\frac{1}{3}-2\int_{-\infty}^{\infty}K(z)\int_{-\infty}^{\infty}F(x-zh)F(x)f(x)\,dx\,dz+\frac{2}{n}\int_{-\infty}^{\infty}f(x)\int_{-\infty}^{\infty}K(u)F(x-uh)\int_{-\infty}^{u}K(v)\,dv\,du\,dx+$$

$$+\frac{2}{n^2}\sum_{j=1}^{n-1}(n-j)\int_{-\infty}^{\infty}f(x)\left(\int_{-\infty}^{\infty}K(u)\,du\int_{-\infty}^{\infty}K(v)F_j(x-uh,x-vh)\,dv\right)dx. \tag{9}$$

Доказательство. Разобьем (6) на три интеграла:

$$E\int_{-\infty}^{\infty}(F(x)-F_h(x))^2 f(x)\,dx=E\int_{-\infty}^{\infty}(F^2(x)f(x)-2F(x)F_h(x)f(x)+F_h^2(x)f(x))\,dx=$$

$$=\int_{-\infty}^{\infty}F^2(x)f(x)\,dx-2E\int_{-\infty}^{\infty}F(x)F_h(x)f(x)\,dx+E\int_{-\infty}^{\infty}F_h^2(x)f(x)\,dx=::I_1-2I_2+I_3. \tag{10}$$

Вычислим I_1. Для этого воспользуемся заменой $t=F(x)$. Тогда

$$I_1=\int_{-\infty}^{\infty}F^2(x)f(x)\,dx=\int_{-\infty}^{\infty}F^2(x)\,dF(x)=\int_0^1 t^2dt=\frac{1}{3}. \tag{11}$$

Запишем выражение для $E\{F_h(x)\}$. Так как

$$F_h(x)=\int_{-\infty}^{x}f_h(t)\,dt=\frac{1}{nh}\sum_{i=1}^{n}\int_{-\infty}^{x}K\left(\frac{t-X_i}{h}\right)dt=\frac{1}{n}\sum_{i=1}^{n}\int_{-\infty}^{(x-X_i)/h}K(y)\,dy=\frac{1}{n}\sum_{i=1}^{n}IK\left(\frac{x-X_i}{h}\right),$$

то

$$E\{F_h(x)\}=\frac{1}{n}\sum_{i=1}^{n}\int_{-\infty}^{\infty}IK\left(\frac{x-y}{h}\right)f(y)\,dy=\int_{-\infty}^{\infty}IK\left(\frac{x-y}{h}\right)f(y)\,dy=\int_{-\infty}^{\infty}f(y)\int_{-\infty}^{(x-y)/h}K(t)\,dt\,dy=$$

$$=\left[F(y)\int_{-\infty}^{(x-y)/h}K(t)\,dt\right]_{-\infty}^{\infty}+\frac{1}{h}\int_{-\infty}^{\infty}F(y)K\left(\frac{x-y}{h}\right)dy=\int_{-\infty}^{\infty}F(x-zh)K(z)\,dz.$$

Отсюда следует, что явное выражение для I_2 может быть записано в виде

$$I_2=\int_{-\infty}^{\infty}F(x)f(x)\left(\int_{-\infty}^{\infty}F(x-zh)K(z)\,dz\right)dx=\int_{-\infty}^{\infty}K(z)\left(\int_{-\infty}^{\infty}F(x-zh)F(x)f(x)\,dx\right)dz \tag{12}$$

или

$$I_2=\int_{-\infty}^{\infty}F(x)f(x)\left(\int_{-\infty}^{\infty}IK\left(\frac{x-y}{h}\right)f(y)\,dy\right)dx=h\int_{-\infty}^{\infty}F(x)f(x)\left(\int_{-\infty}^{\infty}IK(z)f(x-hz)\,dz\right)dx. \tag{13}$$

Далее

$$I_3=\frac{1}{nh^2}\int_{-\infty}^{\infty}f(x)\left(\int_{-\infty}^{\infty}\left(\int_{-\infty}^{x}K\left(\frac{s-y}{h}\right)ds\int_{-\infty}^{x}K\left(\frac{t-y}{h}\right)dt\right)f(y)\,dy\right)dx+$$

$$+\frac{2}{n^2h^2}\sum_{j=1}^{n-1}(n-j)\int_{-\infty}^{\infty}f(x)\left(\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\left(\int_{-\infty}^{x}K\left(\frac{s-y}{h}\right)ds\int_{-\infty}^{x}K\left(\frac{t-z}{h}\right)dt\right)f_j(y,z)\,dydz\right)dx=:: \tag{14}$$

$$=::\frac{1}{nh^2}I_{31}+\frac{2}{n^2h^2}\sum_{j=1}^{n-1}(n-j)I_{32}(j). \tag{15}$$

Упростим I_{31} с использованием определения $IK(x)$ (7):

$$I_{31}=h^2\int_{-\infty}^{\infty}f(x)\left(\int_{-\infty}^{\infty}IK^2\left(\frac{x-y}{h}\right)f(y)\,dy\right)dx=h^3\int_{-\infty}^{\infty}f(x)\left(\int_{-\infty}^{\infty}IK^2(z)f(x-zh)\,dz\right)dx= \tag{16}$$

$$=2h^2\int_{-\infty}^{\infty}f(x)\left(\int_{-\infty}^{\infty}K(u)F(x-uh)IK(u)\,du\right)dx. \tag{17}$$

Осталось получить явное выражение для $I_{32}(j)$:

$$I_{32}(j)=h^2\int_{-\infty}^{\infty}f(x)\left(\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}IK\left(\frac{x-y}{h}\right)IK\left(\frac{x-z}{h}\right)f_j(y,z)\,dydz\right)dx= \tag{18}$$

$$=h^4\int\limits_{-\infty}^{\infty} f(x)\left(\int\limits_{-\infty}^{\infty}\int\limits_{-\infty}^{\infty} IK(u)IK(v)f_j(x-uh,x-vh)\,dudv\right)dx.$$

Из (14), меняя порядок интегрирования, также можно получить, что

$$I_{32}(j)=h^2\int\limits_{-\infty}^{\infty} f(x)\left(\int\limits_{-\infty}^{\infty} K(u)\,du\int\limits_{-\infty}^{\infty} K(v)\,dv\int\limits_{-\infty}^{x-uh}\int\limits_{-\infty}^{x-vh} f_j(y,z)\,dydz\right)dx=$$

$$=h^2\int\limits_{-\infty}^{\infty} f(x)\left(\int\limits_{-\infty}^{\infty} K(u)\,du\int\limits_{-\infty}^{\infty} K(v)F_j(x-uh,x-vh)\,dv\right)dx. \tag{19}$$

Таким образом, выражение (8) следует из (10), (11), (13) – (16) и (18), а выражение (9) – из (10) – (12), (15), (17) и (19). ■

Ограниченность функции $Q(h)$. Из (6) $Q(h)\ge 0$. Получим также оценку сверху.

Теорема 2. *Если $K(z)$ удовлетворяет (2), то $Q(h)\le 4/3$ для любых $n\in\mathbf{N}$ и $h>0$.*

Доказательство. Получим оценки для составляющих $Q(h)$, пользуясь теоремой 1. Для I_2 с учетом (2) и свойств функции распределения из (12) имеем

$$I_2\le\int\limits_{-\infty}^{\infty} F(x)f(x)\,dx\int\limits_{-\infty}^{\infty} K(z)\,dz=\int\limits_{-\infty}^{\infty} F(x)f(x)\,dx=\frac{1}{2}. \tag{20}$$

Из выражения (17) с использованием замены переменных $t=IK(u)$ следует, что

$$\frac{1}{nh^2}I_{31}\le\frac{2}{n}\int\limits_{-\infty}^{\infty} f(x)\,dx\int\limits_{-\infty}^{\infty} K(u)IK(u)\,du=\frac{2}{n}\int\limits_{-\infty}^{\infty} f(x)\,dx\int\limits_{-\infty}^{\infty} IK(u)\,dIK(u)=$$

$$=\frac{2}{n}\int\limits_{-\infty}^{\infty} f(x)\,dx\int\limits_{0}^{1} t\,dt=\frac{1}{n}. \tag{21}$$

Для оценки $I_{32}(j)$ воспользуемся выражением (19):

$$\frac{2}{n^2h^2}\sum_{j=1}^{n-1}(n-j)I_{32}(j)\le\frac{2}{n^2}\sum_{j=1}^{n-1}(n-j)\int\limits_{-\infty}^{\infty} f(x)\,dx\int\limits_{-\infty}^{\infty} K(u)\,du\int\limits_{-\infty}^{\infty} K(v)\,dv=\frac{n-1}{n}. \tag{22}$$

Из (10), (11), (15), (20) – (22) и очевидного неравенства $I_2\ge 0$ следует, что

$$E\int\limits_{-\infty}^{\infty}(F(x)-F_h(x))^2 f(x)\,dx\le\frac{1}{3}+\frac{1}{n}+\frac{n-1}{n}=\frac{4}{3}. \tag{23}$$

Теорема доказана. ■

Таким образом, интегральное взвешенное среднеквадратичное расстояние для функции распределения (6) существует для всех абсолютно непрерывных распределений вероятностей, ядерных функций, удовлетворяющих (2), $n\in\mathbf{N}$ и $h>0$.

Существование минимума функции $Q(h)$.

Предположение 1. *Функции $K(z)$ и $f(x)$ являются непрерывными на области определения (под областью определения здесь и далее понимаются те подмножества числовой прямой, на которых указанные функции не равны нулю).*

Из (23) в условиях предположения 1 следует, что функция (6) является непрерывной. Поэтому с использованием (2), (7), (9) и (20) – (22) получаем, что

$$E\int\limits_{-\infty}^{\infty}(F(x)-F_h(x))^2 f(x)\,dx\xrightarrow[h\to 0]{}\frac{1}{3}-2\int\limits_{-\infty}^{\infty} K(z)\,dz\int\limits_{-\infty}^{\infty} F^2(x)f(x)\,dx+$$

$$+\frac{2}{n}\int\limits_{-\infty}^{\infty} f(x)F(x)\,dx\int\limits_{-\infty}^{\infty} K(u)IK(u)\,du+\frac{2}{n^2}\int\limits_{-\infty}^{\infty} K(u)\,du\int\limits_{-\infty}^{\infty} K(v)\,dv\sum_{j=1}^{n-1}(n-j)\int\limits_{-\infty}^{\infty} f(x)F_j(x,x)\,dx=$$

$$=\frac{1}{3}-\frac{2}{3}+\frac{1}{2n}+\frac{2}{n^2}\sum_{j=1}^{n-1}(n-j)\int\limits_{-\infty}^{\infty} f(x)F_j(x,x)\,dx\le$$

$$\le\frac{1}{3}-\frac{2}{3}+\frac{1}{2n}+\frac{2}{n^2}\sum_{j=1}^{n-1}(n-j)\int\limits_{-\infty}^{\infty} f(x)F(x)\,dx=-\frac{1}{3}+\frac{1}{2n}+\frac{n-1}{2n}=\frac{1}{6}. \tag{24}$$

Если мы рассмотрим случай, когда $X_1, \ldots, X_n$ – независимые одинаково распределенные случайные величины, то $F_j(x, x) = F^2(x)$ и

$$Q(h) \xrightarrow[h\to 0]{} -\frac{1}{3}+\frac{1}{2n}+\frac{n-1}{n}\int_{-\infty}^{\infty} f(x)F^2(x)\,dx = -\frac{1}{3}+\frac{1}{2n}+\frac{n-1}{3n}=\frac{1}{6n}. \tag{25}$$

Заметим, что при выполнении (2) $IK(0)=\int_{-\infty}^{0} K(t)\,dt=\frac{1}{2}$. Тогда из (8) получаем

$$Q(h) \xrightarrow[h\to \infty]{} \frac{1}{3}-2\int_{-\infty}^{\infty} F(x)f(x)\left(\int_{-\infty}^{\infty} IK(0)f(y)\,dy\right)dx+\frac{1}{n}\int_{-\infty}^{\infty} f(x)\left(\int_{-\infty}^{\infty} IK^2(0)f(y)\,dy\right)dx+$$

$$+\frac{2}{n^2}\sum_{j=1}^{n-1}(n-j)\int_{-\infty}^{\infty} f(x)\left(\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} IK^2(0)f_j(y,z)\,dydz\right)dx=\frac{1}{3}-\frac{1}{2}+\frac{1}{4n}+\frac{n-1}{4n}=\frac{1}{12}. \tag{26}$$

Таким образом, из (24) – (26) мы можем заключить, что, по крайней мере, для случая независимых одинаково распределенных случайных величин значение $h=\infty$ не является точкой минимума функции (6).

Предположение 2. *Выполняется условие* $\int_{-\infty}^{\infty} |z|K(z)\,dz<\infty$.

Предположение 3. *Функция* $f(x)$ *ограничена, а функции* $f_j(y\mid z)=f_j(y,z)/f(z)$ *и* $f_j(z\mid y)=$ $=f_j(y,z)/f(y)$, $j=\overline{1,n-1}$, *являются непрерывными на области определения и ограниченными.*

Теорема 3. *При выполнении предположений 1–3 и условий (2) функция* $Q(h)$ *в точке* $h=0$ *является убывающей.*

Доказательство. Вычислим производную функции $Q(h)$ (6). При выполнении предположений 1–3 допустимо дифференцирование под знаком интеграла. Воспользуемся выражениями (10), (12), (15), (17), (18). Тогда получим

$$Q'(h) = 2\int_{-\infty}^{\infty} zK(z)\left(\int_{-\infty}^{\infty} f(x-zh)F(x)f(x)\,dx\right)dz-\frac{2}{n}\int_{-\infty}^{\infty} f(x)\left(\int_{-\infty}^{\infty} uK(u)f(x-uh)IK(u)\,du\right)dx-$$

$$-\frac{2}{n^2h^2}\sum_{j=1}^{n-1}(n-j)\int_{-\infty}^{\infty} f(x)\left(\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} K\left(\frac{x-y}{h}\right)(x-y)IK\left(\frac{x-z}{h}\right)f_j(y,z)\,dydz\right)dx-$$

$$-\frac{2}{n^2h^2}\sum_{j=1}^{n-1}(n-j)\int_{-\infty}^{\infty} f(x)\left(\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} IK\left(\frac{x-y}{h}\right)K\left(\frac{x-z}{h}\right)(x-z)f_j(y,z)\,dydz\right)dx. \tag{27}$$

Заметим, что

$$\frac{1}{h^2}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} K\left(\frac{x-y}{h}\right)(x-y)IK\left(\frac{x-z}{h}\right)f_j(y,z)\,dydz=$$

$$=\frac{1}{h^2}\int_{-\infty}^{\infty} K\left(\frac{x-y}{h}\right)(x-y)\,dy\int_{-\infty}^{\infty}\int_{-\infty}^{(x-z)/h} K(t)\,dt f_j(y,z)\,dz=\int_{-\infty}^{\infty} K(v)v\,dv\int_{-\infty}^{\infty} K(t)\,dt\int_{-\infty}^{x-th} f_j(y,x-vh)\,dy.$$

Отсюда и из (2), (27) следует

$$Q'(h)\big|_{h=0} = 2\int_{-\infty}^{\infty} zK(z)\left(\int_{-\infty}^{\infty} f(x)F(x)f(x)\,dx\right)dz-\frac{2}{n}\int_{-\infty}^{\infty} f(x)\left(\int_{-\infty}^{\infty} uK(u)f(x)IK(u)\,du\right)dx-$$

$$-\frac{2}{n^2}\sum_{j=1}^{n-1}(n-j)\int_{-\infty}^{\infty} f(x)\left(\int_{-\infty}^{\infty} K(u)u\,du\int_{-\infty}^{\infty} K(t)\,dt\int_{-\infty}^{x} f_j(x,z)\,dz\right)dx-$$

$$-\frac{2}{n^2}\sum_{j=1}^{n-1}(n-j)\int_{-\infty}^{\infty} f(x)\left(\int_{-\infty}^{\infty} K(v)v\,dv\int_{-\infty}^{\infty} K(t)\,dt\int_{-\infty}^{x} f_j(y,x)\,dy\right)dx=$$

$$=-\frac{2}{n}\int_{-\infty}^{\infty} f^2(x)\,dx\int_{-\infty}^{\infty} uK(u)IK(u)\,du =:: -\frac{2}{n}G(f)\int_{-\infty}^{\infty} uK(u)IK(u)\,du=$$

$$=-\frac{2G(f)}{n}\left(\int_{0}^{\infty} uK(u)IK(u)\,du+\int_{-\infty}^{0} uK(u)IK(u)\,du\right)=-\frac{2G(f)}{n}\int_{0}^{\infty} uK(u)(IK(u)-IK(-u))\,du=$$

$$= -\frac{2G(f)}{n}\int_0^{\infty} uK(u)\left(\int_{-\infty}^{u} K(v)\,dv - \int_{-\infty}^{-u} K(v)\,dv\right)dz = -\frac{2G(f)}{n}\int_0^{\infty} uK(u)\left(\int_{-u}^{u} K(v)\,dv\right)du.$$

Очевидно, что подынтегральная функция непрерывна, неотрицательна и не равна тождественно нулю. Отсюда и из неравенства $G(f) > 0$ получаем $Q'(h)\big|_{h=0} < 0$, т. е. при $h = 0$ функция $Q(h)$ (6) является убывающей. Теорема доказана. ■

Объединяя результаты, полученные в теореме 3, и оценки (24) – (26), можно утверждать, что, по крайней мере, для случая независимых одинаково распределенных случайных величин интегральное взвешенное среднеквадратичное расстояние для функции распределения достигает своего минимума при конечном положительном значении параметра размытости.

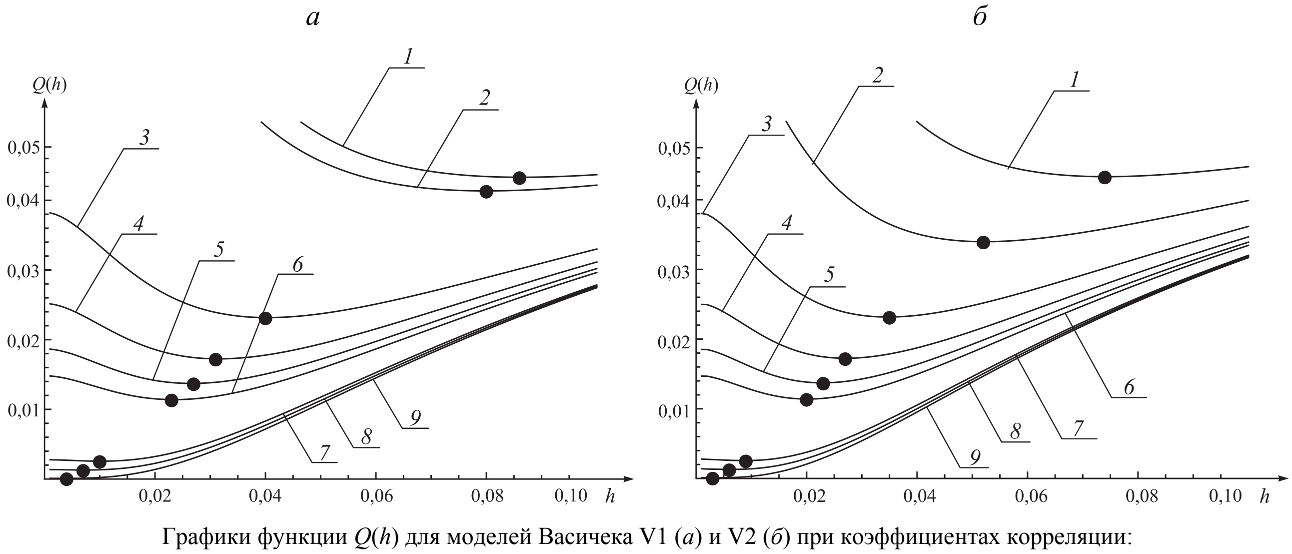
Поведение функции $Q(h)$ в зависимости от коэффициента корреляции в модели Васичека. Рассмотрим одну из моделей процессов финансового рынка – модель Васичека [5], описываемую стохастическим дифференциальным уравнением $dX = k(\theta - X)dt + \sigma dW(t)$, где $W(t)$ – стандартный винеровский процесс, а k, θ, σ – постоянные параметры. Случайный процесс, порождаемый данным уравнением, является нормальным марковским с маргинальной плотностью вероятностей $f(x) = \exp\{-0{,}5(x-\theta)^2 / D\} / \sqrt{2\pi D}$, где $D = \sigma^2/(2k)$, $x \in \mathbf{R}$. Совместная плотность

$$f_j(t,x) = \frac{1}{2\pi D\sqrt{1-\rho^{2j}}}\exp\left\{-\frac{(t-\theta)^2 - 2(t-\theta)(x-\theta)\rho^j + (x-\theta)^2}{2D(1-\rho^{2j})}\right\},\ t \in \mathbf{R}, x \in \mathbf{R},$$

где $\rho = e^{-k\Delta}$ – коэффициент корреляции отсчетов X_i и X_{i+1}, $i = \overline{1, n-1}$.

Как уже упоминалось, мы принимаем, что наблюдения производятся через некоторые интервалы времени Δ. Очевидно, что чем больше Δ, тем меньше зависимость между соседними наблюдениями. Длина интервала времени Δ входит в формулу для коэффициента корреляции ρ. Поэтому зависимость от Δ здесь соответствует зависимости от ρ.

Пусть $n = 1500$. В качестве ядерной функции возьмем $K(z) = e^{-|z|}/2$, $z \in \mathbf{R}$, – плотность распределения Лапласа. На рисунке представлены графики функции $Q(h)$ (6) при различных коэффициентах корреляции для моделей Васичека V1 и V2. В модели V1 $k = 0{,}0484$, $\theta = 0{,}1132$, $\sigma = 0{,}0112$ [6] (данные значения параметров были получены при построении модели Васичека, характеризующей изменение реальных процентных ставок доходности одномесячных векселей Казначейства США с декабря 1946 г. по февраль 1991 г.), в модели V2 $k = 0{,}261$, $\theta = 0{,}02237$, $\sigma = 0{,}0717$ [7] (данные значения параметров были получены при построении модели Васичека, характеризующей изменение ставки по федеральным фондам США с января 1963 г. по декабрь 1998 г.).



Графики функции $Q(h)$ для моделей Васичека V1 (*а*) и V2 (*б*) при коэффициентах корреляции: *1* – 1; *2* – ρ (для V1 $\rho = 0{,}9998$, для V2 $\rho = 0{,}9990$); *3* – 0,997; *4* – 0,995; *5* – 0,993; *6* – 0,991; *7* – 0,95; *8* – 0,9; *9* – 0. Точками обозначены минимальные значения функции

Как видно из рисунка, при $\rho \le 0{,}95$ оптимальные значения параметра h для зависимых и независимых выборочных данных различаются незначительно, однако при $\rho > 0{,}95$ наблюдается значительный рост отклонения оптимального параметра размытости от аналогичного значения h при $\rho = 0$.

В частности, для модели V1 при $\rho = 1$ оптимальный параметр размытости в 21,5 раза больше оптимального h при $\rho = 0$ (для модели V2 отношение указанных параметров размытости равно 24,7).

Так как процентные ставки котируются каждый рабочий день в течение года, то для них $\Delta = 1/250$. В этом случае для модели V1 $\rho = 0{,}9998$ и для модели V2 $\rho = 0{,}9990$. Поэтому для данных моделей при определении оптимального h нельзя пренебрегать зависимостью между отсчетами наблюдаемого процесса изменения процентных ставок.

Полученные результаты говорят о том, что интегральное взвешенное среднеквадратичное расстояние может использоваться для вычисления оптимального параметра размытости в задаче непараметрического ядерного оценивания плотности вероятностей для случая как независимых, так и зависимых выборочных данных.

1. Silverman B. W. Density estimation for statistics and data analysis. London, 1986.
2. Hardle W., Linton O. Handbook of Econometrics. Vol. IV. Chap. 38. Elsevier Science. 1994. P. 2297.
3. Turlach B. A. Bandwidth selection in kernel density estimation: a review. Technical report. Louvain-la-Neuve, 1993.
4. Красногир Е. Г. // Вестн. БГУ. Сер. 1. 2010. № 3. С. 79.
5. Vasicek O. // J. of Financial Economics. 1977. Vol. 5. P. 177.
6. Ahn D.-H., Gao B. // Rev. Financial Studies. 1999. Vol. 12. № 4. P. 721.
7. Ait-Sahalia Y. // J. of Finance. 1999. Vol. 54. № 4. P. 1361.

Поступила в редакцию 28.12.11.

Евгений Григорьевич Красногир – кандидат физико-математических наук, доцент кафедры теории вероятностей и математической статистики.

УДК 519.2

В. А. ВОЛОШКО

РИСК ПРОГНОЗИРОВАНИЯ НА ОСНОВЕ МОДЕЛИ БЛУМФИЛДА ПРИ НАЛИЧИИ ОБУЧАЮЩЕЙ ПОСЛЕДОВАТЕЛЬНОСТИ

The risk of forecasting is analyzed when the observed training and forecasted stationary Gaussian time series are independent and identically distributed. Using the Bloomfield model of the spectral density, the asymptotic expansion of risk of forecasting is constructed in the asymptotic of infinite length of forecasted sample and increasing length of training sample.

Задача прогнозирования временных рядов в случае известной их модели хорошо изучена [1–3]. В случае же неизвестной модели возникает задача ее идентификации по обучающим данным. При этом необходимы методы идентификации, робастные к наличию искажений в обучающих данных [4, 5] или к их малому объему [6–9]. В настоящей статье рассматривается задача прогнозирования стационарного гауссовского временного ряда с идентификацией модели Блумфилда [10, 11] по независимой обучающей последовательности.

Математическая модель и постановка задачи. Пусть $\{x_t : t \in \mathbb{Z}\}$ – гауссовский стационарный временной ряд с нулевым средним $\mathrm{E}\{x_t\} = 0$ и спектральной плотностью (сокращенно – спектром) $S : \Pi \to \mathbb{R}_+$, $\Pi = [-\pi, \pi]$, которая предполагается неизвестной. Для функций $K : \Pi^n \to \mathbb{R}$ будем обозначать:

$$[K]_j = (2\pi)^{-n} \int_{\Pi^n} K(z) \prod\nolimits_{r=1}^{n} \cos\left(j_r z_r\right) dz, \quad j = \left(j_r\right) \in \mathbb{Z}^n. \tag{1}$$

Далее будем предполагать, что у нас имеется подлежащая прогнозированию реализация $X = \left(x_j\right)_{j=1}^{T_f}$ длины $T_f \in \mathbb{N}$ и обучающая реализация $\tilde{X} = \left(\tilde{x}_j\right)_{j=1}^{T}$ длины $T \in \mathbb{N}$ временного ряда $\{\tilde{x}_t\}$, не зависящего от временного ряда $\{x_t\}$ и одинаково распределенного с ним.

Поскольку при обращении времени ($t' = -t$) сохраняются стационарность, гауссовость и спектральная плотность временного ряда [2], мы будем без потери общности рассматривать следующую задачу ретроспективного прогнозирования: по последовательностям $\tilde{X}$ и X построить прогноз для x_0 и исследовать его асимптотические свойства. Эта задача эквивалентна традиционной задаче прогнозирования [4] будущего значения x_{T_f+1} по $\tilde{X}$, X.