УДК 519.7

Непараметрическая оценка квадрата плотности вероятности и её свойства в условиях больших выборок

© Т. К. Юлдашев [1]

Аннотация. Эффективность непараметрических алгоритмов обработки информации, основанных на ядерных оценках квадрата плотности вероятности и коэффициента размытости, во многом определяется объемом статистических данных. Рассмотрены непараметрические оценки квадрата плотности вероятности типа Розенблатта-Парзена. Изучены асимптотические свойства непараметрической оценки квадрата плотности вероятности. Проверена сходимость непараметрической оценки квадрата плотности вероятности с увеличением объема экспериментальных данных к искомой функции квадрата плотности вероятности. Получена формула асимптотической несмещенности искомой оценки. Доказана сходимость в среднеквадратическом и состоятельность оценки квадрата плотности. При этом произведен вычислительный эксперимент для выявления зависимости среднеквадратической ошибки аппроксимации непараметрической оценки квадрата плотности вероятности от объемов статистических данных. Полученная в данной работе непараметрическая оценка квадрата плотности вероятности может быть использовано при построении алгоритмов принятия решений, когда основной исходной информацией являются статистические данные.

Ключевые слова: квадрат плотности вероятности, непараметрическая оценка, асимптотические свойства, оценка типа Розенблатта-Парзена, ядерная функция

Часто изучение технологических, экономических и социально-организационных систем связано с усложнением процессов принятия решений, что в значительной мере характерно для условий априорной неопределенности в закономерностях функционирования систем и их целевых установках. Использование традиционных методов моделирования и управления путем введения последовательности допущений не всегда позволяет получать удовлетворительные результаты.

Теоретическую основу построения обучающихся алгоритмов синтеза и анализа структуры сложных систем в условиях априорной неопределенности составляет задача непараметрического оценивания плотности вероятности.

Непараметрическая оценка квадрата плотности вероятности может быть использовано при построении алгоритмов принятия решений, когда основной исходной информацией являются статистические данные.

Применение непараметрической оценки плотности вероятности и методов восстановления законов распределения случайных величин позволяет получить статистики с улучшенными аппроксимационными свойствами при увеличении количества наблюдений [1] – [5]. Здесь важную роль играют методы, основанные на оценках плотности вероятности типа Розенблатта-Парзена [6] , [7].

1. Постановка задачи

Пусть имеется выборка $V = \left(x^i, i = \overline{1,n}\right)$ статистически независимых наблюдений случайной величины x, распределенной в соответствии с неизвестной плотностью вероятности $p(x)$.

[1] Доцент кафедры высшей математики, Сибирский государственный аэрокосмический университет имени академика М. Ф. Решетнева, г. Красноярск, tursunbay@rambler.ru

В качестве приближения по эмпирическим данным V плотности вероятности $p(x)$ примем статистику [6], [7]:

$$\overline{p}(x) = \frac{1}{nc}\sum_{i=1}^{n}\Phi\Big(\frac{x-x^i}{c}\Big), \tag{1.1}$$

где $\Phi(u)$ – ядерная функция, c – коэффициент размытости ядерной функции.

Коэффициенты размытости $c = c(n)$ ядерной функции в непараметрической оценке плотности вероятности (1.1) убывают с ростом n.

Объектом исследования является статистика:

$$\overline{p}^2(x) = \left(\frac{1}{nc}\sum_{i=1}^{n}\Phi\Big(\frac{x-x^i}{c}\Big)\right)^2. \tag{1.2}$$

Цель настоящей работы заключается в исследовании асимптотических свойств непараметрической оценки квадрата плотности вероятности (1.2) и их сравнении со свойствами статистики (1.1).

2. Асимптотические свойства непараметрической оценки $\overline{p}^2(x)$

Т е о р е м а 2.1. *Пусть выполняются следующие условия:*

1). Функция $p(x)$ ограничена и непрерывна со всеми своими производными до второго порядка включительно;

2). Ядерная функция $\Phi(u)$ является положительным и симметричным, а также $\int\limits_{-\infty}^{\infty}\Phi(u)du = \int\limits_{-\infty}^{\infty}u^2\Phi(u)du = 1$ *;*

3). $b^{\nu,2\mu} = \int\limits_{-\infty}^{\infty}u^{2\mu}\Phi^{\nu}(u)du < \infty$, $1 \le \mu \in \mathbb{N}$, $\nu \ge 1$*;*

4). $\int\limits_{-\infty}^{\infty}u^{2\mu-1}\Phi^{\nu}(u)du = 0$, $1 \le \mu \in \mathbb{N}$, $\nu \ge 1$*;*

5). $\lim_{n\to\infty} c(n) = 0$, $\lim_{n\to\infty} n\cdot c(n) = \infty$.

Тогда:

1) Математическое ожидание M непараметрической оценки $\overline{p}^2(x)$ при увеличении объема экспериментальных данных сходится к искомой функции квадрата плотности вероятности $p^2(x)$, т.е. справедлива формула асимптотической несмещенности оценки $\overline{p}^2(x)$:

$$\lim_{n\to\infty} M\Big(\overline{p}^2(x) - p^2(x)\Big) = 0; \tag{2.1}$$

2) Имеют место сходимость в среднеквадратическом:

$$\lim_{n\to\infty} M\Big(\overline{p}^2(x) - p^2(x)\Big)^2 = 0; \tag{2.2}$$

и состоятельность оценки квадрата плотности $\overline{p}^2(x)$:

$$\lim_{n\to\infty} M\Big(\overline{p}^2(x) - M\Big(\overline{p}^2(x)\Big)\Big)^2 = 0. \tag{2.3}$$

Д о к а з а т е л ь с т в о. Так как

$$\lim_{n\to\infty} M\Big(\overline{p}^{\,2}(x)-p^2(x)\Big)=\lim_{n\to\infty} M\Big(\overline{p}^{\,2}(x)\Big)-\lim_{n\to\infty} M\Big(p^2(x)\Big)=\lim_{n\to\infty} M\Big(\overline{p}^{\,2}(x)\Big)-p^2(x),$$

то для доказательства формулы (2.1) достаточно показать, что

$$\lim_{n\to\infty} M\Big(\overline{p}^{\,2}(x)\Big)=p^2(x). \tag{2.4}$$

Для $M\Big(\overline{p}^{\,2}(x)\Big)$ с помощью статистики (1.2) имеем

$$M\Big(\overline{p}^{\,2}(x)\Big)=M\left(\frac{1}{nc}\sum_{i=1}^{n}\Phi\Big(\frac{x-x^i}{c}\Big)\right)^2=M\left(\frac{1}{n^2c^2}\sum_{i=1}^{n}\Phi^2\Big(\frac{x-x^i}{c}\Big)\right)+$$

$$+M\left(\frac{1}{n^2c^2}\sum_{i=1}^{n}\sum_{\substack{j=1\\ j\neq i}}^{n}\Phi\Big(\frac{x-x^i}{c}\Big)\Phi\Big(\frac{x-x^j}{c}\Big)\right). \tag{2.5}$$

Для первого слагаемого в (2.5) имеем

$$M\left(\frac{1}{n^2c^2}\sum_{i=1}^{n}\Phi^2\Big(\frac{x-x^i}{c}\Big)\right)=\frac{1}{n^2c^2}\sum_{i=1}^{n}\int\limits_{-\infty}^{\infty}\Phi^2\Big(\frac{x-x^i}{c}\Big)p(x^i)dx^i.$$

Так как выборка независимых наблюдений $\Big(x^i, i=\overline{1,n}\Big)$ из генеральной совокупности X есть наблюдения одной и той же случайной величины, то

$$p(x^1)=p(x^2)=\ldots=p(x^n)=p(t).$$

Отсюда

$$M\left(\frac{1}{n^2c^2}\sum_{i=1}^{n}\Phi^2\Big(\frac{x-x^i}{c}\Big)\right)=\frac{1}{nc}\left[\frac{1}{c}\int\limits_{-\infty}^{\infty}\Phi^2\Big(\frac{x-t}{c}\Big)p(t)dt\right].$$

Производим замену переменных

$$\frac{x-t}{c}=u,\ t=x-cu,\ dt=-cdu.$$

Тогда, в силу условий теоремы, получаем

$$\frac{1}{c}\int\limits_{-\infty}^{\infty}\Phi^2\Big(\frac{x-t}{c}\Big)p(t)dt=\frac{1}{c}\int\limits_{-\infty}^{\infty}\Phi^2(u)p(x-cu)du.$$

Разложим функцию $p(x-cu)$ в ряд Тейлора в точке x:

$$p(x-cu)=p(x)+(x-cu-x)p^{(1)}(x)+\frac{(x-cu-x)^2}{2}p^{(2)}(x)+\ldots+O(c^4),$$

где $p^{(k)}(x)$ – производная k- го порядка функции $p(x)$.

Тогда имеем

$$\frac{1}{c}\int\limits_{-\infty}^{\infty}\Phi^2\Big(\frac{x-t}{c}\Big)p(t)dt =$$

$$= \int\limits_{-\infty}^{\infty}\Phi^2(u)\Big[p(x) - cup^{(1)}(x) + \frac{c^2u^2}{2!}p^{(2)}(x) + \ldots + O(c^4)\Big]du.$$

Поэтому, в силу условий теоремы, справедливо соотношение

$$M\left(\frac{1}{n^2c^2}\sum_{i=1}^{n}\Phi^2\Big(\frac{x-x^i}{c}\Big)\right) = \frac{1}{nc}\left[p(x)b^{2,0} + \frac{c^2}{2}p^{(2)}(x)b^{2,2} + \ldots + O(c^4)\right], \tag{2.6}$$

где $b^{2,0} = \int\limits_{-\infty}^{\infty}\Phi^2(u)du$, $b^{2,2} = \int\limits_{-\infty}^{\infty}u^2\Phi^2(u)du$.

Для второго слагаемого в (2.5) имеем

$$M\left(\frac{1}{n^2c^2}\sum_{i=1}^{n}\sum_{\substack{j=1\\ j\neq i}}^{n}\Phi\Big(\frac{x-x^i}{c}\Big)\Phi\Big(\frac{x-x^j}{c}\Big)\right) =$$

$$= \frac{1}{n^2c^2}\sum_{i=1}^{n}\sum_{\substack{j=1\\ j\neq i}}^{n}\left(\int\limits_{-\infty}^{\infty}\int\limits_{-\infty}^{\infty}\Phi\Big(\frac{x-x^i}{c}\Big)\Phi\Big(\frac{x-x^j}{c}\Big)p\big(x^i,x^j\big)dx^idx^j\right) =$$

$$= \frac{1}{n^2c^2}\sum_{i=1}^{n}\sum_{\substack{j=1\\ j\neq i}}^{n}\int\limits_{-\infty}^{\infty}\int\limits_{-\infty}^{\infty}\Phi^2\Big(\frac{x-t}{c}\Big)p^2(t)dt^2 = \frac{n(n-1)}{n^2}\left[\frac{1}{c}\int\limits_{-\infty}^{\infty}\Phi\Big(\frac{x-t}{c}\Big)p(t)dt\right]^2 =$$

$$= \frac{n-1}{n}\left[p(x) + \frac{c^2}{2}p^{(2)}(x) + \ldots + O(c^4)\right]^2 =$$

$$= \frac{n-1}{n}\left[p^2(x) + c^2p(x)p^{(2)}(x) + \frac{c^4}{4}\Big(p^{(2)}(x)\Big)^2 + O(\cdot)\right]. \tag{2.7}$$

Подстановка формул (2.6) и (2.7) в (2.5) дает

$$M\Big(\overline{p}^2(x)\Big) \sim \frac{1}{nc}\left[p(x)b^{2,0} + \frac{c^2}{2}p^{(2)}(x)b^{2,2}\right] +$$

$$+p^2(x) + c^2p(x)p^{(2)}(x) + \frac{c^4}{4}\Big(p^{(2)}(x)\Big)^2. \tag{2.8}$$

Переход к пределу при $n \to \infty$ в (2.8) с учетом (1.2) дает (2.4). Итак, мы доказали асимптотическую несмешенность непараметрической оценки $\overline{p}^2(x)$. Таким образом, доказана справедливость формулы (2.1).

Теперь докажем, что математическое ожидание непараметрической оценки $\overline{p}^4(x)$ при увеличении объема экспериментальных данных сходится к искомой функции плотности вероятности $p^4(x)$:

$$\lim_{n\to\infty} M\Big(\overline{p}^4(x)\Big) = p^4(x). \tag{2.9}$$

Действительно, имеем

$$M\Big(\overline{p}^{\,4}(x)\Big)=M\left(\frac{1}{nc}\sum_{i=1}^{n}\Phi\Big(\frac{x-x^i}{c}\Big)\right)^4=M\left(\frac{1}{n^4c^4}\sum_{i=1}^{n}\Phi^4\Big(\frac{x-x^i}{c}\Big)\right)+$$

$$+M\left(\frac{1}{n^4c^4}\sum_{i=1}^{n}\sum_{\substack{j=1\\ j\neq i}}^{n}\Phi\Big(\frac{x-x^i}{c}\Big)\Phi^3\Big(\frac{x-x^j}{c}\Big)\right)+$$

$$+M\left(\frac{1}{n^4c^4}\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{\substack{k=1\\ k\neq j}}^{n}\Phi\Big(\frac{x-x^i}{c}\Big)\Phi\Big(\frac{x-x^j}{c}\Big)\Phi^2\Big(\frac{x-x^k}{c}\Big)\right)+$$

$$+M\left(\frac{1}{n^4c^4}\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{n}\sum_{\substack{m=1\\ m\neq k}}^{n}\Phi\Big(\frac{x-x^i}{c}\Big)\Phi\Big(\frac{x-x^j}{c}\Big)\Phi\Big(\frac{x-x^k}{c}\Big)\Phi\Big(\frac{x-x^m}{c}\Big)\right). \tag{2.10}$$

Для первого слагаемого в (2.10) используем определение математического ожидания

$$M\left(\frac{1}{n^4c^4}\sum_{i=1}^{n}\Phi^4\Big(\frac{x-x^i}{c}\Big)\right)=\frac{1}{n^4c^4}\sum_{i=1}^{n}\int\limits_{-\infty}^{\infty}\Phi^4\Big(\frac{x-x^i}{c}\Big)p(x^i)dx^i=$$

$$=\frac{1}{n^4c^3}\sum_{i=1}^{n}\left[\frac{1}{c}\int\limits_{-\infty}^{\infty}\Phi^4\Big(\frac{x-t}{c}\Big)p(t)dt\right]=$$

$$=\frac{1}{n^3c^3}\left[p(x)b^{4,0}+\frac{c^2}{2}p^{(2)}(x)b^{4,2}+\ldots+O(c^4)\right]. \tag{2.11}$$

Для второго слагаемого в (2.10), в силу определения математического ожидания, имеем

$$M\left(\frac{1}{n^4c^4}\sum_{i=1}^{n}\sum_{\substack{j=1\\ j\neq i}}^{n}\Phi\Big(\frac{x-x^i}{c}\Big)\Phi^3\Big(\frac{x-x^j}{c}\Big)\right)=$$

$$=\frac{1}{n^4c^4}\sum_{i=1}^{n}\sum_{\substack{j=1\\ j\neq i}}^{n}\int\limits_{-\infty}^{\infty}\int\limits_{-\infty}^{\infty}\Phi\Big(\frac{x-x^i}{c}\Big)\Phi^3\Big(\frac{x-x^j}{c}\Big)p(x^i,x^j)dx^idx^j=$$

$$=\frac{1}{n^4c^4}\sum_{i=1}^{n}\sum_{\substack{j=1\\ j\neq i}}^{n}\int\limits_{-\infty}^{\infty}\int\limits_{-\infty}^{\infty}\Phi^4\Big(\frac{x-t}{c}\Big)p^2(t)dt^2=$$

$$=\frac{n(n-1)}{n^4c^2}\left[\frac{1}{c}\int\limits_{-\infty}^{\infty}\Phi^2\Big(\frac{x-t}{c}\Big)p(t)dt\right]^2\sim\frac{1}{n^2c^2}\left[p(x)b^{2,0}+\frac{c^2}{2}p^{(2)}(x)b^{2,2}\right]^2=$$

$$=\frac{1}{n^2c^2}\left[p^2(x)\Big(b^{2,0}\Big)^2+c^2p^{(2)}(x)p(x)b^{2,2}b^{2,0}+\frac{c^4}{4}\Big(p^{(2)}(x)\Big)^2\Big(b^{2,2}\Big)^2\right]. \tag{2.12}$$

Аналогично и для третьего слагаемого в (2.10), в силу определения математического ожидания, получаем асимптотическое разложение

$$M\left(\frac{1}{n^4c^4}\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{\substack{k=1\\k\neq j}}^{n}\Phi\Big(\frac{x-x^i}{c}\Big)\Phi\Big(\frac{x-x^j}{c}\Big)\Phi^2\Big(\frac{x-x^k}{c}\Big)\right)=$$

$$=\frac{1}{n^4c^4}\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{\substack{k=1\\k\neq j}}^{n}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\Phi^4\Big(\frac{x-t}{c}\Big)p^3(t)dt^3=$$

$$=\frac{n-1}{n^2c}\left[\frac{1}{c}\int_{-\infty}^{\infty}\Phi^{\frac{4}{3}}\Big(\frac{x-t}{c}\Big)p(t)dt\right]^3\sim\frac{1}{nc}\left[p(x)b^{\frac{4}{3},0}+\frac{c^2}{2}p^{(2)}(x)b^{\frac{4}{3},2}\right]^3=$$

$$=\frac{1}{nc}\left[p^3(x)\Big(b^{\frac{4}{3},0}\Big)^3+\frac{3c^2}{2}p^{(2)}(x)p^2(x)b^{\frac{4}{3},2}\Big(b^{\frac{4}{3},0}\Big)^2+\right.$$

$$\left.+\frac{3c^4}{4}p(x)\Big(p^{(2)}(x)\Big)^2\Big(b^{\frac{4}{3},2}\Big)^2b^{\frac{4}{3},0}+O(\cdot)\right]. \tag{2.13}$$

Для четвертого слагаемого в (2.10) имеем следующую апроксимационную оценку

$$M\left(\frac{1}{n^4c^4}\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{n}\sum_{\substack{m=1\\m\neq k}}^{n}\Phi\Big(\frac{x-x^i}{c}\Big)\Phi\Big(\frac{x-x^j}{c}\Big)\Phi\Big(\frac{x-x^k}{c}\Big)\Phi\Big(\frac{x-x^m}{c}\Big)\right)=$$

$$=\frac{n-1}{nc^4}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\Phi^4\Big(\frac{x-t}{c}\Big)p^4(t)dt^4\sim$$

$$\sim\left[\frac{1}{c}\int_{-\infty}^{\infty}\Phi\Big(\frac{x-t}{c}\Big)p(t)dt\right]^4\sim\left[p^2(x)+c^2p^{(2)}(x)p(x)+\frac{c^4}{4}\Big(p^{(2)}(x)\Big)^2\right]^2\sim$$

$$\sim p^4(x)+2c^2p^{(2)}(x)p^3(x)+\frac{c^4}{2}p^2(x)\Big(p^{(2)}(x)\Big)^2. \tag{2.14}$$

С учетом формул (2.11) – (2.14) из (2.10) получаем следующую асимптотическую формулу

$$M\Big(\overline{p}^4(x)\Big)\sim\frac{1}{n^3c^3}\left[p(x)b^{4,0}+\frac{c^2}{2}p^{(2)}(x)b^{4,2}\right]+$$

$$+\frac{1}{n^2c^2}\left[p^2(x)\Big(b^{2,0}\Big)^2+c^2p^{(2)}(x)p(x)b^{2,2}b^{2,0}+\frac{c^4}{4}\Big(p^{(2)}(x)\Big)^2\Big(b^{2,2}\Big)^2\right]+$$

$$+\frac{1}{nc}\left[p^3(x)\Big(b^{\frac{4}{3},0}\Big)^3+\frac{3c^2}{2}p^{(2)}(x)p^2(x)b^{\frac{4}{3},2}\Big(b^{\frac{4}{3},0}\Big)^2+\frac{3c^4}{4}p(x)\Big(p^{(2)}(x)\Big)^2\Big(b^{\frac{4}{3},2}\Big)^2b^{\frac{4}{3},0}\right]+$$

$$+p^4(x)+2c^2p^{(2)}(x)p^3(x)+\frac{c^4}{2}p^2(x)\Big(p^{(2)}(x)\Big)^2. \tag{2.15}$$

Переходя к пределу в (2.15) при $n \to \infty$, в силу условий теоремы, получаем (2.9). С помощью формулы (2.9) нетрудно показать, что имеет место сходимость в среднеквадратическом для оценки плотности $\overline{p}^2(x)$. Действительно, из следующего соотношения

$$\lim_{n\to\infty} M\Big(\overline{p}^2(x)-p^2(x)\Big)^2 = \lim_{n\to\infty} M\Big(\overline{p}^4(x)\Big) - 2\lim_{n\to\infty} M\Big(\overline{p}^2(x)\Big)p^2(x)+p^4(x)$$

с учетом (2.4) и (2.9) получаем, что справедлива формула (2.2).

Поскольку $\overline{p}^2(x)$ является асимптотически несмещенной оценкой для $p^2(x)$ и сходится в среднеквадратическом, то она обладает свойством состоятельности, т.е. справедлива формула (2.3).

Д о к а з а т е л ь с т в о з а к о н ч е н о.

3. Мера близости между искомой плотностью вероятностей $p^2(x)$ и ее оценкой $\overline{p}^2(x)$

С учетом формул (2.8) и (2.15) имеем

$$M\Big(\overline{p}^2(x)-p^2(x)\Big)^2 = M\Big(\overline{p}^4(x)\Big) - 2M\Big(\overline{p}^2(x)\Big)p^2(x)+p^4(x) \sim$$

$$\sim \frac{1}{n^3c^3}\left[p(x)b^{4,0}+\frac{c^2}{2}p^{(2)}(x)b^{4,2}\right]+$$

$$+\frac{1}{n^2c^2}\left[p^2(x)\Big(b^{2,0}\Big)^2+c^2p^{(2)}(x)p(x)b^{2,2}b^{2,0}+\frac{c^4}{4}\Big(p^{(2)}(x)\Big)^2\Big(b^{2,2}\Big)^2\right]+$$

$$+\frac{1}{nc}\left[p^3(x)\Big(b^{\frac{4}{3},0}\Big)^3+\frac{3c^2}{2}p^{(2)}(x)p^2(x)b^{\frac{4}{3},2}\Big(b^{\frac{4}{3},0}\Big)^2+\frac{3c^4}{4}p(x)\Big(p^{(2)}(x)\Big)^2\Big(b^{\frac{4}{3},2}\Big)^2b^{\frac{4}{3},0}\right]+$$

$$+p^4(x)+2c^2p^{(2)}(x)p^3(x)+\frac{c^4}{2}p^2(x)\Big(p^{(2)}(x)\Big)^2-2p^2(x)\left\{\frac{1}{nc}\left[p(x)b^{2,0}+\frac{c^2}{2}p^{(2)}(x)b^{2,2}\right]+\right.$$

$$\left.+p^2(x)+c^2p(x)p^{(2)}(x)+\frac{c^4}{4}\Big(p^{(2)}(x)\Big)^2\right\}+p^4(x)$$

или, пренебрегая слагаемыми малости некоторых величин, получаем

$$M\Big(\overline{p}^2(x)-p^2(x)\Big)^2 \sim \frac{1}{nc}\left\{p^3(x)\left[\Big(b^{\frac{4}{3},0}\Big)^3-2b^{2,0}\right]+\right.$$

$$\left.+c^2p^2(x)p^{(2)}(x)\left[\frac{3}{2}\Big(b^{\frac{4}{3},0}\Big)^2b^{\frac{4}{3},2}-2b^{2,2}\right]+\frac{3c^4}{4}p(x)\Big(p^{(2)}(x)\Big)^2\Big(b^{\frac{4}{3},2}\Big)^2b^{\frac{4}{3},0}\right\}. \tag{3.1}$$

Для определения условий сходимости на всей области изменения x проинтегрируем (3.1)

$$M\int_{-\infty}^{\infty}\Big(\overline{p}^2(x)-p^2(x)\Big)^2dx \sim \frac{1}{nc}\left\{\Big\|p^3(x)\Big\|\left[\Big(b^{\frac{4}{3},0}\Big)^3-2b^{2,0}\right]+\right.$$

$$\left.+c^2\Big\|p^2(x)p^{(2)}(x)\Big\|\left[\frac{3}{2}\Big(b^{\frac{4}{3},0}\Big)^2b^{\frac{4}{3},2}-2b^{2,2}\right]+\frac{3c^4}{4}\Big\|p(x)\Big(p^{(2)}(x)\Big)^2\Big\|\Big(b^{\frac{4}{3},2}\Big)^2b^{\frac{4}{3},0}\right\}. \tag{3.2}$$

где $\Big\|p(x)\Big\| = \int\limits_{-\infty}^{\infty} p(x)dx$.

Величина критерия (3.2) представляет собой меру близости между искомой плотностью вероятностей $p^2(x)$ и ее асимптотической оценкой $\overline{p}^2(x)$. Эта величина при конечном объеме выборки зависит от коэффициента размытости c и ядерной функции $\Phi(u)$.

4. Анализ свойств статистики $\overline{p}^2(x)$

Рассмотрим следующую функцию [8]

$$W_1(n,c) = \frac{M\int\limits_{-\infty}^{\infty}\Big(\overline{p}(x)-p(x)\Big)^2 dx}{\int\limits_{-\infty}^{\infty} p^2(x)dx} \sim \frac{\frac{1}{nc}\|\Phi(u)\|^2 + \frac{c^4\|p^{(2)}(x)\|^2}{4}}{\|p^2(x)\|}. \tag{4.1}$$

Проведем для конкретных условий анализ асимптотических выражений относительных среднеквадратических ошибок аппроксимаций (4.1). С этой целью рассмотрим функцию

$$W_2(n,c) = \frac{M\int\limits_{-\infty}^{\infty}\Big(\overline{p}^2(x)-p^2(x)\Big)^2 dx}{\int\limits_{-\infty}^{\infty} p^4(x)dx} \sim \frac{1}{nc\|p^4(x)\|}\times$$

$$\times\left\{\Big\|p^3(x)\Big\|\left[\Big(b^{\frac{4}{3},0}\Big)^3 - 2b^{2,0}\right] + c^2\Big\|p^2(x)p^{(2)}(x)\Big\|\left[\frac{3}{2}\Big(b^{\frac{4}{3},0}\Big)^2 b^{\frac{4}{3},2} - 2b^{2,2}\right] + \right.$$

$$\left. + \frac{3c^4}{4}\Big\|p(x)\Big(p^{(2)}(x)\Big)^2\Big\|\Big(b^{\frac{4}{3},2}\Big)^2 b^{\frac{4}{3},0}\right\}. \tag{4.2}$$

Приведённые выше отношения зависят от $p(x)$, вида ядерных функций и объёма n исходных статистических данных.

В качестве восстанавливаемой плотности вероятности рассмотрим нормированную функцию Лапласа

$$p(x) = \frac{1}{\sqrt{2\pi}}\exp\Big[-\frac{x^2}{2}\Big].$$

В качестве ядерной функции $\Phi(u)$ будем использовать оптимальное ядро Епанечникова [1]

$$\Phi(u) = \begin{cases} \dfrac{3}{4\sqrt{5}} - \dfrac{3u^2}{20\sqrt{5}}, \ |u| < \sqrt{5}, \\ \qquad\qquad 0, \ |u| \geq \sqrt{5}. \end{cases}$$

В качестве коэффициента размытости c в (4.1) и (4.2) берем оптимальный коэффициент размытости [8]

$$c = \left\{\frac{b^{2,0}}{n\Big\|\Big(p^{(2)}(x)\Big)^2\Big\|}\right\}^{\frac{1}{5}}.$$

Тогда небольшой вычислительный эксперимент показывает, что среднеквадратические ошибки аппроксимаций (4.1) и (4.2) становятся малыми для больших значений n и стремятся к нулю при $n \to \infty$. При этом среднеквадратическая ошибка аппроксимации (4.2) квадрата плотности вероятности стремится к нулю быстрее, чем среднеквадратическая ошибка аппроксимации (4.1) плотности вероятности. Отсюда следует, что статистика (1.2) обладает лучшими асимптотическими свойствами, чем статистика (1.1).

5. Заключение

Непараметрическая оценка квадрата плотности вероятности используется при построении обучающихся алгоритмов синтеза и анализа структуры сложных систем в условиях априорной неопределенности.

В данной работе проверяется сходимость непараметрической оценки $\overline{p}^2(x)$ с увеличением объема экспериментальных данных к искомой функции квадрата плотности вероятности $p^2(x)$. Доказываются: асимптотическая несмещенность, среднеквадратическая сходимость и состоятельность оценки $\overline{p}^2(x)$.

Вычислительная эффективность непараметрических алгоритмов обработки информации, основанных на ядерных оценках квадрата плотности вероятности, во многом определяется объемом статистических данных. Определяется критерий меры близости между искомой плотностью вероятностей $p^2(x)$ и ее оценкой $\overline{p}^2(x)$ и эта величина при конечном объеме выборки зависит от коэффициента размытости c и ядерной функции $\Phi(u)$. В условиях большого объема выборки доказывается, что статистика (1.2) обладает лучшими асимптотическими свойствами, чем статистика (1.1).

Список литературы

1. Епанечников В. А., “Непараметрическая оценка многомерной плотности вероятности”, *Теория вероятностей и ее применения*, **14**:1 (1969), 156 – 161.

2. Мания Г. М., *Статистическое оценивание распределения вероятностей*, ТбГУ, Тбилиси, 1974, 238 с.

3. Надарая Э. А., “Об оценке плотностей распределения случайных величин”, *Сообщ. АН ГрССР*, **32** (1964), 277 – 280.

4. Лапко А. В., Лапко В. А., *Непараметрические системы обработки неоднородной информации*, Наука, Новосибирск, 2007, 174 с.

5. Тарасенко Ф. П., Дмитриев Ю. Г., “Об одном классе непараметрических оценок нелинейных функционалов плотности”, *Теория вероятностей и ее применения*, **19**:2 (1974), 404 – 409.

6. Parzen E., “On estimation of a probability density function and mode”, *Ann. Math. Statistic*, **33** (1962), 1065 – 1076.

7. Rosenblatt M., “Remarks on some nonparametric estimates of a density function”, *Ann. Math. Statistic*, **27** (1956), 642 – 669.

8. Лапко А. В., Лапко В. А., “Свойства непараметрической оценки плотности вероятности многомерных случайных величин в условиях больших выборок”, *Информатика и системы управления*, 2012, №2 (32), 121 – 126.

Nonparametric estimate of quadrate of the probability densities and its properties under the large sample

© T. K. Yuldashev[2]

Abstract. Efficiency of nonparametric data processing algorithms, based on kernel estimators of square of probability density and blur factor, largely determined by the volume of statistical data. It is considered in this article the nonparametric estimate of quadrate of the probability densities of Rosenblatt-Parzen type. It is studied the asymptotic properties of the nonparametric estimate of quadrate of the probability densities. It is tested the convergence of nonparametric estimates of the probability densities with increasing amount of experimental data to the desired probability density function. It is obtained the formula of asymptotically unbiased of the desired estimates. It is proved the mean-square convergence and the solvency of quadrate density. It is produced the computer experiment to determine the dependence of mean square error of nonparametric estimate approximation of the probability densities quadrate from the volume of statistical data. The nonparametric estimate of quadrate of the probability densities, obtained in this article can be used for constructing decision algorithms when the primary source of information is statistical data.

Key Words: quadrate of probability density, nonparametric estimate, asymptotic properties, Rosenblatt-Parzen type estimation, kernel function

[2] Associate professor of Higher Mathematics Chair, M. F. Reshetnev Siberian State Aerospace University, Krasnoyarsk, tursunbay@rambler.ru