УДК 519. 7
ББК 22.17

ИНТЕГРАЛЬНЫЕ ОЦЕНКИ КВАДРАТА ПЛОТНОСТИ ВЕРОЯТНОСТИ И ИХ АИСМПТОТИЧЕСКИЕ СВОЙСТВА

Юлдашев Т. К.[1]
(Сибирский государственный аэрокосмический университет, *Красноярск*)

Рассматриваются методы получения интегральных оценок для квадрата плотности вероятности. При ограниченном объеме обучающей выборки оценки плотности вероятности типа Розенблатта-Парзена могут оказаться негладкой. Интегральные оценки сглаживают эти оценки плотности вероятности таким образом, чтобы улучшить их аппроксимационные свойства. С помощью двух положительных, нормированных и симметричных ядерных функций изучаются асимптотические свойства интегральной оценки квадрата плотности вероятности. Проверяется сходимость интегральной оценки квадрата плотности вероятности с увеличением объема экспериментальных данных к искомой функции квадрата плотности вероятности. Доказываются асимптотическая несмещенность и среднеквадратическая сходимость интегральной оценки квадрата плотности вероятности с увеличением объема статистики. Также устанавливается состоятельность интегральной оценки квадрата плотности вероятности с увеличением объема экспериментальных данных.

Ключевые слова: квадрат плотности вероятности, интегральная оценка, асимптотическая несмещенность, асимптотическая состоятельность, ядерные функции.

---

[1] *Турсун Камалдинович Юлдашев, кандидат физико-математических наук, доцент* (*tursunbay@ramblerl.ru*).

1. Введение

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 в данной ситуации 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 - 9].

При ограниченном объеме обучающей выборки оценка плотности вероятности типа Розенблатта-Парзена может оказаться негладкой, скачкообразной, что негативно сказывается при решении некоторых задач распознавания образов либо автоматической классификации. Здесь дополнительно возникает вопрос о сглаживании оценки плотности вероятности таким образом, чтобы по возможности улучшить аппроксимационные свойства оценки.

В данной работе рассматриваются интегральные оценки квадрата плотности вероятности. Изучаются асимптотические свойства интегральной оценки квадрата плотности вероятности. Проверяется сходимость интегральной оценки квадрата плотности вероятности с увеличением объема экспериментальных данных к искомой функции квадрата плотности вероятности.

2. 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 $p(x)$.

В качестве приближения по эмпирическим данным V плотности вероятности $p(x)$ примем статистику [8 , 9]:

$$\bar{p}(x)=\frac{1}{nc}\sum_{i=1}^{n}\Phi\left(\frac{x-x^i}{c}\right), \tag{1}$$

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

Коэффициенты размытости $c=c(n)$ ядерной функции в непараметрической оценке плотности вероятности (1) убывают с ростом n.

Интегральная оценка плотности вероятности в точке x представляется статистикой

$$\bar{p}_1(x)=\frac{1}{nc}\sum_{i=1}^{n}\int_D\Phi\left(\frac{x-\beta u-x^i}{c}\right)h(u)\,du, \tag{2}$$

где $h(u)-$ функция ядерного типа, $\beta-$ параметр.

Отметим, что оценка типа Розенблатта-Парзена (1) является частным случаем интегральной оценки плотности вероятности (2) в случае, когда $\beta=0$.

Объектом исследования является статистика

(3) $$\bar{p}_1^{\,2}(x) = \left(\frac{1}{nc} \sum_{i=1}^{n} \int_D \Phi\left(\frac{x - \beta u - x^i}{c} \right) h(u)\, du \right)^2 .$$

Цель настоящей работы заключается в исследовании асимптотических свойств интегральной оценки квадрата плотности вероятности (3) и в сравнении полученных результатов при различных значениях параметра β относительно оптимального коэффициента размытости c .

3. Асимптотические свойства интегральной оценки $\bar{p}_1^{\,2}(x)$.

Теорема. Пусть выполняются условия:

1). $\lim\limits_{n\to\infty} c(n) = \lim\limits_{n\to\infty} \beta(n) = 0,\ \lim\limits_{n\to\infty} n \cdot c(n) = \infty;$

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

3). Ядерные функции $\Phi(u)$, $h(u)$ являются положительными и симметричными, $\int_D h(u)\, du = 1$, $\int_{-\infty}^{\infty} \Phi(u) du =$

$$= \int_{-\infty}^{\infty} u^2 \Phi(u) du = 1;$$

4). $b^{\nu,2\mu} = \int_{-\infty}^{\infty} u^{2\mu} \Phi^{\nu}(u) du < \infty,\ 1 \le \mu \in N,\ \nu \ge 1;$

5). $\int_{-\infty}^{\infty} u^{2\mu-1} \Phi^{\nu}(u) du = 0,\ 1 \le \mu \in N,\ \nu \ge 1;$

6). $a^{2\mu} = \int_D u^{2\mu} h(u)\, du < \infty,\ 1 \le \mu \in N;$

7). $\int_D u^{2\mu-1} h(u)\, du = 0,\ 1 \le \mu \in N,\ N \equiv \{1, 2, 3, \ldots\}.$

Тогда: 1) Математическое ожидание M интегральной оценки $\overline{p}_1^2(x)$ при увеличении объема экспериментальных данных сходится к искомой функции квадрата плотности вероятности $p^2(x)$, т.е. справедлива формула асимптотической несмещенности оценки $\overline{p}_1^2(x)$:

(4) $$\lim_{n\to\infty} M\left(\overline{p}_1^2(x) - p^2(x)\right) = 0;$$

2) Имеют место сходимость в среднеквадратическом для оценки квадрата плотности $\overline{p}_1^2(x)$:

(5) $$\lim_{n\to\infty} M\left(\overline{p}_1^2(x) - p^2(x)\right)^2 = 0$$

и состоятельность оценки квадрата плотности $\overline{p}_1^2(x)$, т.е. справедливо предельное соотношение

(6) $$\lim_{n\to\infty} M\left(\overline{p}_1^2(x) - M\left(\overline{p}_1^2(x)\right)\right)^2 = 0.$$

Доказательство. Так как

$$\lim_{n\to\infty} M\left(\overline{p}_1^2(x) - p^2(x)\right) =$$
$$= \lim_{n\to\infty} M\left(\overline{p}_1^2(x)\right) - \lim_{n\to\infty} M\left(p^2(x)\right) = \lim_{n\to\infty} M\left(\overline{p}_1^2(x)\right) - p^2(x),$$

то для доказательства формулы (4) достаточно показать, что

(7) $$\lim_{n\to\infty} M\left(\overline{p}_1^2(x)\right) = p^2(x).$$

С помощью интегральной оценки (3) имеем

(8) $$M\left(\overline{p}_1^2(x)\right) = M\left(\frac{1}{nc}\sum_{i=1}^{n}\int_D \Phi\left(\gamma(x^i)\right)h(u)\,du\right)^2 =$$
$$= A_1 + A_2 = M\left(\frac{1}{n^2c^2}\sum_{i=1}^{n}\int_D \Phi^2\left(\gamma(x^i)\right)h(u)\,du\right) +$$
$$+ M\left(\frac{1}{n^2c^2}\sum_{i=1}^{n}\sum_{\substack{j=1\\ i\neq j}}^{n}\int_D \Phi\left(\gamma(x^i)\right)\Phi\left(\gamma(x^j)\right)h(u)\,du\right),$$

где $\gamma(x^i) = \dfrac{x - \beta u - x^i}{c}$.

Согласно определению математического ожидания для первого слагаемого в (8) имеем

$$A_1 = \frac{1}{n^2 c^2} \sum_{i=1}^{n} \int_{-\infty}^{\infty} \int_{D} \Phi^2\left(\gamma(x^i)\right) h(u)\, p(x^i)\, du\, dx^i.$$

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

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

Поэтому

$$A_1 = \frac{1}{nc}\left[\frac{1}{c} \int_{-\infty}^{\infty} \int_{D} \Phi^2\left(\gamma(t)\right) h(u)\, p(t)\, du\, dt\right].$$

В последнем выражении производим замену переменных

$$\gamma(t) = \frac{x - \beta u - t}{c} = z,\; t = x - \beta u - cz,\; dt = -c\, dz.$$

Отсюда получаем, что

$$A_1 = \frac{1}{nc} \int_{-\infty}^{\infty} \int_{D} \Phi^2(z)\, h(u)\, p(x - \beta u - cz)\, du\, dz\ .$$

Теперь функцию $p(x - \beta u - cz)$ разложим в ряд Тейлора в точке x:

$$p(x - \beta u - cz) = p(x) - (\beta u + cz)\, p^{(1)}(x) +$$
$$+ \frac{(\beta u + cz)^2}{2!}\; p^{(2)}(x) + \ldots + O\left(c^4, \beta^4\right),$$

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

Тогда имеем

$$A_1 = \frac{1}{nc}\left\{ p(x) \int_{-\infty}^{\infty} \Phi^2(z) \int_{D} h(u)\, du\, dz - \right. \tag{9}$$

$$-p^{(1)}(x)\left[\beta\int_{-\infty}^{\infty}\Phi^2(z)\int_D u\,h(u)\,du\,dz+\right.$$

$$\left.+c\int_{-\infty}^{\infty}z\,\Phi^2(z)\int_D h(u)\,du\,dz\right]+\frac{p^{(2)}(x)}{2}\left[\beta^2\int_{-\infty}^{\infty}\Phi^2(z)\int_D u^2h(u)\,du\,dz+\right.$$

$$+2\,\beta\,c\int_{-\infty}^{\infty}z\,\Phi^2(z)\int_D u\,h(u)\,du\,dz+$$

$$\left.\left.+c^2\int_{-\infty}^{\infty}z^2\,\Phi^2(z)\int_D h(u)\,du\,dz\right]+\ldots+O\left(c^4,\beta^4\right)\right\}.$$

Из условий теоремы следует, что

$$\int_{-\infty}^{\infty}\Phi^2(z)\int_D u\,h(u)\,du\,dz=\int_{-\infty}^{\infty}z\,\Phi^2(z)\int_D h(u)\,du\,dz=$$

$$=\int_{-\infty}^{\infty}z\,\Phi^2(z)\int_D u\,h(u)\,du\,dz=0\,.$$

С учетом это из (9) имеем

(10) $$A_1\sim\frac{1}{nc}\left[b^{2,0}p(x)+\frac{b^{2,0}a^2\beta^2+b^{2,2}c^2}{2}p^{(2)}(x)\right],$$

где

$$a^k=\int_D u^k\,h(u)\,du<\infty,\;\; b^{\nu,k}=\int_{-\infty}^{\infty}z^k\,\Phi^{\nu}(z)\,dz<\infty\,,\;\; k=0,2,4,6,\ldots\,.$$

Для второго слагаемого в (8) имеем

(11) $$A_2=M\left(\frac{1}{n^2c^2}\sum_{i=1}^{n}\sum_{\substack{j=1\\ i\neq j}}^{n}\int_D\Phi\left(\gamma(x^i)\right)\Phi\left(\gamma(x^j)\right)h(u)\,du\right)=$$

$$=\frac{1}{n^2c^2}\sum_{i=1}^{n}\sum_{\substack{j=1\\ i\neq j}}^{n}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_D\Phi\left(\gamma(x^i)\right)\Phi\left(\gamma(x^i)\right)h(u)\,p(x^i,x^j)\,du\,dx^i dx^j=$$

$$=\frac{n-1}{nc^2}\sum_{i=1}^{n}\sum_{\substack{j=1\\ i\neq j}}^{n}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_D\Phi^2\left(\gamma(t)\right)h(u)\,p^2(t)\,du\,dt^2=$$

$$= \frac{n-1}{n} \int_D h(u) \left[\int_{-\infty}^{\infty} \Phi(z) p(x - \beta u - cz) dz \right]^2 du \sim \int_D h(u) \times$$

$$\times \left[\int_{-\infty}^{\infty} \Phi(z) \left(p(x) - (\beta u + cz) p^{(1)}(x) + \frac{(\beta u + cz)^2}{2!} p^{(2)}(x) \right) dz \right]^2 du \sim$$

$$\sim p^2(x) + \left(a^2 \beta^2 + c^2\right) p(x) p^{(2)}(x) + a^2 \beta^2 \left(p^{(1)}(x)\right)^2 .$$

Подстановка формул (10) и (11) в (8) и учет малости величин (c^4, β^4) дают

(12) $$M\left(\overline{p}_1^2(x)\right) = A_1 + A_2 \sim$$

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

$$+ p^2(x) + \left(a^2 \beta^2 + c^2\right) p(x) p^{(2)}(x) + a^2 \beta^2 \left(p^{(1)}(x)\right)^2 .$$

Переход к пределу при $n \to \infty$ в (12) дает (7).

Итак, мы доказали асимптотическую несмешенность интегральной оценки $\overline{p}_1^2(x)$, т.е. доказана справедливость формулы (4).

Теперь докажем, что математическое ожидание интегральной оценки $\overline{p}_1^4(x)$ при увеличении объема экспериментальных данных сходится к искомой функции плотности вероятности $p_1^4(x)$, т.е. справедлива формула

(13) $$\lim_{n \to \infty} M\left(\overline{p}_1^4(x)\right) = p^4(x).$$

Действительно, используя формулу (2), имеем

(14) $$M\left(\overline{p}_1^4(x)\right) = M\left(\frac{1}{nc} \sum_{i=1}^{n} \int_D \Phi\left(\gamma(x^i)\right) h(u) du \right)^4 =$$

$$= B_1 + B_2 + B_3 + B_4 = \frac{1}{n^4 c^4} M\left(\sum_{i=1}^{n} \int_D \Phi^4\left(\gamma(x^i)\right) h(u) du \right) +$$

$$+\frac{1}{n^4c^4}M\left(\sum_{i=1}^{n}\sum_{\substack{j=1\\ i\neq j}}^{n}\int_D \Phi\Big(\gamma(x^i)\Big)\Phi^3\Big(\gamma(x^j)\Big)h(u)\,du\right)+$$

$$+\frac{1}{n^4c^4}M\left(\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{\substack{k=1\\ j\neq k}}^{n}\int_D \Phi\Big(\gamma(x^i)\Big)\Phi\Big(\gamma(x^j)\Big)\Phi^2\Big(\gamma(x^k)\Big)h(u)\,du\right)+$$

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

$$\left.\times\Phi\Big(\gamma(x^k)\Big)\Phi\Big(\gamma(x^m)\Big)h(u)\,du\right).$$

Для первого слагаемого в (14) используем определение математического ожидания

(15) $$B_1=\frac{1}{n^4c^4}\sum_{i=1}^{n}\int_{-\infty}^{\infty}\int_D \Phi^4\Big(\gamma(x^i)\Big)h(u)\,p(x^i)\,du\,dx^i=$$

$$=\frac{1}{n^4c^3}\sum_{i=1}^{n}\left[\frac{1}{c}\int_{-\infty}^{\infty}\int_D \Phi^4\big(\gamma(t)\big)h(u)\,p(t)\,du\,dt\right]\sim$$

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

Для второго слагаемого в (14) используем определение математического ожидания

(16) $$B_2=\frac{1}{n^4c^4}\sum_{i=1}^{n}\sum_{\substack{j=1\\ i\neq j}}^{n}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_D \Phi\Big(\gamma(x^i)\Big)\Phi^3\Big(\gamma(x^j)\Big)\times$$

$$\times h(u)\,p(x^i,x^j)\,du\,dx^i dx^j=$$

$$=\frac{1}{n^4c^4}\sum_{i=1}^{n}\sum_{\substack{j=1\\ i\neq j}}^{n}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_D \Phi^4\big(\gamma(t)\big)h(u)\,p^2(t)\,du\,dt^2=$$

$$=\frac{n(n-1)}{n^4c^2}\int_D h(u)\left[\frac{1}{c}\int_{-\infty}^{\infty}\Phi^2\big(\gamma(t)\big)p(t)\,dt\right]^2du\sim\frac{1}{n^2c^2}\times$$

$$\times\left[\left(b^{2,0}\right)^2 p^2(x)+\left(a^2\beta^2+b^{2,2}c^2\right)b^{2,0}\,p(x)\,p^{(2)}(x)+a^2\beta^2\left(p^{(1)}(x)\right)^2\right].$$

Для третьего слагаемого в (14) используем определение математического ожидания

(17) $$B_3=\frac{1}{n^4c^4}\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{\substack{k=1\\ j\neq k}}^{n}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{D}\Phi\left(\gamma(x^i)\right)\Phi\left(\gamma(x^j)\right)\times$$

$$\times\Phi^2\left(\gamma(x^k)\right)h(u)\,p(x^i,x^j,x^k)\,du\,dx^i dx^j\,dx^k=$$

$$=\frac{1}{n^4c^4}\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{\substack{k=1\\ j\neq k}}^{n}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{D}\Phi^4\left(\gamma(t)\right)h(u)\,p^3(t)\,du\,dt^3=$$

$$=\frac{n-1}{n^2c}\int_D h(u)\left[\frac{1}{c}\int_{-\infty}^{\infty}\Phi^{\frac{4}{3}}\left(\gamma(t)\right)p(t)\,dt\right]^3 du\sim\frac{1}{nc}\int_D h(u)\times$$

$$\times\left[\int_{-\infty}^{\infty}\Phi^{\frac{4}{3}}(z)\left(p(x)-(\beta u+cz)\,p^{(1)}(x)+\frac{(\beta u+cz)^2}{2!}p^{(2)}(x)\right)dz\right]^3 du\sim$$

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

Для четвертого слагаемого в (14) имеем оценку

(18) $$B_4=\frac{1}{n^4c^4}\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{n}\sum_{\substack{m=1\\ k\neq m}}^{n}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{D}\Phi\left(\gamma(x^i)\right)\Phi\left(\gamma(x^j)\right)\times$$

$$\times\Phi\left(\gamma(x^k)\right)\Phi\left(\gamma(x^m)\right)h(u)\times$$

$$\times\,p\,(x^i,x^j,x^k,x^m)\,du\,dx^i dx^j\,dx^k\,dx^m=$$

$$=\frac{n-1}{nc^4}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{D}\Phi^4\left(\gamma(x^i)\right)h(u)\,p^4(t)\,du\,dt^4\sim$$

$$\sim\int_D h(u)\left[\frac{1}{\mathrm{c}}\int_{-\infty}^{\infty}\Phi\left(\gamma(t)\right)p(t)\,dt\right]^4 du\sim$$

$$\sim\left[p^2(x)+\left(a^2\beta^2+c^2\right)p(x)\,p^{(2)}(x)+a^2\beta^2\left(p^{(1)}(x)\right)^2\right]^2\sim$$

$$= p^4(x) + 2\left(a^2\beta^2 + c^2\right)p^3(x)\,p^{(2)}(x) + 2\,a^2\beta^2 p^2(x)\left(p^{(1)}(x)\right)^2 +$$
$$+\left[\left(a^2\beta^2 + c^2\right)p(x)\,p^{(2)}(x) + a^2\beta^2\left(p^{(1)}(x)\right)^2\right]^2.$$

С учетом формул (15) - (18) из (14) получаем

$$M\left(\bar{p}_1^4(x)\right) \sim \frac{1}{n^3c^3}\left[b^{4,0}\,p(x) + \frac{a^2\beta^2 + b^{4,2}c^2}{2}\,p^{(2)}(x)\right] + \tag{19}$$
$$+\frac{1}{n^2c^2}\left[\left(b^{2,0}\right)^2 p^2(x) + \right.$$
$$\left. +\left(a^2\beta^2 + b^{2,2}c^2\right)b^{2,0}\,p(x)\,p^{(2)}(x) + a^2\beta^2\left(p^{(1)}(x)\right)^2\right] +$$
$$+\frac{1}{nc}\left[p^3(x)\left(b^{\frac{4}{3},0}\right)^3 + \frac{3}{2}\left(b^{\frac{4}{3},0}\right)^2\left(a^2\beta^2 + b^{\frac{4}{3},2}c^2\right)p^3(x)p^{(2)}(x)\right] +$$
$$+ p^4(x) + 2\left(a^2\beta^2 + c^2\right)p^3(x)\,p^{(2)}(x) + 2\,a^2\beta^2 p^2(x)\left(p^{(1)}(x)\right)^2 +$$
$$+\left[\left(a^2\beta^2 + c^2\right)p(x)\,p^{(2)}(x) + a^2\beta^2\left(p^{(1)}(x)\right)^2\right]^2.$$

Переходя к пределу в (19) при $n \to \infty$, в силу условий теоремы, получаем (13).

С помощью формулы (13) покажем, что имеет место сходимость в среднеквадратическом для оценки плотности $\bar{p}_1^2(x)$. Так как

$$M\left(\bar{p}_1^2(x) - p^2(x)\right)^2 = \tag{20}$$
$$= M\left(\bar{p}_1^4(x) - 2\,\bar{p}_1^2(x)\,p^2(x) + p^4(x)\right) =$$
$$= M\left(\bar{p}_1^4(x)\right) - 2M\left(\bar{p}_1^2(x)\right)p^2(x) + p^4(x),$$

то, переходя к пределу в (20) при $n \to \infty$, с учетом (7) и (13) получаем (5):

$$\lim_{n\to\infty} M\left(\bar{p}_1^2(x) - p^2(x)\right)^2 =$$

$$= \lim_{n\to\infty} M\left(\overline{p}_1^{\,4}(x)\right) - 2\lim_{n\to\infty} M\left(\overline{p}_1^{\,2}(x)\right)p^2(x) + p^4(x) =$$

$$= p^4(x) - 2p^2(x)\,p^2(x) + p^4(x) = 0.$$

Поскольку $\overline{p}_1^{\,2}(x)$ является асимптотически несмещенной оценкой $p^2(x)$ и сходится в среднеквадратическом, то она обладает свойством состоятельности, т.е. справедлива формула (6). Теорема доказана.

Обращаемся к формуле (20). Подставляя (12) и (19) в (20) и пренебрегая слагаемыми малости некоторых величин, получаем

$$M\left(\overline{p}_1^{\,2}(x) - p^2(x)\right)^2 \sim \frac{1}{nc}\left\{\left[\left(b^{\frac{4}{3},0}\right)^3 - 2\,b^{2,0}\right]p^3(x) + \right.$$
$$+\frac{3}{2}\left(a^2\beta^2 + b^{\frac{4}{3},2}c^2\right)\left(b^{\frac{4}{3},0}\right)^2 p^3(x)p^{(2)}(x) -$$
$$\left. -2\left(a^2\beta^2 + b^{2,2}c^2\right)p^2(x)\,p^{(2)}(x)\right\} +$$
$$+\left[\left(a^2\beta^2 + c^2\right)p(x)\,p^{(2)}(x) + a^2\beta^2\left(p^{(1)}(x)\right)^2\right]^2. \tag{21}$$

Для определения условий сходимости на всей области изменения x проинтегрируем (21)

$$M\int_{-\infty}^{\infty}\left(\overline{p}_1^{\,2}(x) - p^2(x)\right)^2 dx \sim \tag{22}$$
$$\sim \frac{1}{nc}\left\{\left[\left(b^{\frac{4}{3},0}\right)^3 - 2\,b^{2,0}\right]\left\|p^3(x)\right\| + \right.$$
$$+\frac{3}{2}\left(a^2\beta^2 + b^{\frac{4}{3},2}c^2\right)\left(b^{\frac{4}{3},0}\right)^2\left\|p^3(x)p^{(2)}(x)\right\| -$$
$$\left. -2\left(a^2\beta^2 + b^{2,2}c^2\right)\left\|p^2(x)\,p^{(2)}(x)\right\|\right\} +$$

$$+\left[\left(a^2\beta^2+c^2\right)\left\| p(x)\,p^{(2)}(x)\right\| + a^2\beta^2\left\| \left(p^{(1)}(x)\right)^2 \right\|\right]^2,$$

где $\quad \left\| p^2(x)\right\| = \int\limits_{-\infty}^{\infty} p^2(x)\,dx, \qquad a^k = \int\limits_D u^k\,h(u)\,du < \infty,$

$$b^{\nu,k} = \int\limits_{-\infty}^{\infty} z^k\,\Phi^{\nu}(z)\,dz < \infty\,,\ \ k = 0,2,4,6,\ldots\ .$$

Величина критерия (22) представляет собой меру близости между искомой плотностью вероятностей $p^2(x)$ и ее интегральной оценкой $\bar{p}_1^2(x)$. При конечном объеме выборки (22) зависит от коэффициента размытости c и ядерных функций $\Phi(u)$, $h(u)$.

4. Анализ свойств статистики $\bar{p}_1^2(x)$**.**

Проведем для конкретных условий анализ асимптотических выражений относительно среднеквадратической ошибки аппроксимации

$$W(n,c,\beta) = \frac{M\int\limits_{-\infty}^{\infty}\left(\bar{p}_1^2(x) - p^2(x)\right)^2 dx}{\int\limits_{-\infty}^{\infty} p^4(x)\,dx}\,, \tag{23}$$

где числитель в (23) определяется из (22). Исследуем зависимость функции $W(n,c,\beta)$ от значения параметра β. С этой целью рассмотрим: 1) $W_1(n,c) = W(n,c,\frac{c}{2})$; 2) $W_2(n,c) = W(n,c,c)$ и 3) $W_3(n,c) = W(n,c,2c)$.

В качестве ядерной функции $h(u)$ берем ступенчатую функцию

$$h(u)=\begin{cases}\dfrac{1}{2}\,, & |u|<1\,,\\ 0\,, & |u|\geq 1\,.\end{cases}$$

Тогда $a^2=\int\limits_{-1}^{1}u^2\,h(u)\,d\,u=\dfrac{1}{3}$.

В качестве коэффициента размытости c в (23) берем оптимальный коэффициента размытости [4]

(24) $$c=\sqrt[5]{b^{2,0}\left(n\left\|\left(p^{(2)}(x)\right)^2\right\|\right)^{-1}}\,.$$

В качестве восстанавливаемой плотности вероятности рассмотрим нормированную функцию Лапласа $p(x)=\dfrac{1}{\sqrt{2\pi}}\exp\left\{-\dfrac{x^2}{2}\right\}$. Нетрудно убедиться, что

$$\left\|p^3(x)\right\|=\frac{\sqrt{3}}{6\pi}\,,\quad \left\|\left(p^{(1)}(x)\right)^2\right\|=\left\|p^{(2)}(x)\;p(x)\right\|=\frac{1}{4\sqrt{\pi}}\,,$$

$$\left\|p^{(2)}(x)\;p^2(x)\right\|=\frac{1}{6\sqrt{3\pi}}\,,$$

$$\left\|p^4(x)\right\|=\frac{1}{2\sqrt{(2\pi)^3}}\,,\quad \left\|\left(p^{(2)}(x)\right)^2\right\|=\frac{3}{8\sqrt{\pi}}\,,$$

$$\left\|p^3(x)\,p^{(2)}(x)\right\|^2=\frac{1}{32\sqrt{2}\,\pi\sqrt{\pi}}\,.$$

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

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

для которого

$$b^{2,0} = \frac{3}{5\sqrt{5}}\,,\ b^{2,2} = -\frac{12\sqrt{5}}{7}\,,\ b^{\frac{4}{3},0} = 0.638\,,\ \left(b^{\frac{4}{3},0}\right)^2 = 0.407\,,$$

$$\left(b^{\frac{4}{3},0}\right)^3 = 0.259\,,\ b^{\frac{4}{3},2} = 0.563\,.$$

Ниже в рисунке 1 изображена зависимость среднеквадратической ошибки аппроксимации (23) от объемов статистических данных при различных значениях параметра β. Из этого рисунка видно, что среднеквадратическая ошибка аппроксимаций (23) становится малым для больших значений n и стремится к нулю при $n \to \infty$. Среднеквадратическая ошибка аппроксимации (23) при $\beta = c$ стремится к нулю быстрее, чем при $\beta > c$. Самый плохой результат получился при $\beta < c$.

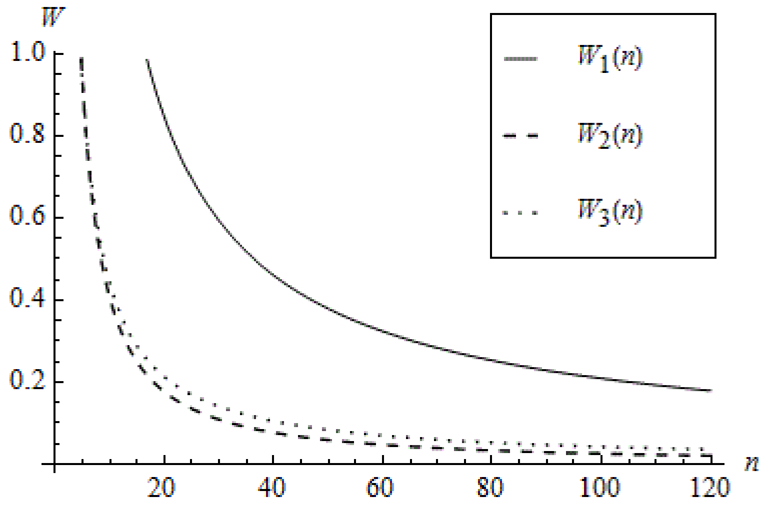

Рис. 1. Зависимость среднеквадратической ошибки аппроксимации (23) от объемов статистических данных при различных значениях параметра β.

4. Заключение.

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

В данной работе доказано, что 1) математическое ожидание интегральной оценки $\bar{p}_1^2(x)$ при увеличении объема экспериментальных данных сходится к искомой функции квадрата плотности вероятности $p^2(x)$; 2) имеют место сходимость в среднеквадратическом для интегральной оценки квадрата плотности $\bar{p}_1^2(x)$ и состоятельность оценки квадрата плотности $\bar{p}_1^2(x)$.

Вычислительная эффективность непараметрических алгоритмов обработки информации, основанных на интегральных оценках квадрата плотности вероятности, во многом определяется объемом статистических данных и значением параметра β. В условиях большого объема выборки оптимальным образом улучшаются аппроксимационные свойства статистики, когда параметр β равняется оптимальному коэффициенту размытости (24).

Автор выражает свою благодарность профессору А. В. Лапко за полезные советы при обсуждении работы.

Литература

1. ЕПАНЕЧНИКОВ В. А. *Непараметрическая оценка многомерной плотности вероятности* // Теория вероятностей и её применения. – 1969. – Том 14. – Вып. 1. – С. 156 - 161.
2. ЛАПКО А. В., ЛАПКО В. А. *Непараметрические системы обработки неоднородной информации.* – Новосибирск: Наука, 2007. – 174 с.
3. ЛАПКО А. В., ЛАПКО В. А. *Анализ свойств непараметрических оценок смеси плотностей вероятности при различных условиях распределения статистических данных* // Информатика и системы управления. – 2013. – № 1 (35). – С. 119 - 126.
4. 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). – С. 121 - 126.
5. МАНИЯ Г. М. *Статистическое оценивание распределения вероятностей.* – Тбилиси: ТбГУ, 1974. – 238 с.
6. НАДАРАЯ Э. А. *Об оценке плотностей распределения случайных величин* // Сообщ. АН ГрССР. – 1964. – Том 32. – С. 277 - 280.
7. ТАРАСЕНКО Ф. П., ДМИТРИЕВ Ю. Г. *Об одном классе непараметрических оценок нелинейных функционалов плотности* // Теория вероятностей и её применения. – 1974. – Том 19. – Вып. 2. – С. 404 - 409.
8. PARZEN E. *On estimation of a probability density function and mode* // Ann. Math. Statistic. – 1962. – Vol. 33. – P. 1065 - 1076.
9. ROSENBLATT M. *Remarks on some nonparametric estimates of a density function* // Ann. Math. Statistic. – 1956. – Vol. 27. – P. 642 - 669.

INTEGRAL ESTIMATES OF QUADRATE OF THE PROBABILITY DENSITIES AND ITS PROPERTIES

Tursun Yuldashev, Siberian state aerospace university, Krasnoyarsk, Cand.Sc., assistant professor (tursunbay@rambler.ru).

Abstract: It is considered in this article the methods of integral estimates of quadrate of the probability densities. In the case of a limited volume of training sample the nonparametric estimates of quadrate of the probability densities of Rosenblatt - Parzen type may be unsmooth. The integral estimates are smoothed these estimates of probability densities in order to improve their approximation properties. By the aid of two positive normalized and symmetric kernel functions it is studied the asymptotic properties of the integral estimates of quadrate of the probability densities. It is tested the convergence of integral estimates of the probability densities with increasing amount of experimental data to the desired probability density. It is proved the asymptotic unbiasedness and RMS (mean squared) convergence of integral estimates of quadrate of the probability densities. It also establishes the asymptotic viability of the integral estimates of the probability densities with increasing amount of experimental data.

Key words: Quadrate of probability density, integral estimate, asymptotic unbiasedness, asymptotic viability, kernel function.

Статья представлена к публикации членом редакционной коллегии ...заполняется редактором…

Поступила в редакцию ...заполняется редактором…
Опубликована ...заполняется редактором…