УДК 519.21

Ю. И. Петунин, Н. П. Тупко (Нац. ун-т им. Т. Шевченко, Киев)

ТЕОРИЯ КВАДРАТИЧНЫХ ОЦЕНОК ДИСПЕРСИИ

A class of quadratic estimates is constructed for the second order moment and for the variance of a random value. These estimates are found on the basis of sample values obtained by simple sampling. The best quadratic estimates are found for the second order moment and for the variance in the case of known mathematical expectation. The exactness of biased and unbiased estimates of variance is investigated in the case of unknown mathematical expectation.

Побудовано клас квадратичних оцінок для моменту другого порядку, а також для дисперсії випадкової величини, які знаходяться на підставі вибіркових значень, що одержані в результаті простого випадкового вибору. Знайдені найкращі квадратичні оцінки для моменту другого порядку, а також для дисперсії у випадку відомого математичного сподівання. Досліджується точність зміщених, а також незміщених оцінок для дисперсії у випадку невідомого математичного сподівання.

Введение. В классической теории вероятностей основными характеристиками случайной величины являются ее математическое ожидание и дисперсия. Это объясняется тем, что математическое ожидание представляет собой характеристику расположения возможных значений случайной величины, а дисперсия является мерой разброса этих значений относительно математического ожидания (среднего значения). На основании этих характеристик можно решить многие задачи теории вероятностей и математической статистики, например построить очень важные для приложений доверительные интервалы, содержащие основную распределенную массу значений случайной величины и оценить их уровень значимости. Кроме того, математическое ожидание и дисперсия имеют великолепные математические свойства (см., например, [1]), что делает их незаменимым инструментом в статистических исследованиях. В связи с этим проблема оценки этих характеристик на основании выборочных значений является очень актуальной. Для оценки неизвестного математического ожидания построены многочисленные теории и получены фундаментальные результаты (см., например, [2] и имеющуюся там литературу), однако проблеме оценки неизвестной дисперсии уделялось значительно меньше внимания и теория оценок дисперсии находится еще в зачаточном состоянии. В этой работе предпринята попытка построить теорию квадратичных оценок неизвестной дисперсии и получить первоначальные результаты в этой области.

Пусть x — случайная величина, а $m(x)$, $D(x)$ — ее математическое ожидание и дисперсия. Как известно,

$$D(x) = m[x - m(x)]^2 = m(y^2), \tag{1}$$

где $y = m(x) - x$ — отклонение случайной величины x от ее математического ожидания. Легко видеть, что

$$D(x) = m(x^2) - [m(x)]^2, \tag{2}$$

откуда следует, что для оценки дисперсии $D(x)$ необходимо оценить момент второго порядка $m(x^2)$ случайной величины x и ее математическое ожидание. Пусть $x_1, x_2, \ldots, x_n$ — выборка возможных значений случайной величины x, полученная путем простого случайного выбора (т. е. случайные величины x_k имеют одинаковую функцию распределения $F(u)$ и независимы в совокупности). Как показано в работе [3], в этом случае наилучшей оценкой неизвестного математического ожидания $m(x)$ является выборочное среднее

$$\bar{x} = \frac{1}{n} \sum_{k=1}^{n} x_k ; \tag{3}$$

согласно практическим рекомендациям [4], оно практически совпадает с $m(x)$, когда объем выборки $x_1, x_2, \ldots, x_n$ превышает число тридцать, $n \geq 30$. В связи с этим рассмотрим вначале задачу оценки момента второго порядка $m(x^2)$; кроме того, из изложенного выше вытекает, что задачу оценки неизвестной дисперсии можно исследовать в случае известного математического ожидания $m(x)$ (когда $n \geq 30$), а также в случае неизвестного $m(x)$ (когда $n < 30$).

1. Оценка неизвестного момента второго порядка $m(x^2)$. Момент второго порядка $L_2 = m(x^2)$ представляет собой математическое ожидание случайной величины $z = x^2$; случайные величины $z_1 = x_1^2$, $z_2 = x_2^2, \ldots, z_n = x_n^2$ имеют одинаковое распределение и независимы в совокупности, так что выборка $z_1, z_2, \ldots, z_n$ получена путем простого случайного выбора, поэтому наилучшей линейной оценкой математического ожидания $m(z) = L_2$ является выборочное среднее $\bar{z}$:

$$\bar{z} = \frac{1}{n} \sum_{k=1}^{n} z_k = \frac{1}{n} \sum_{k=1}^{n} x_k^2 . \tag{4}$$

Это означает, что в классе

$$\Omega = \left\{ u = \sum_{k=1}^{n} c_k z_k = \sum_{k=1}^{n} c_k x_k^2, \quad \sum_{k=1}^{n} c_k = 1 \right\}$$

оценка $\bar{z}$ имеет наименьшую дисперсию: $D(\bar{z}) \leq D(u)$ $\forall u \in \Omega$. Легко видеть, что по отношению к случайным величинам x_k оценки u представляют значения приведенной квадратичной формы, поэтому будем называть u приведенными несмещенными квадратичными оценками, и оценка $\bar{z}$ является наилучшей в классе приведенных несмещенных квадратичных оценок. Однако можно рассмотреть более общий класс таких оценок момента второго порядка L_2:

$$B = \left\{ u = \sum_{i,k=1}^{n} a_{ik} x_i x_k, \;\; m(u^2) = L_2 \right\}.$$

Без ограничения общности можно считать, что матрица $A = \|a_{ij}\|_{i,j=\overline{1,n}}$, определяющая квадратичную форму

$$u = \sum_{i,k=1}^{n} a_{ik} x_i x_k = (Ax, x), \qquad x = (x_1, x_2, \ldots, x_n),$$

является эрмитовой $a_{ij} = a_{ji}$; поскольку момент $L_2 = m(x^2) > 0$, то естественно считать квадратичную форму (Ax, x) положительно определенной: $(Ax, x) \geq 0$ $\forall x \in R^n$. Найдем условие на коэффициенты a_{ij}, при выполнении которых оценки $u \in B$ будут несмещенными. Имеем

$$m(u) = m\left(\sum_{i,k=1}^{n} a_{ik} x_i x_k \right) = m\left(\sum_{\substack{i,k=1 \\ i \neq k}}^{n} a_{ik} x_i x_k + \sum_{i=1}^{n} a_{ii} x_i^2 \right) =$$

$$= m^2 \sum_{\substack{i,k=1 \\ i \neq k}}^{n} a_{ik} + m(x^2) \sum_{i=1}^{n} a_{ii} = m(x^2);$$

так как это равенство должно быть справедливо при всех положительных $m(x^2)$, то

$$\sum_{\substack{i,k=1 \\ i \neq k}}^{n} a_{ik} = 0, \qquad \sum_{i=1}^{n} a_{ii} = 1. \tag{5}$$

Таким образом, класс всех несмещенных квадратичных оценок имеет вид

$$B = \Biggl\{ u = \sum_{i,k=1}^{n} a_{ik} x_i x_k = (Ax, x), \ \ A = \|a_{ik}\|, \ \ \sum_{\substack{i,k=1 \\ i \neq k}}^{n} a_{ik} = 0,$$

$$\sum_{i=1}^{n} a_{ii} = 1, \ \ x = (x_1, \ldots, x_n), \ \ (Ax, x) \geq 0 \ \ \forall x \in R^n \Biggr\}. \tag{6}$$

Теорема 1. *Пусть* $x_1, x_2, \ldots, x_n$ *— выборка, полученная путем простого случайного выбора из генеральной совокупности* G, *распределение которой имеет конечный момент четвертого порядка. Тогда оценка*

$$\bar{z} = \frac{1}{n} \sum_{k=1}^{n} x_k^2$$

является наилучшей в классе B *всех несмещенных квадратичных оценок.*

Доказательство. Докажем, что $D(\bar{z}) \leq D(u) \ \ \forall u \in B$, т. е. оценка $\bar{z}$ имеет наименьшую дисперсию.

Действительно, для $u \in B$

$$D(u) = m(u^2) - L^2, \tag{7}$$

и поэтому

$$m(u^2) = m^4 \sum_{\substack{i,k,j,m=1 \\ i \neq k \neq j \neq m}}^{n} a_{ik} a_{jm} - 2L_2 m^2 \sum_{\substack{i,k,j=1 \\ i \neq k \neq j}}^{n} a_{ii} a_{kj} +$$

$$+ 4L_3 m \sum_{\substack{i,k=1 \\ i \neq k}}^{n} a_{ii} a_{ik} + L_4 \sum_{i=1}^{n} a_{ii}^2 + 2L_2 m^2 \sum_{\substack{i,k,j=1 \\ i \neq k \neq j}}^{n} a_{ik} a_{ij} +$$

$$+ 2L_2^2 \sum_{\substack{i,k=1 \\ i \neq k}}^{n} a_{ik}^2 + 2L_2 m^2 \sum_{\substack{i,k,j=1 \\ i \neq k \neq j}}^{n} a_{ik} a_{kj} + L_2^2 \sum_{\substack{i,k=1 \\ i \neq k}}^{n} a_{ii} a_{kk}.$$

Учитывая условие (5), окончательно имеем

$$m(u^2) = m^4 \left[2 \sum_{\substack{j,m=1 \\ j \neq m}}^{n-2} a_{jm} \left(- \sum_{\substack{i,k=1 \\ i \neq k}}^{n} a_{ik} \right) + \sum_{\substack{i,k,j,m=1 \\ i \neq k \neq j \neq m \\ (i,k) \neq (n,n-1) \\ (i,m) \neq (n,n-1)}}^{n} a_{ik} a_{jm} \right] +$$

$$
+\ 2L_2 m^2 \left[\sum_{\substack{k,j=1 \\ k \neq j}}^{n-1} a_{kj} \left(1 - \sum_{i=1}^{n-1} a_{ii} \right) + \right.
$$

$$
\left. + \sum_{i=1}^{n-2} a_{ii} \left(- \sum_{\substack{i,k=1 \\ i \neq k \\ (i,k) \neq (n,n-1)}}^{n} a_{ik} \right) + \sum_{\substack{i,k,j=1 \\ i \neq k \neq j \\ (k,j) \neq (n,n-1)}}^{n} a_{ii} a_{kj} \right] +
$$

$$
+\ 4L_3 m \left[\sum_{k=1}^{n-2} a_{nk} \left(1 - \sum_{i=1}^{n-1} a_{ii} \right) + \left(1 - \sum_{i=1}^{n-1} a_{ii} \right) \left(- \sum_{\substack{i,k=1 \\ i \neq k \\ (i,k) \neq (n,n-1)}}^{n} a_{ik} \right) + \right.
$$

$$
\left. + \sum_{\substack{i,k=1 \\ i \neq k,\ i \neq n}}^{n} a_{ii} a_{ik} \right] + L_4 \left[\left(1 - \sum_{i=1}^{n-1} a_{ii} \right)^2 + \sum_{i=1}^{n-1} a_{ii}^2 \right] +
$$

$$
+\ 2L_2 m^2 \left[2 \sum_{k=1}^{n-2} a_{nk} \left(- \sum_{\substack{i,k=1 \\ i \neq k \\ (i,k) \neq (n,n-1)}}^{n} a_{ik} \right) + \sum_{\substack{i,k,j=1 \\ i \neq k \neq j \\ (i,k) \neq (n,n-1) \\ (i,j) \neq (n,n-1)}}^{n} a_{ik} a_{ij} \right] +
$$

$$
+\ 2L_2^2 \left[\left(- \sum_{\substack{i,k=1 \\ i \neq k \\ (i,k) \neq (n,n-1)}}^{n} a_{ik} \right)^2 + \sum_{\substack{i,k=1 \\ i \neq k \\ (i,k) \neq (n,n-1)}}^{n} a_{ik}^2 \right] +
$$

$$
+\ 2L_2 m^2 \left[\sum_{j=1}^{n-2} a_{n-1\,j} \left(- \sum_{\substack{i,k=1 \\ i \neq k \\ (i,k) = (n,n-1)}}^{n} a_{ik} \right) + \sum_{i=1}^{n-2} a_{in} \left(- \sum_{\substack{i,k=1 \\ i \neq k \\ (i,k) \neq (n,n-1)}}^{n} a_{ik} \right) + \right.
$$

$$
\left. + \sum_{\substack{i,k,j=1 \\ i \neq k \neq j \\ (i,k) \neq (n,n-1) \\ (i,j) \neq (n,n-1)}}^{n} a_{ik} a_{kj} \right] + L_2^2 \left[2 \sum_{k=1}^{n-1} a_{kk} \left(1 - \sum_{i=1}^{n-1} a_{ii} \right) + \sum_{\substack{i,k=1 \\ i \neq k}}^{n-1} a_{ii} a_{kk} \right]. \tag{8}
$$

Таким образом, задачу поиска условного экстремума мы свели к задаче исследования безусловного экстремума.

Из формулы (7) следует, что функция $f(a_{ik}) = D(u)$ достигает минимума в точке $u^* = \sum_{i,k=1}^{n} a_{ik}^{*} x_i x_k$, в которой функция $q(a_{ik}) = m(u^2)$ принимает наименьшее значение. Далее на основании формулы (8) функция $q(a_{ik})$ является положительно определенной квадратичной формой относительно переменных a_{ik}, поэтому $q(a_{ik})$ достигает минимума в точке, где обращаются в нуль ее частные производные, которые имеют следующий вид:

1) при $i = \overline{1, n-1}$, $k = \overline{1, n-1}$, $i \neq k$

$$\frac{\partial q}{\partial a_{ik}} = m^4 \left[2\left(- \sum_{\substack{j,m=1 \\ j \neq m \\ (j,m) \neq (n,n-1)}}^{n} a_{jm} \right) + 2\left(- \sum_{\substack{j,m=1 \\ j \neq m}}^{n-2} a_{jm} \right) + 2 \sum_{\substack{j,m=1 \\ i \neq k \neq j \neq m \\ (j,m) \neq (n,n-1)}}^{n} a_{jm} \right] +$$

$$+ 2L_2 m^2 \left[1 - \sum_{i=1}^{n-1} a_{ii} - \sum_{i=1}^{n-2} a_{ii} + \sum_{\substack{i=1 \\ j \neq i \neq k}}^{n-1} a_{ii} \right] +$$

$$+ 4L_3 m \left[-\left(1 - \sum_{i=1}^{n-1} a_{ii} \right) + a_{ii} \right] + 2L_2 m^2 \left[-2 \sum_{i=1}^{n-1} a_{nk} + 2 \sum_{\substack{j=1 \\ j \neq i \neq k \\ (i,j) \neq (n,n-1)}}^{n} a_{ij} \right] +$$

$$+ 2L_2^2 \left[2 \sum_{\substack{j=1,\ l=1 \\ j \neq l \\ (j,l) \neq (n,n-1)}}^{n} a_{jl} + 2a_{ik} \right] +$$

$$+ 2L_2 m^2 \left[-\sum_{j=1}^{n-2} a_{n-1\,j} - \sum_{i=1}^{n-2} a_{in} + \sum_{\substack{j=1 \\ j \neq i \neq k \\ (j,i) \neq (n,n-1)}}^{n} a_{ji} + \sum_{\substack{j=1 \\ j \neq i \neq k \\ (k,j) \neq (n,n-1)}}^{n} a_{kj} \right],$$

где $L_k = m(x^k)$ для всех натуральных k;

2) при $k = \overline{1, n-2}$

$$\frac{\partial q}{\partial a_{kn}} = m^4 \left[-2 \sum_{\substack{j,l=1 \\ j \neq l}}^{n-2} a_{jl} + 2 \sum_{\substack{j=1 \\ j \neq k}}^{n-1} a_{jk} \right] + 2L_2 m^2 \left[-\sum_{i=1}^{n-2} a_{ii} + \sum_{\substack{i=1 \\ i \neq k}}^{n-1} a_{ii} \right] +$$

$$+ 4L_3 m \left[-\left(1 - \sum_{i=1}^{n-1} a_{ii} \right) + a_{kk} \right] + 2L_2 m^2 \left[-2 \sum_{k=1}^{n-2} a_{nk} + 2 \sum_{\substack{j=1 \\ j \neq k}}^{n-1} a_{kj} \right] +$$

$$+\ 2L_2^2\left[2\sum_{\substack{i,j=1\\ i\neq j\\ (i,j)\neq(n,n-1)}}^{n} a_{ij}\ +2a_{kn}\right]+$$

$$+\ 2L_2 m^2\left[-\sum_{j=1}^{n-2}a_{n-1j}-\sum_{\substack{i,j=1\\ i\neq j\\ (i,j)\neq(n,n-1)}}^{n} a_{ij}-\sum_{i=1}^{n-2}a_{in}+\sum_{\substack{j=1\\ j\neq k}}^{n-1}a_{nj}+\sum_{\substack{j=1\\ j\neq k}}^{n-1}a_{jk}\right];$$

3) при $k=\overline{1,n-2}$

$$\frac{\partial q}{\partial a_{nk}} = m^4\left[2\left(-\sum_{\substack{j,l=1\\ j\neq l}}^{n}a_{jl}\right)+2\sum_{\substack{j,l=1\\ j\neq l\neq k}}^{n-1}a_{jl}\right]+2L_2 m^2\left[-\sum_{i=1}^{n-2}a_{ii}+\sum_{\substack{i=1\\ i\neq k}}^{n-1}a_{ii}\right]+$$

$$+\ 4L_3 m\left[\left(1-\sum_{i=1}^{n-1}a_{ii}\right)-1+\sum_{i=1}^{n-1}a_{ii}\right]+$$

$$+\ 2L_2 m^2\left[2\left(-\sum_{\substack{i,j=1\\ i\neq j\\ (i,j)\neq(n,n-1)}}^{n} a_{ij}\right)+2\left(-\sum_{i=1}^{n-2}a_{ii}\right)+2\sum_{\substack{j=1\\ j\neq k}}^{n-1}a_{kj}\right]+$$

$$+\ 2L_2 m^2\left[-\sum_{j=1}^{n-2}a_{n-1j}-\sum_{i=1}^{n-1}a_{in}+\sum_{\substack{j=1\\ j\neq k}}^{n-1}a_{kj}+\sum_{\substack{j=1\\ j\neq k}}^{n-1}a_{jk}\right]+$$

$$+\ 2L_2^2\left[-2\sum_{\substack{i,j=1\\ i\neq j\\ (i,j)\neq(n,n-1)}}^{n} a_{ij}\ +2a_{nk}\right];$$

4) при $i=n-1,\ k=n$

$$\frac{\partial q}{\partial a_{n-1n}} = m^4\left[2\left(-\sum_{\substack{j,l=1\\ j\neq l}}^{n-2}a_{jl}\right)+\sum_{\substack{j,l=1\\ j\neq l}}^{n-2}a_{jl}\right]+2L_2 m^2\left[-\sum_{i=1}^{n-2}a_{ii}+\sum_{i=1}^{n-2}a_{ii}\right]+$$

$$+\ 4L_3 m\left[-1+\sum_{i=1}^{n-1}a_{ii}\ +a_{n-1n-1}\right]+$$

$$+\ 2L_2 m^2\left[2\left(-\sum_{k=1}^{n-1}a_{nk}\right)+\sum_{j=1}^{n-2}a_{n-1j}\right]+2L_2^2\left[2\sum_{\substack{i,k=1\\ i\neq k\\ (i,k)\neq(n,n-1)}}^{n} a_{ik}\ +2a_{n-1n}\right]+$$

$$+\ 2L_2 m^2 \left[-\sum_{k=1}^{n-2} a_{n-1k} - \sum_{i=1}^{n-2} a_{in} + \sum_{i=1}^{n-2} a_{ni} + \sum_{i=1}^{n-2} a_{in-1}\right];$$

5) при $k = \overline{1, n-1}$

$$\frac{\partial q}{\partial a_{kn-1}} = m^4 \left[2\left(-\sum_{\substack{j,l=1\\ j\neq l}}^{n-2} a_{jl}\right) + 2 \sum_{\substack{j,l=1\\ j\neq l\neq k\neq n-1\\ (i,j)\neq(n,n-1)}}^{n} a_{jl}\right] +$$

$$+\ 2L_2 m^2 \left[1 - \sum_{i=1}^{n-1} a_{ii} - \sum_{i=1}^{n-2} a_{ii} + \sum_{\substack{i=1\\ i\neq k}}^{n-2} a_{ii}\right] +$$

$$+\ 4L_3 m \left[-1 + \sum_{i=1}^{n-1} a_{ii} + a_{kk}\right] +$$

$$+\ 2L_2 m^2 \left[2\left(-\sum_{j=1}^{n-2} a_{nj}\right) + 2 \sum_{\substack{j=1\\ j\neq k\neq n-1}}^{n} a_{kj}\right] +$$

$$+\ 2L_2^2 \left[2 \sum_{\substack{i,j=1\\ i\neq j\\ (i,j)\neq(n,n-1)}}^{n} a_{ij} + 2a_{kn-1}\right] +$$

$$+\ 2L_2 m^2 \left[-\sum_{j=1}^{n-2} a_{n-1j} - \sum_{i=1}^{n-2} a_{ik} + \sum_{\substack{j=1\\ j\neq k\neq n-1}}^{n} a_{n-1j} + \sum_{\substack{j=1\\ j\neq k\neq n-1}}^{n} a_{jk}\right];$$

6) при $k = \overline{1, n-2}$

$$\frac{\partial q}{\partial a_{n-1k}} = m^4 \left[2\left(-\sum_{\substack{j,l=1\\ j\neq l}}^{n} a_{jl}\right) + 2 \sum_{\substack{j,l=1\\ j\neq l\neq k\neq n-1}}^{n} a_{jl}\right] +$$

$$+\ 2L_2 m^2 \left[1 - \sum_{i=1}^{n-1} a_{ii} - \sum_{i=1}^{n-2} a_{ii} + \sum_{\substack{i=1\\ i\neq k}}^{n-2} a_{ii}\right] +$$

$$+\ 4L_3 m \left[-1 + \sum_{i=1}^{n-1} a_{ii} + a_{n-1n-1}\right] +$$

$$+\ 2L_2 m^2 \left[2\left(-\sum_{i=1}^{n-2} a_{ni}\right) + 2 \sum_{\substack{j=1\\ j\neq k\neq n-1}}^{n} a_{n-1j}\right] +$$

$$+\ 2L_2^2\left[2\sum_{\substack{i,j=1\\ i\neq j\\ (i,j)\neq(n,n-1)}}^{n} a_{ij}+2a_{n-1k}\right]+2L_2m^2\left[-\sum_{\substack{j,j=1\\ i\neq j\\ (i,j)\neq(n,n-1)}}^{n} a_{ij}-\right.$$

$$\left.-\sum_{j=1}^{n-2}a_{n-1j}-\sum_{i=1}^{n-2}a_{in}+\sum_{\substack{j=1\\ j\neq k\neq n-1}}^{n}a_{kj}+\sum_{\substack{j=1\\ j\neq k\neq n-1}}^{n-2}a_{jn-1}\right];$$

7) при $i=\overline{1,n-1}$

$$\frac{\partial q}{\partial a_{ii}}=2L_2m^2\left[-\sum_{\substack{k,j=1\\ k\neq j}}^{n-1}a_{kj}-\sum_{\substack{j,k=1\\ j\neq k\\ (j,k)\neq(n,n-1)}}^{n}a_{jk}+\sum_{\substack{k,j=1\\ k\neq j\neq k\\ (k,j)\neq(n,n-1)}}^{n}a_{kj}\right]+$$

$$+\ 4L_3m\left[-\sum_{k=1}^{n-2}a_{nk}+\sum_{\substack{j,k=1\\ j\neq k\\ (j,k)\neq(n,n-1)}}^{n}a_{jk}+\sum_{\substack{k=1\\ k\neq i}}^{n}a_{ik}\right]+$$

$$+\ L_4\left[-2\left(1-\sum_{i=1}^{n-1}a_{ii}\right)+2a_{ii}\right]+$$

$$+\ L_2^2\left[2\left(1-\sum_{k=1}^{n-1}a_{kk}\right)-2\sum_{k=1}^{n-1}a_{kk}+2\sum_{\substack{k=1\\ k\neq i}}^{n-1}a_{kk}\right].$$

Оценка $\bar{z}$ (4) в классе B имеет коэффициенты $a_{ik}^*=0$, $i\neq k$, и $a_{ii}^*=\dfrac{1}{n}$ $\forall\, i=\overline{1,n}$. Подставляя значения этих коэффициентов в формулы (1)–(7), получаем $\dfrac{\partial q}{\partial a_{ik}}=0$ при $a_{ik}=a_{ik}^*$, поэтому функция $q(a_{ik})$ достигает минимума в точке a_{ik}^*, так что оценка (4) является наилучшей в классе всех квадратичных несмещенных оценок. Заметим, что все приведенные вычисления имеют смысл, когда у случайной величины x_k (выборочных значений) существует конечный момент четвертого порядка $L_4=m(x^4)$. Теорема доказана.

2. Оценка дисперсии случайной величины в случае известного математического ожидания. Перейдем теперь к изучению квадратичных оценок дисперсии на основании выборочных значений, когда математическое ожидание $m(x)=m$ случайной величины x точно известно. В этом случае в силу формулы (2) класс B_{σ^2} несмещенных квадратичных оценок u для $D(u)=\sigma^2$ имеет вид

$$B_{\sigma^2}=\left\{u=\sum_{i,k=1}^{n}a_{ik}x_ix_k-m^2,\ \sum_{i\neq k}a_{ik}=0,\ \sum_{i=1}^{n}a_{ii}=1,\right.$$

$$\left.A=\|a_{ik}\|,\ (Ax,x)\geq m^2\quad\forall\, x\in R^n\right\},$$

причем оценка u^* будет наилучшей, если $u^* = \bar{z} - m^2$, т. е.

$$u^* = S_0^2 = \frac{1}{n}\sum_{k=1}^{n} x_k^2 - m^2. \tag{9}$$

Однако в класс B_{σ^2} не попадают классические оценки неизвестной дисперсии (в случае известного математического ожидания)

$$S^2 = \frac{1}{n}\sum_{k=1}^{n}(x_k - m)^2, \tag{10}$$

так как в выражение S^2 входят линейные члены относительно x_k. В связи с этим следует рассмотреть более широкий класс $B^*_{\sigma^2}$ квадратичных несмещенных оценок

$$u = \sum_{i,k=1}^{n} a_{ik}\, x_i\, x_k + 2\sum_{i=1}^{n} b_i\, x_i + c,$$

представляющих значения квадратичного многочлена относительно переменных $x_1, x_2, \ldots, x_n$, так что $S^2 \in B^*_{\sigma}$.

Используя аналогичные рассуждения, можно отыскать условие несмещенности оценок u из класса $B^*_{\sigma^2}$, однако мы не будем сейчас этим заниматься. Рассмотрим задачу точности оценок S_0^2 и S^2. Легко видеть, что оценки S_0^2 и S^2 являются несмещенными: $m(S^2) = m(S_0^2) = \sigma^2$, поэтому их точность характеризуется дисперсиями $D(S_0^2)$, $D(S^2)$. На основании формулы (2) для этого достаточно проверить неравенство для их моментов второго порядка: $m(S_0^4) = L_2(S_0^2)$, $m(S^4) = L_2(S^2)$. С этой целью найдем момент второго порядка для выборочного среднего $\bar{x}$:

$$m(\bar{x})^2 = m\left(\frac{1}{n}\sum_{k=1}^{n} x_k\right)^2 = \frac{1}{n^2}\, m\left(\sum_{\substack{k,i=1\\k\neq i}}^{n} x_k\, x_i - \sum_{k=1}^{n} x_k^2\right) =$$

$$= \frac{1}{n^2}\sum_{\substack{i,k=1\\i\neq k}}^{n} m(x_i\, x_k) + \frac{1}{n} m\left(\frac{1}{n}\sum_{k=1}^{n} x_k^2\right) =$$

$$= \frac{1}{n^2}\sum_{\substack{k,i=1\\k\neq i}}^{n} m(x_k)\, m(x_i) + \frac{1}{n} L_2 = \frac{1}{n^2}\, n(n-1)\, m^2 + \frac{1}{n} L_2 =$$

$$= \frac{(n-1)\, m^2}{n} + \frac{L_2}{n}.$$

Кроме того, проводя очевидные преобразования, можно показать, что

$$m(\bar{x})^2 = m^2 + \frac{\sigma^2}{n}.$$

Отсюда следует $\lim m(\bar{x}^2) = m^2$, $n \to \infty$, так что $\bar{x}^2$ представляет асимптотически несмещенную оценку m^2. Далее

$$L_2(S_0^2) = m(S_0^2) = m\left(\frac{1}{n}\sum_{k=1}^{n} x_k^2 - m^2\right) =$$

$$= m\left\{\frac{1}{n^2}\left(\sum_{k=1}^{n} x_k^2\right)^2 - 2n^2\frac{1}{n}\sum_{k=1}^{n} x_k^2 + m^4\right\} =$$

$$= m\left\{\frac{1}{n^2}\left[\sum_{\substack{k,i=1\\ k\neq i}}^{n} x_k^2 x_i^2 + \sum_{k=1}^{n} x_k^4\right]\right\} - 2m^2 L_2 + m^4 =$$

$$= \frac{1}{n^2} n(n-1) L_2^2 + \frac{1}{n} L_4 - 2m^2 L_2 + m^4 =$$

$$= \frac{n-1}{n} L_2^2 + \frac{1}{n} L_4 - 2m^2 L_2 + m^4,$$

поэтому

$$D(S_0^2) = \frac{n-1}{n} L_2^2 + \frac{1}{n} L_4 - 2m^2 L_2 + m^4 - \sigma^4.$$

Аналогично

$$m(S^2) = m\left(\frac{1}{n}\sum_{k=1}^{n}(x_k - m)^2\right)^2 =$$

$$= m\left(\frac{1}{n}\sum_{k=1}^{n}(x_k - 2x_k m + m^2)\right)^2 = m\left(\frac{1}{n}\sum_{k=1}^{n} x_k - 2\bar{x}m + m^2\right)^2 =$$

$$= m\{S_0^4 + 4\bar{x}m^2 + 4m^4 + 4S^2m^2 - 8\bar{x}m^3 - 4S^2\bar{x}m\}^2.$$

В случае нормального распределения случайной величины x оценки $\bar{x}$ и S^2 являются независимыми [5], поэтому

$$m(S^4) = m(S_0^2) + 4m^2 m(\bar{x}^2) + 4m^4 + 4m^2 m(S_0^2) - 8m^3 m(\bar{x}) -$$

$$- 4m\,m(S_0^2)m(\bar{x}) = m(S_0^4) - \frac{4}{n}m^4 + 4\frac{L_2}{n}m^2 =$$

$$= m(S_0^4) + \frac{4m^2}{n}(L_2 - m^2).$$

В силу неравенства моментов [6]

$$m^2 \le L_2,$$

поэтому

$$m(S^4) > m(S_0^4) \quad \text{и} \quad D(S^2) > D(S_0^2).$$

Таким образом, для нормально распределенной величины x оценка S_0^2 для неизвестной дисперсии является более предпочтительной, чем классическая оценка S^2. С целью исследования точности оценок S_0^2 и S^2 для распределений, отличающихся от нормального, вычислим выражение

$$m(S_0^2\,\bar{x}) = m\left\{\left(\frac{1}{n}\sum_{k=1}^{n} x_k^2 - m^2\right)^2 \frac{1}{n}\sum_{i=1}^{n} x_i\right\} =$$

$$= m\left\{\frac{1}{n^2}\left(\sum_{k=1}^{n} x_k^2 \sum_{i=1}^{n} x_i\right) - \frac{1}{n} m^2 \sum_{i=1}^{n} x_i\right\} =$$

$$= m\left\{\frac{1}{n^2}\left(\sum_{\substack{k,i=1\\ k\neq i}}^{n} x_k^2 x_i + \sum_{i=1}^{n} x_i^3\right) - \frac{1}{n} m^2 \sum_{i=1}^{n} x_i\right\} =$$

$$= \frac{1}{n^2}(L_2 n(n-1) + nL_3) - \frac{1}{n} m^2 nm =$$

$$= \frac{L_2}{n}(n-1) + \frac{L_3}{n} - m^3,$$

поэтому

$$m(S^4) - m(S_0^4) = 4m^2 m(S_0^2) - 4m\,m(S_0^2\,\bar{x}) - \frac{4m^4}{n} + \frac{4m^2 L_2}{n} =$$

$$= 4m^2(L_2 - m^2) - 4m\left(m\frac{L_2}{n}(n-1) + \frac{L_3}{n} - m^3\right) + \frac{4m^2 L_2}{n} - \frac{4m^4}{n} =$$

$$= 4m^2 L_2 - 4m^4 - \frac{4m^2 L_2}{n}(n-1) + \frac{4mL_3}{n} + 4m^4 + \frac{4m^2 L_2}{n} - \frac{4m^4}{n} =$$

$$= 4m^2 L_2 \frac{2}{n} + 4m\frac{L_3}{n} - \frac{4m^4}{n} = \frac{4}{n}(2mL_2 + L_3 - m^3). \qquad (11)$$

В случае равномерного распределения на интервале $[a, b]$ последний член равенства (11) имеет вид

$$\frac{1}{n}\frac{(b+a)^2}{2}\frac{10b^2 + (b+a)^2 + 2a^2}{24} > 0,$$

поэтому $m(S^4) > m(S_0^4)$, так что классическая оценка (10) снова будет иметь меньшую точность по сравнению с оценкой S_0^2.

Полученные результаты точности оценок S_0^2 и S^2 показывают, что оценка S^2 не будет наилучшей в классе B_σ^*. Действительно, для классов B_σ и B_σ^* справедливо вложение $B_\sigma \subset B_\sigma^*$; поскольку оценка $S_0^2 \in B_\sigma$ имеет меньшую дисперсию, чем S^2, по крайней мере в случае нормального и равномерного распределения, то она является более точной по сравнению с S^2 и в классе B_σ^* (по крайней мере для этих типов распределений).

3. Оценка дисперсии случайной величины в случае неизвестного математического ожидания. Если математическое ожидание случайной величины x неизвестно, то имеет смысл в формулах (9) и (10) заменить m на ее оценку $\bar{x}$; однако две получающиеся при этом оценки совпадают, так как

$$S_*^2 = \frac{1}{n}\sum_{k=1}^{n}(x_k - \bar{x})^2 = \frac{1}{n}\sum_{k=1}^{n}(x_k^2 - 2x_k\bar{x} + \bar{x}^2) =$$

$$= \frac{1}{n}\sum_{k=1}^{n} x_k^2 - 2\frac{1}{n}\bar{x}\sum_{k=1}^{n} x_k + \bar{x}^2 = \frac{1}{n}\sum_{k=1}^{n} x_k^2 - 2\bar{x}^2 + \bar{x} =$$

$$= \frac{1}{n}\sum_{k=1}^{n} x_k^2 - \bar{x}^2 .$$

Однако оценка S_*^2 будет смещенной, поскольку $m(S_*^2) = \frac{n-1}{n}\sigma^2$; если же положить

$$\tilde{S}^2 = \frac{1}{n-1}\sum_{k=1}^{n}(x_k - \bar{x})^2 = \frac{n}{n-1}S_*^2,$$

то оценка $\tilde{S}_*^2$ будет несмещенной. Для того чтобы исследовать точность оценок $\tilde{S}^2$ и S_*^2, необходимо сравнить дисперсию оценки $\tilde{S}^2$ и среднеквадратичную ошибку $\delta^2(S_*^2) = m(S_*^2 - \sigma^2)^2$ оценки S_*^2. С этой целью рассмотрим вначале выражение

$$V = D(\tilde{S}^2) - \delta^2(S_*^2) = m(\tilde{S}^2 - \sigma^2)^2 - m(S_*^2 - \sigma^2)^2 =$$

$$= m[(\tilde{S}^2 - \sigma^2)^2 - (S_*^2 - \sigma^2)^2] =$$

$$= m[(\tilde{S}^2 - \sigma^2 - S_*^2 + \sigma^2)(\tilde{S}^2 - \sigma^2 + S_*^2 - \sigma^2)] =$$

$$= m[(\tilde{S}^2 - S_*^2)(\tilde{S}^2 + S_*^2 - 2\sigma^2)] =$$

$$= m\left[\left(\tilde{S}^2 - \frac{n-1}{n}\tilde{S}^2\right)\left(\tilde{S}^2 + \frac{n-1}{n}\tilde{S}^2 - 2\sigma^2\right)\right] =$$

$$= \frac{1}{n} m\left[\tilde{S}^2\left(\frac{2n-1}{n}\tilde{S}^2 - 2\sigma^2\right)\right] =$$

$$= \frac{1}{n} m\left(\frac{2n-1}{n}\tilde{S}^4 - 2\sigma^2\tilde{S}^2\right) = \frac{2n-1}{n^2} m(\tilde{S}^4) - 2\frac{\sigma^2}{n} m(\tilde{S}^2) =$$

$$= \frac{2n-1}{n^2} m(\tilde{S}^4) - 2\frac{\sigma^4}{n}.$$

Далее вычислим

$$\tilde{S}^4 = \left[\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2\right]^2 =$$

$$= \left[\frac{1}{n-1}\sum_{i=1}^{n}\left(x_i^2 - 2\frac{x_i}{n}\sum_{i=1}^{n} x_i + \frac{1}{n^2}\sum_{i,k=1}^{n} x_i x_k\right)\right]^2 =$$

$$= \left[\frac{1}{n-1}\left(\sum_{i=1}^{n} x_i^2 - \frac{1}{n}\sum_{i,k=1}^{n} x_i x_k\right)\right]^2 = \left[\frac{1}{n}\sum_{i=1}^{n} x_i^2 - \frac{1}{n(n-1)}\sum_{\substack{i,k=1\\ i\neq k}}^{n} x_i x_k\right]^2 =$$

$$= \frac{1}{n^2}\left[\left(\sum_{i=1}^{n} x_i^2\right)^2 - \frac{2}{n-1}\left(\sum_{i=1}^{n} x_i^2\right)\left(\sum_{\substack{i,k=1\\ i\neq k}}^{n} x_i x_k\right) + \frac{1}{(n-1)^2}\left(\sum_{\substack{i,k=1\\ i\neq k}}^{n} x_i x_k\right)^2\right] =$$

$$= \frac{1}{n^2}\left[\sum_{i=1}^{n} x_i^4 + \sum_{\substack{i,k=1\\ i\neq k}}^{n} x_i^n x_k^2 - \frac{2}{n-1}\sum_{\substack{i,k,j=1\\ i\neq k\neq j}}^{n} x_i^2 x_k x_j - \frac{4}{n-1}\sum_{\substack{i,k=1\\ i\neq k}}^{n} x_i^3 x_k + \right.$$

$$\left. + \frac{2}{(n-1)^2}\sum_{\substack{i,k=1\\ i\neq k}}^{n} x_i^2 x_k^2 + \frac{4}{(n-1)^2}\sum_{\substack{i,k,j=1\\ i\neq k\neq j}}^{n} x_i^2 x_k x_j + \frac{1}{(n-1)^2}\sum_{\substack{i,k,j,m=1\\ i\neq k\neq j\neq m}}^{n} x_i x_k x_j x_m\right] =$$

$$= \frac{1}{n^2}\left[\sum_{i=1}^{n} x_i^4 + \frac{(n-1)^2+2}{(n-1)^2}\sum_{\substack{i,k=1\\ i\neq k}}^{n} x_i^2 x_k^2 + \frac{4-2(n-1)}{(n-1)^2}\sum_{\substack{i,k,j=1\\ i\neq k\neq j}}^{n} x_i^2 x_k x_j - \right.$$

$$\left. - \frac{4}{n-1}\sum_{\substack{i,k,j,m=1\\ i\neq k\neq j\neq m}}^{n} x_i x_k x_j x_m\right];$$

$$m(\tilde{S}^4) = \frac{1}{n^2}\left[n\,m(x^4) + \frac{(n-1)^2+2}{(n-1)^2}\,n(n-1)(m^2+\sigma^2) + \right.$$

$$+ \frac{4-2(n-1)}{(n-1)^2}\,n(n-1)(n-2)m^2(m^2+\sigma^2) -$$

$$\left. - \frac{4}{n-1}\,n(n-1)m(x^3)m + \frac{1}{(n-1)^2}\,n(n-1)(n-2)(n-3)m^4\right] =$$

$$= \frac{1}{n}m(x^4) + \frac{(n-1)^2+2}{n(n-1)}(m^2+\sigma^2) + \frac{4-2(n-1)}{n(n-1)}(n-2)m^2(m^2+\sigma^2) -$$

$$- \frac{4}{n}m(x^3)m + \frac{1}{n(n-1)}(n-2)(n-3)m^4;$$

$$V = \frac{2n-1}{n^2}\frac{1}{n}\left[m(x^4) + \frac{(n-1)^2+2}{(n-1)^2}(m^2+\sigma^2)^2 + \frac{6-2n}{n-1}(n-2)m^2(m^2+\sigma^2) - \right.$$

$$\left. - 4m(x^3)m + \frac{(n-2)(n-3)}{n-1}m^4\right] + \frac{2\sigma^4}{n} =$$

$$= \frac{2n-1}{n^2}\left[\frac{1}{n}m(x^4) + \frac{(n-1)^2+2}{(n-1)n}(m^4 + 2m^2\sigma^2 + \sigma^4) + \right.$$

$$\left. + \frac{2(3-n)}{(n-1)n}(n-2)(m^4 + m^2\sigma^2) - \frac{4}{n}m(x^3)m - \frac{(n-2)(n-3)}{n(n-1)}m^4\right] -$$

$$- \frac{2n-1}{n^2}\frac{2n^2}{2n-1}\sigma^4 =$$

$$= \frac{2n-1}{n^2}\left[\frac{1}{n}m(x^4) + m^4\left(\frac{n^2-2n+3}{n(n-1)} + \frac{2(3-n)(n-2)}{n(n-1)} + \frac{(n-3)(n-3)}{n(n-1)}\right) + \right.$$

$$+ 2m^2\sigma^2\left(\frac{n^2-2n+3}{n(n-1)} + \frac{(3-n)(n-2)}{n(n-1)}\right) +$$

$$+ \ \sigma^4 \left(\frac{n^2-2n+3}{n(n-1)} - \frac{2n}{2n-1} \right) - \frac{4}{n} m(x^3)m \Bigg] =$$

$$= \frac{2n-1}{n^2} \left[\frac{1}{n} m(x^4) + m^4 \left(\frac{n^2-2n+3-6-n^2+5n}{n(n-1)} \right) + 2m^2\sigma^2 \frac{3n-3}{n(n-1)} + \right.$$

$$\left. + \ \sigma^4 \frac{(n^2-2n+3)(2n-1) - 2n^2(n-1)}{n(n-1)(2n-1)} - \frac{4}{n} m(x^3)m \right] =$$

$$= \frac{2n-1}{n^2} \left[\frac{1}{n} m(x^4) + m^4 \frac{3(n-1)}{n(n-1)} + 2m^2\sigma^2 \frac{3}{n} + \right.$$

$$\left. + \ \sigma^4 \frac{2n^3-n^2-4n^2+2n+6n-3-2n^3+2n}{n(n-1)(2n-1)} - \frac{4}{n} m(x^3)m \right] =$$

$$= \frac{2n-1}{n^3} \left[m(x^4) + 3m^4 + \frac{6}{n} m^2\sigma^2 + \sigma^4 \frac{-3n^2+8n-3}{n(n-1)(2n-1)} - \frac{4}{n} m(x^3)m \right] =$$

$$= \frac{2n-1}{n^3} \left[m(x^4) + 3m^4 + 6m^2\sigma^2 + \sigma^4 \frac{-3n^2+8n-3}{n(n-1)(2n-1)} - 4m(x^3)m \right].$$

Используя последнюю формулу, окончательно находим выражения для разности V в случае нормального распределения с нулевым математическим ожиданием:

$$V = \frac{2n-1}{n^3} \sigma^4 n(6n^2 - 9n + 2n),$$

поэтому при $n \geq 2$ $V > 0$, так что смещенная оценка S_*^2 является более предпочтительной по сравнению с несмещенной.

Для равномерного распределения

$$V = \frac{(b-a)^4}{5 \cdot 16 \cdot 9} [3(n^3-5) + n(3n-7)^2],$$

поэтому при $n \geq 2$ $V > 0$, и оценка S_*^2 также является более точной по сравнению с $\tilde{S}^2$. Таким образом, использование смещенных оценок дисперсии является более эффективным, чем классических несмещенных оценок.

Перейдем к исследованию класса C квадратичных несмещенных оценок в случае оценки дисперсии при неизвестном математическом ожидании. Класс C состоит из всевозможных квадратичных оценок $u = \sum_{i,j=1}^n a_{ij} x_i x_j$, которые являются несмещенными $m(u) = \sigma^2$, причем матрица $A = \|a_{ij}\|_{i,j=1}^n$, определяющая оценку, считается положительно определенной.

Нетрудно заметить, что из условия несмещенности следует

$$\sum_{i=1}^n a_{ii} = 1, \qquad \sum_{\substack{i,j=1 \\ i \neq j}}^n a_{ij} = -1.$$

Легко видеть, что классическая оценка дисперсии

$$\tilde{S}^2 = \frac{1}{n-1} \sum_{k=1}^n (x_k - \bar{x})^2$$

принадлежит классу C, при этом матрица A имеет вид

$$A = \begin{pmatrix} \frac{1}{n} & -\frac{1}{n(n-1)} & \dots & -\frac{1}{n(n-1)} \\ -\frac{1}{n(n-1)} & \frac{1}{n} & \dots & -\frac{1}{n(n-1)} \\ \vdots & \vdots & \ddots & \vdots \\ -\frac{1}{n(n-1)} & -\frac{1}{n(n-1)} & \dots & \frac{1}{n} \end{pmatrix}.$$

Теорема 2. *Пусть $x_1, x_2, \dots, x_n$ — выборка из генеральной совокупности G с конечным моментом четвертого порядка, которая получена в результате простого случайного выбора. Тогда оценка $\tilde{S}^2$ является наилучшей в классе C всех несмещенных квадратичных оценок.*

Доказательство этой теоремы аналогично доказательству теоремы 1.

1. *Байдак Г. И., Браверман М. Ш., Петунин Ю. И.* Аддитивность дисперсии — характеристическое свойство гильбертова пространства // Функцион. анализ и его прил. – 1983. – **17**, вып. 3. – С. 66–68.
2. *Петунин Ю. И.* Приложение теории случайных процессов в биологии и медицине. – Киев: Наук. думка, 1981. – 320 с.
3. *Курицын Ю. Г., Петунин Ю. И.* К теории линейных оценок математического ожидания случайного процесса // Теория вероятностей и мат. статистика. – 1970. – Вып. 3. – С. 80–92.
4. *Крамер Г.* Математические методы статистики. – М.: Мир, 1975. – 648 с.
5. *Ван дер Варден Б. Л.* Математическая статистика. – М.: Изд-во иностр. лит., 1960. – 436 с.
6. *Романовский В. И.* Математическая статистика. – Ташкент: Изд-во АН УзССР, 1961. – 638 с.

Получено 03.04.98