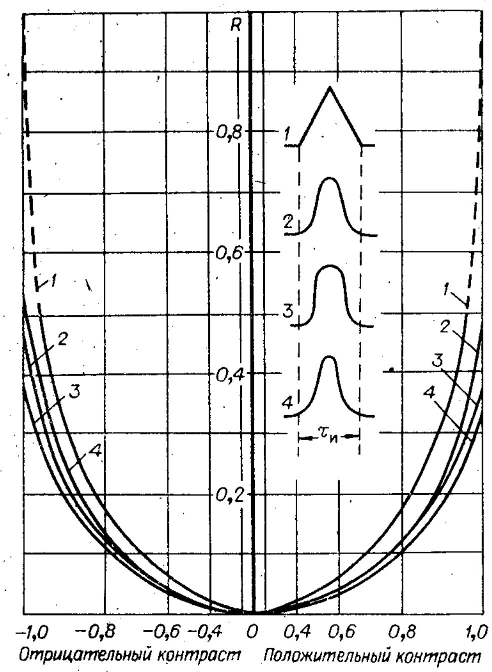
предельные значения $R=R_{\text{пр}}$ четырех наиболее употребительных типов видеоимпульсов, описываемых функциями: *1* — линейной, *2* — квадратичной, *3* — интегралом вероятностей, *4* — Гаусса (рисунок). Кроме того, на рисунке даны графики зависимостей $R$ от коэффициента электрического контраста, из которых видно, что заметного выигрыша в точности при оптимальной обработке можно достигнуть только для высококонтрастного изображения.




Заметим, что значение $R_{\text{пр}}=1$ ($\sigma_{\text{опт}}=0$) для сигнала с линейными фронтами получено вследствие чрезмерной идеализации. Действительно, в случае $k=1$ фоновая составляющая тока равна нулю, и вне промежутка, занимаемого сигналом, шумы также отсутствуют. Так как для линейного фронта его производная на концах импульса изменяется скачком, временное положение сигнала теоретически можно было бы определить абсолютно точно, например, по моменту первого отклонения выходного тока от нуля. Поэтому соответствующие участки кривой *1* на рисунке показаны штрихами.

## ЛИТЕРАТУРА

1. **Михалков К. В.** Основы телевизионной автоматики.— Л.: Энергия, 1967.
2. **Тихонов В. И.** Статистическая радиотехника.— М.: Сов. радио, 1966.
3. **Ильин А. Г., Казанцев Г. Д., Пустынский И. Н.** Об обнаружении сигналов известной формы на фоне нестационарного шума телевизионных датчиков.— Вопросы радиоэлектроники. Сер. Техника телевидения, 1972, вып. 3.

*Поступила в редакцию 6 мая 1980 г.;*
*окончательный вариант — 11 мая 1982 г.*

УДК 519.21

**А. Н. ЕФИМОВ, Е. В. КРИВОРУКОВ**

*(Москва — Харьков)*

# ИССЛЕДОВАНИЕ ЭФФЕКТИВНОСТИ ОЦЕНОК И «ПСЕВДООЦЕНОК» СРЕДНЕГО ПРИ МАЛОМ ЧИСЛЕ НАБЛЮДЕНИЙ

Обработка результатов косвенных наблюдений в системах автоматизации эксперимента и переработки данных зачастую сводится к функциональным преобразованиям и статистическим операциям над ними (например, детектирование и фильтрация). При этом возможны два способа получения конечных результатов:

а) предварительное проведение статистических операций над результатами наблюдений с последующим функциональным преобразованием оценок;

б) первоначальное функциональное преобразование сигналов с получением в дальнейшем статистических оценок.

Иногда выбор той или иной процедуры зависит от наблюдателя, и тогда он заинтересован в более точной из них; в других случаях последовательность диктуется условиями проведения эксперимента, тогда наблюдателю важно оценить свой возможный проигрыш по сравнению с альтернативной процедурой. При небольшом числе наблюдений ситуация обостряется: проигрыш растет.

Задача выбора более точного из алгоритмов, различающихся последовательностью функциональных и статистических операций, ставилась неоднократно [1—5]. В работе [6] в рассмотрение введен новый фактор — объем выборки, от которого, как было показано, существенно зависит правило выбора более точного алгоритма косвенного оценивания среднего. Данная работа является обобщением предыдущей. Здесь найдены условия, при которых несостоятельные «псевдооценки» оказываются точнее состоятельных несмещенных оценок в среднеквадратическом смысле.

Пусть по выборке наблюдений случайной величины $x$ необходимо оценить математическое ожидание величины $y$, связанной с $x$ известной функциональной зависимостью $y=f(x)$. Пусть, далее, число наблюдений $N$ невелико.

Результат может быть получен двумя путями: методом «а» —

$$\bar{x}=\sum_{i=1}^{N} x_i/N, \quad \hat{y}=f(\bar{x}), \tag{1}$$

либо методом «б», когда последовательность операций обратна и

$$y_i=f(x_i), \ \bar{y}=\sum_{i=1}^{N} y_i/N. \tag{2}$$

Оценка $\bar{y}$ (2) как выборочное среднее состоятельна и не смещена, а оценка $\hat{y}$ (1) в общем случае несостоятельна, смещена и фигурирует под названием «псевдооценка» [7].

Для экспериментов, когда число $N$ невелико, несмещенность и состоятельность оценки (2) не гарантируют достаточной точности, и для повышения эффективности оценивания может оказаться целесообразным переход к псевдооценкам, если среднеквадратическая ошибка их будет меньше дисперсии оценки (1).

Чтобы показать, в каких случаях и при каких объемах выборки псевдооценка может оказаться более эффективной, сравним среднеквадратическую точность оценок и псевдооценок. Для этого запишем разницу их среднеквадратических ошибок как функцию от $N$ и исследуем ее поведение:

$$\Delta(N)=M\left[(\hat{y}-m_y)^2\right]-M\left[(\bar{y}-m_y)^2\right]=(m_y-m_{\hat{y}})^2+\sigma_{\hat{y}}^2-\sigma_y^2/N. \tag{3}$$

Здесь $m_y$, $\sigma_y^2$, $m_{\hat{y}}$, $\sigma_{\hat{y}}^2$ — математическое ожидание и дисперсия $y$ и $\hat{y}$. Знак $\Delta(N)$ свидетельствует о преимуществе того или иного метода оценивания: при $\Delta(N)<0$ лучше псевдооценка, и наоборот.

Для исследования поведения $\Delta(N)$ выпишем моменты из (3), обозначив через $F(x)$, $\overset{*}{F}(\bar{x})$, $F(\overset{\circ}{x})$ и $\overset{*}{F}\left(\overset{\circ}{\bar{x}}\right)$ функции распределения величин $x$, $\bar{x}$, $\overset{\circ}{x}=x-m$ и $\overset{\circ}{\bar{x}}=\bar{x}-m$, где $m$ — математическое ожидание величин $x$ и $\bar{x}$:

$$m_y=\int_{x} f(x)\,dF(x)=\int_{\overset{\circ}{x}} f\left(m+\overset{\circ}{x}\right)dF\left(\overset{\circ}{x}\right);\ m_{\hat{y}}=\int_{\bar{x}} f\left(\bar{x}\right)d\overset{*}{F}\left(\bar{x}\right)=$$

$$=\int_{\overset{\circ}{\bar{x}}} f\left(m+\overset{\circ}{\bar{x}}\right)d\overset{*}{F}\left(\overset{\circ}{\bar{x}}\right);$$

$$\sigma_y^2=\int\limits_{\mathring{x}} f^2(m+\mathring{x})\,dF(\mathring{x})-\left[\int\limits_{\mathring{x}} f(m+\mathring{x})\,dF(\mathring{x})\right]^2; \tag{4}$$

$$\sigma_{\hat{y}}^2=\int\limits_{\mathring{\bar{x}}} f^2(m+\mathring{\bar{x}})\,d\overset{*}{F}(\mathring{\bar{x}})-\left[\int\limits_{\mathring{\bar{x}}} f(m+\mathring{\bar{x}})\,d\overset{*}{F}(\mathring{\bar{x}})\right]^2.$$

Потребуем, чтобы функции $f(x)$ и $f^2(x)$ можно было разложить в ряд Тейлора относительно $m$ на всем интервале существования величин $\mathring{x}$ и $\mathring{\bar{x}}$. Кроме того, потребуем, чтобы центральные моменты распределений $F(x)$ и $F(\bar{x})$ были конечны. Тогда, разложив функции $f(\mathring{x})$, $f(\mathring{\bar{x}})$, $f^2(\mathring{x})$ и $f^2(\mathring{\bar{x}})$ из соотношений (4) в ряд Тейлора и меняя местами операции интегрирования и суммирования, получим:

$$m_y=\int\limits_{\mathring{x}}\left[f(m)+\sum_{i=1}^{\infty} f^{(i)}(m)\,\mathring{x}^i/i!\right]dF(\mathring{x})=f(m)+\sum_{i=2}^{\infty} f^{(i)}(m)\,\mu_i/i!;$$

$$m_{\hat{y}}=f(m)+\sum_{i=2}^{\infty} f^{(i)}(m)\,\tilde{\mu}_i/i!;$$

$$\sigma_y^2=\mu_2[f'(m)]^2+\mu_3[f'(m)f''(m)]^2+\mu_4 f'(m)f'''(m)/3+$$
$$+\left(\mu_4-\mu_2^2\right)[f''(m)]^2/4+\sum_{k=5}^{\infty}\left\{2\mu_i f^{(i-1)}(m)\,f'(m)/(i-1)!+\right.$$
$$\left.+\sum_{k=2}^{i-2} f^{(i-k)}(m)\,f^{(k)}(m)\,(\mu_i-\mu_{i-k}\mu_k)/(i-k)!k!\right\};$$

$$\sigma_{\hat{y}}^2=\tilde{\mu}_2[f'(m)]^2+\tilde{\mu}_3^2[f'(m)f''(m)]^2+\tilde{\mu}_4 f'(m)f''(m)/3+\left(\tilde{\mu}_4-\tilde{\mu}_2^2\right)[f''(m)]^2/4+$$
$$+\sum_{i=5}^{\infty}\left\{2\tilde{\mu}_i f^{(i-1)}(m)f'(m)/(i-1)!+\sum_{k=2}^{i-2} f^{(i-k)}(m)\,f^{(k)}(m)\,(\mu_i-\mu_{i-k}\mu_k)/(i-k)!k!\right\}, \tag{5}$$

где $\mu_i$ и $\tilde{\mu}_i$ — $i$-е центральные моменты величин $x$ и $\bar{x}$. Подставим полученные выражения в (3):

$$\Delta(N)=(\tilde{\mu}_3-\mu_3/N)[f'(m)f''(m)]^2+(\tilde{\mu}_4-\mu_4/N)f'(m)f'''(m)/3+$$
$$+\left(\tilde{\mu}_4-\mu_4/N+(N-1)\mu_2^2/N\right)[f''(m)]^2/4+\sum_{i=5}^{\infty}\left\{2(\tilde{\mu}_i-\mu_i/N)f^{(i-1)}(m)\times\right.$$
$$\times f'(m)/(i-1)!+\sum_{k=2}^{i-2} f^{(i-k)}(m)\,f^{(k)}(m)\,[\tilde{\mu}_i-\mu_i/N+(N-1)\mu_{i-k}\mu_k/N-$$
$$\left.-\tilde{\mu}_{i-k}\mu_k-\mu_{i-k}\tilde{\mu}_k]/(i-k)!k!\right\}. \tag{6}$$

Выражение (6) представляет собой разность среднеквадратических ошибок оценки $\bar{y}$ и псевдооценки $\hat{y}$ и является функцией высших центральных моментов величин $x$ и $\bar{x}$, значений высших производных преобразования $y=f(x)$ в точке $m$, а также объема выборки $N$.

В общем случае, не предъявляя новых требований к $F(x)$ и $y=f(x)$, исследовать $\Delta(N)$ по (6) практически невозможно. Поэтому в дальнейшем мы будем выяснять поведение $\Delta(N)$ для определенных классов $F(x)$.

**Распределение $F(x)$ устойчиво.** Задача исследования выражения $\Delta(N)$ существенно упрощается, если центральные моменты $\tilde{\mu}_i$ представить в виде линейной функции от $\mu_i$.

Подобное представление возможно в тех случаях, когда распределение $F(\overset{\circ}{x})$ устойчиво, т. е. когда сумма $\sum \overset{\circ}{x}_i$ распределена так же, как $C_N\overset{\circ}{x}$, где $C_N$ — константа вида $N^{1/\alpha}$, $0<\alpha\leqslant 2$ [8]. С учетом требований, предъявленных выше к законам распределений $F(x)$, для нас представляет интерес среди устойчивых распределений лишь нормальный закон. При этом выборочное среднее $\bar{\overset{\circ}{x}}$ распределено как $\overset{\circ}{x}N^{-1/2}$, и $\tilde{\mu}_i=\mu_i=0$ при $i=2h+1$, $\tilde{\mu}_i=\mu_i N^{-i/2}$ при $i=2h$. Выражение (6) в этом случае принимает вид

$$\Delta(N)=(1-N)\,\mu_4 f'(m)\,f'''(m)/3N^2+(1-N)\left(\mu_4-N\mu_2^2\right)[f''(m)]^2/4N^2+$$
$$+\sum_{\substack{i=2h,\\ h=3,4\ldots}}^{\infty}\Biggl\{2\left(1-N^{h-1}\right)\mu_i f^{(i-1)}(m)\,f'(m)/N^h(i-1)!+\sum_{k=2}^{i-2}f^{(i-k)}(m)\,f^{(k)}(m)\times$$
$$\times\left[\left(1-N^{h-1}\right)\mu_i+\left(N^h+N^{h-1}-N^{k/2}-N^{(i-k)/2}\right)\mu_{i-k}\mu_k\right]/N^h(i-k)!k!\Biggr\}. \quad (7)$$

Выше упоминалось, что псевдооценка $\hat{y}$ может быть точнее оценки $\bar{y}$ при объемах выборок, для которых $\Delta(N)$ отрицательно. Найдем условия, когда одновременно все слагаемые (7) неположительны и, следовательно, $\Delta(N)\leqslant 0$:

$$(1-N)\mu_4 f'(m)f'''(m)/3N^2\leqslant 0 \text{ при } N\geqslant 2,\ \operatorname{sign}[f'(m)]=\operatorname{sign}[f'''(m)];$$
$$(1-N)\left(\mu_4-N\mu_2^2\right)[f''(m)]^2/4N^2\leqslant 0 \text{ при } N\leqslant 3, \text{ так как } \mu_4=3\mu_2^2\,[9]; \quad (8)$$
$$2(1-N^{h-1})\mu_i f^{(i-1)}(m)f'(m)/N^h(i-1)!<0 \text{ при } N\geqslant 2,\ \operatorname{sign}[f^{(i-1)}(m)]=$$
$$=\operatorname{sign}[f'(m)];$$
$$f^{(i-k)}(m)f^{(k)}(m)[(1-N^{h-1})\mu_i+(N^h+N^{h-1}-N^{k/2}-N^{(i-k)/2})\mu_{i-k}\mu_k]/$$
$$/N^h(i-k)!k!\leqslant 0 \quad \text{при} \quad N\leqslant 5,$$
$$\operatorname{sign}[f^{(i-k)}(m)]=\operatorname{sing}[f^{(k)}(m)].$$

Последнее утверждение основывается на следующем: так как $i=2h$, то $i-k$ и $k$ имеют одинаковую четность. При одинаковых знаках $f^{(i-k)}(m)$ и $f^{(k)}(m)$ знак последнего выражения из (8) определяется знаком суммы в квадратных скобках:

$$S=\mu_i-N^{(h-1)}\mu_i+N^h\mu_{i-k}\mu_k+N^{h-1}\mu_{i-k}\mu_k-N^{k/2}\mu_{i-k}\mu_k-N^{(i-k)/2}\mu_{i-k}\mu_k. \quad (9)$$

Поскольку $i=2h$, а $k=2,\ldots,i-2$, то $S=\mu_i(1-N^{h-1})<0$ при $k=2l+1$. При $k=2l$ запишем выражение для $S$ по убывающим степеням $N$, учитывая, что $\mu_i=(i-1)!!\,\mu_2^h$ (9):

$$S=\mu_2^h\{(i-k-1)!!\,(k-1)!!N^h-[(i-1)!!-(i-k-1)!!\,(k-1)!!]\,N^{h-1}-$$
$$-(i-k-1)!!\,(k-1)!!N^l-(i-k-1)!!\,(k-1)!!\,N^{h-l}+(i-1)!!\}. \quad (10)$$

Распространяя область изменения $N$ на всю вещественную ось, выражение (10) можно рассматривать как полином от $N$ и исследовать распределение его корней. Коэффициенты при $N^i$ имеют два изменения знаков, следовательно, полином $S$, согласно теореме Декарта [10], имеет два положительных корня, один из которых $N_1=1$. Второй корень $N_2>1$, и из выражения (10) видно, что с увеличением $N$ он растет, что объясняется возрастанием отрицательного коэффициента при $N^{h-1}$. Минимальный второй положительный корень $N_2=5$ получается при $i=6$, что нетрудно показать, решая уравнение (10) для $h=3$.

Таким образом, при нормальном законе распределения $F(x)$ $\Delta N$ принимает отрицательные значения, если функциональное преобразование $y=f(x)$ таково, что знаки его высших производных одной и той же четности одинаковы. Напомним, что отрицательность $\Delta(N)$ означает, что псевдооценка оказывается точнее состоятельной несмещенной оценки, и, следовательно, имеет место «эффект малых выборок», исчезающий с увеличением $N$.

Функции с одинаковыми знаками производных одной и той же четности известны в классическом анализе. Это так называемые абсолютно и вполне монотонные функции и функции, обратные им [8].

В найденный класс попадают показательные, степенные, гиперболические и другие функции.

Пример. Пусть $y=c^x$, $c>0$, $x$ — нормальная случайная величина с неизвестными параметрами распределения $m$ и $\sigma^2$. Показательные функции попадают в найденный класс функций. Отыщем интервал значений $N$, для которого псевдооценка $\widehat{y}=c^{\bar{x}}$ точнее выборочного среднего $\bar{y}$. Можно показать, что в интервале $1<N\leqslant 7$ $\Delta(N)<0$, т. е. более точной всегда будет псевдооценка $\widehat{y}$. Подробный анализ выражения дан в работе [6].

**Распределение $F(x)$ симметричное и одномодовое.** Теперь не требуем устойчивости $F(x)$, сохраняя, однако, его симметричность. При этом распределение выборочного среднего $\overset{\circ}{\bar{x}}$ не совпадает с распределением случайной величины $C_N\overset{\circ}{x}$. Учитывая, что $\widetilde{\mu}_i=\mu_i=0$ при $i=2h+1$, запишем выражение для $\Delta(N)$:

$$\Delta(N)=(\widetilde{\mu}_4-\mu_4/N)\,f'(m)\,f'''(m)/3+(\widetilde{\mu}_4-\mu_4/N+(N-1)\,\mu_2^2/N)[f''(m)]^2/4+$$
$$+\sum_{\substack{i=2h,\\ h=3,4\ldots}}^{\infty}\Big\{2\,(\widetilde{\mu}_i-\mu_i/N)\,f^{(i-1)}(m)\,f'(m)/(i-1)!+\sum_{k=2}^{i-2}f^{(i-k)}(m)\,f^{(k)}(m)\times$$
$$\times[\widetilde{\mu}_i-\mu_i/N+(N-1)\,\mu_{i-k}\mu_k/N-\widetilde{\mu}_{i-k}\mu_k-\mu_{i-k}\widetilde{\mu}_k]/(i-k)!k!\Big\}. \qquad (11)$$

Для обнаружения эффекта малых выборок достаточно показать, что при $N=2$ существует класс преобразований $y=f(x)$, для которого $\Delta(2)<0$. Сравнение моментов $\mu_i$ и $\widetilde{\mu}_i$ проведем, используя известное соотношение между моментами $\mu_i$ и семиинвариантами $\varkappa_i$ [9]:

$$\exp(\varkappa_2t^2/2!+\varkappa_3t^3/3!+\ldots+\varkappa_rt^r/r!+\ldots)=1+\mu_2t^2/2!+\mu_3t^3/3!+\ldots$$
$$\ldots+\mu_rt^r/r!+\ldots. \qquad (12)$$

Поскольку для характеристической функции выборочного среднего $\overset{\circ}{\bar{x}}$ справедливо соотношение $\varphi_{\overset{\circ}{\bar{x}}}(t)=\varphi^N_{\overset{\circ}{x}}(t/N)$ [8], то, учитывая, что у симметричных распределений $\varkappa_i=0$, $i=2h+1$ [9],

$$\varphi_{\overset{\circ}{\bar{x}}}(t)=\exp[N(\varkappa_2(it)^2/N^22!+\varkappa_4(it)^4/N^44!+\ldots+\varkappa_{2k}(it)^{2k}/N^{2k}2k!+\ldots)]=$$
$$=\exp(\widetilde{\varkappa}_2(it)^2/2!+\widetilde{\varkappa}_4(it)^4/4!+\ldots+\widetilde{\varkappa}_{2k}(it)^{2k}/2k!+\ldots), \qquad (13)$$

где $\widetilde{\varkappa}_i=\varkappa_i/N^{i-1}$.

Полученное выражение показывает связь между семиинвариантами распределений $F(x)$ и $\overset{*}{F}(\bar{x})$. Возвращаясь к выражению (12) и задавая $N=2$, можно сравнить моменты $\mu_i$ и $\widetilde{\mu}_i$. Так, из соотношения $\widetilde{\mu}_4=\widetilde{\varkappa}_4+3\widetilde{\varkappa}_2=\mu_4/8+3\mu_2^2/8$ нетрудно оценить зависимость между $\mu_4$ и $\widetilde{\mu}_4$. Используя таблицу коэффициентов эксцесса $\beta_2=\mu_4/\mu_2^2$ для наиболее распространенных симметричных одномодальных распределений [11], выясняем, что, как правило, $\beta_2\geqslant 2{,}4$, т. е. $\widetilde{\mu}_4<\mu_4/3$. Теперь, рассматривая слагаемые выражения (11), получаем следующее распределение их знаков:

$$(\widetilde{\mu}_4-\mu_4/2)f'(m)f'''(m)/3<0,\ \operatorname{sign}[f'(m)]=\operatorname{sign}[f'''(m)];$$
$$(\widetilde{\mu}_4-\mu_4/2+\mu_2^2/2)\,[f''(m)]^2/4<0,\ \ \beta>2,3;\ \ 2\,(\widetilde{\mu}_i-\mu_i/2)\,f^{(i-1)}(m)\times$$
$$\times f'(m)/(i-1)!<0,\quad \operatorname{sign}[f^{(i-1)}(m)]=\operatorname{sign}[f'(m)]. \qquad (14)$$

Справедливость неравенства $\tilde{\mu}_i < \mu_i/2$ доказывается аналогично соотношению между $\tilde{\mu}_4$ и $\mu_4$ с использованием выражений (12) и (13):

$$f^{(i-k)}(m)f^{(k)}(m)[\tilde{\mu}_i - \mu_i/2 + \mu_{i-k}\mu_k/2 - \tilde{\mu}_{i-k}\mu_k - \mu_{i-k}\tilde{\mu}_k]/(i-k)!k! < 0,$$
$$\operatorname{sign}[f^{(i-k)}(m)] = \operatorname{sign}[f^{(k)}(m)]. \qquad (15)$$

Последнее утверждение основывается на следующем: индекс $i$, как видно из формулы (11), всегда четный; при $k = 2l+1$ $\mu_{i-k} = \mu_k = 0$ и числитель выражения (15) превращается в $\tilde{\mu}_i - \mu_i/2 < 0$, как в (14). При $k = 2l$ слагаемые $\tilde{\mu}_i$ и $\mu_i/2$ не зависят от $k$, а член $\mu_{i-k}\mu_k/2$ максимален при $k = 2$ (так как при этом $\mu_{i-k}$ имеет наибольший возможный порядок для данного $i$). Учитывая, что в этом случае последнее слагаемое $\mu_{i-2}\tilde{\mu}_2 = \mu_{i-2}\mu_2/2$, получаем окончательный вид числителя выражения (15):

$$\tilde{\mu}_i - \mu_i/2 - \tilde{\mu}_{i-2}\mu_2 < \tilde{\mu}_i - \mu_i/2 < 0,$$

доказав тем самым утверждение о знаке (15).

Итак, мы убедились, что при $N=2$ разница в точности $\Delta(2) < 0$ и псевдооценка предпочтительней, если функции $y = f(x)$ удовлетворяют условиям в выражениях (14) и (15). Нетрудно заметить, что это тот же класс функций, что был найден для предыдущего случая: подкласс абсолютно и вполне монотонных функций. Однако в данном случае этот подкласс шире, так как теперь требование бесконечности радиусов сходимости рядов Тейлора необязательно, если распределение $F(x)$ финитное (равномерное, треугольное, параболическое и т. д.). Для подобных $F(x)$ найденный подкласс может, наряду с приведенными ранее, содержать и функции типа $y = x^{\alpha}$, где $\alpha$ — рациональная дробь.

Пример. Пусть $y = x^{1/2}$ — функция, попадающая в найденный класс. Аргумент $x$ равномерно распределен на интервале $[0, a]$, параметр $a > 0$ неизвестен. Будем сравнивать эффективность псевдооценки $\hat{y} = \bar{x}^{1/2}$ и выборочного среднего $\bar{y} = \sum x_i^{1/2}/N$ для $N = 2$. Так как $\bar{x}$ при $N = 2$ имеет симметричное треугольное распределение [8]

$$d\overset{*}{F}(\bar{x}) = \begin{cases} 4\bar{x}/a^2, & 0 \leqslant \bar{x} \leqslant a/2, \\ 4(a-\bar{x})/a^2, & a/2 < \bar{x} < a, \end{cases}$$

соответствующие моменты $\Delta(N)$ (3) таковы:

$$m_y = 2a^{1/2}/3;\ m_{\hat{y}} = \left(16 - 4\sqrt{2}\right)a^{1/2}/15;\ \ \sigma_y^2 = a/18;\ \ \sigma_{\hat{y}}^2 = a/45.$$

Подставляя полученные моменты в (3), находим $\Delta(2) \approx -0{,}005a$. И в данном случае псевдооценка $\hat{y}$ точнее, причем разность в ошибках $\hat{y}$ и $\bar{y}$ весьма существенна: около 25% дисперсии самой величины $y$.

**Распределение $F(x)$ с неотрицательными семиинвариантами $\varkappa_i$.** Наложим на $F(x)$ иное условие: потребуем, чтобы семиинварианты $\varkappa_i$ были неотрицательны. Поэтому в общем случае $\mu_i \neq 0$ при $i = 2h+1$, и в выражении (6) для $\Delta(N)$ будут присутствовать все члены. Класс подобных распределений весьма важен. В него входят, например, гамма-распределения и распределение Пуассона.

Используем тот же метод определения знака $\Delta(N)$, что и в предыдущих случаях, и рассмотрим слагаемые выражения (6). Для сравнения $\mu_i$ и $\tilde{\mu}_i$ опять учтем соотношение между моментами и семиинвариантами (13) и выражение (14):

$$(\tilde{\mu}_4 - \mu_4/N)f'(m)f'''(m)/3 < 0,\ \operatorname{sign}[f'(m)] = \operatorname{sign}[f'''(m)];$$
$$\left(\tilde{\mu}_4 - \mu_4/N + (N-1)\mu_2^2/N\right)[f''(m)]^2/4 < 0,\ \ N \leqslant 3; \qquad (16)$$
$$(\tilde{\mu}_i - \mu_i/N)f^{(i-1)}(m)f'(m)/(i-1)! < 0,\ \operatorname{sign}[f^{(i-1)}(m)] = \operatorname{sign}[f'(m)].$$

Для анализа последнего слагаемого выражения (6)

$$S = f^{(i-k)}(m) f^{(k)}(m) [\widetilde{\mu}_i - \mu_i/N + (N-1)\mu_{i-k}\mu_k/N - \widetilde{\mu}_{i-k}\mu_k - \mu_{i-k} \times \times \widetilde{\mu}_k]/(i-k)!k! \quad (17)$$

возвратимся к последнему члену из (8), который при $N \leqslant 5$ и $\operatorname{sign}[f^{(i-k)}(m)] = \operatorname{sign}[f^{(k)}(m)]$ был отрицательным. Учитывая соотношения (12) и (13) и тот факт, что все $\varkappa_i > 0$, можно утверждать, что $\widetilde{\mu}_i < \mu_i/N^{i/2}$ (в формулах (8) было равенство), поэтому для выражения (17) интервал значений $N$, при котором $S < 0$, будет не меньше, чем в случае устойчивых $F(x)$.

Из условия неположительности выражений (16) и (17) следует, что все производные функции $y = f(x)$ в точке $m$ должны иметь одинаковые знаки, т. е. $f(x)$ составляют подкласс абсолютно монотонных функций. Примерами подобных функций могут служить степенные и показательные.

Пример. Пусть $y = x^n$, $x$ — случайная величина, распределенная по гамма-закону $dF(x) = \eta^\nu x^{\nu-1} e^{-\eta x}/\Gamma(\nu)$, $x > 0$. Параметры $\eta > 0$, $\nu > 0$ неизвестны. Функции $y = x^n$ попадают в найденный класс, являясь при $x > 0$ абсолютно монотонными. Сравним точность псевдооценки $\widehat{y} = \bar{x}^n$ и выборочного среднего $\bar{y} = \sum x_i^n/N$.

Рассматривая $\Delta(N)$ как полином от $\alpha = 1/N$ и исследуя поведение его корней [6], можно показать, что при $\alpha \in [1/7, 1]$ $\Delta(\alpha) \leqslant 0$, т. е. при $1 < N \leqslant 7$, псевдооценка всегда точнее.

**Обсуждение. Что такое «малая» выборка?** Нами показано, как в случае, когда число наблюдений не превосходит десяти, среднее арифметическое — оценка, обладающая такими несомненными асимптотическими достоинствами, как несмещенность и состоятельность, оказалась менее точной, чем псевдооценка. По этому поводу можно сделать несколько замечаний. Во-первых, указание на классы преобразований, для которых этот эффект возможен, не является еще достаточным условием. Выражение (6) нуждается в более тщательном анализе: оно может быть отрицательным и при положительности некоторых слагаемых.

Далее, в последнее время в жертву точности оценок (в среднеквадратическом смысле) уже принесено такое их свойство, как несмещенность. Смещенность не только допускается, но и вводится порой искусственно [12]. Здесь в тех же интересах отказываемся и от состоятельности.

Наконец, полученные результаты могут помочь оговорить, что же такое «малая» выборка. Действительно, «малая» выборка должна отличаться от «большой» некоторым скачком в качестве вычисленных по ней оценок, а не только числом элементов. В нашем случае мы наблюдаем именно такое явление. При минимальном числе наблюдений присутствует «эффект малой выборки», точнее, псевдооценка. По мере увеличения $N$ ее преимущество уменьшается, наконец, асимптотические качества вступают в силу, оценка становится более точной. В этой ситуации естественно провести черту между «большой» и «малой» выборками именно здесь, при равенстве точности оценки и псевдооценки. Очевидно, эта граница подвижна, зависит от законов распределений, классов преобразований, значений оцениваемых параметров.

## ЛИТЕРАТУРА


1. **Ефимов А. Н., Рубанов В. Г.** Некоторые особенности экстраполяции при суммировании значений случайных функций.— Кибернетика, 1970, № 5.
2. **Колев Л. В.** Прогнозиране на сума от нестанционарни случайни процеси.— Изв. ВМЕИ «Ленин», 1976, т. 32, № 4.
3. **Ицкович Э. Л., Подвальный С. Л.** Сравнение методов учета динамических связей, существующих между измеряемыми и искомыми величинами.— Автоматика и телемеханика, 1967, № 8.
4. **Гурин Л. С.** О перестановочности осреднения и приведения.— УМН, 1952, т. VII, вып. 1 (47).

5. **Ефимов А. Н., Криворуков Е. В.** О выборе рациональной последовательности оптимальных линейных статистических операций и функциональных преобразований случайных процессов.— Автоматика и телемеханика, 1977, № 7.
6. **Ефимов А. Н., Криворуков Е. В.** Некоторые особенности усреднения малых выборок при косвенных наблюдениях.— В кн.: Вопросы атомной науки и техники. Сер. Автоматизация и математическое обеспечение физического эксперимента. Харьков, 1977, вып. 1 (7).
7. **Кульман Н. К., Титаренко Н. В.** Интегрирование непрерывного марковского процесса в случае косвенного наблюдения.— Автометрия, 1977, № 1.
8. **Феллер В.** Введение в теорию вероятности и ее приложения.— М.: Мир, 1967, т. 2.
9. **Кендалл М. Дж., Стьюарт А.** Теория распределений.— М.: Наука, 1966.
10. **Корн Г., Корн Т.** Справочник по математике.— М.: Физматгиз, 1974.
11. **Grow E. L., Siddiqvi M. M.** Robust Estimation of Location.— J. American Statistical Association, 1967, vol. 62, N 318, p. 353—389.
12. **Виленкин С. Я.** Статистическая обработка результатов исследования случайных функций.— М.: Энергия, 1979.

*Поступила в редакцию 26 марта 1981 г.;*
*окончательный вариант — 15 июля 1982 г.*

---

УДК 681.3.068

**Ю. А. ГОРИЦКИЙ, С. Е. ЖАРИНОВ**
*(Петропавловск-Камчатский)*

# ОРГАНИЗАЦИЯ ГРАФИЧЕСКОЙ ДИАЛОГОВОЙ СИСТЕМЫ АНАЛИЗА ДАННЫХ

**Введение.** Задачи анализа многомерных данных, где признаки носят количественный характер и их число невелико, типичны для естественнонаучных исследований [1]. В настоящее время их решение зачастую связывают с диалоговым подходом [2], когда главная роль в анализе (визуальном) структуры и оценке результатов отводится самому исследователю, а ЭВМ выполняет всю вспомогательную работу по преобразованию информации к виду, удобному для наглядного представления. Такой подход в различной степени реализован в ряде систем распознавания образов, классификации и отображения информации [3—7]. Однако, как указано в [2], большей частью они не отвечают требованиям, которым должна удовлетворять диалоговая система анализа данных, и в первую очередь требованиям по организации, поскольку основу их составляют заранее определенные наборы алгоритмов со своими специфическими взаимосвязями.

В качестве альтернативы может быть предложена предварительная разработка вопросов организации с учетом всех аспектов задачи анализа данных, в том числе получение соответствующего «скелета» системы. В работе приводится описание структурной схемы, организации данных и диалога системы «Вулканит» Института вулканологии ДВНЦ АН СССР [8]. Первый вариант системы реализован на ЭВМ «Минск-32», второй — на ЭВМ ЕС-1033, однако рассматриваемые вопросы являются машинно-независимыми.

**Структура системы.** Состав системы и ее структура определяются классом решаемых задач. Как известно, процесс анализа данных обычно носит итерационный характер, где каждая итерация состоит из трех шагов: выбор данных для анализа, выбор способа анализа из имеющегося набора и формы представления результатов, оценка результатов.

Первые два выполняются совместно исследователем и ЭВМ, поэтому система должна содержать возможности манипуляций над данными, графического отображения и собственно способы анализа, которые условно можно разделить на три группы в зависимости от имеющейся априорной информации. Так, при наличии модели изучаемой структуры (например, статистической) интерес представляет определение параметров этой мо-