

УДК 389.629

Л.А. Назаренко<sup>1</sup>, П.Ф. Шапов<sup>2</sup><sup>1</sup>Национальный научный центр «Институт метрологии», Харьков<sup>2</sup>Национальный технический университет «Харьковский политехнический институт»

## СТАТИСТИЧЕСКИЕ МЕТОДЫ АНАЛИЗА НЕОПРЕДЕЛЕННОСТИ РЕЗУЛЬТАТОВ КОСВЕННЫХ ИЗМЕРЕНИЙ

*Рассмотрены методы дисперсионного и дискриминантного анализа статистических моделей косвенного измерения, когда остаточная неопределенность моделей является мерой неопределенности результатов измерений. Показана возможность планирования косвенных измерений по числу аргументов и по количеству уровней выходной величины.*

*методы дисперсионного анализа, косвенные измерения*

### Введение

**Постановка проблемы.** Неопределенность результатов измерений характерна для любых процедур получения информации о физических явлениях и процессах в любой области человеческой деятельности, не исключая строго регламентированных процедур метрологического измерительного эксперимента. Измерение – это комплексная информационная процедура, занимающая важное место среди экспериментальных методов познания и от точности измерений зависит эффективность информационных технологий контроля, диагностики, идентификации. Наиболее проблемными здесь являются не прямые измерения, использующие функциональные преобразования рода исходных (входных) измеряемых величин. Проблема уменьшения неопределенности выходной величины, в этом случае, – это проблема выбора адекватной модели измерительного преобразования и решить ее невозможно без изучения вероятностно-статистической модели объекта измерения.

**Анализ литературы.** Стохастическая связь априорно неопределенной выходной величины  $Y$  с вектором входных величин (аргументов)  $\bar{X} = (X_1, \dots, X_p)$  определяет два варианта математической модели объекта косвенных измерений.

Аналоговый объект – функция измерительного преобразования представлена в виде множественной регрессии, а аргументы являются неслучайными величинами. Случайная величина – это  $Y$  [1]. Дискретный объект – уровни  $U_1, \dots, U_k$  величины  $Y$

неслучайны и фиксированы, случайным же является вектор аргументов [2].

Функция измерительного преобразования представляет собой алгоритм, включающий процедуры принятия решений. Обе модели априори предполагают известными плотности распределения как  $Y$ , так и вектора  $\bar{X}$  (случай бесконечно больших выборок). Влияние же ограниченности информации на точность оценивания плотностей распределения величин  $Y$  и  $\bar{X}$  и на правильность выбора модели их стохастической связи – практически не исследовано.

**Цель статьи** – показать возможности методов дисперсионного анализа для изучения вероятностной структуры стохастического влияния косвенно измеряемой величины  $Y$  на значения аргументов  $X_1, \dots, X_p$  при ограничениях на число измерительных экспериментов с целью последующего синтеза градуировочной характеристики, обеспечивающей минимальную неопределенность результатов измерения величины  $Y$ .

### Оценка значимости стохастической связи

Такую оценку необходимо проводить отдельно по каждому аргументу, тестируя их значимость с учетом вероятностей ошибок, как первого ( $\alpha$ ), так и второго ( $\beta$ ). Единственная модель, которая подходит для статистического анализа результата измерения  $X_{ij}^{(j)}$  для  $j$ -го аргумента – это модель компонент дисперсий [3], в которой влияющим фактором является

выходная величина

$$X_{ti}^{(j)} = \bar{X}^{(j)} + U_t + Z_{ti},$$

где  $\bar{X}^{(j)}$  – общее среднее всех измерений по K уровням величины Y;  $j = 1, \dots, P$ ;

$U_t$  – отклонение от  $\bar{X}^{(j)}$ , обусловленное влиянием Y;  $t = 1, \dots, K$ ;

$Z_{ti}$  – случайный остаток модели;  $i = 1, \dots, n$ ;

$n$  – число многократных измерений аргумента  $X_j$ , одинаковое для всех K уровней величины Y.

Дисперсионное разложение для  $X_{ti}^{(j)}$  включает два слагаемых (по  $U_t$  и  $Z_{ti}$ )

$$S = S_u + S_z.$$

Величина критериальной F-статистики, определяемой центральным линейнопреобразованным  $F_{k-1, k(n-1)}$  – распределением, может использоваться для ранжирования аргументов по убыванию стохастической связи с Y (по убыванию F-статистики) при заданных  $\alpha$  и  $\beta$ .

### Оптимизация числа аргументов

Используется более сложная модель ковариационного анализа стохастического влияния Y на подмножества аргументов, размерностью 2, 3, ..., P. Для каждого из аргументов результат измерения представлен моделью, включающей 5 слагаемых. Дисперсионное разложение, соответственно, имеет вид [3]

$$S = S_o + S_{WG} + S_G + S_W + S_R,$$

где  $S_{WG}$  и  $S_G$  – отвечают за эффекты аддитивного изменения неопределенности;

$S_W$  – за эффект мультипликативного влияния;

$S_o$  – за влияние величины Y;

$S_R$  – за влияние случайного остатка модели.

Вычленение из модели результатов измерения слагаемых, обусловленных суммами  $S_{WG}$ ,  $S_G$  и  $S_W$ , уменьшает неопределенность в оценке значимости влияния Y на подмножество аргументов ( $S_R < S_z$ ). Подмножество из  $P_o$  аргументов ( $P_o \leq P$ ), для которого критериальная статистика  $F_{1, N-2k} = S_o(N - 2k)/S_R$  максимальна, обеспечивает минимум неопределенности для функциональной модели измерительного преобразования в виде множественной регрессии ( $N$  – общее число измерений):

$$\hat{Y} = b_o + \sum_{j=1}^{P_o} b_j X_j.$$

Преимущества модели – планирование измерительного эксперимента по K и P. Недостаток – модель влияния Y на  $X_1, \dots, X_p$  должна быть линейной, однако при нарушениях предположений о линейности для всех аргументов оптимум для  $P_o$  не меняется. Следует добавить, что найденное оптимальное число из  $P_o$  аргументов может использо-

ваться для формирования алгоритмической модели косвенного измерения, причем число уровней такого дискретного объекта измерения также возможно оптимизировать [4]. При этом неопределенность результатов косвенного измерения еще более уменьшится, хотя теперь это произойдет за счет избыточности по числу уровней измеряемых величин.

### Оптимизация числа уровней выходной величины

Рассмотрим общий случай, когда аргументы  $X_1, \dots$ ; образуют вектор  $\bar{x}$ , функцией правдоподобия которого является условная p-мерная нормальная плотность распределения вероятности

$$W(\bar{x}|y_k) = (2\pi)^{-1/2} |\Sigma|^{-1/2} \times \exp\left[-\frac{1}{2}(\bar{x} - \mu^{(k)})' \Sigma^{-1}(\bar{x} - \mu^{(k)})\right], \quad (1)$$

где  $\bar{x}$  – вектор столбец измеренных значений;

$\mu^{(k)}$  – вектор условных средних;

$\Sigma$  – матрица ковариаций;

$|\Sigma|$  – определитель ковариационной матрицы, который соответствует условию невырожденности распределения  $W(\bar{x}/y_k)$ , то есть  $|\Sigma| > 0$ .

Вектор  $\mu^{(k)}$  и матрица  $\Sigma$  имеют вид:

$$\mu^{(k)} = \begin{pmatrix} \mu_1^{(k)} \\ \mu_2^{(k)} \\ \dots \\ \mu_p^{(k)} \end{pmatrix}; \quad \Sigma = \begin{pmatrix} \sigma_{11} & \sigma_{12} & \dots & \sigma_{1p} \\ \sigma_{21} & \sigma_{22} & \dots & \sigma_{2p} \\ \dots & \dots & \dots & \dots \\ \sigma_{p1} & \sigma_{p2} & \dots & \sigma_{pp} \end{pmatrix}.$$

Оценку элементов вектора  $\mu^{(k)}$  и матрицы  $\Sigma$  получают на этапе определения градуировочной характеристики, когда для каждого из уровней  $y_1, \dots, y_m$  используют  $n$  образцов, воспроизводящих каждый из этих уровней.

Заменяя  $\mu^{(k)}$  и  $\Sigma$  на их оценки  $M^{(k)}$  и  $S$ , где

$$M^{(k)} = \begin{pmatrix} m_1^{(k)} \\ m_2^{(k)} \\ \vdots \\ m_p^{(k)} \end{pmatrix}; \quad S = \begin{pmatrix} s_{11} & s_{12} & \dots & s_{1p} \\ s_{21} & s_{22} & \dots & s_{2p} \\ \dots & \dots & \dots & \dots \\ s_{p1} & s_{p2} & \dots & s_{pp} \end{pmatrix}. \quad (2)$$

Получим функцию правдоподобия в виде статистики:

$$W(\bar{x}|y_k) = (2\pi)^{-p/2} |S|^{-1/2} \times \exp\left[-\frac{1}{2}(\bar{x} - M^{(k)})' S^{-1}(\bar{x} - M^{(k)})\right]. \quad (3)$$

Статистика  $Z$  являється случайной величиной  $\xi^{(k)}$ , принимающей в каждой из  $m$  состояний свое значение, т.е.

$$\xi^{(k)} \in \{ \xi^{(1)}, \dots, \xi^{(m)} \}, k = \overline{1, m}.$$

В зависимости от вида ковариационной матрицы  $S$  (или  $\Sigma$ ) будем получать два варианта анализа.

1. Аргументы – независимы друг от друга, что соответствует случаю диагональной матрицы  $S$

$$S = \begin{pmatrix} s_{11} & 0 & \dots & 0 \\ 0 & s_{22} & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & s_{pp} \end{pmatrix}.$$

2. Аргументы – стохастически связаны друг с другом, что соответствует случаю матрицы (2).

Независимо от варианта анализа выбор решения  $\gamma_j \in \{ \gamma_1, \dots, \gamma_m \}$  производится в соответствии с правилом выбора по статистике  $\xi^{(j)}$

$$\gamma(\bar{x}) = \gamma_j, \text{ если } \xi^{(j)} = \sup \{ \xi^{(1)}, \dots, \xi^{(m)} \}.$$

Из выражения (3) видно, что максимум статистики  $\xi^{(k)}$  соответствует минимуму показателя экспоненты:

$$T^{(k)} = (\bar{x} - M^{(k)})' S^{-1} (\bar{x} - M^{(k)}). \quad (4)$$

Пусть  $Y \in y_k$  и пусть  $f(T^{(k)})$  – плотность распределения статистики (4).

Вероятность принятия решения  $\gamma_k$  равна

$$P(\gamma_k | Y \in y_k) = \int_0^{T_H} f(T^{(k)}) dT^{(k)} = 1 - \alpha. \quad (5)$$

Вероятность принятия решения  $\gamma_j, j \neq k$ , равна

$$P(\gamma_j | Y \in y_k) = \int_{T_H}^{\infty} f(T^{(k)}) dT^{(k)} = \alpha. \quad (6)$$

Предел интегрирования  $T_H$  определим, исходя из значений ширины зон допуска  $\Delta_i, i = \overline{1, p}$  по каждому из аргументов  $X_1, \dots, X_p$ , считая, что

$$\Delta_i = A_{xi} / m,$$

где  $A_{xi}$  – диапазон изменения показателя  $X_i$ .

Тогда

$$T_H = \Delta' S^{-1} \Delta, \quad (7)$$

где  $\Delta$  – вектор-столбец нормированных отклонений

$$\Delta = \begin{pmatrix} A_{X1} \\ 2m \\ A_{X2} \\ 2m \\ \vdots \\ A_{Xp} \\ 2m \end{pmatrix}.$$

Если  $\Delta_y$  – ширина зоны допуска, зависящая от заданного числа уровней  $m$  при фиксированном диапазоне  $A_y$ , т.е.

$$\Delta_y = A_y / m,$$

а  $\overline{\Delta_y}$  – средняя величина отклонения от уровня  $y_k$  при неправильной классификации этого уровня, то математическое ожидание ширины зоны допуска равно, с учетом выражений (5) – (7):

$$M[\Delta] = \Delta_y (1 - \alpha) + \overline{\Delta_y} \alpha.$$

Минимизация величины  $M[\Delta]$  по числу уровней не эквивалентна минимизации неопределенности результатов косвенного алгоритмического измерения [4] (по максимуму функции правдоподобия (1)).

## Выводы

Уменьшение неопределенности результатов косвенных измерений возможно за счет усложнения статистической модели влияния  $Y$  на  $\bar{X}$  и оптимизации числа аргументов, уровней выходной величины, повторных измерений.

## Список литературы

1. Захаров И.П., Кукуш В.Д. Теория неопределенности измерений: Учеб. пособие. – Х.: Консум, 2002. – 256 с.
2. Володарський Є.Т., Кухарчук В.В., Поджаренко В.О., Сердюк Г.Б. Метрологічне забезпечення вимірювань і контролю: Навч. посіб. – Вінниця: Велес, 2001. – 219 с.
3. Джонсон Н., Лион Ф. Статистика и планирование эксперимента в технике и науке: Методы планирования эксперимента: Пер. с англ. – М.: Мир, 1981. – 520 с.
4. Назаренко Л.А., Щапов П.Ф. Підвищення точності непрямих електричних методів вимірювання вологості при використанні метрологічно невизначених, вимірювальних сигналів // Український метрологічний журнал. – 2006. – Вип. 2. – С. 54-57.

Поступила в редколлегию 7.05.2007

**Рецензент:** д-р техн. наук, доц. И.П. Захаров, Харьковский национальный университет внутренних дел, Харьков.