УДК 389.629

Л.А. Назаренко[1], П.Ф. Щапов[2]

[1]*Национальный научный центр «Институт метрологии», Харьков*
[2]*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 непрямые 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) $\overline{X} = (X_1, \ldots, X_p)$ 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 $У_1, \ldots, У_к$ 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 $\overline{X}$ (случай бесконечно больших выборок). Влияние же ограниченности информации на точность оценивания плотностей распределения величин Y и $\overline{X}$ 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 $X_1, \ldots, X_p$ 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 (α), так и второго (β). Единственная модель, которая подходит для статистического анализа результата измерения $X_{ti}^{(j)}$ для j-го аргумента – это модель компонент дисперсий [3], в которой влияющим фактором является выходная величина

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

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

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

Z_{ti} – случайный остаток модели; i = 1, … , 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 α и β.

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 мультиплитативного 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 ($P_o \le P$), 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, … , 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 использоваться 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 $X_1,\ldots;$ образуют вектор $\overline{x}$, функцией правдоподобия которого является условная p-мерная нормальная плотность распределения вероятности

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

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

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

Σ – матрица ковариаций;

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

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

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

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

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

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

Получим функцию правдоподобия в виде статистики:

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

Статистика 3 является случайной величиной $\xi^{(k)}$, принимающей в каждой из m состояний свое значение, т.е.

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

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

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

$$S = \begin{vmatrix} s_{11} 0 \cdots 0 \\ 0 s_{22} \cdots 0 \\ \cdots\cdots\cdots \\ 0 0 \cdots s_{pp} \end{vmatrix}.$$

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

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

$$\gamma(\overline{x}) = \gamma_j,\ \text{если}\ \xi^{(j)} = \sup\left\{\xi^{(1)}, \ldots, \xi^{(m)}\right\}.$$

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

$$T^{(k)} = \left(\overline{x} - M^{(k)}\right)' S^{-1} \left(\overline{x} - M^{(k)}\right). \qquad (4)$$

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

Вероятность принятия решения γ_k равна

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

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

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

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

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

где A_{xi} – диапазон изменения показателя X_i.

Тогда

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

где Δ – вектор-столбец нормированных отклонений

$$\Delta = \begin{vmatrix} \dfrac{A_{X1}}{2m} \\ \dfrac{A_{X2}}{2m} \\ \vdots \\ \dfrac{A_{XP}}{2m} \end{vmatrix}.$$

Если Δ_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 $\overline{X}$ и оптимизации числа аргументов, уровней выходной величины, повторных измерений.

Список литературы

1. *Захаров И.П., Кукуш В.Д. Теория неопределенности измерениях: 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.