

# МАТЕМАТИЧЕСКАЯ ТЕОРИЯ ИЗМЕРИТЕЛЬНЫХ ЗАДАЧ

Левин С.Ф. (S. Levin), д. т. н., профессор, МИЗИ Ростехрегулирования

## ЧАСТЬ 10. МЕТОД СОВМЕСТНЫХ ИЗМЕРЕНИЙ

Метод совместных измерений применяют для решения измерительных задач идентификации математических моделей зависимостей между физическими величинами, характеризующими свойства объектов измерений [1]. Математическим аппаратом решения этих измерительных задач являются различные методы статистического анализа, так как методы идентификации моделей, как правило, сочетают с методом многократных измерений [2].

При идентификации модели в статических измерительных задачах: 1) устанавливают согласно плану значения физических величин, входных переменных модели, и измеряют физическую величину — выходную переменную; 2) по данным измерений переменных находят характеристики положения и рассеяния модели — параметры положения и рассеяния выходной переменной как функции входных переменных; 3) нормализуют данные измерений характеристиками положения и рассеяния в схеме перекрестного наблюдения [1], находя распределения выходной переменной или функции погрешности неадекватности относительно характеристики положения модели алгоритмами метода многократных измерений.

Применение метода совместных измерений связано с понятиями случайной, неопределенной и стохастической величин. Этими понятиями часто оперируют как равнозначными, причем последним из них — как эквивалентом первых двух [3].

В теории вероятностей случайная величина определена как однозначная действительная функция  $\Xi(\omega)$  на множестве значений  $\Omega$  с системой подмножеств и вероятностной мерой  $P\{\omega: \Xi(\omega) < \xi\} = P_X\{-\infty, \xi\} = F_X(\xi)$  — функцией распределения вероятностей [4]. На практике же случайную называют величину  $\Xi$ , принимающую в опыте непредсказуемые значения [2].

Неопределенной называют величину  $\Psi$ , значение которой не определено или неизвестно, но может быть описано распределением вероятностей. Так описывают ненаблюдаемую составляющую результата решения измерительной задачи или величину, статистическое распределение значений которой эквивалентно данному распределению вероятностей.

Далее стохастической величиной будем называть композицию случайной и неопределенной величин. Для суммы наблюдаемой величины с функцией распределения  $F_*(\xi)$  вида «\*» и ненаблюдаемой величины с равномерным «\*»= $R$ » распределением на интервале

$[a, b]$ , представляющих отклонения данных совместных измерений от характеристики положения модели в качестве функции погрешности неадекватности, параметры которой могут быть функциями входных переменных, будем использовать семейство распределений Леви [2, 5-7]:

$$f_{*R}(\delta) = \frac{F_*(\delta - a) - F_*(\delta - b)}{b - a}. \quad (1)$$

Основными проблемами практического применения метода совместных измерений являются нарушения условий применимости математического аппарата статистического анализа, неполный учет ненаблюдаемой составляющей функции погрешности неадекватности модели и представление модели только зависимостью параметра положения распределения вероятностей выходной переменной от входных переменных — характеристикой положения, что ведет к завышению реальной точности моделей объектов измерений.

### 10.1. ИЗМЕРИТЕЛЬНАЯ ЗАДАЧА ИДЕНТИФИКАЦИИ МОДЕЛИ ФИЗИЧЕСКОГО ОБЪЕКТА

Решение измерительной задачи идентификации модели имеет ряд особенностей [1]:

1) если перечень переменных модели задан, а ее аналитическая структура известна с точностью до параметров, то задачу решают методами параметрической идентификации;

2) если аналитическая структура модели задана набором вариантов (моделью максимальной сложности), то применяют методы структурно-параметрической идентификации;

3) если аналитическая структура модели и перечень ее переменных неизвестны, то задачу решают методами полной идентификации как начальную задачу математической статистики.

Обычно структуру модели объекта устанавливают из физических соображений:

$y = \Phi_s(\Theta_M; X_1, X_2, \dots, X_Q) + \Delta_{*s}(\Theta_*; X_1, X_2, \dots, X_Q)$ , где  $\Theta_* = \{\theta_{*1}, \theta_{*2}, \dots\}$  и  $\Theta_M = \{\theta_m; m = 0, M-1\}$  — множества параметров распределения вероятностей и характеристики положения модели,  $\Delta_{*s}(\Theta_*; X_1, X_2, \dots, X_Q)$  — функция погрешности неадекватности с наблюдаемой (случайной)  $\Xi_{*s}$  и ненаблюдаемой (неопределенной или неисключенной систематической)  $\Psi_{*s}$  составляющими [2], — код структуры модели [1].

При отсутствии данных о структуре модели применяют линейные по параметрам интерполяционные модели с матрицей базисных функций  $\Omega^T = [\Omega_{i,j,k}(X_1^i, \dots, X_Q^j, \dots, X_Q^k)] -$

$$y_g = \theta_{0,\dots,0,0} + \sum_{i=1}^I \dots \sum_{j=1}^J \dots \sum_{k=1}^K \theta_{i,j,k} \cdot \Omega_{i,j,k}(X_1^i, \dots, X_Q^j, \dots, X_Q^k) + \Delta_g, \\ y_g = \theta_0 + \sum_{m=1}^M \theta_m \cdot (x - x_0)^m + \Delta_g. \quad (2)$$

Классическая идея метода совместных измерений заключается в измерении переменных модели объекта при изменении множества входных переменных  $X$  в порядке, вызывающем изменение выходной переменной  $Y$ . Данные измерений представляют матрицами

$$\hat{Y}^T = [\hat{y}_1 \hat{y}_2 \dots \hat{y}_{N-1} \hat{y}_N], \\ \hat{X} = [\hat{x}_{qn}], q = 1, Q, n = 1, N,$$

и интерпретируют как значения одинаково распределенных и независимых случайных величин, для которых вид «\*» совместной плотности распределения вероятностей известен:

$$\prod_{n=1}^N f_{Y_1, X_2, \dots, X_Q}(\hat{y}_n, \hat{x}_{1n}, \dots, \hat{x}_{Qn}; \Theta_*, \Theta_M) = \\ = \ell(\hat{X} | \Theta_*, \Theta_M) \quad (3)$$

Функцию (3), аргументами которой являются неизвестные параметры модели, называют функцией правдоподобия [3], и если распределение вероятностей погрешности неадекватности принадлежит классу экспоненциальных [8], то используют ее логарифм — логарифмическую функцию правдоподобия. Так, в случае двух переменных она может иметь вид

$$L_S(\theta_{s1}, \theta_{s2}; \Theta_M) = N \cdot \ln \frac{S}{2 \cdot \Gamma(1/S)} - N \cdot \\ \cdot \ln \theta_{s2} - \frac{1}{\theta_{s2}^S} \sum_{n=1}^N \left| \hat{y}_n - \theta_0 - \sum_{m=1}^{M-1} \theta_m \cdot \hat{x}_n^m \right|^S, \quad (4)$$

и по ее максимуму находят параметры и наблюдаемую составляющую функции погрешности неадекватности модели объекта, а соответствующий метод параметрической идентификации называют методом максимального правдоподобия (ММП).

Вычислительные схемы ММП при  $S=1$  (распределение Лапласа) и  $S=2$  (распределение Гаусса) называют соответственно методами наименьших модулей (МНМ) и квадратов (МНК). При  $S \neq 2$  в схему МНК вводят весовые коэффициенты, обратно пропорциональные степеням  $(2-S)$  отклонений данных от характеристики положения, и получают оценки итерациями [9].

Широкое распространение для параметрической идентификации получила вычислительная схема МНК при представлении характеристики положения модели в матричной форме

$$Y = R\Theta + \Xi_*, \quad (5)$$

где матрицы параметров и случайных отклонений от характеристики положения с распределением вида «\*», а также

регрессионная матрица соответственно имеют вид

$$\Theta = \begin{bmatrix} \theta_0 \\ \dots \\ \theta_{M-1} \end{bmatrix}, \Xi^* = \begin{bmatrix} \xi_1 \\ \dots \\ \xi_N \end{bmatrix} \text{ и}$$

$$R = \begin{bmatrix} 1 & \Omega_1(\hat{x}_{11}, \dots, \hat{x}_{Q1}) & \dots & \Omega_{M-1}(\hat{x}_{11}, \dots, \hat{x}_{Q1}) \\ \dots & \dots & \dots & \dots \\ 1 & \Omega_1(\hat{x}_{1N}, \dots, \hat{x}_{QN}) & \dots & \Omega_{M-1}(\hat{x}_{1N}, \dots, \hat{x}_{QN}) \end{bmatrix}$$

Минимум суммы квадратов  $\Xi^T \Xi$  случайных отклонений для модели (5) достигается при

$$\frac{\partial(\Xi^T \Xi)}{\partial \Theta} = -2R^T(Y - R\Theta) = 0, \quad (6)$$

причем решение уравнения (6) имеет вид  $\hat{\Theta} = (R^T R)^{-1}(R^T Y)$ . Из-за этой вычислительной простоты и возможности получения в некоторых случаях аналитических оценок МНК относится к наиболее исследованным методам параметрической идентификации [10-11].

Применение МНК предполагает выполнение ряда условий для случайных отклонений  $\Xi$  и их статистической функции распределения  $F_N(\xi)$ :

- I) статистической устойчивости — существования предела  $\lim_{N \rightarrow \infty} F_N(\xi) = F_*(\xi)$ ;
- II) центрированности —  $M\{\xi_n\} = 0$ ,  $n = 1, N$ ;
- III) некоррелированности —  $M\{\xi_i \xi_j\} |_{i \neq j} = 0$ ,  $i, j = 1, N$ ;
- IV) гомоскедастичности —  $M\{\xi_i \xi_j\} |_{i \neq j} = D\{\xi\} = \theta_{G2}^2$ ,  $i, j = 1, N$ ;
- V) гауссовости —  $F_*(\xi) = F_G(\xi)$ ;
- VI) неконфлюэнтности — элементы регрессионной матрицы являются точные числа;
- VII) неограниченности — значения искомых переменных не ограничены;
- VIII) замкнутости —  $\text{rang} R = M$ ;
- IX) адекватности — структура модели является правильной.

Статистический анализ в этих условиях называют классическим линейным регрессионным, и его оценки обладают согласно теореме Гаусса-Маркова рядом привлекательных свойств:

- 1) состоятельности —  $\lim_{N \rightarrow \infty} P\{|\hat{\theta}_m - \theta_m| < \varepsilon\} = 1$ ,  $m = \overline{0, M-1}$ ;
- 2) несмещенности —  $M\{\hat{\theta}_m\} = \theta_m$ ,  $m = \overline{0, M-1}$ ;
- 3)  $M\{\hat{\theta}_k, \hat{\theta}_l\} |_{k \neq l} = a_{kl} \cdot \theta_{2f}^2$ ,  $k, l = \overline{0, M-1}$ , где  $a_{kl}$  — элемент матрицы  $A = (R^T R)^{-1}$ ;
- 4) эффективности,  $D\{\hat{\theta}_m\} = M_{mm} \cdot \theta_{G2}^2 / \det A$ , где  $M_{mm}$  — минор элемента  $a_{mm}$  матрицы  $A$ ;

5) дисперсия оценки характеристики положения  $D\{\hat{\Phi}(\theta, X)\} = \Omega^T A \Omega \theta_{G2}^2$ ;

6)  $M\{\Xi^T \Xi\} = (N-M) \cdot \theta_{G2}^2$ .

Свойства 2), 4) и 6) оценок МНК не зависят от вида распределения случайной составляющей погрешности неадекватности [10], если для него дисперсия существует.

Следует заметить, что в МНМ минимизируют сумму модулей случайных отклонений, что требует численных методов, из-за чего МНМ не получил широкого распространения.

Простейшей реализацией МНК и МНМ является интерполяционная формула Лагранжа

$$y = \sum_{m=1}^M \hat{y}_m \cdot \prod_{i=1, i \neq m}^M \frac{x - \hat{x}_i}{\hat{x}_m - \hat{x}_i} + \Delta_g,$$

которая сразу дает решение задачи в явном виде. Однако при структурно-параметрической идентификации дефект формулы Лагранжа ограничивает ее возможности уже при  $M=2$ .

Пример 1. Найдем при доверительной вероятности  $P=0,90$  МНК оценку  $\bar{\eta}(x)$  характеристики положения случайной функции вида  $y = \theta_0 + \theta_1 \cdot x + \Xi_G$  по данным таблицы 1 [12]. В этом случае

$$\hat{\theta}_0 = \bar{y} - \bar{x} \cdot \hat{\theta}_1 = 3,010,$$

$$\hat{\theta}_1 = (\bar{y} \cdot \bar{x} - \bar{y} \cdot \bar{x}) / (\bar{x}^2 - \bar{x} \cdot \bar{x}) = 1,599,$$

$$\theta_{G2}^2 = (\hat{y}_n - \theta_0 - \theta_1 \cdot \hat{x}_n)^2;$$

$$\bar{\eta}(x) = 4,609 + 1,599 \cdot (x-1).$$

При  $c = -3 < x < 5 = d$  доверительная область характеристики положения

$$|y(x) - \bar{\eta}(x)| < u_{N-2}(P, \lambda) \cdot \frac{\hat{\theta}_{G2}}{\sqrt{N}} \cdot \sqrt{1 + \frac{(x - \bar{x})^2}{\bar{x}^2 - \bar{x}^2}},$$

где  $\hat{\theta}_{G2} = 2,616$ ;  $u_{N-2}(P, \lambda)$  —  $P$ -квантиль для границ доверительной области (см. табл. 4.6а [12]),

$$\lambda = \sqrt{\frac{1}{2} \left[ 1 - \frac{1 + NCD}{\sqrt{(1 + NC^2)(1 + ND^2)}} \right]},$$

$$C = \frac{c - \bar{x}}{\sqrt{\sum_{n=1}^N (\hat{x}_n - \bar{x})^2}},$$

$$D = \frac{d - \bar{x}}{\sqrt{\sum_{n=1}^N (\hat{x}_n - \bar{x})^2}},$$

$$D = -C = 4 / \sqrt{13,86} = 0,2886;$$

$\lambda = 0,9790 u_{18}(0,9; 0,979) = 2,288$  (линейная интерполяция с погрешностью  $< 2 \cdot 10^{-3}$ ).

С доверительной вероятностью 0,90 неизвестная линия регрессии на интервале  $-3 < x < 5$  будет заключена в пределах (сплошные линии на рис. 1)

$$4,61 + 1,60 \cdot (x-1) \pm 5,99 \cdot \sqrt{0,05 + 0,0722 \cdot (x-1)^2}.$$

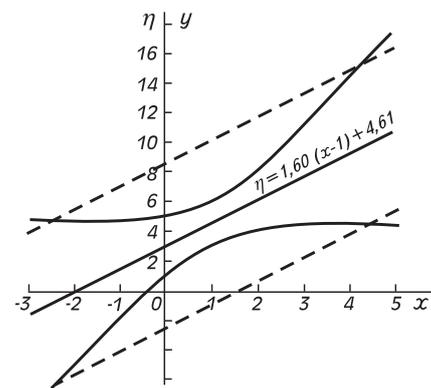


Рис. 1. Доверительные зоны [12, с. 55]

Для доверительной области с прямыми границами по доверительным интервалам на границах области (штриховые линии на рис. 1) используют таблицу 4.6б [12] и  $u_{18}(0,90; 0,9790) \approx 1,891$ :

$$4,61 + 1,60 \cdot (x-1) \pm 1,891 \cdot 2,616 \cdot \sqrt{0,05 + 0,0722 \cdot 4^2}.$$

## 10.2. НАРУШЕНИЕ УСЛОВИЙ ПРИМЕНИМОСТИ ЛИНЕЙНОГО РЕГРЕССИОННОГО АНАЛИЗА

Нарушение условий применимости регрессионного анализа, как и нарушение условий измерений, неизбежно ведет к снижению точности искомого результата [13].

I. Нарушение условия статистической устойчивости ставит под вопрос возможность применения теории вероятностей и математической статистики вообще [10, 14].

II. Нарушение условия центрированности связано с использованием смещенных оценок, не соответствующих с позиций метода максимального правдоподобия виду распределения случайной составляющей погрешности неадекватности модели [2].

III. Нарушение условия некоррелированности приводит к явлению автокорреляции:

- оценки параметров, оставаясь несмещенными, теряют эффективность;
- дисперсии оценок параметров становятся смещенными, что ведет к признанию статистически значимыми входных переменных, которые таковыми могут и не являться;
- дисперсия оценки регрессионной модели во многих случаях занижается;
- нарушаются условия применимости критериев Стьюдента и Фишера.

Продолжение следует

The author continues the consideration of the mathematical theory of measurement. This article is devoted to the method of combined measurements.

### ПРОТОКОЛ ИДЕНТИФИКАЦИИ ХАРАКТЕРИСТИКИ ПОЛОЖЕНИЯ

Таблица 1

n	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
$\hat{X}$	0,0	0,0	0,0	0,0	0,1	0,1	0,1	0,5	0,5	1,0	1,0	1,5	1,5	1,9	1,9	1,9	2,0	2,0	2,0	2,0
$\hat{Y}$	3,32	4,40	2,01	5,07	0,27	3,50	3,42	5,00	0,44	5,42	7,87	5,64	1,27	7,45	7,58	0,00	8,32	10,25	6,56	4,21

Примечание: в таблице исходных данных [12, с. 54] имеет место опечатка, поэтому принято  $x_{12}=1,5$ .

# МАТЕМАТИЧЕСКАЯ ТЕОРИЯ ИЗМЕРИТЕЛЬНЫХ ЗАДАЧ MATHEMATICAL THEORY OF MEASUREMENT PROBLEMS

Левин С.Ф. (S. Levin), д. т. н., профессор, МИЭИ Ростехрегулирования

(Продолжение, начало см. №3-2006)

IV. Нарушение условия гомоскедастичности проявляется как снижение эффективности оценок параметров регрессионной модели и как существенное занижение оценок дисперсий и ковариаций параметров. Анализ этого явления проводят на основе схемы «ε-загрязнения» Дж. Тьюки с долей ε и параметром рассеяния  $\beta\theta_{G2}$  возмущающего распределения [15]:

$$f_{\pm}(\xi) = (1-\varepsilon) \cdot f_G(\theta_{G1}, \theta_{G2}; \xi) + \varepsilon \cdot f_{\pm}(\theta_{G1}, \beta\theta_{G2}; \xi). \quad (7)$$

Применение в этом случае метода взвешенных наименьших квадратов с использованием в представлении функций правдоподобия вместо «долей» весовых коэффициентов, обратно пропорциональных параметрам рассеяния компонентов смеси (7), породило «схему обработки неравноточных измерений» [16], создающей видимость повышения точности результата за счет введения «поправки» в незначимые разряды числового представления оценки [17].

Пример 2. В условиях «ε-загрязнения» стандартизованным распределением Гаусса относительная эффективность оценок параметров положения приведена в таблице 2 [18].

ОТНОСИТЕЛЬНАЯ ЭФФЕКТИВНОСТЬ ОЦЕНОК

Таблица 2

Оценка	ε	Возмущающее распределение вида «*»			
		Гаусса (*=G) β=3	Лапласа (*=L) β=1	Равномерное (*=R) β=3	Коши (*=K), β=1
среднее	0	1	1	1	1
медиана	0,05	0,637	0,637	0,637	0,637
среднее		0,808	0,682	0,362	0,282
медиана	1	0,669	0,894	0,781	0,765
среднее		1	0,5	0,063	0,032
медиана		0,637	1,0	0,835	0,811

В задаче Эддингтона-Фишера для оценок параметра рассеяния смеси распределений в случае β=3 среднее квадратическое отклонение (СКО) при ε=0 требует на 12,4 % меньшего объема выборки, чем среднее абсолютное отклонение (САО) (таблица 3). При ε>0,0018 отношение относительной эффективности меняется в пользу САО, превышая при ε=0,05...0,06 уровень 203 % [19]!

V. Нарушение условия гауссовости является жестким ограничением для применения регрессионного анализа. Оценки МНК теряют свойства 1), 3) и 5), хотя простота получения остается. Но если совокупность случайных отклонений подчиняется распределению Коши, то его теоретическая дисперсия становится бесконечно большой, а выборочная дисперсия как оценка параметра рассеяния становится несостоятельной [10].

VI. Нарушение условия неконфлю-

энтности приводит к возникновению смещения, занижению оценок дисперсии и потере состоятельности оценок параметров модели. Однако конфлюэнтный анализ, математический аппарат которого достаточно подробно описан [20-21], существенно усложняет процедуру оценивания погрешностей получаемых результатов.

VII. Нарушение условия неограниченности ведет к стохастической мультиколлинеарности — корреляции между столбцами регрессионной матрицы [13], которая вызывает:

- неустойчивость оценок (добавление или исключение даже одного отсчета может сильно изменить оценки параметров модели и резко уменьшить точность предсказания по модели);
- численную неустойчивость оценивания из-за погрешностей округления и их накопления;
- сильную корреляцию оценок параметров модели, что лишает смысла их интерпретацию;
- резкое увеличение дисперсии оценок параметров модели;
- ненадежность критериев значимости, т. к. значения СКО входят в их формулы.

VIII. Нарушение условия замкнутости делает задачу вырожденной.

IX. При нарушении условия адекватности неполнота модели ведет к смещению остатков, а избыточность — к завышению дисперсий оценок параметров [13].

Проблемы применимости регрессионного анализа оказались в круге вопросов дискуссии о применимости вероятностно-статистических методов [22-28]. К тому же компьютеризация регрессионного анализа стала, по словам Ю.П. Адлера, путем «вверх по лестнице, ведущей вниз» и низвела его до эвристического метода идентификации моделей, в котором решающую роль играет анализ

остатков и здравый смысл интерпретатора (см. Предисловие к [29]).

Однако развитие анализа данных [29], робастной [19, 30] и непараметрической статистики [31-32] поставило новые вопросы [26, 33]: существует ли более эффективные методы параметрической идентификации, чем метод максимального правдоподобия? что предпочтительнее — робастность или воспроизводимость оценок?

Сомнения в предельной эффективности оценок ММП зародила теорема Гливленко [35]. Ведь аналогичная теорема для распределения относительных частот и плотности распределения вероятностей не была доказана. Эти сомнения усилил критерий Колмогорова, статистика которого  $D_{\Psi\psi}$  как точная верхняя граница модуля разности теоретической  $F_*(\psi)$  и статистической  $F_N(\xi)$  функций распределения [36] согласно лемме о критерии воспроизводимости [37-38, 1] строго связана с вероятностью согласия распределений

$$\alpha_{\Psi\psi} = \int_{-\infty}^{\infty} \inf_{\xi} \{f_{\pm}(\psi), f_{\psi}(\xi)\} d\psi = 1 - D_{\Psi\psi}$$

при единственной точке пересечения графиков плотностей  $f_{\pm}(\xi)$  и  $f_{\psi}(\psi)$ . Позже удалось вычислить статистику критерия воспроизводимости для дискретного распределения FN(o):

$$\hat{\alpha}_* = 1 - \sum_{r=1}^R |D_{\tau(r)}^+ - D_{\tau(r)}^-| \quad (8)$$

где  $D_{\tau(r)}^+ = |\sup F_N(\xi_{\tau(r)}) - F_*(\psi_{\tau(r)})|$  и  $D_{\tau(r)}^- = |\inf F_N(\xi_{\tau(r)}) - F_*(\psi_{\tau(r)})|$  при  $\psi_{\tau(r)} = \xi_{\tau(r)}$  [38].

При  $F_*(\psi_{\tau(r)}) = [\sup F_N(\xi_{\tau(r)}) + \inf F_N(\xi_{\tau(r)})]/2$  абсолютный максимум тождества (8) соответствует абсолютному минимуму САО разностей  $F_*(\psi) - F_N(\xi)$  в точках роста:

$$\hat{\alpha}_* = \frac{1}{2R} \sum_{r=1}^R \left\{ \left| D_{\tau(r)}^- - D_{\tau(r)}^+ \right|, F_*(\psi_{\tau(r)}) \in F_N(\xi_{\tau(r)}) \right. \\ \left. \left| D_{\tau(r)}^+ + D_{\tau(r)}^- \right|, F_*(\psi_{\tau(r)}) \notin F_N(\xi_{\tau(r)}) \right\}. \quad (9)$$

Переход от функций правдоподобия вида (3) к представлению погрешности неадекватности теоретической функции распределения вероятностей как характеристики положения статистической функции распределения привел к методу максимума вероятности согласия [38].

Пример 3 [2]. По данным измерений электрического сопротивления R контрольной меры исходным эталоном локальной поверочной схемы (таблица 4) были получены аналитические оценки ММП и оценки ММС параметров распределений Гаусса, Лапласа и Коши (таблица 5). При упрощенной проверке

АСИМПТОТИЧЕСКАЯ ОТНОСИТЕЛЬНАЯ ЭФФЕКТИВНОСТЬ ОЦЕНОК

Таблица 3

Отношение	Возмущающее распределение Гаусса (G)												
	ε	0	0,001	0,002	0,005	0,01	0,02	0,05	0,10	0,15	0,25	1	
СКО/САО		0,876	0,948	1,016	1,198	1,439	1,752	2,035	1,903	1,689	1,371	1,017	0,876

структурных гипотез для распределений с равным числом параметров ММС обеспечил более высокую вероятность согласия за счет нахождения оценок параметра рассеяния функции распределения вероятностей как характеристики положения численными методами по функции правдоподобия (8) даже в случае распределения Гаусса (рис. 2).

Оценки ММП параметров свертки (1) для гипотезы « $\ll=L$ » [2], Ом: неопределенности  $\theta_{LR0} = 1,3279 \cdot 10^{-4}$ , положения  $\theta_{LR1} = 100,00005$ , рассеяния  $\theta_{LR2} = 4,7245 \cdot 10^{-5}$  Ом. При доверительной вероятности  $P = 0,99$  и вероятности согласия  $\alpha_L = 0,53$  искомое сопротивление  $R = (100,00005 \pm 0,00027)$  Ом.

#### 10.4. СХЕМА ПЕРЕКРЕСТНОГО НАБЛЮДЕНИЯ ПОГРЕШНОСТИ НЕАДЕКВАТНОСТИ

Схема перекрестного наблюдения погрешности неадекватности [25, 27] является обобщением схемы кросс-валидации или «метода складного ножа» [24] для получения несмещенных оценок параметра положения выборки. Согласно этой схеме из выборки поочередно удаляют каждый элемент, оценивают параметр положения, а частные оценки усредняют.

Таблица 4  
**СОПРОТИВЛЕНИЕ КОНТРОЛЬНОЙ МЕРЫ  
НОМИНАЛА 100 Ом**

$r$	$\xi_{[r]} \cdot \text{Ом}$	$\omega_{[r]}$	$F_N^-(\xi_{[r]})$	$F_N^+(\xi_{[r]})$
1	100,00003	1	0	0,1
2	100,00004	3	0,1	0,4
3	100,00005	2	0,4	0,6
4	100,00006	1	0,6	0,7
5	100,00009	3	0,7	1

Схема перекрестного наблюдения для идентификации погрешностей неадекватности моделей объектов измерений играет ту же роль, что и эталонны для идентификации погрешностей средств измерений. Эта схема предусматривает [1]:

- 1) деление пространства переменных на  $(M+1)$  блоков;
- 2) формирование из  $M$  блоков  ${}^{\Pi}X_m$  пробных частей данных измерений с построением соответствующей совокупности характеристик положения рассматриваемой структуры, параметры  ${}^{\Pi}\Theta_m$  которых оценены каким-либо методом по данным этих частей;
- 3) использование в каждой комбинации оставшегося блока данных  ${}^K X_m$  в качестве контрольного [39] и экстраполяции на него соответствующей характеристики положения;



Рис. 2. Параметрическая идентификация функции распределения Гаусса

4) объединение всех экстраполяций в экстраполяционный функционал модели (рис. 3);

5) наблюдаемую составляющую погрешности неадекватности определяют как совокупности значений случайной величины — разности расчетного значения экстраполяционного функционала по рассматриваемой выходной переменной и ее измеренного значения при соответствующих расчету значениях входных переменных

$$\xi_g(\hat{X}) = \hat{Y}_p(\hat{X}) - \sum_{m=1}^{M+1} \Phi_g({}^{\Pi}\Theta_m; \hat{X}) \cdot \Pi({}^K X_m; \hat{X}), \quad (10)$$

где  $\Pi({}^K X_m; X)$  — индикаторная функция [4]  $m$ -й контрольной части данных измерений;

6) ненаблюдаемую составляющую погрешности неадекватности определяют как совокупность погрешностей измерений, данные которых были использованы для построения характеристики положения модели, а также погрешностей статистического оценивания параметров и выбора вида распределения случайной составляющей;

7) окончательные оценки параметров характеристик положения и рассеяния модели получают по всей совокупности данных совместных измерений без деления на блоки.

Общим критерием идентификации распределения случайной составляющей погрешности неадекватности характеристики положения модели является критерий воспроизводимости или максимум вероятности согласия (8) двух распределений отклонений данных совместных измерений — от характеристики положения модели и от

ее экстраполяционного функционала. Критерии выбора характеристики положения модели используют в следующих формах:

а) максимума компактности распределения при минимуме функции компактности [40]

$$K_*(\varepsilon) = 1 - P\{|\Xi - \theta_{*1}| < \varepsilon\} = 1 - [F_*(\theta_{*1} + \varepsilon) - F_*(\theta_{*1} - \varepsilon)] \leq Q(\theta_{*1} - \varepsilon, F_*),$$

где  $Q(I, F)$  — функция концентрации Леви [41];

б) минимума среднего модуля погрешности неадекватности (СМПН)

$$\bar{E}_g(\hat{X}) = \frac{1}{R} \sum_{m=1}^{M+1} \hat{Y}_p({}^K X_m) - \Phi_g({}^{\Pi}\Theta_m; {}^K X_m); \quad (11)$$

в) минимума среднего модуля невязки экстраполяционного функционала

$$\bar{\xi}_g(\hat{X}) = \frac{1}{2R} \sum_{Q, \vartheta} [\sup_m \Phi_g({}^{\Pi}\Theta_m; X) - \inf_m \Phi_g({}^{\Pi}\Theta_m; {}^K X_m)].$$

Метод структурной идентификации моделей объектов измерений по критериям воспроизводимости с использованием схемы перекрестного наблюдения погрешности неадекватности представляет собой метод максимума компактности (ММК) [25, 39, 42, 1].

Обязательным элементом идентификации модели объекта измерений независимо от полноты априорных данных является нахождение функции погрешности неадекватности  $E_p(X)$  по выходной переменной  $Y_p$  как свертки (1) в виде функции входных переменных.

Структурно-параметрическую идентификацию осуществляют только в схеме перекрестного наблюдения перебором вариантов структуры модели максимальной сложности, методов определения параметров и правил сегментации данных с проверкой гипотез вы-

СРАВНИТЕЛЬНАЯ ХАРАКТЕРИСТИКА ОЦЕНОК ММП И ММС

Таблица 5

« $\ll$ »	$L_{ММП}$	$G_{ММП}$	$K_{ММП}$	$L_{ММС}$	$G_{ММС}$	$K_{ММС}$
$\theta_{*1}, \text{Ом}$	100,00005	100,000058	100,0000466	100,00005	100,0000597	100,00005
$\theta_{*2}, \text{Ом}$	$1,8 \cdot 10^{-5}$	$2,227105745 \cdot 10^{-5}$	$1,13358318 \cdot 10^{-5}$	$2,803673252 \cdot 10^{-5}$	$2,92239355 \cdot 10^{-5}$	$1,962610506 \cdot 10^{-5}$
$D_*$	0,2458159884	0,24028	0,2321721315	0,25	0,23002	0,25
$\alpha_*$	<b>0,5346146025</b>	0,36972	0,3127948871	0,6401	0,599958	<b>0,6903888803</b>
$\bar{d}_*$	0,06208249646	0,084794	0,1021634429	0,06499	0,076178	0,06036044267

рожденности  $H_0$ , непрерывности  $H_{00}$  и композиционной однородности  $H_{000}$ .

Гипотеза вырожденности  $H_0 = \{\theta_{0\dots0\dots0} \neq 0; \theta_{i\dots j\dots k} = 0 \forall i\dots j\dots k \neq 0\dots0\dots0\}$  соответствует отсутствию зависимости (постоянный уровень,  $y = \text{const}$ ). Альтернативными являются гипотезы о ненулевых значениях хотя бы одного из параметров  $\theta_{i\dots j\dots k}$  для  $i\dots j\dots k \neq 0\dots0\dots0$ .

Гипотезе непрерывности  $H_{00}$  характеристики положения непрерывной функции на всем интервале наблюдения альтернативными являются гипотезы о кусочно-непрерывной характеристике положения, разделенной на интервале наблюдения точками «разладки», в которых модель характеристики положения претерпевает структурно-параметрические изменения.

Гипотезе композиционной однородности  $H_{000}$  соответствует общая (единая) модель характеристики положения

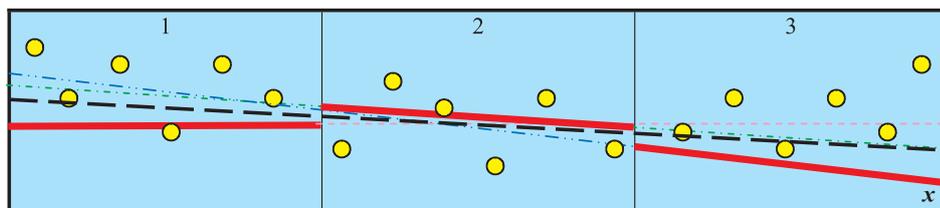


Рис. 3. Экстраполяционный функционал линейной характеристики положения модели

в условиях мультиколлинеарности (таблица 6), когда определитель информационной матрицы  $\sim 1,03 \cdot 10^{-19}$ .

В работе [13] исследовались различные алгоритмы идентификации в условиях мультиколлинеарности — метод Гаусса, ортогонального разложения по Хаузхолдеру, регуляризации по А.Н. Тихонову, гребневой регрессии, «сжатых» оценок Стейна и др., об эффективности которых при решении реальных задач судить трудно, хотя авторами для

— данные протокола измерений выходной переменной модели с кодом структуры 9 [1] центрируют ее экстраполяционным функционалом;

— для последовательностей положительных и отрицательных значений погрешности неадекватности строят характеристику рассеяния и проводят нормализацию отклонений;

— последовательности нормализованных значений погрешности неадекватности модели подвергают обработке вычислительными схемами метода многократных измерений;

— полученное распределение путем обратного преобразования характеристиками рассеяния и положения приводят к характеристике положения рассматриваемой модели.

Пример 5. Рассмотрим применение схемы перекрестного наблюдения для идентификации вида распределения случайной составляющей погрешности неадекватности по данным Примера 3. ММКМНМ для двухпараметрических распределений вероятностей приводит к схеме перекрестного наблюдения погрешности неадекватности с 3 блоками данных (таблица 8).

Оценки параметров гипотетических распределений и соответствующие перекрестные оценки статистик критериев (8) и (9) приведены в таблице 9.

ВЫХОД ХИМИЧЕСКОЙ РЕАКЦИИ У ЗА ВРЕМЯ  $x_1$  ПРИ ТЕМПЕРАТУРЕ  $x_2$

Таблица 6

№ опыта	1	2	3	4	5	6	7	8	9	10
$x_1$ , мин	112	93	115	104	99	120	126	90	97	109
$x_2$ , °C	201	202	203	204	205	206	207	208	209	210
$y$ , %	59,50	81,34	59,57	70,59	76,88	52,46	49,68	85,57	77,59	65,14

для ансамбля данных совместных измерений. Альтернативными являются гипотезы о делении множества реализаций статистического ансамбля данных совместных измерений на подмножества (кластеры), каждому из которых соответствует модель характеристики положения, отличающаяся структурой или (и) значениями параметров.

Структурно-параметрическая идентификация основана на проверке системы нулевых гипотез  $H_0$  и  $H_{00}$ , а также альтернативных гипотез о наличии

большинства перечисленных методов и сделан неутешительный вывод.

Примем в качестве модели максимальной сложности  $y(x_1, x_2) = A + Bx_1 + Cx_2 + Dx_1x_2 + Ex_1^2 + Fx_2^2$  и проведем анализ СМПН ее возможных структур ( $Z = A \vee B \vee C \vee D \vee E \vee F$ ) с помощью системы метрологического сопровождения статических измерительных задач «ММК-стат М» [1, 42] (таблица 7).

Анализ показывает, что датчик псевдослучайных чисел имеет смещение, полная в этих условиях модель обладает

СРЕДНИЙ МОДУЛЬ ПОГРЕШНОСТИ НЕАДЕКВАТНОСТИ МОДЕЛЕЙ

Таблица 7

Z	Структурная гипотеза $H(Z)$ /алгоритм идентификации – ММКМНК/МКМЕДС											
	$H(Z)$		$H(AZ)$		$H(ABZ)$		$H(ABCZ)$		$H(ABCDZ)$		$H(ABCDEZ)$	
	ММКМНК	МКМЕДС	ММКМНК	МКМЕДС	ММКМНК	МКМЕДС	ММКМНК	МКМЕДС	ММКМНК	МКМЕДС	ММКМНК	МКМЕДС
A	<b>14,73</b>	15,75	–	–	–	–	–	–	–	–	–	–
B	23,87	25,63	<b>0,985</b>	6,47	–	–	–	–	–	–	–	–
C	15,07	16,60	21,14	78,40	2,10	3,10	–	–	–	–	–	–
D	24,53	25,63	2,82	42,37	1,91	3,22	1,71	3,28	–	–	–	–
E	33,24	36,04	1,45	5,95	<b>1,25</b>	5,23	1,35	1,17	1,89	5,02	–	–
F	15,89	17,60	21,22	78,41	2,11	3,10	<b>0,804</b>	0,806	2,19	6,15	3,16	29,37

функциональной зависимости и о кусочно-непрерывном характере модели (наличии «разладок» — разрывов 1-го рода).

При использовании данных в виде ансамбля серий измерений проверяют дополнительно гипотезу  $H_{000}$  и альтернативные — о существовании индивидуальных моделей для подмножеств реализаций, данные которых являются статистически однородными.

Пример 4. Рассмотрим параметрическую идентификацию по данным [13] имитационного моделирования регрессионной модели  $\hat{y}(x_1, x_2) = 175 - 0,9x_1 + 0,05x_2 - 0,003x_1x_2 + 0,002x_1^2 + 0,0005x_2^2 + \xi$

существенной структурной составляющей погрешности неадекватности, а лучший по критерию (11) результат дает ММКМНК:  $\hat{y}_{\text{ММКМНК}}(x_1, x_2) = -3941,1650 - 1,0450677x_1 + 39,934086x_2 - 0,096740127x_2^2$ .

Алгоритмы ММК на основе МНК или МНМ обозначают как ММКМНК или ММКМНМ, а при интерполяции по блочным средним и медианам — ММКМНКС и МКМЕДС.

Распределение погрешности неадекватности модели идентифицируют по перекрестному критерию воспроизводимости (алгоритм ММКМП), обеспечивая условия эргодичности:

Оценки статистик (8) и (9) перекрестного критерия максимума статистической воспроизводимости приведены в таблице 8. При этом необходимо иметь в виду, что перекрестные оценки, в отличие от оценок ММС в Примере 3, учитывают структурную составляющую погрешности неадекватности.

*Продолжение следует*

*The author continues the consideration of the mathematical theory of measurement. This article is devoted to the method of combined measurements.*

# МАТЕМАТИЧЕСКАЯ ТЕОРИЯ ИЗМЕРИТЕЛЬНЫХ ЗАДАЧ MATHEMATICAL THEORY OF MEASUREMENT PROBLEMS

Левин С.Ф. (S. Levin), д. т. н., профессор, МИЭИ Ростехрегулирования

(Окончание, начало см. № 3, 4-2006)

Значения статистической функции распределения и перекрестной функции распределения вероятностей Гаусса с параметрами, указанными в таблице 7, при перекрестных оценках вероятности согласия, САО и расстояния Колмогорова приведены в таблице 10.

В этом случае параметры свертки (1), Ом:  $\theta_{GR0}=1,3431 \cdot 10^{-3}$ ,  $\theta_{GR1}=100,0000543$ ,  $\theta_{GR2}=4,7911 \cdot 10^{-5}$ , а соответствующий наикратчайший 99%-толерантный интервал при доверительной вероятности  $P=0,99$  и вероятности согласия  $\alpha_G=0,39$  составляет  $R=(100,00005 \pm 0,00021)$  Ом.

Таким образом, численные методы позволили повысить точность идентификации электрического сопротивления контрольной меры ОмЭС по сравнению с аналитическими оценками ММП на 28 %.

## ЗАКЛЮЧЕНИЕ

Метод совместных измерений является важнейшим методом решения измерительных задач, в т. ч. задач идентификации метрологических характеристик функционального типа,

измерений. Это обстоятельство оказывается решающим при аттестации методик решения практически всех метрологических измерительных задач и существенным образом ограничивает применение нормативных документов при метрологической аттестации методик прикладных измерительных задач, связанных с идентификацией распределений вероятностей и не использующих схему перекрестного наблюдения погрешности неадекватности математической модели объекта измерений.

## ЛИТЕРАТУРА

1. Р 50.2.004-2000 ГСИ. Определение характеристик математических моделей зависимостей между физическими величинами при решении измерительных задач. Основные положения
2. МИ 2916-2005 ГСИ. Идентификация распределений вероятностей при решении измерительных задач
3. Математическая энциклопедия. Т. 5. — М.: «Советская энциклопедия», 1984.
4. Колмогоров А.Н. Основные понятия теории вероятностей. — М.: Наука, 1974.

10. Худсон Д. Статистика для физиков. — М.: Мир, 1967.

11. Линник Ю.В. Метод наименьших квадратов и основы теории обработки наблюдений. — М.: Государственное издательство физико-математической литературы, 1962.

12. Большев Л.Н., Смирнов Н.В. Таблицы математической статистики. — М.: Наука. Главная редакция физико-математической литературы, 1983.

13. Вучков И.Н., Бояджиева Л.Н., Солаков Е.Б. Прикладной линейный регрессионный анализ. — М.: Финансы и статистика, 1987.

14. Гнеденко Б.В. Курс теории вероятностей. — М.: Наука. Главная редакция физико-математической литературы, 1988.

15. Tukey J.W. A survey of sampling from contaminated distributions//Contributions to Prob. and Statist. — Stanford Univ. Press, 1960, p. 448-485.

16. Бурдун Г.Д., Марков Б.Н. Основы метрологии. — М.: Изд-во стандартов, 1985.

17. Левин С.Ф. Легенда о неравноточности. — Партнеры и конкуренты. — 2000. — № 3.

ДАННЫЕ ММКМНМ-ИДЕНТИФИКАЦИИ РАСПРЕДЕЛЕНИЙ ВЕРОЯТНОСТЕЙ

Таблица 8

$r$	$\xi_{[r]}, Ом$	$F_N^{\pm}(\xi_{[r]})$	$\hat{\theta}_{L1}^x, Ом$	$\hat{\theta}_{L2}^x, 10^{-5} Ом$	$\hat{\theta}_{G1}^x, Ом$	$\hat{\theta}_{G2}^x, 10^{-5} Ом$	$\hat{\theta}_{K1}^x, Ом$	$\hat{\theta}_{K2}^x, 10^{-5} Ом$
1	100,00003	0, 0,1	100,00004459	4,46825396				
2	100,00004	0,1, 0,4			100,00005232	1,96825396825	100,0000519	1,53968254
3	100,00005	0,4, 0,6	100,00005531	1,268253968	100,00005448	1,507936508	100,0000563	0,6984126984
4	100,00006	0,6, 0,7	100,00005	0,9285714286				
5	100,00009	0,7, 1			100,00005313	1,71428571	100,000052143	0,71428571426

функций погрешности неадекватности математических моделей объектов измерений, градуировочных характеристик средств измерений, которые составляют теоретическую и экспериментальную основу аттестации программного обеспечения измерительных систем [1-2].

Отличительной особенностью этого метода решения измерительных задач является невозможность определения характеристик погрешностей искомого результата до выполнения совместных

5. Levy P. Calcul des probabilités. — Paris, 1925.

6. Крамер Г. Математические методы статистики. — М.: Мир, 1975.

7. Левин С.Ф. Обеспечение единства измерений при проверке средств измерений. — Измерительная техника. — 2005. — № 8.

8. МИ 83-76 ГСИ. Методика расчета параметров поверочных схем

9. Мудров В.И., Кушко В.Л. Методы обработки измерений. — М.: Радио и связь, 1983.

18. Айвазян С.А., Енюков И.С., Мешалкин Л.Д. Прикладная статистика: Основы моделирования и первичная обработка данных. — М.: Финансы и статистика, 1983.

19. Хьюбер П. Робастность в статистике. — М.: Мир, 1984.

20. Кендалл М. Дж., Стьюарт А. Статистические выводы и связи. — М.: Наука, 1973.

21. Грешиллов А.А. Анализ и синтез стохастических систем. — М.: Радио и связь, 1990.

РЕЗУЛЬТАТЫ ММКМНМ-ИДЕНТИФИКАЦИИ РАСПРЕДЕЛЕНИЙ ВЕРОЯТНОСТЕЙ

Таблица 9

Структура ФРВ	Оценки параметров ФРВ		$\hat{d}_*^x$ МНМ	$\hat{d}_*^x$ ММП	Оценка вероятности согласия, $\hat{\alpha}_*^x$		
	$\hat{\theta}_{*1}, Ом$	$\hat{\theta}_{*2}, Ом$			ММКМНМ	МНМ	ММП
$L_{ММКМНМ}^*$	100,0000532	$1,4426 \cdot 10^{-5}$	0,0697	0,0621	0,0135	0,5625	0,5346
$G_{ММКМНМ}^*$	100,0000543	$1,8254 \cdot 10^{-5}$	<b>0,0640</b>	0,0848	<b>0,3867</b>	0,4444	0,3697
$K_{ММКМНМ}^*$	100,0000531	$9,5238 \cdot 10^{-6}$	0,0687	0,0842	0,3377	0,6110	0,5289

**СТАТИСТИКИ КРИТЕРИЯ ВОСПРОИЗВОДИМОСТИ ДЛЯ РАСПРЕДЕЛЕНИЯ ГАУССА**

Таблица 10

r	$\xi'_{[r]}$	$F_N^-(\xi_{[r]})$	$F_N^+(\xi_{[r]})$	$F_G^x(\xi'_{[r]})$	$D_{Gr}^-$	$D_{Gr}^+$
1	100,00003	0	0,1	0,12862	0,12862	0,02862
2	100,00004	0,1	0,4	0,26602	0,16602	0,13398
3	100,00005	0,4	0,6	0,38269	0,01731	0,21731
4	100,00006	0,6	0,7	0,65634	0,05634	0,04366
5	100,00009	0,7	1	0,98432	0,28432	0,01568

22. Алимов Ю.И. О логическом построении математической статистики как прикладной дисциплины. – Автоматика. — 1971. — № 3.

23. Тутубалин В.Н. Теория вероятностей. — М.: Изд-во МГУ, 1972.

24. Алимов Ю.И. Альтернатива методу математической статистики. — М.: Знание, 1980.

25. Вопросы кибернетики, ВК-94: Статистические методы в теории обеспечения эксплуатации. — М.: АН СССР, Научный совет по комплексной проблеме «Кибернетика», 1982.

26. Эльясберг П.Е. Измерительная информация: сколько ее нужно? как ее обрабатывать? — М.: Наука, 1983.

27. Левин С.Ф., Блинов А.П. Научно-методическое обеспечение гарантированности решения метрологических задач вероятностно-статистическими методами. — Измерительная техника. — 1988. — №12.

28. Статистическая идентификация,

прогнозирование и контроль РЭА. — Севастополь: Знание, 1990.

29. Мостеллер Ф., Тьюки Дж. Анализ данных и регрессия/С предисловием Ю.П. Адлера. — М.: Финансы и статистика, 1982.

30. Хампель Ф. и др. Робастность в статистике: Подход на основе функций влияния. — М.: Мир, 1989.

31. Тарасенко Ф.П. Непараметрическая статистика. — Томск: Изд-во ТГУ, 1976.

32. Холлендер М., Вулф Д.А. Непараметрические методы статистики. — М.: Финансы и статистика, 1983.

33. Статистическая идентификация, прогнозирование и контроль. — Севастополь: Знание, 1991.

34. МИ 1317-2004 ГСИ. Результаты и характеристики погрешности измерений. Формы представления. Способы использования при испытаниях образцов продукции и контроле их параметров

35. Glivenko V.I. Sulla determinazione

empirica delle leggi di probabilita. — Giorn. ist. ital. attuari. — 1933. — V.4. — № 1. — P. 1-10.

36. Kolmogoroff A.N. Sulla determinazione empirica di una legge di distribuzione. — Giorn. ist. ital. attuari. — 1933. — V.4. — № 1. — P. 83-91.

37. Левин С.Ф. Основы теории контроля. — М.: МО СССР, 1983.

38. Левин С.Ф. Идентификация распределений вероятностей. — Измерительная техника. — 2005. — № 2.

39. Ивахненко А.Г. Метод группового учета аргументов — конкурент метода стохастической аппроксимации. — Автоматика. — 1968. — № 3.

40. Левин С.Ф. Теоретические основы метрологии. — М.: ВВИА, 1995.

41. Levy P. Theorie de l`addition des variables aleatoires. — Paris, 1937.

42. Левин С.Ф. Метод максимума компактности и комплексные измерительные задачи. — Измерительная техника. — 1995. — № 7.

43. Левин С.Ф., Лисенков А.Н., Сенко О.В., Харатьян Е.И. Система метрологического сопровождения статических измерительных задач «ММК-стат М». Руководство пользователя. — М.: Госстандарт России, ВЦ РАН, 1998. ☒

*The author continues the consideration of the mathematical theory of measurement. This article is devoted to the method of combined measurements.*