

УДК 006.91.001

М.В. Чеховой, А.О. Матвійчук

Національний технічний університет України «КПІ», Київ

АНАЛІЗ МЕТОДІВ ДОДАННЯ АНОМАЛЬНИХ ВІДЛІКІВ ДО СИГНАЛУ ВИМІРЮВАЛЬНОЇ ІНФОРМАЦІЇ

В статті розглянуто способи побудови тренажерів, що імітують сигнали вимірювальної інформації засобів вимірювальної техніки при налагодженні систем інтернет-калібрування. При цьому враховано необхідність генерації відліків, що відповідають аномальним результатам вимірювання. Моделювання реалізовано шляхом використання двох генераторів псевдовипадкових величин з різними запропонованими параметрами розподілів. Досліджено та порівняно ефективність генерації аномальних результатів вимірювань при різних параметрах розподілень.

Ключові слова: віртуальні тренажери, інтернет-калібрування, імітація сигналів.

Вступ

На даний час використання мережевих та інтернет-технологій в метрології є перспективним та має важливе значення. Зокрема, характеризується стрімким розвитком галузь віддаленого (дистанційного) калібрування [1, 2]. При розробці проектів в даній галузі на першому етапі відбувається моделювання процесу взаємодії всіх учасників інтернет-калібрування (вимірювальних та допоміжних приладів). При цьому використовуються тренажери (симулятори), що відтворюють метрологічні моделі засобів вимірювальної техніки (ЗВТ). В окремих випадках (наприклад, при перевірці алгоритмів визначення аномальних результатів та стійкості програмного забезпечення до завад) постає необхідність використання відповідних тренажерів.

Постановка задачі: дослідження та вдосконалення наявних алгоритмів додання до вибірки відліків, що відповідають аномальним результатам вимірювання.

Наявні алгоритми додання до вибірки аномальних відліків

Найпростіша модель додання аномальних результатів – модель засмічення (модель грубих похибок): розподілення $p_\varepsilon(x)$ результуючої вибірки представляють в вигляді суміші основного розподілення Гауса $\phi(x)$ та іншого розподілення $h(x)$ (розподілення засмічення) [3]:

$$p_\varepsilon(x) = (1 - \varepsilon)\phi(x) + \varepsilon \cdot h(x), \quad (1)$$

де ε – константа, $0 \leq \varepsilon \leq 1$.

Розподілення $h(x)$ може бути розподіленням Гауса зі значно більшою дисперсією, ніж дисперсія $\phi(x)$ (часто при використанні оцінок приймають співвідношення σ_h/σ_ϕ рівним 3 або 10, де σ_h , σ_ϕ – середньоквадратичні відхилення розподілення засмічен-

ня та основного розподілення відповідно). В іншому випадку $h(x)$ відрізняється від розподілення Гауса та має затягнуті „хвости”. Більшість робастних методів розробляють та використовують саме на таких простих моделях, а розширення моделей дає можливість врахувати багато важливих особливостей „некласичних” ситуацій та суттєво наблизити методи оцінки до властивостей реальних даних.

В [4] наведений опис моделі додання аномальних результатів – моделі К-викидів з ідентичним (1) аналітичним представленням, причому $\varepsilon \cdot h(x)$ є достатньо малою складовою $p_\varepsilon(x)$. Це пояснюється тим фактом, що кількість аномальних результатів вимірювання в результуючій вибірці зазвичай на декілька порядків менша за розмір результуючої вибірки.

При моделюванні інформативного параметру ЗВТ з наявністю аномальних результатів було досліджено заміну $n_{\text{зам}}$ елементів початкової вибірки (вибірки нормальних результатів, ВНР розміром $N_{\text{ВНР}}$) на елементи, які відповідають аномальним значенням інформативного параметру. Значення $n_{\text{зам}}$ визначається з вірогідності $P_{\text{АР}}$ появи аномальних результатів вимірювання:

$$n_{\text{зам}} = N_{\text{ВНР}} \cdot P_{\text{АР}}$$

і в реальних випадках $n_{\text{зам}} \ll N_{\text{ВНР}}$. З урахуванням того, що в реальних випадках $n_{\text{зам}}$ складає одиниці, елементи, які відповідають аномальним значенням інформативного параметру виділяються зі спеціально згенерованої вибірки аномальних результатів (ВАР) розміром $N_{\text{ВАР}}$. Тоді результуюча вибірка з наявними аномальними відліками у випадку розташування аномальних результатів в кінцевій її частині буде мати вигляд

$$X_{\text{рез}} = \left[x_1^{\text{HP}}, x_2^{\text{HP}}, \dots, x_{N_{\text{ВНР}} - n_{\text{зам}}}^{\text{HP}}, x_1^{\text{AP}}, x_2^{\text{AP}}, x_{n_{\text{зам}}}^{\text{AP}} \right],$$

де x_i^{HP} - i -й елемент з множини елементів вибірки, які відповідають результатам з нормальними похибками; x_j^{AP} - i -й елемент з множини елементів вибірки, які відповідають результатам з аномальними похибками.

Після генерації вибірки $X_{рез}$ необхідно виконати процедуру декореляції для усунення статистичних взаємозв'язків між елементами результуючої вибірки. Позначимо густину ймовірності розподілу початкової ВНР як $f_{ВНР}(x)$. Множина ВНР' $[x_1^{HP}, x_2^{HP}, \dots, x_{N_{ВНР}-n_{зам}}^{HP}]$ розподілена з густиною ймовірності $f'_{ВНР}(x)$. Різниця $f_{ВНР}(x)$ та $f'_{ВНР}(x)$ обумовлена ненульовим $n_{зам}$ і зменшується зі збільшенням $N_{ВНР}$ або зменшенням $P_{АР}$. Густину ймовірності розподілу ВАР позначимо як $f_{ВАР}(x)$, густину ймовірності $X_{рез}$ позначимо як $f_{рез}(x)$.

Опис розробленого алгоритму додання аномальних результатів

Початкові дані: вибірка $[x_1^{HP}, x_2^{HP}, \dots, x_{N_{ВНР}}^{HP}]$ результатів, $P_{АР}$, параметри густини $f_{ВАР}(x)$ (тип закону розподілення, СКВ ВАР $\sigma_{ВАР}(x)$, граничні значення $f_{ВАР}(x)$ - x_{min}^{AP} та x_{max}^{AP} , граничні значення діапазонів $[x_{min1}^{AP}, x_{max1}^{AP}]$, $[x_{min2}^{AP}, x_{max2}^{AP}]$ в випадку двомодального розподілу густини ймовірності ВАР).

Перелік дій, необхідних для додання аномальних результатів в вибірку, наведений нижче.

1. Згідно початкового параметру $N_{ВНР}$ та $P_{АР}$ необхідно визначити $n_{зам}$.

2. З використанням програмного генератора випадкових чисел (ГВЧ) згенерувати ВАР з дотриманням $f_{ВАР}(x)$, $\sigma_{ВАР}(x)$ та граничних значень ВАР. Необхідно дотримуватись умови $N_{ВАР} > n_{зам}$, для точного дотримання $f_{ВАР}(x)$ рекомендовано встановлювати $N_{ВАР}$ достатньо великим (десятки або сотні).

3. Видалити з ВНР $n_{зам}$ елементів, сформувавши таким чином ВНР'. Порядкові номери елементів, що видаляються, визначити довільно з використанням спеціального ГВЧ.

4. Вилучити з ВАР $n_{зам}$ елементів і долучити їх до ВНР', сформувавши таким чином $X_{рез}$. По-

рядкові номери елементів, що вилучаються, визначити довільно з використанням спеціального ГВЧ.

5. Виконати з використанням спеціального ГВЧ процедуру декореляції $X_{рез}$ для усунення статистичних взаємозв'язків між елементами результуючої вибірки.

Запропоновані різновиди способів додання аномальних результатів

Перелік запропонованих та досліджених різновидів при заміні елементів ВНР значеннями з ВАР:

а) закон розподілення ВАР – нормальний, математичні очікування ВНР та ВАР співпадають;

б) закон розподілення ВАР – рівномірний, математичні очікування ВНР та ВАР співпадають;

в) закон розподілення ВАР – рівномірний, математичні очікування ВНР та ВАР не співпадають (густини розподілу ВНР та ВАР не перекриваються);

г) закон розподілення ВАР – двомодальна композиція двох рівномірних законів, математичні очікування ВНР та ВАР співпадають (густини розподілу ВНР та ВАР не перекриваються).

Результати моделювання представлені в таблицях 1 – 3.

Для перевірки наявності аномальних результатів в кінцевій вибірці були використані критерії Граббса:

1) G_1 та G_n для перевірки на аномальність одного елемента вибірки (окремо найменшого та найбільшого) [5];

2) $G_{1,2}$ та $G_{n-1,n}$ для перевірки на аномальність двох елементів вибірки (окремо найменших та найбільших);

3) $G_{1,2,3}$ та $G_{n-2,n-1,n}$ для перевірки на аномальність трьох елементів вибірки (окремо найменших та найбільших);

4) $G_{1,n}$ для перевірки на аномальність двох елементів вибірки (найменшого та найбільшого).

Порівняння результатів додання до вибірки аномальних відліків графічно представлено на рис. 1, 2.

Особливості програмної реалізації додання до вибірки відліків, які відповідають аномальним результатам

Для генерації ВНР та ВАР був використаний спеціальний ГВЧ заданого розмаху та форми закону розподілення. Для визначення порядкових номерів $n_{зам}$ елементів, що необхідно видалити з ВНР та елементів, що необхідно вилучити з ВАР, були використані стандартні програмні ГВЧ з рівномірним законом розподілення вихідної величини з наступним її нормуванням в межах $N_{ВНР}$ та $N_{ВАР}$ відповідно.

Таблиця 1

Результати додання до вибірки відліків,
які відповідають аномальним похибкам з використанням способу а (для ансамблю з 50-ти реалізацій).
Діапазон значень ВНР – [-3..3], СКВ ВНР = 1. Діапазон значень ВАР – [-10..10], $\sigma_{\text{ВАР}}(x) = 3$

N _{ВНР}	n _{зам}	Кількість реалізацій (з 50-ти) з визначеними аномальними результатами згідно критеріїв						
		G ₁	G _n	G _{1,2}	G _{n-1,n}	G _{1,2,3}	G _{n-2,n-1,n}	G _{1,n}
22	1	9	9	6	8	5	7	11
	3	15	19	13	17	12	14	25
	5	16	18	19	14	17	13	29
52	1	8	8	8	8	7	8	15
	3	22	16	20	17	20	16	30
	5	26	24	22	22	21	18	42
126	1	6	4	6	4	5	3	9
	3	16	18	16	17	14	17	25
	5	20	30	17	29	16	26	41

Таблиця 2

Результати додання до вибірки відліків,
які відповідають аномальним похибкам з використанням способу в (для ансамблю з 50-ти реалізацій).
Діапазон значень ВНР – [-3..3], СКВ ВНР = 1. Діапазон значень ВАР – [-10..-3]

N _{ВНР}	n _{зам}	Кількість реалізацій (з 50-ти) з визначеними аномальними результатами згідно критеріїв						
		G ₁	G _n	G _{1,2}	G _{n-1,n}	G _{1,2,3}	G _{n-2,n-1,n}	G _{1,n}
22	1	48	0	48	0	48	0	48
	3	45	0	50	0	50	0	16
	5	12	0	46	0	50	0	4
52	1	46	0	46	0	45	0	45
	3	50	0	50	0	50	0	50
	5	50	0	50	0	50	0	36
126	1	47	0	44	0	44	0	44
	3	50	0	50	0	50	0	50
	5	50	0	50	0	50	0	49

Таблиця 3

Результати додання до вибірки відліків,
які відповідають аномальним похибкам з використанням способу г (для ансамблю з 50-ти реалізацій).
Діапазон значень ВНР – [-3..3], СКВ ВНР = 1. Діапазони значень ВАР – [-10..-3], [3..10]

N _{ВНР}	n _{зам}	Кількість реалізацій (з 50-ти) з визначеними аномальними результатами згідно критеріїв						
		G ₁	G _n	G _{1,2}	G _{n-1,n}	G _{1,2,3}	G _{n-2,n-1,n}	G _{1,n}
22	1	24	26	24	26	24	26	49
	3	28	31	22	28	21	24	38
	5	19	21	25	29	21	16	36
52	1	20	30	20	30	19	30	47
	3	37	37	33	36	30	35	50
	5	37	41	34	39	32	35	49
126	1	31	17	30	17	30	17	47
	3	42	36	41	33	37	31	50
	5	46	44	44	44	43	43	50

Також було використано окремий ГВЧ для розрахунку оновлених індексів (рангів) елементів $X_{\text{рез}}$ при виконанні декореляції результуючої вибірки. При функціонуванні ГВЧ необхідно забезпечити генерацію ним всіх необхідних рангів з заданого діапазону (обсягу вибірки) та унікальність кожного згенерованого рангу.

Для автоматизації використання критеріїв Граббса було реалізовано спеціальну базу критичних значень $G_{\text{кр}}$ з подальшим використанням інтерполяції сплайнами залежностей $G_{\text{кр}}(n, \alpha)$ при автоматичному визначенні $G_{\text{кр}}$ для заданого n та рівня довіри α .

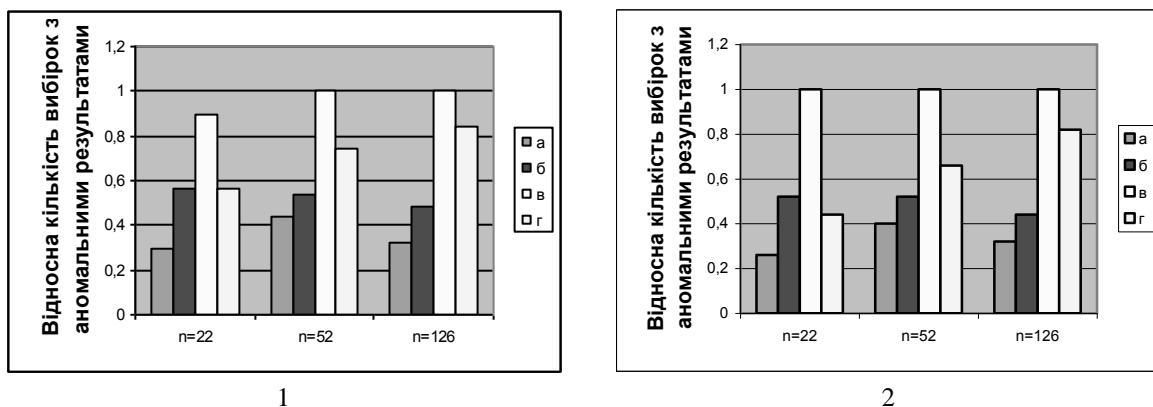


Рис. 1. Порівняння результатів додавання до вибірки аномальних відліків згідно критерію G_1 (1) та $G_{1,2}$ (2): $n_{зам} = 3$ при діапазоні значень ВНР [-3..3], СКВ ВНР = 1. а –діапазон значень ВАР – [-10..10], $\sigma_{ВАР}(x) = 3$; б – діапазон значень ВАР – [-10..10]; в –діапазони значень ВАР – [-10..-3]; г –діапазони значень ВАР – [-10..-3], [3..10]

Висновки

Було виконано моделювання роботи тренажера, що генерує вибірку даних вимірювальної інформації з аномальними результатами. Результуючу вибірку було згенеровано як суміш двох вибірок з різними видами розподілення та середньоквадратичними відхиленнями. Проведено дослідження з різними типами розподілення, середньоквадратичними відхиленнями та математичними очікуваннями.

Розроблений алгоритм зберігає об'єм результуючої вибірки (зміненої шляхом додавання аномальних значень вибірки нормальних результатів) при зміні рівня засміченості. Вперше було проведено дослідження залежності відносної кількості реалізацій, які містять аномальні результати, від форми та середньоквадратичного відхилення закону розподілення вибірки аномальних результатів.

Як критерій якості роботи тренажера для генерації відповідної вибірки з аномальними результатами було обрано відносну кількість вибірок з аномальними результатами до загальної кількості вибірок ансамблю. Найбільша ефективність алгоритму досягається у випадку використання двомодального

або одномодального зміщеного законів розподілення вибірки аномальних результатів.

Список літератури

1. Dudley R.A. Traceability via Internet for Microwave Measurement using Vector Network Analyzers / R.A. Dudley, N.M. Ridler // IEEE Trans. Instrum.Meas. – 2003. – Vol 52, № 1. – P. 130-134.
2. Величко О.М. Калібрівка засобів вимірювальної техніки через internet: стан і перспективи впровадження / О.М. Величко // Український метрологічний журнал. – 2006. – №1. – С. 15-19.
3. Грановский В.А. Методы обработки экспериментальных данных при измерениях / В.А. Грановский, Т.Н. Сирая. – Л.: Энергоатомиздат, 1990 – 288 С.
4. Давид Х.А. Помехоустойчивые оценки при наличии выбросов / Х.А. Давид; пер. с англ. Ю.И. Малахова; под ред. Н.Г. Волкова – М.: Машиностроение, 1984. – 320 с.
5. Grubbs F.E. Extension of sample sizes and percentage points for significance tests of outlying observations / F.E. Grubbs, G. Beck // Technometrics, 1972. – Vol. 14, No. 4. – P. 847-854.

Надійшла до редколегії 3.10.2012

Рецензент: д-р мед. наук, проф. В.П. Яценко, Національний технічний університет України «КПІ», Київ.

АНАЛИЗ МЕТОДОВ ДОБАВЛЕНИЯ АНОМАЛЬНЫХ ОТСЧЕТОВ В СИГНАЛ ИЗМЕРИТЕЛЬНОЙ ИНФОРМАЦИИ

Н.В. Чеховой, А.О. Матвийчук

В статье рассмотрены способы построения тренажеров, имитирующих сигналы измерительной информации средств измерений при наладке систем интернет-калибровки. При этом учтена необходимость генерации отсчетов, соответствующих аномальным результатам измерения. Моделирование реализовано путем использования двух генераторов псевдослучайных величин с различными предложенными параметрами распределений. Исследовано и выполнено сравнение эффективности генерации аномальных результатов измерений при различных параметрах распределений.

Ключевые слова: виртуальные тренажеры, интернет-калибровка, имитация сигналов.

ANALYSIS OF WAYS TO ADD ANOMALOUS RESULTS INTO MEASUREMENT INFORMATION SIGNAL

M.V. Chekhovoi, A.O. Matviychuk

In this article the ways of building simulators that mimic the measurement information signals of measuring instruments for internet-enabled calibration systems adjustment were observed. The need to generate samples corresponding anomalous readings was taken into account. Simulation was realized by using two generators of pseudorandom values with different parameters of proposed distributions. The efficiencies of anomalous measurement results generation with different parameters of distributions was investigated and compared.

Keywords: virtual simulators, Internet-enabled calibration, signal simulation.