

## Література

1. Скляр Б. Цифровая связь. Теоретические основы и практическое применение / Б. Скляр. – Издательский дом «Вильямс», 2007, 1104 с.
2. Кичак В.М. Оцінка впливу кількісних характеристик зміни інформаційного параметру на завадостійкість каналів зв'язку з КАМн / В.М. Кичак, В.С. Белов, А.С. Белов. Науковий журнал «Вісник Хмельницького національного університету». – 2012. - №4. - с. 59-62
3. Кичак В.М. Вимірювання параметрів фазоманіпульованих сигналів при частотному мультиплексуванні / В.М. Кичак, В.С. Белов, А.С. Белов, Б.С. Савчук. XI міжнародна конференція Контроль і управління в складних системах (КУСС-2012): Тези доповідей, м.Вінниця, 9-11 жовтня 2012 р.– Вінниця:ВНТУ.–282 с.
4. Кичак В.М. Визначення бітових спотворень в каналах з прямою корекцією помилок / В.М. Кичак, В.С. Белов, А.С. Белов. Міжнародний науково-технічний журнал «Вимірювальна та обчислювальна техніка в технологічних процесах». – 2012. - №1.- с. 121-124
5. Морелос-Сарагоса Р. Искусство помехоустойчивого кодирования. Методы, алгоритмы, применение: учебное пособие для вузов: Пер. с англ. В.Б. Афанасьева. – М.: Техносфера, 2006. – 319 с.

## References

1. B. Skljär. Cifrovaja svjaz'. Teoreticheskie osnovy i prakticheskoe primenenie. Izdatel'skij dom «Vil'jams», 2007, 1104 s.
2. V.M. Kychak, V.S. Belov, A.S. Belov. Otsinka vplyvu kil'kisnykh kharakterystyk zminy informatsiynoho parametru na zavadostiykist' kanaliv zv'yazku z KAMn. Naukovyy zhurnal «Visnyk Khmel'nyts'koho natsional'noho universytetu». – 2012. - №4.- s. 59-62
3. V.M. Kychak, V.S. Belov, A.S. Belov, B.S. Savchuk. Vymiryuvannya parametriv fazomanipul'ovanykh syhnaliv pry chastotnomu mul'typleksuvanni. Khl mizhnarodna konferentsiya Kontrol' i upravlinnya v skladnykh systemakh (KUSS-2012): Tezy dopovidey, m.Vinnytsya, 9 - 11 zhovtnya 2012 r. – Vinnytsya: VNTU. – 282 s.
4. V.M. Kychak, V.S. Belov, A.S. Belov. Vyznachennya bitovykh spotvoren' v kanalakh z pryamoju korektsiyeyu pomylok // Virjuval'na ta obchisljuval'na tehnika v tehnologichnih procesah.-2012.-№1.- S 121-124
5. Morelos-Saragosa R. Iskustvo pomehoustojchivogo kodirovaniya. Metody, algoritmy, primenenie: uchebnoe posobie dlja vuzov: Per. s angl. V.B. Afanas'eva. – M.: Tehnosfera, 2006. – 319 s.

Рецензія/Peer review : 8.7.2013 р. Надрукована/Printed :17.10.2013 р.

Рецензент:

УДК 519.25

А.В. ГОРОШКО, В.П. РОЙЗМАН

Хмельницький національний університет

## ПРО ЗАСТОСУВАННЯ ГАУСОВОЇ СУМІШІ ДЛЯ СТАТИСТИЧНОЇ ОБРОБКИ ЕМПІРИЧНИХ ДАНИХ

Показано необхідність пошуку методів обробки даних, що підкоряються багатомодальним законам розподілу імовірностей. Запропоновано метод обробки таких даних, що базується на розщепленні гаусових сумішей. Розроблено метод побудови гістограм з обґрунтованим вибором кроку розбиття. Представлені результати застосування розроблених методів для визначення обґрунтованих значень допустимих дисбалансів роторів компресорів авіадвигунів і руйнуючих зусиль керамічних резисторів.

Ключові слова: статистична обробка, густина розподілу, багатомодальний розподіл, гаусова суміш, гістограма, дисбаланс

A.V. GOROSHKO, V.P. ROYZMAN

Khmelnitsky National University

### THE USE OF GAUSSIAN MIXTURE MODELS FOR STATISTICAL PROCESSING OF EMPIRICAL DATA

The necessity of finding methods of data processing, multimodal obeying the laws of probability distributions. We propose a method for processing such data, based on the cleavage of Gaussian mixture. A method for constructing histograms with reasonable step selection of the partition. The results of application of the developed methods for the determination of reasonable values for the valid Rotor aero-engine compressors and destructive forces of ceramic resistors.

Keywords: statistical analysis, density distribution, multimodal probability, Gaussian mixture, histogram, imbalance

## Вступ

Проблему дослідження законів розподілу доводиться вирішувати при ідентифікації технологічних процесів, розробці нормативної документації, контролі якості продукції, що випускається, прогнозуванні ресурсу виробів, і в низці інших задач забезпечення якості виробів, де значення контрольованих величин визначають, випробовуючи дослідні зразки з наступною обробкою експериментальних матеріалів методами математичної статистики.

Проведений авторами аналіз показав, що існуючі підходи та методи вирішення задачі обробки статистичних даних, які в загальному випадку не підкоряються унімодальним законам розподілу, мають суттєві недоліки, що обмежує їх застосування в задачах обробки емпіричних даних. Авторами запропонований метод обробки таких даних [1], що базується на представленні багатомодальних законів у вигляді суміші нормальних законів розподілу імовірностей. Поряд з вирішенням проблеми розділення сумішей розподілів випадкових величин потребують вирішення й інші завдання, зокрема створення методу побудови гістограм таких розподілів.

### Постановка задачі

Проведені авторами вивчення гістограм, побудованих за результатами вимірювань дисбалансів досить великої кількості однотипних роторів авіадвигунів після виготовлення або після експлуатації, показали, що густина розподілу ймовірностей (надалі ГР) добре наближається одновершинною кривою лише після виготовлення і складання в ідентичних умовах, а експлуатаційних - в однакових умовах експлуатації. Дисбаланси ж однотипних роторів, викликані, наприклад, різними виробничими чи експлуатаційними причинами, мають чітко виражені багатопікові гістограми (рис. 1). Причини багатомодальності розподілу виміряних дисбалансів у наведеному прикладі детально викладені у [2]. Однорідна у вихідному стані вибірка роторів в процесі експлуатації розпадається на декілька підвибірок, кожна із яких об'єднана типом домінуючої причини, яка викликає появу експлуатаційного дисбалансу, і, потрапляючи на завод, наприклад, для міжресурсного ремонту, ці ротори утворюють партії з багатомодальними законами розподілу дисбалансів.

Аналогічно, проведене авторами вивчення гістограм, побудованих за результатами вимірювань руйнуючого зусилля деякої кількості однотипних резисторів, показало, що закон розподілу не унімодальний і має чітко виражену двовершинну гістограму (рис.2). Тільки вивчивши процес виробництва цих резисторів, стало можливим пояснити причини появи багатомодальності [2]. Виявилось, що резистори одного і того ж типу виготовляються на заводі на двох однотипних лініях, кожна з яких має специфічні похибки виготовлення. Вироблені на всіх лініях деталі сортують за радіотехнічними ознаками. При цьому в одну партію резисторів, відібраних за однаковими радіотехнічними параметрами, потрапляють деталі, виготовлені на різних лініях, і вони при механічних випробуваннях утворюють стільки однотипних за механічними властивостями груп, скільки різноманітних ліній брали участь у їх виготовленні. Дослідження резисторів з однієї партії показали чітко виражений одномодальний розподіл.

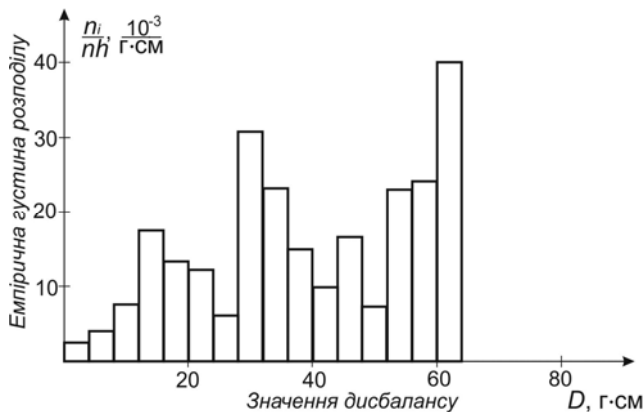


Рис. 1. Розподіл дисбалансів роторів компресорів при низькочастотному балансуванні для лівої опори

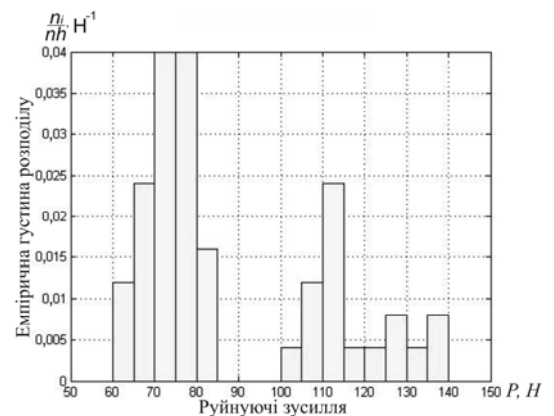


Рис. 2. Розподіл руйнуючих зусиль кераміки резисторів ОМЛТ-0,125

Суть запропонованого авторами методу обробки емпіричних даних, що не підкоряються унімодальним законам розподілу, полягає у представленні емпіричної ГР в вигляді дискретної функції ГР імовірностей суміші  $N$  нормальних розподілів з дисперсіями  $\sigma_i^2$  і математичними сподіваннями  $M_i$ ,  $i = 1, 2, \dots, N$ ,  $2 \leq N < \infty$ ,

$$f(x, M_i, \sigma_i, \rho_i) = (2\pi)^{-\frac{1}{2}} \sum_{i=1}^N \sigma_i^{-1} \rho_i \exp\left[-(x - M_i)^2 (2\sigma_i^2)^{-1}\right], \quad (1)$$

де  $x \in \mathbb{R}$ ,  $\mathbb{R} = (-\infty, +\infty)$ ,  $\rho_i$  - апіорна імовірність  $i$ -ї компоненти суміші,  $\rho_i \in (0, 1)$ ,  $\sum_{i=1}^N \rho_i = 1$ .

Фізична суть багатомодальності і невиконання центральної граничної теореми тут полягає у існуванні  $N$  домінуючих причин, які формують емпіричну функцію  $f(x, M_i, \sigma_i, \rho_i)$ .

Для подальшої обробки результатів експерименту, перш за все, необхідно визначити невідомі параметри, шукаючи їх з умови збігу значень функції (1) у деяких точках (наприклад вершинах і западинах) зі значеннями апроксимуючої функції, графік якої плавною кривою огинає побудовану гістограму. Для пошуку  $3N - 1$  невідомих  $M_i$ ,  $\sigma_i$  і  $\rho_i$  можна використати інтерполяцію на деякій точковій множині,

метод найменших квадратів (МНК) або метод максимальної правдоподібності. В будь-якому разі апроксимація буде тим кращою, чим менші ділянки розбиття при побудові гістограми, тобто чим точніше гістограма і плавна крива, що її огинає. Через цей факт виникає питання, яким повинен бути крок розбиття  $h$  при побудові гістограм. При малому  $h$  гістограма стає невиразною, а при великому дослідник може не виявити моди, що впливають на точність запису (1). Для вибору оптимального кроку авторами запропонований наступний ітераційний метод.

Крок має бути мінімальним, але не меншим за точність вимірювання параметра, а число  $N$  має дорівнювати кількості отриманих при побудові вершин. Далі одним із запропонованих раніше методів необхідно визначити невідомі параметри  $M_i$ ,  $\sigma_i$  і  $\rho_i$ . Якщо в результаті розрахунків один або декілька вагових коефіцієнтів  $\rho_i$  виявляться менше деякої наперед заданої величини  $\beta$ , то відповідними членами у лінійній комбінації (1) можна знехтувати. Дійсно, інтегральна функція розподілу з ГР (1) має вигляд

$$F(x) = (2\pi)^{-\frac{1}{2}} \sum_{i=1}^N \sigma_i^{-1} \rho_i \int_{-\infty}^x \exp\left[-(x - M_i)^2 (2\sigma_i^2)^{-1}\right] dx. \quad (2)$$

Нехай, наприклад,  $\rho_1 < \beta$ . Тоді після відкидання першого доданку в лінійній комбінації (2) нова інтегральна функція розподілу може бути записана у вигляді

$$\bar{F}(x) = (2\pi)^{-\frac{1}{2}} \sum_{i=2}^N \sigma_i^{-1} \rho_i \int_{-\infty}^x \exp\left[-(x - M_i)^2 (2\sigma_i^2)^{-1}\right] dx.$$

Оцінимо різницю

$$\begin{aligned} |F(x) - \bar{F}(x)| &= (\sigma_1 \sqrt{2\pi})^{-1} \rho_1 \int_{-\infty}^x \exp\left[-(x - M_1)^2 (2\sigma_1^2)^{-1}\right] dx \leq \\ &\leq (\sigma_1 \sqrt{2\pi})^{-1} \rho_1 \int_{-\infty}^{\infty} \exp\left[-(x - M_1)^2 (2\sigma_1^2)^{-1}\right] dx \leq \beta, \end{aligned}$$

причому, отримана оцінка справедлива для довільного  $x$ .

Наприклад, якщо функція розподілу імовірностей повинна вимірюватись з точністю до 0,01, то значення  $\beta$  достатньо взяти 0,005. Далі крок гістограми можна збільшувати до тих пір, поки кількість вершин не стане дорівнювати кількості членів в лінійній комбінації (2) після відкидання її малих членів. Знову застосовуючи той же метод розв'язку, але вже для меншої кількості невідомих, можна визначити їх уточнене значення і відкинути малі члени. Такий процес слід продовжувати до тих пір, поки всі  $\rho_i$  не стануть порівнювані з вибраною точністю  $\beta$ . Отриманий при цьому крок може бути взятий за оптимальний. Фізично цей процес означає, що підвибірки з малим  $\rho_i$  вносять вельми незначний внесок у загальну вибірку і тому їх можна об'єднати з однією із підвбірок виробів з близькими величинами досліджуваного параметра.

Одержання закону розподілу імовірності досліджуваного параметра у вигляді (2) дозволяє перейти до вирішення однієї із важливих практичних задач - призначення допустимого значення цього параметра з певною надійністю.

У разі нормального закону розподілу параметра, із довірчою імовірністю  $\gamma$  його допустиме значення встановлюється на основі отриманих його реалізацій як

$$x_{cp} - t_\gamma \cdot \sigma / \sqrt{n} \leq x \leq x_{cp} + t_\gamma \cdot \sigma / \sqrt{n}, \quad (3)$$

де  $x_{cp}$  - вибіркове середнє арифметичне,  $t_\gamma$  - квантиль розподілу Ст'юдента для заданої довірчої

імовірності  $\gamma$  і числа ступенів вільності  $n-1$ ,  $\sigma = \sqrt{(n-1)^{-1} \cdot \sum_{j=1}^n (x_{cp} - x_j)^2}$  - вибіркове середнє квадратичне відхилення,  $n$  - обсяг вибірки.

При ГР результатів спостережень, відмінній від нормальної, слід застосовувати непараметричний підхід, згідно чого у (3) замість квантилі розподілу Ст'юдента слід підставляти число  $U(p)$ , задане рівністю  $\Phi(U(p)) = (1+p)/2$ , де  $\Phi(x)$  - функція стандартного нормального розподілу з математичним сподіванням 0 і дисперсією 1. При цьому отримані довірчі інтервали будуть дещо вужчими.

У разі, якщо отримана гістограма описується багатомодальним законом розподілу, подальші дії з призначення допустимого значення досліджуваного параметра можуть здійснюватись двома шляхами.

1. Розглядається підвбірка з мінімальним (максимальним) значенням  $M_i$ . Очевидно, що допустиме значення параметра виробів цієї підгрупи мінімально (максимально). Призначене допустиме значення параметра для таких виробів може бути прийнято і для всієї партії. У цьому випадку подальша обробка експериментальних даних може відбуватися тільки для зазначеної нормально розподіленої підвбірки значень з параметрами розподілу  $M_i, \sigma_i$ , як описано вище.

Якщо є можливість розділити вихідну вибірку виробів на підвибірки, об'єднані однією з домінуючих причин появи розкиду значень, то аналогічні операції з обробки експериментальних даних слід проводити для кожної підвибірки.

2. Визначені параметри дозволяють записати інтегральну функцію розподілу (2), яку, як і Гаусову випадкову величину, за допомогою комп'ютера можна задати таблицею наступним чином. Для кожного значення величини  $x$ , яке змінюється з певним числовим інтервалом, наприклад, 0,1, за таблицею функції

розподілу нормованого нормального розподілу  $\Phi^x(x) = (2\pi)^{-\frac{1}{2}} \int_{-\infty}^x \exp(-x^2/2) dx$  можна визначити

імовірність  $\gamma_i = (\sigma_i \sqrt{2\pi})^{-1} \int_{-\infty}^x \exp\left[-(x - M_i)^2 (2\sigma_i^2)^{-1}\right] dx = (2\pi)^{-\frac{1}{2}} \int_{-\infty}^{(x - M_i)/\sigma_i} \exp(-x^2/2) dx$  для всіх

$i = 1, 2, \dots, N$  і далі значення інтегральної функції

$$F^x(x) = \sum_{i=1}^N \rho_i \gamma_i. \quad (4)$$

Необхідно відзначити, що, по-перше, другий шлях більш точний, оскільки він враховує функції розподілу всіх підвбірок, а по-друге, більш універсальний, адже з його допомогою можна вирішувати поставлену задачу у випадку довільного розподілу, якщо попередньо скласти для нього таблицю залежності довірчої ймовірності і аргументу інтегральної функції розподілу досліджуваної величини. Також необхідно відзначити, що другий спосіб призначення допусків при багатомодальному розподілі параметра поширюється як частинний випадок і на унімодальний закон. Більше того, призначення допуску за допомогою інтегральної функції розподілу в цьому окремому випадку може слугувати навіть доповненням і уточненням способу розв'язання аналогічної задачі при унімодальних законах розподілу параметра, описаного раніше в припущенні, що істинне значення вимірюваної величини збігається з її математичним очікуванням.

Далі наведені приклади застосування розроблених методів.

## Результати досліджень

### 1. Визначення дисбалансів роторів компресорів авіадвигунів.

Ефективність розроблених методів була перевірена при визначенні початкових дисбалансів роторів компресорів. Розрахунок відбувався з використанням пакету MATLAB. Для відшукування параметрів доданків закону розподілу дисбалансів, представленого на рис. 1, був застосований МНК. Результати розрахунків параметрів функції (1) наступні:  $M_1=15,13$ ,  $M_2=33,77$ ,  $M_3=58,49$ ,  $\sigma_1=5,12$ ,  $\sigma_2=6,98$ ,  $\sigma_3=3,54$ ,  $\rho_1=0,18$ ,  $\rho_2=0,51$ ,  $\rho_3=0,31$ , ГР суміші і одержаних складових представлені на рис. 3.

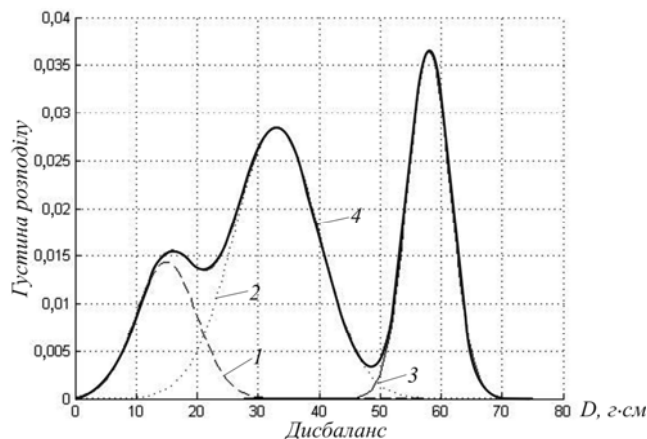


Рис. 3 - ГР суміші (4) і її шуканих складових:

1 - ГР з параметрами  $M_1, \sigma_1$ ;

2 - ГР з параметрами  $M_2, \sigma_2$ ;

3 - ГР з параметрами  $M_3, \sigma_3$ ;

вибірці дисбалансу, що не перевищує 64,9 г·см, дорівнює 0,99. У даній ситуації доцільно вибрати в якості допустимого значення дисбалансу 64,9 г·см, тоді можлива похибка піде в запас міцності виробу, оскільки вироби з більшим допуском, а саме 64,9 г·см, будуть володіти більшою міцністю.

Був також проведений наступний експеримент. Той із вагових коефіцієнтів, який виявився істотно менше двох інших, вважали рівним нулю, і інтегральний закон розподілу вважали двомодальним. Для цих випадків знову визначали відповідні шість параметрів  $M_i, \sigma_i, \rho_i, i = 1, 2$  і знову розраховували таблиці інтегральної функції розподілу.

Порівняльний аналіз таблиць, розрахованих вищевказаним чином, для трьох-і двомодальних законів

розподілу показав, що вони відрізняються неістотно. Це дає можливість при певному аналізі на комп'ютері автоматично знаходити не тільки практично достатню кількість мод закону розподілу для будь-якого кроку  $h$ , обраного при побудові гістограм, а й уточнювати значення самого кроку.

## 2. Визначення руйнуючих зусиль для керамічних резисторів

З метою встановлення граничних напружень, що руйнують керамічні основи резисторів, були проведені вимірювання розтягуючих зусиль, розподіл яких представлений на рис. 2. Для розщеплення суміші був застосований МНК. Результати розрахунків параметрів функції (1) наступні:  $M_1=74$ ,  $M_2=114$ ,  $\sigma_1=6,6$ ,  $\sigma_2=5,7$ ,  $\rho_1=0,66$ ,  $\rho_2=0,34$ , густина розподілу суміші представлена на рис. 4.

Для подальшої обробки розглянемо підвибірку резисторів, руйнуючі зусилля яких підпорядковуються нормальному закону розподілу з параметрами  $M_1=74$ Н,  $\sigma_1=6,6$ Н, тобто найменш

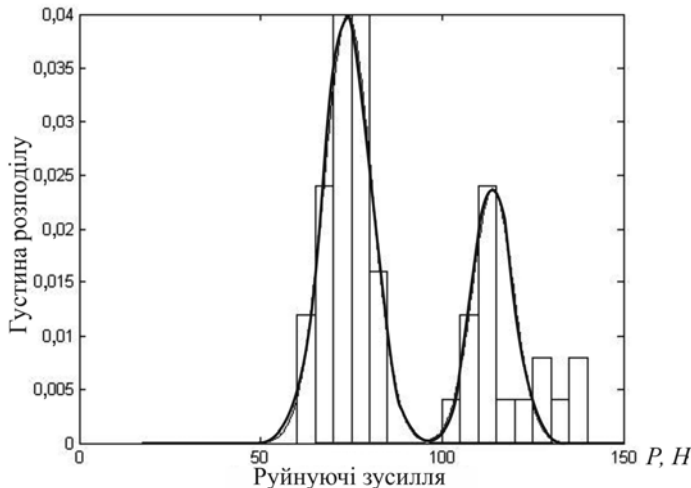


Рис. 4 - Графік шуканої суміші ГР

міцних резисторів, тоді можлива помилка для всієї вибірки піде в запас міцності.

Нехай імовірність того, що вказані зусилля лежать на симетричній відносно  $M_1$  ділянці довжиною  $2\delta$  дорівнює  $\gamma=0,995$ , тобто  $P\{|P_{pvi} - M_1| < \delta\} = \gamma$ . Тоді відношення  $\delta/S_1$  є аргументом функції Лапласа, що дорівнює  $\gamma/2$ . За таблицею знаходимо  $\gamma/2=2,8$ , звідки  $\delta=6,6*2,8=18,48$  Н. Це означає, що з імовірністю 0,995 за граничне значення руйнуючого навантаження можна прийняти  $P_{sp}=74-18,48=55,52$ Н.

## Висновки

Запропонований метод статистичної обробки даних експериментів із вимірними параметрами технічних об'єктів, властивостей і т.п. дозволяє, по-перше, розкрити внутрішню структуру даних з урахуванням можливої багатомодальності законів їх розподілу, і, по-друге, дає правила роботи з такими статистичними матеріалами, зокрема, методи визначення обґрунтованих допустимих значень досліджуваних параметрів.

Розроблений ітераційний метод побудови гістограм багатомодальних в загальному випадку розподілів імовірностей дає дозволяє обґрунтовано вибрати оптимальний крок їх побудови.

За допомогою розроблених методів статистичної обробки даних визначено обґрунтовані значення допустимих дисбалансів роторів компресорів авіадвигунів і руйнуючих зусиль керамічних резисторів.

## Література

1. Горошко А.В. Методи обробки емпіричних даних, що підпорядковуються багатомодальним законам розподілу / А.В. Горошко, В.П. Ройзман // Вісник Хмельницького національного університету. - 2013. №4. -С. 195-201.

2. Горошко А.В. Про задачу обробки статистичних матеріалів, що не підкоряються одномодальним законам розподілу. / А.В. Горошко, В.П. Ройзман // Современные достижения в науке и образовании : сб. тр. VII междунар. Науч. Конф., посвящ. 50-летию Хмельниц. Нац. Ун-та, 25 авг.-1 сент.2012 г., г. Опатия (Хорватия). - Хмельницкий : ХНУ; ФОП Сторожук О.В.-2012.-107 с. (укр., рус., англ.). ISBN 966-96180-35-43. С. 58-66.

## References

1. Goroshko A.V. Metodi obrobki yempirichnikh danikh, shcho pidporядkovuyut'sya bagatomodal'nim zakonam rozpodilu / A.V. Goroshko, V.P. Royzman // Visnik Khmel'nits'kogo natsional'nogo universitetu. -2013. №4. -S. 195-201.

2. Goroshko A.V. Pro zadachu obrobki statistichnikh materialiv, shcho ne pidkoryayut'sya odnomodal'nim zakonam rozpodilu. / A.V. Goroshko, V.P. Royzman // Sovremennyye dostizheniya v nauke i obrazovanii : sb. tr. VII mezhdunar. Nauch. Konf., posvyashch. 50-letiyu Khmel'nits. Nats. Un-ta, 25 avg.-1 sent.2012 g., g. Opatiya (Khorvatiya). - Khmel'nitskiy : KHNU; FOP Storozhuk O.V.-2012.-107 s. (ukr., rus., ang.). ISBN 966-96180-35-43. S. 58-66.

Рецензія/Peer review : 24.8.2013 р. Надрукована/Printed :17.10.2013 р.  
Рецензент: д.т.н., проф. кафедри інженерної механіки і комп'ютерних застосувань  
Хмельницького національного університету Шалапко Ю.І.