Горошко А. В.

Ройзман В. П.

Хмельницький національний університет

Goroshko A. V.

Royzman V. P.

Khmelnitsky National University

УДК 519.25:53.088

РОЗРАХУНОК ДОПУСТИМИХ ЗНАЧЕНЬ ПАРАМЕТРІВ ОБ'ЄКТІВ У ВИПАДКУ ПОЛІМОДАЛЬНОСТІ ЇХ ІМОВІРНІСНИХ РОЗПОДІЛІВ

Анотація. *Для задач статистичної обробки емпіричних даних із полімодальними законами розподілу імовірностей обґрунтовано доцільність їх представлення і обробки у вигляді суміші імовірнісних розподілів. Запропоновано метод обробки таких емпіричних даних, що базується на декомпозиції сумішей імовірнісних розподілів шляхом апроксимації емпіричної густини розподілу. Розроблено алгоритм обґрунтованого пошуку оптимального кроку побудови гістограм розподілу емпіричних даних. Запропоновано метод розрахунку допустимих значень параметрів і характеристик об'єктів у випадку полімодальності їх законів розподілу імовірностей. Наведено приклад застосування методу для статистичної обробки і визначення допустимих значень дисбалансів роторів компресорів авіадвигунів АИ-20.*

Ключові слова: *допуски, гістограма, суміш законів розподілів, дисбаланс, ротор.*

Вступ. Відомо, що допустиме значення параметра, який характеризує властивості або якість роботи нових виробів і матеріалів, які не мають вивчених аналогів, встановлюється шляхом випробувань однієї партії. При цьому для виробів (матеріалів) створюють критичні, найбільш несприятливі для їх роботи ситуації, за яких ці вироби (матеріали) ще здатні виконувати покладені на них функції, і визначають значення досліджуваного параметра. Допустимі значення параметрів з певною надійністю визначаються методами математичної статистики на основі отриманих експериментальних даних.

Найчастіше дослідники обробляють емпіричні дані, виходячи із параметричних статистичних гіпотез. Перевага застосування типових законів розподілу (нормального, логарифмічно нормального, експоненціального закону, закону Вейбулла, гамма-розподілу тощо) полягає в їх достатній вивченості та можливості отримання спроможних, незміщених і відносно високоефективних оцінок параметрів. Однак вказані вище типові закони розподілу не володіють необхідним різноманітністю форм, тому їх застосування не дає необхідної загальності подання випадкових величин, які зустрічаються при дослідженні систем.

Якщо апроксимація на основі типових розподілів не дає бажаної точності статистичних оцінок, у нагоді може стати непараметричний підхід. Методи непараметричної статистики є досить ефективними у багатьох задачах, які достатньо часто виникають на практиці, коли дослідник обробляє відносно малочисельні вибірки, нічого не знаючи про параметри досліджуваної генеральної сукупності. Одним із недоліків непараметричних критеріїв є низька статистична потужність у порівнянні зі стандартними параметричними критеріями. Ріст статистичної потужності можливий лише з ростом об'єму вибірки.

Серед всіх задач статистичної оцінки параметрів можна виокремити клас задач, в якому оцінці підлягають емпіричні дані, сформовані під дією декількох домінуючих причин, причому виявити ці причини і розділити вибірку на відповідні до них підвибірки не видається можливим. Зокрема, такі задачі часто виникають на виробництві при статистичній оцінці параметрів деякої вибірки деталей, які потрапили на підприємство із різних партій. Густина розподілу (ГР) імовірностей досліджуваних параметрів може бути полімодальною. В роботі [1] наведені приклади полімодальних гістограм розподілу фізико-механічних характеристик деяких технічних об'єктів та причини появи полімодальності.

Очевидно, що через ненормальність ГР застосування параметричного підходу в цьому разі буде не ефективним. Непараметричний підхід не дасть відповіді про причини полімодальності і не зможе розкрити внутрішню структуру даних з урахуванням можливої полімодальності законів їх розподілу. Ці проблеми тягнуть за собою труднощі із встановленням допусків експериментально досліджуваних параметрів.

Постановка задачі. Розв'язок задач статистичної обробки даних з полімодальними імовірнісними розподілами автори пропонують здійснювати, апроксимуючи емпіричну ГР сумішшю законів розподілу імовірностей, з наступною її декомпозицією і статистичною оцінкою параметрів компонентів суміші. Для цього є принаймні два важливі аргументи.

1. На відміну від непараметричного підходу, запропонований підхід дає правила роботи з такими статистичними матеріалами, зокрема, методи визначення обґрунтованих допустимих значень досліджуваних параметрів.
2. Запропонований підхід в умовах серійного виробництва і змішування партій деталей дозволяє реалізувати селективний підбір матеріалів і деталей для їх ефективнішого використання, що призводить до збільшення економічної ефективності виробництва.

Суть запропонованого авторами методу обробки емпіричних даних, що не підкоряються унімодальним законам розподілу, полягає у представлені і обробці емпіричної ГР у вигляді суперпозиції k функцій ГР f_i з вектором параметрів $\mathbf{и}_i$ (компонент суміші), $i = 1, 2, ..., k$, $2 \le k < \infty$ у вигляді

$$f(x) = \sum_{i=1}^{k} \rho_i f_i(x, \mathbf{и}_i), \qquad (1)$$

де $x \in \mathbb{R}$, ρ_i - апріорна імовірність (ваговий коефіцієнт) i-ї компоненти суміші, $\rho_i \in (0,1)$, $\sum_{i=1}^{k} \rho_i = 1$. В загальному випадку умова приналежності $\forall i, f_i(X, \mathbf{и}_i)$ до одного параметричного сімейства не ставиться.

Нехай в результаті експерименту одержана вибірка значень $x = (x_1, x_2 ..., x_n)$. Для подальшої обробки результатів експерименту, перш за все, необхідно провести декомпозицію (розщеплення) суміші, тобто визначити невідомі параметри $\rho_1, \rho_2, ... \rho_{i-1}, \mathbf{и}_1, \mathbf{и}_2, ..., \mathbf{и}_k$, наприклад, максимізуючи функцію максимальної правдоподібності [2].

$$W(\rho, \theta, x) = \prod_{j=1}^{n} \sum_{i=1}^{k} \rho_i f_i(x_j, \mathbf{и}_i), \qquad (2)$$

прирівнюючи до нуля її частинні похідні по шуканих параметрах. Як правило, замість пошуку максимуму функції $W(\rho, \theta, x)$ простіше шукати максимум її логарифму

$$\ln W(\rho, \theta, x) = \sum_{j=1}^{n} \log\left(\sum_{i=1}^{k} \rho_i f_i(x_j, \theta_i)\right), \qquad (3)$$

але навіть така постановка задачі без застосування спеціальних прийомів викликає значні труднощі. Тому для декомпозиції суміші (1) застосовують спеціальні методи: ЕМ-алгоритм і його модифікації - SEM, CEM, MCEM, SAEM тощо; наближені методи, такі як метод фіксованих компонент з використанням МНК та методу найменшим модулей, а також Баєсовський класифікатор, описані, наприклад, в роботах [2-4].

Запропонований авторами метод декомпозиції сумішей базується на апроксимації функції ГР імовірностей функцією типу (1) за допомогою МНК або інтерполяції на деякій точковій множині. В той же час відомо, що емпіричні дані вибірки $x = (x_1, x_2 ..., x_n)$ можуть бути представлені лише варіаційним рядом, гістограмою або емпіричною функцією розподілу імовірностей

$$F_n(x) = \frac{1}{n} \sum_{j=1}^{n} 1(x_j < x). \qquad (4)$$

Оскільки нормалізована гістограма (густина відносної частоти) при певних умовах є емпіричною ГР, побудованою для вибірки, апроксимація буде тим точнішою, чим краще побудована нормалізована гістограма буде наближатись до функції ГР імовірностей генеральної сукупності.

При побудові нормалізованих гістограм виникають певні труднощі. Оскільки вибір статистичної моделі розподілу визначається видом гістограми, який, у свою чергу, залежить від способу її побудови, і, особливо, від обраного кроку інтервалу значень, перед дослідниками природно постає питання, яким повинен бути крок розбиття h при побудові гістограм.

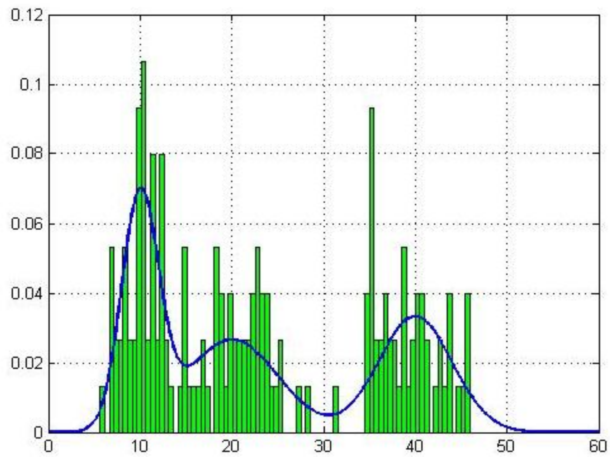
Важливість обґрунтованого вибору кількості інтервалів при побудові гістограми підкреслена, наприклад у [5], де зазначається, що задача вибору оптимальної кількості інтервалів при побудові гістограми – це задача оптимальної фільтрації, а оптимальною кількістю інтервалів є таке, коли максимально можливе згладжування випадкових флуктуацій даних сполучається з мінімальним спотворенням від згладжування самої кривої шуканого розподілу.

Рекомендації щодо вибору кроку розбиття інтервалу значень досліджуваної випадкової величини, які є в літературі з теорії ймовірностей і математичної статистики, носять чисто емпіричний характер (наприклад, правило Штюргеса). Е.С. Вентцель [6] відмічає, що кількість кроків не має бути занадто великою (тоді ряд розподілу стає невиразним, і частоти виявляють в ньому нерівномірні коливання); з іншої сторони вона не має бути занадто малою (при малій кількості кроків властивості розподілу описуються статистичним рядом занадто грубо). В багатьох роботах (див. наприклад огляд [1]) теж дані рекомендації із вибору кількості інтервалів, але незважаючи на широке практичне використання вказаних рекомендацій, лишається відкритим питання обґрунтування вибору кроку розбиття інтервалу значень досліджуваної випадкової величини таким чином, щоб побудований статистичний ряд і гістограма відповідали дійсній структурі даних і забезпечували розкриття цієї структури, зокрема, наявності суміші розподілів.

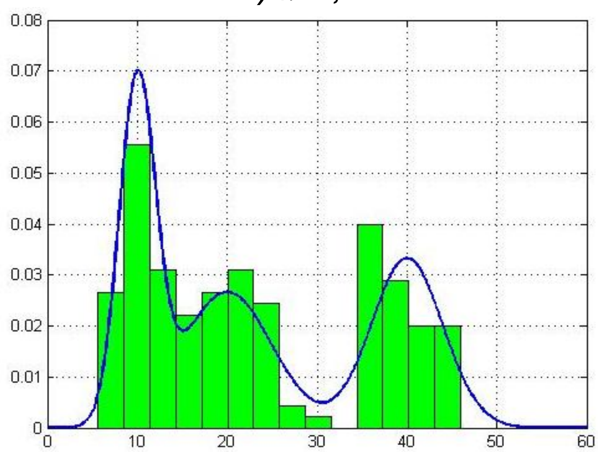
Результати досліджень. На рис. 1. проілюстровано вплив кроку на характер гістограми, зокрема кількість мод, для однієї і тієї ж вибірки емпіричних даних ($n=150$) суміші трьох гаусіан $N_i\left(\mu,\sigma^2\right)$, $i=1,2,3$, з параметрами відповідно $\mu_1=10$, $\sigma_1=2$, $\mu_2=20$, $\sigma_2=5$, $\mu_3=40$, $\sigma_3=4$, $\rho_1=\rho_2=\rho_3=1/3$.

З рис. 1 видно, що з ростом кількості інтервалів гістограма не буде наближатись до ГР. Отже, існує деякий оптимальний крок $h_{опт}$ побудови гістограм, при якому її апроксимація функцією (1) дасть оцінки параметрів ρ_i, θ_i, найближчі до їх справжніх значень.

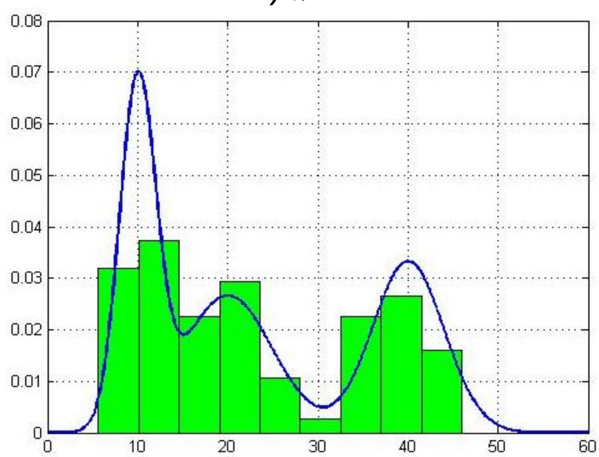
Також є очевидним, що кількість компонент суміші k, визначених за гістограмою, залежить від кроку h. Для визначення h пропонується розглянути різні варіанти його значень, а отже і різні значення кількості компонент суміші k, і надалі вибрати оптимальне значення $h_{опт}$, виходячи із деяких критеріїв якості. Одним із способів може бути перебір всіх можливих значень h, k і оцінка отриманої моделі, максимізуючи деякі критерії, але такий спосіб є занадто ресурсозатратним.



а) h ***=0,5***



б) h ***=3***



в) h ***=5***

Рис. 1. Функція ГР суміші гаусіан і гістограми розподілу емпіричних даних з різними кроками h ***їх побудови***

Слід також пам'ятати, що з ростом k буде збільшуватись і правдоподібність моделі (3), оскільки більш гнучка модель може краще пояснити досліджувані дані, тому цю задачу неможливо розв'язати, просто шукаючи k із умови максимуму правдоподібності і включивши його до шуканих параметрів.

Для вибору оптимального кроку $h_{опт}$ авторами запропонований наступний ітераційний алгоритм. Крок має бути мінімальним, але не менше за точність вимірювання параметра ε. Оскільки справжня кількість мод є невідомою, пропонується вибирати початкову кількість компонент k в (1) наперед більшою, наприклад такою, що дорівнює кількості локальних максимумів функції $f(x)$.

Далі запропонованим раніше методом необхідно визначити невідомі параметри $\rho_1, \rho_2, ...\rho_{i-1}, \mathbf{и}_1, \mathbf{и}_2, ..., \mathbf{и}_n$. Якщо в результаті розрахунків один або декілька вагових коефіцієнтів ρ_i виявляться менше деякої наперед заданої порогової величини β, то відповідними членами у лінійній комбінації (1) можна знехтувати. Дійсно, інтегральна функція розподілу з ГР (1) має вигляд

$$F(x) = \sum_{i=1}^{k} \rho_i \int_{-\infty}^{x} f_i(x, \mathbf{и}_i) dx. \qquad (5)$$

Нехай, наприклад, $\rho_1 < \beta$. Тоді після відкидання першого доданку в лінійній комбінації (1) нова інтегральна функція розподілу може бути записана у вигляді

$$\overline{F}(x) = \sum_{i=2}^{k} \rho_i \int_{-\infty}^{x} f_i(x, \mathbf{и}_i) dx.$$

Оцінимо різницю

$$\left| F(x) - \overline{F}(x) \right| =$$

$$= \rho_1 \int_{-\infty}^{x} f_1(x, \mathbf{и}_1) dx \le \rho_1 \int_{-\infty}^{\infty} f_1(x, \mathbf{и}_1) dx \le \beta.$$

Причому, отримана оцінка справедлива для довільного x. Правомірність застосування такого підходу до емпіричної функції розподілу імовірностей $F_n(x)$ підтверджується тим, що в силу теореми Гливенко при великих n рівномірно по $x \in \mathbb{R}$ виконується співвідношення

$$F_n(x) \approx F(x).$$

Наприклад, якщо функція розподілу імовірностей повинна вимірюватись з точністю до 0,01, то значення β достатньо взяти 0,005. Далі крок гістограми можна збільшувати до тих пір, поки кількість вершин (локальних максимумів) не стане дорівнювати кількості членів k в лінійній комбінації (1) 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 ρ_i не стануть порівнювані з вибраною точністю β. Отриманий при цьому крок може бути взятий за оптимальний $h_{опт}$. Фізично цей процес означає, що підвибірки з малим ρ_i вносять вельми незначний внесок у загальну вибірку і тому їх можна об'єднати з однією із підвибірок виробів з близькими величинами досліджуваного параметра.

Модифікацією запропонованого алгоритму може бути попереднє визначення кількості компонент суміші k одним із відомих методів, наприклад за Байєсовським інформаційним критерієм [3], і виконання ітераційного алгоритму до його зупинки при досягненні рівності кількості компонент знайденому значенню k. При цьому одержаний крок побудови гістограми може бути взятий за оптимальний $h_{опт}$.

Запропонований метод декомпозиції сумішей імовірнісних розподілів можна застосовувати як окремо, так і разом із відомими методами, що підвищує точність і вірогідність знайдених оцінок.

Одержання закону розподілу імовірності досліджуваного параметра у вигляді (1) дозволяє перейти до вирішення однієї із важливих практичних задач - призначення допустимого значення цього параметра з певною надійністю. Введемо деякі обмеження, а саме будемо вважати, що компоненти суміші (1) мають нормальний закон розподілу імовірностей.

Як відомо, розсіювання значень досліджуваного параметра залежить від прийнятого способу виготовлення виробу. Межі інтервалів розсіювання визначаються законами розподілу параметра, який розглядається як випадкова величина, що є сумою випадкових величин, кожна з яких викликається одним з нездоланних чинників. Якщо кількість доданків у сумі досить велике, то може виникнути два варіанти при призначенні функції розподілу параметра.

У разі, коли величина кожної зі складових у описаної раніше сумі мала в порівнянні з її величиною, за центральною граничною теоремою [6] розподіл суми близький до нормального. Фізично це умова малості кожного доданка означає, що жоден з факторів, що обумовили появу відповідної випадкової величини, не має переважаючого значення.

Якщо ж серед зазначених факторів з'являються один або кілька домінуючих, то відповідні доданки мають переважне значення в сумі і закон розподілу суми стає полімодальним.

У разі нормального закону розподілу параметра, його допустиме значення встановлюється на основі отриманих його реалізацій з таких міркувань.

Точковою оцінкою для математичного сподівання в силу закону великих чисел є вибіркове середнє арифметичне x_{cp}. Нижня и верхня довірчі границі для математичного сподівання мають вигляд $x_{cp} - t_\gamma \cdot \sigma / \sqrt{n}$ и $x_{cp} + t_\gamma \cdot \sigma / \sqrt{n}$ відповідно. Тут x_{cp} - вибіркове середнє арифметичне, t_γ - квантиль розподілу Ст'юдента взятий із таблиці для заданої довірчої імовірності γ 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 (виправлене), n – об'єм вибірки.

Тоді із довірчою імовірністю γ можна стверджувати, що допустиме значення параметра знаходиться у межах

$$x_{cp} - t_\gamma \cdot \sigma / \sqrt{n} \le x \le x_{cp} + t_\gamma \cdot \sigma / \sqrt{n} . \qquad (6)$$

Якщо результати спостережень не підпорядковуються нормальному закону розподілу імовірностей, слід застосовувати непараметричний підхід, згідно чого у (11) 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 [7], і можлива помилка піде в запас точності.

За допустиме значення параметра доцільно взяти значення на одному із кінців довірчого інтервалу (8), наприклад

$$\left[x \right] = x_{cp} - t_\gamma \cdot \sigma / \sqrt{n} . \qquad (7)$$

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

1. Розглядається підвибірка з мінімальним (максимальним) значенням μ_i. Очевидно, що характеристика цієї підгрупи мінімальна (максимальна). Отже, визначена характеристика для таких виробів може бути прийнята і для всієї партії, оскільки отримані при цьому похибки підуть у запас. У цьому випадку подальша обробка експериментальних даних може відбуватися тільки для зазначеної нормально розподіленої підвибірки значень з параметрами розподілу μ_i, σ_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 з «вагами»

$$F(x) = \sum_{i=1}^{k} \frac{\rho_i}{\sigma_i \sqrt{2\pi}} \int_{-\infty}^{x} \exp\left(-\frac{\left(x - \mu_i\right)^2}{2\sigma_i^2} \right) dx , \qquad (8)$$

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

$$\Phi^x(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} \exp\left(-\frac{x^2}{2} \right) dx$$

можна визначити імовірність

$$\gamma_i = \frac{1}{\sigma_i \sqrt{2\pi}} \int_{-\infty}^{x} \exp\left(-\frac{\left(x - \mu_i\right)^2}{2\sigma_i^2} \right) dx = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{(x-\mu_i)/\sigma_i} \exp\left(-\frac{x^2}{2} \right) dx \ , \ i = 1, 2, ..., k$$

і далі значення інтегральної функції з «вагами»

$$F^x(x) = \sum_{i=1}^{k} \rho_i \gamma_i \ . \tag{9}$$

Це означає, що функція $F^x(x)$ буде задана таблицею. Отримана таблиця дозволяє не тільки за значеннями x визначати величину функції $F^x(x)$, але і навпаки – за заданими значеннями функції визначати величину аргументу. Таким чином, для заданої довірчої імовірності можна визначити шукане допустиме значення параметра x із співвідношення виду

$$\gamma = P\{x < [x]\} = F^x([x]) = \sum_{i=1}^{k} \frac{\rho_i}{\sigma_i \sqrt{2\pi}} \int_{-\infty}^{[x]} \exp\left(-\frac{(x-\mu_i)^2}{2\sigma_i^2} \right) dx \ ,$$

складеного на основі визначення інтегральної функції розподілу випадкової величини параметра з ГР (1).

Необхідно відзначити, що, по-перше, другий шлях більш точний, оскільки він враховує функції розподілу всіх підвибірок, 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 полімодальному 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 сподіванням.

Результати практичного застосування. Далі наводяться результати перевірки ефективності викладених методу декомпозиції суміші розподілів Гауса для гістограми розподілу значень дисбалансів однотипних роторів компресорів авіадвигунів АИ-20 після експлуатації і методу призначення допусків для значень дисбалансів. Як виявилось [1], одні авіадвигуни з досліджуваної партії працювали на літаках в умовах Крайньої Півночі, інші – в умовах польових, погано обладнаних аеродромів Півдня, треті – в умовах великих перепадів температур при перельотах з Північної півкулі у Південну, четверті – у морських, корозійних умовах і т.п. Тому поява експлуатаційних дисбалансів роторів першого типу викликано в основному попаданням у двигун мілких частинок льоду і снігу, у роторів другого типу – попаданням мілких камінців та інших твердих частинок, третій тип дисбалансу викликаний в основному температурними процесами, четвертий – корозійнимиі т.п. Таким чином, 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 змішуються і утворюють партії з полімодальними законами розподілу дисбалансів, представленим на рис. 2.

Доречно відмітити, що такі значення дисбалансів не приводять до перевищення вібрацій двигунів, вище обумовлених технічними умовами і не викликають будь-яких порушень у роботі авіаційної техніки.

Для пошуку параметрів законів розподілу дисбалансів був застосований спосіб інтерполяції. Кожна компонента у (1) була прийнята за ГР нормального закону $N(\mu, \sigma^2)$. Розрахунок проводився з використанням пакету MATLAB. В таблиці 1. і на рис. 3 наведені результати вказаних розрахунків. Визначені параметри трьох гаусіан дозволили перейти до статистичної обробки емпіричних даних про дисбаланси роторів компресорів авіадвигунів з наступним визначенням допустимих значень дисбалансів.

Таблиця 1

***Шукані параметри суміші імовірнісних розподілі*в**

μ_1	μ_2	μ_3	σ_1	σ_2	σ_3	ρ_1	ρ_2	ρ_3
15,13	33,77	58,49	5,12	6,98	3,54	0,18	0,51	0,31

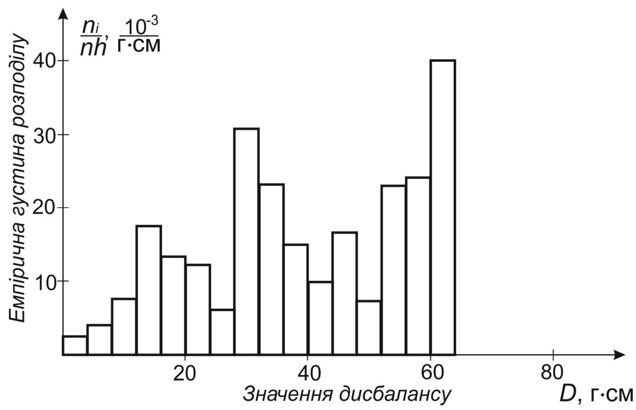


Рис. 2. Розподіл дисбалансів D ротора компресора при низькочастотному балансуванні

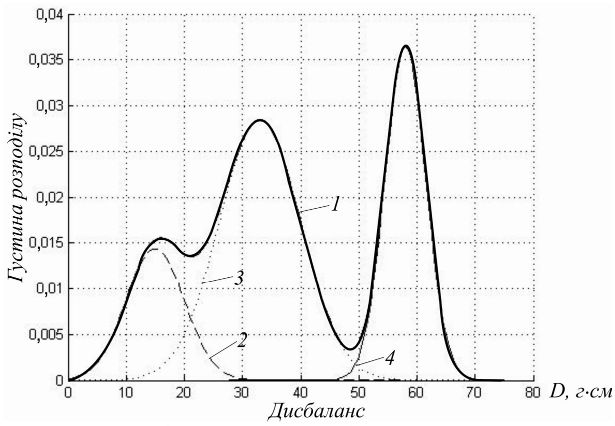


Рис. 3. Графіки шуканої функцій ГР суміші 1 и її складових компонент з параметрами μ_1, σ_1 - 2, μ_2, σ_2 - 3 і μ_3, σ_3 - 4

У відповідності з інтервалом (6) за викладеною вище методикою аналізувалась третя мода суміші розподілів, оскільки для неї характерні більші (а отже, більш небезпечні) дисбаланси. В результаті з довірчою імовірністю 0,95 було встановлено, що допустиме значення функціонального дисбалансу знаходиться в межах $51,1 \le D \le 64,9$ (г·см). Отже, для всієї досліджуваної партії роторів компресорів авіадвигунів за допустиме значення функціонального дисбалансу, тобто такого, який може мати місце під час довгої експлуатації, було прийнято 64,9 г·см, оскільки імовірність його перевищення складає не більше 0,05.

Вміщені в таблиці 1 параметри дозволили розрахувати таблицю інтегральної функції розподілу (9). Аналіз цих таблиць, отриманих за різними розділами вузлів інтерполяції, показує, що незважаючи на певну відмінність параметрів складових законів, значення інтегральної функції відрізняються неістотно, що важливо в способі призначення допусків, де застосовується саме інтегральна функція розподілу.

На основі запропонованого методу інтегральних характеристик за допомогою пакету MATLAB була записана інтегральна функція (9) і з довірчою ймовірністю 0,95 встановлено, що максимальне допустиме значення функціонального дисбалансу не перевищує 61,6 г·см, а імовірність появи у виборці дисбаланса, що не перевищує 64,9 г·см, дорівнює 0,99.

Був також проведений наступний експеримент. Той з вагових коефіцієнтів, який виявився істотно менше двох інших, вважали рівним нулю, і інтегральний закон розподілу вважали двомодальним. Для цих випадків знову визначали відповідні параметрів μ_i, σ_i, ρ_i, $i = 1, 2$, і знову розраховували таблиці інтегральної функції розподілу з «вагами».

Порівняльний аналіз таблиць, розрахованих вищевказаним чином, для трьох- і двомодальних законів розподілу показав, що вони відрізняються неістотно. Це відкриває 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.

Висновки

Запропонований метод декомпозиції сумішей імовірнісних розподілів може використовуватись як узагальнений метод статистичної обробки емпіричних даних, і застосовуватись, як частинний випадок, і для унімодальних законів розподілу. Метод дозволяє розкрити внутрішню структуру даних із врахуванням можливої полімодальності закону їх розподілу і поряд з обґрунтованим вибором кроку побудови гістограм дає правила роботи з такими даними.

Запропонований метод визначення допустимих значень параметрів об'єктів у випадку полімодальності законів розподілу імовірностей їх вибіркових емпіричних значень дозволяє одержати з певною надійністю границі допусків цих параметрів у генеральній сукупності.

Методи застосовані для статистичної обробки значень дисбалансів роторів компресорів авіадвигунів АИ-20 та для визначення їх допустимих значень.

Список використаних джерел

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. Прикладная статистика: Классификация и снижение размерности: Справ. изд. / С.А. Айвазян, В.М. Бухштабер, И.С. Енюков, Л.Д. Мешалкин; Под ред. С. А. Айвазяна. – М.: Финансы и статистика, 1989. – 607 с.: ил.
3. Королев В.Ю. ЕМ-алгоритм, его модификации и их применение к задаче разделения смесей вероятностных распределений. Теоретический обзор. – М.: Изд-во ИПИ РАН, 2007.
4. S.F. Nielsen. The stochastic EM algorithm: estimation and asymptotic results. – Bernoulli, 2000, vol. 6, No. 3, p. 457-489.
5. Новицкий П.Ф., Зограф И.А. Оценка погрешностей результатов измерений. – Л.: Энергоатомиздат. Ленингр. отд-ние, 1985. – 248 с.
6. Вентцель Е.С. Теория вероятностей / Е.С. Вентцель – М.: Наука, 1969. – 576с.
7. Орлов А.И. Прикладная статистика. Учебник. / А.И.Орлов. – М.: Издательство «Экзамен», 2004. – 656 с.

Список джерел в транслітерації

1. Goroshko A.V. Processing methodology of empirical data, under the authority of the law multimodal distributions / A.V. Goroshko, V.P. Roizman / Bulletin of the Khmelnitsky National University. – 2013. Number 4. – P. 195-201.
2. Applied statistics: Classification and reduction of dimension: ref. ed. / S.A. Aivazyan, V.M. Bukhshtaber, I.S. Eniukov, L.D. Meshalkin, ed. S.A. Aivazyan. - Moscow: Finances and Statistics, 1989. - 607 p.
3. Korolev V.Y. EM-algorithm modifications and their application to the separation of mixtures of probability distributions. Theoretical review. - Moscow: Publishing House of the IPI RAN, 2007.
4. S.F. Nielsen. The stochastic EM algorithm: estimation and asymptotic results. – Bernoulli, 2000, vol. 6, No. 3, p. 457-489.
5. Novitsky P.F., Zograph I.A. Evaluation results of measurement errors. – L. Energoatomizdat. Leningrad. Dep-tion, 1985. – 248 p.
6. Wentzel E.S. Probability theory/ ES Wentzel – Moscow: Nauka, 1969. – 576 p.
7. Alexander Orlov. Applied Statistics. Textbook / A.I.Orlov. – M.: Publishing house "Exam", 2004. – 656 p.

РАСЧЕТ ДОПУСТИМЫХ ЗНАЧЕНИЙ ПАРАМЕТРОВ ОБЪЕКТОВ В СЛУЧАЕ ПОЛИМОДАЛЬНОСТИ ИХ ВЕРОЯТНОСТНОГО РАСПРЕДЕЛЕНИЯ

Аннотация. *Для задач статистической обработки эмпирических данных с полимодальными законами распределения вероятностей обоснована целесообразность их представления и обработки в виде смеси вероятностных распределений. Предложен метод обработки таких эмпирических данных, основанный на декомпозиции смесей вероятностных распределений путем аппроксимации эмпирической плотности распределения. Разработан алгоритм обоснованного поиска оптимального шага построения гистограмм распределения эмпирических данных. Предложен метод расчета допустимых значений параметров и характеристик объектов в случае полимодальности их законов распределения вероятностей. Приведен пример применения метода для статистической обработки и определения допустимых значений дисбалансов роторов компрессоров авиадвигателей АИ -20.*

Ключевые слова: *допуски, гистограмма, смесь законов распределений, дисбаланс, ротор.*

CALCULATION OF TOLERANCES FOR PARAMETERS OF OBJECTS IN THE EVENT OF POLYMODALITY PROBABILITY DISTRIBUTIONS

Annotation. *For problems of statistical analysis of empirical data from the polymodal distribution laws of probability the expediency of their submission and processing in the form of a mixture of probability distributions. We propose a method of treatment of such empirical data, based on the decomposition of probability distributions by approximating the empirical density function. An algorithm for finding the optimal sound pitch histogram distribution of the empirical data. The method of calculation of permissible values of the parameters and characteristics of the objects in the event of their polymodality laws of probability distribution . An example of application of the method for statistical analysis and the definition of acceptable values imbalances compressor rotor aircraft engine AI-20.*

Key words: *tolerance, histogram, a mixture distribution law, imbalance, rotor.*