УДК 519.224: 620.179.12 (045)

О.В. Самойліченко, к.т.н.,
Є.Ф. Суслов

МЕТОДИ ОЦІНЮВАННЯ ЗАКОНІВ РОЗПОДІЛУ ІНФОРМАТИВНИХ ОЗНАК ДЛЯ ІМПЕДАНСНОЇ ДЕФЕКТОСКОПІЇ

Національний авіаційний університет, м. Київ, e-mail: nau_307.ukr.net

У статті проаналізовані основні методи оцінювання функції щільності імовірності випадкових величин з негауссівськими розподілами, показано їх переваги та недоліки. Розроблено методики оцінювання функції щільності розподілу імовірності (ЩРІ) на основі спеціальних рядів Грамма-Шарльє і Еджворта та на основі розподілів Джонсона, розроблено рекомендації для вибору оптимального методу залежно від наявної апріорної інформації. Показано результати апробації методик при обробці результатів неруйнівного контролю стільникових панелей імпедансним акустичним методом.

***Ключові слова**: функція щільності розподілу імовірності, закон розподілу, імітаційне моделювання, негауссовість, апроксимація, контроль, дефектоскопія*

Вступ

Аналіз процесів, що протікають в інформаційно-діагностичних системах, потребує використання моделей, до яких ставляться вимоги адекватності реальним даним. Широко відомі моделі на основі гауссівських випадкових процесів. Проте, в реальних інформаційно-діагностичних системах процеси часто носять негауссівський характер, і тому повинні описуватися нагауссівськими законами розподілу миттєвих значень. Для високої достовірності вирішення задач з використанням статистичних методів необхідно мати аналітичний запис статистичних моделей розподілів. Чим точніше оцінена функція ЩРІ, тим більш ефективними та незміщеними будуть статистичні оцінки, отримані з її використанням.

Метою даної роботи є розгляд існуючих методів оцінювання функцій ЩРІ та розроблення рекомендацій для вибору оптимального методу оцінювання даних з негауссівськими розподілами в залежності від наявної апріорної інформації щодо форми закону розподілу емпіричних даних.

Виклад основного матеріалу

Методи оцінювання функції ЩРІ можна поділити на три групи.

Непараметричні методи: гістограмний, метод локального оцінювання, на основі методу найменших квадратів. Загальний вигляд функції ЩРІ невідомий, відомі тільки деякі характеристики. Непараметричні методи ґрунтуються на знаходженні обмеженої області в просторі, що містить всі вектори вибірки, вони можуть бути використані для задач знаходження функції ЩРІ в обмеженому околі [1].

Методи відновлення сумішей розподілів застосовують у випадку, коли "форма" класів має досить складний вигляд, що не піддається опису одним будь-яким розподілом. Припускається, що всередині класу щільність розподілу являє собою суміш декількох розподілів. При цьому використовують ЕМ-алгоритм [1]. ЕМ-алгоритм складається з ітераційного повторення двох кроків. На Е-кроці обчислюється очікуване значення вектора прихованих змінних G за поточним наближенням вектора параметрів Θ. На М-кроці вирішується задача максимізації правдоподібності і знаходиться наступне наближення вектора Θ за поточним значенням векторів G та Θ. Недоліком даного методу є проблема вибору початкового наближення.

До *параметричних методів* належать: метод на основі типових розподілів, метод з використанням ортогональних подань, метод на основі універсальних сімейств розподілів, метод на основі спеціальних рядів. Загальний вигляд функції ЩРІ відомий з точністю до набору параметрів, які можна оцінити за вибіркою.

Перевага застосування типових законів розподілу полягає в їх повній вивченості і можливості отримання незміщених і відносно високоефективних оцінок параметрів [2]. Однак, типові закони розподілу не володіють необхідним різноманіттям форм, тому їх застосування не дає необхідної узагальненості представлення випадкових величин, які зустрічаються при дослідженні інформаційно-діагностичних систем.

Метод з використанням ортогональних подань дає можливість враховувати теоретичні чи експериментальні моменти всіх порядків, що дозволяє підвищити точність оцінювання функцій

ЩРІ [3]. Практичне застосування ортогональних подань щільності імовірностей потребує вирішення ряду проблем. Так, здебільшого для знаходження функцій щільності імовірностей використовують поліноми Ерміта або Лагерра, причому їх вибір строго не обґрунтовується, крім схожості форми щільності імовірностей та вагової функції; враховують лише кілька перших складових ортогональних рядів без аналізу впливу решти складових ряду; не вирішена проблема появи від'ємних значень ортогонального ряду; не дослідженні властивості оцінок коефіцієнтів розкладу та імовірнісних характеристик оцінки щільності імовірностей загалом.

Типові ряди, відомі із математичного аналізу (ряди Тейлора, Фур'є), не підходять для опису функцій розподілів, оскільки не володіють властивостями, характерними цьому виду функцій. Для подібного опису використовують спеціальні функції, наприклад, які базуються на поліномах Чебишева-Ерміта [4]. До них належать ряди Грамма-Шарльє та Еджворта. Їх доцільно використовувати для опису розподілів, близьких до нормального. В інших випадках проявляються серйозні недоліки: ряд може вести себе нерегулярно, помилки апроксимації збільшуються з віддаленням від центру розподілу, сума кінцевого числа членів ряду при великій асиметрії розподілу веде до від'ємних значень функцій, особливо на краях розподілу.

Методи на основі універсальних сімейств розподілів. Питання створення сімейств кривих, що описували б широкий клас реальних розподілів, і залежали від більш, ніж двох параметрів розглядалося в роботах Пірсона, а пізніше Джонсона та Тьюкі [2, 5, 6]. Всі ці роботи ґрунтуються на створенні такого розподілу (або декількох розподілів), які б містили одночасно параметри центру, масштабу та форми, оцінені за емпіричними даними.

Найбільше широке різноманіття форм розподілів запропоноване Пірсоном [5]. В [2] запропоновано область апроксимації експериментальних даних сімейством розподілів Пірсона для кожного типу кривої. Апроксимуючі розподіли Пірсона дають цілком задовільні результати. Однак процедура знаходження відповідної кривої Пірсона громіздка та не завжди може використовуватися на практиці, оскільки розв'язок рівняння для отримання параметрів розподілу в багатьох випадках подається в комплексному вигляді.

Сімейство gh розподілів Тьюкі було докладно вивчено в роботі [6]. Щільність gh-розподілу не може бути виражена в явній формі, це вимагає чисельних обрахунків для отримання оцінок зазначених параметрів. Використання розподілу Тьюкі показано для визначених (симетричних) класів емпіричних розподілів. Використання розподілу Тьюкі має обмеження, викликані єдиною формою розподілу і обмеження таким чином області емпіричних даних, для яких цей розподіл може бути використаний. Іншою недоліком є відсутність загального аналітичного запису функції ЩРІ, що не дозволяє перевірити відповідність емпіричних даних оціненій кривій функції ЩРІ.

Сімейство розподілів Джонсона включає три форми кривих Su, Sl та Sb. Достатній математичний опис сімейства дозволяє знаходить функцію ЩРІ в явному виді, забезпечує високу гнучкість розв'язку задачі підбору розподілів експериментальним даним. Параметри розподілу для отримання відповідної кривої обчислюють як квантильним, так і моментним методом [2].

Враховуючи вказані вище переваги та недоліки кожного методу, проведено порівняння точності оцінювання функції ЩРІ експериментальних даних рядами Грамма-Шарльє та Еджворта а також розподілами Джонсона.

Методика оцінювання функції ЩРІ рядами Грамма-Шарльє та Еджворта полягає в наступному.

Оцінюють вибіркові значення центральних моментів та їх незміщені оцінки μ_i. Отримані значення підставляють у формулу функції ЩРІ, представлену спеціальними рядами Грамма-Шарльє (1) та Еджворта (2) [4]:

$$f(x)=f_n(x)-\frac{1}{6}\sqrt{\frac{\mu_3^2}{\mu_2^3}}\cdot f_3(x)+\frac{1}{24}\left(\frac{\mu_4}{\mu_2^2}\right)f_4(x)+\frac{1}{72}\cdot\frac{\mu_3^2}{\mu_2^3}\cdot f_6(x)+\ldots. \tag{1}$$

$$f(x)=f_n(x)-\frac{1}{3!}\frac{\mu_3}{\sigma^3}f_3(x)+\frac{1}{4!}\left(\frac{\mu_4}{\sigma^4}-3\right)f_4(x)+\frac{10}{6!}\left(\frac{\mu_3}{\sigma^3}\right)^2 f_6(x)+\ldots, \tag{2}$$

де $f_n(x)$ – функція нормального закону центрованої і нормованої випадкової величини, $f_n(x)=(2\pi)^{0.5}\cdot e^{-x^2/2}$; $f_i(x)$ – i-та похідна від функції нормального розподілу.

Оцінювання функції ЩРІ кривими Джонсона зводиться до наступного: на основі оцінок асиметрії та ексцесу вибирають відповідний тип кривої [3]. Квантильним або моментним методом оцінюють чотири параметри кривої (γ, η, ξ і λ), які підставляють у кінцевий вираз для кривої Sl (3), Su (4) або Sb (5):

$$f(x) = \frac{1}{\sigma(x-\xi)\sqrt{2\pi}} \exp\left\{-\frac{1}{2\sigma^2}\left[\ln(x-\xi)-m\right]^2\right\}, \tag{3}$$

$$x \ge \xi,\ -\infty < m < \infty,\ \sigma > 0,\ -\infty < \xi < \infty,\ \eta = 1/\sigma.$$

$$f(x) = \frac{\eta}{\sqrt{2\pi}} \cdot \frac{1}{(x-\xi)^2+\lambda^2} \exp\left[-\frac{1}{2}\gamma + \eta \ln\left\{\left(\frac{x-\xi}{\lambda}\right) + \left[\left(\frac{x-\xi}{\lambda}\right)^2 + 1\right]^{1/2}\right\}^2\right], \tag{4}$$

$$\xi \le x \le \xi + \lambda,\ \eta > 0,\ -\infty < \gamma < \infty,\ \lambda > 0,\ -\infty < \xi < \infty.$$

$$f(x) = \frac{\eta}{\sqrt{2\pi}} \cdot \frac{\lambda}{(x-\xi)(\lambda - x + \xi)} \exp\left\{-\frac{1}{2}\left[\gamma + \eta \ln\left(\frac{x-\xi}{\lambda - x + \xi}\right)\right]^2\right\},$$

$$\xi \le x \le \xi + \lambda,\ \eta > 0,\ -\infty < \gamma < \infty,\ \lambda > 0,\ -\infty < \xi < \infty. \tag{5}$$

Експеримент. Результати

Застосування методу на основі спеціальних рядів має обмеження, викликані тим, що залежно від вихідних значень моментів сума кінцевого числа членів ряду при великій асиметрії розподілу веде до від'ємних значень функцій (1), (2).

З метою визначення меж застосування спеціальних рядів для оцінювання функції ЩРІ було проведено моделюючий експеримент. Для цього функцію, представлену рядами Грамма-Шарльє та Еджворта було виражено через асиметрію as та ексцес sk, попередньо припустивши, що середньоквадратичне відхилення $\sigma = 1$:

$$f(x) = f_n(x) - \frac{1}{6} \cdot as \cdot f_3(x) + \frac{1}{24} \cdot sk \cdot f_4(x) + \frac{10}{726}(as)^2 f_6(x). \tag{6}$$

Отриманий вираз (6) є однаковим як для ряду Грамма-Шарльє, так і для ряду Еджворта при виконанні умови $\sigma = 1$. Змінюючи значення асиметрії від -2 до 2, а ексцесу – від -4 до 4 циклічно з кроком 0,2 отримано результати дослідження поведінки функції ЩРІ, представленні на рис. 1 (цифрою 1 позначено область від'ємних значень функції $f(x)$, цифрою 2 – область додатніх значень).

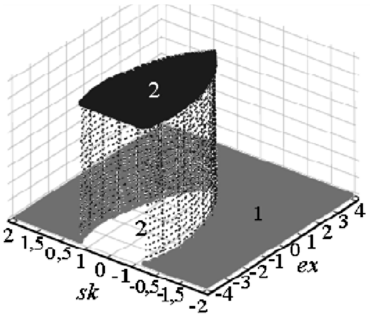


Рис. 1 Межі застосування рядів Грамма-Шарльє та Еджворта

Для визначення найкращого методу оцінювання функції ЩРІ проводилось дослідження точності апроксимації експериментальних даних з різним значенням асиметрії та ексцесу за допомогою методу імітаційного моделювання. З використанням генератора [7] генерувалась вибірка обсягом $n = 50000$ значень з різними значеннями коефіцієнту асиметрії та ексцесу. Проводилося оцінювання функції ЩРІ за наведеними методиками. Для порівняння проводилася апроксимація гауссівським законом розподілу. За допомогою статистики χ^2 оцінювалась якість апроксимації; додатково знаходилась оптимальна в сенсі мінімуму середньоквадратичної похибки функція наближення.

Отримані результати зображені на рис. 2, 3 та приведені в табл. 1, 2. На рис. 2, 3: 1 (точками) – гауссівський закон розподілу; 2 (суцільною) – розподіл Джонсона; 2 а (суцільною сірою) – розподіл Джонсона за умови вибору квантилів в околі центру; 2 б (суцільною чорною) – розподіл Джонсона за умови рівномірного вибору квантилів на всій області визначення; 3 (пунктиром) –спеціальні ряди.

а) (*as*= - 0,5) б) (*as*= - 0,9)

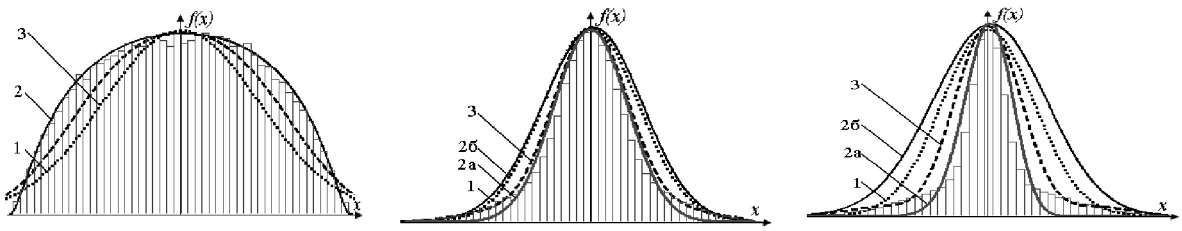
Рис. 2 Апроксимація експериментальних даних з різними значеннями асиметрії

Таблиця 1

Розраховані значення статистик за результатами апроксимації даних зі різними значеннями асиметрії

Асиметрія	Методи оцінювання функції ЩРІ	Статистика χ^2	Середньоквадратична похибка наближення
-0,1	рядом Грамма-Шарльє	11,003	0,056
	розподілом Джонсона	10,297	0,039
	розподілом Гауса	11,032	0,068
-0,3	рядом Грамма-Шарльє	13,854	0,087
	розподілом Джонсона	12,742	0,047
	розподілом Гауса	19,232	0,134
-0,5	рядом Грамма-Шарльє	15,073	0,105
	розподілом Джонсона	13,311	0,067
	розподілом Гауса	20,092	0,162
-0,9	рядом Грамма-Шарльє	18,354	0,121
	розподілом Джонсона	15,494	0,074
	розподілом Гауса	22,853	0,195

Проаналізувавши дані табл.1 можна зробити висновок, що для апроксимації вибірок великих обсягів з різними значеннями асиметрії більшу точність апроксимації забезпечує метод з використанням розподілів Джонсона завдяки широкому різноманіттю форм розподілів, ряди дозволяють отримати задовільну апроксимацію за умови невеликих значень асиметрії.



а) (*sk*= -0,95) б) (*sk*= 1,2) в) (*sk*= 2)

Рис. 3 Апроксимація експериментальних даних з різними значеннями ексцесу

Для вибірок з різними значеннями ексцесу у випадках додатного значення ексцесу оптимальним є метод на основі спеціальних рядів Грамма-Шарльє та Еджворта, хоча точність апроксимації при ексцесі, рівним 2 є найкращою, однак не оптимальною. При апроксимації даних з від'ємним значенням ексцесу найточнішим є метод з використанням розподілів Джонсона.

При оцінюванні функції ЩРІ за допомогою розподілів Джонсона слід враховувати, що точність апроксимації залежить від того, яким методом (квантильним або моментним) оцінювалися параметри а також від вибору рівнів квантилів за квантильного методу. Метод моментів доцільно використовувати, якщо обсяг вибірки не перевищує 100 значень [8]. В даній роботі для оцінювання параметрів розподілів Джонсона використовувався квантильний метод: чотири емпіричних квантиля прирівнювалися до чотирьох квантилів гауссівського закону розподілу, отриману систему з чотирьох нелінійних рівнянь розв'язують чисельними методами. Якщо обирати квантилі по всій області визначення (рівня 0,05; 0,35; 065; 0,95), то отримують апроксимацію кінців розподілу, однак форма кривих не дозволяє точно апроксимувати центр (крива 2 б, рис. 3). Щоб отримати задовільну апроксимацію центру необхідно задати квантилі рівня 0,3; 0,45; 0, 55 та 0,7, однак кінці розподілу при цьому не апроксимуються (крива 2 а, рис. 3). В табл. 2 наведені значення розраховані по кривій 2 б.

Таблиця 2

Розраховані значення статистик за результатами апроксимації даних зі різними значеннями ексцесу

Ексцес	Методи оцінювання функції ЩРІ	Статистика χ^2	Середньоквадратична похибка наближення
-0.95	рядом Грамма-Шарльє	38,9480	1,124
	розподілом Джонсона	<u>16,344</u>	<u>0,081</u>
	розподілом Гауса	41,979	1,866
-0,2	рядом Грамма-Шарльє	<u>31,264</u>	<u>0,795</u>
	розподілом Джонсона	32,063	0,882
	розподілом Гауса	31,877	0,905
1,2	рядом Грамма-Шарльє	<u>31,908</u>	<u>0,851</u>
	розподілом Джонсона	34,699	1,156
	розподілом Гауса	33,084	1,118
2	рядом Грамма-Шарльє	<u>39,460</u>	<u>1,304</u>
	розподілом Джонсона	40,215	1,414
	розподілом Гауса	41,816	1,456

Отже, під час вибору методу оцінювання даних з негауссівськими розподілами, якщо наявна апріорна інформація щодо форми закону розподілу, слід обирати метод з використанням спеціальних рядів Грамма-Шарльє та Еджворта, якщо відомо, що вибірки мають додатне значення ексцесу; у випадках, коли вибірка має асиметрію чи від’ємне значення ексцесу більшу точність забезпечить метод з використанням розподілів Джонсона.

Метод з використанням спеціальних рядів характеризується простотою застосування та зручністю у реалізації. Метод з використанням розподілів Джонсона потребує складних програм, які дозволяють вирішувати систему нелінійних рівнянь за допомогою вбудованих функцій для знаходження параметрів розподілу квантильним методом, що ускладнює застосування даного методу під час розробки портативного обладнання, яке не має можливості працювати зі складними програмами.

Досліджені методи оцінювання функцій ЩРІ було використано при опрацюванні результатів неруйнівного контролю стільникових панелей імпедансним акустичним методом. Цей метод базується на оцінці відмінностей значень механічного імпедансу в дефектних та бездефектних зонах контрольованих багатошарових конструкцій, що визначаються з поверхні виробу за допомогою збудження в ньому згинних коливань низьких частот. При імпульсному імпедансному контролі рішення про наявність пошкодженої ділянки у контрольованій області приймається випадку перевищення граничного значення для одного або декількох інформативних параметрів, яке, в свою чергу, встановлюється після налаштування дефектоскопу на стандартних зразках з нанесеними моделями дефектів. Найбільш широко уживаним інформативним параметром в дефектоскопах, що реалізують імпульсний метод збудження є амплітуда сигналу перетворювача.

Амплітуда вихідного аналогового сигналу перетворювача, що являє собою послідовність радіоімпульсів, умовно складається з суми корисної складової сигналу та випадкової складової, обумовленої відхиленнями фізико-механічних характеристик контрольованого об’єкту від середніх значень та впливом фрикційних шумів. Задача виключення впливу випадкової складової, обумовленої природними неоднорідностями структури композиційного матеріалу, змінами товщини виробу, що не можуть бути проконтрольовані в умовах виробництва, а також іншими факторами, що є наслідками складної структури композиційного матеріалу залишається досить складною. Вплив випадкових складових на інформативний параметр у випадку, коли вирішальне правило базується виключно на порівняні із деяким пороговим значенням може привести до ситуацій коли область контролю буде прийнята за помилково дефектну, або навпаки, дефект буде пропущено.

В ситуації коли на інформативний параметр, за змінами якого приймається рішення про наявність дефекту, впливає випадкова величина, рішення про наявність дефекту може бути прийняте на основі статистичних критеріїв, що включають в себе інформацію про закони розподілу вибірок результатів оцінки інформативних параметрів отриманих у апріорно дефектних та бездефектних областях виробу, а також допустимі значення похибок першого та другого роду. До таких статистичних критеріїв відносяться метод Неймана-Пірсона, метод максимальної правдоподібності, метод мінімаксу, та ін.

Основною проблемою при застосуванні зазначених критеріїв для прийняття рішення при імпульсному імпедансному методі контролю є припущення, що отримані значення інформативних параметрів розподілені за гауссівським законом. Однак при експериментальній оцінці часто виявляється, що закони розподілу амплітуд в деяких випадках можуть мати форму, відмінну від нормального, що знижує ефективність застосування даних критеріїв.

За допомогою стандартного перетворювача імпедансного дефектоскопу з дефектної та бездефектної зон була отримана вибірка з приблизно 5000 реалізацій інформаційних сигналів, з яких було виділено пікові значення амплітуд. За отриманою вибіркою пікових значень було побудовано гістограми та оцінено їх функцію ЩРІ в бездефектній зоні (рис. 4, крива 1- гауссівський закон розподілу) та в зоні з дефектом (рис. 4 крива 2 а – розподіл Sb Джонсона, крива 2 б – гауссівський закон розподілу (для порівняння)). Закони розподілу амплітуд в області дефектів суттєво відрізняються від нормального, а перетин законів розподілу у бездефектній та дефектній областях створює зону невизначеності, коли порогове значення має визначатися вже виходячи із ймовірностей похибок першого та другого роду.

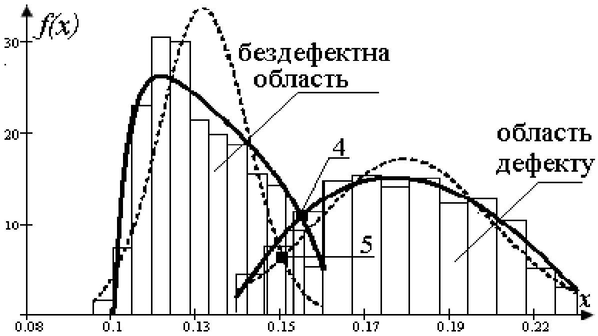


Рис. 4 Апроксимація емпіричних законів розподілів амплітуд

Отримані рівняння кривих, апроксимуючих закони розподілу, можна використовувати для реалізації статистичних правил прийняття рішень при неруйнівному контролі враховуючи величини ризиків замовника та виробника. На рис.4 показано порогове значення, отримане у випадку застосування методу оцінювання функції ЩРІ за допомогою розподілів Джонсона (суцільною) (точка 4) та при заміні закону розподілу дефектної зони гауссівським (курсивом) (точка 5).

Висновки

Отримані результати апробації методу оцінювання функції ЩРІ з використанням розподілів Джонсона свідчать про актуальність задачі оцінювання функцій ЩРІ в інформаційно-діагностичних системах. Розроблені рекомендації щодо вибору розглянутих методів залежно від апріорної інформації дає можливість отримати уточнені оцінки із застосуванням вирішальних правил в імпедансній дефектоскопії.

Список літературних джерел

1. Лепский А. Е. Математические методы распознавания образов [Електронний ресурс]: курс лекцій./ А. Е. Лепский, А. Г. Броневич. – Таганрог: ТТИ ЮФУ, 2009. – 155 с. – Режим дост. до джерела: http://www.lepskiy.ucoz.com/lect_Lepskiy_Bronevich_pass.pdf.

2. Хан Г. Статистические модели в инжинерных задачах / Г. Хан, С.Шапиро. – М. : Мир, 1969. – 395 с.

3. Берегун В. С. Дослідження щільностей імовірностей акустичних сигналів методом ортогональних подань: автореф. дис. на здобуття наук. ступеня канд. тех. наук: спец. 05.09.08 “Прикладна акустика та звукотехніка”/ В. С. Берегун. – Київ, 2010. – 20 с.

4. Кендалл М. Теория распределений / М. Кендалл, А. Стюарт. – М. : Наука, 1966. – 587 с.

5. Пирсон К. Грамматика науки / К. Пирсон; пер.с англ. // СПб. : Шиповник, 1911. – 655 с.

6. He Y. Tukey’s gh distribution for multilpe imputation / Y. He, T.E. Raghunathan // The American statistican. – 2006. – №3 (60). – P. 251–513.

7. Єременко В.С. Алгоритм генерації псевдовипадкових послідовностей з довільно заданим законом розподілу / В.С. Єременко, В.М. Мокійчук, О.В. Самойліченко // Вісник НАУ. – 2005. – №4 (26). – С.24–26.

8. Самойліченко О.В. Дослідження точності оцінювання розширеної невизначеності з використанням перетворення Джонсона в залежності від обсягу даних / О.В. Самойліченко // «ІРК» матеріали ІІІ МНПК : тези доп. – К. : НАУ, 2010. – С. 95–97.