

Еквівалентування Щільностей Розподілу Даних Випадкового Характеру

Борис Мокін

кафедра системного аналізу та інформаційних
технологій
Вінницький національний технічний університет
Вінниця, Україна
borys.mokin@gmail.com

Ольга Войцеховська

кафедра системного аналізу та інформаційних
технологій
Вінницький національний технічний університет
Вінниця, Україна
olgav1085@gmail.com

Наталія Собчук

кафедра електричних станцій та систем
Вінницький національний технічний університет
Вінниця, Україна
natashasobchuk37@gmail.com

Олексій Бондарчук

кафедра системного аналізу та інформаційних
технологій
Вінницький національний технічний університет
Вінниця, Україна
alexey.bondarchuk@aleax.me

Equivalence of Random Data Distribution Densities

Borys Mokin

dept. of System Analysis and Information Technologies
Vinnytsia National Technical University
Vinnytsia, Ukraine
borys.mokin@gmail.com

Olha Voitsekhovska

dept. of System Analysis and Information Technologies
Vinnytsia National Technical University
Vinnytsia, Ukraine
olgav1085@gmail.com

Nataliia Sobchuk

dept. of Electrical Power Stations and Systems
Vinnytsia National Technical University
Vinnytsia, Ukraine
natashasobchuk37@gmail.com

Oleksii Bondarchuk

dept. of System Analysis and Information Technologies
Vinnytsia National Technical University
Vinnytsia, Ukraine
alexey.bondarchuk@aleax.me

Анотація — Запропоновано спосіб синтезу статистичної оцінки еквівалентної щільності розподілу даних випадкового характеру, який не вимагає «вирівнювання гістограм» з використанням χ^2 -розподілу Пірсона. Спосіб використовує подрібнення гістограми з наступним кумулятивним сумуванням для визначення еквівалентної функції розподілу та її інтерполяцією кубічними сплайнами у поєднанні з класичним диференціюванням.

Abstract — A method for the synthesis of a statistical estimate of the equivalent density of the distribution of random data is proposed, which does not require "equalization of histograms" using the Pearson distribution. The method uses histogram shredding followed by cumulative summation to determine the equivalent distribution function and its interpolation by cubic splines in combination with classical differentiation.

Ключові слова — випадкова величина; гістограма; подрібнення гістограми; кумулятивна сума; функція

розподілу; щільність розподілу; інтерполяція; кубічні сплайни; Python-програми.

Keywords — random variable; histogram; histogram shredding; cumulative amount; distribution function; distribution density; interpolation; cubic splines; Python programs.

I. ВСТУП

В задачах оброблення даних випадкового характеру часто виникає необхідність в побудові статистичних оцінок диференціальних законів розподілу цих даних, або, що одне і те ж, в побудові статистичних оцінок їх щільностей розподілу, для реалізації яких в класичній математичній статистиці пропонується спосіб, що починається з формування гістограм з наступним їх «вирівнюванням» з використанням відомих теоретичних моделей для щільностей розподілу випадкових величин та χ^2 -розподілу Пірсона в якості критерія адекватності процесу «вирівнювання». При достатній кількості

експериментальних даних цей спосіб «вирівнювання гістограм», як показано в роботах [1], [2], приводить до адекватних статистичних оцінок, але при їх малих вибірках, як правило, мають місце або довірчі ймовірності зі значеннями, нижчими за поріг довіри, або виникає неоднозначність у тому, в бік якого теоретичного розподілу схилиться статистична його оцінка в припущенні збільшення кількості цих даних.

В нашій доповіді «Еквівалентування законів розподілу», тези якої опубліковані в роботі [3], запропоновано інший спосіб побудови статистичних оцінок функції розподілу випадкової величини, який не використовує процес «вирівнювання гістограм» з застосуванням χ^2 -розподілу Пірсона в якості критерія адекватності процесу «вирівнювання». За цим способом ми запропонували синтезувати еквівалентні функції розподілу, які у більшій мірі віддзеркалюють властивості конкретного набору експериментальних даних у порівнянні з їх «вирівняними за Пірсоном» еквівалентами.

II. ВИКЛАДЕННЯ ОСНОВНОГО МАТЕРІАЛУ

Суть запропонованого способу еквівалентування полягає в подрібненні базової гістограми, зображеної на рис. 1, з наступним кумулятивним сумуванням за програмою, написаною мовою Python [4], для отримання еквівалентної функції розподілу, зображеної на рис. 2 після трикратного подрібнення з наступним сумуванням. При цьому треба не забути кожен ординату поділити на три, щоб на правому кінці графіка вийти на одиницю.

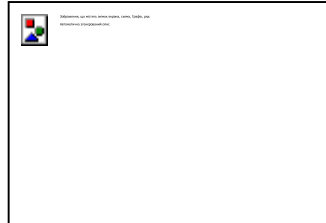
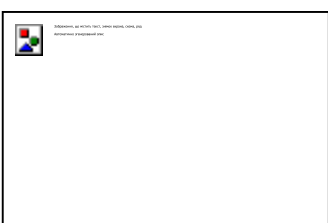


Рис. 1. Гістограма випадкових даних Рис. 2. Подрібнення і кумуляція

В даній доповіді ми пропонуємо на другому етапі еквівалентування здійснити інтерполяцію останньої версії статистичної оцінки кумулятивного варіанту функції розподілу випадкової величини, приведеної у вигляді табл. 1, за допомогою кубічних сплайнів, скориставшись методикою сплайн-інтерполяції, описаною в роботі [5].

ТАБЛИЦЯ 1.

i	0	1	2	3	4	5
x_i	x_0	x_1	x_2	x_3	x_4	x_5
y_0	(*)					
y_1		(*)				
y_2			(*)			
y_3				(*)		
y_4					(*)	
y_5						(*)

В таблиці 1 (*) - певне число.

В процесі сплайн-інтерполяції необхідно розв'язати систему із 20 рівнянь з 20 невідомими, складену з дотриманням умов рівності в дотичних точках базових для

сусідніх відрізків сплайнів, умов рівності їх перших та других похідних у цих точках, а також умов рівності нулю других похідних на кінцях поля базової гістограми, для розв'язання якої нами розроблена програма мовою Python.

Після отримання інтерполяційної формули для статистичної оцінки еквівалентної функції розподілу статистична оцінка еквівалентної щільності розподілу випадкової величини визначається згідно з базовими постулатами класичної теорії ймовірностей операцією диференціювання інтерполяційної формули, отриманої для статистичної оцінки еквівалентної функції розподілу.

І оскільки кожен кубічний сплайн віддзеркалює адекватно оцінку щільності розподілу випадкової величини лише в межах свого базового відрізка, то при обчисленні інтегралу від щільності розподілу для визначення ймовірності попадання випадкової величини в певний діапазон значень, потрібно не забувати розбивати діапазон інтегрування цього інтегралу на піддіапазони, в яких адекватно віддзеркалюють щільність розподілу саме базові для нього сплайни.

Повний текст нашої доповіді з усіма детальними математичними викладками, текстами Python-програм і конкретними числовими характеристиками, окрім виголошення в онлайн-варіанті під час роботи конференції, буде опублікований в №4 за 2024 рік наукового журналу «Вісник Вінницького політехнічного інституту», де він уже отримав позитивний відгук рецензента.

ЛІТЕРАТУРА REFERENCES

- [1] Б. І. Мокін, В. Б. Мокін, та О. Б. Мокін. Практикум для самостійної роботи студентів з навчальної дисципліни «Методологія та організація наукових досліджень». Частина 1: від постановки задачі до синтезу та ідентифікації математичної моделі. Вінниця: ВНТУ, 2018.
- [2] Б. І. Мокін, О. Б. Мокін, та О. М. Косарук. Ідеологія дуальності в вищій технічній освіті на основі інтеграції навчання з виробництвом. Вінниця: ВНТУ, 2019.
- [3] Б. І. Мокін, О. О. Войцеховська, Д. О. Шалагай, та О. В. Бондарчук, «Еквівалентування законів розподілу», на ЛІІІ Всеукраїнській науково-технічній конференції факультету інтелектуальних інформаційних технологій та автоматизації, Вінниця, Україна, 20-22 березня 2024. [Онлайн]. Доступно: <https://conferences.vntu.edu.ua/index.php/all-fksa/all-fksa-2024/paper/view/20844/17254>
- [4] Python. [Електронний ресурс]. Режим доступу: <https://www.python.org/downloads/>
- [5] Р. Н. Кветний, Я. В. Іванчук, І. В. Богач, О. Ю. Софіна, та М. В. Барабан. Методи та алгоритми комп'ютерних обчислень. Теорія і практика: підручник. Вінниця: ВНТУ, 2023.