ІНФОРМАЦІЙНІ ТЕХНОЛОГІЇ ТА КОМП'ЮТЕРНА ТЕХНІКА

https://doi.org/10.31649/1997-9266-2024-173-2-53-60

УДК 519.2+519.6]:004

Б. І. Мокін[1]
О. Б. Мокін[1]
О. О. Войцеховська[1]
Д. О. Шалагай[1]
О. В. Бондарчук[1]

ЕКВІВАЛЕНТНІ МОДЕЛІ ЗАКОНІВ РОЗПОДІЛУ ВИПАДКОВИХ ВЕЛИЧИН

[1]Вінницький національний технічний університет

Обґрунтовано, що визначення будь-яких характеристик випадкових величин можна звести до комп'ютерних обчислень з використанням відповідних чисельно заданих їхніх інтегральних законів розподілу, а тому реалізація етапу «вирівнювання» гістограм класичними функціями, якими в класичній математичній статистиці пропонується апроксимувати закони розподілу випадкових величин, стає непотрібною. Запропоновано метод синтезу еквівалентних моделей законів розподілу випадкових величин, в якому замість процедури «вирівнювання» гістограм реалізується набагато простіша процедура їхнього наближення до чисельно заданих інтегральних функцій розподілу, які після інтерполяції та диференціювання перетворюються в еквівалентні моделі диференціальних законів розподілу в разі, якщо досліднику ці закони для подальших досліджень потрібні в диференціальній формі. Показано, що використання еквівалентних моделей законів розподілу випадкових величин у разі підрахування їхніх чисельних характеристик дає результати, більшою мірою наближені до характеристик реалізацій випадкових величин, що породжують ці закони, порівняно з тими результатами, які мають місце після «вирівнювання» гістограм відомими класичними функціями з використанням процедури Пірсона. Ефективність запропонованого методу продемонстрована на прикладі побудови бази даних, придатної для ідентифікації законів розподілу оцінок студентів, котрі отримують ці оцінки, навчаючись в малоформатній групі. Вихідні умови для цього прикладу взяті з конкретної попередньої публікації авторів, в якій створено базу даних оцінок студентів в групі з 10 осіб під час вивчення ними дисципліни: «Методологія та організація наукових досліджень в галузі інформаційних технологій».

Ключові слова: випадкові величини, закони розподілу, еквівалентні моделі, вирівнювання гістограм, чисельні методи, інтерполяція, Python-програми реалізації.

Вихідні передумови і постановка задачі

У будь-якому підручнику чи навчальному посібнику з теорії ймовірностей, виданому як в Україні, так і за її межами, ми знайдемо інформацію про те, що одними із базових понять цієї науки є поняття законів розподілу випадкових величин X в їхньому інтегральному $F(x)$ чи диференціальному $f(x)$ вираженні, що зв'язані між собою відомим співвідношенням

$$f(x) = \frac{dF}{dx}. \tag{1}$$

Відсилаючи для пересвідчення до будь-якого з цих підручників чи навчальних посібників, нагадаємо, що інтегральний закон $F(x)$ розподілу випадкової величини X у вигляді

$$F(x) = P(X \le x) \tag{2}$$

задає ймовірність $P(\ )$ події, що очікувана випадкова величина X набуде значення, не більшого конкретно заданого нами числа x, а ймовірність $P(x_1 < X \le x_2)$ того, що це значення попаде в діапазон $X \in [x_1, x_2]$ може бути визначено зі співвідношення

$$P(x_1 < X \le x_2) = F(x_2) - F(x_1), \tag{3}$$

якщо відомим є її інтегральний закон розподілу $F(x)$. А якщо відомим є її диференціальний закон розподілу $f(x)$, який ще називають густиною (чи щільністю) її розподілу, то ця ймовірність може бути визначеною зі співвідношення

$$P(x_1 < X \le x_2) = \int_{x_1}^{x_2} f(x)dx. \tag{4}$$

Отже, щоб обчислити ймовірність $P(x_1 < X \le x_2)$ того, що очікуване значення випадкової величини X попаде в діапазон $X \in [x_1, x_2]$, нам потрібно знати або інтегральний закон її розподілу $F(x)$, щоби скористатися виразом (3), або диференціальний закон розподілу $f(x)$, щоби скористатись виразом (4).

А синтез та ідентифікація математичних моделей інтегрального $F(x)$ та диференціального $f(x)$ законів розподілу випадкової величини X, заданої обмеженою множиною її відомих значень, є змістом іншої навчальної дисципліни, яка називається математичною статистикою, і з навчальних посібників з якої, виданих як в Україні, так і за кордоном, легко пересвідчитись, що процедура синтезу диференціального $f(x)$закону розподілу випадкової величини X починається з побудови гістограми, що є сходинковою лінією над віссю x, а висота кожної її сходинки дорівнює відношенню кількості значень випадкової величини, що попали в межі цієї сходинки, до загальної кількості значень випадкової величини, заданої обмеженою множиною усіх її відомих значень. А другим етапом цієї процедури синтезу, тобто, етапом ідентифікації диференціального $f(x)$ закону розподілу випадкової величини X, є алгоритм «вирівнювання» сходинок гістограми гіпопотетичними неперервними кривими розподілу, запропонованими відомими математиками різних епох, починаючи з Гаусса, з використанням χ^2-розподілу Пірсона. З прикладами ідентифікації гістограми кривими нормального закону розподілу Гаусса можна ознайомитись, наприклад, в роботах [1], [2].

І ось саме цей другий етап процедури синтезу диференціального закону розподілу $f(x)$ випадкової величини X — етап «вирівнювання» сходинок гістограми гіпопотетичними неперервними кривими розподілу є найпроблематичнішим, оскільки за малих потужностей відомих значень множин випадкових величин, з використанням яких здійснюється ідентифікація, важко домогтись прийнятного значення довірчої ймовірності для «вирівнюючої» кривої, а за великих потужностей цих множин не вдається ідентифікувати гістограму однією з відомих теоретичних кривих, запропонованих різними математиками, а об'єднання кількох кривих в один ансамбль для ідентифікації гістограми складного профілю не відповідає принципу коректного розв'язання цієї задачі.

Тож подоланню цієї проблемної ситуації шляхом синтезу еквівалентних моделей законів розподілу випадкових величин, придатних для обчислення ймовірностей попадання цих величин в заданий діапазон значень, для побудови яких не потрібна процедура «вирівнювання» гістограм відомими функціями розподілу з використанням χ^2-розподілу Пірсона, і присвячено цю нашу роботу.

Розв'язання поставленої задачі

Почнемо розв'язання поставленої задачі з аналізу виразів (2), (3), (4).

З виразу (2) випливає, що функція, яка апроксимує інтегральний закон розподілу $F(x)$ випадкової величини X, є функцією, неперервно-зростаючою в межах від 0 до 1 для будь-якої множини її значень. Для констатації цього відомого з теорії ймовірностей факту надалі будемо використовувати термін «Властивість 1»

З виразу (3) випливає, що ця функція на лівій межі заданої множини значень випадкової величини X дорівнює нулю. Для констатації цього відомого з теорії ймовірностей факту надалі ми використовуватимемо термін «Властивість 2»

А з третього з цих виразів випливає, що площа під кривою функції, яка апроксимує диференціальний закон розподілу $f(x)$ випадкової величини X, дорівнює одиниці. А оскільки гістограма є статистичною оцінкою функції, яка апроксимує диференціальний закон розподілу $f(x)$ випадкової величини X, то площа фігури, обмеженою зверху сходинковою лінією гістограми, а знизу віссю абсцис, теж дорівнює одиниці. Для констатації цього відомого з математичної статистики факту надалі ми використовуватимемо термін «Властивість 3».

Для подальших викладок, скориставшись Python-програмою 1, створеною на підставі рекомендацій робіт [3]—[5] і поданою нижче, побудуємо гіпотетичну гістограму типового характеру випадкової величини X, не конкретизуючи множину значень цієї випадкової величини, але з дотриманням для її гістограми «Властивості 3».

Python-програма 1:

```
In [1]: import numpy as np
In [2]: import matplotlib
In [3]: import matplotlib.pyplot as plt
In [4]: locs=np.arange(5)
In [5]: L=[0.1,0.15,0.35,0.25,0.15]
In [6]: data=np.array(L)
In [7]: plt.bar(locs,data,width=0.99,color='blue')
```

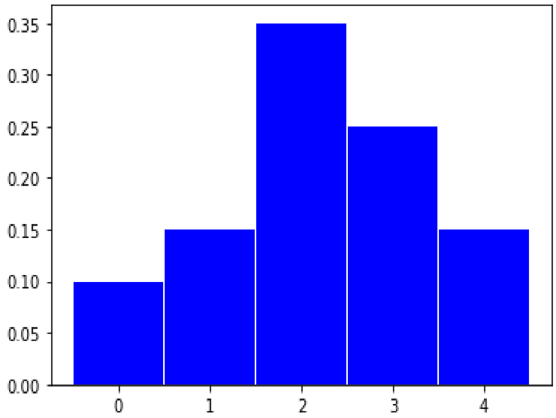


Рис. 1. Гіпотетична гістограма випадкової величини X, номер кожного розряду одиничної ширини якої

Аналізуючи показану на рис. 1 гістограму, бачимо, що навіть у цьому структурно і контурно простому її варіанті важко передбачити, в який бік вона деформуватиметься зі збільшенням кількості використаних для її побудови значень випадкової величини — чи у бік «підтягування» до вищого порогу перших двох розрядів з лівого боку від максимуму, що демонструватиме її наближення в граничному варіанті до нормального розподілу Гаусса, що описується осьовим зрізом «дзвіннице-вої» поверхні, чи у бік «опущення» до ще менших значень порогів перших двох розрядів з лівого боку від максимуму та «підвищення» до ще більших значень порогів останніх двох розрядів з правого боку від максимуму, що демонструватиме її наближення до β-розподілу. Тож, намагаючись «вирівняти» цю гістограму одним з цих двох розподілів, можемо отримати довірчу ймовірність апроксимації на рівні, меншому того, якому можна довіряти. І, якщо немає можливості додати до множини значень випадкової величини, згідно з якою будувалась гістограма, додаткових її значень, то з позицій класичної теорії ймовірностей та класичної математичної статистики задача переходить в категорію таких, що не мають коректного розв'язання.

Але в нашу комп'ютерну епоху, коли не виникає ускладнень у здійсненні будь-яких обсягів цифрових обчислень, визначення будь-яких характеристик випадкових величин можна звести до комп'ютерних обчислень з використанням відповідних чисельно заданих їхніх законів розподілу. Тож реалізація етапу «вирівнювання» гістограм класичними функціями, якими в класичній математичній статистиці пропонується апроксимувати закони розподілу випадкових величин, стає непотрібною. І замість процедури «вирівнювання» гістограм доцільно переходити до процедури чисельного їхнього наближення до чисельно заданих функцій, які ми назвали еквівалентними моделями законів розподілу. Далі ми покажемо, як здійснити це наближення.

Отже, почнемо з першого еквівалентного наближення функції $F(x)$ (рис. 2), яке отримаємо кумулятивним підсумовуванням площ усіх розрядів гістограми з рис. 1. Для цього використаємо Python-програму 2.

Python-програма 2:

```
In [1]: import numpy as np
In [2]: import matplotlib
In [3]: import matplotlib.pyplot as plt
In [4]: locs=np.arange(5)
In [5]: L=[0.1,0.15,0.35,0.25,0.15]
In [6]: data=np.array(L)
In [7]: data1=np.cumsum(data)
In [8]: plt.bar(locs,data1,width=0.99,color='blue')
```

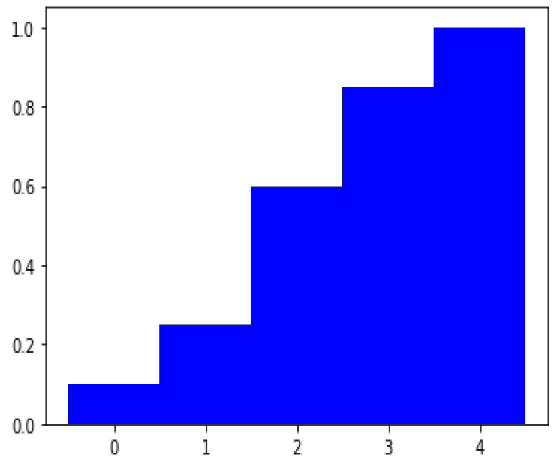


Рис. 2. Перше графічне еквівалентне наближення функції розподілу $F(x)$ випадкової величини X, отримане з гістограми рисунка1 у вигляді сходинкової кривої з числом сходинок, що дорівнює числу розрядів гістограми

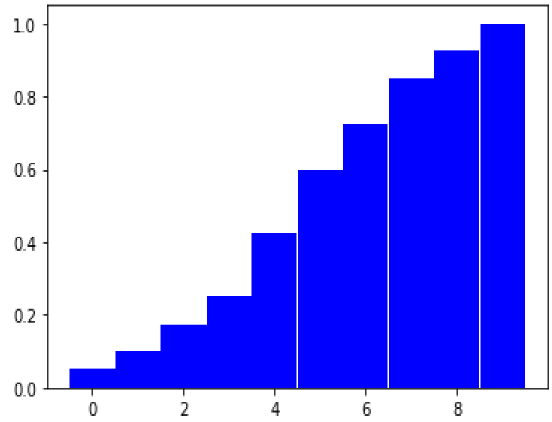


Рис. 3. Друге графічне еквівалентне наближення функції розподілу $F(x)$ випадкової величини X, отримане з гістограми рис. 1 у вигляді сходинкової кривої з числом сходинок, що дорівнює подвоєному числу розрядів гістограми

А далі, використавши Python-програму 3, розіб'ємо кожний розряд гістограми (рис. 1), на два підрозряди і за допомогою кумулятивного підсумовування площ усіх підрозрядів отримаємо друге еквівалентне наближення функції $F(x)$, показане на рис. 3.

Python-програма 3:

```
In [1]: import numpy as np
In [2]: import matplotlib
In [3]: import matplotlib.pyplot as plt
In [4]: locs1=np.arange(10)
In [5]: L=[0.1,0.15,0.35,0.25,0.15]
In [6]: data=np.array(L)
In [7]: data1=np.repeat(data,2);data1
Out[7]:
array([0.1, 0.1, 0.15, 0.15, 0.35, 0.35, 0.25, 0.25, 0.15, 0.15])
In [8]: data2=np.cumsum(data1/2)
In [8]: plt.bar(locs1,data2,width=0.99,color='blue')
```

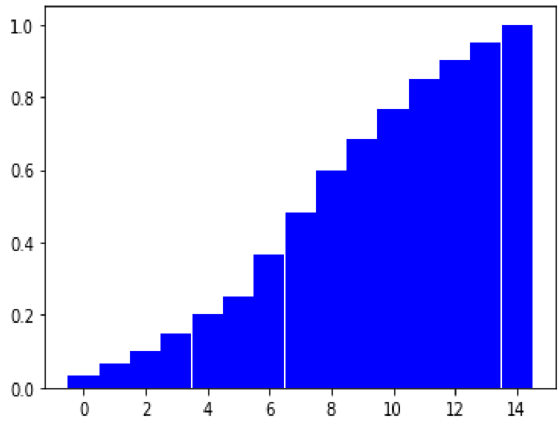


Рис. 4. Третє графічне еквівалентне наближення функції розподілу $F(x)$, отримане з гістограми рис. 1

Для однозначної інтерпретації процедури еквівалентування, використавши Python-програму 4, отримаємо третє еквівалентне наближення функції $F(x)$, показане на рис. 4.

Python-програма 4:

```
In [1]: import numpy as np
In [2]: import matplotlib
In [3]: import matplotlib.pyplot as plt
In [4]: locs2=np.arange(15)
In [5]: L=[0.1,0.15,0.35,0.25,0.15]
In [6]: data=np.array(L)
In [7]: data3=np.repeat(data,3)
In [8]: data4=np.cumsum(data3/3)
In [8]: plt.bar(locs2,data4,width=0.99,color='blue')
```

З рис. 4 та формули (3) легко бачити, що ймовірність попадання масштабованої випадкової величини X^* в діапазон значень [2, 6] дорівнює 0,3, тобто у відповідності до формули (3) маємо

$$P\left(2 < X^* \leq 6\right) = F(6) - F(2) = 0{,}4 - 0{,}1 = 0{,}3. \tag{5}$$

Якщо ж нам в подальших математичних перетвореннях законів розподілу випадкової величини необхідно мати аналітичну модель її диференціального закону розподілу $f(x)$, то необхідно спочатку здійснити інтерполяційну ідентифікацію сходинкової кривої функції розподілу $F(x)$, а потім до результату інтерполяції застосувати вираз (1), за яким визначимо математичну модель диференціального закону розподілу $f(x)$ теж в еквівалентному варіанті.

Оскільки алгоритм розв'язання задачі інтерполяції з використанням усіх варіантів інтерполяційних поліномів, до сплайнів включно, дуже детально розписаний і проілюстрований прикладами в роботі [6], то ми у цій статті про етап інтерполяції сходинкової кривої функції розподілу $F(x)$ більше нічого додавати не будемо, вважаючи, що його, скориставшись алгоритмом, викладеним в роботі [6], виконати зможе самостійно кожен науковець, який читатиме цю статтю.

А завершимо цю частину нашої статті зауваженням, що запропонований нами метод синтезу та ідентифікації еквівалентних моделей законів розподілу випадкових величин окрім своєї простоти в реалізації сприяє використанню в подальшому саме тих характеристик цієї випадкової величини, які проявились в тому обмеженому масиві її значень, використовуючи який ми побудували гістограму, і не нав'язує нам гіпопотетичного віднесення цієї випадкової величини шляхом «вирівнювання» гістограм відомими теоретичними розподілами до класу тих, що визначені на нескінченних множинах.

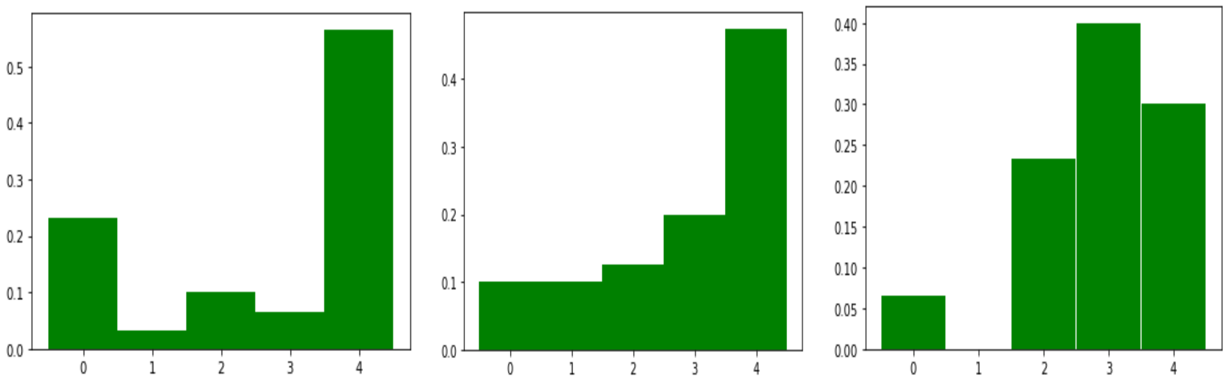
Приклад практичного застосування отриманих результатів

В нашій статті [7], написаній півтора роки тому, в якій розроблено спосіб побудови бази даних, придатної для ідентифікації законів розподілу оцінок студентів, котрі отримують ці оцінки під час вивчення певної навчальної дисципліни, зроблено такі висновки — цитуємо: «Запропоновано спосіб приведення різнодіапазонних оцінок якості різних аспектів вивчення студентами навчальних дисциплін до однорідних множин, придатних для статистичного аналізу, що дозволяє створювати достатньо інформативну базу даних навіть для малоформатних студентських груп, наявність яких в університетах є характерною особливістю існуючого нині в університетах процесу підготовки спеціалістів зі спеціальностей, які не є привабливими, незважаючи на їхню необхідність». Закінчується ця багатосторінкова стаття побудовою трьох гістограм, одна з яких є статистичною оцінкою диференціального закону розподілу оцінок студентів, виставлених викладачем за знання теорії навчальної дисципліни в діапазоні A, B, C, D, E, FX, F системи ECTS, друга — за практику застосування теорії, а третя — за засвоєння навчальної дисципліни в цілому, та обіцянкою в подальшій статті здійснити ідентифікацію цих гістограм шляхом їхнього «вирівнювання» за процедурою Пірсона будь-якими відомими теоретичними функціями розподілу.

Але, як зсувалося, дві з цих трьох гістограм «вирівнюванню» відомими теоретичними функціями за Пірсоном не піддаються, а тому покажемо, як побудувати еквівалентні закони розподілу цих оцінок, скориставшись запропонованим у цій статті методом.

І першим кроком на шляху побудови еквівалентних законів розподілу для множин оцінок, створених в роботі [7], буде приведення гістограм, побудованих для оцінок в діапазоні A, B, C, D, E, FX, F системи ECTS, до вигляду, поданому нами на рис. 1.

Ці гістограми, побудовані за використання даних табл. 3 роботи [7], показані на рис. 5, причому на рис. 5*a* зпоказана гістограма оцінок, отриманих малокомплектною групою студентів, за знання теорії навчальної дисципліни; на рисунку 5*b* зображена гістограма оцінок, отриманих цією ж групою студентів, за практичне застосування теорії цієї навчальної дисципліни, а на рис. 5*c* зображена гістограма оцінок, отриманих цими ж студентами, за підсумкове оволодіння знаннями з цієї навчальної дисципліни загалом.



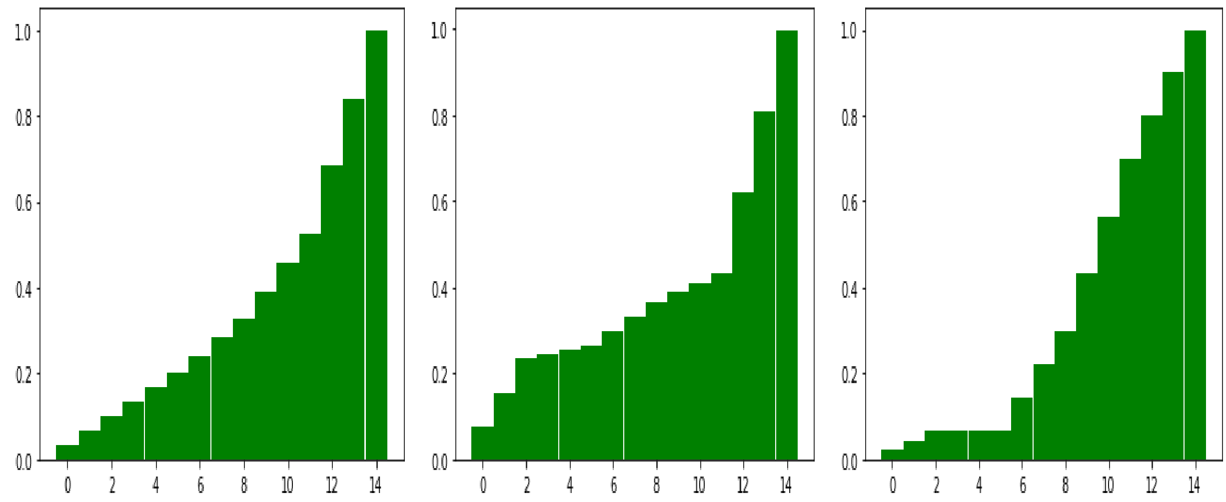
a *b*) *c*)

Рис. 5. Гістограми оцінок, отриманих студентами малокомплектної групи за знання теорії, практики та навчальної дисципліни в цілому, побудовані за даними табл. 3 роботи [7]

З вигляду гістограм на рис. 5, легко бачити, що «вирівняти» жодну з них, використовуючи відомі теоретичні моделі диференціального закону розподілу, з прийнятною довірчою ймовірністю здійснити не вдасться.

А тому є сенс побудувати з їхнім використанням еквівалентні моделі функцій розподілу $F_T(x), F_p(x), F_{nd}(x)$, застосовуючи алгоритм та Python-програми приведеного вище методу.

Реалізуючи для кожної з гістограм на рис. 5 Python-програму 4, отримаємо еквівалентні моделі функцій розподілу $F_T(x), F_p(x), F_{nd}(x)$, геометричне зображення яких показано на рис. 6 у тій же послідовності, що й зображення, які породили їх (див. рис. 5). Тобто на рис. 6*a* показано геометричне зображення еквівалентної функції розподілу $F_T(x)$; на рисунку 6*b* показано геометричне зображення еквівалентної функції розподілу $F_p(x)$, а на рис. 6*c* показано геометричне зображення еквівалентної функції розподілу $F_{nd}(x)$.



a *b*) *c*)

Рис. 6. Графічне відображення еквівалентних функцій розподілу $F_T(x), F_p(x), F_{nd}(x)$ оцінок, отриманих студентами малокомплектної групи за знання теорії, практики та навчальної дисципліни в цілому, побудовані з використанням запропонованого методу еквівалентування на основі бази даних табл. 3 роботи [7]

Використовуючи еквівалентні функції розподілу для обчислення ймовірності отримання студентом підсумкової оцінки з діапазону A, B, C, D, E, FX, F системи ECTS, слід пам'ятати, що створюючи ці еквівалентні моделі, ми спочатку приводили кількісну характеристику групи студентів до відрізка [0, 1], потім повторювали кожну ординату гістограми кілька разів (у нашому випадку 3 рази), приводячи кожний повтор знову до відрізку [0, 1], тож для повернення до кількісної характеристики випадкової величини у вигляді кількості студентів, що отримали оцінку певної категорії, необхідно кожну абсцису графіка еквівалентної функції розподілу поділити на кількість розрядів базової гістограми (у нашому випадку на 5) та на кількість повторень кожної ординати (у нашому випадку на 3) і помножити результат ділення на кількість даних, за якими будувалась гістограма (у нашому випадку 30). Наприклад, абсцисі графіка на рис. 6, яка дорівнює 12, відповідатиме кількісна оцінка

$$\frac{12}{15} \cdot 30 = 24 . \qquad (6)$$

Якщо ж нас цікавлять еквівалентні моделі диференціального закону розподілу, отримані на базі інтегральних, показаних на рис. 6, то, як ми уже про говорили вище, необхідно, скориставшись розписаною в роботі [6] процедурою інтерполяції, отримати інтерполяційні поліноми для наших сходинкових кривих, які потім продиференціювати згідно з виразом (1).

Висновки

Обґрунтовано, що в нашу комп'ютерну епоху, коли не виникає ускладнень у здійсненні будь-яких обсягів цифрових обчислень, визначення будь-яких характеристик випадкових величин можна звести до комп'ютерних обчислень з використанням відповідних чисельно заданих їхніх законів розподілу, а тому реалізація етапу «вирівнювання» гістограм класичними функціями, якими в класичній математичній статистиці пропонується апроксимувати закони розподілу випадкових величин, стає непотрібною.

Запропоновано метод синтезу еквівалентних моделей законів розподілу випадкових величин, в якому замість процедури «вирівнювання» гістограм реалізується набагато простіша процедура їхнього наближення до чисельно заданих інтегральних функцій розподілу, які після інтерполяції та диференціювання перетворюються в еквівалентні моделі диференціальних законів розподілу.

Ефективність запропонованого методу продемонстрована на прикладі побудови бази даних, придатної для ідентифікації законів розподілу оцінок студентів, котрі отримують ці оцінки, навчаючись в малоформатній групі, під час вивчення конкретної навчальної дисципліни.

СПИСОК ВИКОРИСТАНОЇ ЛІТЕРАТУРИ

[1] Б. І. Мокін, В. Б. Мокін, і 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.* ВНТУ, Вінниця, 2018, 179 с. [Електронний ресурс]. Режим доступу: http://mokin.com.ua/files/articles/65/46/Mokin_SRS_MOND.pdf .

[2] Б. І. Мокін, О. Б. Мокін, і О. М. Косарук, *Ідеологія дуальності в вищій технічній освіті на основі інтеграції навчання з виробництвом*, моногр. Вінниця, Україна: ВНТУ, 2019, 224 с.

[3] *Python*. [Електронний ресурс]. Режим доступу: https://www.python.org//downloads/ .

[4] Б. І. Мокін, В. Б. Мокін, і О. Б. Мокін, *Навчальний посібник для опанування студентами способів розв'язання задач з функціонального аналізу мовою Python, Частина 1.* ВНТУ, Вінниця, 2022, 124 с.

[5] Б. І. Мокін, В. Б. Мокін, і О. Б. Мокін, *Навчальний посібник для опанування студентами способів розв'язання задач з функціонального аналізу мовою Python, Частина 2.* ВНТУ, Вінниця, 2023, 144 с.

[6] Р. Н. Квєтний, Я. В. Іванчук, І. В. Богач, О. Ю. Софина, і М. В. Барабан, *Методи та алгоритми комп'ютерних обчислень. Теорія і практика*, підручн. ВНТУ, Вінниця, 2023, 280 с.

[7] О. О. Войцеховська, Б. І. Мокін, і Д. О. Шалагай, «Про один спосіб створення бази даних для системного аналізу якості засвоєння студентами навчальної дисципліни,» *Вісник Вінницького політехнічного інституту,* № 5, с. 58-67, 2022. https://doi.org/10.31649/1997-9266-2022-164-5-58-67 .

Рекомендована кафедрою системного аналізу та інформаційних технологій ВНТУ.

Стаття надійшла до редакції 14.02.2024

Мокін Борис Іванович — академік НАПН України, д-р техн. наук, професор, професор кафедри системного аналізу та інформаційних технологій, e-mail: borys.mokin@gmail.com ;

Мокін Олександр Борисович — д-р техн. наук, професор, професор кафедри системного аналізу та інформаційних технологій, e-mail: abmokin@gmail.com ;

Войцеховська Ольга Олександрівна — д-р філософії, старший викладач кафедри системного аналізу та інформаційних технологій, e-mail: olgav1085@gmail.com ;

Шалагай Дмитро Олександрович — аспірант кафедри системного аналізу та інформаційних технологій, e-mail: d.shalagai@gmail.com ;

Бондарчук Олексій Валерійович — аспірант кафедри системного аналізу та інформаційних технологій, e-mail: alexey.bondarchuk@aleax.me .

Вінницький національний технічний університет, Вінниця

B. I. Mokin[1]
O. B. Mokin[1]
O. O. Voitsekhovska[1]
D. O. Shalagai[1]
O. V. Bondarchuk[1]

Equivalent Models of the Random Variable Distribution Laws

[1]Vinnytsia National Technical University

It is substantiated that the determination of any characteristics of random variables can be reduced to computer calculations using the corresponding numerically determined integral laws of their distribution, and therefore the implementation of the stage of "equalization" of histograms with classical functions, which are proposed in classical mathematical statistics to approximate the random variable distribution laws, becomes unnecessary. A method of synthesis of equivalent models of the random variables distribution laws is proposed, in which instead of the procedure of "equalization" of histograms, a much simpler procedure of their approximation to numerically given integral distribution functions is implemented, which, after interpolation and differentiation, are transformed into equivalent models of differential laws of distribution in the event that the researcher has these laws for further studies are needed in a differential form. It is shown that *the use of equivalent models of the random variable distribution laws when calculating their numerical characteristics gives results that are closer to the characteristics of the implementations of random variables generating these laws compared to the results that occur after the "equalization" of histograms by known classical functions using the Pirson procedure The effectiveness of the proposed method is demonstrated on the example of building a database suitable for identifying the laws of the distribution of grades of students who receive these grades while studying in a small-format group. The starting conditions for this example are taken from a specific previous publication of the authors, in which they created a database of student evaluations in a group of 10 people during their study: "Methodology and organization of scientific research in the field of information technologies."*

Keywords: random variables, distribution laws, equivalent models, histogram alignment, numerical methods, interpolation, Python implementation programs.

Mokin Borys I. — Academician of NAPS of Ukraine, Dr. Sc. (Eng.), Professor, Professor of the Chair of System Analysis and Information Technologies, e-mail: borys.mokin@gmail.com ;

Mokin Oleksandr B. — Dr. Sc. (Eng.), Professor, Professor of the Chair of System Analysis and Information Technologies, e-mail: abmokin@gmail.com ;

Voitsekhovska Olha O. — PhD, Senior Lecturer of the Chair of System Analysis and Information Technologies, e-mail: olgav1085@gmail.com ;

Shalagai Dmytro O. — Post-Graduate Student of the Chair of System Analysis and Information Technologies, e-mail: d.shalagai@gmail.com ;

Bondarchuk Oleksii V. — Post-Graduate Student of the Chair of System Analysis and Information Technologies e-mail: alexey.bondarchuk@aleax.me