УДК 004

Б. І. Мокін[1]
О. О. Войцеховська[1]
Д. О. Шалагай[2]
О. В. Бондарчук[1]

ЕКВІВАЛЕНТУВАННЯ ЗАКОНІВ РОЗПОДІЛУ

[1]Вінницький національний технічний університет
[2]ТОВ "Промавтоматика Вінниця"

Анотація

Обґрунтовано, що визначення будь-яких характеристик випадкових величин можна звести до комп'ютерних обчислень з використанням відповідних чисельно заданих їх інтегральних законів розподілу. Запропоновано метод синтезу еквівалентних моделей законів розподілу випадкових величин, в якому замість процедури «вирівнювання» гістограм реалізується набагато простіша процедура їх наближення до чисельно заданих інтегральних функцій розподілу.

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

Abstract

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. A method of synthesis of equivalent models of the random variable 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.

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

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 їх інтегральному $F(x)$ чи диференціальному $f(x)$ виразженні, 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

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

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

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

якщо відомим є її інтегральний закон розподілу $\mathrm{F}(\mathrm{x})$. 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

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

Отже, щоб обчислити ймовірність $P(x_1<X\leq x_2)$ того, що очікуване значення випадкової величини

X попаде в діапазон $X \in [x_1, x_2]$, нам потрібно знати або інтегральний закон її розподілу $F(x)$, щоб скористатись виразом (3), або диференціальний закон розподілу $f(x)$, щоб 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 являє собою сходинкову лінію над віссю x , висота кожної сходинки якої 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 найбільш проблематичним, оскільки при малих потужностях відомих значень множин випадкових величин, з використанням яких здійснюється ідентифікація, важко домогтись прийнятного значення довірчої ймовірності для «вирівнюючої» кривої, а при великих потужністях цих множин не вдається ідентифікувати гістограму однією із 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 присвячена ця наша робота.

Розв'язання поставленої задачі

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

Із першого з цих виразів витікає, 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 в подальшому тексті ми будемо використовувати термін «Властивість 1».

Із другого з цих виразів витікає, що ця функція на лівій межі заданої множини значень випадкової величини X дорівнює нулю - для констатації цього відомого з теорії ймовірностей факту в-подальшому тексті ми будемо використовувати термін «Властивість 2».

А із третього з цих виразів витікає, 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 в-подальшому тексті ми будемо використовувати термін «Властивість 3».

Для подальших викладок, скориставшись відповідною Python-програмою 1, створеною на підставі рекомендацій робіт [3],[4],[5], побудуємо гіпопотетичну гістограму (рис. 1) типового характеру випадкової величини X, не конкретизуючи множину значень цієї випадкової величини, але з дотриманням для її гістограми «Властивості 3».

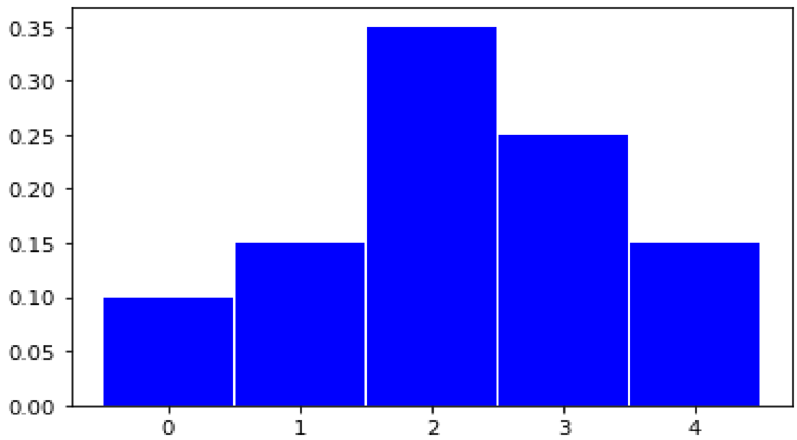

Рис. 1. Гіпопотетична гістограма випадкової величини X, номер кожного одиничної ширини розряду якої проставлено в його середині

Аналізуючи приведену на рисунку 1 гістограму, бачимо, що навіть у цьому структурно і контурно простому її варіанті важко передбачити, в який бік вона деформуватиметься при збільшенні кількості використаних для її побудови значень випадкової величини – чи у бік «підтягування» до більш високого порогу перших двох розрядів з лівого боку від максимуму, що демонструватиме її наближення в граничному варіанті до нормального розподілу Гаусса, що описується осьовим зрізом «дзвінницевої» 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 із цих двох розподілів, ми 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'язку.

Але в нашу комп'ютерну епоху, коли не виникає ускладнень у здійсненні будь-яких цифрових об'ємів обчислень, визначення будь-яких характеристик випадкових величин можна звести до комп'ютерних обчислень з використанням відповідних чисельно заданих їх законів розподілу. Тож реалізація етапу «вирівнювання» гістограм класичними функціями, якими в класичній математичній статистиці пропонується апроксимувати закони розподілу випадкових величин, стає не потрібною. І замість процедури «вирівнювання» гістограм доцільно переходити до процедури чисельного їх 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 $F(x)$, яке отримаємо кумулятивним сумуванням площ усіх розрядів гістограми, приведеної на рисунку 1. Для цього використаємо відповідну Python-програму 2.

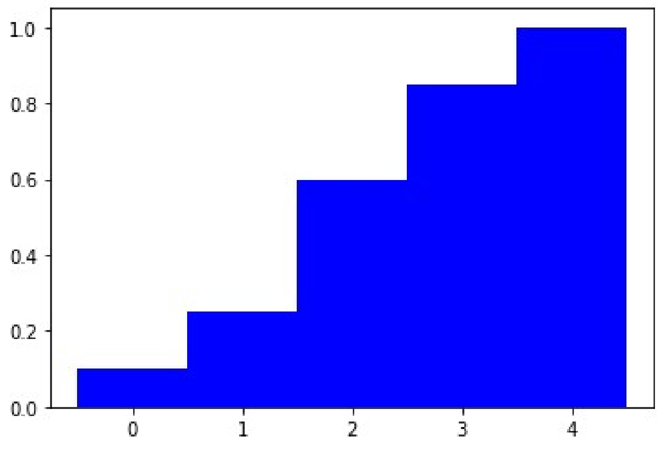

Рис. 2. Перше графічне еквівалентне наближення функції розподілу $F(x)$ випадкової величини X, отримане з гістограми рис.1 у вигляді сходинкової кривої з числом сходинок, що дорівнює числу розрядів гістограми

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

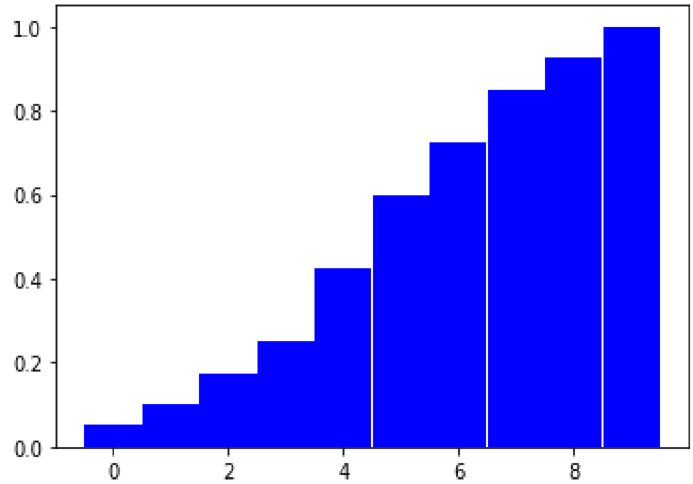

Рис. 3. Друге графічне еквівалентне наближення функції розподілу $F(x)$ випадкової величини X, отримане з гістограми рис.1 у вигляді сходинкової кривої з числом сходинок, що дорівнює подвоєному числу розрядів гістограми

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

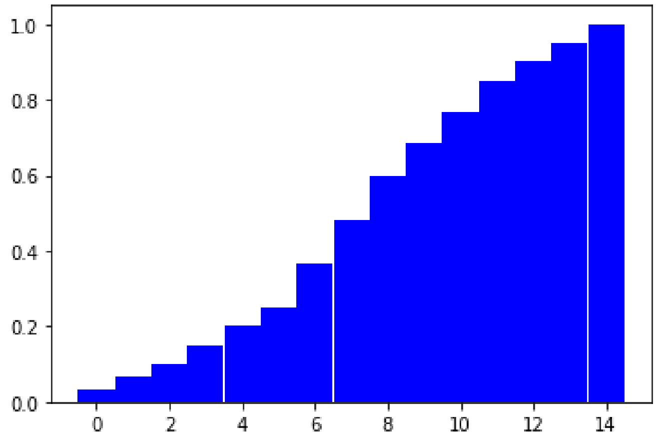

Рис. 4. Третє графічне еквівалентне наближення функції розподілу $F(x)$, отримане з гістограми рис.1

Із 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 з формулою (3) маємо

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

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), яким ми 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 нашій роботі 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 нашу роботу.

А завершимо ми цю нашу роботу 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, з використанням якого 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.

СПИСОК ВИКОРИСТАНОЇ ЛІТЕРАТУРИ

1. Мокін Б. І. Практикум для самостійної роботи студентів з навчальної дисципліни «Методологія та організація наукових досліджень». Частина 1: від постановки задачі до синтезу та ідентифікації математичної моделі / Б. І. Мокін, В. Б. Мокін, О. Б. Мокін.– Вінниця: ВНТУ, 2018. – 179 с.
2. Мокін Б. І. Ідеологія дуальності в вищій технічній освіті на основі інтеграції навчання з виробництвом / Б. І. Мокін, О. Б. Мокін, О.М.Косарук. – Вінниця: ВНТУ, 2019. – 224 с.
3. Python. [Електронний ресурс]. Режим доступу: https://www.python.org//downloads/
4. Мокін Б.І., Мокін В. Б., Мокін О. Б. Навчальний посібник для опанування студентами способів розв'язання задач з функціонального аналізу мовою Python, частина 1 / Вінниця: ВНТУ, 2022. – 124 с.
5. Мокін Б.І., Мокін В. Б., Мокін О. Б. Навчальний посібник для опанування студентами способів розв'язання задач з функціонального аналізу мовою Python, частина 2 / Вінниця: ВНТУ, 2023. – 144 с.
6. Кветний Р.Н., Іванчук Я.В., Богач І.В.. та ін Методи та алгоритми комп'ютерних обчислень. Теорія і практика. Підручник./ Вінниця: ВНТУ, 2023. – 280 с.

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

Войцеховська Ольга Олександрівна — PhD, старший викладач кафедри системного аналізу та інформаційних технологій, Вінницький національний технічний університет, м. Вінниця, e-mail: olgav1085@gmail.com;

Шалагай Дмитро Олександрович — інженер-проектувальник ТОВ "Промавтоматика Вінниця", м. Вінниця, e-mail: d.shalagai@gmail.com;

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

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

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

Shalagai Dmytro O. — design-engineer Ltd "Promavtomatika Vinnytsia", Vinnytsia, 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.