П. КОСОБУЦЬКИЙ

Національний університет "Львівська політехніка"
(Львів 79012; e-mail: petkosob@gmail.com)

УДК 539

ОПТИМАЛЬНІ ЗАКОНОМІРНОСТІ НОРМАЛЬНОГО РОЗПОДІЛУ З ТОЧКИ ЗОРУ ОЦІНКИ СТАТИСТИКИ ВИБІРКИ РЕЗУЛЬТАТІВ ФІЗИЧНОГО ЕКСПЕРИМЕНТУ

На підставі аналізу літературних джерел синтезовані базові ймовірнісні і принципи формування нормального розподілу випадкових розсіянь значень фізичних величин в умовах незалежних випадкових дій на фізичну систему. Зроблений наголос на комплексному підході ймовірнісно-статистичного аналізу вибірки результатів експериментальних вимірювань.

Ключові слова: нормальний розподіл, математичне сподівання, дисперсія, випадкові величини.

1. Вступ

Як відомо, складна система, як ідеальний газ із N частинок, в стані теплової рівноваги підпорядковується лапласівському детермінізму. Але як показав Крилов, зіткнення частинок одна з одною призводить до нестійкості руху багаточастинкової системи за Ляпуновим – рух стає хаотичним і детермінована модель для опису руху не підходить.

Насправді роль випадковостей у формуванні рівноважного стану складної системи зрозумів ще Максвелл і обґрунтував фізичний зміст функції ймовірності. Правильність її змісту і суті було підтверджено дослідами Броуна і Перрена, які встановили, що незважаючи на те, що середнє значення колективного параметра зміщення блукаючої частинки дорівнює нулю, $\langle \Delta x \rangle = 0$, її середнє квадратичне значення відмінне від нуля $\sqrt{\overline{(\Delta x)^2}} \neq 0$. Після фундаментальних теоретичних робіт Максвелла і Больцмана остаточно стало ясно, наскільки важлива роль ймовірнісних підходів у фізичному моделюванні.

Базовою в теорії ймовірностей і математичної статистики є центральна гранична теорема (ЦГТ). Вона стверджує, що коли на вимірювану фізичну величину діє багато випадкових незалежних (або слабо залежних) факторів, кожний із яких зокрема відіграє в загальному результаті недомінантну роль, то статистичний розподіл значень вибірки прямує до нормального (розподілу Гауса) із параметрами m_X (математичне сподівання (МС) E_X або $M[X]$) і σ_X (середньоквадратичне відхилення (СКВ) σ_X: $\sigma_X^2 = D_X$ (дисперсія, як характеристика розсіяння ВВ відносно МС, має розмірність її квадрата, що не дуже зручно для статистики, тому введений параметр σ_X) [1].

Нормальний $N(m_X, \sigma_X)$ розподіл найчастіше використовується для побудови фізичних моделей методом неперервних ВВ і йому присвячено дуже багато робіт [2]. Ця теоретична модель найкраще опрацьована аналітично і на її основі розвинута теорія статистичних оцінок похибок [3]; метод Аленцева–Фока розкладу спектрів на індивідуальні моди [4]; ймовірнісні підходи дослідження динаміки лінійних і нелінійних систем [5]; явищ квантової механіки [6, 7]; створені потужні обчислювальні ресурси комп'ютерної симуляції динаміки атомно-молекулярних сполук [8], тощо [9]. На основі гаусово розподілених ВВ розвинутий цілий напрям статистичної оптики [10]. Доведено також, що нормальний розподіл реалізується у багатьох хаотичних системах, як більярд Больцмана, броунівський рух у гармонічному потенціалі тощо.

Нормальний розподіл стійкий і здатний до самовідтворення. Тому успішно був використаний для статистичного моделювання сучасних високих технологій, як лазерне охолодження атомів [11], стохастичні явища в радіотехнічних системах [12, 13]. Крім цього, розподіл Гауса відноситься до розподілів із "важкими хвостами" [1]. Це дозволяє на підставі ЦГТ допустити, що сума достатньо значної кількості малих ВВ розподілена нормально і цей розподіл граничний для багатьох розподілів, яким би законам не були підпорядковані окремі похибки, оскільки особливості їх у сумі великої кількості доданків нівелюються [14]. Тому клас нормальних функцій найчастіше вибирають за наближену модель генеральної сукупності випадкових даних і за їх вибіркою досить оцінити два параметри m_X і σ_X, а суміші розподілів на їх основі, що володіють безмежною диференційованістю, гладкістю тощо, широко використовуються для різноманітного моделювання [15–19].

Мотивацією даної статті стало застереження про те, що статистичний аналіз закономірностей вибірок експериментальних результатів і вибір теоретичної моделі [20, 21] не завжди супроводжується належною увагою до базових засад теорії ймовірностей та математичної статистики, що відзначалось автором ще в роботі [22].

2. Базові принципи аналізу даних фізичної моделі на предмет нормальності розподілу випадкових значень вимірюваної величини

Фізичні величини чи умови їх вимірювання, завжди зазнають випадкових флуктуацій. Тому коректно вести мову про їх статистично усереднені значення, для чого використовують ймовірнісні методи опрацювання результатів, в тому числі на основі лінійних і нелінійних перетворень ВВ. Флуктуації – це випадкові відхилення макроскопічних величин від їх середніх (наприклад, термодинамічно рівноважних) значень. У макросистемах, флуктуації спричинені наявністю великої кількості ступеней вільності, які мають макроскопічну природу. В мікросистемах, флуктуації зумовлені корпускулярною (атомною) природою будови речовини та електрики (найменша частка електрики дорівнює електрону).

Джерелом флуктуацій систем атомістичної природи є теплові коливання мікрочастинок, електронів, іонів, тощо, із якими пов'язаний броунівський рух, молекулярне розсіяння світла в середовищі, теплові шуми в електричних колах, теплове випромінювання тіл, дробовий ефект в електроніці, спричинені варіаціями електронного потоку в процесах перенесення. Хаотичний характер має процес перемагнічування доменів у феромагнітних середовищах.

Оскільки в реальних умовах вимірювання розподіл зростаючої кількості виміряних фізичних даних завжди прямує до нормального, то наведемо найбільш вагомі аргументи важливості у фізичному моделюванні застосування закону Гауса: 1. Розподіл типу $N(m_X;\sigma_X^2)$ – це добра математична модель для опису явищ і процесів з використанням випадкових чисел. 2. Довільні лінійні комбінації нормально розподілених ВВ, розподілені нормально. 3. Гаусовий випадковий процес цілком піддається теоретичному моделюванню лише з використанням першого і другого моментів, що не справджується для інших розподілів. 4. При проходженні нормально розподіленого випадкового сигналу через лінійну систему, сигнал на виході також розподілений нормально. Тому якщо відсутня теоретична та експериментальна інформація про характер розподілу ВВ (наприклад, шуму вимірювань), то найчастіше допускають, що неперервна ВВ розподілена нормально, а аналогічно дискретна – за законом Пуасона, який прямує до розподілу Гауса при зростанні обсягу вибірки.

Але в статистиці актуальні й інші зв'язані з нормальним розподіли, як розподіл χ^2-квадрат, розподіл Стьюдента (t-розподіл) та розподіл Фішера (F-розподіл). Якщо кожна із n незалежних ВВ $\xi_1, \xi_2, ..., \xi_n$ розподілені за стандартним законом $N(0,1)$, то ВВ типу $\eta = \xi_1^2 + \xi_2^2 + ... + \xi_n^2$ вже розподілена за законом χ^2 із n ступенями вільності. Якщо кожна із $n+1$ незалежних ВВ $\xi_0, \xi_1, \xi_2, ..., \xi_n$ розподілена за законом $N(0,\sigma^2)$, то ВВ типу

$$\eta = \frac{\xi_0}{\sqrt{\frac{1}{n}\sum\limits_{i=1}^{n}\xi_i^2}}$$

[1] Суть розподілу з "важкими хвостами " полягає в тому, що не досить часті події, як, зіткнення літаків, насправді не досить рідкі з точки зору безпекового функціонування систем, щоб їх ймовірністю можна було нехтувати. Якщо в гаусовій моделі гіпотетично аварія відсутня, то в показниковій моделі вона хоч і рідка, але ймовірна: $\mathrm{P}(X \rangle x) \cong x^{-\alpha}$, оскільки $x \to \infty, 0 \langle \alpha$, де параметр α описує швидкість сповільнення "хвоста" розподілу. Асимптотична модель розподілу – гіперболічна, тому повільніша, ніж експоненційна і більша частина ймовірнісної міри може бути зосереджена в хвості розподілу.

розподілена за законом Стьюдента із n ступенями вільності. Якщо кожна із $n+m$ незалежних ВВ $\xi_1, \xi_2, ..., \xi_n, \xi_{n+1}, ..., \xi_{n+m}$ розподілена за законом $N(0, \sigma^2)$, то ВВ типу відношення

$$\eta = \frac{\frac{1}{n}\sum\limits_{i=1}^{n}\xi_i^2}{\frac{1}{m}\sum\limits_{i=n+1}^{n+m}\xi_i^2}$$

розподілене за законом Фішера із n і m ступенями вільності. Часто ВВ вдається перетворити в нормально розподілену за допомогою нескладних математичних перетворень. Наприклад, якщо неперервна ВВ ξ розподілена логнормально з параметрами m_X і σ_X^2, то ВВ $\eta = \ln\xi$ матиме нормальний двопараметричний логнормальний $LN(m_X, \sigma_X^2) = LN2$ розподіл. Якщо $LN2$-розподіл також не підпорядковується нормальному, то перед взяттям логарифма вводиться граничний параметр Σ, який перетворює цей розподіл в трипараметричний логнормальний $LN3$-розподіл [23], так щоб ВВ $Y = \ln(X - \Sigma)$ мала нормальний розподіл.

Аналіз експериментальних даних на предмет перевірки нормальності розподілу з точки зору оцінки дисперсії $\sigma_X^2 = D_X$ і зміщення m_X за даними вибірки вимірювань $\Omega_n = \{x_1, x_2, ..., x_n\}$ рекомендується здійснювати за таким алгоритмом. Методом моментів обчислити параметри вибірки:

$$\overline{X} = m_X = \frac{1}{n}\sum_{i=1}^{n} x_i$$

і

$$S^2 = \sigma_X^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{X})^2.$$

Обчислене середнє арифметичне вибірки є спроможною, незсуненою та ефективною оцінкою МС. Оцінку дисперсії вибірки здійснюють за формулою

$$S^2 = \sigma_X^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{X})^2,$$

яка зсунена, оскільки

$$M\left[S^2\right] = \left(1 - \frac{1}{n}\right)\sigma_X^2 \neq \sigma_X^2.$$

Зсув оцінки зумовлений множником Больцмана $1 - \frac{1}{n}$, однак для обсягів вибірки $n > 30$ його роль нівелюється. При $n < 30$ незсунуту оцінку треба обчислювати за формулою

$$D_X = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{X})^2.$$

Тому для нормально розподіленої вибірки, оцінка

$$S^2 = \sigma_X^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{X})^2$$

не ефективна. Однак, якщо МС m_X заздалегідь відоме, то оцінка

$$\sigma_X^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - m_X)^2$$

незсунута, спроможна і ефективна. Якість одержаних оцінок параметрів оцінюється на підставі міри точності оцінок, серед яких найбільш поширеною є середня квадратична похибка. Ця міра породжує відповідний критерій оптимальності оцінок – критерій мінімуму середньоквадратичної похибки. Для незміщених оцінок мірою їх точності є дисперсія, а критерієм оптимальності – критерій мінімуму дисперсії.

Але параметри положення m_X екстремуму кривої щільності $f_X(x)$ розподілу та її розмах σ_X визначають не фізичну величину, а характеризують вигляд функції $f_X(x)$ розподілу щільності ймовірностей її випадкових значень. Ці поняття зручні для нормального, експоненційного, трапецеїдального та інших розподілів, що використовують метод моментів, який справджується лише тоді, коли інтеграли відповідних обчислень не розбігаються. Більш універсальним є поняття про центр розподілу, який визначається як центр тяжіння розподілу або 50% квантиль. Центр тяжіння розподілу ВВ – це механічний аналог МС, якщо прийняти, що ймовірності значень – це маси точок.

У фізиці на основі моделі про центр тяжіння обґрунтовується, що довільне тіло у невизначеному стані намагається зайняти рівноважний стан. Точно так довільна ВВ за умови значної кількості вимірювань прямує до свого рівноважного (в розумінні середнього). Такий підхід вимагає лише існування моменту нульового порядку та параметра, що характеризує ширину розподілу. Для симетричного розподілу, центр тяжіння збігається із модою. Але на відміну від моди, поняття про центр тяжіння розподілу правомірне для усіх розподілів. Так, для розподілу Коші МС не існує, тоді як визначення центра тяжіння кривої розподілу для нього правомірне. Відсутня мода і для рівномірного розподілу.

Характеристикою флуктуації в хаотичних системах є дисперсія, що пов'язує між собою перший і другий початкові моменти та характеризує її інтенсивність, тому випадок $D_X = 0$ фізичного змісту не має. Наприклад, в електричному колі дисперсія, переважно виражає середню електричну потужність, що виділяється на активному електричному опорі змінною складовою електричної напруги чи струму. В змінному електричному колі параметру σ_X будуть відповідати покази вольтметра або амперметра, якщо шляхом включення в коло конденсатора, усунена стала складова електричного сигналу .

З точки зору теоретичного моделювання дисперсії в фізичній системі, важливо встановити статистично залежні чи незалежні фізичні величини, значення яких змінюються випадково. В безмежному $X \in [-\infty; +\infty]$ інтервалі розсіяння значень ВВ, дисперсія обчислюється як другий центральний момент:

$$D_X = \int_{-\infty}^{\infty} \left(x - \overline{X}\right)^2 C_X f_X(x) dx =$$
$$= \int_{-\infty}^{\infty} \left((x)^2 + \overline{X}^2 - 2x\overline{X}\right) C_X f_X(x) dx \tag{1}$$

і характеризує розкид значень відносно початку відліку на осі абсциси [2]. Для статистично незалежних ВВ, інтеграл (1) дорівнює

$$D_X = C_X \left\{ \int_{-\infty}^{\infty} (x)^2 f_X(x) dx + \overline{X}^2 \int_{-\infty}^{\infty} f_X(x) dx - \right.$$
$$\left. - 2\overline{X} \int_{-\infty}^{\infty} x f_X(x) dx \right\} = \overline{X^2} + \overline{X}^2 \int_{-\infty}^{\infty} C_X f_X(x) dx - 2\overline{X}^2. \tag{2}$$

[2] В теорії ймовірностей, дисперсія – це міра розсіяння від середнього, тоді як у математичній статистиці, вона характеризує ступінь розсіяння кількісних значень статистичної вибірки відносно середнього – МС квадрату відхилення ВВ від її МС.

Оскільки при цьому треба задовільнити умову нормування, то для статистично незалежних ВВ, рівняння дисперсії матиме вигляд:

$$D_X = \overline{X^2} - \overline{X}^2. \tag{3}$$

Тут за оцінку МС m_X ВВX взято середнє $\overline{X}$, як незміщена, обґрунтована спроможна оцінка, яка за наявності в розподілі дисперсії, ще є асимптотично нормальною оцінкою.

Статистичні середні $\overline{X}$ і $\overline{X^2}$ також відіграють важливу роль в фізичних системах. Наприклад, якщо в електричному колі середній квадрат $\overline{X^2}$ можна пов'язати із усередненим по часу середнім квадратом випадкової напруги або струму, то середній квадрат буде пропорційний середній електричній потужності, яка виділяється на активному опорі $\sqrt{\overline{x(t)^2}}$. В електротехніці, таке значення відоме як ефективне.

З іншого боку, в ймовірнісних моделях для фізичного моделювання треба враховувати, що в дійсності абсолютні величини невід'ємні і інтервал розсіяння їх випадкових значень насправді напівобмежений $X \in (0, +\infty)$. Оскільки, як свідчить досвід, щільність розподілу ймовірності $f_X(x)$ великих відхилень x від середнього m_X спадає пропорційно експоненті $\exp\left(-x^2\right)$ з екстремумом в точці $x = 0$, то для практичного застосування придатніший закон розподілу щільності ймовірностей, в якому шляхом віднімання від змінної середнього значення m_X ВВ X центрована:

$$g(X) = X - m_X. \tag{4}$$

Тоді на шкалі центрованої змінної, екстремум експоненти $f_X(x) \approx \exp\left(-(x - m_X)^2\right)$ зсувається в точку з координатою $x = m_X$.

Перехід на осі абсцис від шкали абсолютних значень X фізичної величини до шкали центрованих значень $X - m_X$ не деформує самої кривої Гауса, тому центровану змінну можна піддати нормуванню:

$$g(X) = U = \frac{X - m_X}{\sqrt{2}\sigma_X} = \sqrt{p_X}\,(X - m_X), \tag{5}$$

що дозволяє забезпечити в розрахунках безвимірність показника експоненти і нормаль $N(m_X, \sigma_X)$ спростити до стандартного з нульовим МС і одиничною дисперсією $N(0, 1)$. Він симетричний і для

ISSN 0372-400X. Укр. фіз. журн. 2018. Т. 63, № 7

нього точкою симетрії є координата моди, значення якої збігається із центром тяжіння – МС, а непарні центральні моменти дорівнюють нулю, хоч зворотне твердження не завжди справджується. Так, рівність нулю третього центрального моменту – це лише необхідна умова симетричності розподілу.

Нормування (5) не змінює координати максимуму експоненти $\exp\left(-\frac{(x-m_X)^2}{\sigma_X^2}\right)$, тому функція щільності $f_X(x)$ матиме аналітичний вигляд:

$$f_X(x) = \frac{C_X}{\sqrt{2\pi\sigma_X^2}} e^{-\left(\frac{x-m_X}{\sqrt{2}\sigma_X}\right)^2} =$$
$$= C_X\sqrt{\frac{p_X}{\pi}}\exp\left(-p_X(x-m_X)^2\right). \tag{6}$$

Напівобмеженому інтервалі $X \in (0, +\infty)$ розсіяння значень нормально розподіленої ВВ X, відповідає область $U \in (-\sqrt{p}m_X, +\infty)$ для центровано-нормованої ВВ:

$$x = 0 : u|_{x=0} = -\sqrt{p}m_X, \quad x = +\infty : u|_{x=+\infty} = +\infty \tag{7}$$

і інтеграл нормування для визначення сталої C_X запишеться у вигляді:

$$C_{X\in(0,+\infty)}\int\limits_0^\infty f_X(x)dx = C_{X\in(0,+\infty)}\times$$
$$\times\int\limits_0^\infty\sqrt{\frac{p_X}{\pi}}\exp\left(-p_X(x-m_X)^2\right)dx =$$
$$= \frac{C_{X\in(0,+\infty)}}{\sqrt{\pi}}\int\limits_{-\sqrt{p_X}m_X}^{\infty} e^{-u^2}du = 1. \tag{8}$$

З точки зору геометричної інтерпретації інтеграла, як площі під кривою $f_X(x)$, значення інтеграла $\int_0^\infty f_{X\in(0,+\infty)}(x)dx$ менше інтеграла $\int_{-\infty}^\infty f_{X\in(-\infty,+\infty)}(x)dx$, тобто інтервал $X \in \in (0, +\infty)$ – це усічений безмежний інтервал $(-\infty, +\infty)$, в якому усунуті від'ємні значення аргумента. Тому усічений нормальний розподіл застосовують для моделювання надійності фізико-технічних систем [24–27], фізичних процесів перенесення заряду в електронних приладах [28, 29], тощо, хоч насправді теорія випадкових процесів із обмеженою областю розсіяння значень ВВ була побудована давно, ще в роботах Ейнштейна і Смолуховського [30, 31]. Саме вони запропонували один із перших алгоритмів математичної обробки даних з метою оцінки функції розподілу і щільності ймовірностей емпіричних залежностей за вибіркою експериментальних даних. Зв'язок броунівського руху із гаусовим розподілом зручно продемонструвати за допомогою відомого у фізиці рівняння Фоккера–Планка (або рівняння Крамерса) [32, 33]:

$$\frac{\partial}{\partial t}f(t,x) = D\frac{\partial^2}{\partial x^2}f(t,x). \tag{9}$$

Воно описує процес руху броунівських частинок щільністю розподілу $f(t,x) = \frac{\partial}{\partial t}\mathrm{P}(x,t)$ під дією випадкових сил і має розв'язок

$$f(t,x) = \frac{C}{\sqrt{4\pi D}}\frac{1}{\sqrt{t}}\exp\left(-\frac{x^2}{4Dt}\right). \tag{10}$$

Розв'язок (10) формально нагадує функцію Гауса, якщо величину $2Dt$ трактувати як дисперсію $\sigma^2 = 2Dt$. Тут $\mathrm{P}(x,t)$ – ймовірність перебування частинки в точці з координатою x в момент часу t, де D – коефіцієнт дифузії.

Для усіченого інтервалу розсіяння значень ВВ X, інтеграли Гауса для обчислення статистичних середніх $\overline{X}$ і $\overline{X^2}$ не зводяться до квадратур аналітично, тобто не виражаються через елементарні функції. Для їх обчислення введена спеціальна неелементарна табульована функція $\mathrm{erf}(x)$ похибок [34]:

$$\mathrm{erf}\,(x) = \frac{2}{\sqrt{\pi}}\int\limits_0^x\exp\left(-\frac{u^2}{2}\right)du. \tag{11}$$

Якщо вибірка ВВ підпорядковується нормальному розподілу із стандартним відхиленням σ_X, то $\mathrm{erf}\left(\frac{m}{\sqrt{2}\sigma}\right)$ виражає ймовірність того, що її значення відхилиться від середнього не більше ніж на m_X. Функція похибок зв'язана з функцією стандартного $N(0,1)$розподілу через нормовану функцію Лапласа:

$$\Phi(x) = \frac{1}{\sqrt{2\pi}}\int\limits_{-\infty}^x\exp\left(-\frac{u^2}{2}\right)du. \tag{12}$$

За допомогою функції erf (x) за алгоритмом

$$\mathrm{P}\,(X \leqslant x_a) = \begin{cases} 0{,}5 - \mathrm{erf}\left(\frac{m_X - x_a}{\sigma_X}\right) & \text{для } x_a \leqslant m_X \\ 0{,}5 + \mathrm{erf}\left(\frac{x_a - m_X}{\sigma_X}\right) & \text{для } x_a \geqslant m_X \end{cases} \tag{13}$$

розроблений так званий стандарт "3σ" довірчих інтервалів і сформульовані правила:

$$\begin{aligned} &\mathrm{P}\,(|X - m_X| \leqslant \sigma_X) = 2\,\mathrm{erf}(1) = 0{,}685, \\ &\mathrm{P}\,(|X - m_X| \leqslant \sigma_X) = 2\,\mathrm{erf}(2) = 0{,}991, \\ &\mathrm{P}\,(|X - m_X| \leqslant \sigma_X) = 2\,\mathrm{erf}(3) = 0{,}998, \end{aligned} \tag{14}$$

які дозволяють із заданою довірчою ймовірністю проектувати параметри фізичної моделі із випадковими флуктуаціями.

Важливість функцій Гауса в статистичному моделюванні фізичних систем підтверджена відомим у фізиці ймовірнісним розподілом Максвелла. Незважаючи на те, що декартові складові вектора швидкості хаотичного руху частинок рівноважної системи підлягають нормальному розподілу, розподіл абсолютних швидкостей частинок підпорядковується розподілу Максвелла:

$$\begin{aligned} f(\vartheta_z) &= \sqrt{\frac{1}{\pi}\frac{m}{2k_\mathrm{B}T}}\exp\left(-\frac{m\vartheta_z^2}{2k_\mathrm{B}T}\right) = \\ &= \sqrt{\frac{p_{\vartheta Z}}{\pi}}\exp\left(-p_{\vartheta Z}\vartheta_z^2\right) \Rightarrow\; p_{\vartheta Z} = \frac{m}{2k_\mathrm{B}T}. \end{aligned} \tag{15}$$

Отже, прирівнявши між собою функцію (15) з функцією Гауса, одержимо, що

$$\frac{1}{2\sigma_{\vartheta_Z}^2} = \frac{m}{2k_\mathrm{B}T} \Rightarrow \sigma_{\vartheta_Z} = \frac{k_\mathrm{B}T}{m}.$$

Таким чином, середньо-квадратичне відхилення (СКВ) проекції швидкості у відповідному напрямку зростає лінійно абсолютній температурі та обернено пропорційне масі частинки.

Ймовірнісним аналогом центра мас

$$y_c = \frac{\sum y_k m_k}{\sum m_k}$$

і моменту інерції відносно нього

$$I_{m_X} = \sum\,(y_k - a)^2 m_k$$

є МС і дисперсія. Тому зручно переносити аналогію оптимізації закономірностей динаміки обертання системи матеріальних точок масами m_k на невагомому стрижні на алгоритм мінімізації флуктуаційних процесів (дисперсії) у фізичній системі при русі її в напрямку рівноважного стану. Згідно із даною аналогією, в задачі оптимізації динаміки обертового руху системи матеріальних точок навколо вибраної осі обертання означає, що треба знайти оптимальне значення параметра a її просторового розташування, для якого момент інерції (а отже дисперсія флуктуацій) мінімальний. В обертовій системі кожна з точок розташована вздовж осі на відстані y_k від одного з її кінців, взятого за початок відліку, перпендикулярного до осі обертання.

У процесі обертання радіус-вектор y_k описує круг площею $\pi\, y_k^2$. Тому обчислимо щільність розподілу площі круга $f_{S_k}(s_k)$ з параметрами цього m_S, σ_S, якщо значення y_k розсіяні за нормальним законом $N(m_Y, \sigma_Y)$:

$$y_k = \frac{1}{\sqrt{\pi}}\sqrt{s_k}, \quad \left|\frac{dy_k}{ds_k}\right| = \frac{1}{2\sqrt{\pi}}\frac{1}{\sqrt{s_k}}. \tag{16}$$

Обернена функція $s_k = \pi\, y_k^2$ двозначна, тому із рівності $f_{Y_k}(y_k)dy_k = f_{S_k}(s_k)ds_k$ одержимо, що

$$f_{S_k}(s_k) = f_{Y_k}(y_k)\left|\frac{dy_k}{ds_k}\right| = f_{Y_k}(y_k)\frac{1}{2\sqrt{\pi\, s_k}}. \tag{17}$$

Модуль від відношення похідних взятий тому, щоб усунути вплив зміни знака похідної, тому

$$\begin{aligned} f_{S_k}(s_k) &= \frac{\sqrt{p_{S_k}}}{2\pi\sqrt{s_k}}\Big\{\exp\left(-p_Y(\sqrt{s_k} - \sqrt{\pi}m_Y)^2\right) + \\ &+ \exp\left(-p_Y(\sqrt{s_k} + \sqrt{\pi}m_Y)^2\right)\Big\}, \end{aligned} \tag{18}$$

де

$$p_Y = \frac{1}{2\pi\sigma_Y^2}, \quad p_{S_k} = \frac{1}{2\sigma_{S_k}^2}.$$

Площа набуває невід'ємних значень, тому статистичне середнє m_{S_k} буде дорівнювати

$$\begin{aligned} &m_{S_k} = \overline{S_k} = \\ &= \frac{\int\limits_0^{+\infty} s_k\left\{e^{-p_Y(\sqrt{s_k}-\sqrt{\pi}m_Y)^2} + e^{-p_Y(\sqrt{s_k}+\sqrt{\pi}m_Y)^2}\right\}ds_k}{\int\limits_0^{+\infty}\left\{e^{-p_Y(\sqrt{s_k}-\sqrt{\pi}m_Y)^2} + e^{-p_Y(\sqrt{s_k}+\sqrt{\pi}m_Y)^2}\right\}ds_k} = \\ &= \frac{\sqrt{2}}{\pi\sigma_Y}\left\{\frac{\sqrt{\pi}}{4p_Y\sqrt{p_Y}} + \pi\, m_Y^2\frac{\sqrt{\pi}}{2\sqrt{p_Y}}\right\} = \pi\left(\sigma_Y^2 + m_Y^2\right). \end{aligned} \tag{19}$$

Обчислимо аналогічно статистичне середнє $\overline{S_k^2}$:

$$\overline{S_k^2} = \\ = \frac{\int\limits_0^{+\infty} s_k^2 \left\{ e^{-p_Y(\sqrt{s_k}-\sqrt{\pi}m_Y)^2} + e^{-p_Y(\sqrt{s_k}+\sqrt{\pi}m_Y)^2} \right\} ds_k}{\int\limits_0^{+\infty} \left\{ e^{-p_Y(\sqrt{s_k}-\sqrt{\pi}m_Y)^2} + e^{-p_Y(\sqrt{s_k}+\sqrt{\pi}m_Y)^2} \right\} ds_k} = \\ = \pi^2 \left(3\sigma_Y^4 + m_Y^4 + 6\sigma_Y^2 m_Y^2\right). \quad (20)$$

Таким чином, СКВ буде дорівнювати

$$\sigma_{S_k} = \pi(3\sigma_Y^4 + m_Y^4 + 6\sigma_Y^2 m_Y^2 - m_Y^4 - \\ - \sigma_Y^4 - 2\sigma_Y^2 m_Y^2)^{1/2} = \sqrt{2}\pi\sigma_Y(\sigma_Y^2 + 2m_Y^2)^{1/2}. \quad (21)$$

Оскільки СКВ σ_{S_k} і математичне сподівання m_{S_k} вже не можуть змінюватись незалежно один від одного, а залежні від статистично незалежних m_Y, σ_Y, то лише оптимізацією значень m_Y, σ_Y гаусових ВВ представляється можливим оптимізувати стан із мінімальним СКВ.

І на завершення відзначимо таке. Для моделювання фізичних систем із випадковими впливами, актуальним питанням є обсяг вибірки експериментальних результатів, який взятий для проведення статистичного аналізу. За теоремою Ляпунова [35] не можна використовувати ймовірнісні обчислення, якщо обсяг вибірки $n \leqslant 20 - 30$. При малому обсязі вибірки $n < 30$ оцінка СКВ розподілу ненадійна, тому застосовують розподіл Стьюдента. За його допомогою обчислюють ймовірність відхилення $\overline{X}$ від m_X, не перевищує деяку величину $\sigma_{\overline{X}}$:

$$p\left(-\sigma_{\overline{X}} < \overline{X} - m_X < +\sigma_{\overline{X}}\right) = \\ = \frac{1}{\sigma_{\overline{X}}}\sqrt{\frac{2}{\pi}} \int\limits_{m_X-\sigma_{\overline{X}}}^{m_X+\sigma_{\overline{X}}} \exp\left(-\frac{\left(\overline{X} - m_X\right)^2}{2\sigma_{\overline{X}}^2}\right) d\overline{X}. \quad (22)$$

Тоді відхилення середнього арифметичного від дійсного значення вимірюваної величини не буде перевищувати

$$\Delta p = t_p S_{\overline{X}} = t_p \frac{S_X}{\sqrt{n}}. \quad (23)$$

При цьому точність наближеної рівності $\overline{X} \approx m_X$ зростає із збільшенням n. Однак, цей висновок ще не означає, що для збільшення точності кінцевих результатів треба намагатись в першу чергу збільшити точність окремих вимірювань. Коефіцієнти Стьюдента t_p табульовані для різних значень ймовірностей P і ступеней вільності $k = n - 1$. Граничним випадком розподілу Стьюдента при $k = 1$ є розподіл Коші (відсутнє СКВ і дисперсія, оскільки інтеграл розходиться і не може бути зроблена ефективна оцінка центра розподілу, тому для визначення центра використовують медіану), а при $k \to \infty$ – розподіл Гауса. Розподіл Коші, або Лорентца, або Вігнера–Брейта відіграє важливу роль у фізичному моделюванні. Незважаючи на те, що у відомих законах пропорційності Ома, Гука чи інших при достатньому обсязі незалежних вимірювань спаду напруги U_R на активному опорі R та сили струму крізь нього, чи деформації тіла Δx та прикладеної до нього сили F_X у законі Гука, що забезпечує нормальний розподіл їх флуктуацій, розподіл випадкових змін коефіцієнтів пропорційності $R = \frac{U_R}{I_R}$, $k = \frac{F_X}{\Delta x}$ підлягатиме розподілу Коші, на що не завжди звертається достатня увага.

3. Висновки

З точки зору досягнення максимальної завершеності в побудові статистичного портрету фізичної моделі, викладені вище закономірності можуть слугувати як перші кроки для встановлення характеру розподілу випадкових значень величин, одержаних в результаті експериментальних вимірювань. Але вони потрібні.

Оцінка закону розподілу з заданою довірчою ймовірністю вимагає розширення спектра застосування статистично-ймовірнісних досліджень, як ступеня перекриття смуг розподілів [36, 37], параметра відстані між ними [38], моделювання методами узагальнених [39, 40] і змішаних [41, 42] розподілів, тощо. Саме комплексність опрацювання експериментальних даних, дозволило сьогодні успішно реалізувати такі високотехнологічні методи візуалізації складних об'єктів як атомно-силова і тунельна мікроскопія, ядерно-магнітна резонансна спектроскопія [43, 44], тощо.

Однак, в читача не повинна скластись думка про те, що випадкові закономірності фізичних величин обов'язково повинні підпорядковуватись нормальному розподілу. Відомі фізичні процеси, в яких випадкові закономірності підтверджують правомірність інших законів, наприклад, експоненціального в радіоактивному розпаді [45] (але якщо експеримент повторити n разів і кожного разу визначати сталу розпаду, то при $n \to \infty$ її статистика прямуватиме до гаусової), показникового в явищах синергетики та фрактальної динаміки [46, 47], логнормального в прогнозах оптимальної локалізації корисних копалин за геофізичними спостереженнями [48], тощо. Але це тема іншої роботи.

1. G. Bohm, G. Zech. *Introduction to Statistics and Data Analysis for Physicists* (Verlag Deutsches Electronen-Synchrotron).
2. J.K. Patel, C.B. Read. *Handbook of the Normal Distribution* (Dekker, 1982) [ISBN: 0-8247-1541-1].
3. D.S. Sivia, J. Skilling. *Data Analysis: A Bayesian Tutorial* (Oxford University Press, 2006) [ISBN: 978-0-19-856831-5].
4. М.В. Фок. Разделение сложных спектров на индивидуальные полосы при помощи метода Аленцева. *Труды ФИАН СССР* **63**, 3 (1972).
5. А.С. Гусев. *Вероятностные методы в механике машин и конструкций* (МГТУ им. Баумана, 2009) [ISBN: 978-5-7038-3160-1].
6. E. Wichmann. *Quantum Physics.Berkeley Physics Course* (McGraw-Hill Book Company, 2010), Vol. 4 [ISBN: 0070048614].
7. F. Reif. *Fundamentals Statistics and Thermal Physics* (Waveland Press, Inc. 1965) [ISBN: 1-57766-612-7].
8. http://www.d.umn.edu/∼psiders/courses/chem5650/ gaussviewtutorial /tutorial.html
9. В.М. Золотарев. *Одномерные устойчивые распределения* (Наука, 1983).
10. П.В. Короленко, М.С. Маганова. *Основы статистических методов в оптике* (Университетская книга, 2010) [ISBN: 978-5-91304-126-5].
11. Ф. Барду, Ж.-Ф. Бушо, А. Аспе, К. Коэн-Таннуджи. *Статистика Леви и лазерное охлаждение. Как редкие события останавливают атомы.* (Физматлит, 2006) [ISBN: 978-5-9221-0670-2].
12. О.Н. Маслов.*Устойчивые распределения и их применение в радиотехнике* (Радио и связь, 1994) [ISBN: 5-256-01187-1].
13. G. Samorodnitsky, M.S. Taqqu. *Stable Non-Gaussian Random Processes* (Chapman and Hall, 1994) [ISBN: 9780412051715].
14. S. Holmes. *Sums of Random Variables: Statistics 116* (Stanford, 1998).
15. D. Titterington, A. Smith, U. Makov. *Statistical Analysis of Finite Mixture Distributions* (Wiley, 1985) [ISBN: 0-471-90763-4].
16. N.A. Carreira-Perpiñán. Mode-finding for mixture of Gaussian distributions. *IEEE Trans. On Pattern Analys. and Mach. Intell* **22** (11), 1318 (2000).
17. Я. АФомин, Г.Р. Тарловский. *Статистическая теория распознавания образов* (Радио и связь, 1986).
18. Цж. А. Хартиган. *Задачи, связанные с функциями распределения в кластер-анализе. В кн. Классификация и кластер* Пер. с англ. П.Кольцова. Под ред. Ю. Журавлёва (Мир, 1980), с. 42.
19. N.N. Aprausheva, I.A. Gorlach, A.A. Zhelnin, S.V. Sorokin. An experiment on automated statistical recognition of clouds. *J. Computational Mathematics and Mathematical Physics* **38**, 101715 (1998).
20. М. Подригало, О. Ісакова, А. Коробко. Новий спосіб оцінювання збігу результатів теоретичних та експериментальних досліджень. *Метрологія та прилади* **5**, 48 (2017).
21. Д.З. Шматко, О.В. Кочнева. Обчислення вірогідності визначення дефектів деталей методом неруйнівного контролю. *Математичне моделювання* **1** (36), 51 (2017).
22. П. Кособуцький. Стосовно моделювання математичного сподівання і дисперсії вибірок гаусово розподілених випадкових величин. *УФЖ* **61**, 823 (2017).
23. J.R. Stedinger, R.M. Vogel, E. Foufoula-Georgiou. *Frequency anlysis of extreme events.* In Handbook of Applied Hydrology. Edited by D. Maidment (McGraw-Hill, 1993), Chapter 18, Part 3 [ISBN: 978-0-07-171177-7].
24. A. Hald. Maximum likelihood estimation of the parameters of a normal distribution which is truncated at a known point. *Scandinavian Actuarial Journal* **1949** (1), 119 (1949).
25. B. Gnedenko, I. Ushakov. *Probabilistic Reliability Engineering* (A Wiley-interscience Publication John Wiey & Sons Inc., 1995) [ISBN: 0-471-30502-2].
26. В.Р. Матвиевский. *Надежность технических систем* (МГИЭМ, 2002) [ISBN: 5-230-22198-4].
27. A. O'Connor , M. Modarres, A. Mosleh. *Probability Distributions Used in Reliability Engineering* (CRE University of Maryland, College Park, 2016) [ISBN: 978-0-9966468].
28. J. Xinzhang, L. Tao. An empirical formula for yield estimation from singly truncated performance data of qualified semiconductor devices. *J. Semicond.* **33** (12), 125008-1 (2012).
29. K. Gu, X. Jia, H. You, T. Liang. The yield estimation of semiconductor products based on truncated samples. *Int. J. Metrol. Qual. Eng.* **4**, 215 (2013).
30. А. Эйнштейн. *О движении взвешенных в покоящейся жидкости частиц, требуемом молекулярно-кинетической теории теплоты* (ОНТИ-ГРОТЛ, 1936).
31. А. Эйнштейн, М. Смолуховский. *Броуновское движение* (ОНТИ-ГРОТЛ, 1936).
32. H. Risken, T. Frank. *The Fokker-Planck quation: methods of solutions and applications* (Springer, 1989) [ISBN: 0-387-50498-2].
33. I. Gihman, A. Skorochod. *The Theory of Statistic Processes I* (Springer, 2004) [ISBN: 3-540-20284-6].
34. E. Ng, M. Geller. A table of integrals of the error functions. *J. Research of the National Bureau of Standerds-B Mathematical Sciences B* **73** (1), 1 (1969).
35. Ш. Сулаберидзе. *Методы анализа и обработки измеренных значений величин* (Балт. ГТУ, 2013) [ISBN: 978-5-85546-742-0].

36. E. Inman, E. Bradley. The overlapping coefficient as a measure of agreement between probability distributions and point estimation of the overlap of two normal densities. *Commun Statist Theory Methods* **18**, 3851 (1989).
37. Z. Karian, E. Dudewicz. *Handbook of Fitting Statistical Distributions with R* (Taylor and Francis, 2011) [ISBN: 978-1-58488-711-9].
38. D.C. Dovson, B.V. Landau. The Frechet distance between multivariate normal distributions. *J. Multivariate Analysis* **12**, 450 (1982).
39. M. Varanasi, B. Aazhang. Parametric generalized Gaussian density estimation. *J. Acoustical Society of America* **86** (4), 1404 (1989).
40. S. Nadarajah. A generalized normal distribution. *J. Appl. Stat.* **32** (7), 685 (2005).
41. J. Behboodian. On a Mixture of normal distributions. *Biometrica* **57** (1), 215 (1970).
42. Н.Н. Апраушева, С.В. Сорокин. *Заметки о гауссовых смесях* (ВЦ РАН, 2015).
43. B. Voigtlander. *Scanning Probe Microscopy. Atomic Force Microscopy and Acanning Tunneling Microscopy* (Springer, 2015) [ISBN: 978-3-662-45239-4].
44. Н.М. Сергеев. *Спектроскопия ЯМР* (Изд-во МГУ, 1981).
45. E. Rutherford, F. Soddy. A comparative study of the radioactivity of radium and thorium. *Philosophical Magazine.* **5** (28), 445 (1903).
46. Т. Ахромеева, С. Курдюмов, Г. Малинецкий, А. Самарский. *Структуры и хаос в нелинейных средах* (Физматлит, 2007) [ISBN: 978-5-9221-0887-4].
47. M. Schroder. *Fractals, Chaos, Power Laws: Minutes from an Infinite Paradise* (W.H. Freeman and Company, 1991) [ISBN: 0-7167-2136-8].
48. А.А. Никитин. *Статистические методы выделения геофизических аномалий* (Недра, 1979).

Одержано 14.03.18

P. Kosobutskyy

OPTIMAL REGULARITIES
OF THE NORMAL DISTRIBUTION
FOR ESTIMATING THE SAMPLE STATISTICS
OF THE RESULTS OF A PHYSICAL EXPERIMENT

S u m m a r y

Basic probabilistic principles for the formation of the normal distribution for random fluctuations of physical quantities under the action of independent random factors on the physical system have been formulated. The emphasis is made on the integrated approach to the probabilistic statistical analysis of a sample of experimental results.