УДК 519.2 https://doi.org/10.15407/itm2024.02.076

Е. Г. ГЛАДКИЙ, В. І. ПЕРЛИК

СПОСОБИ ПОБУДОВИ РОЗПОДІЛУ МАКСИМАЛЬНОЇ ВІДСТАНІ МІЖ ВИПАДКОВИМИ НОРМАЛЬНИМИ ТОЧКАМИ НА ПЛОЩИНІ

Державне підприємство «Конструкторське бюро «Південне» ім. М. К. Янгеля»;
вул. Криворізька,3, м. Дніпро, Україна; e-mail: edgladky@gmail.com

У багатьох практичних задачах виникає потреба у побудові розподілу максимальної відстані між випадковими точками на площині. У літературі, зазвичай, розглядається випадок великої кількості таких точок, для якого визначається асимптотичний розподіл. В статті розглянуто задачу побудови розподілу максимальної відстані між незначною кількістю випадкових нормальних точок на площині, у яких координати є незалежними випадковими величинами, що підпорядковуються стандартному нормальному розподілу. Як базовий, розглянуто окремий випадок трьох випадкових нормальних точок на площині. Для нього досліджено три напрямки побудови розподілу максимальної відстані між точками.

Згідно із першим, функція розподілу будується, виходячи з геометричних міркувань. Для цього розглянуто геометричні місця розташування трьох точок, виходячи із умови, щоб максимальна відстань між ними не перевищила певної величини. Місцезнаходження третьої точки на площині визначається відносно двох інших – самої лівої точки та самої нижньої. Побудова функції розподілу в цьому випадку передбачає послідовне розв’язання кількох інтегралів, для чого використовуються чисельні методи. Отримані результати мають гарний збіг із результатами статистичного моделювання.

Дослідження відстаней між парами випадкових нормальних точок на площині складає основу другого напрямку. Окремо відстані між кожною парою випадкових нормальних точок підпорядковуються одновимірним розподілам Релея, проте у сукупності такі випадкові величини виявляються корельованими, адже визначаються через одні і ті ж координати точок. Побудовано спільний розподіл для квадратів відстаней між трьома точками, для чого використано тривимірний розподіл Морана–Даунтона. З його допомогою отримано функцію розподілу квадрату максимальної відстані між трьома випадковими нормальними точками, що тотожна розподілу максимальної відстані. З’ясовано, що в області невеликих значень вона недооцінює реальну імовірність того, що максимальна відстань не перевищить певну величину. Для великих значень відстаней зазначені імовірності збігаються.

В межах третього напрямку розглянуто використання розподілу Райса (узагальнення розподілу Релея) для апроксимації невідомого розподілу максимальної відстані між трьома випадковими нормальними точками на площині. Визначено параметри розподілу Райса, виходячи із методу найменших квадратів, і отримано непогану збіжність із результатами статистичного моделювання.

Результати, які отримано для трьох випадкових нормальних точок, узагальнено на випадок більшої (до n = 30) кількості точок. Показано, що третій напрямок у цьому випадку є найбільш ефективним.

Ключові слова: *випадкові точки на площині, максимальна відстань між точками, функція розподілу, розподіл Морана–Даунтона, розподіл Райса.*

Many practical problems call for constructing the maximum interpoint distance distribution for random pints in a plane. In the literature, the case of a great number of points is considered, for which an asymptotic distribution is determined. This paper addresses the problem of constructing the maximum interpoint distance distribution for a small number of random points in a plane whose coordinates are independent random quantities that obey the standard normal distribution. The special case of three random points in a plane is considered as the basic one, for which three ways to construct the maximum interpoint distance distribution are studied.

The first way is to construct the distribution function from geometrical considerations. To do this, the loci of three points are considered from the condition that the maximum distance between them shall not exceed a certain value. The position of the third point in the plane is determined relative to the two other points: the leftmost and the lowermost one. In this case, the construction of the distribution function involves the successive evaluation of several integrals using numerical methods. The obtained results are in good agreement with those of statistical simulation.

The second way is based on studying the distance between pairs of random normal points in a plane. Taken separately, the distances between each pair of random normal points obey one-dimensional Rayleigh distributions, but in the aggregate they prove to be correlated because they are determined from the same pint coordinates. A joint distribution of the squared distances between three points is constructed using the three-dimensional Moran-Downton distribution. Using it, a distribution function of the squared maximum distance between three random normal points, which is identical with the maximum interpoint distance distribution, is obtained. It is found that for small values it underestimates the actual probability of the maximum distance not exceeding a certain value. For great distance values, the above probabilities coincide.

The third way uses the Rice distribution (a generalization of the Rayleigh distribution) to approximate the unknown maximum interpoint distance distribution for three random normal points in a plane. The Rice distribution parameters found by the least-squares method are in good agreement with those obtained by statistical simulation.

The results for three random normal points are generalized to a greater number of points (up to 30). It is shown that in this case the third way is most efficient.

Keywords: *random points in a plane, maximum interpoint distance, distribution function, Moran–Downton distribution, Rice distribution.*

Вступ. У багатьох практичних задачах (теорія стрільби, контроль якості, астрономія, медицина та інш.) виникає потреба визначити імовірнісні характеристики, насамперед, функцію розподілу та функцію щільності, максимальної відстані для деякої множини випадкових точок (в англомовній літературі отримала назву *maximum interpoint distance*).

Простий розв'язок отримано в одновимірному випадку, коли певна кількість випадкових точок з відомим законом розподілу «кидається» на пряму. Нехай $X_1,\ldots, X_n$ взаємно незалежні випадкові величини з функцією розподілу $F(x)$ і щільністю розподілу $f(x)$, які будемо розглядати як вибірку (фактично це випадкові точки, «кинуті» на пряму). У цьому випадку максимальна відстань між випадковими точками – розмах W_n ($W_n \geq 0$) – являє собою різницю між найбільшим і найменшим значеннями вибірки

$$W_n = X_{(n)} - X_{(1)},$$

де $X_{(1)}$, $X_{(n)}$ – відповідно мінімальне та максимальне значення (порядкові статистики). Функція розподілу розмаху визначається як [2, 3]

$$F_n(w) = \Pr(W_n \leq w) = n \int_{-\infty}^{+\infty} \left[F(x+w) - F(x)\right]^{n-1} f(x)dx\,. \qquad (1)$$

Для її отримання розглядається ймовірність того, що за відомого x рівно одна з випадкових величин (ВВ) X_i належатиме околу (x; x+dx) (таких варіантів буде n), а решта (n-1) ВВ опиняться в інтервалі (x; x+w). Зазначена ймовірність має вигляд

$$nf(x)dx\left[F(x+w) - F(x)\right]^{n-1}.$$

Шляхом інтегрування останнього виразу за x отримуємо функцію розподілу розмаху (1).

Для стандартного нормального розподілу ВВ $X_1,\ldots, X_n$ з математичним сподіванням рівним нулю та середнім квадратичним відхиленням, що дорівнює одиниці, відповідно матимемо

$$F_n(w) = n \int_{-\infty}^{+\infty} \left[\Phi\left(x+w\right) - \Phi\left(x\right)\right]^{n-1} \frac{1}{\sqrt{2®}} \exp\left\{-\frac{1}{2}x^2\right\} dx\,,$$

де Ф(•) – функція Лапласа.

Отже, ймовірність того, що максимальна відстань між n випадковими точками, «кинутими» на пряму, не перевищить d, буде дорівнювати $F_n(d)$.

Щільність розподілу ВВ W_n визначається шляхом диференціювання [3]

$$f_n(w) = \frac{\partial F_n(w)}{\partial w} = n(n-1) \int_{-\infty}^{+\infty} f(x)\left[F(x+w) - F(x)\right]^{n-2} f(x+w)dx\,.$$

Розглянемо тепер двовимірний випадок. Тут у більшості випадків авторів (дивись, наприклад, [5]) цікавить питання щодо граничного (асимптотичного) розподілу максимальної відстані між n випадковими точками у разі $n \to \infty$

$$D_n^{(2)} = \max_{1 \leq i \leq j \leq n} \left|P_i - P_j\right|,$$

де P_i, P_j – точки на площині; $|\bullet|$ – Евклідова відстань між випадковими точками на площині. Досить часто, під час розгляду такої задачі, випадкові точки вважаються рівномірно розподіленими всередині деякої обмеженої області, наприклад, одиничного кола або квадрата. Розглядаються також точки, розподіли яких мають необмежені області (наприклад, нормальні). Для них, зокрема, показано [9], що асимптотичні розподіли $D_n^{(2)}$ відповідають розподілам екстремальних значень, наприклад, подвійному експоненціальному. У той же час у багатьох практичних задачах дослідникам доводиться мати справу з невеликою кількістю випадкових точок n і відповідно бажано мати задовільний математичний апарат для дослідження ймовірнісних характеристик максимальної відстані між випадковими точками у цьому випадку. Подібний запит досить часто зустрічається у мережі InterNet. Метою статті є формування рекомендацій щодо розподілу максимальної відстані між випадковими точками на площині для розв’язання прикладних задач.

1. Постановка задачі. Розглянемо n розташованих на площині точок, координати яких (X_i, Y_i) $i=\overline{1,n}$ є взаємно незалежними ВВ, що підпорядковуються стандартному нормальному розподілу $N(0,1)$. Надалі такі точки називатимемо випадковими нормальними точками на площині. Нажаль, у загальному випадку для певної кількості (більше двох) випадкових нормальних точок на площині одержати такий розподіл проблематично, тому спочатку розглянемо окремий випадок для трьох випадкових нормальних точок ($n = 3$) і спробуємо узагальнити отримані результати на більші значення n. Тобто визначимо ймовірність того, що максимальна відстань між трьома випадковими нормальними точками на площині не перевищить величину d. Розв’язання такої задачі може бути отримано кількома способами.

2. Геометричний спосіб отримання функції розподілу максимальної відстані між трьома випадковими нормальними точками на площині. Застосуємо прийом, аналогічний до того, за допомогою якого було отримано функцію розподілу розмаху в одновимірному випадку (дивись формулу (1)). Розглянемо спільну ймовірність того, що двовимірна величина (X, Y) належить певній елементарній області $(x; x+dx)\times(y; y+dy)$. Це буде точка (x_1, y_1). Кількість варіантів, з урахуванням того, що ВВ X і Y є незалежними, буде дорівнювати 3×3. Вважатимемо, що це буде сама ліва точка (гранична точка в горизонтальному напрямку, вздовж осі ОX). Тепер розглянемо другу точку (x_2, y_2), яка буде найнижчою. Для цього вона повинна розташуватися нижче і правіше від уже наявної точки (x_1, y_1). При цьому відстань від точки (x_1, y_1) до (x_2, y_2) не повинна перевищувати d. Тобто вона має перебувати у правій нижній чверті кола радіусу d із центром у точці (x_1, y_1) (рис. 1). Геометричним місцем розташування третьої точки, виходячи із умови неперевищення відстані між нею та іншими точками величини d, буде зафарбована область, яку показано на рис. 2. Ця область обмежується вертикаллю, що проходить через точку **1**(x_1, y_1), горизонталлю, що проходить через точку **2**(x_2, y_2), а також дугами двох кіл радіусу d, які побудовані із точок **1** і **2** відповідно й перетинаються у точці **C**. Точки **A** і **B** мають координати: $\mathbf{A}(x_1, y_2+\sqrt{d^2-(x_1-x_2)^2})$; $\mathbf{B}(x_1+\sqrt{d^2-(y_2-y_1)^2}, y_2)$. Координати точки **C**

визначаються як перетин двох кіл або перетин одного з кіл і перпендикуляра, що проходить через середину відрізка **12**.

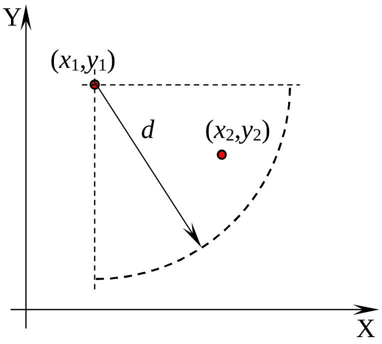

Рис. 1 – Область розташування точки (x_2,y_2)

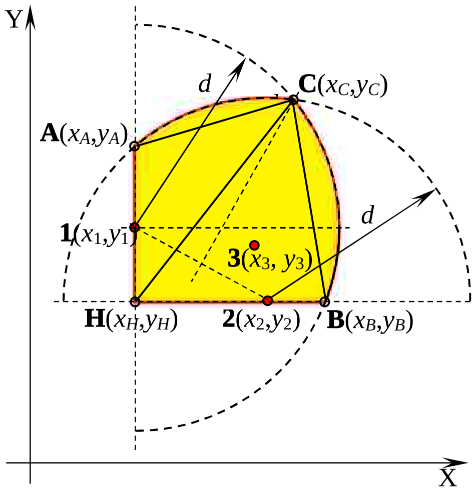

Рис. 2 – Область для третьої точки

Таким чином, необхідно визначити ймовірність того, що третя точка потрапить у вказану область. Позначимо її $P_3(d;x_1,y_1,x_2,y_2)$. Ця ймовірність залежить від координат вже наявних першої та другої точок. Тобто це буде умовна ймовірність того, що третя випадкова нормальна точка потрапить у зафарбовану область, за умови, що перша та друга точки мають координати (x_1, y_1) і (x_2, y_2). У результаті ймовірність того, що відстань між трьома нормальними точками на площині не перевищить d, визначиться шляхом послідовного інтегрування:

$$P_1(d;x_1,y_1)=\int_{x_1}^{x_1+d}\left[\int_{y_1-\sqrt{d^2-(x_2-x_1)^2}}^{y_1}P_3(d;x_1,y_1,x_2,y_2)N(y_2)dy_2\right]N(x_2)dx_2\ ,$$

$$P(d)=3^2\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}P_1(d,x_1,y_1)N(x_1)N(y_1)dx_1dy_1\ , \qquad (2)$$

де $N(\bullet)$ – щільність стандартного розподілу.

Для визначення $P_3(d;x_1,y_1,x_2,y_2)$ розіб'ємо зафарбовану на рис. 2 область на два трикутники ΔHAC і ΔHCB та два сегменти AC і CB, отже

$$P_3(d;x_1,y_1,x_2,y_2)=P_{\Delta HAC}+P_{\Delta HCB}+P_{segAC}+P_{segCB}\ ,$$

де $P_{\Delta HAC}$, $P_{\Delta HCB}$, P_{segAC}, P_{segCB} – відповідно ймовірності потрапляння точки (x_3, y_3) у зазначені фігури. Співвідношення для їх визначення мають вигляд:

$$P_{\Delta HAC}=\int_{x_1}^{x_C}\left[\Phi\left(\frac{y_C-y_A}{x_C-x_A}(x_3-x_A)+y_A\right)-\Phi\left(\frac{y_C-y_H}{x_C-x_H}(x_3-x_H)+y_H\right)\right]N(x_3)dx_3\ ;$$

$$P_{\Delta HCB} = \int_{y_2}^{y_C} \left[\Phi\left(\frac{x_C - x_B}{y_C - y_B}(y_3 - y_B) + x_B \right) - \Phi\left(\frac{x_C - x_H}{y_C - y_H}(y_3 - y_H) + x_H \right) \right] N(x_3) dx_3 \,;$$

$$P_{segAC} = \int_{x_1}^{x_C} \left[\Phi\left(y_2 + \sqrt{d^2 - (x_3 - x_2)^2} \right) - \Phi\left(\frac{y_C - y_A}{x_C - x_A}(x_3 - x_A) + y_A \right) \right] N(x_3) dx_3 \,;$$

$$P_{segCB} = \int_{y_2}^{y_3} \left[\Phi\left(x_1 + \sqrt{d^2 - (x_3 - y_1)^2} \right) - \Phi\left(\frac{x_C - x_B}{y_C - y_B}(y_3 - y_B) + x_B \right) \right] N(x_3) dx_3 \,.$$

Отримана згідно з формулою (2) ймовірність $P(d)$ являє собою функцію розподілу максимальної відстані між трьома випадковими нормальними точками на площині. Для її визначення потрібні чисельні процедури. Результати розрахунку $P(d)$ для значень d із інтервалу [0,5; 4,75] із кроком 0,25, отримані з використанням пакета математичних обчислень MathCAD, наведено в таблиці 1. Для порівняння у таблиці також показано результати статистичного моделювання (симуляції). Кількість експериментів тут і нижче, де наводяться результати симуляції, становить $N = 5\times10^5$. З результатів таблиці 1 випливає досить непоганий збіг між значеннями, отриманими з використанням (2), і результатами симуляції.

Таблиця 1 – Імовірності $P(d)$, отримані згідно з (2)

d	$P(d)$	Монте-Карло
0,50	0,0030	0,0029
0,75	0,0141	0,0148
1,00	0,0405	0,0392
1,25	0,0879	0,0855
1,50	0,1584	0,1542
1,75	0,2496	0,2465
2,00	0,3555	0,3483
2,25	0,4675	0,4621
2,50	0,5769	0,5737
2,75	0,6775	0,6714
3,00	0,7632	0,7624
3,25	0,8333	0,8275
3,50	0,8869	0,8830
3,75	0,9261	0,9236
4,00	0,9535	0,9523
4,25	0,9718	0,9707
4,50	0,9834	0,9825
4,75	0,9906	0,9901

Слід також зазначити, що для одержання функції розподілу $P(d)$ може бути використаний і ще один схожий підхід. Нехай перша точка має координату (x_1, y_1). Щоб відстань між першою та другою випадковими точками (x_2, y_2) не перевищила величини d, друга точка повинна перебувати всередині кола радіуса d із центром у точці (x_1, y_1). Відстані від третьої точки (x_3, y_3) до двох уже наявних точок **1**(x_1, y_1) і **2**(x_2, y_2) не перевищать d, якщо вона буде перебувати в області, що утворюється перетинанням першого та другого кіл (зафарбована область на рис. 3). За аналогією позначимо умовну ймовірність влучання третьої точки в зазначену область $P_3(d; x_1, y_1, x_2, y_2)$. У підсумку ймовірність того, що відстань між трьома нормальними

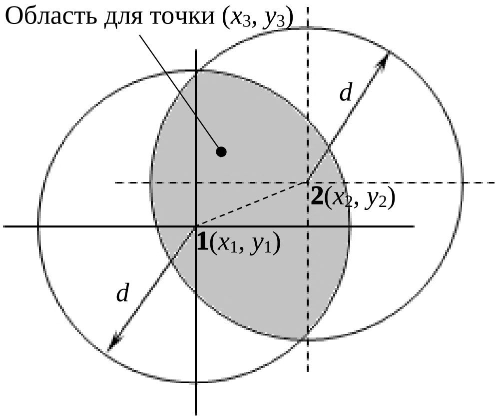

Рис. 3 – Області розташування трьох нормальних точок на площині точками на площині не перевищить величини d, визначиться шляхом послідовного інтегрування:

$$P_1(d;x_1,y_1)=\int_{x_1-d}^{x_1+d}\left[\int_{y_1-\sqrt{d^2-(x_2-x_1)^2}}^{y_1+\sqrt{d^2-(x_2-x_1)^2}}P_3(d;x_1,y_1,x_2,y_2)N(y_2)dy_2\right]N(x_2)dx_2\ ;$$

$$P(d)=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}P_1(d,x_1,y_1)N(x_1)N(y_1)dx_1dy_1\ . \qquad (3)$$

Імовірність $P_3(d;x_1,y_1,x_2,y_2)$ того, що третя точка (x_3, y_3) потрапить в область, яка є перетинанням двох кіл (рис. 3), можна визначити за аналогією до попереднього випадку, але при цьому процедура виявляється набагато складнішою. Як результат, суттєво зростає (в кілька разів у порівнянні з попереднім випадком) тривалість розрахунків. Імовірності $P(d)$ згідно з (3), отримані з використанням пакета математичних обчислень MathCAD, наведено в таблиці 2.

Таблиця 2 – Імовірності $P(d)$, отримані згідно з (3)

d	$P(d)$	Монте-Карло
1,0	0,0395	0,0392
1,5	0,1550	0,1542
2,0	0,3493	0,3483
2,5	0,5701	0,5737
3,0	0,7608	0,7624

Недоліком геометричного способу є те, що з використанням (2) або (3) одержати розподіл щільності операцією диференціювання практично неможливо.

3. Аналітичний метод одержання функції розподілу максимальної відстані між трьома випадковими нормальними точками на площині

Дістанемо розв’язок, розглядаючи відстані між трьома випадковими нормальними точками на площині. Випадкові величини, що являють відстані між парами точок (кількість пар точок у розглянутому випадку n = 3 дорівнює трьом) запишуться наступним чином:

$$R_1=\sqrt{(X_1-X_2)^2+(Y_1-Y_2)^2}\ ;$$

$$R_2=\sqrt{(X_1-X_3)^2+(Y_1-Y_3)^2}\ ;$$

$$R_3=\sqrt{(X_2-X_3)^2+(Y_2-Y_3)^2}\ .$$

Випадкові величини X_i і Y_i ($i=\overline{1,3}$), які входять у кожний із виразів для R_i ($i=\overline{1,3}$), є незалежними, отже ВВ R_i окремо підпорядковуються одновимірним законам Релея [1, 4]. Таким чином, імовірність того, що кожна ВВ R_i не перевищить величину d, дорівнюватиме

$$P(d)=1-\exp\left(-\frac{d^2}{2\pm_u^2}\right), \qquad (4)$$

де $\pm_u=\sqrt{2}$.

Припустивши, що всі ВВ R_i ($i=\overline{1,3}$) є незалежними, ймовірності того, що максимальна відстань між трьома випадковими нормальними точками на площині не перевищить певну величину *d* (ймовірності події $\bigcap_{i=1}^{3} R_i < d$), представлено в таблиці 3. Виявляємо істотну недооцінку отриманими значеннями результатів симуляції. Погрішність у цьому випадку пов'язана з тим, що існує кореляція між ВВ R_i ($i=\overline{1,3}$), оскільки у вирази для їхнього визначення входять одні і ті самі ВВ. Отже, для трьох залежних ВВ функція розподілу максимуму утворюється в такий спосіб [3]

Таблиця 3 – Імовірність того, що максимальна відстань між трьома випадковими нормальними точками не перевищить *d* для незалежних R_i ($i=\overline{1,3}$)

d	$\prod_{i=1}^{3} \Pr\left(R_i < d\right)$	Монте-Карло
1,0	0,0108	0,0392
1,5	0,0796	0,1542
2,0	0,2526	0,3483
2,5	0,4938	0,5737
3,0	0,7160	0,7624

$$F_{\max}(R_{\max} \le d) = \Phi_R^{(3)}(d,d,d),$$

де $\Phi_R^{(3)}(\bullet)$ – спільна функція розподілу ВВ R_1, R_2 і R_3. Тобто, для отримання коректного результату необхідно побудувати спільний розподіл ВВ R_i ($i=\overline{1,3}$).

Отримаємо розв'язок в інший спосіб. Для цього розглянемо ВВ, які являють квадрати відстаней між трьома випадковими нормальними точками:

$$W_1 = \left(X_1 - X_2\right)^2 + \left(Y_1 - Y_2\right)^2 = U_1 + V_1;$$

$$W_2 = \left(X_1 - X_3\right)^2 + \left(Y_1 - Y_3\right)^2 = U_2 + V_2;$$

$$W_3 = \left(X_2 - X_3\right)^2 + \left(Y_2 - Y_3\right)^2 = U_3 + V_3,$$

де позначено $U_1 = \left(X_1 - X_2\right)^2$; $U_2 = \left(X_1 - X_3\right)^2$; $U_3 = \left(X_2 - X_3\right)^2$; $V_1 = \left(Y_1 - Y_2\right)^2$; $V_2 = \left(Y_1 - Y_3\right)^2$; $V_3 = \left(Y_2 - Y_3\right)^2$. Оскільки функція квадрата є зростаючою, подія, яка полягає в тому, що максимальна відстань між випадковими нормальними точками на площині не перевищить *d*, еквівалентна події, що квадрат максимальної відстані не перевищить d^2. Тобто можемо записати

$$F_{\max}(R_{\max} \le d) = F_{\max}(W_{\max} \le d^2) = \Phi_W^{(3)}(d^2, d^2, d^2),$$

де $\Phi_W^{(3)}(t_1, t_2, t_3)$ – спільна функція розподілу ВВ W_1, W_2 і W_3. Отже треба побудувати $\Phi_W^{(3)}(t_1, t_2, t_3)$.

Відзначимо деякі особливості ВВ U_i і V_i ($i=\overline{1,3}$). Окремо ці ВВ підпорядковуються розподілу хі-квадрат з одним степінем вільності, щільність якого має вигляд [4]

$$f(u) = \frac{1}{\sqrt{2} \pm_u \Gamma\left(\frac{1}{2}\right)\sqrt{u}} \exp\left(-\frac{u}{2 \pm_u^2}\right) \quad (u \ge 0).$$

Числові характеристики зазначених ВВ: $m_{U_i} = m_{V_i} = 2$; $D_{U_i} = D_{V_i} = 8$ ($i = \overline{1,3}$). Між парами ВВ U_1 і U_2 (і відповідно U_1, U_3; U_2, U_3; а за аналогією V_1, V_2; V_1, V_3 та V_2, V_3) існує лінійна залежність. Коефіцієнт лінійної кореляції між ВВ U_1 і U_2 дорівнює

$$^\circ_{U_1U_2} = \frac{\sim^{(1,1)}_{U_1,U_2} - m_{U_1}m_{U_2}}{\pm_{U_1}\pm_{U_2}} = \frac{6-2\cdot 2}{(2\sqrt{2})^2} = \frac{1}{4}$$

(значення для $\sim^{(1,1)}_{U_1,U_2} = \int\limits_{-\infty}^{\infty}\int\limits_{-\infty}^{\infty}\int\limits_{-\infty}^{\infty} (x_1 - x_2)^2 (x_1 - x_3)^2 N(x_1)N(x_2)N(x_3)dx_1dx_2dx_3 =$

$$= \int\limits_{-\infty}^{\infty}\int\limits_{-\infty}^{\infty}\int\limits_{-\infty}^{\infty} (x_1^4 + x_1^2x_3^2 + x_1^2x_2^2 + x_2^2x_3^2)N(x_1)N(x_2)N(x_3)dx_1dx_2dx_3 = 3+1+1+1=6).$$

Розглядаючи тепер ВВ W_i ($i = \overline{1,3}$), легко з'ясувати, що окремо вони мають розподіли хі-квадрат із двома степінями вільності, які мають вигляд

$$f(w_i) = \frac{1}{2\pm_u^2\Gamma(1)}\exp\left(-\frac{w_i}{2\pm_u^2}\right) = \frac{1}{4}\exp\left(-\frac{w_i}{4}\right) \quad (w_i \geq 0) \qquad (5)$$

(фактично являють собою експоненціальні розподіли).

Числові характеристики ВВ W_i можна отримати інтегруванням, безпосередньо виходячи із ВВ X_i і Y_i ($i = \overline{1,3}$). Так, наприклад, для ВВ W_1 маємо (викладки не наведено)

$$m_{W_1} = \int\limits_{-\infty}^{\infty}\underset{(4)}{\ldots}\int\limits_{-\infty}^{\infty}\left((x_1 - x_2)^2 + (y_1 - y_2)^2\right)N(x_1)N(x_2)N(y_1)N(y_2)dx_1dx_2dy_1dy_2 = 4;$$

$$\sim^{(2)}_{W_1} = \int\limits_{-\infty}^{\infty}\underset{(4)}{\ldots}\int\limits_{-\infty}^{\infty}\left((x_1 - x_2)^2 + (y_1 - y_2)^2\right)^2 N(x_1)N(x_2)N(y_1)N(y_2)dx_1dx_2dy_1dy_2 = 32;$$

$$D_{W_1} = \sim^{(2)}_{W_1} - m_{W_1}^2 = 32 - 4\cdot 4 = 16.$$

Коваріаційний момент між W_1 і W_2 (а за аналогією між парами W_1, W_3 і W_2, W_3) становить

$$K_{W_1W_2} = \int\limits_{-\infty}^{\infty}\underset{(6)}{\ldots}\int\limits_{-\infty}^{\infty}\left((x_1 - x_2)^2 + (y_1 - y_2)^2\right)\times$$

$$\times\left((x_1 - x_3)^2 + (y_1 - y_3)^2\right)\prod_{i=1}^{3} N(x_i)N(y_i)dx_idy_i = 20.$$

Виходячи з нього, коефіцієнт лінійної кореляції між W_1 і W_2 дорівнює

$$^\circ_{W_1,W_2} = \frac{K_{W_1,W_2} - m_{W_1}m_{W_2}}{\sqrt{D_{W_1}D_{W_2}}} = \frac{20 - 4\cdot 4}{4\cdot 4} = \frac{1}{4}.$$

Отже виявляється, що коефіцієнт лінійної кореляції між парами ВВ W_i ($i=\overline{1,3}$) збігається із коефіцієнтом лінійної кореляції між ВВ U_1 і U_2.

Для опису пар ВВ W_1 і W_2 (а відповідно W_1, W_3 та W_2, W_3) можна скориста-тися двовимірним розподілом хі-квадрат з двома степінями вільності (також має назву двовимірного експоненціального розподілу Морана–Даунтона), який має вигляд [8]

$$f(w_1,w_2)=\frac{\lambda^2}{(1-\rho^2)}\exp\left[-\frac{\lambda}{(1-\rho^2)}(w_1+w_2)\right]\cdot I_0\left(\frac{2\rho\lambda\sqrt{w_1w_2}}{(1-\rho^2)}\right), \qquad (6)$$

де ρ – параметр кореляції ($|\rho|\le 1$); $\lambda=\frac{1}{4}$; $I_0(z)$ – модифікована функція Бесе-ля першого роду нульового порядку, що визначається співвідношенням $I_0(z)=\sum_{k=0}^{\infty}\frac{\left(\frac{z}{2}\right)^{2k}}{k!\Gamma(k+1)}$. Параметр кореляції ρ пов’язаний з коефіцієнтом лінійної кореляції γ_{W_1,W_2} наступним чином $\rho=\sqrt{\gamma_{W_1,W_2}}$.

Розподіл (6) дозволяє коректно описати кожну із пар ВВ W_1, W_2; W_1, W_3 та W_2, W_3. Зауважимо, що у деяких джерелах (дивись, наприклад, [6]) функція щільності (6) визначається за допомогою гіпергеометричної функції ${}_0F_1\left[1,\frac{\rho^2 w_1w_2}{16(1-\rho^2)^2}\right]$, де ${}_0F_1(b;z)=1+\sum_{k=1}^{\infty}\left(\prod_{l=0}^{k-1}\frac{1}{(1+l)(b+l)}\right)z^k$, замість функції Бесселя.

Безпосередньо описати тривимірну ВВ (W_1, W_2, W_3) дозволяє тривимір-ний аналог розподілу (6) [8]. Для випадку $\lambda_1=\lambda_2=\lambda_3=\lambda$ і $\rho=\sqrt{\gamma_{W_1,W_2}}=\sqrt{\gamma_{W_1,W_3}}=\sqrt{\gamma_{W_2,W_3}}$ тривимірний експоненціальний розподіл Морана–Даунтона матиме вигляд

$$f_W^{(3)}(\overline{w})=\frac{\lambda^3}{(1-\rho^2)^2}\exp\left[-\frac{\lambda}{(1-\rho^2)}(w_1+w_2+w_3)\right]\cdot\sum_{i=0}^{\infty}\frac{1}{(i!)^3}\left(\frac{\rho^2\lambda^3 w_1w_2w_3}{(1-\rho^2)^3}\right)^i. \qquad (7)$$

Запишемо щільність розподілу (7) наступним чином

$$f_W^{(3)}(\overline{w})=\frac{\lambda^3}{(1-\rho^2)^2}\sum_{i=0}^{\infty}\frac{\rho^{2i}\lambda^{3i}}{(i!)^3(1-\rho^2)^{3i}}\left(\prod_{k=1}^{3}w_k^i\exp\left[-\frac{\lambda w_k}{(1-\rho^2)}\right]\right).$$

Тепер визначимо інтеграл

$$\Phi_W^{(3)}(t,t,t)=\int_0^t\int_0^t\int_0^t f(w_1,w_2,w_3)dw_1dw_2dw_3=$$

$$=\sum_{i=0}^{\infty}\frac{\rho^{2i}\lambda^{3(i+1)}}{(i!)^3(1-\rho^2)^{3i+2}}\left(\int_0^t w^i\exp\left[-\frac{\lambda w}{(1-\rho^2)}\right]dw\right)^3=$$

$$=\sum_{i=0}^{\infty}\frac{(1-\rho^2)\rho^{2i}}{(i!)^3}\left(\gamma\left(i+1,\frac{\lambda t}{(1-\rho^2)}\right)\right)^3, \qquad (8)$$

де $¥(a,x)=\int_0^x t^{a-1}e^{-t}dt$ – неповна гама-функція; $-=\dfrac{1}{4}$ (дивись (5) і (6)).

Отриманий вираз (8) являє собою функцію розподілу квадрата максимальної відстані між трьома нормальними точками на площині $F_W^{(3)}(t)=\Phi_W^{(3)}(t,t,t)$. Беручи похідну від функції $F_W^{(3)}(t)$ за t, одержимо функцію щільності розподілу квадрата відстані між трьома випадковими нормальними точками, яка має вигляд

$$f_W^{(3)}(t)=\sum_{i=0}^{\infty}\frac{3^{-\,¨\,2i}}{(i!)^3}\left(¥\left(i+1,\frac{-t}{(1-¨^2)}\right)\right)^2\left(\frac{-t}{(1-¨^2)}\right)^i\exp\left(-\frac{-t}{(1-¨^2)}\right).$$

На рис. 4 показано графіки функції розподілу та функції щільності квадрата максимальної відстані між трьома випадковими нормальними точками на площині.

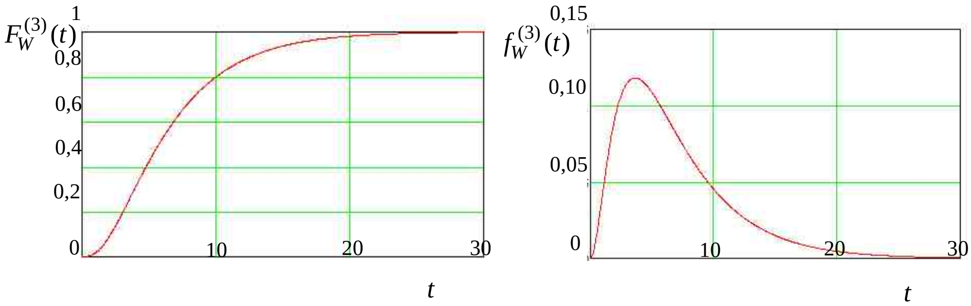

Рис. 4 – Функція розподілу та функція щільності квадрата максимальної відстані між трьома випадковими нормальними точками на площині

Визначимо імовірності $F_{\max}(W_{\max}\le d^2)$, скориставшись виразом (8). Результати розрахунків наведено у таблиці 4 і для наочності на рис. 5. Для порівняння в таблиці і на рисунку також наведено значення $P(d)$, отримані згідно з формулою (2).

Порівняння отриманих кривих на рис. 5 показує, що у разі невеликих d співвідношення (8) недооцінює реальну ймовірність того, що максимальна відстань між трьома випадковими нормальними точками виявиться меншою за d. При збільшенні d дві криві $F_W^{(3)}(d^2)$ й $P(d)$ практично збігаються. Відмінність $F_W^{(3)}(d^2)$ і $P(d)$ для невеликих d, очевидно, обумовлена наступним.

Таблиця 4 – Імовірності $F_W^{(3)}(d^2)$

d	$F_W^{(3)}(d^2)$	$P(d)$
0,50	0,0004	0,0030
0,75	0,0037	0,0141
1,00	0,0171	0,0405
1,25	0,0504	0,0879
1,50	0,1111	0,1584
1,75	0,2001	0,2496
2,00	0,3104	0,3555
2,25	0,4305	0,4675
2,50	0,5491	0,5769
2,75	0,6572	0,6775
3,00	0,7493	0,7632
3,25	0,8235	0,8333
3,50	0,8803	0,8869
3,75	0,9217	0,9261
4,00	0,9506	0,9535
4,25	0,9699	0,9718
4,50	0,9823	0,9834
4,75	0,9899	0,9906

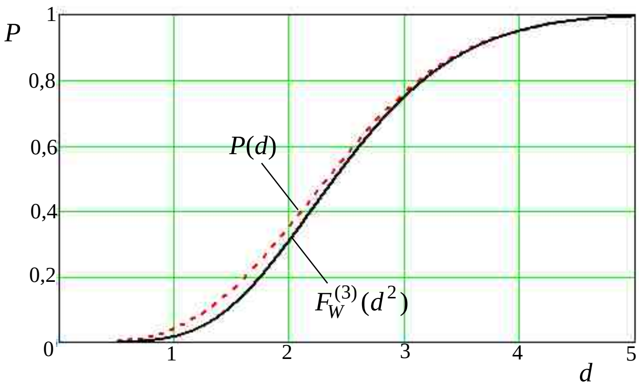

Рис. 5 – Імовірність того, що максимальна відстань між трьома випадковими нормальними точками не перевищить d

Тривимірний розподіл Морана–Даунтона (7), що описує спільну поведінку ВВ W_1, W_2 і W_3, дає можливість коректно оцінити числові характеристики одномірних і двовимірних складових, проте є похибка при визначенні спільних тривимірних характеристик зв'язку. Наприклад, шляхом прямого інтегрування можна встановити, що початковий мішаний момент третього порядку дорівнює

$$\sim^{(1,1,1)}_{W_1W_2W_3} = \int\limits_{-\infty}^{\infty} \underset{(6)}{\dots} \int\limits_{-\infty}^{\infty} \Big((x_1 - x_2)^2 + (y_1 - y_2)^2\Big)\Big((x_1 - x_3)^2 + (y_1 - y_3)^2\Big)\times$$

$$\times\Big((x_2 - x_3)^2 + (y_2 - y_3)^2\Big)\prod_{i=1}^{3} N(x_i)N(y_i)dx_i dy_i = 96.$$

Аналогічне значення отримуємо і з використанням статистичного моделювання. В той же час, безпосереднім інтегруванням розподілу ВВ W_1, W_2 і W_3 (7) отримуємо величину 132.

4. Апроксимація функції розподілу максимальної відстані між трьома випадковими нормальними точками на площині. Розглядаючи відстань між двома випадковими нормальними точками, послідовний розв'язок інтегралів (за аналогією із трьома точками, дивися формулу (4))

$$P'(d, x_1, y_1) = \int\limits_{x_1}^{x_1+d} \left[\Phi(y_1) - \Phi\left(y_1 - \sqrt{d^2 - (x_2 - x_1)^2}\right)\right] N(x_2)dx_2 \ ;$$

$$P(d) = 2^2 \int\limits_{-\infty}^{\infty}\int\limits_{-\infty}^{\infty} P'(d, x_1, y_1)N(x_1)N(y_1)dx_1dy_1$$

призводить до результату, що є ідентичним розподілу Релея (4).

Можна припустити, що деякий розподіл, який узагальнює розподіл Релея, буде реалізуватися й у випадку максимальної відстані між трьома випадковими нормальними точками. В одновимірному випадку узагальненням розподілу Релея є розподіл Райса (Rise) [1, 7], який має вигляд

$$f(x;\neg,\pm) = \frac{x}{\pm^2}\exp\left(-\frac{x^2+\neg^2}{2\pm^2}\right) I_0\left(\frac{x\neg}{\pm^2}\right) \quad (x \geq 0),$$

де ±, ¬ – параметри розподілу (ν, ± > 0); $I_0(z)$ – модифікована функція Беселя першого роду нульового порядку. У разі ¬ = 0 розподіл Райса перетворюється на розподіл Релея.

Отже, знайдемо такі значення параметрів розподілу Райса ¬ і ±, щоб отримані результати були близькими до значень $P(d)$, які визначено з використанням (2). Для функції розподілу Райса маємо:

$$F(x;\neg,\pm) = \int_0^x f(u;\neg,\pm)du = \int_0^x \frac{u}{\pm^2}\exp\left(-\frac{u^2+\neg^2}{2\pm^2}\right)\sum_{k=0}^{\infty}\frac{(u\neg)^{2k}}{(2\pm^2)^{2k}k!\Gamma(k+1)}du .$$

Шляхом заміни змінної $\frac{u^2}{2\pm^2} = t$ інтеграл запишемо у кінцевому вигляді, використовуючи неповну гама-функцію

$$F(x;\neg,\pm) = \exp\left(-\frac{\neg^2}{2\pm^2}\right)\sum_{k=0}^{\infty}\frac{\neg^{2k}¥\left(k+1,\frac{x^2}{2\pm^2}\right)}{(2\pm^2)^k k!\Gamma(k+1)} . \qquad (9)$$

З використанням метода найменших квадратів шляхом мінімізації функції

$$L(\pm,\neg) = \sum_{i=1}^{k}\left(P(d_i) - F(d_i;\pm,\neg)\right)^2 \to \min , \qquad (10)$$

де k – кількість точок, для яких визначалися значення функції розподілу $P(d)$ (таблиця 1), було отримано значення параметрів розподілу Райса для випадку трьох випадкових нормальних точок

$$\pm_* = 0{,}9391; \; \neg_* = 2{,}1485.$$

На рис. 6 показано відповідний графік функції розподілу Райса $F(d;\neg_*,\pm_*)$, яка описує максимальну відстань між трьома випадковими нормальними точками на площині.

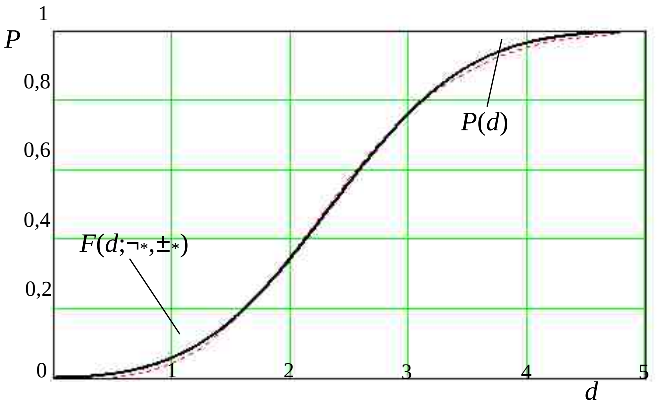

Рис. 6 – Функція розподілу Райса $F(d;\neg_*,\pm_*)$

Слід відзначити близькість апроксимуючої функції розподілу Райса до значень, отриманих за допомогою геометричного способу, у всьому діапазоні можливих значень d.

5. Випадок *n* випадкових нормальних точок на площині. Отримаємо узагальнення на випадок, коли кількість випадкових нормальних точок на площині $n > 3$.

Використовуючи геометричний підхід, можна припустити, що (n-2) точки повинні потрапити в зафарбовану область, яку показано на рис. 2. Тобто, для визначення ймовірності неперевищення максимальною відстанню між n випадковими нормальними точками величини d використати співвідношення (2), в якому при визначенні $P_1(d;x_1,y_1)$ в інтегралі застосовується імовірність $\left(P_3(d;x_1,y_1,x_2,y_2)\right)^{n-2}$ (називатимемо модифікованою формулою (2)). Проте ймовірність, яку отримано таким шляхом, буде перевищувати реальну величину. Це обумовлено таким: якщо випадкових нормальних точок буде більше, ніж три, то у межах зазначеної області (рис. 2) можлива ситуація, коли відстань між наступними (починаючи із третьої) точками може перевищити d. Наприклад, нехай третя точка близька до точки **H**. Якщо тепер наступна, четверта за рахунком, точка буде близькою до точки **C**, то відстань між третьою та четвертою точками перевищить d. Аналогічна ситуація буде, якщо третя точка буде перебувати біля точки **A**, а четверта – в околі точки **B**. У таблиці 5 наведено значення, отримані для зазначеної області з використанням модифікованого співвідношення (2), а також для порівняння результати статистичного моделювання.

Таблиця 5 – Імовірності неперевищення максимальною відстанню між n випадковими нормальними точками величини d

d	n = 4		n = 10		n = 15	
	$P(d)$	Монте-Карло	$P(d)$	Монте-Карло	$P(d)$	Монте-Карло
1,00	$6{,}7\cdot10^{-3}$	$6{,}7\cdot10^{-3}$	$<10^{-4}$	$<10^{-4}$	$<10^{-4}$	$<10^{-4}$
1,25	0,0217	0,0209	$<10^{-4}$	$<10^{-4}$	$<10^{-4}$	$<10^{-4}$
1,50	0,0534	0,0506	$<10^{-4}$	$<10^{-4}$	$<10^{-4}$	$<10^{-4}$
1,75	0,1076	0,1040	$<10^{-4}$	$<10^{-4}$	$<10^{-4}$	$<10^{-4}$
2,00	0,1856	0,1783	$2{,}3\cdot10^{-3}$	$1{,}9\cdot10^{-3}$	$<10^{-4}$	$<10^{-4}$
2,25	0,2864	0,2756	$9{,}2\cdot10^{-3}$	$7{,}8\cdot10^{-3}$	$<10^{-4}$	$<10^{-4}$
2,50	0,4000	0,3916	0,0286	0,0256	$2{,}5\cdot10^{-3}$	$2{,}3\cdot10^{-3}$
2,75	0,5184	0,5067	0,0701	0,0644	0,0109	0,0102
3,00	0,6304	0,6205	0,1423	0,1323	0,0345	0,0300
3,25	0,7280	0,7215	0,2465	0,2317	0,0854	0,0757
3,50	0,8096	0,8032	0,3746	0,3565	0,1723	0,1564
3,75	0,8720	0,8663	0,5117	0,4934	0,2937	0,2754
4,00	0,9168	0,9128	0,6416	0,6246	0,4360	0,4174
4,25	0,9488	0,9448	0,7525	0,7403	0,5798	0,5630
4,50	0,9696	0,9667	0,8386	0,8278	0,7076	0,6940
4,75	0,9824	0,9806	0,9004	0,8907	0,8094	0,7982
5,00	0,9904	0,9886	0,9416	0,9355	0,8831	0,8745
5,25	0,9952	0,9939	0,9674	0,9641	0,9323	0,9264
5,50	0,9968	0,9967	0,9826	0,9801	0,9628	0,9591
5,75	0,9986	0,9985	0,9911	0,9899	0,9805	0,9780
6,00	1	1	0,9956	0,9948	0,9902	0,9890
6,25	1	1	1	1	0,9954	0,9953

Отже перевищення для модифікованого співвідношення (2) результатів статистичного моделювання несуттєве, другий-третій знак після коми. Відповідно, можна використовувати модернізоване співвідношення (2) і для випадку $n > 3$, при цьому отримані значення представлятимуть оцінки зверху.

У рамках аналітичного підходу можна одержати узагальнення розподілу (8) для n випадкових нормальних точок на площині, яке має вигляд

$$F_W^{(n)}(t)=\sum_{i=0}^{\infty}\frac{(1-¨^{2})^{¨\,2i}}{(i!)^{N}}\left(¥\left(i+1,\frac{{}^{-}t}{(1-¨^{2})}\right)\right)^{N}, \qquad (11)$$

де $N=\dfrac{n(n-1)}{2}$ – кількість пар точок, що розглядаються (відстані між точками). Наприклад, для чотирьох випадкових нормальних точок ($n = 4$) маємо $N = 6$. Результати розрахунків функції розподілу для $n = 4$ і $n = 7$ показано на рис. 7. Для порівняння на графіках наведено емпіричні функції розподілу, що отримано з використанням статистичного моделювання.

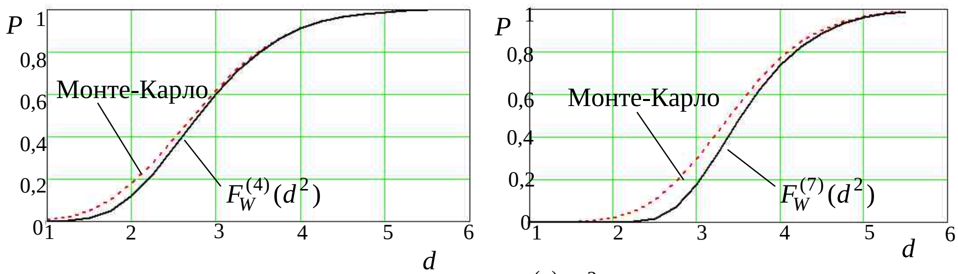

Рис. 7 – Функція розподілу $F_W^{(n)}(d^2)$ для різних n

Таким чином, виявляємо аналогічну ситуацію, як і у випадку $n = 3$, пов'язану з недооцінкою $F_W^{(n)}(d^2)$ реальної ймовірності для невеликих відстаней d та їх практичний збіг у разі великих d. При цьому різниця між $F_W^{(n)}(d^2)$ і результатами статистичного моделювання стає більш істотною у разі збільшення кількості випадкових нормальних точок n.

За аналогією з випадком $n = 3$ можна побудувати апроксимуючий розподіл Райса і для випадку більшої кількості випадкових нормальних точок n. У якості значень функції розподілу, відносно яких буде відбуватися визначення параметрів розподілу (9), скористаємося емпіричною функцію розподілу, яку отримано за результатами статистичного моделювання. Значення параметрів розподілу Райса, отримані з використанням функції (10) на основі методу найменших квадратів для $n \le 30$, наведено в таблиці 6.

Таблиця 6 – Параметри розподілу Райса, апроксимуючого максимальну відстань між n випадковими нормальними точками на площині

Параметр	n						
	4	5	7	10	15	20	30
¬*	2,6140	2,9188	3,3175	3,7128	4,1071	4,3822	4,7175
±*	0,8810	0,8397	0,7897	0,7388	0,6924	0,6647	0,6119

Як приклад на рис. 8 показано графіки функції апроксимуючого розподілу Райса для $n = 4, 7, 10$, а також результати статистичного моделювання.

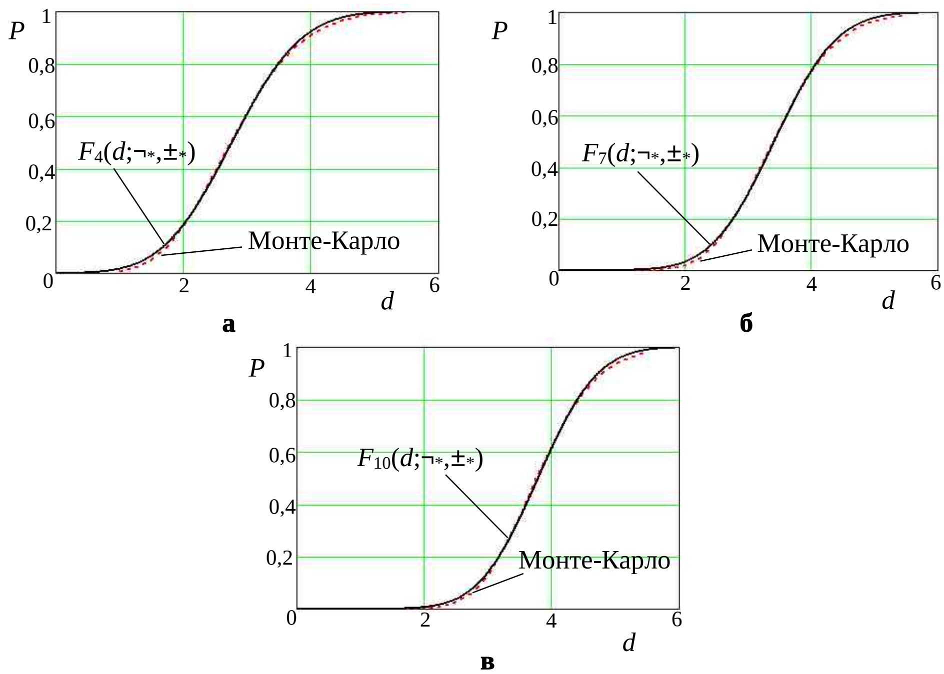

Рис. 8 – Апроксимуючі розподіли Райса: **а)** – *n* = 4; **б)** – *n* = 7; **в)** – *n* = 10

Розрахунки й графіки показують, що у всіх випадках отримано досить непогану збіжність між значеннями апроксимуючої функції Райса й результатами статистичного моделювання.

На рис. 9 показано графіки зміни значень параметрів ¬* і ±* у залежності від кількості випадкових точок *n* в інтервалі *n* = 3…30.

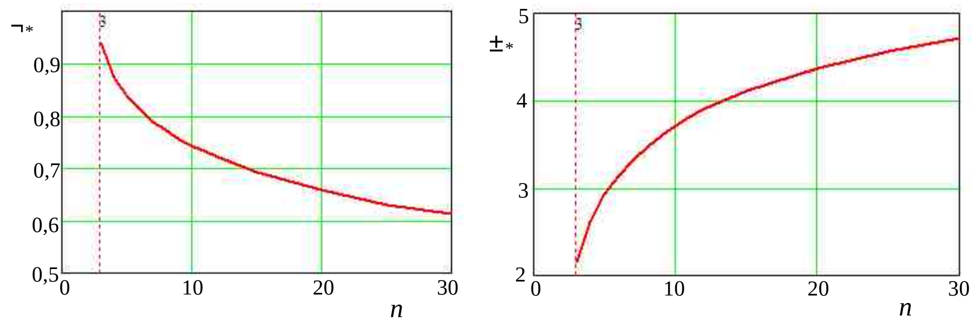

Рис. 9 – Залежність параметрів ¬* і ±* від *n*

Зміну параметрів ¬* і ±* у залежності від кількості випадкових точок *n*, можна описати наступною загальною наближеною функцією

$$f(n;a,b,c,d) = a - \exp\left(\frac{b}{n^c + d}\right),$$

де для ¬* необхідно використовувати значення параметрів *a* = 6,5047; *b* = 13,1922; c = 0,8; *d* = 6,622; а для ±* – *a* = 1,3964; *b* = -1,3589; c = 0,5; *d* = 0,0544 відповідно.

Висновки та рекомендації. Виходячи із проведеного аналізу, для практичних розрахунків можна рекомендувати таке:

- співвідношення (2) або (3), що побудовані виходячи із геометричних міркувань, дозволяють достатньо точно отримати ймовірності неперевищення максимальною відстанню між трьома випадковими нормальними точками на площині деякої величини. Проте для них потрібно використання чисельних методів визначення інтегралів, а функцію густини одержати неможливо. Модифіковане співвідношення (2), у якому при визначенні $P_1(d;x_1,y_1)$ використовується ймовірність $\left(P_3(d;x_1,y_1,x_2,y_2)\right)^{n-2}$, може бути використано також і для $n > 3$ випадкових точок на площині, проте результати, що отримуємо, перевищуватимуть точні значення. Перевищення при цьому незначне. Отже, такий спосіб можна використовувати для отримання оцінок зверху;
- для визначення ймовірності неперевищення максимальною відстанню між випадковими нормальними точками на площині деякої величини для невеликих n (3-5) і значних d (у разі, коли отримувана імовірність буде більшою за 0,6 – 0,7) можна використовувати функцію розподілу (8) і (11);
- більш широке практичне застосування може мати апроксимація розподілу максимальної відстані між випадковими нормальними точками на площині з використанням розподілу Райса. Придатний розв'язок отримуємо, як для $n = 3$, так і для випадку $n > 3$. Для $n = 3…30$ отримано значення параметрів апроксимуючого розподілу Райса.

1. *Губарев В. В.* Таблицы характеристик случайных величин и векторов. Новосибирск: Изд-во НЭТИ, 1980. 225 с. Деп. в ВИНИТИ №3146–81.
2. *Гумбель Э.* Статистика экстремальных значений. М.: Мир, 1965. 452 с.
3. *Дэйвид Г.* Порядковые статистики. М.: Наука, 1979. 336 с.
4. *Крамер Г.* Математические методы статистики. М.: ГИИЛ, 1975. 648 с.
5. *Appel M. J. B., Najim C. A., Russo R. P.* Limit laws for the diameter of a random point set. Adv. in Appl. Probab. 2002. Vol. 34. Pp. 1–10. https://doi.org/10.1017/S0001867800011356
6. *Joarder A. H., Omar M., Gupta A. K.* The distribution of a Linear Combination of Two Correlated Chi-Square Variables. Vol. 36. No. 2. Pp. 211–221.
7. *Jonson N. L., Kotz S., Balakrishnan N.* Continuous Univariate Distributions. Vol. 1. N.Y.e.a. John Wiley and Sons, 2000. 756 p.
8. *Kotz S., Balakrishnan N., Johnson N. L.* Continuous Multivariate Distributions. Vol. 1: Models and Applications. N.Y.e.a. John Wiley and Sons, 2000. 722 p. https://doi.org/10.1002/0471722065
9. *Matthews P. C., Rukhin A. L.* Asymptotic distribution of the normal sample range. Ann. Appl. Probab. 1993. Vol. 3. Pp. 454–466. https://doi.org/10.1214/aoap/1177005433

Отримано 05.04.2024,
в остаточному варіанті 02.06.2024