Лекція 25.

Моделювання нормально розподілених випадкових величин

Нормальний закон розподілу зустрічається в природі досить часто, тому для нього розроблені окремі ефективні методи моделювання. Формула розподілу ймовірності значень випадкової величини x за нормальним законом має вигляд:

$$y = \frac{1}{\sigma_x \cdot \sqrt{2\pi}} \cdot e^{-\frac{(x - m_x)^2}{2\sigma_x^2}}$$

Як видно, нормальний розподіл має два параметри: математичне очікування m_x і середньоквадратичне відхилення σ_x величини x від цього математичного очікування.

x — випадкова величина;

$y(x)$ — імовірність прийняття випадковою величиною значення x;

m_x — математичне очікування;

σ_x — середнє квадратичне відхилення.

Нормалізованим нормальним розподілом називається такий нормальний розподіл, у якого $m_x = 0$ й $\sigma_x = 1$. З нормалізованого розподілу можна одержати будь-який інший нормальний розподіл із заданими m_x й σ_x по формулі: $z = m_x + x \cdot \sigma_x$.

Розглядаючи останню формулу, згадаєте формули комп'ютерної графіки: операція масштабування виражається в математичній моделі через множення (це відповідає зміні розкиду величини, розтягуванню геометричного образа), операція зсуву виражається через додавання (це відповідає зміні значення найбільш імовірної величини, зсуву геометричного образа).

Функція нормального розподілу має вигляд дзвона. На мал. 25.1 показане нормалізований нормальний розподіл.

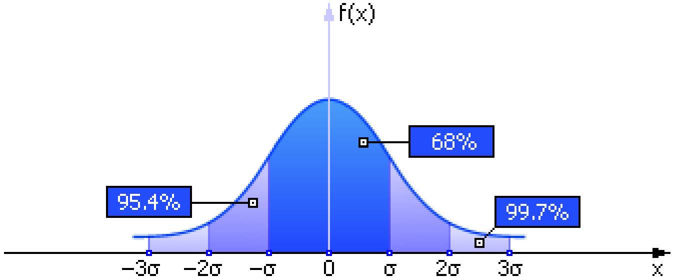

Рис. 25.1. Графічний вид нормального закону розподілу випадкової величини x з параметрами $m_x = 0$ й $\sigma_x = 1$ (розподіл нормалізований)

Графік на мал. 25.1 показує, що в області $-\sigma < x < \sigma$ на графіку зосереджено 68% площі розподілу, в області $-2\sigma < x < 2\sigma$ на графіку зосереджено 95.4% площі розподілу, в області $-3\sigma < x < 3\sigma$ на графіку зосереджено 99.7% площі розподілу («правило трьох сигм»). Згадаєте, будь ласка, мал. 2.7 з лекції 02.

Приклад. По нормальному розподілі розподілений ріст людей, що перебувають одночасно у великій аудиторії. А саме: досить мало людей дуже великого росту, і настільки ж мала ймовірність зустріти людей дуже малого росту. В основному, легше зустріти людей середнього росту - і ймовірність цього велика.

Наприклад, середній ріст людей становить, в основному, 170 див, тобто $m_x = 170$. Відомо також, що $\sigma_x = 20$. На мал. 25.1 показано, що частка людей з ростом від 150 до 190 ($170 - 20 < 170 < 170 + 20$) становить у суспільстві 68%. Частка людей від 130 див до 210 див ($170 - 2 \cdot 20 < 170 < 170 + 2 \cdot 20$) становить у суспільстві 95.4%. Частка людей від 110 див до 230 ($170 - 3 \cdot 20 < 170 < 170 + 3 \cdot 20$) становить у суспільстві 99.7%. Наприклад, імовірність того, що людина виявиться ростом менше 110 див або більше 230 див становить усього 3 чоловік на 1000.

Властивості нормального розподілу

Зміна параметра нормального розподілу m_x приводить до зрушення кривої по осі x (див. мал. 25.2).

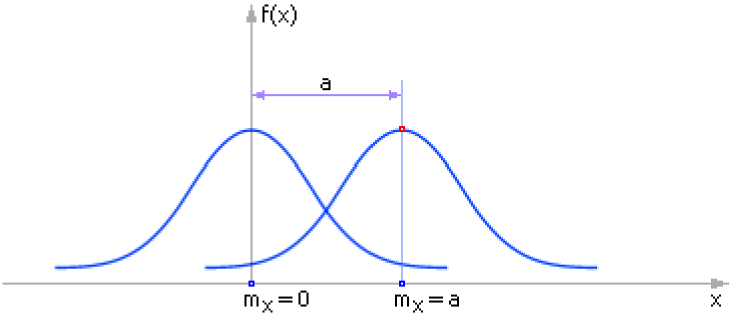

Рис. 25.2. Вплив параметра «математичне очікування»
на вид закону нормального розподілу випадкової величини x

Зміна параметра нормального розподілу σ_x приводить до масштабування форми по осі x (нагадуємо, у кожному разі завжди площа під кривої щільності ймовірності незмінна й дорівнює 1).

Ніж більше не випадковий процес, тим менше його середньоквадратичне відхилення, тим уже й вище дзвін на графіку. Зміна параметра нормального розподілу σ_x приводить до масштабування форми (див. мал. 25.3) по осі x (нагадуємо, у кожному разі завжди площа під кривої щільності ймовірності незмінна й дорівнює 1).

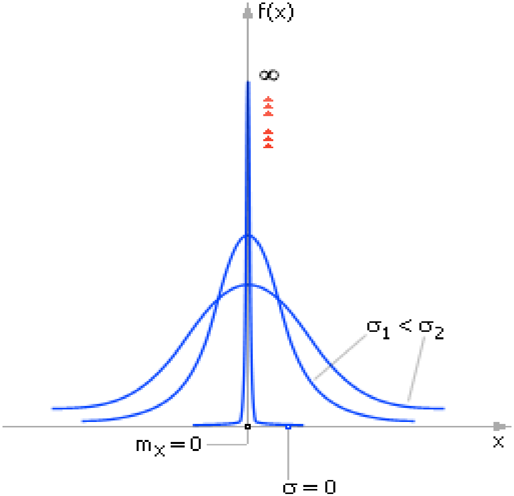

Рис. 25.3. Вплив параметра «середньоквадратичне відхилення» на вид закону нормального розподілу випадкової величини x

Ніж більше не випадковий процес, тим менше його середньоквадратичне відхилення, тим уже й вище дзвін на графіку. Дійсно, розкид випадковості щодо математичного очікування стає усе більше мінімальним. У межі детермінований процес має вигляд, показаний на мал. 25.4.

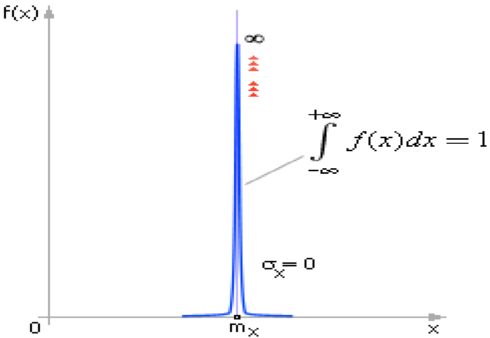

Рис. 25.4. Вид закону нормального розподілу ймовірності
при переході його до детермінованого випадку в межі ($\sigma_x = 0$).
Випадкова подія стає детермінованим: $x = m_x \pm 0$ (розкиду немає)

Вивчати детерміновані процеси простіше. Чим більше величина σ_x, тим менш закономірне поводження досліджуваного об'єкта, тому що можливі будь-які значення його параметрів, що характеризують, розкид величин щодо середньої очікуваної збільшується. Прогнозування й керування поводженням об'єкта в цьому випадку утрудняється.

Розглянемо вид інтегральної кривої щільності розподілу випадкової величини, розподіленої за нормальним законом. Вид її наведений на мал. 25.5. F — інтегральна функція Лапласа. Зміст інтегральної функції — імовірність того, що випадкова величина прийме значення з діапазону від $-\infty$ до x. Наприклад, запис $F(170) = 0.5$ для нашого приклада означає: імовірність того, що випадково обраний з аудиторії людина буде ростом не вище 170 див, становить 0.5 (тобто кожен другий).

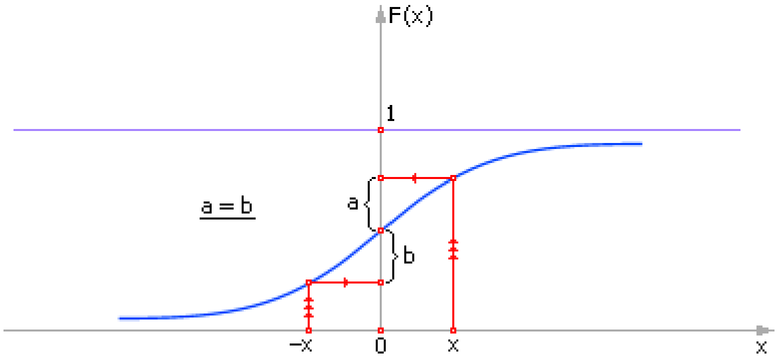

Рис. 25.5. Вид інтегральної функції Лапласа F(x)

Дана функція задана інтегралом від щільності ймовірності нормального розподілу:

$$F(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{t^2}{2}} dt$$

На жаль, цей інтеграл не береться в загальному виді, тому функція Лапласа задана у вигляді таблиці для $m_x = 0$ й $\sigma_x = 1$. Оскільки функція Лапласа симетрична щодо крапки $(x = 0, y = 0.5)$ (як і функція самого нормального розподілу), $F(-x) = 1 - F(x)$, те в таблиці втримується тільки одна з її симетричних частин.

Якщо задається інтервал інтегрування функції Лапласа $[a; b]$, то:

$$P(a < x < b) = \int_{a}^{b} \frac{1}{\sigma\sqrt{2\pi}} \cdot e^{-\frac{(x-m)^2}{2\sigma^2}} dx = \int_{\frac{a-m}{\sigma}}^{\frac{b-m}{\sigma}} \frac{1}{\sigma\sqrt{2\pi}} \cdot e^{-\frac{t^2}{2}} dt = F(b) - F(a)$$

Імовірність влучення X в інтервал, симетричний відносно m_x:

$$P(|x - m_x| < L) = P((m_x - L) < x < (m_x + L)) = F((b - m_x)/\sigma) - F((a - m_x)/\sigma) =$$
$$= F(L/\sigma) - F(-L/\sigma) = 2F(L/\sigma) - 1$$

Наприклад, для правила «трьох сигм»: $P(|x - m_x| < 3\sigma) = 2 \cdot F(3) - 1 = 2 \cdot 0.9987 - 1 = 0.9973$ (як раніше ми й указували). Число $F(3) = 0.9987$ узято з таблиці Лапласа.

Приклад. Знайти ймовірність виготовлення деталі з помилкою в її розмірах не більше 15 мм, якщо відомо, що виготовлення деталі з помилкою розподілено за нормальним законом $m = 0$ й $\sigma = 10$ мм.

$P(|x| < 15) = P(-15 < x < 15) = F((15 - 0)/10) - F((-15 - 0)/10) = F(1.5) - F(-1.5) = F(1.5) - (1 - F(1.5)) = 2 \cdot F(1.5) - 1 = 2 \cdot 0.9332 - 1 = 0.8664$. Тобто 8664 деталей з 10000 будуть мати помилку в розмірах не більше 15 мм.

Важливою властивістю закону є те, що нормальний розподіл є межею для різного виду розподілів, що випливає із центральної граничної теореми: «Для великого числа N випадкових величин X з будь-яким законом розподілу їхня сума є випадкове число з нормальним законом розподілу».

$$\lim_{N\to\infty} P\{\alpha < (\sum_{i=1}^{N} x_i - N\cdot a)\cdot\frac{1}{\sqrt{N}\cdot\sigma} < \beta\} = \frac{1}{\sqrt{2\pi}}\int_{\alpha}^{\beta} e^{-\frac{t^2}{2}}dt = F(\beta) - F(\alpha)$$

де a — математичне очікування в законі розподілу випадкової величини X; σ — середньоквадратичне відхилення в законі розподілу випадкової величини X; N — кількість випадкових чисел.

Згадаємо досвід «Дошка Гальтона» з фізики (див. мал. 25.6).

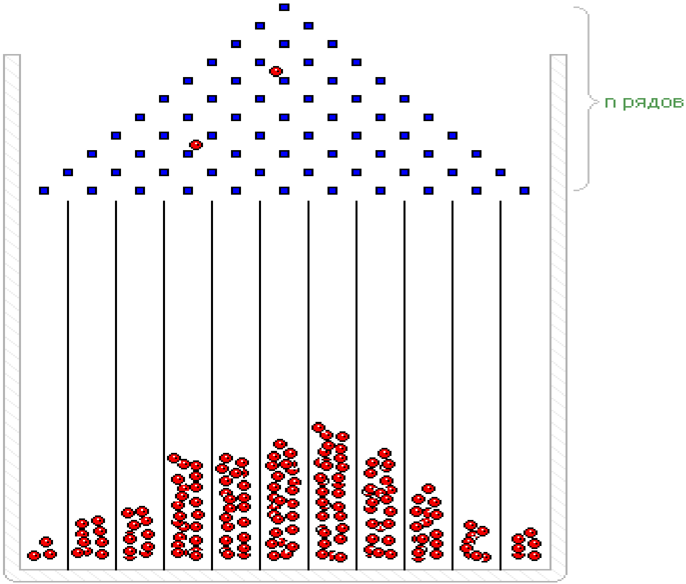

Рис. 25.6. Дошка Гальтона. Кульки, що падають зверху в посудину випадково розподіляються в ньому згідно з нормальним законом розподілу

Дошка розділена на секції; у верхній частині дошки перебувають особливим образом розташовані стрижні, ударяючись об які, множина падаючих зверху під впливом сили ваги кульок, випробовуючи зіткнення також і між собою, міняють свою траєкторію польоту. У результаті в різні секції попадає різна кількість кульок. Якщо дочекатися закінчення досвіду, то в кожній секції буде певна кількість кульок, звичайно, щораз різне, тому що процес їхнього зіткнення випадковий. Але цікаво те, що розподіл кульок по секціях буде утворювати нормальний закон розподілу. Начебто б дошка не міняється, кульки падають ті самі, і проте, по-перше, форма розподілу злегка коливається (випадковість), і різні кульки попадають у різні секції, по-друге, на макрорівні, де проявляється організація кульок як сукупності, завжди виходить нормальний закон розподілу (закономірність).

Для допитливого розуму досвід породжує дуже цікаві питання. Чому малі відхилення в поводженні елементів системи (кульок) ведуть до більших викидів у їхніх координатах? Чому випадки не компенсуються? Наскільки вони не компенсуються?

Позначимо через D відстань між стрижнями дошки Гальтона, r — діаметр стрижня (див. мал. 25.7, а, б). Очевидно, що відхилення β кульки від стрімкої траєкторії при малому куті α (зміна напрямку внаслідок удару об стрижень) можна порахувати як $\mathrm{tg}(\alpha_1) = \beta_1/r$ або $\alpha_1 \approx \beta_1/r$ (при малих кутах α_1). Тобто, далі кулька летить під кутом α_1 до траєкторії вільного падіння на відстань D, поки не зустріне новий стрижень. За час перельоту по прямій від стрижня до стрижня кулька відхилиться від стрімкої вертикальної лінії на величину $\beta_2 = \alpha_1 \cdot D$. Далі знову, випробувавши зіткнення зі стрижнями, кулька відхилиться на кут $\alpha_2 \approx \beta_2/r$ й $\beta_3 \approx \alpha_2 \cdot D = \beta_1 \cdot (D/r)^2$ і так далі.

Рис. 25.7. Взаємодія кульок зі стрижнями дошки (деталізація)

Якщо кулька випробує n зіткнень, то він відхилиться від стрімкої траєкторії на відстань $\beta_n \approx \beta_1 \cdot (D/r)^{n-1}$ (звичайно, якщо йому «повезе» і він буде відхилятися увесь час в одну сторону).

Тепер подивимося, скільки це означає в цифрах, наприклад, $D = 20r$, нехай стрижнів, від яких відхилилася кулька в польоті (зіткнень), $n = 10$, $\beta_1 = 1$ нм (відстань менше атома!!), те тоді легко обчислити, що $\beta_{10} = 10.24$ км. Тобто, видно, що надмалі відхилення (1 нм і число зіткнень $n = 10$), фактично випадковості, приводять до макроефектів. Насправді такий розкид наступає набагато швидше, тому що кульки зіштовхуються ще й між собою. До речі, видалення якоїсь кульки на 10 км — це теж неймовірну подію, тому що ймовірність, що кульці «повезе» 20 разів підряд, становить украй малоймовірне число $P = 0.520 = 10^{-6}$.

Розглянемо ще один експеримент «Розподіл Максвелла». У посудині перебуває 40 синіх й 40 червоних куль, кожні у своїй половинці посудини. Половинки посудини розділені перегородкою з отвором, через яке можуть проникати кулі з однієї частини посудини в іншу (див. мал. 25.8). Кількість куль, як синіх, так і червоних, у кожній із частин посудини підраховуються.

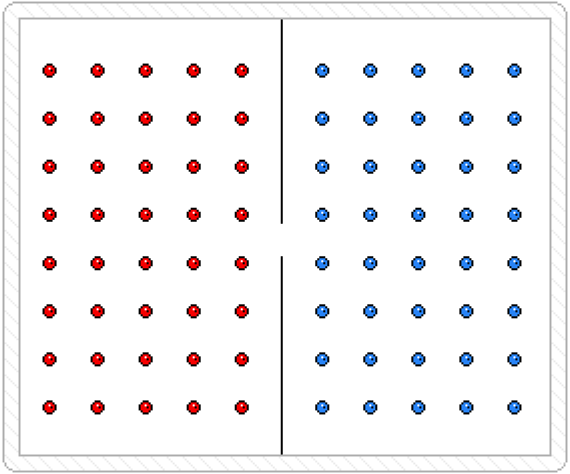

Рис. 25.8. Схема експерименту, що демонструє дифузію (поява хаосу) у складних системах

Кулі імітують броунівське рух молекул, зіштовхуючись один з одним пружно, обмінюючись при ударі енергіями, імпульсами, напрямками швидкості й також взаємодіючи зі стінками посудини. При змодельованому в досвіді пружному ударі усереднюється середня кінетична енергія часток, що зустрічаються. Незважаючи на те, і в цьому суть дифузії, що кулі перебували кожні кольори строго у своїй половині, згодом приблизно половина (50%) червоних куль виявиться в першій половині посудини, а половина (50%) — у другій його частині, то ж стосується й синіх куль (50% : 50%). Ужите тут слово «приблизно» означає, що час від часу сукупність куль буде ділитися не рівно на половини, а в пропорції (50% ± h : 50% ± h), тому що випадкова кількість куль буде переходити в довільні моменти часу з однієї частини посудини в іншу, і навпаки. Виявлена в наших досвідах відносна флуктуація середніх енергій досить велика, порядку 20%. Незважаючи на те, що кулі кожних кольорів перебували строго у своїй половині (у цьому й складається суть дифузії), згодом приблизно половина (50%) червоних куль виявиться в першій половині посудини, а половина (50%) — у другій його частині, то ж стосується й синіх куль (50%:50%) (див. мал. 25.9).

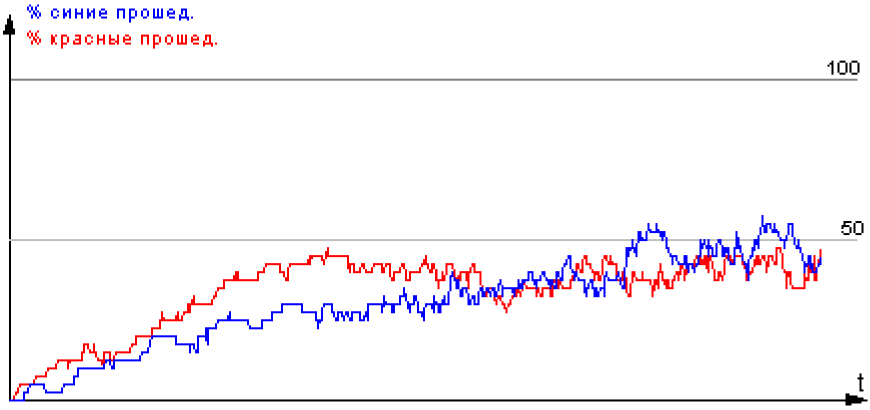

Рис. 25.9. Типовий вид поводження характеристик у системах з дифузією. На малюнку показана зміна кількості синіх і червоних молекул в одній з половинок посудини згодом

Особливість 1. Можете продовжувати спостереження як завгодно довго, але при цьому ніколи первісного розподілу цих куль досягнуто не буде. І, навпаки, при будь-якому початковому порядку куль система буде переходити до тому самому хаотичного стану. Чому? Пояснення цього треба шукати в тому, що час незворотній. Операція усереднення також необоротна. Якщо я вам повідомлю, що сума двох чисел дорівнює 8, то навряд чи ви зможете сказати, які я вибрав для цього доданки. І множина варіантів доданків ведуть до одній і тій же сумі.

Хаотична система не пам'ятає своє минуле. А саме: стан такої системи по закінченні часу не залежить від того, як і з якого стану система прийшла в хаотичне. З будь-якого початкового стану наша система досить швидко переходить у цей хаотичний стан, і перебуває в ньому далі скільки завгодно довго. Тобто хаотичний стан — стійкий стан. Цікаво, що в цьому стійкому стані можна відшукати такий ряд параметрів, що буде незмінний згодом, такі параметри описують систему в цілому й називаються макропараметрами. Наприклад, у нашому прикладі стійким макропараметром є розподіл кількості молекул n по модулі швидкості v, що демонструє закон Максвелла $n(v) = r \cdot v^2 \cdot e^{-c \cdot v2/T}$ (r, c — коефіцієнти, T — температура сукупності молекул).

Особливість 2. Якщо розділити посудина на більше число частин, то кулі розподіляться між ними приблизно в рівних частках.

Особливість 3. Слово «приблизно» навпіл, говорить нам про те, що у відповіді 50% : 50% завжди є розкид, дисперсія, флуктуація. Цей розкид зменшується, якщо збільшити число молекул у посудині.

Особливість 4. Якщо запустити модель знову й спробувати кулі спочатку впорядкувати, наприклад, установити їх рівно рядами в одній половині посудини, або вгорі, або в одному з кутів, то дуже швидко кулі «приблизно» рівномірно розподіляться по посудині, і наступить хаос, тобто порядок зникне й, що цікаво, ніколи знову не наступить. Звідки виникає хаос, як система переходить від порядку до хаосу? Причина хаосу, як ми вже раніше показали в досвіді «Дошка Гальтона», з'являється в момент **невеликих відхилень** куль (нехай, наприклад, 10^{-9} м) від ідеальної теоретичної траєкторії, але це невелике відхилення на прольоті відстані D приводить до істотного відхилення траєкторії від ідеалу й зіткненню куль, що **різко міняє** траєкторії сусідів. Миттєво з'являється лавина зіткнень, і хаос наступає дуже швидко.

Особливість 5. Раніше, уважалося, що системи з малим числом елементів не здатні проявляти випадковість, хаотичне поводження. Переважало міркування фізиків про газ із мільярдів молекул. Проведіть цей же експеримент із п'ятьма кулями, потім із трьома й, нарешті, однією кулею. Розподіл Максвелла в результаті експерименту проявляється досить уперто, щоправда, для цього потрібно чекати в кілька разів довше. (Виключення складе досвід з однією кулею!) **Досвід показує, що усереднення по ансамблі молекул і за часом проводить до однакового результату**. Так що хаос може виникнути й на невеликих множинах.

Отже, і досвід «Дошка Гальтона» й експериментальне зняття розподілу Максвелла приводять нас емпірично до того, що нормальний розподіл широкий поширене в навколишньому нас світі. Підсумовування випадкових величин, сума дій, спільне накладення ефектів на мікрорівні часто приводить, усереднившись, до появи нормального розподілу на макрорівні.

Розглянемо питання імітації випадкових величин, заданих нормальним законом розподілу.

Табличний метод генерації нормально розподілених чисел

Для цього нормальне число можна взяти з довідника в таблиці функції Лапласа й одержати випадкове число по методу узяття зворотної функції (див. лекцію 24): $x = F^{-1}(r)$, де F — інтегральна функція Лапласа.

Технічно це означає, що треба розіграти випадкове рівномірно розподілене число r з інтервалу [0; 1] стандартним ГВЧ (див. таблицю абсолютно випадкових перевірених чисел), знайти рівне йому число в таблиці значень функції Лапласа в стовпці F і по рядку визначити випадкову величину x, що відповідає цьому числу.

Недоліком методу є необхідність зберігання в пам'яті комп'ютера всієї таблиці чисел функції Лапласа.

Метод генерації нормально розподілених чисел, що використає центральну граничну теорему

Загальна ідея методу наступна: потрібно скласти випадкові числа з будь-яким законом розподілу, нормалізувати їх і перевести в потрібний діапазон нормального розподілу.

Допустимо, що нам треба з метою імітації одержати ряд випадкових чисел x, розподілених за нормальним законом із заданими математичним очікуванням m_x і середньоквадратичним відхиленням σ_x.

1. Складемо n випадкових чисел, використовуючи стандартний ГВЧ:

$$V = \sum_{i=1}^{n} r_i$$

   Згідно ЦПТ числа V утворять ряд значень, розподілений за нормальним законом. Ці числа тим краще описують нормальний закон, чим більше параметр n. На практиці n беруть рівними 6 або 12. Помітимо, що закон розподілу чисел V має математичне очікування $m = n/2$, $\sigma_V = \mathrm{sqrt}(n/12)$. Тому він є зміщеним щодо заданого довільного.
2. За допомогою формули $z = (V - m)/\sigma_V$ нормалізуємо цей ряд. Одержимо нормалізований закон нормального розподілу чисел Z. Тобто $m_z = 0$, $\sigma_z = 1$.
3. Формулою (зрушення на m_x і масштабування на σ_x) перетворимо ряд Z у ряд x: $x = z \cdot \sigma_x + m_x$.

Приклад. Змоделювати потік заготівель для обробки їх на верстаті. Відомо, що довжина заготівлі коливається випадковим образом. Середня довжина заготівлі становить 35 див, а середньоквадратичне відхилення реальної довжини від середньої становить 10 див. Тобто за умовами задачі $m_x = 35$, $\sigma_x = 10$. Тоді значення випадкової величини буде розраховуватися по формулі: $V = r_1 + r_2 + r_3 + r_4 + r_5 + r_6$, де r — випадкові числа із ГВЧ$_{pp}$ [0; 1], $n = 6$.

$$X = \sigma_x \cdot (\mathrm{sqrt}(12/n) \cdot (V - n/2)) + m_x = 10 \cdot \mathrm{sqrt}(2) \cdot (V - 3) + 35$$

або

$$X = 10 \cdot \mathrm{sqrt}(2) \cdot ((r_1 + r_2 + r_3 + r_4 + r_5 + r_6) - 3) + 35.$$

Метод Мюлера

Зовсім простим методом одержання нормальних чисел є метод Мюлера, що використає формули: $Z = \sqrt{(-2 \cdot \mathrm{Ln}(r_1))} \cdot \cos(2\pi \cdot r_2)$, де r_1 й r_2 — випадкові числа із ГВЧ$_{\text{рр}}$ [0; 1].

Можна також скористатися аналогічною формулою $Z = \sqrt{(-2 \cdot \mathrm{Ln}(r_1))} \cdot \sin(2\pi \cdot r_2)$, де r_1 й r_2 — випадкові числа із ГВЧ$_{\text{рр}}$ [0; 1].

Приклад. Матеріал надходить у цех один раз у добу по 10 штук відразу. Витрата матеріалу із цеху випадковий за нормальним законом з математичним очікуванням $m = 10$ і середньоквадратичним відхиленням $\sigma = 3.5$. Обчислити ймовірність дефіциту на складі при запасі матеріалу в початковий момент часу 20 штук.

При реалізації в середовищі моделювання Stratum рішення задачі буде виглядати в такий спосіб.

```
z := sqrt(–2 · ln(rnd)) · cos(2 · PI · rnd)
x := σ · z + m
запас := запас + 10
запас := (запас – x) · ed(запас – x)
дефицит := дефицит + not(ed(запас –
x))
N := N + 1
P := дефицит/N
stop(N > k)
```

z — нормальне нормалізоване випадкове число;
x — нормальне число, щоденна витрата матеріалів;
запас — стан складу: початок дня, моделювання приходу;
запас — стан складу: кінець дня, моделювання витрати;
дефіцит — лічильник днів, протягом яких спостерігався дефіцит;
N — кількість днів;
P — імовірність дефіциту;
k — моделювання протягом k днів.