Лекція 26.

Моделювання системи випадкових величин

Часто на практиці зустрічаються системи випадкових величин, тобто такі дві (і більше) різні випадкові величини X, Y (і інші), які залежать друг від друга. Наприклад, якщо відбулася подія X і прийняло якесь випадкове значення, то подія Y відбувається хоча й випадково, але з обліком того, що X уже прийняло якесь значення.

Наприклад, якщо в якості X випало велике число, то Y повинне випасти теж досить велике число (якщо кореляція позитивна). Досить імовірно, що якщо людина має велику вагу, то він, швидше за все, буде й великого росту. Хоча це НЕ ОБОВ'ЯЗКОВО, це НЕ ЗАКОНОМІРНІСТЬ, а кореляція випадкових величин. Тому що бувають, хоча й рідко, люди з більшою вагою, але невеликого росту або з маленькою вагою й високі. І всі таки основна маса гладких людей - високі, а низьких людей - мають малу вагу.

По визначенню, якщо випадкові величини незалежні, то $f(\mathbf{x}) = f(x_1) \cdot f(x_2) \cdot \ldots \cdot f(x_n)$...

x_i — випадкова незалежна величина;

$f(x_i)$ — щільність імовірності випадання випадкової незалежної величини x_i;

$f(\mathbf{x})$ — щільність імовірності випадання вектора $\mathbf{x}$ випадкових незалежних величин $x_1, x_2, \ldots, x_n$...

Якщо випадкові величини залежні, то $f(\mathbf{x}) = f(x_1) \cdot f(x_2 \mid x_1) \cdot f(x_3 \mid x_2, x_1) \cdot \ldots \cdot f(x_n \mid x_{n-1}, x_{n-2}, \ldots, x_2, x_1)$...

$x_j \mid x_{j-1}, \ldots, x_1$ — випадкові залежні величини: випадання x_j за умови, що випали $x_{j-1}, \ldots, x_1$;

$f(x_j \mid x_{j-1}, \ldots, x_1)$ — щільність умовної ймовірності появи x_j, якщо випали $x_{j-1}, \ldots, x_1$;

$f(\mathbf{x})$ — імовірність випадання вектора $\mathbf{x}$ випадкових залежних величин.

Нехай, приміром, є дві залежних події — X й Y, розподілених за нормальним законом. X має математичне очікування m_x і середньоквадратичне відхилення σ_x. Y має математичне очікування m_y і середньоквадратичне відхилення σ_y. Коефіцієнт кореляції — q — показує, наскільки тісно зв'язані події X й Y. Якщо коефіцієнт кореляції дорівнює одиниці, то залежність подій X й Y взаємно однозначна: одному значенню X відповідає одне значення Y (див. мал. 26.1).

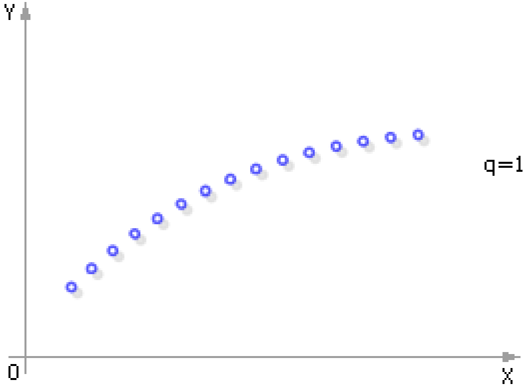

Рис. 26.1. Вид залежності двох випадкових величин при позитивному коефіцієнті кореляції (q=1)

При q близьких до одиниці виникає картина, показана на мал. 26.2, тобто одному значенню X можуть відповідати вже кілька значень Y (точніше, одне з декількох значень Y, обумовлене випадковим способом); у цьому випадку події X й Y менш корельовані, менш залежні друг від друга.

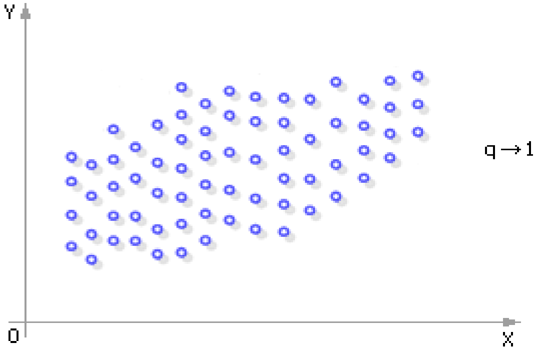

Рис. 26.2. Вид залежності двох випадкових величин
при позитивному коефіцієнті кореляції (0 < q < 1)

І, нарешті, коли коефіцієнт кореляції прагне до нуля, виникає ситуація, при якій будь-якому значенню *X* може відповідати будь-яке значення *Y*, тобто події *X* й *Y* не залежать або майже не залежать друг від друга, не корелюють один з одним (див. мал. 26.3).

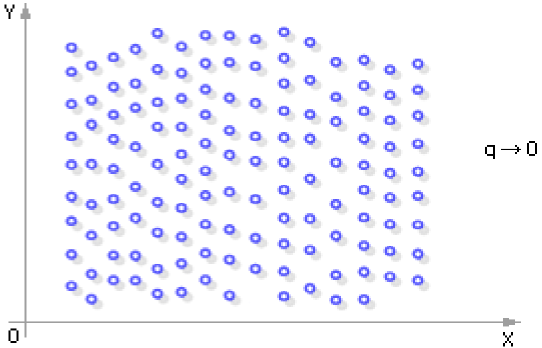

Рис. 26.3. Вид залежності двох випадкових величин при коефіцієнті кореляції близькому до нуля (q –> 0)

На всіх графіках кореляція була прийнята позитивною величиною. Якщо $q < 0$, то графіки будуть виглядати так, як показано на мал. 26.4.

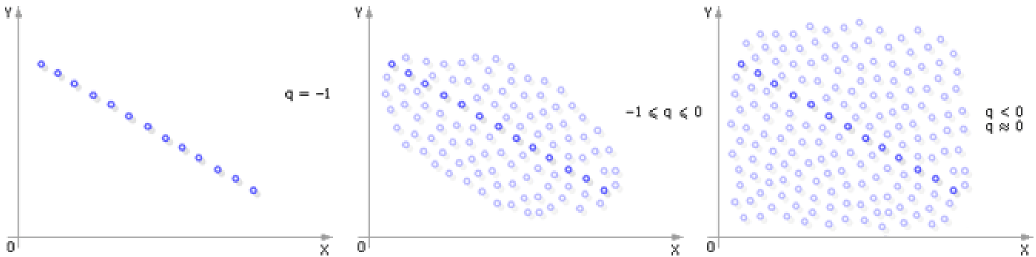

Рис. 26.4. Вид залежності двох випадкових величин
при негативному коефіцієнті кореляції
а) q = –1; б) –1 < q < 0; в) q –> 0

Насправді випадкові події (X й Y) не можуть приймати з рівною ймовірністю будь-які значення, як це має місце на мал. 26.2. Приміром, у групі студентів не може бути людей понадмалого або надвеликого росту; в основному, люди мають якийсь середній ріст і розкид навколо цього середнього росту. Тому на одних ділянках осі X кількість подій розташована густіше, на інші — рідше. (Щільність випадкових подій, кількість крапок на графіках більше поблизу величин m_x). Те ж саме вірно й для Y. І тоді мал. 26.2 можна зобразити більш точно, так, як показано на мал. 26.5.

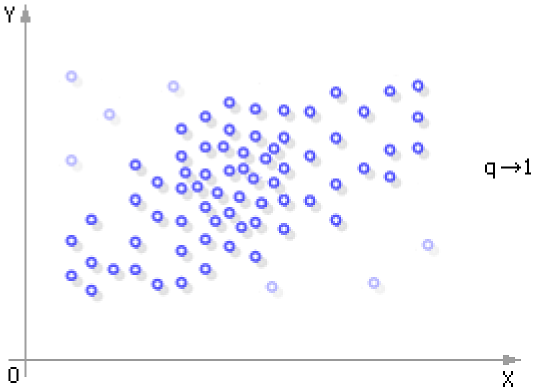

Рис. 26.5. Ілюстрація системи випадкових залежних величин

Для приклада візьмемо нормальний розподіл, як найпоширеніше. Математичне очікування вказує на самі ймовірні події, тут число подій більше й графік подій густіше. Позитивна кореляція вказує (див. мал. 26.2), що більші випадкові величини X викликають до генерації більші Y. Негативна кореляція вказує (див. мал. 26.4), що більші випадкові величини X стимулюють до генерації менші випадкові величини Y. Нульова й близька до нуля кореляція показує (див. мал. 26.3), що величина випадкової величини X ніяк не пов'язана з певним значенням випадкової величини Y. Легко зрозуміти сказане, якщо уявити собі спочатку розподілу $f(X)$ і $f(Y)$ окремо, а потім зв'язати їх у систему (див. мал. 26.6, мал. 26.7 і мал. 26.8).

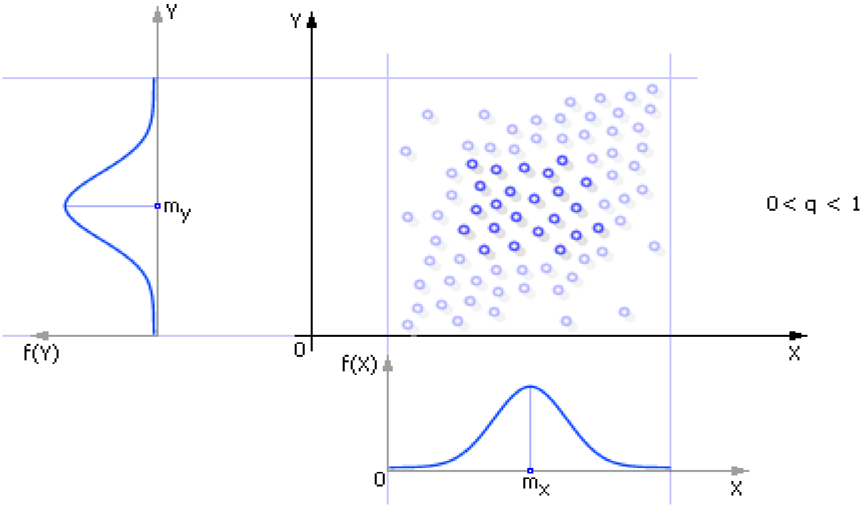

Рис. 26.6. Генерація системи випадкових величин при позитивному коефіцієнті кореляції

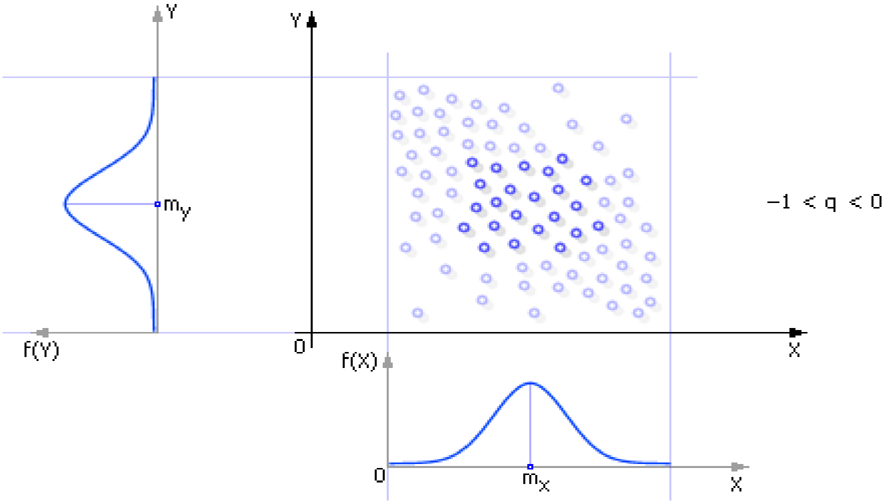

Рис. 26.7. Генерація системи випадкових величин
при негативному коефіцієнті кореляції

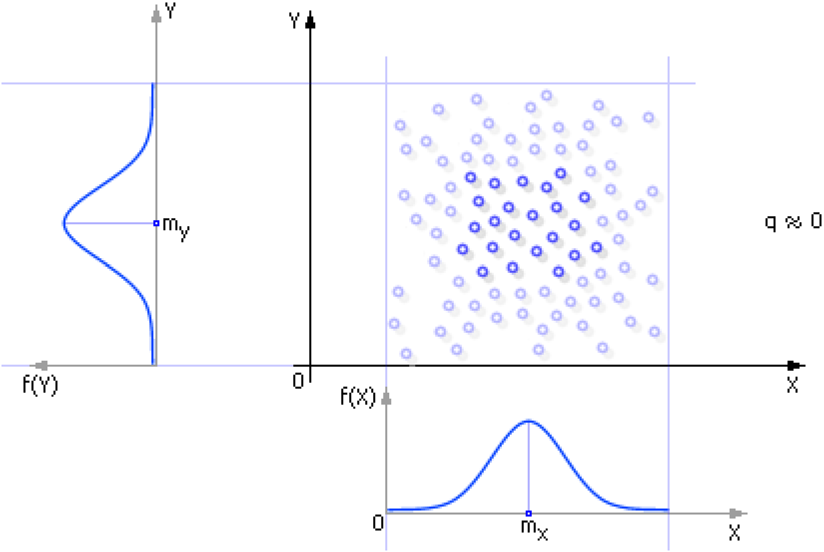

Рис. 26.8.

Приклад реалізації алгоритму моделювання двох залежних випадкових подій X й Y.

Умова: допустимо, що X й Y розподілено за нормальним законом з відповідними значеннями m_x, σ_x й m_y, σ_y. Задано коефіцієнт кореляції двох випадкових подій q, тобто випадкові величини X й Y залежні друг від друга, Y не зовсім випадково.

Тогда можливий алгоритм реалізації моделі буде наступним.

1. Розігрується шість випадкових рівномірно розподілених на інтервалі $[0; 1]$ чисел b_1, b_2, b_3, b_4, b_5, b_6; перебуває їхня сума S: $S = b_1 + b_2 + b_3 + b_4 + b_5 + b_6$. Перебуває нормально розподілене випадкове число x по наступній формулі: $x = \mathrm{sqrt}(2) \cdot \sigma_x \cdot (S - 3) + m_x$, див. лекцію 25.
2. По формулі $m_{y/x} = m_y + q \cdot \sigma_y/\sigma_x \cdot (x - m_x)$ перебуває математичне очікування $m_{y/x}$ (знак y/x означає, що y буде приймати випадкові значення з урахуванням умови, що x уже прийняв якісь певні значення).
3. По формулі $\sigma_{y/x} = \sigma_y \cdot \mathrm{sqrt}(1 - q^2)$ перебуває середньоквадратичне відхилення $\sigma_{y/x}$ (знак y/x означає, що y буде приймати випадкові значення з урахуванням умови, що x уже прийняв якісь певні значення).
4. Розігрується шість випадкових рівномірно розподілених на інтервалі $[0; 1]$ чисел r_1, r_2, r_3, r_4, r_5, r_6; перебуває їхня сума k: $k = r_1 + r_2 + r_3 + r_4 + r_5 + r_6$. Перебуває нормально розподілене випадкове число y по наступній формулі: $y = \mathrm{sqrt}(2) \cdot \sigma_{y/x} \cdot (k - 3) + m_{y/x}$.