

Основные статистики

Среднее значение

Среднее значение μ_x набора из n измеренных значений для случайной величины X находится как

$$\mu_x = \sum_{i=1}^N \frac{x_i}{n}. \quad (1)$$

Среднее значение также называют первым центральным моментом или центром гравитации, функции плотности вероятности, которая может быть вместе с дисперсией или вторым центральным моментом, единственно практически доступной информацией по данным свойств грунтов.

Среднее значение случайной величины, также определяемое как ожидаемое или центральное значение, – это число, которое случайная величина принимает в среднем и находится, если все возможные значения случайной величины умножить на их вероятность возникновения, а затем просуммировать:

$$E(x) = \int xf(x)dx \approx \sum xp(x_i),$$

(2)

где $f(x)$ – функция плотности вероятности x для непрерывных случайных величин; $p(x_i)$ – вероятность значения x_i для дискретных случайных величин.

Среднее значение может быть рассчитано из репрезентативных данных, оно обеспечивает объективную оценку ожидаемого значения параметра; следовательно, среднее и ожидаемое значение численно совпадают.

Дисперсия

Дисперсия $\text{var}(x)$ случайной величины X представляет собой ожидаемое значение квадрата разницы между случайной величиной и ее средним значением. Там, где доступны фактические данные, дисперсия данных может быть рассчитана следующим образом:

$$\begin{aligned} \text{var}(x) &= E\left[(x - \mu_x)^2\right] = \int (x - \mu_x)^2 f(x) dx = \\ &= \frac{\left[(x - \mu_x)^2\right]}{n}. \end{aligned} \quad (3)$$

Приведенная выше форма суммирования, включающая член x_i , дает дисперсию выборки, содержащей ровно n элементов. Обычно выборка размера n используется для получения оценки дисперсии связанной случайной величины, которая представляет всю совокупность элементов или континуум материала. Чтобы получить объективную оценку, n заменяется на $n - 1$:

$$\text{var}(x) = \frac{\left[(x - \mu_x)^2 \right]}{n - 1}. \quad (4)$$

Дисперсия (или стандартное отклонение) является мерой того, разбросаны ли значения случайной величины относительно ее среднего значения. Если различные возможные значения случайной величины сосредоточены близко к среднему значению, то дисперсия мала, а если достаточно разбросаны, то дисперсия велика (рис. 1).

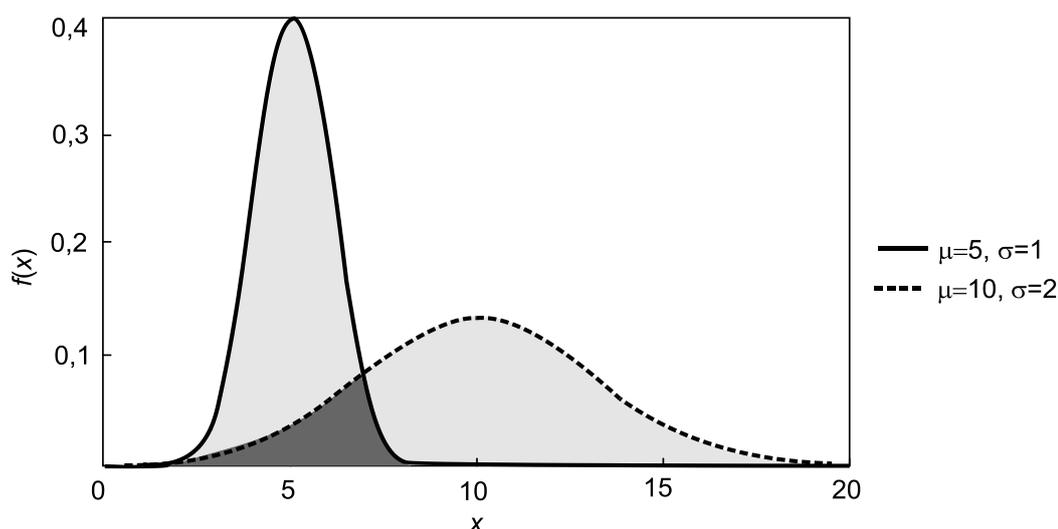


Рис. 1. Влияние среднего значения и типичной дисперсии для двух нормальных распределений

Стандартное отклонение

Для того, чтобы выразить разброс или дисперсию случайной величины относительно ее ожидаемого значения в тех же единицах, что и сама случайная величина, стандартное отклонение σ_x находится как квадратный корень из дисперсии; таким образом:

$$\sigma_x = \sqrt{\text{var}(x)}. \quad (5)$$

Коэффициент вариации

Для того чтобы обеспечить удобное безразмерное выражение неопределенности, присущей случайной величине, стандартное отклонение делится на среднее, чтобы получить коэффициент вариации COV или относительное стандартное отклонение, который обычно выражается в процентах. Коэффициент вариации определяется как отношение стандартного отклонения σ к среднему μ :

$$\text{COV}(x) = \frac{\sigma_x}{\mu_x} 100\%. \quad (6)$$

Асимметрия

В теории вероятностей и статистике асимметрия является мерой несимметричного распределения вероятностей действительно значимой случайной величины. Значение асимметрии может быть положительным или отрицательным или даже неопределенным. Качественно отрицательная асимметрия указывает на то, что хвост в левой части функции плотности вероятности длиннее, чем в правой части, и основная часть значений (возможно, включая медиану) лежит справа от среднего. Положительная асимметрия указывает на то, что хвост справа длиннее, чем слева, и основная часть значений лежит слева от среднего (рис. 2 а,б). Нулевое значение указывает на то, что значения относительно равномерно распределены по обе стороны от среднего, что обычно, но не обязательно подразумевает симметричное распределение.

Асимметрия случайной величины X – это третий стандартизированный момент, обозначаемый v и определяемый как

$$v_x = \int_{-\infty}^{+\infty} \frac{(x - \mu_x)^3}{\sigma_x^3} f(x) dx, \quad (7)$$

где $f(x)$ — функция плотности вероятности случайной величины x .

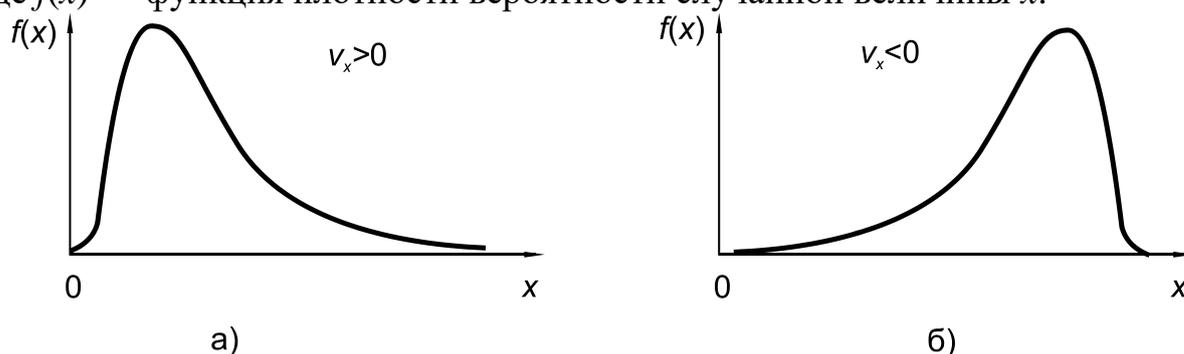


Рис. 2. Пример положительно (а) и отрицательно (б) асимметричных распределений

На рис. 2б показан случай положительно асимметричного распределения, которое круче для низких значений случайной величины и скошенное для больших значений. Отрицательно скошенное распределение, как на рис. 2а, является плоским для малых значений случайной величины и крутым для больших значений.

Среднее значение, стандартное отклонение и коэффициент вариации взаимосвязаны: зная любые два из них, можно найти и третий. На практике удобным способом оценки параметров при наличии небольшого количества данных является допущение, что коэффициент вариации аналогичен ранее измеренным значениям из других наборов данных того же параметра.

Коэффициент автокорреляции

Коэффициент автокорреляции, ρ_k , оценивает корреляцию между любыми двумя случайными наблюдениями, разделенными расстоянием (лагом) в k единиц:

$$r_k = \frac{c_k^*}{c_0^*} = \frac{\sum_{i=1}^{n-k} (X_i - \bar{X})(X_{i+k} - \bar{X})}{\sum_{i=1}^n (X_i - \bar{X})^2}, \quad (8)$$

где (\bar{X}) – среднее значение наблюдений X_1, X_2, \dots, X_n ; и $0 \leq k < n$.

Масштаб флуктуации или длина корреляции

Длина корреляции θ , известная также как масштаб оценивает степень пространственной зависимости параметров грунта посредством различных автокорреляционных функций (см. таблицу 1). Длина корреляции определяет расстояние, в пределах которого значения (точки в пространстве)

Таблица 1

Тип автокорреляционной функции и длина корреляции (Vanmarcke, 1983a)

Автокорреляционная функция θ		
$\rho(\tau) = \begin{cases} 1 - \frac{ \tau }{a} & \text{при } \tau < a \\ 0 & \text{при } \tau \geq a \end{cases}$	Билинейная модель	a
$\rho(\tau) = \exp\left(\frac{- \tau }{b}\right)$	Модель Маркова	$2b$
$\rho(\tau) = \exp\left(-\left(\frac{ \tau }{c}\right)^2\right)$	Модель Гаусса	$\sqrt{\pi c}$

выборки значительно коррелируют (обычно более чем на 10%). Функция автокорреляции – это математическое выражение, определяющее корреляцию между двумя точками в пространстве. Точки, разделенные расстоянием, большее θ , становится значительно некоррелированным. Поскольку длина корреляции должна быть конечной, чтобы быть значимой для инженерных приложений, корреляционные функции должны быть строго неотрицательными. В таблице 1 приведены три обычно используемые функции автокорреляции, которые удовлетворяют этому критерию. Модель Маркова иногда полезна при двумерном анализе, когда предполагается, что вертикальные и горизонтальные корреляции независимы. В этом случае автокорреляция функция может быть выражена в разделяемой форме:

$$\rho(\tau) = \exp\left(-\frac{\tau_h}{\theta_h} - \frac{\tau_v}{\theta_v}\right), \quad (9)$$

где τ_h и τ_v – расстояния в горизонтальном и вертикальном направлениях между двумя точками выборки, а θ_h или δ_x , θ_v или δ_z – длина корреляции (масштаб флуктуации) в горизонтальном и вертикальном направлениях, соответственно.

Из-за особенностей формирования отложений грунтов длина горизонтальной корреляции обычно больше длины вертикальной корреляции для большинства свойств грунтов.