Нормальное распределение

На этом уроке мы

- Определим, что такое нормальное распределение и посмотрим на его свойства
- Познакомимся с правилом двух и трех сигм — оно часто бывает полезно в статистике
- А также изучим логику арифметических операций с нормальным распределением

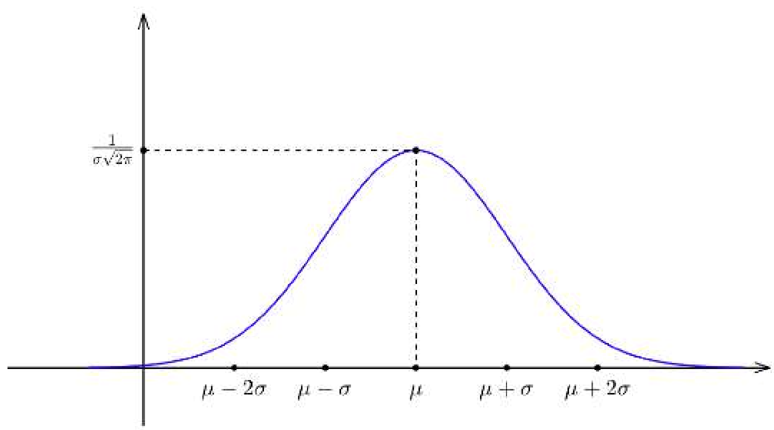
Материал этого урока будет полезен и сам по себе, безотносительно следующих уроков. Часто на практике приходится иметь дело с суммой многих случайных величин. Многие случайные величины, которые вы встретите на практике, имеют нормальное распределение — поэтому мы будем учиться с ним работать.

Нормальное распределение

В жизни многие величины имеют *нормальное распределение*. Например, близкое к нормальному распределение имеет

- рост взрослых женщин и рост взрослых мужчин,
- погрешность измерения приборов,
- оценки школьников на ЕГЭ по математике.

На следующем шаге мы дадим определение нормального распределения, выписав его формулу плотности. А пока ограничимся картинкой:



Нормальное распределение имеет два параметра – μ и σ^2, где μ – математическое ожидание (среднее), а σ^2 – дисперсия. Нормальное распределение обозначается $N(\mu, \sigma^2)$.

Нормальное распределение также называют *распределением Гаусса*.

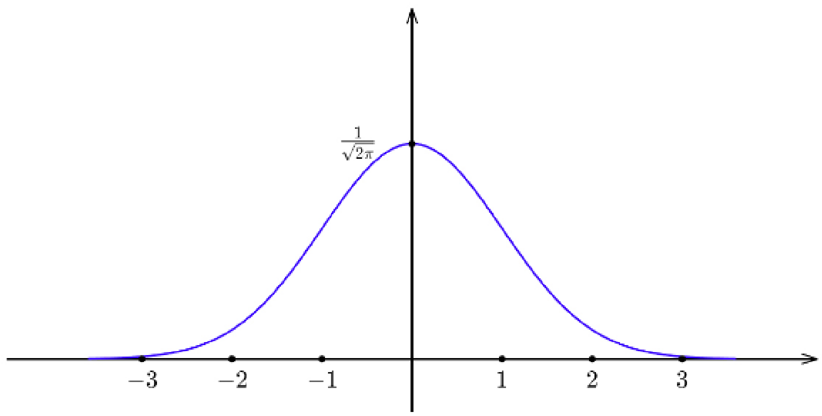
Распределение $N(0,1)$ — стандартное нормальное распределение

Давайте начнём со случая, когда математическое ожидание равно 0, а дисперсия равна 1.

Нормальное распределение с математическим ожиданием 0 и дисперсией 1 обозначается $N(0,1)$. Его называют *стандартным* нормальным распределением, потому что с ним удобнее всего работать. Вот его функция плотности распределения:

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{\left(-\frac{1}{2}x^2\right)},$$

вот график этой функции плотности:



Плотность нормального распределения в общем случае

Комментарий. На практике мы никогда не будем самостоятельно брать интеграл, использующий функцию плотности нормального распределения. За нас это сделает компьютер. Нам же достаточно понять, как правильно выписать формулу: куда какие коэффициенты ставятся и почему.

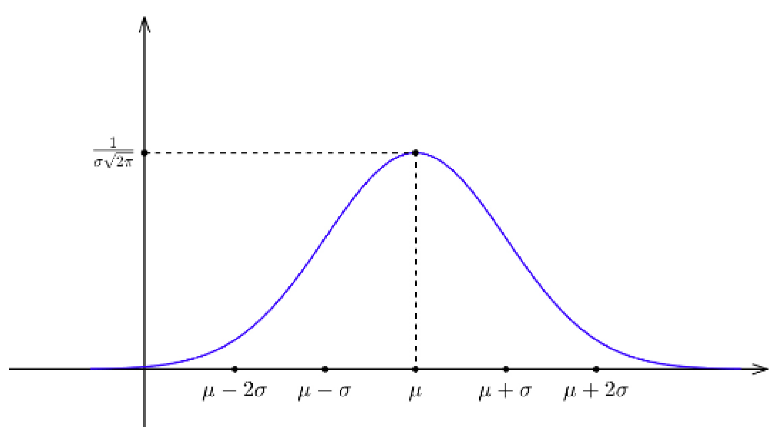
Формула плотности вероятности нормального распределения со средним μ и дисперсией σ^2 такая:

$$p_\xi(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

На первый взгляд выглядит пугающе. Но μ и σ — это константы для фиксированно распределения, поэтому по существу формула мало отличается от $N(0,1)$. Проверьте, что если подставить в эту формулу $\mu = 0$ и $\sigma^2 = 1$, то получится формула из предыдущего параграфа.

Думать про нормальное распределение $N(\mu, \sigma^2)$ можно так. Если X имеет распределение $N(0,1)$, то $\sigma X + \mu$ имеет распределение $N(\mu, \sigma^2)$. В формуле $\sigma X + \mu$ умножение на σ изменяет дисперсию с 1 на σ^2, а прибавление μ изменяет математическое ожидание с 0 на μ.

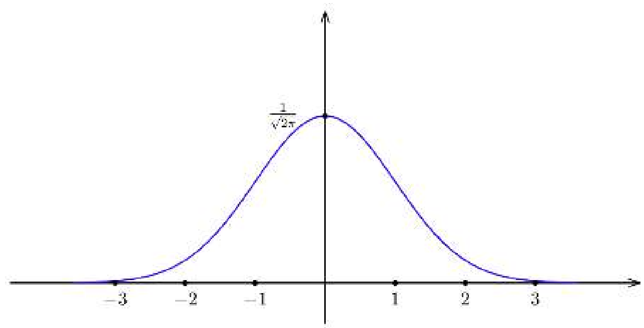
Вот так выглядит график функции плотности нормального распределения:

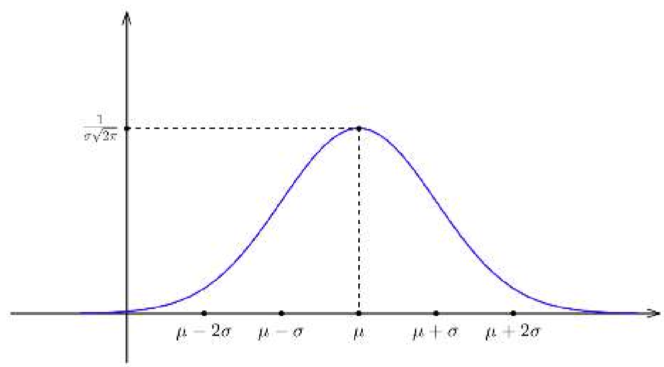


На самом деле, по картинке видно, что в данном случае $\mu > 0$, но в общем случае это, конечно, не обязательно так.

Как думать про график функции плотности $N(\mu, \sigma^2)$?

Снова нарисуем графики $N(0, 1)$ и $N(\mu, \sigma^2)$:



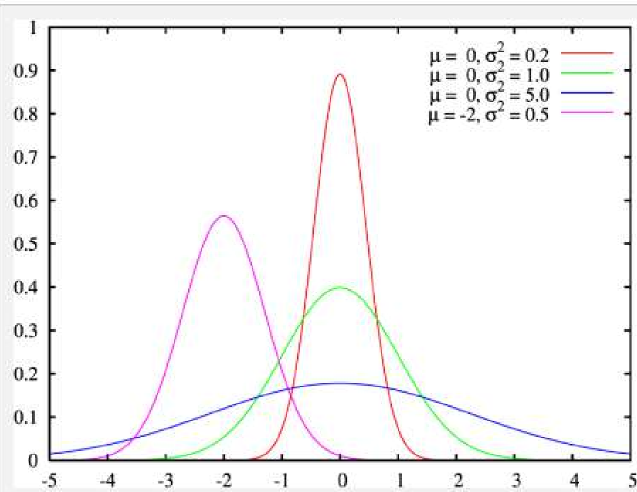


Ещё раз напомним формулу функции плотности $N(\mu, \sigma^2)$:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}.$$

1. Сначала мы взяли график плотности распределения $N(0, 1)$, то есть график функции $\frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}$.
2. Затем растянули его по оси OX в σ раз (за это отвечает знаменатель в $\frac{x-\mu}{\sigma}$)
3. Потом по оси OY сжали в σ раз (за это отвечает σ в выражении $\frac{1}{\sigma\sqrt{2\pi}}$)
4. И сдвинули его на μ вправо (за это отвечает μ в $\frac{x-\mu}{\sigma}$)).
5. Получился график плотности распределения $N(\mu, \sigma^2)$.

На картинке ниже изображены сразу несколько плотностей нормального распределения с различными значениями параметров μ и σ^2 для сравнения:



Плотность нормального распределения
в зависимости от параметров. Источник: Википедия

Снова напомним формулу для стандартного нормального распределения:

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{\left(-\frac{1}{2}x^2\right)}$$

Вычислите значения плотности стандартного нормального распределения в следующих точках, округлив ответы до тысячных:

Заполните пропуски

1. $f(0) =$ ______ .
2. $f(2) =$ ______ .
3. $f(3) =$ ______ .
4. $f(-3) =$ ______ .

Формула плотности нормального распределения в общем случае:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

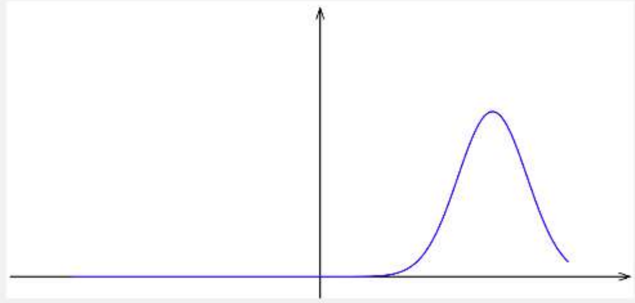
Пусть $\mu = 3, \sigma = 2$. Чему равна плотность в точке $x = 5$?

Ответ округлите до тысячных.

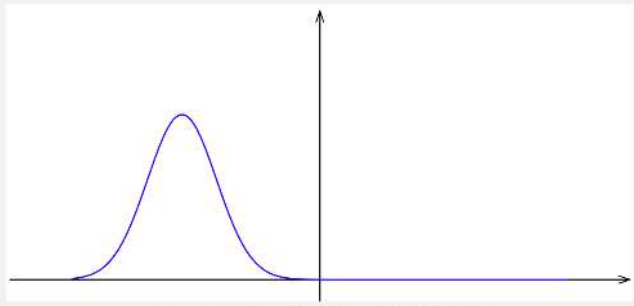
Введите численный ответ

Введите число

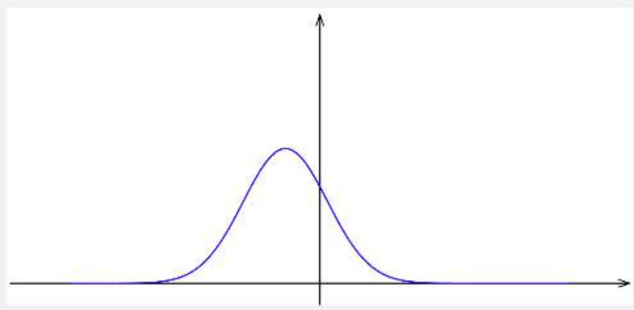
Вот плотности нескольких нормальных распределений. Сопоставьте их со значениями параметров:



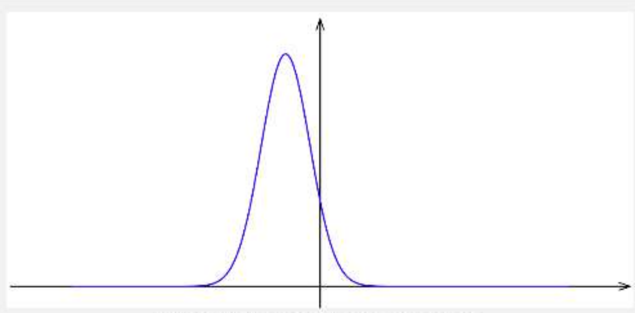
Нормальное распределение Вася



Нормальное распределение Артур



Нормальное распределение Моника



Нормальное распределение Лена

Сопоставьте значения из двух списков

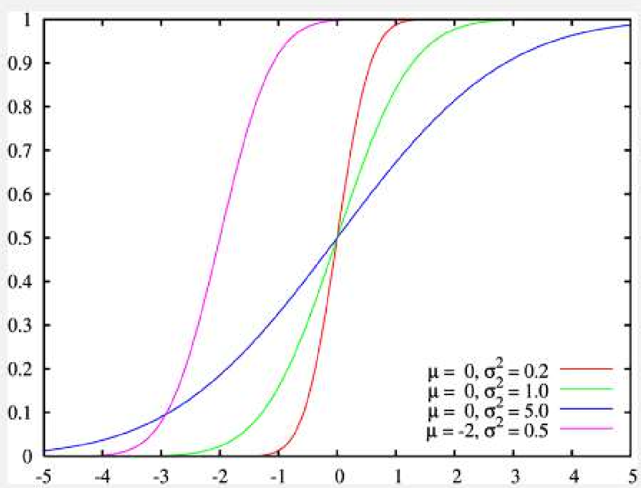
Нормальное распределение Вася	$N(-4, 1)$
Нормальное распределение Лена	$N(5, 1)$
Нормальное распределение Артур	$N(-1, 1.5)$
Нормальное распределение Моника	$N(-1, 0.5)$

Функция распределения нормального распределения

А вот с функцией распределения ситуация интересная. Как мы помним, функция распределения это интеграл плотности. Интеграл плотности нормального распределения не берётся в элементарных функциях. То есть плотность интегрируема, но интеграл невозможно выразить через арифметические, показательные, степенные, логарифмические, тригонометрические, обратные тригонометрические функции и их конечную композицию. То, что это нельзя сделать — строго доказанный факт, можно не пытаться это сделать :) То есть выписать формулу для функции распределения нормального распределения не получится.

Впрочем, при работе с нормальным распределением часто достаточно его свойств. В случае необходимости на помощь приходят численные методы.

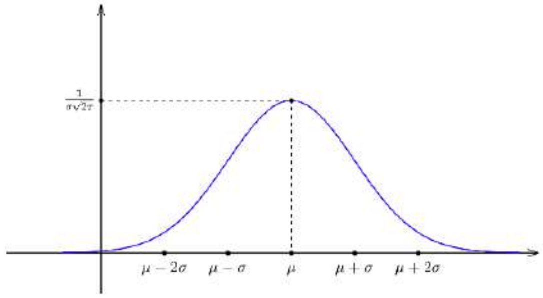
А вот графики функций распределения для разных параметров μ и σ:



Функция распределения нормального распределения
в зависимости от параметров. Источник: Википедия

Свойства нормального распределения

Обычно по плотности вероятности легче представить себе случайную величину, чем по функции распределения. Помогает держать в голове аналогию с гистограммами. Там где график плотности выше, значения более вероятны. Посмотрим ещё раз на плотность $N(\mu, \sigma^2)$



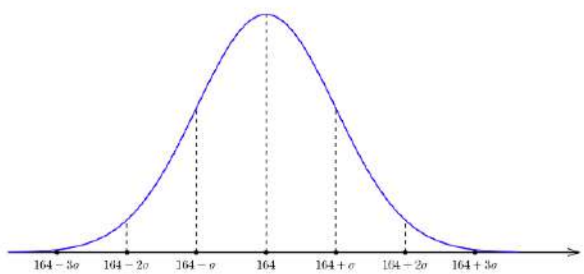
Как мы помним, дисперсия σ^2 это квадрат отклонения. Значит, σ в таком случае – это стандартное отклонение (мы предполагаем, что σ неотрицательна). По графику плотности видно, что:

- Нормальное распределение симметрично: то есть отклонение от среднего в большую сторону так же вероятно, как и отклонение в меньшую сторону
- Бо́льшая часть значений сконцентрирована вокруг среднего – мат.ожидания μ
- При этом небольшие отклонения от среднего вполне вероятны
- Чем меньше дисперсия σ^2, тем ближе к мат.ожиданию μ сконцентрированы значения случайной величины – впрочем, это уже скорее свойство дисперсии, чем нормального распределения.

Пример

Как мы уже упоминали, рост взрослых женщин и рост взрослых мужчин имеет нормальное распределение. Соотнося это со свойствами, описанными выше:

- У большинства рост достаточно близок к среднему по популяции
- Чем дальше рост от среднего, тем меньше доля людей с таким ростом

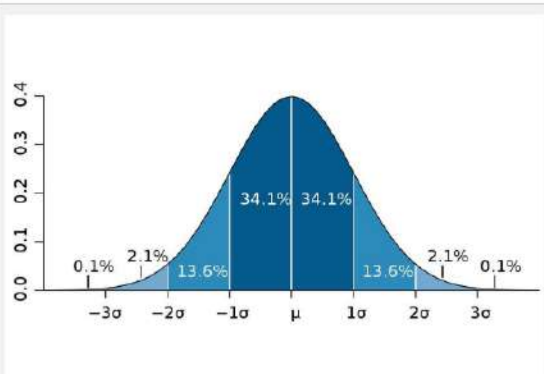


Правила двух и трех сигм

Вот одно из важных свойств нормального распределения, которые помогают с ним работать. Пусть случайная величина ξ имеет нормальное распределение со средним μ и дисперсией σ^2. Тогда:

- $P(\mu - \sigma < \xi < \mu + \sigma\ ) = 0.682 \cdots \approx 0.68$
- $P(\mu - 2\sigma < \xi < \mu + 2\sigma\ ) = 0.954 \cdots \approx 0.95$. Другими словами, вероятность получить результат, отклоняющийся от μ хотя бы на 2σ, меньше 0.05. Это называется *правилом двух сигм*. Или правилом двух стандартных отклонений (если в вашей задаче стандартное отклонение обозначено другой буквой).
- $P(\mu - 3\sigma < \xi < \mu + 3\sigma\ ) = 0.997 \cdots \approx 0.99$. Другими словами, вероятность получить результат, отклоняющийся от μ хотя бы на 3σ, меньше 0.01. Это называется *правилом трёх сигм*. Или правилом трёх стандартных отклонений (если в вашей задаче стандартное отклонение обозначено другой буквой).

Применять его мы будем в уроке про ЦПТ. Вот картинка:

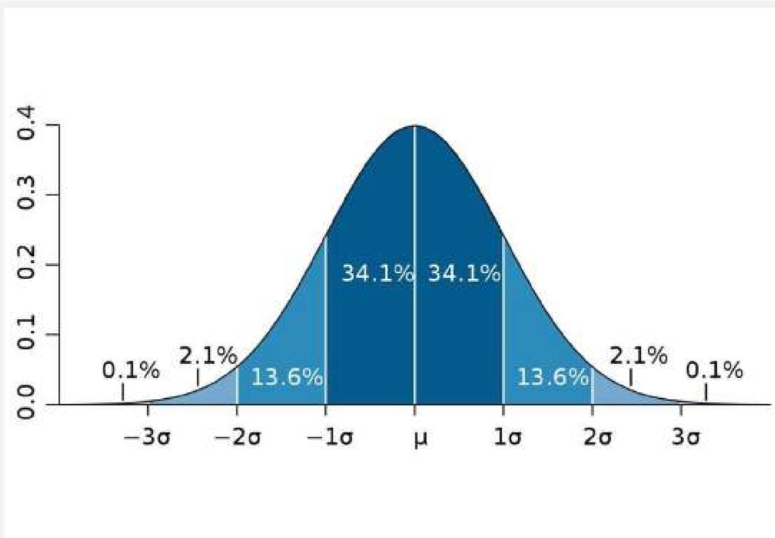


Источник: news.mit.edu

Выпишем ещё раз правила двух и трёх сигм с предыдущего шага.

Пусть случайная величина ξ имеет нормальное распределение со средним μ и дисперсией σ^2. Тогда:

- $P(\mu - \sigma < \xi < \mu + \sigma\ ) = 0.682\cdots \approx 0.68$
- $P(\mu - 2\sigma < \xi < \mu + 2\sigma\ ) = 0.954\cdots \approx 0.95$. Другими словами, вероятность получить результат, отклоняющийся от μ хотя бы на 2σ, меньше 0.05. Это называется *правилом двух сигм*. Или правилом двух стандартных отклонений (если в вашей задаче стандартное отклонение обозначено другой буквой).
- $P(\mu - 3\sigma < \xi < \mu + 3\sigma\ ) = 0.997\cdots \approx 0.99$. Другими словами, вероятность получить результат, отклоняющийся от μ хотя бы на 3σ, меньше 0.01. Это называется *правилом трёх сигм*. Или правилом трёх стандартных отклонений (если в вашей задаче стандартное отклонение обозначено другой буквой).



Источник: news.mit.edu

Выберите все подходящие ответы из списка

Пусть ξ имеет распределение $N(-5, 9)$. Тогда вероятность того, что ξ лежит в интервале $(-23, 13)$, примерно равна 0.68.

Пусть ξ имеет распределение $N(6, 4)$. Тогда вероятность того, что ξ лежит в интервале $(0, 12)$, примерно равна 0.99.

Пусть ξ имеет распределение $N(10, 4)$. Тогда вероятность того, что ξ лежит в интервале $(8, 10)$, примерно равна 0.34.

Пусть ξ имеет распределение $N(3, 25)$. Тогда вероятность того, что ξ лежит в инервале $(3, 8)$, примерно равна 0.68.

Операции над нормально распределёнными случайными величинами

Обозначение. Случайная величина, имеющая распределение $N(\mu, \sigma^2)$ для каких-то μ и σ, называется *нормально распределённой.*

Вот важное свойство нормально распределённых случайных величин.

Утверждение. Сумма нескольких совместно независимых нормально распределённых случайных величин – это тоже нормально распределенная случайная величина.

Доказывать это утверждение мы не будем. Благодаря этому утверждению и формулам для математического ожидания и дисперсии из предыдущего урока мы можем проводить вычисления с нормально распределёнными величинами. Давайте посмотрим сначала на более общий, а потом на более частный пример.

Примеры

Пример 1. Даны две независимые нормально распределённые величины X и Y. Их распределения это $N(\mu_1, \sigma_1^2)$ и $N(\mu_2, \sigma_2^2)$ соответственно. Давайте найдём распределение суммы этих величин $X + Y$.

- По свойству математического ожидания $E[X + Y] = E[X] + E[Y]$, то есть $E[X + Y] = \mu_1 + \mu_2$.
- Так как X и Y независимы, выполнено $Var(X + Y) = Var(X) + Var(Y) = \sigma_1^2 + \sigma_2^2$.
- Благодаря утверждению из начала параграфа мы знаем, что $X + Y$ имеет нормальное распределение. Так как мы уже нашли $E[X + Y]$ и $Var(X + Y)$, мы доказали, что $X + Y$ имеет распределение $N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$.

Пример 2. Случайные величины X и Y независимы и имеют распределения $N(3, 5)$ и $N(2, 8)$. Найдём распределение величины $X + Y$. Пользуясь результатом Примера 1 мы получаем, что случайная величина $X + Y$ имеет распределение $N(3 + 2, 5 + 8) = N(5, 13)$.

Пример 3. Случайные величины X и Y имеют распределения $N(3, 5)$ и $N(2, 8)$. Найдём распределение случайной величины $2X + Y + 5$.

- Случайная величина $2X$ нормально распределена и имеет $E[2X] = 2E[X] = 2 \cdot 3 = 6$, $Var(2X) = 2^2 Var(X) = 20$. То есть её распределение это $N(6, 20)$.
- Значит, случайная величина $2X + Y + 5$ имеет распределение $N(E[2X] + E[Y] + E[5], Var(2X) + Var(Y) + Var(5)) = N(6 + 2 + 5, 20 + 8 + 0) = N(13, 28)$.

Как про это думать? Пусть мы знаем, что какая-то величина имеет нормальное распределение. Тогда нам достаточно найти её математическое ожидание и дисперсию, чтобы полностью определить её распределение. А вычислять математическое ожидание и дисперсию мы уже умеем.

Задача с проверкой. Нормальное распределение 1

Задача. Независимые случайные величины X и Y имеют распределения $N(4, 5)$ и $N(3, 9)$ соответственно.

1. Найдите распределение случайной величины $X + Y$.
2. Найдите распределение случайной величины $X + \frac{Y}{3} - 4$.

Проверка. Введите математическое ожидание и дисперсию случайных величин.

Заполните пропуски

1. $E[X + Y] =$ ____ , $Var[X + Y] =$ ____ .
2. $E[X + \frac{Y}{3} - 4] =$ ____ , $Var(X + \frac{Y}{3} - 4) =$ ____ .

Задача с проверкой. Нормальное распределение 2

Задача. Независимые случайные величины X, Y и Z имеют распределения $N(1,2), N(3,4)$ и $N(5,3)$ соответственно. Найдите распределение случайной величины $\frac{X+Y+Z}{3}+2.$

Проверка. Введите математическое ожидание и дисперсию случайной величины.

Заполните пропуски

$E[\frac{X+Y+Z}{3}+2] =$ ______ .

$Var(\frac{X+Y+Z}{3}+2) =$ ______ .

Задача с проверкой. Нормальное распределение 3

Задача. Все совместно независимые случайные величины $X_1, \ldots, X_n$ имеют одинаковое распределение $N(\mu, \sigma^2)$. Найдите распределение случайной величины $X_1 + \ldots + X_n$.

Проверка. Пусть независимые случайные величины $X_1, \ldots, X_{10}$ имеют одинаковое распределение $N(5, 2)$. Найдите распределение случайной величины $X_1 + \ldots + X_{10}$.

Заполните пропуски

$E[X_1 + \ldots + X_{10}] =$ ______ .

$Var(X_1 + \ldots + X_{10}) =$ ______ .

Задача с проверкой. Нормальное распределение 4

Задача. Все совместно независимые случайные величины $X_1, \ldots, X_n$ имеют одинаковое распределение $N(\mu, \sigma^2)$. Найдите распределение случайной величины $\frac{X_1 + \ldots + X_n}{n}$.

Проверка. Пусть независимые случайные величины $X_1, \ldots, X_{10}$ имеют одинаковое распределение $N(5, 2)$. Найдите распределение случайной величины $\frac{X_1 + \ldots + X_{10}}{10}$.

Заполните пропуски

$E[\frac{X_1 + \ldots + X_{10}}{10}] =$ ______ .

$Var(\frac{X_1 + \ldots + X_{10}}{10}) =$ ______ .

Что мы прошли на этом уроке

- Узнали, что такое нормальное распределение
- Поняли, как ведут себя нормально распределённые случайные величины при сложении и умножении на число
- В частности, обсудили, что сумма совместно независимых нормально распределённых случайных величин – это тоже нормально распределенная случайная величина

Что нас ждёт на следующем уроке

На следующем уроке мы

- познакомимся со статистическими тестами и проверкой гипотез
- узнаем, что такое статистика, уровень значимости, критическое множество и статистический критерий
- поймём, как проверить, насколько построенный тест хорош