Теория вероятностей и математическая статистика

Нормальное распределение.

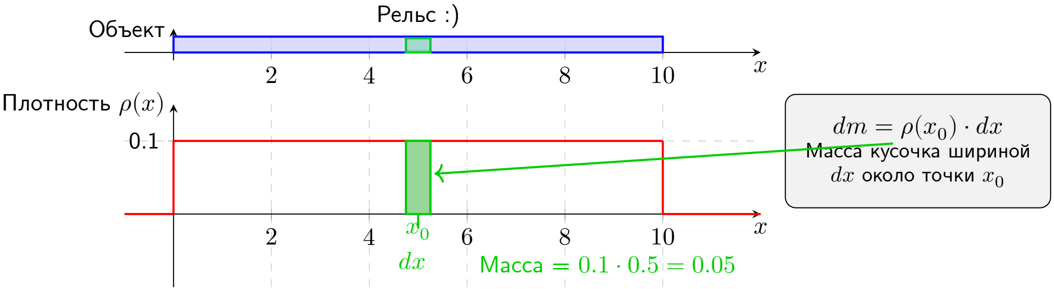
Глеб Карпов

ФКН ВШЭ

В прошлых сериях...

Физическая аналогия: значение плотности вещества в точке

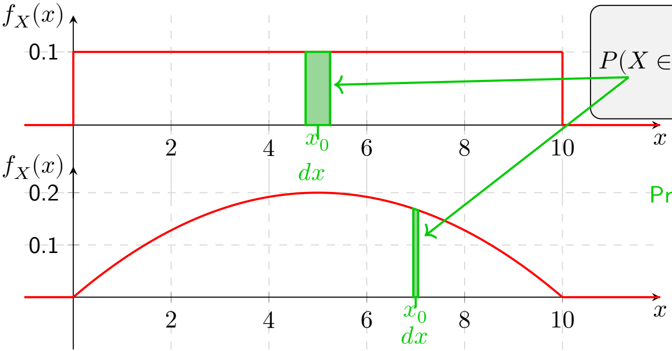
- Само по себе значение функции плотности $\rho(x_0)$ не обозначает массу в точке x_0. Смысл несет именно произведение плотности на длину, вспомним, чтобы сократились размерности $[kg] = \left[\frac{kg}{m}\right][m]$



В прошлых сериях...

Значение плотности вероятности в точке

- По аналогии с физикой: само по себе значение функции плотности $f_X(x_0)$ не обозначает вероятность в точке x_0. Смысл несет именно произведение плотности на длину интервала.

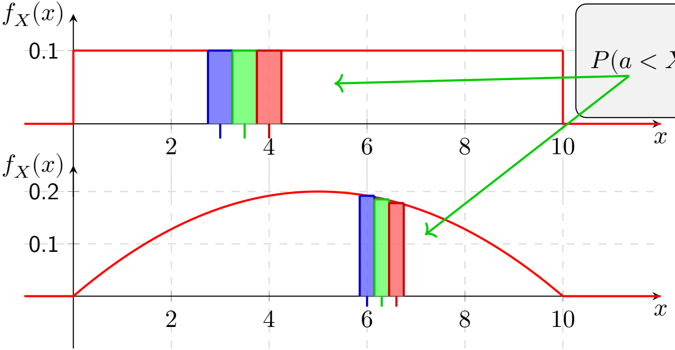


$$P(X \in (x_0 - \tfrac{dx}{2}, x_0 + \tfrac{dx}{2})) = f_X(x_0) \cdot dx$$

Probability = area = $f_X(x_0) \cdot dx$

Вероятность попадания в интервал

- Последовательно складывая такие вероятности попадания в окрестность, получаем площадь под графиком функции плотности от точки a до точки b, которая и интерпретируется нами как $P(a < X < b)$.



$$P(a < X < b) = \sum f_X(x_i)dx = \int\limits_a^b f_X(x)dx$$

Математическое ожидание

- Напомним, что для дискретной СВ мы вычисляем математическое ожидание как 'взвешенную сумму':

$$E[X] = \sum_{i=1}^{n} x_i P(X = x_i), \text{ вычисляется для } \forall x_i \in \Omega_X.$$

- Если у нас есть другая СВ G, являющаяся функцией от X: $G = g(X)$:

$$E[g(X)] = \sum_{i=1}^{n} g(x_i) P(X = x_i)$$

- Для непрерывного случая мы заменяем сумму на интеграл:

$$\sum_{i=1}^{n} x_i \, P(X = x_i) \Rightarrow \int_{-\infty}^{+\infty} x \, f_X(x)dx \qquad \sum_{i=1}^{n} g(x_i) \, P(X = x_i) \Rightarrow \int_{-\infty}^{+\infty} g(x) \, f_X(x)dx$$

- Финально:

$$\boxed{E[X] = \int_{-\infty}^{+\infty} x f_X(x)dx} \qquad \boxed{E[g(X)] = \int_{-\infty}^{+\infty} g(x) f_X(x)dx}$$

Нормальное распределение

- Функция плотности вероятности параметризована двумя константами μ и σ^2 и записывается как:

$$f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \ \forall x \in \mathbb{R}$$

- Параметры влияют на вид функции плотности и, как следствие, на "поведение" нормальной случайной величины. Чтобы подчеркнуть конкретную нормальную величину, с конкретно рассматриваемым набором параметров, мы пишем:

$$X \sim \mathcal{N}\left(\mu, \sigma^2\right)$$

Характеристики нормальной величины

Имеет место интересное совпадение параметров и характеристик нормальной случайной величины:

$$E[X] = \mu, \ Var[X] = \sigma^2$$

Будьте внимательны, в общем случае параметры и характеристики случайной величины совпадать не обязаны.

Влияние параметров на функцию плотности

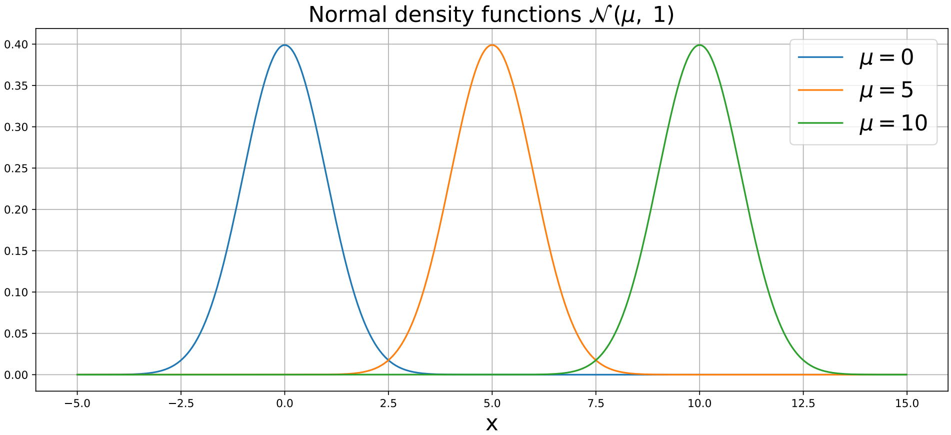


Рис. 1: Влияние параметра μ

Влияние параметров на функцию плотности

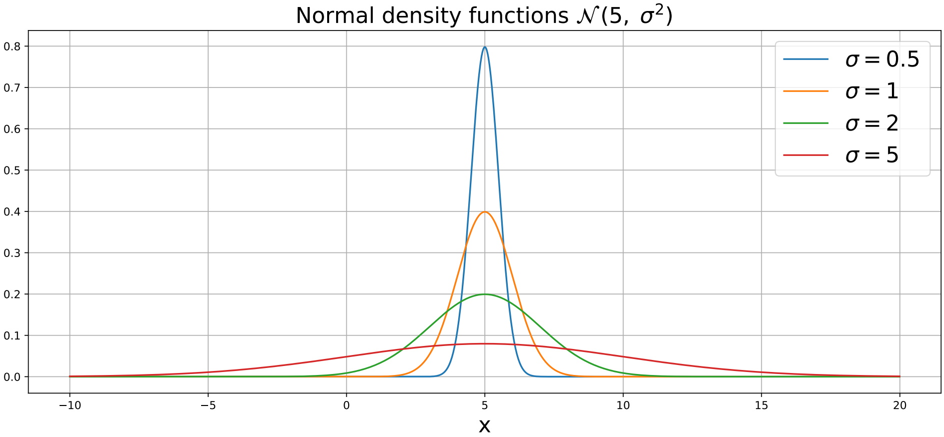


Рис. 2: Влияние параметра σ

Влияние параметров на функцию плотности

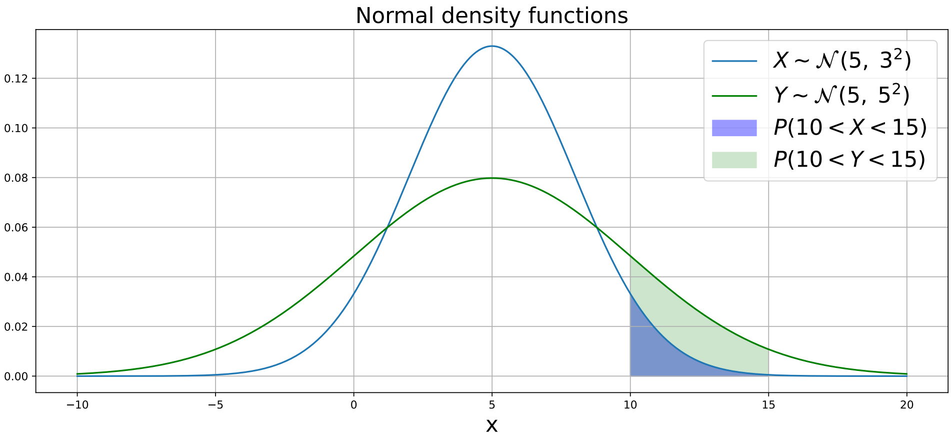


Рис. 3: Влияние параметра σ

Стандартная нормальная величина

Все нормальные равны, но одно равнее других

- Поскольку существует бесконечное число пар (μ, σ^2), существует также бесконечное число различных нормальных распределений.
- Это потребовало бы вычисления десятков сложных вероятностных интегралов каждый раз.
- Что если вычисление вероятности **любой** нормальной СВ можно свести к знанию всего одной переменной?

Стандартная нормальная величина

Все нормальные равны, но одно равнее других

- Встречайте стандартную нормальную случайную величину:

$$Z \sim \mathcal{N}(0,1), \quad f_Z(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}, \ \forall x \in \mathbb{R}$$

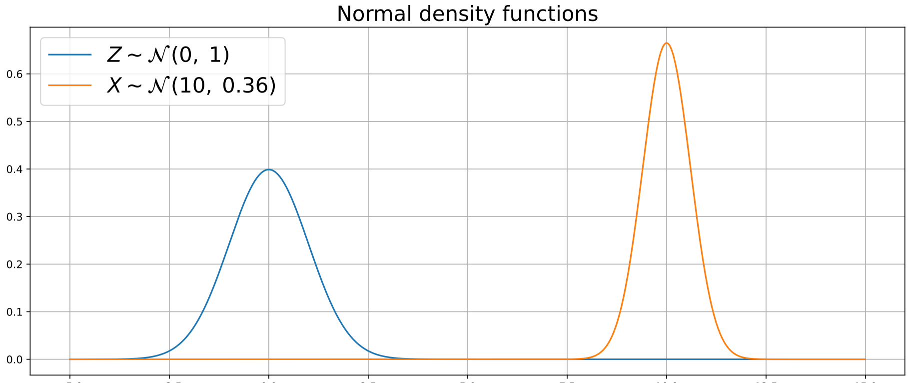
- Давным-давно математики вычислили десятки вероятностных интегралов и создали таблицу стандартного нормального распределения, которая показывает значения функции распределения $F_Z(x)$ для многих возможных значений x.
- Мы можем найти любую вероятность любой нормальной случайной величины, преобразовав её в стандартную нормальную.

Стандартная нормальная величина

Формула преобразования

- Пусть X — нормальная случайная величина и $X \sim \mathcal{N}(\mu, \sigma^2)$.
- Тогда следующая функция (преобразование) преобразует X в стандартную нормальную величину:

$$Z = \frac{X - \mu}{\sigma}$$

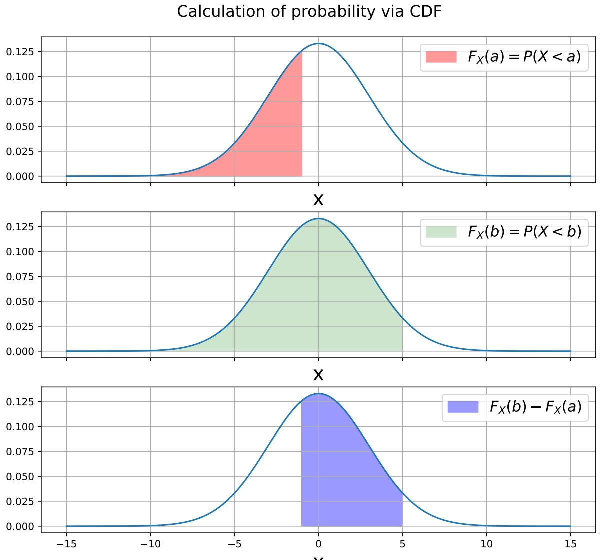


Стандартная нормальная величина

- Пусть X — нормальная случайная величина и $X \sim \mathcal{N}\left(\mu, \sigma^2\right)$.
- Нас интересует некоторая вероятность $P(a < X < b)$, $\forall a,\, b : a \leq b$.
- Применим преобразование к каждой части неравенства, чтобы не нарушить его:

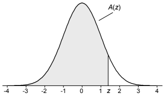
$$P(a < X < b) = P\left(\underbrace{\frac{a-\mu}{\sigma}}_{\tilde{a}} < \underbrace{\frac{X-\mu}{\sigma}}_{Z} < \underbrace{\frac{b-\mu}{\sigma}}_{\tilde{b}}\right) = P\left(\tilde{a} < Z < \tilde{b}\right) = F_Z(\tilde{b}) - F_Z(\tilde{a}).$$

Стандартная нормальная величина



Стандартная нормальная величина

Cumulative Standardized Normal Distribution



$A(z)$ is the integral of the standardized normal distribution from $-\infty$ to z (in other words, the area under the curve to the left of z). It gives the probability of a normal random variable not being more than z standard deviations above its mean. Values of z of particular importance:

z	$A(z)$	
1.645	0.9500	Lower limit of right 5% tail
1.960	0.9750	Lower limit of right 2.5% tail
2.326	0.9900	Lower limit of right 1% tail
2.576	0.9950	Lower limit of right 0.5% tail
3.090	0.9990	Lower limit of right 0.1% tail
3.291	0.9995	Lower limit of right 0.05% tail

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986
3.0	0.9987	0.9987	0.9987	0.9988	0.9988	0.9989	0.9989	0.9989	0.9990	0.9990
3.1	0.9990	0.9991	0.9991	0.9991	0.9992	0.9992	0.9992	0.9992	0.9993	0.9993
3.2	0.9993	0.9993	0.9994	0.9994	0.9994	0.9994	0.9994	0.9995	0.9995	0.9995
3.3	0.9995	0.9995	0.9995	0.9996	0.9996	0.9996	0.9996	0.9996	0.9996	0.9997
3.4	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9998
3.5	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998
3.6	0.9998	0.9998	0.9999							

Пример

Если X — нормально распределенная случайная величина со средним 6 и дисперсией 25, найти:

1. $P(6 \leq X \leq 12)$,
2. $P(0 \leq X \leq 8)$,
3. $P(-2 < X \leq 0)$.

Пример

Вопрос 1

$P(6 \leq X \leq 12)$

- $X \sim \mathcal{N}(6, 25)$, значит $\mu = 6$, $\sigma = 5$
- Преобразование: $Z = \frac{X-6}{5}$
- $P(6 \leq X \leq 12) = P\left(\frac{6-6}{5} \leq Z \leq \frac{12-6}{5}\right) = P(0 \leq Z \leq 1.2)$
- $= F_Z(1.2) - F_Z(0) = 0.8849 - 0.5 = 0.3849$

Пример

Визуализация вопроса 1

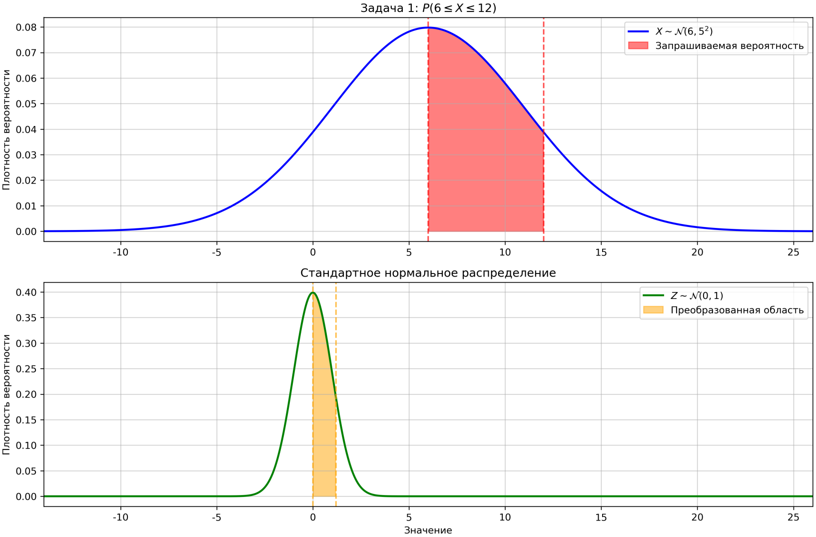


Рис. 6

Пример

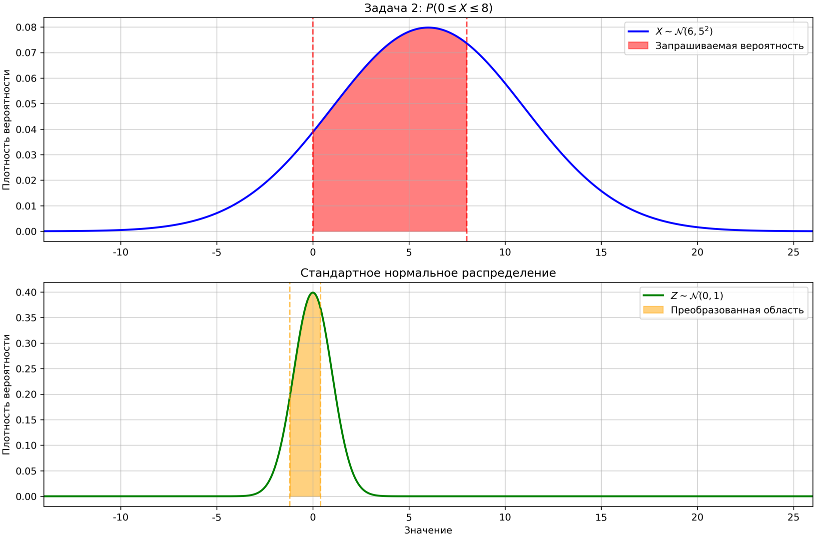
Вопрос 2

$P(0 \leq X \leq 8)$

- Преобразование: $Z = \frac{X-6}{5}$
- $P(0 \leq X \leq 8) = P\left(\frac{0-6}{5} \leq Z \leq \frac{8-6}{5}\right) = P(-1.2 \leq Z \leq 0.4)$
- Используем свойство симметрии: $F_Z(-1.2) = 1 - F_Z(1.2) = 1 - 0.8849 = 0.1151$
- $= F_Z(0.4) - F_Z(-1.2) = 0.6554 - 0.1151 = 0.5403$

Пример

Визуализация вопроса 2



Пример

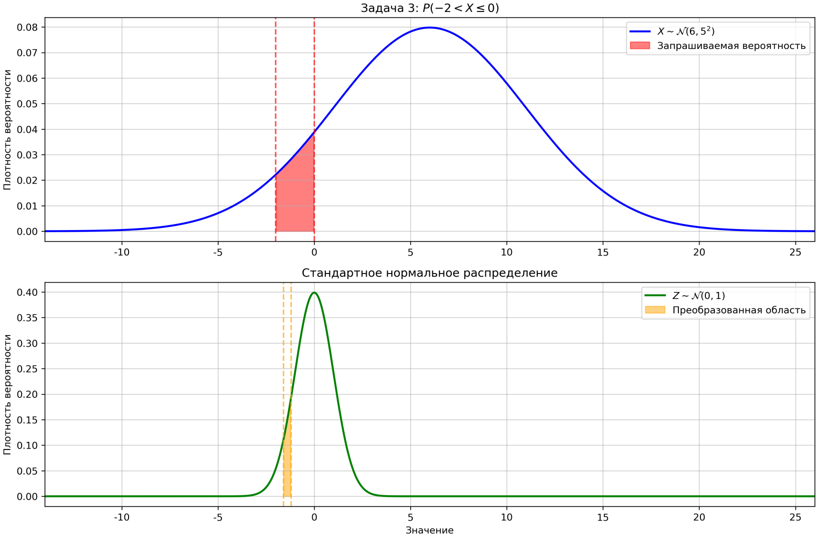
Вопрос 3

$P(-2 < X \leq 0)$

- Преобразование: $Z = \frac{X-6}{5}$
- $P(-2 < X \leq 0) = P\left(\frac{-2-6}{5} < Z \leq \frac{0-6}{5}\right) = P(-1.6 < Z \leq -1.2)$
- Используем свойство симметрии: $F_Z(-1.2) = 1 - F_Z(1.2) = 1 - 0.8849 = 0.1151$
- $F_Z(-1.6) = 1 - F_Z(1.6) = 1 - 0.9452 = 0.0548$
- $= F_Z(-1.2) - F_Z(-1.6) = 0.1151 - 0.0548 = 0.0603$

Пример

Визуализация вопроса 3



Многомерное нормальное распределение

Definition

Пара случайных величин (X, Y) имеет **двумерное нормальное распределение**, если для всех $a, b \in \mathbb{R}$ линейная комбинация $aX + bY$ имеет одномерное нормальное распределение.

Definition

Вектор случайных величин $\mathbf{X} = (X_1, X_2, \dots, X_n)$ имеет **многомерное нормальное распределение**, если для всех векторов $\mathbf{a} \in \mathbb{R}^n$ линейная комбинация $\mathbf{a}^\top \mathbf{X}$ имеет одномерное нормальное распределение.
$\mathbf{a}^\top \mathbf{X} = a_1 X_1 + a_2 X_2 + \dots + a_n X_n$

Устойчивость нормального распределения

- **Ключевое свойство**: Нормальное распределение **устойчиво** относительно линейных преобразований.
- Если $X \sim \mathcal{N}(\mu, \sigma^2)$, то для любых констант $a \neq 0$ и b:

$$Y = aX + b \sim \mathcal{N}(a\mu + b, a^2\sigma^2)$$

- **Интерпретация**: Линейное преобразование нормальной случайной величины даёт нормальную случайную величину.

Устойчивость нормального распределения

Линейная комбинация двух независимых нормальных величин

- Частный случай: если $X_1 \sim \mathcal{N}(\mu_1, \sigma_1^2)$ и $X_2 \sim \mathcal{N}(\mu_2, \sigma_2^2)$ независимы, то:

$$X_1 + X_2 \sim \mathcal{N}(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$$

- **Связь с многомерным нормальным**: Независимые нормальные величины (X_1, X_2) формируют двумерное нормальное распределение, поэтому любая линейная комбинация $aX_1 + bX_2$ имеет нормальное распределение.

Устойчивость нормального распределения

Обобщение на n переменных

- Пусть $X_1, \dots, X_n$ — независимые нормальные случайные величины:

$$X_i \sim \mathcal{N}(\mu_i, \sigma_i^2), \quad i = 1, \dots, n$$

- **Важно**: Вектор (коллекция, набор) независимых нормальных случайных величин имеет **многомерное нормальное распределение**.
- По определению многомерного нормального распределения, **любая** линейная комбинация имеет нормальное распределение:

$$Y = \sum_{i=1}^{n} a_i X_i \sim \mathcal{N}\left(\sum_{i=1}^{n} a_i \mu_i, \sum_{i=1}^{n} a_i^2 \sigma_i^2\right)$$