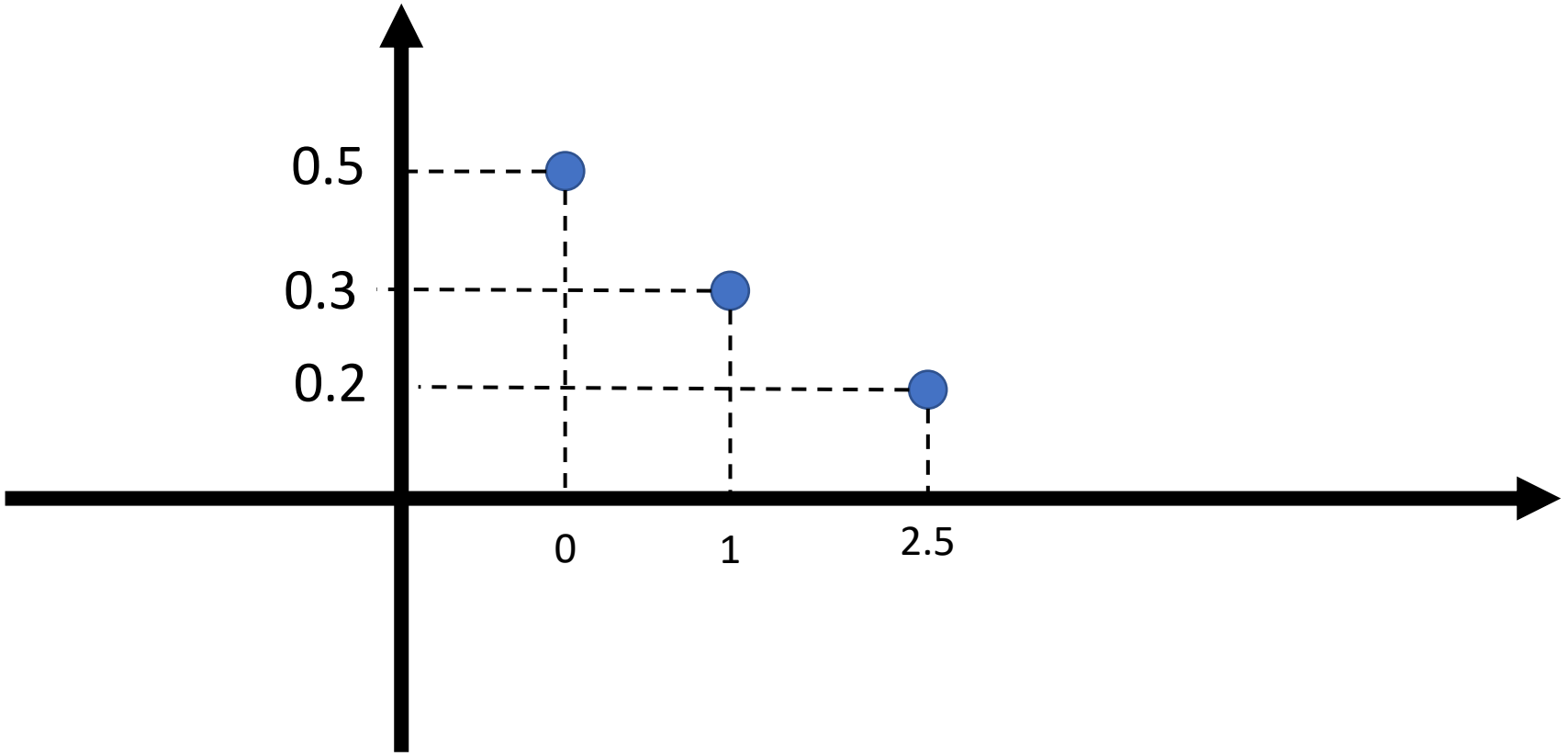
Непрерывные случайные величины и ЦПТ

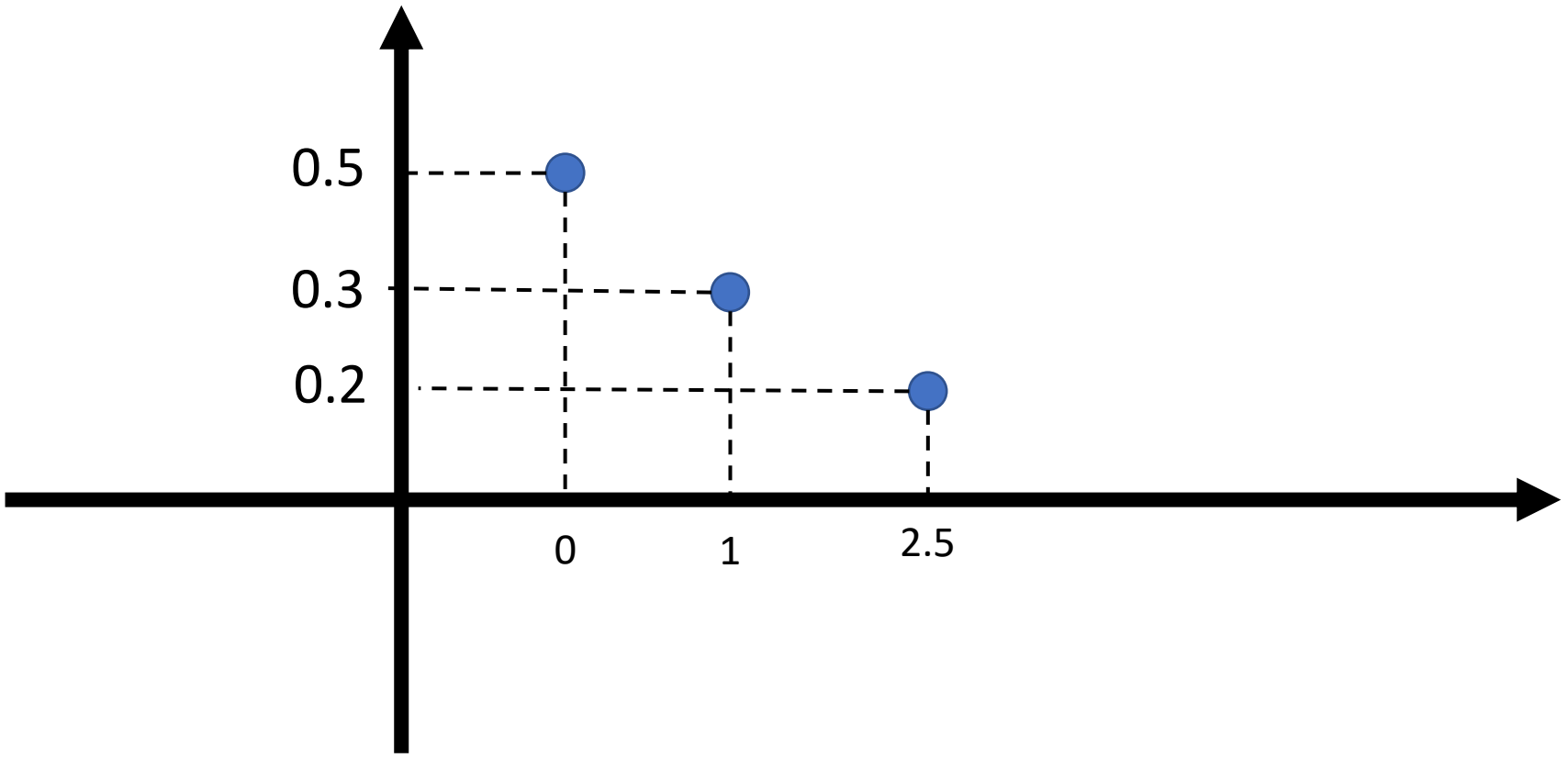
Probability mass function (функция вероятности)

По оси откладываем X откладываем значения, которые может принимать переменная, по Y - вероятности



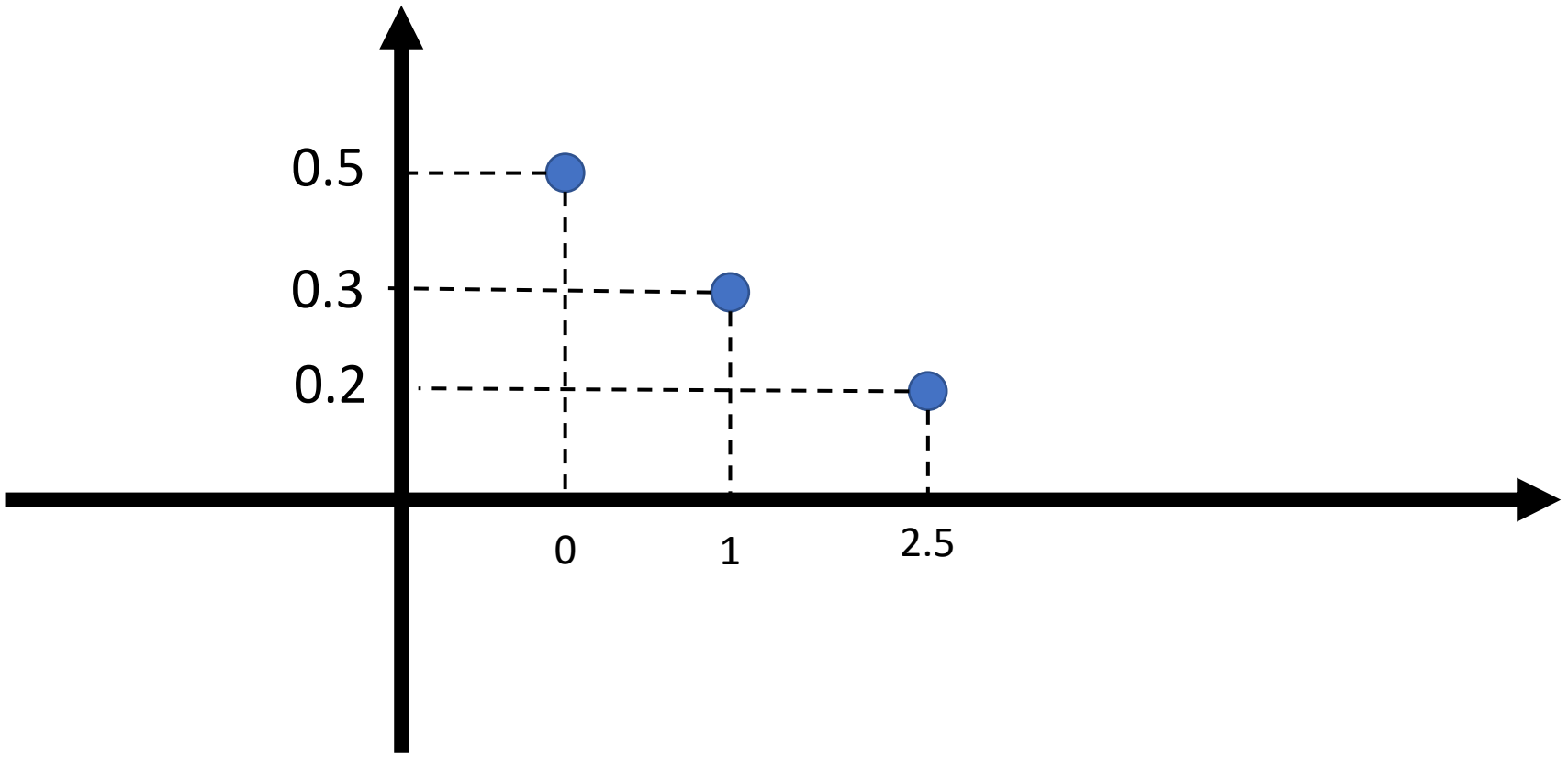
Probability mass function (функция вероятности)

Какое матожидание этой случайной величины?



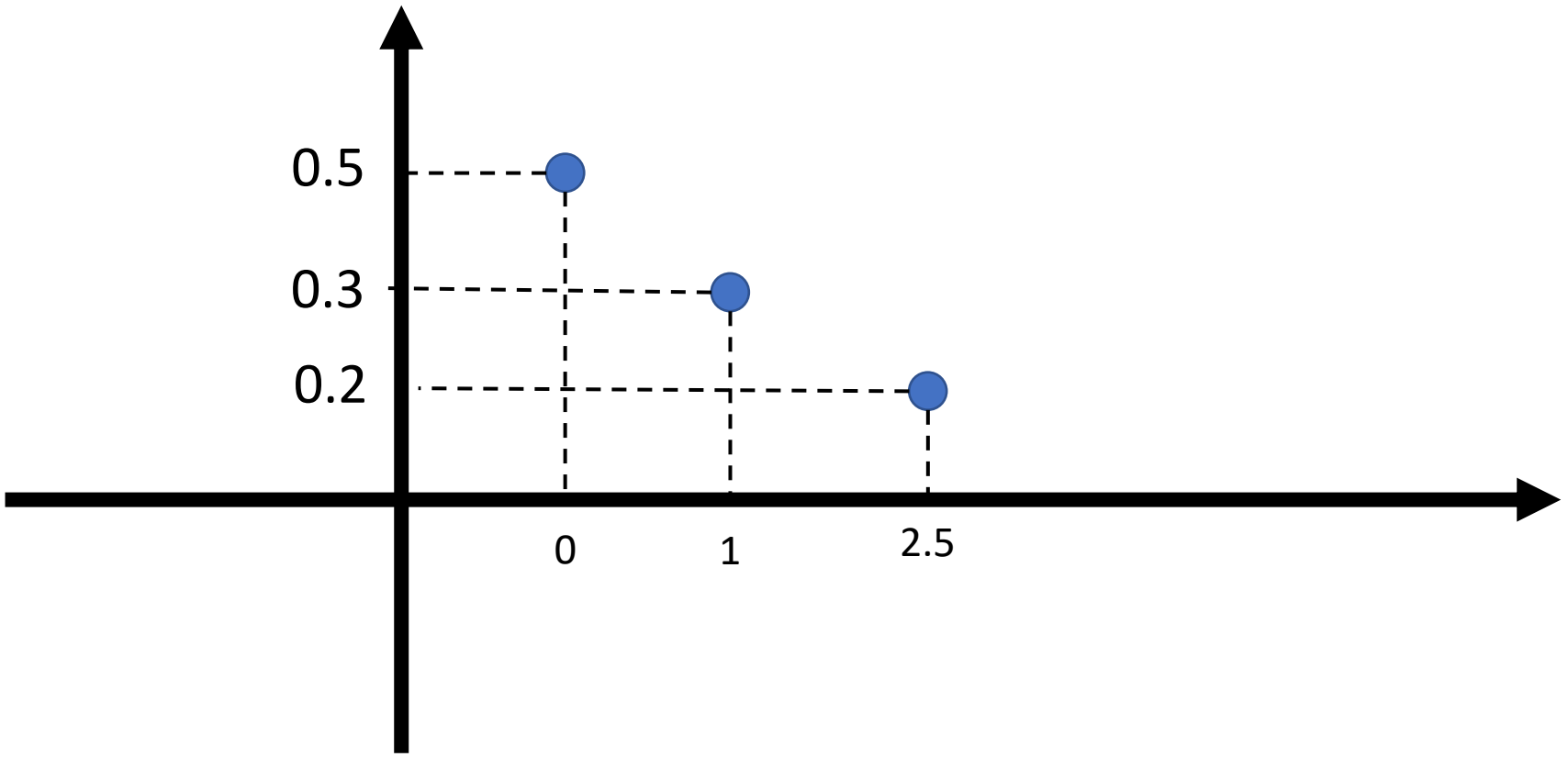
Probability mass function (функция вероятности)

$$\mathrm{E}(\mathrm{X}) = 0 \cdot 0.5 + 0.3 \cdot 1 + 0.2 \cdot 2.5 = 0.8$$



Probability mass function (функция вероятности)

Какая дисперсия этой случайной величины?

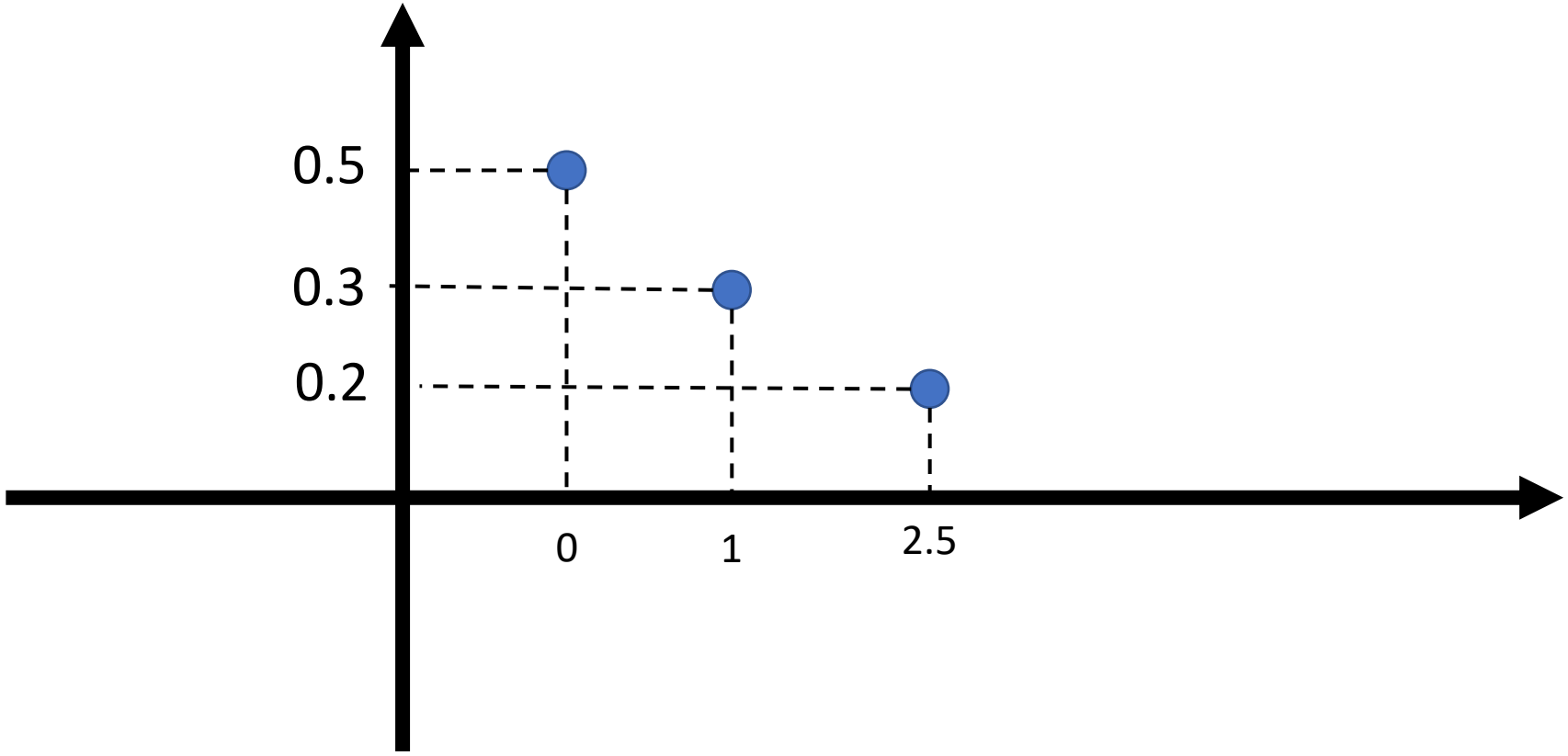


Probability mass function (функция вероятности)

$$\mathrm{E}(X) = 0 \cdot 0.5 + 0.3 \cdot 1 + 0.2 \cdot 2.5 = 0.8$$

$$\mathrm{E}(X^2) = 0 \cdot 0.5^2 + 0.3 \cdot 1^2 + 0.2 \cdot 2.5^2 = 1.55$$

$$\mathrm{D}(X) = 1.55 - 0.8^2 = 0.91$$



Непрерывные случайные величины

Бросаем случайно точку на отрезок – какова вероятность, что попадем в заданную точку?

Непрерывные случайные величины

Бросаем случайно точку на отрезок – какова вероятность, что попадем в заданную точку?

Теоретически – 0

Непрерывные случайные величины

Берем случайное число из отрезка от 0 до 1. Какова вероятность, что оно в точности равно 0.5?

Теоретически – 0. Потому если мы захотели б построить probability mass function для такой случайной величины – у нас бы ничего не получилось

Непрерывные случайные величины

Берем случайное число из отрезка от 0 до 1. Какова вероятность, что оно оно будет не больше 0.5?

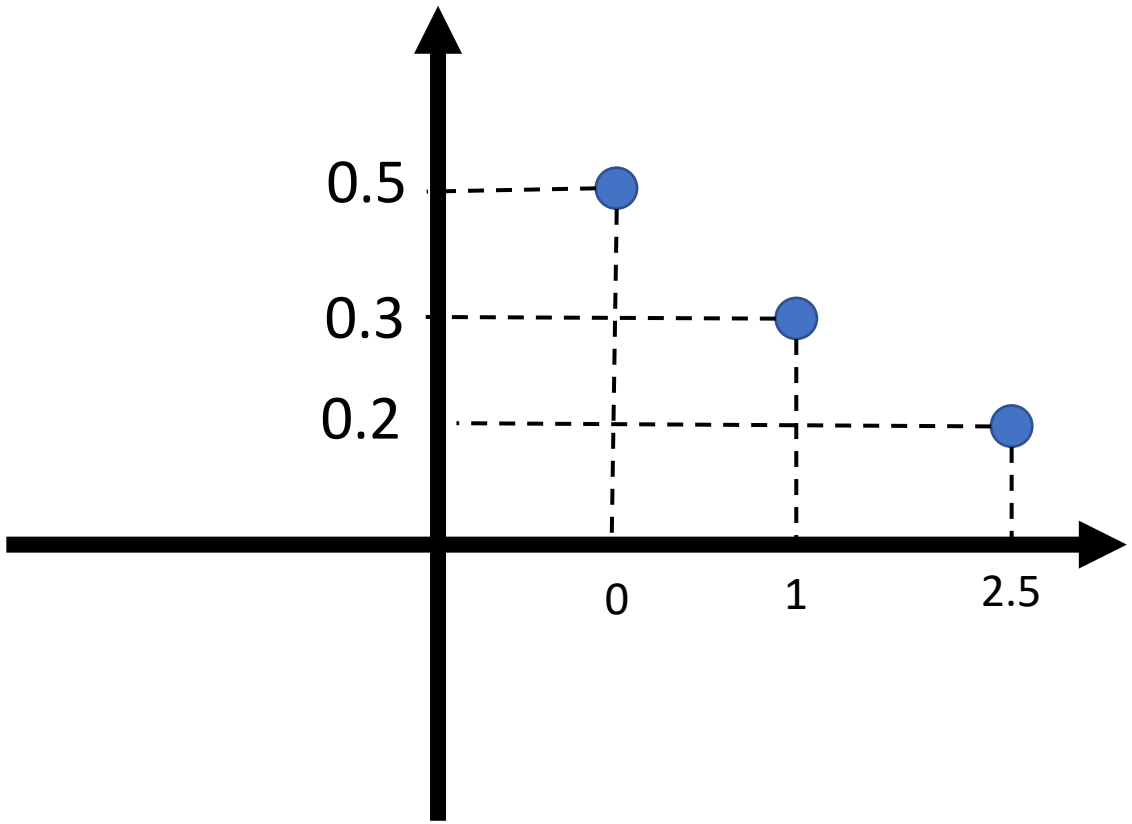
0.5

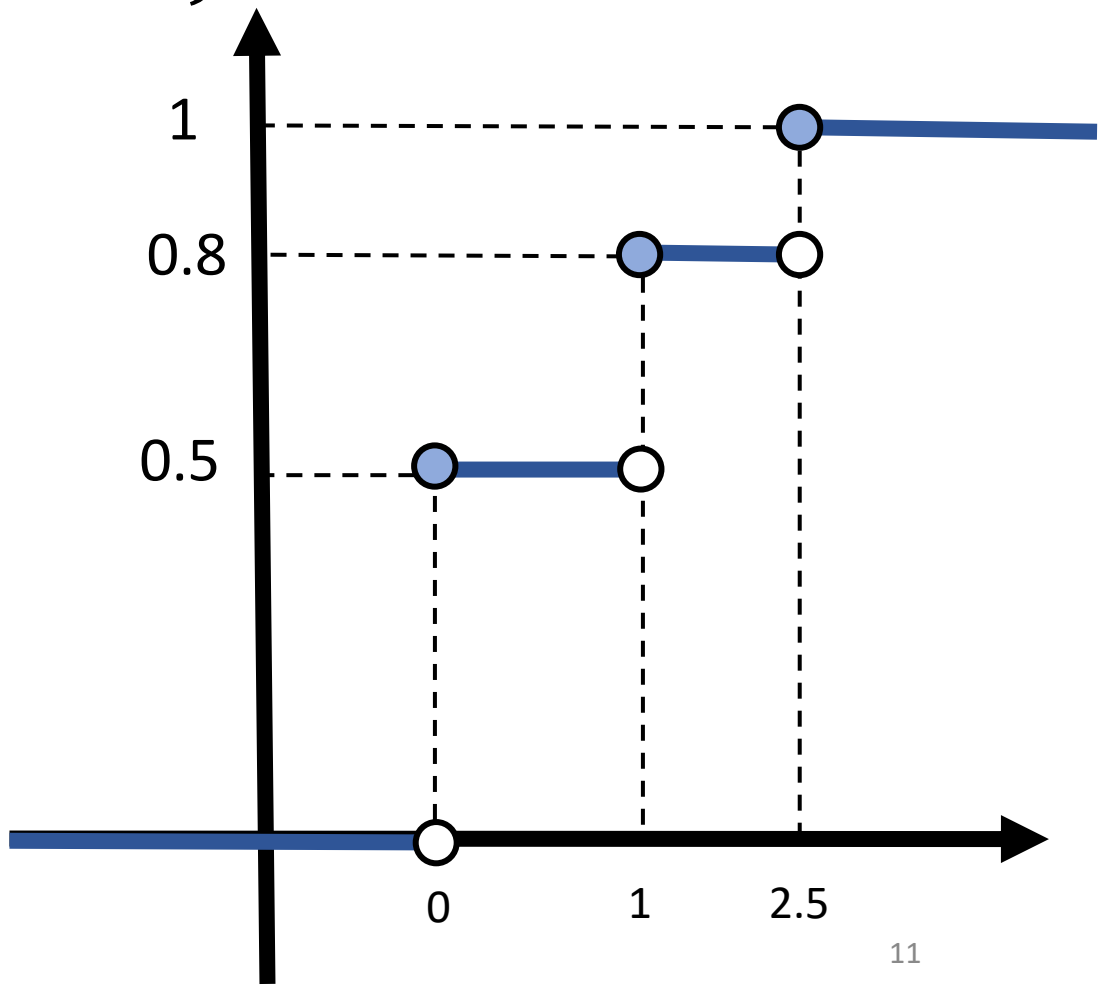
$$P(X \leq 0.5) = 0.5$$

Функция распределения

Функция, возвращающая вероятность того, что случайная величина не примет значения больше заданного

$$F_X(x) = P(X \leq x)$$

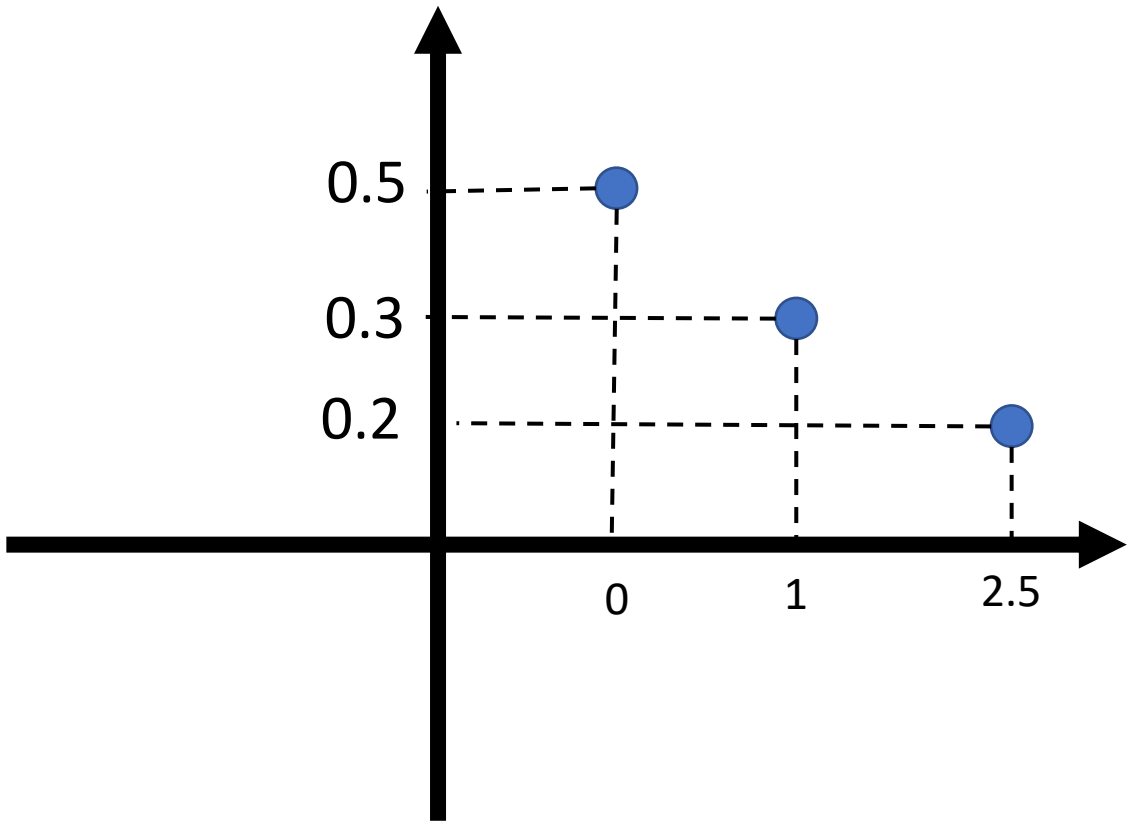


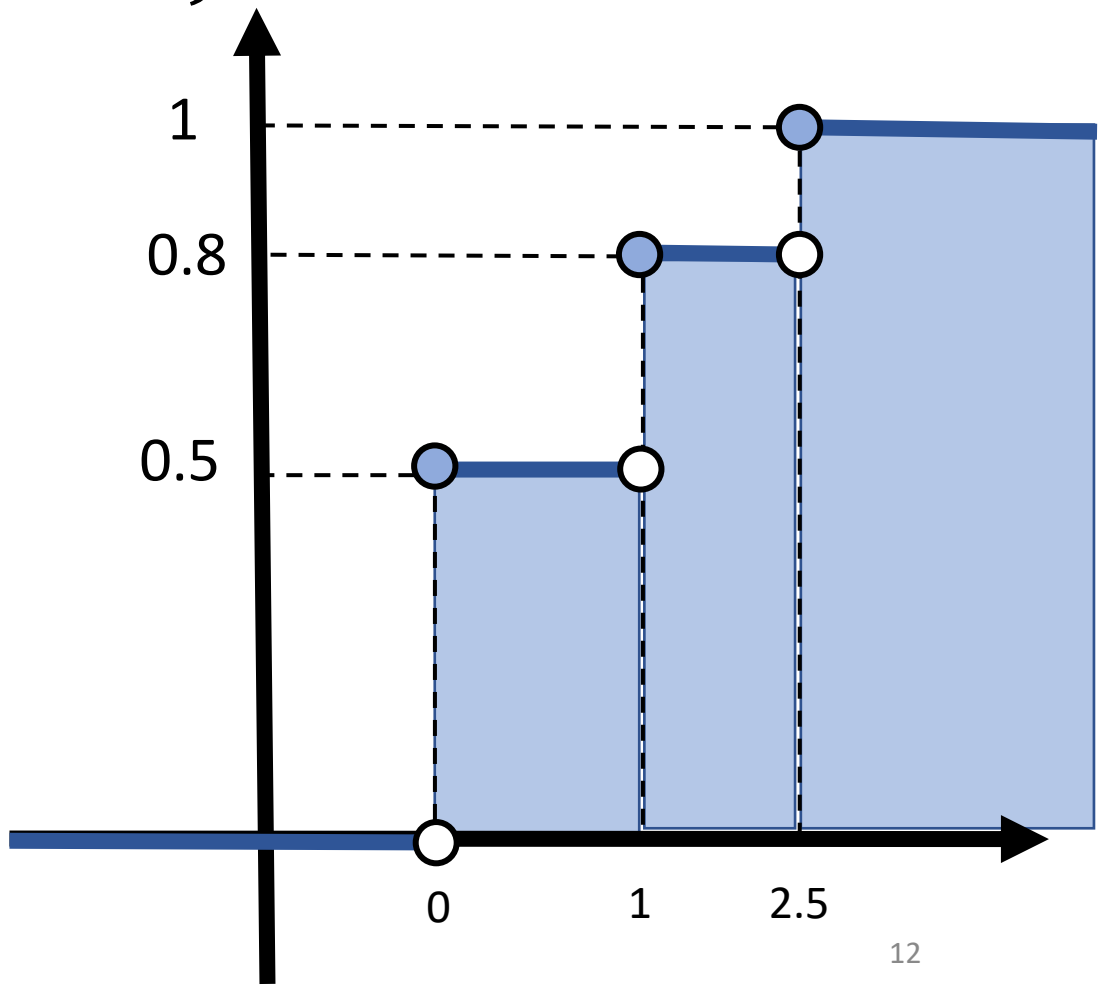


Функция распределения

Функция, возвращающая вероятность того, что случайная величина не примет значения больше заданного

$$F_X(x) = P(X \leq x)$$

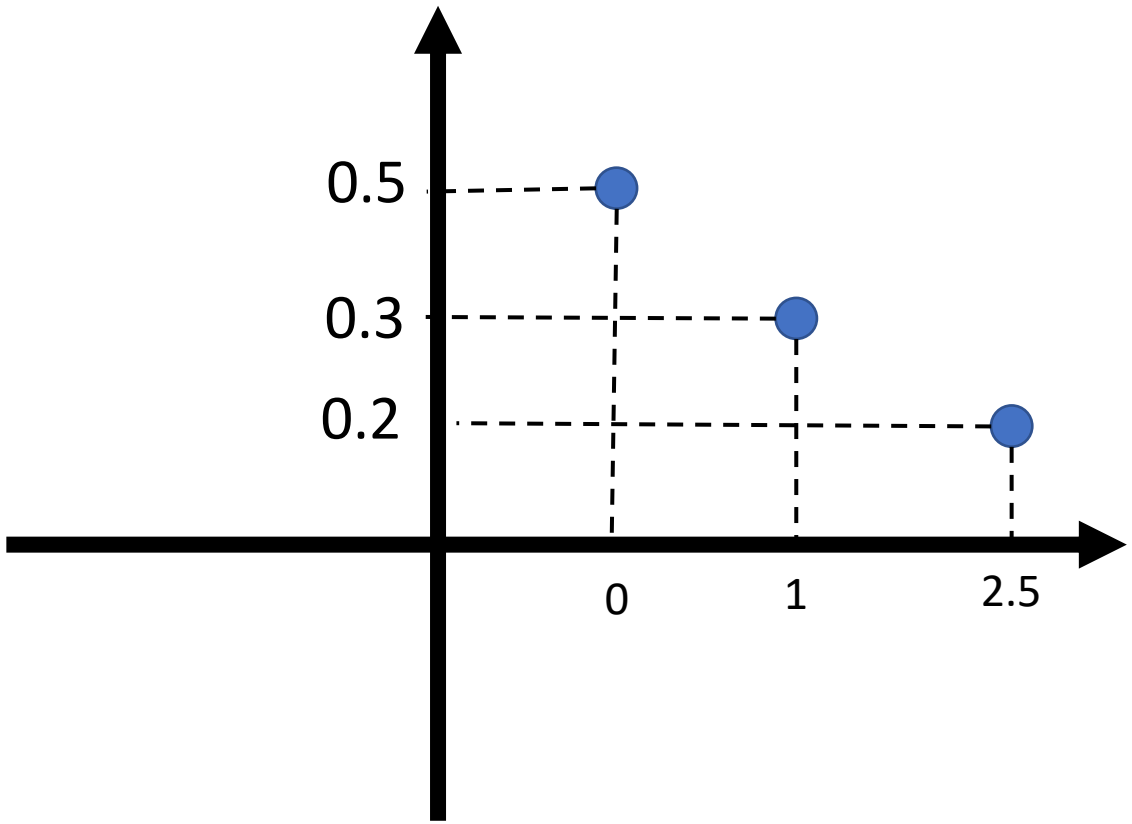


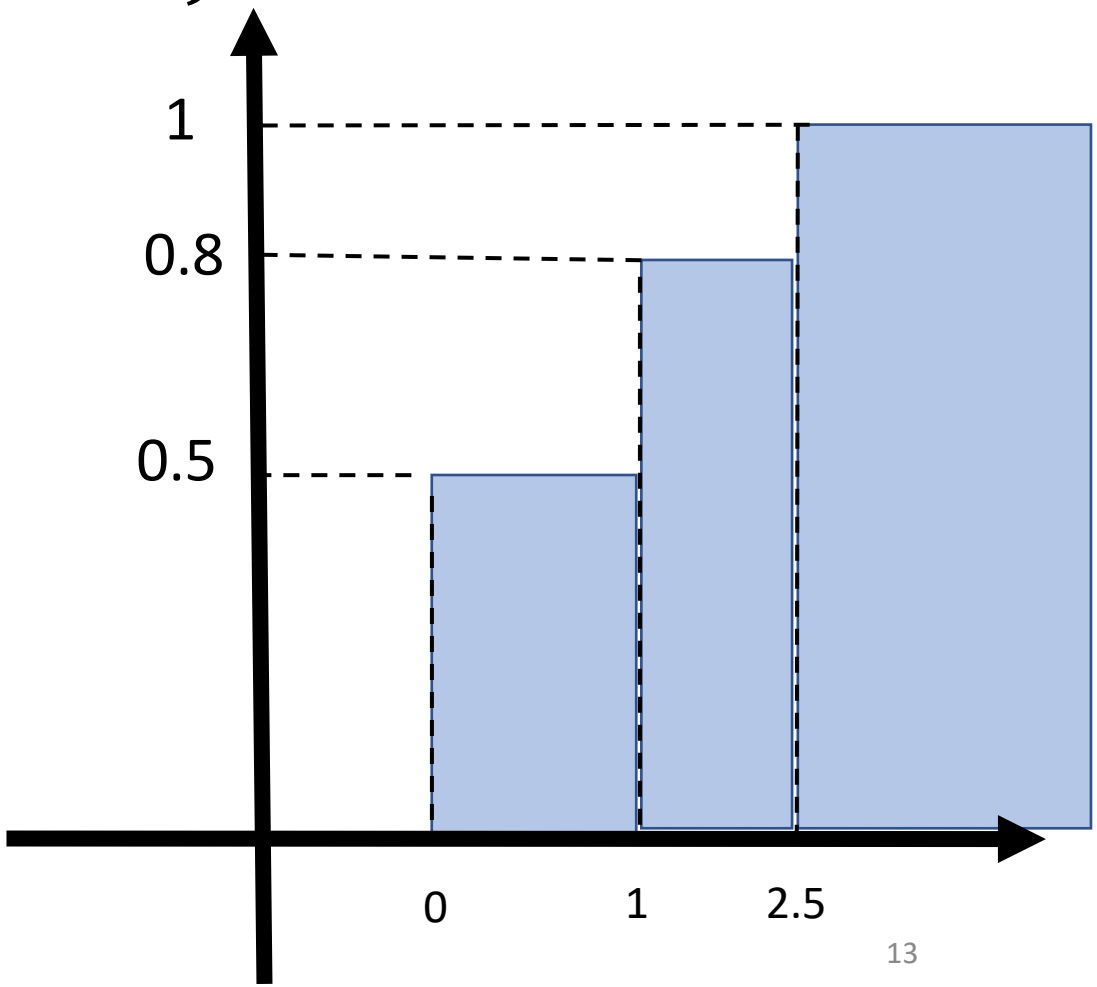


Функция распределения

Функция, возвращающая вероятность того, что случайная величина не примет значения больше заданного

$$F_X(x) = P(X \leq x)$$





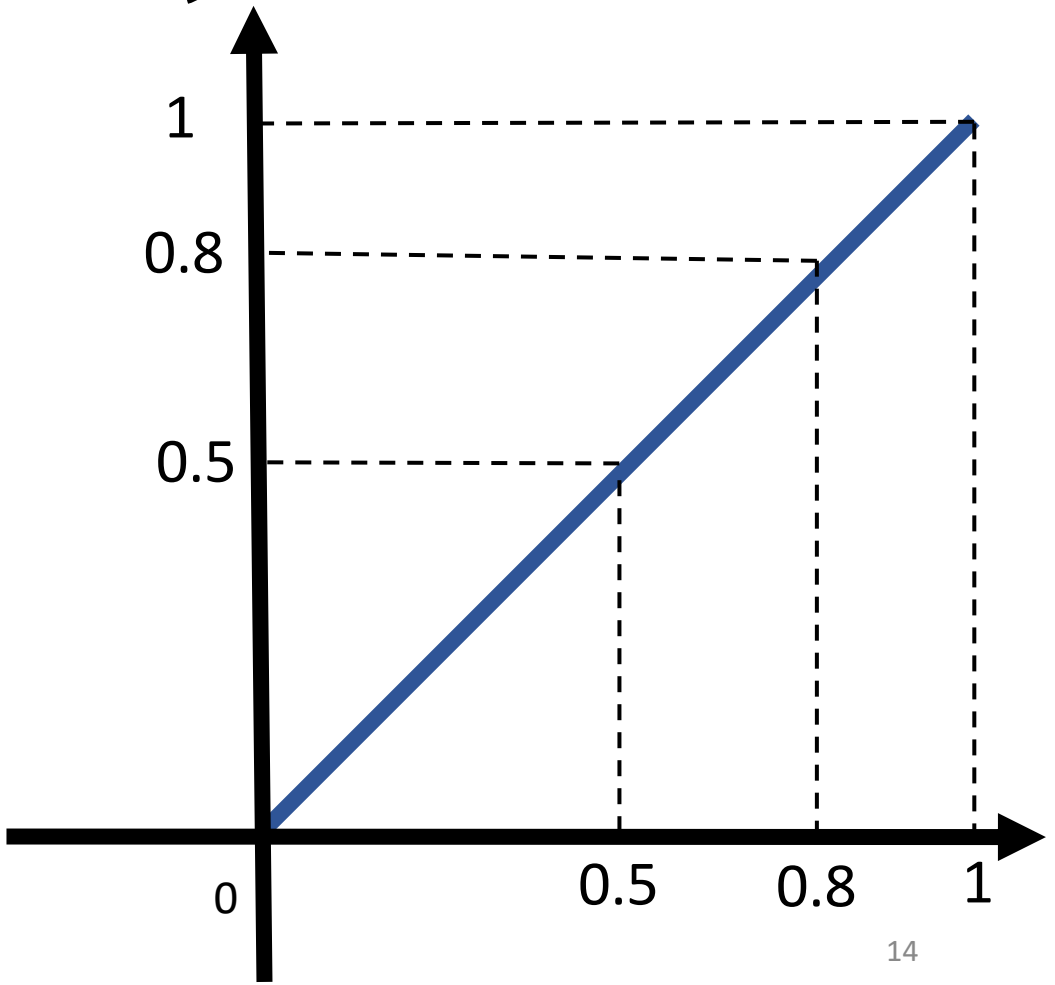
Функция распределения

В отличии от функции вероятности, можем определить и для непрерывной величины

$$F_X(x) = P(X \leq x)$$

Генерируем любое вещественное число из отрезка от 0 до 1 с равной вероятностью – **равномерное распределение** на отрезке [0, 1]

Можно задать и на любом отрезке [a, b]



Равномерное распределение

Берем случайное число из отрезка от 0 до 1. Какова вероятность, что оно оно будет не больше 0.5?

0.5

$$\mathrm{P}(\mathrm{X} \leq 0.5) = F_X(0.5) = 0.5$$

Равномерное распределение

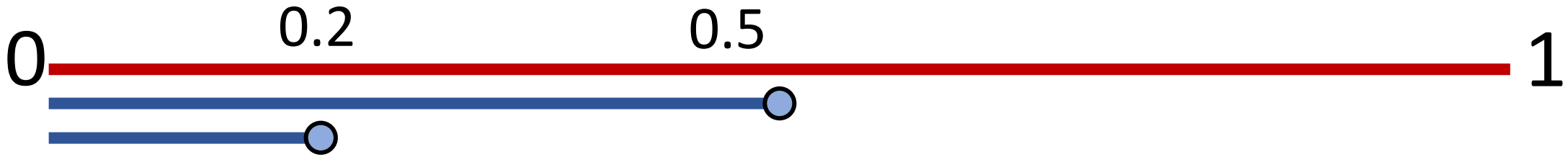
Берем случайное число из отрезка от 0 до 1. Какова вероятность, что оно оно будет **меньше** 0.5?

0.5

$$\mathrm{P}(0.2 \leq \mathrm{X} \leq 0.5) =$$
$$\mathrm{P}(\mathrm{X} \leq 0.5) - \mathrm{P}(\mathrm{X} = 0.5) =$$
$$F_X(0.5) \; - \; 0 =$$
$$0.3$$

Равномерное распределение

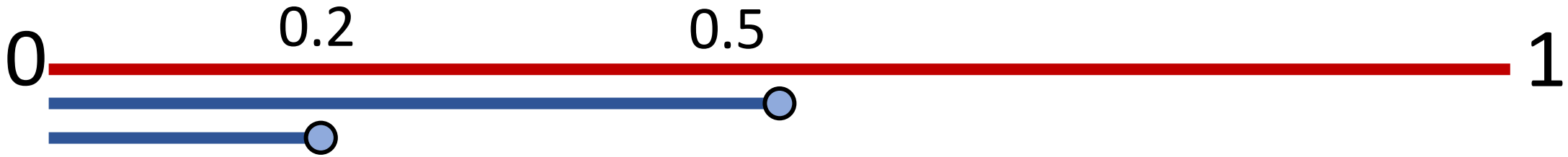
Берем случайное число из отрезка от 0 до 1. Какова вероятность, что оно будет находиться в отрезке [0.2, 0.5]?



$$\mathrm{P}(0.2 \leq \mathrm{X} \leq 0.5) =$$
$$\mathrm{P}(\mathrm{X} \leq 0.5) - \mathrm{P}(\mathrm{X} \leq 0.2) + \mathrm{P}(\mathrm{X} = 0.2) =$$
$$F_X(0.5) - F_X(0.2) + 0 =$$
$$0.3$$

Равномерное распределение

Берем случайное число из отрезка от 0 до 1. Какова вероятность, что оно будет находиться в отрезке [0.2, 0.5]?



$$\mathrm{P}(0.2 \leq \mathrm{X} \leq 0.5) =$$
$$\mathrm{P}(\mathrm{X} \leq 0.5) - \mathrm{P}(\mathrm{X} \leq 0.2) + \mathrm{P}(\mathrm{X} = 0.2) =$$
$$F_X(0.5) - F_X(0.2) + 0 =$$
$$0.3$$

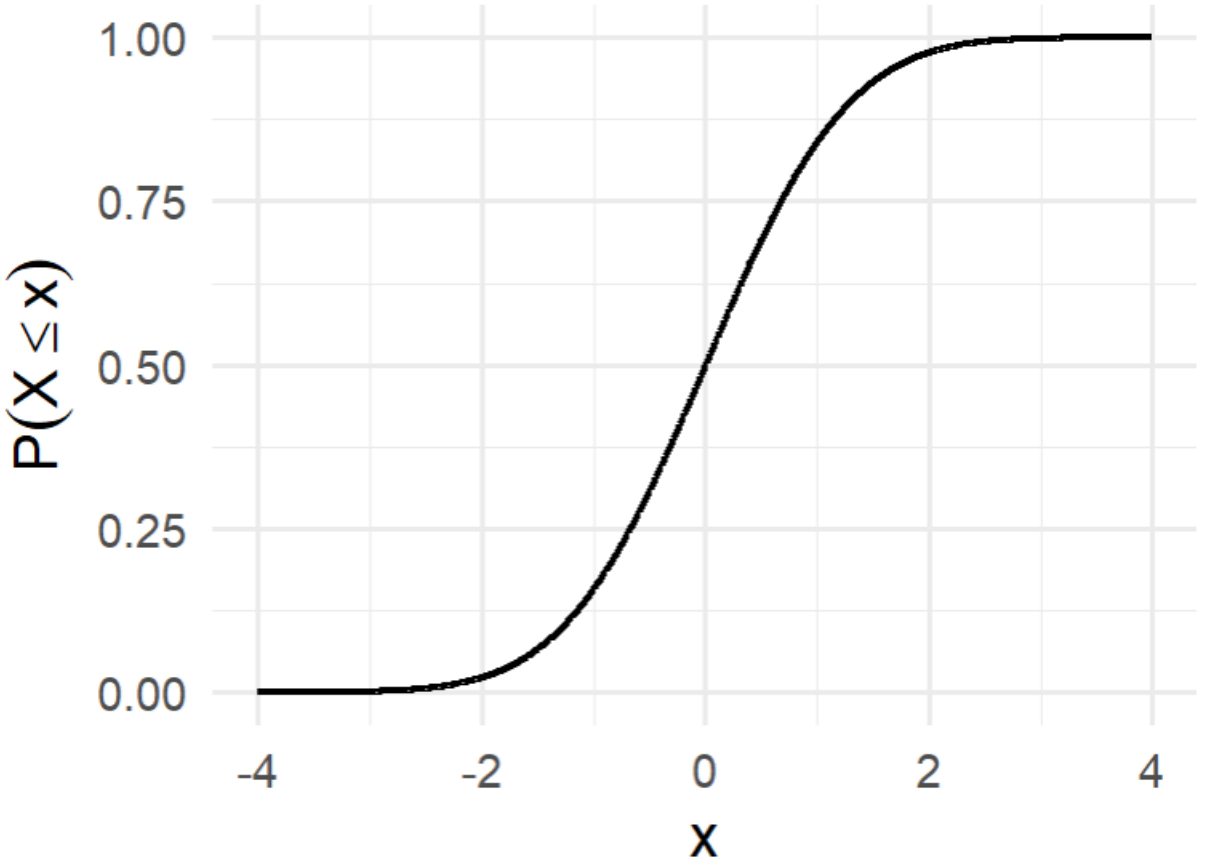
Равномерное распределение

Берем случайное число из отрезка от 0 до 1. Какова вероятность, что оно будет находиться в отрезке [0.2, 0.5]?

Можем просто разделить длину отрезка [0.2, 0.5] на длину всего отрезка

Функция распределения

Допустим, у нас такая функция распределения.
Какая вероятность больше - что наша величина попадет на отрезок [-2, 0] или [2, 4]?

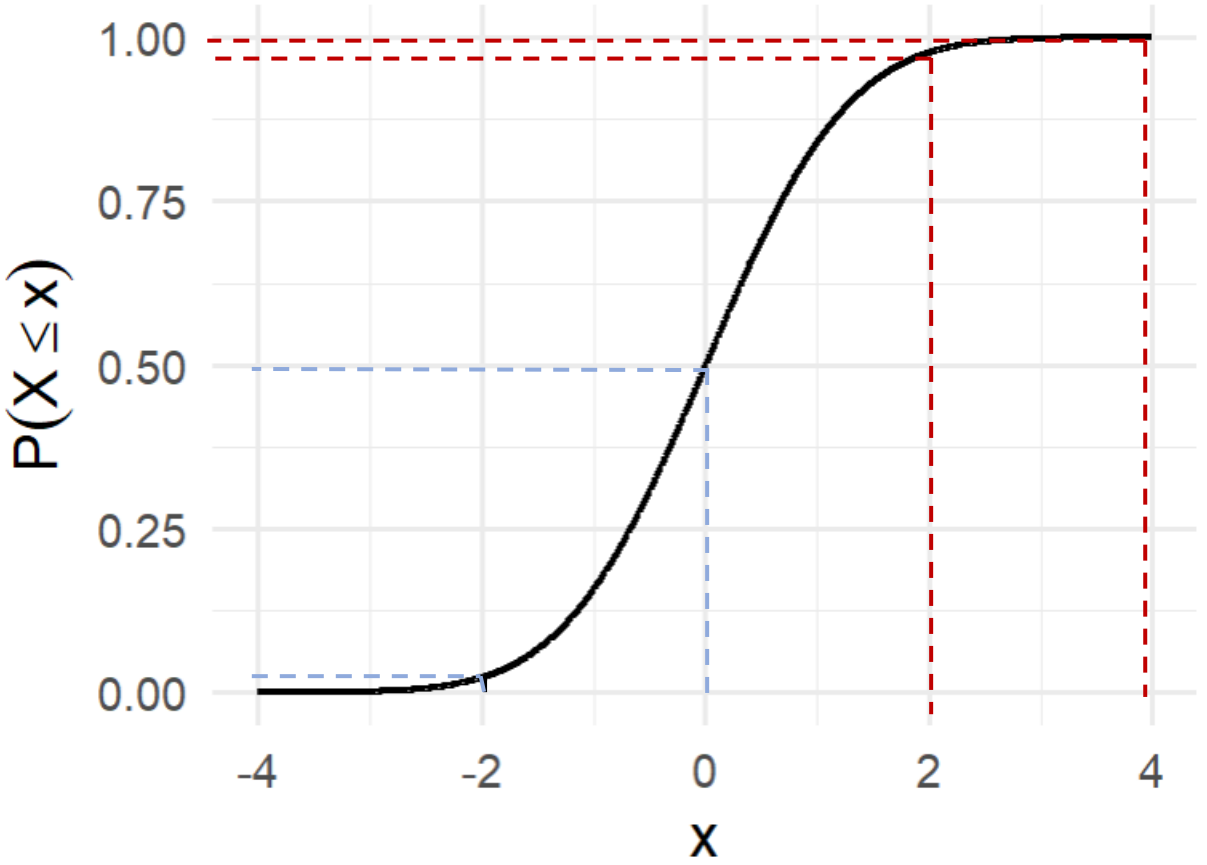


Функция распределения

Допустим, у нас такая функция распределения.
Какая вероятность больше - что наша величина попадет на отрезок [-2, 0] или [2, 4]?

На глаз видно, что в первый. Заметим, что отрезки одинаковой длины.

А если отрезки [-4, -2] и [2, 4]?

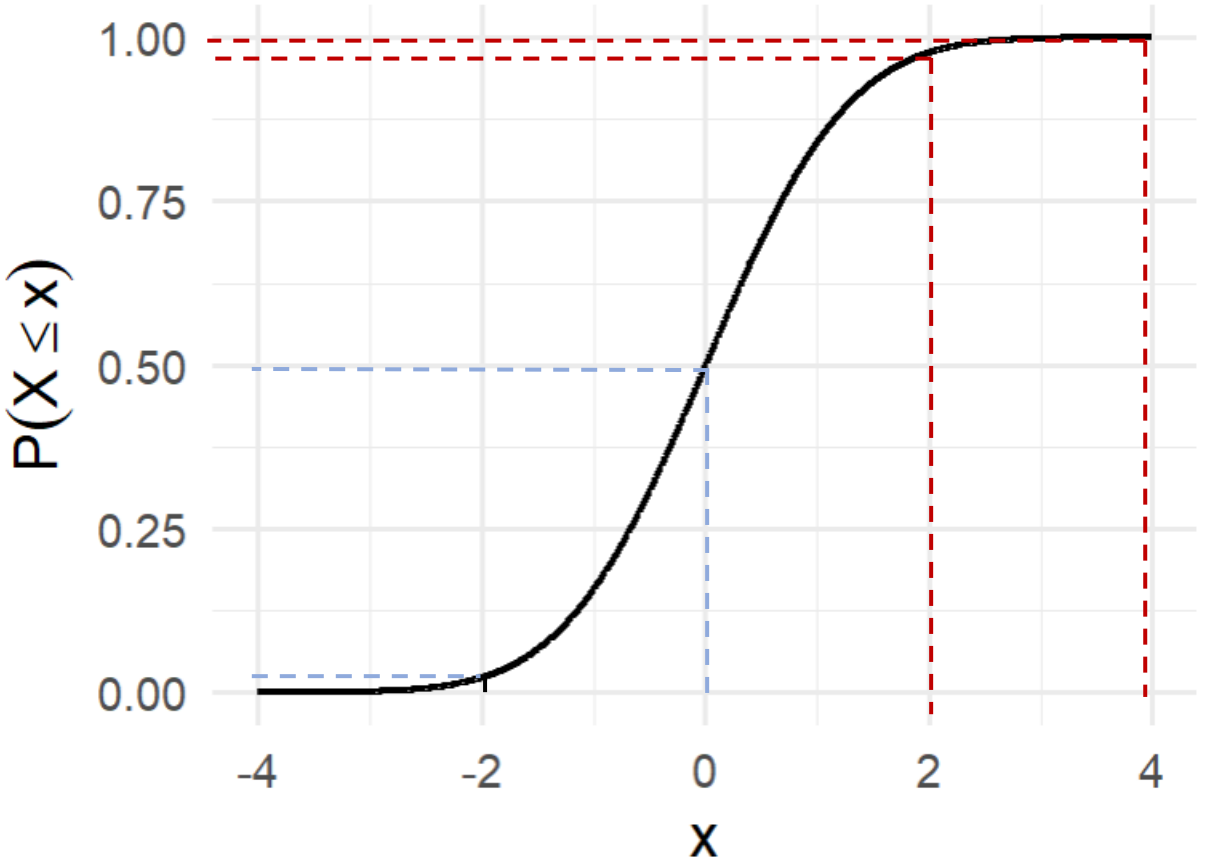


Функция распределения

Допустим, у нас такая функция распределения.
Какая вероятность больше - что наша величина попадет на отрезок [-2, 0] или [2, 4]?

На глаз видно, что в первый. Заметим, что отрезки одинаковой длины.

А если отрезки [-4, -2] и [2, 4]?



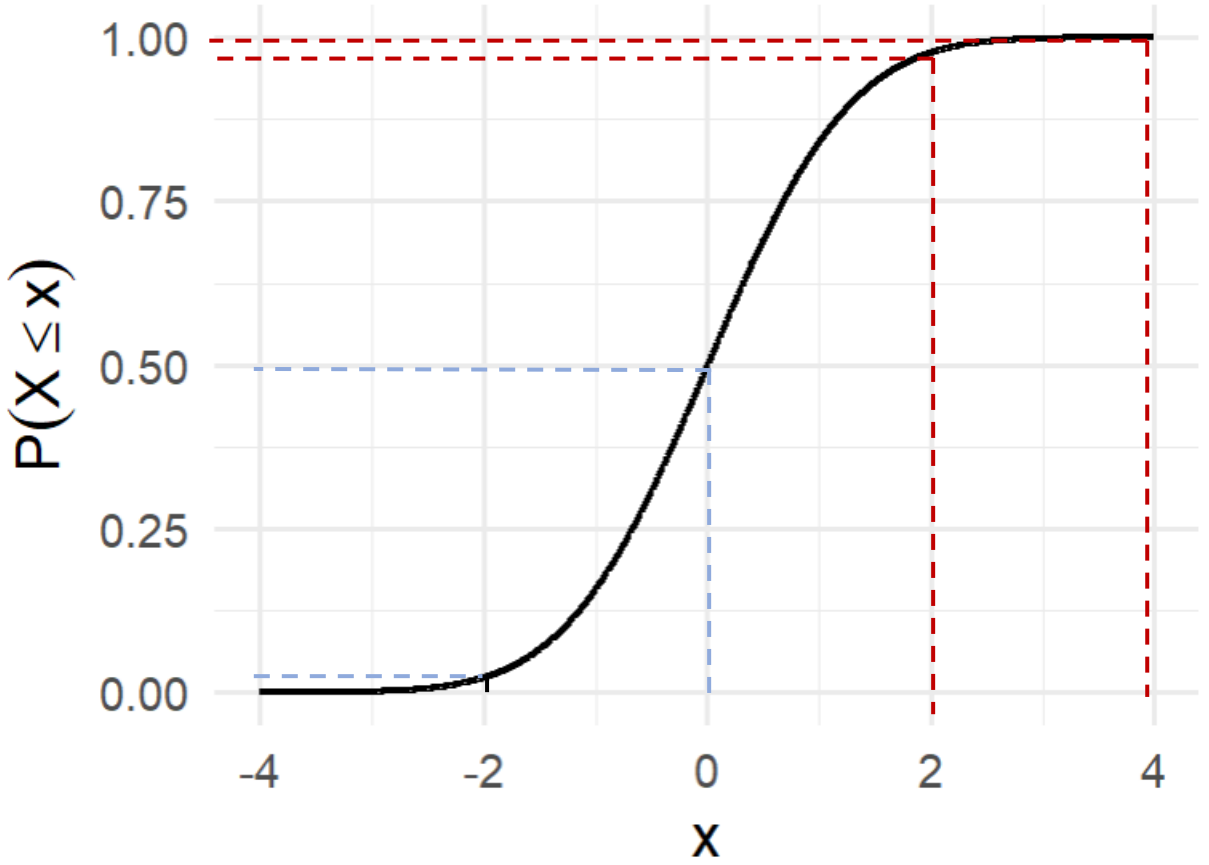
Функция распределения

Допустим, у нас такая функция распределения.
Какая вероятность больше - что наша величина попадет на отрезок [-2, 0] или [2, 4]?

На глаз видно, что в первый. Заметим, что отрезки одинаковой длины.

А если отрезки [-4, -2] и [2, 4]?

Вероятности одинаковые. График функции распределения в этом плане немного контринтуитивен, а геометрическую интерпретации отыскать еще сложнее.



Функция плотности

Проблема – с одной стороны у нас есть интуиция насчет того, как выражать вероятности для непрерывных величин геометрически, с другой – пока интерпретировать функцию распределения геометрически мы не можем.

Также, охарахтеризовать какой-то отрезок $[a,\ b]$ с точки зрения того, насколько вероятно случайная величина попадет именно в него, мы можем только через вычисление функцию распределения на его концах

Да и как считать матожидания и прочие характеристики непонятно

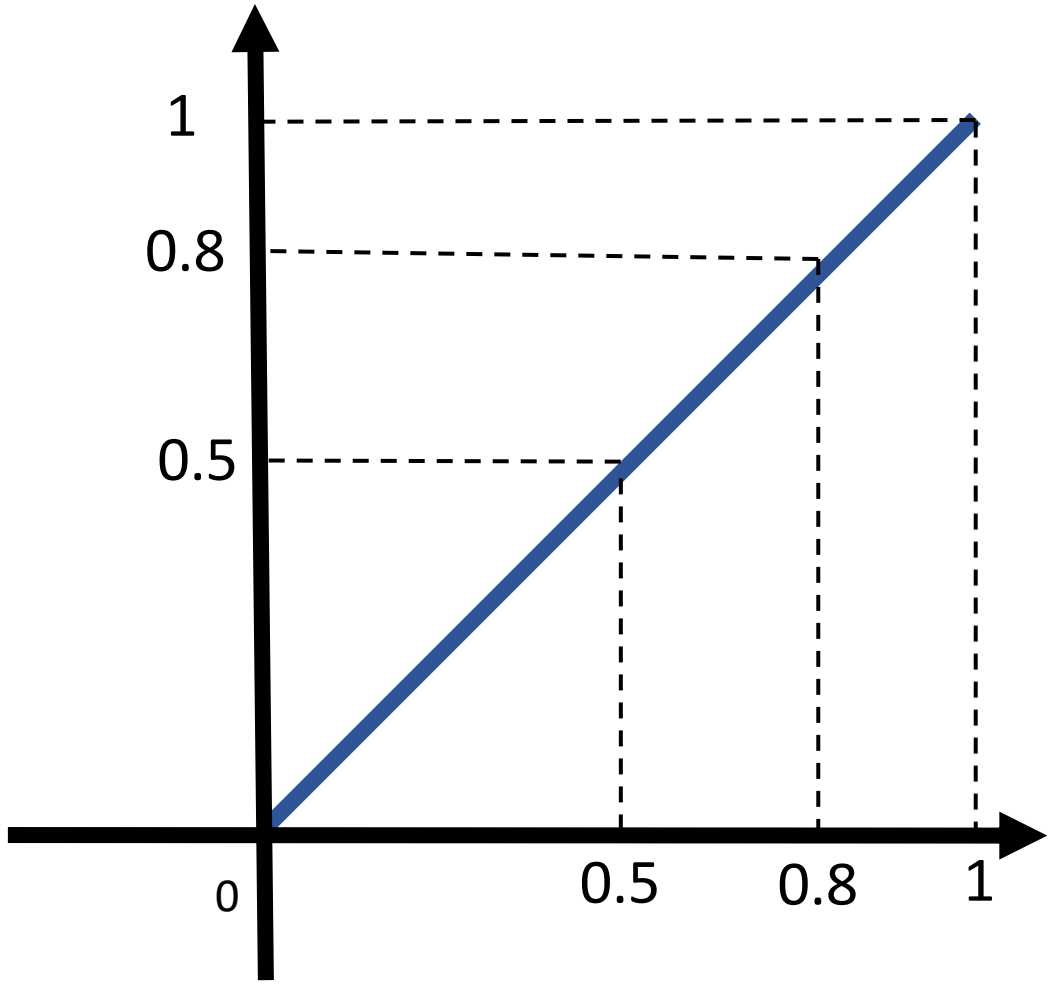
$$E(X) = \sum_i x_i p_i \ - \text{для дискретной случайной величины}$$

Нам нужен ***аналог функции вероятности*** для непрерывных распределений, обладающий хотя бы частью ее свойств и имеющий геометрическую интерпретацию

Функция плотности

Да, вероятность того, что $X = a$ для непрерывной величины всегда равна 0.

Но – вероятность, что $X \in [a - \delta; a + \delta]$ не обязательно равна нулю. В частности, для равномерного распределения на отрезке [0, 1] для любого отрезка $[a - \delta; a + \delta]$, пересекающегося с $[0,1]$, она не нулевая



$$P(a - \delta \leq X \leq a + \delta) = F_X(a + \delta) - F_x(a - \delta)$$

Функция плотности

$$P(a - \delta \leq X \leq a + \delta) = F_X(a + \delta) - F_x(a - \delta)$$

Полученное число зависит от δ. Очевидно, нам это не нравится. Потому сделаем две вещи

1. Разделим полученное число на размер отрезка $[a - \delta, a + \delta]$.

$$\frac{P(a - \delta \leq X \leq a + \delta)}{2\delta} = \frac{F_X(a + \delta) - F_x(a - \delta)}{2\delta}$$

2. Устремим $\delta \to 0$

$$\lim_{\delta \to 0} \frac{F_X(a + \delta) - F_x(a - \delta)}{2\delta}$$

Ничего не напоминает?

Функция плотности

Получается производная функции распределения непрерывной случайной величины*

$$\lim_{\delta \to 0} \frac{F_X(a+\delta) - F_x(a-\delta)}{2\delta} = F'_X = \frac{dF_X(x)}{dx}$$

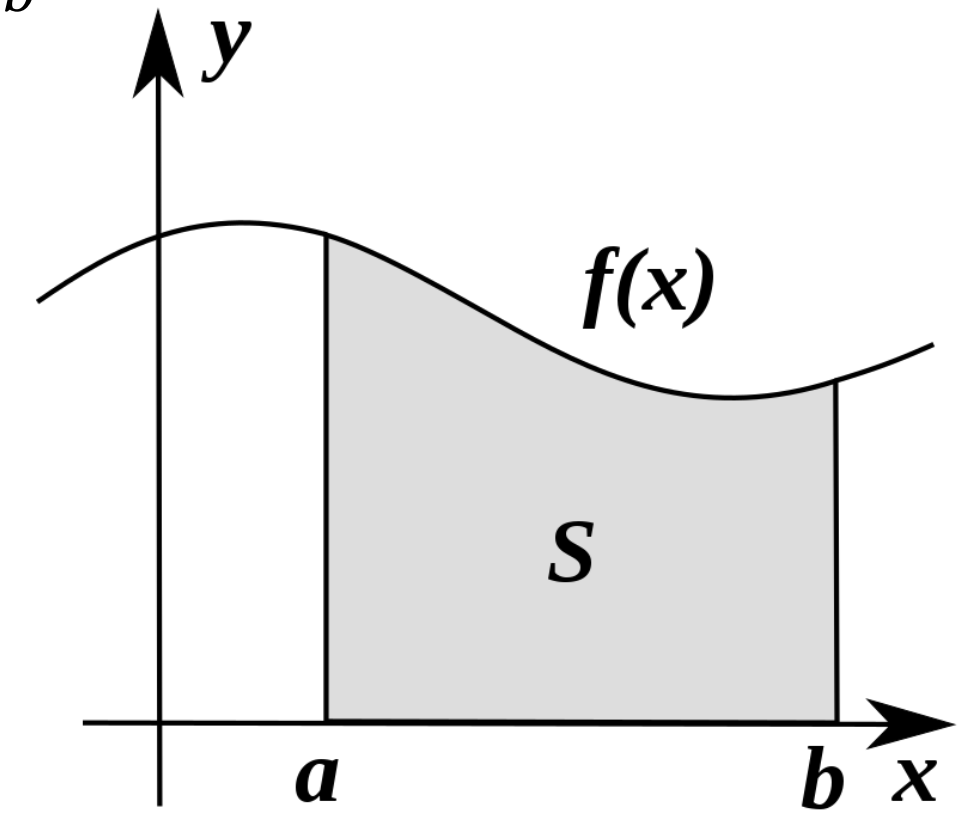
$$f_X = \frac{dF_x}{dx} - \text{функция плотности вероятности} *$$

*Есть некоторые нюансы, которые выходят за рамки лекции

Функция плотности

Аналог вероятности для непрерывных величин, но **вероятностью не является**. Отношение $\frac{f_X(a)}{f_X(b)}$ равно отношению вероятностей получить, близкую к a и величину, близкую к b

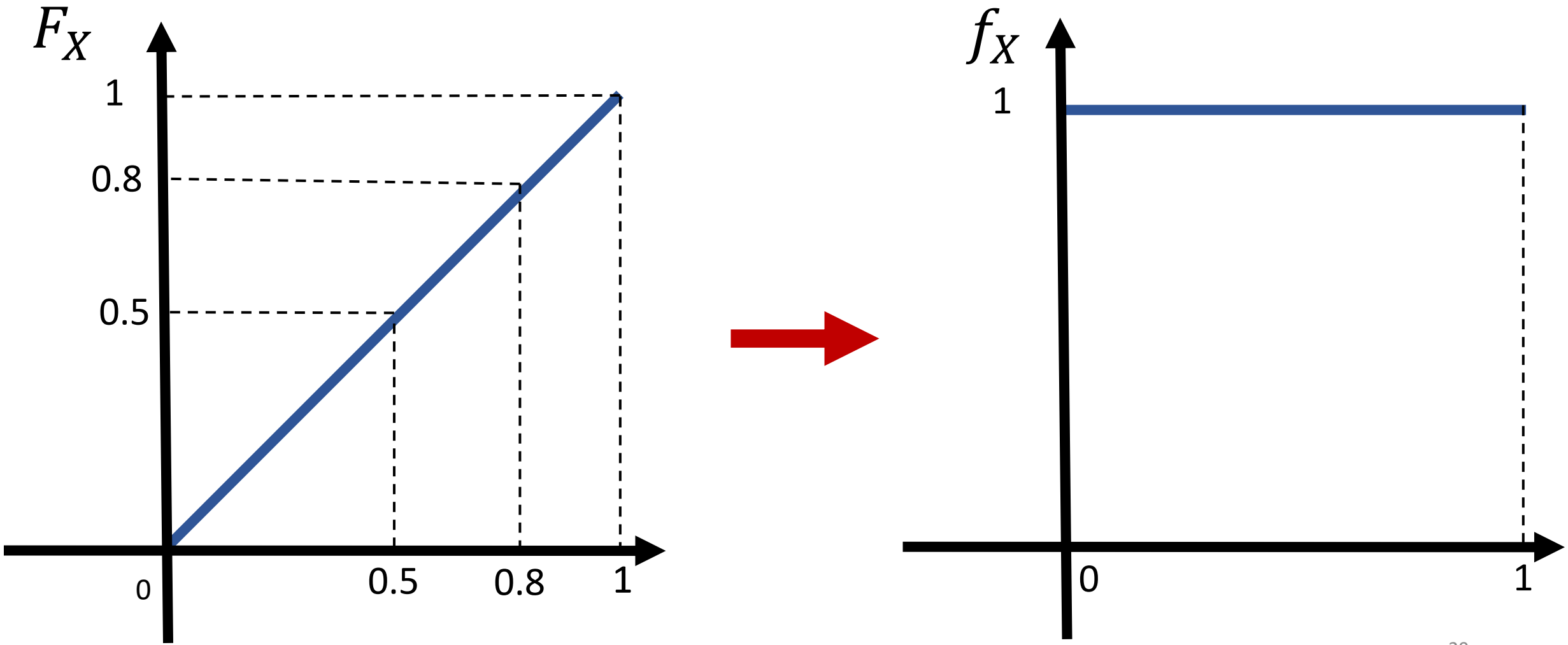
1. $f_X \geq 0$
2. $\int f_X(x)dx = 1$
3. $\int_{-\infty}^{A} f_X(x)dx = F_X(A)$



Вероятность того, что случайная величина, имеющая функцию плотности f_X примет значение от a до b – площадь под графиком этой функции

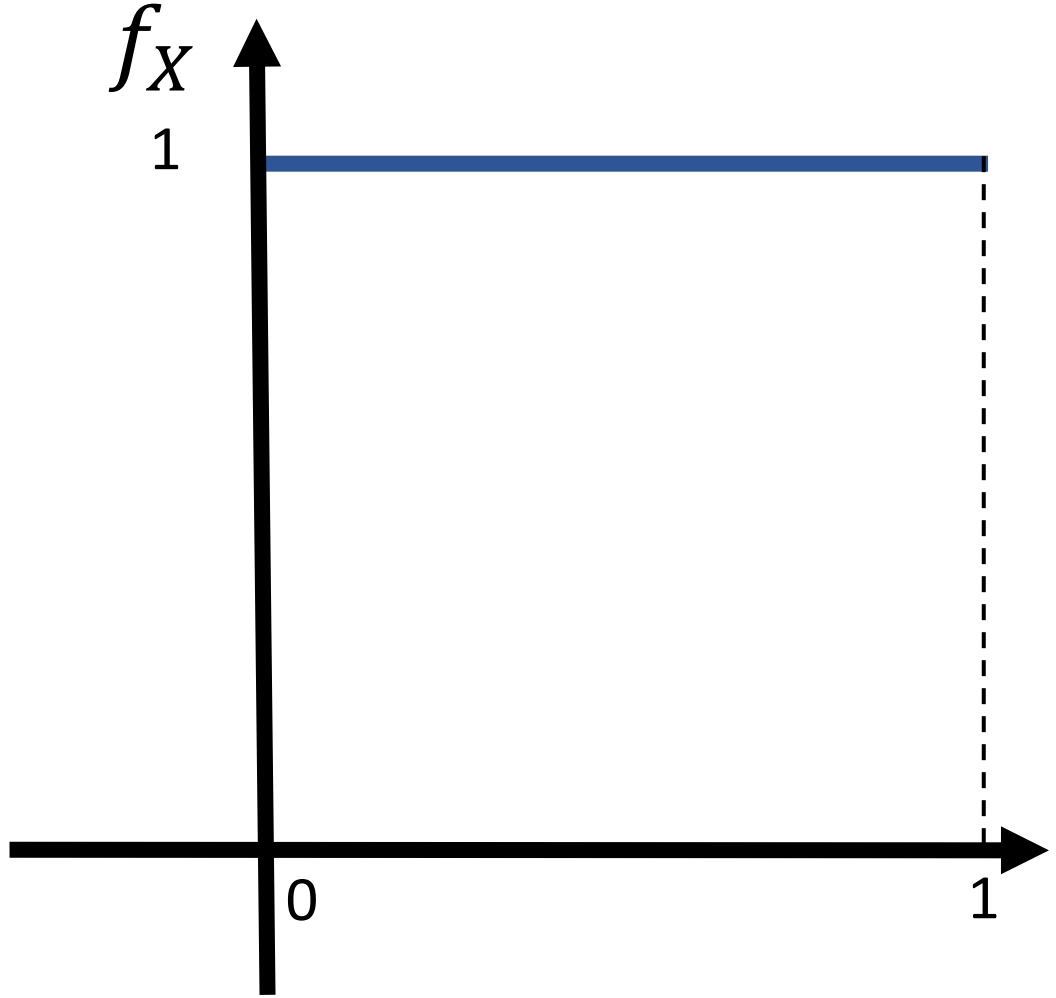
$$P(a \leq X \leq b) = F_X(b) - F_X(a) = \int_{-\infty}^{b} f_X(x)dx - \int_{-\infty}^{a} f_X(x)dx = \int_{a}^{b} f_X(x)dx$$

Функция плотности равномерного распределения



Функция плотности равномерного распределения

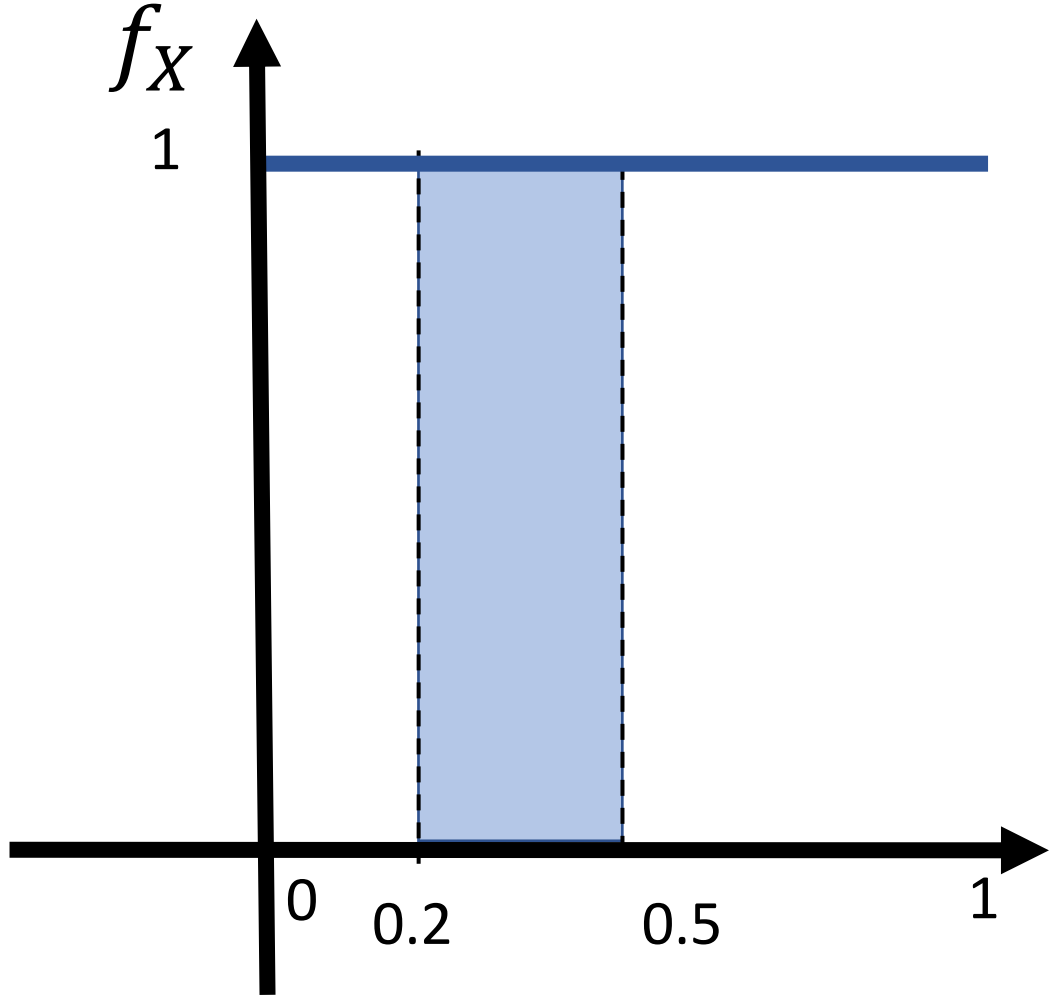
Берем случайное число из отрезка от 0 до 1. Какова вероятность, что оно будет находиться в отрезке [0.2, 0.5]?



Функция плотности равномерного распределения

Берем случайное число из отрезка от 0 до 1. Какова вероятность, что оно будет находиться в отрезке [0.2, 0.5]?

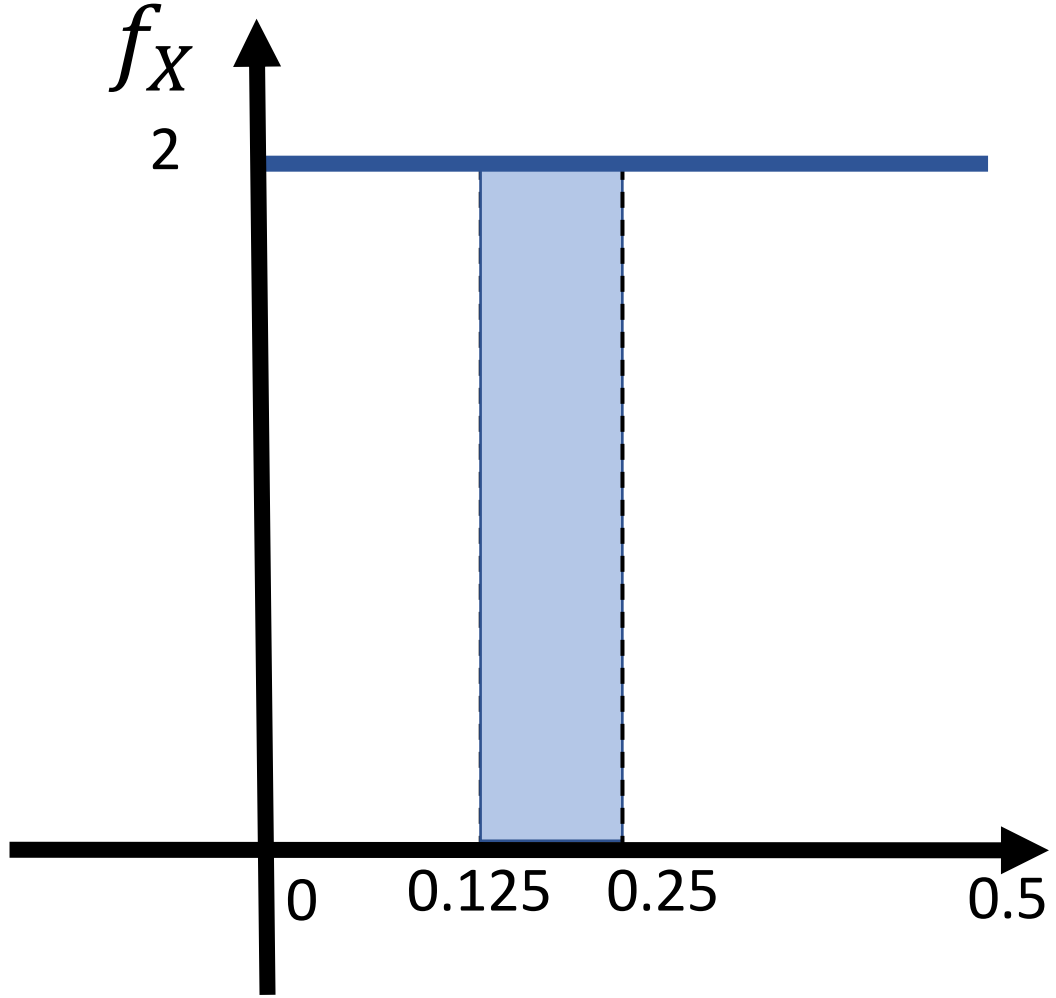
Просто площадь отмеченного прямоугольника



Функция плотности равномерного распределения

Берем случайное число из отрезка от 0 до 0.5. Какова вероятность, что оно будет находиться в отрезке [0.125, 0.25]?

Просто площадь отмеченного прямоугольника.



Среднее случайной величины

Для дискретного случая

$$E(X) = \sum_i x_i p_i$$

Для непрерывного случая

$$E(X) = \int_{-\infty}^{+\infty} x f_x(x) dx$$

Чему равно среднее суммы n независимых одинаково распределенных случайных величин?

Чему равна сумма n независимых одинаково распределенных случайных величин?

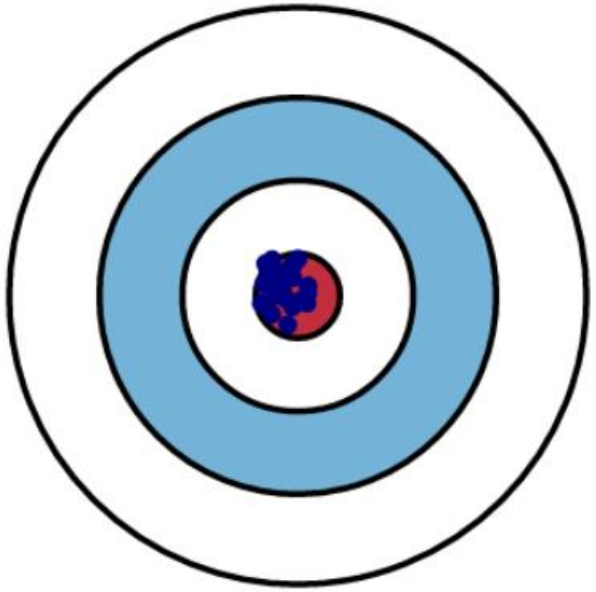
Пусть среднее каждой случайной величины μ. Тогда

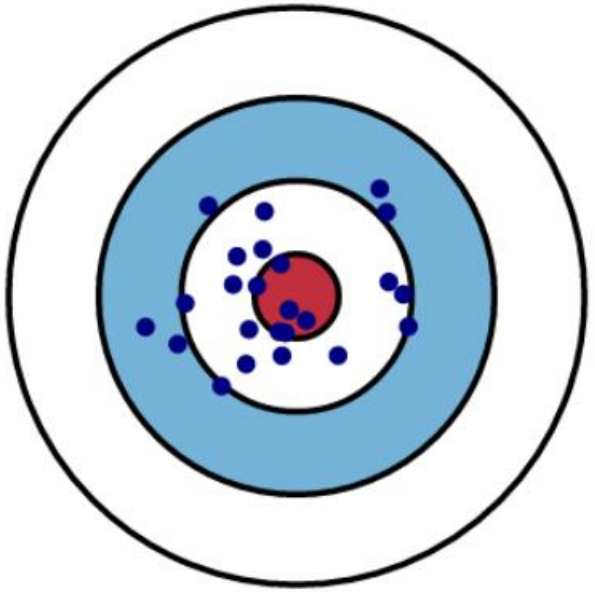
$$E(\sum_i X_i) = \sum_i E(X_i) = n\mu$$

Дисперсия случайной величины

$$\mathrm{D}(X) = E(X - EX)^2$$

Характеризует, насколько сильно величина склонна отклоняться от среднего. Формула не меняется





Чему равна дисперсия суммы n независимых одинаково распределенных случайных величин?

Чему равна дисперсия суммы n независимых одинаково распределенных случайных величин?

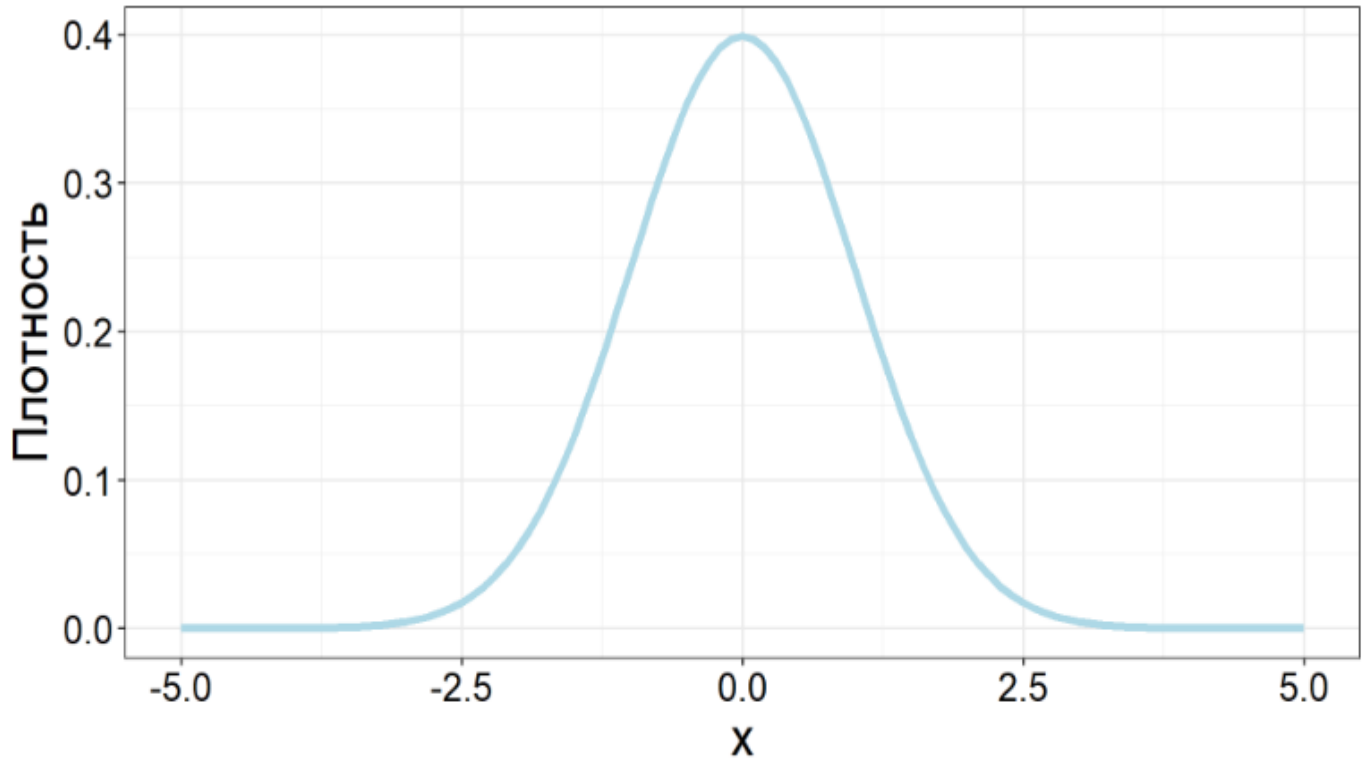
Пусть среднее каждой случайной величины μ. Тогда

$$\mathrm{D}(\sum_i X_i) = \sum_i D(X_i) = n\sigma^2$$

Нормальное распределение

Более всего знакомо именно по своей функции плотности

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



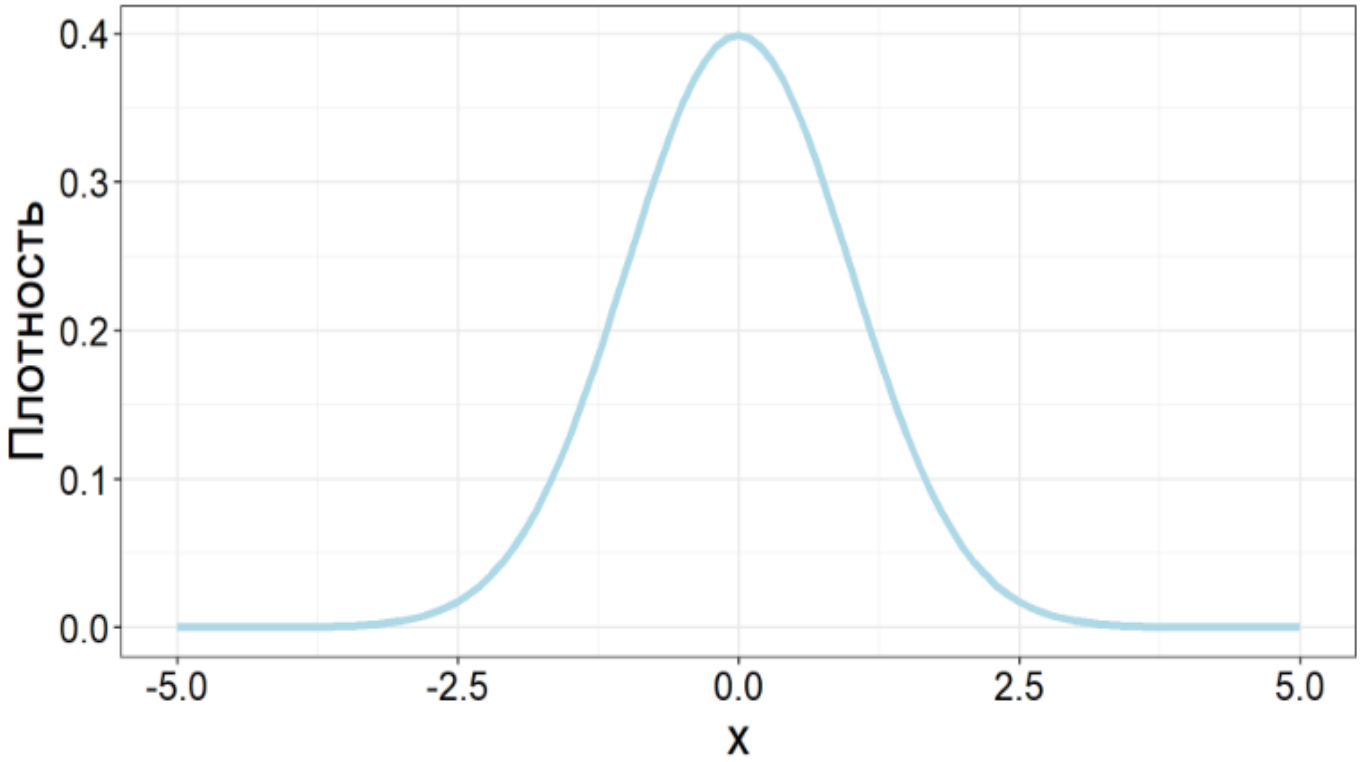
Принимает значения на $(-\infty, +\infty)$

Два параметра: среднее μ и дисперсия - σ

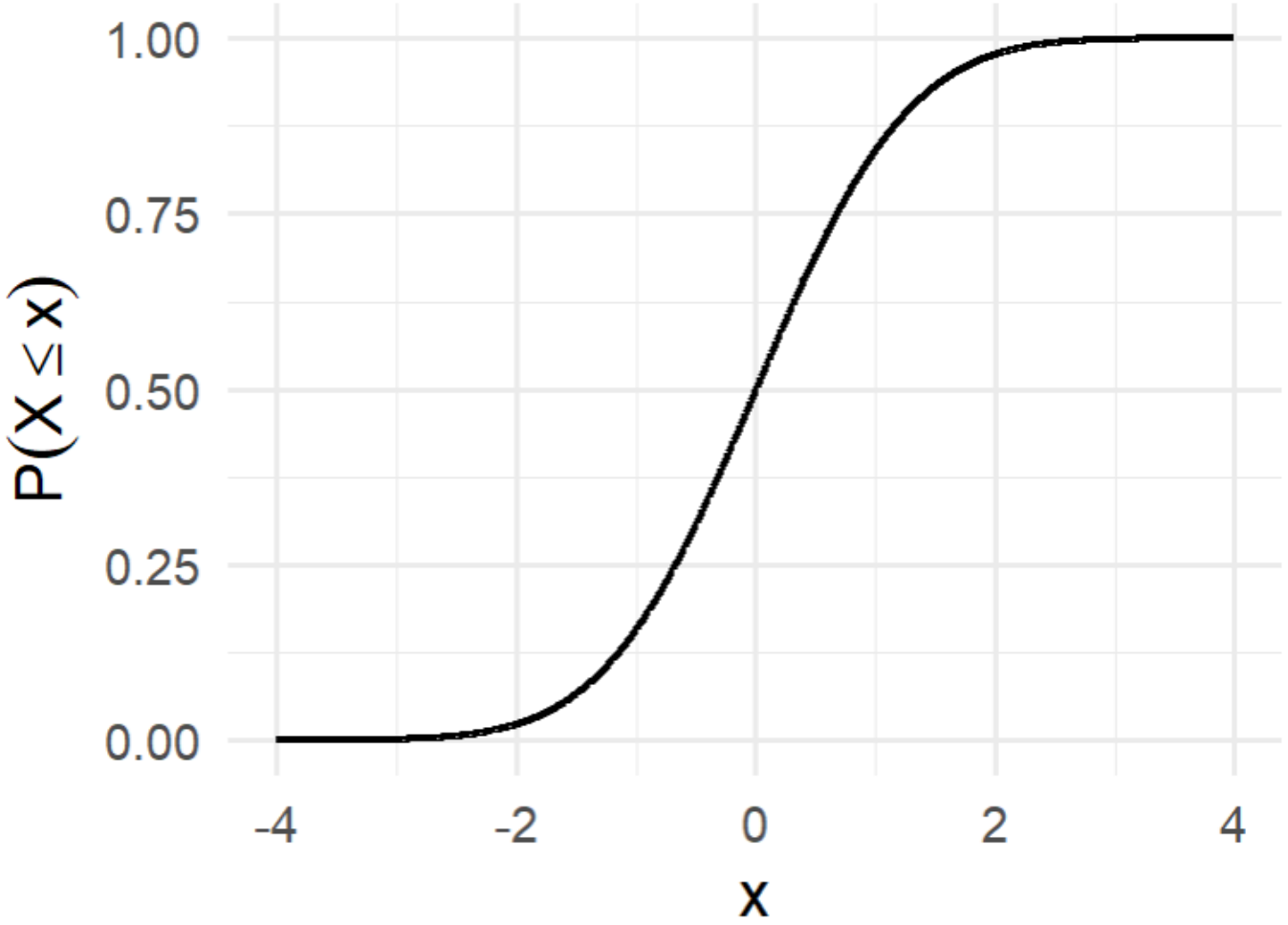
Стандартное нормальное распределение

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

среднее $\mu = 0$ и
дисперсия - $\sigma = 1$



Функция распределения нормального распределения



Нормальное распределение

Формула сложная, выглядит сложно, считать сложно, зачем оно??

Оно обладает рядом интересных свойств, которые позволяют его использовать в задачах, где другие не используешь. :

Пусть X распределено $N(\mu_a, \sigma_b^2)$, Y распределено $N(\mu_b, \sigma_b^2)$, тогда $Z = X + Y$ распределено $N(\mu_a + \mu_b, \sigma_a^2 + \sigma_b^2)$

Нормальное распределение

Оно обладает рядом интересных свойств, которые позволяют его использовать в задачах, где другие распределения не используешь (часто оно появляется в них автоматически):

Пусть X распределено $N(\mu_a, \sigma_b^2)$, Y распределено $N(\mu_b, \sigma_b^2)$, тогда $Z = X + Y$ распределено $N(\mu_a + \mu_b, \sigma_a^2 + \sigma_b^2)$

Пусть X распределено $N(\mu, \sigma^2)$, тогда $Z = \frac{X-\mu}{\sigma}$ распределено $N(0, 1)$.

Центральная предельная теорема

Пусть у нас есть n независимых случайных величин - $X_1, X_2, X_3, \dots X_n$. Помимо этого пусть эти случайные величины распределены одинаково, и пусть у каждой есть конечные матожидание μ и дисперсия σ^2. Пусть $S_n = \sum_i X_i$. Тогда при $n \to \infty$

$$\frac{S_n - n\mu}{\sigma\sqrt{n}} \xrightarrow[n\to\infty]{} N(0,1)$$

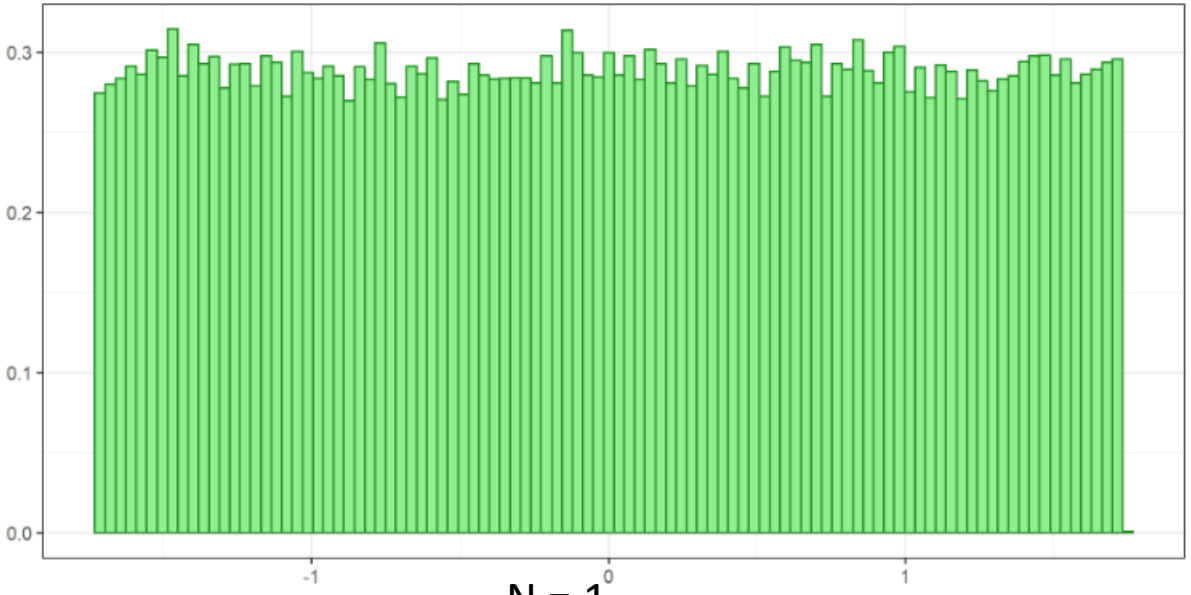
Центральная предельная теорема

Пусть у нас есть n независимых случайных величин - $X_1, X_2, X_3, \dots X_n$. Помимо этого пусть эти случайные величины распределены одинаково, и пусть у каждой есть конечные матожидание μ и дисперсия σ^2. Пусть $S_n = \sum_i X_i$. Тогда при $n \to \infty$

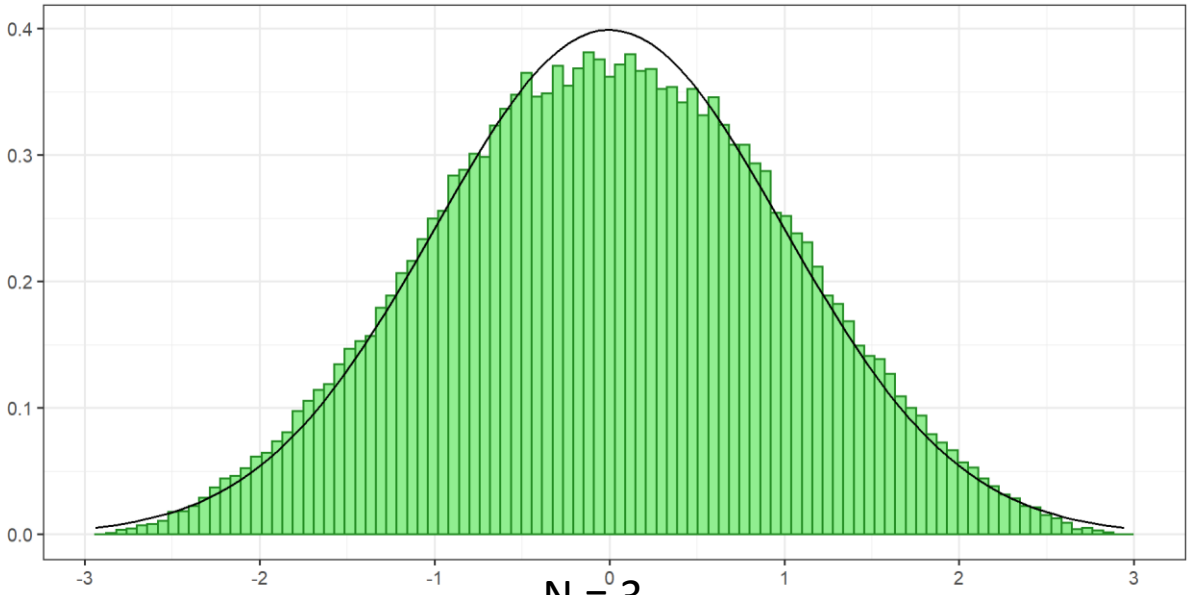
$$\frac{S_n - n\mu}{\sigma\sqrt{n}} \xrightarrow[n\to\infty]{} N(0,1)$$

Почему эта теорема так важна?

ЦПТ для равномерного распределения



N = 1

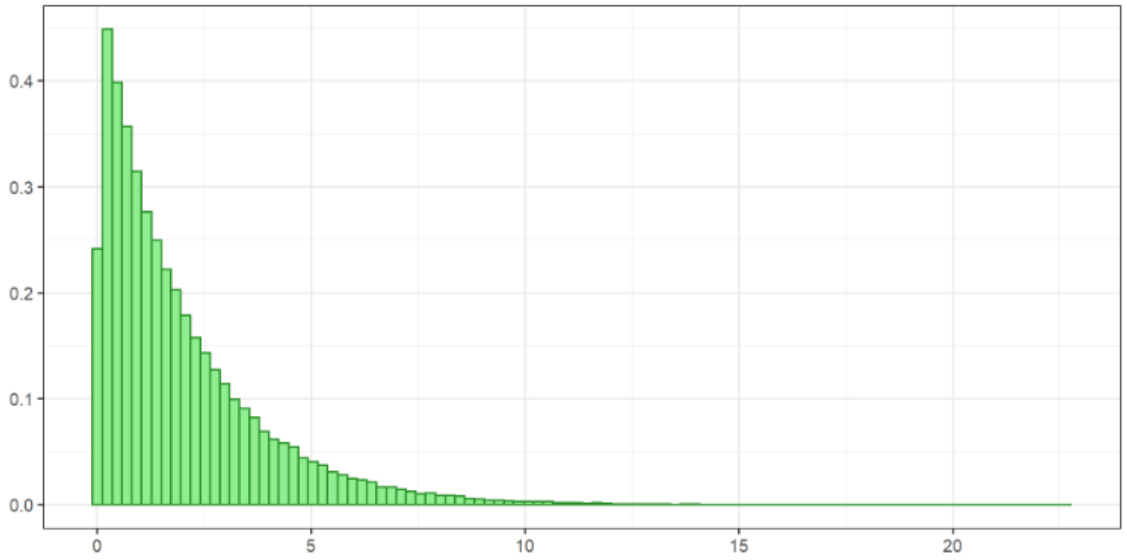


N = 3

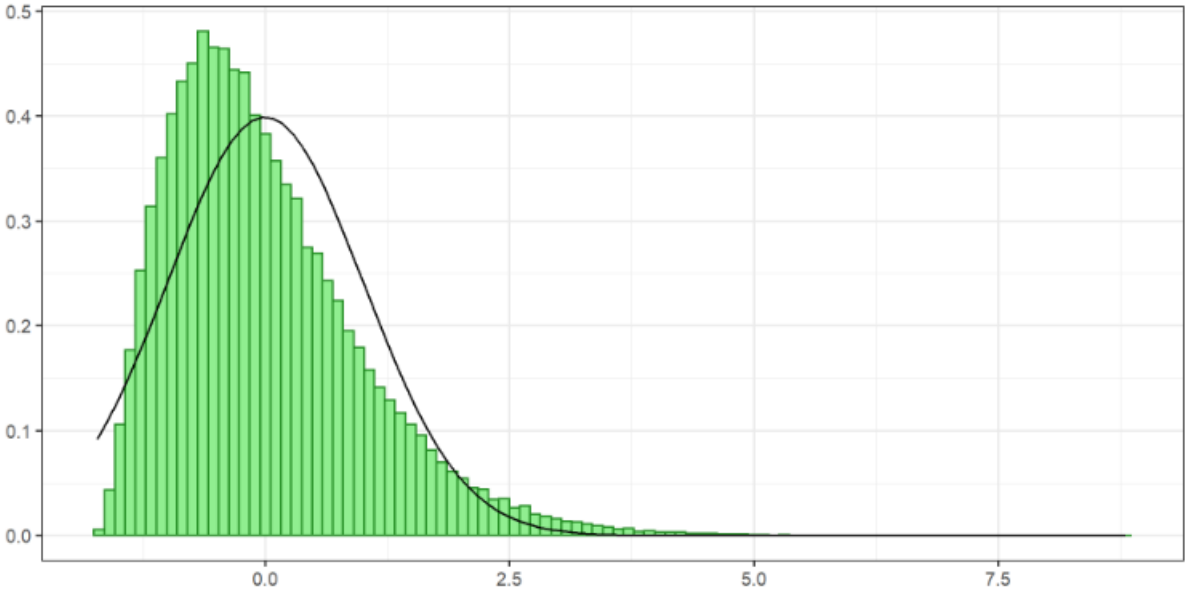
N = 5

N = 10

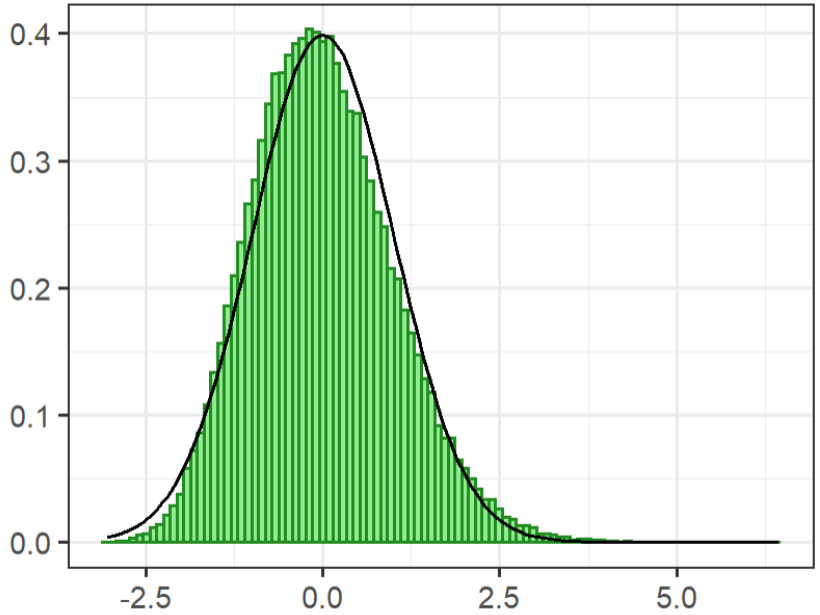
ЦПТ для экспоненциального распределения



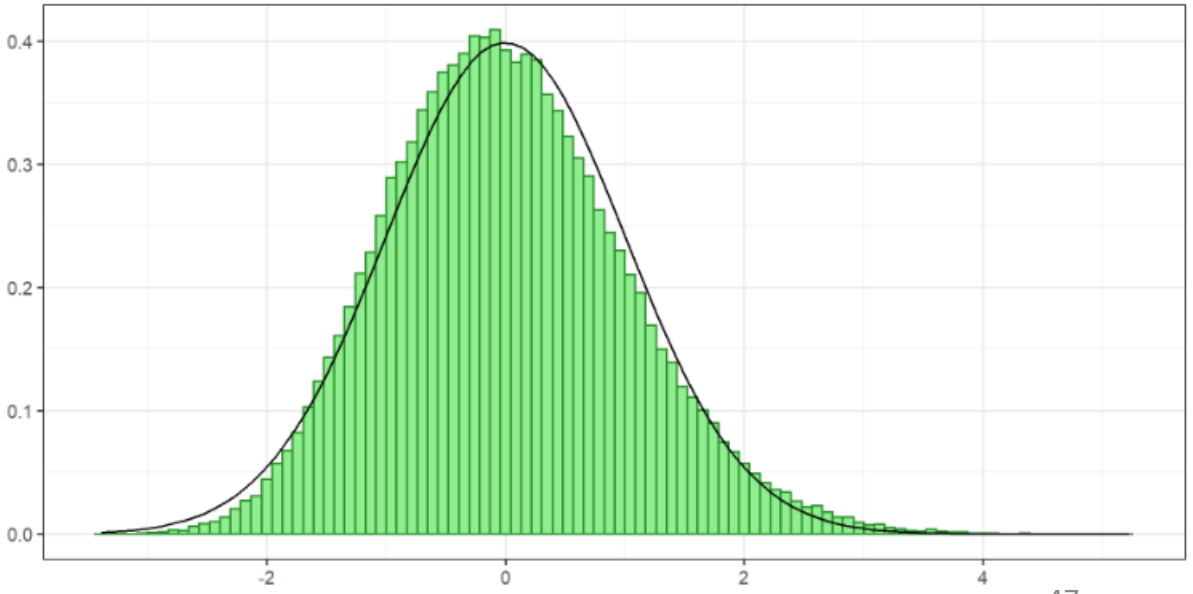
N = 1



N = 3

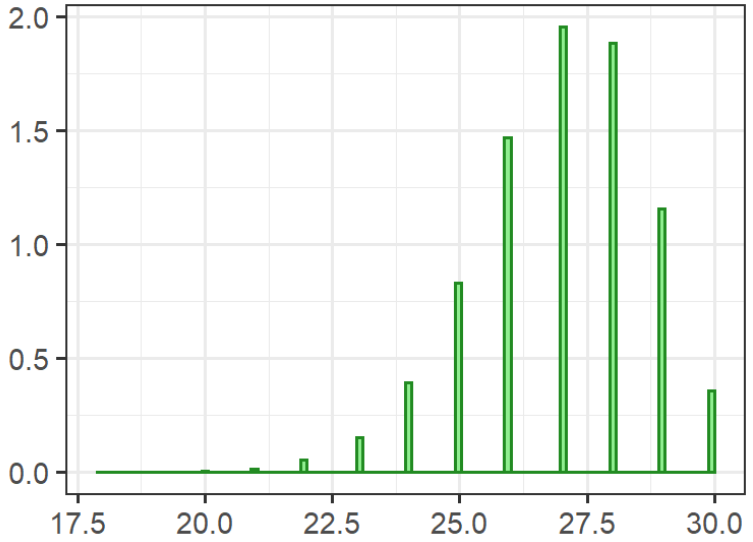


N = 25

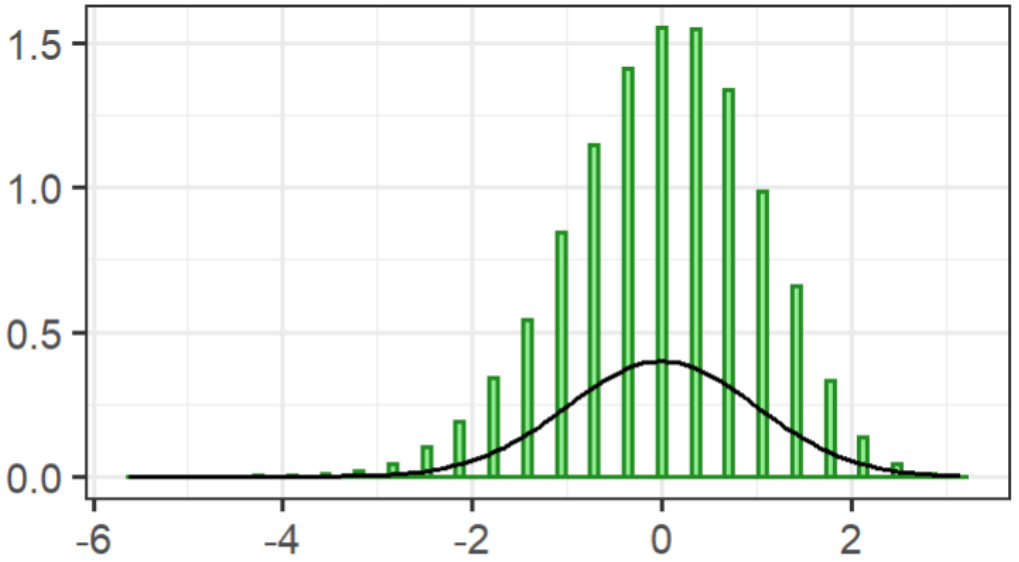


N = 40

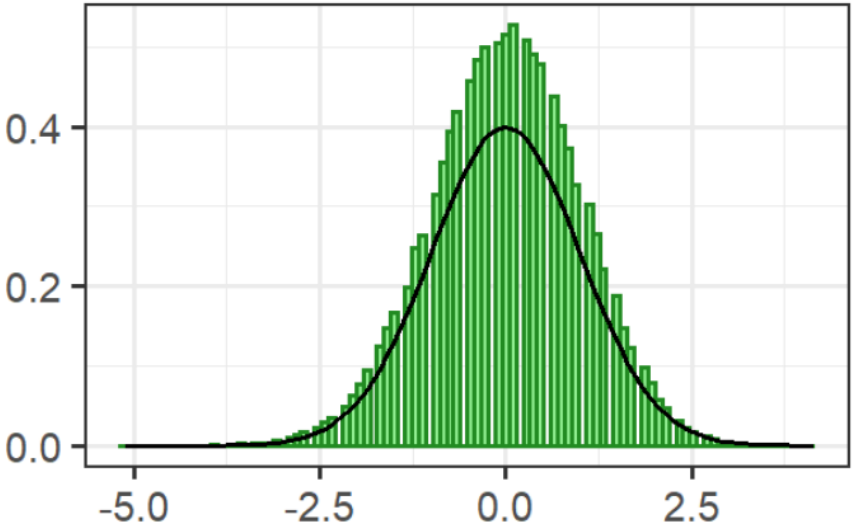
ЦПТ для биномиального распределения (p=0.9, число испытаний 50)



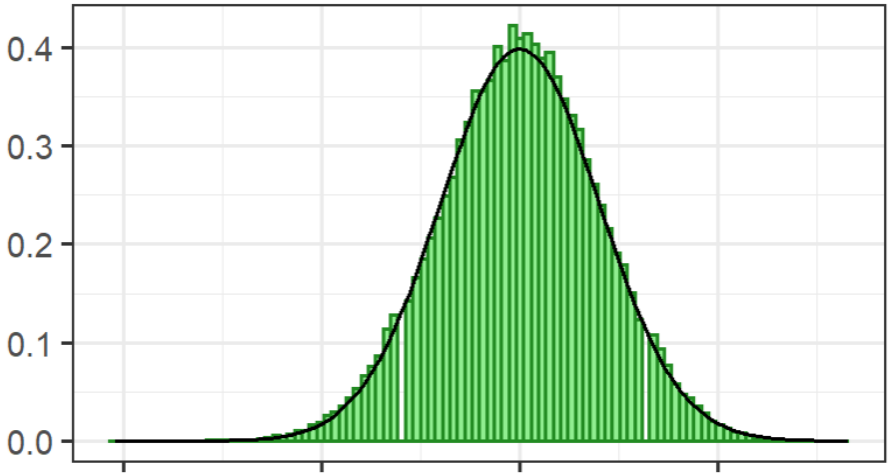
N = 1



N = 3



N = 25



N = 40

Центральная предельная теорема

Сходимость как правило быстрая.
При $n \geq 40^*$ уже можем ожидать нормального распределения.
Неформально – сумма достаточного числа
независимых одинаково распределенных
случайных величин распределена близко
к нормальному распределению. Нам **не особо важно знать**, как именно распределены эти величины.

$$S_n \sim N(n\mu, n\sigma^2)$$

*разумеется – n – «магическое» число. Может и не сойтись на 40

Центральная предельная теорема и среднее

Сходимость как правило быстрая.

При $n \geq 40$ уже можем ожидать нормального распределения.

Нам **не особо важно знать**, как именно распределены эти величины.

$$S_n \sim N(n\mu, n\sigma^2)$$

$\overline{X} = \frac{S_n}{n}$ - выборочное среднее. Например – средний результат 50 измерений

$$\overline{X} \sim N(\mu, \frac{\sigma^2}{n})$$

Центральная предельная теорема и шум в данных

Сходимость как правило быстрая.

Нам **не особо важно знать**, как именно распределены эти величины.

Шум в данных – результат **суммарного** действия большого числа мелких случайных величин с 0 средним. Потому мы можем предполагать, что

$$noise \sim N(0, \sigma^2)$$

Центральная предельная теорема и шум в данных

Сходимость как правило быстрая.

Нам **не особо важно знать**, как именно распределены эти величины.

Шум в данных – результат **суммарного** действия большого числа мелких независимых случайных величин с 0 средним. Потому мы можем предполагать, что

$$noise \sim N(0, \sigma^2)$$

Но нам еще нужно, чтобы распределение этих величин было одинаковым, разве нет?

Центральная предельная теорема и шум в данных

Сходимость как правило быстрая.

Нам **не особо важно знать**, как именно распределены эти величины.

Шум в данных – результат **суммарного** действия большого числа мелких независимых случайных величин с 0 средним. Потому мы можем предполагать, что

$$noise \sim N(0, \sigma^2)$$

Но нам еще нужно, чтобы распределение этих величин было одинаковым, разве нет?

Оказывается, есть вариант теоремы для по-разному распределенных величин. Среди них «просто» не должно быть величины, «забивающей» остальные.

Центральная предельная теорема и шум в данных

Сходимость как правило быстрая.

Нам **не особо важно знать**, как именно распределены эти величины.

Шум в данных – результат **суммарного** действия большого числа мелких независимых случайных величин с 0 средним. Потому мы можем предполагать, что

$$noise \sim N(0, \sigma^2)$$

А что если величины зависимы?

Центральная предельная теорема и шум в данных

Сходимость как правило быстрая.

Нам **не особо важно знать**, как именно распределены эти величины.

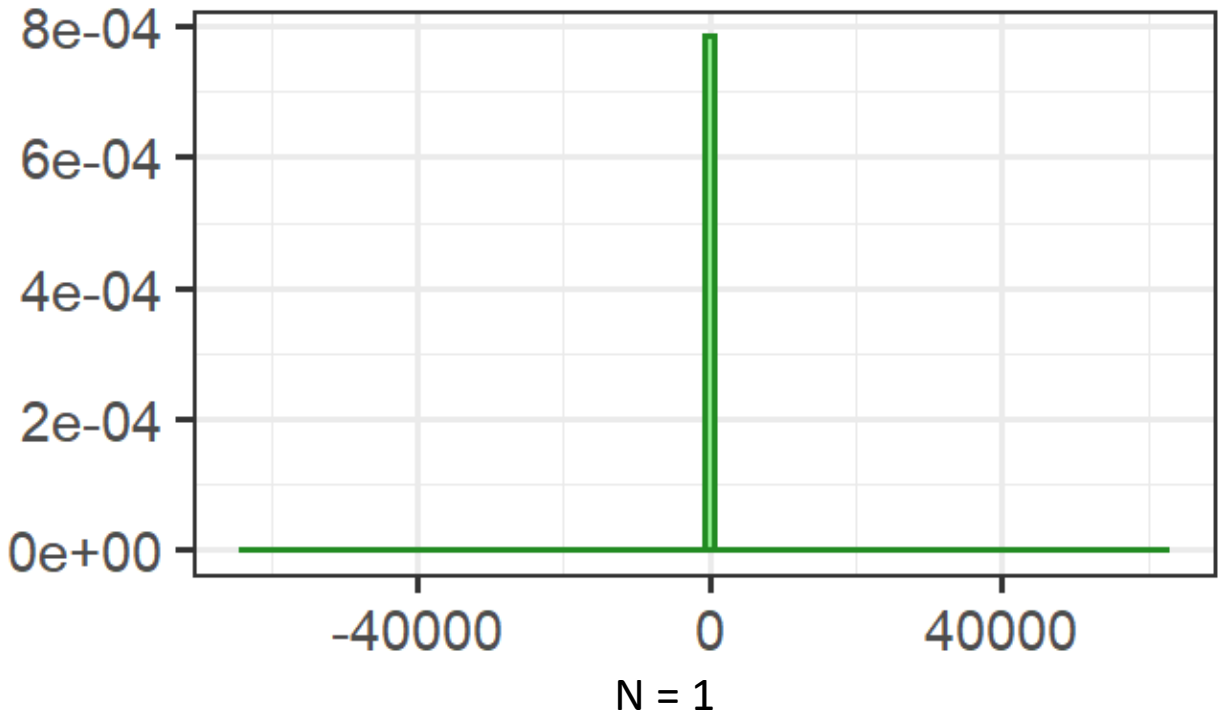
Шум в данных – результат **суммарного** действия большого числа мелких независимых случайных величин с 0 средним. Потому мы можем предполагать, что

$$noise \sim N(0, \sigma^2)$$

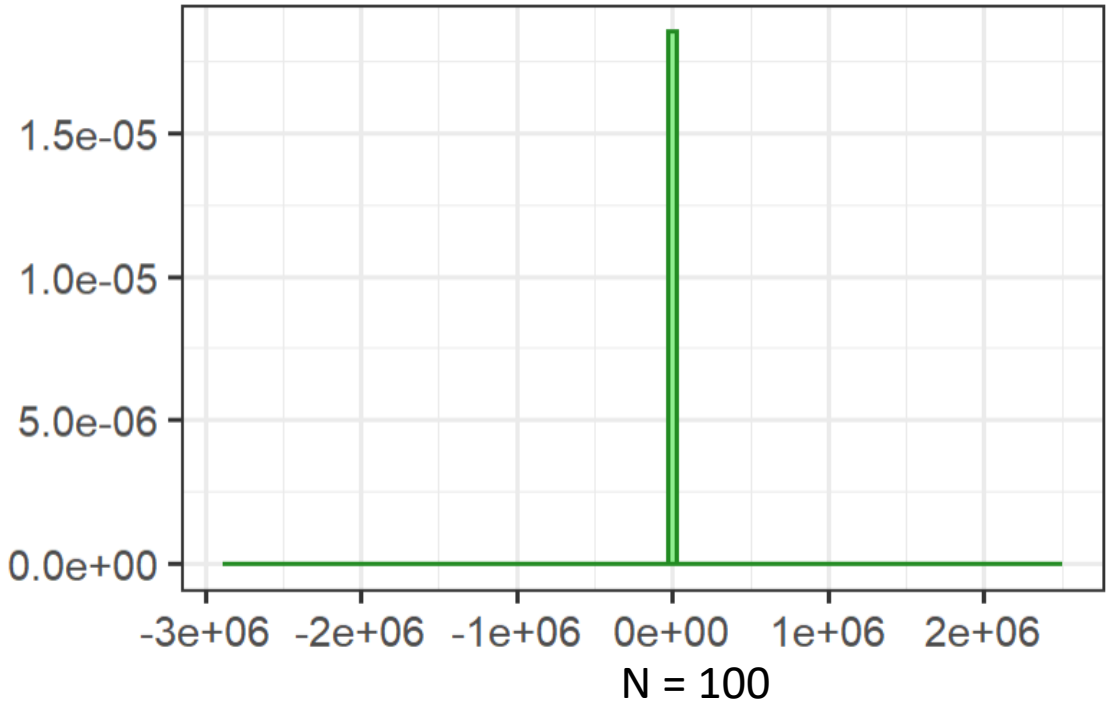
А что если величины зависимы?

Для слабой зависимости тоже получится свой вариант ЦПТ

Центральная предельная теорема и распределение Коши



N = 1

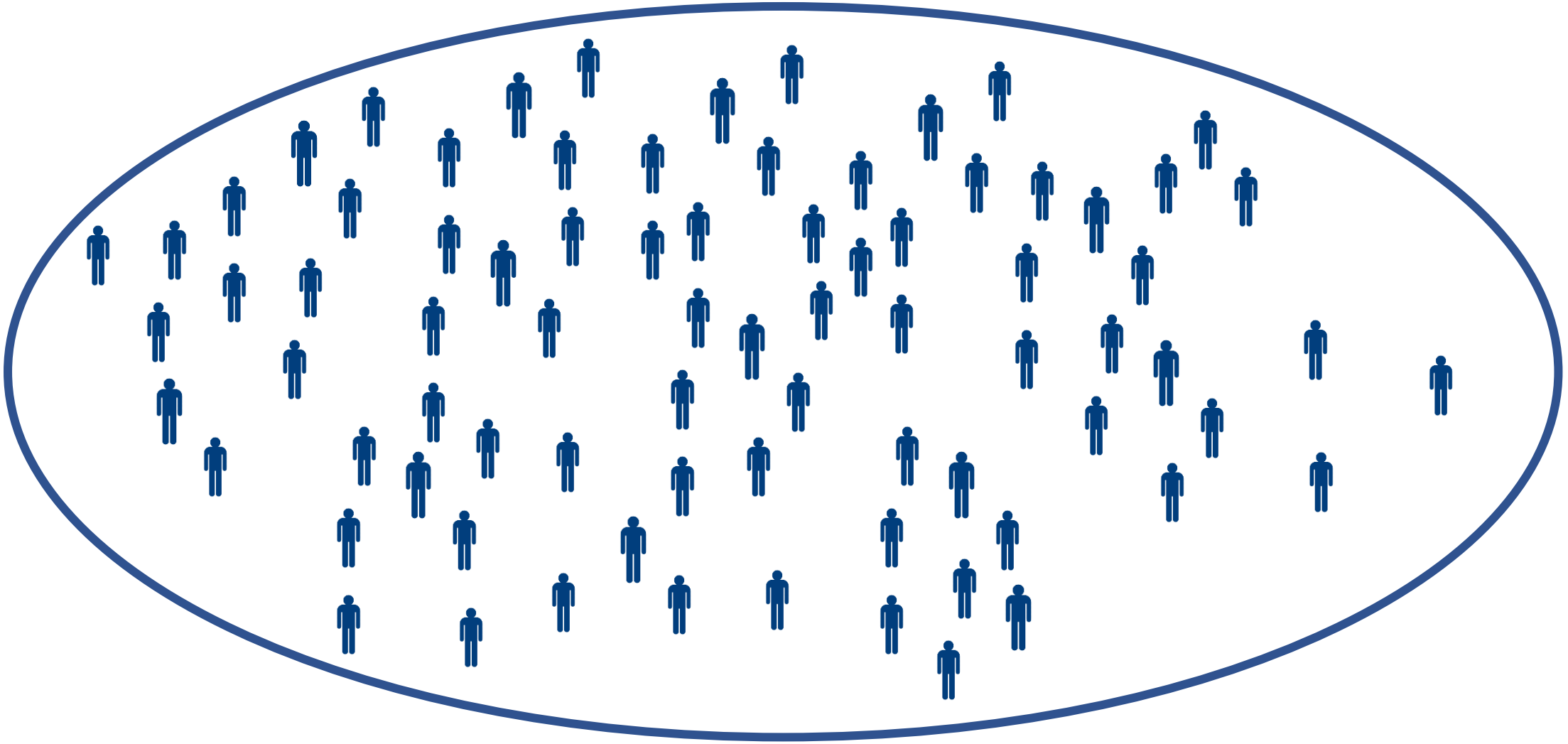


N = 100

У распределения Коши нет даже матожидания. Потому для него ЦПТ не выполнится. Но вероятность встретить подобные распределения в реальном эксперименте очень-очень мала

Введение в статистику

Генеральная совокупность



Набор всех возможных объектов, реальных или гипотетических, которые нас интересуют в данном исследовании и о которых мы хотим сделать какие-то выводы

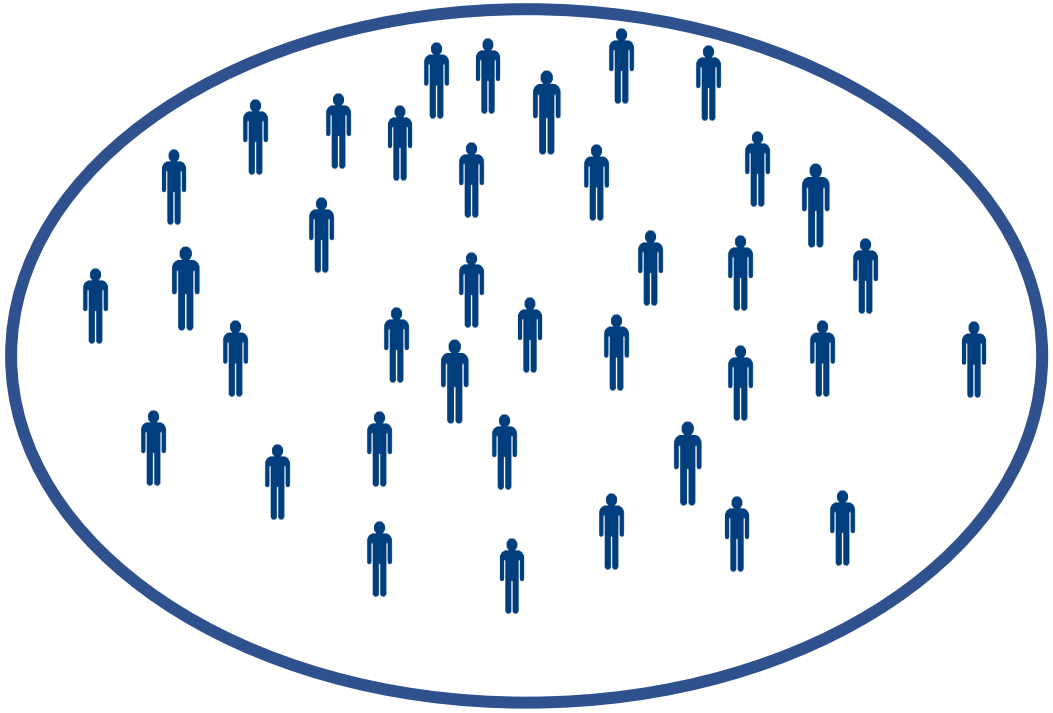
Генеральная совокупность: примеры

Набор всех возможных объектов, реальных или гипотетических, которые нас интересуют в данном исследовании и о которых мы хотим сделать какие-то выводы

1. Ответы людей в возрасте от 30 до 49 лет, проживающие на территории РФ в данный момент, на вопрос «Поддерживаете ли вы политику, проводимую X?»
2. Зарплаты всех мужчин в возрасте от 19 до 29 лет в США
3. Оценки IQ на основе определенного теста **T** всех людей, реальных и гипотетических
4. Оценки, выставленные учителями данной методике
5. Результаты всех возможных подбрасываний данной монетки
6. Результаты терапии, примененной ко всем возможным пациентам с данным диагнозом
7. …

Генеральная совокупность: параметры

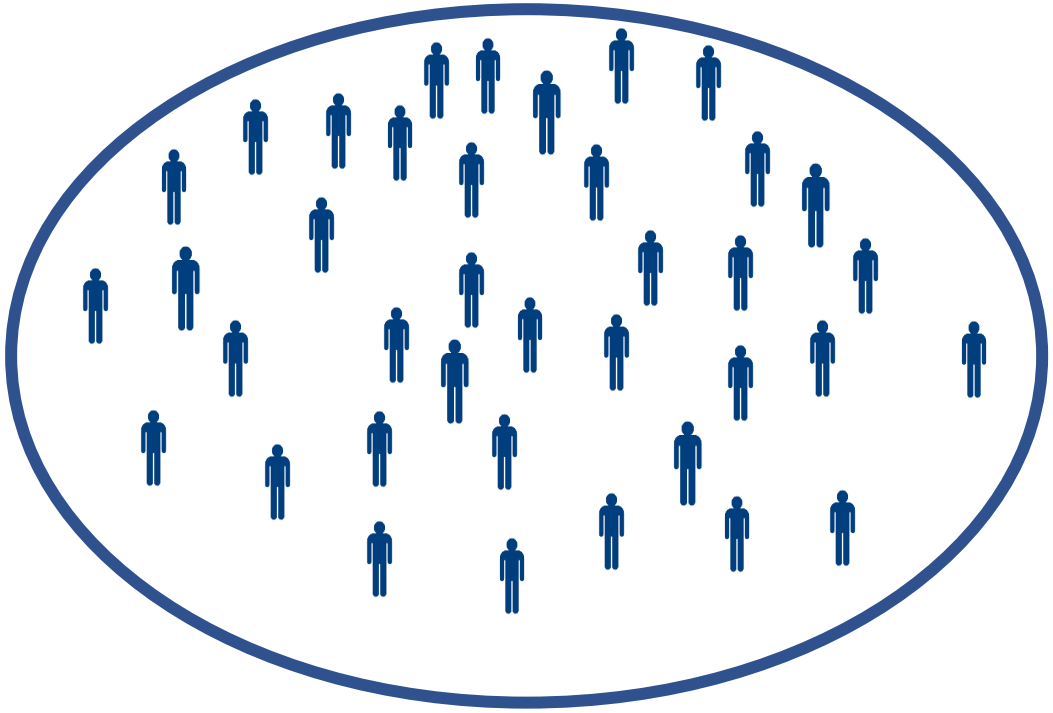
Мы предполагаем, что данная совокупность описывается неким набором параметров. То есть, зная все эти параметры, мы можем сами генерировать объекты генеральной совокупности, неотличимые от реальных



$$Model(\mu, \sigma^2, \alpha, \beta, \gamma, \theta, \dots)$$

Генеральная совокупность: параметры

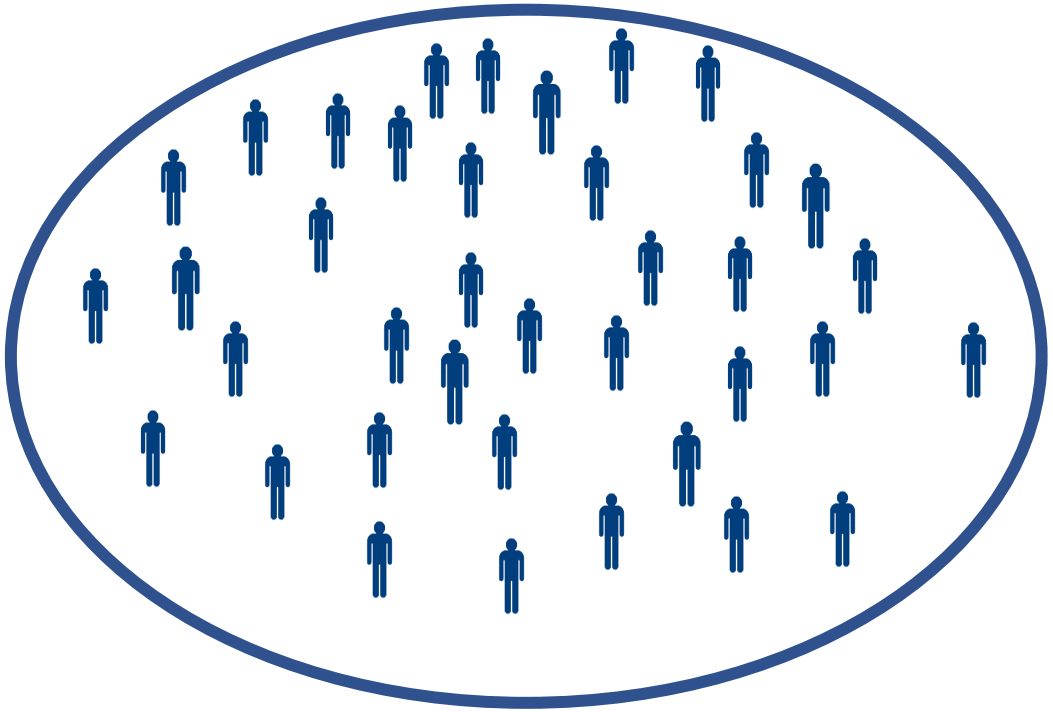
Мы предполагаем, что данная совокупность описывается неким набором параметров. То есть, зная все эти параметры, мы можем сами генерировать объекты генеральной совокупности, неотличимые от реальных



$$Model(\mu, \sigma^2, \alpha, \beta, \gamma, \theta, \dots)$$

В статистике вывода (inference statistics) мы предполагаем,
что эти параметры строго заданы и константны.
Например, данное заболевание при такой терапии и таких данных пациента заканчивается благополучно в 95% случаев

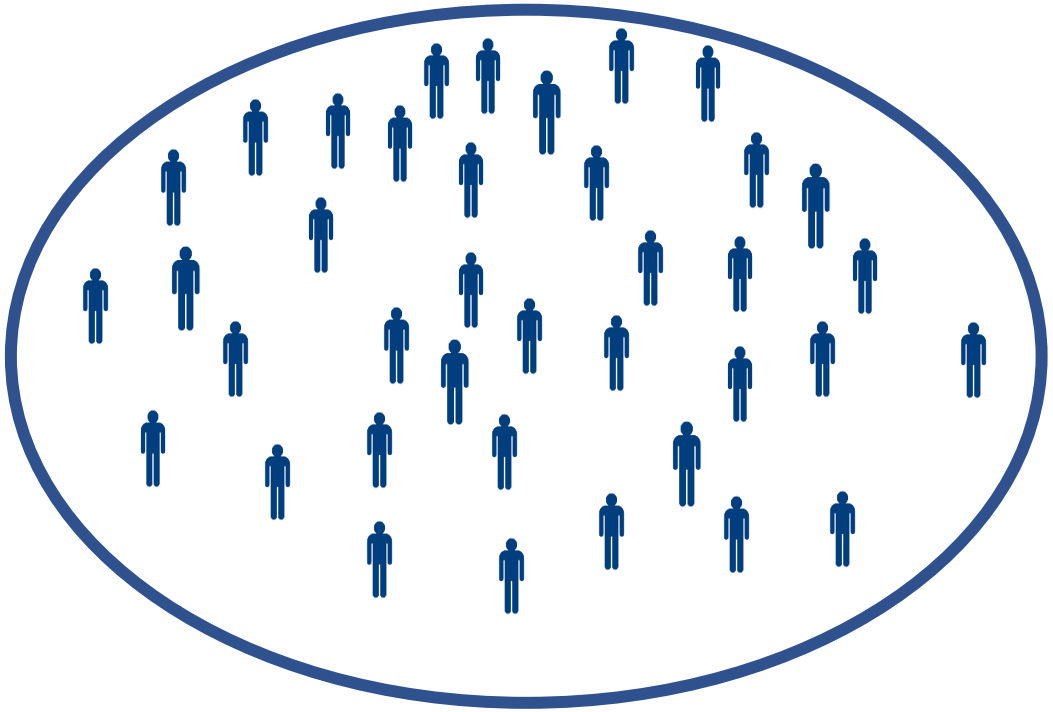
Генеральная совокупность: оценка параметров



$$Model(\mu, \sigma^2, \alpha, \beta, \gamma, \theta, \dots)$$

Если нам известна все генеральная совокупность – мы можем в точности оценить все ее параметры.

Генеральная совокупность: оценка параметров



$$Model(\mu, \sigma^2, \alpha, \beta, \gamma, \theta, \dots)$$

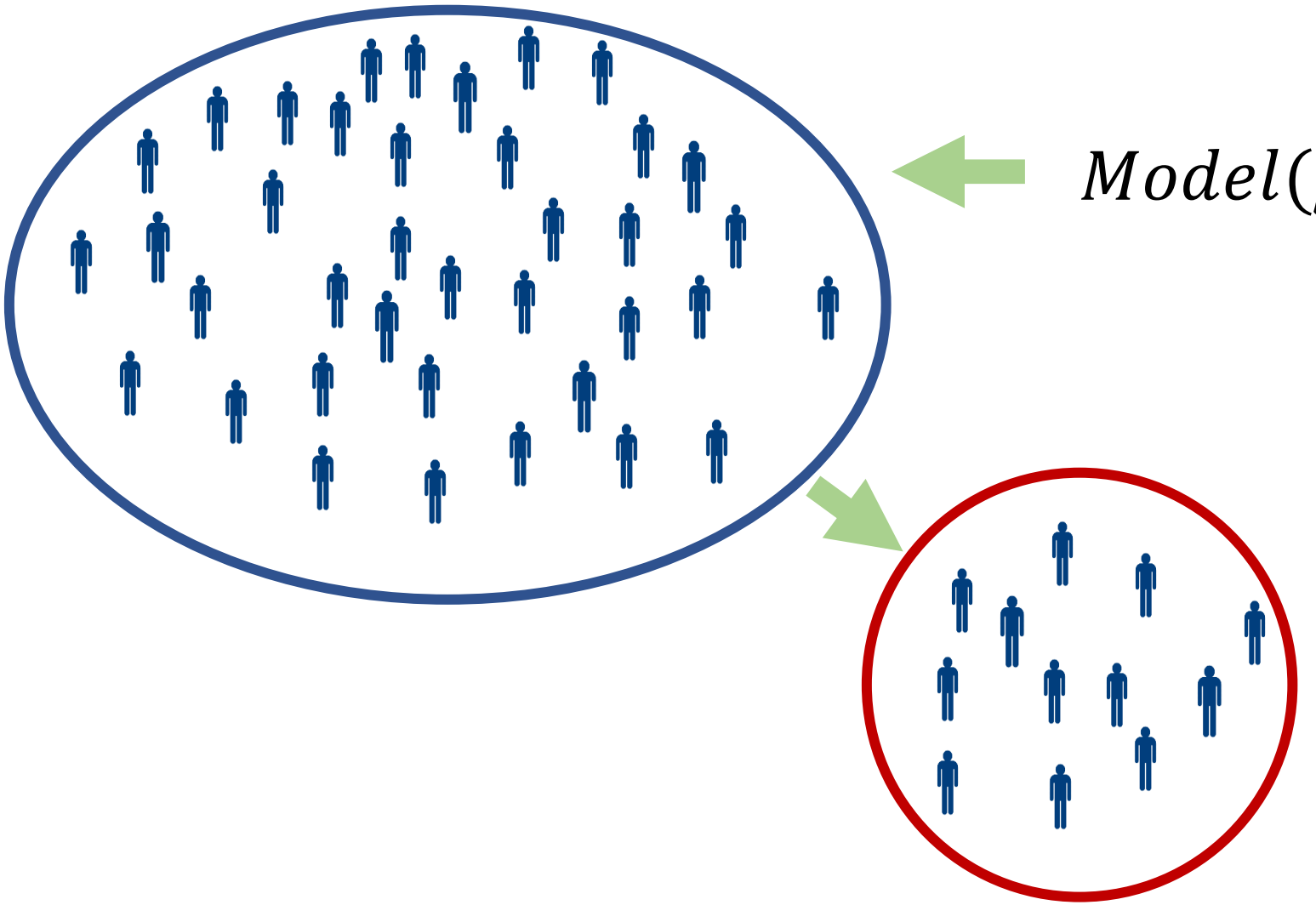
Если нам известна все генеральная совокупность – мы можем в точности оценить все ее параметры.

Однако в большинстве случаев получить информацию обо ВСЕХ объектах генеральной совокупности очень дорого или невозможно.

Генеральная совокупность: пример бесконечной совокупности

Результаты всех возможных бросков данной монетки

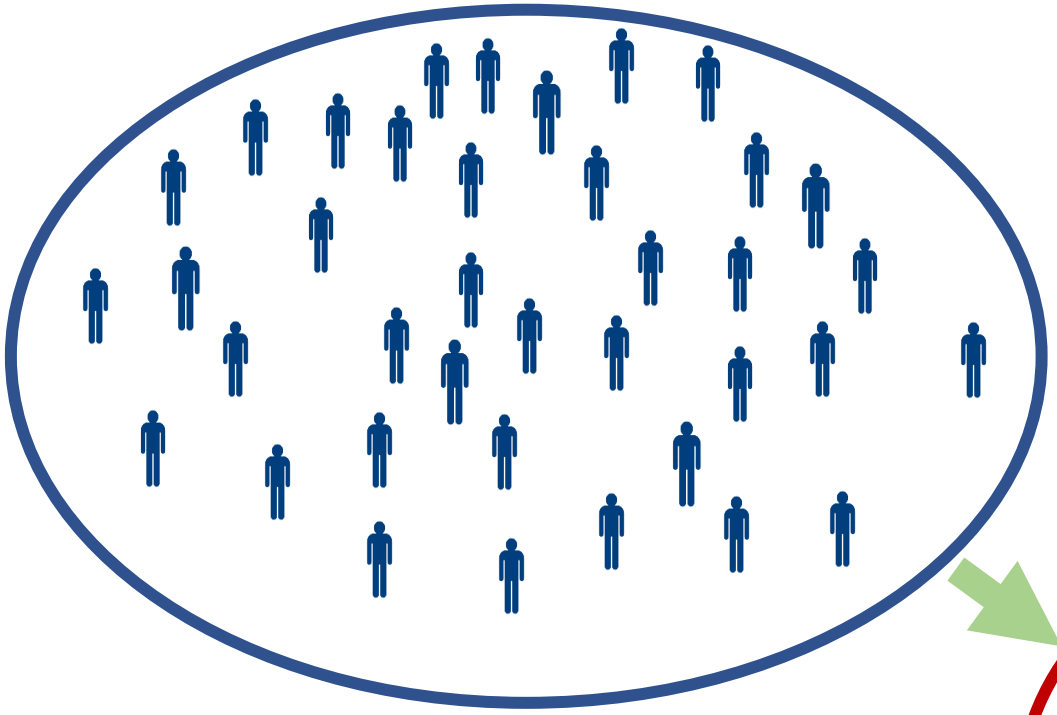
Выборка из генеральной совокупности



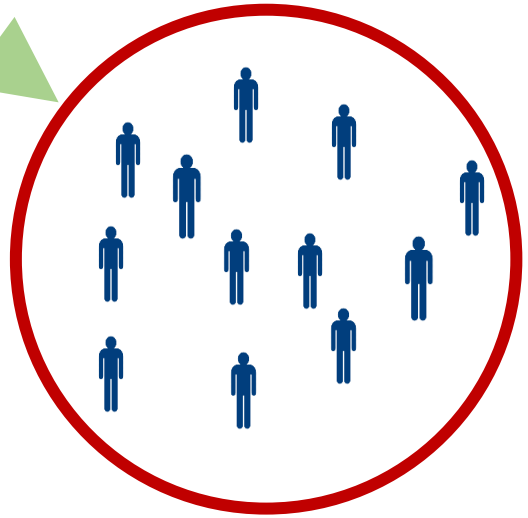
$$Model(\mu, \sigma^2, \alpha, \beta, \gamma, \theta, \dots)$$

Мы выбираем только конечный и обычно сравнительно малый поднабор из генеральной совокупности

Выборка из генеральной совокупности



$$Model(\mu, \sigma^2, \alpha, \beta, \gamma, \theta, \dots)$$



$$\hat{\mu}, \widehat{\sigma^2}, \hat{\alpha}, \hat{\beta}, \hat{\gamma}, \hat{\theta}, \dots)$$

Наши оценки параметров генеральной совокупности

Оценка параметра

Пусть у нас есть монетка и мы хотим оценить для нее вероятность выпадения орла. В реальности монетка абсолютно честная, p=0.5

Оценка параметра

Пусть у нас есть монетка и мы хотим оценить для нее вероятность выпадения орла. В реальности монетка абсолютно честная, p=0.5

$$\hat{p} = \frac{2}{4} = 0.5$$

Оценка параметра

Пусть у нас есть монетка и мы хотим оценить для нее вероятность выпадения орла. В реальности монетка абсолютно честная, p=0.5

$$\hat{p} = \frac{1}{4} = 0.25$$

Оценка параметра

Пусть у нас есть монетка и мы хотим оценить для нее вероятность выпадения орла. В реальности монетка абсолютно честная, p=0.5

$$\hat{p} = \frac{4}{4} = 1$$

Оценка параметра

Пусть у нас есть монетка и мы хотим оценить для нее вероятность выпадения орла. В реальности монетка абсолютно честная, p=0.5

Чем является $\hat{p}$?

Оценка параметра

Пусть у нас есть монетка и мы хотим оценить для нее вероятность выпадения орла. В реальности монетка абсолютно честная, p=0.5

Чем является $\hat{p}$?

Число выпавших орлов – случайная величина, распределенная по биномиальному закону

$$k \sim Binomial(n = 4, p = 0.5)$$

Тогда и $\hat{p} = \frac{k}{n}$ также случайная величина.

Оценка параметра

Пусть у нас есть монетка и мы хотим оценить для нее вероятность выпадения орла. В реальности монетка абсолютно честная, p=0.5

Чем является $\hat{p}$?

Число выпавших орлов – случайная величина, распределенная по биномиальному закону

$$k \sim Binomial(n = 4, p = 0.5)$$

Тогда и $\hat{p} = \frac{k}{n}$ также случайная величина.

Из того, что наша выборка является случайной величиной (если выборка != генеральной совокупности), следует, что и оценка параметра является случайной величиной, не обязательно равной реальному значению параметра

Оценка параметра

Если оценка параметра – случайная величина, то у нее можно пытаться посчитать матожидание и дисперсию

$$E(\hat{p}) = E\left(\frac{k}{n}\right) = \frac{E(k)}{n} = \frac{np}{n} = p$$

То есть в среднем наша оценка равна оцениваемому параметру

$$D(\hat{p}) = D\left(\frac{k}{n}\right) = \frac{D(k)}{n^2} = \frac{npq}{n^2} = \frac{pq}{n}$$

То есть наша оценка может отличаться от оцениваемого параметра, при этом с ростом размера выборки n вероятность сильного отклонения становится все меньше.

Оценка параметра – матожидание и дисперсия

Если оценка параметра – случайная величина, то у нее можно пытаться посчитать матожидание и дисперсию

$$E(\hat{p}) = E\left(\frac{k}{n}\right) = \frac{E(k)}{n} = \frac{np}{n} = p$$

То есть в среднем наша оценка равна оцениваемому параметру

$$D(\hat{p}) = D\left(\frac{k}{n}\right) = \frac{D(k)}{n^2} = \frac{npq}{n^2} = \frac{pq}{n}$$

То есть наша оценка может отличаться от оцениваемого параметра, при этом с ростом размера выборки n вероятность сильного отклонения становится все меньше.

Оценка параметра - смещение

Строго говоря – нас интересует не собственно матожидание оценки, а матожидание того, насколько она отличается от реального значения параметра

$$Bias = E(\hat{p} - p) = E(\hat{p}) - p = p - p = 0$$

Это матожидание называется **смещением** оценки. В данном случае смещение равно 0 – наша оценка **несмещенная**

Оценка – любая функция над выборкой

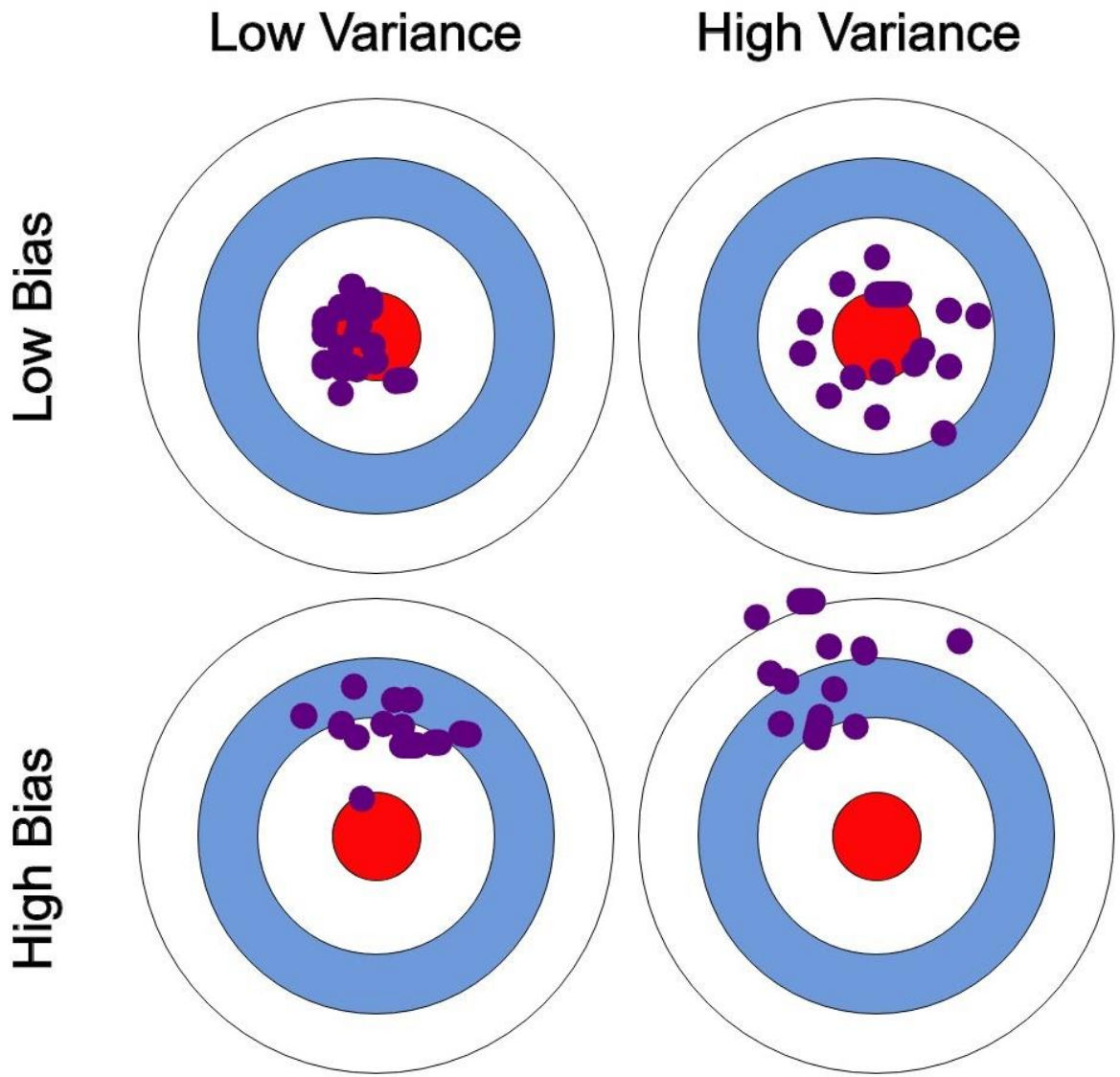
Можно сказать, что оценкой параметра является любая функция от выборки. Например:

$\hat{p}_1 = \frac{(\sum_{i=1}^{3} x_i)}{3}$ - то есть обращаем внимание только на первые 3 наблюдения. Всегда

$$\text{Bias}(\hat{p}_1) = E\left(\frac{k_1}{3}\right) - p = \frac{3p}{3} - p = 0$$

$$\text{D}(\hat{p}_1) = D\left(\frac{k_1}{3}\right) = \frac{3pq}{9} = \frac{pq}{3}$$

Bias и variance



В иделе мы хотим иметь оценку и с маленьким смещением, и с маленькой дисперсией.

Bias и variance

Реальность жестока – часто мы можем уменьшать одно, увеличивая другое. Например, пусть у нас такая оценка:

$\hat{p}_2 = \frac{k+3}{n+6}$ - то есть фактически мы добавляем 6 испытаний, из которых 3 окончились выпадением орла

$$\text{Bias}(\hat{p}_2) = E\left(\frac{k+3}{n+6}\right) - p = \frac{3+np}{n+6} - p \neq 0$$

$$\text{D}(\hat{p}_1) = D\left(\frac{k+3}{n+6}\right) = \frac{D(k+3)}{(n+6)^2} = \frac{D(k)}{(n+6)^2} = \frac{npq}{(n+6)^2} \leq \frac{\text{pq}}{\text{n}}$$

Выборочное среднее

Оценка среднего генеральной совокупности

$$\bar{x} = \frac{\sum_i x_i}{n}$$

Выборочная дисперсия

Оценка дисперсии генеральной совокупности

$$s^2 = \frac{1}{n-1}\sum(x_i - \bar{x})^2$$

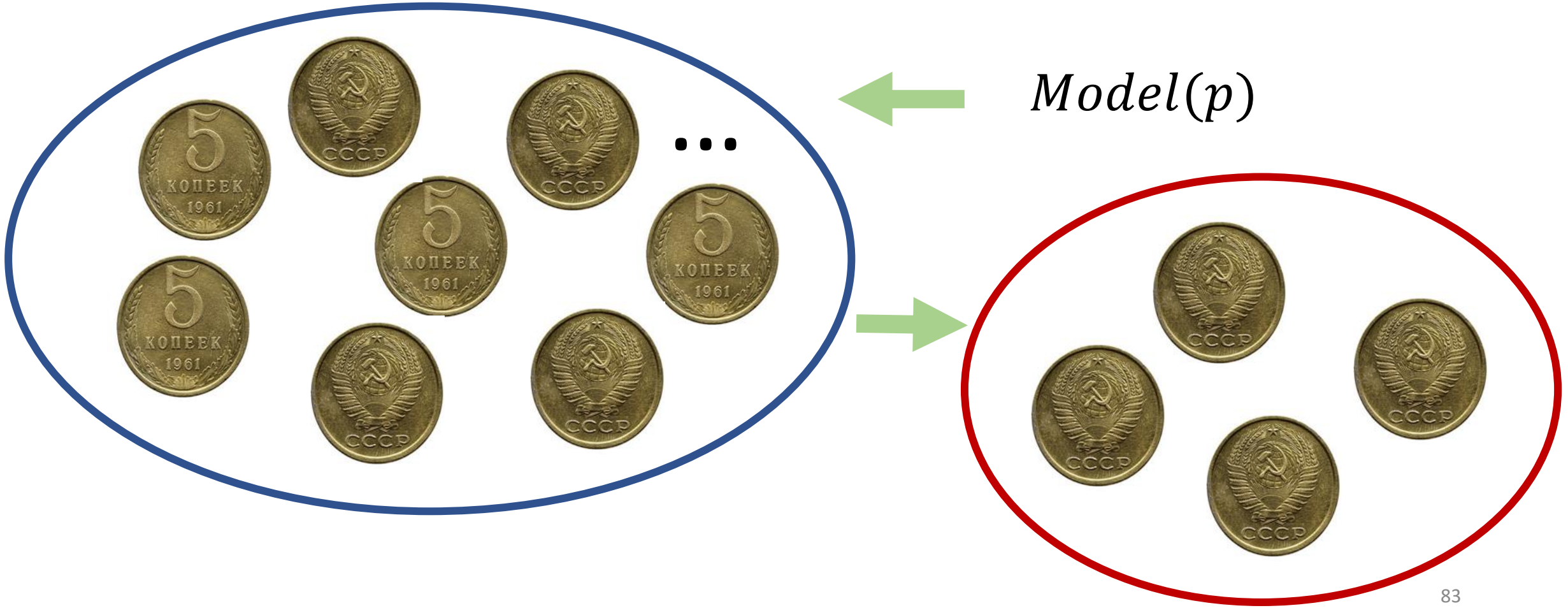
Выборочное стандартное отклонение

Оценка стандартного отклонения (коря из дисперсии) генеральной совокупности

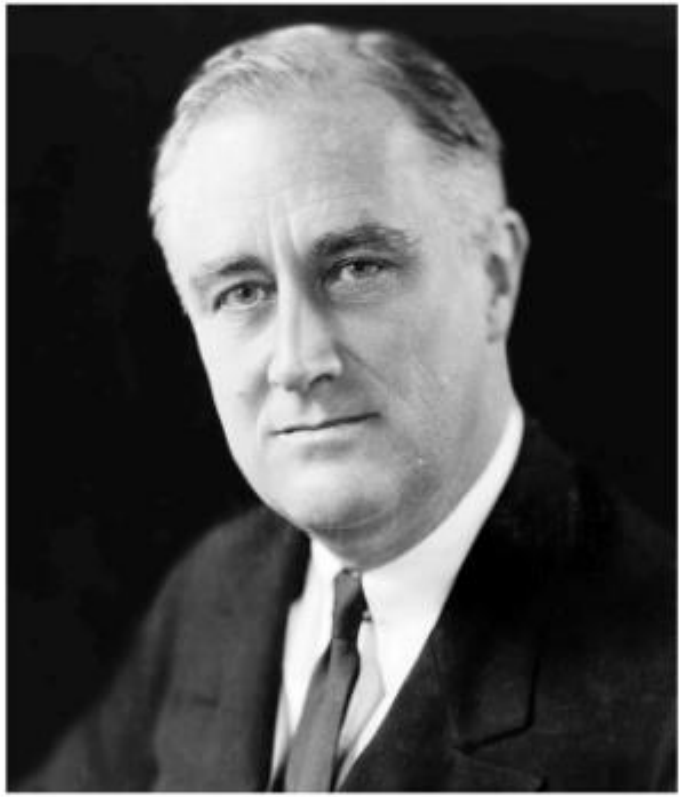
$$s = \sqrt{\frac{1}{n-1}\sum(x_i - \bar{x})^2}$$

Проблемы с выборкой

Что мешает мне брать в выборку только выпадения орла? Или брать их чаще, игнорируя часть выпадений решки?



Репрезентативность выборки



Журнал «Литерари Дайджес» - 2 млн собранных подписей,
победа Альфреда Лондона
Гэллап – 3000 подписей – победа второго кандидата.
Кому верить?

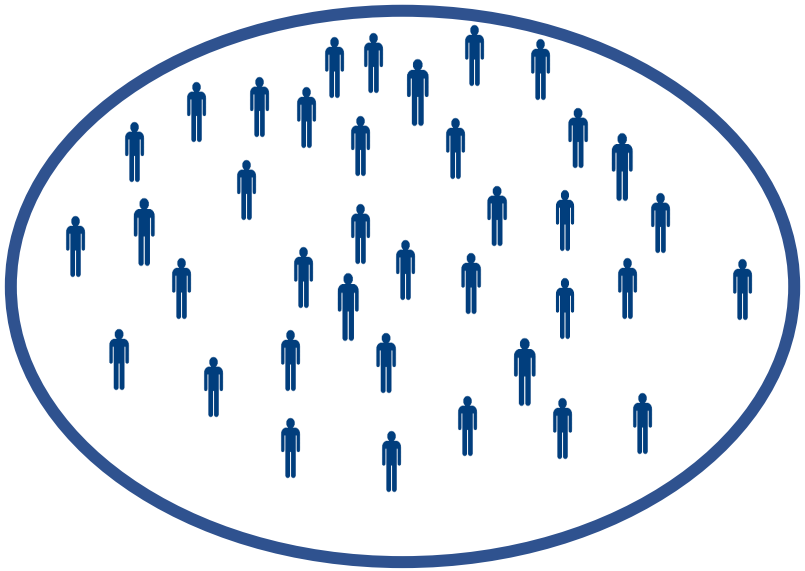
Репрезентативность выборки

Журнал «Литерари Дайджес» - 2 млн собранных подписей у владельцев телефонов (роскошь на тот момент), победа Альфреда Лондона.
Гэллап – 3000 подписей – победа Рузвельта. Правильно предсказал и то, что “Литерали Дайжджес” ошибочно предскажет победу Лондона

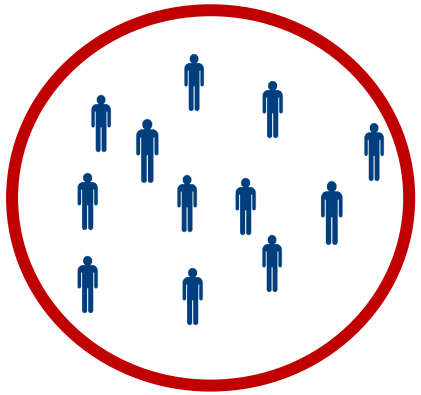

переиграл
и
УНИЧТОЖИЛ

Гипотезы

Так как у любой оценки, посчитанной по выборке, есть дисперсия, мы никогда не можем утверждать что-то наверняка. Мы можем лишь на основе наших данных оценивать достоверность наших предположений – гипотез – о параметрах генеральной совокупности



$$Model(\mu, \sigma^2, \alpha, \beta, \gamma, \theta, \dots)$$



$$\hat{\mu}, \widehat{\sigma^2}, \hat{\alpha}, \hat{\beta}, \hat{\gamma}, \hat{\theta}, \dots)$$

Наши оценки параметров генеральной совокупности

Гипотезы

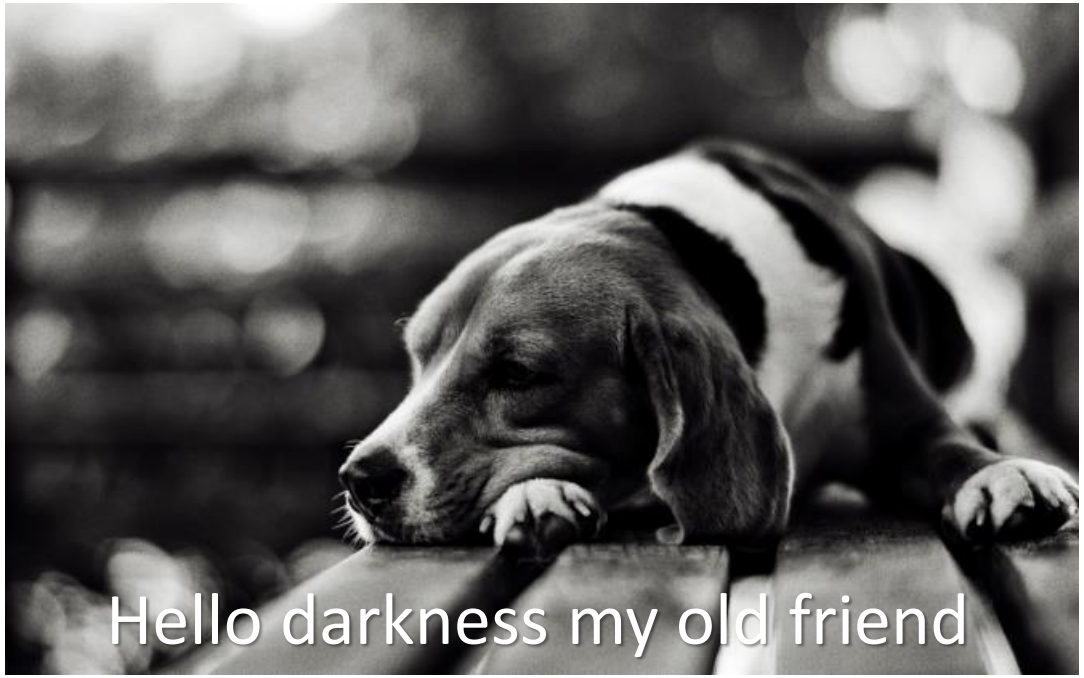
$\mu = 2?$

$\sigma^2 = 1?$

…

Гипотезы: H_0

Нулевая гипотеза - это проверяемое предположение, которое обычно заявляет, что наш мир сер, скучен, в нем нет места интересным эффектам, и чего-то значимого в результатах нашего эксперимента – нет. Обычно формулируется как отсутствие различий, отсутствие влияние фактора, отсутствие эффекта, равенство нулю значений параметров и т.п. Нулевая гипотеза – это гипотеза, с которой, как правило, мы **хотим** не согласиться.



Гипотезы: H_0, примеры

Есть две группы людей, одни в числе близких друзей имеют людей с повышенным весом, вторые нет. Известны веса людей в первой и второй группах;

Примеры гипотез:

H_0: в среднем веса людей с друзьями с повышенным весом и без таких друзей – не отличаются

H_0: дисперсия распределения веса людей с друзьями с повышенным весом и без таких друзей – не отличаются

H_0: в среднем веса людей с друзьями с повышенным весом и без таких друзей – отличаются не больше, чем на 2 кг

Гипотезы: $\mathbf{H_0}$, примеры

Есть две группы программистов - пишущие только на Python, и пишущие только на C/C++ . Известна статистика самоубийств среди первых и вторых в течении года;

Гипотезы: $\mathbf{H_0}$, примеры

Есть две группы программистов - пишущие только на Python, и пишущие только на C/C++ . Известна статистика самоубийств среди первых и вторых в течении года;

Примеры гипотез:

H_0: в среднем число самоубийств среди программистов, пишущих на $C++$ и на Python $-$ не отличается

H_0: в среднем число самоубийств среди программистов, пишущих на $C++$, не меньше, чем у пишущих на Python

Гипотезы: $\mathbf{H_1}$

Альтернативная гипотеза H_1 - это гипотеза, которая, как правило, является отрицанием нулевой гипотезы и состоит в том, что эффект есть, результаты наших экспериментов значимы и т.д.

Гипотезы: H_1, примеры

Есть две группы людей, одни в числе близких друзей имеют людей с повышенным весом, вторые нет. Известны веса людей в первой и второй группах;

Примеры гипотез:

H_0: в среднем веса людей с друзьями с повышенным весом и без таких друзей не отличаются

H_1: в среднем веса людей с друзьями с повышенным весом и без таких друзей отличаются

H_0: в среднем веса людей с друзьями с повышенным весом и без таких друзей – отличаются не больше, чем на 2 кг

H_1: в среднем веса людей с друзьями с повышенным весом и без таких друзей – отличаются больше, чем на 2 кг

Гипотезы: H_1, примеры

Есть две группы программистов - пишущие только на Python, и пишущие только на C/C++ . Известна статистика самоубийств среди первых и вторых в течении года;

Примеры гипотез:

H_0: в среднем число самоубийств среди программистов, пишущих на C + + и на Python − не отличается

H_0: в среднем число самоубийств среди программистов, пишущих на C + +, не меньше, чем у пишущих на Python

Гипотезы: H_1, примеры

Есть две группы программистов - пишущие только на Python, и пишущие только на C/C++ . Известна статистика самоубийств среди первых и вторых в течении года;

Примеры гипотез:

H_0: в среднем число самоубийств среди программистов, пишущих на C + + и на Python не отличается

H_1: в среднем число самоубийств среди программистов, пишущих на C + + и на Python отличается

H_0: в среднем число самоубийств среди программистов, пишущих на C + +, не меньше, чем у пишущих на Python

H_1: в среднем число самоубийств среди программистов, пишущих на C + +, больше, чем у пишущих на Python

Гипотезы – о параметрах

H_0: выборочные средние двух групп не отличаются

Гипотезы – о параметрах

H_0: выборочные средние двух групп не отличаются

Такая формулировка гипотезы неверна – тут нечего проверять.
Посчитали одно выборочное среднее, посчитали второе - сравнили

Польза витамина С

Представим, что мы хотим проверить, насколько хорошо витамин С помогает в лечении простуды. Для этого мы делим пациентов на пары (на основе пола, возраста, здоровья и т.д.). Далее считаем сколько, в скольких парах люди, принимавшие витамин С, выздоровели от простуды раньше.

$$H_0: P(\text{витамин С помогает}) = \frac{1}{2}$$

$$H_1: P(\text{витамин } C \text{ не помогает}) = \frac{1}{2}$$

Польза витамина С

Это биномиальное распределение. Число испытаний – число пар пациентов. Число успехов – число пар, в которых пациент, принимавший витамин С выздоровел раньше.

Можем посчитать вероятность* нашего наблюдения при условии, что $H_0: p = 0.5$ верна.

$$C_{17}^{13} p^{13} (1 - p)^4 = 0.018$$

* не совсем – мы уже получили результаты эксперимента, их вероятность – 1. Мы считаем вероятность наблюдать подобные результаты в принципе

Польза витамина С

Это биномиальное распределение. Число испытаний – число пар пациентов. Число успехов – число пар, в которых пациент, принимавший витамин С выздоровел раньше.

Можем посчитать вероятность* нашего наблюдения при условии, что $H_0: p = 0.5$ верна.

$$C_{17}^{13} p^{13} (1 - p)^4 = 0.018$$

Однако сама по себе нам эта вероятность ни о чем не говорит. Чем больше пар наберем, тем меньше будет вероятность любого конкретного наблюдения

Польза витамина С

Это биномиальное распределение. Число испытаний – число пар пациентов. Число успехов – число пар, в которых пациент, принимавший витамин С выздоровел раньше.

Можем посчитать вероятность* нашего наблюдения при условии, что $H_0: p = 0.5$ верна.

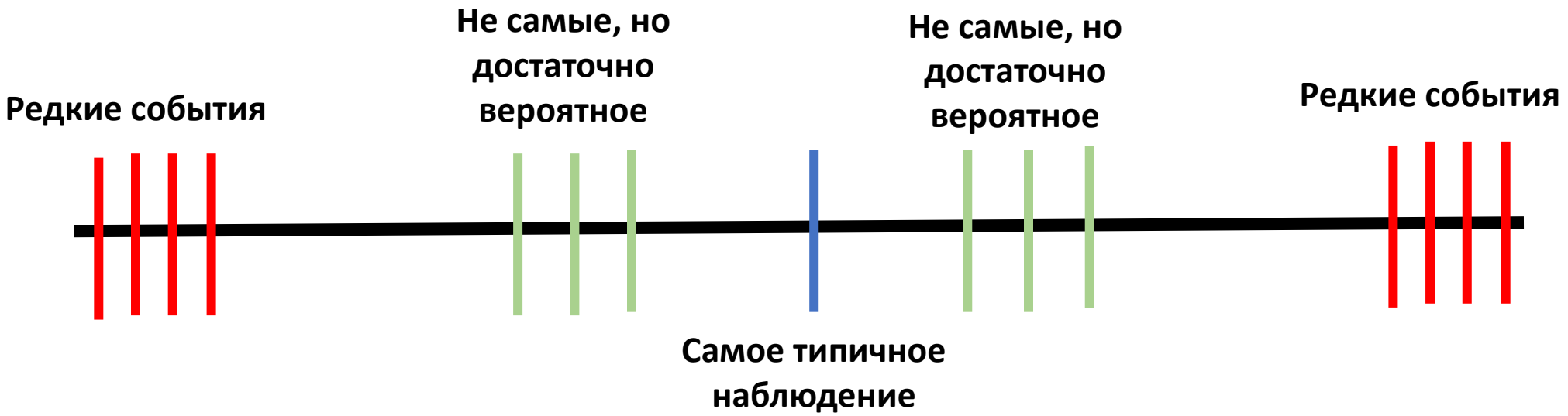
$$C_{17}^{13} p^{13} (1 - p)^4 = 0.018$$

Однако сама по себе нам эта вероятность ни о чем не говорит. Чем больше пар наберем, тем меньше будет вероятность любого конкретного наблюдения

Отчасти здесь и происходит разделение – что делать дальше. Один подход применяет фреквентисты, другой – байесиане. Мы рассмотрим первый

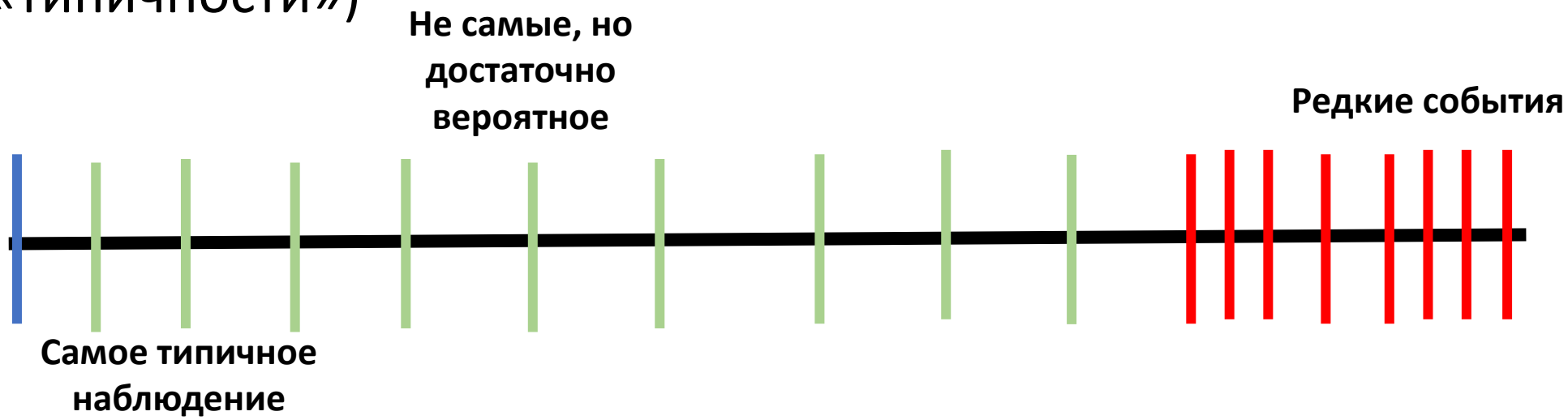
P-value

Посмотрим на то, а какие вообще могут быть результаты и то, насколько они вероятны



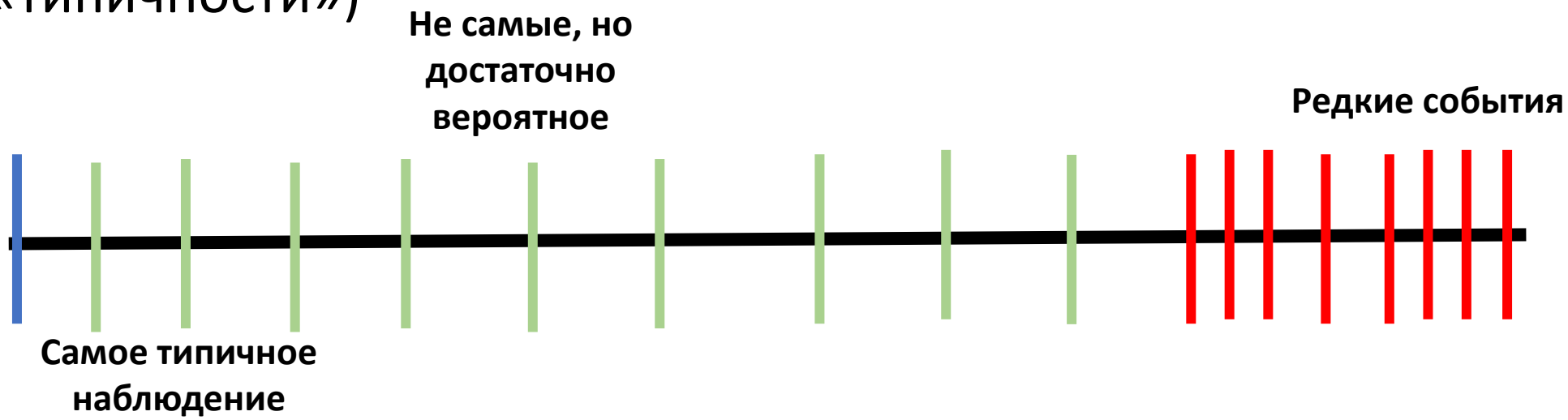
P-value

Посмотрим на то, а какие вообще могут быть результаты и то, насколько они вероятны (ну или можно отсортировать их по «типичности»)



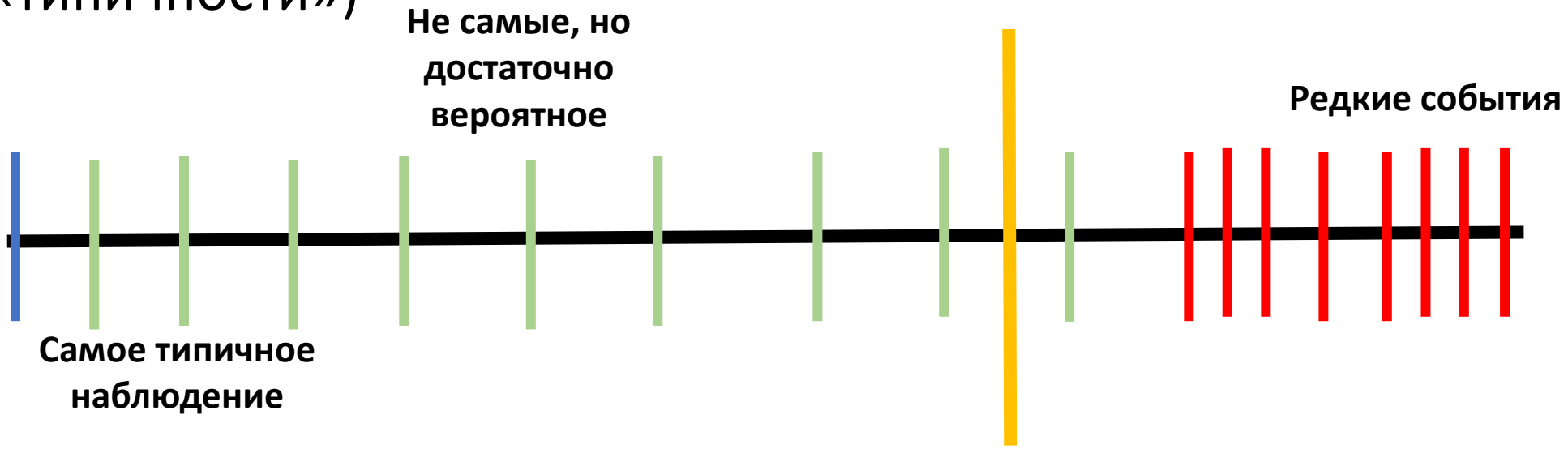
P-value

Посмотрим на то, а какие вообще могут быть результаты и то, насколько они вероятны (ну или можно отсортировать их по «типичности»)



P-value

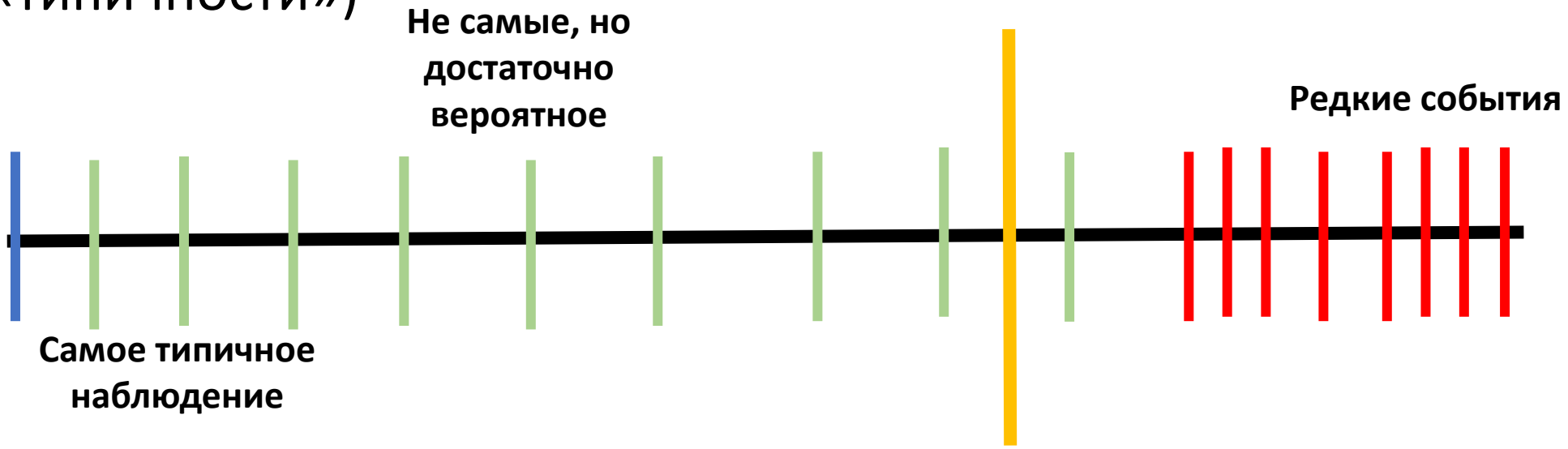
Посмотрим на то, а какие вообще могут быть результаты и то, насколько они вероятны (ну или можно отсортировать их по «типичности»)



Допустим, наше находится где-нить здесь

P-value

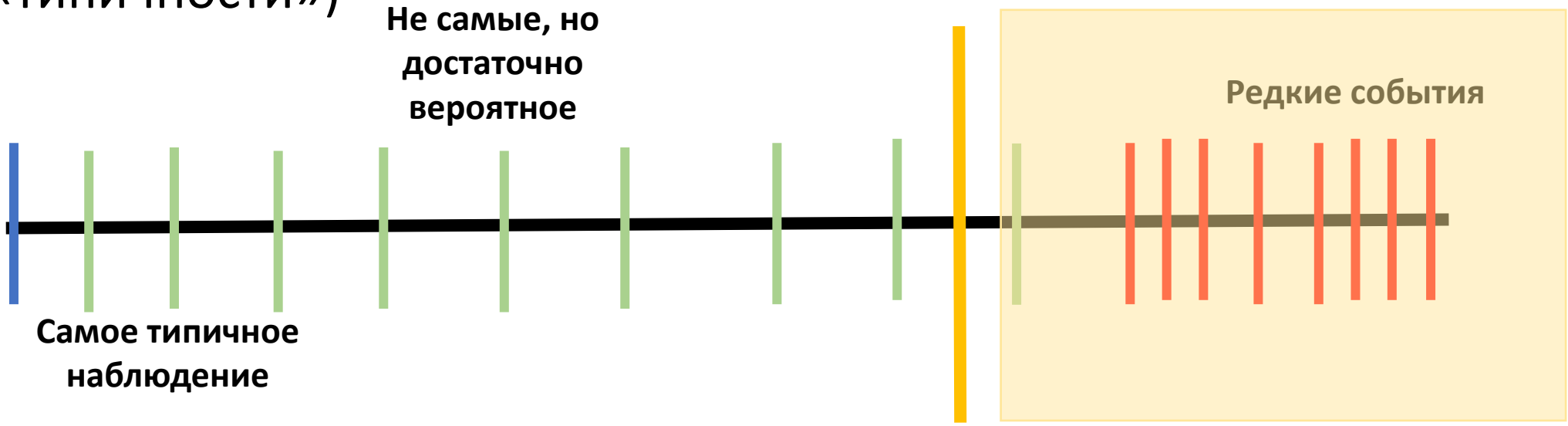
Посмотрим на то, а какие вообще могут быть результаты и то, насколько они вероятны (ну или можно отсортировать их по «типичности»)



Допустим, наше находится где-нить здесь. А давайте учитывать не только его «вес», но и вес всех событий, которые еще более редкие

P-value

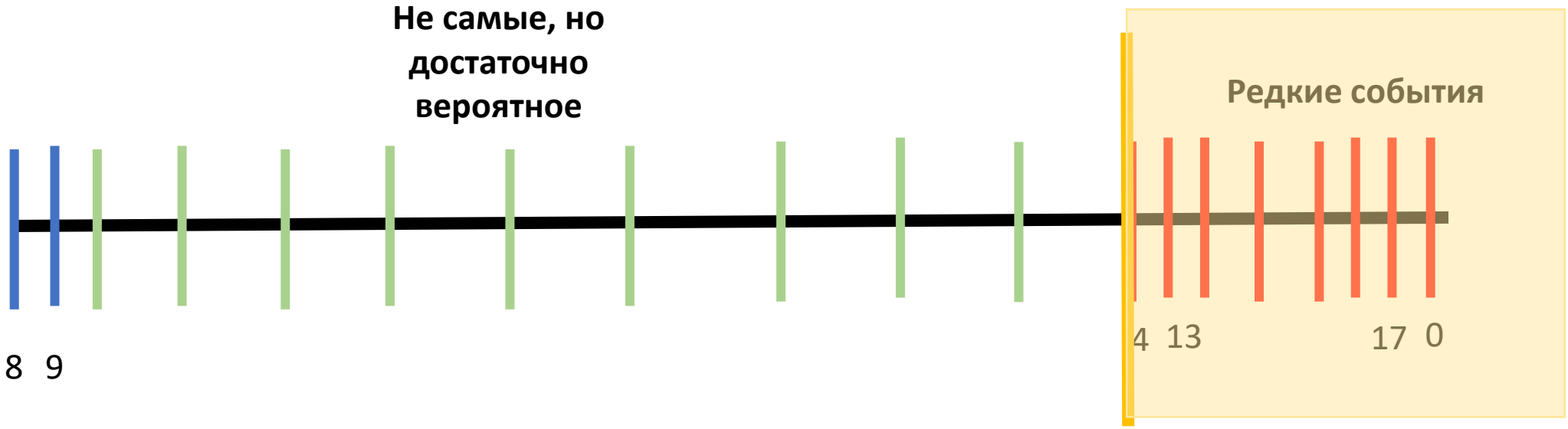
Посмотрим на то, а какие вообще могут быть результаты и то, насколько они вероятны (ну или можно отсортировать их по «типичности»)



Допустим, наше находится где-нить здесь. А давайте учитывать не только его «вес», но и вес всех событий, которые еще более редкие

P-value

Для нашей задачи

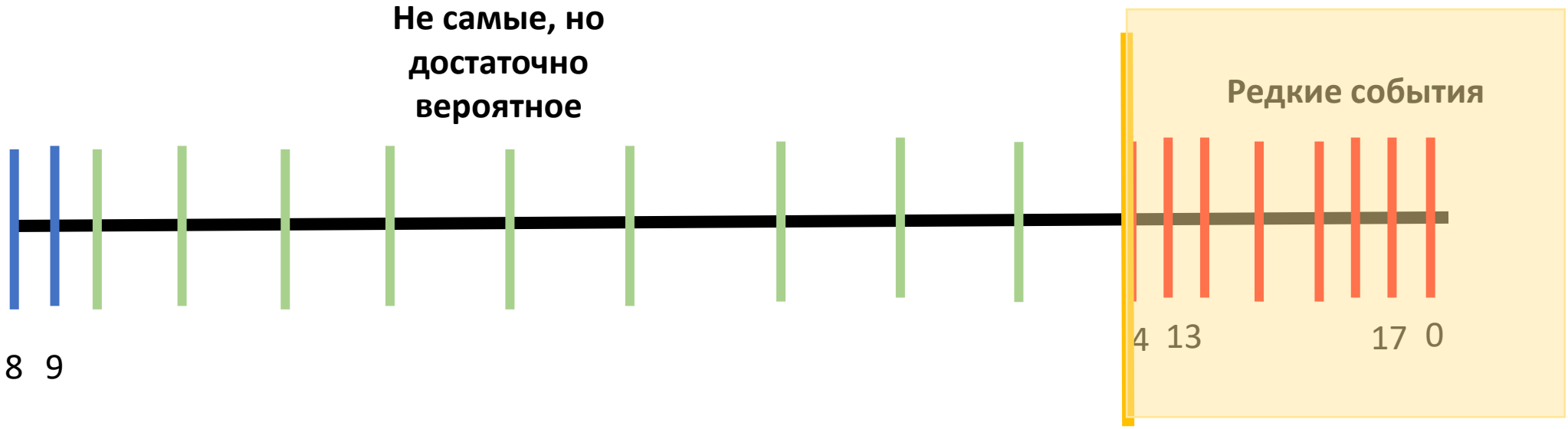


Самое типичное наблюдение

$$\sum_{i=13}^{17} p(i) + \sum_{i=0}^{4} p(i) = 0.049$$

P-value

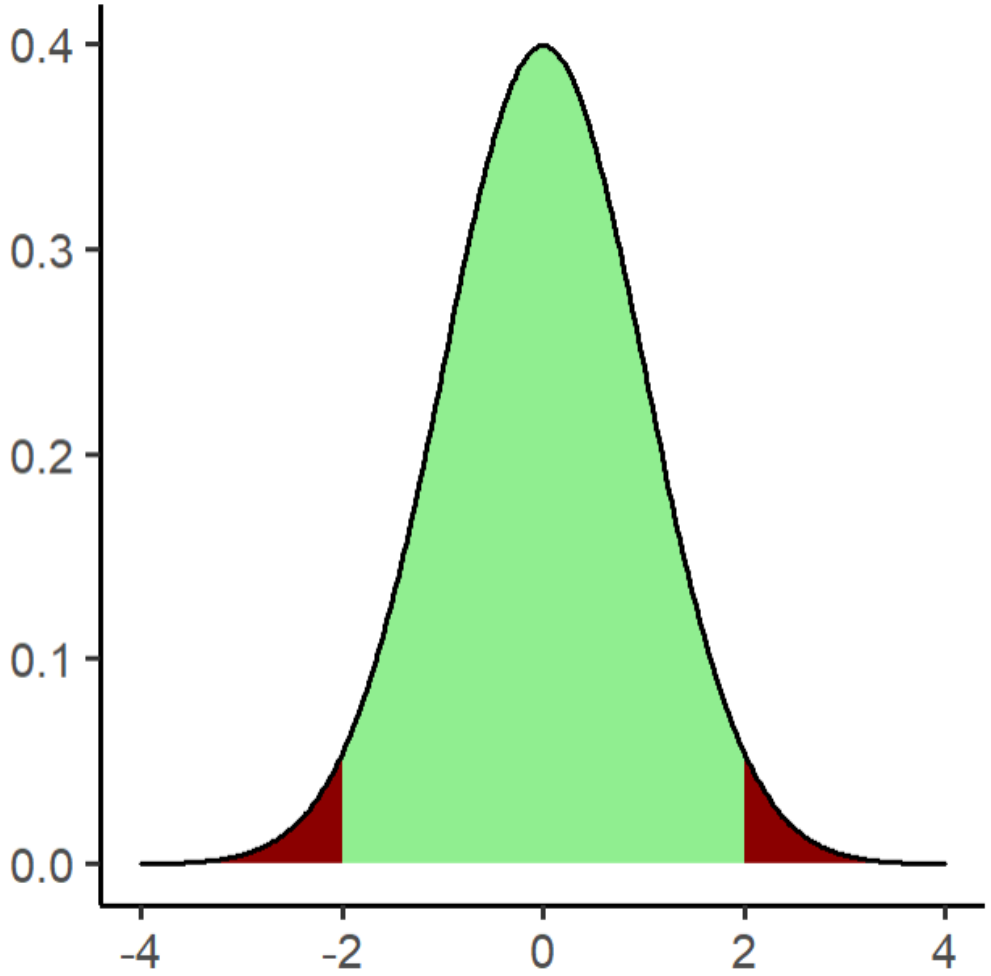
Для нашей задачи



$$\sum_{i=13}^{17} p(i) + \sum_{i=0}^{4} p(i) = 0.049 = p - value$$

P-value

$p\text{-}value$ — вероятность наблюдать результат такой же или более критичный результат при условии верности H_0

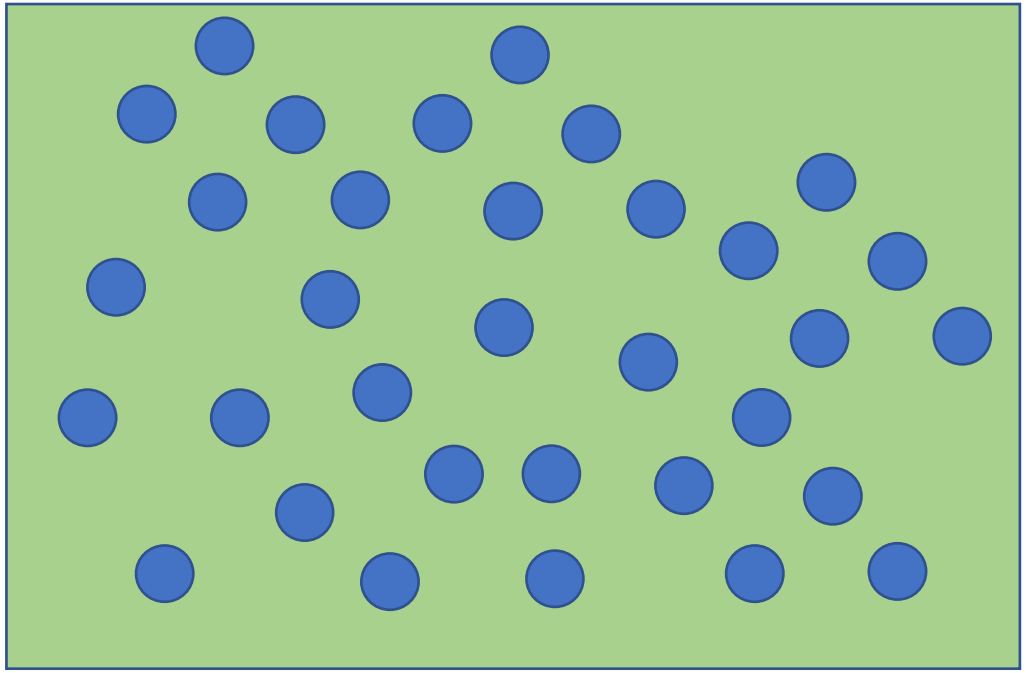


P-value: немного другой взгляд

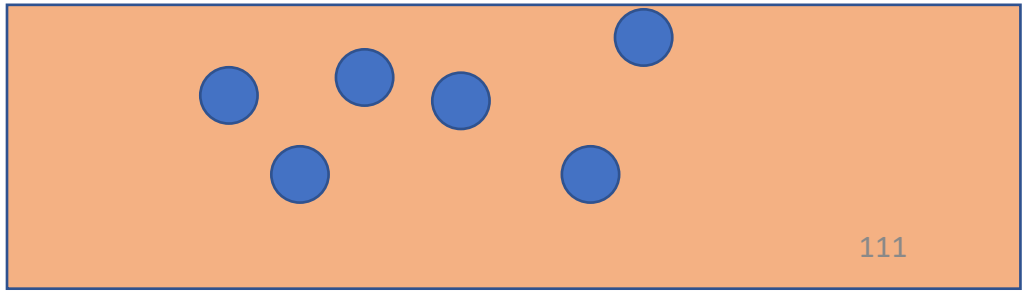
Мы проверили эксперимент и получили что оценка, величина Y, которую мы вычисляем, равна a.

Теперь предположим, что мы проведем еще 9999+ (в идеале – бесконечно много) таких же в точности экспериментов и что H_0 – верна.

$Y < a$



$Y \geq a$

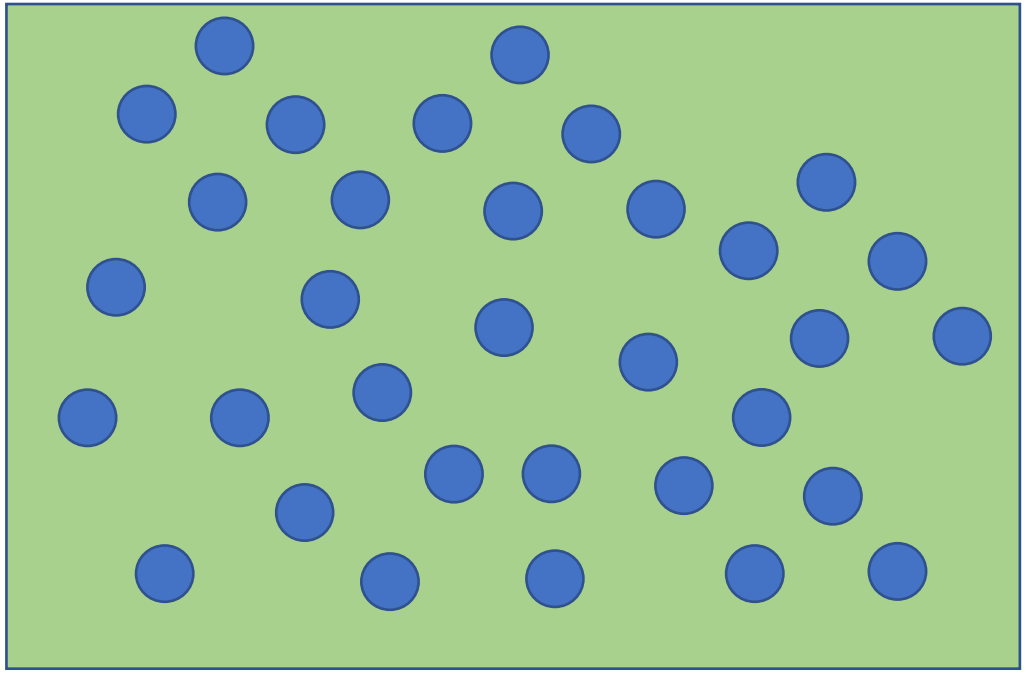


P-value: немного другой взгляд

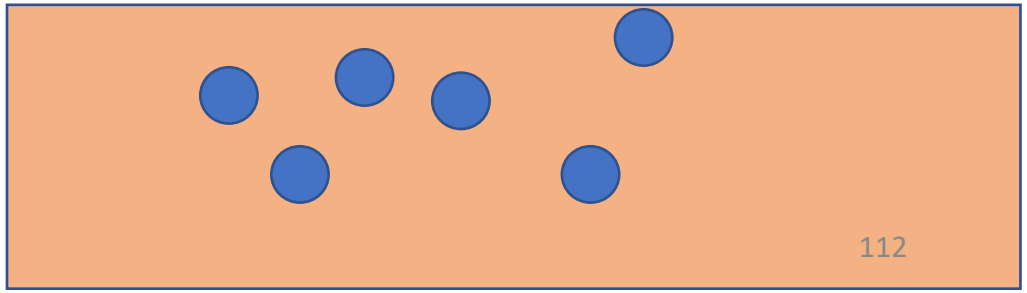
Мы проверили эксперимент и получили что оценка, величина Y, которую мы вычисляем, равна a.

Доля шариков в оранжевой коробочке – и есть p-value

$Y < a$

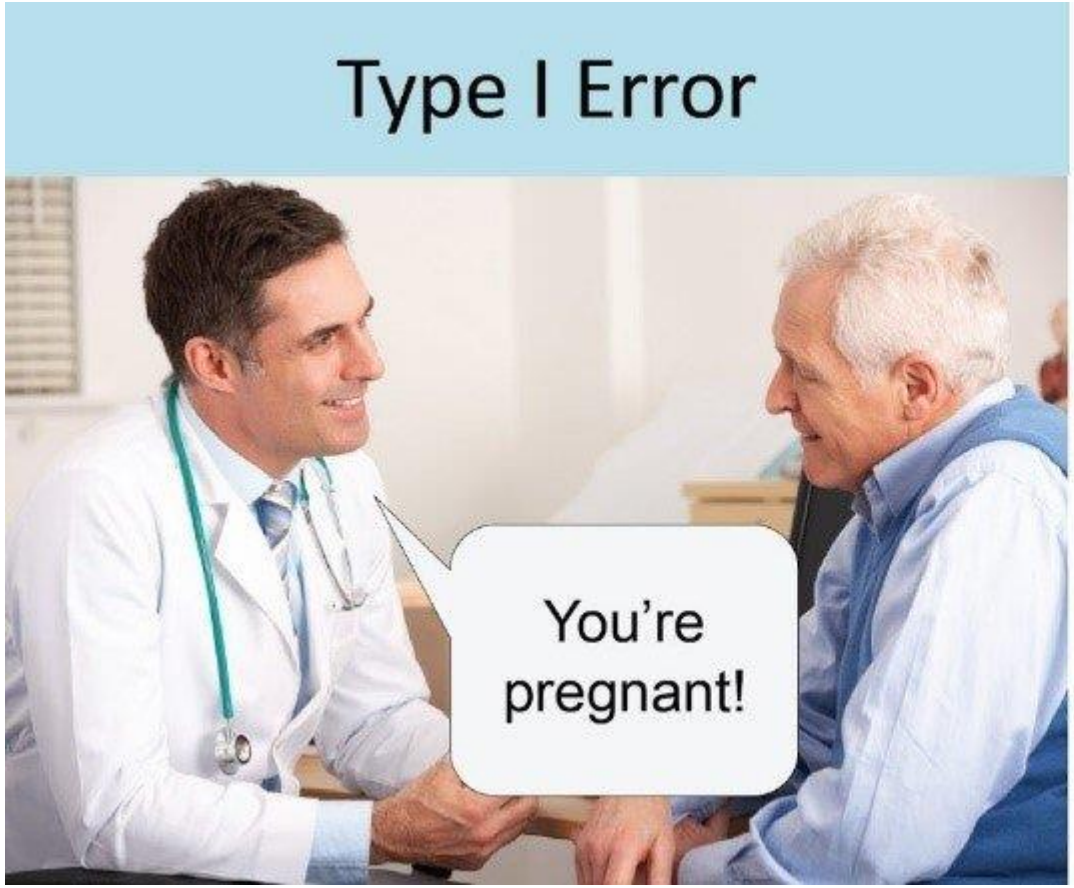


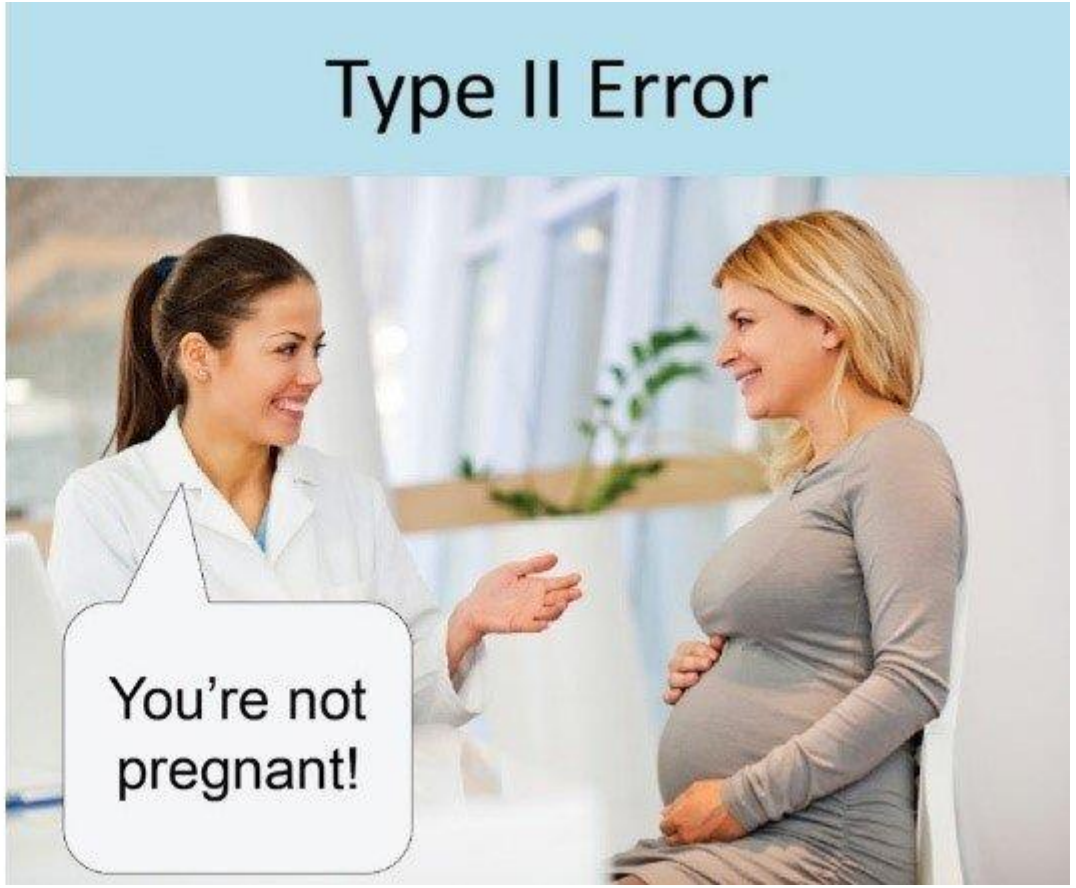
$Y \geq a$



Ошибка первого и второго рода

Теперь у нас есть «вес» нашего наблюдения. Осталось понять, какой порог на этот вес ввести





Ошибка первого и второго рода

Теперь у нас есть «вес» нашего наблюдения. Осталось понять, какой порог на этот вес ввести

H0	**верна**	**не верна**
Отклоняется	Ошибка первого рода	Решение верное
Не отклоняется	Решение верное	Ошибка второго рода

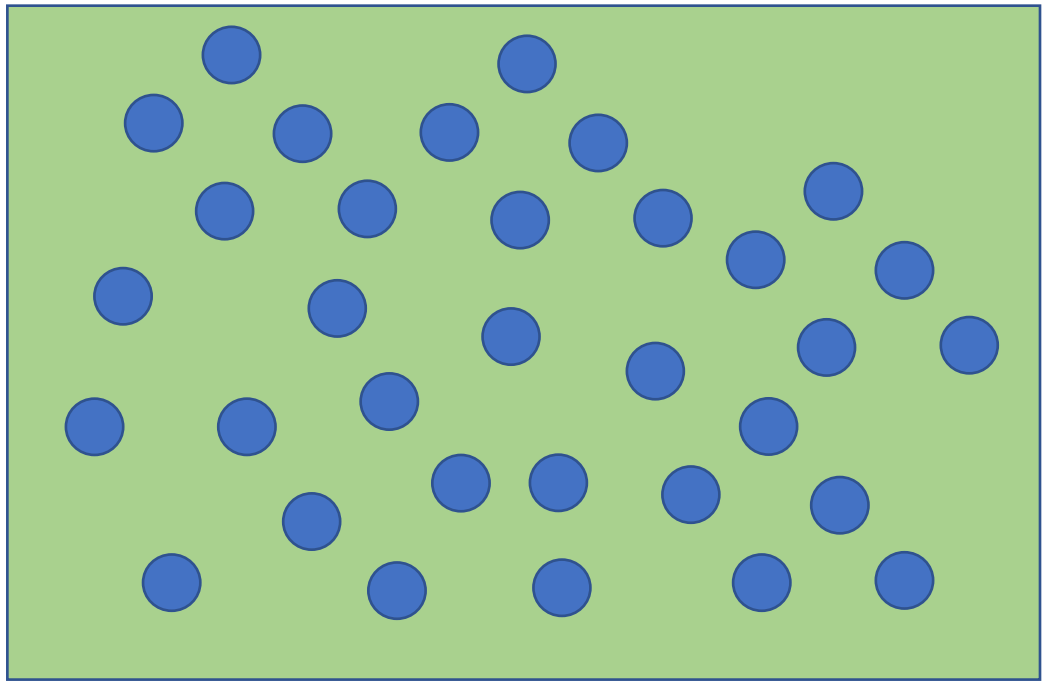
Вероятность ошибки первого рода

Пусть мы ввели некую процедуру T, которая на основе выборки отвергает или не отвергает H_0.

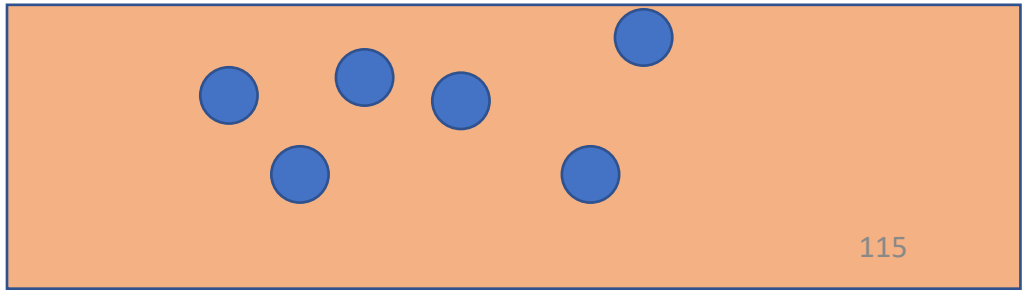
Допустим, мы провели бесчисленное множество экспериментов, для которых H_0 - верна, и оказалось, что наша процедура ошибочно отвергает H_0 в доле случаев α.

Тогда мы говорим, что вероятность ошибки первого рода равна α

$\boldsymbol{H_0}$ не отвергнута ошибочно



$\boldsymbol{H_0}$ ошибочно отвергнута



Вероятность ошибки первого рода

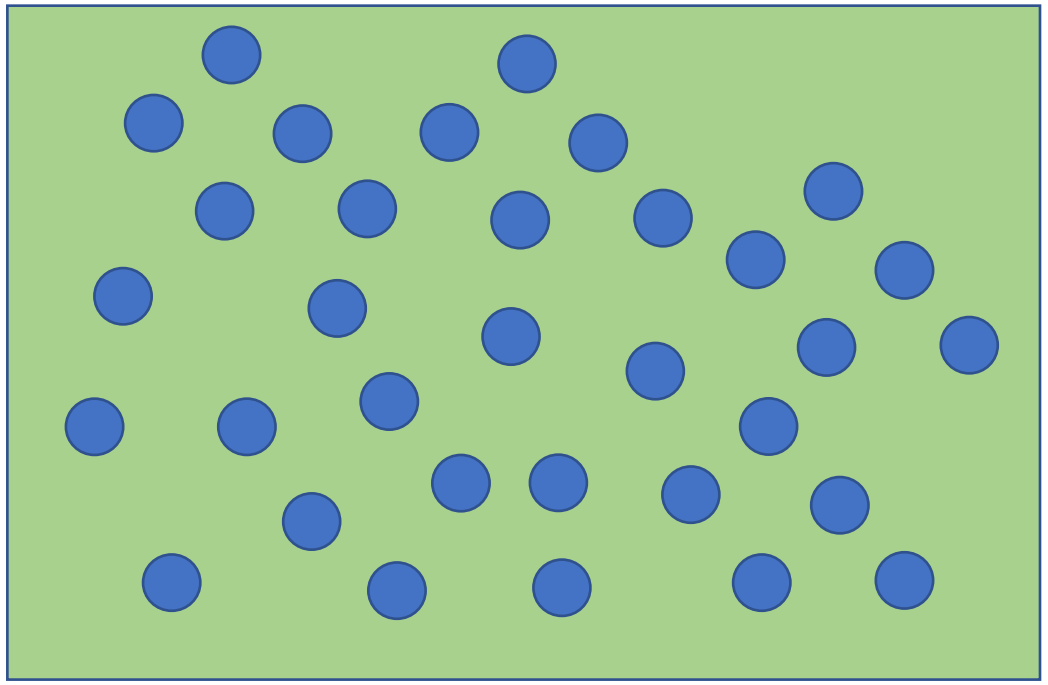
Пусть мы ввели некую процедуру T, которая на основе выборки отвергает или не отвергает H_0.

Допустим, мы провели бесчисленное множество экспериментов, для которых H_0 - верна, и оказалось, что наша процедура ошибочно отвергает H_0 в доле случаев α.

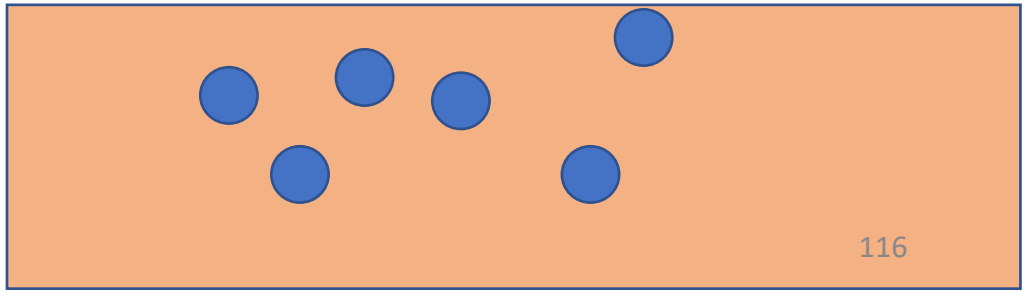
Тогда мы говорим, что вероятность ошибки первого рода равна α

Что-то напоминает?

$\boldsymbol{H_0}$ не отвергнута ошибочно



$\boldsymbol{H_0}$ ошибочно отвергнута



P-value и вероятность ошибки первого рода

Процедура T, которая состоит в том, что мы считаем $p\text{-}value$ нашего наблюдения и если

$$p\text{-}value \leq \alpha$$

отвергаем гипотезу H_0, будет поддерживать ошибку первого рода на уровне α (уровень значимости α)

P-value и вероятность ошибки первого рода

Процедура T, которая состоит в том, что мы считаем $p\text{-}value$ нашего наблюдения и если

$$p\text{-}value \leq \alpha$$

отвергаем гипотезу H_0, будет поддерживать ошибку первого рода на уровне α

Статистический фреймворк

На основе своих знаний о предмете изучение, сформулируй нулевую и альтернативную гипотезу и выбери уровень значимости α

Проведи эксперимент и получи наблюдения

Проведи эксперимент и получи наблюдения

Получи значение $p\text{-}value$ и сравни его с α

$p\text{-}value$ vs α

$p\text{-}value$ – это значение, вычисляемое **на основе выборки**. По сути – это тоже оценка. Еще можно называеть его *выборочной статистикой*

Уровень значимости α – это то, какую долю ложноположительных результатов мы согласны терпеть. Он фиксируется **до** эксперимента

Одновыборочный $Z\text{-}test$

Пусть у нас есть выборка размера $n \geq 40$. Мы хотим проверить гипотезу о том, что среднее выборки равно a.

$$\overline{X} \sim N(\mu, \frac{\sigma^2}{n})$$