Математическое ожидание абсолютно непрерывных случайных величин

Нестрого говоря, математическое ожидание это среднее значение случайной величины.

Мы с вами уже научились вычислять математическое ожидание дискретных случайных величин — то есть случайных величин, принимающих не более чем счётное число различных значений. Напомним, как это делается. Пусть ξ — дискретная случайная величина. Обозначим её множество значений за X. Тогда

$$E[\xi] \ = \sum_{x \in X} x P(\xi = x)$$

Мы можем позволить себе такую форму записи, потому что для конечных множеств сумма не зависит от последовательности слагаемых. А в случае счётного числа значений математическое ожидание определено только в случае, если такой ряд сходится абсолютно — то есть результат не зависит от порядка слагаемых.

В случае с непрерывной случайной величиной мы уже не можем так писать, потому что сумма более чем счётного числа слагаемых не определена. Но, как и в случае с плотностью распределения, нам поможет непрерывный аналог суммы — интеграл.

В этом уроке мы

- Поймём интуицию, стоящую за определением математического ожидания для абсолютно непрерывных случайных величин.
- Определим математическое ожидание. А также разберёмся, когда оно существует.
- Найдём математическое ожидания для
    - равномерного распределения на отрезке,
    - экспоненциального распределения,
    - распределения Парето (которое мы определим в этом уроке).

Плотность: напоминание

Начнём с небольшого напоминания. Абсолютно непрерывные случайные величины — это случайные величины, у которых есть плотность. То есть такая функция $p_\xi : \mathbb{R} \to \mathbb{R}$, что для всех $a \in \mathbb{R}$ выполнено

$$F_\xi(a) = \int\limits_{-\infty}^{a} p_\xi(x)dx.$$

Отсюда следует, что для всех $a \in \mathbb{R}$, в которых $F'_\xi(a)$ определена, выполнено $F'_\xi(a) = p_\xi(a)$.

А так же, что для любого отрезка $[a, b]$ выполнено $P(\xi \in [a, b]) \ = \int\limits_{a}^{b} p_\xi(x)dx$.

Математическое ожидание – идея

Напомним, в случае случайной величины с конечным числом значений математическое ожидание можно вычислять так: $E[\xi] = \sum x_i\, P(\xi = x_i)$, где x_i пробегает все возможные значения ξ. То есть у нас есть конечное количество возможных значений, и мы суммируем слагаемые вида "значение, умноженное на вероятность получить это значение".

Теперь пусть у нас есть абсолютно непрерывная случайная величина ξ, которая принимает значения от 0 до 10. Давайте подумаем, как построить её математическое ожидание. Мы хотим сымитировать такую же конструкцию, как и в конечном случае.

Сначала притворимся, что ξ принимает конечное число значений. Делать мы будем это так: разобьём отрезок $[0, 10]$ на 10 равных отрезков. Каждый из этих отрезков как бы имитирует одно из значений величины ξ. Посмотрим на то, какой вклад в матожидание даёт каждый из отрезков.

- Вот отрезок $[0, 1]$. Какова вероятность того, что ξ попадет на этот отрезок? Она равна $\int\limits_0^1 p_\xi(x)dx$. А какое значения ξ принимает на этом отрезке? Очевидно, значения от 0 до 1 (это тавтология). Поэтому мы приблизим эти значения числом 1. Тогда этот отрезок даст приблизительно такой вклад в математическое ожидание: $1 \cdot \int\limits_0^1 p_\xi(x)dx = \int\limits_0^1 1 \cdot p_\xi(x)dx$
- Вот отрезок $[1, 2]$. Какова вероятность того, что ξ попадет на этот отрезок? Она равна $\int\limits_1^2 p_\xi(x)dx$. А какое значения ξ принимает на этом отрезке? Очевидно, значения от 1 до 2. Поэтому мы приблизим эти значения числом 2. Тогда этот отрезок даст приблизительно такой вклад в математическое ожидание: $2 \cdot \int\limits_1^2 p_\xi(x)dx = \int\limits_1^2 2 \cdot p_\xi(x)dx$
- $\ldots$
- Вот отрезок $[9, 10]$. Какова вероятность того, что ξ попадет на этот отрезок? Она равна $\int\limits_9^{10} p_\xi(x)dx$. А какое значения ξ принимает на этом отрезке? Очевидно, значения от 9 до 10. Поэтому мы приблизим эти значения числом 10. Тогда этот отрезок даст приблизительно такой вклад в математическое ожидание: $10 \cdot \int\limits_9^{10} p_\xi(x)dx = \int\limits_9^{10} 10 \cdot p_\xi(x)dx$

Итого мы оценили матожидание ξ как

$$\int\limits_0^1 1 \cdot p_\xi(x)dx + \int\limits_1^2 2\ \cdot p_\xi(x)dx + \cdots + \int\limits_9^{10} 10 \cdot p_\xi(x)dx$$

Ясно, что сумма этих десяти слагаемых приблизительно равна $\int\limits_0^{10} x \cdot p_\xi(x)\ dx$. Теперь давайте брать разбиение не на 10 отрезков, а на 20, на 30 и так далее (это соответствует стремлению ранга разбиения к нулю). Для каждого из этих разбиений мы будем вычислять аналогичные суммы. Ясно, что они будут всё ближе и ближе к $\int\limits_0^{10} x \cdot\ p_\xi(x)dx$. Поэтому естественно считать, что по определению $E[\xi] := \int\limits_0^{10}\ x \cdot p_\xi(x)dx$.

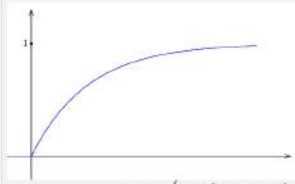
Числа 0 и 10 мы поставили, только чтобы пример был более наглядным. Ясно, что в общем случае значения не обязаны лежать в отрезке от 0 до 10. Значения могут лежать на прямой от $-\infty$ до $+\infty$. То есть естественно считать $E[\xi] := \int\limits_{-\infty}^{+\infty} x \cdot\ p_\xi(x)dx$.

Для наглядности на следующем шаге мы приведём похожее рассуждение, но с гистограммами.

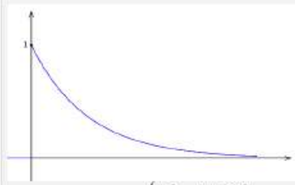
Комментарий. Заметьте, что нам не нужно доказывать, что $E[\xi] = \int\limits_{-\infty}^{+\infty} x \cdot\ p_\xi(x)$. Мы скажем, что мы определяем $E[\xi]$ как $\int\limits_{-\infty}^{+\infty} x \cdot\ p_\xi(x)dx$. На этом шаге мы только давали мотивацию для этого определения.

Псевдо-мат.ожидание

Аналогично тому, как мы строили псевдо-гистограмму, найдём приближенное значение мат.ожидания – псевдо-мат.ожидание. В качестве примера рассмотрим хорошо знакомое нам экспоненциальное распределение с параметром $\lambda = 1$. Вот его функция распределения и плотность:



Функция распределения $F_\xi(x) = \begin{cases} 1 - e^{-x}, & \text{если } x > 0 \\ 0, & \text{иначе} \end{cases}$

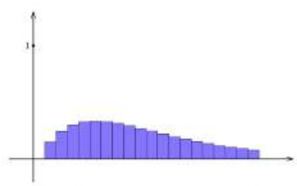


Плотность $p_\xi(x) = \begin{cases} e^{-x}, & \text{если } x > 0 \\ 0, & \text{иначе} \end{cases}$

Разобьём числовую ось на интервалы шириной 0.1. Над каждым интервалом (a, b) нарисуем столбик такой высоты, чтобы площадь столбика равнялась $a \cdot P(a < \xi < b)$. То есть высота столбика должна равняться

$$a \cdot \frac{F_\xi(b) - F_\xi(a)}{b - a} = a \cdot \frac{\int_a^b p_\xi(x)dx}{b - a}.$$

Получим вот такую картинку:

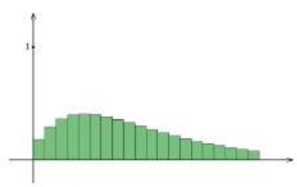


Заметим, что суммарная площадь столбиков приближенно соответствует тому, что мы бы хотели считать математическим ожиданием. А если точнее, то это будет оценка на мат.ожидание снизу, так как мы заменили все значения случайной величины ξ из интервала (a, b) на a.

Аналогично мы можем построить оценку на мат.ожидание сверху. Для этого нужно нарисовать над каждым интервалом столбик такой высоты, чтобы площадь столбика равнялась $b \cdot P(a < \xi < b)$. То есть высота столбика должна равняться

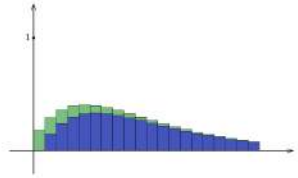
$$b \cdot \frac{F_\xi(b) - F_\xi(a)}{b - a} = b \cdot \frac{\int_a^b p_\xi(x)dx}{b - a}.$$

Получим вот такую картинку:

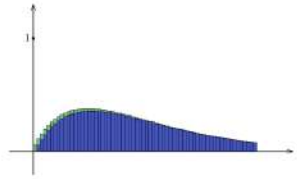


Суммарная площадь таких столбиков будет нашей оценкой сверху на мат.ожидание.

Если нарисовать эти столбики на одной картинке, то можно увидеть разницу между полученными значениями (это суммарная площадь зелёного, не закрашенного синим):

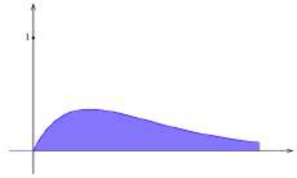


Если уменьшать ширину интервала в разбиении, то зазор между оценкой снизу и сверху будет сокращаться, а верхушки столбиков будут всё больше похожи на непрерывную линию:



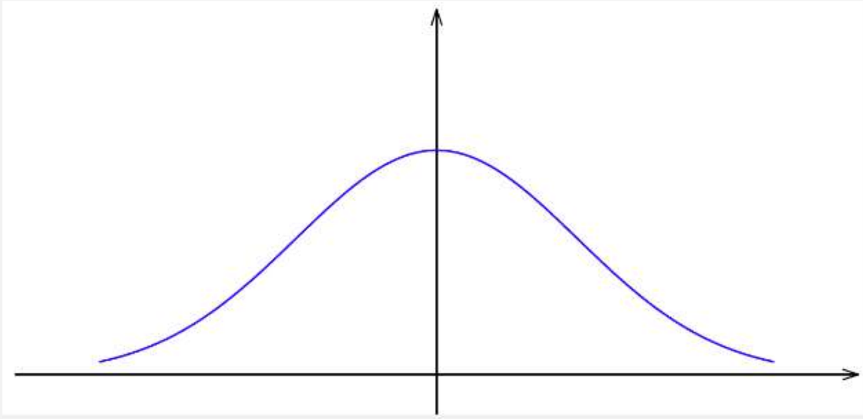
Предел псевдо-мат.ожидания

Если устремить ширину интервалов в разбиении $\mathbb{R}$ к нулю, то столбики действительно превратятся в непрерывную линию. Её формула – произведение $xp_\xi(x)$. А верхняя и нижняя оценки на мат.ожидание сравняются и превратятся в площадь под графиком $xp_\xi(x)$, то есть $\int\limits_{-\infty}^{+\infty} xp_\xi(x)dx$.

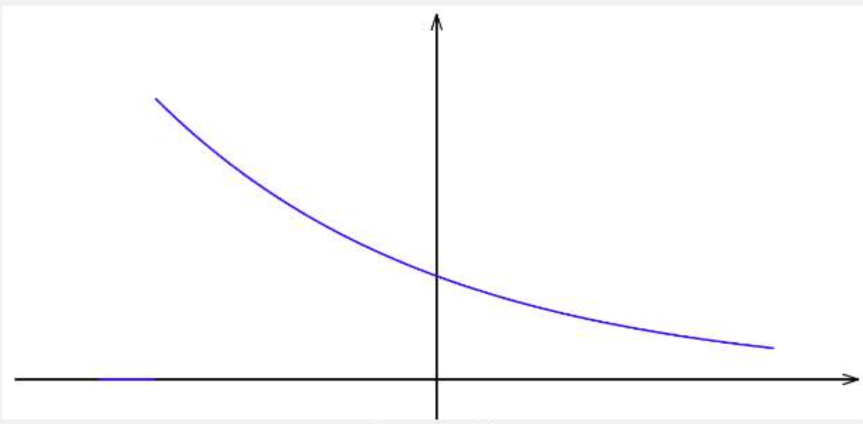


Прежде, чем мы перейдём к формальному определению, решим задачу на интуитивное понимание мат.ожидания.

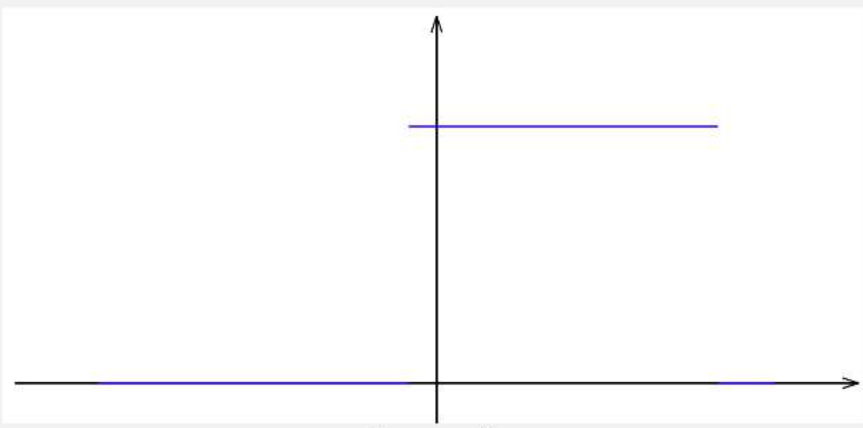
Сопоставьте графики плотности случайной величины с её мат.ожиданием.



Плотность 1. (Забегая вперед, это — нормальное распределение.)



Плотность 2.



Плотность 3.

Сопоставьте значения из двух списков

Плотность 1	Математическое ожидание равно 0
Плотность 2	Математическое ожидание отрицательно
Плотность 3	Математическое ожидание положительно

Математическое ожидание абсолютно непрерывной случайной величины

Смысл математического ожидания абсолютно непрерывной случайной величины мы уже разобрали. Чтобы дать полностью формальное определение, нужно лишь оговориться, что нужный нам интеграл существует.

Определение. Пусть ξ — абсолютно непрерывная случайная величина с плотностью $p(\xi)$. Если интеграл $\int\limits_{-\infty}^{+\infty} x p_\xi(x) dx$ сходится, то его значение называют *математическим ожиданием* ξ. В противном случае говорят, что математическое ожидание не определено.

Как вы уже знаете из этого шага, интеграл $\int\limits_{-\infty}^{+\infty} f(x) dx$ по всей числовой прямой (то есть от $-\infty$ до $+\infty$) существует только в случае, если сходятся интегралы с каждого из концов:

$$\int\limits_{-\infty}^{0} f(x) dx \quad \text{и} \quad \int\limits_{0}^{+\infty} f(x) dx.$$

То есть существуют конечные пределы:

$$\lim_{a \to -\infty} \int\limits_{a}^{0} f(x)\, dx \quad \text{и} \quad \lim_{b \to +\infty} \int\limits_{0}^{b} f(x)\, dx.$$

Примеры

На следующих шагах мы найдём математическое ожидания для

- равномерного распределения на отрезке,
- экспоненциального распределения,
- распределения Парето (которое мы определим).

Пример. Пусть ξ равномерно распределена на отрезке $[3, 7]$. То есть $p_\xi(x) = \frac{1}{4}$ если $x \in [3, 7]$, и $p_\xi(x) = 0$ если $x \notin [3, 7]$. Найдём $E[\xi]$.

Интуитивно, ответ очевиден – это число 5, являющееся серединой отрезка $[3, 7]$. Но давайте найдём его строго, чтобы потренироваться применять формулу для математического ожидания.

Решение. По определению,

$$E[\xi] = \int\limits_{-\infty}^{+\infty} x p_\xi(x) dx = \int\limits_{3}^{7} x p_\xi(x) dx,$$

здесь мы воспользовались тем, что вне отрезка $[3, 7]$ функция p_ξ обращается в ноль. Так как при $x \in [3, 7]$ имеем $p_\xi(x) = \frac{1}{4}$, выполнено такое равенство:

$$\int\limits_{3}^{7} x p_\xi(x) dx = \int\limits_{3}^{7} x \cdot \frac{1}{4}\, dx$$

Первообразная функции $x \cdot \frac{1}{4}$ это $(\frac{x^2}{2} \cdot \frac{1}{4})$, поэтому

$$\int\limits_{3}^{7} x \cdot \frac{1}{4}\, dx = \left(\frac{7^2}{2} \cdot \frac{1}{4}\right) - \left(\frac{3^2}{2} \cdot \frac{1}{4}\right) = \frac{49 - 9}{2 \cdot 4} = \frac{40}{8} = 5.$$

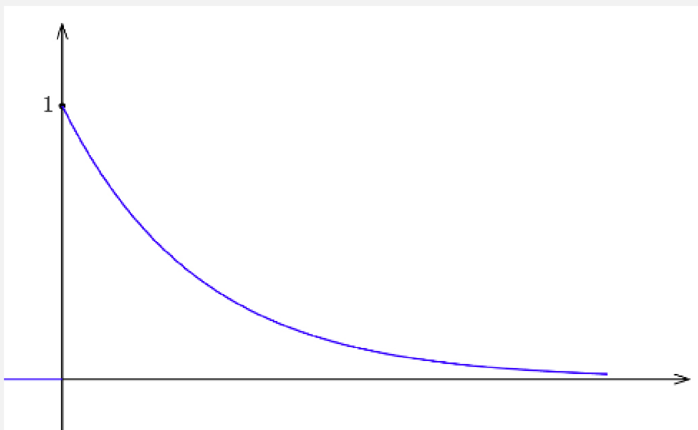
Задача. Пусть ξ равномерно распределена на отрезке $[2, 10]$. Если математическое ожидание ξ определено, то запишите в ответ его значение, округленное с точностью до 3 знаков после запятой. Если не определено, введите `no`

Введите математическую формулу

Напишите ваш ответ здесь...

Мы начинали знакомство с математическим ожиданием абсолютно непрерывных случайных величин с экспоненциального распределения, но его математическое ожидание пока не вычислили. Давайте сделаем это.

Найдите математическое ожидание случайной величины ξ, имеющей экспоненциальное распределение с параметром $\lambda = 1$ или убедитесь, что оно не определено. Её плотность:



Плотность $p_\xi(x) = \begin{cases} e^{-x}, & \text{если } x > 0 \\ 0, & \text{иначе} \end{cases}$

Чтобы найти неопределённый интеграл, нужный для решения этой задачи, материала нашего курса недостаточно. Поэтому мы предлагаем вам воспользоваться для этих целей WolframAlpha. Вот пример, как с его помощью можно найти неопределённый интеграл. Осталось только понять, какую функцию нужно интегрировать.

Задача. Если математическое ожидание ξ определено, то запишите в ответ его значение, округленное с точностью до 3 знаков после запятой. Если не определено, введите `no`

Введите математическую формулу

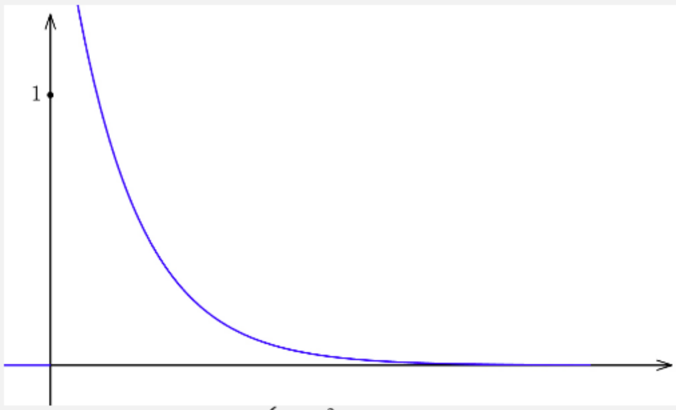
Напишите ваш ответ здесь...

Задача с проверкой. Математическое ожидание 1

Всё это время мы работали с экспоненциальным распределением с параметром $\lambda = 1$. В общем случае плотность случайной величины ξ с экспоненциальным распределением с параметром $\lambda > 0$ равняется

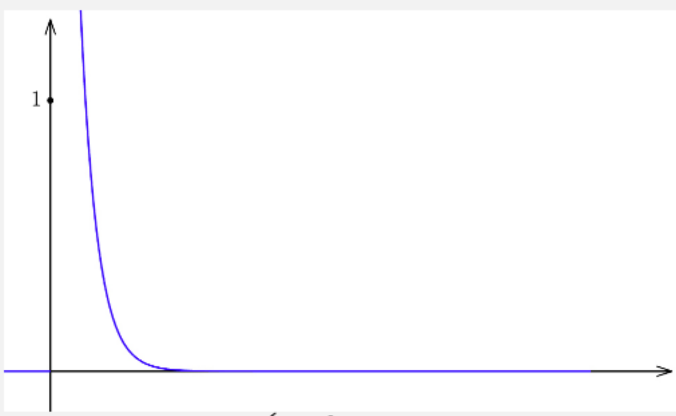
$$p_\xi(x) = \begin{cases} \lambda e^{-\lambda x}, & \text{если } x > 0 \\ 0, & \text{иначе} \end{cases}$$

1. Найдите математическое ожидание экспоненциального распределения с параметром $\lambda = 2$



$$p_\xi(x) = \begin{cases} 2e^{-2x}, & \text{если } x > 0 \\ 0, & \text{иначе} \end{cases}$$

2. Найдите математическое ожидание экспоненциального распределения с параметром $\lambda = 8$



$$p_\xi(x) = \begin{cases} 8e^{-8x}, & \text{если } x > 0 \\ 0, & \text{иначе} \end{cases}$$

Попробуйте угадать ответ для произвольного λ.

Комментарий. Как мы помним, экспоненциальное распределение можно использовать, чтобы моделировать время до следующего звонка в службу поддержки. Если в среднем это время большое – то нужно взять λ поменьше. Если в среднем это время маленькое – то нужно взять λ побольше. Очень неформально говоря, λ отвечает за вероятность генерации звонка в каждый момент времени. Если среднее время между звонками равно $\frac{1}{8}$ часа, то какое λ стоит взять?

Проверка. Введите ответы с точностью до 3 знаков после запятой. Если математическое ожидание не определено, введите `no`

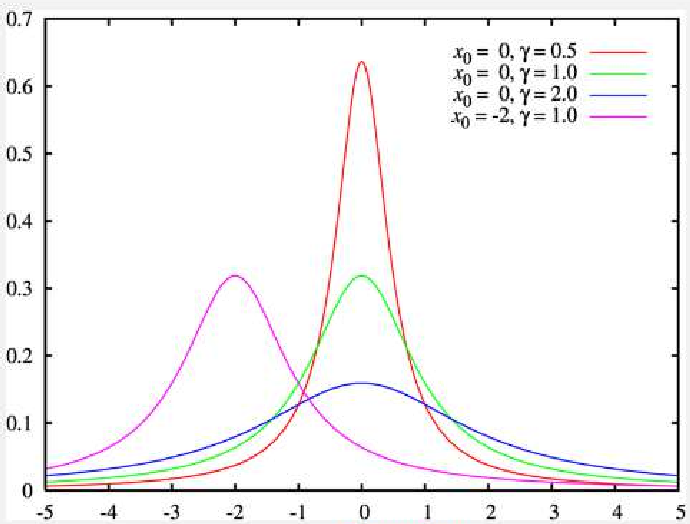
Заполните пропуски

1. Математическое ожидание экспоненциального распределения с параметром $\lambda = 2$ равно

2. Математическое ожидание экспоненциального распределения с параметром $\lambda = 8$ равно

А всё-таки бывает так, что математическое ожидание не определено?

Да, бывает. Например, распределение Коши — абсолютно непрерывно, но математическое ожидание у него не определено. Вот графики его плотности для различных значений параметров:



Источник: Википедия

Оно похоже по виду на нормальное распределение, с которым мы ещё познакомимся в следующих уроках. Однако, по сравнению с нормальным распределением у распределения Коши "более тяжелые хвосты". То есть больше вероятности сконцентрировано на лучах, уходящих в минус бесконечность и плюс бесконечность. Неформально говоря, поэтому нужный интеграл и не сходится.

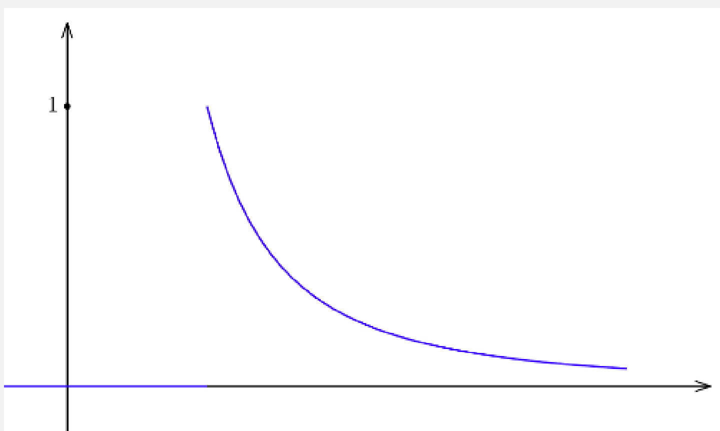
Распределение Парето

Распределение Парето хорошо описывает такие величины:

- размеры метеоритов (очень много маленьких, мало больших)
- размеры поселений (много маленьких деревень и маленьких городов, мало больших городов)
- число читателей у разных писателей (много писателей с маленьким числом читателей, мало писателей с большим числом читателей)

Как и у экспоненциального распределения, распределение Парето имеет параметры. То есть это не одно распределение, а целое семейство распределений.

У распределения Парето параметров два: x_m и α. Если взять $x_m = \alpha = 1$, то функция распределения будет такой: $1 - \frac{1}{x}$ (при $x > 0$). А плотность распределения будет такой:



$$p_\xi(x) = \begin{cases} \frac{1}{x^2}, & \text{если } x > 1 \\ 0, & \text{иначе} \end{cases}$$

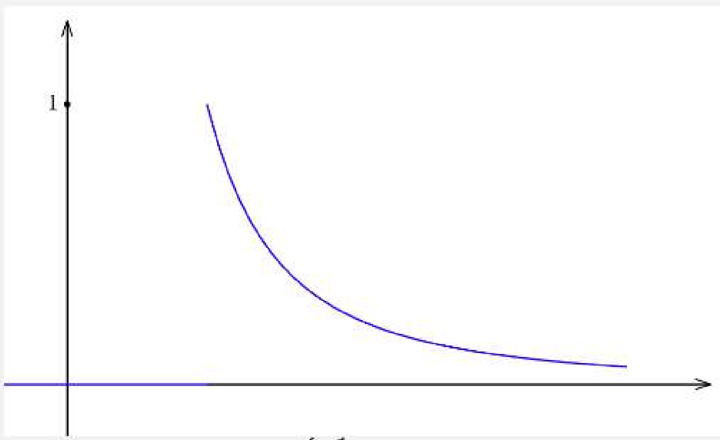
Дополнительный материал: Принцип Парето или $20/80$

Принцип Парето это эмпирическое правило, которое иногда работает. Вот его разнообразные варианты:

- 20% работы приносят 80% результата
- 20% клиентов приносят 80% дохода
- 20% заданий занимают 80% времени
- 20% самых богатых людей владеют 80% мирового капитала
- на 20% самых известных стримеров приходится 80% всех просмотров стримеров

Естественно, на практике пропорция редко оказывается в точности равна $20/80$

Распределение Парето с параметрами $x_m = \alpha = 1$ имеет такую плотность:



$$p_\xi(x) = \begin{cases} \frac{1}{x^2}, & \text{если } x > 1 \\ 0, & \text{иначе} \end{cases}$$

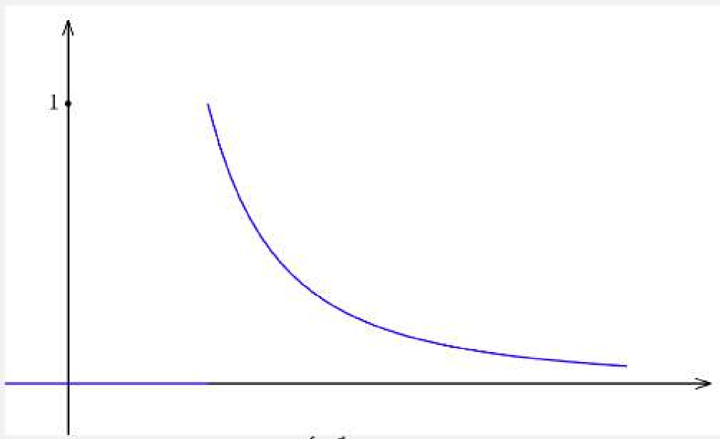
Для случайной величины ξ, имеющей распределение Парето с параметрами $x_m = \alpha = 1$ вычислите следующие вероятности. Ответы округлите с точностью до 3 знаков после запятой.

Заполните пропуски

1. $P(\xi < 2) =$
2. $P(\xi > 3)$
3. $P(2 < \xi < 4)$

Задача с проверкой. Математическое ожидание 2

Давайте найдём математическое ожидание случайной величины, имеющей распределение Парето. Распределение Парето с параметрами $x_m = \alpha = 1$ имеет такую плотность:



$$p_\xi(x) = \begin{cases} \dfrac{1}{x^2}, & \text{если } x > 1 \\ 0, & \text{иначе} \end{cases}$$

Задача. Найдите математическое ожидание случайной величины ξ или убедитесь, что оно не определено.

Проверка. Введите ответ с точностью до 3 знаков после запятой. Если математическое ожидание не определено, введите `no`

Введите математическую формулу

Напишите ваш ответ здесь...

Задача с проверкой. Математическое ожидание 3

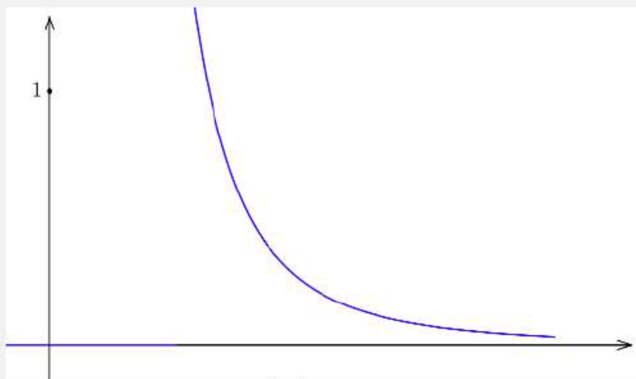
Посмотрим, как ведёт себя распределение Парето с другими параметрами. Если зафиксировать $x_m = 1$, то у распределения Парето останется один параметр, который мы будем менять – это $\alpha > 0$. Функция распределения в этом случае будет такой: $1 - \frac{1}{x^\alpha}$ (при $x > 0$). Плотность такого распределения будет равняться:

$$p_\xi(x) = \begin{cases} \dfrac{\alpha}{x^{\alpha+1}}, & \text{если } x > 1 \\ 0, & \text{иначе} \end{cases}$$

Задача. Найдите математическое ожидание случайной величины, имеющей распределение Парето с параметрами $x_m = 1$ и произвольным α.

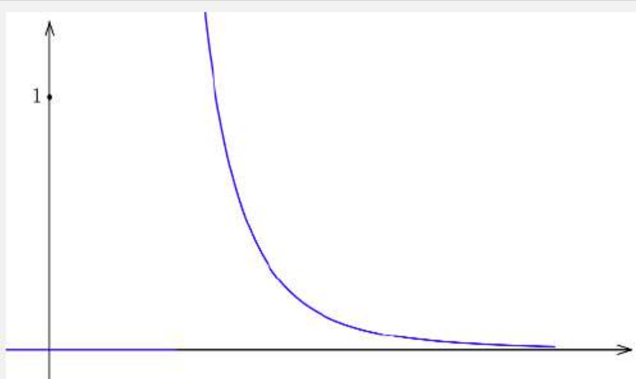
Задача для проверки.

$1.$ Найдите математическое ожидание случайной величины, имеющей распределение Парето с параметрами $x_m = 1$ и $\alpha = 2$



$$p_\xi(x) = \begin{cases} \dfrac{2}{x^3}, & \text{если } x > 1 \\ 0, & \text{иначе} \end{cases}$$

$2.$ Найдите математическое ожидание случайной величины, имеющей распределение Парето с параметрами $x_m = 1$ и $\alpha = 3$



$$p_\xi(x) = \begin{cases} \dfrac{3}{x^4}, & \text{если } x > 1 \\ 0, & \text{иначе} \end{cases}$$

Введите ответ с точностью до 3 знаков после запятой. Если математическое ожидание не определено, введите `no`

Заполните пропуски

$1.$ Математическое ожидание случайной величины, имеющей распределение Парето с параметрами $x_m = 1$ и $\alpha = 2$ равняется

$2.$ Математическое ожидание случайной величины, имеющей распределение Парето с параметрами $x_m = 1$ и $\alpha = 3$ равняется

Что мы прошли на этом уроке

- Мы узнали главный итог этого урока — с помощью плотности p_ξ мы можем посчитать мат.ожидание случайной величины ξ:

$$E\xi = \int\limits_{-\infty}^{+\infty} x p_\xi(x) dx$$

- Также мы посчитали мат.ожидание для нескольких конкретных распределений: экспоненциального и Парето

Что нас ждёт на следующем уроке

На следующем уроке мы

- узнаем, как считается дисперсия абсолютно непрерывной случайной величины
- порешаем задачки на нахождение дисперсии