Математика для Data Science. Теория вероятностей. Решения задач

Содержание

Замечание. Вот этим цветом отмечены ссылки на страницы внутри этого файла.

Определённый интеграл

Задача 1

Дана постоянная функция $f(x) = 3$, определённая на отрезке $[0, 100]$. Найдите $\int\limits_0^{100} f(x)\,dx$.

Решение.

По определению $\int\limits_0^{100} f(x)\,dx$ — это предел при ранге разбиения стремящемся к нулю от интегральной суммы $\sum\limits_{i=0}^{k-1} f(c_i)(x_{i+1} - x_i) = \sum\limits_{i=0}^{k-1} 3(x_{i+1} - x_i) = 3\sum\limits_{i=0}^{k-1}(x_{i+1} - x_i) = 3\left(x_1 - x_0 + x_2 - x_1 + x_3 - x_2 + \cdots + x_k - x_{k-1}\right)$. В последнем выражении все слагаемые кроме x_0 и x_k сокращаются и остаётся $3(x_k - x_0) = 3(100 - 0) = 300$.

Задача 2

Дана функция $f(x) = x$, определённая на отрезке $[0, 1]$. Построим последовательность разбиений. Разбиение номер k будет состоять из точек $0 < \frac{1}{k} < \frac{2}{k} < \cdots < \frac{k-1}{k} < 1$. В качестве c_i на каждом из отрезков мы выбираем самую правую точку отрезка.

1. Докажите, что ранг этих разбиений стремится к нулю.
2. Найдите предел соответствующих интегральных сумм.

Заметьте, что мы нашли предел только одной последовательности разбиений. Чтобы доказать, что найденное число действительно является интегралом функции $f(x) = x$ на отрезке $[0, 1]$, нам бы пришлось рассмотреть всевозможные другие последовательности разбиений.

Подсказка. Полезна будет формула $1 + 2 + \cdots + k = \frac{k(k-1)}{2}$.

Решение.

1. Посчитаем ранг разбиения: $\frac{m}{k} - \frac{m-1}{k} = \frac{1}{k}$. То есть все длины отрезков разбиения равны $\frac{1}{k}$. А значит, ранг разбиения тоже равен $\frac{1}{k}$. А, как мы помним, $\lim\limits_{k\to\infty} \frac{1}{k} = 0$.
2. Рассмотрим теперь предел интегральных сумм:

$$\lim_{k\to\infty} \sum_{m=1}^{k} \frac{1}{k} \cdot f\left(\frac{m}{k}\right) = \lim_{k\to\infty} \sum_{m=1}^{k} \frac{1}{k} \cdot \frac{m}{k} = \lim_{k\to\infty} \frac{1}{k^2} \sum_{m=1}^{k} m = \lim_{k\to\infty} \frac{1}{k^2} \cdot \frac{k(k-1)}{2} = \lim_{k\to\infty} \left(\frac{k^2 - k}{2k^2}\right) = \lim_{k\to\infty} \left(\frac{1}{2} - \frac{1}{2k}\right) = \frac{1}{2},$$

поскольку в предпоследнем выражении предел $\lim\limits_{k\to\infty} \frac{1}{2k}$ равен нулю.

Неопределённый интеграл

Дополнительная задача

Задача. Формализуйте доказательство с предыдущего шага.
Вам понадобятся определения непрерывности из курса матана, вот они.
Определение [по Коши]. Функция f называется *непрерывной* в точке x_0, если

- $x_0 \in D$ (где D это область определения f),
- для любого $\varepsilon > 0$ найдётся $\delta > 0$, такое что выполнено неравенство $|f(x) - f(x_0)| < \varepsilon$ для всех $x \in D$, удовлетворяющих $|x - x_0| < \delta$.

Определение [по Гейне]. Функция f называется *непрерывной* в точке x_0, если

- $x_0 \in D$,

- предел $\lim\limits_{x\to x_0} f(x)$ существует и равен $f(x_0)$.

Определения непрерывности в точке по Коши и по Гейне эквивалентны.
Определение. Функция f называется непрерывной если она непрерывна в каждой точке D.

Решение. Обоснуем, почему для непрерывной функции $f(x)$ выполнено $\int\limits_{l}^{l+\Delta l} f(x)\,dx \approx \int\limits_{l}^{l+\Delta l} f(l)\,dx$.
Правая часть этого примерного равенства равна $f(l)\Delta l$.
Нам нужно доказать, что

$$\int\limits_{l}^{l+\Delta l} f(x)\,dx - f(l)\Delta l \stackrel{?}{=} o(\Delta l)$$

при $\Delta l \to 0$.

Знак вопроса над равенством мы используем здесь и далее, чтобы не забыть, что это равенство мы хотим доказать.

То есть что

$$\lim_{\Delta l\to 0} \frac{1}{\Delta l}\left(\int\limits_{l}^{l+\Delta l} f(x)\,dx - f(l)\Delta l\right) \stackrel{?}{=} 0$$

Запишем это по-другому:

$$\lim_{\Delta l\to 0} \frac{\int\limits_{l}^{l+\Delta l} f(x)\,dx}{\Delta l} \stackrel{?}{=} f(l).$$

Воспользуемся определением непрерывности по Коши: поскольку функция f непрерывна в точке l, то для любого $\varepsilon > 0$ найдётся $\delta > 0$, такое что выполнено неравенство $|f(x) - f(l)| < \varepsilon$ для всех x, удовлетворяющих $|x - l| < \delta$.

Раскрыв в неравенстве модуль, получим, что для таких x, что $|x-l| < \delta$, выполнено $f(l)-\varepsilon < f(x) < f(l)+\varepsilon$.

А тогда при $\Delta l < \delta$ выполнено $\int\limits_{l}^{l+\Delta l} (f(l) - \varepsilon)\,dx < \int\limits_{l}^{l+\Delta l} f(x)\,dx < \int\limits_{l}^{l+\Delta l} (f(l) + \varepsilon)\,dx$

В левой и правой части неравенства интеграл берётся от константы, а значит он равен подынтегральному выражению, умноженному на длину отрезка, по которому ведётся интегрирование. Итак, неравенство равносильно такому: $(f(l) - \varepsilon)\Delta l < \int\limits_{l}^{l+\Delta l} f(x)\,dx < (f(l) + \varepsilon)\Delta l$

А тогда, поделив на Δl, мы получим

$$f(l) - \varepsilon < \frac{\int\limits_{l}^{l+\Delta l} f(x)\,dx}{\Delta l} < f(l) + \varepsilon$$

Теперь вспомним, что ε мы можем брать произвольным положительным числом. А следовательно

$$\lim_{\Delta l\to 0} \frac{\int\limits_{l}^{l+\Delta l} f(x)\,dx}{\Delta l} = f(l).$$

А ровно это мы и хотели доказать, мы молодцы!

Задача 1

1. Найдите неопределённый интеграл функции x^n.

2. Найдите определённый интеграл $\int\limits_{a}^{b} x^n\,dx$.

Каждый раз, когда мы пишем "найдите неопределённый интеграл" мы имеем в виду "найдите какой-нибудь неопределённый интеграл".

Подсказка. Нужно найти первообразные данной в условии функции и воспользоваться формулой Ньютона-Лейбница.

Решение. Будем пользоваться формулой Ньютона-Лейбница: если $F'(x) = f(x)$, то

$$\int\limits_a^b f(x)\,dx = F(b) - F(a).$$

1. Нам нужно найти такую функцию $F(x)$, что $F'(x) = x^n$. Вспомним формулу из матана: для любого $m \in \mathbb{Z} \setminus \{0\}$ выполнено $(x^m)' = mx^{m-1}$. То есть взятие производной уменьшает степень на единицу. Поскольку мы хотим получить x^n, то будем искать $F(x)$ в виде $F(x) = cx^{n+1}$. Тогда $F'(x) = c(n+1)x^n = x^n$. Итого $c = \frac{1}{n+1}$ и, значит,
$$F(x) = \frac{x^{n+1}}{n+1}.$$
   А тогда $\int x^n dx = \frac{x^{n+1}}{n+1} + C$, где C — константа.

2. По формуле Ньютона-Лейбница $\int\limits_a^b x^n\,dx = F(b) - F(a) = \frac{b^{n+1}}{n+1} - \frac{a^{n+1}}{n+1} = \frac{b^{n+1}-a^{n+1}}{n+1}$.

   По только что полученной формуле $\int\limits_0^2 x^5\,dx = \frac{2^{5+1}-0^{5+1}}{5+1} = \frac{2^6}{6} = \frac{32}{3}$.

Задача 2

1. Найдите неопределённый интеграл функции $\cos(x)$.

2. Найдите определённый интеграл $\int\limits_a^b \cos(x)\,dx$.

Подсказка. Как и в прошлой задаче нужно найти первообразные данной в условии функции и воспользоваться формулой Ньютона-Лейбница.

Решение. Как и в прошлой задаче будем пользоваться формулой Ньютона-Лейбница: если $F'(x) = f(x)$, то

$$\int\limits_a^b f(x)\,dx = F(b) - F(a).$$

1. Нам нужно найти $F(x)$ такую, что $F'(x) = \cos(x)$. Вспоминая формулы производных, получаем, что можно взять $F(x) = \sin(x)$.

   То есть $\int \cos(x) dx = \sin(x) + C$, где C — константа.

2. По формуле Ньютона-Лейбница $\int\limits_a^b \cos(x)\,dx = \sin(b) - \sin(a)$.

Задача 3

Пусть f и g – непрерывные функции на отрезке $[a, b]$.

1. Докажите, что $\int\limits_a^b cf(x)\,dx = c\int\limits_a^b f(x)\,dx$ для любого $c \in \mathbb{R}$.

2. Докажите, что $\int\limits_a^b f(x)\,dx + \int\limits_a^b g(x)\,dx = \int\limits_a^b f(x)+g(x)\,dx$.

Тем самым операция взятия определённого интеграла линейна на множестве всех непрерывных функций.

Подсказка. Можно решать эту задачу, пользуясь определением интеграла, а можно — через формулу Ньютона-Лейбница.

Решение.

1. Поскольку по следствию с шага ранее мы знаем, что достаточно рассмотреть одну последовательность разбиений, то рассмотрим разбиения, где длины отрезков одинаковы и равны Δx. Тогда $\int\limits_a^b cf(x)\,dx = \lim\limits_{\Delta x\to 0}\sum\limits_{i=0}^{k-1} cf(\xi_i)\Delta x$, где за ξ_i обозначена точка, выбранная на отрезке $[x_i, x_{i+1}]$.
   По свойствам операции суммирования и взятия предела
   $\lim\limits_{\Delta x\to 0}\sum\limits_{i=0}^{k-1} cf(\xi_i)\Delta x = \lim\limits_{\Delta x\to 0} c\cdot\left(\sum\limits_{i=0}^{k-1} f(\xi_i)\Delta x\right) = c\cdot\left(\lim\limits_{\lambda_R\to 0}\sum\limits_{i=0}^{k-1} f(\xi_i)\Delta x\right) = c\int\limits_a^b f(x)\,dx$
2. Аналогично предыдущему пункту воспользуемся линейнойстью суммирования и взятия предела:
   $\int\limits_a^b (f(x)+g(x))\,dx = \lim\limits_{\Delta x\to 0}\sum\limits_{i=0}^{k-1}(f(\xi_i)+g(\xi_i))\Delta x = \lim\limits_{\Delta x\to 0}\left(\sum\limits_{i=0}^{k-1} f(\xi_i)\Delta x + \sum\limits_{i=0}^{k-1} g(\xi_i)\Delta x\right) =$
   $= \lim\limits_{\Delta x\to 0}\sum\limits_{i=0}^{k-1} f(\xi_i)\Delta x + \lim\limits_{\Delta x\to 0}\sum\limits_{i=0}^{k-1} g(\xi_i)\Delta x = \int\limits_a^b f(x)\,dx + \int\limits_a^b g(x)\,dx$

Задача 4

Найдите несобственный интеграл $\int\limits_0^{+\infty} x\,dx$ или докажите, что он расходится

Решение. По определению $\int\limits_0^{+\infty} x\,dx = \lim\limits_{b\to+\infty}\int\limits_0^b x\,dx = \lim\limits_{b\to+\infty}\left(\frac{b^2}{2}-\frac{0}{2}\right) = +\infty$, то есть ряд расходится.

Задача 5

Найдите несобственный интеграл $\int\limits_1^{+\infty}\frac{1}{x^3}\,dx$ или докажите, что он расходится

Решение. $\int\limits_1^{+\infty}\frac{1}{x^3}\,dx = \int\limits_1^{+\infty} x^{-3}\,dx = \lim\limits_{b\to+\infty}\int\limits_1^b x^{-3}\,dx = \lim\limits_{b\to+\infty}\left(\frac{b^{-2}}{-2}-\frac{1^{-2}}{-2}\right) = \lim\limits_{b\to+\infty}\left(\frac{b^{-2}}{-2}-\frac{1^{-2}}{-2}\right) = \lim\limits_{b\to+\infty}\left(\frac{1}{-2b^2}+\frac{1}{2}\right) = \frac{1}{2}$.

Задача 4

Найдите несобственный интеграл $\int\limits_{-\infty}^{+\infty}\sin x\,dx$ или докажите, что он расходится

Решение. $\int\limits_{-\infty}^{+\infty}\sin x\,dx = \int\limits_{-\infty}^{c}\sin x\,dx + \int\limits_{c}^{+\infty}\sin x\,dx$. При этом $\int\limits_{c}^{+\infty}\sin x\,dx = \lim\limits_{b\to+\infty}\int\limits_c^b \sin x\,dx = \lim\limits_{b\to+\infty}(-\cos(b)+\cos(c)) = \cos(c) - \lim\limits_{b\to+\infty}\cos(b)$. Но предел $\lim\limits_{b\to+\infty}\cos(b)$ не существует: например, если взять $b_n = \pi n$, то $\lim\limits_{n\to\infty}\cos(b_n) = \lim\limits_{n\to\infty}\cos(\pi n) = \lim\limits_{n\to\infty}(-1)^n$, а этого предела не существует.

Непрерывные вероятностные пространства

Задача 1

Докажите, что вероятность прихода Светы в течение вечеринки действительно равна 1. То есть, что $P([0,3]) = \int\limits_0^3 g(t)\,dt = 1$.

Можете строго найти интеграл, используя первообразную. А можете считать, что мы уже убедились, что интеграл это площадь, и просто найти площадь нужных частей подграфика (используя формулы площади прямоугольника и площади прямоугольного треугольника).

Подсказка. Если решать через интеграл, то полезно воспользоваться его линейностью.

Решение. Напомним, что $g : [0,3] \to \mathbb{R}$ и функция определена так: $g(t) = \frac{2}{9}(3-t)$. Разберём оба способа решения:

1. Через интеграл.

$$P([0,3]) = \int\limits_0^3 g(t)\,dt = \int\limits_0^3 \left(\frac{2}{9}(3-t)\right)\,dt = \int\limits_0^3 \left(\frac{2}{3} - \frac{2t}{9}\right)\,dt = \int\limits_0^3 \frac{2}{3}\,dt - \int\limits_0^3 \frac{2t}{9}\,dt = \frac{2}{3}\int\limits_0^3 1\,dt - \frac{2}{9}\int\limits_0^3 t\,dt$$

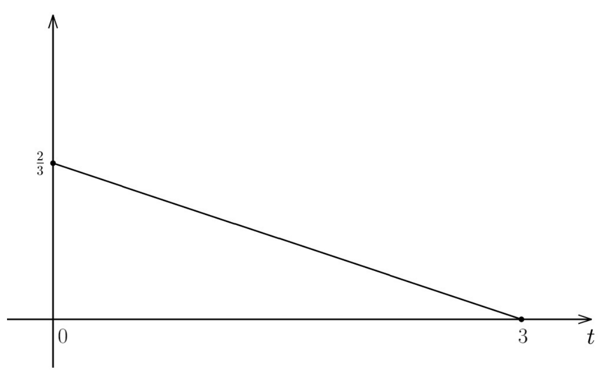
   В последних двух равенствах мы воспользовались линейностью интеграла, которую доказали в этой устной задаче.

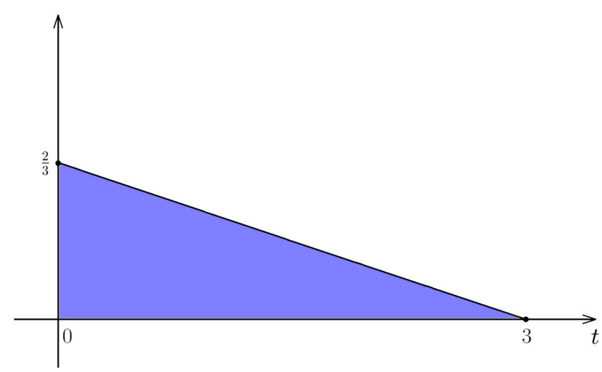
   Теперь вспоминим, что в одной из задач прошлого урока мы доказали, что $\int\limits_a^b t^n\,dt = \frac{b^{n+1}-a^{n+1}}{n+1}$. Воспользуемся этой формулой в нашем случае: в первом интеграле степень $n = 0$ (ведь $t^0 = 1$), во втором — $n = 1$

$$P([0,3]) = \frac{2}{3}\int\limits_0^3 t^0\,dt - \frac{2}{9}\int\limits_0^3 t\,dt = \frac{2}{3}\cdot\frac{3^1-0^1}{1} - \frac{2}{9}\cdot\frac{3^2-0^2}{2} = \frac{2}{3}\cdot\frac{3}{1} - \frac{2}{9}\cdot\frac{9}{2} = 2 - 1 = 1.$$

2. Через площадь.

   Напомним, что график функции $g(t)$ выглядит так:





   На графике справа выделена интересующая нас область, соответствующая отрезку $[0,3]$.

   Мы видим, что подграфик образует прямоугольный треугольник с катетами длины $\frac{2}{3}$ и 3. А значит, его площадь равна половине произведения катетов: $\frac{1}{2}\cdot\left(\frac{2}{3}\cdot 3\right) = 1$.

Задача 2

Оказывается, что и Костя, и Лена приходят на вечеринку ровно на час. То есть каждый из них приходит на вечеринку, проводит на ней час времени и уходит. Если вечеринка закончилась раньше, чем пройдёт этот час, то человек просто уходит вместе с остальными гостями в конце вечеринки. Найдите вероятность того, что вы сможете представить друг другу Костю и Лену.

Другими словами, найдите вероятность того, что в какой-то момент на вечеринке одновременно будут и Костя, и Лена.

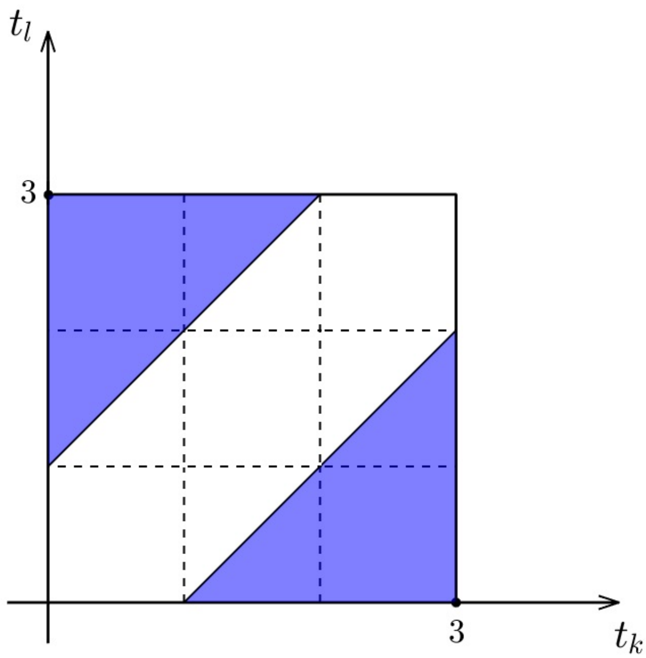
Подсказка. Это событие соответствует условию $|t_k - t_l| < 1$.

Решение. Костя и Лена не пересекутся, если придут с разницей не меньше часа. То есть либо $t_k - t_l \geq 1$, либо $t_l - t_k \geq 1$. Другими словами, они не пересекутся, если $|t_k - t_l| \geq 1$. А значит, в какой-то момент на вечеринке одновременно будут и Костя, и Лена, если, наоборот, $|t_k - t_l| < 1$.

На предыдущем шаге мы уже нашли вероятность того, что Лена придёт хотя бы на час раньше Кости: $P(t_l + 1 < t_k) = \frac{2}{9}$. Поскольку времена прихода Кости и Лены независимы и одинаково распределены, то вероятность того, что Костя придёт хотя бы на час раньше Лены, такая же: $P(t_k + 1 < t_l) = \frac{2}{9}$.

Итак, искомая $P(|t_k - t_l| < 1) = \frac{2}{9} + \frac{2}{9} = \frac{4}{9}$.

На рисунке множество точек, для которых выполнено $|t_k - t_l| \geq 1$ выглядит так:



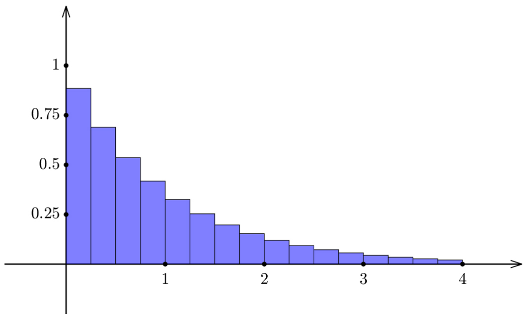
Площадь каждого из выделенных треугольников равна $\frac{1}{2} \cdot 2 \cdot 2 = 2$. Тогда их суммарная площадь равна 4. Площадь всего квадрата равна $3 \cdot 3 = 9$. Тогда площадь искомого участка квадрата равна $9 - 4 = 5$, а искомая вероятность тогда равна $\frac{5}{9}$.

Плотность вероятности

Задача 1

Чему будет равняться сумма площадей всех столбиков псевдо-гистограммы для экспоненциального распределения с параметром $\lambda = 1$?

Решение. Пусть каждый из наших столбиков имеет ширину Δx.



Мы уже выяснили, что площадь столбца равняется вероятности того, что случайная величина попадёт на этот интервал. Тогда в нашем случае сумма площадей столбиков равна $\sum\limits_{k=0}^{\infty} P(k \cdot \Delta x < \xi < (k+1) \cdot \Delta x)$, где ξ — случайная величина с экспоненциальным распределением с параметром $\lambda = 1$.

Тогда, пользуясь предыдущей задачей, мы получаем, что сумма площадей столбиков равна

$$\sum_{k=0}^{\infty} P(k \cdot \Delta x < \xi < (k+1) \cdot \Delta x) = \sum_{k=0}^{\infty} P(k \cdot \Delta x \leq \xi < (k+1) \cdot \Delta x) = P(\xi \in [0, +\infty)) = 1,$$

ведь экспоненциальное распределение как раз определено на луче $[0, +\infty)$.

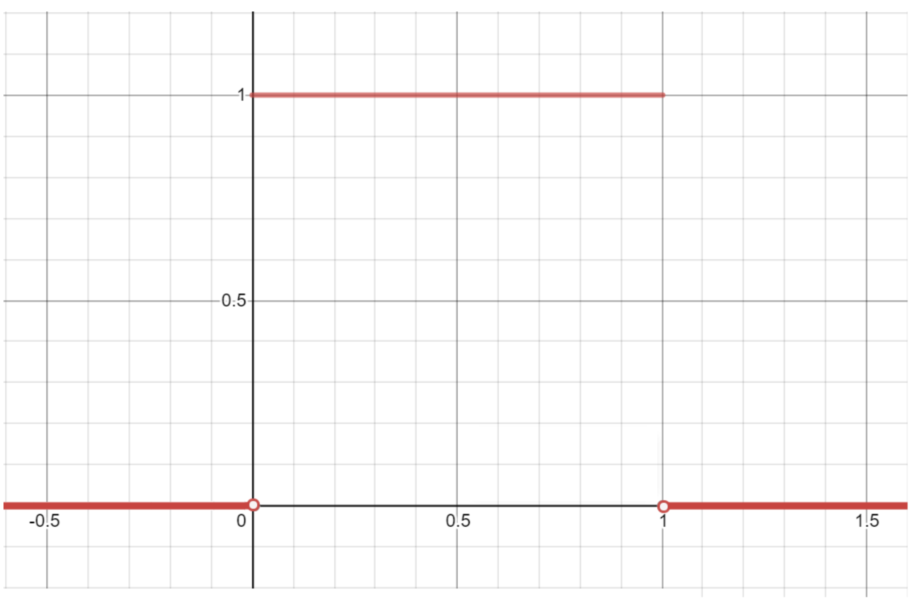
Задача 3

Нарисуйте график плотности вероятности для равномерного распределения

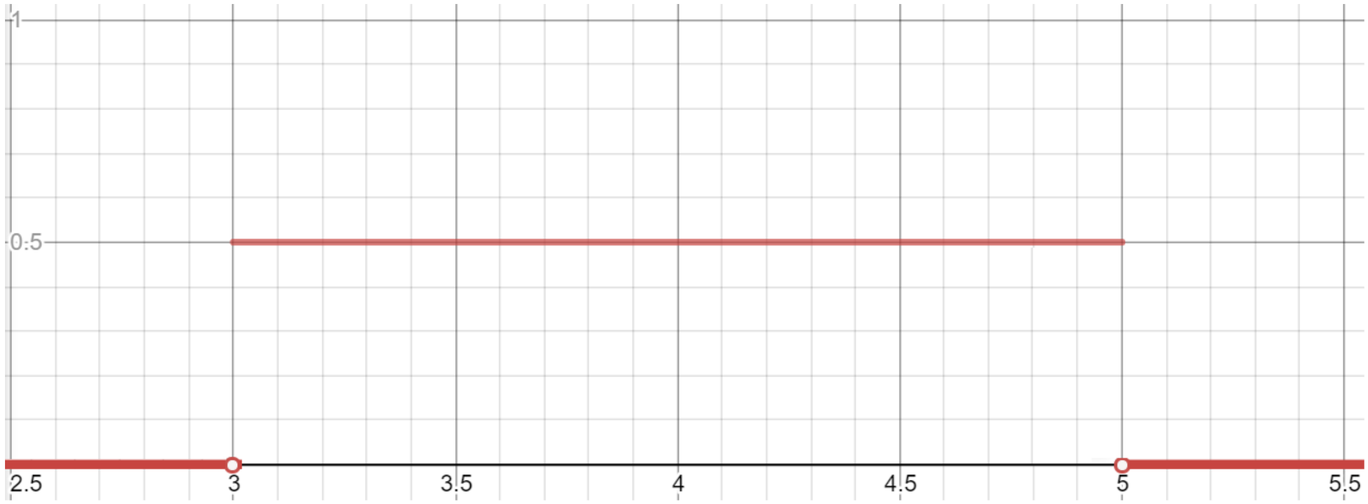
1. на отрезке $[0, 1]$
2. на отрезке $[3, 5]$
3. на отрезке $[1, x]$, где $x > 1$

Решение. Поскольку функция вероятности для равномерного распределения линейна на отрезке, на котором задано равномерное распределение, то её производная на этом отрезке постоянна. А вне этого отрезка функция вероятности постоянна (равна нулю слева от отрезка и единице — справа от него), а значит производная на этих лучах равна нулю. Итак, мы получаем вот такие графики плотности вероятности для равномерного распределения

1. на отрезке $[0, 1]$



2. на отрезке $[3, 5]$



3. на отрезке $[1, x]$, где $x > 1$

   График будет выглядеть так: на отрезке $[1, x]$ функция постоянна и равна $\frac{1}{x-1}$, а на лучах $(-\infty, 1)$ и $(x, +\infty)$ функция равняется нулю: