Теория вероятности. Непрерывные случайные величины

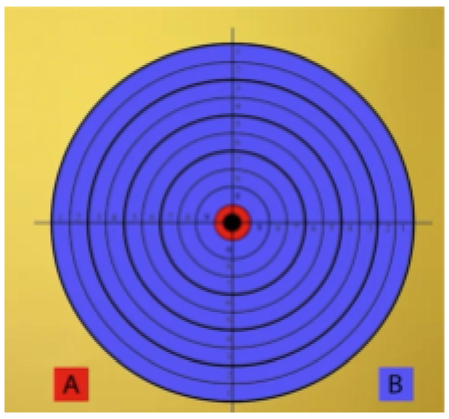
Содержание:

Условная вероятность
Непрерывные случайные величины
Многомерные распределения
Энтропия

Условная вероятность

$$A|B : P(A|B) = \frac{P(AB)}{P(B)}, \quad P(B) > 0$$

- Формула Байеса одна из основных теорем элементарной теории вероятностей, которая позволяет определить вероятность какого-либо события при условии, что произошло другое статистически взаимозависимое с ним событие. Другими словами, по формуле Байеса можно более точно пересчитать вероятность, взяв в расчёт как ранее известную информацию, так и данные новых наблюдений.



$$P(B) = 0.8, P(A) = 0.05$$
$$P(AB) = P(A) = 0.05 \Rightarrow$$
$$P(A|B) = 0.05/0.8 = 0.0625$$

Событие B – вероятность попадания в мишень
Событие A – вероятность попадания в десятку
Событие AB – вероятность, того что мы попадем и в десятку, и в мишень
Событие A|B – условная вероятность, что при попадании в мишень, мы попали в десятку

Формула полной вероятности:

$$P(A) = P(A|B)P(B) + P(A|\bar{B})P(\bar{B})$$

Позволяет найти вероятность события, если имеется полная группа событий. Так для примера с мишенью $\bar{B}$ – это вероятность события, когда мы не попали в мишень.

Непрерывные случайные величины

Вспомним, как задаются дискретные случайные величины.
Пусть X случайная величина, которая принимает счетное множество значений A.
Каждому значению присваиваем свою вероятность большую/ равную нулю и меньше 1.

- X принимает счётное множество значений $A = \{a_1, a_2, a_3, \dots\}$ с вероятностями $p_1, p_2, p_3, \dots$ где $p_i \geq 0 \ \forall i$ и $\sum_{i=1}^{\infty} p_i = 1$
- $P(X = a_i) = p_i$ – функция вероятности

Ключевой момент: в силу счётности X мы можем определить функцию вероятности для каждого фиксированного ai из A.

В случае абсолютно непрерывных случайных величин так сделать нельзя, потому что вероятность каждого значения случайной величины будет нулевой! Поэтому непрерывные случайные величины нельзя задавать с помощью функции вероятности.

Один из способов задания непрерывной случайной величины является **функция распределения** (**функция**, характеризующая распределение **случайной величины** или **случайного** вектора; вероятность того, что случайная **величина** X примет значение, меньшее или равное x, где x — произвольное действительное число).

Другим способом задания непрерывной случайной величины является **плотность распределения** случайной величины f(x), которая тесно связана с функцией распределения непрерывной случайной величины.

$$f(x) : \int_a^b f(x)dx = P(a \leq X \leq b) \text{ – плотность распределения}$$

$$F(x) = \int_{-\infty}^{x} f(u)du$$

$$\int_{-\infty}^{+\infty} f(u)du = P(-\infty \leq X \leq +\infty) = 1$$

Значение функции распределения в точке x для непрерывной случайной величины мы не можем вычислить суммой функций плотности распределения, так как значений случайной величины несчетное количество.

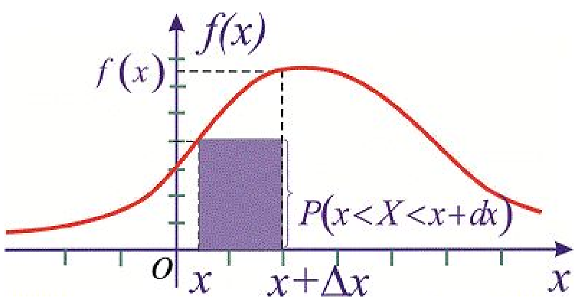
Вероятность попадания случайной величины в промежуток $[\alpha; \beta]$ определяется зависимостью

$$P(\alpha < X < \beta) = \int_\alpha^\beta f(x)dx.$$

Функция распределения вероятностей непрерывной случайной величины определяется через плотность распределения вероятностей интегрированием

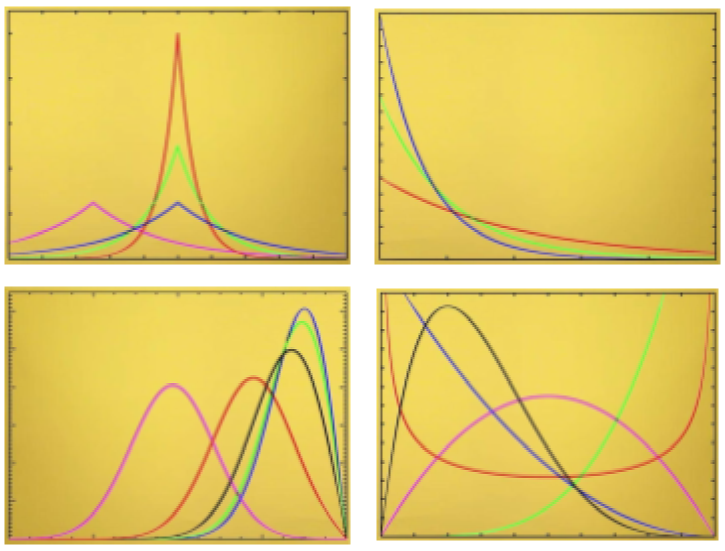
$$F(x) = \int_{-\infty}^{x} f(x)dx.$$

Геометрически на графике плотности вероятностей $f(x)dx$ соответствует площадь прямоугольника с основанием dx и высотой $f(x)$

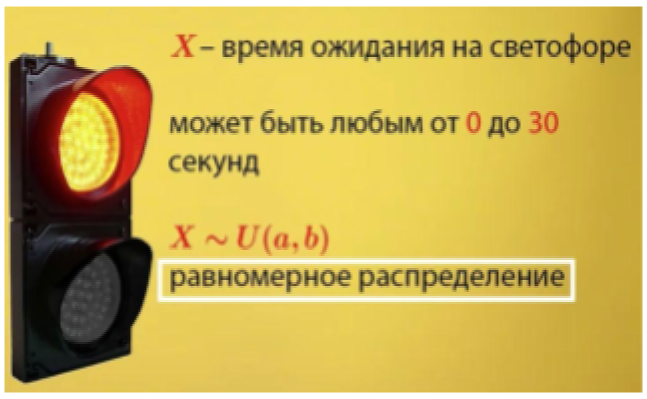


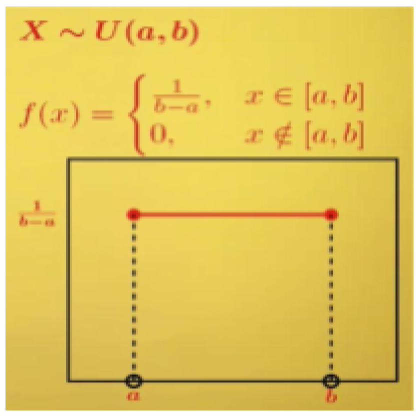
Функция плотности распределения, в отличие от неубывающей функции распределения, может вести себя совершенно по-разному. В этом их достоинство: проще отличать семейства распределений по плотностям.

Примеры непрерывных случайных величин (равномерное распределение)



Ярким примером непрерывной случайной величины, распределённой **равномерно**, является время ожидания перехода дороги со светофором без секунд.

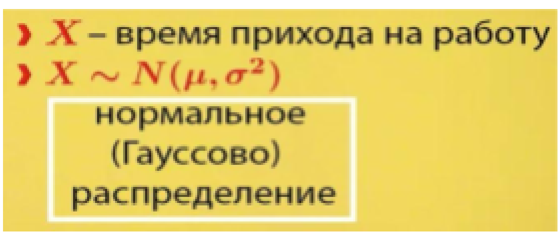


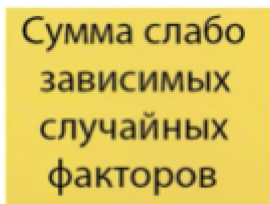


Значение случайной величины в данном примере – это количество секунд (доли секунд), которое осталось ждать до появления зеленого сигнала светофора. Количество секунд - может быть любое значение в интервале времени от 0 до 30 секунд и все эти значения являются равновероятными, поэтому значения функции плотности на данном интервале времени будут равны константе 1/(b-a) – это следует из того, что плотность равна 1, а одна из сторон прямоугольника равна b-a. Можно предсказать, что чем больше будет становиться интервал от a до b, тем меньше будет значение константы 1/(b-a).

Примеры непрерывных случайных величин (нормальное распределение)

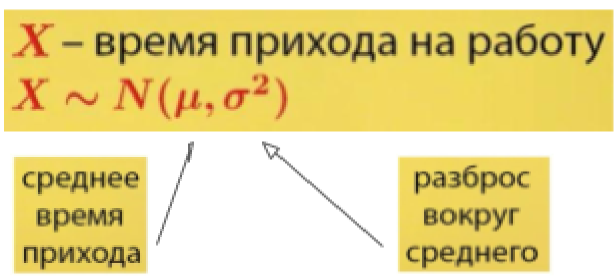
Рисунок. Пример на основе времени прихода на работу, если вы всегда старайтесь приходить в офис, например, около 12:00.





где μ - это среднее значение (в данном пример 12.00), σ^2- величина отклонения от среднего значения

Так же можно моделировать нормальным распределением следующие величины:
Погрешность барометра и длину листьев одного дерева.



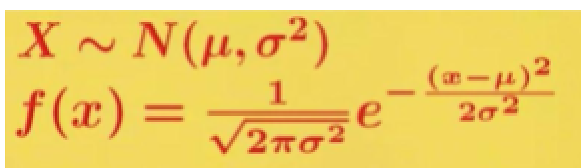
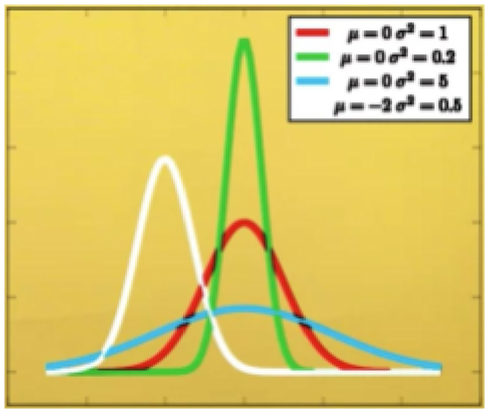
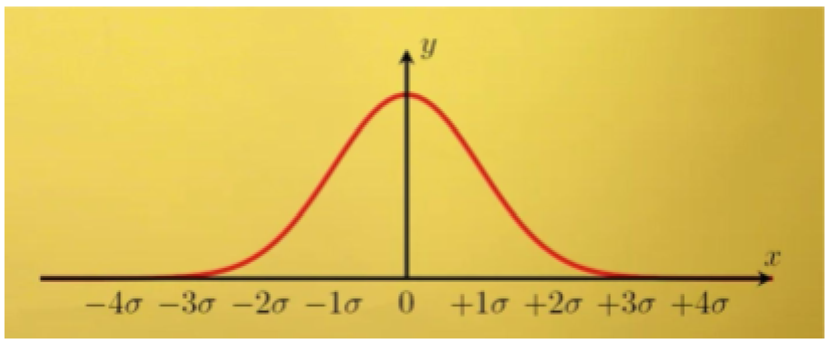


График распределения плотности будет выглядеть следующим образом:





Примеры непрерывных случайных величин (экспоненциальное распределение)

Ещё одним наиболее часто встречающимся непрерывным распределением является **экспоненциальное** распределение случайных величин.

$$F(x) = \begin{cases} 0, \ x < 0 \\ 1 - e^{-\lambda x}, \ x \geq 0 \end{cases} \qquad f(x) = \begin{cases} 0, \ x < 0 \\ \lambda e^{-\lambda x}, \ x \geq 0 \end{cases}$$

Здесь **λ**-единственный параметр данного распределения, полностью определяющий его свойства.
В частности, числовые характеристики выражаются через этот параметр: E(X)=1/λ (мат ожидание),D(X)=1/λ^2(дисперсия).
В случае нормального распределения плотность распределения никогда не 0. В случае экспоненциального распределения плотность распределения равна 0, когда значение непрерывной случайной величины имеет отрицательное значение.

Экспоненциальное распределение моделирует время между двумя последовательными свершениями события, а параметр **λ** описывает среднее число наступлений события в единицу времени. Обычно с помощью этого закона описывают:

- продолжительность обслуживания покупателя
- время жизни оборудования до отказа
- промежуток времени между поломками

Примеры непрерывных случайных величин (распределение Стьюдента)

Некоторые распределения связаны между собой.Одним из таких семейств для непрерывных случайных величин является **распределение Стьюдента**.
Пусть Yi -независимые стандартные нормальные случайные величины, тогда

$$Y_i \sim \mathrm{N}(0,1), \ i = 1, \ldots n$$

$$t = \frac{Y_0}{\sqrt{\frac{1}{n}\sum_{i=1}^{n} Y_i^2}} \qquad f_t(y) = \frac{\Gamma\left(\frac{n+1}{2}\right)}{\sqrt{\pi n}\,\Gamma\left(\frac{n}{2}\right)} \left(1 + \frac{y^2}{n}\right)^{-\frac{n+1}{2}}$$

Имеет распределение Стьюдента с n степенями свободы (количество случайных величин, взятых для подсчета суммы).

Распределение Стьюдента симметрично. В частности если t имеет распределение Стьюдента с n степенями свободы, то «–t» имеет то же распределение.
Все виды распределений есть в Phyton.

Многомерные распределения

Зачастую наш эксперимент зависит далеко не от одного параметра, и хочется каким-то образом построить распределение над векторами параметров.
Случай дискретных переменных (таблица совместного распределения):

	c=0	c=1	c=2
g=0	0.1	0.1	0.1
g=1	0.2	0.4	0.1

$$\sum_{i=1}^{\infty}\sum_{j=1}^{\infty} \mathrm{P}(\xi_1 = a_i, \ \xi_2 = b_j) = 1.$$

Пусть имеется два параметра c, g. В таблице указаны вероятности. Сумма по всем элементам таблицы должна равняться 1.

Если мы хотим посчитать с какой вероятностью наш первый параметр примет конкретное фиксированное значение, то мы считаем объединение по всем событиям для второго параметра (т.е. первый параметр примет значение при любом значении второго параметра)

$$\{\xi_1 = a_i\} = \bigcup_{j=1}^{\infty} \{\xi_1 = a_i, \xi_2 = b_j\}.$$

Маргинальное распределение

$$\mathsf{P}(\xi_1 = a_i) = \sum_{j=1}^{\infty} \mathsf{P}(\xi_1 = a_i, \xi_2 = b_j), \quad \mathsf{P}(\xi_2 = b_j) = \sum_{i=1}^{\infty} \mathsf{P}(\xi_1 = a_i, \xi_2 = b_j).$$

Аналогично можно определить совместное распределение для случая абсолютно непрерывных случайных величин (сумма заменяется интегралом, а вероятность – плотностью вероятности):

$$\mathsf{P}((\xi_1, \xi_2) \in B) = \iint\limits_B f_{\xi_1,\xi_2}(s_1, s_2) ds_1 ds_2.$$

$$F_{\xi_1,\xi_2}(x_1, x_2) = \mathsf{P}(\xi_1 < x_1, \xi_2 < x_2) = \int\limits_{-\infty}^{x_1} \left(\int\limits_{-\infty}^{x_2} f_{\xi_1,\xi_2}(s_1, s_2)\, ds_2 \right) ds_1.$$

Свойства плотности ничем не отличаются от случая одномерного распределения:

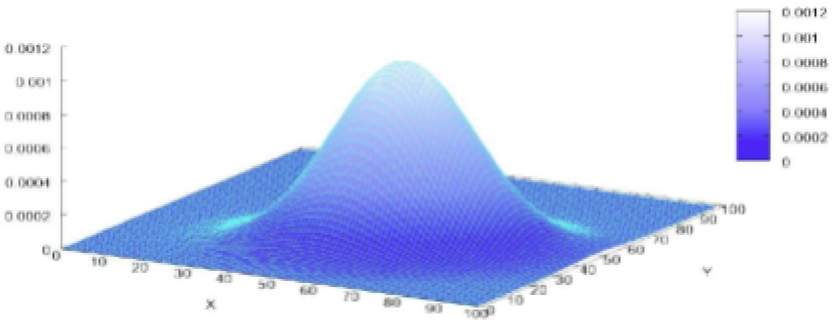
$$f_{\xi_1,\xi_2}(x_1, x_2) \geqslant 0 \text{ для любых } x_1, x_2 \in \mathbb{R}; \qquad \iint\limits_{\mathbb{R}^2} f_{\xi_1,\xi_2}(x_1, x_2)\, dx_1\, dx_2 = 1.$$

Многомерные распределения(примеры)

Многомерное нормальное:

$$X \sim N(\mu, \Sigma),\ \mu \in \mathbb{R}^k$$
$$\Sigma \in \mathbb{R}^{k \times k} \text{ положительно определена,}$$

$$f(x) = (2\pi)^{-\frac{k}{2}} |\Sigma|^{-\frac{1}{2}} e^{-\frac{1}{2}(x-\mu)^T |\Sigma^{-1}|(x-\mu)}$$



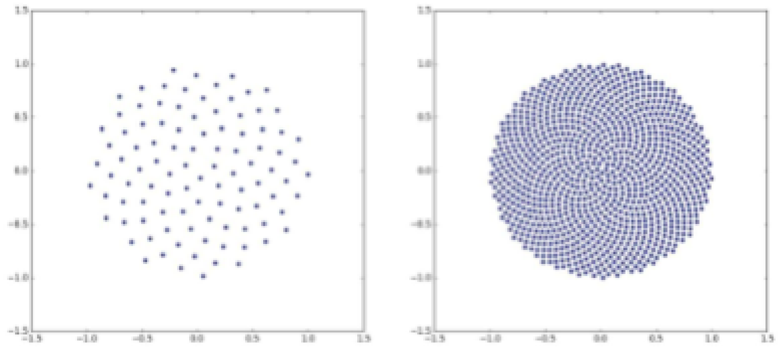
В данном случает х- это уже вектор с координатами (x, y).

Многомерное равномерное:

$$f_{\xi_1,\ldots,\xi_n}(x_1, \ldots, x_n) = \begin{cases} \dfrac{1}{\lambda(S)}, & \text{если } (x_1, \ldots, x_n) \in S, \\ 0, & \text{если } (x_1, \ldots, x_n) \notin S. \end{cases}$$

λ(S) - площадь круга (в двумерном распределении)

$$\int\limits_{\mathbb{R}^n} f_{\xi_1,\ldots,\xi_n}(x_1, \ldots, x_n)\, dx_1 \ldots\, dx_n = \frac{1}{\lambda(S)} \int\limits_S dx_1 \ldots\, dx_n = \frac{1}{\lambda(S)} \lambda(S) = 1.$$



Энтропия

Одно из важнейших понятий теории информации, напрямую связанное с теорией вероятности.
Информационная энтропия — мера неопределённости некоторой системы, в частности непредсказуемость появления какого-либо символа первичного алфавита.
Например, в последовательности букв, составляющих какое-либо предложение на русском языке, разные буквы появляются с разной частотой, поэтому неопределённость появления для некоторых букв меньше, чем для других.

Информационная двоичная энтропия для независимых случайных событий x с n возможными состояниями, распределённых с вероятностями p_i, рассчитывается по формуле Шеннона:

$$H(x) = -\sum_{i=1}^{n} p_i \log_2 p_i.$$

Энтропия(пример)

В случае равновероятных событий формула Шеннона упрощается до формулы Хартли:

$I = -\log_2 p = \log_2 N,$ (p=1/N => p=1*N^{-1})

где I – количество передаваемой информации, p – вероятность события, N – возможное количество различных (равновероятных) сообщений.

Пример: В колоде 36 карт. Какое количество информации (энтропия) содержится в сообщении, что из колоды взята карта с портретом “туз”; “туз пик”?

Вероятность p_1 = 4/36 = 1/9 (4 туза в колоде), а p_2 = 1/36 (туз пик – 1 в колоде). Используя формулу Хартли имеем:

$$I_1 = -\log_2 p_1 = \log_2 \frac{1}{\frac{1}{9}} = \log_2 9 \approx 3.17$$
$$I_2 = -\log_2 p_2 = \log_2 \frac{1}{\frac{1}{36}} = \log_2 36 \approx 5.17$$

Заметим (из второго результата), что для кодирования всех карт, необходимо 6 бит.

Пример: В колоде 36 карт. Из них 12 карт с “портретами”. Поочередно из колоды достается и показывается одна из карт для определения изображен ли на ней портрет. Карта возвращается в колоду. Определить количество информации, передаваемой каждый раз, при показе одной карты.

$$I = -(p_{ic} \log_2 p_{ic} + p_{ot} \log_2 p_{ot}) = \frac{12}{36} \log_2 \frac{1}{\frac{12}{36}} + \frac{36-12}{36} \log_2 \frac{1}{\frac{36-12}{36}} = \frac{\ln 3}{3 \ln 2} + \frac{2 \ln \frac{3}{2}}{3 \ln 2} \approx 0.91$$

где I – количество информации которую мы получаем при доставании одной карты

В данном случаем речь идет о разновероятных событиях: вероятность 12/36 – карта с портретом, вероятность 24/36 – карта без портрета, поэтому должна использоваться формула Шеннона. Так как мы возвращаем карты в колоду, вероятности не меняются.

Пример: Документация некоторого учреждения размещена в 4-х комнатах. В каждой комнате находится 16 шкафов. Каждый шкаф имеет 8 полок. Определить количество информации, которое несет сообщение о том, что нужный документ находится в третьей комнате, в тринадцатом шкафу на пятой полке.

Для независимых $x_1, \ldots, x_n$ справедливо:

$$I(x_1, x_2, \ldots, x_n) = \sum_{i=1}^{n} I(x_i)$$

$$I = \log_2 4 + \log_2 16 + \log_2 8 = 9$$

Документация может лежать равновероятно в любой из четырех комнат, поэтому используем формулу Хартли. I – количество информации об адресе документации.

Программное представление

In [1]:
```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import scipy.stats as sts

%matplotlib inline
```

Нормальное распределение

In [2]:
```
mu = 2.0
sigma = 0.5

# зададим нормально распределенную случайную величину
norm_rv = sts.norm(loc=mu, scale=sigma)

# сгенерируем 10 значений
norm_rv.rvs(size=10)
```

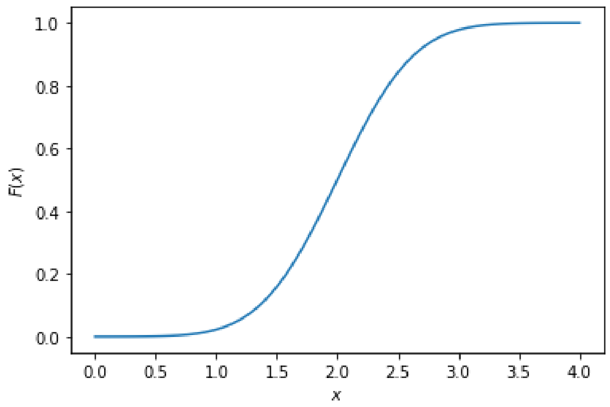
Out[2]:
```
array([1.32075542, 1.87925099, 3.19329038, 1.50120935, 2.01523973,
       1.66441037, 1.40808052, 2.17976504, 1.91338985, 2.78322131])
```

Параметр `loc` задаёт μ, `scale` — среднеквадратичное отклонение σ, `size` — размер выборки. Имя параметра `size` при вызове функции `rvs` можно не писать.

In [5]:
```
norm_rv.cdf(3)
```

Out[5]:
```
0.9772498680518208
```

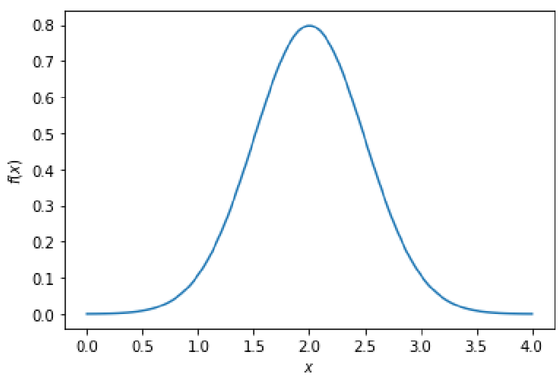
In [7]:
```
x = np.linspace(0,4,100)
cdf = norm_rv.cdf(x) # функция может принимать и вектор (x)
plt.plot(x, cdf)
plt.ylabel('$F(x)$')
plt.xlabel('$x$')
plt.show()
```



In [8]:
```
norm_rv.pdf(3)
```

Out[8]:
```
0.10798193302637613
```

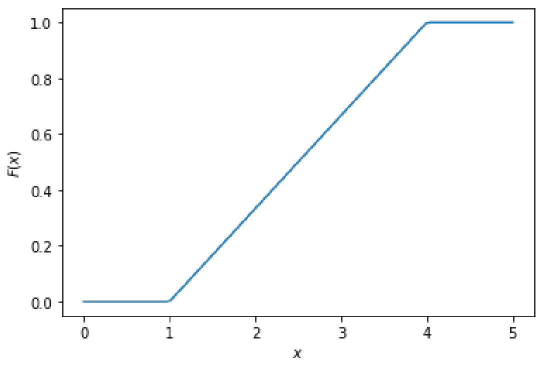
In [10]:
```
x = np.linspace(0,4,100)
pdf = norm_rv.pdf(x)
plt.plot(x, pdf)

plt.ylabel('$f(x)$')
plt.xlabel('$x$')
plt.show()
```



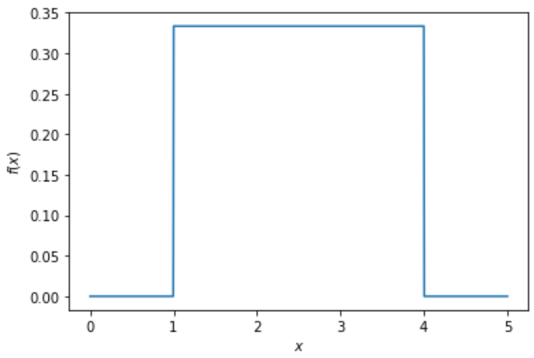
Равномерное распределение на отрезке

```
In [11]: a = 1
         b = 4

         # обратите внимание, что в этой функции задается левая граница и масштаб, а не левая и правая границы:
         uniform_rv = sts.uniform(a, b-a)

         uniform_rv.rvs(10)

Out[11]: array([1.2853015 , 1.76217901, 1.4837676 , 3.11608716, 2.48071534,
                2.19564752, 2.26084079, 3.37152618, 3.41740713, 1.85511775])
```

```
In [12]: x = np.linspace(0,5,100)
         cdf = uniform_rv.cdf(x)
         plt.plot(x, cdf)

         plt.ylabel('$F(x)$')
         plt.xlabel('$x$')
         plt.show()
```



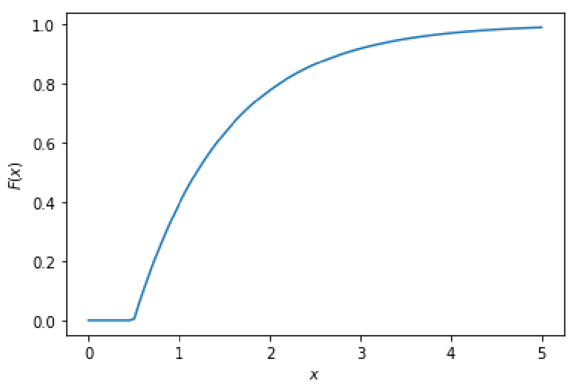

```
In [13]: x = np.linspace(0,5,1000)
         pdf = uniform_rv.pdf(x)
         plt.plot(x, pdf)

         plt.ylabel('$f(x)$')
         plt.xlabel('$x$')
         plt.show()
```



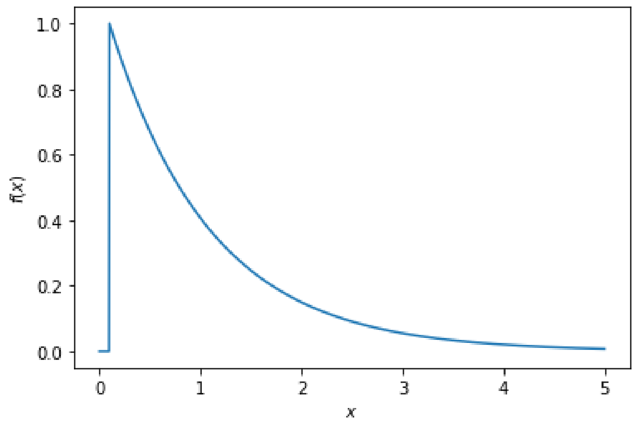
Экспоненциальное распределение

```
In [19]: lam = 0.1

         expon_rv = sts.expon(lam)

         expon_rv.rvs(10)

Out[19]: array([0.17460432, 0.71076815, 0.20756784, 2.00334463, 0.13374002,
                0.88763609, 2.33902922, 0.25857695, 0.35758519, 0.53382361])
```

```
In [20]: x = np.linspace(0,5,100)
         cdf = expron_rv.cdf(x)
         plt.plot(x, cdf)

         plt.ylabel('$F(x)$')
         plt.xlabel('$x$')
         plt.show()
```



In [21]:
```
x = np.linspace(0,5,1000)
pdf = expon_rv.pdf(x)
plt.plot(x, pdf)

plt.ylabel('$f(x)$')
plt.xlabel('$x$')
plt.show()
```

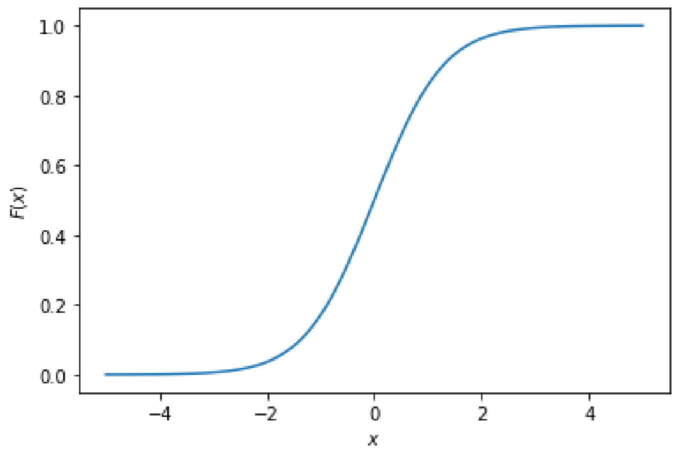


Распределение Стьюдента

In [26]:
```
n = 10

t_rv = sts.t(n)

t_rv.rvs(10)
```

Out[26]:
```
array([-1.79256774, -0.4293779 , -1.12665525, -1.22751042, -1.16395528,
        0.42538999, -0.97521322,  1.30606174, -0.41304665, -0.17785034])
```

In [28]:
```
x = np.linspace(-5,5,100)
cdf = t_rv.cdf(x)
plt.plot(x, cdf)

plt.ylabel('$F(x)$')
plt.xlabel('$x$')
plt.show()
```



In [29]:
```
x = np.linspace(-5,5,1000)
pdf = t_rv.pdf(x)
plt.plot(x, pdf)

plt.ylabel('$f(x)$')
plt.xlabel('$x$')
plt.show()
```