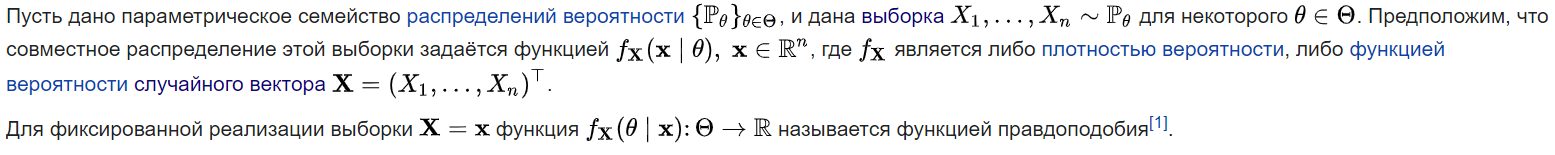
# Machine learning notes

## Параметрическое распределение



Логарифмическая функция правдоподобия

Graphical user interface, text, application, email

Description automatically generated

Пример

Graphical user interface, text

Description automatically generated

Важно понимать, что по абсолютному значению правдоподобия нельзя делать никаких вероятностных суждений. Правдоподобие позволяет сравнить несколько вероятностных распределений с разными параметрами и оценить в контексте какого из них наблюдаемые события наиболее вероятны.

## Теорема Байеса

Graphical user interface, text, application, email

Description automatically generated

Смысл теоремы Байеса

A picture containing diagram

Description automatically generated

## Batch size

The **batch size** defines the number of samples that will be propagated through the network.

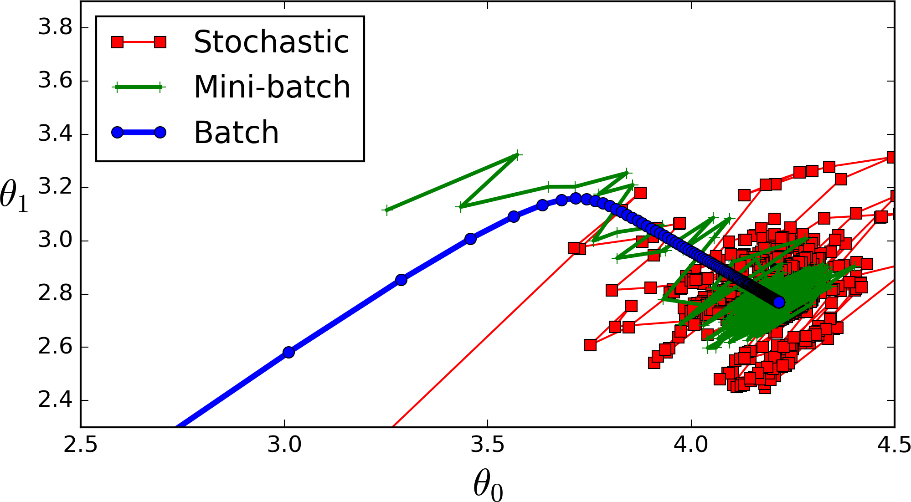
For instance, let's say you have 1050 training samples and you want to set up a batch\_size equal to 100. The algorithm takes the first 100 samples (from 1st to 100th) from the training dataset and trains the network. Next, it takes the second 100 samples (from 101st to 200th) and trains the network again. We can keep doing this procedure until we have propagated all samples through of the network. Problem might happen with the last set of samples. In our example, we've used 1050 which is not divisible by 100 without remainder. The simplest solution is just to get the final 50 samples and train the network.

Advantages of using a batch size < number of all samples:

* It requires less memory. Since you train the network using fewer samples, the overall training procedure requires less memory. That's especially important if you are not able to fit the whole dataset in your machine's memory.
* Typically networks train faster with mini-batches. That's because we update the weights after each propagation. In our example we've propagated 11 batches (10 of them had 100 samples and 1 had 50 samples) and after each of them we've updated our network's parameters. If we used all samples during propagation we would make only 1 update for the network's parameter.

Disadvantages of using a batch size < number of all samples:

* The smaller the batch the less accurate the estimate of the gradient will be. In the figure below, you can see that the direction of the mini-batch gradient (green color) fluctuates much more in comparison to the direction of the full batch gradient (blue color).



## Train, test, validation sets

There is always a need to validate the stability of your machine learning model. I mean you just can’t fit the model to your training data and hope it would accurately work for the real data it has never seen before. **You need some kind of assurance that your model has got most of the patterns from the data correct, and its not picking up too much on the noise, or in other words its low on bias and variance.**

Validation

This process of deciding whether the numerical results quantifying hypothesized relationships between variables, are acceptable as descriptions of the data, is known as **validation**. Generally, an error estimation for the model is made after training, better known as evaluation of residuals. In this process, a numerical estimate of the difference in predicted and original responses is done, also called the training error. However, this only gives us an idea about how well our model does on data used to train it. Now its possible that the model is underfitting or overfitting the data. So, the **problem with this evaluation technique is that it does not give an indication of how well the learner will generalize to an independent/ unseen data set**. **Getting this idea about our model is known as Cross Validation.**

Holdout Method

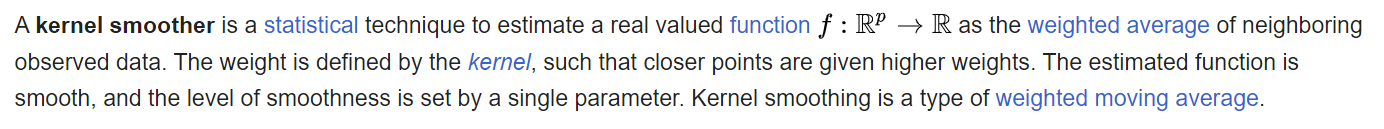
Now a basic remedy for this involves removing a part of the training data and using it to get predictions from the model trained on rest of the data. The error estimation then tells how our model is doing on unseen data or the validation set. **This is a simple kind of cross validation technique, also known as the holdout method**. Although this method doesn’t take any overhead to compute and is better than traditional validation, **it still suffers from issues of high variance**. **This is because it is not certain which data points will end up in the validation set and the result might be entirely different for different sets.**

K-Fold Cross Validation

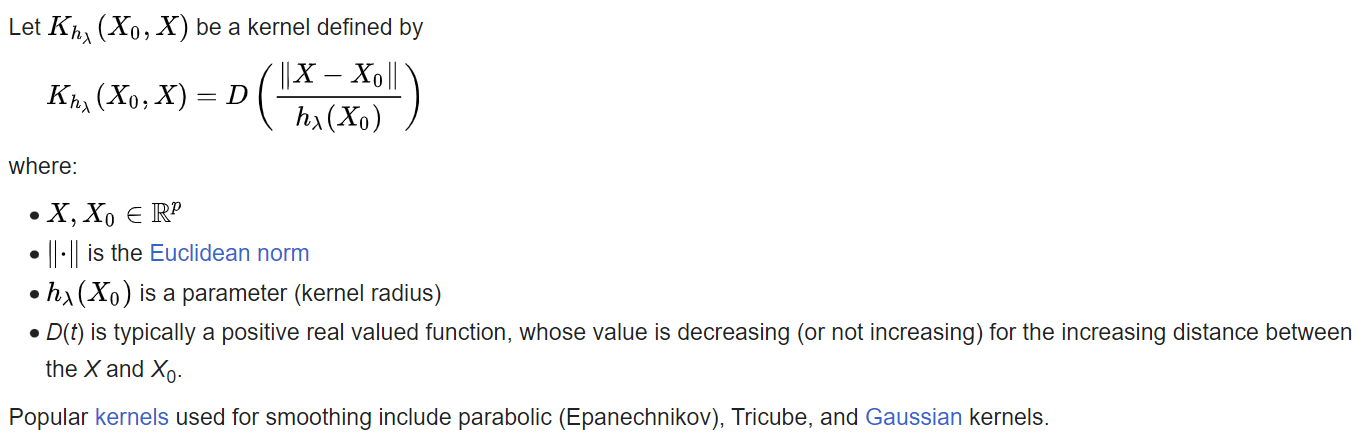
As there is never enough data to train your model, removing a part of it for validation poses a problem of underfitting. **By reducing the training data**, **we risk losing important patterns/ trends in data set, which in turn increases error induced by bias.** So, what we require is a method that provides ample data for training the model and also leaves ample data for validation. K Fold cross validation does exactly that.

In **K Fold cross validation**, the data is divided into k subsets. Now the holdout method is repeated k times, such that **each time, one of the k subsets is used as the test set/ validation set and the other k-1 subsets are put together to form a training set**. The error estimation is averaged over all k trials to get total effectiveness of our model. As can be seen, every data point gets to be in a validation set exactly once, and gets to be in a training set k-1 times. **This significantly reduces bias as we are using most of the data for fitting, and also significantly reduces variance as most of the data is also being used in validation set.** Interchanging the training and test sets also adds to the effectiveness of this method. **As a general rule and empirical evidence, K = 5 or 10 is generally preferred**, but nothing’s fixed and it can take any value.

## Kernel Density Estimator



Definition:

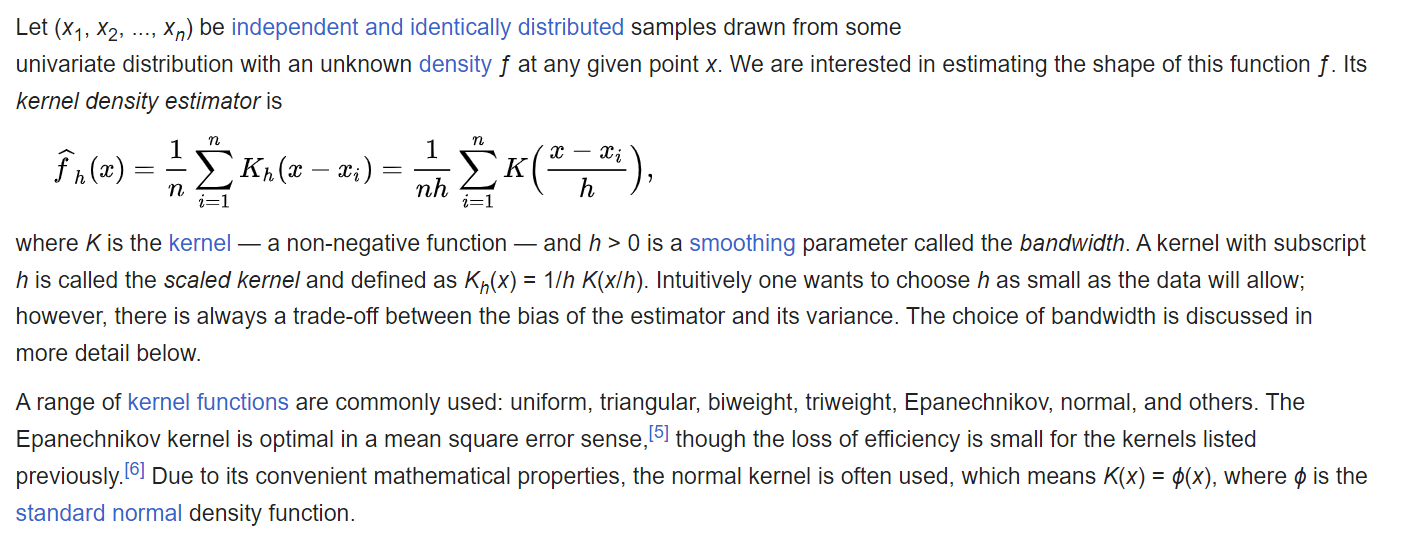


Gaussian kernel smoother:

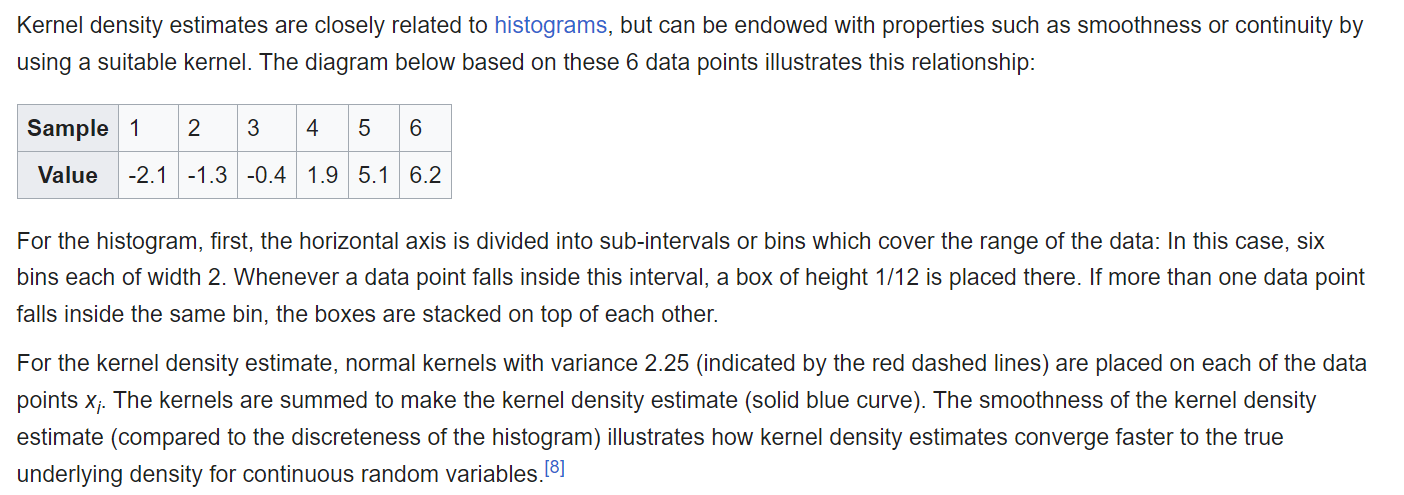
Chart

Description automatically generated

KDE definition:



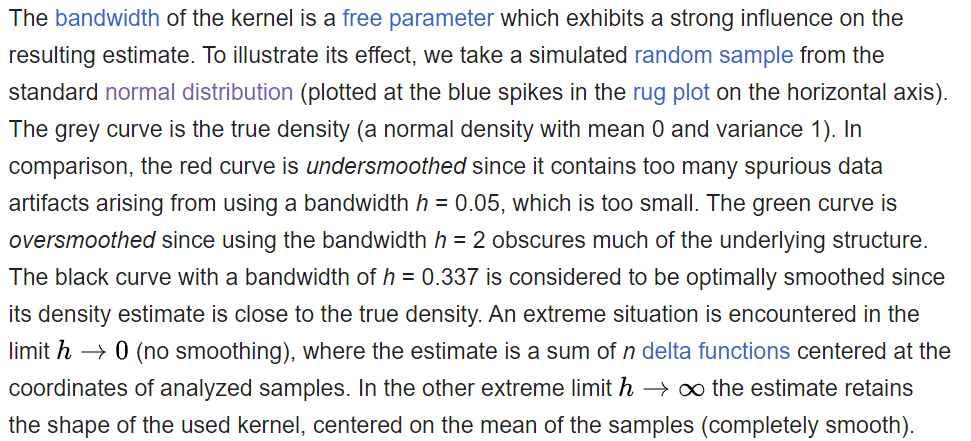
Example:



Chart, histogram

Description automatically generated

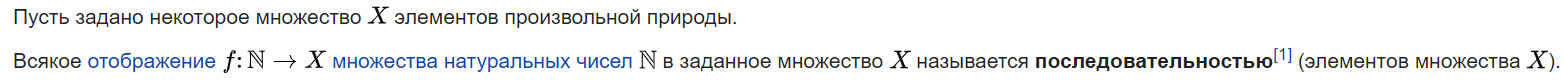
Bandwidth selection:

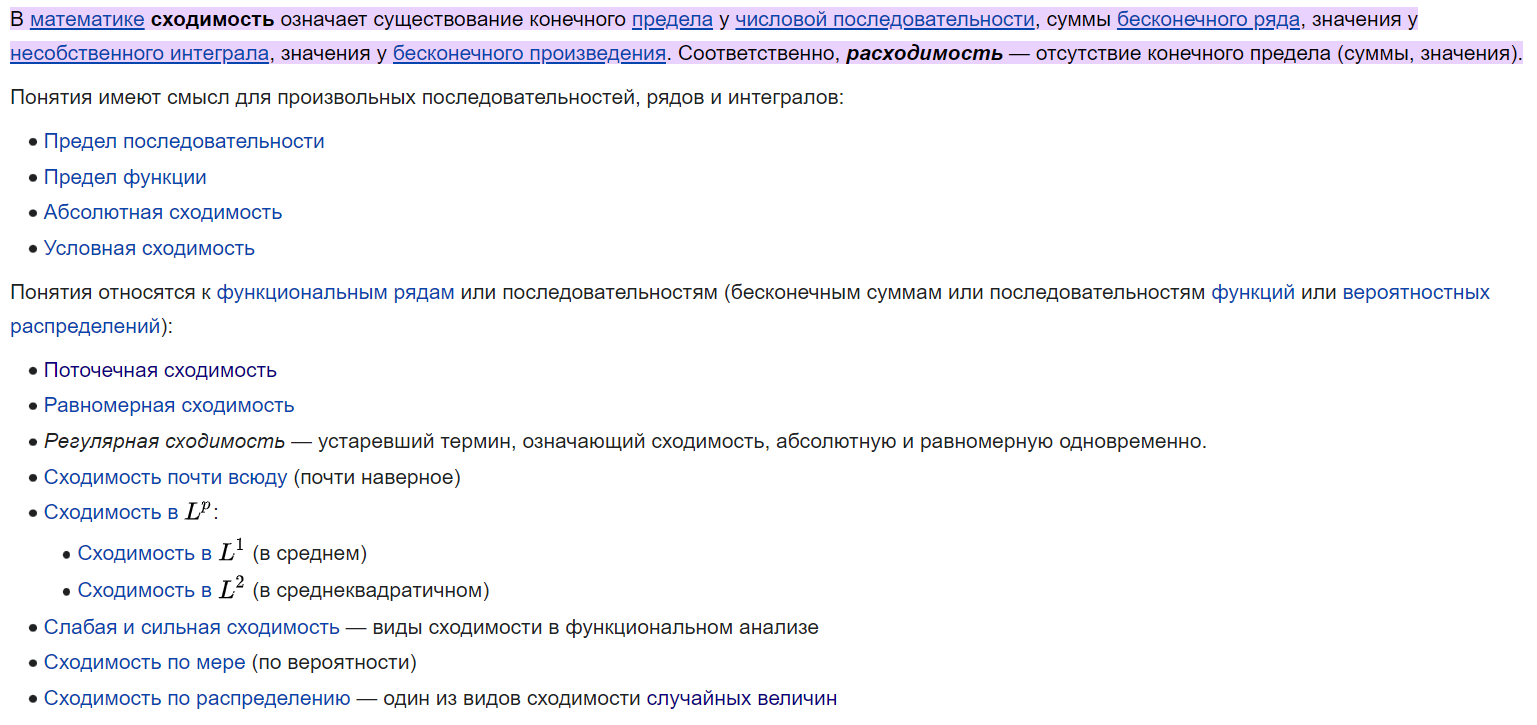


Chart, histogram

Description automatically generated

## Последовательность, сходимость





Предел последовательности:

Text

Description automatically generated

Предел функции:

Graphical user interface, text, application, email

Description automatically generated

Абсолютная сходимость:

A picture containing diagram

Description automatically generated

Поточечная сходимость:

A picture containing application

Description automatically generated

Равномерная сходимость:

Graphical user interface, text, application

Description automatically generated

Сходимость почти всюду:

Graphical user interface, text, application

Description automatically generated

Сходимость в

Graphical user interface, text, application, email

Description automatically generated

Сходимость по мере (по вероятности):

Graphical user interface, text, application, email

Description automatically generated

Сходимость по распределению:

Graphical user interface, text, application

Description automatically generated

## Экспонента, логарифм

Определение:

Application

Description automatically generated with low confidence

Chart, line chart

Description automatically generated

Свойства:

Text

Description automatically generated with medium confidence

Определение логарифма:

Graphical user interface, text, application, email

Description automatically generatedChart

Description automatically generated

Свойства:

Table

Description automatically generated

## Производная

Graphical user interface, text, application, email

Description automatically generated

Text

Description automatically generated with medium confidence

## Сумма ряда, интегрирование

Graphical user interface, text, application

Description automatically generated

Graphical user interface, text, application, email

Description automatically generated

## Случайные величины

Graphical user interface, text, email

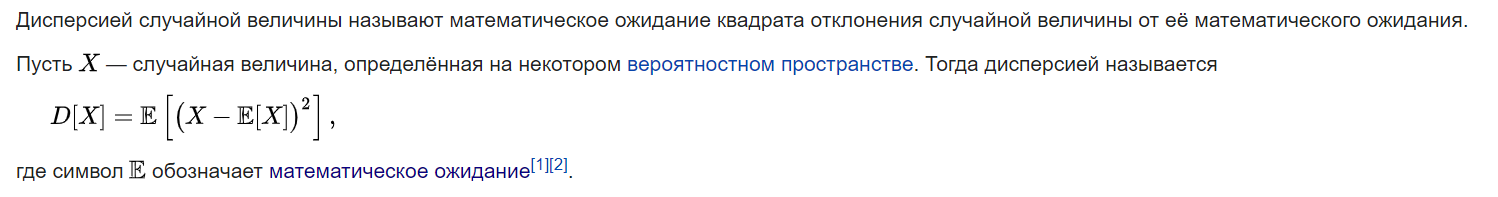
Description automatically generated

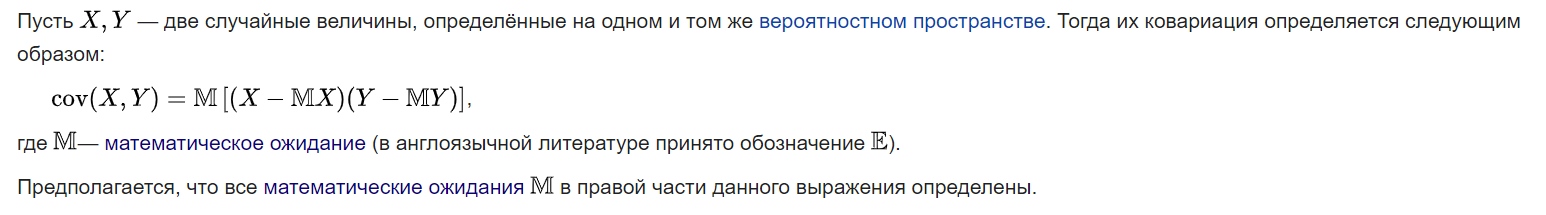
A picture containing diagram

Description automatically generated

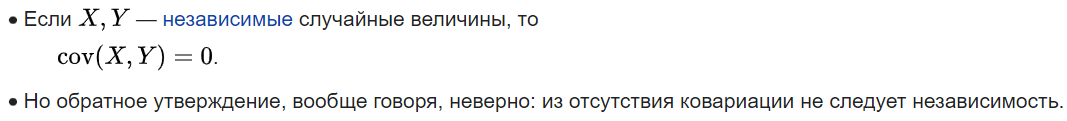
Graphical user interface, text, application, email

Description automatically generated

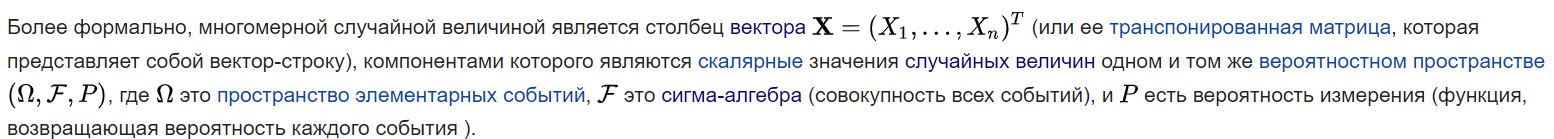




Graphical user interface, text, application

Description automatically generated

## Случайный вектор, ковариационная матрица



Text

Description automatically generated

## Условная верояность, условная независимость

Graphical user interface, text, application, Word

Description automatically generated

Graphical user interface, text, application, email

Description automatically generated

## Law of total expectation

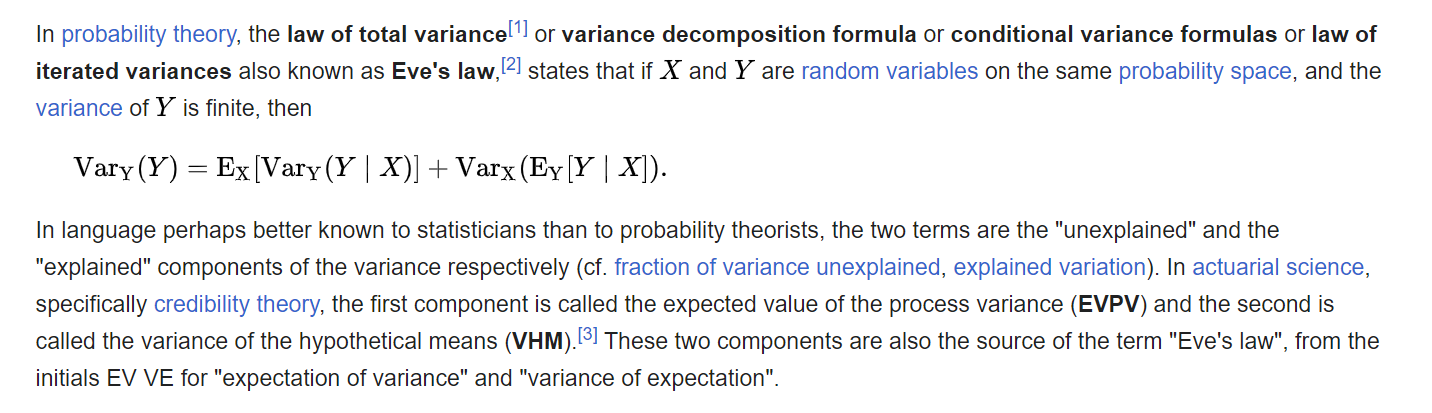
Graphical user interface, text, application

Description automatically generated

Graphical user interface, text, application, email

Description automatically generated

## Law of total variance



## Дельта-функция

Text

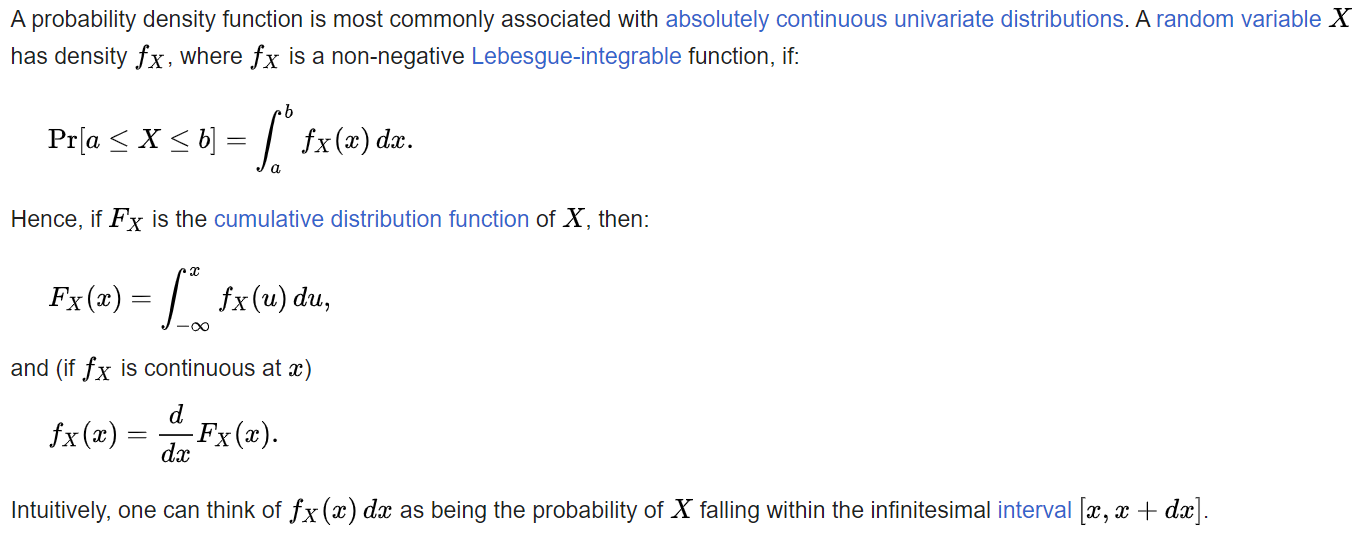
Description automatically generated with low confidence

A picture containing diagram

Description automatically generated

Плотность вероятности вырожденного распределения есть дельта-функция.

## PDF and CDF



## Формула полной вероятности

Graphical user interface, text, application, email

Description automatically generated

## Product rule & chain rule for probability

The product rule of probability can be applied to this phenomenon of the independent transmission of characteristics. It states that the probability of two independent events occurring together can be calculated by multiplying the individual probabilities of each event occurring alone.

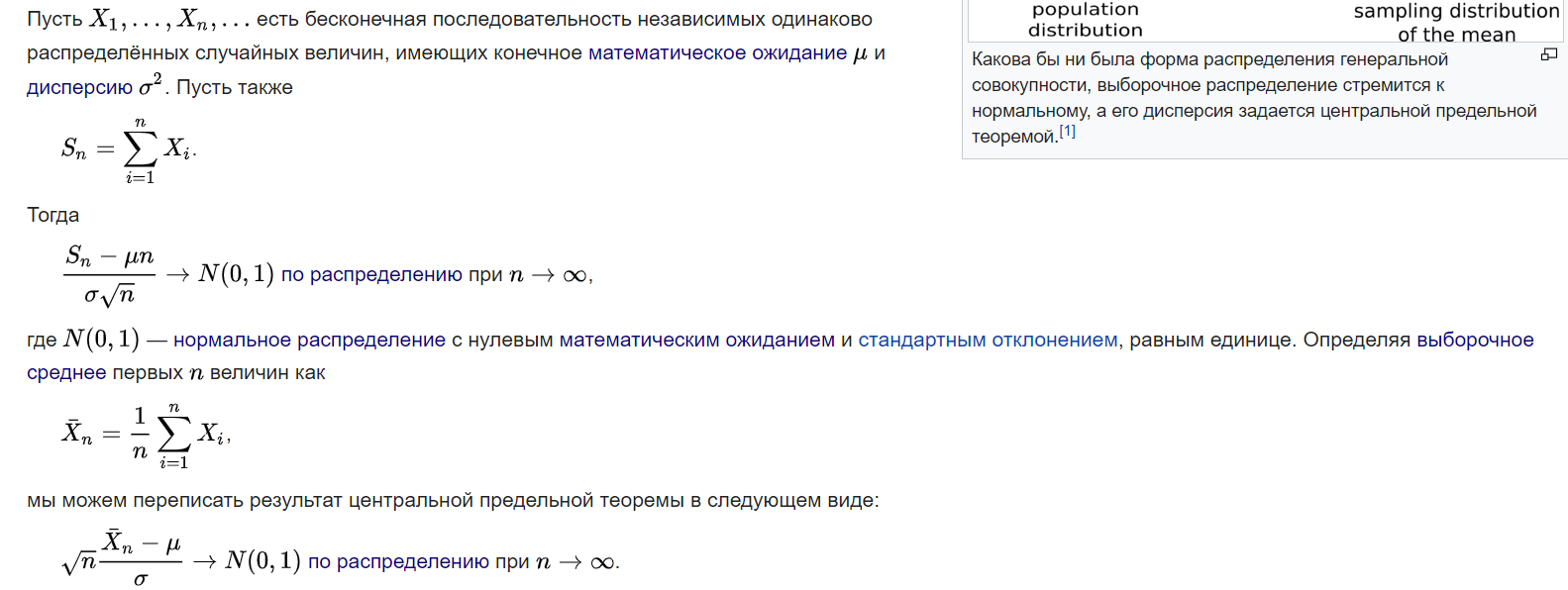
Graphical user interface, text

Description automatically generated

Graphical user interface, text, email

Description automatically generated

## Центральная предельная теорема



## Закон больших чисел

Text

Description automatically generated

Graphical user interface, text, application, email

Description automatically generated

## Квантиль, перцентиль



Graphical user interface, text, application

Description automatically generated

## Аффинное отображение

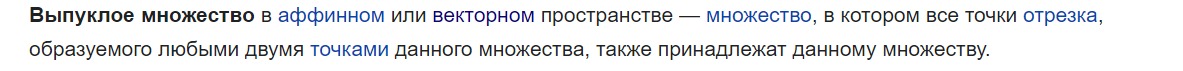
Graphical user interface, text

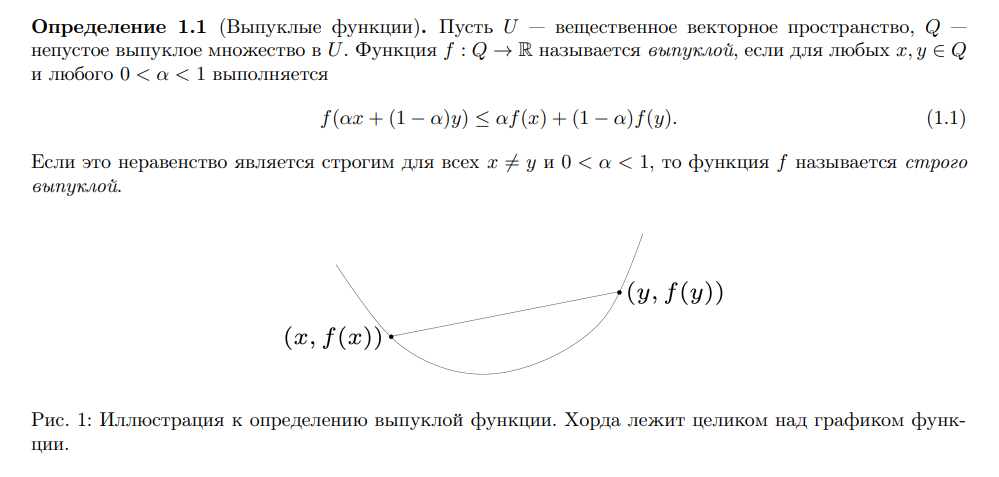
Description automatically generated

Chart

Description automatically generated with medium confidence

## Выпуклые функции



Text

Description automatically generated with medium confidence

Text

Description automatically generated

Text

Description automatically generated

## Квадратичная форма

Определение: 