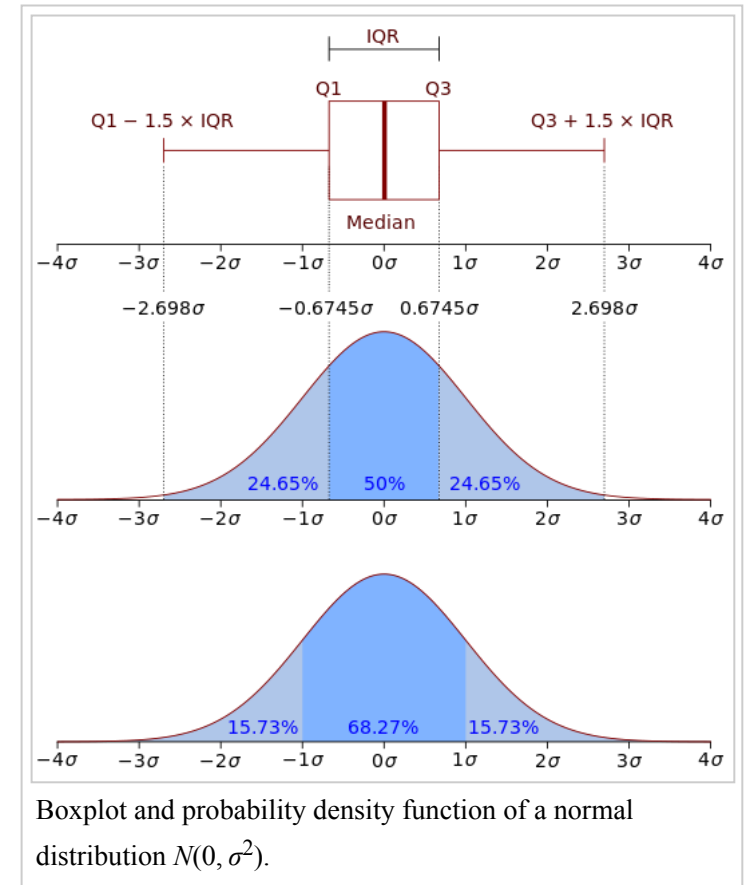# Probability density function

From Wikipedia, the free encyclopedia

In probability theory, a **probability density function** (**PDF**), or **density** of a continuous random variable, is a function, whose value at any given sample (or point) in the sample space (the set of possible values taken by the random variable) can be interpreted as providing a *relative likelihood* that the value of the random variable would equal that sample. In other words, while the *absolute likelihood* for a continuous random variable to take on any particular value is 0 (since there are an infinite set of possible values to begin with), the value of the PDF at two different samples can be used to infer that, in any particular draw of the random variable, how much more likely it is that the random variable would equal one sample compared to the other sample.

In a more precise sense, the PDF is used to specify the probability of the random variable falling *within a particular range of values*, as opposed to taking on any one value. This probability is given by the integral of this variable's PDF over that range—that is, it is given by the area under the density function but above the horizontal axis and between the lowest and greatest values of the range. The probability density function is nonnegative everywhere, and its integral over the entire space is equal to one.
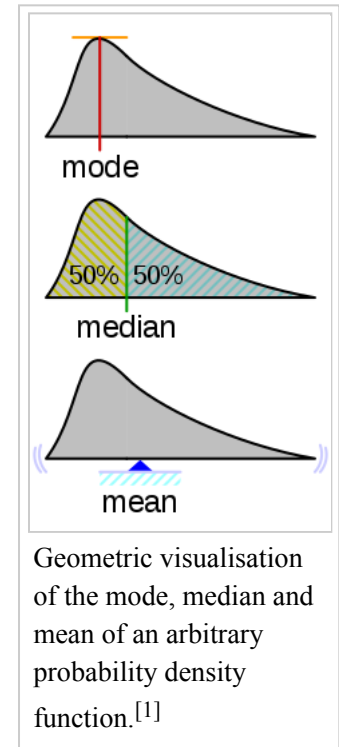
The terms "*probability distribution function*"[2] and "*probability function*"[3] have also sometimes been used to denote the probability density function. However, this use is not standard among probabilists and statisticians. In other sources, "probability distribution function" may be used when the probability distribution is defined as a function over general sets of values, or it may refer to the cumulative distribution function, or it may be a probability mass function (PMF) rather than the density. Further confusion of terminology exists because *density function* has also been used for what is here called the "probability mass function" (PMF).[4] In general though, the PMF is used in the context of discrete random variables (random variables that take values on a discrete set), while PDF is used in the context of continuous random variables.

Boxplot and probability density function of a normal distribution $N(0, \sigma^2)$.

Contents

Geometric visualisation of the mode, median and mean of an arbitrary probability density function.[1]

Example

Suppose a species of bacteria typically lives 4 to 6 hours. What is the probability that a bacterium lives *exactly* 5 hours? The answer is 0%. A lot of bacteria live for *approximately* 5 hours, but there is no chance that any given bacterium dies at *exactly* 5.0000000000... hours.

Instead we might ask: What is the probability that the bacterium dies between 5 hours and 5.01 hours? Let's say the answer is 0.02 (i.e., 2%). Next: What is the probability that the bacterium dies between 5 hours and 5.001 hours? The answer is probably around 0.002, since this is 1/10th of the previous interval. The probability that the bacterium dies between 5 hours and 5.0001 hours is probably about 0.0002, and so on.

In these three examples, the ratio (probability of dying during an interval) / (duration of the interval) is approximately constant, and equal to 2 per hour (or 2 hour^{-1}). For example, there is 0.02 probability of dying in the 0.01-hour interval between 5 and 5.01 hours, and (0.02 probability / 0.01 hours) = 2 hour^{-1}. This quantity 2 hour^{-1} is called the *probability density* for dying at around 5 hours.

Therefore, in response to the question "What is the probability that the bacterium dies at 5 hours?", a literally correct but unhelpful answer is "0", but a better answer can be written as (2 hour^{-1}) *dt*. This is the probability that the bacterium dies within a small (infinitesimal) window of time around 5 hours, where *dt* is the duration of this window.

For example, the probability that it lives longer than 5 hours, but shorter than (5 hours + 1 nanosecond), is (2 hour^{-1})×(1 nanosecond) $\simeq 6\times10^{-13}$ (using the unit conversion 3.6×10^{12} nanoseconds = 1 hour).

There is a *probability density function f* with $f(5 \text{ hours}) = 2 \text{ hour}^{-1}$. The integral of f over any window of time (not only infinitesimal windows but also large windows) is the probability that the bacterium dies in that window.

Absolutely continuous univariate distributions

A probability density function is most commonly associated with absolutely continuous univariate distributions. A random variable X has density f_X, where f_X is a non-negative Lebesgue-integrable function, if:

$$\Pr[a \leq X \leq b] = \int_a^b f_X(x)\, dx.$$

Hence, if F_X is the cumulative distribution function of X, then:

$$F_X(x) = \int_{-\infty}^x f_X(u)\, du,$$

and (if f_X is continuous at x)

$$f_X(x) = \frac{d}{dx} F_X(x).$$

Intuitively, one can think of $f_X(x)\, dx$ as being the probability of X falling within the infinitesimal interval $[x, x + dx]$.

Formal definition

(*This definition may be extended to any probability distribution using the measure-theoretic definition of probability.*)

A random variable X with values in a measurable space $(\mathcal{X}, \mathcal{A})$ (usually $\mathbf{R}^n$ with the Borel sets as measurable subsets) has as probability distribution the measure X_*P on $(\mathcal{X}, \mathcal{A})$: the **density** of X with respect to a reference measure μ on $(\mathcal{X}, \mathcal{A})$ is the Radon–Nikodym derivative:

$$f = \frac{dX_*P}{d\mu}.$$

That is, f is any measurable function with the property that:

$$\Pr[X \in A] = \int_{X^{-1}A} dP = \int_A f \, d\mu$$

for any measurable set $A \in \mathcal{A}$.

Discussion

In the continuous univariate case above, the reference measure is the Lebesgue measure. The probability mass function of a discrete random variable is the density with respect to the counting measure over the sample space (usually the set of integers, or some subset thereof).

Note that it is not possible to define a density with reference to an arbitrary measure (e.g. one can't choose the counting measure as a reference for a continuous random variable). Furthermore, when it does exist, the density is almost everywhere unique.

Further details

Unlike a probability, a probability density function can take on values greater than one; for example, the uniform distribution on the interval [0, ½] has probability density $f(x) = 2$ for $0 \le x \le \frac{1}{2}$ and $f(x) = 0$ elsewhere.

The standard normal distribution has probability density

$$f(x) = \frac{1}{\sqrt{2\pi}} \, e^{-x^2/2}.$$

If a random variable X is given and its distribution admits a probability density function f, then the expected value of X (if the expected value exists) can be calculated as

$$\mathbf{E}[X] = \int_{-\infty}^{\infty} x \, f(x) \, dx.$$

Not every probability distribution has a density function: the distributions of discrete random variables do not; nor does the Cantor distribution, even though it has no discrete component, i.e., does not assign positive probability to any individual point.

A distribution has a density function if and only if its cumulative distribution function $F(x)$ is absolutely continuous. In this case: F is almost everywhere differentiable, and its derivative can be used as probability density:

$$\frac{d}{dx} F(x) = f(x).$$

If a probability distribution admits a density, then the probability of every one-point set $\{a\}$ is zero; the same holds for finite and countable sets.

Two probability densities f and g represent the same probability distribution precisely if they differ only on a set of Lebesgue measure zero.

In the field of statistical physics, a non-formal reformulation of the relation above between the derivative of the cumulative distribution function and the probability density function is generally used as the definition of the probability density function. This alternate definition is the following:

If dt is an infinitely small number, the probability that X is included within the interval $(t, t + dt)$ is equal to $f(t)\, dt$, or:

$$\Pr(t < X < t + dt) = f(t)\, dt.$$

Link between discrete and continuous distributions

It is possible to represent certain discrete random variables as well as random variables involving both a continuous and a discrete part with a generalized probability density function, by using the Dirac delta function. For example, let us consider a binary discrete random variable having the Rademacher distribution—that is, taking −1 or 1 for values, with probability ½ each. The density of probability associated with this variable is:

$$f(t) = \frac{1}{2}(\delta(t + 1) + \delta(t - 1)).$$

More generally, if a discrete variable can take n different values among real numbers, then the associated probability density function is:

$$f(t) = \sum_{i=1}^{n} p_i\, \delta(t - x_i),$$

where $x_1, \ldots, x_n$ are the discrete values accessible to the variable and $p_1, \ldots, p_n$ are the probabilities associated with these values.

This substantially unifies the treatment of discrete and continuous probability distributions. For instance, the above expression allows for determining statistical characteristics of such a discrete variable (such as its mean, its variance and its kurtosis), starting from the formulas given for a continuous distribution of the probability.

Families of densities

It is common for probability density functions (and probability mass functions) to be parametrized—that is, to be characterized by unspecified parameters. For example, the normal distribution is parametrized in terms of the mean and the variance, denoted by μ and σ^2 respectively, giving the family of densities

$$f(x; \mu, \sigma^2) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}.$$

It is important to keep in mind the difference between the domain of a family of densities and the parameters of the family. Different values of the parameters describe different distributions of different random variables on the same sample space (the same set of all possible values of the variable); this sample space is the domain of the family of random variables that this family of distributions describes. A given set of parameters describes a single distribution within the family sharing the functional form of the density. From the perspective of a given distribution, the parameters are constants, and terms in a density function that contain only parameters, but not variables, are part of the normalization factor of a distribution (the multiplicative factor that ensures that the area under the density—the probability of *something* in the domain occurring— equals 1). This normalization factor is outside the kernel of the distribution.

Since the parameters are constants, reparametrizing a density in terms of different parameters, to give a characterization of a different random variable in the family, means simply substituting the new parameter values into the formula in place of the old ones. Changing the domain of a probability density, however, is trickier and requires more work: see the section below on change of variables.

Densities associated with multiple variables

For continuous random variables $X_1, \ldots, X_n$, it is also possible to define a probability density function associated to the set as a whole, often called **joint probability density function**. This density function is defined as a function of the n variables, such that, for any domain D in the n-dimensional space of the values of the variables $X_1, \ldots, X_n$, the probability that a realisation of the set variables falls inside the domain D is

$$\Pr\left(X_1, \cdots, X_n \in D\right) = \int_D f_{X_1,\ldots,X_n}\left(x_1, \cdots, x_n\right) dx_1 \cdots dx_n.$$

If $F(x_1, \ldots, x_n) = \Pr(X_1 \leq x_1, \ldots, X_n \leq x_n)$ is the cumulative distribution function of the vector $(X_1, \ldots, X_n)$, then the joint probability density function can be computed as a partial derivative

$$f(x) = \left. \frac{\partial^n F}{\partial x_1 \cdots \partial x_n} \right|_x$$

Marginal densities

For $i=1, 2, \ldots,n$, let $f_{X_i}(x_i)$ be the probability density function associated with variable X_i alone. This is called the "marginal" density function, and can be deduced from the probability density associated with the random variables $X_1, \ldots, X_n$ by integrating on all values of the $n-1$ other variables:

$$f_{X_i}\left(x_i\right) = \int f(x_1, \cdots, x_n) \, dx_1 \cdots dx_{i-1} \, dx_{i+1} \cdots dx_n.$$

Independence

Continuous random variables $X_1, \ldots, X_n$ admitting a joint density are all independent from each other if and only if

$$f_{X_1,\ldots,X_n}(x_1,\cdots,x_n) = f_{X_1}(x_1)\cdots f_{X_n}(x_n).$$

Corollary

If the joint probability density function of a vector of n random variables can be factored into a product of n functions of one variable

$$f_{X_1,\ldots,X_n}(x_1,\cdots,x_n) = f_1(x_1)\cdots f_n(x_n),$$

(where each f_i is not necessarily a density) then the n variables in the set are all independent from each other, and the marginal probability density function of each of them is given by

$$f_{X_i}(x_i) = \frac{f_i(x_i)}{\int f_i(x)\,dx}.$$

Example

This elementary example illustrates the above definition of multidimensional probability density functions in the simple case of a function of a set of two variables. Let us call $\vec{R}$ a 2-dimensional random vector of coordinates (X, Y): the probability to obtain $\vec{R}$ in the quarter plane of positive x and y is

$$\Pr(X > 0, Y > 0) = \int_0^\infty \int_0^\infty f_{X,Y}(x,y)\,dx\,dy.$$

Dependent variables and change of variables

If the probability density function of a random variable X is given as $f_X(x)$, it is possible (but often not necessary; see below) to calculate the probability density function of some variable $Y = g(X)$. This is also called a "change of variable" and is in practice used to generate a random variable of arbitrary shape $f_{g(X)} = f_Y$ using a known (for instance uniform) random number generator.

If the function g is monotonic, then the resulting density function is

$$f_Y(y) = \left| \frac{d}{dy}(g^{-1}(y)) \right| \cdot f_X(g^{-1}(y)).$$

Here g^{-1} denotes the inverse function.

This follows from the fact that the probability contained in a differential area must be invariant under change of variables. That is,

$$|f_Y(y)\,dy| = |f_X(x)\,dx|,$$

or

$$f_Y(y) = \left|\frac{dx}{dy}\right| f_X(x) = \left|\frac{d}{dy}(x)\right| f_X(x) = \left|\frac{d}{dy}(g^{-1}(y))\right| f_X(g^{-1}(y)) = \frac{f_X(g^{-1}(y))}{|g'(g^{-1}(y))|}.$$

For functions which are not monotonic the probability density function for y is

$$\sum_{k=1}^{n(y)} \left|\frac{d}{dy} g_k^{-1}(y)\right| \cdot f_X(g_k^{-1}(y))$$

where $n(y)$ is the number of solutions in x for the equation $g(x) = y$, and $g^{-1}{}_k(y)$ are these solutions.

It is tempting to think that in order to find the expected value $E(g(X))$ one must first find the probability density $f_{g(X)}$ of the new random variable $Y = g(X)$. However, rather than computing

$$E(g(X)) = \int_{-\infty}^{\infty} y f_{g(X)}(y)\,dy,$$

one may find instead

$$E(g(X)) = \int_{-\infty}^{\infty} g(x) f_X(x)\,dx.$$

The values of the two integrals are the same in all cases in which both X and $g(X)$ actually have probability density functions. It is not necessary that g be a one-to-one function. In some cases the latter integral is computed much more easily than the former. See Law of the unconscious statistician.

Multiple variables

The above formulas can be generalized to variables (which we will again call y) depending on more than one other variable. $f(x_1, \ldots, x_n)$ shall denote the probability density function of the variables that y depends on, and the dependence shall be $y = g(x_1, \ldots, x_n)$. Then, the resulting density function is

$$\int_{r=g(x_1,\cdots,x_n)} \frac{f(x_1,\cdots,x_n)}{\sqrt{\sum_{j=1}^{n} \frac{\partial g}{\partial x_j}(x_1,\cdots,x_n)^2}}\, dV$$

where the integral is over the entire (n-1)-dimensional solution of the subscripted equation and the symbolic dV must be replaced by a parametrization of this solution for a particular calculation; the variables x_1, ..., x_n are then of course functions of this parametrization.

This derives from the following, perhaps more intuitive representation: Suppose x is an n-dimensional random variable with joint density f. If $y = H(x)$, where H is a bijective, differentiable function, then y has density g:

$$g(\mathbf{y}) = f(\mathbf{x}) \left| \det\left(\frac{d\mathbf{x}}{d\mathbf{y}} \right) \right|$$

with the differential regarded as the Jacobian of the inverse of H, evaluated at y.

For example, in the 2-dimensional case $\mathbf{x}=(x_1, x_2)$, suppose the transform H is given as $y_1=H_1(x_1,x_2)$, $y_2=H_2(x_1,x_2)$ with inverses $x_1=H_1^{-1}(y_1,y_2)$, $x_2=H_2^{-1}(y_1,y_2)$. The joint distribution for $\mathbf{y}=(y_1, y_2)$ has density[5]

$$g(y_1, y_2) = f_{X_1,X_2}\left(H_1^{-1}(y_1,y_2), H_2^{-1}(y_1,y_2)\right) \left| \frac{\partial H_1^{-1}}{\partial y_1} \frac{\partial H_2^{-1}}{\partial y_2} - \frac{\partial H_1^{-1}}{\partial y_2} \frac{\partial H_2^{-1}}{\partial y_1} \right|$$

Using the delta-function (and assuming independence) the same result is formulated as follows.

If the probability density function of independent random variables X_i, $i = 1, 2, \ldots n$ are given as $f_{X_i}(x_i)$, it is possible to calculate the probability density function of some variable $Y = G(X_1, X_2, \ldots X_n)$. The following formula establishes a connection between the probability density function of Y denoted by $f_Y(y)$ and $f_{X_i}(x_i)$ using the Dirac delta function:

$$f_Y(y) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f_{X_1}(x_1) f_{X_2}(x_2) \cdots f_{X_n}(x_n) \delta(y - G(x_1, x_2, \cdots, x_n))\, dx_1\, dx_2 \cdots dx_n$$

Sums of independent random variables

Not to be confused with Mixture distribution

The probability density function of the sum of two independent random variables U and V, each of which has a probability density function, is the convolution of their separate density functions:

$$f_{U+V}(x) = \int_{-\infty}^{\infty} f_U(y) f_V(x-y)\, dy = (f_U * f_V)(x)$$

It is possible to generalize the previous relation to a sum of N independent random variables, with densities $U_1, \ldots, U_N$:

$$f_{U_1+\cdots+U_N}(x) = (f_{U_1} * \cdots * f_{U_N})(x)$$

This can be derived from a two-way change of variables involving $Y=U+V$ and $Z=V$, similarly to the example below for the quotient of independent random variables.

Products and quotients of independent random variables

Given two independent random variables U and V, each of which has a probability density function, the density of the product $Y=UV$ and quotient $Y=U/V$ can be computed by a change of variables.

Example: Quotient distribution

To compute the quotient $Y=U/V$ of two independent random variables U and V, define the following transformation:

$$Y = U/V$$
$$Z = V$$

Then, the joint density $p(Y,Z)$ can be computed by a change of variables from U,V to Y,Z, and Y can be derived by marginalizing out Z from the joint density.

The inverse transformation is

$$U = YZ$$
$$V = Z$$

The Jacobian matrix $J(U, V | Y, Z)$ of this transformation is

$$\begin{vmatrix} \frac{\partial U}{\partial Y} & \frac{\partial U}{\partial Z} \\ \frac{\partial V}{\partial Y} & \frac{\partial V}{\partial Z} \end{vmatrix} = \begin{vmatrix} Z & Y \\ 0 & 1 \end{vmatrix} = |Z|.$$

Thus:

$$p(Y, Z) = p(U, V)\, J(U, V | Y, Z) = p(U)\, p(V)\, J(U, V | Y, Z) = p_U(YZ)\, p_V(Z)\, |Z|.$$

And the distribution of Y can be computed by marginalizing out Z:

$$p(Y) = \int_{-\infty}^{\infty} p_U(YZ)\, p_V(Z)\, |Z|\, dZ$$

Note that this method crucially requires that the transformation from U, V to Y, Z be bijective. The above transformation meets this because Z can be mapped directly back to V, and for a given V the quotient U/V is monotonic. This is similarly the case for the sum $U+V$, difference $U-V$ and product UV.

Exactly the same method can be used to compute the distribution of other functions of multiple independent random variables.

Example: Quotient of two standard normals

Given two standard normal variables U and V, the quotient can be computed as follows. First, the variables have the following density functions:

$$p(U) = \frac{1}{\sqrt{2\pi}} e^{-\frac{U^2}{2}}$$

$$p(V) = \frac{1}{\sqrt{2\pi}} e^{-\frac{V^2}{2}}$$

We transform as described above:

$$Y = U/V$$
$$Z = V$$

This leads to:

$$p(Y) = \int_{-\infty}^{\infty} p_U(YZ)\, p_V(Z)\, |Z|\, dZ$$

$$= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}Y^2 Z^2} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}Z^2} |Z|\, dZ$$

$$= \int_{-\infty}^{\infty} \frac{1}{2\pi} e^{-\frac{1}{2}(Y^2+1)Z^2} |Z|\, dZ$$

$$= 2 \int_{0}^{\infty} \frac{1}{2\pi} e^{-\frac{1}{2}(Y^2+1)Z^2} Z\, dZ$$

$$= \int_{0}^{\infty} \frac{1}{\pi} e^{-(Y^2+1)u}\, du \qquad\qquad u = \tfrac{1}{2} Z^2$$

$$= -\frac{1}{\pi(Y^2+1)} e^{-(Y^2+1)u} \bigg]_{u=0}^{\infty}$$

$$= \frac{1}{\pi(Y^2+1)}$$

This is a standard Cauchy distribution.

See also

- Density estimation
- Likelihood function
- List of probability distributions
- Probability mass function
- Secondary measure

- Uses as *position probability density*:
  - Atomic orbital
  - Home range

Bibliography

- Pierre Simon de Laplace (1812). *Analytical Theory of Probability*.

    The first major treatise blending calculus with probability theory, originally in French: *Théorie Analytique des Probabilités*.

- Andrei Nikolajevich Kolmogorov (1950). *Foundations of the Theory of Probability*.

The modern measure-theoretic foundation of probability theory; the original German version (*Grundbegriffe der Wahrscheinlichkeitsrechnung*) appeared in 1933.

- Patrick Billingsley (1979). *Probability and Measure*. New York, Toronto, London: John Wiley and Sons. ISBN 0-471-00710-2.

- David Stirzaker (2003). *Elementary Probability*. ISBN 0-521-42028-8.

    Chapters 7 to 9 are about continuous variables.

1. "AP Statistics Review - Density Curves and the Normal Distributions" (http://apstatsreview.tumblr.com/post/50058615236/density-curves-and-the-normal-distributions?action=purge). Retrieved 16 March 2015.
2. Probability distribution function (http://planetmath.org/?method=png&from=objects&id=2884&op=getobj) PlanetMath
3. Probability Function (http://mathworld.wolfram.com/ProbabilityFunction.html) at MathWorld
4. Ord, J.K. (1972) *Families of Frequency Distributions*, Griffin. ISBN 0-85264-137-0 (for example, Table 5.1 and Example 5.4)
5. David., Stirzaker, (2007-01-01). *Elementary probability* (https://www.worldcat.org/oclc/851313783). Cambridge University Press. ISBN 0521534283. OCLC 851313783 (https://www.worldcat.org/oclc/851313783).

External links

- Ushakov, N.G. (2001), "Density of a probability distribution" (https://www.encyclopediaofmath.org/index.php?title=D/d031110), in Hazewinkel, Michiel, *Encyclopedia of Mathematics*, Springer, ISBN 978-1-55608-010-4
- Weisstein, Eric W. "Probability density function" (http://mathworld.wolfram.com/ProbabilityDensityFunction.html). *MathWorld*.