

A (real) random variable x is defined by its *probability density function* (or *probability distribution function*) $p(x) \geq 0$.

The probability for x to be in the interval $[x, x + dx]$ is $p(x)dx$. So the probability that x is in the interval $[a, b]$ is $\int_a^b p(x)dx$. The random variable must be in $] - \infty, \infty[$ meaning that the probability to be in that infinite interval is one: $\int_{-\infty}^{\infty} p(x)dx = 1$.

Those two properties $p(x) \geq 0$ and $\int_{-\infty}^{\infty} p(x)dx = 1$ are the defining properties of a pdf. A function that does not satisfy those cannot be a pdf.

$$\text{Mean} = \bar{x} = \int_{-\infty}^{\infty} x p(x) dx$$

$$\text{Variance} = \int_{-\infty}^{\infty} (x - \bar{x})^2 p(x) dx$$

Examples:

Uniform distribution:

$$p(x) = 1 \text{ if } 0 < x < 1, p(x) = 0 \text{ otherwise}$$

Mean=0.5, Standard deviation (STD)=sqrt(variance)= $1/(2\sqrt{3})$. Matlab's **rand**.

Gaussian distribution:

$$p(x) = \frac{e^{-(x-x_0)^2/2\sigma^2}}{\sqrt{2\pi}\sigma}$$

Mean= x_0 , variance= σ^2 , STD= σ . Matlab's **randn** gives the normalized distribution with mean=0 and STD=1. To get mean x_0 and std σ , use $x = x_0 + \sigma * \text{randn}$.

The probability to be in a certain interval $[a, b]$ can be reduced to an integral of $\exp(-u^2)$ by the change of variables $x = x_0 + \sqrt{2}\sigma u$. That integral is directly related to the *Error function* which is a Matlab built-in special function (**erf**, see “help erf”, or “help specfun”). The error function is defined as

$$\text{erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-u^2} du.$$

This integral cannot be done analytically. Matlab computes it numerically.

For example, the probability to be within 2σ from the mean is

$$P(|x - x_0| \leq 2\sigma) = \int_{x_0-2\sigma}^{x_0+2\sigma} \frac{e^{-(x-x_0)^2/2\sigma^2}}{\sqrt{2\pi}\sigma} dx = \frac{1}{\sqrt{\pi}} \int_{-\sqrt{2}}^{\sqrt{2}} e^{-u^2} du = \text{erf}(\sqrt{2}) \approx 0.9545.$$

This is a noteworthy fact about Gaussian distributions: there is about 95% chance to be within 2 STD of the mean.