Cumulative Distribution Functions

The cumulative density function $F(x)$ is the probability that the RV X is less than the number x.

$$F(a) = Prob(X \leq a)$$
$$F(b) = Prob(X \leq b)$$

If $a < b$ then $Prob(a < X \leq b) = F(b) - F(a)$

Continuous Density Function

If $F(x)$ is a continuous cumulative density function (i.e. $F(-\infty) = 0,\ F(\infty) = 1$, and F is monotonically increasing) then

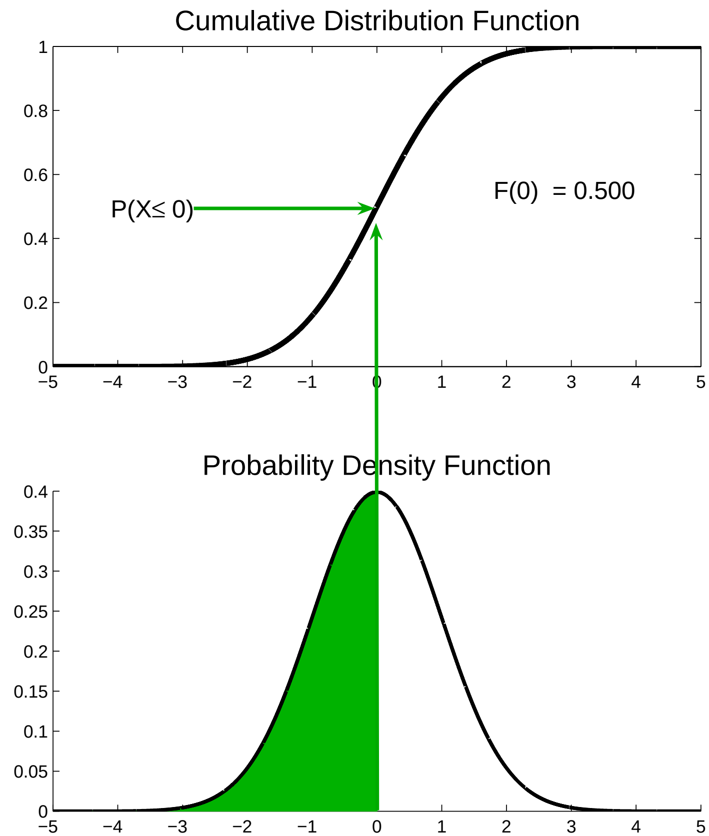
$$f(x) = \frac{dF(x)}{dx}$$

is the continuous density function.
f exists by the Fundamental Theorem of Calculus.
Also,

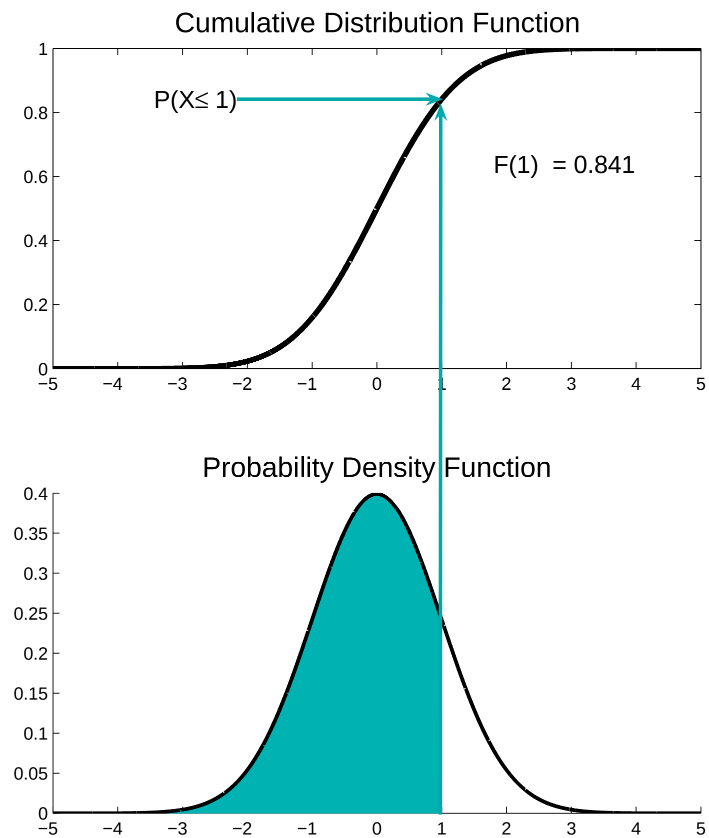
$$P(a < X \leq b) = \int_a^b f(x)dx = F(b) - F(a)$$

Also by the Fundamental Theorem of Calculus.

Example

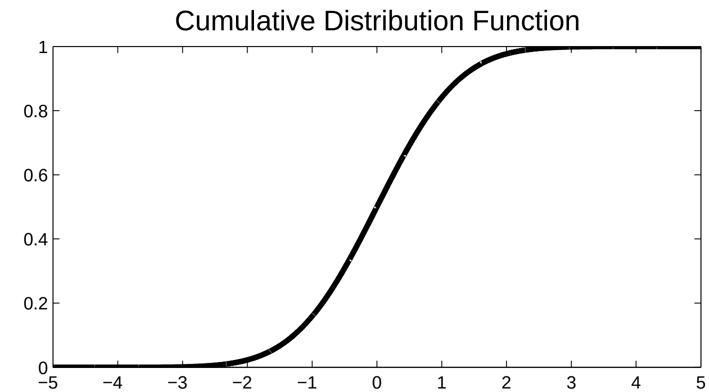

Cumulative Distribution Function
P(X≤ 0)
F(0) = 0.500
Probability Density Function
1
0.8
0.6
0.4
0.2
0
0.35
0.3
0.25
0.15
0.1
0.05
−5
−4
−3
−2
−1
1
2
3
4
5

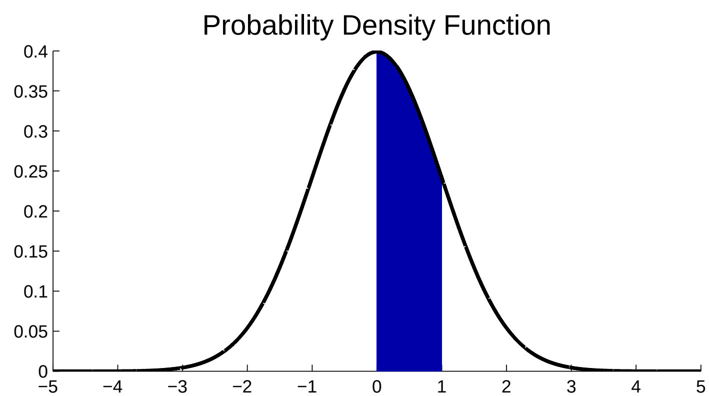
Example



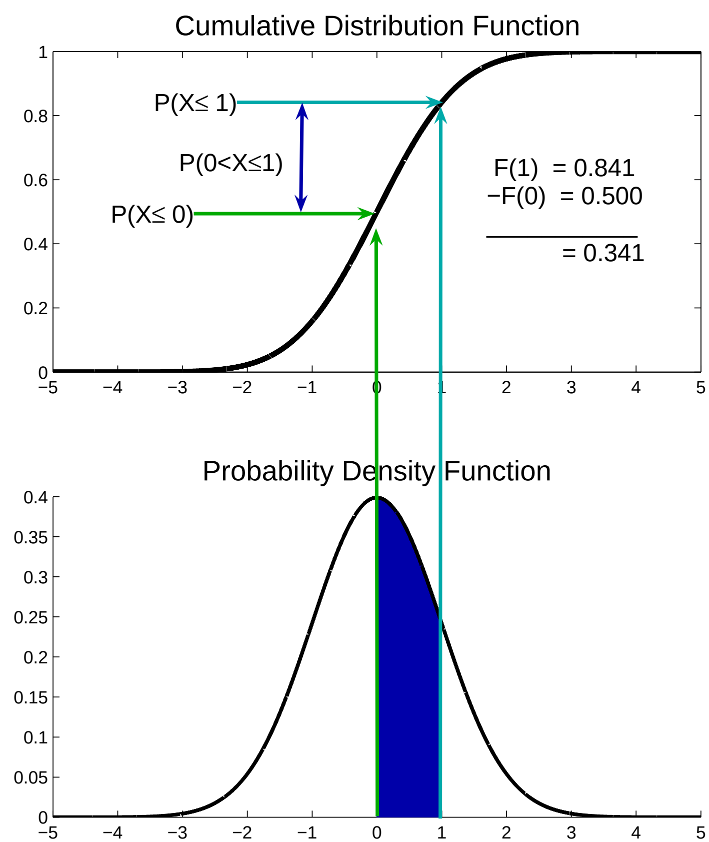
Cumulative Distribution Function
P(X≤ 1)
F(1) = 0.841
1
0.8
0.6
0.4
0.2
0
−5
−4
−3
−2
−1
0
1
2
3
4
5
Probability Density Function
0.4
0.35
0.3
0.25
0.2
0.15
0.1
0.05
0

Example


Cumulative Distribution Function
1
0.8
0.6
0.4
0.2
0
−5 −4 −3 −2 −1 0 1 2 3 4 5


Probability Density Function
0.4
0.35
0.3
0.25
0.2
0.15
0.1
0.05
0
−5 −4 −3 −2 −1 0 1 2 3 4 5

Example


Cumulative Distribution Function
P(X≤ 1)
P(0<X≤1)
P(X≤ 0)
F(1) = 0.841
−F(0) = 0.500
= 0.341
1
0.8
0.6
0.4
0.2
0
−5 −4 −3 −2 −1 0 1 2 3 4 5
Probability Density Function
0.4
0.35
0.3
0.25
0.2
0.15
0.1
0.05
0
−5 −4 −3 −2 −1 0 1 2 3 4 5

CDF-PDF relation

$$f(x) = \frac{dF(x)}{dx}$$

$$P(a < X \leq b) = \int_a^b f(x)dx = F(b) - F(a)$$

Note, the PDF does not strictly give the probability
$P(X = a) = P(a \leq X \leq a) = \int_a^a f(x)dx = F(a) - F(a) = 0!$

Interpretation: The probability of a small interval around x is proportional to $f(x)$
$P(a < X < a + dx) = F(a + dx) - F(a) = f(x)dx$
Imagine centering a small rectangle around $f(x)$
The area of the rectangle is proportional to $f(x)$

A few questions

We know that $F(-\infty) = 0, F(\infty) = 1$ and F is monotonically nondecreasing.

- What is $\int_{-\infty}^{\infty} f(x)dx$?

- Is $f(x)$ positive, negative or perhaps both?

A few questions

We know that $F(-\infty) = 0, F(\infty) = 1$ and F is monotonically nondecreasing.

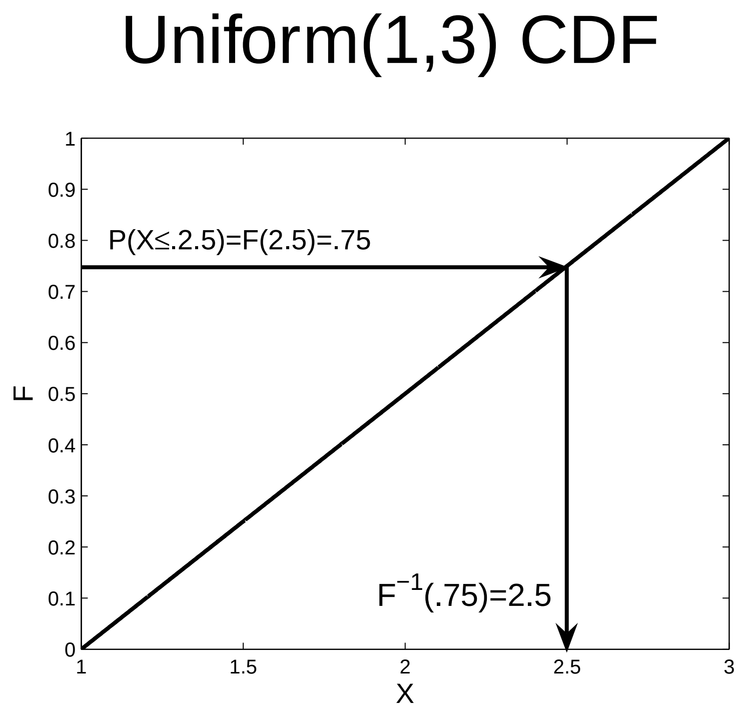
- What is $\int_{-\infty}^{\infty} f(x)dx$?
  $F(\infty) - F(-\infty) = 1 - 0 = 1$
  A probability density function always integrates to 1
- Is $f(x)$ positive, negative or perhaps both?

A few questions

We know that $F(-\infty) = 0, F(\infty) = 1$ and F is monotonically nondecreasing.

- What is $\int_{-\infty}^{\infty} f(x)dx$?
  $F(\infty) - F(-\infty) = 1 - 0 = 1$
  A probability density function always integrates to 1
- Is $f(x)$ positive, negative or perhaps both?
  $f(x) \geq 0$
  $f(x) = \frac{dFx}{dx}$ and the slope of F is always ≥ 0

Quantiles



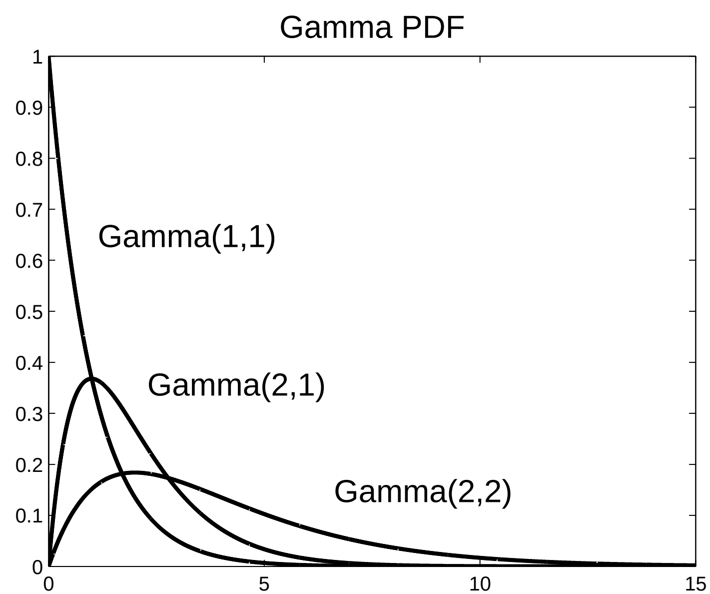
The p-th quantile of F is the number x for which

$$P(X \leq x) = F(X) = p$$

Definition: The 50-th quantile is the median.

Gamma Distribution

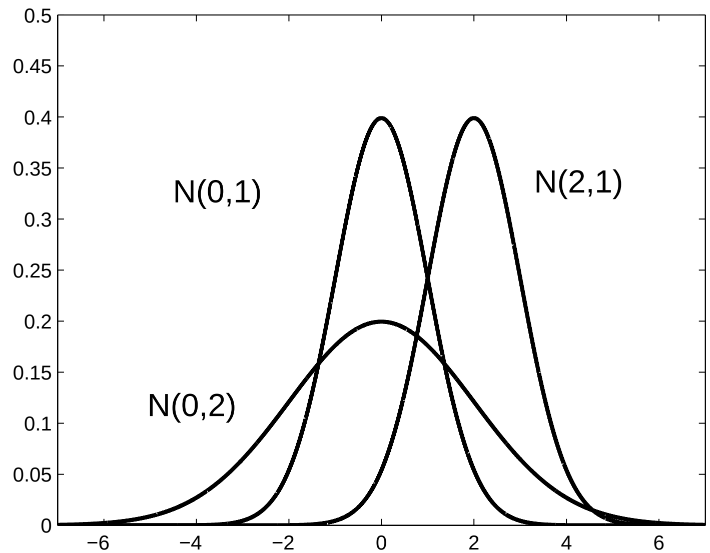
$$f(x) = \frac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\lambda x}$$



Gamma is a very flexible distribution
α is the “shape” parameter
λ is the “scale” parameter

Gaussian Distribution

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



μ = "location" parameter, σ = scale parameter

μ = mean, σ^2 = variance

The not so normal distribution

The mean of large number of random data has a Gaussian distribution (Central Limit Theorem)
Wait, I thought the mean was one number, how can it have a distribution?

- The data is random, so the mean is random
- Collect the data again, and your mean will be different
- These means have a normal distribution

It was at first thought that most phenomena had a Gaussian distribution, hence they were normal
Everything else was an aberration
Assuming data is “normal” can be dangerous

The Standard Normal Distribution

To add insult to injury, the normal density with mean 0 and variance 1 (N(0,1)) is the standard normal density.

The standard normal pdf is denoted by ϕ.
The standard normal cdf is denoted by Φ

Example: $\Phi(0)$ is the probability that a standard normal variable is ≤ 0
$\Phi(0) = F(0) = .5$ Example: $\Phi(-1)$ is the probability that a standard normal variable is ≤ -1
$\Phi(-1) = F(-1) = .16$