Statistics and Machine Learning Toolbox

Analyze and model data using statistics and machine learning

Statistics and Machine Learning Toolbox™ provides functions and apps to describe, analyze, and model data using statistics and machine learning. You can use descriptive statistics and plots for exploratory data analysis, fit probability distributions to data, generate random numbers for Monte Carlo simulations, and perform hypothesis tests. Regression and classification algorithms let you draw inferences from data and build predictive models.

For analyzing multidimensional data, Statistics and Machine Learning Toolbox lets you identify key variables or features that impact your model with sequential feature selection, stepwise regression, principal component analysis, regularization, and other dimensionality reduction methods. The toolbox provides supervised and unsupervised machine learning algorithms, including support vector machines (SVMs), boosted and bagged decision trees, k-nearest neighbor, k-means, k-medoids, hierarchical clustering, Gaussian mixture models, and hidden Markov models.

Examples
Functions and Other Reference
Release Notes
PDF Documentation

Getting Started
Learn the basics of Statistics and Machine Learning Toolbox

Descriptive Statistics and Visualization
Data import and export, descriptive statistics, visualization

Probability Distributions
Data frequency models, random sample generation, parameter estimation

Hypothesis Tests
t-test, F-test, chi-square goodness-of-fit test, and more

Cluster Analysis
Unsupervised learning techniques to find natural groupings and patterns in data

ANOVA
Analysis of variance and covariance, multivariate ANOVA, repeated measures ANOVA

Regression
Linear, generalized linear, nonlinear, and nonparametric techniques for supervised learning

Classification
Supervised learning algorithms for binary and multiclass problems

Dimensionality Reduction
PCA, factor analysis, nonnegative matrix factorization, sequential feature selection, and more

Industrial Statistics
Design of experiments (DOE); survival and reliability analysis; statistical process control
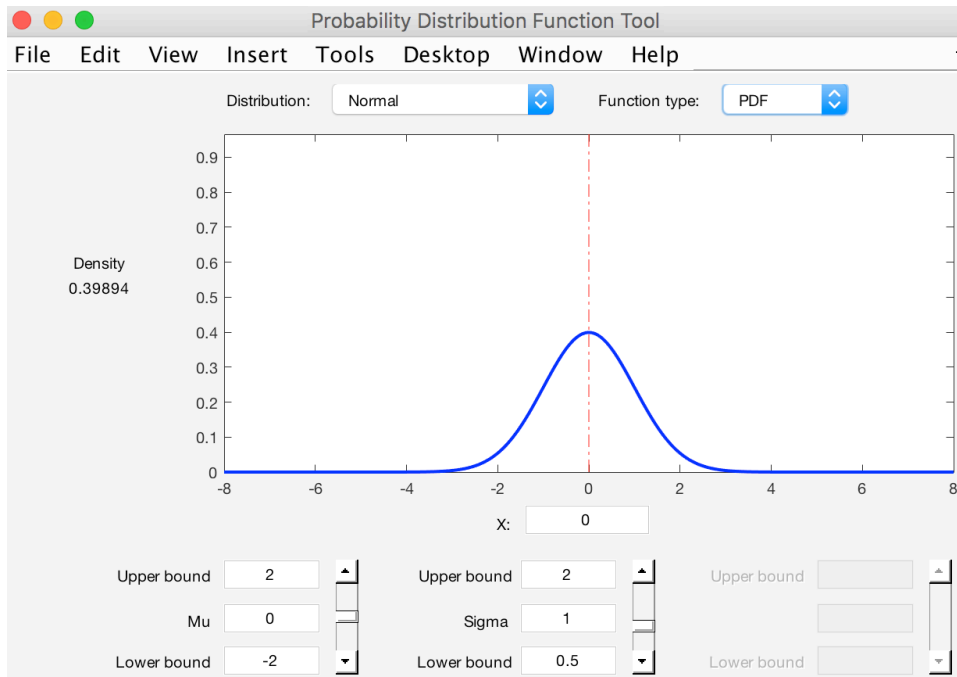
Speed Up Statistical Computations
Parallel or distributed computation of statistical functions
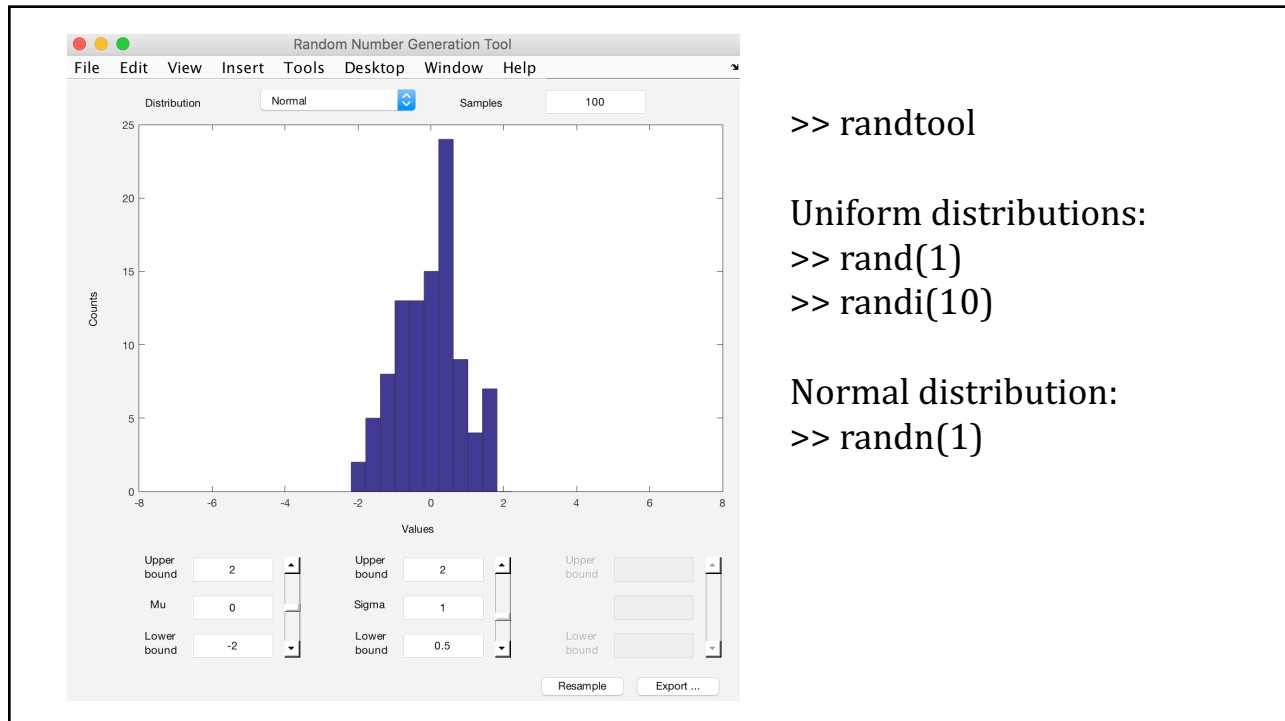

>> disttool

- adjust parameters
- drag red line to view values
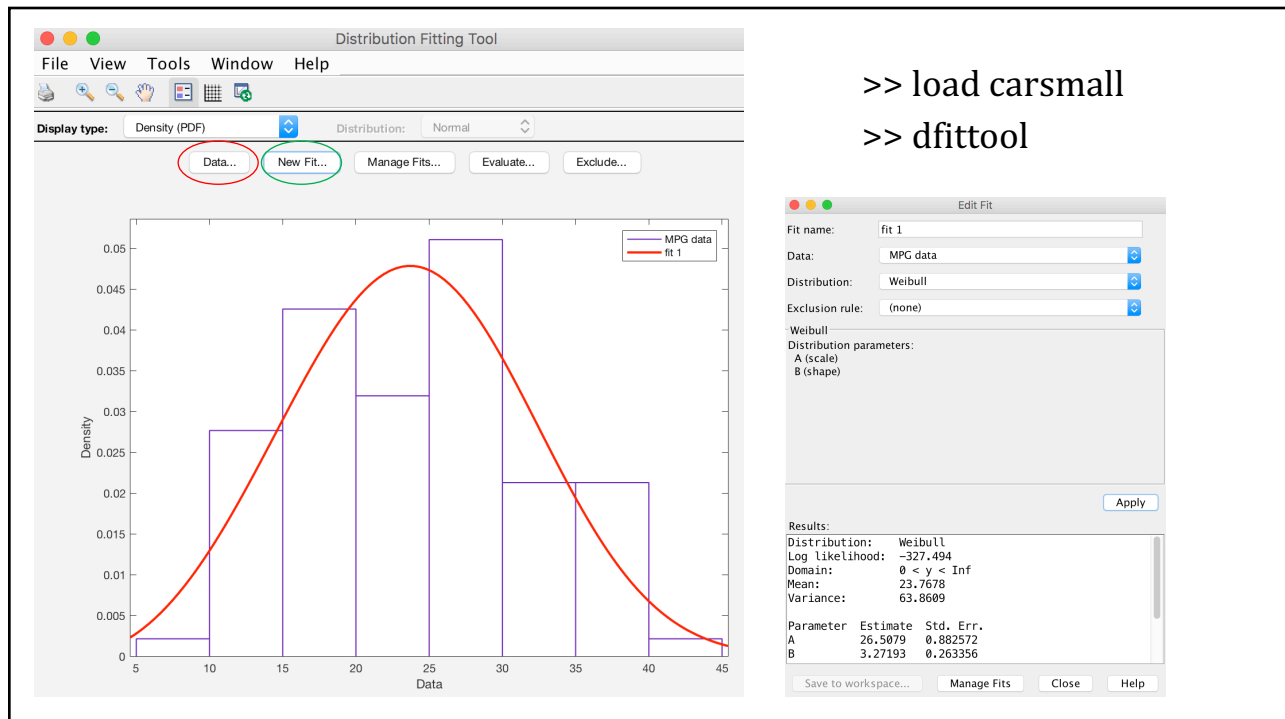- PDF vs. CDF

>> randtool

Uniform distributions:
>> rand(1)
>> randi(10)

Normal distribution:
>> randn(1)

>> load carsmall
>> dfittool