Wiki Activity: Monte carlo simulations – choosing distributions

For Monte Carlo simulation, choosing the right probability distribution for your data is a critical step. The accuracy of your results, predictions, and insights depends heavily on how well the chosen distribution matches the underlying characteristics of your data.

The exact number of probability distributions available depends on how specific and nuanced you get with categorisation. There are dozens of widely recognised distributions in use, and with extensions, modifications, and specialised applications, the number increases substantially.

These distributions can be grouped into several families based on data type (continuous vs. discrete), shape, and application. Below is a summary of some widely used distributions:

Type of Distribution	**Data**	**Description / Purpose**	**Real-World Example**	**Shape of Histogram**
Poisson Distribution	Data represents the number of events occurring in a fixed interval of time or space	Models how often an event is likely to happen over a fixed amount of time or space.	Example: Number of car accidents at a specific intersection, the number of food delivery orders in one hour.	Skewed to the right, with a long tail
Normal Distribution	Data is continuous and symmetrically distributed, often found in nature	Represents natural variations and errors.	Example: Heights of people, test scores, measurement errors.	Bell-shaped, symmetric
Uniform Distribution	Data is uniformly distributed between a minimum and maximum value, all outcomes are equally likely	Models situations where every outcome in a range is equally probable.	Example: Rolling a fair die, generating random numbers in a given range.	Flat, even, all values are equally likely
Triangular Distribution	Data is continuous, with a known minimum, maximum, and most likely value	Models scenarios with known bounds and a peak at the most likely value.	Example: Estimating time for project completion (minimum, most likely, and maximum time).	Triangular shape with a peak at the mode (most likely value)
Negative Binomial Distribution	Data represents the number of trials until a fixed number of successes occurs in a sequence of	Models the number of trials needed for a fixed	Example: Number of trials until you get 10 heads in coin flips.	Skewed to the right, with a peak to the left

Type of Distribution	Data	Description / Purpose	Real-World Example	Shape of Histogram
	independent Bernoulli trials	number of successes.		
Log-Normal Distribution	Data is continuous and positively skewed, where the logarithm of the data is normally distributed	Models situations with multiplicative effects or non-negative outcomes.	Example: Stock prices, incomes, size of particles.	Skewed to the right (positively skewed)
Weibull Distribution	Data represents the time until a failure event occurs, such as in reliability analysis or survival analysis	Used for modeling lifetimes and reliability, especially in engineering.	Example: Time until failure of a machine or product.	Skewed to the right, with a shape that depends on the shape parameter (a)

By understanding these characteristics, you can select the most appropriate distribution, improving the accuracy of your simulation and the reliability of your decisions.

Here is a way to decide which distribution to choose:

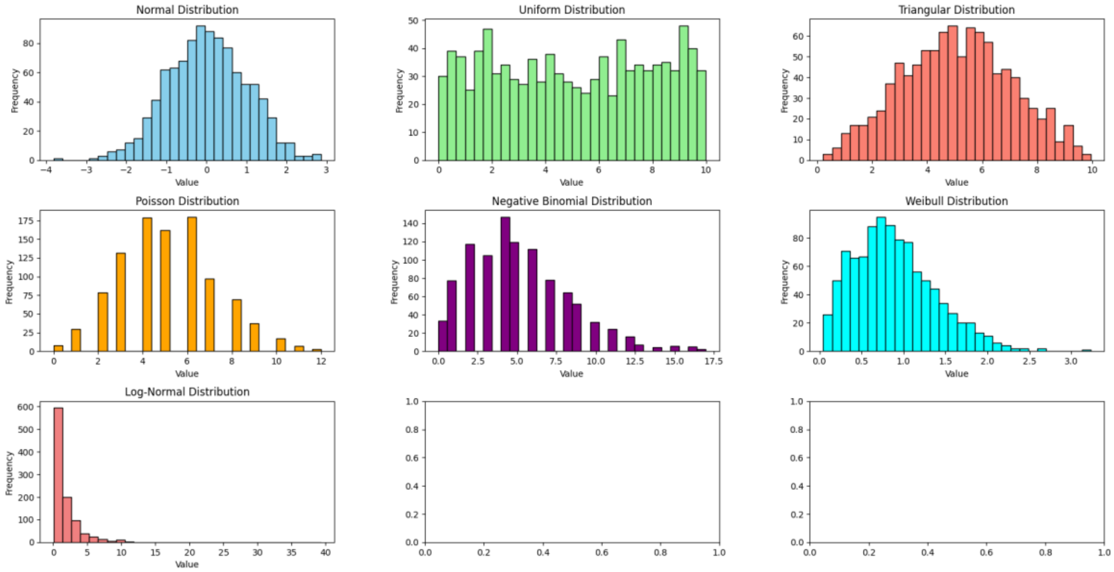
How to Choose a Probability Distribution

Step 1. Understand your data
Begin by identifying the nature of your data:

- Continuous data: Can take any value within a range (e.g. height, weight, temperature).
- Discrete data: Can only take specific values (e.g. number of students, defective items).
- Binary data: Only two outcomes are possible (e.g. success/failure, yes/no).

Step 2. Examine the data's distribution shape
Plot your data to get an idea of its shape. Histograms and density plots help reveal whether the data:

- Symmetric bell-shaped: Normal distribution is a strong candidate.
- Right-skewed (long tail to the right): Consider Exponential, Log-Normal, or Gamma.
- Left-skewed (long tail to the left): Consider Weibull or certain Log-Normal cases.
- Flat or even: Uniform distribution may be appropriate.



To support this process, you can use Python libraries:

- NumPy and SciPy to generate and fit distributions.
- Matplotlib to plot histograms and visualize probability density functions.

Step 3. Fit and test distributions
Export your simulation results from Excel to Python, fit the assumed distribution, and run formal tests (e.g., Kolmogorov–Smirnov, Anderson–Darling, Shapiro–Wilk) using Python libraries such as scipy.stats and statsmodels.

References

Matplotlib (2024) *Matplotlib: Python Plotting — Matplotlib 3.1.1 Documentation*. Available at: https://matplotlib.org/.

Pelin Okutan (2024) *Understanding Common Probability Distributions in Monte Carlo Simulations*. Available at: https://medium.com/@pelinokutan/understanding-common-probability-distributions-in-monte-carlo-simulations-5dbbbf995a45

Polanitzer, R. (2023) *Monte Carlo Simulation and Triangular Distribution in Python; Predict Next Year's Sales Volume*. Available at: https://medium.com/@polanitzer/monte-carlo-simulation-and-triangular-distribution-in-python-predict-next-years-sales-volume-d37f419f2dfc

The SciPy community (2019) *Statistical functions (scipy.stats) — SciPy v1.3.3 reference guide*. Available at: https://docs.scipy.org/doc/scipy/reference/stats.html.

Thomopoulos, N.T. (2012) *Essentials of Monte Carlo Simulation*. Springer Nature. Available at: https://doi.org/10.1007/978-1-4614-6022-0

W3schools (2025) *Random Data Distribution*. Available at: https://www.w3schools.com/python/numpy/numpy_random_distribution.asp