# Probability Basics

Review Goldsman, D. & Goldsman, P. (2020). A First Course in Probability and Statistics. Lulu. [Download link](https://www.lulu.com/search?page=1&q=goldsman&pageSize=10&adult_audience_rating=00)

Other resources that are helpful when reviewing probability and statistics are:

<https://www.probabilitycourse.com/>

[Khan Academy](https://www.khanacademy.org/math/statistics-probability)

[YouTube](https://www.youtube.com/results?search_query=probability+and+statistics)

<https://stats.libretexts.org/Bookshelves>

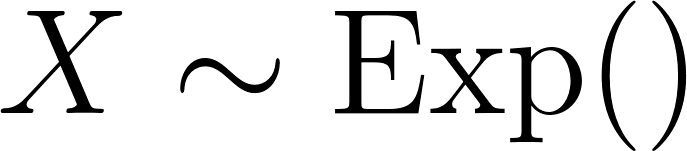
# Why is Probability Important?

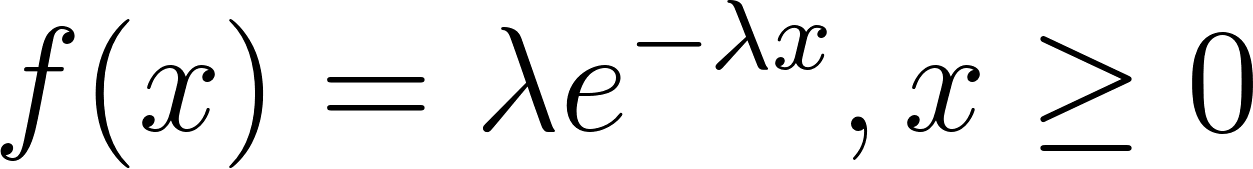
A solid and fundamental understanding of probability is central to many simulation methodologies and applications. Probability provides the framework to model and interpret uncertainty, variability, and randomness in systems. Here's why its understanding is pivotal to the concept of simulation:

* Modeling Uncertainty: Many real-world systems and processes are inherently uncertain. Probability allows us to quantify and model this uncertainty, whether it's in the context of the weather, stock market fluctuations, the behavior of subatomic particles, or the spread of diseases.
* Stochastic Simulations: Many simulations are stochastic in nature, meaning they incorporate randomness. The correct application and interpretation of this randomness requires a foundation in probability theory.
* Monte Carlo Methods: One of the most widely used simulation techniques, Monte Carlo methods rely on repeated random sampling to estimate outcomes. Understanding probability is essential to design, interpret, and ensure the validity of Monte Carlo simulations.
* Statistical Interpretation: Simulations often produce vast amounts of data. Probability provides the basis for statistical methods to analyze, summarize, and make inferences from this data.
* Risk Assessment: Probability is the bedrock of risk assessment in fields like finance, engineering, and medicine. Simulations often aim to understand the likelihood of certain outcomes, such as system failures or financial losses, and to make decisions accordingly.
* Queueing Theory: In operations research and computer science, the study of queues (like those in customer service scenarios or computer networks) is underpinned by probability theory.
* Reliability Engineering: In fields like aerospace or electronics, simulations are used to estimate the reliability and potential failure modes of systems. These estimations are grounded in probability.
* Genetics and Evolutionary Biology: Simulations in these fields might involve understanding the probabilistic nature of gene inheritance or the spread of genetic traits in populations.
* Probabilistic Algorithms: In computer science, some algorithms are inherently probabilistic, providing "good enough" solutions faster than deterministic counterparts. Understanding their behavior and guarantees requires knowledge of probability.
* Validation & Calibration: When using simulations to model real-world systems, it's essential to validate and calibrate the models against actual data. Probability and statistics offer the tools to measure how well the simulation aligns with reality and to adjust model parameters accordingly.
* Decision Making Under Uncertainty: Simulations are often used to aid decision-making. Probability provides the framework to weigh different outcomes and make informed decisions in the presence of uncertainty.
* Understanding Convergence: In iterative simulations, understanding when a simulation has run long enough or when it has "converged" often requires probabilistic reasoning.
* Random Number Generation: At a fundamental level, many simulations require the generation of random numbers or sequences. Understanding their properties, distributions, and potential biases is crucial, and this understanding comes from probability theory.

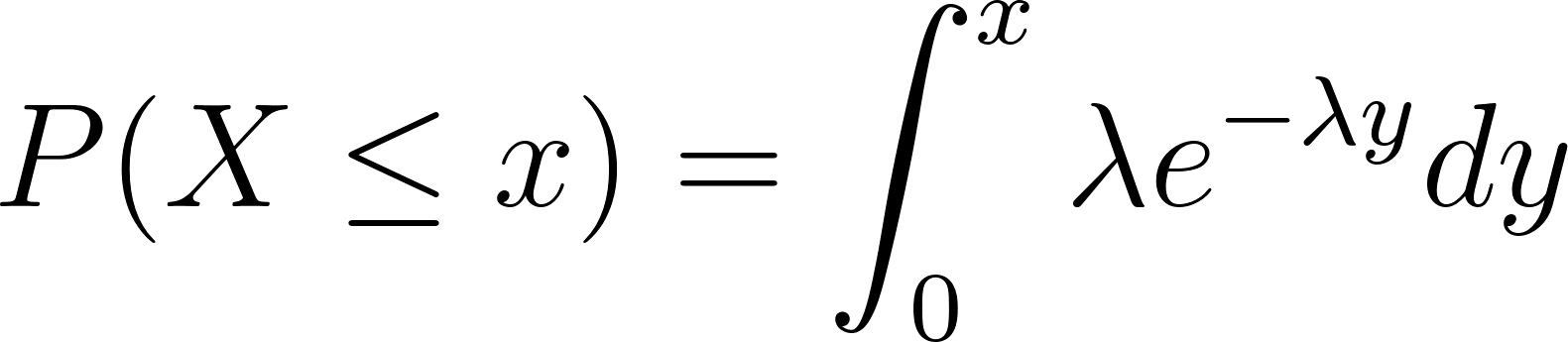
Probability theory offers the tools, language, and concepts to grapple with the inherent uncertainties and variabilities in countless systems. It's the foundation on which much of simulation science is built, allowing practitioners to model, predict, and make decisions in complex, dynamic, and uncertain environments.

# Example: Find CDF of Exponential Random Variable

Let [](https://latex-staging.easygenerator.com/eqneditor/editor.php?latex=X%20%5Csim%20%5Ctext%7BExp%7D(%5Clamba)#0), then the probability density function of [](https://www.codecogs.com/eqnedit.php?latex=X#0) is:

[](https://www.codecogs.com/eqnedit.php?latex=f(x)%3D%5Clambda%20e%5E%7B-%5Clambda%20x%7D%2C%20x%20%5Cgeq%200#0)

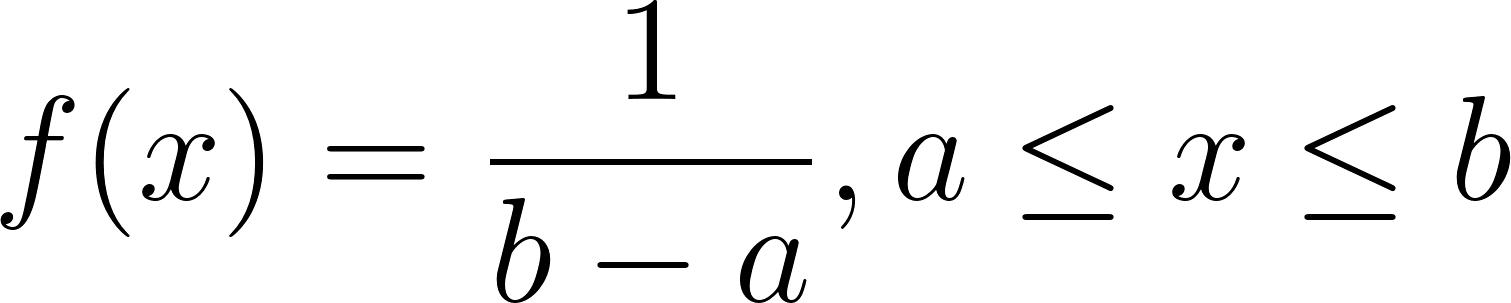
The cumulative distribution function is by definition the following; the change of variable using [](https://www.codecogs.com/eqnedit.php?latex=y#0) is for clarity when doing the integration.

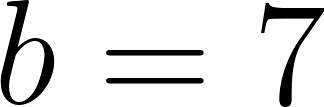
[](https://www.codecogs.com/eqnedit.php?latex=P(X%20%5Cleq%20x)%3D%5Cint_0%5E%7Bx%7D%20%5Clambda%20e%5E%7B-%5Clambda%20y%7D%20dy#0)

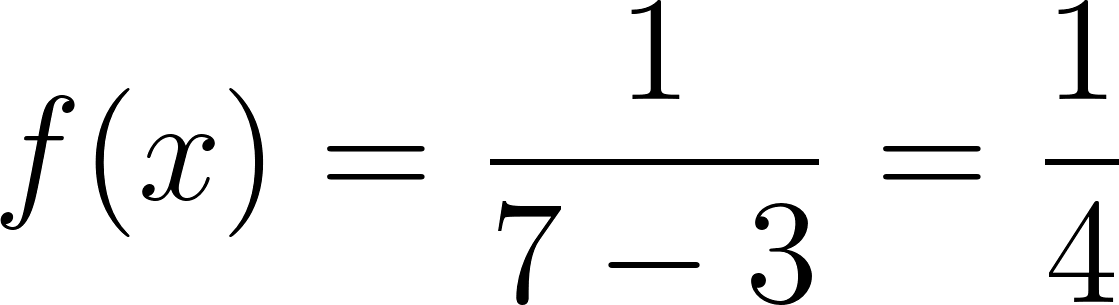
[](https://www.codecogs.com/eqnedit.php?latex=%3D-e%5E%7B-%5Clambda%20y%7D%20%5Cbig%20%5Crvert_0%5Ex%3D-e%5E%7B-%5Clambda%20x%7D-%5Cbig(-e%5E0%20%5Cbig)%3D-e%5E%7B-%5Clambda%20x%7D%2B1%3D1-e%5E%7B-%5Clambda%20x#0)

# Example: Continuous random variable

A continuous random variable with parameters [](https://www.codecogs.com/eqnedit.php?latex=a#0) (minimum) and [](https://www.codecogs.com/eqnedit.php?latex=b#0) (maximum) has a probability density function:

[](https://www.codecogs.com/eqnedit.php?latex=f(x)%3D%5Cdfrac%7B1%7D%7Bb-a%7D%2C%20a%20%5Cleq%20x%20%5Cleq%20b#0)

When [](https://www.codecogs.com/eqnedit.php?latex=a%3D3#0) and [](https://www.codecogs.com/eqnedit.php?latex=b%3D7#0) this is

[](https://www.codecogs.com/eqnedit.php?latex=f(x)%3D%5Cdfrac%7B1%7D%7B7-3%7D%3D%5Cdfrac%7B1%7D%7B4%7D#0)

# Probability Density

Imagine you have a long, narrow stretch of beach. The entire stretch represents all possible values a continuous random variable might take. Now, think of sand on that beach.

Probability Density: This is like taking a tiny, tiny segment of that beach and examining how densely packed the sand grains are in that segment. It gives you a measure of how dense the sand is at a specific point, but because the segment is infinitely small, the actual number of sand grains (which would be analogous to probability) is technically zero. The value of the density at that point is given by the probability density function (PDF).

Probability: Now, instead of just looking at one tiny segment, you decide to take a whole scoop of sand from a certain segment of the beach, say between two points A and B. The total number of sand grains in that scoop is like the probability that the continuous random variable falls between points A and B. This is found by integrating the probability density (or summing up all those tiny segments) over the interval [A, B]. The result of this integration is given by the cumulative distribution function (CDF).

For a continuous random variable:

* Probability Density (given by the PDF) at a specific point is NOT a probability. It just gives a measure of how "dense" the probability is around that point.
* Probability that the variable falls in a specific interval is obtained by integrating the PDF over that interval. This is equivalent to evaluating the difference in the CDF values at the endpoints of the interval.

A key point to remember is that for continuous random variables, the probability at a single, exact point is always zero. That's because there are infinitely many points on the real number line, so the chance of landing on one exact point is zero. Instead, we talk about the probability over an interval or range of values. The PDF helps describe how those probabilities are distributed across different intervals.