OTTO VON GUERICKE UNIVERSITÄT MAGDEBURG

INF FAKULTÄT FÜR INFORMATIK

Introduction to Simulation

Random Variables and Random Numbers

Motivation, Contents

Motivation:

- Discrete-event systems usually contain random variables
- These have a significant effect on discrete-event simulations
- Our simulator has to be able to sample random variables

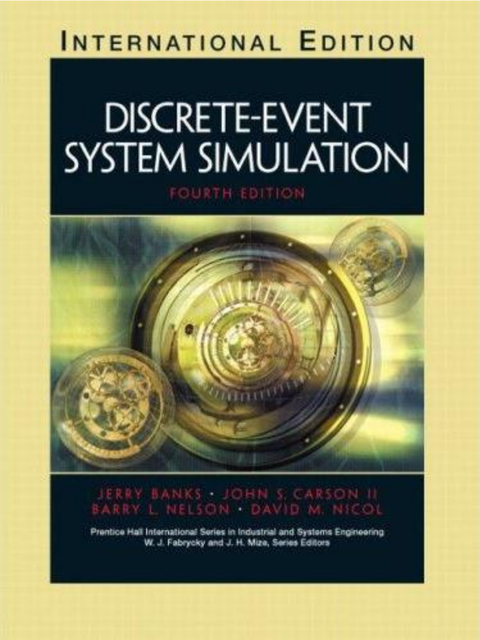
Contents:

- Explain what a random variable (RV) is
- Explain some important properties of RVs
- Present some of the most important RVs in practice
- Describe how to generate random numbers

Background Reading

Relevant sections of the book:

- 5.1
- 5.2
- 5.3
- 5.4
- 5.6
- 7.1
- 7.2
- 7.3.1
- 8.1



Random Variables

A simple definition:

- Random variables take on values at random

Examples:

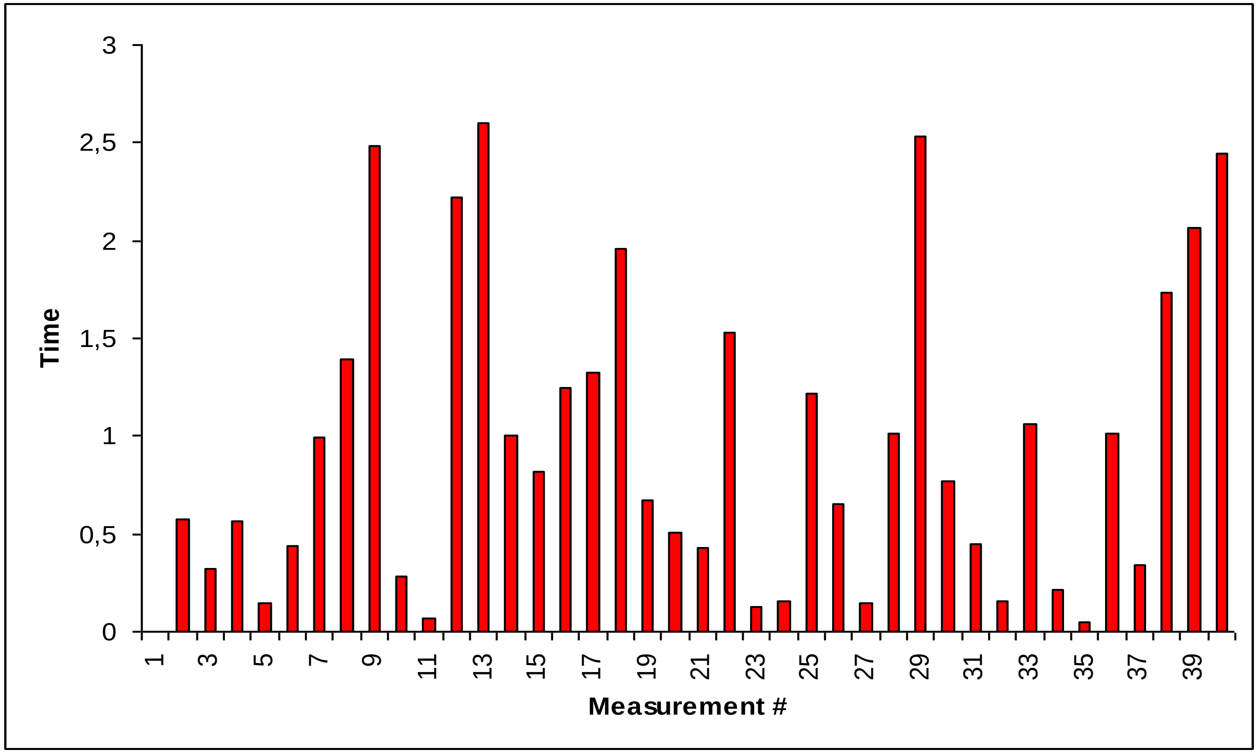
- Service times at a bank
- Time to Failure (TTF) of a piece of equipment
- Human lifetimes

However, the randomness usually has a structure

- This structure is described using a probability distribution

Random Variables

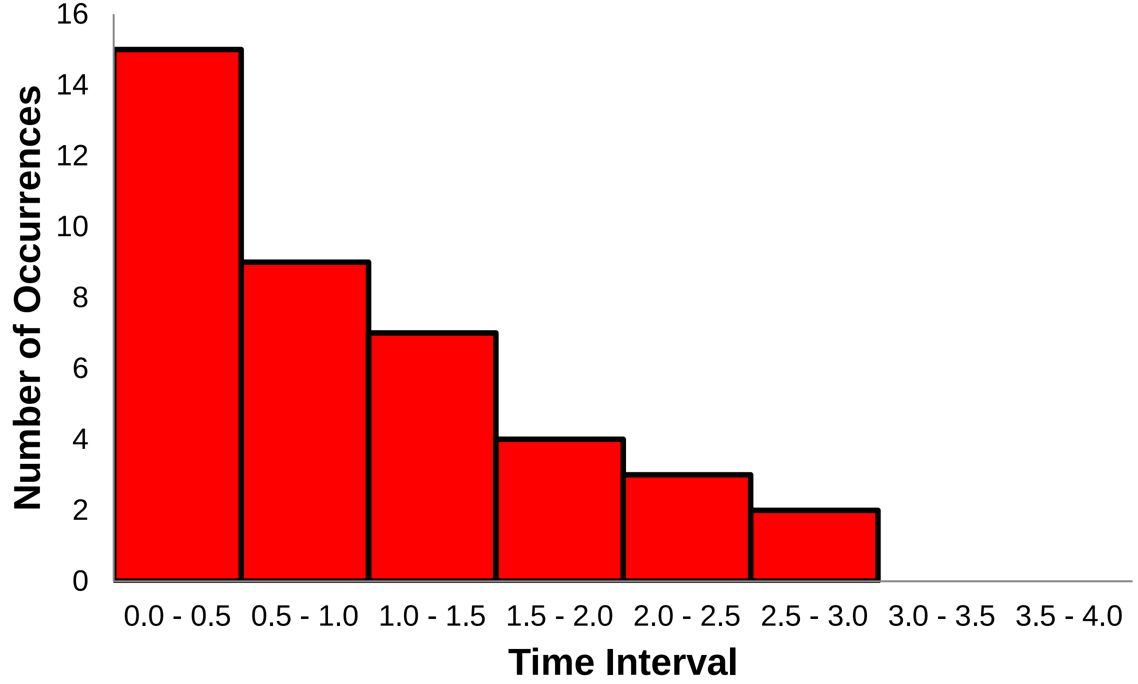
Example: inter-arrival times of customers at a bank:



Presented in this way, the values have no apparent structure

Random Variables

Group the measurements into half-minute intervals:



A certain structure in the randomness becomes visible

Discrete Distributions

Let X be a random variable

If the range R_X of X is finite or countably infinite, then

- X is a discrete random variable

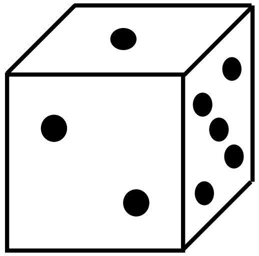
Denote the possible values of X by x_i

- Then we have probabilities $p(x_i) = P(X = x_i)$

$p(x_i)$ is called the probability mass function

Discrete Distributions

Example:

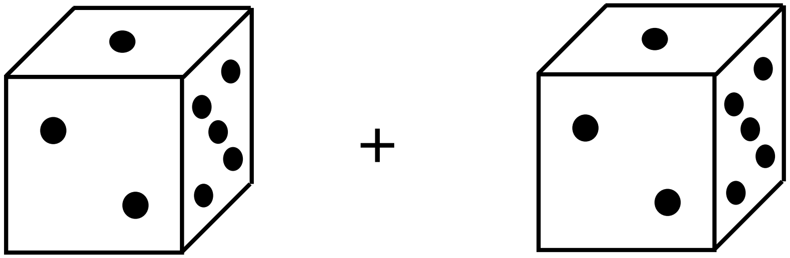


Then we have $R_X = \{1, 2, 3, 4, 5, 6\}$

x_i	1	2	3	4	5	6	
$p(x_i)$	1	1	1	1	1	1	/6

Discrete Distributions

Example: Sum of two dice



We have: $R_X = \{2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12\}$

x_i	2	3	4	5	6	7	8	9	10	11	12	
$p(x_i)$	1	2	3	4	5	6	5	4	3	2	1	/36

Discrete Distributions

Example: Choice of meal at a university canteen

$R_X = \{\text{Line1}, \text{Line2}, \text{Line3}\}$

Probabilities:

x_i	Line1	Line2	Line3	
$p(x_i)$	800	600	300	/1700

Continuous Distributions

Let X be a random variable

If the range R_X of X is uncountably infinite, then

- X is a continuous random variable

Continuous distributions can be described using

- Probability density functions (pdf)
- Cumulative distribution functions (cdf)

Continuous Distributions

Examples:

- Inter–arrival times at a bank:

  $R_X = (0, \infty)$ seconds

- Heights of students in ItS:

  $R_X = (155, 204)$ cm

PDF and CDF

A function $f(x)$ is a probability density function (pdf), if

$$f(x) \geq 0 \quad \forall x \qquad\qquad \int_{-\infty}^{\infty} f(s)\,ds = 1$$

A function $F(x)$ is a cumulative distribution function (cdf), if

$$\lim_{x \to \infty} F(x) = 1 \qquad\qquad x_1 > x_2 \Rightarrow F(x_1) \geq F(x_2)$$

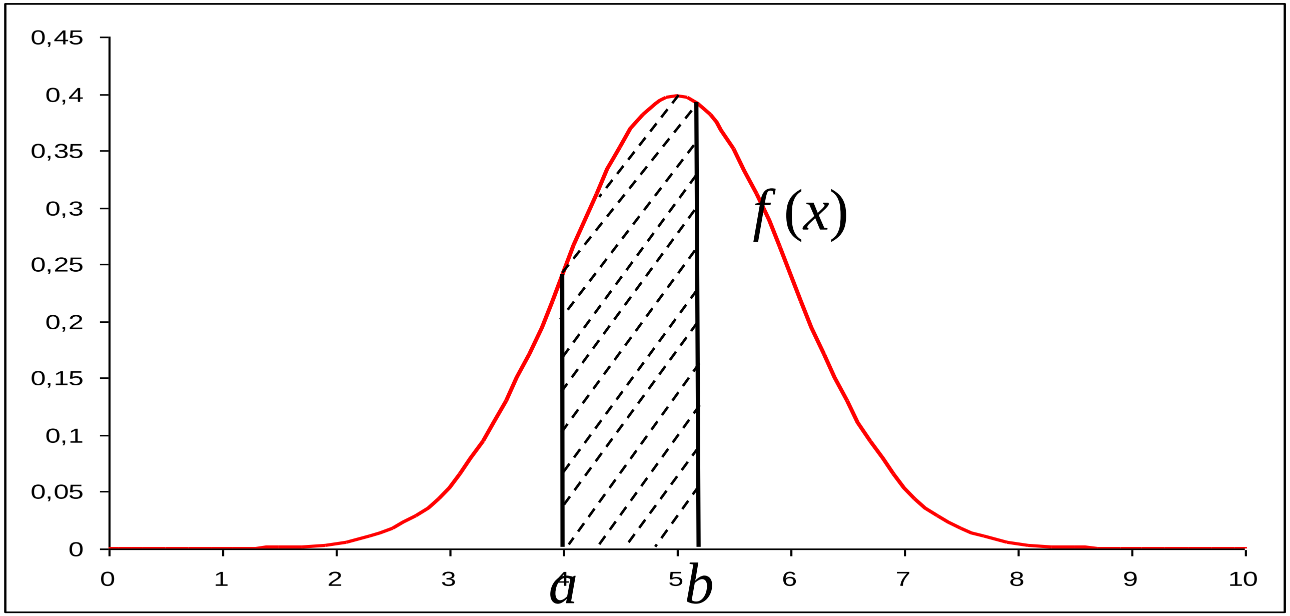
A cdf $F(x)$ is always related to a pdf $f(x)$ via

$$F(x) = \int_{-\infty}^{x} f(s)\,ds$$

PDF and CDF

Meaning of the pdf:

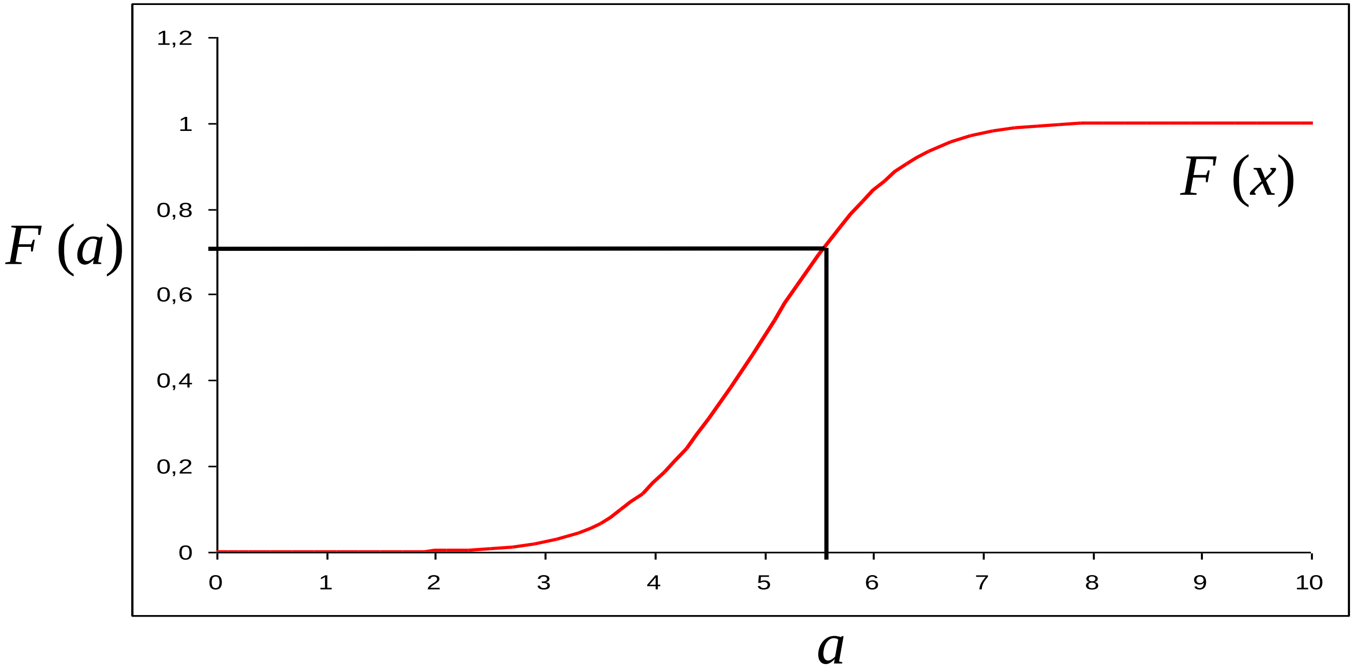
$$P(a < X < b) = \int_a^b f(s)\,ds$$



PDF and CDF

Meaning of the cdf:

$$P(X \leq a) = F(a)$$



Sampling

What is sampling?

- Obtaining a random number from a given distribution

Example:

- Sum of two dice rolls

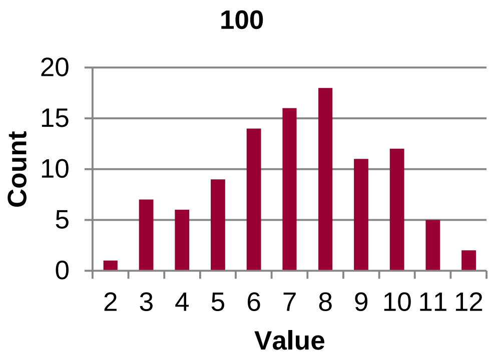
One possible sequence of samples:

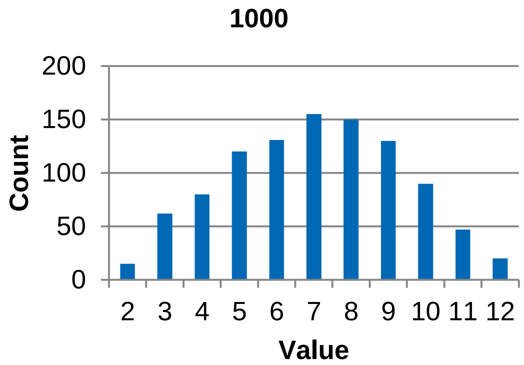
- 7, 4, 10, 2, 5, 11, 8, 4

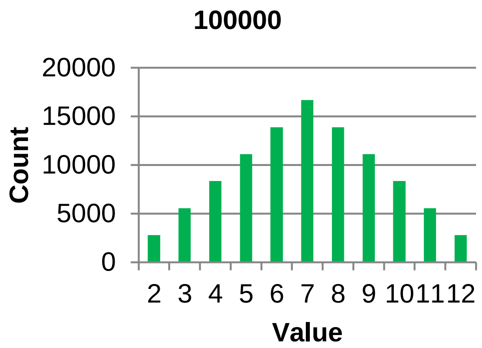
Sampling

If you collect enough samples, …

- you will begin to see the distribution







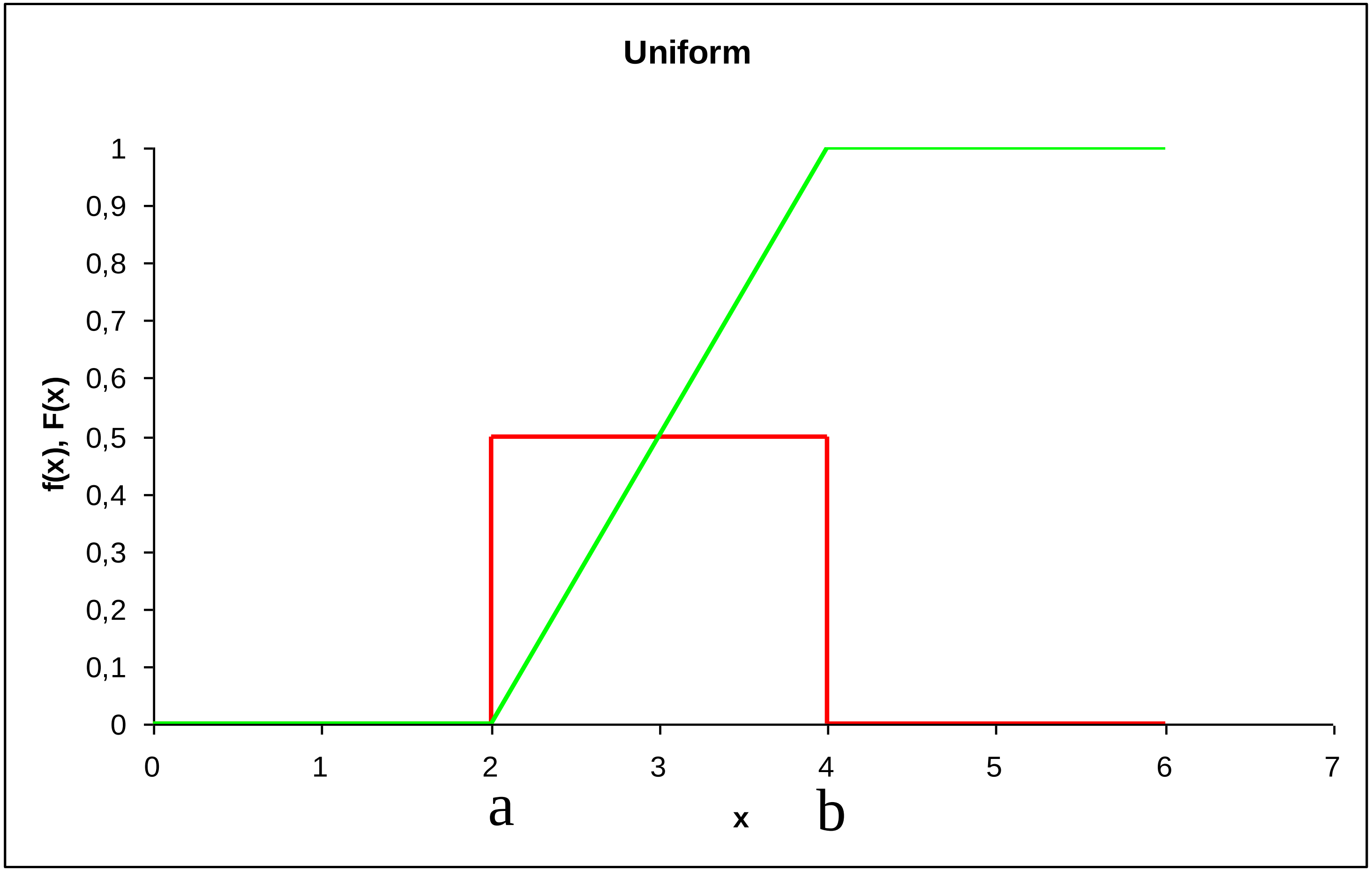
Common Distributions

There are many distributions in common use:

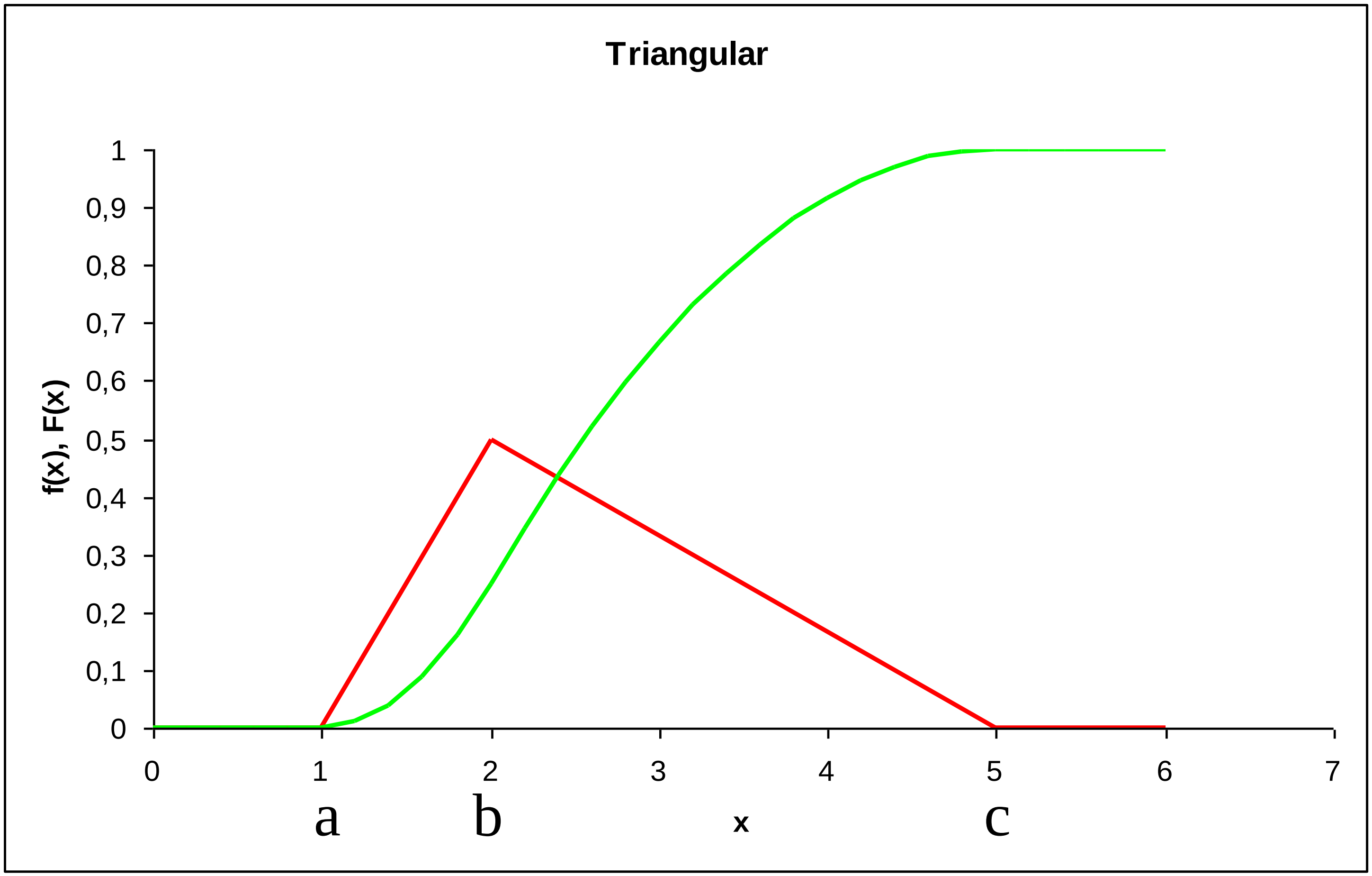
- Uniform
- Exponential
- Normal
- Triangular
- Weibull
- Lognormal
- Beta
- Gamma
- Hyperexponential
- Cox

We will just look at a few of these

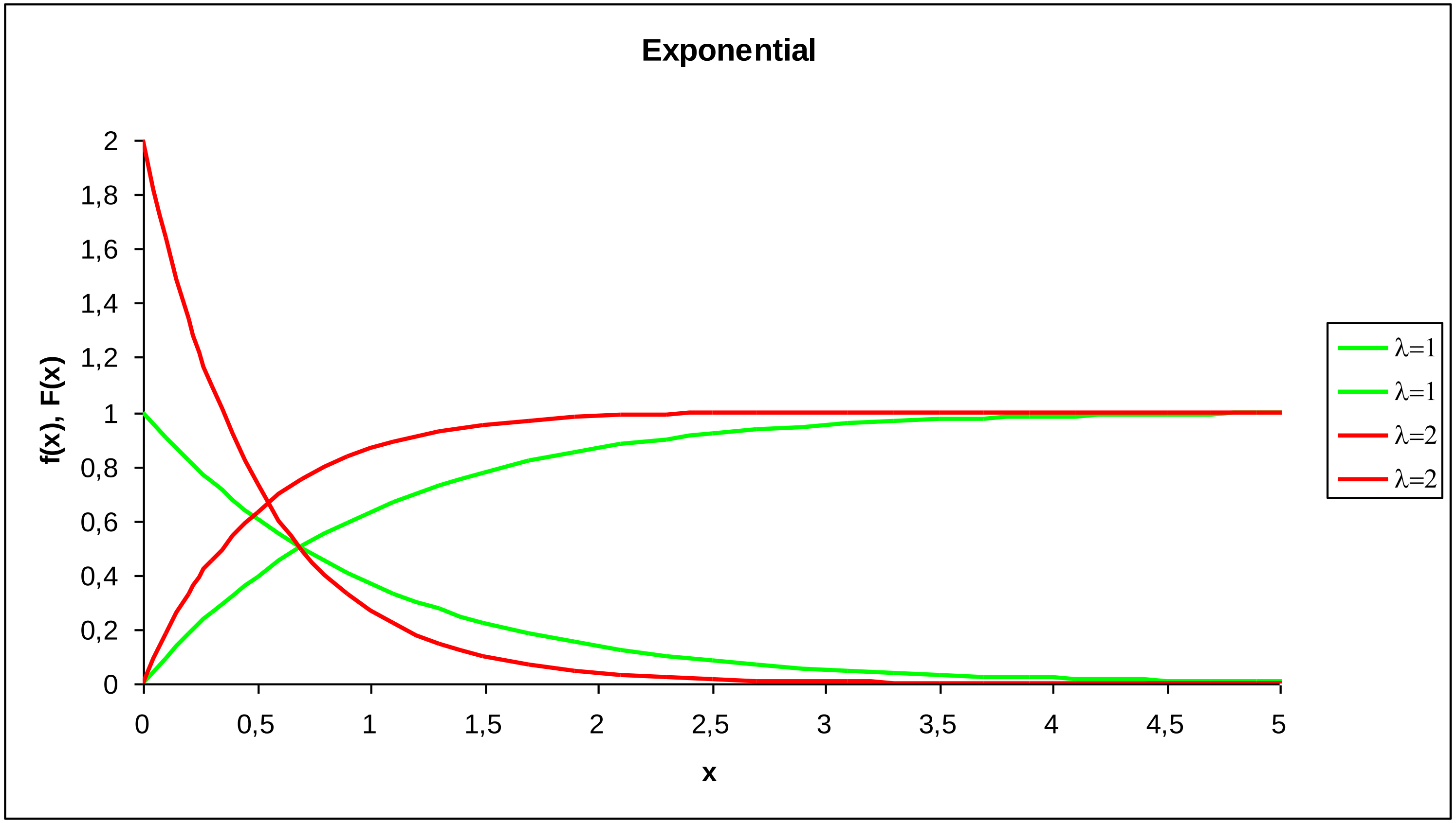
The Uniform Distribution



The Triangular Distribution



The Exponential Distribution



The Exponential Distribution

Definition:

$$f(x) = \lambda e^{-\lambda x} \qquad\qquad F(x) = 1 - e^{-\lambda x}$$

Applications:

- Independent arrivals from an infinite population
- Lifetimes of electronic components

The Exponential Distribution

The exponential distribution is memoryless:

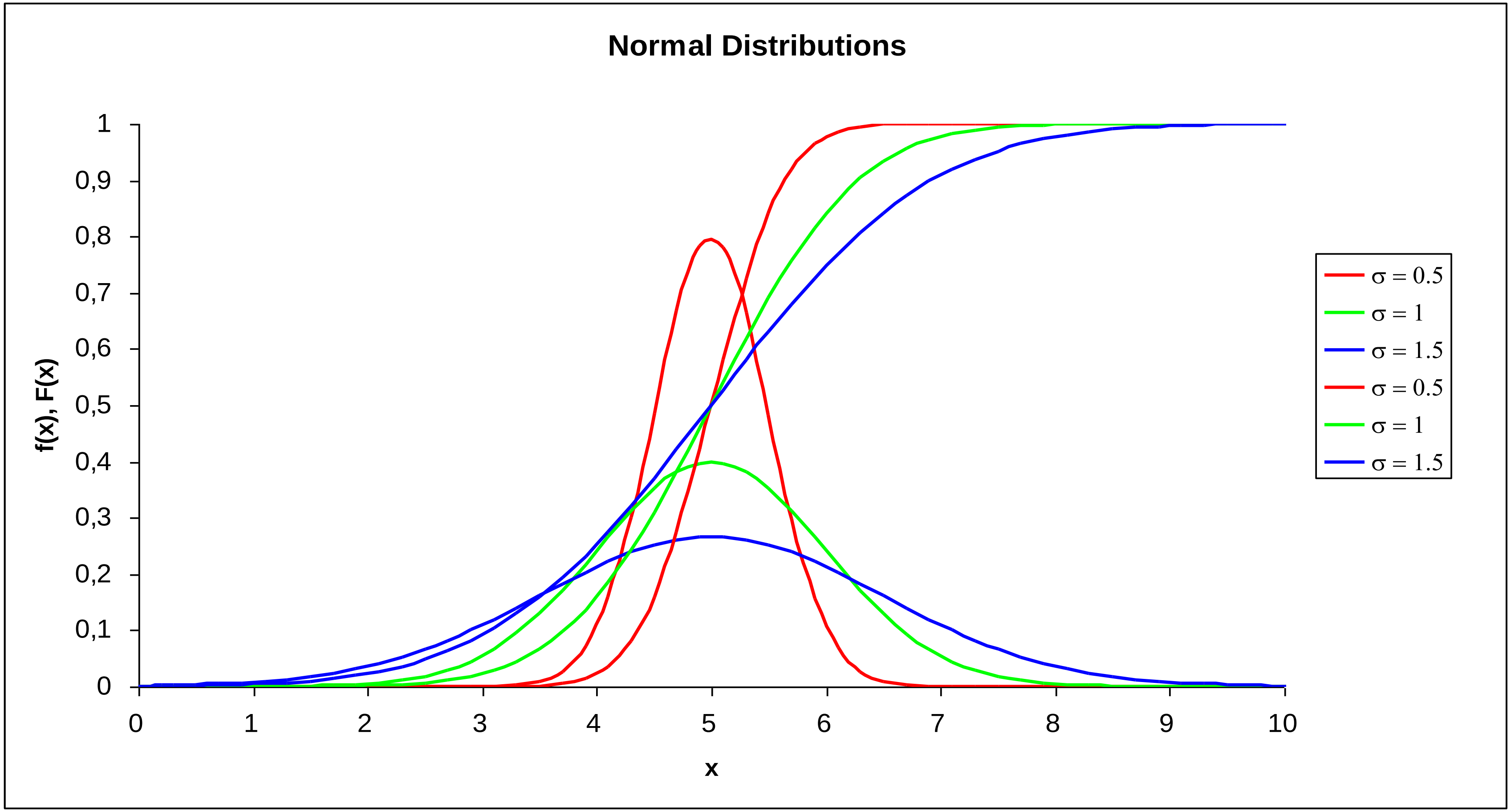
$$P(X > s + t \mid X > s) = P(X > t)$$

Intuitive interpretation of memorylessness:

- The probability that the event will occur within a given future time interval does not depend on how long you have been waiting for it to happen.

This property is unique and very important!

The Normal Distribution



The Normal Distribution

Definition:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \qquad F(x) = ?$$

σ^2 is the variance

μ is the mean

$\sigma^2 = 1$, $\mu = 0$ yields the standard normal distribution

$F(x)$ cannot be expressed analytically

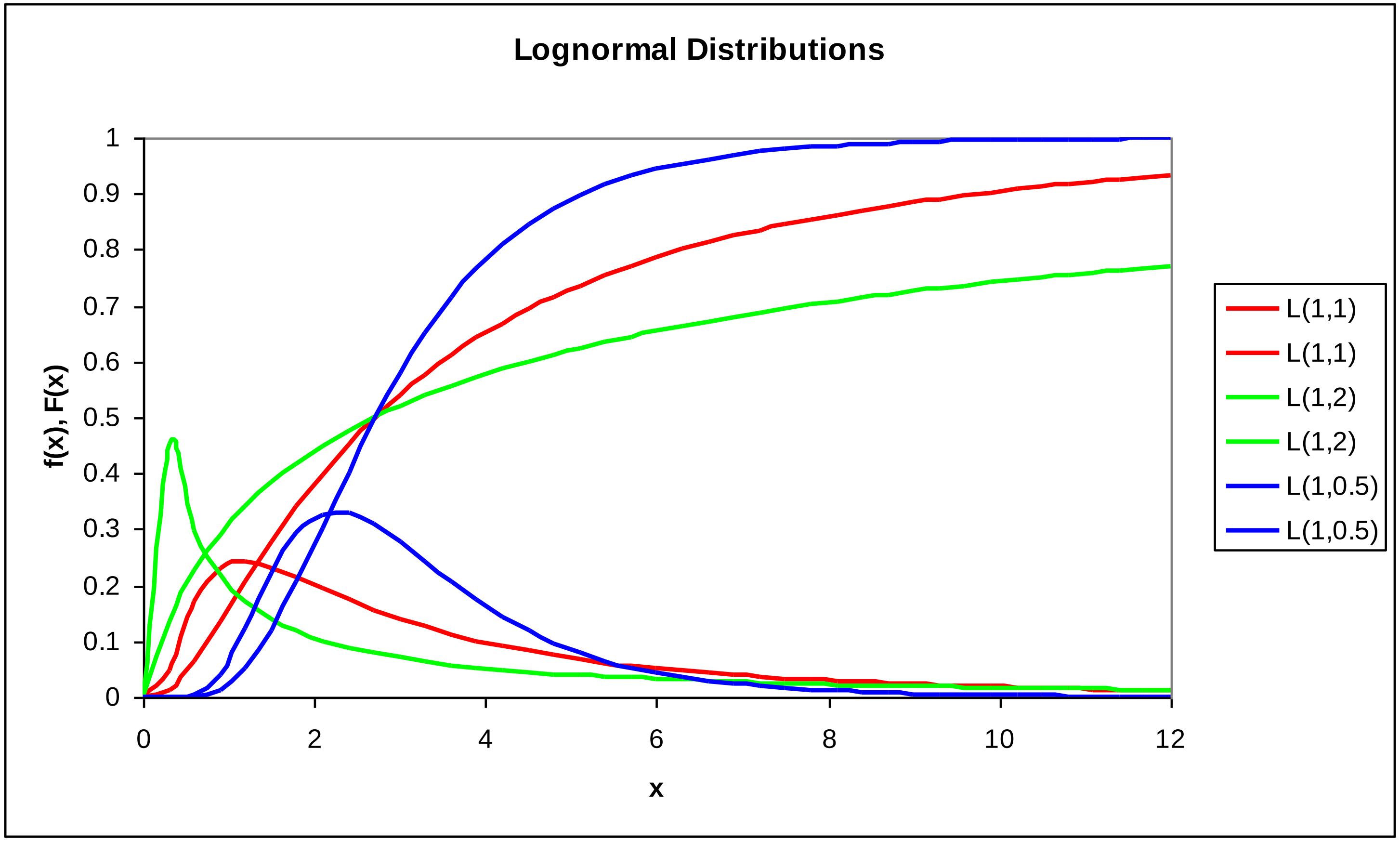
The Normal Distribution

Applications:

- Shooting at a target
- Common natural attributes of a population
- Service times

Also known as the Gaussian distribution

The Lognormal Distribution



The Lognormal Distribution

Lognormal density function:

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

$F(x)$ cannot be expressed analytically

$$X \sim \mathrm{LN}(\mu, \sigma^2) \quad \Leftrightarrow \quad \ln X \sim \mathrm{N}(\mu, \sigma^2)$$

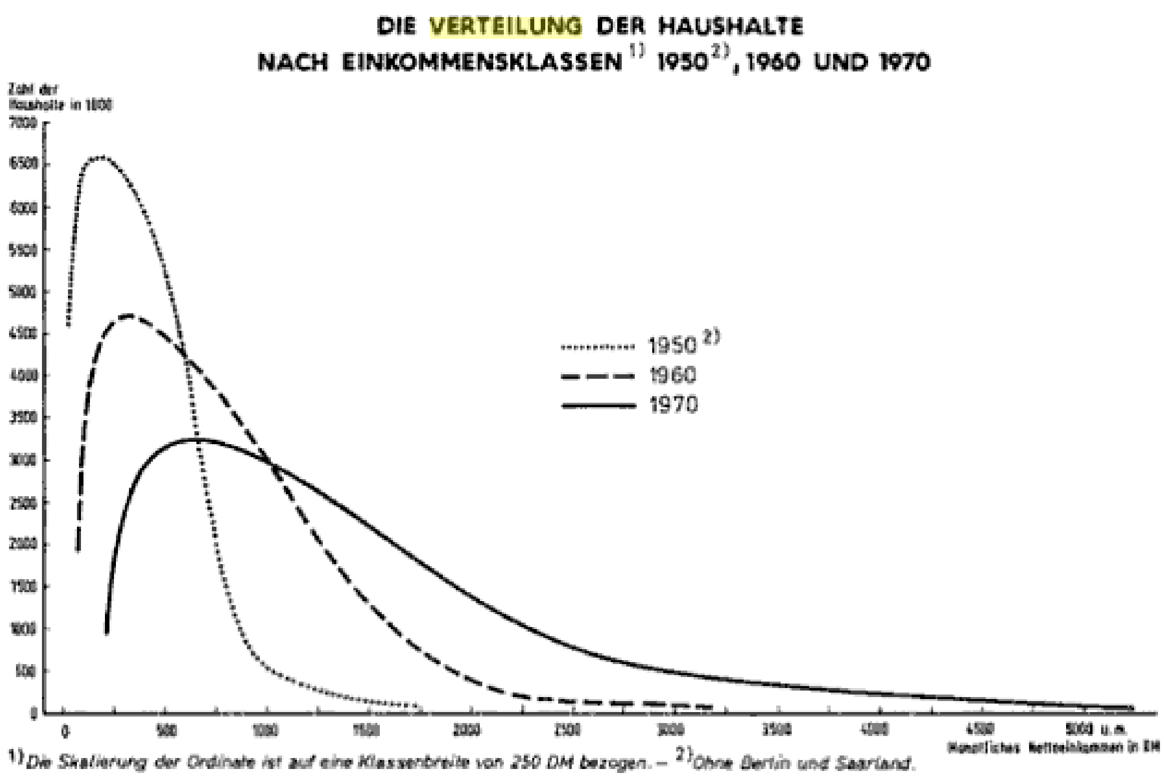
Applications: Many empirically determined distributions

- Household incomes
- Age at first marriage

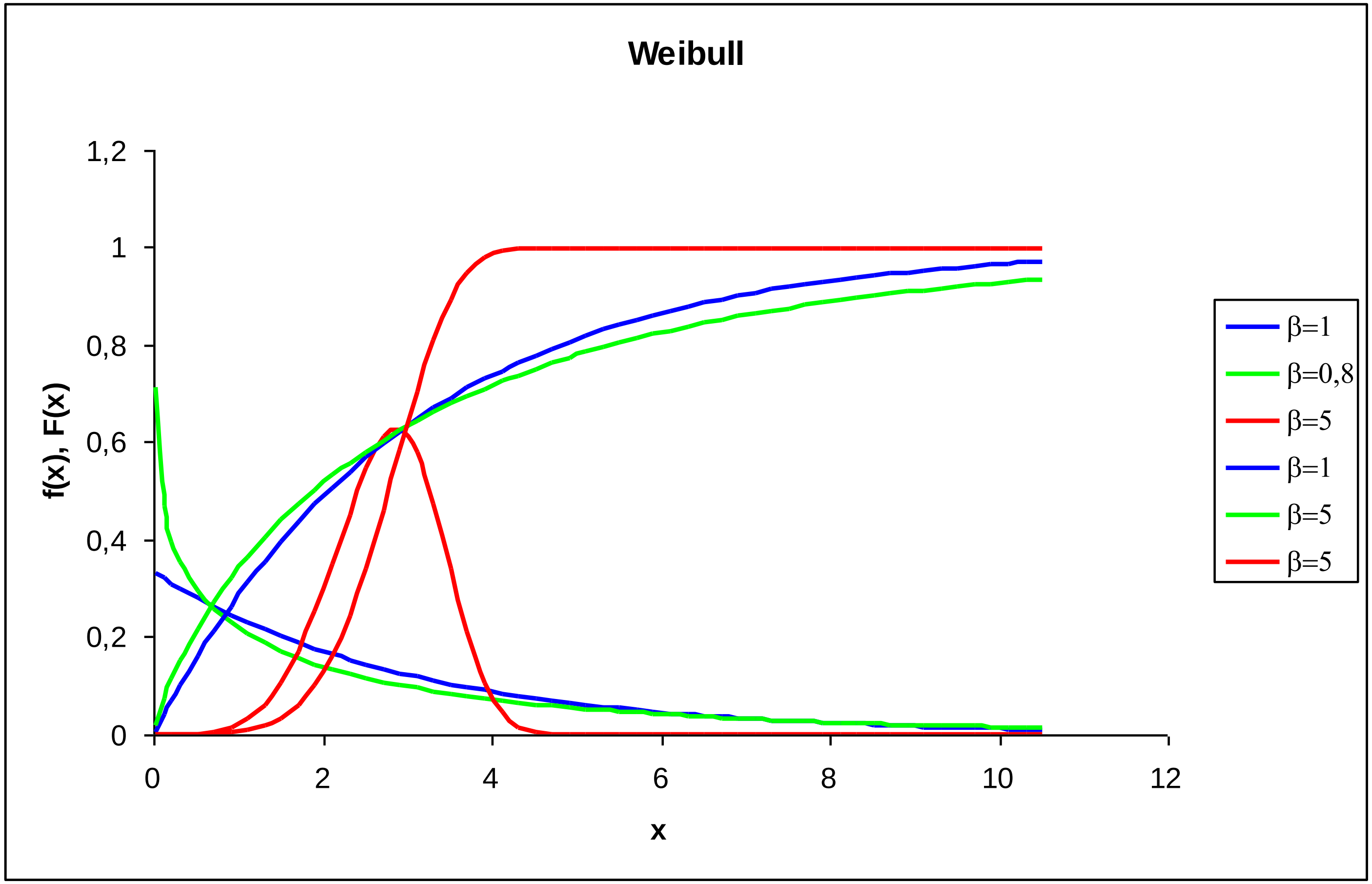
The Lognormal Distribution

Aus

- L. Rebeggiani: Persönliche Einkommensverteilung, privater Konsum und Wachstum



The Weibull Distribution



The Weibull Distribution

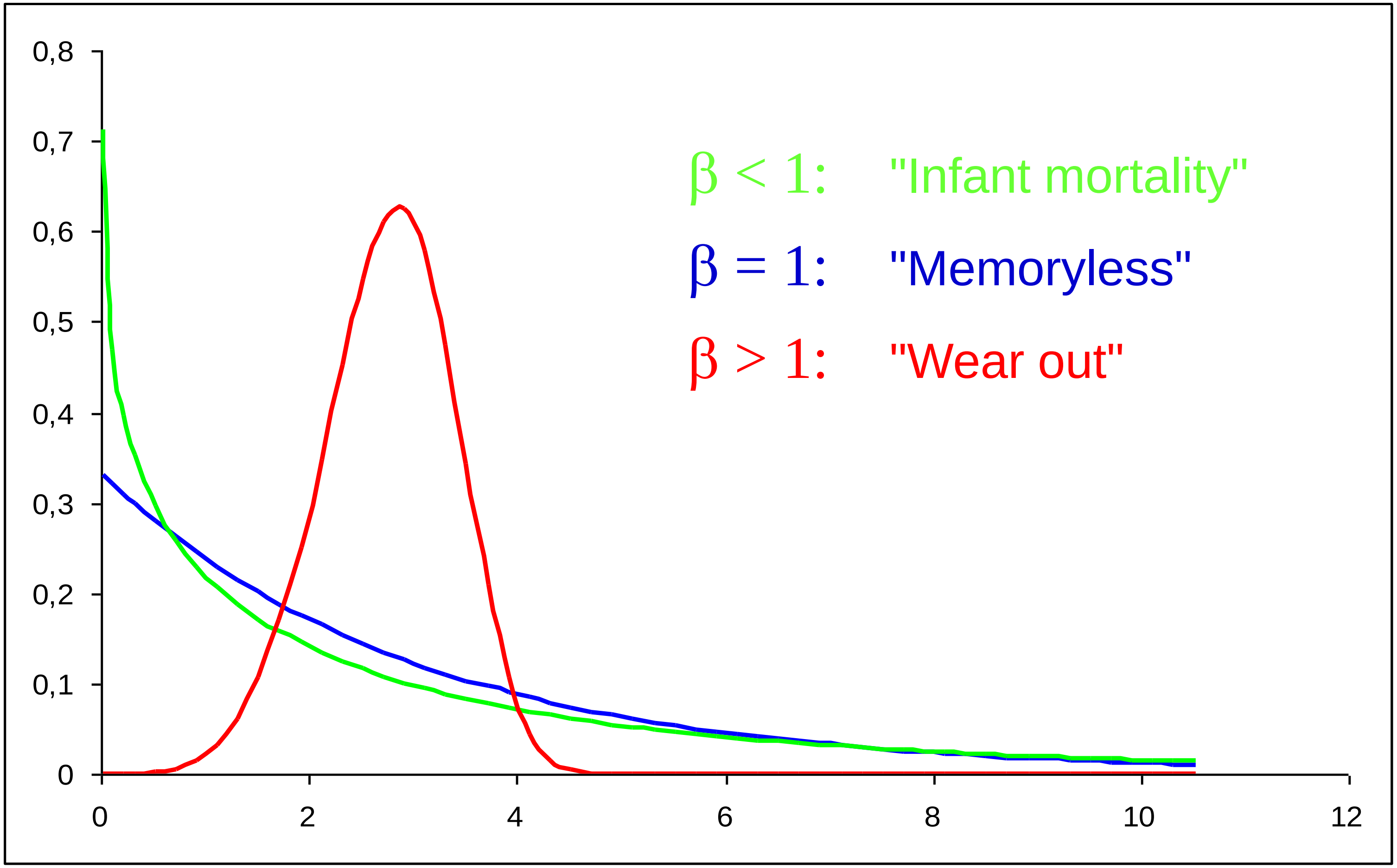
Definition:

$$f(x) = \frac{\beta}{\alpha}\left(\frac{x}{\alpha}\right)^{\beta-1} e^{-\left(\frac{x}{\alpha}\right)^{\beta}} \qquad F(x) = 1 - e^{-\left(\frac{x}{\alpha}\right)^{\beta}}$$

Application: lifetimes of different types of component:

- $\beta < 1$: "Infant mortality" (Manufacturing errors)
- $\beta = 1$: "Memoryless" (Electronic parts)
- $\beta > 1$: "Wear" (Mechanical parts)

The Weibull Distribution

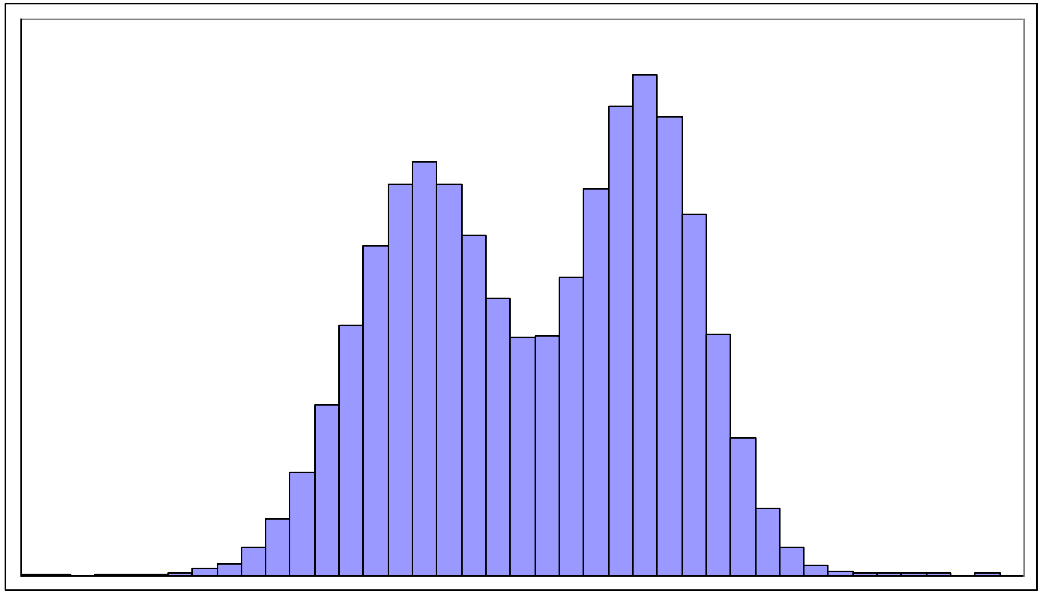


Empirical Distributions

In practice, measured distributions often do not correspond to standard mathematical functions

Example:

- Time to repair for a welding robot



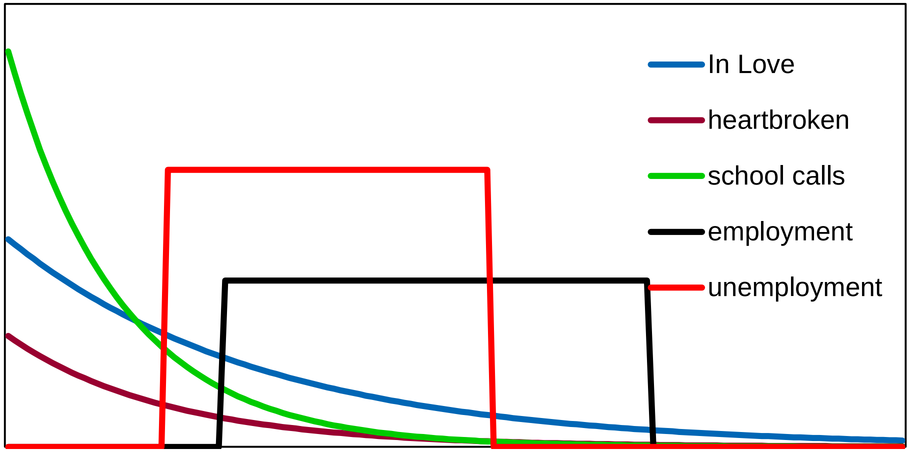
The Sims – Almost Normal Family Life

The situation contains five random variables:

- Duration of being in love
- Duration of being heartbroken
- Time to the next call from school
- Duration of an employment period
- Duration of an unemployment period

N.B.:

- The curves are not drawn to the same scale!



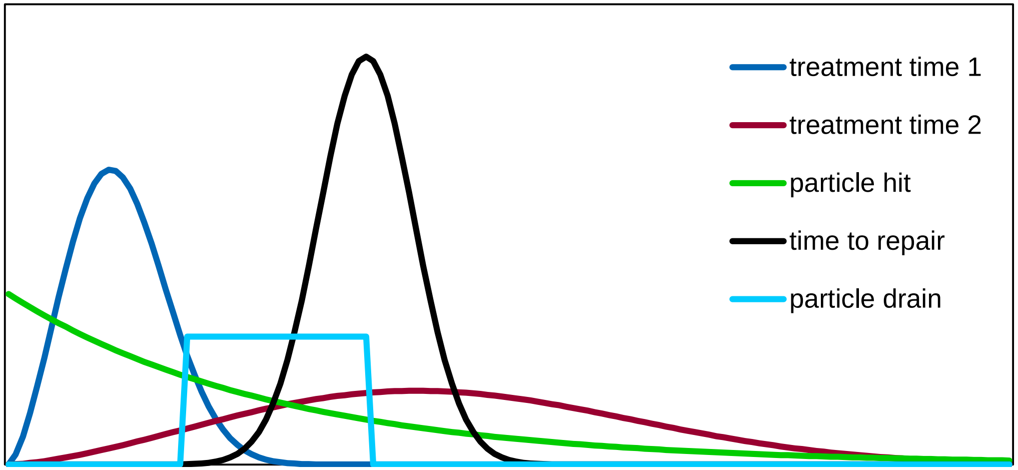
Star Trek – USS Enterprise in Danger

The situation contains five random variables:

- Time for medical treatment stage 1
- Time for medical treatment stage 2
- Time between two particle hits
- Total time that will be needed for the repair
- Drain on the shields of one particle hit

N.B.:

- The curves are not drawn to the same scale!



Random Numbers

Our simulator uses random variables

- For example in order to model arrival intervals & service times

We need samples from these random variables

- In order to obtain appropriately distributed random numbers
- For example individual arrival intervals & service times

We must therefore be able to generate random numbers

We will use a deterministic algorithm

- In fact, we generate *pseudo*–random numbers

Properties of RNG

Desirable properties of a pseudo-random number generator:

- Independence (contains no pattern)
- Uniform distribution
- Speed
- Portability
- Long cycle
- Replicability

Linear Congruential Method

The most important type of RNG is called

- The Linear Congruential Method (LCM)

The LCM is used in almost all simulators

- It (can) fulfil all the desired properties
- It is very easy to implement

Linear Congruential Method

The LCM needs three integer parameters a, c and m

It generates a sequence of integers $0 < X_i < m$:

$$X_i = \left(a \cdot X_{i-1} + c\right) \bmod m$$

The random numbers $0 < R_i < 1$ that we need are then

$$R_i = \frac{X_i}{m}$$

The starting value X_0 is called the *seed*

Linear Congruential Method

The sequence of RN is periodic

The quality of the RN produced depends on a, c and m

How to best choose a, c and m?

- The choice is made in order to achieve maximum period P

Different choices for a, c and m yield strongly differing Ps

Linear Congruential Method

Example: $a = 13$, $m = 64$, $c = 0$, $X_0 = 1, 2, 3$, and 4

i	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
X_i	1	13	41	21	17	29	57	37	33	45	9	53	49	61	25	5	1
X_i	2	26	18	42	34	58	50	10	2								
X_i	3	39	59	63	51	23	43	47	35	7	27	31	19	55	11	15	3
X_i	4	52	36	20	4												

Problems with the LCM

Consider the LCM with $a=10205$, $c=0$, $m=2^{15}$

$$
\begin{aligned}
x_0 &= 12345 = 01100000011\textcolor{red}{1001}\\
x_1 &= 20533 = 10100000011\textcolor{red}{0101}\\
x_2 &= 20673 = 10100001100\textcolor{red}{0001}\\
x_3 &= 7581 = 00111011001\textcolor{red}{1101}\\
x_4 &= 31625 = 11110111000\textcolor{red}{1001}\\
x_5 &= 1093 = 00001000100\textcolor{red}{0101}\\
x_6 &= 12945 = 01100101001\textcolor{red}{0001}\\
x_7 &= 15917 = 01111100010\textcolor{red}{1101}
\end{aligned}
$$

Look at the periodicity of the last bits!

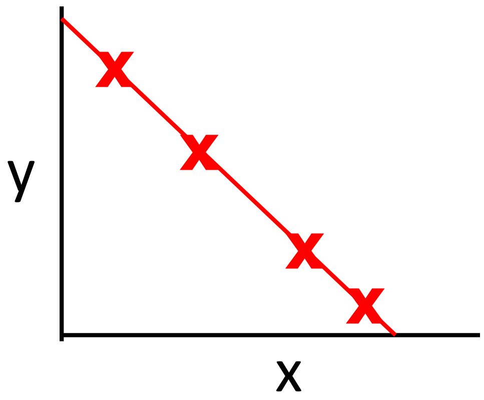
Problems with the LCM

Consider the sequence

- 0.2, 0.8, 0.7, 0.3, 0.4, 0.6, 0.9, 0.1

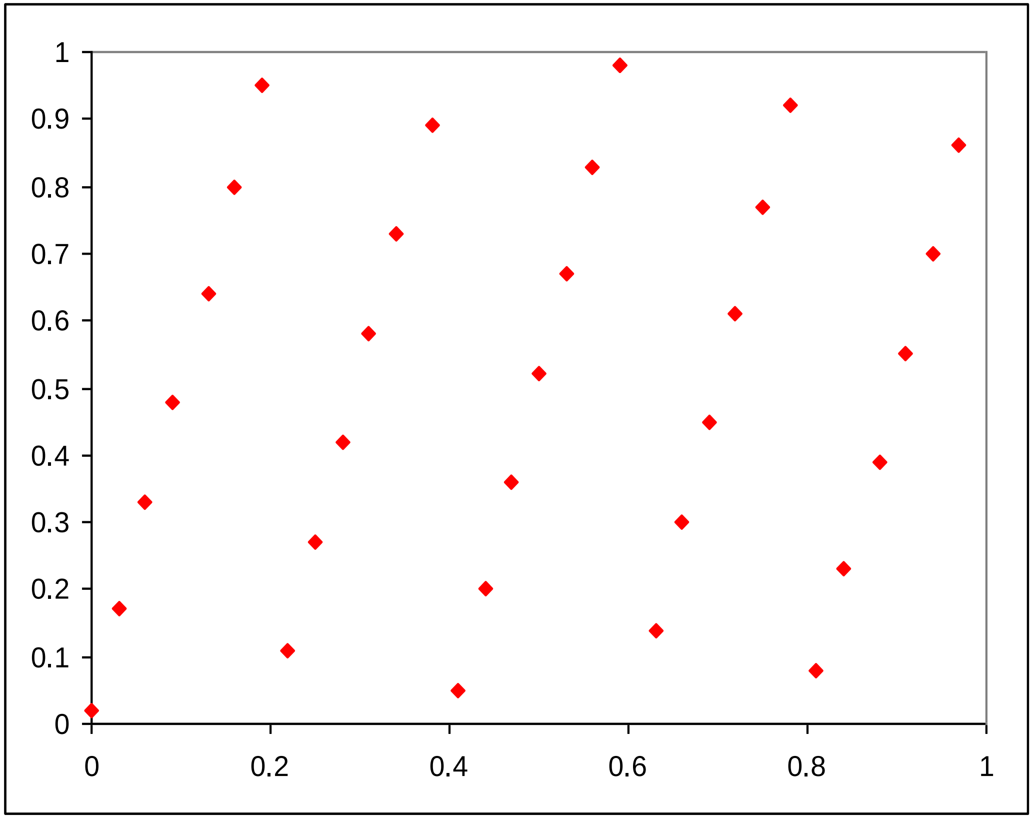
Interpret the numbers as (x,y) pairs:

(0.2, 0.8) (0.7, 0.3) (0.4, 0.6) (0.9, 0.1)



Problems with the LCM

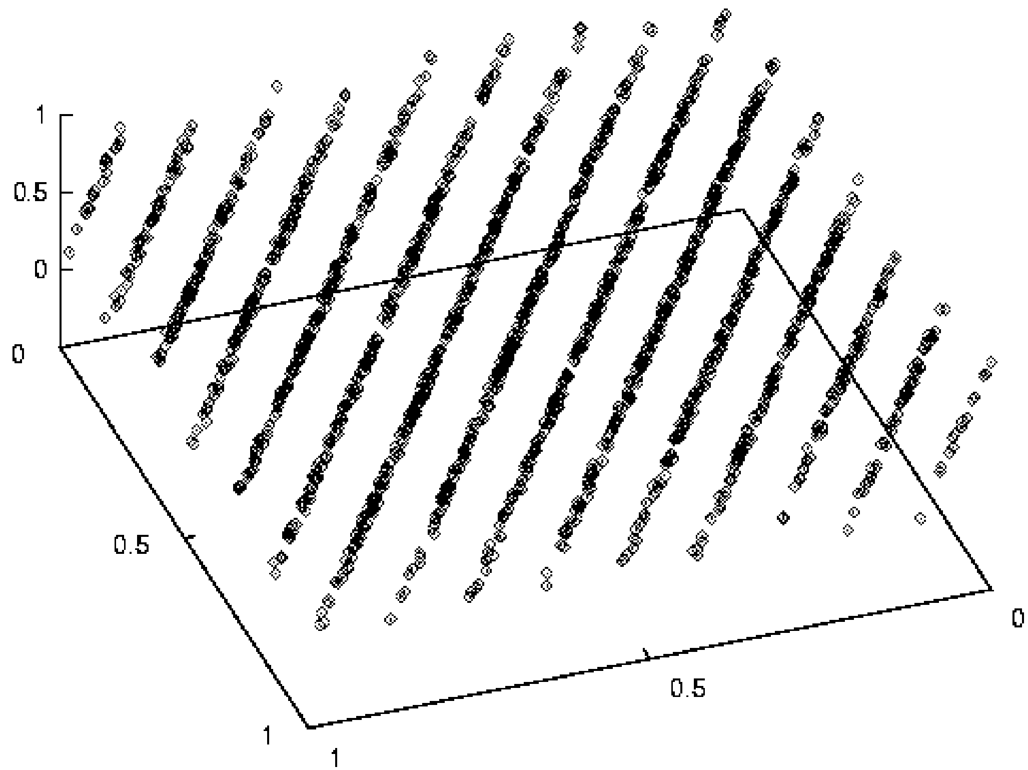
Plot results for LCM $a = 37$, $c = 1$, $m = 64$ in (x, y) pairs:



Problems with the LCM

The IBM generator RANDU: $a = 2^{16}+3$, $c = 0$, $m = 2^{31}$

- Plot results in (x, y, z) triplets:



Generating Random Variates

How to generate random numbers of any distribution:

Given:

- a $U(0,1)$ random number generator
- a cdf $F(x)$ that describes the desired distribution

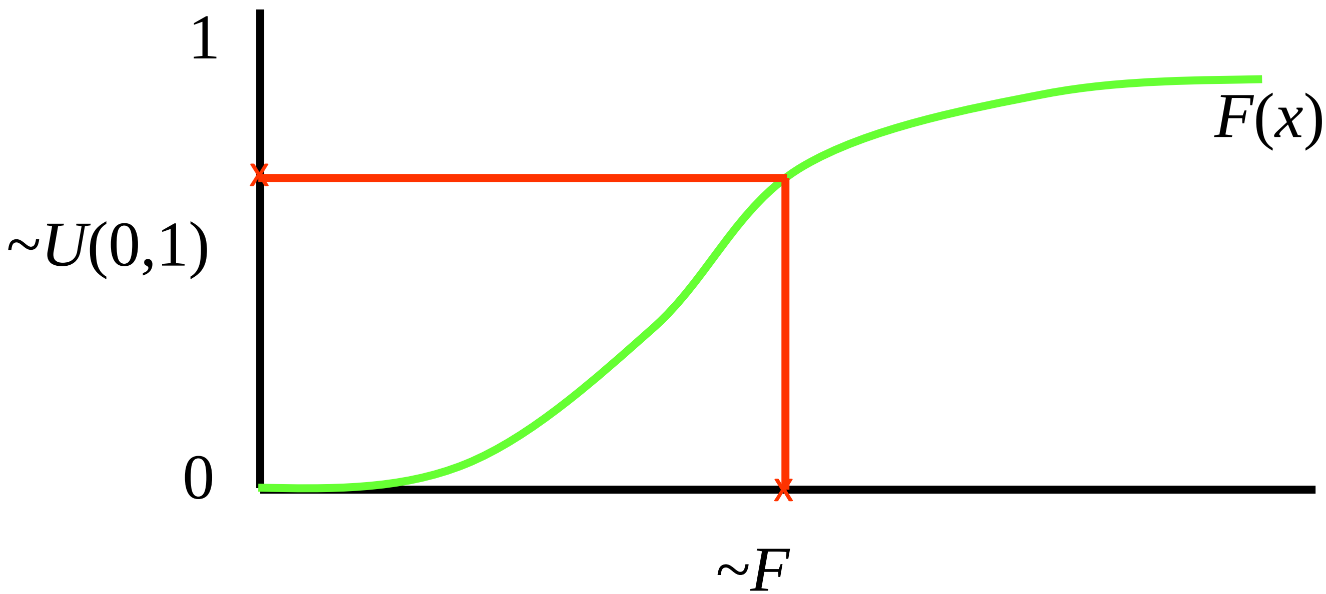
Then:

1. Generate $y \sim U(0,1)$
2. Compute $x = F^{-1}(y)$

Required when your simulator doesn't provide the distribution you need

Generating Random Variates

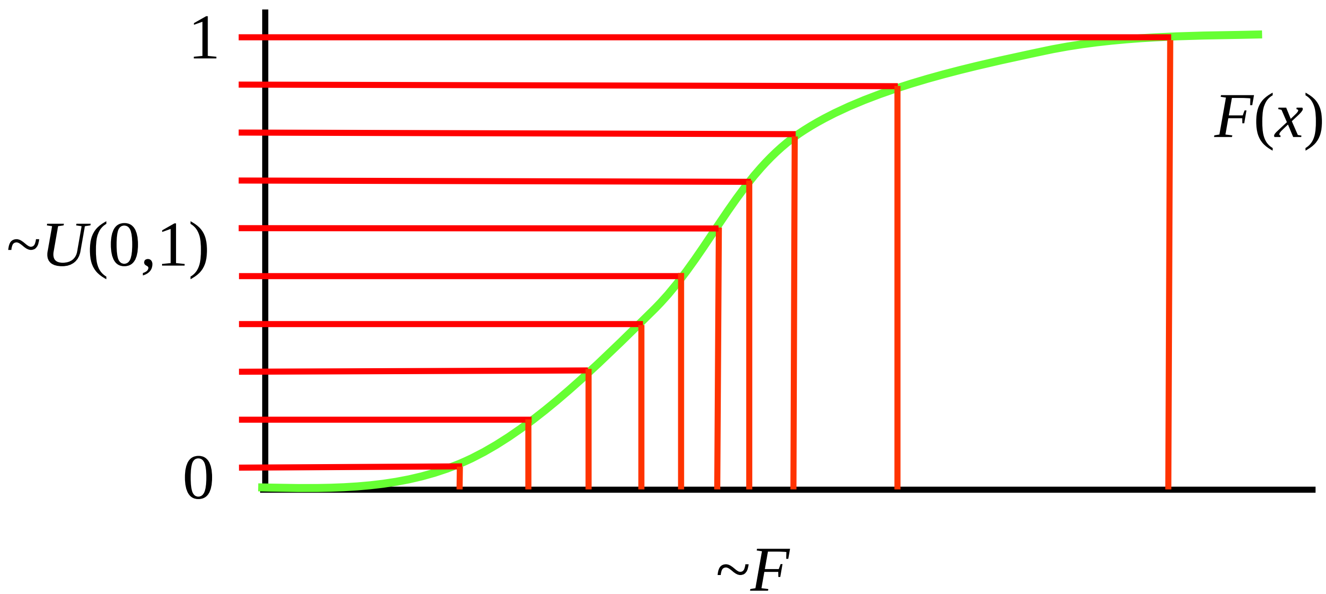
How to generate random numbers of any distribution:



Generating Random Variates

Effect of the method:

- To redistribute the uniformly distributed random numbers



Learning Goals

Learning questions:

- What is a probability density function?
- What is a probability distribution function?
- How are the two related?
- How are the PDF and CDF interpreted?
- How are the Exponential, Normal and Weibull distributions defined?
- Which probability distribution has the memoryless property? What does this mean in practice?
- What properties should a random number generator have?
- Explain how the LCM works