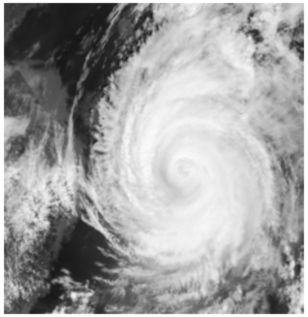











COST TU1402 Training School
on the Value of
Structural Health Monitoring Information

Brief Introduction to Probability Theory

Jochen Köhler

6.11.2017

NTNU
Norwegian University of
Science and Technology

Lecture overview

Lecture practicalities

Motivation

Probability

Bayes' Rule

Random Variables and Distributions

Contacts and Material

Lecturer

Jochen Köhler,
jochen.kohler@ntnu.no,

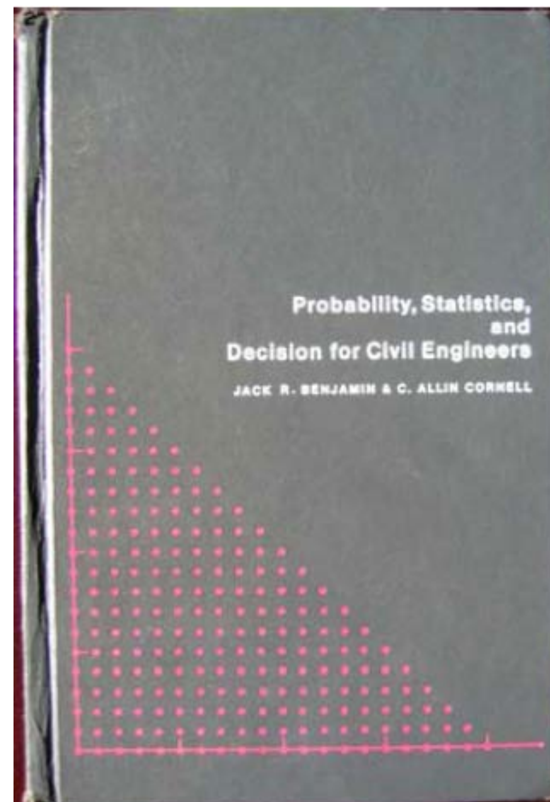
Material

The lecture is brief.
The interested student is invited to find further material in the corresponding *.zip folder.

Material

Benjamin, J. R. and C. A. Cornell (1970). *Probability, Statistics, and Decision for Civil Engineers.*

TOC



Material

Michael Havbro Faber, 2012,

Statistics and Probability Theory

TOC

1. Engineering Decisions Under Uncertainties

2. Basic Probability Theory

3. Descriptive Statistics

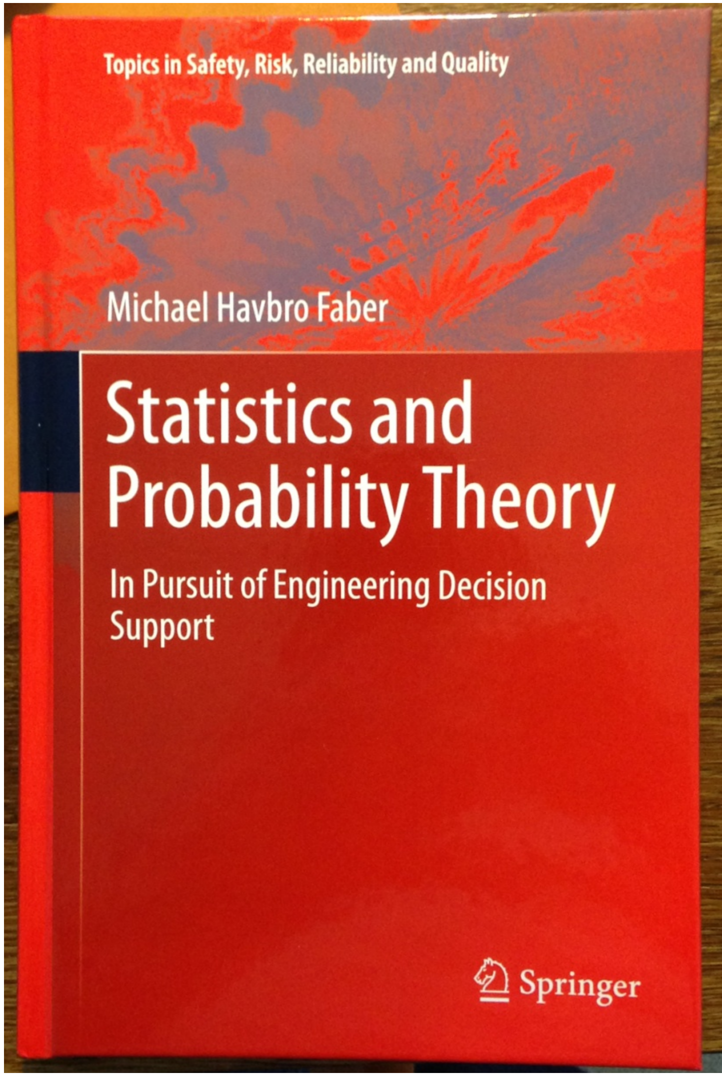
4. Uncertainty Modeling

5. Estimation and Model Building

6. Methods of Structural Reliability

7. Bayesian Decision Theory

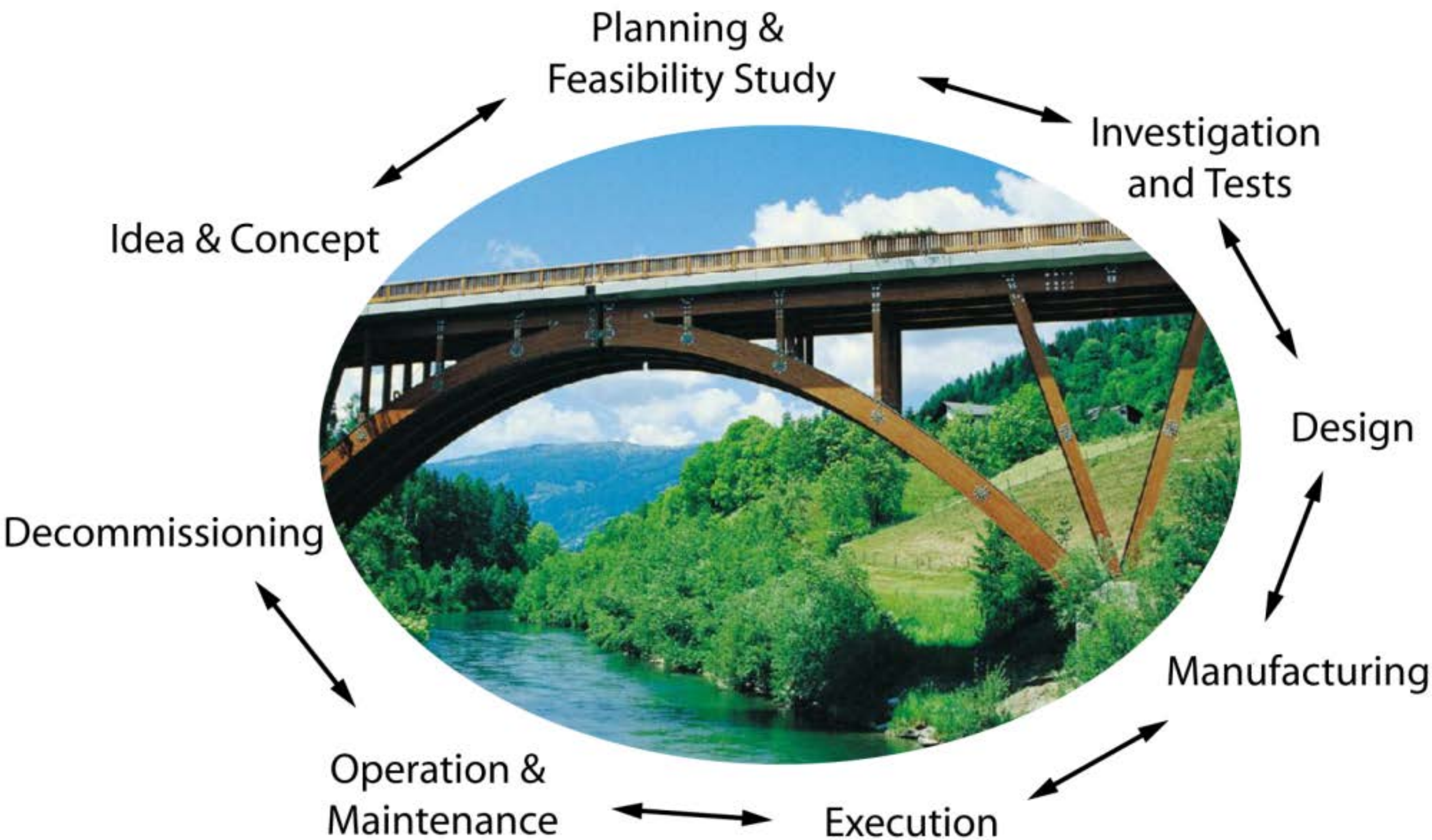
Processes



Build Environment

Build Environment

Structural lifecycle



What Structural Engineers do:

Decision making or support

- plan
- investigate
- dimension / design
- inspect
- maintain
- deconstruct

The build environment: e.g. dwellings, hospitals, schools, office buildings, industrial facilities , dams, bridges, tunnels.

Constrains:

assure

- safety for personnel and
- safety for environment
- cost effectiveness

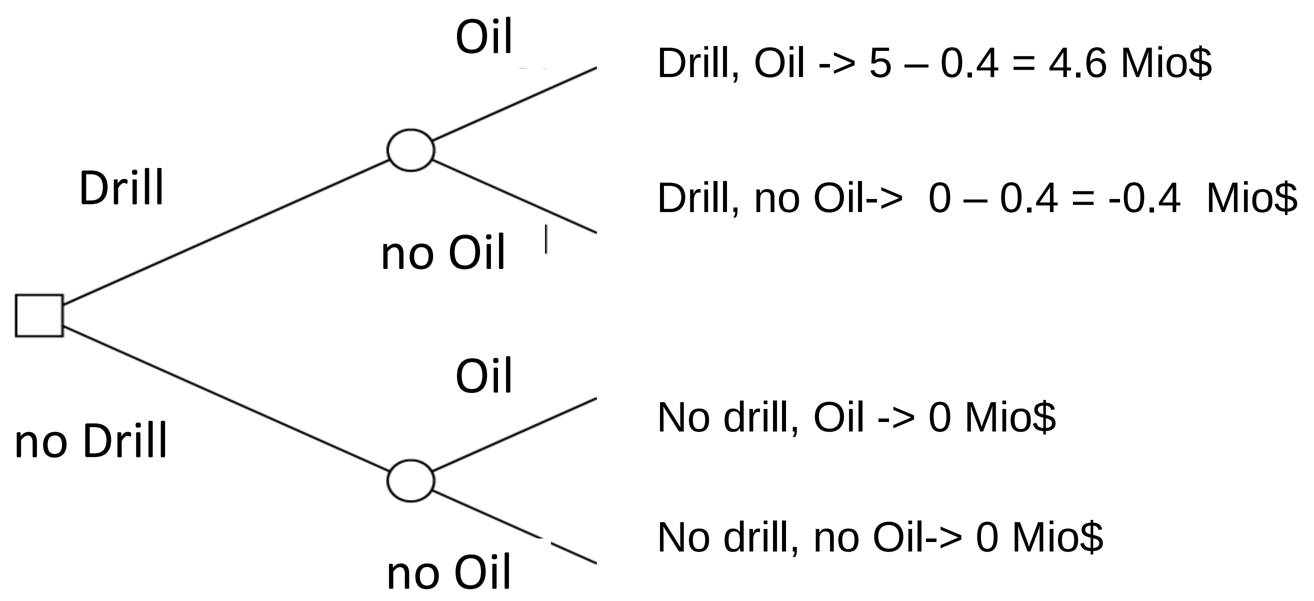
What is a decision problem ?

What is a decision problem ?

The oil wildcatter

$$R_A = 0.1 \cdot 4.6 + 0.9 \cdot (-0.4) = 0.1 = \sum_{i=1}^{n_E} R_{E_i} = \sum_{i=1}^{n_E} P_{E_i} \cdot C_{E_i}$$

$P_{oil} = 0.1$



Engineering =
Answering the basic questions of reasoning

- What can I know ?
- What shall I do ?
- What may I hope ?

Immanuel Kant (1724 – 1804)

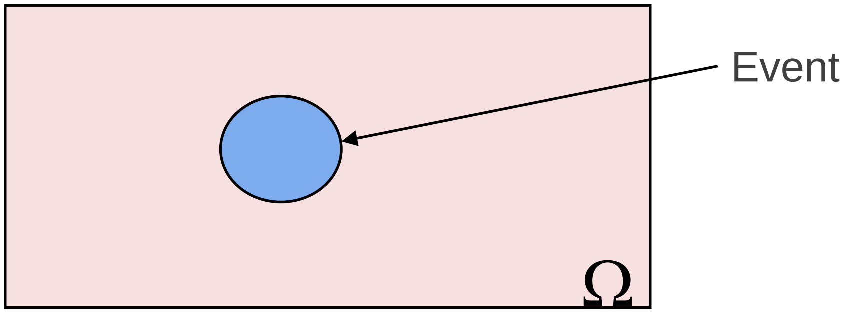
Prologue – Set Theory

The total set of all possible outcomes of an experiment is called event space Ω.

This can be written as follows:

$$\Omega =]-\infty;\infty[$$

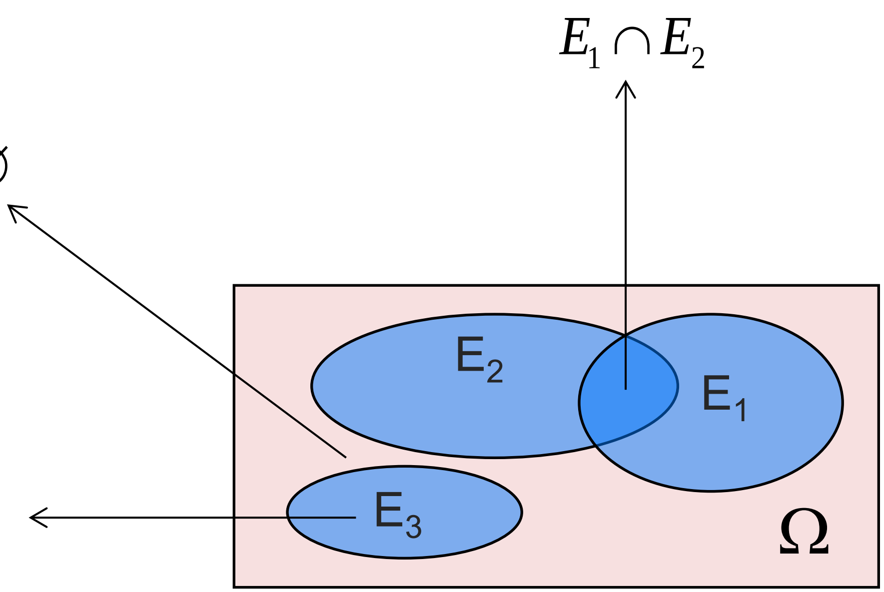
An event contains a specific collection of observations and is a subset of the event space.



Prologue – Set Theory

By using the Venn Diagram we can define different relation between events:

$E_1 \cap E_3 = \varnothing$



Unions:
$E_2 \cup E_3$

Probability - what is it?

Three different interpretations:

- The Frequency Interpretation of Probability

- The Classical Interpretation of Probability

- The Subjective Interpretation of Probability

The ”rules of calculus” are not affected by the interpretation !!

Probability – the Axioms

Axiom 1: For every event A, $Pr(A) \geq 0$, i.e. the probability of every event must be nonnegative.

Axiom 2: The probability of a certain event S is one; $Pr(S) = 1$.

Axiom 3: For every infinite sequence of disjoint events $A_1, A_2, \ldots$, $Pr\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} Pr(A_i)$.

Probability – the Axioms

The 3 axioms allow to explore the following properties of probability:

- $Pr(\varnothing) = 0$, the probability of the empty set is zero.
- $Pr\left(\bigcup_{i=1}^{n} A_i\right) = \sum_{i=1}^{n} Pr(A_i)$, Axiom 3 can be generalized to finite sequences of disjoint events.
- For every event A, $Pr(A^c) = 1 - Pr(A)$, where A^c is the complementary event of A.
- If $A \subset B$, then $Pr(A) \leq Pr(B)$.
- For every event A, $0 \leq Pr(A) \leq 1$.
- For every two events A and B,
  $Pr(A \cup B) = Pr(A) + Pr(B) - Pr(A \cap B)$.

Conditional Probability

A primary use of probability in engineering decision making is associated with updating probabilities based on observed events. The updated probability of the event A after we learn that event B has occurred is the conditional probability of A given B, $Pr\left(A|B\right)$.

Definition:

$$Pr\left(A|B\right) = \frac{Pr\left(A \cap B\right)}{Pr\left(B\right)}$$

Conditional Probability

From the commutative property of multiplication and intersection ($c * d = d * c$ and $A \cap B = B \cap A$)

the **Multiplication Rule for Conditional Probabilities** can be derived:

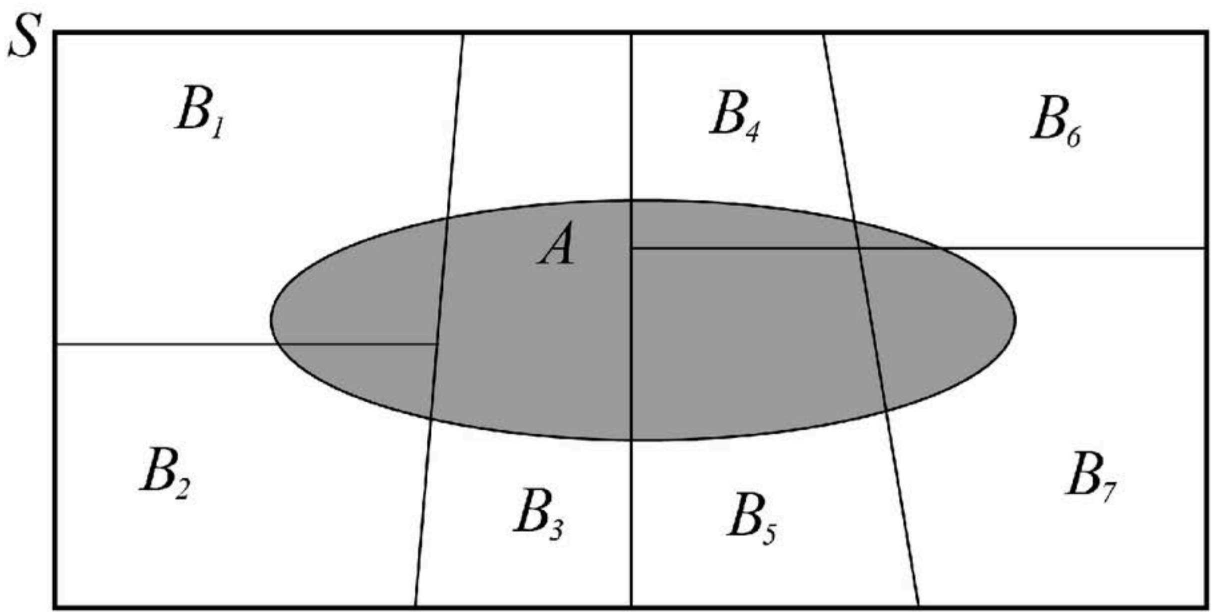
If $Pr(A) > 0$ and $Pr(B) > 0$ then,

$$Pr(A \cap B) = Pr(B)Pr(A|B) = Pr(A)Pr(B|A) = Pr(B \cap A)$$

Total Probability Theorem

A sample space S can be divided in k disjoint events $B_1, B_2, ..., B_k$ such that $\cup_{j=1}^{k} B_i = S$, i.e. B_i are the events form a partition of the sample space. For any event A in S,

$$Pr\left(A\right) = \sum_{j=1}^{k} Pr\left(A|B_j\right)Pr\left(B_j\right)$$



Independence

Definition of Independence

Two events A and B are independent if

$$Pr(A \cap B) = Pr(A)Pr(B)$$

and correspondingly

$Pr(A|B) = Pr(A)$ and $Pr(B|A) = Pr(B)$.

Bayes' Theorem

Suppose that we have k disjoint events $B_1, B_2, ..., B_j, ..., B_k$ and we observe an event A. If $Pr\left(A|B_j\right)$ are known the Bayes Theorem can be utilized to compute the conditional probability of B_j given A, $Pr\left(B_j|A\right)$.

Starting from the general definition of conditional probability and utilizing the Total Probability Theorem for replacing $Pr(A)$ Bayes' Theorem can be derived as:

$$Pr\left(B_i|A\right) = \frac{Pr\left(A \cap B_i\right)}{Pr\left(A\right)} = \frac{Pr\left(A|B_i\right)Pr\left(B_i\right)}{\sum_{j=1}^{k} Pr\left(A|B_j\right)Pr\left(B_j\right)}$$

Random Variables

Probability density function – Probability distribution function

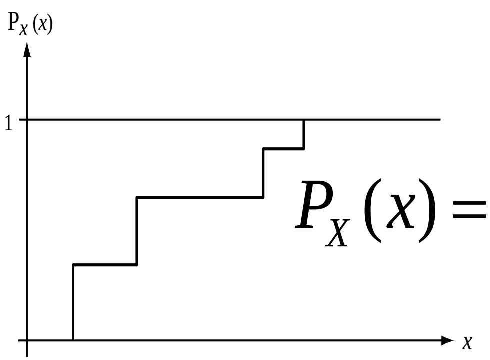
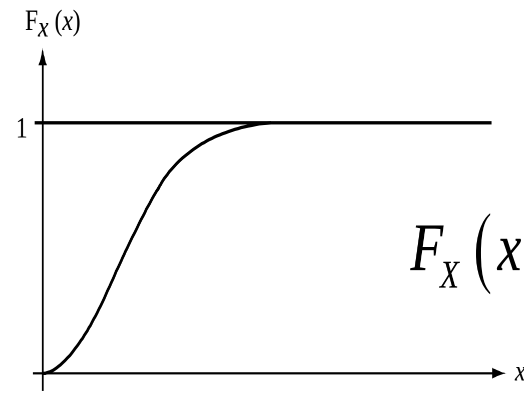
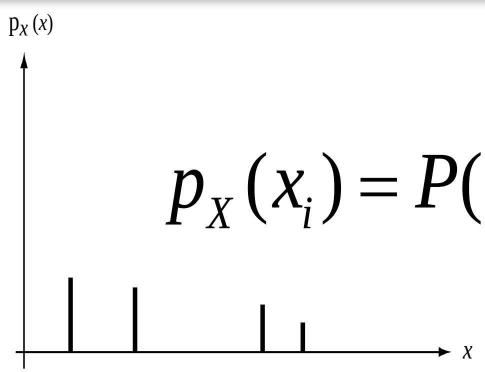
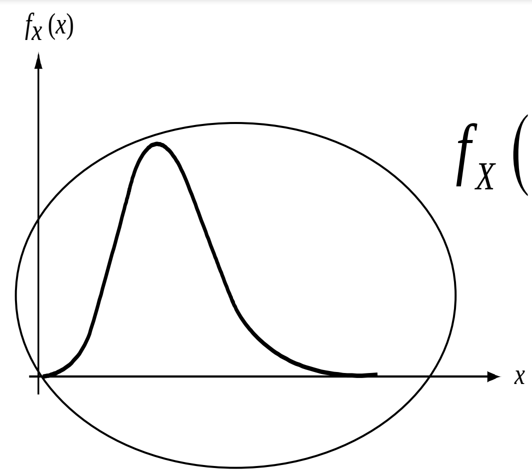
- A random variable is denoted by using large letters:

$$X$$

- A realization of a random variable is denoted by using small letters:

$$x$$

Random Variables

	Discrete	Continous
Distribution cdf	 $P_X(x) = \sum_{x_i < x} p_X(x_i)$	 $F_X(x) = P(X < x)$
Density pdf	 $p_X(x_i) = P(X = x_i)$	 $f_X(x) = \frac{\partial F_X(x)}{\partial x}$

Random Variables

Moments of random variables

- Probability distribution functions (density and cumulative distribution functions) are defined by their parameters or moments:

$$F_X(x, \mathbf{p}) \qquad f_X(x, \mathbf{p}) \qquad f_X(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2 \right)$$

Parameter

- The parameter of a distribution can be expressed by the moments of the distribution function and vice versa.

Moments of variables

- The ith moment of a continuous random variable is defined by:

$$\lambda_i = \int_{-\infty}^{\infty} x^i \cdot f_X(x)\,dx$$

- The ith moment of a discrete random variable is defined by

$$\lambda_i = \sum_{j=1}^{n} x_j^i \cdot p_X(x_j)$$

Central moments

- The i^th **central moment** of a continuous random variable is defined by:

$$\lambda_i = \int_{-\infty}^{\infty} \left(x^i - \mu\right) \cdot f_X\left(x\right) dx$$

- The i^th **central moment** of a discrete random variable is defined by

$$\lambda_i = \sum_{j=1}^{n} \left(x_j^i - \bar{x}\right) \cdot p_X\left(x_j\right)$$

The first moment is the mean - The first central moment is zero.

The second central moment is the variance

The third central moment is the skewness

The fourth central moment is the kurtosis

Expectation Operator

The **expectation operator** facilitates the calculation of the mean value and the variance of random variables.

$$E[X] = \int_{-\infty}^{\infty} x \cdot f_X(x)\,dx$$

$$Var[X] = \int_{-\infty}^{\infty} (x - \mu_X)^2 \cdot f_X(x)\,dx$$

This is especially important for a compact notation and communication among experts and reading of reports.

The expectation operator is often used when dealing with functions of random variables.

Expectation Operator

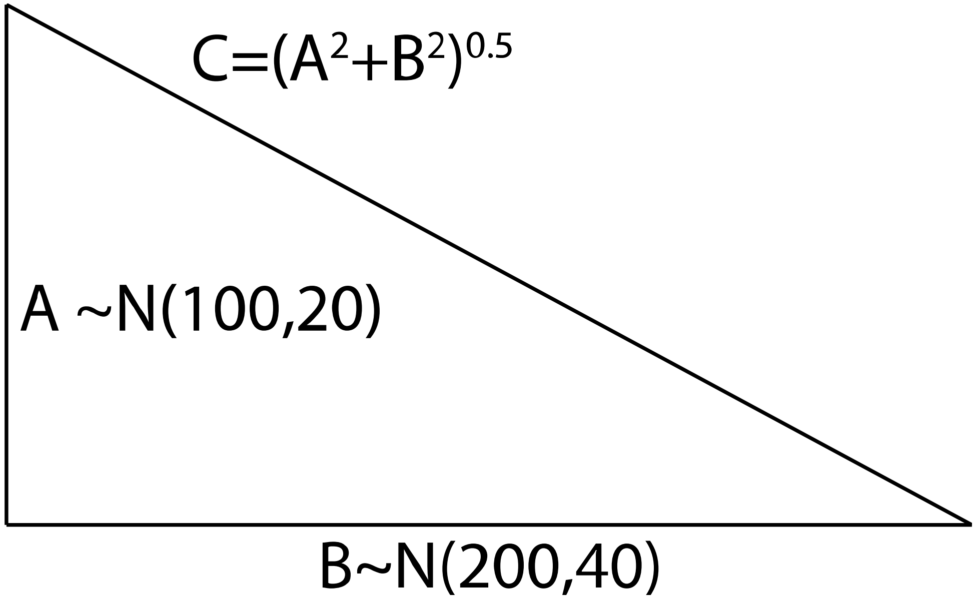
JENSEN'S Inequality!!

$$E[g(X)] \neq g(E[X])$$

Equality only for the rare case of linear functions.

Expectation Operator

$$E\left[\sqrt{A^2+B^2}\right] \approx 225.1 \quad \neq \quad \sqrt{100^2+200^2} = 223.6$$



JENSEN'S Inequality!!

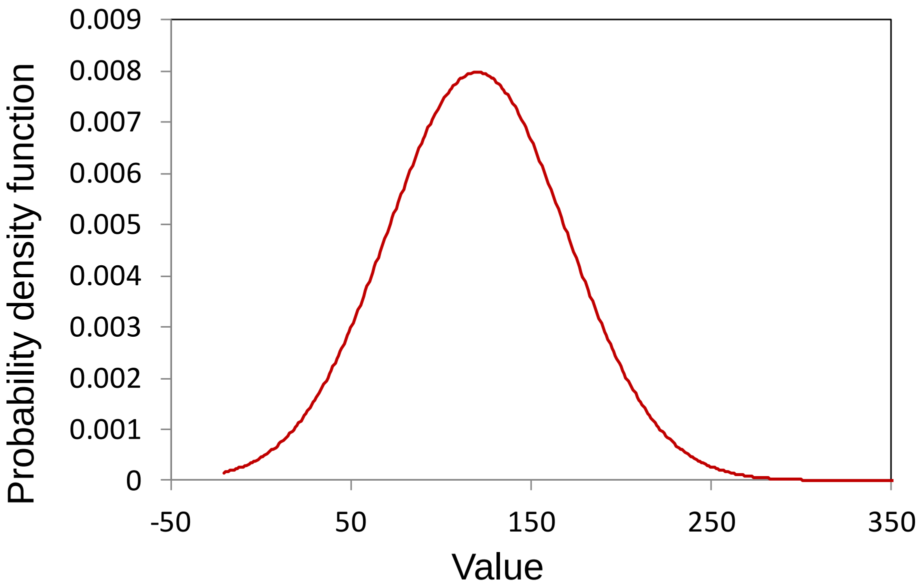
Distribution

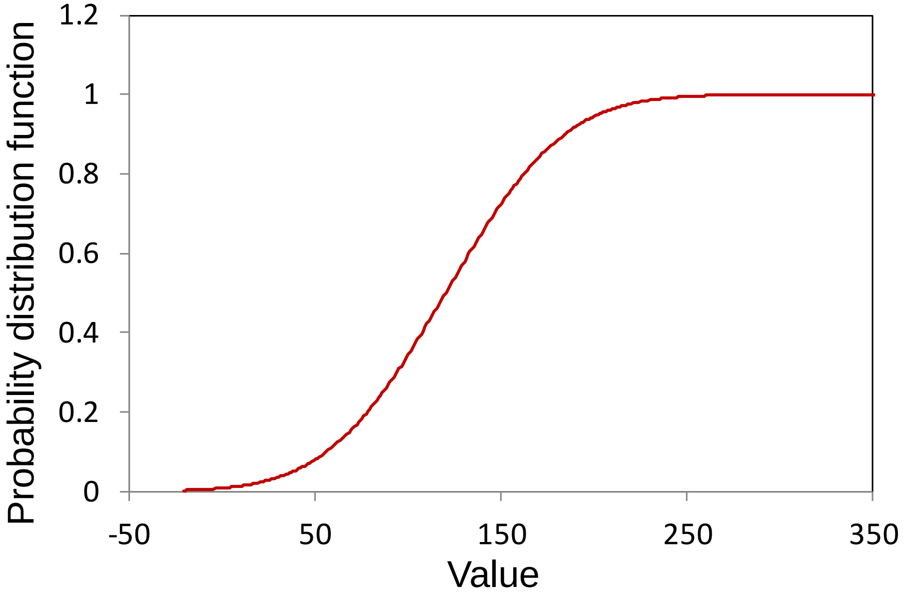
Normal distribution

$$f_X(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2 \right)$$

$$F_X(x) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{x} \exp\left(-\frac{1}{2}\left(\frac{s-\mu}{\sigma}\right)^2 \right) ds$$

μ=120; σ=50





Distribution

Log-Normal distribution

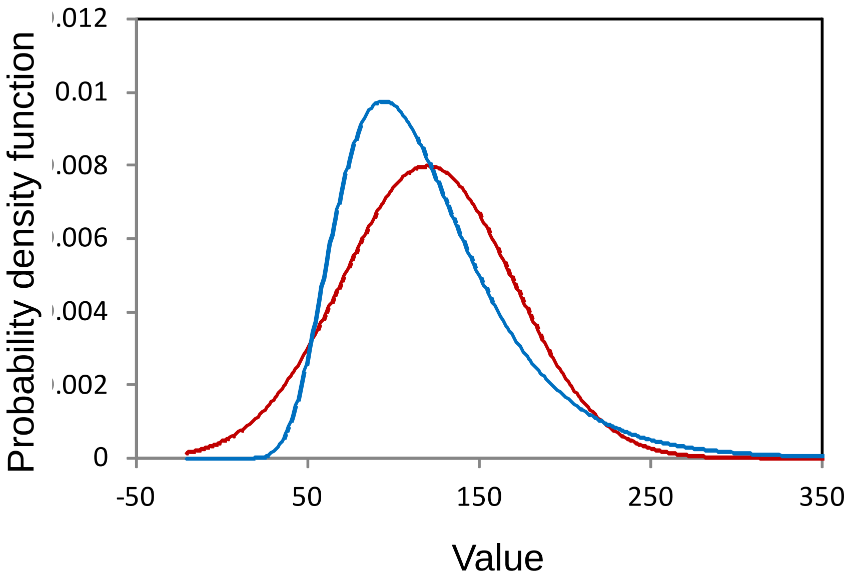
$$f_X(x) = \frac{1}{x\zeta\sqrt{2\pi}} \exp\left(-\frac{1}{2}\left(\frac{\ln(x)-\lambda}{\zeta}\right)^2\right)$$

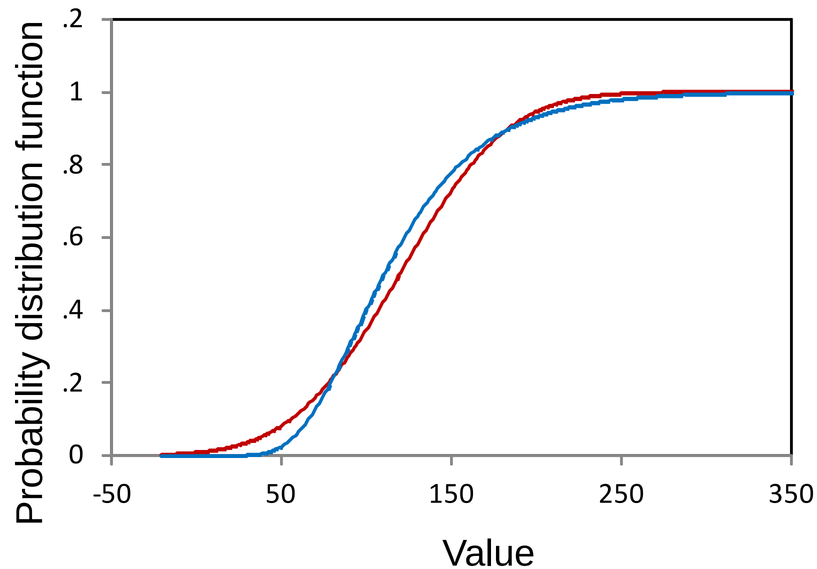
$$F_X(x) = \Phi\left(\frac{\ln(x)-\lambda}{\zeta}\right)$$

$$\mu = \exp\left(\lambda + \frac{\zeta^2}{2}\right)$$

$$\sigma = \mu\sqrt{\exp\left(\zeta^2\right)-1}$$

μ=120; σ=50





Distribution

Gamma distribution

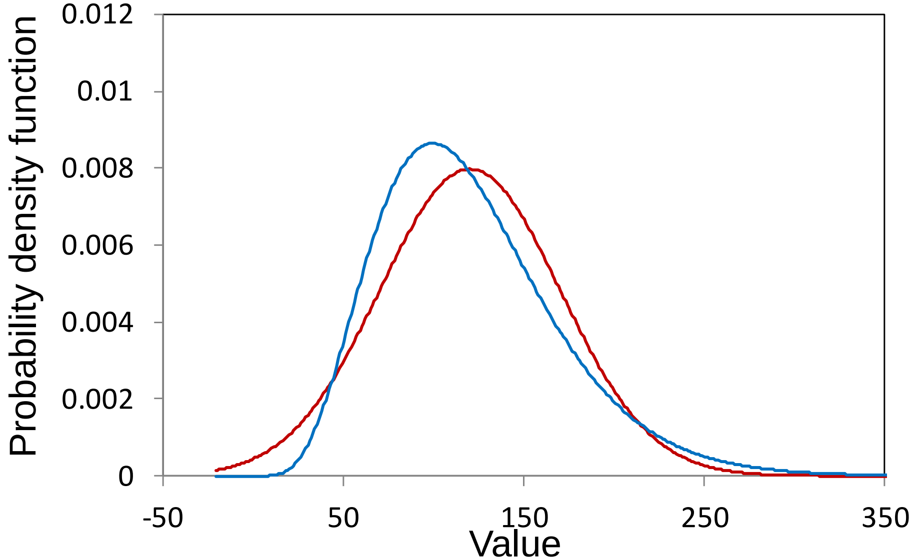
$$f_X(x) = \frac{\lambda(\lambda x)^{k-1}}{\Gamma(k)} \exp(-\lambda x)$$

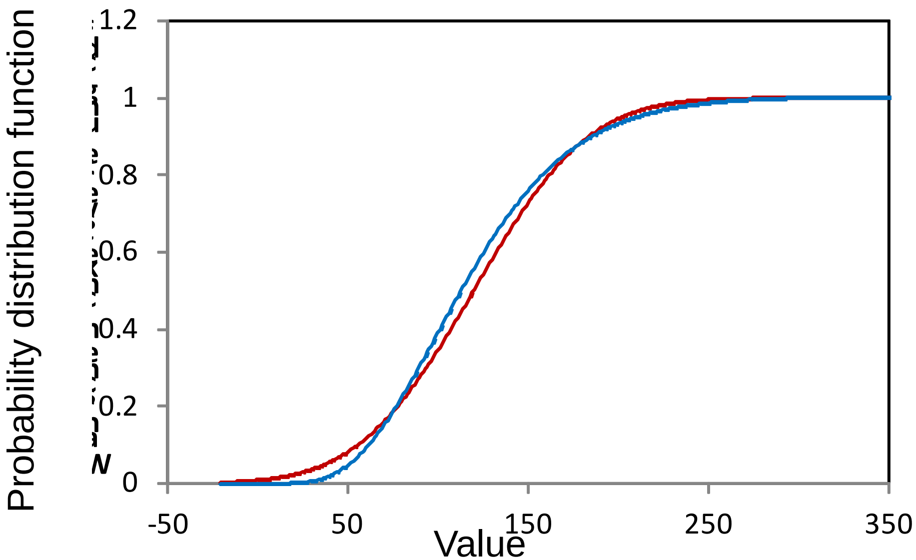
$$F_X(x) = \frac{\Gamma(k,\lambda x)}{\Gamma(k)} \quad , \Gamma(k,t) = \int_0^t \exp(-u)\, u^{k-1}\, du$$

$$\mu = \frac{k}{\lambda}$$

$$\sigma = \frac{\sqrt{k}}{\lambda}$$

μ=120; σ=50





Distribution

Uniform distribution

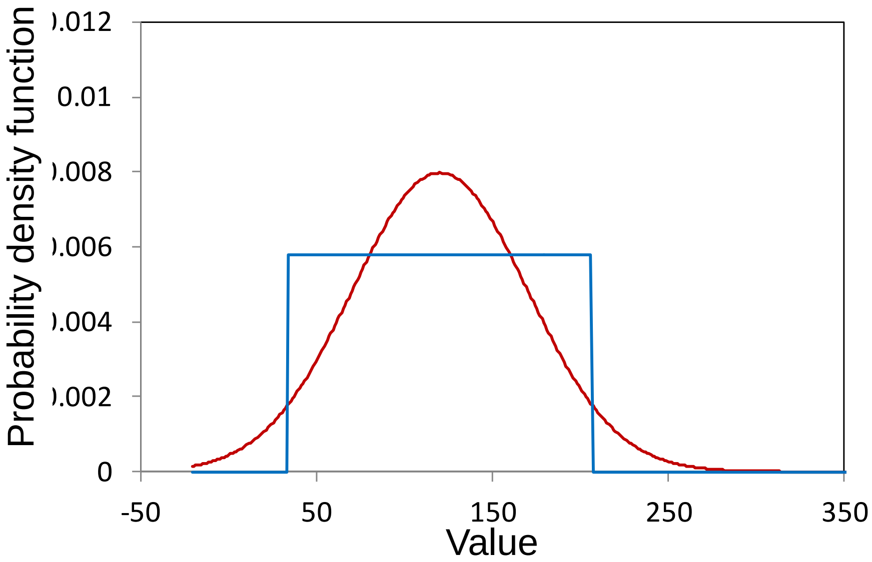
$$f_X(x) = \frac{1}{b-a}$$

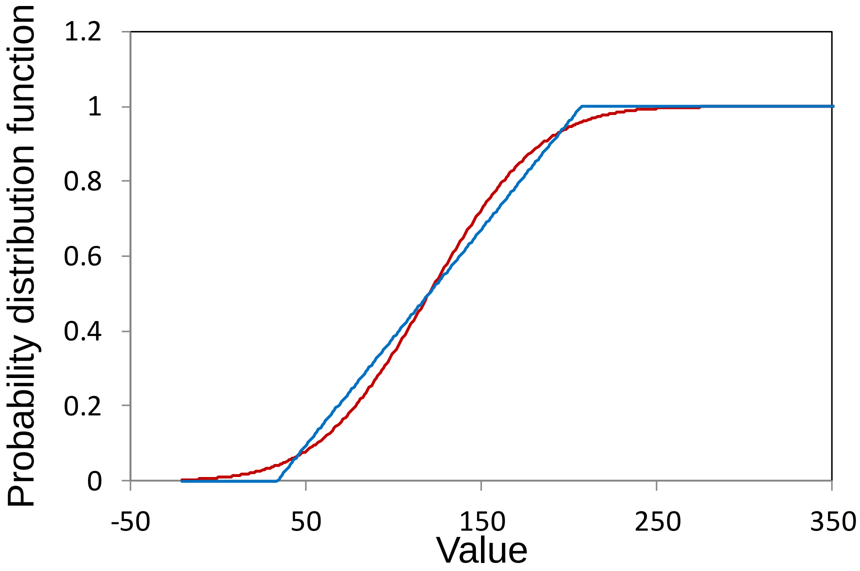
$$F_X(x) = \frac{x-a}{b-a}$$

$$\mu = \frac{a+b}{2}$$

$$\sigma = \frac{b-a}{\sqrt{12}}$$

μ=120; σ=50





Distribution

Exponential distribution

$$f_X(x) = \lambda \exp(-\lambda(x))$$

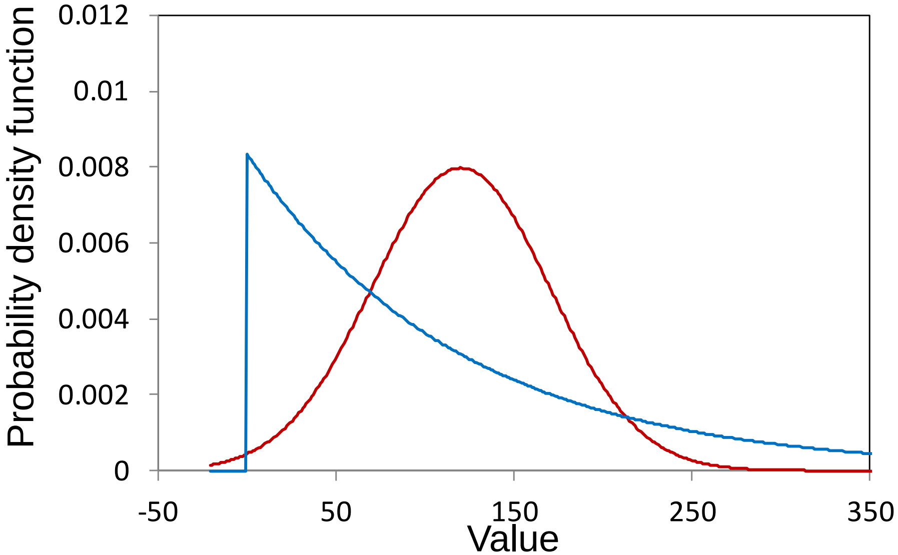
$$F_X(x) = 1 - \exp(-\lambda x)$$

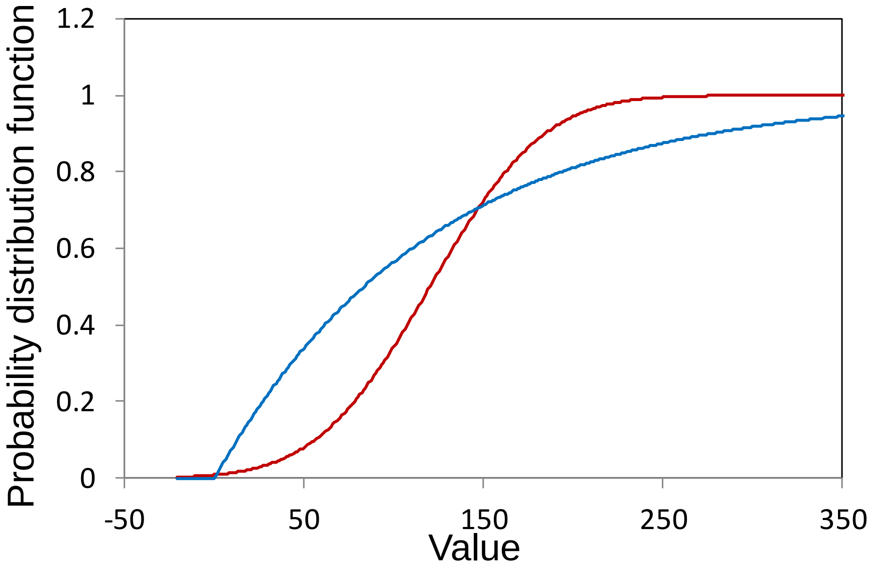
$$\mu = \frac{1}{\lambda}$$

$$\sigma = \frac{1}{\lambda}$$

μ=120; σ=120

only 1 parameter





Distribution

Beta distribution

$$f_X(x) = \frac{\Gamma(r+t)}{\Gamma(r)\cdot\Gamma(t)} \cdot \frac{(x-a)^{r-1}(b-x)^{t-1}}{(b-a)^{r+t-1}}$$

$$F_X(x) = \frac{\Gamma(r+t)}{\Gamma(r)\cdot\Gamma(t)} \int_a^u \frac{(x-a)^{r-1}(b-x)^{t-1}}{(b-a)^{r+t-1}} dx$$

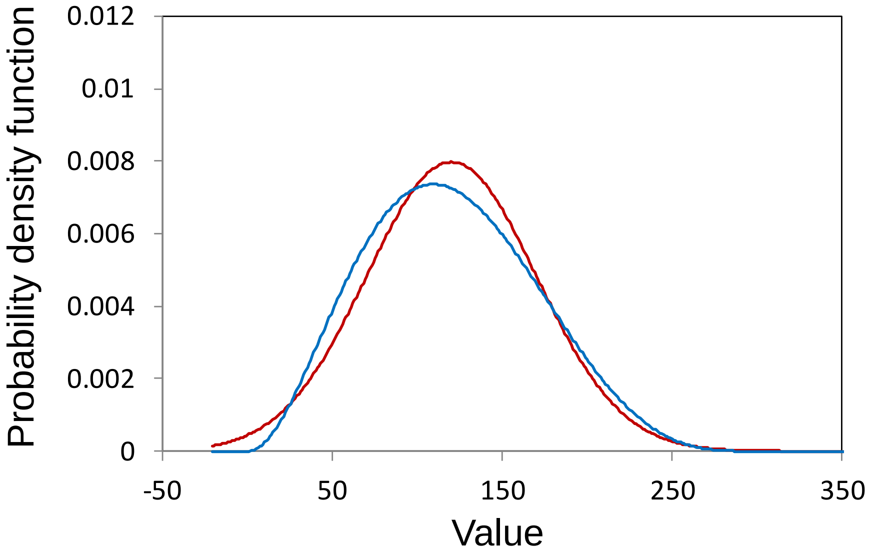
$$\Gamma(x) = \int_0^\infty e^{-t} t^{x-1} dt$$

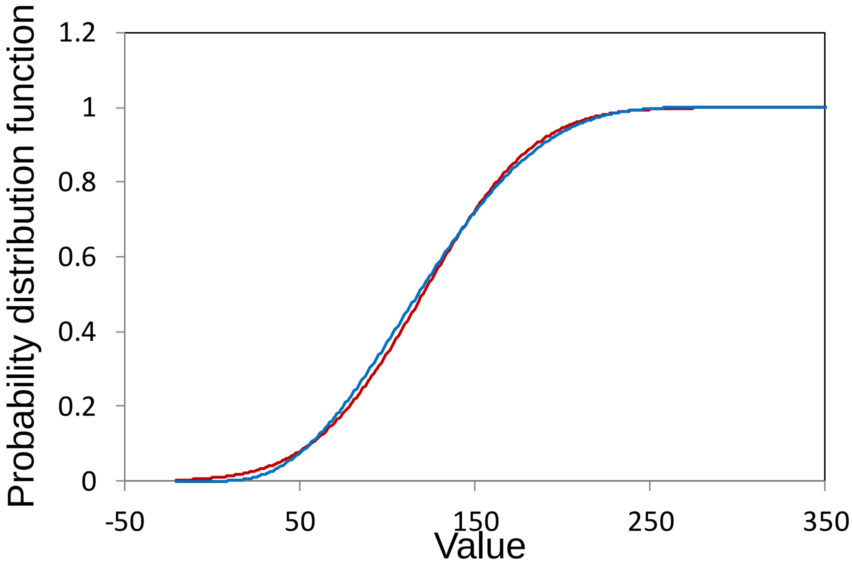
$$\mu = a + (b-a)\frac{r}{r+t}$$

$$\sigma = \frac{b-a}{r+t}\sqrt{\frac{rt}{r+t+1}}$$

μ=120; σ=50; a=0; b=300

4 parameter





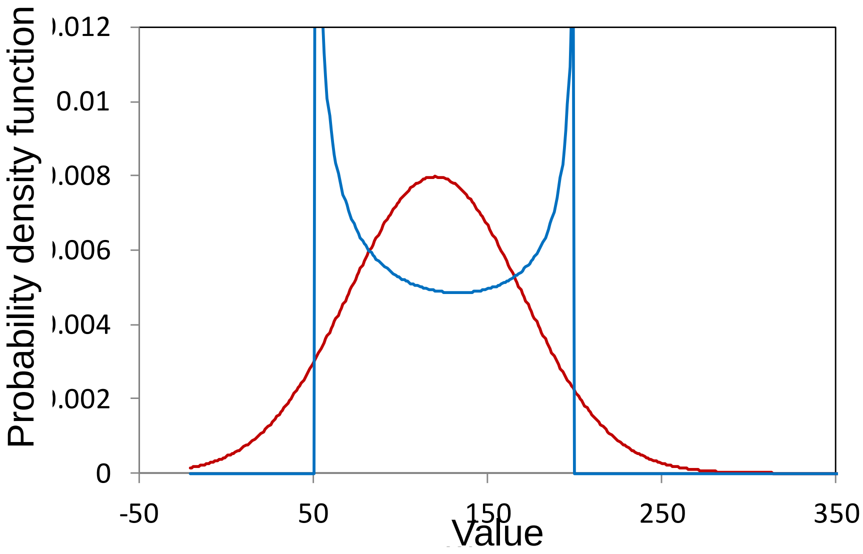
Distribution

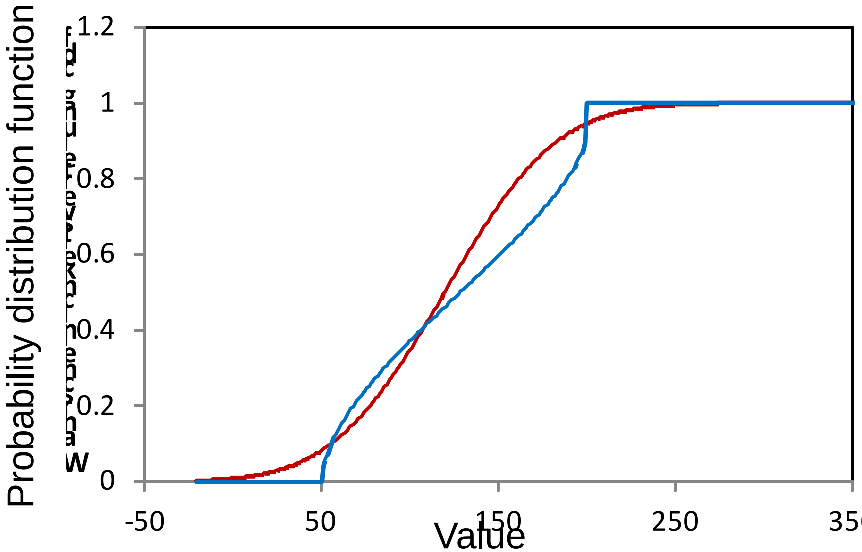
Beta distribution

$$f_X(x) = \frac{\Gamma(r+t)}{\Gamma(r)\cdot\Gamma(t)} \cdot \frac{(x-a)^{r-1}(b-x)^{t-1}}{(b-a)^{r+t-1}}$$

$$F_X(x) = \frac{\Gamma(r+t)}{\Gamma(r)\cdot\Gamma(t)} \int_a^u \frac{(x-a)^{r-1}(b-x)^{t-1}}{(b-a)^{r+t-1}}\, dx$$

μ=120; σ=50; a=50; b=200 → very flexible





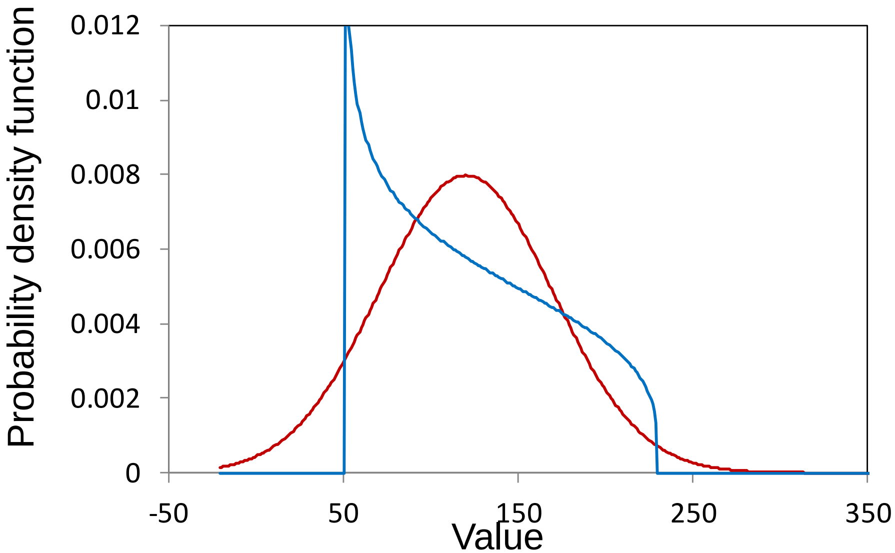
Distribution

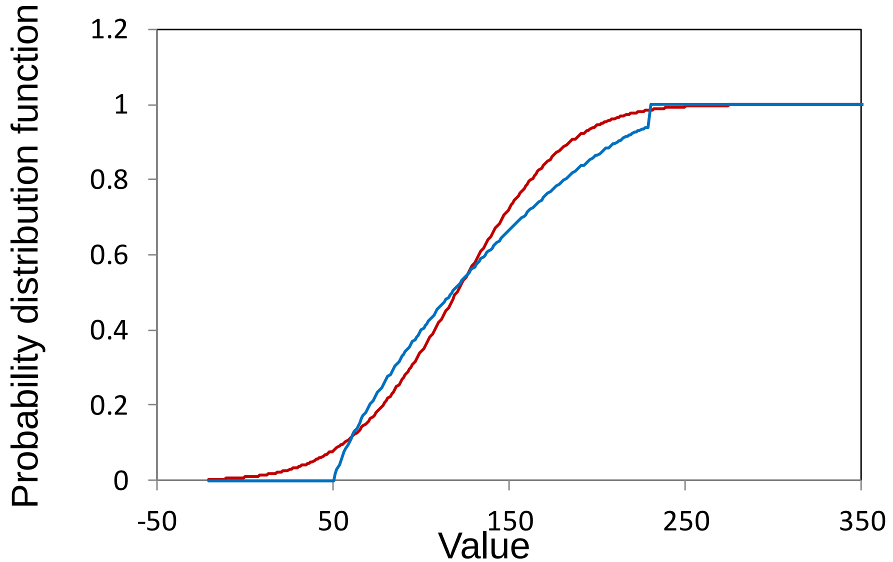
Beta distribution

$$f_X(x) = \frac{\Gamma(r+t)}{\Gamma(r)\cdot\Gamma(t)} \cdot \frac{(x-a)^{r-1}(b-x)^{t-1}}{(b-a)^{r+t-1}}$$

$$F_X(x) = \frac{\Gamma(r+t)}{\Gamma(r)\cdot\Gamma(t)} \int_a^u \frac{(x-a)^{r-1}(b-x)^{t-1}}{(b-a)^{r+t-1}}\, dx$$

μ=120; σ=50; a=50; b=230 → very flexible





Distribution

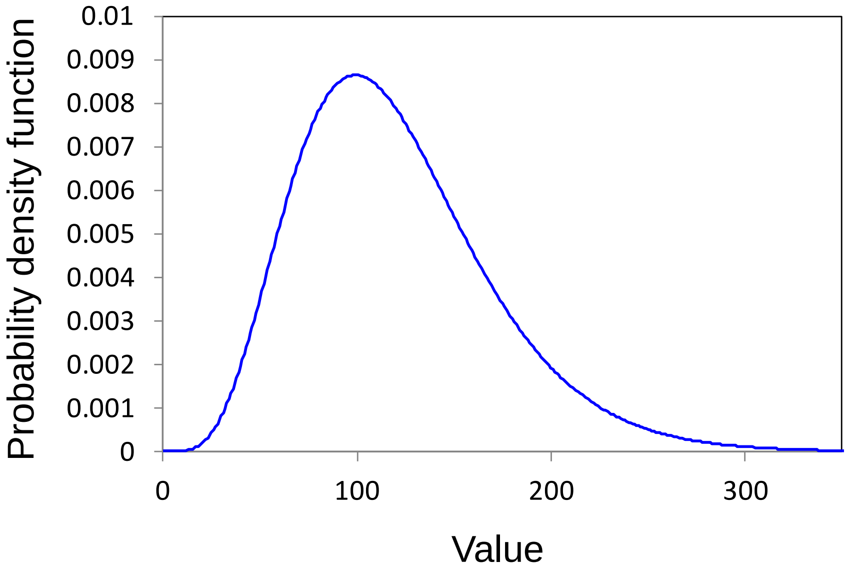
Gamma distribution

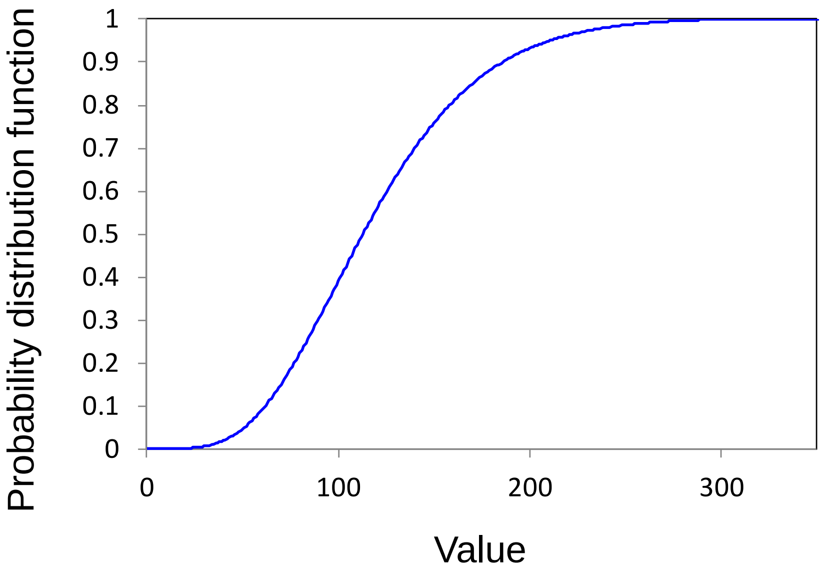
$$f_X(x|k,\theta) = \frac{1}{\theta^k}\frac{1}{\Gamma(k)} x^{k-1} e^{-\frac{x}{\theta}} \qquad x,k,\theta > 0 \qquad F_X(x|k,\theta) = \int_0^x \frac{1}{\theta^k}\frac{1}{\Gamma(k)} u^{k-1} e^{-\frac{u}{\theta}}\, du$$

$$E[X] = k\theta$$

$$Var[X] = k\theta^2$$

μ=120; σ=50





Some guidance on the use of distributions

Normal distribution

The sum of independent random variables converges to the normal distribution (**The central limit theorem CLT**); Defined between – infinity and + infinity.

Lognormal distribution

The product of independent (positive) random variables converges to the log-normal distribution; defined between zero and +infinity. Used e.g. for material strength.

Some guidance on the use of distributions

Exponential distribution

Describes the time between the occurrence of two events which follows a Poisson process.
Used e.g. for modeling the mean time between failures.

Uniform distribution

Used for modeling events which are equal probable in a defined interval.

Some guidance on the use of distributions

Gamma distribution

Describes the time until the k^{th} event of a Poisson process occurred. Frequently used to describe observations.

Beta distribution

Very flexible and used to model observation of any kind in a specific interval.

Conditional Density Functions

A probability density function can be expressed conditional to the parameters θ:

$$f_X(x|\theta)$$

This is especially important if the parameters are not known or uncertain.

If the uncertainty about the parameters is represented by a probability density $\pi(\theta)$ the **Total Probability Theorem** can be applied in order to find the so called **Predictive Density Function**:

$$f_X(x) = \int_\theta f_X(x|\theta)\pi(\theta)d\theta$$

The posterior probability (density) function for θ is

$$\pi(\theta|\boldsymbol{x}) = \frac{\pi(\theta)\, f(\boldsymbol{x}|\theta)}{f(\boldsymbol{x})}$$

where

$$f(\boldsymbol{x}) = \begin{cases} \int_\Theta \pi(\theta)\, f(\boldsymbol{x}|\theta)\, d\theta & \text{if } \theta \text{ is continuous,} \\ \\ \sum_\Theta \pi(\theta)\, f(\boldsymbol{x}|\theta) & \text{if } \theta \text{ is discrete.} \end{cases}$$

Notice that, as $f(\boldsymbol{x})$ is not a function of θ, Bayes Theorem can be rewritten as

$$\pi(\theta|\boldsymbol{x}) \propto \pi(\theta) \times f(\boldsymbol{x}|\theta)$$

$$i.e. \quad \text{posterior} \propto \text{prior} \times \text{likelihood.}$$

Thanks for attention

Jochen Köhler

NTNU
Norwegian University of
Science and Technology