Introduction into Probabilistic Methods and their Application in Engineering Sciences with Focus on Monte-Carlo and Response Surface Methods

David Pusch
André Beschorner
Robin Schmidt

7. Dresdner Probabilistik-Workshop
Dresden 08. – 09.10.2014

DRESDEN concept
Exzellenz aus Wissenschaft und Kultur

Outline

- Introduction
- Part 1: Basics of Statistics
- Part 2: Regression
- Part 3: Probabilistic System Analysis using Monte-Carlo Methods

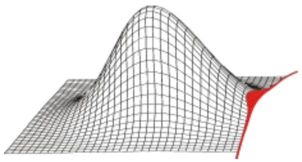
Motivation

Two turbine blades from the same engine...

Massachusetts Institute of Technology, Prof. David L. Darmofal

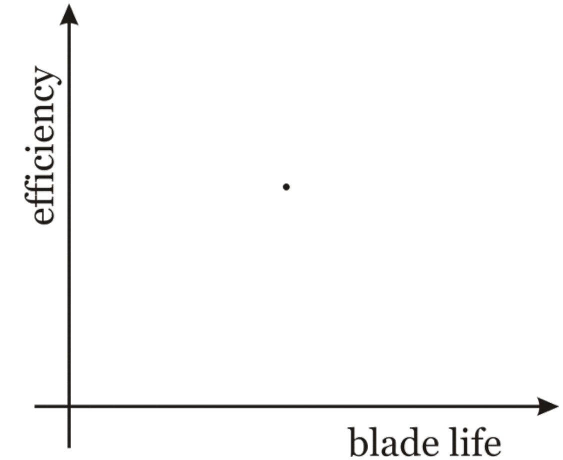
... but with clearly different life

Simulation vs. Reality



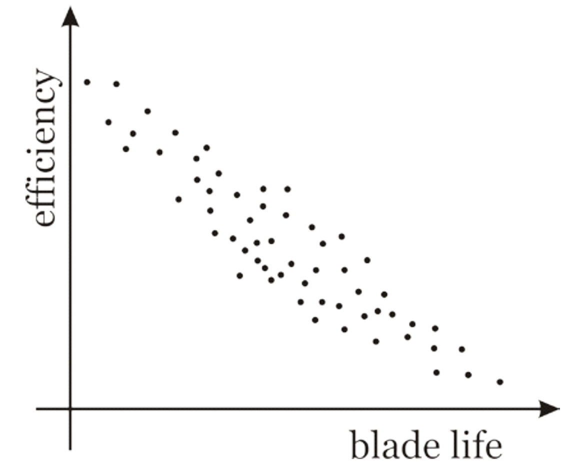
- Classic simulation-based approach used in design

Design Intent → Simulation →



- Real behavior of physical system

Design Intent → Reality →



Stochastic Theory

Combinatorics

How many possibilities are there to arrange elements, or to pick elements from a population?

Statistics

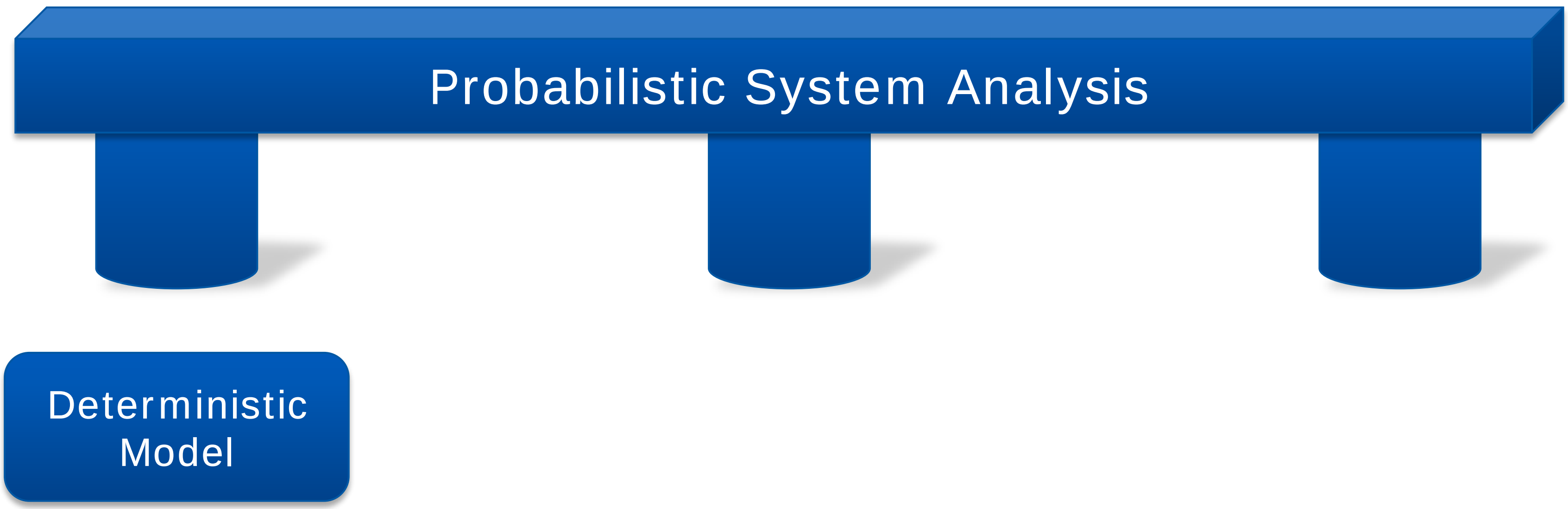
collection, analysis, interpretation, presentation and organization of data

Probabilistic

investigation and modeling of random events

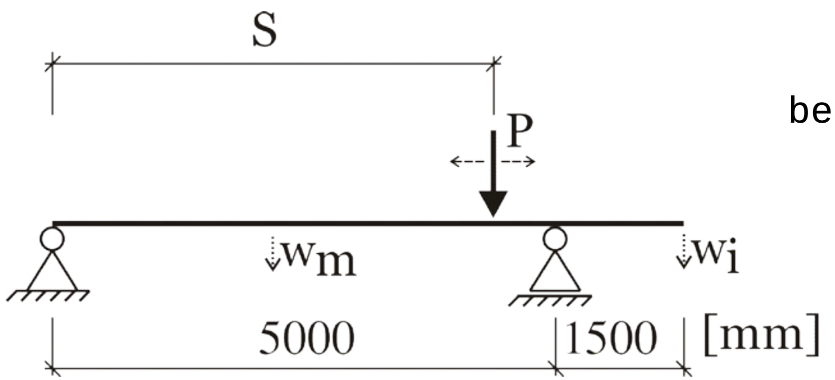
prediction of random events

Requirements for Probabilistic Simulations

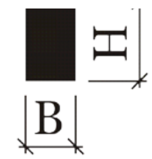


Example of a Deterministic Model

cantilever beam with two supports



beam cross-section

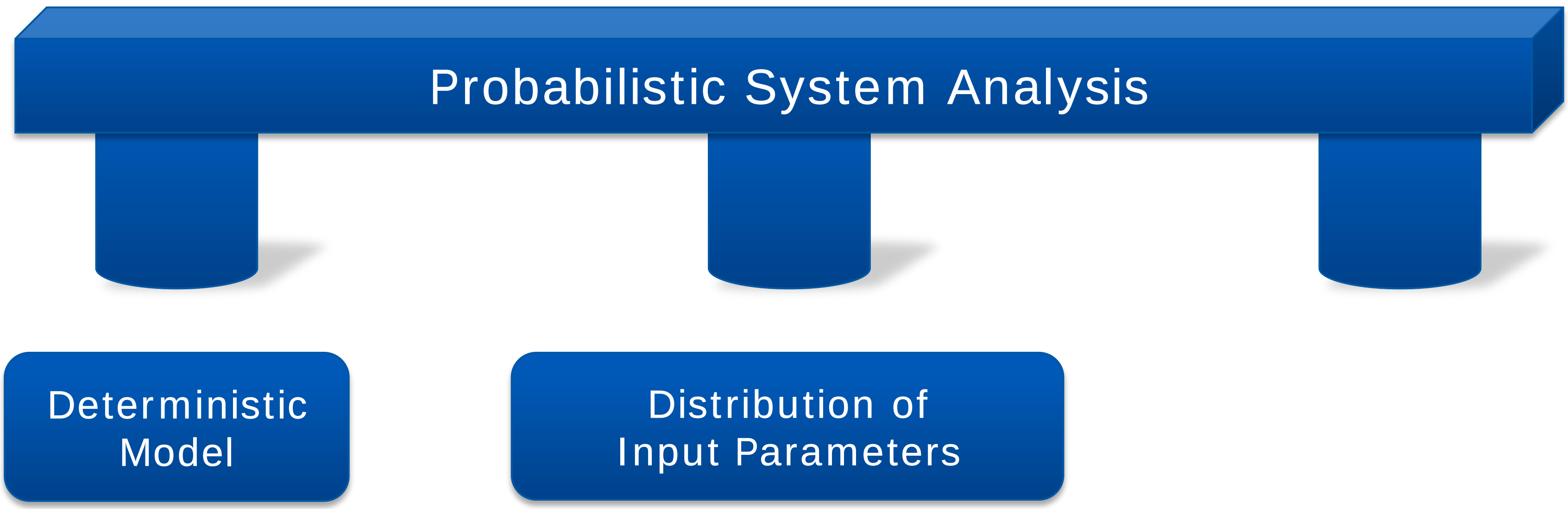


input parameters:
- H ... beam height
- W ...beam width
- P ... point load
- S ... position of point load
- Young's modulus

→ Beam Theory →

result values:
- deflections w_m, w_i

Requirements for Probabilistic Simulations



random experiment

random variable

discrete | continuous

probability mass function (pmf)

probability density function (pdf)

$$F(b) = \sum_{b_{ri} \leq \boldsymbol{b}} f(b_{ri})$$

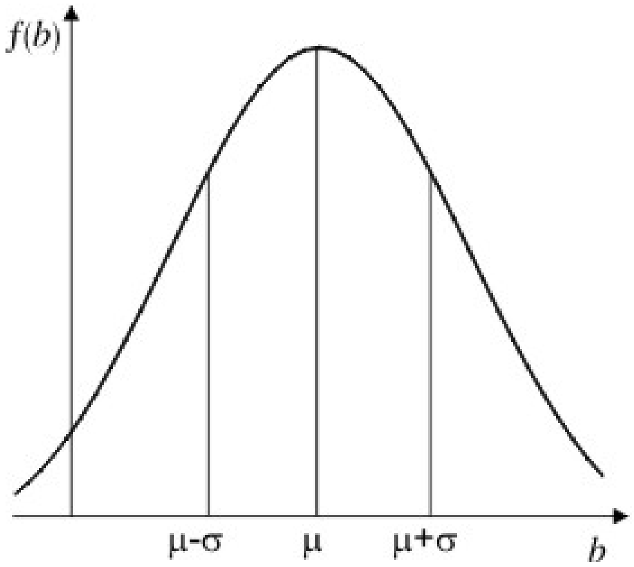
cummulative density function (cdf)

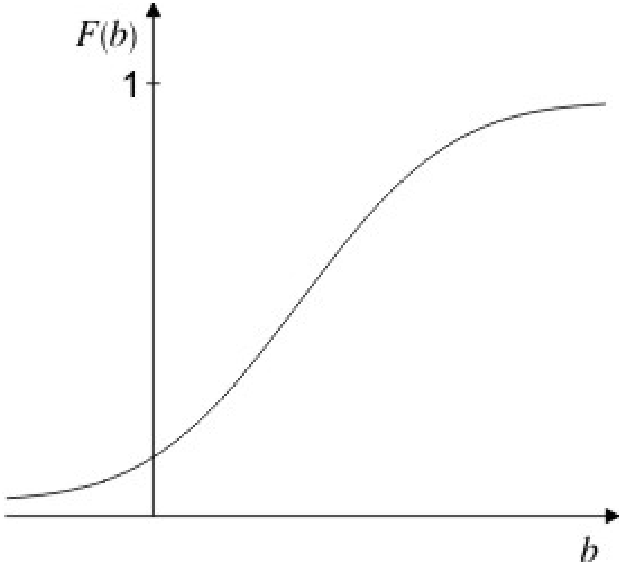
$$F(b) = \int_{-\infty}^{\boldsymbol{b}} f(\tilde{b})\,d\tilde{b}$$

$$F(\boldsymbol{b}) = P(b \leq \boldsymbol{b})$$

[1]

also called "Normal Distribution"





$$f(b) = \frac{1}{\sigma\sqrt{2\pi}} exp\left[-\frac{(b-\mu)^2}{2\sigma^2}\right]$$

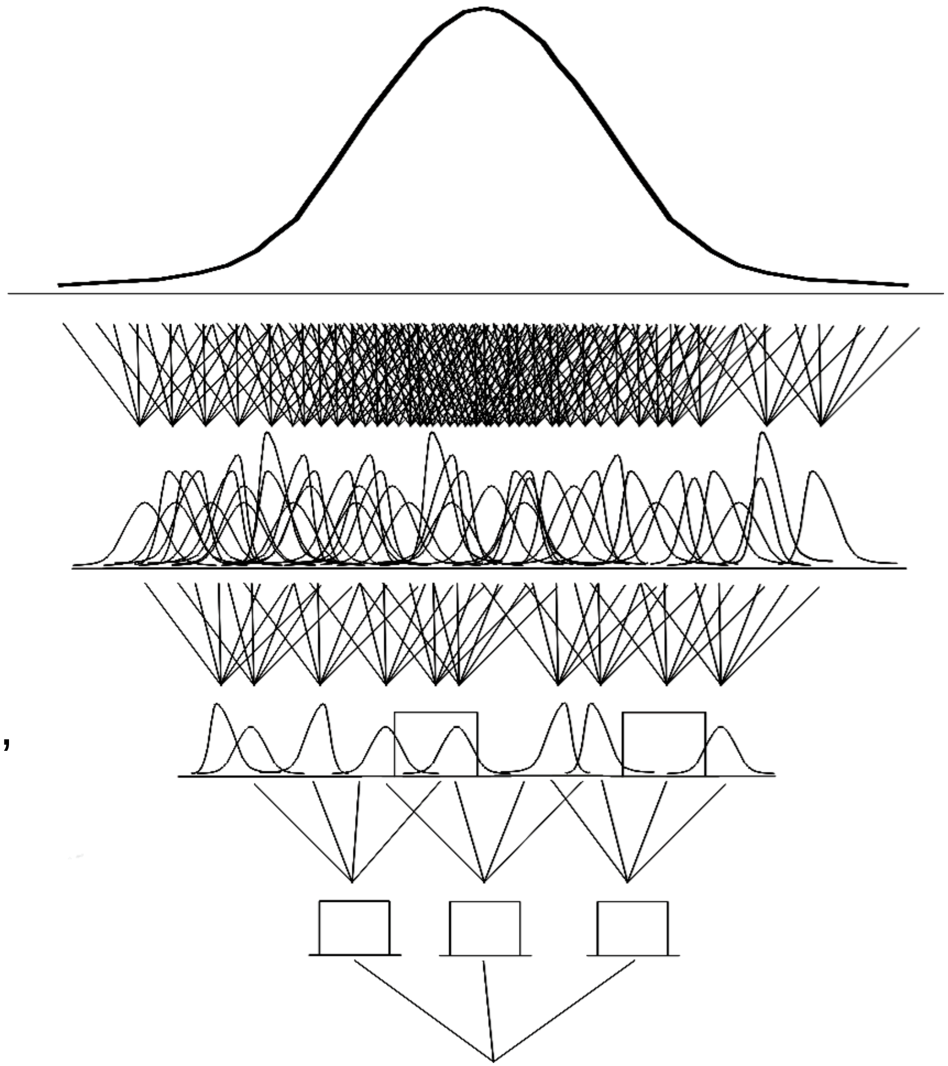
$$F(b) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{b} exp\left[-\frac{(\tilde{b}-\mu)^2}{2\sigma^2}\right] d\tilde{b}$$

Weight

Coating

Die, Configuration, Batch

Supplier

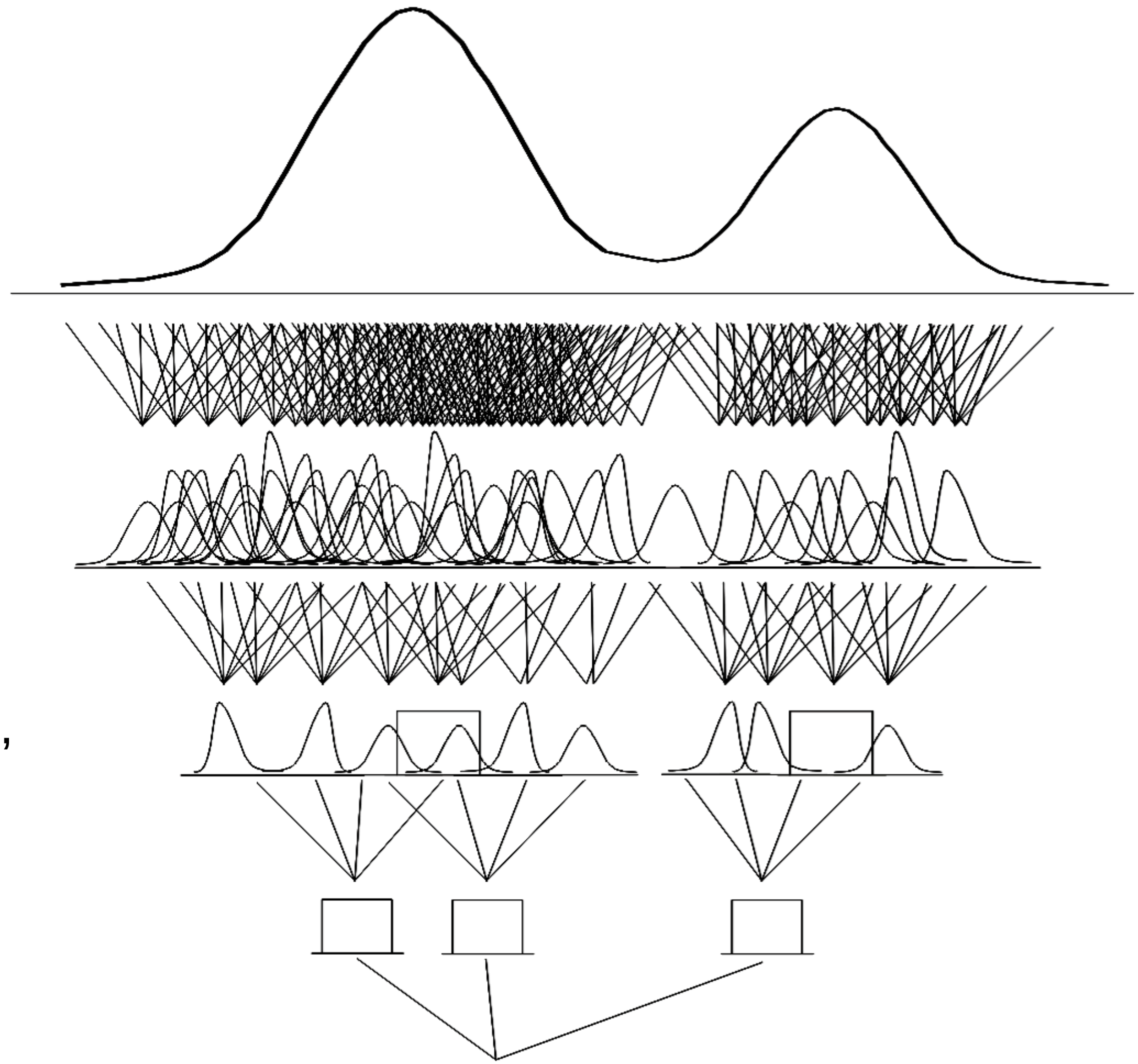


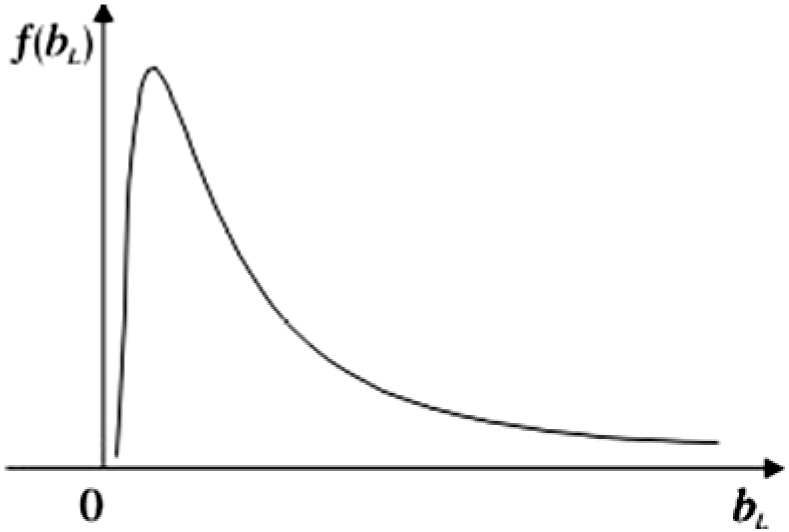
Weight

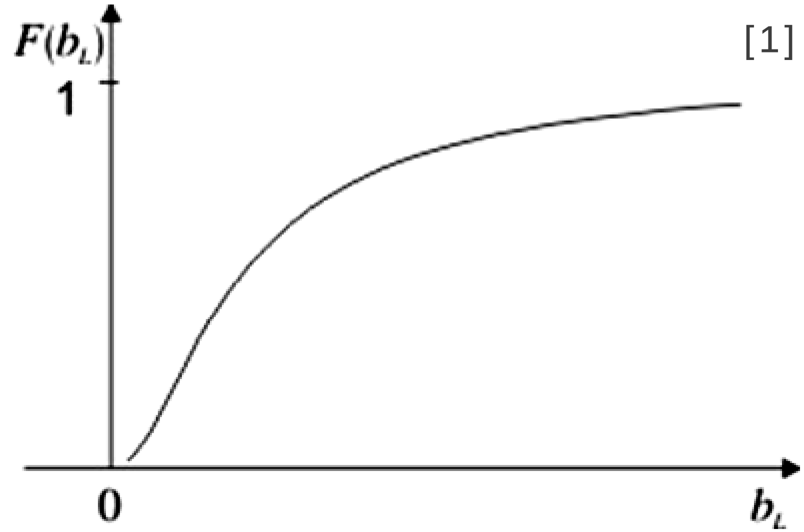
Coating

Die, Configuration,
Batch

Supplier







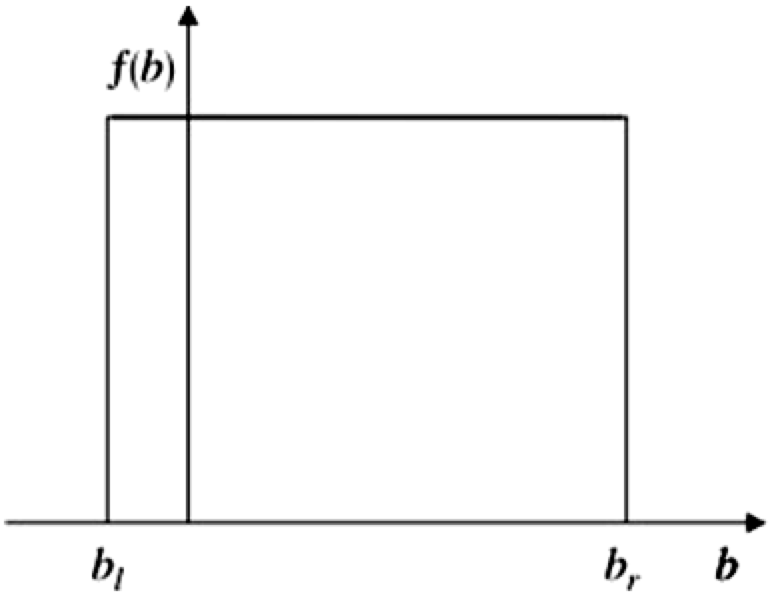
[1]

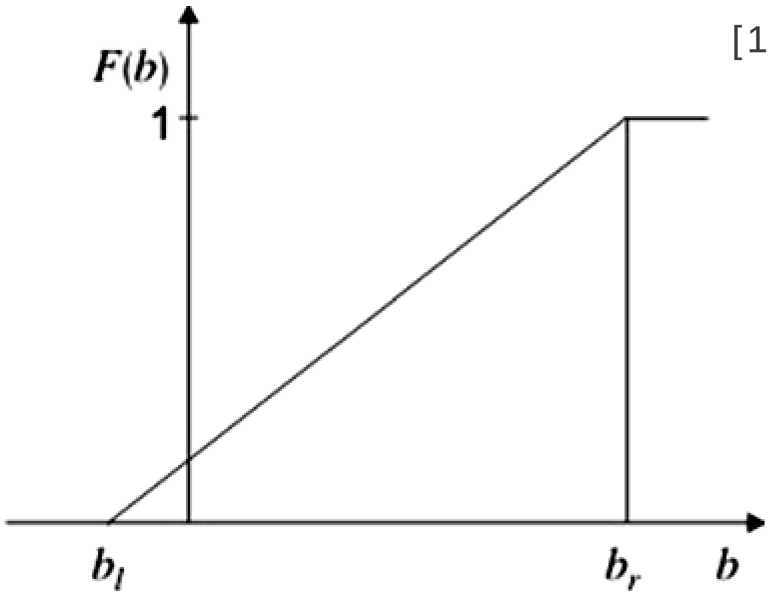
$$f(b_L) = \begin{cases} \dfrac{1}{\zeta\sqrt{2\pi}\, b_L} \exp\{-\dfrac{(\ln b_L - \lambda)^2}{2\,\zeta^2}\} & \text{für } b_L > 0 \\ 0 & \text{sonst} \end{cases}$$

$$F(b_L) = \begin{cases} \dfrac{1}{\zeta\,\sqrt{2\,\pi}} \displaystyle\int_0^{b_L} \dfrac{1}{\widetilde{b}_L} \exp\{-\dfrac{(\ln \widetilde{b}_L - \lambda)^2}{2\,\zeta^2}\}\, d\widetilde{b}_L & \text{für } \widetilde{b}_L > 0 \\ 0 & \text{sonst} \end{cases}$$

Equal Distribution

[1]

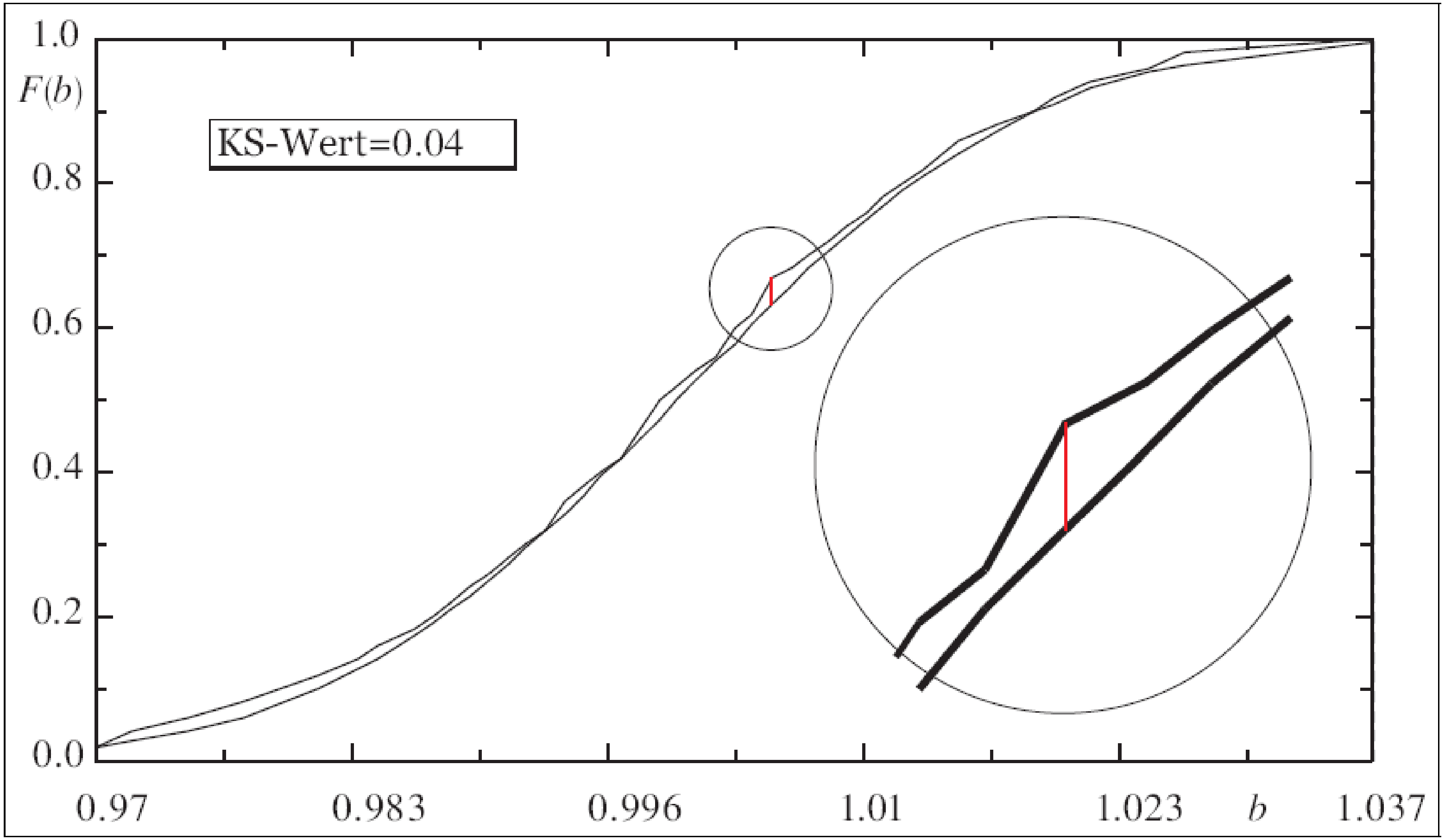




$$f(b) = \begin{cases} \dfrac{1}{b_r - b_l} & b_l \le b \le b_r \\ 0 & \text{sonst} \end{cases}$$

$$F(b) = \begin{cases} 0 & b < b_l \\ \dfrac{b - b_l}{b_r - b_l} & b_l \le b \le b_r \\ 1 & b > b_r \end{cases}$$

Kolmogorow-Smirnow-Test



$$KS = \max_{-\infty < b_k < \infty} |F_d(b_k) - F_s(b_k)|$$

Anderson-Darling-Test

- is a modification of the Kolmogorow-Smirnow-Test
- deviations between test- and target distribution are stronger weighted at the edges than in the midsection of the distribution function [2]

$$A^2 = -n_{sim} - \frac{1}{n_{sim}} \sum_{k}^{n_{sim}} (2k-1)\left(\ln F_s(b_k) + \ln\left[1 - F_s\left(b_{n_{sim}+1-k}\right)\right]\right)$$

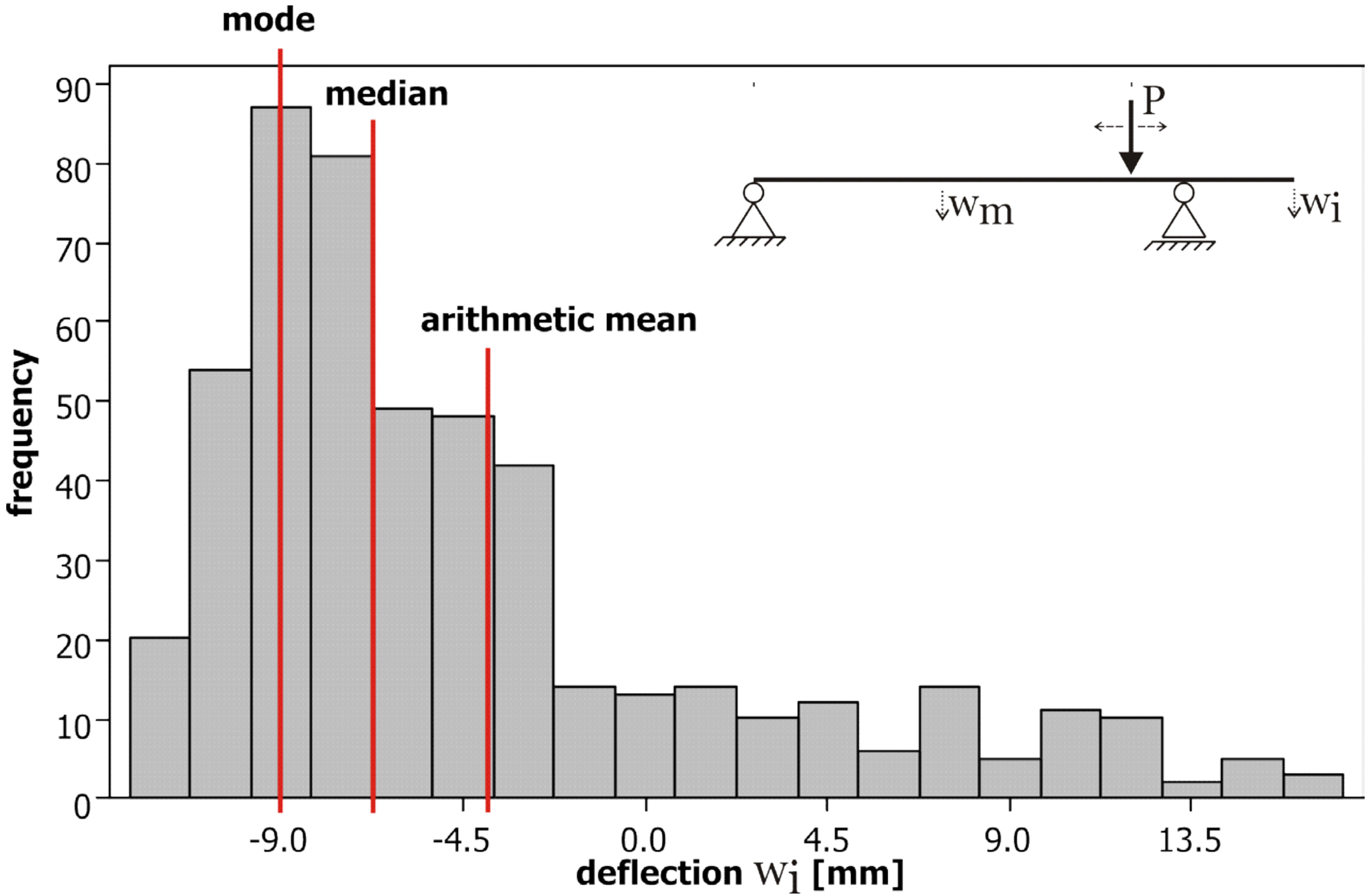
- F_s is the cumulative distribution function of the test data
- tables of critical values of A for different distributions are available in [3] for example

[2] Anderson, T. W., Darling, D.A., 1952, Asymptotic Theory of Certain “Goodness of Fit” Criteria Based on Stochastic Processes, Annals of Mathematical Statistics 23, Pages 193-212
[3] Stephens, M.A., 1974, EDF Statistics for Goodness of Fit and some Comparisons, Journal of the American Statistical Association, Vol. 69, Pages 730-737

Statistical Measures 1

- **arithmetic mean**: $$\bar{b}_{ri} = \frac{1}{n_{sim}} \sum_{k=1}^{n_{sim}} b_{ri,k}$$
  - centroid of the area underneath the density function
  - sensitive towards outliers
- **median**:
  - divides the area below the probability density function into two pieces of equal size
  - robust towards outliers
- **modal value** or **mode**:
  - value of the data set, that occurs with the greatest frequency
  - not necessarily unique.

Statistical Measures 2



Statistical Measures 3

- **standard deviation**:

$$\sigma(b_{ri}) = \sqrt{Var(b_{ri})} = \sqrt{\frac{1}{n_{sim}-1}\sum_{k=1}^{n_{sim}}(b_{ri,k} - \bar{b}_{ri})^2}$$

- **coefficient of variation:**

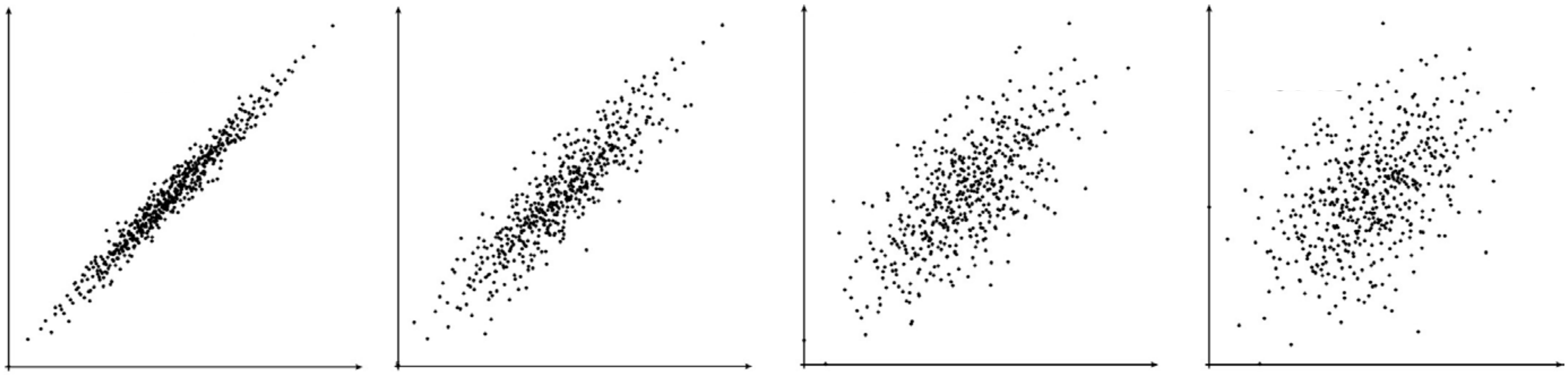
$$\delta(b_{ri}) = \frac{\sigma(b_{ri})}{\bar{b}_{ri}}$$

Requirements for Probabilistic Simulations

Probabilistic System Analysis

- Deterministic Model
- Distribution of Input Parameters
- Correlations between input parameters

Correlations



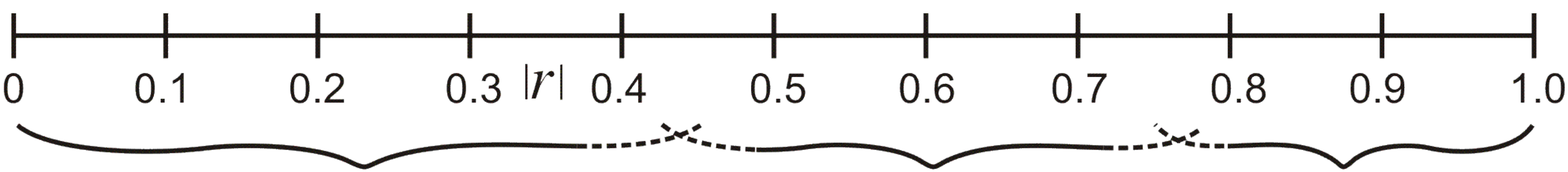
Pearson's Correlation Coefficient:

[1]

$$r_{b_{ri}b_{rj}} = \frac{Cov(b_{ri}, b_{rj})}{\sqrt{Var(b_{ri})}\sqrt{Var(b_{rj})}}$$

$$Cov\big(b_{ri}, b_{rj}\big) = \frac{1}{n_{sim}-1}\sum_{k=1}^{n_{sim}}(b_{ri,k}-\bar{b}_{ri})(b_{rj,k}-\bar{b}_{rj})$$

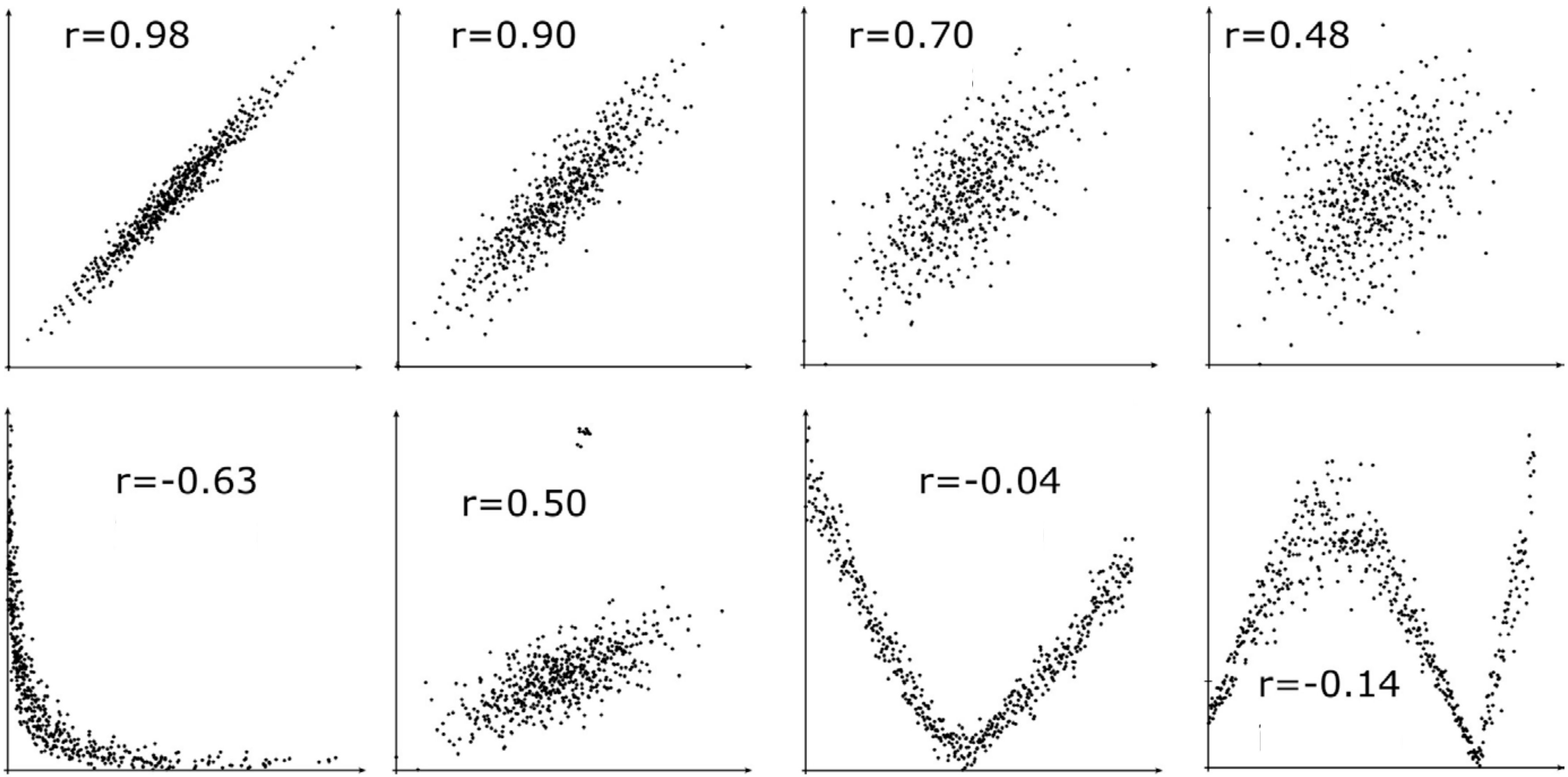
Range: [-1,1]



no or weak... medium... strong...

...linear relation

Correlations



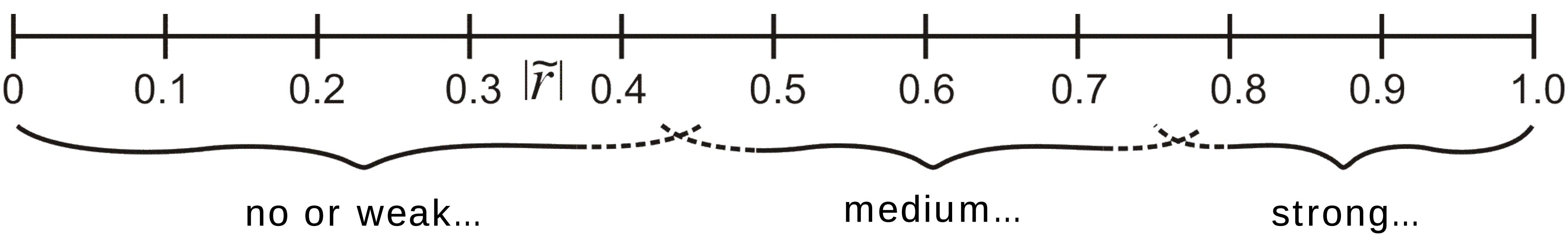
Spearman's Rank Correlation Coefficient

[1]

$$b_{ri} = \begin{bmatrix} b_{ri,1} \\ \vdots \\ b_{ri,n_{sim}} \end{bmatrix} \Rightarrow Rank(b_{ri}) = \begin{bmatrix} R_{b_{ri,1}} = Rank\ of\ b_{ri,1}\ in\ b_{ri} \\ \vdots \\ R_{b_{ri,n_{sim}}} = Rank\ of\ b_{ri,n_{sim}}\ in\ b_{ri} \end{bmatrix}$$

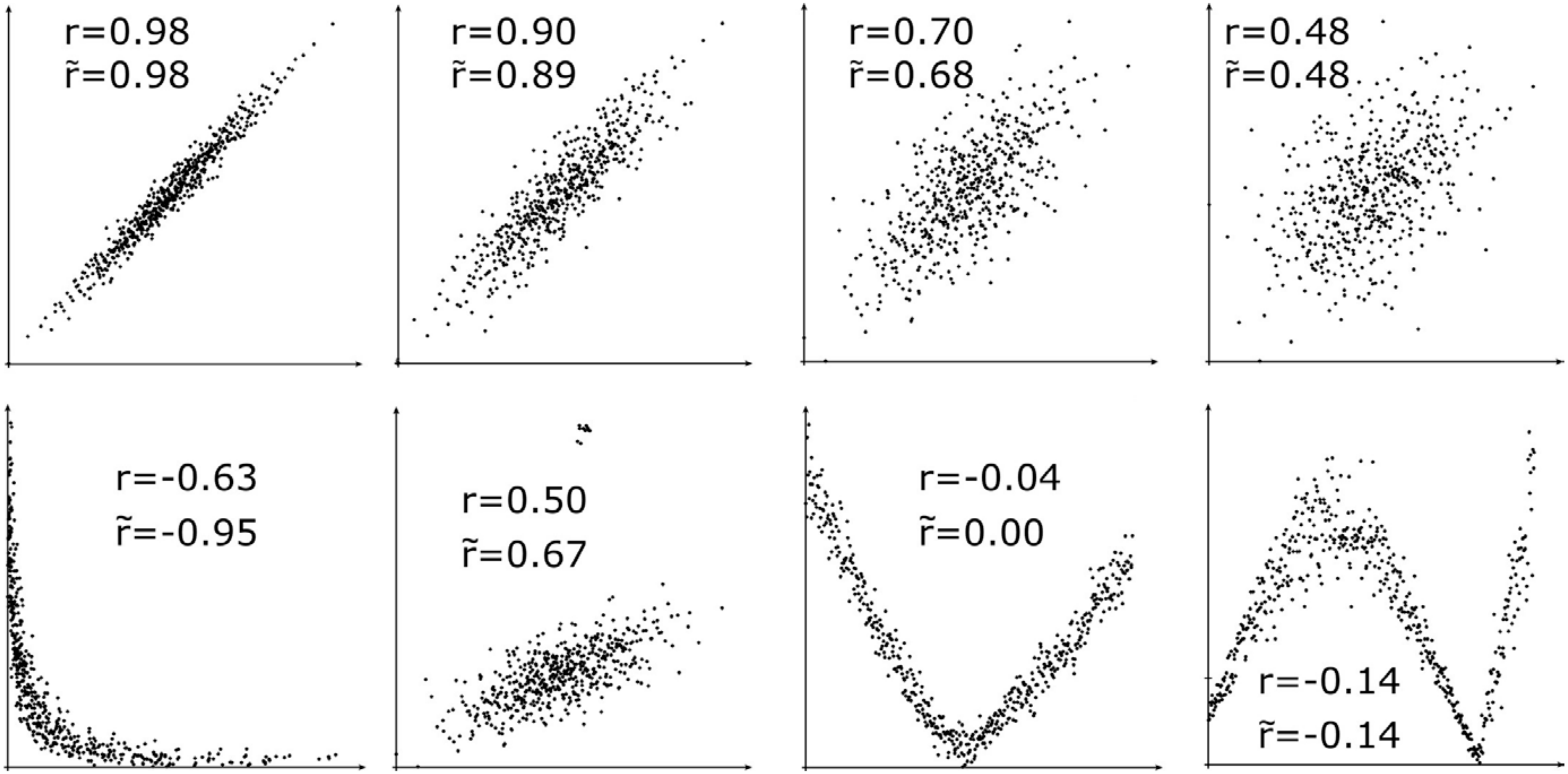
$$\tilde{r}_{b_{ri}b_{rj}} = \frac{\sum_{k=1}^{n_{sim}}(R_{b_{ri,k}} - \bar{R}_{b_{ri}})(R_{b_{rj,k}} - \bar{R}_{b_{rj}})}{\sqrt{\sum_{k=1}^{n_{sim}}(R_{b_{ri,k}} - \bar{R}_{b_{ri}})^2}\sqrt{\sum_{k=1}^{n_{sim}}(R_{b_{rj,k}} - \bar{R}_{b_{rj}})^2}}$$

Range: [-1,1]



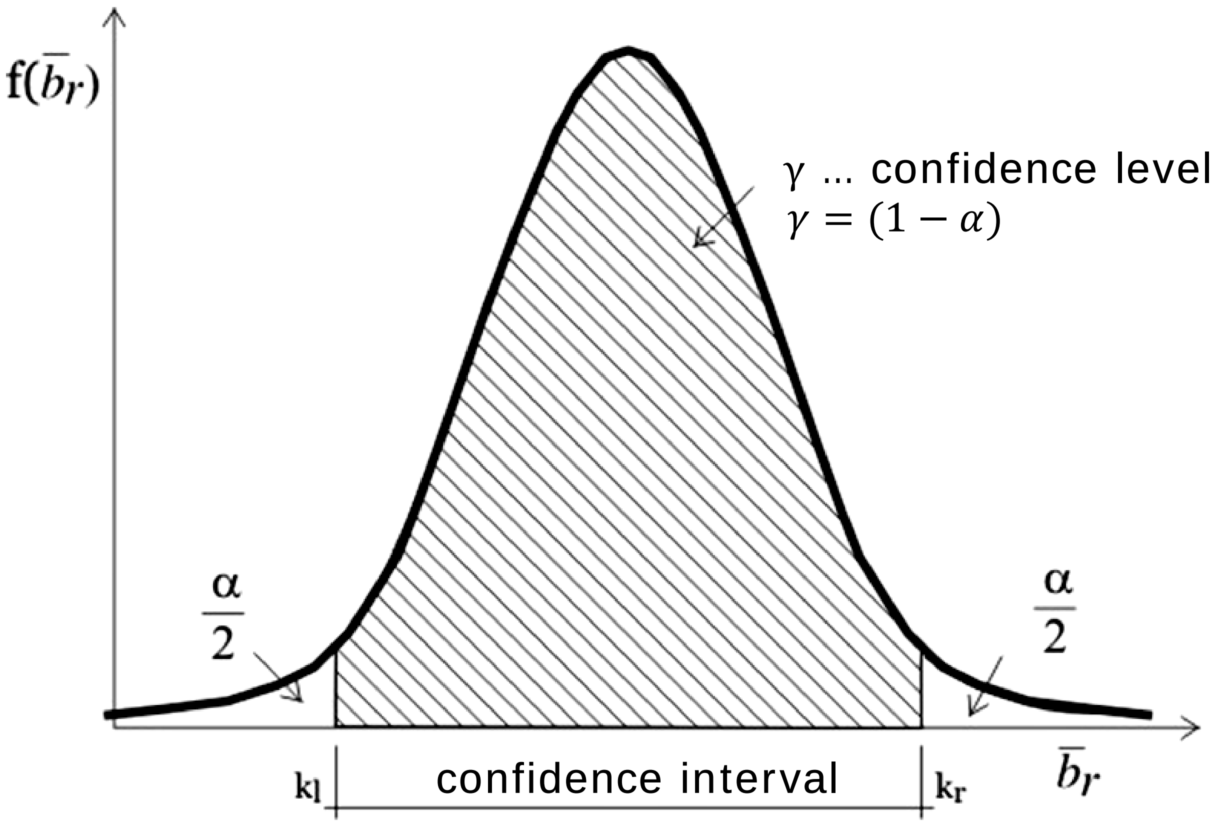
...linear or monotonic nonlinear relation

Correlations



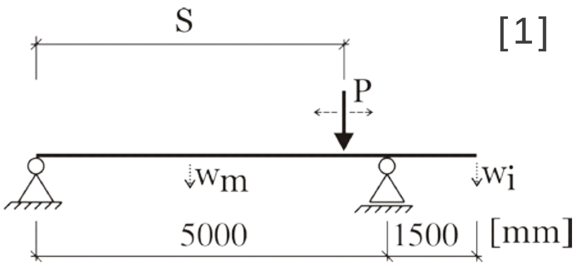
Confidence Interval

- Statistical measures (like mean, standard deviation, ...) are point estimations without any information about the quality of the estimation
- Confidence interval provides this information



- Relative frequency of an attribute/ event f:

(e.g. exceedance of a deflection limit)



[1]

$$\hat{p} = \frac{n_f}{n_{sim}}; \quad \hat{p} \to p_{true} \quad if \quad n_{sim} \to \infty$$

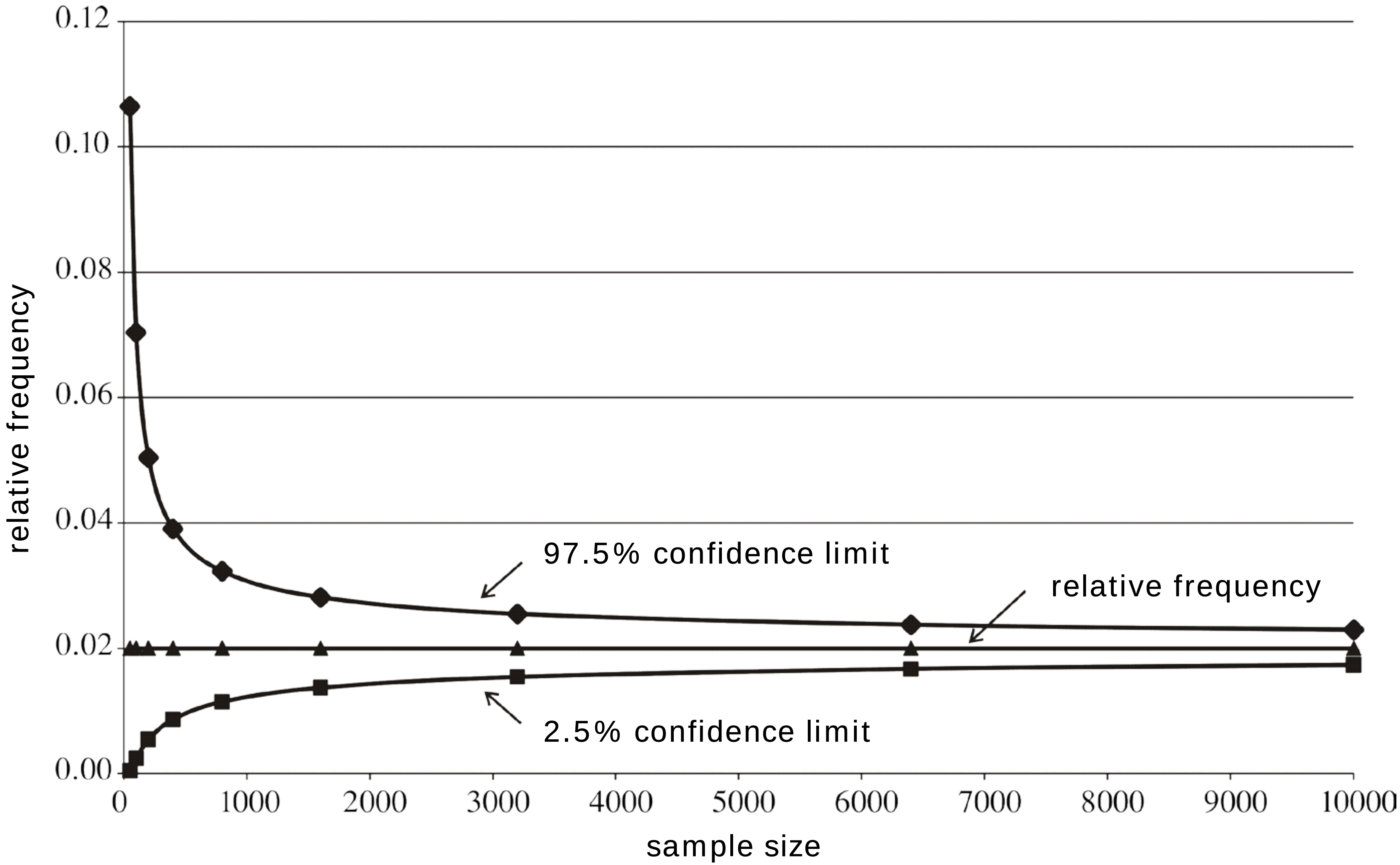
left confidence limit:

$$k_l = \frac{n_f}{n_f + (n_{sim} - n_f + 1) F_{\frac{\alpha}{2};2(n_{sim}-n_f+1);2n_f}}$$

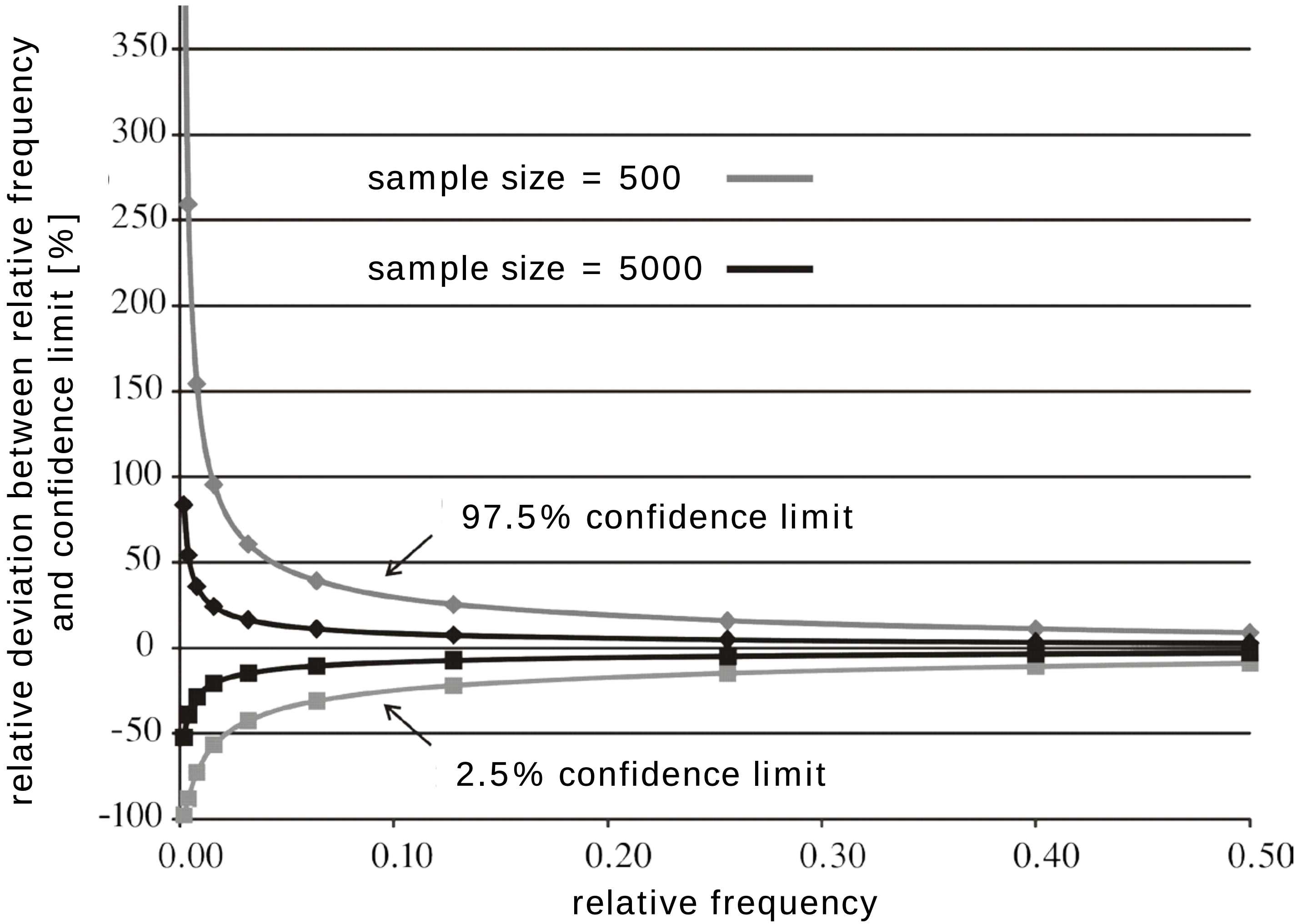
right confidence limit:

$$k_r = \frac{(n_f+1) F_{\frac{\alpha}{2};2(n_f+1);2(n_{sim}-n_f)}}{n_{sim} - n_f + (n_f + 1) F_{\frac{\alpha}{2};2(n_f+1);2(n_{sim}-n_f)}}$$

Confidence Interval – Example 1



Confidence Interval – Example 1



CI of Correlation Coefficients

- The distribution of a correlation coefficient differs from the normal distribution, if its absolute value is significantly greater than zero [1]
- Ronald Aylmer Fisher : normalization via z-transformation in order to calculate the confidence limits [4]

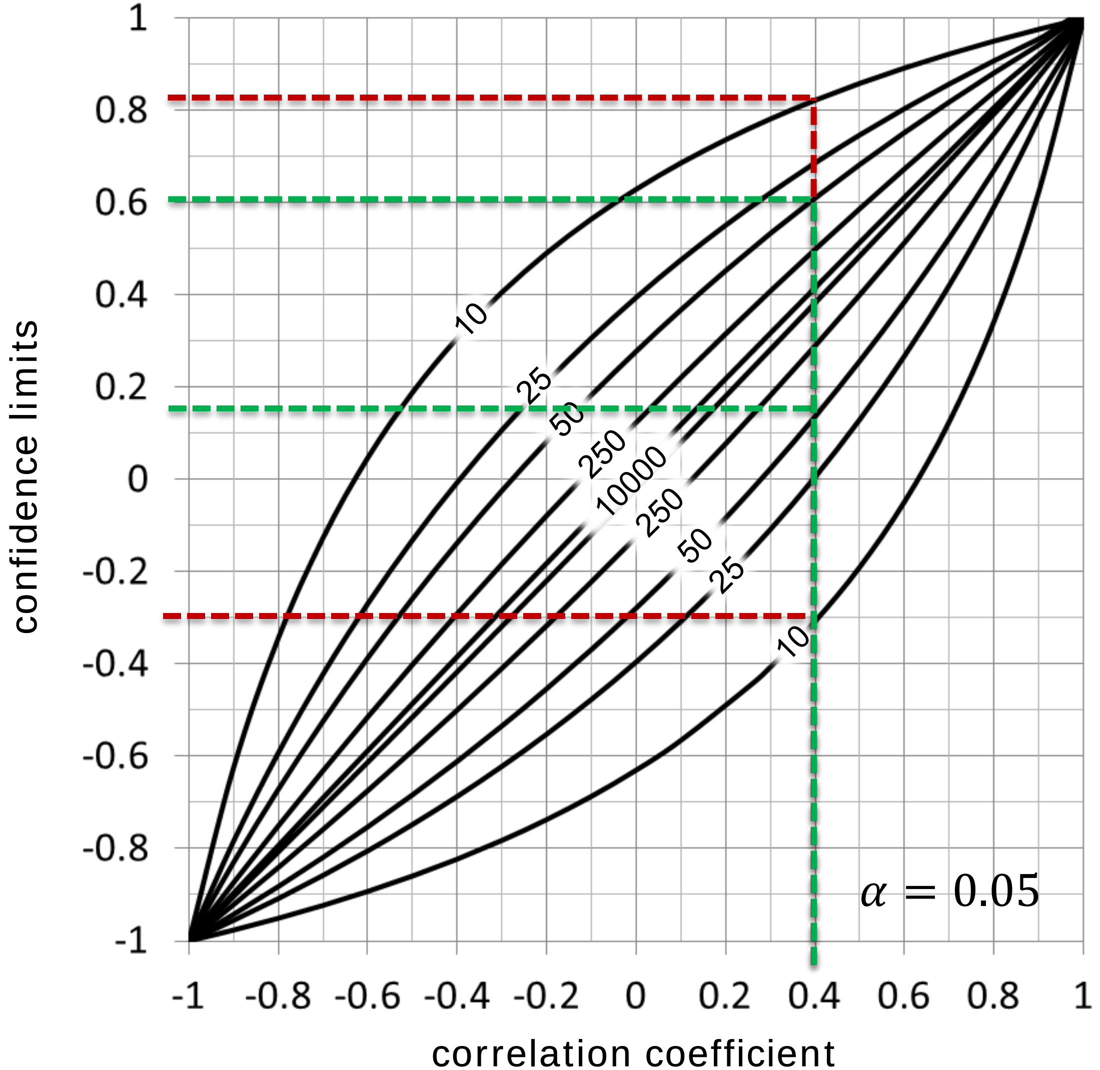
$$\dot{z} = \operatorname{arctanh}(r)$$

left confidence limit: $k_l = \tanh(\dot{k}_l)$ $\quad \dot{k}_l = \dot{z} - \dfrac{z_{1-\alpha/2}}{\sqrt{n_{sim} - 3}}$

right confidence limit: $k_r = \tanh(\dot{k}_r)$ $\quad \dot{k}_r = \dot{z} + \dfrac{z_{1-\alpha/2}}{\sqrt{n_{sim} - 3}}$

[1] Sachs, L., 2004, Angewandte Statistik, Anwendung statistischer Methoden, Springer, Berlin/ Heidelberg/ New York
[4] Fisher, R. A., 1970, Statistical Methods for Research Workers, Oliver & Boyd, Edinburgh

CI of Correlation Coefficients



probabilistic system analysis

Deterministic model

Distribution of input parameters

Correlations between input parameters

Probabilistic methods

References

[1] Sachs, L., 2004, *Angewandte Statistik, Anwendung statistischer Methoden*, Springer, Berlin / Heidelberg/ New York

[2] Anderson, T. W., Darling, D.A., 1952, *Asymptotic Theory of Certain "Goodness of Fit" Criteria Based on Stochastic Processes*, Annals of Mathematical Statistics 23, Pages 193-212

[3] Stephens, M.A., 1974, *EDF Statistics for Goodness of Fit and some Comparisons*, Journal of the American Statistical Association, Vol. 69, Pages 730-737

[4] Fisher, R. A., 1970, *Statistical Methods for Research Workers*, Oliver & Boyd, Edinburgh