Aleatoric or epistemic?
Does it matter?

presentation on the paper of
Kiureghian and Ditlevsen, 2008

Markus K. Sandomeer

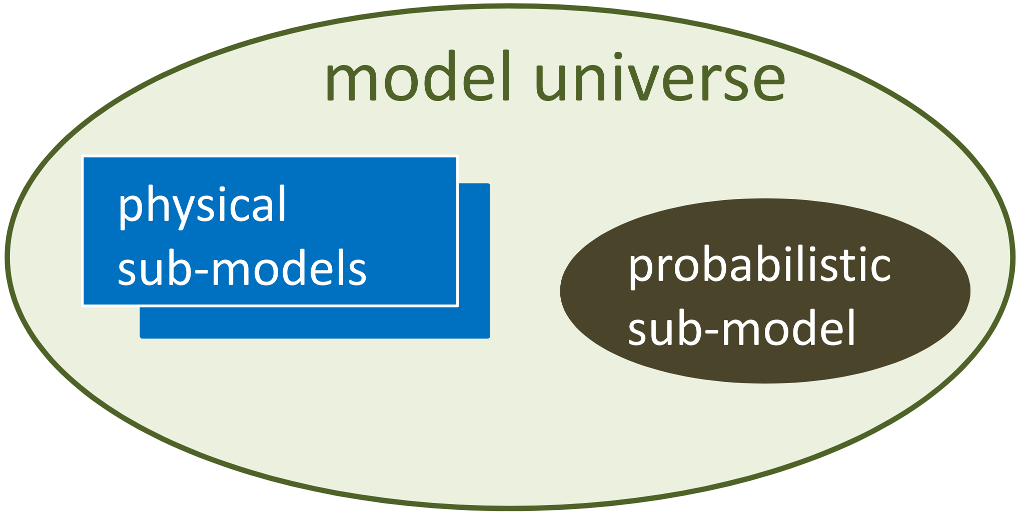
ETH Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zürich

ibk Institute of Structural Engineering
Group Risk and Safety

„Probability is the most important concept in modern science, especially as nobody has the slightest notion of what it means.“

Bertrand Rusell, 1929
philosopher & mathematician

introduction



- model universe contains set of physical and probabilistic sub-models
- mathematical idealizations of reality
- invariably imperfect → additional uncertainties
- Any discussion on the nature and character of uncertainties should be stated within the confines of the model universe.

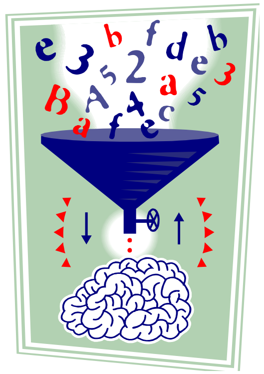
introduction

categorization of uncertainty into two types:

aleatory uncertainties

- intrinsic randomness of a phenomenon
- *e.g. throwing a dice...*
- cannot be reduced

epistemic uncertainties



- episteme = knowledge (Greek)
- lack of knowledge (data)
- can be reduced
- represented in the model by auxiliary non-physical variables (define statistical dependencies, correlations)

uncertainty in characterisation of uncertainty...

characterization of uncertainty
pragmatic choice dependent
on purpose of application,
responsibility of model builder

What are the difficulties?

1. most engineering problems involve both types of uncertainties

2. determination whether particular uncertainty should be considered as aleatory or epistemic

confining model universe

set of input variables: $\mathbf{x} = \left(x_1, x_2, ..., x_n\right)^T$

which take values as outcomes of the corresponding set of basic random variables:

$$\mathbf{X} = \left(X_1, X_2, ..., X_n\right)^T$$

parameterized probabilistic sub-model:

$$f_{\mathbf{X}}\left(\mathbf{x}, \mathbf{\Theta}_f\right)$$ *distribution of random vector* $\mathbf{X}$

set of parameterized physical sub-models:

$$y_i = g_i\left(\mathbf{x}, \mathbf{\Theta}_g\right),\ i = 1, 2, ..., m$$

relation between quantities $\mathbf{x}$ and m derived quantities

$$\mathbf{y} = \left(y_1, y_2, ..., y_m\right)^T$$

confining model universe

random variables $\mathbf{X}$:

- directly observable; empirical data available
- represent quantities like material properties, load characteristics, other environmental effects, geometric dimensions, ...

derived variables $\mathbf{y}$:

- usually not directly observable, except in laboratory or field studies for model development
- describe engineering performance criteria as e.g. stresses, deformations, stability limits, measures of damage, loss, downtime, concentration of toxins, ...

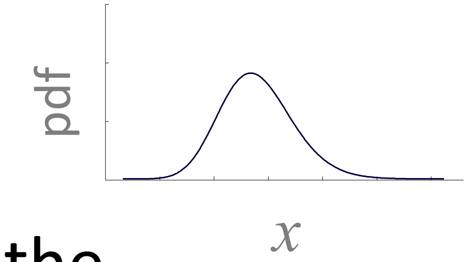
confining model universe

sub-models $f_{\mathbf{x}}\left(\mathbf{x}, \mathbf{\Theta}_{f}\right)$ and $y_{i}=g_{i}\left(\mathbf{x}, \mathbf{\Theta}_{g}\right)$:

- invariably imperfect mathematical idealizations of reality
- contain uncertain errors
- parameters assessed by fitting the sub-models to data
- alternative models may be applicable with different likelihoods of relative validities

1. uncertainty inherent in basic random variables $\mathbf{X}$

uncertainty inherent in directly measured material property constants and load values



2. uncertain model error

resulting from selection of the form of the **probabilistic** sub-model used to describe the distribution of basic variables $f_{\mathbf{x}}\left(\mathbf{x},\boldsymbol{\Theta}_f\right)$

3. uncertain modelling errors

resulting from selection of the form of the **physical** sub-model used to describe the derived variables

$$y_i = g_i\left(\mathbf{x},\boldsymbol{\Theta}_g\right),\ i = 1,2,\ldots,m$$

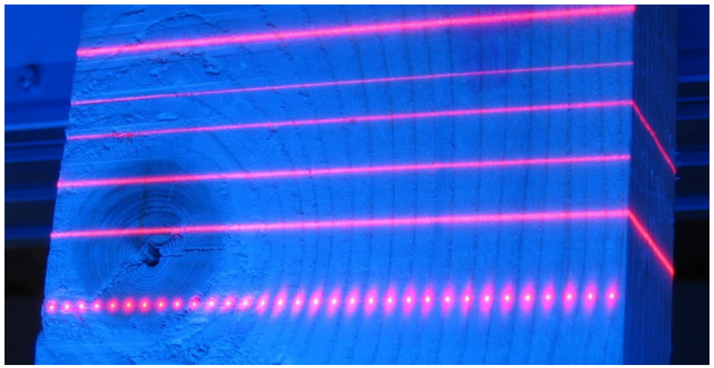
4. statistical uncertainty
   estimation of the parameters $\boldsymbol{\Theta}_f$ of the **probabilistic** sub-model

5. statistical uncertainty
   estimation of the parameters $\boldsymbol{\Theta}_g$ of the **physical** sub-models

6. uncertain errors in measurement and observations

Parameters $\boldsymbol{\Theta}_f$ and $\boldsymbol{\Theta}_g$ estimated based on measurements and observations.

Errors included in indirect measurements, e.g. non-destructive testing of material strength.



non-destructive laser & X-ray measurements for prediction of timber strength
MiCROTEC GmbH

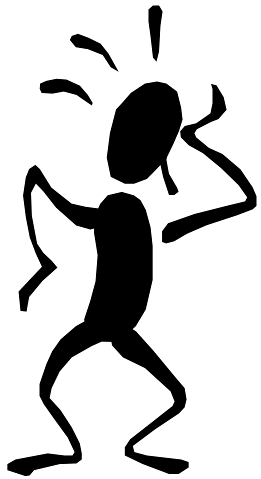
7. uncertainty in random variable $\mathbf{Y}$

in addition to previous described uncertainties:

- uncertain errors from computational errors
- numerical approximations or truncations

8. special case: Human Error

uncertainties arising from human activities and decisions;
tend to be random in nature →
categorized as aleatory uncertainties



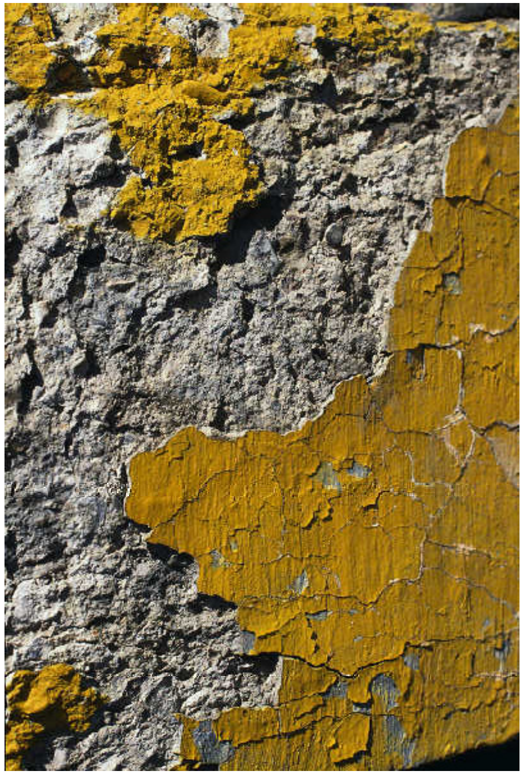
Uncertainty in basic random variables

example: Compressive Strength of Concrete

Millau, France

example: Compressive Strength of Concrete

existing building



- epistemic
- specimens taken from the building can be tested
- testing itself with measurement errors; epistemic uncertainty, if alternative methods would be at hand

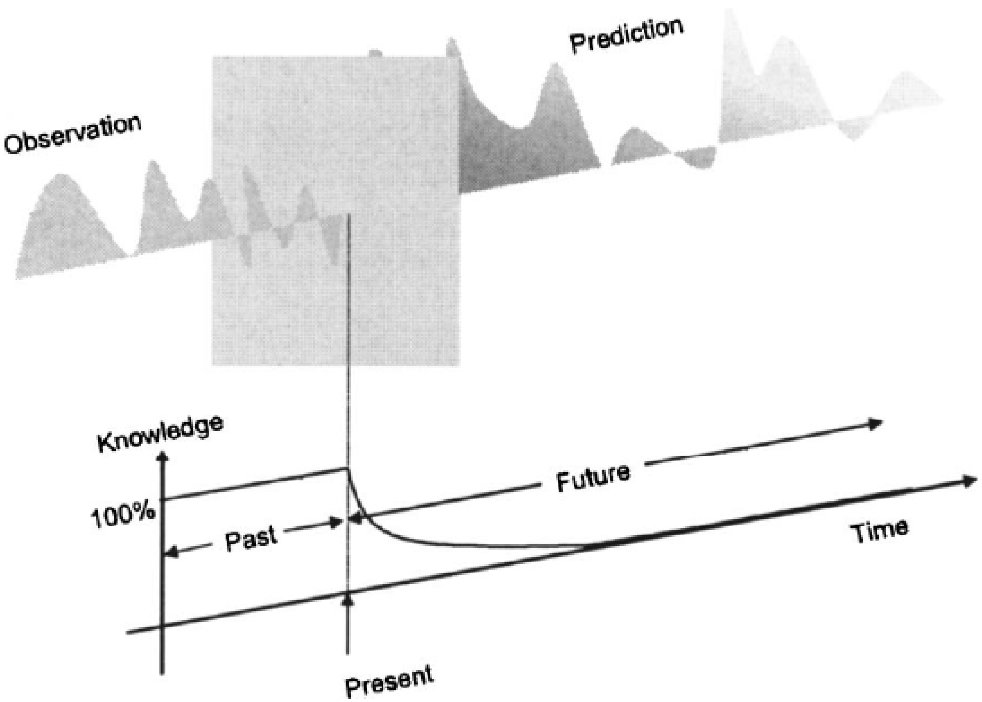
example: Compressive Strength of Concrete

future building

- aleatory
- until building will be realized no additional testing will reduce variability of concrete strength

example: Compressive Strength of Concrete

aleatory uncertainty transforms into epistemic uncertainty as the building is realized



Faber M. H., On the Treatment of Uncertainties and Probabilities in Engineering Decision Analysis, Journal of Offshore Mechanics and Arctic Engineering, 2005

model uncertainty in general:

- development of mathematical model to predict physical quantity y in terms of two sets of basic variables $\mathbf{x}$ and $\mathbf{z}$.
- exact form of relationship between basic variables and derived variable is unknown
- dependency of y on $\mathbf{z}$ is not clear
- e.g. since no observations of $\mathbf{z}$ can be realized, including them into the model would not be useful

for example: Attenuation Model

- intensity at a site is dependent on:
  - earthquake magnitude (measured)
  - distance (measured)
  - propagation velocity of the fault rupture
  - mechanical characteristics of the path of propagation of seismic waves
  - geologic features surrounding the site
  - ...

$\mathbf{x}$

$\mathbf{z}$

- however, not all of these variables can be measured for a given site

➡ exclusion from attenuation model

model uncertainty

predictive model of y: $$y = \boxed{\hat{g}\left(\mathbf{x}, \boldsymbol{\theta}_g\right)} + \boxed{\varepsilon}$$

idealized mathematical model

$$\varepsilon = y - \hat{g}\left(\mathbf{x}, \boldsymbol{\theta}_g\right)$$

model error (residuals)

assumption: $\boldsymbol{\theta}_g$ and ε are realizations of the random variables $\boldsymbol{\Theta}_g$ and $\mathbf{E}$.

$\mathbf{E}$ incorporates:
a) uncertain effect of missing variables
b) effect of potentially inaccurate form of the model.

predictive model of y:

homoskedastic (unbiased) form of the model:

- mean of $\mathbf{E}$ is equal to zero
- normal distributed
- standard deviation Σ_{E} as a measure of the inaccuracy of the model

➡ set of parameters to be estimated is: $\left(\mathbf{\Theta}_g, \Sigma_{\mathrm{E}}\right)$ and correlation coefficients between error terms of different sub-samples

uncertainty in $\mathbf{E}$:

epistemic & **aleatory**

epistemic

- uncertain effects in missing variables $\mathbf{z}$
- potentially inaccurate form of the model
- both are reducible by refining the model

aleatory

- limited state of scientific knowledge sets thresholds for refinement
- inability for additional measurements (additional variables)
- uncertainty due to missing variables is aleatory, if these are characterized as being aleatory random variables

probabilistic model: $f_{\mathbf{X}}\left(\mathbf{x}, \boldsymbol{\Theta}_f\right)$

In computing probabilities, particularly for rare events, an error of uncertain magnitude arises from the assumed distribution model.

This error can be placed in the epistemic category, since gathering of more data would allow a better fit of the distribution and, therefore, a reduction in model uncertainty.

difficult to assess the magnitude of the error arising from choice of distribution model; → 3 approaches...

probabilistic model: $f_{\mathbf{x}}\left(\mathbf{x}, \boldsymbol{\Theta}_f\right)$

1. Compute probability of interest for all viable distribution models and assess variability in the computed probability values.

2. Parameterize the choice of the distribution. Uncertainty in the distribution model is then represented by the uncertainty in the parameter.

3. Cast the model choice as a decision problem.

large computational efforts for all of them

parameter uncertainty

parameters of the sub-models are estimated based on statistical analysis of observed data.

parameters $\left(\boldsymbol{\Theta}_g, \Sigma_{\mathrm{E}}\right)$ of physical sub-models based on pair-wise observations of $\mathbf{X}$ and $\mathbf{Y}$.

parameters $\boldsymbol{\Theta}_f$ of the probabilistic sub-model are based on observations of $\mathbf{X}$.

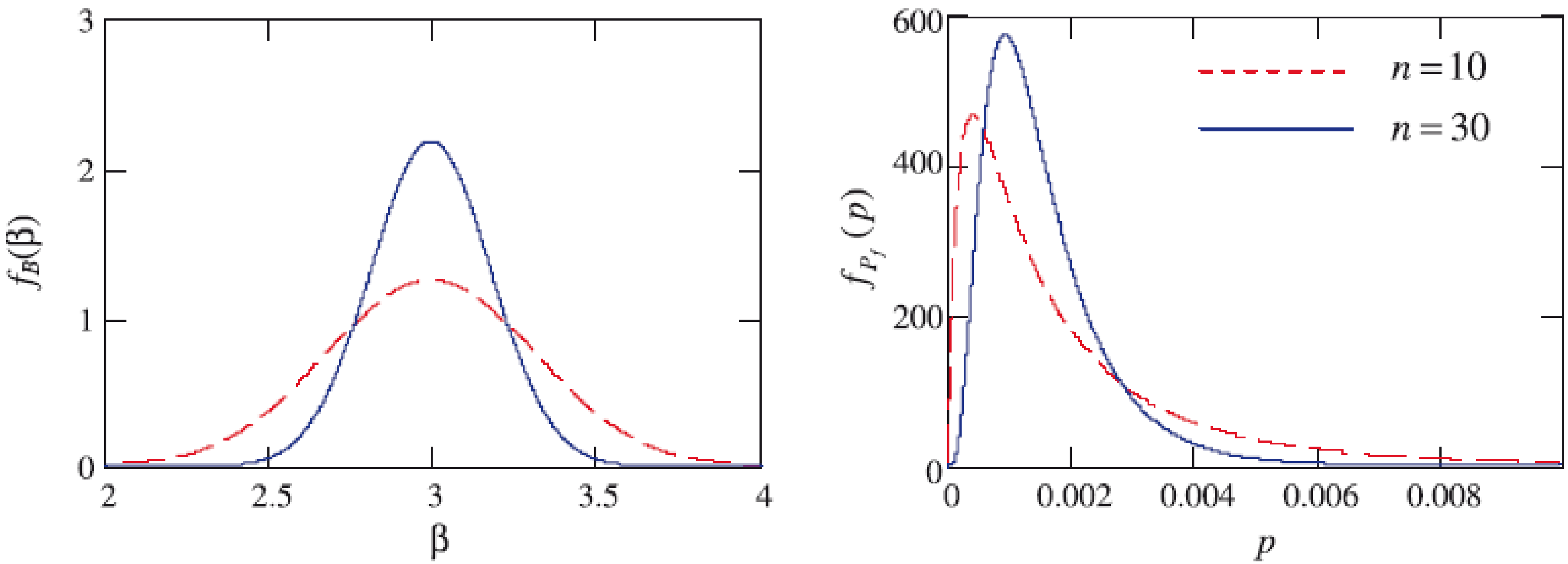
parameter estimation

uncertainty in parameter estimates

- statistical uncertainty
- directly related to amount (sample size) and quality (accuracy of observations, prior information) of available information

➡ parameter uncertainties are epistemic in nature

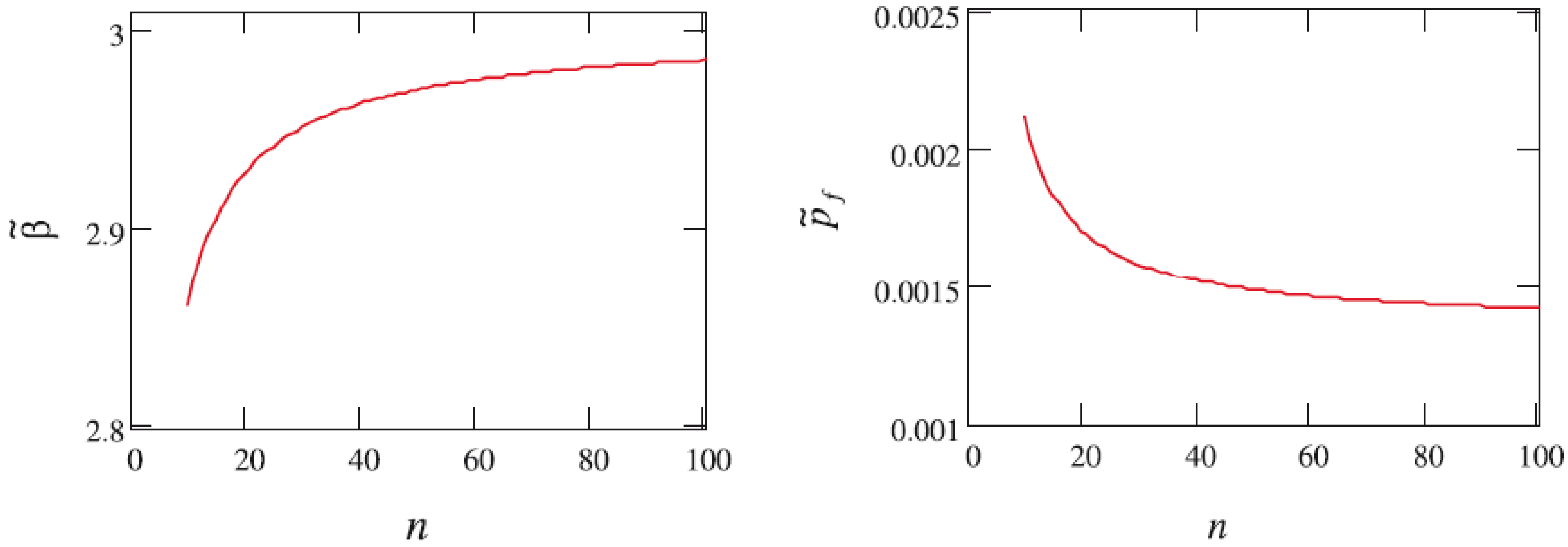
- uncertainty in estimation decreases with increasing quantity and quality of available observed data influence of statistical uncertainty on reliability index (left) and failure probability (right)



Kiureghian AD, Ditlevsen O, Aleatory or epistemic? Does it matter? Struc Saf (2008)

examples – influence of uncertainty

predictive reliability index (left) and predictive failure probability (right) as function of sample size



Kiureghian AD, Ditlevsen O, Aleatory or epistemic? Does it matter? Struc Saf (2008)

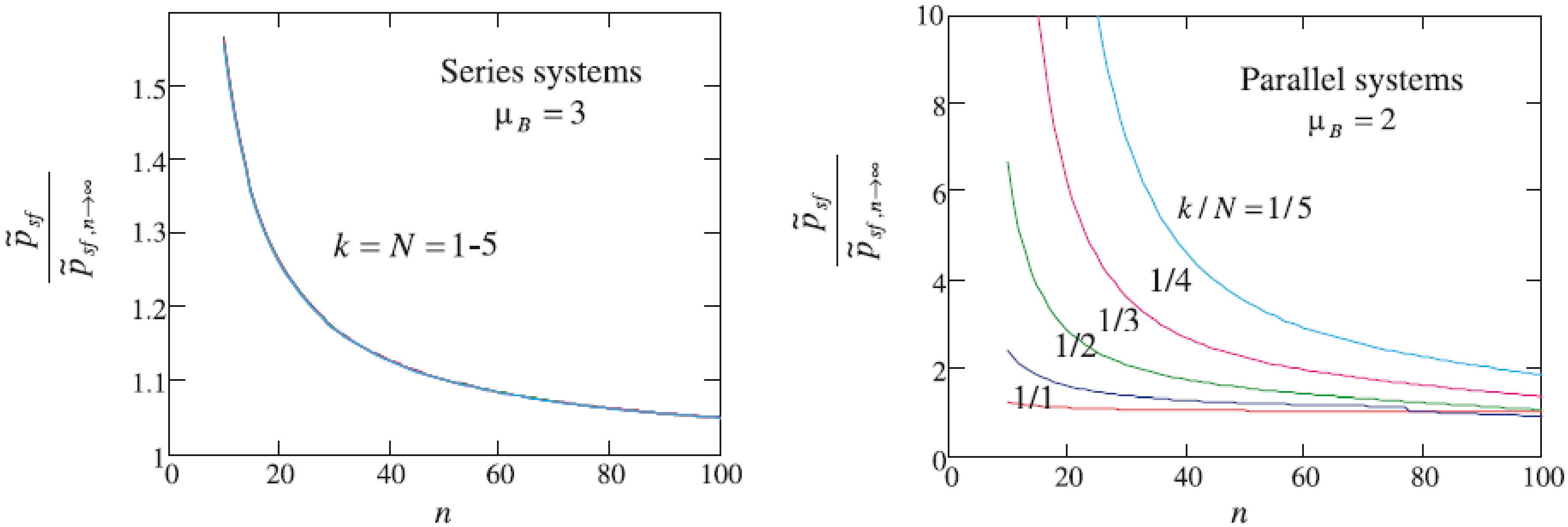
increasing n → $\tilde{\beta}$ approaches $\mu_B = 3$

→ $\tilde{p}_f$ approaches $\Phi(-\mu_B) = 0.00135$

(limiting values without statistical uncertainty)

examples – influence of uncertainty

influence of statistical uncertainty on series (left) and parallel (right) systems



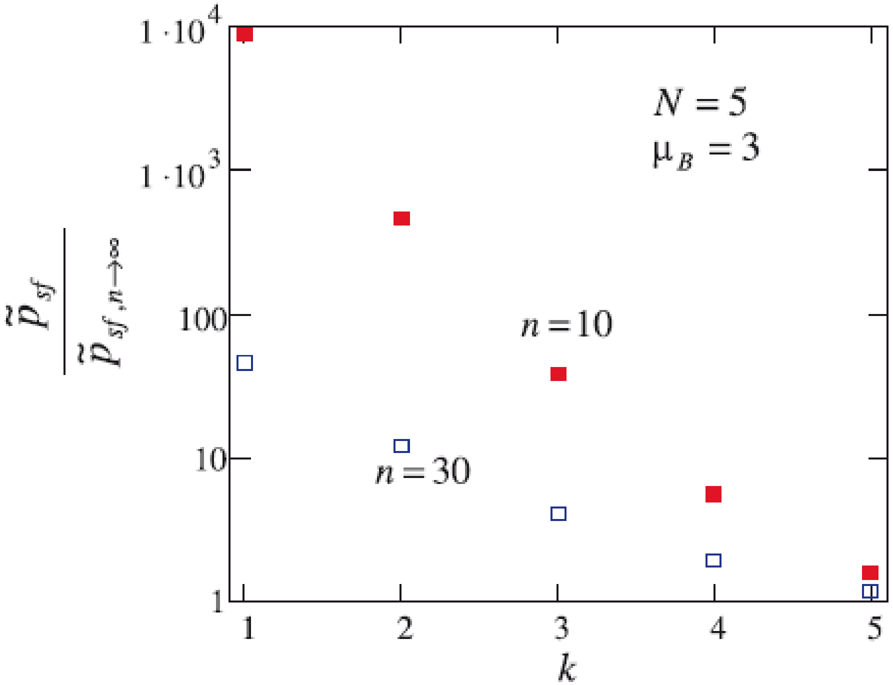
Kiureghian AD, Ditlevsen O, Aleatory or epistemic? Does it matter? Struc Saf (2008)

statistical dependence among the estimated states of the components

increasing statistical uncertainty → decreasing sample size → increase of predictive failure probability for both systems

examples – influence of uncertainty

influence of statistical uncertainty on k-out-of-N system



Kiureghian AD, Ditlevsen O, Aleatory or epistemic? Does it matter? Struc Saf (2008)

increasing statistical uncertainty → decreasing sample size → increase of predictive failure probability