Statistics and Probability Theory in Civil, Surveying and Environmental Engineering

Prof. Dr. Michael Havbro Faber

Swiss Federal Institute of Technology

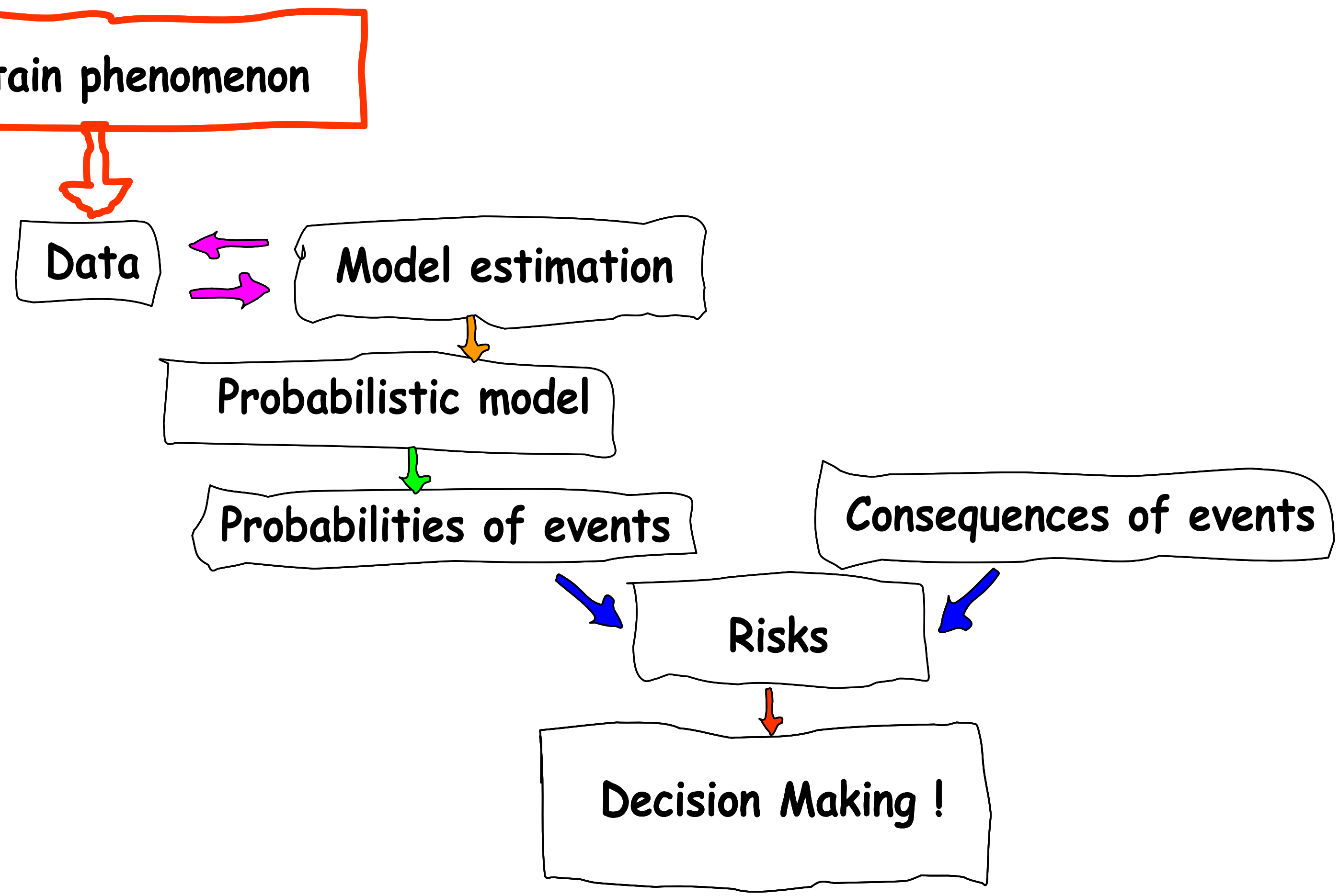
ETH Zürich, Switzerland

Contents of Today's Lecture

- Overview of Uncertainty Modelling
- Uncertainties in Engineering Problems
- Random Variables
  - discrete cumulative distribution and probability density functions
  - continuous cumulative distribution and probability density functions
  - characterization of random variables
  - moments of random variables
  - the expectation and the variance operator

Overview of Uncertainty Modelling

- Why uncertainty modelling



Uncertainties in Engineering Problems

Different types of uncertainties influence decision making

- **Inherent natural variability – aleatory uncertainty**
  - result of throwing dices
  - variations in material properties
  - variations of wind loads
  - variations in rain fall

- **Model uncertainty – epistemic uncertainty**
  - lack of knowledge (future developments)
  - inadequate/imprecise models (simplistic physical modelling)

- **Statistical uncertainties – epistemic uncertainty**
  - sparse information/small number of data

Uncertainties in Engineering Problems

- Consider as an example a dike structure

  - the design (height) of the dike will be determining the frequency of floods

  - if exact models are available for the prediction of future water levels and our knowledge about the input parameters is perfect then we can calculate the frequency of floods (per year) - a deterministic world !

  - even if the world would be deterministic – we would not have perfect information about it – so we might as well consider the world as random

Uncertainties in Engineering Problems

In principle the so-called

inherent physical uncertainty (aleatory – Type I)

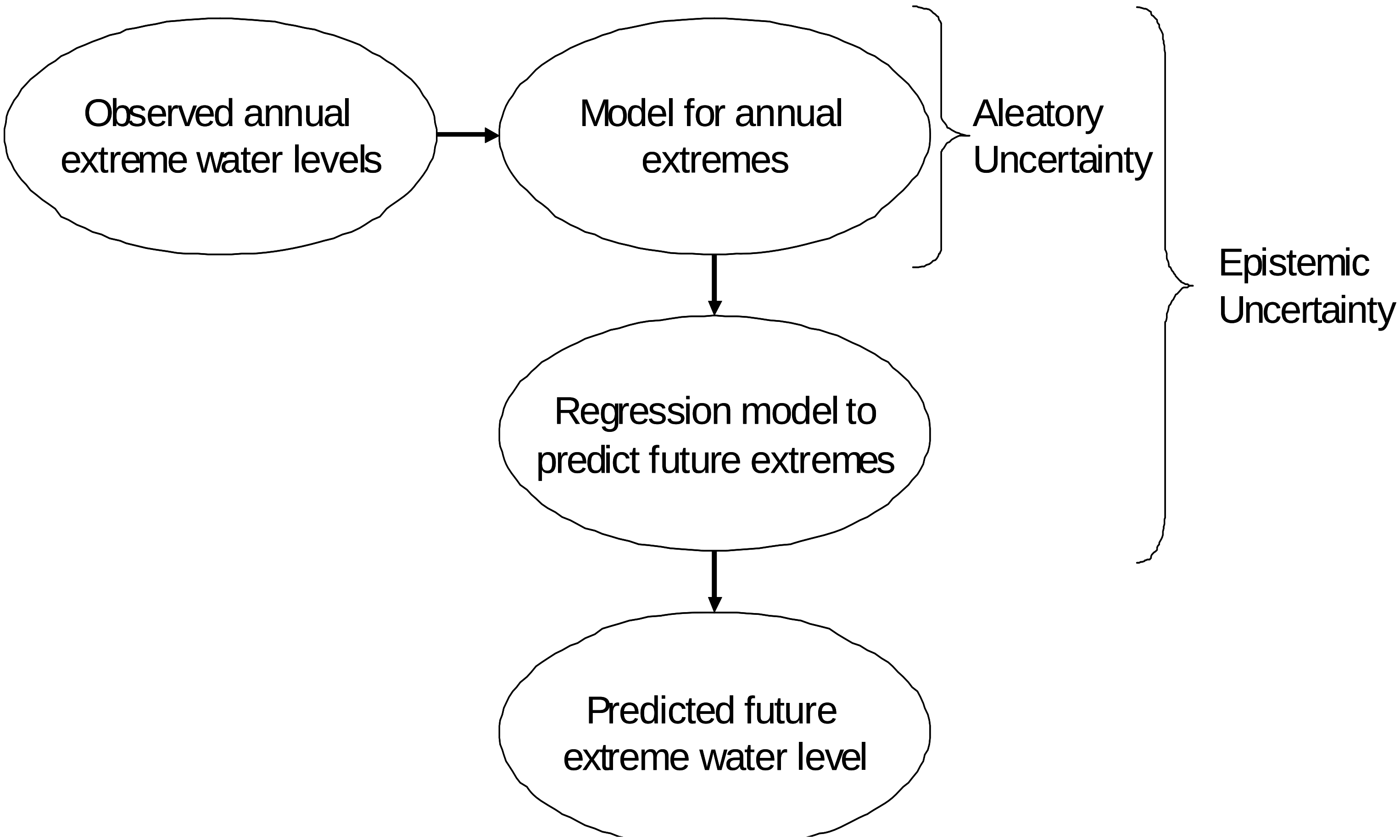
is the uncertainty caused by the fact that the world is random, however, another pragmatic viewpoint is to define this type of uncertainty as

any uncertainty which cannot be reduced by means of collection of additional information

the uncertainty which can be reduced is then the

model and statistical uncertainties (epistemic – Type II)

Uncertainties in Engineering Problems



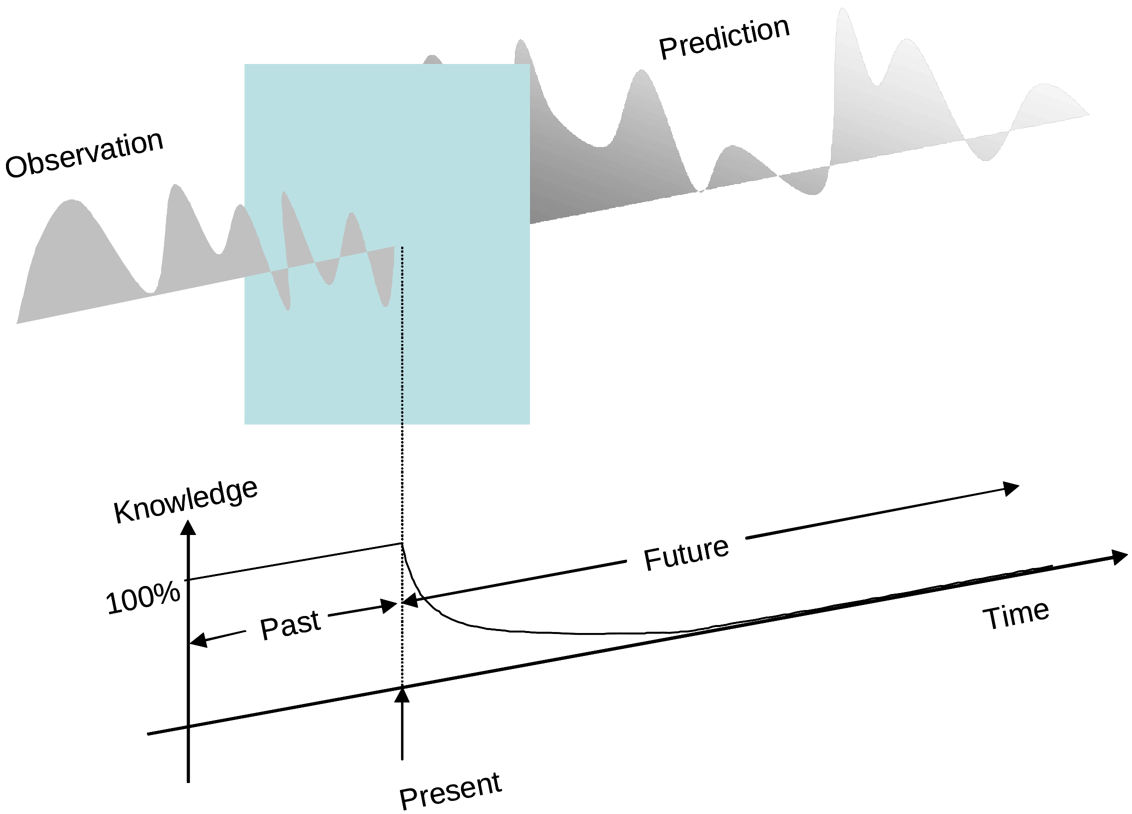
Uncertainties in Engineering Problems

The relative contribution of aleatory and epistemic uncertainty to the prediction of future water levels is thus influenced directly by the applied models

refining a model might reduce the epistemic uncertainty – but in general also changes the contribution of aleatory uncertainty

the uncertainty structure of a problem can thus be said to be scale dependent !

Uncertainties in Engineering Problems



The uncertainty structure changes also as function of time – is thus time dependent !

Random Variables

- Probability density and cumulative distribution functions

  A random variable is denoted with capital letters : X

  A realization of a random variable is denoted with small letters : x

  We distinguish between

  - *continuous random variables* : can take any value in a given range

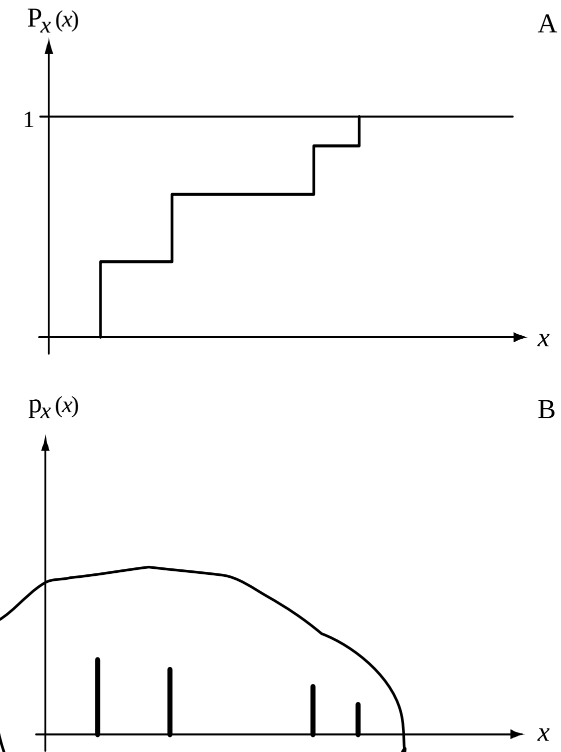
  - *discrete random variables* : can take only discrete values

Random Variables

- Probability density and cumulative distribution functions

The probability that the outcome of a discrete random variable *X* is smaller than *x* is denoted the *cumulative distribution function*

$$P_X(x) = \sum_{x_i < x} p_X(x_i)$$



The *probability density function* for a discrete random variable is defined by

$$p_X(x_i) = P(X = x_i)$$

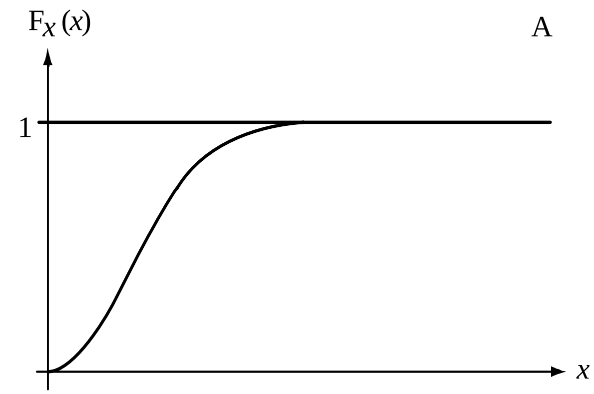
Sum of these must equal 1

Random Variables

- Probability density and cumulative distribution functions

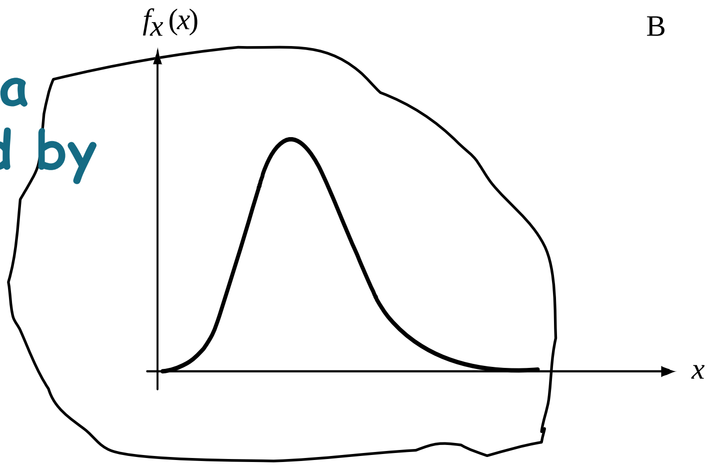
The probability that the outcome of a continuous random variable X is smaller than x is denoted the *cumulative distribution function*

$$F_X\left(x\right) = P\left(X < x\right)$$



The *probability density function* for a continuous random variable is defined by

$$f_X\left(x\right) = \frac{\partial F_X(x)}{\partial x}$$



Integral of this must equal 1

Random Variables

- Moments of random variables and the expectation operator

Probability distributions (cumulative distribution function and probability density function) can be described in terms of their paramaters $\mathbf{p}$ or their moments

Often we write

$$F_X(x, \underbrace{\mathbf{p}}_{\text{Parameters}}) \qquad f_X(x, \underbrace{\mathbf{p}}_{\text{Parameters}})$$

The parameters can be related to the moments and visa versa

Random Variables

- **Moments of random variables and the expectation operator**

  The i'th moment m_i for a continuous random variable X is defined through

$$m_i = \int_{-\infty}^{\infty} x^i f_X(x)dx$$

  The *expected value E[X]* of a continuous random variable X is defined accordingly as the first moment

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

Random Variables

- **Moments of random variables and the expectation operator**

  The i'th moment m_i for a **discrete** random variable X is defined through

$$m_i = \sum_{j=1}^{n} x_j^i \, p_X(x_j)$$

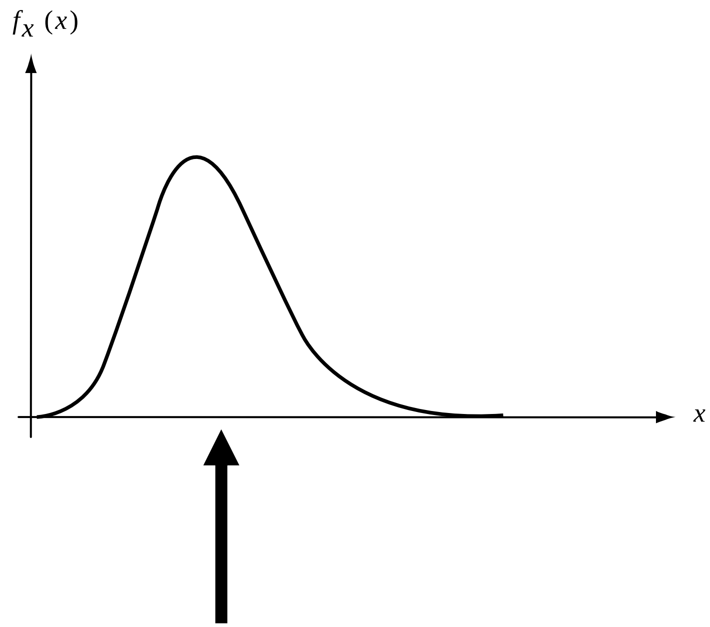
  The *expected value E[X]* of a **discrete** random variable X is defined accordingly as the first moment

$$\mu_X = E[X] = \sum_{j=1}^{n} x_j \, p_X(x_j)$$

Random Variables

- **Moments of random variables and the expectation operator**

The expected value (or mean value) of a random variable can be understood as the *center of gravity* of the probability density function of the random variable !



Random Variables

- **Moments of random variables and the expectation operator**

  The *standard deviation* σ_X of a continuous random variable is defined as the second central moment i.e. for a continuous random variable X we have

$$\underbrace{\sigma_X{}^2 = Var\left[X\right]}_{\text{Variance}} = E\left[(X-\mu_X)^2\right] = \int_{-\infty}^{\infty}(x-\underbrace{\mu_X}_{\text{Mean value}})^2 f_X(x)dx$$

  for a discrete random variable we have correspondingly

$$\sigma_X^2 = Var\left[X\right] = \sum_{j=1}^{n}(x_j-\mu_X)^2 p_X(x_j)$$

Random Variables

- **Moments of random variables and the expectation operator**

The ratio between the standard deviation and the expected value of a random variable is called the *Coefficient of Variation CoV* and is defined as

$$CoV[X] = \frac{\sigma_X}{\mu_X}$$

Dimensionless

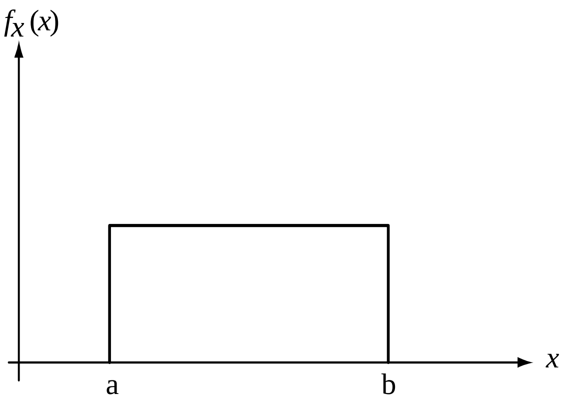
a useful characteristic to indicate the variability of the random variable around its expected value

Random Variables

- **Example – uniformly distributed random variable**

  probability density and cumulative distribution functions

$$f_X(x) = \begin{cases} 0, & x < a \\ \dfrac{1}{b-a}, & a \le x \le b \\ 0, & b < x \end{cases}$$

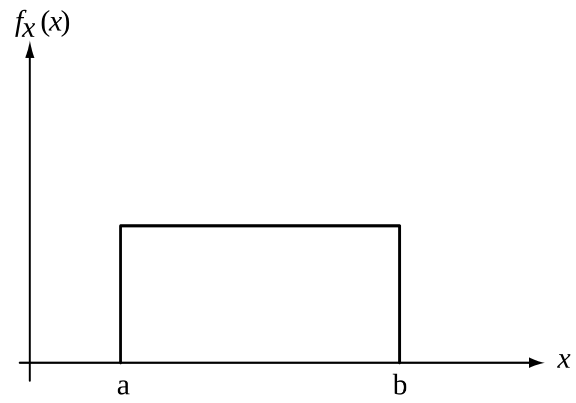


$$F_X(x) = \begin{cases} 0, & x < a \\ \int_a^x f_X(y)dy = \int_a^x \dfrac{1}{b-a} dy = \dfrac{(x-a)}{(b-a)}, & a \le x \le b \\ 1, & b < x \end{cases}$$

Random Variables

- **Example – uniformly distributed random variable**

  expected value and variance



$$\mu_X = E\left[X\right] = \int_a^b x \ f_X(x) dx = \int_a^b \frac{x}{b-a} dx = \left.\frac{x^2}{2(b-a)}\right|_a^b$$

$$= \frac{(b+a)}{2}$$

$$\sigma_X^2 = E\left[(X-\mu)^2\right] = \int_a^b (x-\mu)^2 f_X(x) dx = \int_a^b \frac{(x-\mu)^2}{(b-a)} dx = \left.\frac{\frac{1}{3}x^3 - x^2\mu + x\mu^2}{(b-a)}\right|_a^b$$

$$= \frac{1}{12}(b-a)^2$$