Bayesian Uncertainty Quantification in Engineering Science

Dr. Panagiotis Hatjidoukas, Prof. Costas Papadimitriou,
costasp@uth.gr

Week 1 Notes

Fall Semester 2016-2017

Prerequisites

Elementary knowledge in Probability and Statistics

Course Material

- **Class Notes**
- **Papers distributed as Required Reading (see class web site)**

OVERVIEW OF BAYESIAN UQ

Integrating Data with Models

Definitions

System – The real world (actual thing) to be analyzed

Mathematical Model – A collection of laws and mathematical equations introduced to describe the behavior of the actual system (usually based on physical laws or observations). It is based on theory and assumptions often used to construct a model.

Examples: algebraic equations, ordinary or partial differential equations (ODEs or PDEs), discrete equations

Computational Model – A numerical approximation or discretisation of the mathematical model in a form that can be implemented in computers. Most mathematical models are too complicated to solve them exactly and numerical approximations are most of the time introduced to solve the problem in available computers.

Examples: spatial and temporal discretization of PDEs, numerical integration, truncation of infinite sums

Sources of Uncertainty

- **Modeling (or Structural) Uncertainty**

  Arise from assumptions used to build a mathematical model for

  A. representing the physical system (the real thing)

  B. representing the interactions of the system with the environment

  Comes from the lack of knowledge for the underlying true physics, leading to discrepancies (model bias) between the predictions from the model and the observations (measurements). The model inadequacy is always present and the question is how to select the best models over a family of alternative models introduced to model the same physical phenomenon.

- **Parametric Uncertainty**

  Arise from lack of knowledge of the appropriate values of the parameters of a mathematical model. Examples include the material properties of a continuum such as solid or fluid, the properties involved in constitutive laws, the boundary conditions, etc.

Sources of Uncertainty

- **Computational (or Algorithmic) Uncertainty**

  linked to the numerical uncertainty arising from the numerical approximations introduced to implement the analysis in a computer. Examples include spatial and temporal discretization of PDEs using finite element methods, finite difference methods or particle methods.

- **Measurement uncertainty**

  arises from the variability in the values of the experimental properties due to variability in experimental set up, errors in the measuring equipment, and inaccuracies in the data acquisition system.

Example 1: Solid Mechanics/Structural Dynamics

- **Modeling (or Structural) Uncertainty**

  Selection of linear or nonlinear constitute laws to represent the material behavior (e.g. stress-strain relationship)

  Selection of boundary conditions

- **Parametric Uncertainty**

  The values of the constant parameters involved in the constitutive laws are not completely known (modulus of elasticity, Poisson ratio, etc)

  The values of the stiffness in isolated parts of the structure are unknown

  stiffness and damping values of isolation devices are uncertain (dampers, etc)

  For contact problems, friction, restitution coefficients are not completely known

  Computational (or Algorithmic) Uncertainty

  Spatial discretization of the PDEs using finite element methods

  Temporal discretization of the resulting ODEs

- **Measurement uncertainty**

  Uncertainties in measuring the acceleration, strains, etc, in various locations of the structure due to errors in the measuring equipment, and inaccuracies in the data acquisition system.

Example 2: Fluid Dynamics

- **Modeling (or Structural) Uncertainty**

  Selection of flow model (Filtered Navier Stokes equations + Turbulence model)

  Selection of boundary conditions

- **Parametric Uncertainty**

  The values of the constant parameters involved in the Turbulence model

  The values of the model are not suitable near boundaries

  For some problems (flow in hydrophobic surfaces) the parameters of the boundary conditions are not known.

  Computational (or Algorithmic) Uncertainty

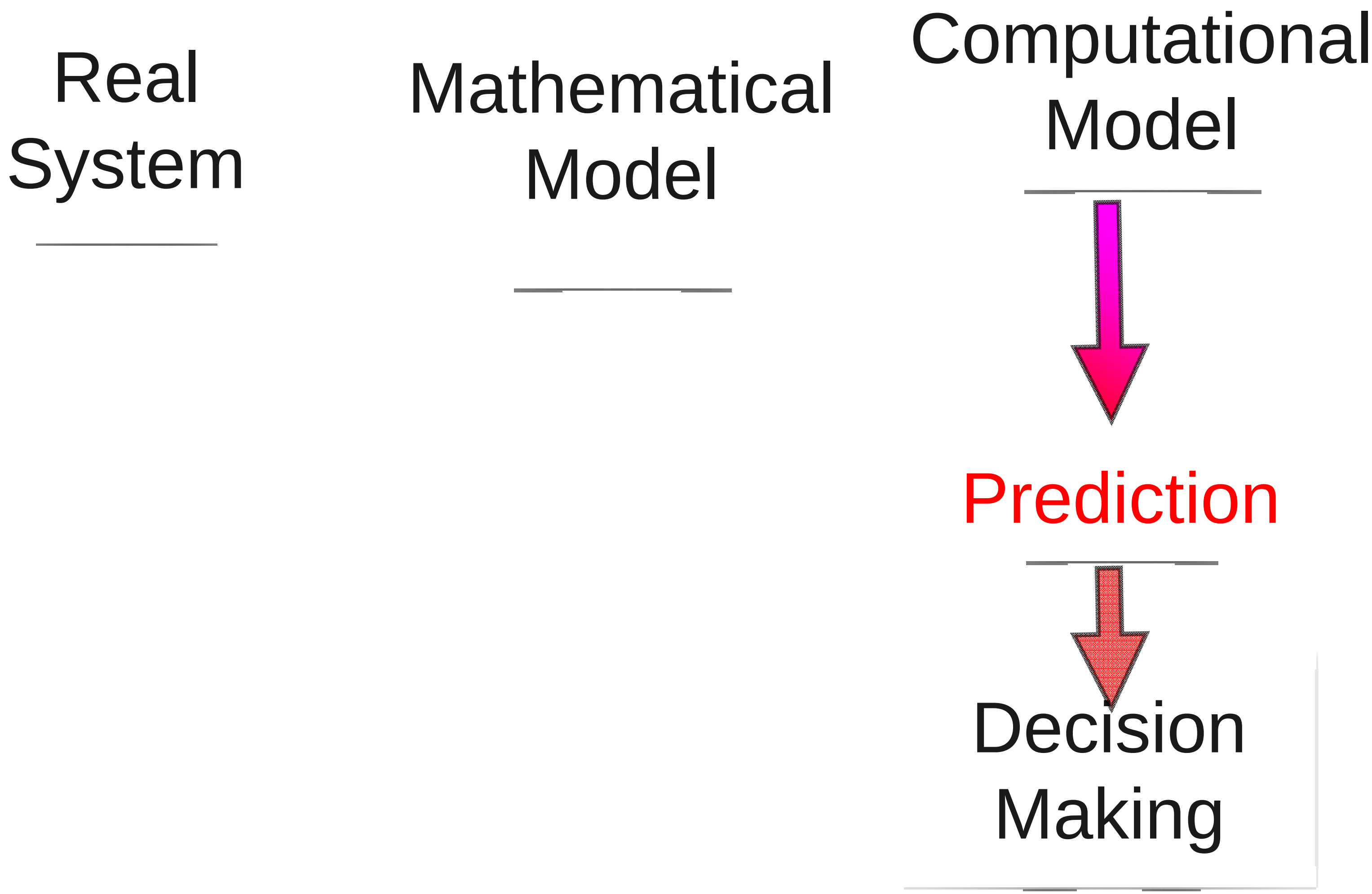
  Spatial discretization of the PDEs using numerical methods (grids, particles, etc.)

  Temporal discretization of the resulting ODEs

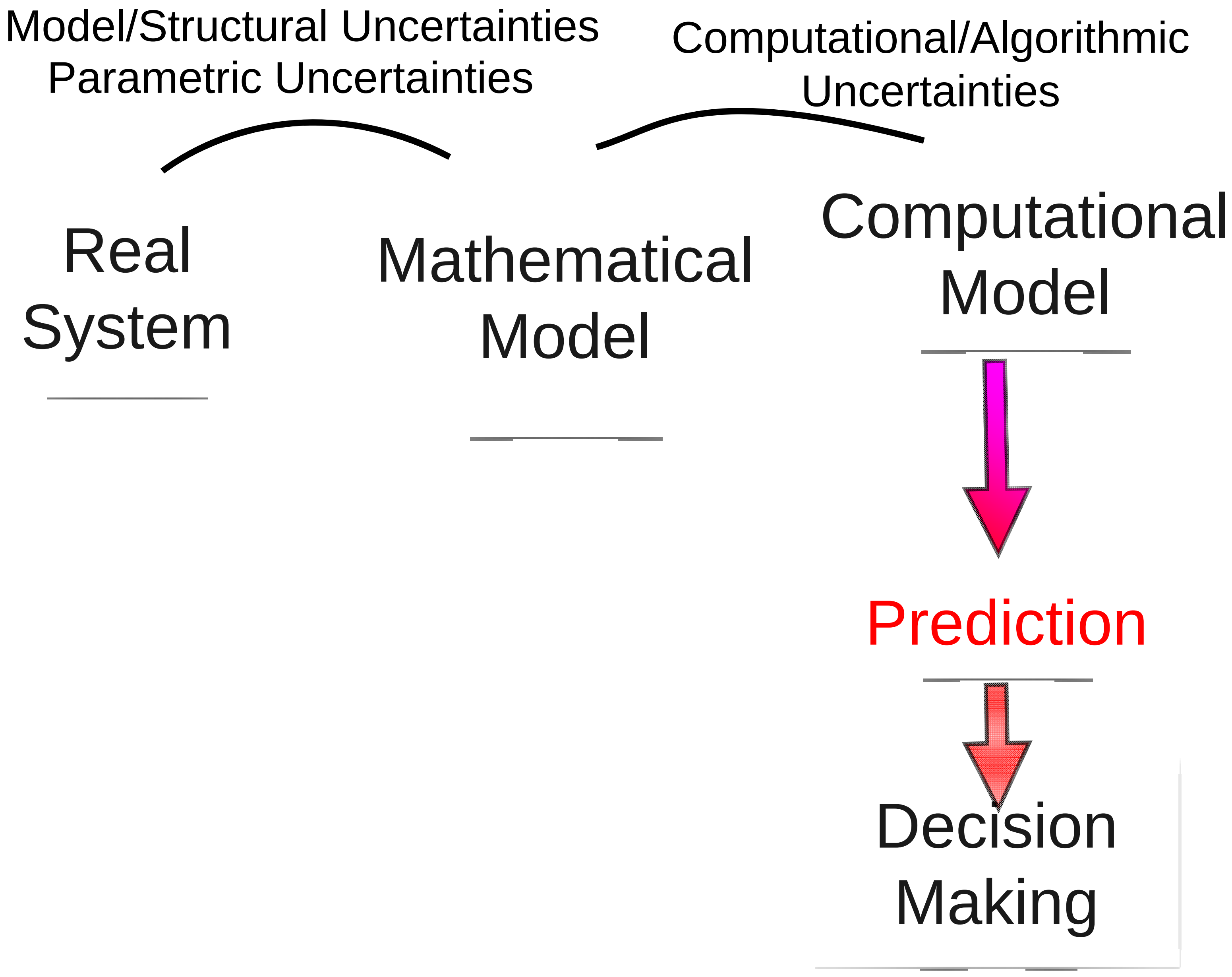
- **Measurement uncertainty**

  Uncertainties in measuring flow quantities such as flow fields and drag coefficients due to errors in the measuring equipment, and inaccuracies in the data acquisition system.

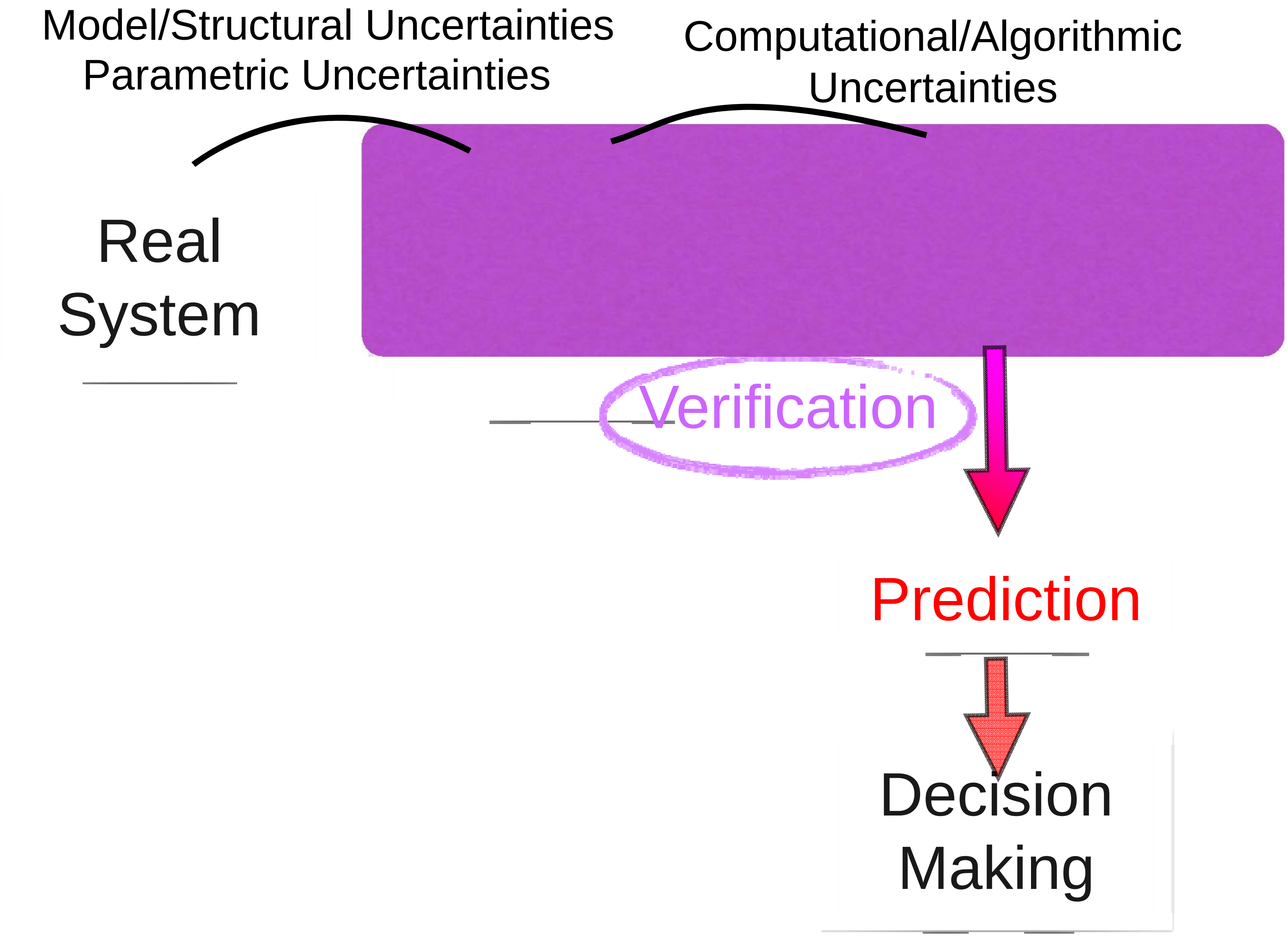
Model-based Decision Making under Uncertainty



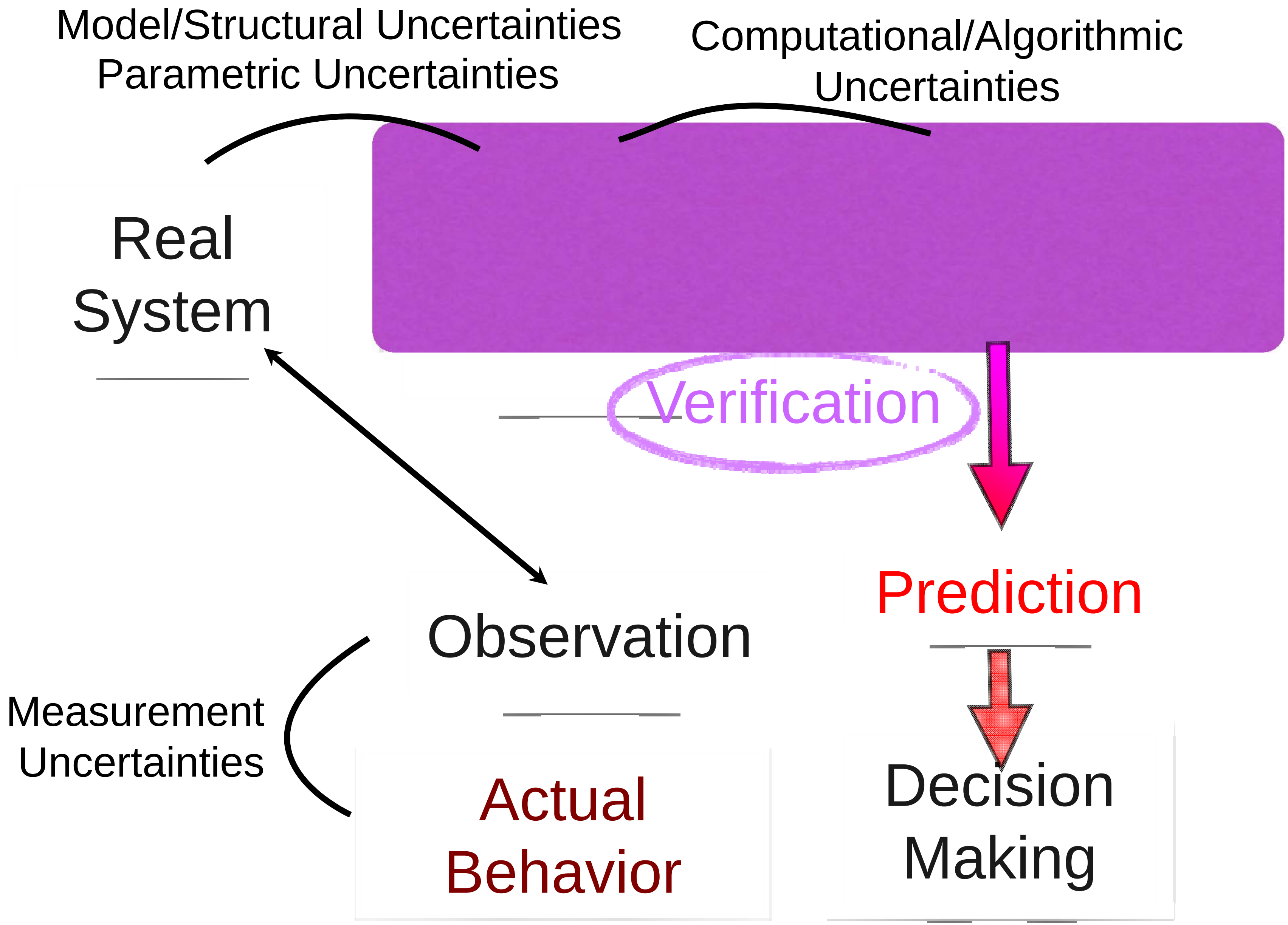
Model-based Decision Making under Uncertainty



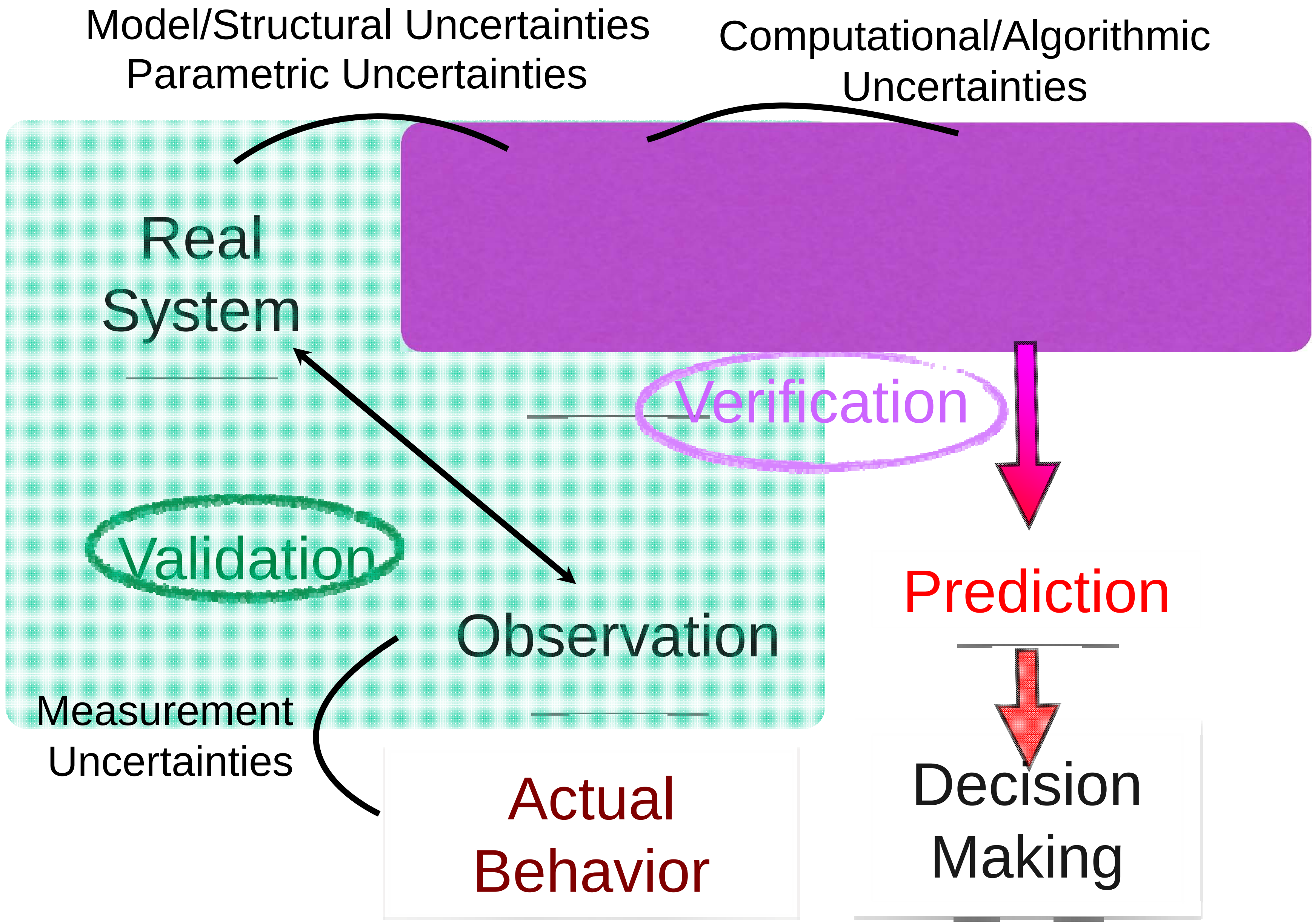
Model-based Decision Making under Uncertainty



Model-based Decision Making under Uncertainty



Model-based Decision Making under Uncertainty



Uncertainty Quantification

Probability is used to quantify uncertainties. Probability models are used to model the missing/incomplete information.

We use **Cox interpretation of probability,** representing the **degree of belief or plausibility of a proposition based on available information.** It expresses our relative belief in the truth of various propositions. It ranks the propositions by assigning a real number to each one. The largest the numerical value associated with a proposition, the more we believe it.

Probabilities are always conditional on information and this conditioning must be stated explicitly.

Cox has shown that for **consistent plausible reasoning** the real number we attach to our beliefs of the propositions have to **obey the usual rules of probability theory.** The **calculus of probability** is thus used to **manage (quantify and propagate) uncertainties** (incomplete information) in system analysis.

Probability density functions (PDF) assigned on a parameter are used to quantify how plausible each possible value of this parameter is.

Probability Logic Fundamentals

Let a , b and c be propositions. Also define

$P(b \mid a) =$ plausibility of proposition b conditioned on the information contained in proposition a

The **axioms** of probability logic are stated as

$$P(b \mid a) \geq 0$$

$$P(b \mid a, b) = 1$$

$$P(b \mid a) + P(\sim b \mid a) = 1 \qquad \text{Sum rule}$$

$$P(c, b \mid a) = P(c \mid b, a)\, P(b \mid a) \qquad \text{Product rule}$$

Properties

$$P(b \mid a) \in [0,1]$$

$$P(c \text{ or } b \mid a) = P(c \mid a) + P(b \mid a) - P(c, b \mid a)$$

$$P(c \text{ or } b \mid a) = P(c \mid a) + P(b \mid a)$$

if b and c cannot both be true (mutually exclusive) conditioned on a

Probability Logic Fundamentals

Properties

If only one of $b_1,\ b_2,\ \ldots\ ,\ b_n$ is true based on the information in a, then

Marginalization Theorem

$$P(c \mid a) = \sum_{k=1}^{n} P(c, b_k \mid a)$$

Total Probability Theorem

$$P(c \mid a) = \sum_{k=1}^{n} P(c \mid b_k, a)\ P(b_k \mid a)$$

Bayes Theorem

$$P(b_k \mid c, a) = \frac{P(c \mid b_k, a)\ P(b_k \mid a)}{\sum_{k=1}^{n} P(c \mid b_k, a)\ P(b_k \mid a)}, \qquad k = 1, \ldots, n$$

Discrete Variables

Let X be an uncertain variable that can take discrete values $x_1, \ldots, x_n$. The following notation

$$p(x_i \mid a) = P(X = x_i \mid a)$$

is used to denote the probability of the proposition $X = x_i$, i.e. the variable X to take the value x_i, given the information in proposition a. Note that the propositions $X = x_1$, $X = x_2$, …, $X = x_n$ are mutually exclusive. Let also Y be another uncertain discrete variable with possible values $y_1, \ldots, y_n$. It can be readily verified that the following hold true.

Marginalization Theorem

$$p(x_i \mid a) = \sum_{k=1}^{m} p(x_i, y_k \mid a)$$

Total Probability Theorem

$$p(x_i \mid a) = \sum_{k=1}^{m} p(x_i \mid y_k, a)\, p(y_k \mid a)$$

Bayes Theorem

$$p(y_k \mid x_i, a) = \frac{p(x_i \mid y_k, a)\, p(y_k \mid a)}{\sum_{k=1}^{m} p(x_i \mid y_k, a)\, p(y_k \mid a)}$$

Continuous Variables

Let X be an uncertain variable that can take values on a continuous domain $x \in [x_{start}, x_{end}]$. The following notation

$$P(X \le x \mid a) \equiv F(x \mid a)$$

is used to denote the probability of the proposition $X \le x$, i.e. the variable X to take value less than x, given the information in proposition a. It is referred as the cumulative probability distribution of a variable X. Define the probability distribution function $f(x)$ of a variable X from the expression

$$P(x < X \le x + dx \mid a) = f(x \mid a)dx \qquad (1)$$

It can be readily derived that

$$f(x \mid a) = \frac{dF(x \mid a)}{dx} \qquad (2)$$

using the fact that the statement $X \le x + dx \mid a$ is the sum of the statement $X \le x \mid a$ and $x < X \le x + dx \mid a$ and that the statements $X \le x \mid a$ and $x < X \le x + dx \mid a$ are mutually exclusive so that using the sum rule

$$P(X \le x + dx \mid a) = P(X \le x \text{ or } x < X \le x + dx \mid a) = P(X \le x \mid a) + P(x < X \le x + dx \mid a)$$

which results in

$$P(x < X \le x + dx \mid a) = P(X \le x + dx \mid a) - P(X \le x \mid a) = F(x + dx \mid a) - F(x \mid a)$$

Using (1), one derives that $F(x + dx) - F(x) = f(x \mid a)dx$ which results in (2).

Continuous Variables

Finally, it can be readily shown that for the probability distribution function $f(x \mid a)$ of a continuous variable X , the following hold true.

Marginalization Theorem

$$f(x \mid a) = \int f(x, y \mid a)\ dy$$

Total Probability Theorem

$$f(x \mid a) = \int f(x \mid y, a)\ f(y \mid a)\ dy$$

Bayes Theorem

$$f(y \mid x, a) = \frac{f(x \mid y, a)\ f(y \mid a)}{\int f(x \mid y, a)\ f(y \mid a)\ dy}$$

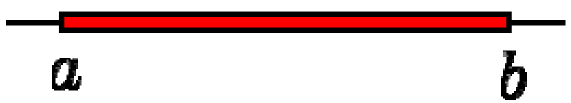
Uncertainty Quantification: Concept Demonstration

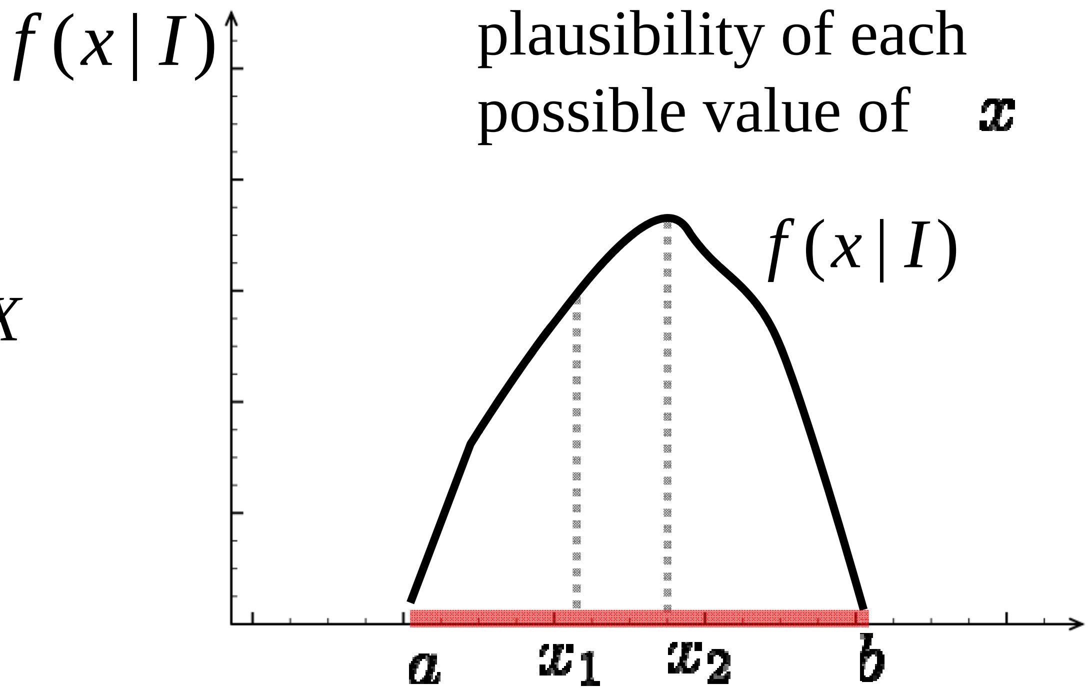
Consider a mathematical model and a single parameter X of this model. We assume that we have incomplete information about the value of the parameter. We know that the parameter can take values in the range [a,b] (range of possible values). In the absence of observations, lets assume that we can specify how plausible is each possible value of the parameter based on theoretical arguments, expert opinions or engineering experience.

A PDF $f(x|I)$ is postulated to specify how plausible is each possible value x of the parameter based on the available information. The value of X is a constant and not a random variable. Its constant value is uncertain to us. A number $f(x|I)$ is assigned to each possible value of the parameter to represent our belief that one value, say x_2 , is more plausible than another value, say x_1 .

feasible values of parameter X

support $x \in [a, b]$





Uncertainty Quantification: Concept Demonstration

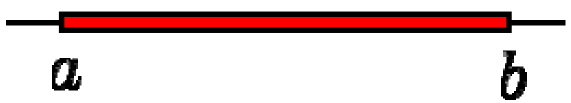
Using the probability theory, the numbers/values $f(x \mid I)$ for all $x \in [a, b]$ have to satisfy

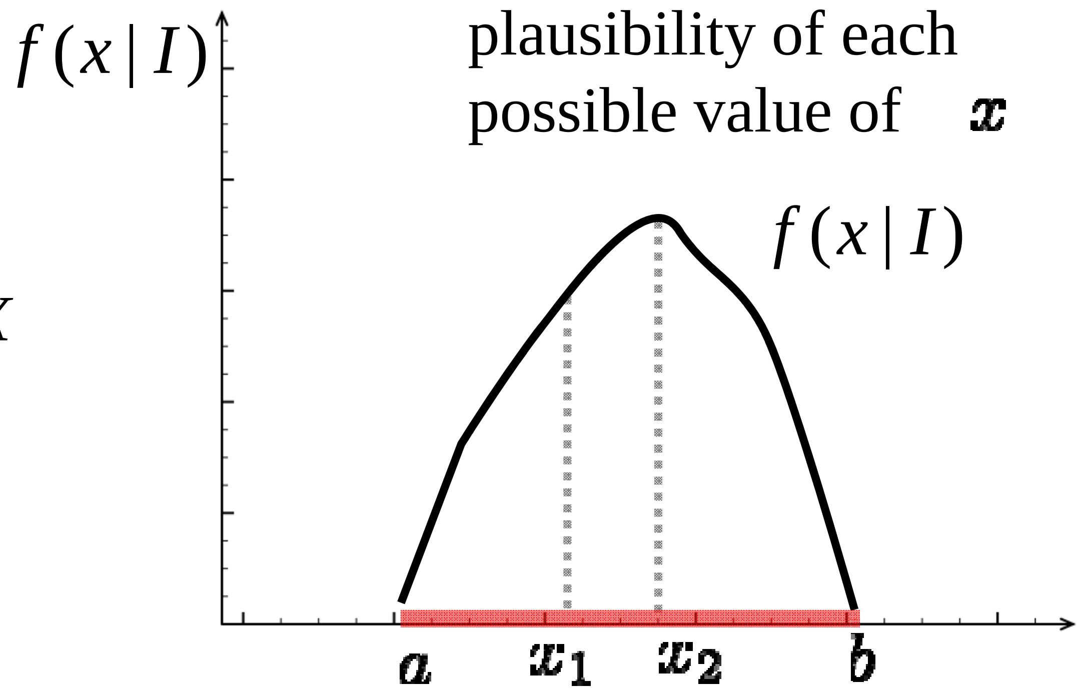
$$\int_a^b f(x \mid I)\, dx = 1$$

A first question that arises is how to propagate this uncertainty through the system. We will use the calculus of probability and we will demonstrate this with an very simple example.

feasible values of parameter X

support $x \in [a, b]$



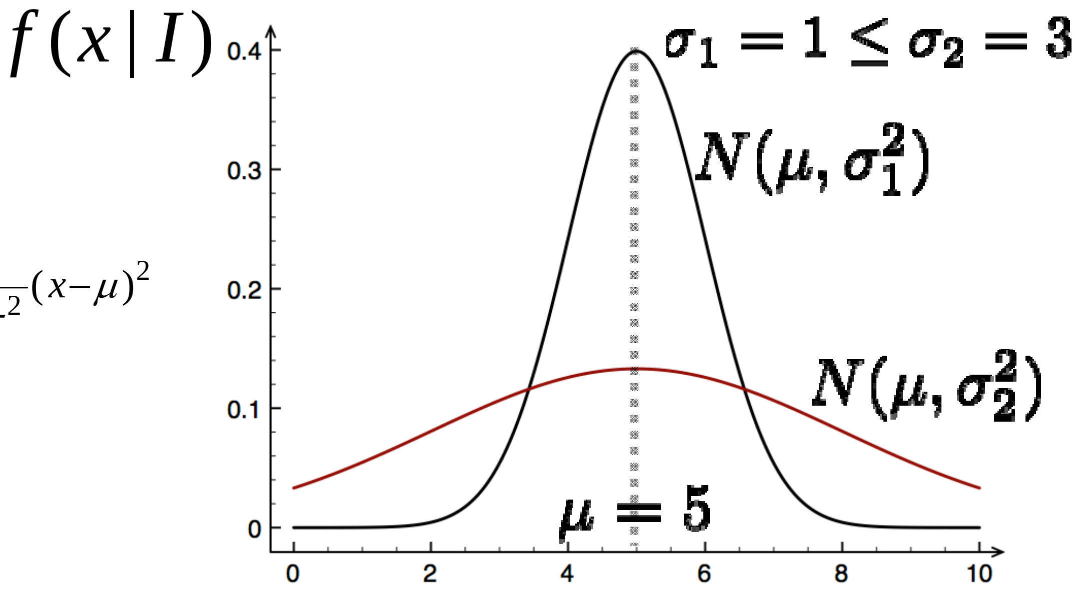


Examples of PDFs: Normal (Gaussian) & Uniform

Normal (Gaussian) Distribution

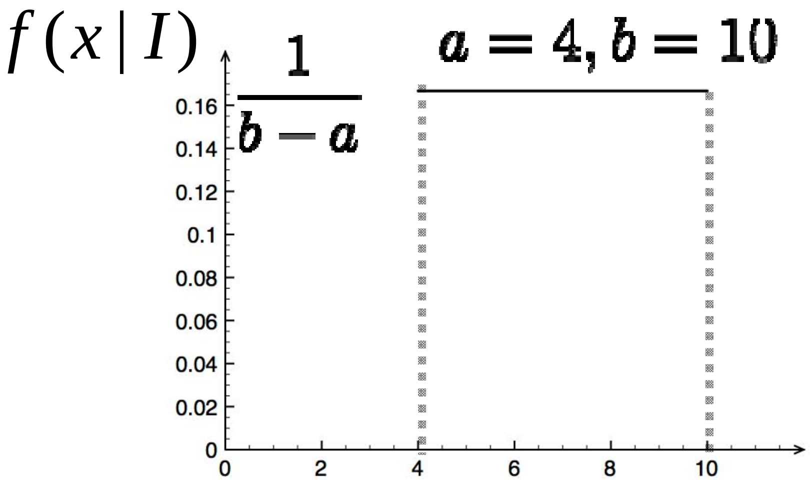
$$X \sim N(\mu, \sigma^2)$$

$$f(x \mid I) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2\sigma^2}(x-\mu)^2}$$

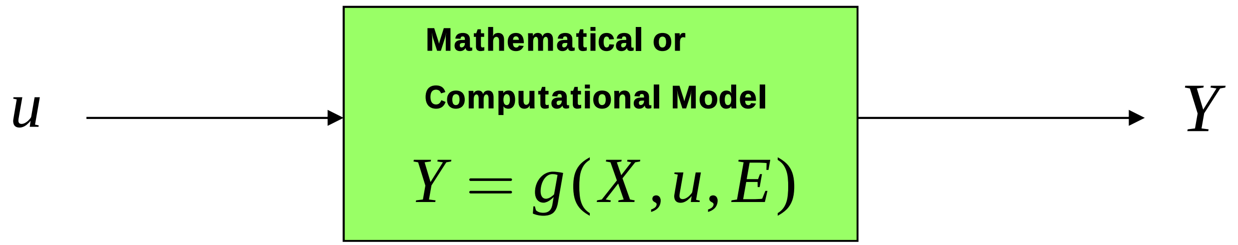


Uniform Distribution

$$X \sim U(a, b)$$



Uncertainty Propagation (Prior Analysis)



X is the uncertain model parameter; x is a possible value of X

Y is the uncertain output quantity of interest (QoI); y is a possible value of Y

E is the prediction error

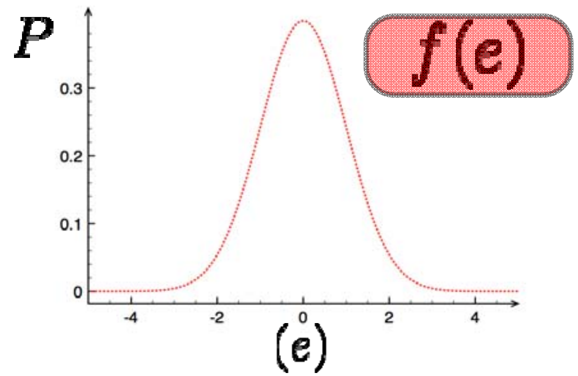
u is the input; Assumed in this example to be known

Example 1 (Special Linear Case): $Y = X + E$

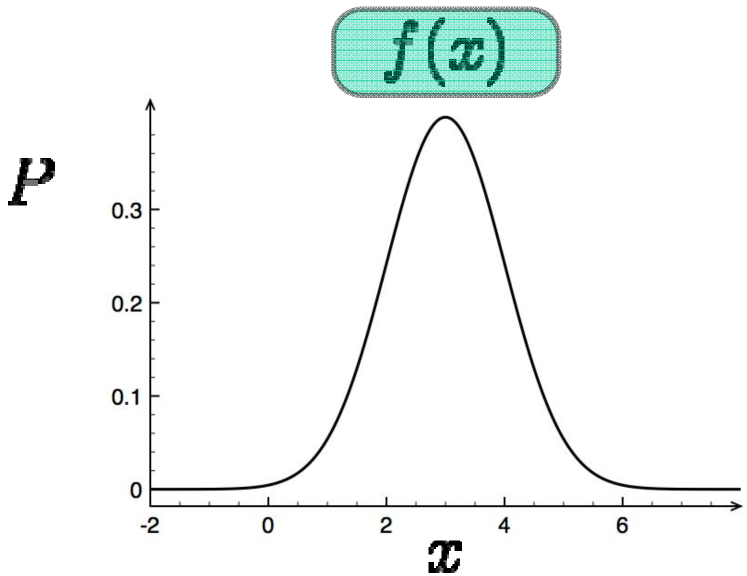
X and E are independent

$f(x) = N(\mu, \sigma^2)$

$E \sim N(0, 1^2)$



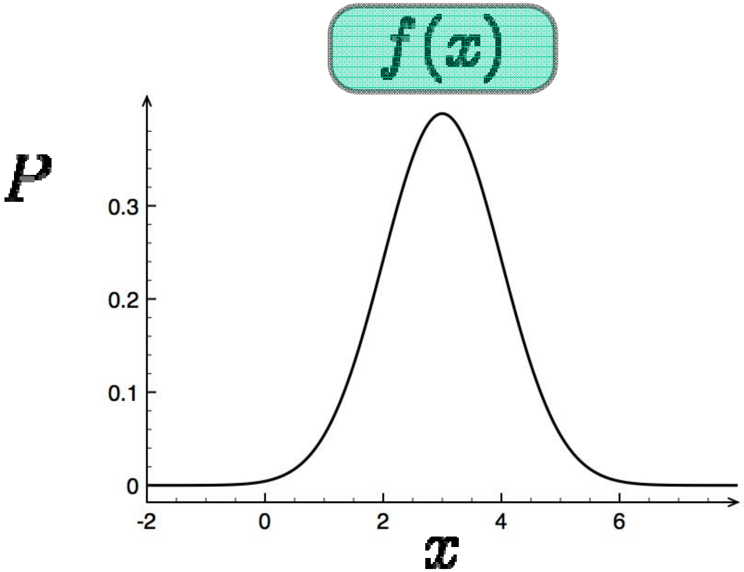
Uncertainty Propagation (Prior Analysis)

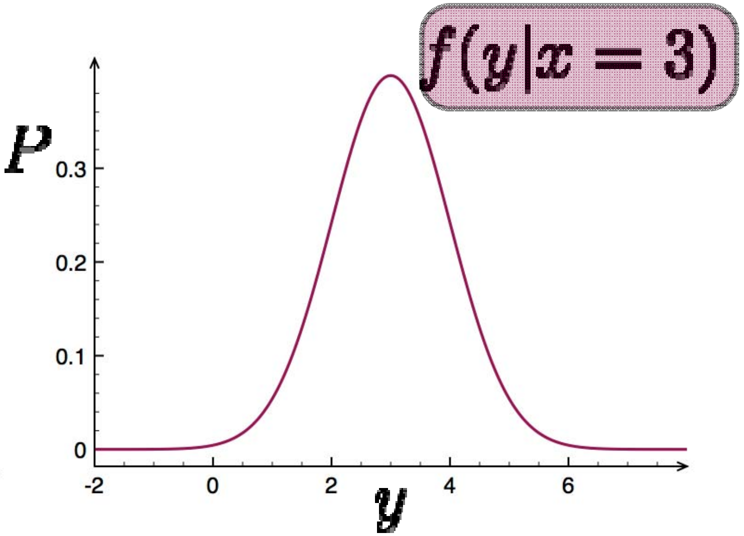


Uncertainty Propagation (Prior Analysis)

Using the calculus of probability, the probability distribution (PDF) of y conditioned on the value of x is

$$f(y \mid x) = N\left(x, 1^2\right)$$





output QoI

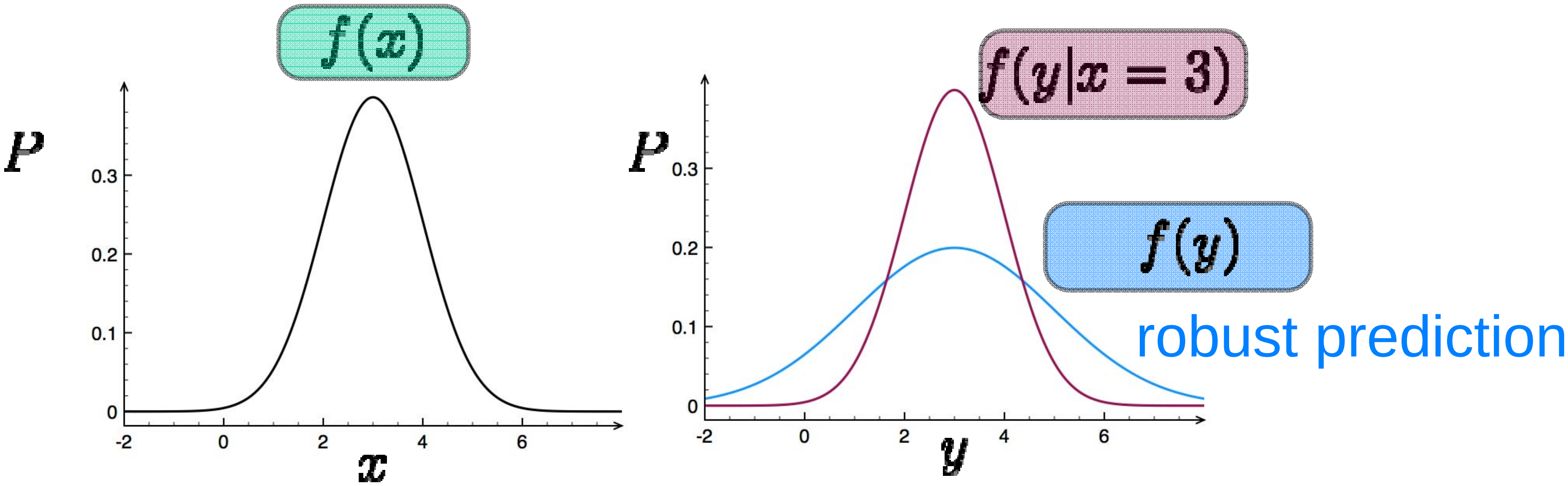
Uncertainty Propagation (Prior Analysis)

Using the calculus of probability, the probability distribution (PDF) of y conditioned on the value of x is

$$f(y|x) = N(\mu, 1^2)$$

The probability distribution of the output QoI Y is given by

$$f(y) = N(\mu, \sigma^2 + 1^2)$$



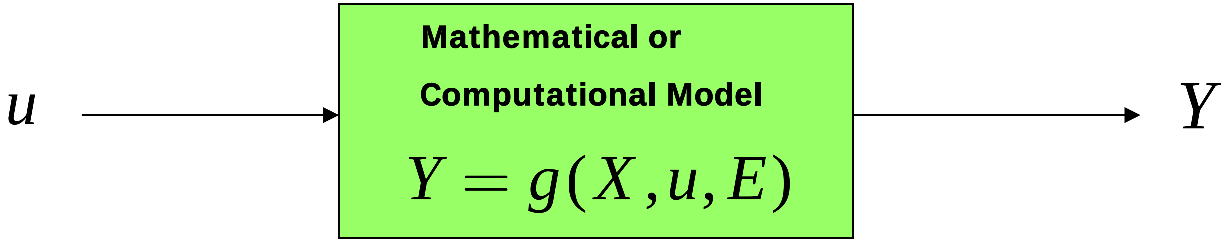
output QoI

Uncertainty Quantification

Measures of Uncertainty in QoI

- PDF
- Mean, std, skewness (asymmetry), curtosis (deviation from normality)
- Confidence intervals
- Probability of QoI lying in a predefined set (failure probability; probability of unacceptable performance, first passage problem)

Uncertainty Propagation



Tools for uncertainty propagation in <u>prior system analysis</u>

- Analytical (Useful for demonstration of theory; not applicable in practical engineering problems)
- Local expansion techniques: Perturbation, Taylor series, etc (small uncertainties)
- Functional expansion methods: Neumann, Polynomial Chaos
- Numerical integration methods: sparse grid methods
- Reliability-based approximate or asymptotic methods: FORM, SORM
- Stochastic simulation methods: Monte Carlo, Importance sampling, adaptive sampling, subset simulation, line sampling, etc

Uncertainty Quantification based on Observations

Observations or measurement data are collected from the system.

Let these observations be denoted by $\hat{Y}$.

The problem is now to update the uncertainty in the parameters using the information contained in the observations. This is achieved using the Bayes theorem.

BAYES THEOREM

Likelihood **Prior**

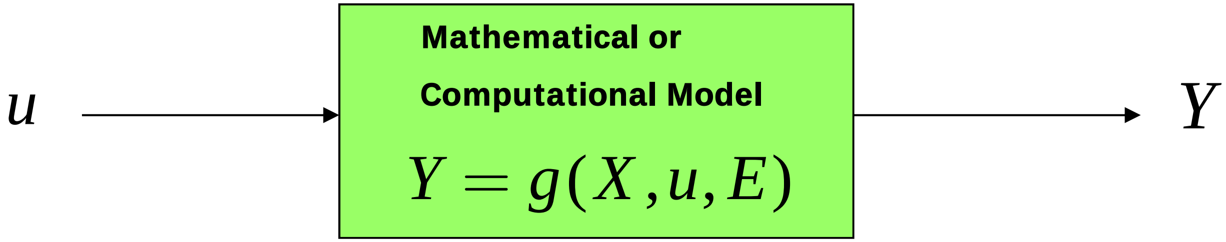
Posterior

$$f(x \mid \hat{Y}, I) = \frac{f(\hat{Y} \mid x, I)\, f(x \mid I)}{f(\hat{Y} \mid I)}$$

Evidence

Bayes theorem gives the **posterior PDF (uncertainty)** of the model parameters which quantifies how plausible is each possible value of the parameter in light of the available observations from the system. This updated PDF of the uncertainty in the parameters is based on two quantities. The first one is called the **likelihood** and gives the probability to observe the data given a possible value of the model parameters. The likelihood is influenced by the data. The second one is the **prior probability** of the model parameters, which contains any information before data are utilized. The term in the denominator is called the **evidence** and for parameter estimation is just a normalization constant (does not depend on the parameters). For model selection, however, this term plays a crucial role.

Uncertainty Quantification (Posterior Analysis)



X is the uncertain model parameter; x is a possible value of X

Y is the uncertain output quantity of interest (QoI); y is a possible value of Y

E is the prediction error

u is the input; Assumed in this example to be known

Example 1 (Special Linear Case): $Y = X + E$

X and E are independent

$f(x) = N(\mu, \sigma^2)$

$E \sim N(0, 1^2)$

$\hat{Y}$ A **single observation** (data) from the system

Uncertainty Quantification (Posterior Analysis)

Given the data $\hat{Y}$, the posterior PDF of the parameter x is obtained from Bayes theorem as

$$f(x|\hat{Y},I) = \frac{f(y=\hat{Y}|x,I)\,f(x|I)}{f(y=\hat{Y}|I)}$$

Using the mathematical model $Y = X + E$ we obtain the likelihood in the form

$$f(y|x,I) = N\left(x,1^2\right)$$

which for $y = \hat{Y}$ gives

$$f(\hat{Y}|x,I) = \frac{1}{\sqrt{2\pi 1}}\exp\left[-\frac{1}{2}\frac{(\hat{Y}-x)^2}{1^2}\right]$$

Substituting in (1) along with the prior PDF

$$f(x|I) = N\left(\mu,\sigma^2\right)$$

one readily obtains that

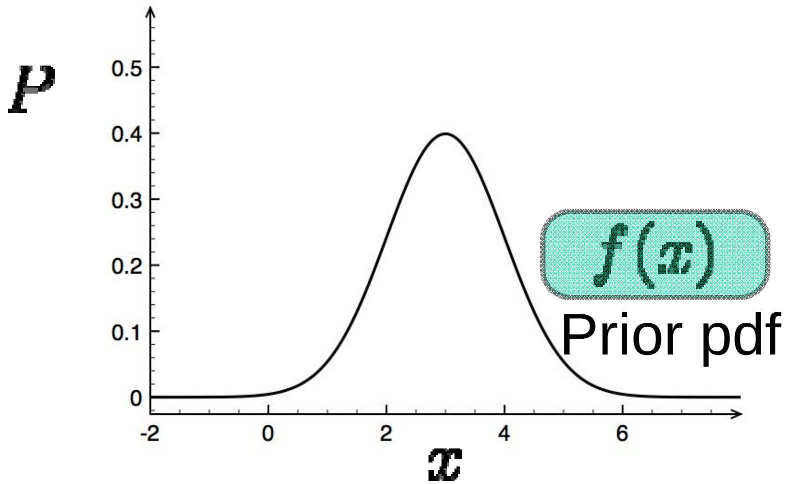
$$f(x|\hat{Y},I) \propto \exp\left[-\frac{1}{2}\left(\frac{(\hat{Y}-x)^2}{1^2}+\frac{(x-\mu)^2}{\sigma^2}\right)\right]$$

which can be shown to simplify to a normal distribution for the **posterior PDF** of the parameter

$$f(x|\hat{Y},I) = N\left(\frac{\mu+\hat{Y}\sigma^2}{1+\sigma^2},\frac{\sigma^2}{1+\sigma^2}\right)$$

Uncertainty Quantification (Posterior Analysis)

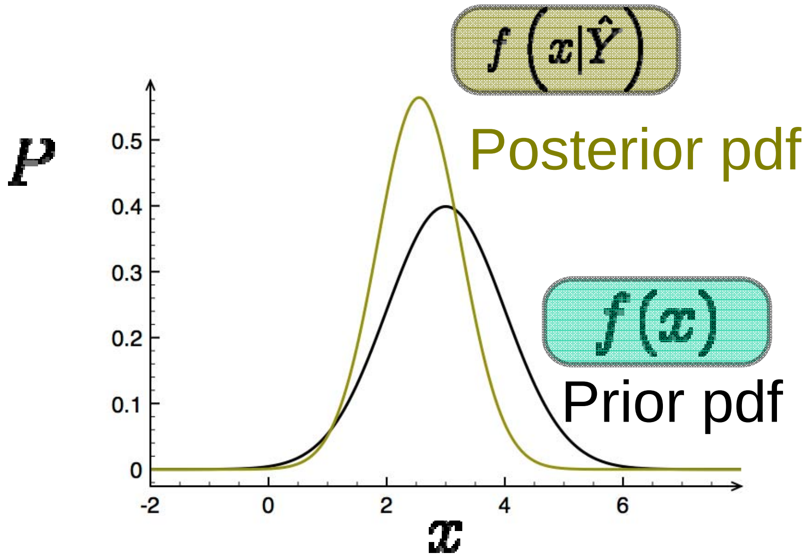
prior PDF $f(x) = N\left(\mu, \sigma^2\right)$



Uncertainty Quantification (Posterior Analysis)

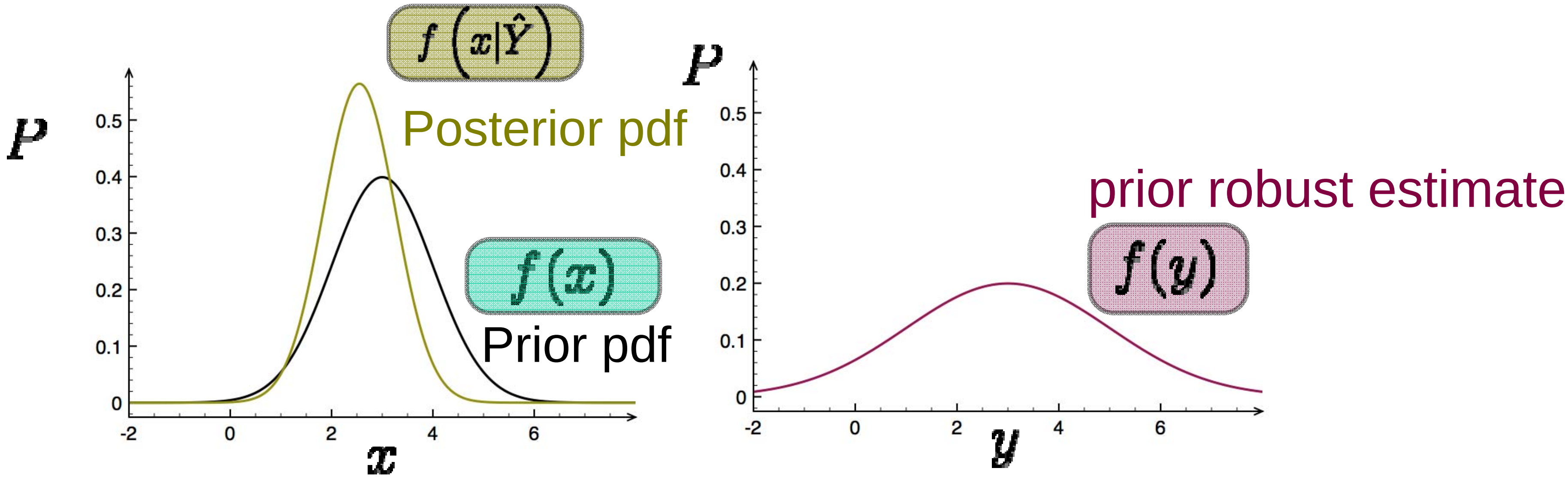
prior PDF $f(x) = N\left(\mu, \sigma^2\right)$

posterior PDF $f(x \mid \hat{Y}) = N\left(\frac{\mu + \hat{Y}\sigma^2}{1+\sigma^2}, \frac{\sigma^2}{1+\sigma^2}\right)$



Uncertainty Propagation (Posterior Analysis)

The **prior robust prediction** for the QoI Y is $f(y \mid I) = N\left(\mu, \sigma^2 + 1\right)$

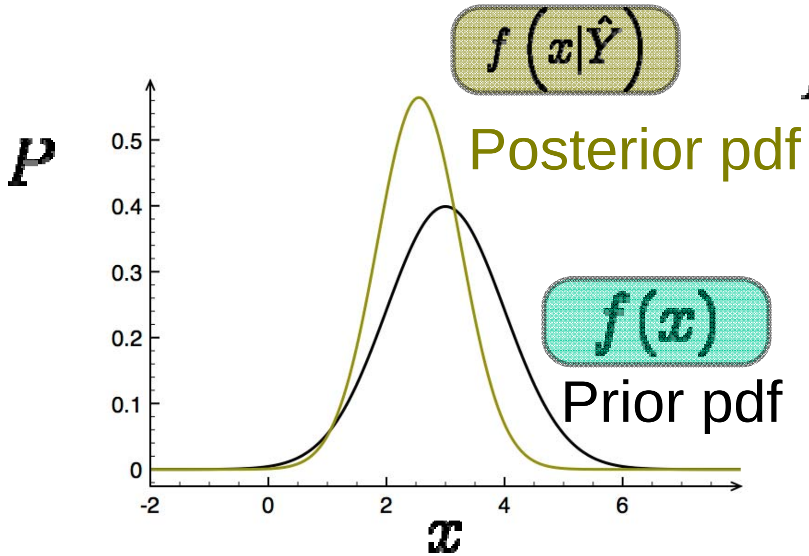


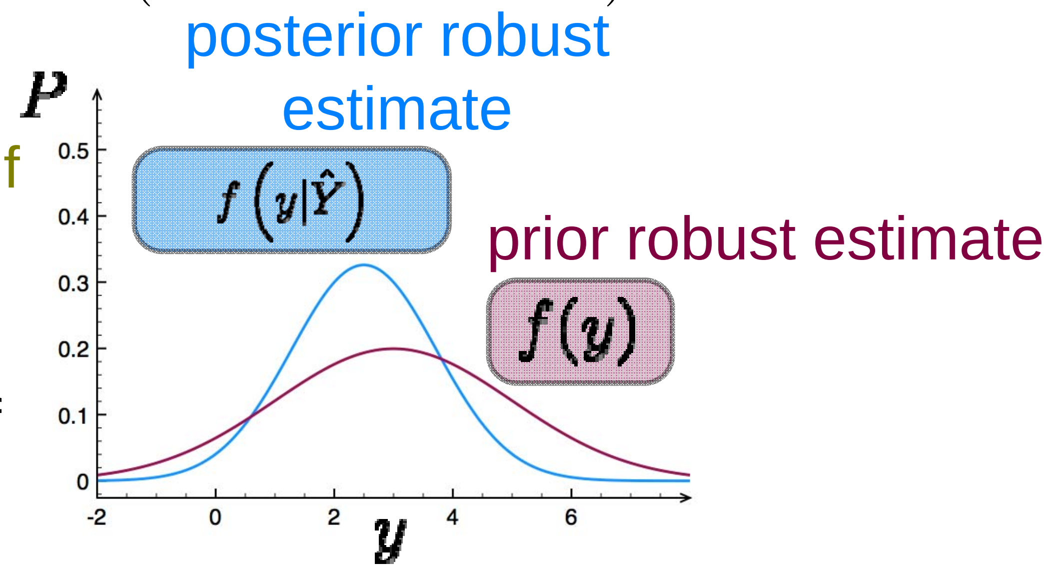
Uncertainty Propagation (Posterior Analysis)

The **prior robust prediction** for the QoI Y is $f(y \mid I) = N\left(\mu, \sigma^2 + 1\right)$

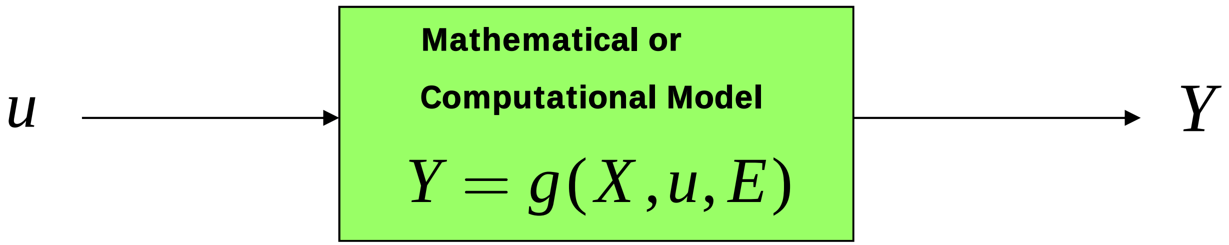
The **posterior robust prediction** for the QoI Y, taking into account the observations (measurements), is readily obtained form the fact that the **posterior PDF of the model parameter is normal**

$$f(y \mid \hat{Y}, I) = N\left(\frac{\mu + \hat{Y}\sigma^2}{1+\sigma^2}, \frac{\sigma^2}{1+\sigma^2} + 1\right)$$





Uncertainty Quantification: Nonlinear Model



X is the uncertain model parameter; x is a possible value of X

Y is the uncertain output quantity of interest (QoI); y is a possible value of Y

E is the prediction error

u is the input; Assumed in this example to be known

Example 2 (Nonlinear Case): $Y_k = g(X, u) + E_k$

$\boldsymbol{X}$ and $\boldsymbol{E}$ are independent

$f(x) = N(\mu, \sigma^2)$

$E_k \sim N(0, 1^2)$ independent and identically distributed (i.i.d)

$\hat{Y} = \{\hat{Y}_1, \cdots, \hat{Y}_N\}$ **Observations** (data) from the system

Uncertainty Propagation (Posterior Analysis)

The **posterior PDF** of the parameter is given by (for analysis see notes written on the board)

$$f(x\,|\,\hat{Y}) \propto \prod_{k=1}^{N} \exp\left[-\frac{1}{2}\left(\hat{Y}_k - g(x,u)\right)^2\right] \exp\left[-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}\right] \qquad \textbf{(1)}$$

The posterior PDF **does not follow a simple known distribution**. This complicates the identification of the region of most plausible values the parameters and the subsequent system analysis such as the **posterior robust prediction** since it can not be computed using the arguments of example 1. Sampling from the posterior PDF is also a challenging problem. **Stochastic simulation methods** such as **Markov Chain Monte Carlo** have been developed to sample from the posterior PDF.

Asymptotic approximations (valid for large number of data) can also be used to approximate the posterior PDF by a normal PDF.

Using the total probability theorem, the **posterior robust prediction** of a QoI Y takes the form

$$f(y\,|\,\hat{Y}) = \int f(y\,|\,x)\; \underbrace{f(x,\hat{Y})}_{\textbf{Posterior PDF}}\; dx$$

This integral can only be evaluated using numerical integration. However, this is **inefficient for more than a few model parameters**. Need to use more efficient techniques to evaluate such integrals. Such tools include **asymptotic approximations** and **stochastic simulation algorithms**.

Derivation of Likelihood

Example 1: Nonlinear Model

Consider the mathematical model

$$Y = g(x,u) + E$$

of a physical process/system, where E is a Gaussian distribution, i.e. $E \sim N(0,\sigma^2)$. Given the values of x and σ^2 the output quantity of interest Y follows the Gaussian distribution $Y \sim N(g(x,u),\sigma^2)$ or, equivalently, the uncertainty in y is given by the PDF

$$p(y \mid \mu, \sigma^2, I) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{1}{2\sigma^2}(y - g(x,u))^2 \right] \qquad (1)$$

Given a set of independent observations/data $D \equiv (\hat{Y}_1, \hat{Y}_2, \ldots, \hat{Y}_N) \equiv \{\hat{Y}_k\}_{1\to N}$, we are interesting in updating the uncertainty in the variables x and σ^2. This involves the estimation of the likelihood.

Bayes Theorem: Using Bayes' theorem, the inference about the values of x and σ^2 given the data and the information I (I includes the selection of the Gaussian model) is expressed by the posterior PDF

$$p(x,\sigma^2 \mid \{\hat{Y}_k\}_{1\to N}, I) \propto p(\{\hat{Y}_k\}_{1\to N} \mid x,\sigma^2, I)\; p(x,\sigma^2 \mid I) \qquad (2)$$

Derivation of Likelihood

Estimation of Likelihood: To estimate the likelihood $p(\{\hat{Y}_k\}_{1\to N} \mid x, \sigma^2, I)$, one can use the fact that the data are independent and apply successively the product rule of the axioms of probability, given by

$$p(b, a \mid I) = p(b \mid a, I)\ \ p(a \mid I) \tag{1}$$

to finally derive that

$$\begin{aligned} p(\{\hat{Y}_k\}_{1\to N} \mid x, \sigma^2, I) &= \prod_{k=1}^{N} p(\hat{Y}_k \mid x, \sigma^2, I) \\ &= \prod_{k=1}^{N} \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{1}{2\sigma^2}(\hat{Y}_k - g(x,u))^2\right] \\ &= \frac{1}{\sqrt{2\pi}\sigma^N} \exp\left[-\frac{1}{2\sigma^2}\sum_{k=1}^{N}(\hat{Y}_k - g(x,u))^2\right] \end{aligned} \tag{2}$$

Bayesian Uncertainty Quantification and Propagation

Tools for uncertainty quantification and propagation in posterior system analysis

- Asymptotic approximations
- Stochastic simulation methods: variants of MCMC (Markov Chain Monte Carlo), Transitional MCMC, Sequential Monte Carlo, DRAM, etc

Bayesian Uncertainty Quantification and Propagation

Issues to be considered

- Multi-dimensional uncertain parameter space (we only discussed the 1-d case)
- Models for which the QoI depends nonlinear on the parameters (we discussed the linear case)
- Selection of prior PDF for the model parameters
- Ranking alternative models introduced to represent the system – Model averaging
- Account for measurement and computational uncertainties
- Approximate methods for posterior system analysis
- Stochastic simulation methods for posterior system analysis: variants of MCMC (Markov Chain Monte Carlo), Transitional MCMC, DRAM, etc
- Optimal experimental design: what quantities to measure in order to get the most information out of the data in order to reduce uncertainties in model parameters and predictions.
- ...