Bayesian Uncertainty Quantification

Part I: Introduction

Petros Koumoutsakos

CSE Laboratory - ETHZ

SMARTER THAN US
THE RISE OF MACHINE INTELLIGENCE
STUART ARMSTRONG

Copyrighted Material
NEW AFTERWORD
NICK BOSTROM
SUPERINTELLIGENCE
Paths, Dangers, Strategies
'I highly recommend this book'
BILL GATES
NEW YORK TIMES BESTSELLER

Copyrighted Material
ARTIFICIAL INTELLIGENCE
AND THE END
OF THE HUMAN ERA
OUR FINAL
INVENTION
JAMES BARRAT
Copyrighted Material

AI is humanity's
biggest existential
threat

Copyrighted Material
THE
SENTIENT
MACHINE
THE COMING AGE OF
ARTIFICIAL INTELLIGENCE
AMIR HUSAIN
Copyrighted Material

CALUM CHACE
SECOND EDITION
SURVIVING
AI
The promise and peril
of artificial intelligence

Copyrighted Material
NEW YORK TIMES BESTSELLER
AI
SUPER-
POWERS
CHINA,
SILICON VALLEY,
AND THE
NEW WORLD ORDER
KAI-FU LEE
Copyrighted Material

AI is transformative, …
worst or best thing that
ever happened to our
civilizations -
we just don't know
GETTY IMAGES

...computers have made arithmetic cheap.

Solving complex equations is done more easily and in less time ..

- **What will AI technology make cheap ? Prediction.**

- **Prediction is central to decision-making under uncertainty**

- **Better prediction under uncertainty -> new opportunities for all companies**

COMPUTERS : A Disruptive Technology

https://www.testnewsonline.com/2012/10/02/deep-blue-beat-kasparov-because-of-bug/

What is Artificial Intelligence ?

THINK **LIKE HUMANS** - Bellman 1978

"The automation of activities that we associate with human thinking, activities such as decision-making, problem solving, learning …

THINK **RATIONALLY** - Winston 1992

"The study of the computations that make it possible to perceive, reason, and act"

ACT **LIKE HUMANS** - Rich and Knight, 1991

"The study of how to make computers do things at which, at the moment, people are better"

ACT **RATIONALLY** - Schalkoff, 1990

"A field of study that seeks to explain and emulate intelligent behavior in terms of computational processes"

A rationalist approach involves a combination of mathematics and engineering.

Artificial Intelligence, A Modern Approach, S. J. Russell and P. Norvig

People in each group sometimes cast aspersions on work done in the other groups, but the truth is that each direction has yielded valuable insights. Let us look at each in more detail.

What is Intelligence ?

John McCarthy: "*Intelligence is the computational part of the ability to achieve goals in the world*".

- A system having a goal or not, is not a property of the system itself. It is in the **relationship between the system and an observer.**

- The system is most usefully understood/predicted/controlled in terms **of its outcomes rather than its mechanisms.**

http://jmc.stanford.edu/artificial-intelligence/what-is-ai/index.html

TWO TYPES OF COMPUTING

I. Use the Computer to execute our instructions

II. Instruct the Computer to achieve our goals

- ***Dimensionality Reduction - Predict - Control - Optimize - Decide***

BUT OUR INSTRUCTIONS AND OUR GOALS ARE SUBJECT TO UNCERTAINTY IN OUR MODELS

THE IMPERFECT PATHS TO KNOWLEDGE

ADAPTED FROM : R. Moser, T. Oden, O. Ghattas, UT Austin

THE QUEST FOR KNOWLEDGE

- *All men by nature desire knowledge (Aristotle)*
- *Man has an intense desire for assured knowledge (Einstein)*
- *Knowledge is Power (Bacon)*

KNOWLEDGE:

- true, justified belief *(Plato)*.
- understanding, as opposed to opinion.
- quantifiable relationships between facts/observations and ideas.
  - "He believes it, but it isn't so," vs. "He knows it, but it isn't so." (Wittgenstein)

Immanuel Kant (1724 – 1804)

"Philosophy needs a science to determine the possibility, the principle, and the scope of our whole prior knowledge."

Our intellect does not draw its laws from nature, but tries – with varying degrees of success – to impose upon nature laws which it freely invents.

Pierre-Simon de Laplace (1749-1827)

*“**The theory of probabilities is at bottom nothing but common sense reduced to calculus;** it enables us to appreciate with exactness that which accurate minds feel with a sort of instinct for which ofttimes they are unable to account.”*

“Life’s most important questions are, for the most part, nothing but probability problems.”

“What we know is not much. What we do not know is immense.”

Hans Reichenbach (1891- 1953) The Rise of Scientific Philosophy, 1951.

The Path to Truth . . .

If error is corrected whenever it is recognized as such,
the path to error is the path of truth.

Sir Arthur Eddington (1882-1944)

“Experimentalists will be surprised to learn that we will not accept any experimental evidence that is not confirmed by theory.”

Werner Heissenberg (1901-1976)

“What we observe is not nature itself,
but nature exposed to our method of questioning.”

HUMAN QUEST FOR KNOWLEDGE

CREDIT: Vladimir Cherkassky

- Knowledge driven design of a series of experiments :
  - Account for all sources of uncertainty in the experimental campaign
  - Have a full description of the system state and its evolution.

The Classic view:

- Not sufficient just to detect patterns in events
- Describe events with **few fundamental, deterministic, causal principles.**

- Describes well “simple” deterministic systems.
  - “simple” refers to conceptual simplicity, rather than ‘technical’ system complexity.
  - e.g. a mechanical system with many moving objects interacting with each other; each object is described by simple equations that involve just a few variables.

Classical Approach to Human Knowledge

Classic Knowledge: Describe many facts with few fundamental principles.

- The number of facts (data,observations) used to derive such knowledge is usually **small**.
- The cost of collecting or generating these observations is **high.**
- The principles may **not be "useful"** (i.e. non-predictive).

Empirical Approach to Knowledge: PRINCIPLES - DATA- COMPLEXITY

Empirical knowledge: useful dependencies (i.e. predictive) estimated from data or derived from experience.

In contrast to first- principle knowledge, empirical knowledge:

- Describes properties of *"complex"* systems that lack credible first-principle models ('complexity' refers to a large number of observed parameters/variables);
- **is statistical**, i.e., allows to make non-deterministic predictions, at best;
- has a quantifiable practical utility for a given application (i.e. the "Oracle")
- **Highly facilitated by the digital processing and acquisition of knowledge**

DEFINITIONS

SYSTEM– The real world problem 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, 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: discretization of PDEs, numerical integration, truncation of infinite sums

STOCHASTIC SYSTEM THEORY

SYSTEM : any part of the world, natural or man-made, that we want to conceptually isolate to study. Inputs and outputs give its connections with the rest of the world
e.g. structural, mechanical, chemical, electrical, biological, economic, and geophysical systems

SYSTEM THEORY : goal is to provide a unified theoretical framework and computational tools to study systems Usually divided: system modeling, system analysis, system identification, system design optimization (including control systems)

STOCHASTIC SYSTEM THEORY : goal is to quantify the effects of both input and modeling uncertainty using probability, leading to both prior (initial) and posterior (updated using system data) stochastic predictions of the system output and performance. Use “stochastic” and “probabilistic” as synonymous.

REAL WORLD vs MODEL WORLD

REAL WORLD

MODEL WORLD

Model-based Decision Making under Uncertainty

Sources of Uncertainty I

• 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 II

- **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.

Why Uncertainty Quantification (UQ)? UQ for Decision Makin

- Example: hurricane forecasting

Nightly News | March 08, 2013

European weather forecasts superior to US models

The predictions from European computer models, which have 10 times the computing ability of the National Weather Service, have increasingly become more accurate than our models with the starkest example being Hurricane Sandy. NBC's Al Roker reports.

Why UQ ?

UQ for Validation

Example: measurements of the speed of light (1870-1960)

Image from Christie et al., Los Alamos Science, #29, 2005

Why UQ ?

UQ for Design/Optimization -

Example: transonic wing shape optimization

Image from T. Zang, 2003

- Choice of design conditions can dramatically affect performance
- Impact of flight conditions uncertainties can lead to unknown/unexpected consequences

Example: 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**
  - 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.

Stochastic system analysis: Predicting system performance under uncertainty

Prior analysis: During stochastic design, use probability models to predict system performance, treating uncertainties in input, system modeling and output

Posterior analysis: During operation, use sensor data to update these probability models and their performance predictions

Example: 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.

Large Scale MD Simulations of Water Transport in CNTs

L = 2μm & R ~ 2nm

$$Q_{HP} = \frac{\pi R^2}{8\mu}\frac{\Delta P}{L}$$

$$E = \frac{Q}{Q_{HP}}$$

Walther JH, Ritos K, Cruz-Chu E, Megaridis CM, Koumoutsakos P. , Barriers to Superfast Water Transport in Carbon Nanotube Membranes. **Nano Letters**, 2013, 13 (5)

Fast Mass Transport Through Sub-2 Nanometer Carbon Nanotubes
Holt, et al., Science 312, 1034 (2006)
Kim et al. Fabrication of flexible, aligned c arbon nanotube/ polymer composite membranes by in-situ polymerization.
J.of Membrane Science 2014;460:91-8.
Enhanced Flow in Carbon Nanotubes
Majumder , et al. Nature, 438, 2005
S.K. Youn, DIAMETER MODULATION AND INTEGRATION OF VERTICALLY ALIGNED SINGLE WALLED CARBON NANOTUBES FOR UNDERSTANDING OF MASS TRANSPORT IN CARBON NANOTUBES, PhD Thesis 2014
Measurement of the Rate of Water Translocation through Carbon Nanotubes
Xingcai Qin, et al., NanoLetters, 11, 2173, 2011
E
L / nm

Sources of Uncertainty in Water-Graphite Systems

MODELLING **PARAMETRIC** **COMPUTATIONAL** **MEASUREMENT**

$$\phi_{\mathrm{LJ}}(r_{ij}) = 4\epsilon_{\mathrm{LJ}}\left[\left(\frac{\sigma_{\mathrm{LJ}}}{r_{ij}}\right)^{12} - \left(\frac{\sigma_{\mathrm{LJ}}}{r_{ij}}\right)^{6}\right]$$

K. Osborne III (2009)

ennard-Jones potential : well depth and cut-off

BAYESIAN UNCERTAINTY FOR NANOSCALE FLOWS

Water Contact Angle

Parameters: ε, r_{cut}

MODEL CALIBRATION

Water Transport in CNTs

Friction coefficient and slip length of water inside CNTs

Example 2: Calibrating MD simulation for water

- **HETEROGENEOUS DATA**
  - *Diffusion coefficient*
  - *Density*
  - *Radial Distribution Function (RDF)*

Data sources: Holz et al. 2000; Jones & Harris 1992; Soper 2013

Bayesian Inference using Individual Data Set

Calibrate for each data set individually…

* Calibrate q and ε_{LJ} only, σ_{LJ} highly correlated with ε_{LJ}

Why blood flow simulations?

CANCER

8M Deaths/Year

90% due to Metastasis

TUMOR INDUCED ANGIOGENESIS

Why interested in blood?

CTC-chip

Isolation of rare circulating tumour cells in cancer patients by microchip technology, Nagrath et al., Nature, 2007

Funnel ratchets

Cell separation based on size and deformability using microfluidic funnel ratchets, McFaul et al., Lab on a Chip, 2012

HB-chip

Isolation of circulating tumor cells using a microvortex generating herringbone-chip, Scott et al.,Proc. Nat. Acad. Sci., 2014

CTC detection → High throughput - mL Samples

DPD FORCES : N-body Interactions + Stochastics

$$\mathbf{F}_i = \sum_{n=1, n\neq i}^{N} \mathbf{F}_{i,n}^{C,\mathrm{DPD}} + \mathbf{F}_{i,n}^{D,\mathrm{DPD}} + \mathbf{F}_{i,n}^{R,\mathrm{DPD}}$$

$$+ \sum_{k=1, k\neq i}^{K} \mathbf{F}_{i,k}^{C,\mathrm{FSI}} + \mathbf{F}_{i,k}^{D,\mathrm{FSI}} + \mathbf{F}_{i,k}^{R,\mathrm{FSI}}$$

$$+ \sum_{m=1, m\neq i}^{M} \mathbf{F}_{i,m}^{C,\mathrm{wall}} + \mathbf{F}_{i,m}^{D,\mathrm{wall}} + \mathbf{F}_{i,m}^{R,\mathrm{wall}}$$

$$\mathbf{F}_{ij}^{C} = \begin{cases} a_{ij}(1 - r_{ij})\mathbf{e}_{ij}, & \text{if } r_{ij} < 1 \\ 0, & \text{if } r_{ij} \geq 1 \end{cases}$$

$$\mathbf{F}_{ij}^{D} = -\gamma w^{D}(r_{ij})(\mathbf{e}_{ij} \cdot \mathbf{v}_{ij})\mathbf{e}_{ij}$$

$$\mathbf{F}_{ij}^{R} = \sigma w^{R}(r_{ij})\theta_{ij}\mathbf{e}_{ij}$$

$$\mathbf{r}_{ij} = \mathbf{r}_i - \mathbf{r}_j$$

$$r_{ij} = ||\mathbf{r}_{ij}||$$

$$\mathbf{e}_{ij} = \mathbf{r}_{ij}/||\mathbf{r}_{ij}||$$

$$\mathbf{v}_{ij} = \mathbf{v}_i - \mathbf{v}_j$$

$$w^{D}(r) = (w^{R}(r))^2$$

$$\sigma^2 = 2\gamma k_B T$$

The Red Blood Cell (RBC) model

$$V(\{\mathbf{x}_i\}) = V_{in-plane} + V_{bending} + V_{area} + V_{volume}$$

$$V_{in-plane} = \sum_{j \in 1 \ldots N_s} \left[\frac{k_B T l_{max} \left(3x_j^2 - 2x_j^3\right)}{4p(1-x_j)} + \frac{k_p}{(n-1) l_j^{n-1}} \right]$$

$$V_{bending} = \sum_{j \in 1 \ldots N_s} k_b \left[1 - cos(\theta_j - \theta_0)\right]$$

$$V_{area} = \frac{k_a (A^{tot} - A_0^{tot})^2}{2A_0^{tot}} + \sum_{j \in 1 \ldots N_t} \frac{k_d (A_j - A_0)^2}{2A_0}$$

$$V_{volume} = \frac{k_v (V^{tot} - V_0^{tot})^2}{2V_0^{tot}}$$

Mechanical Properties

membrane shear modulus

$$\mu_0 = \frac{\sqrt{3} k_B T}{4 p l_{max} x_0} \left(\frac{x_0}{2(1-x_0)^3} - \frac{1}{4(1-x_0)^2} + \frac{1}{4} \right) + \frac{\sqrt{3} k_p (n+1)}{4 l_0^{n+1}}$$

where: $x_0 = \frac{l_0}{l_{max}}$

area compression modulus

$$K = 2\mu_0 + k_a + k_d$$

membrane bending rigidity

$$k_b = \frac{2k_c}{\sqrt{3}}$$

Young's modulus

$$Y = \frac{4K\mu_0}{K + \mu_0}$$

membrane shear viscosity

$$\eta_m = \sqrt{3}\gamma^T + \frac{\sqrt{3}\gamma^C}{4}$$

DPD FORCES: Red Blood Cells

$$
\begin{aligned}
\mathbf{F}^{\text{cell}} &= \sum_{n=1}^{N} \boldsymbol{F}^{\text{dihedral},1}_{0,n-1,n,n+1} + \boldsymbol{F}^{\text{dihedral},2}_{0,n,N+n,n+1} + \boldsymbol{F}^{\text{triangle}}_{0,n,n+1} + \boldsymbol{F}^{\text{bond}}_{0,n} \\
&= \sum_{n=1}^{N} \beta^b_{n,n+1} \left[\frac{\boldsymbol{\xi}_n \times \boldsymbol{a}_{n+1} + \boldsymbol{\xi}_{n+1} \times \boldsymbol{a}_n}{\xi_n \xi_{n+1}} - \cos\theta_{n,n+1} \left(\frac{\boldsymbol{\xi}_n \times \boldsymbol{a}_n}{\xi_n^2} + \frac{\boldsymbol{\xi}_{n+1} \times \boldsymbol{a}_{n+1}}{\xi_{n+1}^2} \right) \right] \\
&+ \beta^b_{n,N+n} \left[\frac{\boldsymbol{\xi}_{N+n} \times \boldsymbol{a}_n}{\xi_n \xi_{N+n}} - \cos\theta_{n,N+n} \frac{\boldsymbol{\xi}_n \times \boldsymbol{a}_n}{\xi_n^2} \right] \\
&+ \left(\frac{qC_q}{A_n^{q+1}} - k_a \frac{A - A_0^{tot}}{A_0^{tot}} \right) \frac{\boldsymbol{\xi}_n \times \boldsymbol{a}_n}{4A_n} \\
&- \frac{k_v}{18} \frac{V - V_0^{tot}}{V_0^{tot}} (\boldsymbol{\xi}_n + (\boldsymbol{r}_0 + \boldsymbol{r}_n + \boldsymbol{r}_{n+1}) \times \boldsymbol{a}_n) \\
&- \frac{k_B T}{p} \left(\frac{1}{4(1 - b_n/l_m)^2} - \frac{1}{4} + \frac{b_n}{l_m} \right) \frac{\boldsymbol{b}_n}{b_n} \\
&+ \sqrt{2k_B T} \left(\sqrt{2\gamma^T} d\overline{\boldsymbol{W}^S_{ij}} + \sqrt{3\gamma^C - \gamma^T} \frac{tr[d\boldsymbol{W}_{ij}]}{3} \boldsymbol{I} \right)
\end{aligned}
$$

$$x_0 = l_0/l_m$$

$$A_0 = \sqrt{3} l_0^2/4$$

$$C_q = \frac{\sqrt{3} A_0^{q+1} k_B T (4x_0^2 - 9x_0 + 6)}{4pql_m(1 - x_0)^2}$$

$$\beta^b_{ij} = k_b \frac{\sin\theta_{ij}\cos\theta_0 - \cos\theta_{ij}\sin\theta_0}{\sqrt{1 - \cos^2\theta_{ij}}}$$

$$d\overline{\boldsymbol{W}^S_{ij}} = d\boldsymbol{W}^S_{ij} - tr[d\boldsymbol{W}^S_{ij}]\mathbf{1}/3$$

VALIDATION

(Karniadakis Group)

Microfluidics:
High-Throughput Separation of CTCs

The CTC-iChip

[Karabacak et al - Nature Protocols - 2014]

THE CTC-ICHIP

The in-silico lab-on-a-chip

μ-Fluidics for CTCs: in silico and in vitro

Simulation of exact replica of the Pachinko device: Toner, Truskey, Kapur, U.S. Patent No. 8,304,230, 2012

- Rise of RBCs in higher rows of device
- Increase of separation distance

Discrepancy with experiment:

- asymmetry in RBC distribution between posts
- underestimation of RBC mass flux

Model validation (in-house)

Solvent: Dissipative Particle Dynamics

Hiemenz flow: Velocity profiles near the stagnation point

Taylor-Couette flow

Oscillatory Stokes flow

Stokes flow around a cylinder

Flow around a cylinder (Re=40)

RBC membrane: DPD-RBC model (Pivkin 2008)

RBC in shear flow: Angle of rotation

RBC in Taylor-Couette flow: Comparison with BEM results

DPD validation: Different experiments - different model parameters

TEST CASE	Solvent Model / RBCmodel	Membrane rigidity	Membrane viscosity	Maximum spring extension	Persistence length	viscosity-contrast
stretching (Fedosov et al., 2010)	DPD / stress-free	4.8E-19 [J]	0.022 [Pa.s]	1.23 e-6 [m]	1.99 e-9 [m]	1
squeeze in micro-channel (Bow et al., 2011)	DPD / constant spring eq. length	7.5 E-19 [J]	varied to study its effect	3.17 e-7 [m]	1.08 e-7 [m]	1
shear of whole blood (Fedosov et al., 2011)	DPD / stress-free	3.0 E-19 [J]	0.0144 [Pa.s]	1.3e-7 [m]	1.99 e-9 [m]	1
DLD device (Henry et al., 2016)	SDPD+a / stress-free	4.8 E-19 [J]	0.022 [Pa.s]	1.23 e-6 [m]	1.99 e-9 [m]	5

Importance of viscosity ratio, $R=\eta_{in}/\eta_{out}$ for Stokes Flow (Re=0.1)

Deviation of trajectory
from front/back
symmetry around the cylinder

Confirmed by other authors

Figure 6. Stroboscopic images of RBCs in section 2, taken from simulations and experiments. (a) RBC lane swapping is promoted by tumbling when $C = \eta_i/\eta_o = 5$. **(b)** Tank-treading type dynamics occurs at $C = 1$ and the RBC favors the displacement mode.

Fedodov's Group: Henry E. et al., Scientific Reports, 2016.

THE PREDICTION GAME: MODELS & DATA

- Models are imperfect representations of reality
- Computational models involve parameters

- Probability as the Logic of Science

- How to choose parameters and how much to trust them?
- How to integrate DATA ?

Bayesian Uncertainty Quantification: Models and Data

"**Theories** have to be judged in terms of their **probabilities** in light of the **evidence**."

$$P(A|B)P(B) = P(B|A)P(A)$$

$$\boxed{\overset{\textbf{posterior}}{P(A|B)} = \frac{\overset{\textit{likelihood}}{P(B|A)}\overset{\textit{prior}}{P(A)}}{P(B)}}$$

$A \rightarrow$ Hypothesis/Model

$B \rightarrow$ DATA

Probability and 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.

Probability and Uncertainty Quantification

- 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.

Interpretation of probability

- The axioms of probability are well-established but after three centuries, there is still disagreement about the meaning of probability

- The interpretation of probability is important in applications to real systems and phenomenon – it governs:
  - perceived domain of its applicability; e.g. is the probability of a model meaningful?
  - understanding of the results of uncertainty analysis;
    - e.g. what does the failure probability mean? (Is it an inherent property of the system OR a property of what we know about the system and its future excitation?)

Example of PDFs

Normal (Gaussian) Distribution

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

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

Uniform Distribution

$$X \sim \mathcal{U}(a, b)$$

Two prevailing interpretations of probability: FREQUENTIST & BAYESIAN

- **FREQUENTIST: Probability of an "inherently random"event is the relative frequency of occurrence of the event in the "long run":**

- Probability distributions are inherent properties of "random" phenomena

- Limited scope, e.g. no meaning for the probability of a model

- "Inherent randomness" is assumed but cannot be proved

- The definition is not an operational one, it is impossible to establish directly a complex probability distribution by applying the definition

- **BAYESIAN: Probability is a measure of the plausibility of a statement conditional on specified information:**

- Probability distributions represent states of plausible knowledge about systems and phenomena, not inherent properties of them

- Probability of a model is a measure of its plausibility relative to other models in a set

- Pragmatically quantifies uncertainty due to missing information without any claim that this is due to nature's "inherent randomness"

- Need procedures for constructing initial probability distributions

Frequentist vs Bayesian

- **Is it "inherent randomness " or does it just "look random"?**

(a) Data-stream from a random number generator "looks random" but it is deterministic if the algorithm and initial condition ("seed") are known;

(b) The outcome of coin tosses "looks random" but it is a deterministic mechanics problem if the initial conditions are known

E.T. Jaynes' answer (2003): (a) is an example of the Mind-Projection Fallacy: **Our models of reality are confused with reality,**

or its more specific application here: Our uncertainty is projected onto nature as its inherent property

Frequentist vs Bayesian: SUMMARY

- Frequentist assumes that there are inherently random events that are controlled by (unknown) probability distributions.
- Bayesian focuses on quantifying uncertainty about propositions due to incomplete information using appropriate probability models.

- For further discussion:
  - D.S. Sivia & J. Skilling, 2005 [Chapter 1 - 3)
  - E.T. Jaynes 1984 [Link at Course Webpage]
  - T.J. Loredo 1990 [Link at Course Webpage: see Sections 1,2]

Bayesian Uncertainty Quantification: Models and Data

"**Theories** have to be judged in terms of their **probabilities** in light of the **evidence**."

$$P(A|B)P(B) = P(B|A)P(A)$$

$$\boxed{\overset{\textbf{posterior}}{P(A|B)} = \frac{\overset{\textit{likelihood}}{P(B|A)}\overset{\textit{prior}}{P(A)}}{P(B)}}$$

$A \rightarrow$ Hypothesis/Model

$B \rightarrow$ DATA

Bayesian Uncertainty Quantification: Calibration and model selection

PARAMETER ESTIMATION

$$p(\theta_i \mid D, M_i) = \frac{p(D \mid \theta_i, M_i)\, p(\theta \mid M_i)}{p(D \mid M_i)}$$

Experiments

Physical limitations

Past studies

MODEL CLASS SELECTION

$$p(M_i \mid D) = \frac{p(D \mid M_i)\, p(M_i)}{p(D)}$$

Evidence of a Model

$$p(D \mid M_i) = \int p(D \mid \theta_i, M_i)\, p(\theta_i \mid M_i)\, d\theta_i$$

- Bayesian inference : **large numbers of model evaluations**
- Each simulation : **computationally intensive**

Bayesian UQ: Propagation

$\theta \in \mathcal{R}^n$ → **M** (Model class) → QoI $p(\theta \mid M)$ → $p(q \mid \theta, M)$ **Conditional PDF**

$p(q \mid \theta, M)$

QUANTITIES OF INTEREST: **Posterior Robust Predictions: PDF**

$$p(q \mid D, M) = \int \underbrace{p(q \mid \theta, M)}_{\text{Conditional PDF}} \underbrace{p(\theta \mid D, M)}_{\text{Posterior PDF}} d\theta$$

SAMPLING HIGH DIMENSIONAL INTEGRALS

Sample Estimate:

$$p(q \mid D, M) = \frac{1}{N} \sum_{i=1}^{N} \underbrace{p(q \mid \theta^{(i)}, M)}_{\text{Conditional PDF}}$$

$$\theta^{(i)} \sim p(\theta \mid D, M)$$

Samples drawn from Posterior PDF

Bayes' theorem

ROADS TO PREDICTION

MACHINE LEARNING	BAYESIAN INFERENCE
• computationally inexpensive,	• computationally more expensive
• usually cannot identify all parameters	• distribution of all parameters
• lack possibilities to quantify uncertainty	• can quantify uncertainty
• can result in ill-posed problems	• well-posed problems
• does not incorporate prior knowledge about parameters	• incorporates prior knowledge
	• requires prior knowledge even when it's not accessible

Probability as a logic - > Bayesian probability

<u>KEY IDEA:</u>

Probability P[b|a] = measure of plausibility of statement *b* based on the information in statement a

Probability as a logic: Rigorous foundation for Bayesian probability

- Extends binary Boolean logic to a multi-valued logic for <u>quantification of plausible reasoning</u> under <u>incomplete information</u>
- Boolean propositional logic deals with the special case of <u>complete information</u> where the truth or falsity of b is known from a

$a \Rightarrow b$ or $a \Rightarrow \sim b$ corresponding to $P(b\,|\,a) = 1$ or $P(b\,|\,a) = 0$

Seminal work on foundations by R.T. Cox: "Probability, Frequency and Reasonable Expectation", Amer. J. Physics 1946 The Algebra of Probable Inference, Johns Hopkins Press 1961

Treatise on theory and applications by E.T. Jaynes: Probability Theory – The Logic of Science, Cambridge U. Press 2003 [Link at Course Webpage has first three chapters]

Probability Fundamentals : Axioms and Properties

- By extending Boolean logic to incomplete information, R.T. Cox derived a minimal set of axioms for probability logic (1946, 1961)
  For any statements a, b, c :

$$P1 : P[b\,|\,a] \geq 0$$ [By convention]

$$P2 : P[\sim b\,|\,a] = 1 - P[b\,|\,a]$$ [Negation Function]

$$P3 : P[c \,\&\, b\,|\,a] = P[c\,|\,b \,\&\, a]\,P[b\,|\,a]$$ [Conjuction Function]

- These axioms and De Morgans' Law of Logic imply <u>Disjunction Function:</u>

$$P[c \text{ or } b\,|\,a] = P[c\,|\,a] + P[b\,|\,a] - P[c \,\&\, b\,|\,a]$$

- They imply <u>Kolmogorov's statement of probability axioms</u> (1933) for probability measure P(A), which has no built-in meaning in the axioms

Probability Fundamentals : Axioms and Properties

Let a, b, c be propositions. Also define:

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

The **<u>axioms</u>** of probability logic are stated as

$$P(b\,|\,a) \geq 0$$
$$P(b\,|\,a, b) = 1$$
$$P(b\,|\,a) + P(\sim b\,|\,a) = 1$$
$$P(c, b\,|\,a) = P(c\,|\,b, a)\,P(b, a)$$

<u>Properties</u>

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

$P(c \text{ or } b\,|\,a) = P(c\,|\,a) + P(b\,|\,a) - P(c, b\,|\,a)$ if b and c cannot both be true (muttualy exclusive) conditioned on a

Probability Logic Fundamentals

<u>Properties</u>

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

<u>Marginalization Theorem:</u>

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

<u>Total Probability Theorem:</u>

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

<u>Bayes Theorem:</u>

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

Discrete Variables

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

$$p(x_i \,|\, a) = P(X = x_i \,|\, a)$$

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 $X = x_i, i = 1, \ldots, n$ 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.

<u>Marginalization Theorem:</u>

$$P(x_i \,|\, a) = \sum_{k=1}^{n} P(x_i, y_k \,|\, a)$$

<u>Total Probability Theorem:</u>

$$P(x_i \,|\, a) = \sum_{k=1}^{n} P(x_i \,|\, y_k, a)\, P(y_k \,|\, a)$$

<u>Bayes Theorem:</u>

$$P(y_k \,|\, x_i, a) = \frac{P(x_i \,|\, y_k, a)\, P(y_k \,|\, a)}{\sum_{\ell=1}^{n} P(x_i \,|\, y_\ell, a)\, P(y_\ell \,|\, a)}, \qquad k = 1, \ldots, n$$

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

$$P(X \leq x \,|\, a) \equiv F(X \,|\, a)$$

is used to denote the probability of the proposition $X \leq x$, i.e. the variable X to take values 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 $f(x)$ from the expression

$$P(x < X \leq x + dx \,|\, a) = f(x \,|\, a)dx$$

It can be readily derived that

$$f(x \,|\, a) = \frac{dF(x \,|\, a)}{dx}$$

using the fact the statement $X \leq x + dx$ is the sum of the statement $X \leq x \,|\, a$ and $x < X \leq x + dx \,|\, a$ and that these statements are mutually exclusive so that using the sum rule

$$P(X \leq x + dx \,|\, a) = P(X \leq x \text{ or } x < X < x + dx \,|\, a) = P(X \leq x \,|\, a) + P(x < X \leq x + dx \,|\, a)$$

$$\Rightarrow P(x < X \leq x + dx) = P(X \leq x + dx \,|\, a) - P(X \leq x \,|\, a) = F(x + dx \,|\, a) - F(x \,|\, a)$$

$$\Rightarrow F(x + dx \,|\, a) - F(x \,|\, a) = f(x \,|\, a)dx$$

Continuous Variables

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

<u>Marginalization Theorem:</u>

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

<u>Total Probability Theorem:</u>

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

<u>Bayes Theorem:</u>

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

Models and Data

Bayes/Laplace+ 300 Years of Advances in Computers and Data

$A \rightarrow$ Hypothesis/Model

$B \rightarrow$ DATA

"**Theories** must be judged in terms of their **probabilities** in light of the **evidence**."

$$\underbrace{P(A|B)}_{\text{posterior}} = \frac{\overbrace{P(B|A)}^{\text{likelihood}}\overbrace{P(A)}^{\text{prior}}}{P(B)}$$

Example of PDFs

Normal (Gaussian) Distribution

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

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

Uniform Distribution

$$X \sim \mathcal{U}(a, b)$$

Bayesian UQ: Estimation

Physical limitations
Past studies
Expert elicitation

Data

$$p(\theta_i \,|\, D, M_i) = \frac{p(D \,|\, \theta_i, M_i)\, p(\theta_i \,|\, M_i)}{f(D \,|\, M_i)}$$

PARAMETER ESTIMATION

Bayesian UQ: Calibration and model selection

Result: a ***distribution*** over all possible parameters of the model

- **How to compare the output of the model with the data?**

Classical approach:

prediction error equation

$$D = F(\theta) + \epsilon$$
$$\epsilon \sim \mathcal{N}(0,\Sigma)$$

likelihood $p(D\,|\,\theta)$

Bayes' Theorem

posterior $$p(\theta\,|\,D) = \frac{p(D\,|\,\theta)p(\theta)}{p(D)}$$

Bayesian UQ: Calibration and model selection

Experimental Data: D

Use observations to select the model classes and estimate their parameter values such that the model predictions best fit the data

PARAMETER ESTIMATION

$$p(\theta_i \mid D, M_i) = \frac{p(D \mid \theta_i, M_i)\, p(\theta \mid M_i)}{p(D \mid M_i)}$$

Experiments

Physical limitations
Past studies
Expert elicitation

MODEL CLASS SELECTION

$$p(M_i \mid D) = \frac{p(D \mid M_i)\, p(M_i)}{p(D)}$$

Evidence of a Model

$$p(D \mid M_i) = \int p(D \mid \theta_i, M_i)\, p(\theta_i \mid M_i)\, d\theta_i$$

- Bayesian inference requires **large numbers of model evaluations**
- Each simulation run is usually **computationally intensive**

Bayesian UQ: Propagation

$\theta \in \mathcal{R}^n$ → **M** (Model class) → QoI $p(\theta \mid M)$ → $p(q \mid \theta, M)$ **Conditional PDF**

$p(q \mid \theta, M)$

QUANTITIES OF INTEREST: **Posterior Robust Predictions: PDF**

$$p(q \mid D, M) = \int \underbrace{p(q \mid \theta, M)}_{\text{Conditional PDF}} \underbrace{p(\theta \mid D, M)}_{\text{Posterior PDF}} d\theta$$

SAMPLING HIGH DIMENSIONAL INTEGRALS

Sample Estimate:

$$p(q \mid D, M) = \frac{1}{N} \sum_{i=1}^{N} \underbrace{p(q \mid \theta^{(i)}, M)}_{\text{Conditional PDF}}$$

$$\theta^{(i)} \sim p(\theta \mid D, M)$$

Samples drawn from Posterior PDF

Bayesian UQ: Model Selection

Which model is better given the data?

Probability of

a model

given the data

Model Model

Evidence Prior

$$p(M_i \mid D) = \frac{p(D \mid M_i)p(M_i)}{p(D)}$$

Model

Evidence

$$p(D \mid M_n) = \int_{\Omega_i} \dots \int_{\Omega_n} p(D \mid \theta_1, \dots, \theta_n) p(\theta_1, \dots, \theta_n) d\theta_1 \dots \theta_n$$

Comparing two models - a 1-parameter and a 2-parameter model

Occam's razor

$$m_{12} = \frac{f(D \mid M_1)}{p(D \mid M_2)} \frac{p(M_1)}{p(M_2)} = \frac{\int_{\Omega_1} p(D \mid \theta_1) p(\theta_1 \mid M_1) d\theta_1}{\int_{\Omega_1} \int_{\Omega_2} p(D \mid \theta_1, \theta_2) p(\theta_1, \theta_2 \mid M_2) d\theta_1 d\theta_2} \frac{p(M_1)}{p(M_2)}$$

Prior and Evidence

How to select a prior?

Maximum entropy principle:

select a probability distribution with maximum entropy (the least informative one) given the prior information about the model parameters

Examples:

- *Bounds of the parameters are known => uniform prior*
- *Mean is known => exponential distribution*
- *Mean and variance of the parameters are known => normal prior*

How to handle the evidence?

Markov chain Monte Carlo method:

don't compute the evidence! Obtain samples from the posterior distribution and use kernel density estimates to reconstruct the underlying distribution

Example:

General rule: accept with probability

$$\zeta = min\left(1, \frac{p(\theta')}{p(\theta)}\right)$$

Markov Chain Monte Carlo

Posterior

Unidentifiable manifolds?

Annealing:

sample intermediate distributions from prior - posterior

Local covariance:

each Markov chain has its own proposal covariance,

Multi-modal?

Many Markov chains:

population-based sampling method

Our Work

Example:

- *sampling intermediate distribution*
- *each of the red points starts a Markov chain*
- *each of the Markov chains has its own covariance*

TMCMC: Transitional Markov Chain Monte Carlo

- target distribution: p
- annealed distribution

$$p^{\gamma}, \gamma \in [0,1]$$

- global exploration
- parallel execution
- estimator of evidence

HPC Challenges

- Bayesian inference needs **large numbers of model evaluations**
- Each simulation run is usually **computationally intensive**
- How to efficiently exploit HPC architectures for UQ:
  - algorithmic level of parallelism
  - simulation level of parallelism

Korali: Uncertainty Quantification Library

- Open-source library distributed under LGPL licence
- Available at http://www.cse-lab.ethz.ch/software/
- Algorithms:
  - ***TMCMC*** (for exact Bayesian inference)
  - ***ABC-SubSim*** (for approximate Bayesian inference)
  - ***CMA-ES*** (for optimisation)
  - ***Subset Simulation*** (for rare events sampling)
  - ***A-PNDL*** (for adaptive parallel numerical differentiation)

✓Platform agnostic task-based parallelism

✓Multi-level parallelism

✓Transparent load balancing

Korali: Applications

Hierarchical Bayesian Model: Future predictions in Pharmacodynamics

multiple patient data

+

model

$$\frac{dC}{dt} = -\vartheta_1 C$$

$$\frac{dP}{dt} = \vartheta_4 P(1 - \frac{P^*}{K}) + \vartheta_5 Q_P - \vartheta_3 P - \vartheta_1 \vartheta_2 CP$$

$$\frac{dQ}{dt} = \vartheta_3 P + \vartheta_1 \vartheta_2 CQ$$

$$\frac{dQ_P}{dt} = \vartheta_1 \vartheta_2 CQ - \vartheta_5 Q_P - \vartheta_6 Q_P$$

$$C(0) = 0,\ P(0) = \vartheta_7,\ Q(0) = \vartheta_8,\ Q_P(0) = 0$$

+

hierarchical Bayesian

prediction per patient

Hierarchical Bayesian Model: reconsidering the Lenard-Jones potential: calibrating the exponent

RDF from multiple thermodynamic conditions

+

model

$$M\ddot{X} = -\nabla V(X)$$

$$V(r) = 4\varepsilon\left[\left(\frac{\sigma}{r}\right)^p - \left(\frac{\sigma}{r}\right)^6\right]$$

+

hierarchical Bayesian

exponent value per data set

Bayesian Model: Granular materials: DEM model selection

Coefficient of restitution of steel beads

+

model

+

Bayesian Inference

Most probable model given the data

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

$$p(x) = \mathcal{N}(x \,|\, \mu, \sigma^2)$$

$$E \sim \mathcal{N}(0,1)$$

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

$$p(y\,|\,x) = \mathcal{N}(y\,|\,x,1)$$

output QoI

Uncertainty Propagation (Prior Analysis)

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

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

The probability distribution of the output QoI Y is given by

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

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 prior system analysis

- 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, etc.

Uncertainty Propagation (Posterior Analysis)

Given the data D, the posterior PDF of the parameter x is obtained from Bayes' theorem as

$$p(x\,|\,D) = \frac{p(y = D\,|\,x)\,p(x)}{p(y = D)}$$

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

$$p(y\,|\,x) = \mathcal{N}(y\,|\,x, 1)$$

which for $y = D$ gives $p(D\,|\,x) = \dfrac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(D-x)^2}$

Substituting in the first line along with the prior PDF $p(x) = \mathcal{N}(x\,|\,\mu, \sigma^2)$

one readily obtains that $p(x\,|\,D) \propto e^{-\frac{1}{2}(D-x)^2 - \frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}}$

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

$$p(x\,|\,D) = \mathcal{N}\left(x\,\middle|\,\frac{1^2\mu + \sigma^2 D}{1^2 + \sigma^2}, \frac{1^2\sigma^2}{1^2 + \sigma^2}\right)$$

Uncertainty Quantification (Posterior Analysis)

prior PDF $p(x) = \mathcal{N}(x \,|\, \mu, \sigma^2)$

Uncertainty Quantification (Posterior Analysis)

prior PDF $p(x) = \mathcal{N}(x \,|\, \mu, \sigma^2)$

posterior PDF $p(x \,|\, D) = \mathcal{N}\left(x \,\middle|\, \frac{1^2\mu + \sigma^2 D}{1^2 + \sigma^2}, \frac{1^2\sigma^2}{1^2 + \sigma^2} \right)$

Uncertainty Quantification (Posterior Analysis)

The **prior robust prediction** for the QoI Y is $p(y) = \mathcal{N}(y \,|\, \mu, \sigma^2 + 1^2)$

Uncertainty Quantification (Posterior Analysis)

The **prior robust prediction** for the QoI Y is $f(y) = \mathcal{N}(y \,|\, \mu, \sigma^2 + 1^2)$

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**

$$p(y \,|\, D) = \mathcal{N}\left(y \,\middle|\, \frac{1^2\mu + \sigma^2 D}{1^2 + \sigma^2}, \frac{1^2\sigma^2}{1^2 + \sigma^2} \right)$$

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):

$$Y_k = f(X, u_k) + E_k$$

X and E are independent

$p(x) = \mathcal{N}(x \,|\, \mu, \sigma^2)$

$E_k \sim \mathcal{N}(0,1)$ independent and identically distributed (i.i.d)

$D = (d_1, \ldots, d_N)$ **Observations** (data) from the system

Uncertainty Quantification (Posterior Analysis)

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

$$p(x\,|\,D) \approx \prod_{i=1}^{N} \exp\left(-\frac{1}{2}(d_i - f(x, u_i))^2\right) \exp\left(-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}\right)$$

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

$$p(y\,|\,D) = \int p(y\,|\,x)\,p(x\,|\,D)\,dx$$

Posterior PDF

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

Estimation of Likelihood: To estimate the likelihood $p(D \mid x, \sigma^2)$, one can use the fact that the data are independent and apply successively the product rule of the axioms of probability, given by

$$p(b, a) = p(b \mid a) p(a)$$

To finally derive that

$$\begin{aligned} p(D \mid x, \sigma^2) &= \prod_{k=1}^{N} p(d_k \mid x, \sigma^2, I) \\ &= \prod_{k=1}^{N} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(d_k - f(x, u_k))^2} \\ &= (2\pi\sigma^2)^{-N/2} e^{-\frac{1}{2\sigma^2}\sum_{k=1}^{N}(d_k - f(x, u_k))^2} \end{aligned}$$

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
- 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.
- …