Mathematical Foundations of Bayesian Inverse Problems

Well-posedness and Statistical Estimates in Bayesian Inverse Problems

Claudia Schillings, Aretha Teckentrup

LMS short course: Introduction to the Bayesian approach to inverse problems

Outline

What are the Challenges in UQ?

- **What is uncertainty quantification (UQ) about?**

- **What is uncertainty?**
- **How can we model the uncertainty in the system?**
- **Which system quantities are uncertain?**
- **What are the effects of the uncertain input quantities on the solution?**
- **How can we efficiently solve the resulting (high dimensional) problems?**
- **What are the challenges?**

What is Uncertainty Quantication (UQ)?

Many aspects of modern life involve uncertainty:

Environmental systems

weather, climate, seismic, subsurface geophysics

Engineering systems

Biological systems

Physical systems

Waste Isolation Pilot Plant (WIPP)

- US DOE repository for radioactive waste situated in New Mexico
- Large amount of publicly available data

What is Uncertainty Quantication (UQ)?

Many aspects of modern life involve uncertainty:

Environmental systems

Engineering systems

automobiles, aircraft, bridges, structures

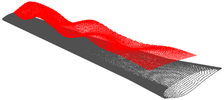
Biological systems

Physical systems

Aerodynamic Shape Optimization

- Optimal shape w.r. to drag
- Influence of manufacturing tolerances
- Reliability, robust behavior

Copyright DLR.



What is Uncertainty Quantication (UQ)?

Many aspects of modern life involve uncertainty:

Environmental systems

Engineering systems

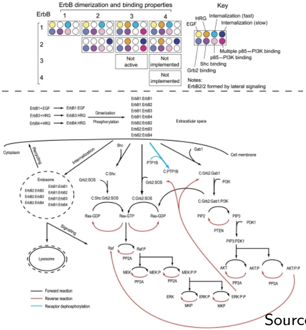
Biological systems

health and medicine, pharmaceuticals, gene expression, cancer research

Physical systems

ErbB signaling pathways

- Regulation of diverse physiological responses
- Mass action model



Source: Chen et al.

What is Uncertainty Quantication (UQ)?

Many aspects of modern life involve uncertainty:

Environmental systems

Engineering systems

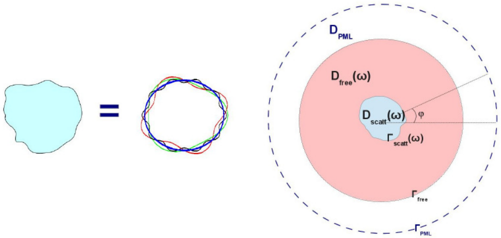
Biological systems

Physical systems

nano-optics, quantum physics, radioactive decay

UQ in Nano-Optics

- Quantification of the influence of defects in fabrication process on the optical response of nano structures
- Stochastic shape of the scatterer data



Quantification and Minimization of Uncertainties

Uncertainty quantification in numerical simulations

Minimization of uncertainties in design and control problems

Identification of uncertain parameters from noisy observations

Numerical analysis

Methods for optimal control / optimization problems

Approximation theory

Stochastics

Statistics

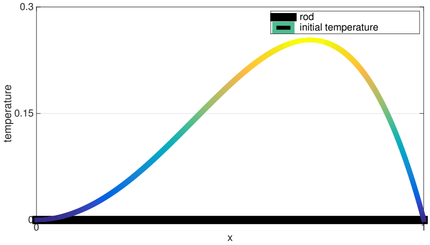
Algorithms and data structures

Introductory Example

We consider a rod of unit length and unit thermal conductivity.
The temperature distribution $u(x,t)$ satisfies the heat equation

$$\frac{\partial^2 u}{\partial x^2} - \frac{\partial u}{\partial t} = 0, \qquad 0 < x < 1,\ t > 0$$

with b.c. $u(0,t) = u(1,t) = 0$ and i.c. $u(x,0) = u_0(x)$.



Introductory Example

We consider a rod of unit length and unit thermal conductivity.
The temperature distribution $u(x,t)$ satisfies the heat equation

$$\frac{\partial^2 u}{\partial x^2} - \frac{\partial u}{\partial t} = 0, \qquad 0 < x < 1,\ t > 0$$

with b.c. $u(0,t) = u(1,t) = 0$ and i.c. $u(x,0) = u_0(x)$.

Given the temperature distribution at time $T > 0$, what is the initial temperature distribution?

Introductory Example

We can express the solution in terms of its Fourier components

$$u(x,t) = \sum_{n=1}^{\infty} c_n e^{-(n\pi)^2 t} \sin(n\pi x) \; ,$$

where c_n are the Fourier sine coefficients of the initial state $u_0(x) = \sum_{n=1}^{\infty} c_n \sin(n\pi x)$.

Introductory Example

We can express the solution in terms of its Fourier components

$$u(x,t) = \sum_{n=1}^{\infty} c_n e^{-(n\pi)^2 t} \sin(n\pi x) \ ,$$

where c_n are the Fourier sine coefficients of the initial state $u_0(x) = \sum_{n=1}^{\infty} c_n \sin(n\pi x)$.

Determine $\mathbf{c_n}$ from the final data in order to determine $\mathbf{u_0}$.

Introductory Example

We can express the solution in terms of its Fourier components

$$u(x,t) = \sum_{n=1}^{\infty} c_n e^{-(n\pi)^2 t} \sin(n\pi x) \ ,$$

where c_n are the Fourier sine coefficients of the initial state $u_0(x) = \sum_{n=1}^{\infty} c_n \sin(n\pi x)$.

Assume that we have two initial states, $u_0^{(1)},\ u_0^{(2)}$, differing only by a single high frequency component

$$u_0^{(1)}(x) - u_0^{(2)}(x) = c_N \sin(N\pi x) \qquad \text{for large } N \ .$$

The solutions at time T will differ by

$$u^{(1)}(x,T) - u^{(2)}(x,T) = c_N e^{-(N\pi)^2 T} \sin(N\pi x) \ .$$

Introductory Example

We can express the solution in terms of its Fourier components

$$u(x,t) = \sum_{n=1}^{\infty} c_n e^{-(n\pi)^2 t} \sin(n\pi x) \ ,$$

where c_n are the Fourier sine coefficients of the initial state $u_0(x) = \sum_{n=1}^{\infty} c_n \sin(n\pi x)$.

Assume that we have two initial states, $u_0^{(1)},\ u_0^{(2)}$, differing only by a single high frequency component

$$u_0^{(1)}(x) - u_0^{(2)}(x) = c_N \sin(N\pi x) \qquad \text{for large } N \ .$$

The solutions at time T will differ by

$$u^{(1)}(x,T) - u^{(2)}(x,T) = c_N e^{-(N\pi)^2 T} \sin(N\pi x) \ .$$

Any information about high-frequency components will be lost in the presence of measurement errors.

Inverse Problem

Physical Model

$$\mathcal{G}(u) \to y$$

- $u \in X$ parameter vector / parameter function
- $\mathcal{G} : X \to Y$ forward response operator
- y result / observations
- Evaluation of $\mathcal{G}$ expensive

Forward Problem

Find the output y for given parameters u

$\to$ **well-posed**

Inverse Problem

Find the unknown data $u \in X$ from noisy observations

$$y = \mathcal{G}(u) + \eta \qquad \text{with } \eta \sim \mathcal{N}(0, \Gamma)$$

- $u \in X$ parameter vector / parameter function
- $\mathcal{G} : X \to Y$ forward response operator
- y result / observations
- Evaluation of $\mathcal{G}$ expensive

Forward Problem

Find the output y for given parameters u

$\to$ **well-posed**

Inverse Problem

Find the parameters u from (noisy) observations y

$\to$ **ill-posed**

Deterministic Optimization Problem

Find the unknown data $u \in X$ from noisy observations

$$y = \mathcal{G}(u) + \eta$$

Deterministic optimization problem

$$\min_u \frac{1}{2} \|y - \mathcal{G}(u)\|^2 + R(u)$$

- $\|y - \mathcal{G}(u)\|$ potential / data misfit
- R regularization term

Deterministic Optimization Problem

Find the unknown data $u \in X$ from noisy observations

$$y = \mathcal{G}(u) + \eta$$

Deterministic optimization problem

$$\min_u \frac{1}{2}\|y - \mathcal{G}(u)\|^2 + R(u)$$

- **Large-scale, deterministic optimization problem**
- **No quantification of the uncertainty in the unknown $\mathbf{u}$**
- **Proper choice of the regularization term $\mathbf{R}$**

Bayesian Inverse Problem

Find the unknown data $u \in X$ from noisy observations

$$y = \mathcal{G}(u) + \eta$$

Bayesian inverse problem

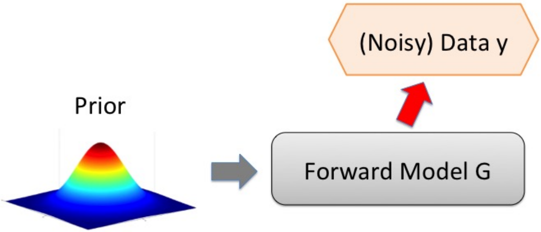
- u, η, y random variables / fields
- Prior μ_0, posterior μ^y
- Goal of computation: moments of system quantities under the posterior w.r. to noisy data

Bayesian Inverse Problem

Find the unknown data $u \in X$ from noisy observations

$$y = \mathcal{G}(u) + \eta$$

Bayesian inverse problem

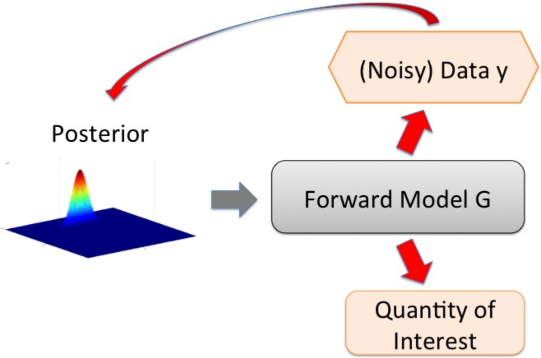


Bayesian Inverse Problem

Find the unknown data $u \in X$ from noisy observations

$$y = \mathcal{G}(u) + \eta$$

Bayesian inverse problem



Bayesian Inverse Problem

Find the unknown data $u \in X$ from noisy observations

$$y = \mathcal{G}(u) + \eta$$

Bayesian inverse problem

- **Quantification of uncertainty in $\mathbf{u}$ and system quantities**
- **Well-posedness of the inverse problem**
- **Incorporation of prior knowledge on the uncertain data $\mathbf{u}$**
- **Need for efficient approximations of the posterior**

Background in Probability Theory

Probability Space $(\Omega, \mathcal{A}, \mu)$

- Ω is a set
- $\mathcal{A}$ is a σ-algebra over Ω
  - the empty set $\{\} \in \mathcal{A}$,
  - the complement $A^c := \{\omega \in \Omega : \omega \notin A\} \in \mathcal{A}$ for all $A \in \mathcal{A}$ and
  - the union $\cup_{j\in\mathbb{N}} A_j \in \mathcal{A}$ for every sequence $(A_j)_{j\in\mathbb{N}} \in \mathcal{A}$.
- A measure μ on a measurable space $(\Omega, \mathcal{A})$ is a mapping from $\mathcal{A}$ to $\mathbb{R}_+ \cup \{\infty\}$ such that
  - the empty set has measure zero, i.e. $\mu(\{\}) = 0$ and
  - $\mu(\cup_{j\in\mathbb{N}} A_j) = \sum_{j\in\mathbb{N}} \mu(A_j)$ if $A_j \in \mathcal{A}$ are disjoint.
- μ is a probability measure iff
  - $\mu(\Omega) = 1$
- $\mathcal{B}(\mathbb{R})$ denotes the Borel σ-algebra and equals the smallest σ-algebra containing all open subsets of $\mathbb{R}$.
- The usual notion of volume of subsets of $\mathbb{R}^d$ gives rise to the Lebesgue measure.

Background in Probability Theory

Probability Space $(\Omega, \mathcal{A}, \mu)$

- Ω is a set
- $\mathcal{A}$ is a σ-algebra over Ω
  - the empty set $\{\} \in \mathcal{A}$,
  - the complement $A^c := \{\omega \in \Omega : \omega \notin A\} \in \mathcal{A}$ for all $A \in \mathcal{A}$ and
  - the union $\cup_{j\in\mathbb{N}} A_j \in \mathcal{A}$ for every sequence $(A_j)_{j\in\mathbb{N}} \in \mathcal{A}$.
- A measure μ on a measurable space $(\Omega, \mathcal{A})$ is a mapping from $\mathcal{A}$ to $\mathbb{R}_+ \cup \{\infty\}$ such that
  - the empty set has measure zero, i.e. $\mu(\{\}) = 0$ and
  - $\mu(\cup_{j\in\mathbb{N}} A_j) = \sum_{j\in\mathbb{N}} \mu(A_j)$ if $A_j \in \mathcal{A}$ are disjoint.
- μ is a probability measure iff
  - $\mu(\Omega) = 1$
- $\mathcal{B}(\mathbb{R})$ denotes the Borel σ-algebra and equals the smallest σ-algebra containing all open subsets of $\mathbb{R}$.
- The usual notion of volume of subsets of $\mathbb{R}^d$ gives rise to the Lebesgue measure.

Background in Probability Theory

Probability Space $(\Omega, \mathcal{A}, \mu)$

- Ω is a set
- $\mathcal{A}$ is a σ-algebra over Ω
  - the empty set $\{\} \in \mathcal{A}$,
  - the complement $A^c := \{\omega \in \Omega : \omega \notin A\} \in \mathcal{A}$ for all $A \in \mathcal{A}$ and
  - the union $\cup_{j \in \mathbb{N}} A_j \in \mathcal{A}$ for every sequence $(A_j)_{j \in \mathbb{N}} \in \mathcal{A}$.
- A measure μ on a measurable space $(\Omega, \mathcal{A})$ is a mapping from $\mathcal{A}$ to $\mathbb{R}_+ \cup \{\infty\}$ such that
  - the empty set has measure zero, i.e. $\mu(\{\}) = 0$ and
  - $\mu(\cup_{j \in \mathbb{N}} A_j) = \sum_{j \in \mathbb{N}} \mu(A_j)$ if $A_j \in \mathcal{A}$ are disjoint.
- μ is a probability measure iff
  - $\mu(\Omega) = 1$
- $\mathcal{B}(\mathbb{R})$ denotes the Borel σ-algebra and equals the smallest σ-algebra containing all open subsets of $\mathbb{R}$.
- The usual notion of volume of subsets of $\mathbb{R}^d$ gives rise to the Lebesgue measure.

Background in Probability Theory

Probability Space $(\Omega, \mathcal{A}, \mu)$

- Ω is a set
- $\mathcal{A}$ is a σ-algebra over Ω
  - the empty set $\{\} \in \mathcal{A}$,
  - the complement $A^c := \{\omega \in \Omega : \omega \notin A\} \in \mathcal{A}$ for all $A \in \mathcal{A}$ and
  - the union $\cup_{j \in \mathbb{N}} A_j \in \mathcal{A}$ for every sequence $(A_j)_{j \in \mathbb{N}} \in \mathcal{A}$.
- A measure μ on a measurable space $(\Omega, \mathcal{A})$ is a mapping from $\mathcal{A}$ to $\mathbb{R}_+ \cup \{\infty\}$ such that
  - the empty set has measure zero, i.e. $\mu(\{\}) = 0$ and
  - $\mu(\cup_{j \in \mathbb{N}} A_j) = \sum_{j \in \mathbb{N}} \mu(A_j)$ if $A_j \in \mathcal{A}$ are disjoint.
- μ is a probability measure iff
  - $\mu(\Omega) = 1$
- $\mathcal{B}(\mathbb{R})$ denotes the Borel σ-algebra and equals the smallest σ-algebra containing all open subsets of $\mathbb{R}$.
- The usual notion of volume of subsets of $\mathbb{R}^d$ gives rise to the Lebesgue measure.

Background in Probability Theory

Probability Space $(\Omega, \mathcal{A}, \mu)$

- Ω is a set
- $\mathcal{A}$ is a σ-algebra over Ω
  - the empty set $\{\} \in \mathcal{A}$,
  - the complement $A^c := \{\omega \in \Omega : \omega \notin A\} \in \mathcal{A}$ for all $A \in \mathcal{A}$ and
  - the union $\cup_{j \in \mathbb{N}} A_j \in \mathcal{A}$ for every sequence $(A_j)_{j \in \mathbb{N}} \in \mathcal{A}$.
- A measure μ on a measurable space $(\Omega, \mathcal{A})$ is a mapping from $\mathcal{A}$ to $\mathbb{R}_+ \cup \{\infty\}$ such that
  - the empty set has measure zero, i.e. $\mu(\{\}) = 0$ and
  - $\mu(\cup_{j \in \mathbb{N}} A_j) = \sum_{j \in \mathbb{N}} \mu(A_j)$ if $A_j \in \mathcal{A}$ are disjoint.
- μ is a probability measure iff
  - $\mu(\Omega) = 1$
- $\mathcal{B}(\mathbb{R})$ denotes the Borel σ-algebra and equals the smallest σ-algebra containing all open subsets of $\mathbb{R}$.
- The usual notion of volume of subsets of $\mathbb{R}^d$ gives rise to the Lebesgue measure.

Background in Probability Theory

Probability Space $(\Omega, \mathcal{A}, \mu)$

- Ω is a set
- $\mathcal{A}$ is a σ-algebra over Ω
  - the empty set $\{\} \in \mathcal{A}$,
  - the complement $A^c := \{\omega \in \Omega : \omega \notin A\} \in \mathcal{A}$ for all $A \in \mathcal{A}$ and
  - the union $\cup_{j\in\mathbb{N}} A_j \in \mathcal{A}$ for every sequence $(A_j)_{j\in\mathbb{N}} \in \mathcal{A}$.
- A measure μ on a measurable space $(\Omega, \mathcal{A})$ is a mapping from $\mathcal{A}$ to $\mathbb{R}_+ \cup \{\infty\}$ such that
  - the empty set has measure zero, i.e. $\mu(\{\}) = 0$ and
  - $\mu(\cup_{j\in\mathbb{N}} A_j) = \sum_{j\in\mathbb{N}} \mu(A_j)$ if $A_j \in \mathcal{A}$ are disjoint.
- μ is a probability measure iff
  - $\mu(\Omega) = 1$
- $\mathcal{B}(\mathbb{R})$ denotes the Borel σ-algebra and equals the smallest σ-algebra containing all open subsets of $\mathbb{R}$.
- The usual notion of volume of subsets of $\mathbb{R}^d$ gives rise to the Lebesgue measure.

Background in Probability Theory

Probability Space $(\Omega, \mathcal{A}, \mu)$

- Ω is a set
- $\mathcal{A}$ is a σ-algebra over Ω
  - the empty set $\{\} \in \mathcal{A}$,
  - the complement $A^c := \{\omega \in \Omega : \omega \notin A\} \in \mathcal{A}$ for all $A \in \mathcal{A}$ and
  - the union $\cup_{j \in \mathbb{N}} A_j \in \mathcal{A}$ for every sequence $(A_j)_{j \in \mathbb{N}} \in \mathcal{A}$.
- A measure μ on a measurable space $(\Omega, \mathcal{A})$ is a mapping from $\mathcal{A}$ to $\mathbb{R}_+ \cup \{\infty\}$ such that
  - the empty set has measure zero, i.e. $\mu(\{\}) = 0$ and
  - $\mu(\cup_{j \in \mathbb{N}} A_j) = \sum_{j \in \mathbb{N}} \mu(A_j)$ if $A_j \in \mathcal{A}$ are disjoint.
- μ is a probability measure iff
  - $\mu(\Omega) = 1$
- $\mathcal{B}(\mathbb{R})$ denotes the Borel σ-algebra and equals the smallest σ-algebra containing all open subsets of $\mathbb{R}$.
- The usual notion of volume of subsets of $\mathbb{R}^d$ gives rise to the Lebesgue measure.

Background in Probability Theory

Probability Space $(\Omega, \mathcal{A}, \mu)$

- Ω is a set
- $\mathcal{A}$ is a σ-algebra over Ω
  - the empty set $\{\} \in \mathcal{A}$,
  - the complement $A^c := \{\omega \in \Omega : \omega \notin A\} \in \mathcal{A}$ for all $A \in \mathcal{A}$ and
  - the union $\cup_{j\in\mathbb{N}} A_j \in \mathcal{A}$ for every sequence $(A_j)_{j\in\mathbb{N}} \in \mathcal{A}$.
- A measure μ on a measurable space $(\Omega, \mathcal{A})$ is a mapping from $\mathcal{A}$ to $\mathbb{R}_+ \cup \{\infty\}$ such that
  - the empty set has measure zero, i.e. $\mu(\{\}) = 0$ and
  - $\mu(\cup_{j\in\mathbb{N}} A_j) = \sum_{j\in\mathbb{N}} \mu(A_j)$ if $A_j \in \mathcal{A}$ are disjoint.
- μ is a probability measure iff
  - $\mu(\Omega) = 1$
- $\mathcal{B}(\mathbb{R})$ denotes the Borel σ-algebra and equals the smallest σ-algebra containing all open subsets of $\mathbb{R}$.
- The usual notion of volume of subsets of $\mathbb{R}^d$ gives rise to the Lebesgue measure.

Background in Probability Theory

Probability Space $(\Omega, \mathcal{A}, \mu)$

- Ω is a set
- $\mathcal{A}$ is a σ-algebra over Ω
  - the empty set $\{\} \in \mathcal{A}$,
  - the complement $A^c := \{\omega \in \Omega : \omega \notin A\} \in \mathcal{A}$ for all $A \in \mathcal{A}$ and
  - the union $\cup_{j\in\mathbb{N}} A_j \in \mathcal{A}$ for every sequence $(A_j)_{j\in\mathbb{N}} \in \mathcal{A}$.
- A measure μ on a measurable space $(\Omega, \mathcal{A})$ is a mapping from $\mathcal{A}$ to $\mathbb{R}_+ \cup \{\infty\}$ such that
  - the empty set has measure zero, i.e. $\mu(\{\}) = 0$ and
  - $\mu(\cup_{j\in\mathbb{N}} A_j) = \sum_{j\in\mathbb{N}} \mu(A_j)$ if $A_j \in \mathcal{A}$ are disjoint.
- μ is a probability measure iff
  - $\mu(\Omega) = 1$
- $\mathcal{B}(\mathbb{R})$ denotes the Borel σ-algebra and equals the smallest σ-algebra containing all open subsets of $\mathbb{R}$.
- The usual notion of volume of subsets of $\mathbb{R}^d$ gives rise to the Lebesgue measure.

Background in Probability Theory

Probability Space $(\Omega, \mathcal{A}, \mu)$

- Ω is a set
- $\mathcal{A}$ is a σ-algebra over Ω
  - the empty set $\{\} \in \mathcal{A}$,
  - the complement $A^c := \{\omega \in \Omega : \omega \notin A\} \in \mathcal{A}$ for all $A \in \mathcal{A}$ and
  - the union $\cup_{j\in\mathbb{N}} A_j \in \mathcal{A}$ for every sequence $(A_j)_{j\in\mathbb{N}} \in \mathcal{A}$.
- A measure μ on a measurable space $(\Omega, \mathcal{A})$ is a mapping from $\mathcal{A}$ to $\mathbb{R}_+ \cup \{\infty\}$ such that
  - the empty set has measure zero, i.e. $\mu(\{\}) = 0$ and
  - $\mu(\cup_{j\in\mathbb{N}} A_j) = \sum_{j\in\mathbb{N}} \mu(A_j)$ if $A_j \in \mathcal{A}$ are disjoint.
- μ is a probability measure iff
  - $\mu(\Omega) = 1$
- $\mathcal{B}(\mathbb{R})$ denotes the Borel σ-algebra and equals the smallest σ-algebra containing all open subsets of $\mathbb{R}$.
- The usual notion of volume of subsets of $\mathbb{R}^d$ gives rise to the Lebesgue measure.

Background in Probability Theory

Probability Space $(\Omega, \mathcal{A}, \mu)$

- Ω is a set
- $\mathcal{A}$ is a σ-algebra over Ω
  - the empty set $\{\} \in \mathcal{A}$,
  - the complement $A^c := \{\omega \in \Omega : \omega \notin A\} \in \mathcal{A}$ for all $A \in \mathcal{A}$ and
  - the union $\cup_{j \in \mathbb{N}} A_j \in \mathcal{A}$ for every sequence $(A_j)_{j \in \mathbb{N}} \in \mathcal{A}$.
- A measure μ on a measurable space $(\Omega, \mathcal{A})$ is a mapping from $\mathcal{A}$ to $\mathbb{R}_+ \cup \{\infty\}$ such that
  - the empty set has measure zero, i.e. $\mu(\{\}) = 0$ and
  - $\mu(\cup_{j \in \mathbb{N}} A_j) = \sum_{j \in \mathbb{N}} \mu(A_j)$ if $A_j \in \mathcal{A}$ are disjoint.
- μ is a probability measure iff
  - $\mu(\Omega) = 1$
- $\mathcal{B}(\mathbb{R})$ denotes the Borel σ-algebra and equals the smallest σ-algebra containing all open subsets of $\mathbb{R}$.
- The usual notion of volume of subsets of $\mathbb{R}^d$ gives rise to the Lebesgue measure.

Background in Probability Theory

Probability Space $(\Omega, \mathcal{A}, \mu)$

- Ω is a set
- $\mathcal{A}$ is a σ-algebra over Ω
  - the empty set $\{\} \in \mathcal{A}$,
  - the complement $A^c := \{\omega \in \Omega : \omega \notin A\} \in \mathcal{A}$ for all $A \in \mathcal{A}$ and
  - the union $\cup_{j\in\mathbb{N}} A_j \in \mathcal{A}$ for every sequence $(A_j)_{j\in\mathbb{N}} \in \mathcal{A}$.
- A measure μ on a measurable space $(\Omega, \mathcal{A})$ is a mapping from $\mathcal{A}$ to $\mathbb{R}_+ \cup \{\infty\}$ such that
  - the empty set has measure zero, i.e. $\mu(\{\}) = 0$ and
  - $\mu(\cup_{j\in\mathbb{N}} A_j) = \sum_{j\in\mathbb{N}} \mu(A_j)$ if $A_j \in \mathcal{A}$ are disjoint.
- μ is a probability measure iff
  - $\mu(\Omega) = 1$
- $\mathcal{B}(\mathbb{R})$ denotes the Borel σ-algebra and equals the smallest σ-algebra containing all open subsets of $\mathbb{R}$.
- The usual notion of volume of subsets of $\mathbb{R}^d$ gives rise to the Lebesgue measure.

Real-valued Random Variables

Let $(\Omega, \mathcal{A}, \mu)$ be a probability space. A mapping $f : \Omega \to \mathbb{R}$ is called a random variable, if f is $\mathcal{A} - \mathcal{B}(\mathbb{R})$-measurable, i.e. for all $B \in \mathcal{B}(\mathbb{R})$

$$f^{-1}(B) = \{\omega \in \Omega : f(\omega) \in B\} \in \mathcal{A} \quad \Rightarrow \quad f^{-1}(\mathcal{B}(\mathbb{R})) \subset \mathcal{A}$$

- Image measure/distribution
  - $\mu_f(B) := (\mu \circ f^{-1}(B)) = \mu(f^{-1}(B)) = \mu(\{f \in B\})$.
  - $(\mathbb{R}, \mathcal{B}(\mathbb{R}), \mu_f)$ is again a probability measure.
- Distribution function
  - $F_f(x) := \mu_f((-\infty, x]) = \mu(f \le x) = \mu(\{\omega \in \Omega : f(\omega) \le x\})$.
- Density function
  - If $\mu_f(B) := \int_B \rho(x) \mathrm{d}x$ for all $B \in \mathcal{B}(\mathbb{R})$ with a measurable function $\rho : \mathbb{R} \to [0, \infty)$.
- If h is a real-valued measurable function on $\mathbb{R}$

$$\Rightarrow h \circ f \quad \mathcal{A} - \mathcal{B}(\mathbb{R})\text{-measurable}$$

Real-valued Random Variables

Let $(\Omega, \mathcal{A}, \mu)$ be a probability space. A mapping $f : \Omega \to \mathbb{R}$ is called a random variable, if f is $\mathcal{A} - \mathcal{B}(\mathbb{R})$-measurable, i.e. for all $B \in \mathcal{B}(\mathbb{R})$

$$f^{-1}(B) = \{\omega \in \Omega : f(\omega) \in B\} \in \mathcal{A} \quad \Rightarrow \quad f^{-1}(\mathcal{B}(\mathbb{R})) \subset \mathcal{A}$$

- Image measure/distribution
  - $\mu_f(B) := (\mu \circ f^{-1}(B)) = \mu(f^{-1}(B)) = \mu(\{f \in B\})$.
  - $(\mathbb{R}, \mathcal{B}(\mathbb{R}), \mu_f)$ is again a probability measure.
- Distribution function
  - $F_f(x) := \mu_f((-\infty, x]) = \mu(f \leq x) = \mu(\{\omega \in \Omega : f(\omega) \leq x\})$.
- Density function
  - If $\mu_f(B) := \int_B \rho(x) \mathrm{d}x$ for all $B \in \mathcal{B}(\mathbb{R})$ with a measurable function $\rho : \mathbb{R} \to [0, \infty)$.
- If h is a real-valued measurable function on $\mathbb{R}$

$$\Rightarrow h \circ f \quad \mathcal{A} - \mathcal{B}(\mathbb{R})\text{-measurable}$$

Real-valued Random Variables

Let $(\Omega, \mathcal{A}, \mu)$ be a probability space. A mapping $f : \Omega \to \mathbb{R}$ is called a random variable, if f is $\mathcal{A} - \mathcal{B}(\mathbb{R})$-measurable, i.e. for all $B \in \mathcal{B}(\mathbb{R})$

$$f^{-1}(B) = \{\omega \in \Omega : f(\omega) \in B\} \in \mathcal{A} \quad \Rightarrow \quad f^{-1}(\mathcal{B}(\mathbb{R})) \subset \mathcal{A}$$

- Image measure/distribution
  - $\mu_f(B) := (\mu \circ f^{-1}(B)) = \mu(f^{-1}(B)) = \mu(\{f \in B\})$.
  - $(\mathbb{R}, \mathcal{B}(\mathbb{R}), \mu_f)$ is again a probability measure.

- Distribution function
  - $F_f(x) := \mu_f((-\infty, x]) = \mu(f \leq x) = \mu(\{\omega \in \Omega : f(\omega) \leq x\})$.

- Density function
  - If $\mu_f(B) := \int_B \rho(x) \mathrm{d}x$ for all $B \in \mathcal{B}(\mathbb{R})$ with a measurable function $\rho : \mathbb{R} \to [0, \infty)$.

- If h is a real-valued measurable function on $\mathbb{R}$

$$\Rightarrow h \circ f \quad \mathcal{A} - \mathcal{B}(\mathbb{R})\text{-measurable}$$

Real-valued Random Variables

Let $(\Omega, \mathcal{A}, \mu)$ be a probability space. A mapping $f : \Omega \to \mathbb{R}$ is called a random variable, if f is $\mathcal{A} - \mathcal{B}(\mathbb{R})$-measurable, i.e. for all $B \in \mathcal{B}(\mathbb{R})$

$$f^{-1}(B) = \{\omega \in \Omega : f(\omega) \in B\} \in \mathcal{A} \quad \Rightarrow \quad f^{-1}(\mathcal{B}(\mathbb{R})) \subset \mathcal{A}$$

- Image measure/distribution
  - $\mu_f(B) := (\mu \circ f^{-1}(B)) = \mu(f^{-1}(B)) = \mu(\{f \in B\})$.
  - $(\mathbb{R}, \mathcal{B}(\mathbb{R}), \mu_f)$ is again a probability measure.
- Distribution function
  - $F_f(x) := \mu_f((-\infty, x]) = \mu(f \leq x) = \mu(\{\omega \in \Omega : f(\omega) \leq x\})$.
- Density function
  - If $\mu_f(B) := \int_B \rho(x)\mathrm{d}x$ for all $B \in \mathcal{B}(\mathbb{R})$ with a measurable function $\rho : \mathbb{R} \to [0, \infty)$.
- If h is a real-valued measurable function on $\mathbb{R}$

$$\Rightarrow h \circ f \quad \mathcal{A} - \mathcal{B}(\mathbb{R})\text{-measurable}$$

Real-valued Random Variables

Let $(\Omega, \mathcal{A}, \mu)$ be a probability space. A mapping $f : \Omega \to \mathbb{R}$ is called a random variable, if f is $\mathcal{A} - \mathcal{B}(\mathbb{R})$-measurable, i.e. for all $B \in \mathcal{B}(\mathbb{R})$

$$f^{-1}(B) = \{\omega \in \Omega : f(\omega) \in B\} \in \mathcal{A} \quad \Rightarrow \quad f^{-1}(\mathcal{B}(\mathbb{R})) \subset \mathcal{A}$$

- Image measure/distribution
  - $\mu_f(B) := (\mu \circ f^{-1}(B)) = \mu(f^{-1}(B)) = \mu(\{f \in B\})$.
  - $(\mathbb{R}, \mathcal{B}(\mathbb{R}), \mu_f)$ is again a probability measure.

- Distribution function
  - $F_f(x) := \mu_f((-\infty, x]) = \mu(f \leq x) = \mu(\{\omega \in \Omega : f(\omega) \leq x\})$.

- Density function
  - If $\mu_f(B) := \int_B \rho(x)\mathrm{d}x$ for all $B \in \mathcal{B}(\mathbb{R})$ with a measurable function $\rho : \mathbb{R} \to [0, \infty)$.

- If h is a real-valued measurable function on $\mathbb{R}$

$$\Rightarrow h \circ f \quad \mathcal{A} - \mathcal{B}(\mathbb{R})\text{-measurable}$$

Real-valued Random Variables

Let $(\Omega, \mathcal{A}, \mu)$ be a probability space. A mapping $f : \Omega \to \mathbb{R}$ is called a random variable, if f is $\mathcal{A} - \mathcal{B}(\mathbb{R})$-measurable, i.e. for all $B \in \mathcal{B}(\mathbb{R})$

$$f^{-1}(B) = \{\omega \in \Omega : f(\omega) \in B\} \in \mathcal{A} \quad \Rightarrow \quad f^{-1}(\mathcal{B}(\mathbb{R})) \subset \mathcal{A}$$

- Image measure/distribution
  - $\mu_f(B) := (\mu \circ f^{-1}(B)) = \mu(f^{-1}(B)) = \mu(\{f \in B\})$.
  - $(\mathbb{R}, \mathcal{B}(\mathbb{R}), \mu_f)$ is again a probability measure.
- Distribution function
  - $F_f(x) := \mu_f((-\infty, x]) = \mu(f \leq x) = \mu(\{\omega \in \Omega : f(\omega) \leq x\})$.
- Density function
  - If $\mu_f(B) := \int_B \rho(x) \mathrm{d}x$ for all $B \in \mathcal{B}(\mathbb{R})$ with a measurable function $\rho : \mathbb{R} \to [0, \infty)$.
- If h is a real-valued measurable function on $\mathbb{R}$

$$\Rightarrow h \circ f \quad \mathcal{A} - \mathcal{B}(\mathbb{R})\text{-measurable}$$

Real-valued Random Variables

Let $(\Omega, \mathcal{A}, \mu)$ be a probability space. A mapping $f : \Omega \to \mathbb{R}$ is called a random variable, if f is $\mathcal{A} - \mathcal{B}(\mathbb{R})$-measurable, i.e. for all $B \in \mathcal{B}(\mathbb{R})$

$$f^{-1}(B) = \{\omega \in \Omega : f(\omega) \in B\} \in \mathcal{A} \quad \Rightarrow \quad f^{-1}(\mathcal{B}(\mathbb{R})) \subset \mathcal{A}$$

- Image measure/distribution
  - $\mu_f(B) := (\mu \circ f^{-1}(B)) = \mu(f^{-1}(B)) = \mu(\{f \in B\})$.
  - $(\mathbb{R}, \mathcal{B}(\mathbb{R}), \mu_f)$ is again a probability measure.
- Distribution function
  - $F_f(x) := \mu_f((-\infty, x]) = \mu(f \leq x) = \mu(\{\omega \in \Omega : f(\omega) \leq x\})$.
- Density function
  - If $\mu_f(B) := \int_B \rho(x)\mathrm{d}x$ for all $B \in \mathcal{B}(\mathbb{R})$ with a measurable function $\rho : \mathbb{R} \to [0, \infty)$.
- If h is a real-valued measurable function on $\mathbb{R}$

$$\Rightarrow h \circ f \quad \mathcal{A} - \mathcal{B}(\mathbb{R})\text{-measurable}$$

Real-valued Random Variables

Let $(\Omega, \mathcal{A}, \mu)$ be a probability space. A mapping $f : \Omega \to \mathbb{R}$ is called a random variable, if f is $\mathcal{A} - \mathcal{B}(\mathbb{R})$-measurable, i.e. for all $B \in \mathcal{B}(\mathbb{R})$

$$f^{-1}(B) = \{\omega \in \Omega : f(\omega) \in B\} \in \mathcal{A} \quad \Rightarrow \quad f^{-1}(\mathcal{B}(\mathbb{R})) \subset \mathcal{A}$$

- Image measure/distribution
  - $\mu_f(B) := (\mu \circ f^{-1}(B)) = \mu(f^{-1}(B)) = \mu(\{f \in B\})$.
  - $(\mathbb{R}, \mathcal{B}(\mathbb{R}), \mu_f)$ is again a probability measure.
- Distribution function
  - $F_f(x) := \mu_f((-\infty, x]) = \mu(f \leq x) = \mu(\{\omega \in \Omega : f(\omega) \leq x\})$.
- Density function
  - If $\mu_f(B) := \int_B \rho(x)\mathrm{d}x$ for all $B \in \mathcal{B}(\mathbb{R})$ with a measurable function $\rho : \mathbb{R} \to [0, \infty)$.
- If h is a real-valued measurable function on $\mathbb{R}$

$$\Rightarrow h \circ f \quad \mathcal{A} - \mathcal{B}(\mathbb{R})\text{-measurable}$$

Expectations

- For $f \in \mathcal{L}^1(\mu)$ is the expectation defined as

$$\mathbb{E}(f) = \int_\Omega f \mathrm{d}\mu = \int_\Omega f(\omega) \mathrm{d}\mu(\omega)$$

- For a real-valued measurable function h
  - discretely distributed random variable: $\mu_f = \sum_{i=1}^n p_i \mathbb{1}_{x_i}$

$$\mathbb{E}(h \circ f) = \int_\Omega h \circ f \mathrm{d}\mu = \int_\mathbb{R} h \mathrm{d}\mu_f = \sum_{i=1}^n p_i f_{x_i}$$

  - continuously distributed random variable: $\mu_f([a,b]) = \int_a^b \rho(x) \mathrm{d}x$.

$$\mathbb{E}(h \circ f) = \int_\Omega h \circ f \mathrm{d}\mu = \int_\mathbb{R} h \mathrm{d}\mu_f = \int_\mathbb{R} h(x)\rho(x) \mathrm{d}x \ .$$

- For $f \in \mathcal{L}^2(\mu)$ is the variance defined as

$$\mathbb{V}(f) = \mathbb{E}((f - \mathbb{E}(f))^2) = \mathbb{E}(f^2) - (\mathbb{E}(f))^2 \ .$$

Expectations

- For $f \in \mathcal{L}^1(\mu)$ is the expectation defined as

$$\mathbb{E}(f) = \int_\Omega f \mathrm{d}\mu = \int_\Omega f(\omega) \mathrm{d}\mu(\omega)$$

- For a real-valued measurable function h
  - discretely distributed random variable: $\mu_f = \sum_{i=1}^n p_i \mathbb{1}_{x_i}$

$$\mathbb{E}(h \circ f) = \int_\Omega h \circ f \mathrm{d}\mu = \int_\mathbb{R} h \mathrm{d}\mu_f = \sum_{i=1}^n p_i f_{x_i}$$

  - continuously distributed random variable: $\mu_f([a,b]) = \int_a^b \rho(x) \mathrm{d}x$.

$$\mathbb{E}(h \circ f) = \int_\Omega h \circ f \mathrm{d}\mu = \int_\mathbb{R} h \mathrm{d}\mu_f = \int_\mathbb{R} h(x)\rho(x) \mathrm{d}x \, .$$

- For $f \in \mathcal{L}^2(\mu)$ is the variance defined as

$$\mathbb{V}(f) = \mathbb{E}((f - \mathbb{E}(f))^2) = \mathbb{E}(f^2) - (\mathbb{E}(f))^2 \, .$$

Expectations

- For $f \in \mathcal{L}^1(\mu)$ is the expectation defined as

$$\mathbb{E}(f) = \int_\Omega f \mathrm{d}\mu = \int_\Omega f(\omega)\mathrm{d}\mu(\omega)$$

- For a real-valued measurable function h
  - discretely distributed random variable: $\mu_f = \sum_{i=1}^n p_i \mathbb{1}_{x_i}$

$$\mathbb{E}(h \circ f) = \int_\Omega h \circ f \mathrm{d}\mu = \int_\mathbb{R} h \mathrm{d}\mu_f = \sum_{i=1}^n p_i f_{x_i}$$

  - continuously distributed random variable: $\mu_f([a, b]) = \int_a^b \rho(x)\mathrm{d}x$.

$$\mathbb{E}(h \circ f) = \int_\Omega h \circ f \mathrm{d}\mu = \int_\mathbb{R} h \mathrm{d}\mu_f = \int_\mathbb{R} h(x)\rho(x)\mathrm{d}x \ .$$

- For $f \in \mathcal{L}^2(\mu)$ is the variance defined as

$$\mathbb{V}(f) = \mathbb{E}((f - \mathbb{E}(f))^2) = \mathbb{E}(f^2) - (\mathbb{E}(f))^2 \ .$$

Expectations

- For $f \in \mathcal{L}^1(\mu)$ is the expectation defined as

$$\mathbb{E}(f) = \int_\Omega f \mathrm{d}\mu = \int_\Omega f(\omega) \mathrm{d}\mu(\omega)$$

- For a real-valued measurable function h
  - discretely distributed random variable: $\mu_f = \sum_{i=1}^n p_i \mathbb{1}_{x_i}$

$$\mathbb{E}(h \circ f) = \int_\Omega h \circ f \mathrm{d}\mu = \int_\mathbb{R} h \mathrm{d}\mu_f = \sum_{i=1}^n p_i f_{x_i}$$

  - continuously distributed random variable: $\mu_f([a,b]) = \int_a^b \rho(x) \mathrm{d}x$.

$$\mathbb{E}(h \circ f) = \int_\Omega h \circ f \mathrm{d}\mu = \int_\mathbb{R} h \mathrm{d}\mu_f = \int_\mathbb{R} h(x)\rho(x) \mathrm{d}x \, .$$

- For $f \in \mathcal{L}^2(\mu)$ is the variance defined as

$$\mathbb{V}(f) = \mathbb{E}((f - \mathbb{E}(f))^2) = \mathbb{E}(f^2) - (\mathbb{E}(f))^2 \, .$$

Expectations

- For $f \in \mathcal{L}^1(\mu)$ is the expectation defined as

$$\mathbb{E}(f) = \int_\Omega f \mathrm{d}\mu = \int_\Omega f(\omega)\mathrm{d}\mu(\omega)$$

- For a real-valued measurable function h
  - discretely distributed random variable: $\mu_f = \sum_{i=1}^n p_i \mathbb{1}_{x_i}$

$$\mathbb{E}(h \circ f) = \int_\Omega h \circ f \mathrm{d}\mu = \int_\mathbb{R} h \mathrm{d}\mu_f = \sum_{i=1}^n p_i f_{x_i}$$

  - continuously distributed random variable: $\mu_f([a,b]) = \int_a^b \rho(x)\mathrm{d}x$.

$$\mathbb{E}(h \circ f) = \int_\Omega h \circ f \mathrm{d}\mu = \int_\mathbb{R} h \mathrm{d}\mu_f = \int_\mathbb{R} h(x)\rho(x)\mathrm{d}x \, .$$

- For $f \in \mathcal{L}^2(\mu)$ is the variance defined as

$$\mathbb{V}(f) = \mathbb{E}((f - \mathbb{E}(f))^2) = \mathbb{E}(f^2) - (\mathbb{E}(f))^2 \, .$$

Product Spaces and Joint Distributions

- Product space $(\times_{i=1}^N \Omega_i, \otimes_{i=1}^N \mathcal{A}_i, \otimes_{i=1}^N \mu_i)$
  - $\times_{i=1}^N \Omega_i = \{(\omega_1, \ldots, \omega_N),\ \omega_i \in \Omega_i,\ \imath = 1, \ldots, N\}$.
  - $\otimes_{i=1}^N \mathcal{A}_i$ smallest σ-algebra that contains all sets of the form $A = A_1 \times \ldots \times A_N$ with $A_i \in \mathcal{A}_i,\ i = 1, \ldots, N$.
  - $\otimes_{i=1}^N \mu_i$ is defined by

$$\otimes_{i=1}^N \mu_i = \mu_1(A_1) \times \ldots \times \mu_N(A_N)\,, \qquad \forall A_1 \in \mathcal{A}_1, \ldots, A_N \in \mathcal{A}_N\,.$$

- For a real-valued random variables $(f_i,\ i = 1, \ldots, N)$, we define $f_1 \otimes \ldots \otimes f_N : \Omega \to \mathbb{R}^N$ via $\omega \mapsto (f_1(\omega), \ldots, f_N(\omega)$ a random variable in $(\mathbb{R}^N, \mathcal{B}(\mathbb{R}^N))$.
  - Joint distribution is for $B \in \mathcal{B}(\mathbb{R}^N)$ given by

$$\mu_{f_1 \otimes \ldots \otimes f_N}(B) = \mu(f_1 \otimes \ldots \otimes f_N \in B)$$

  - Joint distribution function is given by

$$\begin{aligned} F_{f_1 \otimes \ldots \otimes f_N}(x_1, \ldots, x_N) &= \mu_{f_1 \otimes \ldots \otimes f_N}((-\infty, x_1], \ldots, (-\infty, x_N]) \\ &= \mu(f_1 \le x_1, \ldots, X_n \le x_n) \end{aligned}$$

Product Spaces and Joint Distributions

- Product space $(\times_{i=1}^N \Omega_i, \otimes_{i=1}^N \mathcal{A}_i, \otimes_{i=1}^N \mu_i)$
  - $\times_{i=1}^N \Omega_i = \{(\omega_1, \dots, \omega_N),\ \omega_i \in \Omega_i,\ \imath = 1, \dots, N\}$.
  - $\otimes_{i=1}^N \mathcal{A}_i$ smallest σ-algebra that contains all sets of the form $A = A_1 \times \ldots \times A_N$ with $A_i \in \mathcal{A}_i,\ i = 1, \dots, N$.
  - $\otimes_{i=1}^N \mu_i$ is defined by

$$\otimes_{i=1}^N \mu_i = \mu_1(A_1) \times \ldots \times \mu_N(A_N)\,, \qquad \forall A_1 \in \mathcal{A}_1, \dots, A_N \in \mathcal{A}_N\,.$$

- For a real-valued random variables $(f_i,\ i = 1, \dots, N)$, we define $f_1 \otimes \ldots \otimes f_N : \Omega \to \mathbb{R}^N$ via $\omega \mapsto (f_1(\omega), \dots, f_N(\omega)$ a random variable in $(\mathbb{R}^N, \mathcal{B}(\mathbb{R}^N))$.
  - Joint distribution is for $B \in \mathcal{B}(\mathbb{R}^N)$ given by

$$\mu_{f_1 \otimes \ldots \otimes f_N}(B) = \mu(f_1 \otimes \ldots \otimes f_N \in B)$$

  - Joint distribution function is given by

$$\begin{aligned} F_{f_1 \otimes \ldots \otimes f_N}(x_1, \dots, x_N) &= \mu_{f_1 \otimes \ldots \otimes f_N}((-\infty, x_1], \dots, (-\infty, x_N]) \\ &= \mu(f_1 \le x_1, \dots, X_n \le x_n) \end{aligned}$$

Product Spaces and Joint Distributions

- Product space $(\times_{i=1}^N \Omega_i, \otimes_{i=1}^N \mathcal{A}_i, \otimes_{i=1}^N \mu_i)$
  - $\times_{i=1}^N \Omega_i = \{(\omega_1, \dots, \omega_N),\ \omega_i \in \Omega_i,\ \imath = 1, \dots, N\}$.
  - $\otimes_{i=1}^N \mathcal{A}_i$ smallest σ-algebra that contains all sets of the form $A = A_1 \times \ldots \times A_N$ with $A_i \in \mathcal{A}_i,\ i = 1, \dots, N$.
  - $\otimes_{i=1}^N \mu_i$ is defined by

$$\otimes_{i=1}^N \mu_i = \mu_1(A_1) \times \ldots \times \mu_N(A_N)\,, \qquad \forall A_1 \in \mathcal{A}_1, \dots, A_N \in \mathcal{A}_N\,.$$

- For a real-valued random variables $(f_i,\ i = 1, \dots, N)$, we define $f_1 \otimes \ldots \otimes f_N : \Omega \to \mathbb{R}^N$ via $\omega \mapsto (f_1(\omega), \dots, f_N(\omega)$ a random variable in $(\mathbb{R}^N, \mathcal{B}(\mathbb{R}^N))$.
  - Joint distribution is for $B \in \mathcal{B}(\mathbb{R}^N)$ given by

$$\mu_{f_1 \otimes \ldots \otimes f_N}(B) = \mu(f_1 \otimes \ldots \otimes f_N \in B)$$

  - Joint distribution function is given by

$$\begin{aligned} F_{f_1 \otimes \ldots \otimes f_N}(x_1, \dots, x_N) &= \mu_{f_1 \otimes \ldots \otimes f_N}((-\infty, x_1], \dots, (-\infty, x_N]) \\ &= \mu(f_1 \leq x_1, \dots, X_n \leq x_n) \end{aligned}$$

Product Spaces and Joint Distributions

- Product space $(\times_{i=1}^N \Omega_i, \otimes_{i=1}^N \mathcal{A}_i, \otimes_{i=1}^N \mu_i)$
  - $\times_{i=1}^N \Omega_i = \{(\omega_1, \dots, \omega_N),\ \omega_i \in \Omega_i,\ \imath = 1, \dots, N\}$.
  - $\otimes_{i=1}^N \mathcal{A}_i$ smallest σ-algebra that contains all sets of the form $A = A_1 \times \dots \times A_N$ with $A_i \in \mathcal{A}_i,\ i = 1, \dots, N$.
  - $\otimes_{i=1}^N \mu_i$ is defined by

$$\otimes_{i=1}^N \mu_i = \mu_1(A_1) \times \dots \times \mu_N(A_N)\,, \qquad \forall A_1 \in \mathcal{A}_1, \dots, A_N \in \mathcal{A}_N\,.$$

- For a real-valued random variables $(f_i,\ i = 1, \dots, N)$, we define $f_1 \otimes \dots \otimes f_N : \Omega \to \mathbb{R}^N$ via $\omega \mapsto (f_1(\omega), \dots, f_N(\omega)$ a random variable in $(\mathbb{R}^N, \mathcal{B}(\mathbb{R}^N))$.
  - Joint distribution is for $B \in \mathcal{B}(\mathbb{R}^N)$ given by

$$\mu_{f_1 \otimes \dots \otimes f_N}(B) = \mu(f_1 \otimes \dots \otimes f_N \in B)$$

  - Joint distribution function is given by

$$\begin{aligned} F_{f_1 \otimes \dots \otimes f_N}(x_1, \dots, x_N) &= \mu_{f_1 \otimes \dots \otimes f_N}((-\infty, x_1], \dots, (-\infty, x_N]) \\ &= \mu(f_1 \le x_1, \dots, X_n \le x_n) \end{aligned}$$

Product Spaces and Joint Distributions

- Product space $(\times_{i=1}^N \Omega_i, \otimes_{i=1}^N \mathcal{A}_i, \otimes_{i=1}^N \mu_i)$
  - $\times_{i=1}^N \Omega_i = \{(\omega_1, \ldots, \omega_N),\ \omega_i \in \Omega_i,\ \imath = 1, \ldots, N\}$.
  - $\otimes_{i=1}^N \mathcal{A}_i$ smallest σ-algebra that contains all sets of the form $A = A_1 \times \ldots \times A_N$ with $A_i \in \mathcal{A}_i,\ i = 1, \ldots, N$.
  - $\otimes_{i=1}^N \mu_i$ is defined by

$$\otimes_{i=1}^N \mu_i = \mu_1(A_1) \times \ldots \times \mu_N(A_N)\,, \qquad \forall A_1 \in \mathcal{A}_1, \ldots, A_N \in \mathcal{A}_N\ .$$

- For a real-valued random variables $(f_i,\ i = 1, \ldots, N)$, we define $f_1 \otimes \ldots \otimes f_N : \Omega \to \mathbb{R}^N$ via $\omega \mapsto (f_1(\omega), \ldots, f_N(\omega)$ a random variable in $(\mathbb{R}^N, \mathcal{B}(\mathbb{R}^N))$.
  - Joint distribution is for $B \in \mathcal{B}(\mathbb{R}^N)$ given by

$$\mu_{f_1 \otimes \ldots \otimes f_N}(B) = \mu(f_1 \otimes \ldots \otimes f_N \in B)$$

  - Joint distribution function is given by

$$\begin{aligned} F_{f_1 \otimes \ldots \otimes f_N}(x_1, \ldots, x_N) &= \mu_{f_1 \otimes \ldots \otimes f_N}((-\infty, x_1], \ldots, (-\infty, x_N]) \\ &= \mu(f_1 \leq x_1, \ldots, X_n \leq x_n) \end{aligned}$$

Product Spaces and Joint Distributions

- Product space $(\times_{i=1}^{N}\Omega_i, \otimes_{i=1}^{N}\mathcal{A}_i, \otimes_{i=1}^{N}\mu_i)$
  - $\times_{i=1}^{N}\Omega_i = \{(\omega_1, \ldots, \omega_N),\ \omega_i \in \Omega_i,\ \imath = 1, \ldots, N\}$.
  - $\otimes_{i=1}^{N}\mathcal{A}_i$ smallest σ-algebra that contains all sets of the form $A = A_1 \times \ldots \times A_N$ with $A_i \in \mathcal{A}_i,\ i = 1, \ldots, N$.
  - $\otimes_{i=1}^{N}\mu_i$ is defined by

$$\otimes_{i=1}^{N}\mu_i = \mu_1(A_1) \times \ldots \times \mu_N(A_N)\,, \qquad \forall A_1 \in \mathcal{A}_1, \ldots, A_N \in \mathcal{A}_N\ .$$

- For a real-valued random variables $(f_i,\ i = 1, \ldots, N)$, we define $f_1 \otimes \ldots \otimes f_N : \Omega \to \mathbb{R}^N$ via $\omega \mapsto (f_1(\omega), \ldots, f_N(\omega)$ a random variable in $(\mathbb{R}^N, \mathcal{B}(\mathbb{R}^N))$.
  - Joint distribution is for $B \in \mathcal{B}(\mathbb{R}^N)$ given by

$$\mu_{f_1 \otimes \ldots \otimes f_N}(B) = \mu(f_1 \otimes \ldots \otimes f_N \in B)$$

  - Joint distribution function is given by

$$\begin{aligned} F_{f_1 \otimes \ldots \otimes f_N}(x_1, \ldots, x_N) &= \mu_{f_1 \otimes \ldots \otimes f_N}((-\infty, x_1], \ldots, (-\infty, x_N]) \\ &= \mu(f_1 \le x_1, \ldots, X_n \le x_n) \end{aligned}$$

Product Spaces and Joint Distributions

- Product space $(\times_{i=1}^N \Omega_i, \otimes_{i=1}^N \mathcal{A}_i, \otimes_{i=1}^N \mu_i)$
  - $\times_{i=1}^N \Omega_i = \{(\omega_1, \dots, \omega_N),\ \omega_i \in \Omega_i,\ \imath = 1, \dots, N\}$.
  - $\otimes_{i=1}^N \mathcal{A}_i$ smallest σ-algebra that contains all sets of the form $A = A_1 \times \ldots \times A_N$ with $A_i \in \mathcal{A}_i,\ i = 1, \dots, N$.
  - $\otimes_{i=1}^N \mu_i$ is defined by

$$\otimes_{i=1}^N \mu_i = \mu_1(A_1) \times \ldots \times \mu_N(A_N)\ , \qquad \forall A_1 \in \mathcal{A}_1, \dots, A_N \in \mathcal{A}_N\ .$$

- For a real-valued random variables $(f_i,\ i = 1, \dots, N)$, we define $f_1 \otimes \ldots \otimes f_N : \Omega \to \mathbb{R}^N$ via $\omega \mapsto (f_1(\omega), \dots, f_N(\omega)$ a random variable in $(\mathbb{R}^N, \mathcal{B}(\mathbb{R}^N))$.
  - Joint distribution is for $B \in \mathcal{B}(\mathbb{R}^N)$ given by

$$\mu_{f_1 \otimes \ldots \otimes f_N}(B) = \mu(f_1 \otimes \ldots \otimes f_N \in B)$$

  - Joint distribution function is given by

$$\begin{aligned} F_{f_1 \otimes \ldots \otimes f_N}(x_1, \dots, x_N) &= \mu_{f_1 \otimes \ldots \otimes f_N}((-\infty, x_1], \dots, (-\infty, x_N]) \\ &= \mu(f_1 \le x_1, \dots, X_n \le x_n) \end{aligned}$$

Conditional Probabilities and Expectations

Let $(\Omega, \mathcal{A}, \mu)$ be a probability space.

- The conditional probability that an event $A \in \mathcal{A}$ occurs given that an event $B \in \mathcal{A}$ has occurred is defined by

$$\mu(A|B) = \frac{\mu(A \cap B)}{\mu(B)} \qquad \text{for } \mu(B) > 0\ .$$

- Given two random variables f and h with joint density function ρ and $x \in \mathbb{R}$ such that $\int_{-\infty}^{\infty} \rho(x, v)\mathrm{d}v > 0$,
  - the conditional distribution function of h given $f = x$ is given by

$$F_{h|f}(y|x) = \int_{-\infty}^{y} \frac{\rho(x, v)\mathrm{d}v}{\int_{-\infty}^{\infty} \rho(x, v)\mathrm{d}v}\ .$$

  - the conditional density function of h given $f = x$ is given by

$$\rho_{h|f}(y|x) = \frac{\rho(x, y)}{\int_{-\infty}^{\infty} \rho(x, v)\mathrm{d}v}\ .$$

  - the conditional expectation $\mathbb{E}(h|f)$ is defined by

$$\mathbb{E}(h|f) = \int_{-\infty}^{\infty} y \rho_{h|f}(y|x)\mathrm{d}y\ .$$

Conditional Probabilities and Expectations

Let $(\Omega, \mathcal{A}, \mu)$ be a probability space.

- The conditional probability that an event $A \in \mathcal{A}$ occurs given that an event $B \in \mathcal{A}$ has occurred is defined by

$$\mu(A|B) = \frac{\mu(A \cap B)}{\mu(B)} \qquad \text{for } \mu(B) > 0\ .$$

- Given two random variables f and h with joint density function ρ and $x \in \mathbb{R}$ such that $\int_{-\infty}^{\infty} \rho(x, v)\mathrm{d}v > 0$,
  - the conditional distribution function of h given $f = x$ is given by

$$F_{h|f}(y|x) = \int_{-\infty}^{y} \frac{\rho(x, v)\mathrm{d}v}{\int_{-\infty}^{\infty} \rho(x, v)\mathrm{d}v}\ .$$

  - the conditional density function of h given $f = x$ is given by

$$\rho_{h|f}(y|x) = \frac{\rho(x, y)}{\int_{-\infty}^{\infty} \rho(x, v)\mathrm{d}v}\ .$$

  - the conditional expectation $\mathbb{E}(h|f)$ is defined by

$$\mathbb{E}(h|f) = \int_{-\infty}^{\infty} y\rho_{h|f}(y|x)\mathrm{d}y\ .$$

Conditional Probabilities and Expectations

Let $(\Omega, \mathcal{A}, \mu)$ be a probability space.

- The conditional probability that an event $A \in \mathcal{A}$ occurs given that an event $B \in \mathcal{A}$ has occurred is defined by

$$\mu(A|B) = \frac{\mu(A \cap B)}{\mu(B)} \qquad \text{for } \mu(B) > 0 \; .$$

- Given two random variables f and h with joint density function ρ and $x \in \mathbb{R}$ such that $\int_{-\infty}^{\infty} \rho(x,v)\mathrm{d}v > 0$,
  - the conditional distribution function of h given $f = x$ is given by

$$F_{h|f}(y|x) = \int_{-\infty}^{y} \frac{\rho(x,v)\mathrm{d}v}{\int_{-\infty}^{\infty} \rho(x,v)\mathrm{d}v} \; .$$

  - the conditional density function of h given $f = x$ is given by

$$\rho_{h|f}(y|x) = \frac{\rho(x,y)}{\int_{-\infty}^{\infty} \rho(x,v)\mathrm{d}v} \; .$$

  - the conditional expectation $\mathbb{E}(h|f)$ is defined by

$$\mathbb{E}(h|f) = \int_{-\infty}^{\infty} y \rho_{h|f}(y|x)\mathrm{d}y \; .$$

Conditional Probabilities and Expectations

Let $(\Omega, \mathcal{A}, \mu)$ be a probability space.

- The conditional probability that an event $A \in \mathcal{A}$ occurs given that an event $B \in \mathcal{A}$ has occurred is defined by

$$\mu(A|B) = \frac{\mu(A \cap B)}{\mu(B)} \qquad \text{for } \mu(B) > 0 \; .$$

- Given two random variables f and h with joint density function ρ and $x \in \mathbb{R}$ such that $\int_{-\infty}^{\infty} \rho(x, v) \mathrm{d}v > 0$,
  - the conditional distribution function of h given $f = x$ is given by

$$F_{h|f}(y|x) = \int_{-\infty}^{y} \frac{\rho(x, v)\mathrm{d}v}{\int_{-\infty}^{\infty} \rho(x, v)\mathrm{d}v} \; .$$

  - the conditional density function of h given $f = x$ is given by

$$\rho_{h|f}(y|x) = \frac{\rho(x, y)}{\int_{-\infty}^{\infty} \rho(x, v)\mathrm{d}v} \; .$$

  - the conditional expectation $\mathbb{E}(h|f)$ is defined by

$$\mathbb{E}(h|f) = \int_{-\infty}^{\infty} y \rho_{h|f}(y|x)\mathrm{d}y \; .$$

Conditional Probabilities and Expectations

Let $(\Omega, \mathcal{A}, \mu)$ be a probability space.

- The conditional probability that an event $A \in \mathcal{A}$ occurs given that an event $B \in \mathcal{A}$ has occurred is defined by

$$\mu(A|B) = \frac{\mu(A \cap B)}{\mu(B)} \qquad \text{for } \mu(B) > 0 \, .$$

- Given two random variables f and h with joint density function ρ and $x \in \mathbb{R}$ such that $\int_{-\infty}^{\infty} \rho(x, v)\mathrm{d}v > 0$,
  - the conditional distribution function of h given $f = x$ is given by

$$F_{h|f}(y|x) = \int_{-\infty}^{y} \frac{\rho(x, v)\mathrm{d}v}{\int_{-\infty}^{\infty} \rho(x, v)\mathrm{d}v} \, .$$

  - the conditional density function of h given $f = x$ is given by

$$\rho_{h|f}(y|x) = \frac{\rho(x, y)}{\int_{-\infty}^{\infty} \rho(x, v)\mathrm{d}v} \, .$$

  - the conditional expectation $\mathbb{E}(h|f)$ is defined by

$$\mathbb{E}(h|f) = \int_{-\infty}^{\infty} y \rho_{h|f}(y|x)\mathrm{d}y \, .$$

Independent Random Variables

- For independent random variables $(f_i)_{i=1}^N$, it holds

$$\begin{aligned}
\mu_{f_1\otimes\ldots\otimes f_N}(B_1\times\ldots\times B_N) &= \mu(f_1\in B_1,\ldots,f_N\in B_N)) \\
&= \mu(f_1\in B_1)\cdots\mu(f_N\in B_N) \\
&= \mu_{f_1}(B_1)\cdots\mu_{f_N}(B_N) \\
&= \mu_{f_1}\otimes\ldots\otimes\mu_{f_N}(B_1\times\ldots\times B_N)
\end{aligned}$$

- $(f_i)_{i=1}^N$ are independent $\Leftrightarrow$ the joint distribution is equal to the product measure

Independent Random Variables

- For independent random variables $(f_i)_{i=1}^N$, it holds

$$\begin{aligned} \mu_{f_1 \otimes \ldots \otimes f_N}(B_1 \times \ldots \times B_N) &= \mu(f_1 \in B_1, \ldots, f_N \in B_N)) \\ &= \mu(f_1 \in B_1) \cdots \mu(f_N \in B_N) \\ &= \mu_{f_1}(B_1) \cdots \mu_{f_N}(B_N) \\ &= \mu_{f_1} \otimes \ldots \otimes \mu_{f_N}(B_1 \times \ldots \times B_N) \end{aligned}$$

- $(f_i)_{i=1}^N$ are independent $\Leftrightarrow$ the joint distribution is equal to the product measure

Bayesian Inversion in $\mathbb{R}^n$

Find the unknown data $u \in \mathbb{R}^n$ from noisy observations

$$y = \mathcal{G}(u) + \eta \qquad \eta \in \mathbb{R}^K$$

- $u \in \mathbb{R}^n$ random variable with Lebesgue density $\rho_0(u)$.
- $\mathcal{G} : \mathbb{R}^n \to \mathbb{R}^K$ measurable function.
- $\eta \in \mathbb{R}^K$ independent of u ($u \perp \eta$), distributed according to measure μ_η with Lebesgue density $\rho(\eta)$.
- $y|u$ is then distributed according to measure μ_u with Lebesgue density $\rho(y - \mathcal{G}(u))$.
- $(u, y) \in \mathbb{R}^n \times \mathbb{R}^K$ is a random variable with Lebesgue density $\rho(y - \mathcal{G}(u))\rho_0(u)$.

Bayesian Inversion in $\mathbb{R}^n$

Find the unknown data $u \in \mathbb{R}^n$ from noisy observations

$$y = \mathcal{G}(u) + \eta \qquad \eta \in \mathbb{R}^K$$

- $u \in \mathbb{R}^n$ random variable with Lebesgue density $\rho_0(u)$.
- $\mathcal{G} : \mathbb{R}^n \to \mathbb{R}^K$ measurable function.
- $\eta \in \mathbb{R}^K$ independent of u ($u \perp \eta$), distributed according to measure μ_η with Lebesgue density $\rho(\eta)$.
- $y|u$ is then distributed according to measure μ_u with Lebesgue density $\rho(y - \mathcal{G}(u))$.
- $(u, y) \in \mathbb{R}^n \times \mathbb{R}^K$ is a random variable with Lebesgue density $\rho(y - \mathcal{G}(u))\rho_0(u)$.

Bayesian Inversion in $\mathbb{R}^n$

Find the unknown data $u \in \mathbb{R}^n$ from noisy observations

$$y = \mathcal{G}(u) + \eta \qquad \eta \in \mathbb{R}^K$$

- $u \in \mathbb{R}^n$ random variable with Lebesgue density $\rho_0(u)$.
- $\mathcal{G} : \mathbb{R}^n \to \mathbb{R}^K$ measurable function.
- $\eta \in \mathbb{R}^K$ independent of u ($u \perp \eta$), distributed according to measure μ_η with Lebesgue density $\rho(\eta)$.
- $y|u$ is then distributed according to measure μ_u with Lebesgue density $\rho(y - \mathcal{G}(u))$.
- $(u, y) \in \mathbb{R}^n \times \mathbb{R}^K$ is a random variable with Lebesgue density $\rho(y - \mathcal{G}(u))\rho_0(u)$.

Bayesian Inversion in $\mathbb{R}^n$

Find the unknown data $u \in \mathbb{R}^n$ from noisy observations

$$y = \mathcal{G}(u) + \eta \qquad \eta \in \mathbb{R}^K$$

- $u \in \mathbb{R}^n$ random variable with Lebesgue density $\rho_0(u)$.
- $\mathcal{G} : \mathbb{R}^n \to \mathbb{R}^K$ measurable function.
- $\eta \in \mathbb{R}^K$ independent of u ($u \perp \eta$), distributed according to measure μ_η with Lebesgue density $\rho(\eta)$.
- $y|u$ is then distributed according to measure μ_u with Lebesgue density $\rho(y - \mathcal{G}(u))$.
- $(u, y) \in \mathbb{R}^n \times \mathbb{R}^K$ is a random variable with Lebesgue density $\rho(y - \mathcal{G}(u))\rho_0(u)$.

Bayesian Inversion in $\mathbb{R}^n$

Find the unknown data $u \in \mathbb{R}^n$ from noisy observations

$$y = \mathcal{G}(u) + \eta \qquad \eta \in \mathbb{R}^K$$

- $u \in \mathbb{R}^n$ random variable with Lebesgue density $\rho_0(u)$.
- $\mathcal{G} : \mathbb{R}^n \to \mathbb{R}^K$ measurable function.
- $\eta \in \mathbb{R}^K$ independent of u ($u \perp \eta$), distributed according to measure μ_η with Lebesgue density $\rho(\eta)$.
- $y|u$ is then distributed according to measure μ_u with Lebesgue density $\rho(y - \mathcal{G}(u))$.
- $(u, y) \in \mathbb{R}^n \times \mathbb{R}^K$ is a random variable with Lebesgue density $\rho(y - \mathcal{G}(u))\rho_0(u)$.

Bayesian Inversion in $\mathbb{R}^n$

Find the unknown data $u \in \mathbb{R}^n$ from noisy observations

$$y = \mathcal{G}(u) + \eta \qquad \eta \in \mathbb{R}^K$$

Bayes' Theorem

Assume that

$$Z := \int_{\mathbb{R}^n} \rho(y - \mathcal{G}(u))\rho_0(u)\mathrm{d}u > 0 \ .$$

Then, $u|y$ is a random variable with Lebesgue density $\rho^y(u)$ given by

$$\rho^y(u) = \frac{1}{Z}\rho(y - G(u))\rho_0(u) \ .$$

Bayesian Inversion in $\mathbb{R}^n$

Find the unknown data $u \in \mathbb{R}^n$ from noisy observations

$$y = \mathcal{G}(u) + \eta \qquad \eta \in \mathbb{R}^K$$

- $\rho_0(u)$ is the prior density.
- $\rho(y - \mathcal{G}(u))$ is the likelihood.
- $\rho^y(u)$ is the posterior density.
- $$\Phi(u; y) = -\log \rho(y - \mathcal{G}(u))$$

  is called the potential (negative log likelihood).

Bayesian Inversion in $\mathbb{R}^n$

Find the unknown data $u \in \mathbb{R}^n$ from noisy observations

$$y = \mathcal{G}(u) + \eta \qquad \eta \in \mathbb{R}^K$$

- $\rho_0(u)$ is the prior density.
- $\rho(y - \mathcal{G}(u))$ is the likelihood.
- $\rho^y(u)$ is the posterior density.
- $$\Phi(u; y) = -\log \rho(y - \mathcal{G}(u))$$

  is called the potential (negative log likelihood).

Bayesian Inversion in $\mathbb{R}^n$

Find the unknown data $u \in \mathbb{R}^n$ from noisy observations

$$y = \mathcal{G}(u) + \eta \qquad \eta \in \mathbb{R}^K$$

- $\rho_0(u)$ is the prior density.
- $\rho(y - \mathcal{G}(u))$ is the likelihood.
- $\rho^y(u)$ is the posterior density.
- $$\Phi(u; y) = -\log \rho(y - \mathcal{G}(u))$$
  is called the potential (negative log likelihood).

Bayesian Inversion in $\mathbb{R}^n$

Find the unknown data $u \in \mathbb{R}^n$ from noisy observations

$$y = \mathcal{G}(u) + \eta \qquad \eta \in \mathbb{R}^K$$

- $\rho_0(u)$ is the prior density.
- $\rho(y - \mathcal{G}(u))$ is the likelihood.
- $\rho^y(u)$ is the posterior density.
- $$\Phi(u; y) = -\log \rho(y - \mathcal{G}(u))$$

  is called the potential (negative log likelihood).

Bayesian Inversion in $\mathbb{R}^n$

Find the unknown data $u \in \mathbb{R}^n$ from noisy observations

$$y = \mathcal{G}(u) + \eta \qquad \eta \in \mathbb{R}^K$$

- Bayes' formula expresses

$$\mathbb{P}(u|y) = \frac{1}{\mathbb{P}(y)} \mathbb{P}(y|u)\mathbb{P}(u) \; .$$

- Let μ^y be a measure on $\mathbb{R}^n$ with density ρ^y and μ_0 a measure on $\mathbb{R}^n$ with Lebesgue density ρ_0. Then, Bayes' theorem may be written as

$$\frac{\mathrm{d}\mu^y}{\mathrm{d}\mu_0}(u) = \frac{1}{Z} \exp(-\Phi(u;y)) \; , \qquad Z = \int_{\mathbb{R}^n} \exp(-\Phi(u;y))\mu_0(\mathrm{d}u) \; .$$

- The expression for the Radon-Nikodym derivative is to be interpreted as follows: for all measurable function $f : \mathbb{R}^n \to \mathbb{R}$

$$\mathbb{E}_{\mu^y} f(u) = \mathbb{E}_{\mu_0}\left(\frac{\mathrm{d}\mu^y}{\mathrm{d}\mu_0}(u) f(u)\right) \; .$$

Bayesian Inversion in $\mathbb{R}^n$

Find the unknown data $u \in \mathbb{R}^n$ from noisy observations

$$y = \mathcal{G}(u) + \eta \qquad \eta \in \mathbb{R}^K$$

- Bayes' formula expresses

$$\mathbb{P}(u|y) = \frac{1}{\mathbb{P}(y)} \mathbb{P}(y|u)\mathbb{P}(u) \ .$$

- Let μ^y be a measure on $\mathbb{R}^n$ with density ρ^y and μ_0 a measure on $\mathbb{R}^n$ with Lebesgue density ρ_0. Then, Bayes' theorem may be written as

$$\frac{\mathrm{d}\mu^y}{\mathrm{d}\mu_0}(u) = \frac{1}{Z}\exp(-\Phi(u;y)) \ , \qquad Z = \int_{\mathbb{R}^n} \exp(-\Phi(u;y))\mu_0(\mathrm{d}u) \ .$$

- The expression for the Radon-Nikodym derivative is to be interpreted as follows: for all measurable function $f : \mathbb{R}^n \to \mathbb{R}$

$$\mathbb{E}_{\mu^y} f(u) = \mathbb{E}_{\mu_0}\left(\frac{\mathrm{d}\mu^y}{\mathrm{d}\mu_0}(u) f(u)\right) \ .$$

Bayesian Inversion in $\mathbb{R}^n$

Find the unknown data $u \in \mathbb{R}^n$ from noisy observations

$$y = \mathcal{G}(u) + \eta \qquad \eta \in \mathbb{R}^K$$

- Bayes' formula expresses
$$\mathbb{P}(u|y) = \frac{1}{\mathbb{P}(y)}\mathbb{P}(y|u)\mathbb{P}(u) \; .$$
- Let μ^y be a measure on $\mathbb{R}^n$ with density ρ^y and μ_0 a measure on $\mathbb{R}^n$ with Lebesgue density ρ_0. Then, Bayes' theorem may be written as
$$\frac{\mathrm{d}\mu^y}{\mathrm{d}\mu_0}(u) = \frac{1}{Z}\exp(-\Phi(u;y)) \; , \qquad Z = \int_{\mathbb{R}^n} \exp(-\Phi(u;y))\mu_0(\mathrm{d}u) \; .$$
- The expression for the Radon-Nikodym derivative is to be interpreted as follows: for all measurable function $f : \mathbb{R}^n \to \mathbb{R}$
$$\mathbb{E}_{\mu^y} f(u) = \mathbb{E}_{\mu_0}(\frac{\mathrm{d}\mu^y}{\mathrm{d}\mu_0}(u)f(u)) \; .$$

Bayesian Inversion in $\mathbb{R}^n$

1D Gaussian, linear example

Forward response operator $g \in L(\mathbb{R}, \mathbb{R})$

Prior $\mu_0 = \mathcal{N}(0, \sigma_0^2)$, $\sigma_0 \in \mathbb{R}$, Leb. dens. $\rho_0(u) = \frac{1}{\sigma_0\sqrt{2\pi}} \exp(-\frac{\|u\|^2}{\sigma_0^2})$

Noise $\eta \sim \mathcal{N}(0, \gamma^2)$, $\gamma \in \mathbb{R}$, Leb. dens. $\rho_0(\eta) = \frac{1}{\gamma\sqrt{2\pi}} \exp(-\frac{\|\eta\|^2}{2\gamma^2})$.

Bayesian Inversion in $\mathbb{R}^n$

1D Gaussian, linear example

Forward response operator $g \in L(\mathbb{R}, \mathbb{R})$

Prior $\mu_0 = \mathcal{N}(0, \sigma_0^2)$, $\sigma_0 \in \mathbb{R}$, Leb. dens. $\rho_0(u) = \frac{1}{\sigma_0\sqrt{2\pi}} \exp(-\frac{\|u\|^2}{\sigma_0^2})$

Noise $\eta \sim \mathcal{N}(0, \gamma^2)$, $\gamma \in \mathbb{R}$, Leb. dens. $\rho_0(\eta) = \frac{1}{\gamma\sqrt{2\pi}} \exp(-\frac{\|\eta\|^2}{2\gamma^2})$.

Bayes' formula

$$\rho(u|y) = \frac{\rho(y|u)\rho(u)}{\int_{\mathbb{R}^2} \rho(y|u)\rho(u)\mathrm{d}u}$$

Bayesian Inversion in $\mathbb{R}^n$

1D Gaussian, linear example

Forward response operator $g \in L(\mathbb{R}, \mathbb{R})$

Prior $\mu_0 = \mathcal{N}(0, \sigma_0^2)$, $\sigma_0 \in \mathbb{R}$, Leb. dens. $\rho_0(u) = \frac{1}{\sigma_0\sqrt{2\pi}} \exp(-\frac{\|u\|^2}{\sigma_0^2})$

Noise $\eta \sim \mathcal{N}(0, \gamma^2)$, $\gamma \in \mathbb{R}$, Leb. dens. $\rho_0(\eta) = \frac{1}{\gamma\sqrt{2\pi}} \exp(-\frac{\|\eta\|^2}{2\gamma^2})$.

Bayes' formula

$$\rho(u|y) = \frac{\rho(y|u)\rho(u)}{\int_{\mathbb{R}^2} \rho(y|u)\rho(u)\mathrm{d}u}$$

'

Completing the square

$$\mu^y = \mathcal{N}\Big(\frac{\sigma_0^2 g}{\gamma^2 + g^2\sigma_0^2} y, \sigma_0^2 - \frac{\sigma_0^4 g^2}{\gamma^2 + g^2\sigma_0^2}\Big)$$

By assumption, we have for the prior density $\rho_0(u)$

$$\rho_0(u) \propto \exp\left(-\frac{1}{2\sigma_0^2}u^2\right)$$

and the noise has density

$$\rho(\eta) \propto \exp(-\frac{1}{2\gamma^2}\eta^2)\,.$$

By assumption, we have for the prior density $\rho_0(u)$

$$\rho_0(u) \propto \exp\big(-\frac{1}{2\sigma_0^2}u^2\big)$$

and the noise has density

$$\rho(\eta) \propto \exp(-\frac{1}{2\gamma^2}\eta^2)\ .$$

Thus, by Bayes' formula, the posterior is given by

$$\rho^y(u) \propto \exp\big(-\frac{1}{2\sigma_0^2}u^2 - \frac{1}{2\gamma^2}(y-gu)^2\big) = \exp\big(-\frac{1}{2}\big(\frac{1}{\sigma_0^2}+\frac{g^2}{\gamma^2}\big)u^2 - 2\frac{gy}{\gamma^2}u + \frac{y^2}{u^2}\big)\ .$$

By assumption, we have for the prior density $\rho_0(u)$

$$\rho_0(u) \propto \exp\big(-\frac{1}{2\sigma_0^2}u^2\big)$$

and the noise has density

$$\rho(\eta) \propto \exp(-\frac{1}{2\gamma^2}\eta^2)\ .$$

Thus, by Bayes' formula, the posterior is given by

$$\rho^y(u) \propto \exp\big(-\frac{1}{2\sigma_0^2}u^2 - \frac{1}{2\gamma^2}(y-gu)^2\big) = \exp\big(-\frac{1}{2}\big(\frac{1}{\sigma_0^2}+\frac{g^2}{\gamma^2}\big)u^2 - 2\frac{gy}{\gamma^2}u + \frac{y^2}{u^2}\big)\ .$$

Defining

$$a = \frac{1}{\sigma_0^2} + \frac{g^2}{\gamma^2}\ , \quad b = \frac{gy}{\gamma^2}\ , \quad c = \frac{y^2}{\gamma^2}\ ,$$

we are interested in constants m, K, σ, such that

$$\rho^y(u) \propto \exp(-\frac{1}{2\sigma^2}(u-m)^2 + K)\ .$$

By completing the square, we obtain

$$au^2 - 2bu + c \quad = \quad a(u^2 - 2\frac{b}{a}u + \frac{c}{a})$$

By completing the square, we obtain

$$\begin{aligned} au^2 - 2bu + c &= a(u^2 - 2\frac{b}{a}u + \frac{c}{a}) \\ &= a(u^2 - 2\frac{b}{a}u + (\frac{b}{a})^2 - (\frac{b}{a})^2 + \frac{c}{a}) \end{aligned}$$

By completing the square, we obtain

$$\begin{aligned}
au^2 - 2bu + c &= a(u^2 - 2\frac{b}{a}u + \frac{c}{a}) \\
&= a(u^2 - 2\frac{b}{a}u + (\frac{b}{a})^2 - (\frac{b}{a})^2 + \frac{c}{a}) \\
&= a(u - \frac{b}{a})^2 + c - \frac{b^2}{a}
\end{aligned}$$

By completing the square, we obtain

$$\begin{aligned} au^2 - 2bu + c &= a(u^2 - 2\frac{b}{a}u + \frac{c}{a}) \\ &= a(u^2 - 2\frac{b}{a}u + (\frac{b}{a})^2 - (\frac{b}{a})^2 + \frac{c}{a}) \\ &= a(u - \frac{b}{a})^2 + c - \frac{b^2}{a} \end{aligned}$$

and thus,

$$\sigma = \frac{1}{a} = \frac{\sigma_0^2\gamma^2}{\gamma^2 + g^2\sigma_0^2}, \quad m = \frac{b}{a} = \frac{\sigma_0^2 g}{\gamma^2 + \sigma_0^2 g^2} y$$

By completing the square, we obtain

$$
\begin{aligned}
au^2 - 2bu + c &= a(u^2 - 2\frac{b}{a}u + \frac{c}{a}) \\
&= a(u^2 - 2\frac{b}{a}u + (\frac{b}{a})^2 - (\frac{b}{a})^2 + \frac{c}{a}) \\
&= a(u - \frac{b}{a})^2 + c - \frac{b^2}{a}
\end{aligned}
$$

and thus,

$$
\sigma = \frac{1}{a} = \frac{\sigma_0^2\gamma^2}{\gamma^2 + g^2\sigma_0^2}\,, \qquad m = \frac{b}{a} = \frac{\sigma_0^2 g}{\gamma^2 + \sigma_0^2 g^2}y
$$

and for the constant K, which does not depend on u, we obtain

$$
K = c - \frac{b^2}{a} = \frac{y^2}{\gamma^2} - \frac{\sigma_0^2 g^2 y^2}{y^4 + y^2 g^2 \sigma^2}\,.
$$

By completing the square, we obtain

$$
\begin{aligned}
au^2 - 2bu + c &= a(u^2 - 2\frac{b}{a}u + \frac{c}{a}) \\
&= a(u^2 - 2\frac{b}{a}u + (\frac{b}{a})^2 - (\frac{b}{a})^2 + \frac{c}{a}) \\
&= a(u - \frac{b}{a})^2 + c - \frac{b^2}{a}
\end{aligned}
$$

and thus,

$$
\sigma = \frac{1}{a} = \frac{\sigma_0^2\gamma^2}{\gamma^2 + g^2\sigma_0^2}\,, \qquad m = \frac{b}{a} = \frac{\sigma_0^2 g}{\gamma^2 + \sigma_0^2 g^2}y
$$

and for the constant K, which does not depend on u, we obtain

$$
K = c - \frac{b^2}{a} = \frac{y^2}{\gamma^2} - \frac{\sigma_0^2 g^2 y^2}{y^4 + y^2 g^2 \sigma^2}\,.
$$

Hence,

$$
\rho^y(u) \propto \exp\Big(-\frac{1}{2(\frac{\sigma_0^2\gamma^2}{\gamma^2+g^2\sigma_0^2})^2}(u - \frac{\sigma_0^2 g}{\gamma^2 + g^2\sigma_0^2}y)^2\Big)\,.
$$

Bayesian Inversion in $\mathbb{R}^n$

1D Gaussian, linear example

Forward response operator $g \in L(\mathbb{R}, \mathbb{R})$

Prior $\mu_0 = \mathcal{N}(0, \sigma_0^2)$, $\sigma_0 \in \mathbb{R}$, Leb. dens. $\rho_0(u) = \frac{1}{\sigma_0\sqrt{2\pi}} \exp(-\frac{\|u\|^2}{\sigma_0^2})$

Noise $\eta \sim \mathcal{N}(0, \gamma^2)$, $\gamma \in \mathbb{R}$, Leb. dens. $\rho(\eta) = \frac{1}{\gamma\sqrt{2\pi}} \exp(-\frac{\|\eta\|^2}{2\gamma^2})$.

Bayes' formula

$$\rho(u|y) = \frac{\rho(y|u)\rho(u)}{\int_{\mathbb{R}} \rho(y|u)\rho(u)\mathrm{d}u}$$

'

Completing the square

$$\mu^y = \mathcal{N}\Big(\frac{\sigma_0^2 g}{\gamma^2 + g^2\sigma_0^2} y, \sigma_0^2 - \frac{\sigma_0^4 g^2}{\gamma^2 + g^2\sigma_0^2}\Big)$$

Estimators

- Maximum a posteriori estimate (MAP)

$$u_{MAP} = \arg\max_{u \in \mathbb{R}^n} \rho^y(u)$$

- Conditional mean (CM)

$$u_{CM} = \mathbb{E}(u|y) = \int_{\mathbb{R}^n} u \rho^y(u) \mathrm{d}u$$

- Maximum likelihood estimate (ML)

$$u_{ML} = \arg\max_{u \in \mathbb{R}^n} \rho(y - \mathcal{G}(u))$$

- Conditional variance (spread estimator)

$$\mathbb{V}(u|y) = \int_{\mathbb{R}^n} (u - u_{CM})(u - u_{CM})^\top \rho^y(u) \mathrm{d}u$$

Estimators

- Maximum a posteriori estimate (MAP)

$$u_{MAP} = \arg\max_{u \in \mathbb{R}^n} \rho^y(u)$$

- Conditional mean (CM)

$$u_{CM} = \mathbb{E}(u|y) = \int_{\mathbb{R}^n} u\rho^y(u)\mathrm{d}u$$

- Maximum likelihood estimate (ML)

$$u_{ML} = \arg\max_{u \in \mathbb{R}^n} \rho(y - \mathcal{G}(u))$$

- Conditional variance (spread estimator)

$$\mathbb{V}(u|y) = \int_{\mathbb{R}^n} (u - u_{CM})(u - u_{CM})^\top \rho^y(u)\mathrm{d}u$$

Estimators

- Maximum a posteriori estimate (MAP)

$$u_{MAP} = \arg\max_{u \in \mathbb{R}^n} \rho^y(u)$$

- Conditional mean (CM)

$$u_{CM} = \mathbb{E}(u|y) = \int_{\mathbb{R}^n} u\rho^y(u)\mathrm{d}u$$

- Maximum likelihood estimate (ML)

$$u_{ML} = \arg\max_{u \in \mathbb{R}^n} \rho(y - \mathcal{G}(u))$$

- Conditional variance (spread estimator)

$$\mathbb{V}(u|y) = \int_{\mathbb{R}^n} (u - u_{CM})(u - u_{CM})^\top \rho^y(u)\mathrm{d}u$$

Estimators

- Maximum a posteriori estimate (MAP)

$$u_{MAP} = \arg\max_{u \in \mathbb{R}^n} \rho^y(u)$$

- Conditional mean (CM)

$$u_{CM} = \mathbb{E}(u|y) = \int_{\mathbb{R}^n} u \rho^y(u) \mathrm{d}u$$

- Maximum likelihood estimate (ML)

$$u_{ML} = \arg\max_{u \in \mathbb{R}^n} \rho(y - \mathcal{G}(u))$$

- Conditional variance (spread estimator)

$$\mathbb{V}(u|y) = \int_{\mathbb{R}^n} (u - u_{CM})(u - u_{CM})^\top \rho^y(u) \mathrm{d}u$$

Example 1: Linear Gaussian Case

Connection of MAP and CM and optimisation

Forward response operator $A \in L(\mathbb{R}^n, \mathbb{R}^J)$

Prior $\mu_0 = \mathcal{N}(m_0, C_0)$

Noise $\eta \sim \mathcal{N}(0, \Gamma)$

Posterior $\mu^y = \mathcal{N}(m, C)$ with
$m = m_0 + C_0 A^*(AC_0A^* + \Gamma)^{-1}(y - Am_0)$ and
$C = C_0 - C_0 A^*(AC_0A^* + \Gamma)^{-1} AC_0$.

Example 1: Linear Gaussian Case

Connection of MAP and CM and optimisation

Forward response operator $A \in L(\mathbb{R}^n, \mathbb{R}^J)$

Prior $\mu_0 = \mathcal{N}(m_0, C_0)$

Noise $\eta \sim \mathcal{N}(0, \Gamma)$

Posterior $\mu^y = \mathcal{N}(m, C)$ with
$m = m_0 + C_0 A^*(AC_0A^* + \Gamma)^{-1}(y - Am_0)$ and
$C = C_0 - C_0A^*(AC_0A^* + \Gamma)^{-1}AC_0$.

- $u_{MAP} = \arg\max_{u \in \mathbb{R}^n} \rho^y(u) = \arg\min_{u \in \mathbb{R}^n} (u-m)^\top C^{-1}(u-m) = m$
- $u_{CM} = \mathbb{E}(u|y) = m$
- $u_{opt} = \arg\min_{u \in \mathbb{R}^n} \|Ax - y\|_\Gamma^2 + \|u - m_0\|_{C_0}^2 = m$

Example 1: Linear Gaussian Case

Connection of MAP and CM and optimisation

Forward response operator $A \in L(\mathbb{R}^n, \mathbb{R}^J)$

Prior $\mu_0 = \mathcal{N}(m_0, C_0)$

Noise $\eta \sim \mathcal{N}(0, \Gamma)$

Posterior $\mu^y = \mathcal{N}(m, C)$ with
$m = m_0 + C_0 A^*(AC_0A^* + \Gamma)^{-1}(y - Am_0)$ and
$C = C_0 - C_0 A^*(AC_0A^* + \Gamma)^{-1} AC_0$.

- $u_{MAP} = \arg\max_{u \in \mathbb{R}^n} \rho^y(u) = \arg\min_{u \in \mathbb{R}^n} (u-m)^\top C^{-1}(u-m) = m$
- $u_{CM} = \mathbb{E}(u|y) = m$
- $u_{opt} = \arg\min_{u \in \mathbb{R}^n} \|Ax - y\|_\Gamma^2 + \|u - m_0\|_{C_0}^2 = m$

Example 1: Linear Gaussian Case

Connection of MAP and CM and optimisation

Forward response operator $A \in L(\mathbb{R}^n, \mathbb{R}^J)$

Prior $\mu_0 = \mathcal{N}(m_0, C_0)$

Noise $\eta \sim \mathcal{N}(0, \Gamma)$

Posterior $\mu^y = \mathcal{N}(m, C)$ with
$m = m_0 + C_0 A^*(AC_0A^* + \Gamma)^{-1}(y - Am_0)$ and
$C = C_0 - C_0A^*(AC_0A^* + \Gamma)^{-1}AC_0$.

- $u_{MAP} = \arg\max_{u\in\mathbb{R}^n} \rho^y(u) = \arg\min_{u\in\mathbb{R}^n}(u-m)^\top C^{-1}(u-m) = m$
- $u_{CM} = \mathbb{E}(u|y) = m$
- $u_{opt} = \arg\min_{u\in\mathbb{R}^n} \|Ax - y\|_\Gamma^2 + \|u - m_0\|_{C_0}^2 = m$

Example 2: Comparison of CM and MAP (1D)

Let u be a real-valued rv and assume that the posterior density is given by

$$\rho^y(u) = \frac{\alpha}{\sigma_0}\varphi\Big(\frac{u}{\sigma_0}\Big) + \frac{1-\alpha}{\sigma_1}\varphi\Big(\frac{u-1}{\sigma_1}\Big)$$

with $0 < \alpha < 1,\ \sigma_0, \sigma_1 > 0$ and $\varphi(u) = \frac{1}{\sqrt{2\pi}}\exp(-u^2/2)$.

Example 2: Comparison of CM and MAP (1D)

Let u be a real-valued rv and assume that the posterior density is given by

$$\rho^y(u) = \frac{\alpha}{\sigma_0}\varphi\Big(\frac{u}{\sigma_0}\Big) + \frac{1-\alpha}{\sigma_1}\varphi\Big(\frac{u-1}{\sigma_1}\Big)$$

with $0 < \alpha < 1,\ \sigma_0, \sigma_1 > 0$ and $\varphi(u) = \frac{1}{\sqrt{2\pi}}\exp(-u^2/2)$.

Estimators

- $u_{CM} = 1 - \alpha$
- $u_{MAP} = \begin{cases} 0,\ \text{if } \alpha/\sigma_0 > (1-\alpha)/\sigma_1 \\ 1,\ \text{if } \alpha/\sigma_0 < (1-\alpha)/\sigma_1 \\ \{0,1\},\ \text{if } \alpha/\sigma_0 = (1-\alpha)/\sigma_1 \end{cases}$

Example 2: Comparison of CM and MAP (1D)

Let u be a real-valued rv and assume that the posterior density is given by

$$\rho^y(u) = \frac{\alpha}{\sigma_0}\varphi\Big(\frac{u}{\sigma_0}\Big) + \frac{1-\alpha}{\sigma_1}\varphi\Big(\frac{u-1}{\sigma_1}\Big)$$

with $0 < \alpha < 1,\ \sigma_0, \sigma_1 > 0$ and $\varphi(u) = \frac{1}{\sqrt{2\pi}}\exp(-u^2/2)$.

Estimators

- $u_{CM} = 1 - \alpha$
- $u_{MAP} = \begin{cases} 0, \text{ if } \alpha/\sigma_0 > (1-\alpha)/\sigma_1 \\ 1, \text{ if } \alpha/\sigma_0 < (1-\alpha)/\sigma_1 \\ \{0,1\}, \text{ if } \alpha/\sigma_0 = (1-\alpha)/\sigma_1 \end{cases}$

Example 2: Comparison of CM and MAP (1D)

Let u be a real-valued rv and assume that the posterior density is given by

$$\rho^y(u) = \frac{\alpha}{\sigma_0}\varphi\Big(\frac{u}{\sigma_0}\Big) + \frac{1-\alpha}{\sigma_1}\varphi\Big(\frac{u-1}{\sigma_1}\Big)$$

with $0 < \alpha < 1,\ \sigma_0, \sigma_1 > 0$ and $\varphi(u) = \frac{1}{\sqrt{2\pi}}\exp(-u^2/2)$.

Estimators

- $u_{CM} = 1 - \alpha$
- $u_{MAP} = \begin{cases} 0, \text{ if } \alpha/\sigma_0 > (1-\alpha)/\sigma_1 \\ 1, \text{ if } \alpha/\sigma_0 < (1-\alpha)/\sigma_1 \\ \{0,1\}, \text{ if } \alpha/\sigma_0 = (1-\alpha)/\sigma_1 \end{cases}$

Example 2: Comparison of CM and MAP (1D)

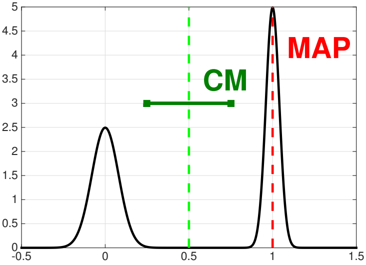


Figure: Posterior density with parameter values $\alpha = 1/2,\ \sigma_0 = 0.08,\ \sigma_1 = 0.04.$

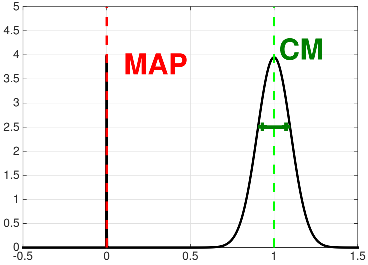


Figure: Posterior density with parameter values $\alpha = 0.01,\ \sigma_0 = 0.001,\ \sigma_1 = 0.1.$

Motivation for the Infinite Setting

- Many inverse problems in the physical sciences require the determination of an unknown field from a finite set of indirect measurements.

- The Bayesian approach leads to a natural well-posedness and stability theory.

- This framework provides a way of deriving and developing efficient algorithms with convergence rates independent of the number of input uncertainties.

Motivation for the Infinite Setting

- Many inverse problems in the physical sciences require the determination of an unknown field from a finite set of indirect measurements.
- The Bayesian approach leads to a natural well-posedness and stability theory.
- This framework provides a way of deriving and developing efficient algorithms with convergence rates independent of the number of input uncertainties.

Motivation for the Infinite Setting

- Many inverse problems in the physical sciences require the determination of an unknown field from a finite set of indirect measurements.
- The Bayesian approach leads to a natural well-posedness and stability theory.
- This framework provides a way of deriving and developing efficient algorithms with convergence rates independent of the number of input uncertainties.

Motivation for the Infinite Setting

We consider steady groundwater flow in a 2D confined aquifer governed by

$$-\nabla \cdot u \nabla q = f$$

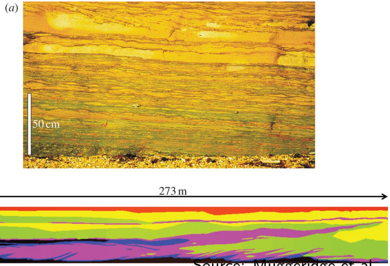
with piezometric head q, source f and hydraulic conductivity u.

Uncertainty in the hydraulic conductivity u

- Typical Models: log-normal prior or multipoint prior

Measurements

- Measurements $q(x_j)$ for some set of points $\{x_j\}_{j=1}^K$ in the physical domain



Source: Muggeridge et al.

Prior Modeling - The General Setting

- The parameter space X is a separable Banach space.
- The physical domain $D \subset \mathbb{R}^d$ is a bounded, open set with Lipschitz boundary.
- $\{\phi_j\}_{j\in\mathbb{N}}$ denotes an infinite sequence in X with norm $\|\cdot\|$ of $\mathbb{R}$-valued functions on the physical domain D with $\|\phi_j\| = 1$ for $j = 1, \dots, \infty$.
- A mean function $u_0 \in X$ is assumed to be given.
- $$u = u_0 + \sum_{j=1}^{\infty} u_j \phi_j$$
- $\{u_j\}_{j\in\mathbb{N}}$ is a sequence of real-valued random functions.
- $\gamma = \{\gamma_j\}_{j\in\mathbb{N}}$ is a deterministic sequence with $\gamma \in l^p$ for some $p > 0$ and $\zeta = \{\zeta_j\}_{j\in\mathbb{N}}$ is an i.i.d. sequence of centred random variables.
- $$u = u_0 + \sum_{j=1}^{\infty} \gamma_j \phi_j \zeta_j$$

Prior Modeling - The General Setting

- The parameter space X is a separable Banach space.
- The physical domain $D \subset \mathbb{R}^d$ is a bounded, open set with Lipschitz boundary.
- $\{\phi_j\}_{j\in\mathbb{N}}$ denotes an infinite sequence in X with norm $\|\cdot\|$ of $\mathbb{R}$-valued functions on the physical domain D with $\|\phi_j\| = 1$ for $j = 1, \ldots, \infty$.
- A mean function $u_0 \in X$ is assumed to be given.
- $$u = u_0 + \sum_{j=1}^{\infty} u_j \phi_j$$
- $\{u_j\}_{j\in\mathbb{N}}$ is a sequence of real-valued random functions.
- $\gamma = \{\gamma_j\}_{j\in\mathbb{N}}$ is a deterministic sequence with $\gamma \in l^p$ for some $p > 0$ and $\zeta = \{\zeta_j\}_{j\in\mathbb{N}}$ is an i.i.d. sequence of centred random variables.
- $$u = u_0 + \sum_{j=1}^{\infty} \gamma_j \phi_j \zeta_j$$

Prior Modeling - The General Setting

- The parameter space X is a separable Banach space.
- The physical domain $D \subset \mathbb{R}^d$ is a bounded, open set with Lipschitz boundary.
- $\{\phi_j\}_{j\in\mathbb{N}}$ denotes an infinite sequence in X with norm $\|\cdot\|$ of $\mathbb{R}$-valued functions on the physical domain D with $\|\phi_j\| = 1$ for $j = 1, \ldots, \infty$.
- A mean function $u_0 \in X$ is assumed to be given.
- $$u = u_0 + \sum_{j=1}^{\infty} u_j \phi_j$$
- $\{u_j\}_{j\in\mathbb{N}}$ is a sequence of real-valued random functions.
- $\gamma = \{\gamma_j\}_{j\in\mathbb{N}}$ is a deterministic sequence with $\gamma \in l^p$ for some $p > 0$ and $\zeta = \{\zeta_j\}_{j\in\mathbb{N}}$ is an i.i.d. sequence of centred random variables.
- $$u = u_0 + \sum_{j=1}^{\infty} \gamma_j \phi_j \zeta_j$$

Prior Modeling - The General Setting

- The parameter space X is a separable Banach space.
- The physical domain $D \subset \mathbb{R}^d$ is a bounded, open set with Lipschitz boundary.
- $\{\phi_j\}_{j \in \mathbb{N}}$ denotes an infinite sequence in X with norm $\|\cdot\|$ of $\mathbb{R}$-valued functions on the physical domain D with $\|\phi_j\| = 1$ for $j = 1, \ldots, \infty$.
- A mean function $u_0 \in X$ is assumed to be given.
- $$u = u_0 + \sum_{j=1}^{\infty} u_j \phi_j$$
- $\{u_j\}_{j \in \mathbb{N}}$ is a sequence of real-valued random functions.
- $\gamma = \{\gamma_j\}_{j \in \mathbb{N}}$ is a deterministic sequence with $\gamma \in l^p$ for some $p > 0$ and $\zeta = \{\zeta_j\}_{j \in \mathbb{N}}$ is an i.i.d. sequence of centred random variables.
- $$u = u_0 + \sum_{j=1}^{\infty} \gamma_j \phi_j \zeta_j$$

Prior Modeling - The General Setting

- The parameter space X is a separable Banach space.
- The physical domain $D \subset \mathbb{R}^d$ is a bounded, open set with Lipschitz boundary.
- $\{\phi_j\}_{j\in\mathbb{N}}$ denotes an infinite sequence in X with norm $\|\cdot\|$ of $\mathbb{R}$-valued functions on the physical domain D with $\|\phi_j\| = 1$ for $j = 1, \dots, \infty$.
- A mean function $u_0 \in X$ is assumed to be given.
- $$u = u_0 + \sum_{j=1}^{\infty} u_j \phi_j$$
- $\{u_j\}_{j\in\mathbb{N}}$ is a sequence of real-valued random functions.
- $\gamma = \{\gamma_j\}_{j\in\mathbb{N}}$ is a deterministic sequence with $\gamma \in l^p$ for some $p > 0$ and $\zeta = \{\zeta_j\}_{j\in\mathbb{N}}$ is an i.i.d. sequence of centred random variables.
- $$u = u_0 + \sum_{j=1}^{\infty} \gamma_j \phi_j \zeta_j$$

Prior Modeling - The General Setting

- The parameter space X is a separable Banach space.
- The physical domain $D \subset \mathbb{R}^d$ is a bounded, open set with Lipschitz boundary.
- $\{\phi_j\}_{j\in\mathbb{N}}$ denotes an infinite sequence in X with norm $\|\cdot\|$ of $\mathbb{R}$-valued functions on the physical domain D with $\|\phi_j\| = 1$ for $j = 1, \dots, \infty$.
- A mean function $u_0 \in X$ is assumed to be given.
- $$u = u_0 + \sum_{j=1}^{\infty} u_j \phi_j$$
- $\{u_j\}_{j\in\mathbb{N}}$ is a sequence of real-valued random functions.
- $\gamma = \{\gamma_j\}_{j\in\mathbb{N}}$ is a deterministic sequence with $\gamma \in l^p$ for some $p > 0$ and $\zeta = \{\zeta_j\}_{j\in\mathbb{N}}$ is an i.i.d. sequence of centred random variables.
- $$u = u_0 + \sum_{j=1}^{\infty} \gamma_j \phi_j \zeta_j$$

Prior Modeling - The General Setting

- The parameter space X is a separable Banach space.
- The physical domain $D \subset \mathbb{R}^d$ is a bounded, open set with Lipschitz boundary.
- $\{\phi_j\}_{j\in\mathbb{N}}$ denotes an infinite sequence in X with norm $\|\cdot\|$ of $\mathbb{R}$-valued functions on the physical domain D with $\|\phi_j\| = 1$ for $j = 1, \ldots, \infty$.
- A mean function $u_0 \in X$ is assumed to be given.
- $$u = u_0 + \sum_{j=1}^{\infty} u_j \phi_j$$
- $\{u_j\}_{j\in\mathbb{N}}$ is a sequence of real-valued random functions.
- $\gamma = \{\gamma_j\}_{j\in\mathbb{N}}$ is a deterministic sequence with $\gamma \in l^p$ for some $p > 0$ and $\zeta = \{\zeta_j\}_{j\in\mathbb{N}}$ is an i.i.d. sequence of centred random variables.
- $$u = u_0 + \sum_{j=1}^{\infty} \gamma_j \phi_j \zeta_j$$

Prior Modeling - The General Setting

- The parameter space X is a separable Banach space.
- The physical domain $D \subset \mathbb{R}^d$ is a bounded, open set with Lipschitz boundary.
- $\{\phi_j\}_{j\in\mathbb{N}}$ denotes an infinite sequence in X with norm $\|\cdot\|$ of $\mathbb{R}$-valued functions on the physical domain D with $\|\phi_j\| = 1$ for $j = 1, \ldots, \infty$.
- A mean function $u_0 \in X$ is assumed to be given.
- $$u = u_0 + \sum_{j=1}^{\infty} u_j \phi_j$$
- $\{u_j\}_{j\in\mathbb{N}}$ is a sequence of real-valued random functions.
- $\gamma = \{\gamma_j\}_{j\in\mathbb{N}}$ is a deterministic sequence with $\gamma \in l^p$ for some $p > 0$ and $\zeta = \{\zeta_j\}_{j\in\mathbb{N}}$ is an i.i.d. sequence of centred random variables.
- $$u = u_0 + \sum_{j=1}^{\infty} \gamma_j \phi_j \zeta_j$$

Uniform Priors

$$u(x,\omega) := u_0(x) + \sum_{j\in\mathbb{N}} \zeta_j(\omega)\gamma_j\phi_j(x)$$

- $\zeta = (\zeta_j)_{j\in\mathbb{N}}$ iid sequence of real-valued random variables $\zeta_j \sim \mathcal{U}[-1,1]$
- $u_0, \phi_j \in X$ with $X = L^\infty(D)$ (work with a separable space X' found as the closure of the linear span of the functions $(u_0, \{\phi_j\}_{j\in\mathbb{N}}$ w.r. to the $\|\cdot\|_\infty$ on X).
- $\gamma = \{\gamma_j\}_{j\in\mathbb{N}}$ is a deterministic sequence with $\gamma \in l^p$ for $0 < p \leq 1$.

Uniform Priors

$$u(x,\omega) := u_0(x) + \sum_{j\in\mathbb{N}} \zeta_j(\omega)\gamma_j\phi_j(x)$$

- $\zeta = (\zeta_j)_{j\in\mathbb{N}}$ iid sequence of real-valued random variables $\zeta_j \sim \mathcal{U}[-1,1]$
- $u_0, \phi_j \in X$ with $X = L^\infty(D)$ (work with a separable space X' found as the closure of the linear span of the functions $(u_0, \{\phi_j\}_{j\in\mathbb{N}}$ w.r. to the $\|\cdot\|_\infty$ on X).
- $\gamma = \{\gamma_j\}_{j\in\mathbb{N}}$ is a deterministic sequence with $\gamma \in l^p$ for $0 < p \le 1$.

We assume that there exists strictly positive constants $\delta > 0$ and $0 < u_{\min} \le u_{\max} < \infty$ such that

$$\begin{aligned} \operatorname*{ess\,inf}_{x\in D} u_0(x) &\ge u_{\min} \qquad \operatorname*{ess\,sup}_{x\in D} u_0(x) \le u_{\max} \\ \sum_{j\in\mathbb{N}} |\gamma_j(x)| &= \frac{\delta}{1+\delta} u_{\min} \,. \end{aligned}$$

Uniform Priors

$$u(x,\omega) := u_0(x) + \sum_{j\in\mathbb{N}} \zeta_j(\omega)\gamma_j\phi_j(x)$$

- $\zeta = (\zeta_j)_{j\in\mathbb{N}}$ iid sequence of real-valued random variables $\zeta_j \sim \mathcal{U}[-1,1]$
- $u_0, \phi_j \in X$ with $X = L^\infty(D)$ (work with a separable space X' found as the closure of the linear span of the functions $(u_0, \{\phi_j\}_{j\in\mathbb{N}}$ w.r. to the $\|\cdot\|_\infty$ on X).
- $\gamma = \{\gamma_j\}_{j\in\mathbb{N}}$ is a deterministic sequence with $\gamma \in l^p$ for $0 < p \leq 1$.

The following holds $\mathbb{P}$-a.s.: The function u satisfies

$$\frac{1}{1+\delta}u_{\min} \leq u(x) \leq u_{\max} + \frac{\delta}{1+\delta}u_{\min} \quad a.e.\ x \in D\ ,$$

i.e. $\mu_0(X') = 1$.

Uniform Priors

Example: 1d $u(x,\omega) = 1 + \sum_{j=1}^{10} 0.25 \frac{1}{j^2} \Xi_{D_j} \zeta_j(\omega)$ with $D_j = [(j-1)\frac{1}{10}, j\frac{1}{10}]$.

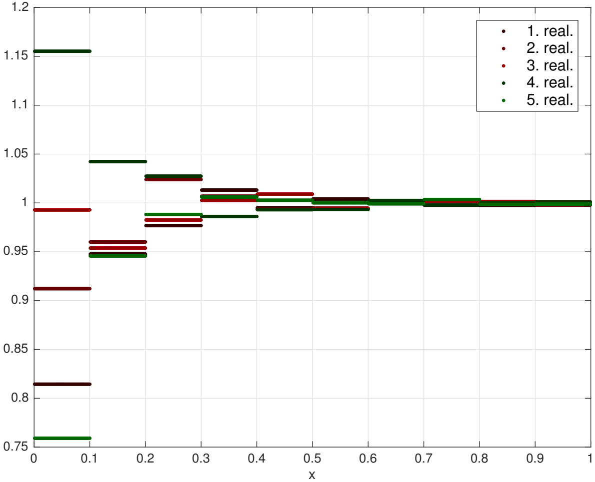


Figure: 5 realization of the random field u.

Uniform Priors

Example: 2d $u((x_1, x_2), \omega) = 1 + \sum_{j=1}^{64} 0.1 \frac{1}{j^2} (\sin(2\pi j x_1) + \cos(2\pi j x_2)) \zeta_j(\omega)$.

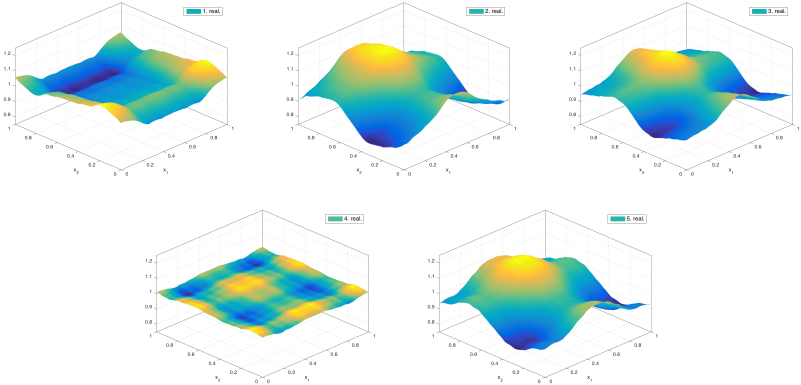


Figure: 5 realization of the random field u.

Uniform Priors

Groundwater flow problem - Variational formulation

$$\int_D u(x,\omega)\nabla q(x,\omega)\cdot\nabla v(x)\mathrm{d}x = \int_D f(x)v(x)\mathrm{d}x\ , \qquad v\in V = H_0^1(D)$$

Existence and uniqueness of $q(\cdot,\omega)$ ensured by UEA and Lax-Milgram. In particular, we have P-a.s.

$$\|q\|_V \le \frac{1}{\tilde{u}_{\min}}\|f\|_{V^*}$$

with $\tilde{u}_{\min} = \frac{1}{1+\delta}u_{\min}$ and with the measurability of q (as a direct consequence of the measurability of u and the local Lipschitz continuity of the solution map, $q\in L^r(\Omega,V)$ for all $1\le r\le\infty$.

Gaussian Priors

$$u(x,\omega) := u_0(x) + \sum_{j\in\mathbb{N}} \zeta_j(\omega)\gamma_j\phi_j(x)$$

- $\zeta = (\zeta_j)_{j\in\mathbb{N}}$ iid sequence of real-valued random variables $\zeta_j \sim \mathcal{N}(0,1)$
- X is a Hilbert space and $\{\phi_j\}_{j\in\mathbb{N}}$ is an orthonormal basis of H.
- $\gamma = \{\gamma_j\}_{j\in\mathbb{N}}$ is a deterministic sequence and u generates random draws from the Gaussian measure $\mathcal{N}(u_0, \mathcal{C})$.
- $\rightarrow$ Karhunen-Loève expansion.

Gaussian Priors

Random Field Perspective

- A random field $u : D \times \Omega \to \mathbb{R}$ is Gaussian if all finite dimensional distributions are Gaussian, i.e. for any $x_1, \ldots, x_n \in D$, the random vector $z := (u(x_1, \cdot), \ldots, u(x_n, \cdot))$ has a multivariate Gaussian distribution with
  - mean $\mu = \mathbb{E}[u(x_1, \cdot), \ldots, u(x_n, \cdot)]$
  - covariance matrix $C = (\mathbb{C}ov_u(x_i, x_j))_{i,j=1}^n$ and
  - probability density function

$$\rho(z) = \frac{1}{(2\pi)^{d/2}\sqrt{\det(C)}} \exp(-\frac{1}{2}(z-\mu)^\top C^{-1}(z-\mu)) \, .$$

- Consider the Karhunen-Loève expansion

$$u(x, \omega) = \mathbb{E}[u](x) + \sum_{j=1}^{\infty} \sqrt{\gamma_j}\phi_j(x)\zeta_j(\omega) \, ,$$

where γ_j and ϕ_j for $j = 1, \ldots, n$ are the eigenvalues and eigenfunctions of the covariance operator $\mathcal{C}$. The random variables $\zeta_j, \; j = 1, \ldots, n$ are independent $\mathcal{N}(0, 1)$ random variables.

Gaussian Priors

Random Field Perspective

- A random field $u: D \times \Omega \to \mathbb{R}$ is Gaussian if all finite dimensional distributions are Gaussian, i.e. for any $x_1, \ldots, x_n \in D$, the random vector $z := (u(x_1, \cdot), \ldots, u(x_n, \cdot))$ has a multivariate Gaussian distribution with
  - mean $\mu = \mathbb{E}[u(x_1, \cdot), \ldots, u(x_n, \cdot)]$
  - covariance matrix $C = (\mathbb{C}ov_u(x_i, x_j))_{i,j=1}^n$ and
  - probability density function

$$\rho(z) = \frac{1}{(2\pi)^{d/2}\sqrt{\det(C)}} \exp(-\frac{1}{2}(z-\mu)^\top C^{-1}(z-\mu)) \, .$$

- Consider the Karhunen-Loève expansion

$$u(x, \omega) = \mathbb{E}[u](x) + \sum_{j=1}^{\infty} \sqrt{\gamma_j}\phi_j(x)\zeta_j(\omega) \, ,$$

where γ_j and ϕ_j for $j = 1, \ldots, n$ are the eigenvalues and eigenfunctions of the covariance operator $\mathcal{C}$. The random variables $\zeta_j, \; j = 1, \ldots, n$ are independent $\mathcal{N}(0,1)$ random variables.

Gaussian Priors

Random Field Perspective

- A random field $u : D \times \Omega \to \mathbb{R}$ is Gaussian if all finite dimensional distributions are Gaussian, i.e. for any $x_1, \ldots, x_n \in D$, the random vector $z := (u(x_1, \cdot), \ldots, u(x_n, \cdot))$ has a multivariate Gaussian distribution with
  - mean $\mu = \mathbb{E}[u(x_1, \cdot), \ldots, u(x_n, \cdot)]$
  - covariance matrix $C = (\mathbb{C}ov_u(x_i, x_j))_{i,j=1}^n$ and
  - probability density function

$$\rho(z) = \frac{1}{(2\pi)^{d/2}\sqrt{\det(C)}} \exp(-\frac{1}{2}(z-\mu)^\top C^{-1}(z-\mu)) .$$

- Consider the Karhunen-Loève expansion

$$u(x, \omega) = \mathbb{E}[u](x) + \sum_{j=1}^{\infty} \sqrt{\gamma_j}\phi_j(x)\zeta_j(\omega) ,$$

where γ_j and ϕ_j for $j = 1, \ldots, n$ are the eigenvalues and eigenfunctions of the covariance operator $\mathcal{C}$. The random variables ζ_j, $j = 1, \ldots, n$ are independent $\mathcal{N}(0, 1)$ random variables.

Gaussian Priors

Random Field Perspective

- A random field $u : D \times \Omega \to \mathbb{R}$ is Gaussian if all finite dimensional distributions are Gaussian, i.e. for any $x_1, \dots, x_n \in D$, the random vector $z := (u(x_1, \cdot), \dots, u(x_n, \cdot))$ has a multivariate Gaussian distribution with
  - mean $\mu = \mathbb{E}[u(x_1, \cdot), \dots, u(x_n, \cdot)]$
  - covariance matrix $C = (\mathbb{C}ov_u(x_i, x_j))_{i,j=1}^n$ and
  - probability density function

$$\rho(z) = \frac{1}{(2\pi)^{d/2}\sqrt{\det(C)}} \exp(-\frac{1}{2}(z-\mu)^\top C^{-1}(z-\mu)) \, .$$

- Consider the Karhunen-Loève expansion

$$u(x, \omega) = \mathbb{E}[u](x) + \sum_{j=1}^{\infty} \sqrt{\gamma_j}\phi_j(x)\zeta_j(\omega) \, ,$$

where γ_j and ϕ_j for $j = 1, \dots, n$ are the eigenvalues and eigenfunctions of the covariance operator $\mathcal{C}$. The random variables $\zeta_j, \; j = 1, \dots, n$ are independent $\mathcal{N}(0, 1)$ random variables.

Gaussian Priors

Random Field Perspective

- A random field $u : D \times \Omega \to \mathbb{R}$ is Gaussian if all finite dimensional distributions are Gaussian, i.e. for any $x_1, \dots, x_n \in D$, the random vector $z := (u(x_1, \cdot), \dots, u(x_n, \cdot))$ has a multivariate Gaussian distribution with
  - mean $\mu = \mathbb{E}[u(x_1, \cdot), \dots, u(x_n, \cdot)]$
  - covariance matrix $C = (\mathbb{C}ov_u(x_i, x_j))_{i,j=1}^n$ and
  - probability density function

$$\rho(z) = \frac{1}{(2\pi)^{d/2}\sqrt{\det(C)}} \exp(-\frac{1}{2}(z-\mu)^\top C^{-1}(z-\mu)) \, .$$

- Consider the Karhunen-Loève expansion

$$u(x,\omega) = \mathbb{E}[u](x) + \sum_{j=1}^{\infty} \sqrt{\gamma_j}\phi_j(x)\zeta_j(\omega) \, ,$$

  where γ_j and ϕ_j for $j = 1, \dots, n$ are the eigenvalues and eigenfunctions of the covariance operator $\mathcal{C}$. The random variables $\zeta_j, \ j = 1, \dots, n$ are independent $\mathcal{N}(0,1)$ random variables.

Gaussian Priors

Let u be a sample from the measure $\mu_0 = \mathcal{N}(0, \mathcal{C})$ where $\mathcal{C} = (-\Delta)^{-s}$ and $s > \frac{d}{2}$. Δ denotes the Laplacian with homogeneous Dirichlet conditions.

Then, draws from μ_0 are almost surely in $C(D)$, i.e. $\mu_0(X) = 1$ with $X = C(D)$.

Gaussian Priors

Example: 2d, Gaussian RF with zero mean and $C = (-\Delta)^{-2}$.

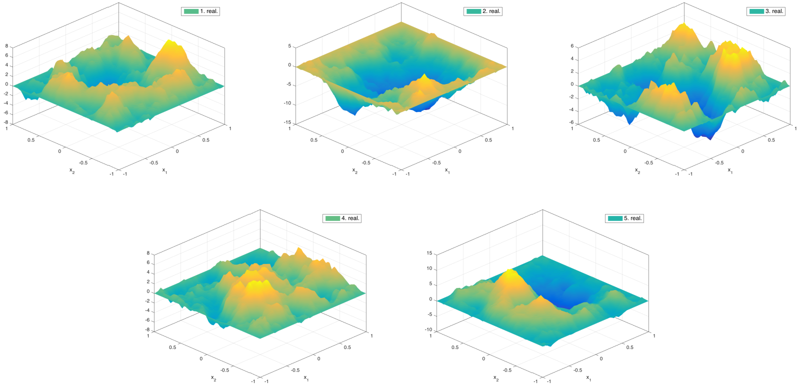


Figure: 5 realization of the random field u.

Gaussian Priors

$$u(x,\omega) = \mathbb{E}[u](x) + \sum_{j=1}^{\infty} \sqrt{\gamma_j}\phi_j(x)\zeta_j(\omega)$$

- $\zeta = (\zeta_j)_{j\in\mathbb{N}}$ iid sequence of real-valued random variables $\zeta_j \sim \mathcal{N}(0,1)$
- $(\gamma_j, \phi_j)_{j\in\mathbb{N}}$ sequence of eigenpairs (in descending order, $\|\phi_j\|_{L^2(D)} = 1$).

Groundwater flow problem - Variational formulation

$$\int_D \exp(u(x,\omega))\nabla q(x,\omega)\cdot\nabla v(x)\mathrm{d}x = \int_D f(x)v(x)\mathrm{d}x\ , \qquad v \in V = H_0^1(D)$$

$0 < u_{\min}(\omega) \leq \exp(u(x,\omega)) \leq u_{\max}(\omega) < \infty$, for almost all $x \in D,\ \omega \in \Omega$

Bayes' Theorem for Inverse Problems

Find the unknown data $u \in X$ from noisy observations

$$y = \mathcal{G}(u) + \eta$$

- $u \sim \mu_0$ measure on X.
- $\mathcal{G} : X \to Y$ measurable function with X, Y separable Banach spaces.
- $\eta \sim \mu_\eta$ measure on Y, $\eta \perp u$.
- $y|u$ is then distributed according to measure μ_u, the translate of μ_η by $\mathcal{G}(u)$.
- Assume $\mu_u \ll \mu_\eta$ for u μ_0-a.s., thus

$$\frac{\mathrm{d}\mu_u}{\mathrm{d}\mu_\eta}(y) = \exp(-\Phi(u; y))$$

  for some potential $\Phi : X \times Y \to \mathbb{R}$. $\Rightarrow$ $\Phi(u; \cdot)$ is measurable and $\mathbb{E}_{\mu_\eta} \exp(-\Phi(u; y)) = 1$.
- ν_0 is the product measure $\nu_0(\mathrm{d}u, \mathrm{d}y) = \mu_0(\mathrm{d}u)\mu_\eta(\mathrm{d}y)$

Bayes' Theorem for Inverse Problems

Find the unknown data $u \in X$ from noisy observations

$$y = \mathcal{G}(u) + \eta$$

- $u \sim \mu_0$ measure on X.
- $\mathcal{G} : X \to Y$ measurable function with X, Y separable Banach spaces.
- $\eta \sim \mu_\eta$ measure on Y, $\eta \perp u$.
- $y|u$ is then distributed according to measure μ_u, the translate of μ_η by $\mathcal{G}(u)$.
- Assume $\mu_u \ll \mu_\eta$ for u μ_0-a.s., thus

  $$\frac{\mathrm{d}\mu_u}{\mathrm{d}\mu_\eta}(y) = \exp(-\Phi(u; y))$$

  for some potential $\Phi : X \times Y \to \mathbb{R}$. $\Rightarrow$ $\Phi(u; \cdot)$ is measurable and $\mathbb{E}_{\mu_\eta} \exp(-\Phi(u; y)) = 1$.
- ν_0 is the product measure $\nu_0(\mathrm{d}u, \mathrm{d}y) = \mu_0(\mathrm{d}u)\mu_\eta(\mathrm{d}y)$

Bayes' Theorem for Inverse Problems

Find the unknown data $u \in X$ from noisy observations

$$y = \mathcal{G}(u) + \eta$$

- $u \sim \mu_0$ measure on X.
- $\mathcal{G} : X \to Y$ measurable function with X, Y separable Banach spaces.
- $\eta \sim \mu_\eta$ measure on Y, $\eta \perp u$.
- $y|u$ is then distributed according to measure μ_u, the translate of μ_η by $\mathcal{G}(u)$.
- Assume $\mu_u \ll \mu_\eta$ for u μ_0-a.s., thus

$$\frac{\mathrm{d}\mu_u}{\mathrm{d}\mu_\eta}(y) = \exp(-\Phi(u; y))$$

  for some potential $\Phi : X \times Y \to \mathbb{R}$. $\Rightarrow \Phi(u; \cdot)$ is measurable and $\mathbb{E}_{\mu_\eta} \exp(-\Phi(u; y)) = 1$.
- ν_0 is the product measure $\nu_0(\mathrm{d}u, \mathrm{d}y) = \mu_0(\mathrm{d}u)\mu_\eta(\mathrm{d}y)$

Bayes' Theorem for Inverse Problems

Find the unknown data $u \in X$ from noisy observations

$$y = \mathcal{G}(u) + \eta$$

- $u \sim \mu_0$ measure on X.
- $\mathcal{G} : X \to Y$ measurable function with X, Y separable Banach spaces.
- $\eta \sim \mu_\eta$ measure on Y, $\eta \perp u$.
- $y|u$ is then distributed according to measure μ_u, the translate of μ_η by $\mathcal{G}(u)$.
- Assume $\mu_u \ll \mu_\eta$ for u μ_0-a.s., thus

$$\frac{\mathrm{d}\mu_u}{\mathrm{d}\mu_\eta}(y) = \exp(-\Phi(u; y))$$

  for some potential $\Phi : X \times Y \to \mathbb{R}$. $\Rightarrow \Phi(u; \cdot)$ is measurable and $\mathbb{E}_{\mu_\eta} \exp(-\Phi(u; y)) = 1$.
- ν_0 is the product measure $\nu_0(\mathrm{d}u, \mathrm{d}y) = \mu_0(\mathrm{d}u)\mu_\eta(\mathrm{d}y)$

Bayes' Theorem for Inverse Problems

Find the unknown data $u \in X$ from noisy observations

$$y = \mathcal{G}(u) + \eta$$

- $u \sim \mu_0$ measure on X.
- $\mathcal{G} : X \to Y$ measurable function with X, Y separable Banach spaces.
- $\eta \sim \mu_\eta$ measure on Y, $\eta \perp u$.
- $y|u$ is then distributed according to measure μ_u, the translate of μ_η by $\mathcal{G}(u)$.
- Assume $\mu_u \ll \mu_\eta$ for u μ_0-a.s., thus

$$\frac{\mathrm{d}\mu_u}{\mathrm{d}\mu_\eta}(y) = \exp(-\Phi(u; y))$$

  for some potential $\Phi : X \times Y \to \mathbb{R}$. $\Rightarrow \Phi(u; \cdot)$ is measurable and $\mathbb{E}_{\mu_\eta} \exp(-\Phi(u; y)) = 1$.
- ν_0 is the product measure $\nu_0(\mathrm{d}u, \mathrm{d}y) = \mu_0(\mathrm{d}u)\mu_\eta(\mathrm{d}y)$

Bayes' Theorem for Inverse Problems

Find the unknown data $u \in X$ from noisy observations

$$y = \mathcal{G}(u) + \eta$$

- $u \sim \mu_0$ measure on X.
- $\mathcal{G} : X \to Y$ measurable function with X, Y separable Banach spaces.
- $\eta \sim \mu_\eta$ measure on Y, $\eta \perp u$.
- $y|u$ is then distributed according to measure μ_u, the translate of μ_η by $\mathcal{G}(u)$.
- Assume $\mu_u \ll \mu_\eta$ for u μ_0-a.s., thus

  $$\frac{\mathrm{d}\mu_u}{\mathrm{d}\mu_\eta}(y) = \exp(-\Phi(u; y))$$

  for some potential $\Phi : X \times Y \to \mathbb{R}$. $\Rightarrow \Phi(u; \cdot)$ is measurable and $\mathbb{E}_{\mu_\eta} \exp(-\Phi(u; y)) = 1$.
- ν_0 is the product measure $\nu_0(\mathrm{d}u, \mathrm{d}y) = \mu_0(\mathrm{d}u)\mu_\eta(\mathrm{d}y)$

Bayes' Theorem for Inverse Problems

Find the unknown data $u \in X$ from noisy observations

$$y = \mathcal{G}(u) + \eta$$

Bayes' Theorem

Assume that $\Phi : X \times Y \to \mathbb{R}$ is ν_0 measurable and that for y μ_η-a.s.,

$$Z := \int_X \exp(-\Phi(u; y))\mu_0(\mathrm{d}u) > 0 \, .$$

Then, the conditional distribution of $u|y$ exists under $\nu(\mathrm{d}u, \mathrm{d}y) = \mu_0(\mathrm{d}u)\mu_u(\mathrm{d}y)$, and is denoted by μ^y. Furthermore $\mu^y \ll \mu_0$ and

$$\frac{\mathrm{d}\mu^y}{\mathrm{d}\mu_0}(u) = \frac{1}{Z} \exp(-\Phi(u; y))$$

Bayes' Theorem for Inverse Problems

Find the unknown data $u \in X$ from noisy observations

$$y = \mathcal{G}(u) + \eta$$

- Define a suitable prior measure μ_0 and noise measure μ_η.
- Determine the potential Φ such that $\frac{\mathrm{d}\mu_u}{\mathrm{d}\mu_\eta}(y) = \exp(-\Phi(u; y))$.
- Show that Φ is ν_0 measurable.
- Show that the normalization constant is Z is positive almost surely w.r. to y.

Groundwaterflow problem

Model Problem

$$-\nabla \cdot u \nabla q = f \quad \text{on } D + \text{hom. Dirichlet-BC.}$$

- $G : X \mapsto H_0^1$ the forward map, the solution operator of the elliptic problem, $f \in L^2(D)$
- $\mathcal{O} : H_0^1 \mapsto \mathbb{R}^K$ bounded, linear observation operator
- $\mathcal{G} : X \mapsto \mathbb{R}^K$ uncertainty-to-observation map, $\mathcal{G} = \mathcal{O} \circ G$
- $\eta \in \mathbb{R}^K$ the observational noise ($\eta \sim \mathcal{N}(0, \Gamma)$)

Groundwaterflow problem

Model Problem

$$-\nabla \cdot u \nabla q = f \quad \text{on } D + \text{hom. Dirichlet-BC.}$$

- $G : X \mapsto H_0^1$ the forward map, the solution operator of the elliptic problem, $f \in L^2(D)$
- $\mathcal{O} : H_0^1 \mapsto \mathbb{R}^K$ bounded, linear observation operator
- $\mathcal{G} : X \mapsto \mathbb{R}^K$ uncertainty-to-observation map, $\mathcal{G} = \mathcal{O} \circ G$
- $\eta \in \mathbb{R}^K$ the observational noise ($\eta \sim \mathcal{N}(0, \Gamma)$)

Find the unknown data u from noisy observations

$$y = \mathcal{G}(u) + \eta$$

Groundwaterflow problem

Uniform Prior

- $u(x,\omega) := u_0(x) + \sum_{j\in\mathbb{N}} \zeta_j(\omega)\gamma_j\phi_j(x)$
- $\frac{1}{1+\delta}u_{\min} \le u(x,\omega) \le u_{\max} + \frac{\delta}{1+\delta}u_{\min}$ a.e. $x \in D$, $\mathbb{P}$ a.s. $\Rightarrow \mu_0(X) = 1$.
- $\mu_\eta = \mathcal{N}(0,\Gamma)$ and $\mu_u = \mathcal{N}(\mathcal{G}(u),\Gamma)$, thus

$$\frac{\mathrm{d}\mu_u}{\mathrm{d}\mu_\eta}(y) = \exp(-\Phi(u;y))$$

  with

$$\Phi(u;y) = \frac{1}{2}|\Gamma^{-\frac{1}{2}}(y - G(u))|^2 - \frac{1}{2}|\Gamma^{-\frac{1}{2}}y|^2$$

- Continuity of Φ and $\mu_0(X) = 1 \Rightarrow \Phi$ is ν_0-measurable.
- Boundedness of $\Phi(\cdot;y)$ ensures $Z > 0$.

Groundwaterflow problem

Uniform Prior

- $u(x,\omega) := u_0(x) + \sum_{j\in\mathbb{N}} \zeta_j(\omega)\gamma_j\phi_j(x)$
- $\frac{1}{1+\delta}u_{\min} \leq u(x,\omega) \leq u_{\max} + \frac{\delta}{1+\delta}u_{\min}$ a.e. $x \in D$, $\mathbb{P}$ a.s. $\Rightarrow \mu_0(X) = 1$.
- $\mu_\eta = \mathcal{N}(0,\Gamma)$ and $\mu_u = \mathcal{N}(\mathcal{G}(u),\Gamma)$, thus

$$\frac{\mathrm{d}\mu_u}{\mathrm{d}\mu_\eta}(y) = \exp(-\Phi(u;y))$$

  with

$$\Phi(u;y) = \frac{1}{2}|\Gamma^{-\frac{1}{2}}(y - G(u))|^2 - \frac{1}{2}|\Gamma^{-\frac{1}{2}}y|^2$$

- Continuity of Φ and $\mu_0(X) = 1 \Rightarrow \Phi$ is ν_0-measurable.
- Boundedness of $\Phi(\cdot;y)$ ensures $Z > 0$.

Groundwaterflow problem

Uniform Prior

- $u(x,\omega) := u_0(x) + \sum_{j\in\mathbb{N}} \zeta_j(\omega)\gamma_j\phi_j(x)$
- $\frac{1}{1+\delta}u_{\min} \leq u(x,\omega) \leq u_{\max} + \frac{\delta}{1+\delta}u_{\min}$ a.e. $x \in D$, $\mathbb{P}$ a.s. $\Rightarrow \mu_0(X) = 1$.
- $\mu_\eta = \mathcal{N}(0,\Gamma)$ and $\mu_u = \mathcal{N}(\mathcal{G}(u),\Gamma)$, thus

$$\frac{\mathrm{d}\mu_u}{\mathrm{d}\mu_\eta}(y) = \exp(-\Phi(u;y))$$

  with

$$\Phi(u;y) = \frac{1}{2}|\Gamma^{-\frac{1}{2}}(y - G(u))|^2 - \frac{1}{2}|\Gamma^{-\frac{1}{2}}y|^2$$

  Note that the second term comes from the normal distribution of the noise, i.e. η has Lebesgue density $\frac{1}{\sqrt{(2\pi)^K|\Gamma|}}\exp(-\frac{1}{2}|\Gamma^{-\frac{1}{2}}y|^2)$.
- Continuity of Φ and $\mu_0(X) = 1 \Rightarrow \Phi$ is ν_0-measurable.
- Boundedness of $\Phi(\cdot;y)$ ensures $Z > 0$.

Groundwaterflow problem

Uniform Prior

- $u(x,\omega) := u_0(x) + \sum_{j\in\mathbb{N}} \zeta_j(\omega)\gamma_j\phi_j(x)$
- $\frac{1}{1+\delta}u_{\min} \leq u(x,\omega) \leq u_{\max} + \frac{\delta}{1+\delta}u_{\min}$ a.e. $x \in D$, $\mathbb{P}$ a.s. $\Rightarrow \mu_0(X) = 1$.
- $\mu_\eta = \mathcal{N}(0,\Gamma)$ and $\mu_u = \mathcal{N}(\mathcal{G}(u),\Gamma)$, thus

$$\frac{\mathrm{d}\mu_u}{\mathrm{d}\mu_\eta}(y) = \exp(-\Phi(u;y))$$

  with

$$\Phi(u;y) = \frac{1}{2}|\Gamma^{-\frac{1}{2}}(y - G(u))|^2 - \frac{1}{2}|\Gamma^{-\frac{1}{2}}y|^2$$

- Continuity of Φ and $\mu_0(X) = 1 \Rightarrow \Phi$ is ν_0-measurable.
- Boundedness of $\Phi(\cdot;y)$ ensures $Z > 0$.

Groundwaterflow problem

Uniform Prior

- $u(x,\omega) := u_0(x) + \sum_{j\in\mathbb{N}} \zeta_j(\omega)\gamma_j\phi_j(x)$
- $\frac{1}{1+\delta}u_{\min} \leq u(x,\omega) \leq u_{\max} + \frac{\delta}{1+\delta}u_{\min}$ a.e. $x \in D$, $\mathbb{P}$ a.s. $\Rightarrow \mu_0(X) = 1$.
- $\mu_\eta = \mathcal{N}(0,\Gamma)$ and $\mu_u = \mathcal{N}(\mathcal{G}(u),\Gamma)$, thus

$$\frac{\mathrm{d}\mu_u}{\mathrm{d}\mu_\eta}(y) = \exp(-\Phi(u;y))$$

  with

$$\Phi(u;y) = \frac{1}{2}|\Gamma^{-\frac{1}{2}}(y - G(u))|^2 - \frac{1}{2}|\Gamma^{-\frac{1}{2}}y|^2$$

- Continuity of Φ and $\mu_0(X) = 1 \Rightarrow \Phi$ is ν_0-measurable.
- Boundedness of $\Phi(\cdot;y)$ ensures $Z > 0$.

Groundwaterflow problem

Uniform Prior

- $u(x,\omega) := u_0(x) + \sum_{j\in\mathbb{N}} \zeta_j(\omega)\gamma_j\phi_j(x)$
- $\frac{1}{1+\delta}u_{\min} \le u(x,\omega) \le u_{\max} + \frac{\delta}{1+\delta}u_{\min}$ a.e. $x \in D$, $\mathbb{P}$ a.s. $\Rightarrow \mu_0(X) = 1$.
- $\mu_\eta = \mathcal{N}(0,\Gamma)$ and $\mu_u = \mathcal{N}(\mathcal{G}(u),\Gamma)$, thus

$$\frac{\mathrm{d}\mu_u}{\mathrm{d}\mu_\eta}(y) = \exp(-\Phi(u;y))$$

with

$$\Phi(u;y) = \frac{1}{2}|\Gamma^{-\frac{1}{2}}(y - G(u))|^2 - \frac{1}{2}|\Gamma^{-\frac{1}{2}}y|^2$$

- Continuity of Φ and $\mu_0(X) = 1 \Rightarrow \Phi$ is ν_0-measurable.
- Boundedness of $\Phi(\cdot;y)$ ensures $Z > 0$.

$$\frac{\mathrm{d}\mu^y}{\mathrm{d}\mu_0}(u) = \frac{1}{Z}\exp(-\frac{1}{2}|\Gamma^{-\frac{1}{2}}(y-G(u))|^2)\ , \quad Z = \int_X \exp(-\frac{1}{2}|\Gamma^{-\frac{1}{2}}(y-G(u))|^2)\mu_0(\mathrm{d}u)\ .$$

Groundwaterflow problem

Gaussian Prior

- $\kappa = \exp(u)$, prior on u is the measure $\mathcal{N}(0, \Delta^{-2})$, $\Rightarrow \mu_0(X) = 1$, $X = C(D)$.
- $\mu_\eta = \mathcal{N}(0, \Gamma)$ and $\mu_u = \mathcal{N}(\mathcal{G}(u), \Gamma)$, thus

$$\frac{\mathrm{d}\mu_u}{\mathrm{d}\mu_\eta}(y) = \exp(-\Phi(u; y))$$

with

$$\Phi(u; y) = \frac{1}{2}|\Gamma^{-\frac{1}{2}}(y - \mathcal{G}(u))|^2 - \frac{1}{2}|\Gamma^{-\frac{1}{2}} y|^2$$

with $\mathcal{G}$ s.t. the diffusion coefficient is $\kappa = \exp(u)$.
- Local Lipschitz continuity of Φ and $\mu_0(X) = 1 \Rightarrow \Phi$ is ν_0-measurable.
- Boundedness of $\mathcal{G}$ on the unit ball in X ensures $Z > 0$.

Groundwaterflow problem

Gaussian Prior

- $\kappa = \exp(u)$, prior on u is the measure $\mathcal{N}(0, \Delta^{-2})$, $\Rightarrow \mu_0(X) = 1$, $X = C(D)$.
- $\mu_\eta = \mathcal{N}(0, \Gamma)$ and $\mu_u = \mathcal{N}(\mathcal{G}(u), \Gamma)$, thus

$$\frac{\mathrm{d}\mu_u}{\mathrm{d}\mu_\eta}(y) = \exp(-\Phi(u; y))$$

  with

$$\Phi(u; y) = \frac{1}{2}|\Gamma^{-\frac{1}{2}}(y - \mathcal{G}(u))|^2 - \frac{1}{2}|\Gamma^{-\frac{1}{2}}y|^2$$

  with $\mathcal{G}$ s.t. the diffusion coefficient is $\kappa = \exp(u)$.
- Local Lipschitz continuity of Φ and $\mu_0(X) = 1 \Rightarrow \Phi$ is ν_0-measurable.
- Boundedness of $\mathcal{G}$ on the unit ball in X ensures $Z > 0$.

Groundwaterflow problem

Gaussian Prior

- $\kappa = \exp(u)$, prior on u is the measure $\mathcal{N}(0, \Delta^{-2})$, $\Rightarrow \mu_0(X) = 1$, $X = C(D)$.
- $\mu_\eta = \mathcal{N}(0, \Gamma)$ and $\mu_u = \mathcal{N}(\mathcal{G}(u), \Gamma)$, thus

$$\frac{\mathrm{d}\mu_u}{\mathrm{d}\mu_\eta}(y) = \exp(-\Phi(u; y))$$

  with

$$\Phi(u; y) = \frac{1}{2}|\Gamma^{-\frac{1}{2}}(y - \mathcal{G}(u))|^2 - \frac{1}{2}|\Gamma^{-\frac{1}{2}}y|^2$$

  with $\mathcal{G}$ s.t. the diffusion coefficient is $\kappa = \exp(u)$.
- Local Lipschitz continuity of Φ and $\mu_0(X) = 1 \Rightarrow \Phi$ is ν_0-measurable.
- Boundedness of $\mathcal{G}$ on the unit ball in X ensures $Z > 0$.

Groundwaterflow problem

Gaussian Prior

- $\kappa = \exp(u)$, prior on u is the measure $\mathcal{N}(0, \Delta^{-2})$, $\Rightarrow \mu_0(X) = 1$, $X = C(D)$.
- $\mu_\eta = \mathcal{N}(0, \Gamma)$ and $\mu_u = \mathcal{N}(\mathcal{G}(u), \Gamma)$, thus

$$\frac{\mathrm{d}\mu_u}{\mathrm{d}\mu_\eta}(y) = \exp(-\Phi(u; y))$$

  with

$$\Phi(u; y) = \frac{1}{2}|\Gamma^{-\frac{1}{2}}(y - \mathcal{G}(u))|^2 - \frac{1}{2}|\Gamma^{-\frac{1}{2}}y|^2$$

  with $\mathcal{G}$ s.t. the diffusion coefficient is $\kappa = \exp(u)$.
- Local Lipschitz continuity of Φ and $\mu_0(X) = 1 \Rightarrow \Phi$ is ν_0-measurable.
- Boundedness of $\mathcal{G}$ on the unit ball in X ensures $Z > 0$.

Groundwaterflow problem

Gaussian Prior

- $\kappa = \exp(u)$, prior on u is the measure $\mathcal{N}(0, \Delta^{-2})$, $\Rightarrow \mu_0(X) = 1$, $X = C(D)$.
- $\mu_\eta = \mathcal{N}(0, \Gamma)$ and $\mu_u = \mathcal{N}(\mathcal{G}(u), \Gamma)$, thus

$$\frac{\mathrm{d}\mu_u}{\mathrm{d}\mu_\eta}(y) = \exp(-\Phi(u;y))$$

  with

$$\Phi(u;y) = \frac{1}{2}|\Gamma^{-\frac{1}{2}}(y - \mathcal{G}(u))|^2 - \frac{1}{2}|\Gamma^{-\frac{1}{2}}y|^2$$

  with $\mathcal{G}$ s.t. the diffusion coefficient is $\kappa = \exp(u)$.
- Local Lipschitz continuity of Φ and $\mu_0(X) = 1 \Rightarrow \Phi$ is ν_0-measurable.
- Boundedness of $\mathcal{G}$ on the unit ball in X ensures $Z > 0$.

$$\frac{\mathrm{d}\mu^y}{\mathrm{d}\mu_0}(u) = \frac{1}{Z}\exp(-\frac{1}{2}|\Gamma^{-\frac{1}{2}}(y - G(u))|^2)\ , \quad Z = \int_X \exp(-\frac{1}{2}|\Gamma^{-\frac{1}{2}}(y - G(u))|^2)\mu_0(\mathrm{d}u)\ .$$

Groundwaterflow problem

Model Problem

$$-\nabla \cdot {\color{red}u}\nabla q = f \quad \text{on } D + \text{hom. Dirichlet-BC.}$$

- $G : X \mapsto H_0^1$ the forward map, the solution operator of the elliptic problem, $f \in L^2(D)$
- $\mathcal{O} : H_0^1 \mapsto \mathbb{R}^K$ bounded, linear observation operator
- $\mathcal{G} : X \mapsto \mathbb{R}^K$ uncertainty-to-observation map, $\mathcal{G} = \mathcal{O} \circ G$
- $\eta \in \mathbb{R}^K$ the observational noise ($\eta \sim \mathcal{N}(0, \Gamma)$)

Find the unknown data u from noisy observations

$$y = \mathcal{G}(u) + \eta$$

Groundwaterflow problem

Well-Posedness

Let μ_0 denote the uniform / lognormal prior and $\mathcal{G}$ be uncertainty-to-observation map of the inverse problem.
Then, for every fixed $r > 0$, there is $C = C(r) > 0$ such that for all $y, y' \in B_Y(0, r)$

$$d_{\mathsf{Hell}}(\mu^y, \mu^{y'}) \leq C\|y - y'\|_Y \ .$$

- The Hellinger distance between two probability measures μ and μ' on a separable Banach space, where $\mu \ll \nu$ and $\mu' \ll \nu$ for a reference measure ν, is

$$d_{\mathsf{Hell}}(\mu, \mu') = \sqrt{\frac{1}{2}\int \left(\sqrt{\frac{\mathrm{d}\mu}{\mathrm{d}\nu}} - \sqrt{\frac{\mathrm{d}\mu'}{\mathrm{d}\nu}}\right)^2 \mathrm{d}\nu} \ .$$

Groundwaterflow problem

Well-Posedness

Let μ_0 denote the uniform / lognormal prior and $\mathcal{G}$ be uncertainty-to-observation map of the inverse problem.
Then, for every fixed $r > 0$, there is $C = C(r) > 0$ such that for all $y, y' \in B_Y(0, r)$

$$d_{\mathsf{Hell}}(\mu^y, \mu^{y'}) \leq C \|y - y'\|_Y \ .$$

- The Hellinger distance between two probability measures μ and μ' on a separable Banach space, where $\mu \ll \nu$ and $\mu' \ll \nu$ for a reference measure ν, is

$$d_{\mathsf{Hell}}(\mu, \mu') = \sqrt{\frac{1}{2} \int \left(\sqrt{\frac{\mathrm{d}\mu}{\mathrm{d}\nu}} - \sqrt{\frac{\mathrm{d}\mu'}{\mathrm{d}\nu}} \right)^2 \mathrm{d}\nu} \ .$$

The Hellinger distance directly translates into bounds on expectations, i.e.

$$\|\mathbb{E}^{\mu} f - \mathbb{E}^{\mu'} f\| \leq 2(\mathbb{E}^{\mu}\|f\|^2 - \mathbb{E}^{\mu'}\|f\|^2)^{\frac{1}{2}} d_{\mathsf{Hell}}(\mu, \mu').$$

Groundwaterflow problem

- μ and μ^N are both measures, which are absolutely continuous w.r. to the prior μ_0.
- $\frac{\mathrm{d}\mu}{\mathrm{d}\mu_0}(u) = \frac{1}{Z}\exp(-\Phi(u)),\ Z = \int_X \exp(-\Phi(u))\mu_0(\mathrm{d}u)$.
- $\frac{\mathrm{d}\mu^N}{\mathrm{d}\mu_0}(u) = \frac{1}{Z^N}\exp(-\Phi^N(u)),\ Z^N = \int_X \exp(-\Phi^N(u))\mu_0(\mathrm{d}u)$.

Well-Posedness

Assume that $\mathcal{G}$ is approximated by a function $\mathcal{G}^N$ such that, for any $\epsilon > 0$ there is a $K(\epsilon) > 0$ with

$$|\mathcal{G}(u) - \mathcal{G}^N(u)| \leq K \exp(\epsilon \|u\|_X^2)\psi(N)$$

where $\psi(N) \to 0$ as $N \to \infty$.
Then there is $C > 0$ such that for all N sufficiently large

$$d_{\mathsf{Hell}}(\mu, \mu^N) \leq C\psi(N)\ .$$

Summary

- Mathematical structure of the Bayesian approach to inverse problems in differential equations.
- Definition of priors using a random series over an infinite set of functions.
- Bayes' theorem in infinite dimensions.
- Well-posedness and approximation theory (for uniform/lognormal prios and the elliptic model problem).

Summary

- Mathematical structure of the Bayesian approach to inverse problems in differential equations.
- Definition of priors using a random series over an infinite set of functions.
- Bayes' theorem in infinite dimensions.
- Well-posedness and approximation theory (for uniform/lognormal prios and the elliptic model problem).

How can we efficiently compute / approximate the posterior?

References

- M. Dashti and A.M.Stuart. The Bayesian Approach to Inverse Problems. Handbook of Uncertainty Quantification, Editors R. Ghanem, D. Higdon and H. Owhadi, Springer, 2015.

- J. Kaipio and E. Somersalo, Statistical and Computational Inverse Problems, Springer.

- A.M. Stuart. Inverse problems: a Bayesian perspective. Acta Numerica 19(2010).