Uncertainty Quantification and Quasi-Monte Carlo

Sommersemester 2025

Vesa Kaarnioja
vesa.kaarnioja@fu-berlin.de

FU Berlin, FB Mathematik und Informatik

Fourth lecture, May 5, 2025

Let $(\Omega, \mathcal{F}, \mu)$ be a probability space. We consider the problem

$$\begin{cases} -\nabla \cdot (a(\boldsymbol{x}, \omega) \nabla u(\boldsymbol{x}, \omega)) = f(\boldsymbol{x}) & \text{for } \boldsymbol{x} \in D, \text{ (a.e.) } \omega \in \Omega, \\ u(\boldsymbol{x}, \omega) = 0 & \text{for } \boldsymbol{x} \in \partial D, \text{ (a.e.) } \omega \in \Omega, \end{cases}$$

where the diffusion coefficient $a(\cdot, \omega)$ is *random*. In consequence, the solution $u(\cdot, \omega)$ is a random function/field.

In order to analyze $u(\cdot, \omega)$, some approaches might be:

- Monte Carlo methods $\rightarrow$ slow convergence rate.
- Sparse grid methods $\rightarrow$ good convergence, poor parallelization.

In certain problems (such as the PDE above) the dependence of u on a tends to be quite smooth (under moderate modeling assumptions). Quasi-Monte Carlo methods take advantage of this fact and can be used to obtain faster-than-Monte Carlo convergence rates.

Probability measures

Let Ω be a set and let $\mathcal{P}(\Omega) := \{B \mid B \subseteq \Omega\}$ denote its power set. A subset $\mathcal{F}$ of $\mathcal{P}(\Omega)$ is called *σ-algebra* (or σ-field) if

1. $\varnothing \in \mathcal{F}$,
2. $\Omega \setminus A \in \mathcal{F}$ for every $A \in \mathcal{F}$, and
3. $\bigcup_{n\in\mathbb{N}} A_n \in \mathcal{F}$ for every countable subset $\{A_n\}_{n\in\mathbb{N}}$ of $\mathcal{F}$.

A pair $(\Omega, \mathcal{F})$ is called a *measurable space*.

An intuitive way of thinking about σ-algebras is that they contain information. The subsets contained in a σ-algebra represent events for which we can decide, after the observation, whether they happened or not. Hence, $\mathcal{F}$ represents all the information we can get from an experiment. For a topological space Ω (e.g., $\mathbb{R}^s$), the smallest σ-algebra containing all open sets in Ω is called *Borel σ-algebra* on Ω and it is denoted by $\mathrm{Bor}(\Omega)$.

A function $\mu\colon \mathcal{F} \to [0,\infty) \cup \{\infty\}$ is called *probability measure* if

(i) $\mu(\varnothing) = 0$,

(ii) for every countable subset $\{A_n\}_{n\in\mathbb{N}} \subset \mathcal{F}$ of pairwise disjoint sets (i.e., $A_i \cap A_j = \varnothing$ if $i \neq j$),

$$\mu\left(\bigcup_{k=1}^{\infty} A_k\right) = \sum_{k=1}^{\infty} \mu(A_n),$$

(iii) and $\mu(\Omega) = 1$.

We call $\mu(A)$ the *probability* of an event $A \in \mathcal{F}$. If $\mu(A) = 1$, we say that the event A occurs *almost surely*. A triple $(\Omega, \mathcal{F}, \mu)$ is called *probability space*. If only properties (i) and (ii) are satisfied, μ is called a *measure*. A measure is called σ-finite if Ω is the countable union of measurable sets with finite measure.

Example

The *Dirac measure* $\delta_{\boldsymbol{m}}$ at a point $\boldsymbol{m} \in \mathbb{R}^s$ is a probability measure on $(\mathbb{R}^s, \mathrm{Bor}(\mathbb{R}^s))$ defined by

$$\delta_{\boldsymbol{m}}(A) = \begin{cases} 1 & \text{if } \boldsymbol{m} \in A, \\ 0 & \text{if } \boldsymbol{m} \notin A \end{cases} \quad \text{for all } A \in \mathrm{Bor}(\mathbb{R}^s).$$

Example

The Lebesgue measure λ on $(\mathbb{R}^s, \mathrm{Bor}(\mathbb{R}^s))$ is σ-finite, but not a probability measure, since $\lambda(\mathbb{R}^s) = \infty$.

Let μ and ν be two measures on the same measure space. Then μ is said to be *absolutely continuous with respect to* ν (or *dominated by* ν) if $\nu(A) = 0$ implies $\mu(A) = 0$ for each $A \in \mathcal{F}$. We denote this by $\mu \ll \nu$. Measures μ and ν are called *equivalent* if $\mu \ll \nu$ and $\nu \ll \mu$. If μ and ν are supported on disjoint sets, they are called *mutually singular*.

Theorem (Radon–Nikodym)

Let μ and ν be two measures on a measure space $(\Omega, \mathcal{F})$. If $\mu \ll \nu$ and ν is σ-finite, then there exists a unique ν-integrable function f such that

$$\mu(A) = \int_A f(\omega)\, \nu(\mathrm{d}\omega) \quad \textit{for all } A \in \mathcal{F}.$$

The function f is called *Radon–Nikodym derivative* (or *density*) of μ with respect to ν and it is denoted by $\frac{\mathrm{d}\mu}{\mathrm{d}\nu}$.

Example

If μ is a measure which is absolutely continuous with respect to the Lebesgue measure λ on $(\mathbb{R}^s, \mathrm{Bor}(\mathbb{R}^s))$, then it has a unique density $p \in L^1(\mathbb{R}^s)$ by the Radon–Nikodym theorem.

Example

Let $\mu_1 = \mathcal{U}([0,1])$ and $\mu_2 = \mathcal{U}([0,2])$ be uniform probability measures on $\mathbb{R}$. Then $\mu_1 \ll \mu_2$ with

$$\frac{\mathrm{d}\mu_1}{\mathrm{d}\mu_2}(t) = \begin{cases} 2 & \text{for } t \in [0,1], \\ 0 & \text{otherwise,} \end{cases}$$

but μ_2 is not absolutely continuous with respect to μ_1 because $\mu_1([1,2]) = 0$, whereas $\mu_2([1,2]) = \frac{1}{2} > 0$.

Random variables

A function $x\colon \Omega \to X$ between a probability space $(\Omega, \mathcal{F}, \mu)$ and a measurable space $(X, \mathcal{X})$ is called a *random variable (with values in X)* if it is measurable, that is, if

$$x^{-1}(A) \in \mathcal{F} \quad \text{for every } A \in \mathcal{X}.$$

Here, $x^{-1}(A) = \{\omega \in \Omega : x(\omega) \in A\}$.

A random variable x induces a probability measure ν on X, defined by

$$\nu(A) := \mu(x^{-1}(A)) \quad \text{for all } A \in \mathcal{X},$$

which is called *probability distribution* (or *law*) of x. We write $x \sim \nu$ if x is distributed according to ν.

A random variable x connects an event $A \in \mathcal{X}$ with a corresponding event $x^{-1}(A) \in \mathcal{F}$ and assigns the probability of $x^{-1}(A)$ to A. This probability is denoted by

$$\mathbb{P}(x \in A) := \nu(A) = \mu(x^{-1}(A)) = \mu(\{\omega \in \Omega : x(\omega) \in A\}).$$

Now, let $\boldsymbol{x}$ be a random variable with values in $(\mathbb{R}^s, \mathrm{Bor}(\mathbb{R}^s))$ and ν its distribution.

If ν is absolutely continuous with respect to the Lebesgue measure λ on $\mathbb{R}^s$, then by the Radon–Nikodym theorem there exists a unique $p \in L^1(\mathbb{R}^s)$ such that

$$\nu(A) = \int_A p(\boldsymbol{x}) \mathrm{d}\boldsymbol{x} \quad \text{for all } A \in \mathrm{Bor}(\mathbb{R}^s).$$

The function p is called *probability density* of $\boldsymbol{x}$.

In what follows, we will assume that $\mathbb{R}^s$-valued random variables have a probability density.

Let $\boldsymbol{x}$, $\boldsymbol{x}_1$, and $\boldsymbol{x}_2$ be $\mathbb{R}^s$-valued random variables.

- The *mean* or *expected value* of $\boldsymbol{x}$ with distribution ν and probability density function p is given by

$$\mathbb{E}[\boldsymbol{x}] := \int_{\mathbb{R}^s} \boldsymbol{x}\, \nu(\mathrm{d}\boldsymbol{x}) = \int_{\mathbb{R}^s} \boldsymbol{x} p(\boldsymbol{x})\, \mathrm{d}\boldsymbol{x}.$$

- A *mode* $\bar{\boldsymbol{x}}$ of a random variable $\boldsymbol{x}$ is defined as a maximizer of its density p, i.e.,

$$\bar{\boldsymbol{x}} \in \operatorname*{arg\,max}_{\boldsymbol{x} \in \mathbb{R}^s} p(\boldsymbol{x}).$$

- The *covariance* (or *covariance matrix*) of two random variables $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ is defined by

$$\mathrm{Cov}(\boldsymbol{x}_1, \boldsymbol{x}_2) = \mathbb{E}\left[(\boldsymbol{x}_1 - \mathbb{E}[\boldsymbol{x}_1])(\boldsymbol{x}_2 - \mathbb{E}[\boldsymbol{x}_2])^{\mathrm{T}}\right].$$

- The *variance* of random variable $\boldsymbol{x}$ is its covariance with itself:

$$\mathrm{Var}(\boldsymbol{x}) = \mathrm{Cov}(\boldsymbol{x}, \boldsymbol{x}).$$

- The *characteristic function* $\varphi_{\boldsymbol{x}}$ of $\boldsymbol{x}$ is defined by

$$\varphi_{\boldsymbol{x}}(\boldsymbol{h}) = \int_{\mathbb{R}^s} \exp(\mathrm{i}\, \boldsymbol{h}^{\mathrm{T}} \boldsymbol{x})\, \nu(\mathrm{d}\boldsymbol{x}) = \int_{\mathbb{R}^s} \exp(\mathrm{i}\, \boldsymbol{h}^{\mathrm{T}} \boldsymbol{x}) p(\boldsymbol{x}) \mathrm{d}\boldsymbol{x} \quad \text{for all } \boldsymbol{h} \in \mathbb{R}^s.$$

  A random variable is uniquely determined by its characteristic function.

Gaussian random variables

Let $\boldsymbol{m} \in \mathbb{R}^s$ and $C \in \mathbb{R}^{s\times s}$ be a symmetric positive semidefinite matrix.[†] An $\mathbb{R}^s$-valued random variable $\boldsymbol{x}$ is said to be *Gaussian* (or *normal*) with mean $\boldsymbol{m}$ and covariance C, denoted by $\boldsymbol{x} \sim \mathcal{N}(\boldsymbol{m}, C)$, if its characteristic function $\varphi_{\boldsymbol{x}}$ is given by

$$\varphi_{\boldsymbol{x}}(\boldsymbol{h}) = \exp\left(\mathrm{i}\,\boldsymbol{h}^{\mathrm{T}}\boldsymbol{m} - \frac{1}{2}\boldsymbol{h}^{\mathrm{T}} C \boldsymbol{h}\right) \quad \text{for all } \boldsymbol{h} \in \mathbb{R}^s.$$

A Gaussian random variable is completely determined by its mean and its covariance.

Remark: Multivariate Gaussian random variables also have the following characterization. A random vector $\boldsymbol{x} = (x_1, \ldots, x_s)^{\mathrm{T}}$ has a multivariate normal distribution iff $y = a_1 x_1 + \cdots + a_s x_s$ is (univariate) normally distributed for all constants $a_1, \ldots, a_s \in \mathbb{R}$.

[†] Recall that this means $\boldsymbol{\xi}^{\mathrm{T}} C \boldsymbol{\xi} \geq 0$ for all $\boldsymbol{\xi} \in \mathbb{R}^s$.

- If, in addition, C is positive definite[†], $\boldsymbol{x} \sim \mathcal{N}(\boldsymbol{m}, C)$ has the probability density

$$p(\boldsymbol{x}) = \frac{1}{(2\pi)^{s/2}\sqrt{\det C}} \exp\left(-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{m})^{\mathrm{T}} C^{-1}(\boldsymbol{x}-\boldsymbol{m})\right)$$
$$= \frac{1}{(2\pi)^{s/2}\sqrt{\det C}} \exp\left(-\frac{1}{2}\|C^{-\frac{1}{2}}(\boldsymbol{x}-\boldsymbol{m})\|^2\right).$$

  Note that C is invertible and $C^{-1/2}$ exists due to our assumptions on C.

- The Dirac measure $\delta_{\boldsymbol{m}}$ at a point $\boldsymbol{m} \in \mathbb{R}^s$ can be understood as a Gaussian distribution with covariance $C = 0$, i.e., $\delta_{\boldsymbol{m}} = \mathcal{N}(\boldsymbol{m}, \boldsymbol{0})$.

- If $\boldsymbol{z}_1 \sim \mathcal{N}(\boldsymbol{m}_1, C_1)$ and $\boldsymbol{z}_2 \sim \mathcal{N}(\boldsymbol{m}_2, C_2)$ are independent and $a_1, a_2 \in \mathbb{R}$, then

$$\boldsymbol{z} = a_1\boldsymbol{z}_1 + a_2\boldsymbol{z}_2 \sim \mathcal{N}(a_1\boldsymbol{m}_1 + a_2\boldsymbol{m}_2, a_1^2 C_1 + a_2^2 C_2).$$

- If $\boldsymbol{z} \sim \mathcal{N}(\boldsymbol{m}, C)$, $L \in \mathbb{R}^{s\times k}$, and $\boldsymbol{a} \in \mathbb{R}^s$, then

$$\boldsymbol{w} = L\boldsymbol{z} + \boldsymbol{a} \sim \mathcal{N}(L\boldsymbol{m} + \boldsymbol{a}, LCL^{\mathrm{T}}).$$

[†]Recall that this means $\boldsymbol{\xi}^{\mathrm{T}} C\boldsymbol{\xi} > 0$ for all $\boldsymbol{\xi} \in \mathbb{R}^s \setminus \{\boldsymbol{0}\}$.

Conditional and marginal probability densities

Let $\boldsymbol{x}$ and $\boldsymbol{y}$ be random variables with values in $\mathbb{R}^s$ and $\mathbb{R}^k$, respectively. If the random variable $(\boldsymbol{x}, \boldsymbol{y})$ has a probability density $p_{\boldsymbol{x},\boldsymbol{y}}$, i.e., if

$$\mathbb{P}(\boldsymbol{x} \in A, \boldsymbol{y} \in B) = \mathbb{P}((\boldsymbol{x}, \boldsymbol{y}) \in A \times B) = \int_{A \times B} p_{\boldsymbol{x},\boldsymbol{y}}(\boldsymbol{u}, \boldsymbol{v}) \mathrm{d}(\boldsymbol{u}, \boldsymbol{v}),$$

for all $A \in \mathrm{Bor}(\mathbb{R}^s)$ and $B \in \mathrm{Bor}(\mathbb{R}^k)$, then $p_{\boldsymbol{x},\boldsymbol{y}}$ is called *joint probability density* of $\boldsymbol{x}$ and $\boldsymbol{y}$. Here $\mathbb{P}(\boldsymbol{x} \in A, \boldsymbol{y} \in B) := \mathbb{P}(\boldsymbol{x} \in A \text{ and } \boldsymbol{y} \in B)$.

Now, the *marginal probability density* $p_{\boldsymbol{x}}$ of $\boldsymbol{x}$ is defined by

$$p_{\boldsymbol{x}}(\boldsymbol{u}) = \int_{\mathbb{R}^k} p_{\boldsymbol{x},\boldsymbol{y}}(\boldsymbol{u}, \boldsymbol{v}) \mathrm{d}\boldsymbol{v} \quad \text{for all } \boldsymbol{u} \in \mathbb{R}^s.$$

Analogously, the marginal density of $\boldsymbol{y}$ is

$$p_{\boldsymbol{y}}(\boldsymbol{v}) = \int_{\mathbb{R}^s} p_{\boldsymbol{x},\boldsymbol{y}}(\boldsymbol{u}, \boldsymbol{v}) \mathrm{d}\boldsymbol{u} \quad \text{for all } \boldsymbol{v} \in \mathbb{R}^k.$$

The marginal density of $\boldsymbol{x}$ is indeed the probability density for $\boldsymbol{x}$ in the situation that we have no information about the random variable $\boldsymbol{y}$, because

$$\begin{aligned}\mathbb{P}(\boldsymbol{x} \in A) = \mathbb{P}(\boldsymbol{x} \in A, \boldsymbol{y} \in \mathbb{R}^k) &= \int_{A \times \mathbb{R}^k} p_{\boldsymbol{x},\boldsymbol{y}}(\boldsymbol{u}, \boldsymbol{v}) \mathrm{d}(\boldsymbol{u}, \boldsymbol{v}) \\ &= \int_A \left(\int_{\mathbb{R}^k} p_{\boldsymbol{x},\boldsymbol{y}}(\boldsymbol{u}, \boldsymbol{v}) \mathrm{d}\boldsymbol{v} \right) \mathrm{d}\boldsymbol{u} = \int_A p_{\boldsymbol{x}}(\boldsymbol{u}) \mathrm{d}\boldsymbol{u}\end{aligned}$$

for every $A \in \mathrm{Bor}(\mathbb{R}^s)$.

The random variables $\boldsymbol{x}$ and $\boldsymbol{y}$ are called *independent* if

$$\mathbb{P}(\boldsymbol{x} \in A, \boldsymbol{y} \in B) = \mathbb{P}(\boldsymbol{x} \in A)\mathbb{P}(\boldsymbol{y} \in B)$$

for all $A \in \mathrm{Bor}(\mathbb{R}^s)$, $B \in \mathrm{Bor}(\mathbb{R}^k)$ or, equivalently, if

$$p_{\boldsymbol{x},\boldsymbol{y}}(\boldsymbol{u}, \boldsymbol{v}) = p_{\boldsymbol{x}}(\boldsymbol{u}) p_{\boldsymbol{y}}(\boldsymbol{v}) \quad \text{almost surely.}$$

Next, we consider the random variable $\boldsymbol{x}$ in the opposite situation that we know everything about the random variable $\boldsymbol{y}$: we have observed it and know what value it has taken.

We say we consider the random variable $\boldsymbol{x}$, *given* that we know the value $\boldsymbol{y}_0$ taken by $\boldsymbol{y}$, and denote this by $\boldsymbol{x}|\boldsymbol{y} = \boldsymbol{y}_0$. For $\boldsymbol{y}_0 \in \mathbb{R}^k$ with $p_{\boldsymbol{y}}(\boldsymbol{y}_0) > 0$, the *conditional probability density* of $\boldsymbol{x}|\boldsymbol{y} = \boldsymbol{y}_0$, $p_{\boldsymbol{x}|\boldsymbol{y}=\boldsymbol{y}_0}$, is then defined by

$$p_{\boldsymbol{x}|\boldsymbol{y}=\boldsymbol{y}_0}(\boldsymbol{u}) = \frac{p_{\boldsymbol{x},\boldsymbol{y}}(\boldsymbol{u}, \boldsymbol{y}_0)}{p_{\boldsymbol{y}}(\boldsymbol{y}_0)}.$$

If $\boldsymbol{x}$ and $\boldsymbol{y}$ are independent and $p_{\boldsymbol{y}}(\boldsymbol{y}_0) > 0$, then

$$p_{\boldsymbol{x}|\boldsymbol{y}=\boldsymbol{y}_0}(\boldsymbol{u}) = p_{\boldsymbol{x}}(\boldsymbol{u}).$$

Representation of random fields

Random field

Definition

Let $D \subset \mathbb{R}^d$ and let $(\Omega, \mathcal{F}, \mu)$ be a probability space. A function $A\colon D \times \Omega \to X$ is called a *random field* if $A(\boldsymbol{x}, \cdot)$ is an X-valued random variable for all $\boldsymbol{x} \in D$.

Definition

We call a random field $A\colon D \times \Omega \to X$ square-integrable if

$$\int_\Omega \int_D |A(\boldsymbol{x}, \omega)|^2 \, \mathrm{d}\boldsymbol{x} \, \mu(\mathrm{d}\omega) < \infty.$$

Our goal will be to model (infinite-dimensional) input random fields using finite-dimensional expansions with s variables.

Comment on notation: In what follows, s will always refer to the "stochastic dimension" (dimension of the stochastic/parametric space) while d will refer to the "spatial dimension" (dimension of the spatial Lipschitz domain $D \subset \mathbb{R}^d$, $d \in \{2, 3\}$).

Remark: separable Hilbert space

A Hilbert space is said to be *separable* if (and only if) there exists a *countable orthonormal basis* $\{\psi_j\}_{j=1}^{\infty}$ of H with respect to the inner product $\langle \cdot, \cdot \rangle_H$, that is,

$$\langle \psi_j, \psi_k \rangle_H = \delta_{j,k} \quad \text{and} \quad \left\| f - \sum_{j=1}^{\ell} \langle f, \psi_j \rangle_H \psi_j \right\|_H \xrightarrow{\ell \to \infty} 0 \quad \text{for all } f \in H.$$

This last condition is often written as

$$f = \sum_{j=1}^{\infty} \langle f, \psi_j \rangle_H \psi_j.$$

Note that $\sum_{j=1}^{\ell} \langle f, \psi_j \rangle_H \psi_j$ is precisely the orthogonal projection onto the subspace spanned by $\psi_1, \ldots, \psi_\ell$.

Mercer's theorem

Let $a(\boldsymbol{x}, \omega)$ be a square-integrable random field with mean

$$\bar{a}(\boldsymbol{x}) = \int_\Omega a(\boldsymbol{x}, \omega)\, \mu(\mathrm{d}\omega), \quad \boldsymbol{x} \in D,$$

and a continuous, symmetric, positive definite[†] covariance

$$K(\boldsymbol{x}, \boldsymbol{x}') = \int_\Omega (a(\boldsymbol{x}, \omega) - \bar{a}(\boldsymbol{x}))(a(\boldsymbol{x}', \omega) - \bar{a}(\boldsymbol{x}'))\, \mu(\mathrm{d}\omega).$$

Mercer's theorem: if $D \subset \mathbb{R}^d$ is a compact, measurable set with positive Lebesgue measure, then the covariance operator $\mathcal{C}\colon L^2(D) \to L^2(D)$,

$$(\mathcal{C}u)(\boldsymbol{x}) = \int_D K(\boldsymbol{x}, \boldsymbol{x}')u(\boldsymbol{x}')\, \mathrm{d}\boldsymbol{x}', \quad \boldsymbol{x} \in D,$$

has a countable sequence of eigenvalues $\{\lambda_k\}_{k\geq 1}$ and corresponding eigenfunctions $\{\psi_k\}_{k\geq 1}$ satisfying $\mathcal{C}\psi_k = \lambda_k\psi_k$ such that $\lambda_1 \geq \lambda_2 \geq \cdots \geq 0$ and $\lambda_k \to 0$ and the eigenfunctions form an orthonormal basis for $L^2(D)$.
Note that this means:

$$\int_D K(\boldsymbol{x}, \boldsymbol{x}')\psi_k(\boldsymbol{x}')\, \mathrm{d}\boldsymbol{x}' = \lambda_k\psi_k(\boldsymbol{x}), \quad \int_D \psi_k(\boldsymbol{x})\psi_\ell(\boldsymbol{x})\, \mathrm{d}\boldsymbol{x} = \delta_{k,\ell}.$$

[†] In this context, positive definite means: for all choices of finitely many points $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_k \in D$, $k \in \mathbb{N}$, the *Gram matrix* $G := [K(\boldsymbol{x}_i, \boldsymbol{x}_j)]_{i,j=1}^k$ is positive semidefinite.

The Karhunen–Loève (KL) expansion of a random field

Theorem

Let $(\Omega, \mathcal{F}, \mu)$ be a probability space, let $D \subset \mathbb{R}^d$ be a compact, measurable set with positive Lebesgue measure, and let $a\colon D \times \Omega \to \mathbb{R}$ be a square-integrable random field with continuous, symmetric, positive definite covariance $K(\boldsymbol{x}, \boldsymbol{x}') = \mathbb{E}[(a(\boldsymbol{x}, \cdot) - \bar{a}(\boldsymbol{x}))(a(\boldsymbol{x}', \cdot) - \bar{a}(\boldsymbol{x}'))]$. Then the eigensystem $(\lambda_k, \psi_k) \in \mathbb{R}_+ \times L^2(D)$ of the covariance operator $\mathcal{C}\colon L^2(D) \to L^2(D)$, as described on the previous slide, can be used to write

$$a(\boldsymbol{x}, \omega) = \bar{a}(\boldsymbol{x}) + \sum_{k=1}^{\infty} \sqrt{\lambda_k} \xi_k(\omega) \psi_k(\boldsymbol{x}),$$

$$\text{where } \xi_k(\omega) = \frac{1}{\sqrt{\lambda_k}} \int_D (a(\boldsymbol{x}, \omega) - \bar{a}(\boldsymbol{x})) \psi_k(\boldsymbol{x}) \, \mathrm{d}\boldsymbol{x},$$

where the convergence is in L^2 w.r.t. the stochastic parameter and uniform in $\boldsymbol{x}$. Furthermore, the random variables ξ_k are zero-mean uncorrelated random variables with unit variance, i.e.,

$$\mathbb{E}[\xi_k] = 0 \quad \textit{and} \quad \mathbb{E}[\xi_k \xi_\ell] = \delta_{k,\ell}.$$

Proof. WLOG, we can assume that $\bar{a}(\boldsymbol{x}) = 0$.[†] By Mercer's theorem, $\{\psi_k\}_{k=1}^{\infty}$ forms an orthonormal basis on $L^2(D)$ and we can write

$$K(\boldsymbol{x}, \boldsymbol{x}') = \sum_{k=1}^{\infty} \underbrace{\left(\int_D K(\boldsymbol{x}, \boldsymbol{t}) \psi_k(\boldsymbol{t}) \, \mathrm{d}\boldsymbol{t} \right)}_{=\lambda_k \psi_k(\boldsymbol{x})} \psi_k(\boldsymbol{x}') = \sum_{k=1}^{\infty} \lambda_k \psi_k(\boldsymbol{x}) \psi_k(\boldsymbol{x}').$$

Moreover, the random field a can be expressed using the same eigenbasis:

$$a(\boldsymbol{x}, \omega) = \sum_{k=1}^{\infty} \sqrt{\lambda_k} \xi_k(\omega) \psi_k(\boldsymbol{x}), \quad \xi_k(\omega) = \frac{1}{\sqrt{\lambda_k}} \int_D a(\boldsymbol{x}, \omega) \psi_k(\boldsymbol{x}) \, \mathrm{d}\boldsymbol{x}.$$

One easily computes that

$$\mathbb{E}[\xi_k] = \mathbb{E}\left[\frac{1}{\sqrt{\lambda_k}} \int_D a(\boldsymbol{x}, \cdot) \psi_k(\boldsymbol{x}) \, \mathrm{d}\boldsymbol{x} \right] = \frac{1}{\sqrt{\lambda_k}} \int_D \mathbb{E}[a(\boldsymbol{x}, \cdot)] \psi_k(\boldsymbol{x}) \, \mathrm{d}\boldsymbol{x} = 0$$

and

$$\mathbb{E}[\xi_k \xi_\ell] = \mathbb{E}\left[\frac{1}{\lambda_k} \int_D \int_D a(\boldsymbol{x}, \cdot) a(\boldsymbol{x}', \cdot) \psi_k(\boldsymbol{x}) \psi_\ell(\boldsymbol{x}') \, \mathrm{d}\boldsymbol{x} \, \mathrm{d}\boldsymbol{x}' \right]$$

$$= \frac{1}{\lambda_k} \int_D \int_D \mathbb{E}[a(\boldsymbol{x}, \cdot) a(\boldsymbol{x}', \cdot)] \psi_k(\boldsymbol{x}) \psi_\ell(\boldsymbol{x}') \, \mathrm{d}\boldsymbol{x} \, \mathrm{d}\boldsymbol{x}'$$

$$= \frac{1}{\lambda_k} \int_D \int_D K(\boldsymbol{x}, \boldsymbol{x}') \psi_k(\boldsymbol{x}) \psi_\ell(\boldsymbol{x}') \mathrm{d}\boldsymbol{x} \mathrm{d}\boldsymbol{x}' = \frac{1}{\lambda_k} \int_D \psi_k(\boldsymbol{x}) \underbrace{\left(\int_D K(\boldsymbol{x}, \boldsymbol{x}') \psi_\ell(\boldsymbol{x}') \mathrm{d}\boldsymbol{x}' \right)}_{=\lambda_\ell \psi_\ell(\boldsymbol{x})} \mathrm{d}\boldsymbol{x} = \delta_{k,\ell},$$

since $\int_D \psi_k(\boldsymbol{x}) \psi_\ell(\boldsymbol{x}) \, \mathrm{d}\boldsymbol{x} = \delta_{k,\ell}$.

[†] Once the claim has been proved for a zero-mean random field $a(\boldsymbol{x}, \omega)$, the general case follows simply by applying the theorem to $a(\boldsymbol{x}, \omega) \leftarrow a(\boldsymbol{x}, \omega) - \bar{a}(\boldsymbol{x})$.

Recall from the previous slide that

$$a(\boldsymbol{x},\omega)=\sum_{k=1}^{\infty}\sqrt{\lambda_k}\xi_k(\omega)\psi_k(\boldsymbol{x}),\ \ \xi_k(\omega)=\frac{1}{\sqrt{\lambda_k}}\int_D a(\boldsymbol{x},\omega)\psi_k(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x},$$

where $\mathbb{E}[\xi_k]=0,$ and $\mathbb{E}[\xi_k\xi_\ell]=\delta_{k,\ell}$. Let

$$a_s(\boldsymbol{x},\omega)=\sum_{k=1}^{s}\sqrt{\lambda_k}\xi_k(\omega)\psi_k(\boldsymbol{x}).$$

$$\mathbb{E}[|a(\boldsymbol{x},\cdot)-a_s(\boldsymbol{x},\cdot)|^2]=\mathbb{E}[a(\boldsymbol{x},\cdot)^2]+\mathbb{E}[a_s(\boldsymbol{x},\cdot)^2]-2\mathbb{E}[a(\boldsymbol{x},\cdot)a_s(\boldsymbol{x},\cdot)]$$

$$=K(\boldsymbol{x},\boldsymbol{x})+\mathbb{E}\bigg[\sum_{k=1}^{s}\sum_{\ell=1}^{s}\sqrt{\lambda_k\lambda_\ell}\xi_k(\cdot)\xi_\ell(\cdot)\psi_k(\boldsymbol{x})\psi_\ell(\boldsymbol{x})\bigg]$$

$$-2\mathbb{E}\bigg[\bigg(\sum_{\ell=1}^{\infty}\sqrt{\lambda_\ell}\xi_\ell(\cdot)\psi_\ell(\boldsymbol{x})\bigg)\bigg(\sum_{k=1}^{s}\sqrt{\lambda_k}\xi_k(\cdot)\psi_k(\boldsymbol{x})\bigg)\bigg]$$

$$=K(\boldsymbol{x},\boldsymbol{x})+\sum_{k=1}^{s}\sum_{\ell=1}^{s}\sqrt{\lambda_k\lambda_\ell}\mathbb{E}[\xi_k\xi_\ell]\psi_k(\boldsymbol{x})\psi_\ell(\boldsymbol{x})-2\mathbb{E}\bigg[\sum_{\ell=1}^{\infty}\sum_{k=1}^{s}\sqrt{\lambda_k\lambda_\ell}\xi_\ell(\cdot)\xi_k(\cdot)\psi_\ell(\boldsymbol{x})\psi_k(\boldsymbol{x})\bigg]$$

$$=K(\boldsymbol{x},\boldsymbol{x})+\sum_{k=1}^{s}\sum_{\ell=1}^{s}\sqrt{\lambda_k\lambda_\ell}\delta_{k,\ell}\psi_k(\boldsymbol{x})\psi_\ell(\boldsymbol{x})-2\sum_{\ell=1}^{\infty}\sum_{k=1}^{s}\sqrt{\lambda_k\lambda_\ell}\mathbb{E}[\xi_\ell\xi_k]\psi_\ell(\boldsymbol{x})\psi_k(\boldsymbol{x})$$

$$=K(\boldsymbol{x},\boldsymbol{x})+\sum_{\ell=1}^{s}\lambda_\ell\psi_\ell(\boldsymbol{x})^2-2\sum_{\ell=1}^{\infty}\sum_{k=1}^{s}\sqrt{\lambda_k\lambda_\ell}\mathbb{E}[\xi_\ell\xi_k]\psi_\ell(\boldsymbol{x})\psi_k(\boldsymbol{x})\qquad(\mathbb{E}[\xi_\ell\xi_k]=\delta_{\ell,k})$$

$$=K(\boldsymbol{x},\boldsymbol{x})-\sum_{\ell=1}^{s}\lambda_\ell\psi_\ell(\boldsymbol{x})^2\to 0\quad\text{in the } L^2 \text{ sense by Mercer's theorem.}\quad\square$$

The Karhunen–Loève (KL) expansion of random field $a(\boldsymbol{x}, \omega)$ can be written as

$$a(\boldsymbol{x}, \omega) = \bar{a}(\boldsymbol{x}) + \sum_{k=1}^{\infty} \sqrt{\lambda_k} \xi_k(\omega) \psi_k(\boldsymbol{x}).$$

Remarks:

- The KL expansion minimizes the mean-square truncation error:

$$\left\| a(\boldsymbol{x}, \omega) - \bar{a}(\boldsymbol{x}) - \sum_{k=1}^{s} \sqrt{\lambda_k} \xi_k(\omega) \psi_k(\boldsymbol{x}) \right\|_{L^2(\Omega, \mu; L^2(D))} = \left(\sum_{k=s+1}^{\infty} \lambda_k \right)^{1/2}.$$

- The random variables ξ_k are centered and uncorrelated, but not necessarily independent.
- If the random field $a(\boldsymbol{x}, \omega)$ is Gaussian – by definition, this means that $(a(\boldsymbol{x}_1, \omega), \ldots, a(\boldsymbol{x}_k, \omega))$ is a multivariate Gaussian random variable for all $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_k \in D$, $k \in \mathbb{N}$ – then the random variables ξ_k are independent.

The utility of the KL expansion comes from the fact that it is an effective method of representing *input* random fields when their covariance structure is known.

Essentially, if the (infinite-dimensional) input random field has a known covariance (which satisfies the conditions of Mercer's theorem), then we can use the KL expansion to find a finite-dimensional approximation, which is optimal in the mean-square error sense.

Example

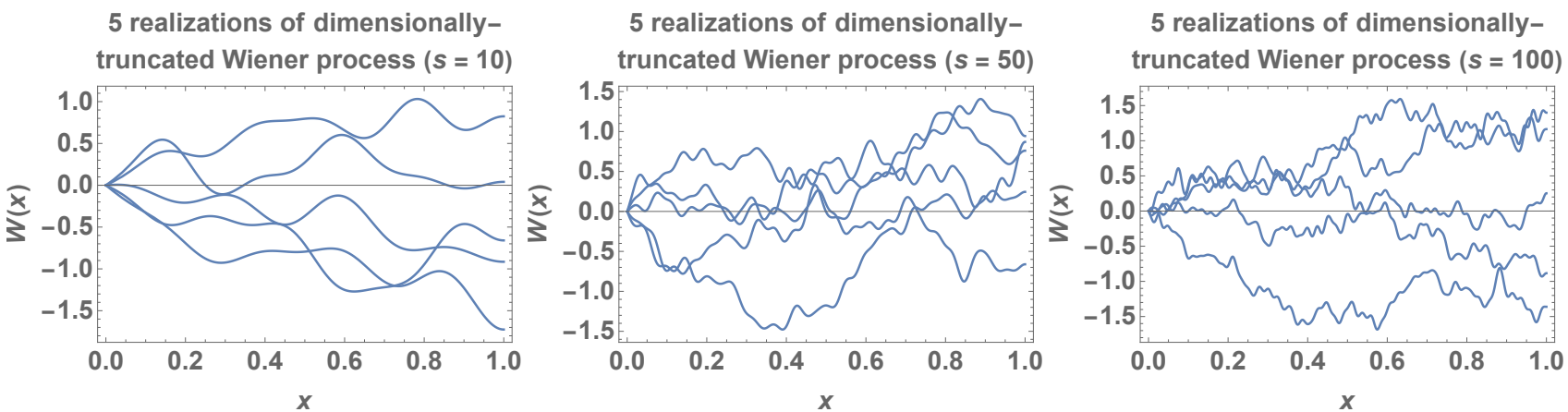
Let us consider the Wiener process over $D = [0, 1]$, which we regard as a centered standard Gaussian random field $W(x, \omega)$ with covariance function $K(x, y) = \min\{x, y\}$, $x, y \in [0, 1]$. It can be shown that

$$\int_0^1 K(x, y)\psi_k(y)\,\mathrm{d}y = \lambda_k \psi_k(x),$$

where $\psi_k(x) = \sqrt{2}\sin((k - \frac{1}{2})\pi x)$, $\lambda_k = \frac{1}{(k-\frac{1}{2})^2\pi^2}$. Then it has the KL expansion

$$W(x, \omega) = \sum_{k=1}^{\infty} \sqrt{\lambda_k} y_k(\omega)\psi_k(x), \quad y_k \sim \mathcal{N}(0, 1).$$

Let us plot some realizations with the series truncated to $s \in \{10, 50, 100\}$ terms.



Modeling assumptions

In engineering and practical applications, the idea is that we have some *a priori* knowledge/belief that the unknown input random field is distributed according to some known probability distribution with a known covariance.

- If the input random field is Gaussian with a known, nice covariance function[†], then we use the KL expansion to find a reasonable finite-dimensional approximation of true input. Since the KL expansion decorrelates the stochastic variables, and uncorrelated jointly Gaussian random variables are independent, we can exploit the independence of the stochastic variables to parameterize the model problem.
- If the input random field is *not Gaussian*, then the stochastic variables in the KL expansion are uncorrelated *but not necessarily independent*. For the purposes of mathematical analysis, we typically assume that the random variables in the input random field are independent so that we can parameterize the model problem. (Transforming dependent random variables into independent random variables can be done using, e.g., the Rosenblatt transformation, but this is computationally expensive.)

Note especially that in the Gaussian setting we do not need to make any "extra" effort to ensure the independence of the stochastic variables in the KL expansion.

[†]Matérn covariance is an especially popular choice.

Example (Lognormal input random field)

Let $D \subset \mathbb{R}^d$, $d \in \{2, 3\}$, be a Lipschitz domain and consider the PDE problem

$$\begin{cases} -\nabla \cdot (a(\boldsymbol{x}, \omega) \nabla u(\boldsymbol{x}, \omega)) = f(\boldsymbol{x}) & \text{for } \boldsymbol{x} \in D, \\ u(\cdot, \omega)|_{\partial D} = 0, \end{cases}$$

where $f : D \to \mathbb{R}$ is a fixed (deterministic) source term. We can model a lognormally distributed random diffusion coefficient $a \colon D \times \Omega \to \mathbb{R}$ using the KL expansion, e.g., as

$$a(\boldsymbol{x}, \omega) = a_0(\boldsymbol{x}) \exp\left(\sum_{k=1}^{\infty} y_k(\omega) \psi_k(\boldsymbol{x}) \right), \quad y_k \sim \mathcal{N}(0, 1),$$

where $a_0 \in L^\infty(D)$ is such that $a_0(\boldsymbol{x}) > 0$ and the random variables y_k are uncorrelated (and thus independent in the Gaussian case).

Due to the independence, we can consider the above as a *parametric PDE* with $a(\boldsymbol{x}, \boldsymbol{y}) \equiv a(\boldsymbol{x}, \boldsymbol{y}(\omega))$ and $u(\boldsymbol{x}, \boldsymbol{y}) \equiv u(\boldsymbol{x}, \boldsymbol{y}(\omega))$, where (formally) $\boldsymbol{y} \in \mathbb{R}^{\mathbb{N}}$ is a *parametric vector* endowed with a product Gaussian measure.

Example (Uniform and affine input random field)

Let $D \subset \mathbb{R}^d$, $d \in \{2, 3\}$, be a Lipschitz domain, $f : D \to \mathbb{R}$ is a fixed (deterministic) source term, and consider the PDE problem

$$\begin{cases} -\nabla \cdot (a(\boldsymbol{x}, \omega) \nabla u(\boldsymbol{x}, \omega)) = f(\boldsymbol{x}) & \text{for } \boldsymbol{x} \in D, \\ u(\cdot, \omega)|_{\partial D} = 0. \end{cases}$$

We can model a uniformly distributed random diffusion coefficient $a : D \times \Omega \to \mathbb{R}$ using the KL expansion, e.g., as

$$a(\boldsymbol{x}, \omega) = a_0(\boldsymbol{x}) + \sum_{k=1}^{\infty} y_k(\omega) \psi_k(\boldsymbol{x}), \quad y_k \sim \mathcal{U}(-\tfrac{1}{2}, \tfrac{1}{2}),$$

where the random variables y_k are uncorrelated. *Unlike the Gaussian setting, the random variables y_k are generally not independent!*

In numerical analysis, the random variables y_k are often **assumed** to be independent – this alllows us to consider the above as a parametric PDE with $a(\boldsymbol{x}, \boldsymbol{y}) \equiv a(\boldsymbol{x}, \boldsymbol{y}(\omega))$ and $u(\boldsymbol{x}, \boldsymbol{y}) \equiv u(\boldsymbol{x}, \boldsymbol{y}(\omega))$, where $\boldsymbol{y} \in [-\frac{1}{2}, \frac{1}{2}]^{\mathbb{N}}$ is a *parametric vector* endowed with a uniform probability measure.

The Monte Carlo method

A simple technique to approximate the integral

$$I(f) := \int_{\mathrm{supp}(p)} f(\boldsymbol{x}) p(\boldsymbol{x}) \, \mathrm{d}\boldsymbol{x}$$

is to use a sample average. If we are able to draw an i.i.d. sample $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n$ from the probability distribution corresponding to the PDF p then one can consider the Monte Carlo estimate

$$I_n^{\mathrm{MC}}(f) := \frac{1}{n} \sum_{i=1}^{n} f(\boldsymbol{x}_i).$$

Generally speaking, the Law of Large Numbers and the Central Limit Theorem imply that

$$\lim_{n \to \infty} I_n^{\mathrm{MC}}(f) = I(f) \quad \text{and} \quad \mathrm{Var}(I_n^{\mathrm{MC}}(f) - I(f)) \approx \frac{\mathrm{Var}(f(X))}{n}$$

provided that $f(X)$ has finite mean and variance with X distributed according to the probability distribution corresponding to p.

Model problem 1: uniform and affine model

For the purposes of numerical analysis, it is often desirable to start by analyzing a simpler model. Fix $f \in L^2(D)$, let $U = [-1/2, 1/2]^{\mathbb{N}}$, and consider the problem of finding, for all $\boldsymbol{y} \in U$, $u(\cdot, \boldsymbol{y}) \in H_0^1(D)$ such that

$$\int_D a(\boldsymbol{x}, \boldsymbol{y}) \nabla u(\boldsymbol{x}, \boldsymbol{y}) \cdot \nabla v(\boldsymbol{x}) \, \mathrm{d}\boldsymbol{x} = \int_D f(\boldsymbol{x}) v(\boldsymbol{x}) \, \mathrm{d}\boldsymbol{x} \quad \text{for all } v \in H_0^1(D),$$

where the diffusion coefficient has the parametrization

$$a(\boldsymbol{x}, \boldsymbol{y}) := a_0(\boldsymbol{x}) + \sum_{j=1}^{\infty} y_j \psi_j(\boldsymbol{x}), \quad \boldsymbol{x} \in D, \ \boldsymbol{y} \in U,$$

where $a_0 \in L^\infty(D)$, there exist $a_{\min}, a_{\max} > 0$ s.t. $0 < a_{\min} \leq a(\boldsymbol{x}, \boldsymbol{y}) \leq a_{\max} < \infty$ for all $\boldsymbol{x} \in D$ and $\boldsymbol{y} \in U$, and the *stochastic fluctuations* $\psi_j : D \to \mathbb{R}$ are functions of the spatial variable such that

- $\psi_j \in L^\infty(D)$ for all $j \in \mathbb{N}$,
- $\sum_{j=1}^{\infty} \|\psi_j\|_{L^\infty(D)} < \infty$.

Goals: compute $\mathbb{E}[G(u)]$ and $\mathrm{Var}[G(u)]$ for some bounded, linear functional $G : H_0^1(D) \to \mathbb{R}$ (*quantity of interest*); alternatively, one might be interested in $\mathbb{E}[u(\boldsymbol{x}, \cdot)]$ and $\mathrm{Var}[u(\boldsymbol{x}, \cdot)]$ (full PDE solution).

Model problem 2: lognormal model

In many practical applications, it is desirable to model the diffusion coefficient as a lognormal random field. Fix $f \in L^2(D)$, let $U = \mathbb{R}^{\mathbb{N}}_*$, and consider the problem of finding, for all $\boldsymbol{y} \in U$, $u(\cdot, \boldsymbol{y}) \in H^1_0(D)$ such that

$$\int_D a(\boldsymbol{x}, \boldsymbol{y}) \nabla u(\boldsymbol{x}, \boldsymbol{y}) \cdot \nabla v(\boldsymbol{x}) \, \mathrm{d}\boldsymbol{x} = \int_D f(\boldsymbol{x}) v(\boldsymbol{x}) \, \mathrm{d}\boldsymbol{x} \quad \text{for all } v \in H^1_0(D),$$

where the diffusion coefficient has the parametrization

$$a(\boldsymbol{x}, \boldsymbol{y}) := a_0(\boldsymbol{x}) \exp\left(\sum_{j=1}^{\infty} y_j \psi_j(\boldsymbol{x}) \right), \quad \boldsymbol{x} \in D, \ \boldsymbol{y} \in U,$$

where $a_0 \in L^\infty(D)$ is such that $a_0(\boldsymbol{x}) > 0$ and the *stochastic fluctuations* $\psi_j : D \to \mathbb{R}$ are functions of the spatial variable such that

- $\psi_j \in L^\infty(D)$ for all $j \in \mathbb{N}$,
- $\sum_{j=1}^{\infty} \|\psi_j\|_{L^\infty(D)} < \infty$.

Goals: compute $\mathbb{E}[G(u)]$ and $\mathrm{Var}[G(u)]$ for some bounded, linear functional $G : H^1_0(D) \to \mathbb{R}$; alternatively, one might be interested in $\mathbb{E}[u(\boldsymbol{x}, \cdot)]$ and $\mathrm{Var}[u(\boldsymbol{x}, \cdot)]$.

Here, $\mathbb{R}^{\mathbb{N}}_* := \{\boldsymbol{y} \in \mathbb{R}^{\mathbb{N}} \mid \sum_{j=1}^{\infty} |y_j| \|\psi_j\|_{L^\infty(D)} < \infty\}$. More on this condition later...

Numerical experiment

Let us consider the problem of calculating the (dimensionally-truncated) $\mathbb{E}[u_s(\boldsymbol{x}, \cdot)]$ using the Monte Carlo method. Fix the spatial domain $D = (0,1)^2$ and source term $f(\boldsymbol{x}) = x_1$. The PDE problem in this case is to find, for all $\boldsymbol{y} \in \mathbb{R}^s$, $u_s(\cdot, \boldsymbol{y}) \in H_0^1(D)$ s.t.

$$\int_D a_s(\boldsymbol{x}, \boldsymbol{y}) \nabla u_s(\boldsymbol{x}, \boldsymbol{y}) \cdot \nabla v(\boldsymbol{x}) \, \mathrm{d}\boldsymbol{x} = \int_D f(\boldsymbol{x}) v(\boldsymbol{x}) \, \mathrm{d}\boldsymbol{x} \quad \text{for all } v \in H_0^1(D)$$

endowed with the (dimensionally-truncated) lognormally parameterized diffusion coefficient

$$a_s(\boldsymbol{x}, \boldsymbol{y}) = \exp\left(\sum_{k=1}^{s} y_k \psi_k(\boldsymbol{x}) \right), \quad y_k \in \mathbb{R},$$

with stochastic fluctuations $\psi_k(\boldsymbol{x}) := k^{-\vartheta} \sin(\pi k x_1) \sin(\pi k x_2)$ and a fixed decay parameter $\vartheta > 1$. We solve the PDE using a first-order finite element method with mesh size $h = 2^{-5}$ and stochastic dimension $s = 100$. We draw a random sample $\boldsymbol{y}_1, \ldots, \boldsymbol{y}_n \sim \mathcal{N}(\boldsymbol{0}, I_s)$ and compute the Monte Carlo approximation

$$\mathbb{E}[u_{s,h}(\boldsymbol{x}, \boldsymbol{y})] \approx \frac{1}{n} \sum_{k=1}^{n} u_{s,h}(\boldsymbol{x}, \boldsymbol{y}_k) = I_n^{\mathrm{MC}}(u_{s,h}(\boldsymbol{x}, \cdot)).$$

We plot the estimated L^2 error by using $I_{n'}^{\mathrm{MC}}(u_{s,h}(\boldsymbol{x}, \cdot))$ for $n' \gg n$ as the reference solution and compute $\|\mathbb{E}[u_{s,h}] - I_n^{\mathrm{MC}}(u_{s,h})\|_{L^2(D)} \approx \|I_{n'}^{\mathrm{MC}}(u_{s,h}) - I_n^{\mathrm{MC}}(u_{s,h})\|_{L^2(D)}$. (To compute the $L^2(D)$-norm of a function $v_h = \sum_j c_j \phi_j \in V_h$ belonging to a FE space, we use the mass matrix $M_{i,j} = \int_D \phi_i(\boldsymbol{x}) \phi_j(\boldsymbol{x}) \, \mathrm{d}\boldsymbol{x}$ as $\|v_h\|_{L^2} = \sqrt{\boldsymbol{c}^{\mathrm{T}} M \boldsymbol{c}}$.)

Appendix

Rosenblatt transformation

In the non-Gaussian setting, the uncorrelated random variables can be made independent using, e.g., the *Rosenblatt transformation*.

The following is an excerpt from "Structural Reliability Analysis and Prediction", 3rd edition, by R. E. Melchers and A. T. Beck (2018).

A dependent random vector $\boldsymbol{X} = (X_1, \ldots, X_s)$ may be transformed to the *independent* uniformly distributed random vector $\boldsymbol{U} = (U_1, \ldots, U_s)$ through the Rosenblatt (1952) transformation $\boldsymbol{U} = T\boldsymbol{X}$ given by

$$
\begin{aligned}
u_1 &= \mathbb{P}(X_1 \le x_1) = F_1(x_1), \\
u_2 &= \mathbb{P}(X_2 \le x_2 | X_1 = x_1) = F_2(x_2|x_1), \\
&\vdots \\
u_s &= \mathbb{P}(X_s \le x_s | X_1 = x_1, \ldots, X_{s-1} = x_{s-1}) = F_s(x_s|x_1, \ldots, x_{s-1}).
\end{aligned}
$$

If the joint PDF $p_{\boldsymbol{X}}$ is known, then the conditional CDF F_s can be determined by

$$
F_s(x_s|x_1, \ldots, x_{s-1}) = \frac{\int_{-\infty}^{x_s} p_{X_1,\ldots,X_s}(x_1, \ldots, x_{s-1}, t)\,\mathrm{d}t}{p_{X_1,\ldots,X_{s-1}}(x_1, \ldots, x_{s-1})}.
$$