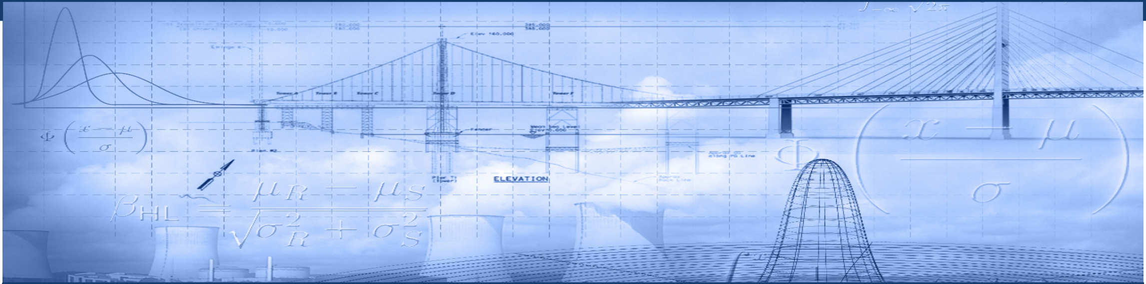
Stochastic polynomial chaos expansions for emulating stochastic simulators

ETH zürich



Stochastic polynomial chaos expansions for emulating stochastic simulators

Xujia Zhu, Bruno Sudret

Chair of Risk, Safety and Uncertainty Quantification, ETH Zurich

March 11, 2021 – MascotNum Workshop on Stochastic Simulators, Institut Henri Poincaré, Paris

Outline

Stochastic simulators

Stochastic polynomial chaos expansions
- Motivations
- Formulation
- Stochastic surrogate

Examples
- Bimodal toy example
- Geometric Brownian motion

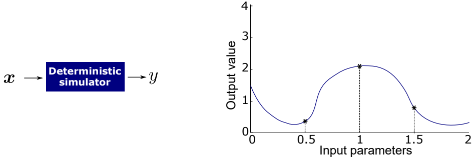
Conclusions & outlook

Deterministic vs. stochastic simulators

Deterministic simulators

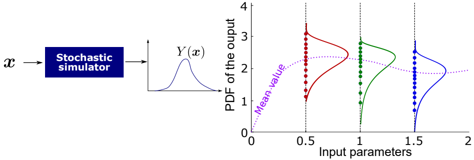
- Each set of input variables has a unique corresponding output

$$\mathcal{M}_d : \mathcal{D}_{\boldsymbol{X}} \subset \mathbb{R}^M \rightarrow \mathbb{R}$$



Stochastic simulators

- A given set of input parameters can lead to different values for the output
- $\mathcal{M}_s(\boldsymbol{x})$ is a random variable
- Source of randomness: $Y(\boldsymbol{x}) = \mathcal{M}_d(\boldsymbol{x}, \boldsymbol{Z})$, $\boldsymbol{Z}$ are latent variables



Stochastic emulators

Questions

- What are the mean/variance/quantiles or the probability density function (PDF) of the output for a given set of input parameters?
- What range (or distribution) of outputs can I expect? $\rightarrow$ uncertainty propagation
- What is the probability that the model response exceeds a threshold? $\rightarrow$ reliability analysis
- Which input variable is most important, which ones hardly affect the output? $\rightarrow$ sensitivity analysis
- ...

Stochastic emulators

Questions

- What are the mean/variance/quantiles or the probability density function (PDF) of the output for a given set of input parameters?
- What range (or distribution) of outputs can I expect? $\rightarrow$ uncertainty propagation
- What is the probability that the model response exceeds a threshold? $\rightarrow$ reliability analysis
- Which input variable is most important, which ones hardly affect the output? $\rightarrow$ sensitivity analysis
- ...

Large number of model runs

Stochastic emulators

Questions

- What are the mean/variance/quantiles or the probability density function (PDF) of the output for a given set of input parameters?
- What range (or distribution) of outputs can I expect? $\rightarrow$ uncertainty propagation
- What is the probability that the model response exceeds a threshold? $\rightarrow$ reliability analysis
- Which input variable is most important, which ones hardly affect the output? $\rightarrow$ sensitivity analysis
- ...

Large number of model runs

Need for surrogate models

- Easy to evaluate
- Accurate for estimating the response distribution
- Non-intrusive: treat the simulator as a black box

Outline

Stochastic simulators

Stochastic polynomial chaos expansions

- Motivations
- Formulation
- Stochastic surrogate

Examples

Conclusions & outlook

Motivations

Model description

- A random variable Y can be represented by

$$Y = \mathcal{M}_d(\boldsymbol{Z}) \overset{\mathrm{d}}{=} g(\tilde{Z}) = \sum_{\alpha \in \mathcal{A}} c_\alpha \Psi_\alpha(\tilde{Z})$$

- $\tilde{Z}$ is a latent random variable with the cumulative distribution function $F_{\tilde{Z}}$
- Ψ_α's are the associated polynomial chaos basis functions, i.e., $\mathbb{E}\left[\Psi_{\alpha_1}(\tilde{Z})\,\Psi_{\alpha_2}(\tilde{Z})\right] = \delta_{\alpha_1 \alpha_2}$
- Weak approximation[1]

Mathematical foundation

- Iso-probabilistic transform $Y \overset{\mathrm{d}}{=} F_Y^{-1}(F_{\tilde{Z}}(\tilde{Z}))$
- Cameron-Martin theorem $Y \overset{\mathrm{d}}{=} F_Y^{-1}(F_{\tilde{Z}}(\tilde{Z})) = \sum_{\alpha=0}^{+\infty} c_\alpha \Psi_\alpha(\tilde{Z})$
- Convergence in L^2 implies convergence in distribution

[1] Xiu (2010). *Numerical methods for stochastic computations: a spectral method approach.* Princeton University Press

However...

Estimation method

- Maximum likelihood estimation (MLE) [2]: $f_Y(y) = \sum_j \frac{f_{\tilde{Z}}(\tilde{z}_j)}{|g'(\tilde{z}_j)|}$ with $\tilde{z}_j$ being the j-th root of the equation $y = g(\tilde{z}) = \sum_{\alpha \in \mathcal{A}} c_\alpha \Psi_\alpha(\tilde{z})$
- The likelihood can reach infinity
- Remedy: forcing $g(\cdot)$ to be strictly monotone, but nontrivial

[2] Desceliers et al. (2006). *Maximum likelihood estimation of stochastic chaos representations from experimental data*, Int. J. Numer. Meth. Engng. (66): 978–1001

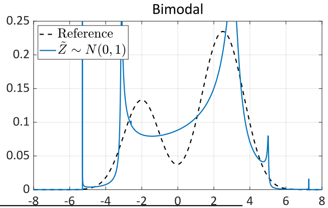
However...

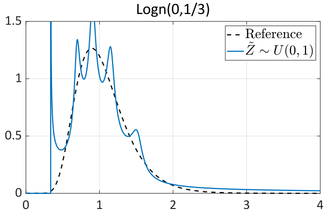
Estimation method

- Maximum likelihood estimation (MLE) [2]: $f_Y(y) = \sum_j \frac{f_{\tilde{Z}}(\tilde{z}_j)}{|g'(\tilde{z}_j)|}$ with $\tilde{z}_j$ being the j-th root of the equation $y = g(\tilde{z}) = \sum_{\alpha \in \mathcal{A}} c_\alpha \Psi_\alpha(\tilde{z})$
- The likelihood can reach infinity
- Remedy: forcing $g(\cdot)$ to be strictly monotone, but nontrivial

Oracle

- The best fit would be to build a surrogate directly from $Y = F_Y^{-1}(F_{\tilde{Z}}(\tilde{Z}))$





[2] Desceliers et al. (2006). *Maximum likelihood estimation of stochastic chaos representations from experimental data*, Int. J. Numer. Meth. Engng. (66): 978–1001

Stochastic polynomial chaos expansions

$$Y \stackrel{\mathrm{d}}{=} g(\tilde{Z}) + \epsilon = \sum_{\alpha \in \mathcal{A}} c_\alpha \Psi_\alpha(\tilde{Z}) + \epsilon$$

- Additional noise variable $\epsilon \sim \mathcal{N}(0, \sigma^2)$
- ϵ is a smoother:

$$f_Y(y) = \int_{-\infty}^{+\infty} f_{Y \mid \tilde{Z}}(y \mid \tilde{z}) \, f_{\tilde{Z}}(\tilde{z}) \mathrm{d}\tilde{z} = \int_{-\infty}^{+\infty} f_\epsilon(y - g(\tilde{z}); \sigma) \, f_{\tilde{Z}}(\tilde{z}) \, \mathrm{d}\tilde{z}$$
$$= \int_{-\infty}^{+\infty} f_\epsilon(y - t; \sigma) \sum_j \frac{f_{\tilde{Z}}(\tilde{z}_j)}{|g'(\tilde{z}_j)|} \, \mathrm{d}t = (f_T * f_\epsilon)(y)$$

  where $T = g(\tilde{Z})$, and $\tilde{z}_j$ is the j-th root of the equation $t = g(\tilde{z})$
- c_α's and σ should be estimated from data

Estimation method

Maximum likelihood estimation

- The likelihood for $Y = y$ can be written as:

$$l(\boldsymbol{c}, \sigma; y) = \int_{\mathcal{D}_{\tilde{Z}}} \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{\left(y - \sum_{\alpha \in \mathcal{A}} c_\alpha \Psi_\alpha(\tilde{z})\right)^2}{2\sigma^2} \right) f_{\tilde{Z}}(\tilde{z}) \mathrm{d}\tilde{z}$$

- The integral is calculated by quadrature:

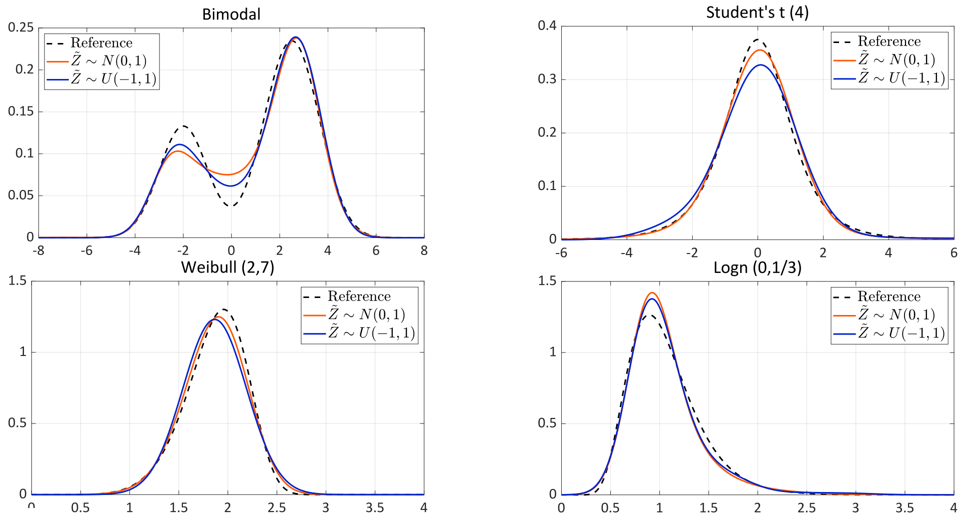
$$l(\boldsymbol{c}, \sigma; y) \approx \tilde{l}(\boldsymbol{c}, \sigma; y) = \sum_{j=1}^{N_Q} \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{\left(y - \sum_{\alpha \in \mathcal{A}} c_\alpha \Psi_\alpha(\tilde{z}_j)\right)^2}{2\sigma^2} \right) w_j$$

- For a given σ, the maximum likelihood estimation is used for estimating the coefficients from data $\mathcal{Y} = \left\{ y^{(1)}, \ldots, y^{(N)} \right\}$: $\hat{\boldsymbol{c}} = \arg\max_{\boldsymbol{c}} \sum_{i=1}^{N} \log\left(\tilde{l}\left(\boldsymbol{c}, \sigma; y^{(i)} \right) \right)$

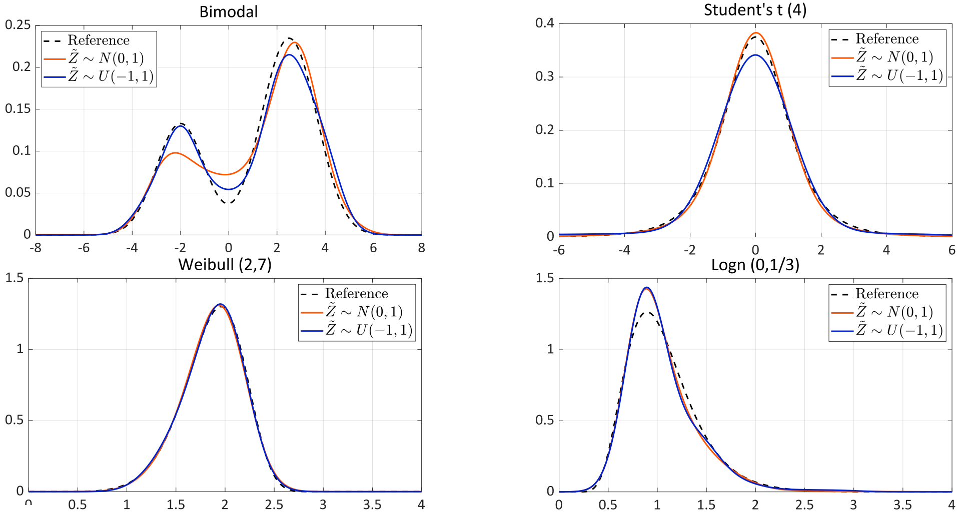
Cross-validation

- For fixed basis functions, cross-validation is applied for σ selection

Preliminary results (500 training points)



Preliminary results (1000 training points)



Remarks

Properties

- The stochastic PCE is a latent variable model, and also a generative model
- The additional noise variable ϵ is used to prevent overfitting
- The equality in $Y \overset{\mathrm{d}}{=} g(\tilde{Z}) = \sum_{\alpha \in \mathcal{A}} c_\alpha \Psi_\alpha(\tilde{Z}) + \epsilon$ is in distribution, i.e., the latent variable $\tilde{Z}$ does not have real physical meanings

Links with machine learning

- Variational autoencoder[3]: variational Bayesian, Evidence Lower BOund (ELBO)
- Normalizing flow[4]: special bijective and easy-to-inverse functions $\boldsymbol{Y} \overset{\mathrm{d}}{=} \boldsymbol{g}(\tilde{\boldsymbol{Z}})$
- GAN[5]: adversarial loss (minimax)

[3] Kingma and Welling (2014). *Auto-encoding variational Bayes.*, Proc. of the 2nd International Conference on Learning Representations (ICLR)

[4] Dinh et. al. (2016). *Density estimation using real NVP.* arXiv:1605.08803

[5] Goodfellow et. al. (2014). *Generative Adversarial Networks.* Proc. of the International Conference on Neural Information Processing Systems (NeurIPS): 2672–2680

Stochastic PCE as a stochastic surrogate

Formulation

$$Y(\boldsymbol{x}) \stackrel{\mathrm{d}}{=} \sum_{\boldsymbol{\alpha}\in\mathcal{A}} c_{\boldsymbol{\alpha}} \Psi_{\boldsymbol{\alpha}}(\boldsymbol{x}, \tilde{Z}) + \epsilon$$

- For a given x, $Y(\boldsymbol{x})$ is a function of the latent variable $\tilde{Z}$ perturbed by a Gaussian additive noise ϵ

Estimation

- The conditional likelihood for a data point $(\boldsymbol{x}, y)$ is

$$l(\boldsymbol{c}, \sigma; \boldsymbol{x}, y) = \int_{\mathcal{D}_{\tilde{Z}}} \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{\left(y - \sum_{\boldsymbol{\alpha}\in\mathcal{A}} c_{\boldsymbol{\alpha}} \Psi(\boldsymbol{x}, \tilde{z})\right)^2}{2\sigma^2}\right) f_{\tilde{Z}}(\tilde{z}) \mathrm{d}\tilde{z}$$

- Numerical integration by quadrature $l(\boldsymbol{c}, \sigma; \boldsymbol{x}, y) \approx \tilde{l}(\boldsymbol{c}, \sigma; \boldsymbol{x}, y)$
- Maximum likelihood to estimate the PC coefficients $\hat{\boldsymbol{c}} = \arg\max_{\boldsymbol{c}} \sum_{i=1}^{N} \log\left(\tilde{l}\left(\boldsymbol{c}, \sigma; \boldsymbol{x}^{(i)}, y^{(i)}\right)\right)$
- Cross-validation for σ selection

Cross-validation

- The data set is divided into N_{cv} equal sized groups $\{V_k : k = 1, \dots, N_{\mathrm{cv}}\}$
- For $k \in \{1, \dots, N_{\mathrm{cv}}\}$, we pick the $k-$th group V_k as the validation set and the other $N_{\mathrm{cv}} - 1$ folds $V_{\sim k}$ as training set to build a stochastic PCE,

$$\hat{\boldsymbol{c}}_k(\sigma) = \arg\max_{\boldsymbol{c}} \sum_{i \in V_{\sim k}} \log\left(\tilde{l}\left(\boldsymbol{c}, \sigma; \boldsymbol{x}^{(i)}, y^{(i)}\right)\right)$$

- We evaluate the likelihood on the validation set, which gives the out-of-sample performance

$$\mathbf{l}_k(\sigma) = \sum_{i \in V_k} \log\left(\tilde{l}\left(\hat{\boldsymbol{c}}_k(\sigma), \sigma; \boldsymbol{x}^{(i)}, y^{(i)}\right)\right)$$

- We repeat the procedure for all $k \in \{1, \dots, N_{\mathrm{cv}}\}$, and the optimal σ^* is selected as

$$\sigma^* = \arg\max_{\sigma} \sum_{k=1}^{N_{\mathrm{cv}}} \mathbf{l}_k(\sigma)$$

Model selection

Adaptivity

- Truncation scheme $\mathcal{A}^{p,q,M} = \left\{ \boldsymbol{\alpha} \in \mathbb{N}^M : \|\boldsymbol{\alpha}\|_q \overset{\text{def}}{=} \left(\sum_{i=1}^{M} \alpha_i^q\right)^{\frac{1}{q}} \leq p \right\}$
  - Choose the appropriate degree (p-adaptivity)
  - Choose the appropriate q-norm (q-norm-adaptivity)
- Choose the appropriate latent variable: normal or uniform

Selection criteria

- Information criterion, e.g., Akaike information criterion (AIC), Bayesian information criterion (BIC)
- Cross-validation, e.g., the same quantity as for σ-selection

Outline

Stochastic simulators

Stochastic polynomial chaos expansions

Examples
- Bimodal toy example
- Geometric Brownian motion

Conclusions & outlook

Error metric

Wasserstein distance (of order 2)

- For continuous random variables, it is the L^2 distance between the quantile functions:

$$d_{\mathrm{WS}}^2(Y, \hat{Y}) = \|Q_Y - Q_{\hat{Y}}\|_{L^2}^2$$

Normalized Wasserstein distance

$$\varepsilon = \frac{\mathbb{E}_{\boldsymbol{X}}\left[d_{\mathrm{WS}}^2\left(Y(\boldsymbol{X}), \hat{Y}(\boldsymbol{X})\right)\right]}{\mathrm{Var}\left[Y\right]}$$

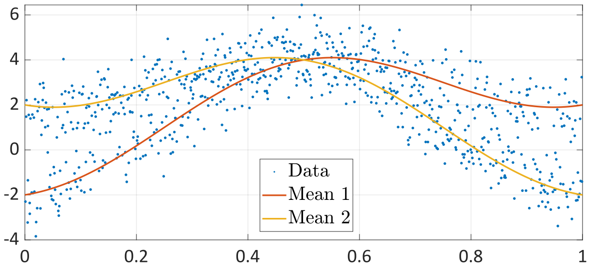
- Shift and scale invariant

Bimodal toy example

Description of the simulator

$$f_{Y|X}(y \mid X = x) = 0.6\, f_n(4\sin^2(\pi \cdot x) + 4x - 2) + 0.4\, f_n(4\sin^2(\pi \cdot x) - 4x + 2)$$

- f_n is the PDF of a normal distribution with mean 0 and standard deviation 0.8.
- The response distribution is a mixture of Gaussian PDFs
- $X \sim \mathcal{U}(0, 1)$



Bimodal toy example

Description of the simulator

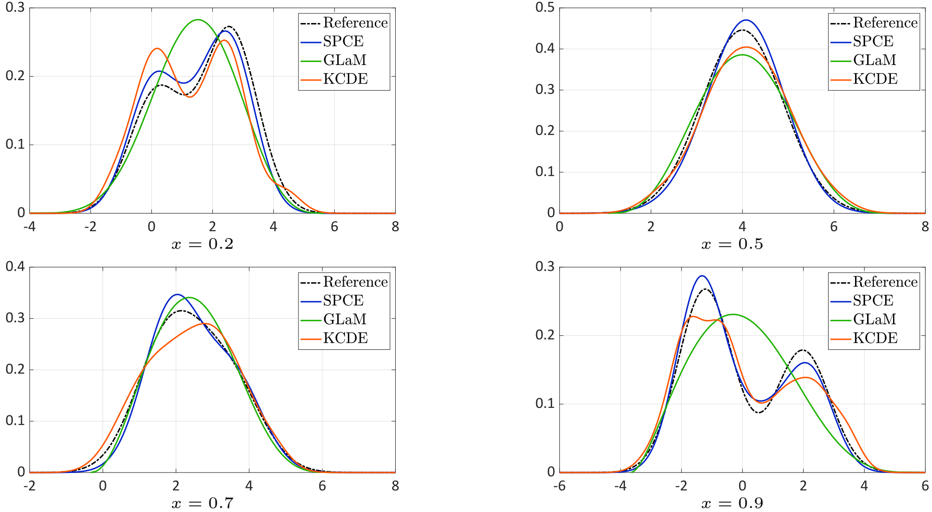
$$f_{Y|X}(y \mid X = x) = 0.6\, f_n(4\sin^2(\pi \cdot x) + 4x - 2) + 0.4\, f_n(4\sin^2(\pi \cdot x) - 4x + 2)$$

- f_n is the PDF of a normal distribution with mean 0 and standard deviation 0.8.
- The response distribution is a mixture of Gaussian PDFs
- $X \sim \mathcal{U}(0, 1)$

Surrogate setting

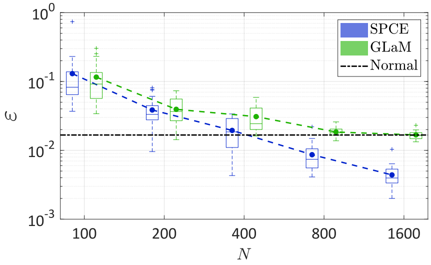
- Experimental design (ED): $\boldsymbol{\mathcal{X}}$ are generated using the Latin hypercube sampling
- Degree and q-norm adaptivity: $p \in \{1, 2, \ldots, 5\}$ and $q \in \{0.5, 0.75, 1\}$
- Latent variable selection: $\tilde{Z} \sim \mathcal{N}(0, 1)$ and $\tilde{Z} \sim \mathcal{U}(-1, 1)$

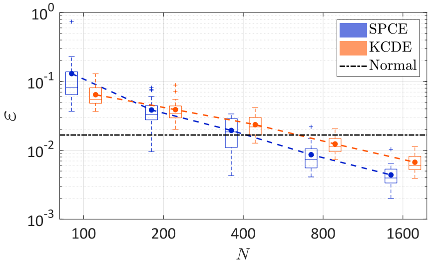
PDF predictions (ED with $N = 800$)



Convergence study

- Experimental design of size 100, 200, 400, 800, 1600
- 20 independent runs for each scenario
- Normalized Wasserstein distance as a performance indicator





Geometric Brownian motion

$$\mathrm{d}S_t = x_1\, S_t\, \mathrm{d}t + x_2\, S_t\, \mathrm{d}W_t$$

- S_t: price process, W_t: Wiener process, x_1: drift, x_2: volatility
- $X_1 \sim \mathcal{U}(0, 0.1)$, $X_2 \sim \mathcal{U}(0.1, 0.4)$, and $Y(\boldsymbol{x}) = S_1(\boldsymbol{x})$
- The analytical distribution of S_t exists (Itô's calculus)

$$S_1(\boldsymbol{x})/S_0 \sim \mathcal{LN}\left(x_1 - \frac{x_2^2}{2}, x_2\right)$$

Geometric Brownian motion

$$\mathrm{d}S_t = x_1\, S_t\, \mathrm{d}t + x_2\, S_t\, \mathrm{d}W_t$$

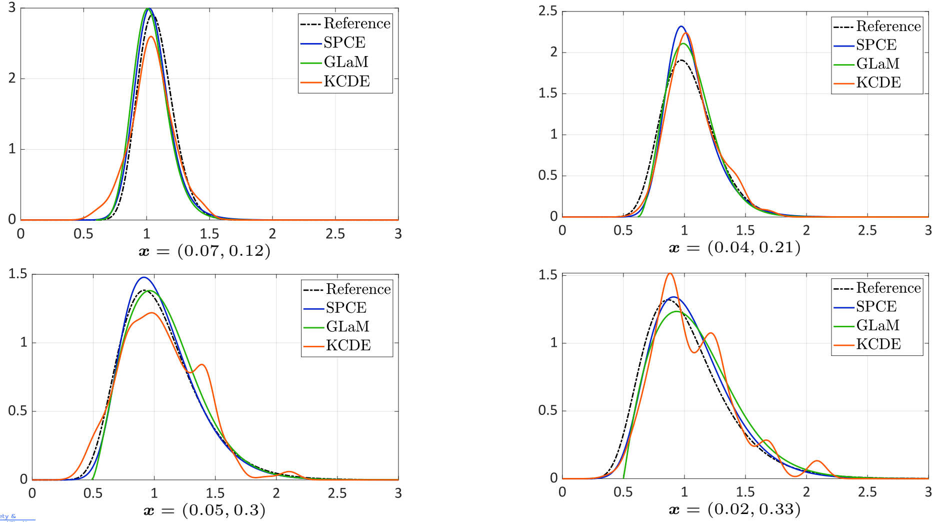
- S_t: price process, W_t: Wiener process, x_1: drift, x_2: volatility
- $X_1 \sim \mathcal{U}(0, 0.1)$, $X_2 \sim \mathcal{U}(0.1, 0.4)$, and $Y(\boldsymbol{x}) = S_1(\boldsymbol{x})$
- The analytical distribution of S_t exists (Itô's calculus)

$$S_1(\boldsymbol{x})/S_0 \sim \mathcal{LN}\left(x_1 - \frac{x_2^2}{2}, x_2\right)$$

Surrogate setting

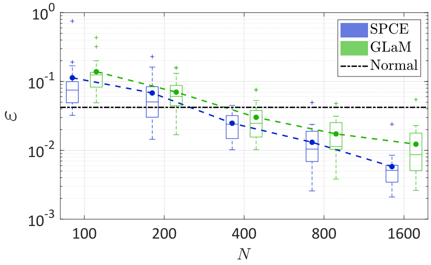
- Experimental design (ED): $\boldsymbol{\mathcal{X}}$ are generated using the Latin hypercube sampling
- Degree and q-norm adaptivity: $p \in \{1, 2, \ldots, 5\}$ and $q \in \{0.5, 0.75, 1\}$
- Latent variable selection: $\tilde{Z} \sim \mathcal{N}(0, 1)$ and $\tilde{Z} \sim \mathcal{U}(-1, 1)$

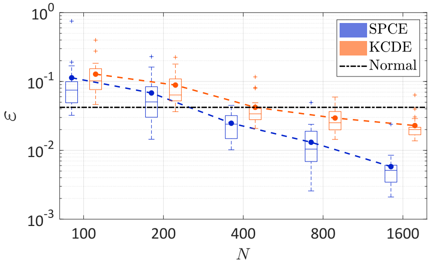
PDF predictions (ED with $N = 400$)



Convergence study

- Experimental design of size 100, 200, 400, 800, 1600
- 20 independent runs for each scenario
- Normalized Wasserstein distance as a performance indicator





Outline

Stochastic simulators

Stochastic polynomial chaos expansions

Examples

Conclusions & outlook

Conclusions & outlook

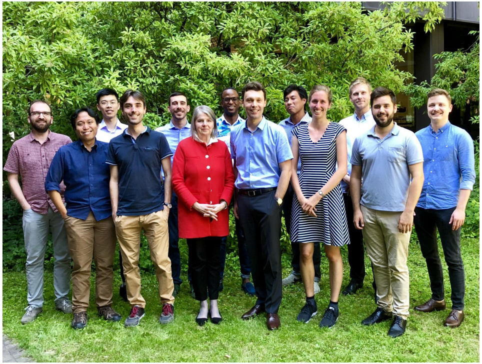
Conclusions

- Stochastic PCE is developed for general statistical inference of distributions
- On top of the input variables, a latent variable and a noise variable are introduced to represent the intrinsic randomness of a stochastic simulator
- The PC coefficients are calibrated by the maximum likelihood estimation
- The noise variable is a smoother, and its standard deviation σ is a hyperparameter
- The proposed algorithm does not require replications

Outlook

- Theoretical properties of the surrogate model (consistency, asymptotic properties)
- More applications
- Sparse stochastic PCE
- Conditional Wasserstein GAN [6]

[6] Ariovsky et al. (2017). *Wasserstein Generative Adversarial Networks*, Proc. of the 34th International Conference on Machine Learning (70): 214–223



Chair of Risk, Safety & Uncertainty Quantification

www.rsuq.ethz.ch

The Uncertainty Quantification Software

www.uqlab.com

UQLab

The Uncertainty Quantification Community

www.uqworld.org

UQWorld

Thank you very much for your attention !

Statistical models

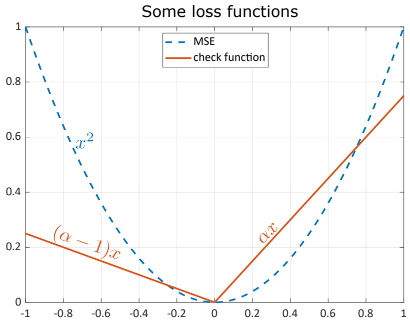
Statistical assumptions

- Data generation process, e.g., linear models

$$Y = a\,X + b + \epsilon$$

Estimation method

- A framework to infer the model
- Loss function, e.g. mean estimation: mean-squared error, quantile estimation: check function loss



Statistical models

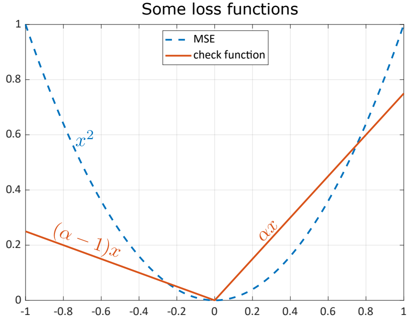
Statistical assumptions

- Data generation process, e.g., linear models

$$Y = a\,X + b + \epsilon$$

Estimation method

- A framework to infer the model
- Loss function, e.g. mean estimation: mean-squared error, quantile estimation: check function loss



Distribution estimation

- Parametric family (e.g. exponential family): maximum likelihood estimation
- Kernel estimation: $f(y \mid \boldsymbol{x}) = f(\boldsymbol{x}, y)/f(\boldsymbol{x})$
- Latent variable models: variational auto-encoder, GAN