Non-Parametric Surrogate Models for Uncertainty Quantification in Structural Vibration

HyeongUk Lim (student of Dr. Lance Manuel)

The University of Texas at Austin

18th Dec 2018
Outline

1. Introduction

2. Surrogate Model: Polynomial Chaos Expansion (PCE)

3. Engineering Applications
   - Fatigue Damage in a Marine Riser
   - Long-term Extreme Response of a Moored Structure
   - Uncertainty in Floor Vibration System
   - Structural Reliability using Surrogate Models

4. Conclusions
About Me; Memory in KICT

- B.E., Kyung Hee University, 2006-2012
- M.S., KAIST, 2012-2014

- Research internship during 10/2014-05/2015
  - Worked in Geotech Department
  - With many senior researchers
  - For on-line rock mass classification

- PhD candidate, UT Austin, 2015-2019 summer (expected)

**Major:** Civil Engineering, focused on numerical simulations using applied probabilities

**Interests:** Structural Vibration and Reliability via **Uncertainty Quantification**

**Applications:** Offshore structures, Wooden floors, and actively searching for others . . .
Engineering complex system requires reliable prediction (e.g., by computer simulations)
Purpose of Computer Simulation

- To inform some decision (for design, operations, or control)
  - Quantities are predicted to inform decision
  - These are Quantities of Interest (QoIs)
  - In most cases, need to predict QoIs when confirming observational data is not available

- Computational model is not scientific theory
  - Their validity depends on their purpose, such as
  - (1) nature of QoIs and (2) required accuracy

We need validation of computational models to predict QoIs well

How? ⇒ by Uncertainty Quantification (UQ)
Validation Requires UQ

- **Validation**: systematic comparison of model outputs to measurements
  - Confirm model is "close" to observations
  - What about the unobserved QoI?

- Such comparisons are necessary but not **sufficient** for prediction
  - Predictions are necessarily an extrapolation from available information
  - Need to assess confidence in the extrapolation

Validation requires **UQ**:
- How close between observations and model outputs?
- How strong is a validation test?
- Uncertainty in data, models and model parameters

(Ref: "Validating Computational Predictions of Unobserved Quantities" by R. D. Moser)
Uncertainty Quantification

- Uncertainties in “inputs” are modeled by PDFs:
  \[ X \sim f_X \]

- A computation model relates inputs to any “response” of interest:
  \[ M : y \equiv M(x) \]

- Uncertainty Quantification seeks statistics of \( Y \)
  e.g., \( \mathbb{E}[Y] \), \( \text{Var}(Y) \), \( f_Y \), \( F_Y \), \( P(Y > b) \)
  
  mean, variance, pdf, cdf, exceedance

- Sensitivity analysis: to assess dominant sources of uncertainty
Sources of Uncertainty

- **Aleatory and Epistemic:**
  - natural randomness (e.g., ocean waves, material variability)
  - natural hazards (earthquakes, floods, landslides)
  - exceptional service loads
  - climate loads (hurricanes, snow storms, hail)
  - accidental human actions (explosions, fire)
  - lack of knowledge, limited data, imperfect models
Computational Model: $\mathcal{M}$

Complex physical systems assessed using computational models

![Diagram](image)

(1) Validation + UQ

(2) Verification

Two considerations:

- Mathematical idealization of physical phenomenon
- Computational implementation of numerical solution for math model
Surrogate Model: $\hat{M}$

A surrogate model, $\hat{M}$, is an approximation of the computational model, $M$, that is constructed, based upon:

- some runs of the computation model (truth/original)
- assumed characteristics of the surrogate (e.g., polynomial form)

$$M(x) \approx \hat{M}(x)$$

minutes or hours per a run
vs.
seconds per $10^6$ runs

Challenge: Surrogates require rigorous validation (against the “truth”)
PCE: Formulation and Computation of Coefficients

Formation of PCE:

$$\hat{M}(x) = \sum_{i=0}^{P} c_i \Psi_i(x),$$

Parameters to be chosen carefully: $c_i$, $P$, and also $\Psi_i$

Estimation of $c_i$:

- Need a set of samples, $D = \{x_j, y_j\}_{j=1}^{N}$
- Post-process of the set, $\{\hat{M}(x_j), y_j\}_{j=1}^{N}$

Estimation of $P$: Heuristics

$\Psi_i$: Askey scheme (parametric)
PCE by Non-parametric Way

Formation of PCE:

$$\hat{M}(x) = \sum_{i=0}^{P} c_i \Psi_i(x),$$

When a limited number of data is available, $\Psi$ can be determined by a non-parametric way, so called Arbitrary PCE (APCE):

- Using the sample raw moments of random variables
- And Gran-Schmidt orthogonalization
- Can be generalized for dependent random variables

Challenge:

- The number of available data matters
Application 1: Fatigue Damage in a Marine Riser
Introduction

- Fatigue damage in risers undergoing vortex-induced vibration (VIV) remains a challenging problem $\Rightarrow$ great uncertainty associated with it.

- A distributed wake oscillator model (Srinil, 2010) has been proposed and its effectiveness demonstrated for different flow conditions and riser configurations.

- An Uncertainty Quantity (UQ) framework is presented – using up to 3 random variables – to assess fatigue damage uncertainty.

- Case studies with different UQ surrogate models are compared vs. Monte Carlo simulation, while addressing efficiency and accuracy.
Structure Considered

Top-tensioned riser in uniform flow

Truth system:
- Governing PDEs ⇒ Standard Galerkin method
- Obtain displacement time series at $x$ from the top
- Estimate fatigue damage using post-processing

Random variable inputs:
- $\Delta A_{max}, \Delta \omega$: empirical coefficients
- $V$: current velocity

Validation:
- 20,000 runs of the truth system
- Exceedance probabilities
Fatigue damage is estimated at $x = L/4$, $L/2$, and $3L/4$ from the top:

- Measurement data (used in PDEs) collected over several references
- Fit data to the empirical curves
Random Variables: $\Delta A_{max}, \Delta \omega$

- Measurement data (used in PDEs) collected over several references
- Fit data to the empirical curves
Random Variables: $\Delta A_{max}$, $\Delta \omega$

- Modeling the residuals by Shifted Generalized Lognormal Distribution (SGLD)
Monte Carlo Simulations

- Monte Carlo simulations with(out) correlation between variables
- Dots go to red as fatigue damage is high
- Fatigue damage is obtained by the rainflow cycle counting
**MCS vs. PCE**

- Comparison of pdf, QQ plots, and exceedance probability plots
- 10 sets of MCS and PCE
- PCE: order-5 ($P$) and $5^3$ quadrature points (for $c_i$)
Application 2: Long-term Extreme Response of a Moored Structure
Introduction

- Long-term analysis using Monte Carlo Simulation (MCS) is expensive.
- Polynomial Chaos Expansion (PCE) framework for predicting long-term surge motion extreme response of a moored floating structure.
- We consider uncertainty in (1) environmental variables and (2) time-varying short-term loads.
- Validation studies involve PCE vs. MCS long-term response on the simple moored floating structure.
Problem Formulation

\[ M \ddot{u}(t) + 2\zeta \sqrt{K} M \dot{u}(t) + Ku(t) = F_{WF}(t) + F_{LF}(t) \]

\[ F^{(1)}(\omega) = T^{(1)}(\omega)\eta(\omega) \]
\[ F^{(2)}(\omega_r, \omega_s) = T^{(2)}(\omega_r, \omega_s)\eta(\omega_r)\eta(\omega_s) \]

\[ u(t) = \sum_{r=1}^{R} \underbrace{H(\omega_r) T^{(1)}(\omega_r)}_{\text{FRF}} \underbrace{A_r e^{i\omega_r t}}_{\text{Wave amplitude}} + \sum_{r=1}^{R} \sum_{s=1}^{R} \sum_{r \neq s} \underbrace{H(\omega_r - \omega_s) T^{(2)}(\omega_r, \omega_s) A_r A_s^* e^{i(\omega_r - \omega_s) t}}_{\text{Transfer}} \]

QoI: \( \max\{u(t), T = 30 \text{ min}\} \)
Comparison of Exceedance Probability

- Lower probability level is of concern.

- More samples for PCE show small error in the estimation of PoEs.
PCE model with order-3 and 100 samples estimate PoEs well.

Importance sampling can be applied to PCE as well.
Application 3: Uncertainty in Floor Vibration System
Introduction

- Wooden multi-story buildings have increased their market share in Europe during the last decades.

- Wooden buildings are more sensitive to dynamic loads at low frequencies due to the variability in the material properties of wood.

- For example, local variations occur even if wooden members are taken from the same batch, because wood is a natural material.

- The uncertainty in modal frequencies is studied.
Introduction
Mode shapes: simply supported at two edges

An wooden floor system with 7 joists:
UQ for modal frequencies: using SS (LHS) and MCS

(SS = Systematic Sampling (Stratified using Latin Hypercube))

- (top) 40 SS vs. (bottom) \( \approx 10,000 \) MCS
- densities from CDFs at 0.0125, 0.0375, ..., 0.9875
- SS - good match with MCS
Extremes - Tail Probabilities

- (top) order-1, (mid) 2, (bottom) 3
- Samples for MCS & PCE: \( \approx 10,000 \)
- 10 sets of MCS & PCE (bootstrap resamplings for MCS)
- order-3 PCE suggests “convergence”
Application 4: Structural Reliability using Surrogate Models
Summary for Examples

Ex 1: \[ g(x) = x_1 + 2x_2 + 2x_3 + x_4 - 5x_5 - 5x_6 + 0.001 \sum_{i=1}^{6} \sin(100x_i) \]

Ex 2: \[ g(x) = 1.1 - 0.00115x_1x_2 + 0.00157x_2^2 + 0.00117x_1^2 + 0.0135x_2x_3 - 0.0705x_2 - 0.00534x_1 - 0.0149x_1x_3 - 0.0611x_2x_4 + 0.071x_1x_4 - 0.226x_3 + 0.0333x_3^2 - 0.558x_3x_4 + 0.998x_4 - 1.339x_4^2 \]

Ex 3: \[ g(x) = x_1 - x_2 \]

Ex 4: \[ g(x) = \frac{6x_1x_3(x_2-x_3)x_5}{x_2(x_6x_3^3-(x_6-x_7)(x_5-2x_8)^3)} - x_4 \]

<table>
<thead>
<tr>
<th>Ex</th>
<th>Ref</th>
<th>Total Degree</th>
<th># of RVs</th>
<th>Dependency</th>
<th>Type of RVs</th>
<th>Purpose</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Liu and Der Kiureghian (1991)</td>
<td>1</td>
<td>6</td>
<td>no</td>
<td>all lognormals</td>
<td>noisy limit state</td>
</tr>
<tr>
<td>2</td>
<td>Liu and Der Kiureghian (1991)</td>
<td>2</td>
<td>4</td>
<td>no</td>
<td>( X_1 ) : type II extreme quadratic function ( X_2,3 ) : normal ( X_4 ) : lognormal</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>Der Kiureghian and Liu (1986)</td>
<td>1</td>
<td>2</td>
<td>yes, ( \rho_{X_1X_2} = 0.5 )</td>
<td>( X_1 ) : uniform ( X_2 ) : exponential</td>
<td>dependent RVs</td>
</tr>
<tr>
<td>4</td>
<td>Huang and Du (2006)</td>
<td>2</td>
<td>8</td>
<td>no</td>
<td>all normals</td>
<td></td>
</tr>
</tbody>
</table>
### Summary for Examples

Ex 5: \( g_5(\mathbf{x}) = 4 - x_1/4 + \sin(5x_1) - x_2 \)

Ex 6: \( g_6(\mathbf{x}) = \sum_{i=1}^{20} x_i - 8.951, \quad g_6(\mathbf{x}) = -\sum_{i=1}^{20} x_i + 36.720 \)

Ex 7: \( g_7(\mathbf{x}) = x_7 - 3x_4 \sqrt{\frac{\pi x_8}{4x_6 A^3}} \left[ \frac{B x_6}{x_5 x_6 (4B^2 + C^2) + DB^2} \frac{(x_5 E^3 + x_6 A^3) E}{4BF^4} \right] \),

\( A, B, C, D, E, F \) are the functions of \( \mathbf{x} \)

<table>
<thead>
<tr>
<th>Ref</th>
<th>Total Degree</th>
<th># of RVs</th>
<th>Dependency</th>
<th>Type of RVs</th>
<th>Purpose</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ex 5</td>
<td>Sundar and Shields (2016)</td>
<td>1</td>
<td>2</td>
<td>no</td>
<td>all normals</td>
</tr>
<tr>
<td>Ex 6</td>
<td>Engelund and Rackwitz (1993)</td>
<td>1</td>
<td>20</td>
<td>no</td>
<td>all exponentials</td>
</tr>
<tr>
<td>Ex 7</td>
<td>Der Kiureghian and De Stefano (1991)</td>
<td>2</td>
<td>8</td>
<td>no</td>
<td>all lognormals</td>
</tr>
</tbody>
</table>
Estimation of Probability of Failure

- Geometrical approximation of a limit state function (FORM/SORM):
  - Optimization algorithms (iterative search)
  - Can be considered as linear/quadratic limit state function approximation

- Advanced sampling:
  - Importance sampling (IS)
  - Subset simulation (SS)
  - and many variants . . .

- Surrogate limit state function: there are cases where each are useful
  - Polynomial Chaos Expansion (PCE)
  - Kriging
  - Support Vector Machine (SVM)
  - Active-subspace based surrogate
  - and many variants . . .

- Moment method:
  - Pearson and Johnson system
  - Generalized Lambda distribution
  - Maximum entropy method (MEM)
  - and many variants . . .

The surrogate form should be chosen carefully depending on the nature of problems dealing with.
Probability of Failure using Surrogate Limit State

Probability of failure using the "truth" limit state function:

\[ P_f = P[g_X \leq 0] = \int_{g_X \leq 0} f_X(x) \, dx \approx \frac{1}{N} \sum_{i=1}^{N} I[g_X(x^{(i)}) \leq 0], \]

Probability of failure using "surrogate" limit state function:

\[ \widehat{P}_f = P[\widehat{g}_Q \leq 0] = \int_{\widehat{g}_Q \leq 0} f_Q(q) \, dq \approx \frac{1}{N} \sum_{i=1}^{N} I[\widehat{g}_Q(q^{(i)}) \leq 0], \]

\( T : X \leftrightarrow Q \), mapping \( T \) is necessary

Propose \( P_f \approx \widehat{P}_f \Rightarrow \) need appropriate validation of \( \widehat{g}_Q \), depending on problems
Example 1: Noisy Limit State Function

Given by Liu and Der Kiureghian (1991),

\[ g(x) = x_1 + 2x_2 + 2x_3 + x_4 - 5x_5 - 5x_6 + 0.001 \sum_{i=1}^{6} \sin(100x_i), \]

\[ X_i = 1, 2, 3, 4 \sim \text{LN}(\lambda_1, \zeta_1), \quad \mu_{X_i} = 120, \quad \sigma_{X_i} = 12, \]

\[ X_5 \sim \text{LN}(\lambda_2, \zeta_2), \quad \mu_{X_5} = 50, \quad \sigma_{X_5} = 15, \]

\[ X_6 \sim \text{LN}(\lambda_3, \zeta_3), \quad \mu_{X_6} = 40, \quad \sigma_{X_6} = 12, \]

\( X_i \) are independent each other.

Solution (given in Ref):

\[ P_f = 1.23 \times 10^{-2}, \quad \beta = 2.3482. \]
Example 1: Noisy Limit State Function – PCE coefficients

For illustration purpose, 1 set of PCE coefficients (HPCE and APCE) is presented.

Truth:

\[ g_x(x) = x_1 + 2x_2 + 2x_3 + x_4 - 5x_5 - 5x_6 + 0.001 \sum_{i=1}^{6} \sin(100x_i) \]

HPCE:

\[ \hat{g}_Q(q) = 269.9987 \cdot 1 + 11.9692q_1 + 23.9414q_2 + 23.9378q_3 + 11.9738q_4 - 73.3819q_5 - 58.7016q_6 + 0.5938(q_1^2 - 1) + \cdots \]

APCE:

\[ \hat{g}_X(x) = 269.9995 \cdot 1 + 1 \cdot (x_1 - 120) + 2 \cdot (x_2 - 120) + 1.9999 \cdot (x_3 - 120) + 0.9999 \cdot (x_4 - 120) - 5 \cdot (x_5 - 50) - 5 \cdot (x_6 - 40) \]
Example 1 Results: APCE vs HPCE – Lower Tail

10 sets of the truth and surrogates; LR; $3 \times (6+p)$ samples.

Hermite polynomial (HPCE) is alternatively selected for lognormal variables.
Arbitrary polynomial (APCE) is constructed using the raw moment sequence of a lognormal variable.

<table>
<thead>
<tr>
<th>$P(g &lt; b)$</th>
<th>Truth</th>
<th>APCE ($p = 1$)</th>
<th>HPCE ($p = 1$)</th>
<th>HPCE ($p = 2$)</th>
<th>HPCE ($p = 3$)</th>
<th>HPCE ($p = 4$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$P(g &lt; b) = 10^{-4}$</td>
<td>$[-293.892, -238.338]$</td>
<td>$[-293.889, -238.337]$</td>
<td>$[-133.442, -96.513]$</td>
<td>$[-259.843, -211.955]$</td>
<td>$[-284.256, -233.526]$</td>
<td>$[-293.353, -238.076]$</td>
</tr>
</tbody>
</table>

Statistics are obtained using 100 sets; RMSE takes the mean value of RMSE of 100 sets.
* the number of samples that are used to make surrogates; the same 100,000 samples are used for the empirical PoEs.
Example 1 Results: APCE vs LPCE (Second Choice for Askey)

Another Legendre polynomial (LPCE) is selected. The convergence looks bad; $p = 4$ is not enough; LPCE is not appropriate for lognormal variables than HPCE.

<table>
<thead>
<tr>
<th></th>
<th>Truth</th>
<th>APCE ($p = 1$)</th>
<th>LPCE ($p = 1$)</th>
<th>LPCE ($p = 2$)</th>
<th>LPCE ($p = 3$)</th>
<th>LPCE ($p = 4$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>RMSE</td>
<td></td>
<td>$2.10 \times 10^{-3}$</td>
<td>$4.25 \times 10^1$</td>
<td>$3.02 \times 10^1$</td>
<td>$2.14 \times 10^1$</td>
<td>$2.37 \times 10^1$</td>
</tr>
<tr>
<td>COV of $P_f$</td>
<td>$2.68 \times 10^{-2}$</td>
<td>$1.39 \times 10^0$</td>
<td>$9.22 \times 10^{-2}$</td>
<td>$3.24 \times 10^{-2}$</td>
<td>$3.15 \times 10^{-2}$</td>
<td></td>
</tr>
<tr>
<td>$\sigma P_f$</td>
<td>$3.28 \times 10^{-4}$</td>
<td>$8.18 \times 10^{-6}$</td>
<td>$9.83 \times 10^{-5}$</td>
<td>$3.26 \times 10^{-4}$</td>
<td>$3.34 \times 10^{-4}$</td>
<td></td>
</tr>
<tr>
<td>$\mu P_f$</td>
<td>$1.23 \times 10^{-2}$</td>
<td>$5.90 \times 10^{-6}$</td>
<td>$1.10 \times 10^{-3}$</td>
<td>$1.01 \times 10^{-2}$</td>
<td>$1.06 \times 10^{-2}$</td>
<td></td>
</tr>
<tr>
<td># of samples</td>
<td>100,000</td>
<td>21*</td>
<td>21*</td>
<td>84*</td>
<td>252*</td>
<td>630*</td>
</tr>
</tbody>
</table>

$P(g < b) = 10^{-4}$ [-293.892 , -238.338] [-293.889 , -238.337] [-7.869 , -0.0346] [-151.247 , -135.693] [-126.485 , -107.228] [-146.9516 , -121.1396]
Example 2: Quadratic Limit State Function

Given by Liu and Der Kiureghian (1991),

\[g_X(x) = 1.1 - 0.00115x_1x_2 + 0.00157x_2^2 + 0.00117x_1^2 + 0.0135x_2x_3 - 0.0705x_2 - 0.00534x_1 - 0.0149x_1x_3 - 0.0611x_2x_4 + 0.0717x_1x_4 - 0.226x_3 + 0.0333x_3^2 - 0.558x_3x_4 + 0.998x_4 - 1.339x_4^2,\]

\[X_1 \sim \text{Type II Extreme}, \quad \mu_{X_1} = 10, \quad \sigma_{X_1} = 5,\]
\[X_2 \sim \text{N}(25, 5^2),\]
\[X_3 \sim \text{N}(0.8, 0.2^2),\]
\[X_4 \sim \text{LN}(\lambda_4, \zeta_4), \quad \mu_{X_4} = 0.0625, \quad \sigma_{X_4} = 0.0625,\]
\[X_i \text{ are independent each other.}\]

Solution (given in Ref): \(\beta = 1.36, \quad P_f = 8.69 \times 10^{-2}\) (by FORM, not MCS).
Example 2: Quadratic Limit State Function – PCE coefficients

For illustration purpose, 1 set of PCE coefficients (HPCE and APCE) is presented.

Truth:

\[ g_X(x) = 1.1 - 0.00115x_1x_2 + 0.00157x_2^2 + 0.00117x_1^2 \\
+ 0.0135x_2x_3 - 0.0705x_2 - 0.00534x_1 - 0.0149x_1x_3 \\
- 0.0611x_2x_4 + 0.0717x_1x_4 - 0.226x_3 + 0.0333x_3^2 \\
- 0.558x_3x_4 + 0.998x_4 - 1.339x_4^2 \]

HPCE:

\[ \hat{g}_Q(q) = 269.9987 \cdot 1 + 11.9692q_1 + 23.9414q_2 + 23.9378q_3 \\
+ 11.9738q_4 - 73.3819q_5 - 58.7016q_6 + 0.5938(q_1^2 - 1) + \cdots \]

APCE:

\[ \hat{g}_X(x) = 0.1238 \cdot 1 - 0.0066(x_1 - 10.1327) + 0.0008(x_2 - 24.8689) - 0.0357(x_3 - 0.6991) \\
- 0.5185(x_4 - 0.0694) + 0.0012(x_1^2 - 27.9777x_1 + 161.8982) + \cdots \]
Example 2 Results: APCE vs HPCE – Lower Tail

10 sets of the truth and surrogates; LR; $3 \times (4 + p)$ samples.

Hermite polynomial (HPCE) is alternatively selected for Type II extreme and lognormal variables. Arbitrary polynomial (APCE) is constructed using the raw moment sequence of the variables.

<table>
<thead>
<tr>
<th></th>
<th>Truth</th>
<th>APCE ($p = 1$)</th>
<th>APCE ($p = 2$)</th>
<th>HPCE ($p = 1$)</th>
<th>...</th>
<th>HPCE ($p = 10$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>RMSE</td>
<td>1.24 × 10^{-2}</td>
<td>3.78 × 10^{-1}</td>
<td>6.90 × 10^{-16}</td>
<td>3.57 × 10^{0}</td>
<td>...</td>
<td>1.40 × 10^{-1}</td>
</tr>
<tr>
<td>COV of $P_f$</td>
<td>6.89 × 10^{-4}</td>
<td>1.48 × 10^{-2}</td>
<td>1.24 × 10^{-2}</td>
<td>1.68 × 10^{-2}</td>
<td>...</td>
<td>1.25 × 10^{-2}</td>
</tr>
<tr>
<td>$\sigma P_f$</td>
<td>5.57 × 10^{-2}</td>
<td>8.09 × 10^{-4}</td>
<td>6.89 × 10^{-2}</td>
<td>7.22 × 10^{-4}</td>
<td>...</td>
<td>6.96 × 10^{-4}</td>
</tr>
<tr>
<td>$\mu P_f$</td>
<td>100,000</td>
<td>5.45 × 10^{-2}</td>
<td>5.57 × 10^{-2}</td>
<td>4.30 × 10^{-2}</td>
<td>...</td>
<td>5.57 × 10^{-2}</td>
</tr>
<tr>
<td># of samples</td>
<td></td>
<td>15*</td>
<td>45*</td>
<td>15*</td>
<td></td>
<td>3003*</td>
</tr>
</tbody>
</table>

$P(g < b) = 10^{-3}$  
$[-0.517, -0.465]$  
$[-0.191, -0.163]$  
$[-0.517, -0.465]$  
$[-0.0465, -0.0414]$  
$[-0.525, -0.473]$  

$P(g < b) = 10^{-4}$  
$[-1.846, -1.276]$  
$[-0.722, -0.431]$  
$[-1.846, -1.276]$  
$[-0.0852, -0.0726]$  
$[-1.703, -1.221]$  

$P(g < b) = 10^{-5}$  
$[-12.563, -2.334]$  
$[-2.553, -1.021]$  
$[-12.563, -2.334]$  
$[-0.165, -0.0998]$  
$[-9.674, -2.656]$  

Statistics are obtained using 100 sets; RMSE takes the mean value of RMSE of 100 sets.

* the number of samples that are used to make surrogates; the same 100,000 samples are used for the empirical PoEs.
Example 3: Limit State Function with Dependent Variables

Given by Der Kiureghian and Liu (1986),

\[ g(x) = x_1 - x_2, \]

\[ X_1 \sim \text{Unif}[0, 100], \quad \mu_{X_1} = 50, \quad \sigma_{X_1} = 28.87, \]

\[ X_2 \sim \text{Exp}(0.08), \quad \mu_{X_1} = 12.5, \quad \sigma_{X_1} = 12.5, \quad \rho_{X_1X_2} = 0.5 \]

Solution : \( P_f = 4.86 \times 10^{-2} \) (by MCS).
Example 3 Results: APCE vs HPCE – Lower Tail

10 sets of the truth and surrogates; LR; $3 \times \left( \frac{2+p}{p} \right)$ samples.

Hermite polynomial (HPCE) is alternatively selected.
Arbitrary polynomial (APCE) is constructed using the raw moment sequence of the variables.

<table>
<thead>
<tr>
<th></th>
<th>Truth</th>
<th>APCE ($p = 1$)</th>
<th>HPCE ($p = 1$)</th>
<th>...</th>
<th>HPCE ($p = 23$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>RMSE</td>
<td>1.54 $\times 10^{-2}$</td>
<td>3.62 $\times 10^{-15}$</td>
<td>2.29 $\times 10^{1}$</td>
<td>...</td>
<td>1.19 $\times 10^{1}$</td>
</tr>
<tr>
<td>COV of $P_f$</td>
<td>7.47 $\times 10^{-4}$</td>
<td>1.54 $\times 10^{-2}$</td>
<td>2.39 $\times 10^{-2}$</td>
<td>...</td>
<td>1.54 $\times 10^{-2}$</td>
</tr>
<tr>
<td>$\sigma P_f$</td>
<td>4.86 $\times 10^{-2}$</td>
<td>7.47 $\times 10^{-4}$</td>
<td>4.18 $\times 10^{-4}$</td>
<td>...</td>
<td>7.48 $\times 10^{-4}$</td>
</tr>
<tr>
<td>$\mu P_f$</td>
<td>100,000</td>
<td>4.86 $\times 10^{-2}$</td>
<td>1.75 $\times 10^{-2}$</td>
<td>...</td>
<td>4.85 $\times 10^{-2}$</td>
</tr>
<tr>
<td># of samples</td>
<td>9*</td>
<td>9*</td>
<td>900*</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

$P(g < b) = 10^{-3}$
$[-24.702, -22.884]$  
$[24.702, 22.884]$  
$[-58.050, -55.812]$  
$[-29.665, -22.884]$

$P(g < b) = 10^{-4}$
$[-46.322, -39.507]$  
$[-46.322, -39.507]$  
$[-79.211, -72.824]$  
$[-79.211, -72.824]$

$P(g < b) = 10^{-5}$
$[-84.501, -52.362]$  
$[-84.501, -52.362]$  
$[-135.081, -86.757]$  
$[-15491.774, -237.843]$

Statistics are obtained using 100 sets; RMSE takes the mean value of RMSE of 100 sets.

* the number of samples that are used to make surrogates; the same 100,000 samples are used for the empirical PoEs.
Example 3 Results: APCE – Coefficients

For illustration purpose, the coefficients for 10 APCE sets (different LR samples) are listed:

\[ \hat{g}_X(x) = 30.123 \times \begin{cases} 1 & \text{if } \Psi_0 \\
 c_0 
\end{cases} + 1 \times \begin{cases} (x_1 - 36.226) - 1 & \text{if } \Psi_1 = x_1 - \mathbb{E}[X_1] \\
 c_1 
\end{cases} \times \begin{cases} (x_2 - 6.103) & \text{if } \Psi_2 = x_2 - \mathbb{E}[X_2] \\
 c_2 
\end{cases} \]

\[ \hat{g}_X(x) = 29.048 + (x_1 - 36.312) - (x_2 - 7.264) \]
\[ \hat{g}_X(x) = 42.315 + (x_1 - 50.968) - (x_2 - 8.654) \]
\[ \hat{g}_X(x) = 43.153 + (x_1 - 59.127) - (x_2 - 15.974) \]
\[ \hat{g}_X(x) = 45.758 + (x_1 - 55.601) - (x_2 - 9.843) \]
\[ \hat{g}_X(x) = 26.434 + (x_1 - 43.758) - (x_2 - 17.325) \]
\[ \hat{g}_X(x) = 39.118 + (x_1 - 49.793) - (x_2 - 10.675) \]
\[ \hat{g}_X(x) = 34.669 + (x_1 - 49.214) - (x_2 - 14.545) \]
\[ \hat{g}_X(x) = 20.141 + (x_1 - 29.709) - (x_2 - 9.568) \]
\[ \hat{g}_X(x) = 34.766 + (x_1 - 47.759) - (x_2 - 12.993) \]

\[ g_X(x) = d_0 + d_1 x_1 + d_2 x_2 \]

\[ \hat{g}_X(x) = \hat{d}_0 + \hat{d}_1 x_1 + \hat{d}_2 x_2 \]

Good approximation by APCE.
Reason Why HPCE Exhibits Bad Convergence

\[ g_X(x) = x_1 - x_2, \]
\[ X_1 \sim \text{Unif}[0, 100] \text{ and } X_2 \sim \text{Exp}(0.08), \]
\[ \rho_{X_1X_2} = 0.5. \]

Because Askey PCE scheme is based on the independence assumption of random variables, we need to identify \( X' = [X'_1, X'_2]^T \) — independent physical space, though it is challenging.
⇒ but it is relatively easy to identify in normal space (by Nataf transformation),
⇒ but \( g_Q(q) \) may be more complicated than \( g_X(x) \); lead to slower convergence of surrogate model.

By Nataf transformation, \( \rho_{Q'_1, Q'_2} \) can be found; \( Q' = [Q'_1, Q'_2]^T \) — dependent normal space.

And we can decorrelate a bivariate normal variable by \( Q = L^{-1}Q' \); \( L : \) lower Cholesky triangle of a covariance matrix \( \Sigma_X = LL^T. \)

Thus, HPCE has to be done in \( Q \) space using nonlinear mappings as follows:

\[ X_1 = F_{X_1}^{-1}(\Phi(Q'_1)) = F_{X_1}^{-1}(\Phi(Q_1)), \]
\[ X_2 = F_{X_2}^{-1}(\Phi(Q'_2)) = F_{X_2}^{-1}\left(\Phi\left(\rho_{Q'_1, Q'_2}Q_1 + \sqrt{1 - \rho_{Q'_1, Q'_2}^2}Q_2\right)\right). \]
Example 4: Limit State Function: I-beam

Given by Huang and Du (2006),

\[ g(x) = \frac{6x_1x_3(x_2 - x_3)x_5}{x_2(x_6x_5^3 - (x_6 - x_7)(x_5 - 2x_8)^3)} - x_4 \]

\[ X_1 \sim N(6070, 200) \]
\[ X_2 \sim N(120, 6) \]
\[ X_3 \sim N(72, 6) \]
\[ X_4 \sim N(170000, 4760) \]
\[ X_5 \sim N(2.3, 1/24) \]
\[ X_6 \sim N(2.3, 1/24) \]
\[ X_7 \sim N(0.16, 1/48) \]
\[ X_8 \sim N(0.26, 1/48) \]

Solution (given in Ref): \( P_f = 8.711 \times 10^{-1} \) (by \( 10^6 \) MCS samples).
\[ \Rightarrow \] large failure probability; different thresholds \( (g(x) < b) \) will be investigated here.
Example 4: Limit State Function: I-beam

Number of function evaluation: 45 \((= 1 \times \binom{8+2}{2})\)

HPCE and APCE give the same result (because \(X_i\) follow normal distributions).
Example 5: Limit State Function: Sinusoidal Function

Given by Sundar and Shields (2006),

\[ g_X(x) = 4 - x_1/4 + \sin(5x_1) - x_2, \]

\[ X_1 \sim N(0, 1), \]

\[ X_2 \sim N(0, 1). \]

Solution (given in Ref): \( P_f = 4.5460 \times 10^{-4} \) (10^8 MCS samples).

- A sinusoidal function may be approximated by a higher order polynomial function (Refer to the fig in the right).
- However, \( \sin(5x_1) \) requires a quite higher order polynomial function (due to high frequency).
- Difficult to get an accurate surrogate by PCE (or aPCE) with a moderate number of samples.
- Even with a large number of samples, it is difficult to avoid overfitting issue.

A PCE surrogate might not be a good choice for oscillatory limit state function, but ...
Example 5: Limit State Function: Sinusoidal Function

Alternative—transforming to a favorable functional form to PCE:

- proceed with $g_1$ and $g_2$, instead of $g$
- note $g_2$ and $x_1$ are functionally dependent
- in general cases, the decomposition (e.g., $g = g_1 + g_2$) is not easily obtained.
- but, assume that the decomposition is available.

Let’s define:

$$P_f = P(g < 0) = \int P(g < 0 \mid g_2) f(g_2) dg_2$$

- MCS: Inner loop

$$P(g < 0 \mid g_2 = a) = P(g_1 < -a) \approx \frac{1}{N_{in}} \sum_{i=1}^{N_{in}} I(g_1^{(i)} < -a)$$

$$= \frac{1}{N_{in}} \sum_{i=1}^{N_{in}} I(4 - x_1^{(i)}/4 - x_2^{(i)} < -a)$$

note $\sin(5x_1) = a$

- MCS: Outer loop

$$P(g < 0) \approx \frac{1}{N_{out}} \sum_{j=1}^{N_{out}} P(g_1 < -a^{(j)})$$

$$= \frac{1}{N_{out}} \sum_{j=1}^{N_{out}} \left[ \frac{1}{N_{in}} \sum_{i=1}^{N_{in}} I(4 - x_1^{(j)}/4 - x_2^{(i)} < -a^{(j)}) \right]$$

$$= \frac{1}{N_{out}} \sum_{j=1}^{N_{out}} \left[ \frac{1}{N_{in}} \sum_{i=1}^{N_{in}} I\left(4 - x_1^{(j)}/4 - x_2^{(i)} < -\sin(5x_1^{(j)})\right) \right]$$
Example 5: Limit State Function: Sinusoidal Function

MCS: Truth vs. Alternative (nested MCS)

Tail Distribution of $g$

- Conventional MCS ($10^5$ samples) vs. Alternative way ($350 \times 350$ samples)

MCS: Truth vs. Taylor series

Tail Distribution of $g$

- A higher order Taylor series ($p = 60$) is accurate down to $P = 10^{-5}$ level.
- A higher order PCE may work or not.
Example 5: Limit State Function: Sinusoidal Function

- Recall alternative formulation:

\[
P(g < 0) \approx \frac{1}{N_{out}} \sum_{j=1}^{N_{out}} \left[ \frac{1}{N_{in}} \sum_{i=1}^{N_{in}} I \left( \frac{4 - x_1(j)}{4 - x_2(i)} < -\sin(5x_1^4) \right) \right]
\]

- Samples for \( g_1 \) and \( g_2 \):

\[3N = 3 \times \left( \frac{2+1}{1} \right) = 9\]

- PCE with \( g_1 \) and Fourier series with \( g_2 \):

\[g_1 \approx \hat{g}_1 \text{ and } g_2 \approx \hat{g}_2\]
Example 5: Limit State Function: Sinusoidal Function

\[ \hat{g}_1 = 4.0000 \times 1 - 0.2500 \times x_1 - 1.0000 \times x_2 \]

\[ g_2 = \sin(5x_1) \]

\[ \hat{g}_1 = 4.0000 \times 1 - 0.2500 \times x_1 - 1.0000 \times x_2 \]

\[ \hat{g}_2 = a_0 + a_1 \cos(x_1 w) + b_1 \sin(x_1 w) \]
\[ + a_2 \cos(2x_1 w) + b_2 \sin(2x_1 w) \]
\[ = 6.098 \times 10^{-12} + 2.197 \times 10^{-12} \cos(5x_1) + 1 \times \sin(5x_1) \]
\[ + 3.951 \times 10^{-13} \cos(10x_1) + 3.017 \times 10^{-12} \sin(10x_1) \]
Example 6: Limit State Function: 20 non-Gaussian variables

Given by Engelund and Rackwitz (1993),

\[
(1) : \quad g_\mathbf{x}(\mathbf{x}) = \sum_{i=1}^{20} x_i - 8.951,
\]

\[
(2) : \quad g_\mathbf{x}(\mathbf{x}) = -\sum_{i=1}^{20} x_i + 36.720,
\]

\[X_{i=1}^{20} \sim \text{Exp}(1).\]

Solutions for the both (given in Ref): \(P_f = 1 \times 10^{-3}\).
Example 6: Limit State Function: 20 non-Gaussian variables

Number of function evaluations: $63 = 3 \times \binom{20+1}{1}$

With 63 evaluations to make a $\hat{g}$, $P(g < b)$ for any threshold value ($b$) can be estimated.
Example 7: Limit State Function: Strong Non-linearity

Given by Der Kiureghian and De Stefano (1991),

\[ g(x) = x_7 - 3x_4 \sqrt{\frac{\pi x_8}{4x_6 A^3} \left[ \frac{Bx_6}{x_5 x_6 (4B^2 + C^2) + DB^2} \frac{(x_5 E^3 + x_6 A^3)E}{4BF^4} \right]}, \]

\[ A = \sqrt{x_4/x_2}, \quad B = (x_5 + x_6)/2, \quad C = (E - A)/F, \]

\[ D = x_2/x_1, \quad E = \sqrt{x_3/x_1}, \quad F = (E + A)/2, \]

\( X_{i=1 \ to \ 20} \sim \text{Exp}(1). \)

Solutions for the both (given in Ref): \( P_f = 4.79 \times 10^{-3}. \)

This limit state function exhibits strong non-linearity.

⇒ may not work with PCE or even APCE.
Conclusions

- Several engineering applications work well with PCE surrogate models.

- A surrogate limit state function by APCE can approximate the truth limit state function well.

- APCE does not need any iso-probabilistic transformations, which can lead to slow convergence to the truth limit state function.

- HPCE with a higher order approximate the truth limit state function well, regarding accurate estimation of CoV of $P_f$; however, other criteria such as lower tail behavior and RMSE is not satisfactory than APCE.

- It should be noted that there are cases where specific surrogate modeling method works best.
Thank you very much for your attention!