Treatment of uncertainties in multi-physics model for wind turbine asset management

Elias FEKHARI

Électricité de France R&D

Chatou, France

&

Côte d'Azur University

Nice, France

This dissertation is submitted for the degree of

Doctor of Philosophy

Pr. Mireille BOSSY,	INRIA, Sophia-Antipolis	Examiner
Dr. Vincent CHABRIDON	EDF R&D, Chatou	Co-advisor
Dr. Sébastien DA VEIGA	ENSAI, Rennes	Examiner
Dr. Bertrand IOOSS	EDF R&D, Chatou	Thesis director
Dr. Joseph MURÉ	EDF R&D, Chatou	Co-advisor
Pr. Franck SCHOEFS	Nantes Université, Nantes	Reviewer
Pr. Daniel STRAUB	TUM, Munich	Reviewer
Pr. Bruno SUDRET	ETH, Zürich	Examiner

Table of contents

List of figures

List of tables

Introduction

Part I

Introduction to treatment of uncertainties and wind energy

Chapter 1

Treatment of uncertainties in computer experiments

1.1 Problem specification (step A)

1.1.1 Black-box computer model

1.1.2 Output quantity of interest

1.2 Input uncertainty quantification (step B)

1.2.1 Joint probability distribution [copulogram package]

1.2.2 Parametric multivariate estimation

1.2.3 Non-parametric multivariate estimation

1.2.4 Goodness-of-fit

1.3 Uncertainty propagation for central tendency estimation (step C)

1.3.1 Numerical integration

"Good" properties

[Curse of dim / Sequential / Deterministic]

Gauss-Kronrod

Monte Carlo

Quasi-Monte Carlo and Koksma-Hlawka inequality

1.3.2 Numerical design of experiments

Space-filling metrics

[MinMax / PhiP / MaxMin / Discrepancies]

"Good" properties

[Curse of dim / Projections in sub-spaces / Sequential / Deterministic]

Monte Carlo, quasi-Monte Carlo, randomized quasi-Monte Carlo designs

LHS, optimized LHS designs

1.3.3 Central tendency estimation

Iso-probabilistic transformation

Central tendency estimation is a probabilistic integration

1.4 Uncertainty propagation for rare event estimation (step C)

1.4.1 Problem formalization

Limit-state function, failure event and domain

Risk measures [Failure probability, quantile, super-quantile]

1.4.2 Rare event estimation methods

FORM/SORM

Monte Carlo

Importance sampling

Adaptive sampling (SS/NAIS/IS-CE/Moving particles)

1.5 Sensitivity analysis (step C')

1.5.1 Global sensitivity analysis

1.5.2 Reliability-oriented sensitivity analysis

1.6 Metamodeling

1.6.1 Global metamodel

1.6.2 Reliability-oriented metamodel

Chapter 2

Introduction to wind turbine modeling and design

2.1 Wind turbine modeling

2.1.1 Synthetic wind generation [TurbSim, Kaimal spectrum]

2.1.2 Synthetic wave generation

2.1.3 Aerodynamic interactions

2.1.4 Servo-Hydro-Aero-Elastic wind turbine simulation [DIEGO]

2.1.5 Soil modeling

2.1.6 Wake modeling [FarmShadow]

2.2 Recommended design practices

2.2.1 Design load cases

2.2.2 Dynamic response design

2.2.3 Fatigue response design

2.3 Uncertain inputs

2.3.1 Environmental inputs

2.3.2 System inputs

2.3.3 Probabilistic fatigue assessment

Part II

Contributions to uncertainty quantification and propagation

Chapter 3

Kernel-based uncertainty quantification

3.1 Nonparametric fit of the environmental inputs (OMAE cpaper 2023)

3.2 Quantifying and clustering the wake-induced perturbations within a wind farm (WAKE cpaper 2023)

Chapter 4

Kernel-based central tendency estimation

4.1 Kernel discrepancy

4.2 Quantization with kernel herding [SIAM UQ talk 2022, RENEW cpaper 2022]

4.3 Gaussian process regression

4.4 Bayesian quadrature [otkerneldesign package]

4.5 Numerical experiments [ctbenchmark package]

4.6 Application to wind turbine mean fatigue estimation (DCE paper)

Part III

Contributions to rare event estimation

Chapter 5

Nonparametric rare event estimation

5.1 Bernstein adaptive nonparametric conditional sampling

5.1.1 Introduction

Reliability analysis of a system is often associated with rare event probability estimation. Considering that the system's performance is modeled by a deterministic scalar function $g : \mathscr{D}_{\mathbf{x}} \subseteq \mathbb{R}^d \to \mathbb{R}$, called *limit-state function* and a critical threshold on the system's output $y_{\text{th}} \in \mathbb{R}$, one can define the *failure domain* as $\mathscr{F}_{\mathbf{x}} := \{\mathbf{x} \in \mathscr{D}_{\mathbf{x}} | g(\mathbf{x}) \leq y_{\text{th}}\}$. Uncertain inputs are represented by a continuous random vector $\mathbf{X} \in \mathscr{D}_{\mathbf{x}}$ assumed to be distributed according to its joint probability density function (PDF) $f_{\mathbf{X}}$. In this context, uncertainty propagation consists in composing the random vector $\mathbf{X}$ by the function g to get an output variable of interest $Y = g(\mathbf{X}) \in \mathbb{R}$. A usual risk measure in reliability analysis is the *failure probability*, denoted by p_{f}, and defined as the probability that the system exceeds the threshold y_{th}: Rare event problems are usually solved in the so-called *standard normal space* after applying an "iso-probabilistic transformation" which can be either the Rosenblatt or the generalized Nataf one [10]. Additionally, the limit-state function g can be viewed as an input-output "black-box" model which can be costly to evaluate (e.g., a complex numerical model), making the failure probability estimation nontrivial. When the limit-state function is a costly computer model, one can build a surrogate model and use specific active learning methods (see, e.g., Moustapha et al. [15]). However, using surrogate models is not always possible for practical engineering applications as they might introduce another level of approximation, which can be prohibitive from safety auditing. Moreover, their validation as well as their behavior with respect to large input dimension case make also their use quite complex (see, e.g., [12].

Going back to the rare event estimation literature, one can consider two major types of techniques for failure probability calculation [14]: (i) Geometric approaches, such as the *first-/second-order reliability method* (FORM/SORM) whose aim is to approximate the limit-state function by a first-/second-order Taylor expansion at the most probable failure point; (ii) Simulation-based techniques such as the *crude Monte Carlo* method. Unfortunately, FORM/SORM methods do not provide a lot of statistical information as they are purely geometric approaches. Meanwhile, estimating a rare event probability by crude Monte Carlo becomes rapidly intractable. To overcome this limit, advanced simulation techniques have been developed: among others, one can mention several "variance reduction methods" such as the non-adaptive and adaptive versions of the *Importance Sampling* [17] (either parametric, using the Cross-Entropy method Kurtz and Song [8], or nonparametric Morio [13]) and splitting techniques [3] such as the *Subset Simulation* (SS) Au and Beck [1]. In these techniques, the idea is to write the rare event p_{f} as a product of larger conditional probabilities, each one of them being easier to estimate. To generate intermediary conditional samples, this method uses Markov chain Monte Carlo (MCMC) sampling, which presents numerous versions [16]. However, MCMC algorithms are known to be highly tunable algorithms which produce non-i.i.d. samples, which consequently, cannot be used for direct statistical estimation (e.g., failure probability or sensitivity indices [5].

The present work proposes a new rare event estimation method, adopting the same sequential structure as SS while using a strictly different sampling mechanism to generate conditional samples. This method intends to fit the intermediary conditional distributions with a nonparametric tool called the *Empirical Bernstein Copula*. Contrarily to SS, the proposed method named "Bernstein adaptive nonparametric conditional sampling" (BANCS), generates i.i.d. samples of the intermediary conditional distributions. For instance, a practical use of such i.i.d. samples can be to estimate dedicated reliability-oriented sensitivity indices (see, e.g., [4, 11]).

In this paper, Section 2 will recall the methodology of subset sampling and probabilistic modeling. Then, Section 3 will introduce the BANCS method for rare event estimation. Section 4 will apply this method to three toy-cases and analyze the results with respect to SS performances. Then, the last section present some conclusions and research perspectives.

5.1.2 Background

Subset sampling

Subset sampling splits the failure event $\mathscr{F}_{\mathbf{x}}$ into an intersection of $k_{\#}$ intermediary events $\mathscr{F}_{\mathbf{x}} = \cap_{k=1}^{k_{\#}} \mathscr{F}_{[k]}$. Each are nested such that $\mathscr{F}_{[1]} \supset \cdots \supset \mathscr{F}_{[k_{\#}]} = \mathscr{F}_{\mathbf{x}}$. The failure probability is then expressed as a product of conditional probabilities:

$$p_{\mathrm{f}} = \mathbb{P}(\mathscr{F}_{\mathbf{x}}) = \mathbb{P}(\cap_{k=1}^{k_\#} \mathscr{F}_{[k]}) = \prod_{k=1}^{k_\#} \mathbb{P}(\mathscr{F}_{[k]} | \mathscr{F}_{[k-1]}). \tag{5.1}$$

From a practical point of view, the analyst tunes the algorithm by setting the intermediary probabilities $\mathbb{P}(\mathscr{F}_{[k]} | \mathscr{F}_{[k-1]}) = p_0, \forall k \in \{1, \dots, k_\#\}$. Then, the corresponding quantiles $q_{[1]}^{p_0} > \cdots > q_{[k_\#]}^{p_0}$ are estimated for each conditional subset samples $\mathbf{X}_{[k],N}$ of size N. Note that the initial quantile is estimated by crude Monte Carlo sampling on the input PDF $f_{\mathbf{X}}$. Following conditional subset samples are generated by MCMC sampling of $f_{\mathbf{X}}(\mathbf{x}|\mathscr{F}_{[k-1]})$, using as seeds initialisation points the $n = Np_0$ samples given by $\mathbf{A}_{[k],n} = \{\mathbf{X}_{[k-1]}^{(j)} \subset \mathbf{X}_{[k-1],N} | g(\mathbf{X}_{[k-1]}^{(j)}) > \widehat{q}_{[k-1]}^{\alpha}\}_{j=1}^{n}$. This process is repeated until an intermediary quantile exceeds the threshold: $\widehat{q}_{[k_\#]}^{p_0} < y_{\mathrm{th}}$. Finally, the failure probability is estimated by:

$$p_{\mathrm{f}} \approx \widehat{p}_{\mathrm{f}}^{\mathrm{SS}} = p_0^{k_\# - 1} \frac{1}{N} \sum_{j=1}^{N} \mathbb{1}_{\{g(\mathbf{x}) \leq y_{\mathrm{th}}\}} (\mathbf{X}_{[k_\#],N}^{(j)}). \tag{5.2}$$

In practice, the subset sample size should be large enough to properly estimate intermediary quantiles, which leads [1] to recommend setting $p_0 = 0.1$. SS efficiency depends on the proper choice and tuning of the MCMC algorithm [16]. Our work uses the SS implementation from OpenTURNS[1] [2] which integrates a component-wise Metropolis-Hastings algorithm. As an alternative to generating samples on a conditional distribution by MCMC, one could try to fit this conditional distribution.

Multivariate modeling using copulas

The Sklar theorem [6] affirms that the multivariate distribution of any random vector $\mathbf{X} \in \mathbb{R}^d$ can be broken down into two objects:

1. A set of univariate marginal distributions to describe the behavior of the individual variables;
2. A function describing the dependence structure between all variables, called a copula.

This theorem states that considering a random vector $\mathbf{X} \in \mathbb{R}^d$, with its distribution F and its marginals $\{F_i\}_{i=1}^{d}$, there exists a copula $C : [0,1]^d \to [0,1]$, such that:

$$F(x_1, \dots, x_d) = C\left(F_1(x_1), \dots, F_p(x_d)\right). \tag{5.3}$$

[1] https://openturns.github.io/www/index.html

It allows us to divide the problem of fitting a joint distribution into two independent problems: fitting the marginals and fitting the copula. Note that when the joint distribution is continuous, this copula is unique. Provided a dataset, this framework allows to combine a parametric (or nonparametric) fit of marginals with a parametric (or nonparametric) fit of the copula. When the distribution's dimension is higher than two, one can perform a parametric fit using vine copulas [7], implying the choice of multiple types of parametric copulas. Otherwise, nonparametric fit by multivariate kernel density estimation (KDE) presents a computational burden as soon as the dimension increases [4]. Since univariate marginals are usually well-fitted with nonparametric tools (e.g., KDE), let us introduce an effective nonparametric method for copula fitting.

5.2 A NEW COPULA-BASED CONDITIONAL SAMPLING METHOD

5.2.1 Empirical Bernstein copula

Copulas are continuous and bounded functions defined on a compact set (the unit hypercube). Bernstein polynomials allow to uniformly approximate as closely as desired any continuous and real-valued function defined on a compact set (Weierstrass approximation theorem). Therefore, they are good candidates to approximate unknown copulas. This concept was introduced as *empirical Bernstein copula* (EBC) by [18] for applications in economics and risk management. Later on, [19] offered further asymptotic studies. Formally, the multivariate Bernstein polynomial for a function $C : [0,1]^d \to \mathbb{R}$ on a grid over the unit hypercube $G := \left\{\frac{0}{m_1}, \dots, \frac{m_1}{m_1}\right\} \times \cdots \times \left\{\frac{0}{m_d}, \dots, \frac{m_d}{m_d}\right\}, \mathbf{m} = (m_1, \dots, m_d) \in \mathbb{N}^d$, writes:

$$B_{\mathbf{m}}(C)(\mathbf{u}) := \sum_{t_1=0}^{m_1} \dots \sum_{t_d=0}^{m_d} C\left(\frac{t_1}{m_1}, \dots, \frac{t_d}{m_d}\right) \prod_{j=1}^{d} P_{m_j,t_j}(u_j), \tag{5.4}$$

with $\mathbf{u} = (u_1, \dots, u_d) \in [0,1]^d$, and the Bernstein polynomial $P_{m,t}(u) := \frac{t!}{m!(t-m)!} u^m (1-u)^{t-m}$. Notice how the grid definition implies the polynomial's order. When C is a copula, then $B_{\mathbf{m}}(C)$ is called "Bernstein copula". Therefore, the empirical Bernstein copula is an application of the Bernstein polynomial in Eq. (5.4) to the so-called "empirical copula".

In practice, considering a sample $\mathbf{X}_n = \left\{\mathbf{x}^{(1)}, \dots, \mathbf{x}^{(n)}\right\} \in \mathbb{R}^{np}$ and the associated ranked sample $\mathbf{R}_n = \left\{\mathbf{r}^{(1)}, \dots, \mathbf{r}^{(n)}\right\}$, the corresponding empirical copula writes:

$$C_n(\mathbf{u}) := \frac{1}{n} \sum_{i=0}^{n} \prod_{j=1}^{p} \mathbb{1}\left\{ \frac{r_j^{(i)}}{n} \leq u_j \right\}, \tag{5.5}$$

with $\mathbf{u} = (u_1, \dots, u_d) \in [0,1]^d$. In the following, the polynomial order is set as equal in each dimension: $\{m_i = m\}_{j=1}^{d}$. Theoretically, the tuning parameter can be optimized to minimize an "Mean Integrated Squared Error" (MISE), leading to a bias-variance tradeoff. Formally, the MISE of the empirical Bernstein copula $B_{\mathbf{m}}(C_n)$ is defined as follows:

$$\mathbb{E}\left[\|B_{\mathbf{m}}(C_n) - C\|_2^2\right] = \mathbb{E}\left[\int_{\mathbb{R}^d} (B_{\mathbf{m}}(C_n)(\mathbf{u}) - C(\mathbf{u})\mathrm{d}\mathbf{u})^2\right]. \tag{5.6}$$

Then, [18] prove in their Theorem 3 that:

- $B_{\mathbf{m}}(C_n)(\mathbf{u}) \to C(\mathbf{u})$ for any $u_j \in]0,1[$ if $\frac{m^{d/2}}{n} \to 0$, when $m, n \to \infty$.
- The optimal order of the polynomial in terms of MISE is: $m \lesssim m_{\text{IMSE}} = n^{2/(d+4)}, \forall u_j \in]0,1[$. The sign $\lesssim$ means "less than or approximately".

Let us remark that in the special case $m = n$, also called the "Beta copula" in [19], the bias is very small while the variance gets large. To illustrate the previous theorem, [9] represents the evolution of the m_{IMSE} for different dimensions and sample sizes (see Fig. 5.1). In high dimension, the values of m_{IMSE} tend towards one, which is equivalent to the independent copula. Therefore, high-dimensional problems should be divided into a product of smaller problems on which the EBC is tractable. Provided a large enough learning set $\mathbf{X}_n$, KDE fitting of marginals combined with EBC fitting of the copula delivers good results even on complex dependence structures. Moreover, EBC provides an explicit expression, making a Monte Carlo generation of i.i.d. samples simple. In the following, this nonparametric tool is used to fit the intermediary conditional distributions present in subset sampling.

5.2.2 Bernstein adaptive nonparametric conditional sampling (BANCS) method

This new method reuses the main idea from SS while employing a different approach to generate conditional samples. Instead of using MCMC sampling, the conditional distribution is firstly fitted by a nonparametric procedure, before sampling on this nonparametric model.

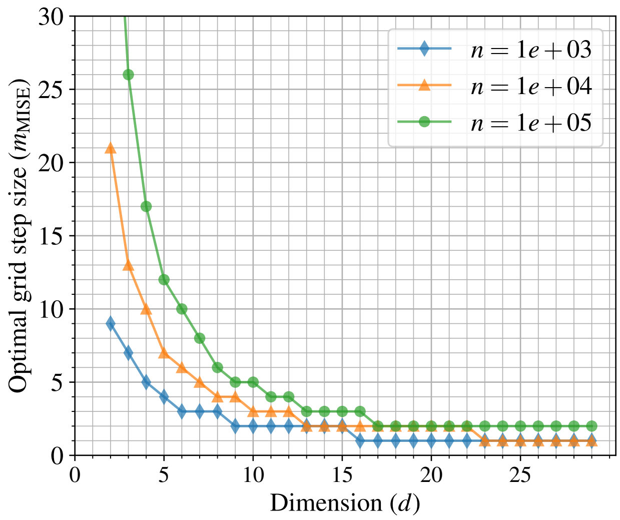

Fig. 5.1 Evolution of m_{IMSE} for different dimensions and sample sizes.

As described in Algorithm 1, conditional sampling is done on a distribution composed by merging marginals $\{\widehat{F}_i\}_{i=1}^{d}$ fitted by KDE, with a copula $B_{\mathbf{m}}(C_n)$ fitted by EBC. Fig. 5.2 illustrate the nonparametric fit and conditional sampling in BANCS method on a two-dimensional reliability problem (later introduced as "toy-case #1"). At iteration k, after estimating the intermediary quantile $\widehat{q}_{[k]}^{p_0}$, a nonparametric model is fitted on $\mathbf{A}_{[k+1],n}$ and used to generate the next N-sized subset sample $\mathbf{X}_{[k+1],N}$. Note that the BANCS method does not require iso-probabilistic transform.

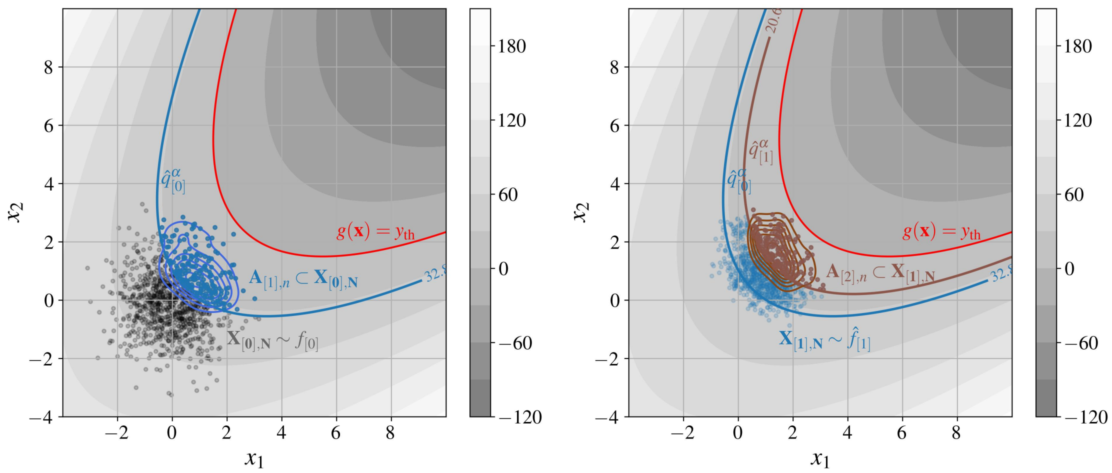

Fig. 5.2 BANCS on toy-case #1: illustration of conditional sampling and nonparametric fit at the first and second iterations.

Algorithm 1 Bernstein adaptive nonparametric conditional sampling (BANCS).

▷ *Inputs:* ◁

$f_{\mathbf{X}}$, joint PDF of the inputs

$g(\cdot)$, limit-state function

$y_{\text{th}} \in \mathbb{R}$, threshold defining the failure event

N, number of samples per iteration

$m \in \mathbb{N}$, parameter of the EBC fitting

$p_0 \in]0, 1[$, empirical quantile order (rarity parameter)

▷ *Algorithm:* ◁

Set $k = 0$ and $f_{[0]} = f_{\mathbf{X}}$

Sample $\mathbf{X}_{[0],N} = \{\mathbf{X}_{[0]}^{(j)}\}_{j=1}^{N} \overset{\text{i.i.d}}{\sim} f_{[0]}$

Evaluate $G_{[0],N} = \{g(\mathbf{X}_{[0]}^{(j)})\}_{j=1}^{N}$

Estimate the empirical p_0-quantile $\widehat{q}_{[0]}^{p_0}$ of the set $G_{[0],N}$

while $\widehat{q}_{[k]}^{p_0} > y_{\text{th}}$ **do**

  Subsample $\mathbf{A}_{[k+1],n} = \{\mathbf{X}_{[k]}^{(j)} \subset \mathbf{X}_{[k],N} | g(\mathbf{X}_{[k]}^{(j)}) > \widehat{q}_{[k]}^{p_0}\}_{j=1}^{n}$

  Fit marginals of the subset $\mathbf{A}_{[k+1],n}$ by KDE $\{\widehat{F}_i\}_{i=1}^{d}$

  Fit the copula of the subset $\mathbf{A}_{[k+1],n}$ by EBC $B_{\mathbf{m}}(C_n)$

  Build a CDF $\widehat{F}_{[k+1]}(\mathbf{x}) = B_{\mathbf{m}}(C_n)(\widehat{F}_1(x_1), \ldots, \widehat{F}_d(x_d))$

  Sample $\mathbf{X}_{[k+1],N} = \{\mathbf{X}_{[k+1]}^{(j)}\}_{j=1}^{N} \overset{\text{i.i.d}}{\sim} \widehat{f}_{[k+1]}$

  Evaluate $G_{[k+1],N} = \{g(\mathbf{X}_{[k+1]}^{(j)})\}_{j=1}^{N}$

  Estimate the empirical p_0-quantile $\widehat{q}_{[k+1]}^{p_0}$ of $G_{[k+1],N}$

  Set $k = k + 1$

Set total iteration number $k_{\#} = k - 1$

Estimate $\widehat{p}_{\text{f}} = (1 - p_0)^{k_{\#}} \cdot \frac{1}{N} \sum_{j=1}^{N} \mathbb{1}_{\{g(\mathbf{X}_{[k_{\#}]}^{(j)}) \geq y_{\text{th}}\}}(\mathbf{X}_{[k_{\#}]^{(j)}})$

▷ *Outputs:* ◁

$\widehat{p}_{\text{f}}$, estimate of p_{f}

As discussed in the previous section, EBC fitting is tuned by the Bernstein polynomial of order m, implying a bias-variance tread off. In Fig. 5.2, conditional distributions fitted by EBC (blue and brown isolines) seem to present a slight bias since they overlay the quantiles. However, reducing this bias implies decreasing the tuning parameter m, until $m = 1$, which is equivalent to an independent copula. Tools to control the goodness of fit of nonparametric conditional distributions are also available. As an example, let us consider the fitted conditional distribution at the first iteration (visible in Fig. 5.2). Its quantile-quantile plot in Fig. 5.3 shows a good fit of the two marginals by KDE. Then, the goodness of fit of copulas can be evaluated by Kendall's plot, represented in Fig. 5.4. This fit is also good, even if a slight bias is again visible.

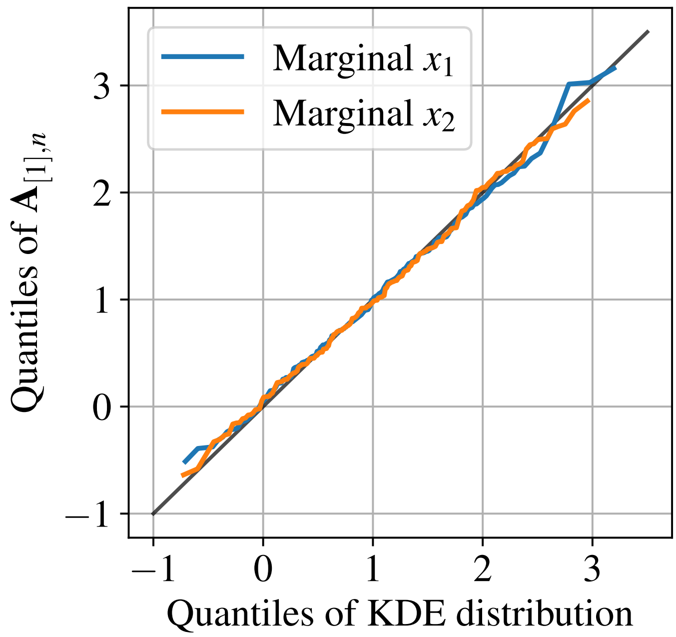

Fig. 5.3 QQ-plot for KDE of marginals of the conditional distribution from Fig. 5.2.

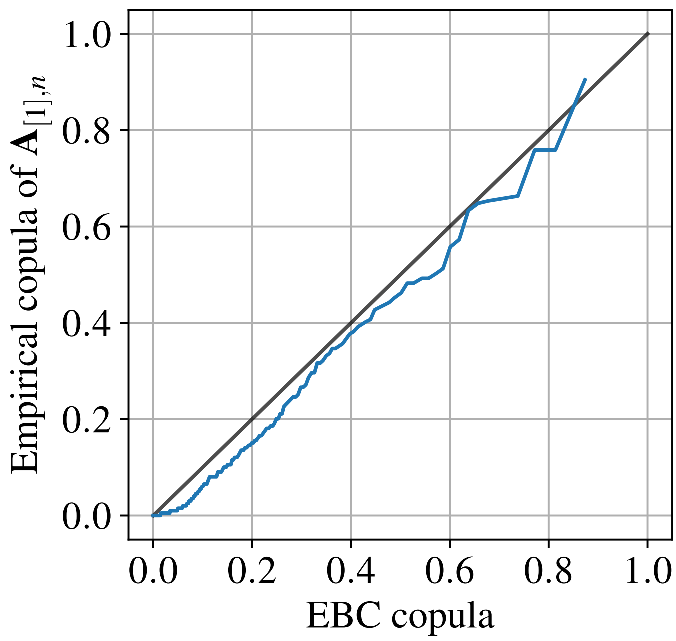

Fig. 5.4 Kendall plot for EBC on the copula of a conditional distribution from Fig. 5.2.

5.3 Numerical experiments

In the following analytical numerical experiments, the intermediary probabilities were set to $p_0 = 0.1$, allowing a fair comparison with subset sampling. Then, the subset sample size is set to $N = 10^4$, in order to get a reasonable sample size $n = Np_0 = 10^3$ to perform the nonparametric fitting. EBC tuning is setup to minimize the MISE in Eq. (5.6): $m = 1 + n^{\frac{2}{d+4}}$. In order to take into account the variability of the method's results, each experiment is repeated 100 times, allowing the computation of a coefficient of variation $\widehat{\delta} = \frac{\sigma_{\widehat{p_\mathrm{f}}}}{\mu_{\widehat{p_\mathrm{f}}}}$. Note that an implementation of the BANCS method and the following numerical experiments are available in a Git repository[2].

Toy-case #1: Parabolic reliability problem Let us define the parabolic reliability problem, considering the function $g_1 : \mathbb{R}^2 \to \mathbb{R}$:

$$g_1(\mathbf{x}) = (x_1 - x_2)^2 - 8(x_1 + x_2 - 5), \tag{5.7}$$

with the input random vector $\mathbf{X} = (X_1, X_2)$ following a standard 2-dimensional normal distribution. The reliability problem consists in evaluating: $p_{\mathrm{f},1} = \mathbb{P}(g_1(\mathbf{X}) \leq 0) = 1.31 \times 10^{-4}$.

Toy-case #2: Four-branch reliability problem Let us define the four-branch reliability problem (originally proposed by [20]), considering the following function $g_2 : \mathbb{R}^2 \to \mathbb{R}$:

$$g_2(\mathbf{x}) = \min \begin{pmatrix} 5 + 0.1(x_1 - x_2)^2 - \frac{(x_1+x_2)}{\sqrt{2}} \\ 5 + 0.1(x_1 - x_2)^2 + \frac{(x_1+x_2)}{\sqrt{2}} \\ (x_1 - x_2) + \frac{9}{\sqrt{2}} \\ (x_2 - x_1) + \frac{9}{\sqrt{2}} \end{pmatrix}, \tag{5.8}$$

with the input random vector $\mathbf{X} = (X_1, X_2)$ following a standard 2-dimensional normal distribution. The reliability problem consists in evaluating: $p_{\mathrm{f},2} = \mathbb{P}(g_2(\mathbf{X}) \leq 0) = 2.21 \times 10^{-4}$.

Toy-case #3: high-dimensional reliability problem Let us define the higher-dimensional reliability problem (proposed by [21]), considering the following function $g_3 : \mathbb{R}^7 \to \mathbb{R}$:

[2] https://github.com/efekhari27/icasp14

$$g_3(\mathbf{x}) = 15.59 \times 10^4 - \frac{x_1 x_3^2}{2x_3^2} \frac{x_2^4 - 4x_5 x_6 x_7^2 + x_4(x_6 + 4x_5 + 2x_6 x_7)}{x_4 x_5 (x_4 + x_6 + 2x_6 x_7)}, \tag{5.9}$$

with the input random vector $\mathbf{X} = (X_1, \dots, X_7)$, following a product of normal distributions defined in [21]. The reliability problem consists in evaluating: $p_{\mathrm{f},3} = \mathbb{P}(g_3(\mathbf{X}) \leq 0) = 8.10 \times 10^{-3}$.

5.3.1 Results analysis

Results of our numerical experiments are presented graphically (for 2-dimensional problems) in Figures 5.5 and 5.6, and numerically in Table 5.1. In the same fashion as the previous illustrations, the figures represent the intermediary quantiles $\widehat{q}_{[k]}^{p_0}$ estimated over conditional samples of size $N = 10^4$. Moreover, samples $\mathbf{A}_{[k+1],n}$ exceeding these quantiles are also represented in the same color. Notice how the last estimated quantile is set to the problem threshold $y_{\mathrm{th}} = 0$. To capture the dispersion of BANCS estimation, 100 repetitions were realized. Let us notice that for each toy-case, BANCS well estimates the failure probabilities' orders of magnitude. Yet the numerical values in Table 5.1 consistently present a positive bias, leading to an overestimated failure probability. This bias is partially explained by the EBC tuning choice and could be reduced at the expense of a slightly higher variance.

The variance obtained with the repetitions is quite large. Although, part of it is due to the fact that the algorithm might compute a different total number of subsets (e.g., toy-case #1 is either solved in four or five subsets). Overall, considering the EBC tuning from Eq. (5.6), BANCS performs worst than SS on toy-cases #1 and #2 but performs as well as SS on the toy-case #3. This might be due to the fact that toy-case #3 has a higher input dimension. However, one can note that SS coefficient of variation is computed by an approximation, tending to underestimate the true coefficient of variation (see e.g., [16]).

Table 5.1 Results of the numerical experiments (subset sample size $N = 10^4$, $p_0 = 0.1$).

	d	$p_{\mathrm{f}}^{\mathrm{ref}}$	$\widehat{p_{\mathrm{f}}}^{\mathrm{BANCS}}$	$\widehat{\delta}^{\mathrm{BANCS}}$	$\widehat{p_{\mathrm{f}}}^{\mathrm{SS}}$	$\widehat{\delta}^{\mathrm{SS}}$
Toy-case #1	2	1.31×10^{-4}	2.67×10^{-4}	24%	1.30×10^{-4}	9%
Toy-case #2	2	2.21×10^{-4}	4.23×10^{-4}	7%	2.24×10^{-4}	6%
Toy-case #3	7	8.10×10^{-3}	9.32×10^{-3}	15%	8.92×10^{-3}	6%

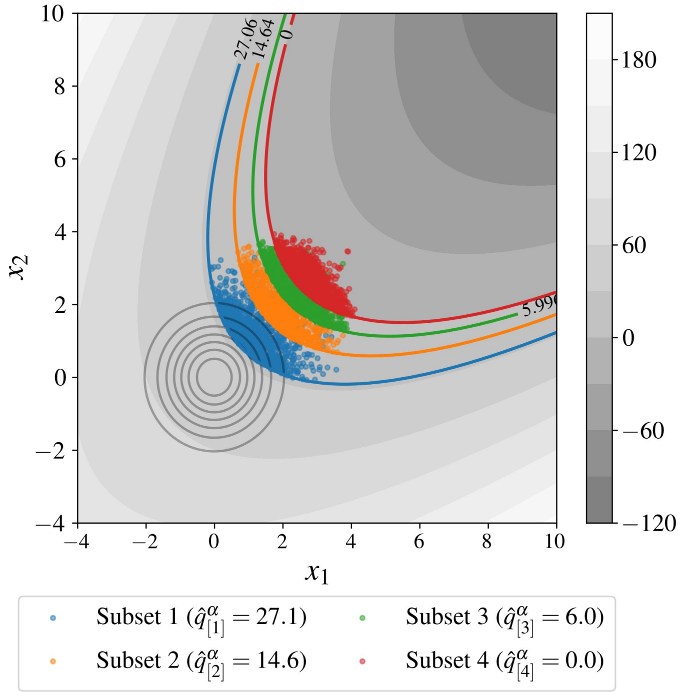

Fig. 5.5 BANCS sampling steps on toy-case #1.

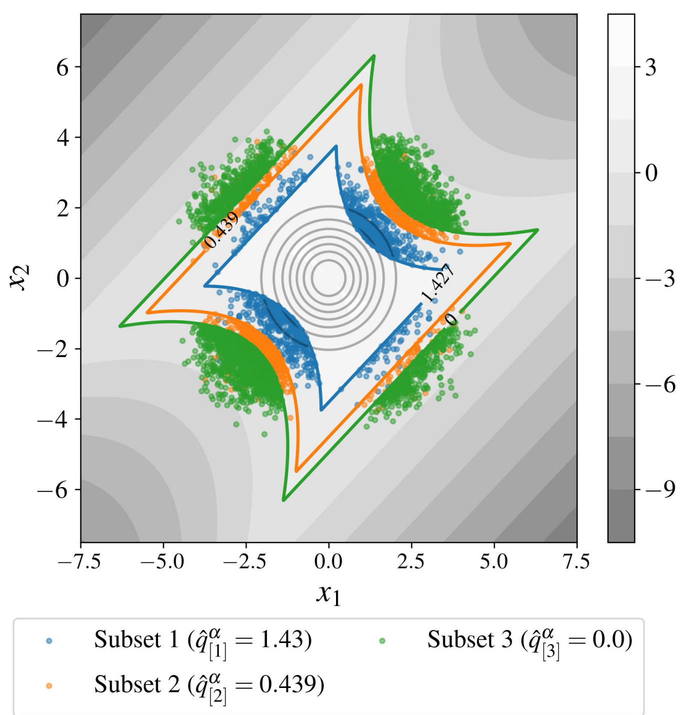

Fig. 5.6 BANCS sampling steps on toy-case #2.

5.4 Application to wind turbine fatigue reliability

5.5 Conclusion

Subset Simulation uses MCMC sampling to generate its intermediary conditional samples. However, MCMC algorithms tends to be complex to tune and does not generate i.i.d. conditional samples. In this work, a new method is proposed, replacing MCMC sampling with a simpler procedure. An intermediary conditional distribution is first fitted by a nonparametric approach, mixing kernel density estimation for fitting the marginals and Empirical Bernstein Copula (EBC) for fitting the copula. Then, the resulting allows to perform direct Monte Carlo sampling. This method is named “Bernstein adaptive nonparametric conditional sampling” (BANCS) and is applied to three toy-cases (two 2-dimensional and one 7-dimensional) and compared with SS.

The method shows promising results, even though a small positive bias consistently appears. This issue results from EBC tuning, creating a bias-variance tradeoff in the copula fit. Theoretical works offer optimal tuning, allowing us to find the optimal compromise. In our numerical experiments, an empirical estimation of BANCS variance is computed over a set of repetitions. BANCS estimated coefficient of variation is higher than SS approximated coefficient of variation. This work can be further explored by building an approximation of BANCS variance and confidence interval. One major advantage remains that the samples generated at each iteration are i.i.d. leading to a possible use of these samples to perform global reliability-oriented sensitivity analysis [11] in order to detect and analyze the most influential input variables leading to failure.

Chapter 6

Sequential reliability oriented sensitivity analysis

6.1 HSIC for GSA

6.2 HSIC for TSA & CSA

6.3 Sequential ROSA

6.4 Application to wind turbine fatigue reliability

Conclusion

References

[1] Au, S.-K. and Beck, J. L. (2001). Estimation of small failure probabilities in high dimensions by subset simulation. *Probabilistic Engineering Mechanics*, 16(4):263–277.

[2] Baudin, M., Dutfoy, A., Iooss, B., and Popelin, A. (2017). Open TURNS: An industrial software for uncertainty quantification in simulation. In Ghanem, R., Higdon, D., and Owhadi, H., editors, *Springer Handbook on Uncertainty Quantification*, pages 2001–2038. Springer.

[3] Cérou, F., Del Moral, P., Furon, T., and Guyader, A. (2012). Sequential monte carlo for rare event estimation. *Statistics and computing*, 22:795–808.

[4] Chabridon, V., Balesdent, M., Perrin, G., Morio, J., Bourinet, J.-M., and Gayton, N. (2021). Global reliability-oriented sensitivity analysis under distribution parameter uncertainty. *Mechanical Engineering under Uncertainties: From Classical Approaches to Some Recent Developments*, pages 237–277.

[5] Da Veiga, S., Gamboa, F., Iooss, B., and Prieur, C. (2021). *Basics and Trends in Sensitivity Analysis: Theory and Practice in R.* Society for Industrial and Applied Mathematics.

[6] Joe, H. (1997). *Multivariate Models and Multivariate Dependence Concepts.* Chapman and Hall.

[7] Joe, H. and Kurowicka, D. (2011). *Dependence modeling: vine copula handbook.* World Scientific.

[8] Kurtz, N. and Song, J. (2013). Cross-entropy-based adaptive importance sampling using Gaussian mixture. *Structural Safety*, 42:35–44.

[9] Lasserre, M. (2022). *Apprentissages dans les réseaux bayésiens à base de copules non-paramétriques.* PhD thesis, Sorbonne Université.

[10] Lebrun, R. (2013). *Contributions à la modélisation de la dépendance stochastique.* PhD thesis, Université Paris-Diderot – Paris VII. (in English).

[11] Marrel, A. and Chabridon, V. (2021). Statistical developments for target and conditional sensitivity analysis: Application on safety studies for nuclear reactor. *Reliability Engineering & System Safety*, 214:107711.

[12] Marrel, A., Iooss, B., and Chabridon, V. (2022). The ICSCREAM methodology: Identification of penalizing configurations in computer experiments using screening and metamodel – Applications in thermal-hydraulics. *Nuclear Science and Engineering*, 196:301–321.

[13] Morio, J. (2011). Non-parametric adaptive importance sampling for the probability estimation of a launcher impact position. *Reliability Engineering and System Safety*, 96(1):178–183.

[14] Morio, J. and Balesdent, M. (2015). *Estimation of Rare Event Probabilities in Complex Aerospace and Other Systems: A Practical Approach*. Woodhead Publishing, Elsevier.

[15] Moustapha, M., Marelli, S., and Sudret, B. (2022). Active learning for structural reliability: Survey, general framework and benchmark. *Structural Safety*, 96:102174.

[16] Papaioannou, I., Betz, W., Zwirglmaier, K., and Straub, D. (2015). MCMC algorithms for Subset Simulation. *Probabilistic Engineering Mechanics*, 41:89–103.

[17] Rubinstein, R. Y. and Kroese, D. P. (2008). *Simulation and the Monte Carlo Method*. Wiley, Second ed. edition.

[18] Sancetta, A. and Satchell, S. (2004). The Bernstein copula and its applications to modeling and approximations of multivariate distributions. *Econometric Theory*, 20(3):535–562.

[19] Segers, J., Sibuya, M., and Tsukahara, H. (2017). The empirical beta copula. *Journal of Multivariate Analysis*, 155:35–51.

[20] Waarts, P. (2000). *Structural reliability using finite element methods: an appraisal of directional adaptive response surface sampling (DARS)*. PhD thesis, Technical University of Delft, The Netherlands.

[21] Yun, W., Lu, Z., Zhang, Y., and Jiang, X. (2018). An efficient global reliability sensitivity analysis algorithm based on classification of model output and subset simulation. *Structural Safety*, 74:49–57.

Appendix A

Multivariate distribution modeling

Appendix B

Nonparametric copula estimation

Appendix C

Rare event estimation algorithms

Appendix D

Résumé étendu de la thèse