Identification of multi-modal random variables through mixtures of polynomial chaos expansions

Anthony Nouy

To cite this version:


HAL Id: hal-00521932
https://hal.archives-ouvertes.fr/hal-00521932
Submitted on 29 Sep 2010

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L’archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d’enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.
Identification of multi-modal random variables through mixtures of polynomial chaos expansions

Anthony Nouy

Abstract

A methodology is introduced for the identification of a multi-modal real-valued random variable from a collection of samples. The random variable is seen as a finite mixture of uni-modal random variables. A functional representation of the random variable is used, which can be interpreted as a mixture of polynomial chaos expansions. After a suitable separation of samples into sets of uni-modal samples, the coefficients of the expansion are identified by using an empirical projection technique. This identification procedure allows for a generic representation of a large class of multi-modal random variables with low-order generalized polynomial chaos representations.

Résumé

Identification de variables aléatoires multi-modales par mélange de décompositions sur la chaos polynômial. Une méthodologie est proposée pour l’identification d’une variable aléatoire multi-modale à partir d’échantillons. La variable aléatoire est vue comme un mélange fini de variables aléatoires uni-modales. Une représentation fonctionnelle de la variable aléatoire est utilisée. Elle peut être interprétée comme un mélange de décompositions sur le chaos polynômial. Après une séparation adaptée des échantillons en sous-ensembles d’échantillons uni-modaux, les coefficients de la décomposition sont identifiés en utilisant une technique de projection empirique. Cette procédure d’identification permet une représentation générique d’une large classe de variables aléatoires multi-modales avec une décomposition sur chaos polynômial généralisé de faible degré.

Key words: Uncertainty quantification; Identification; Multi-modal density; Polynomial Chaos; Finite Mixture Model; Spectral Stochastic Methods

Mots-clés : Quantification d’incertitudes; Identification; Densité multi-modale; Chaos Polynômial; Modèle de mélange fini; Méthodes Spectrales Stochastiques

Preprint 29 septembre 2010
1 Introduction

Uncertainty quantification and propagation in physical systems appear as a critical path for the improvement of the prediction of their response. For the numerical estimation of outputs of stochastic systems driven by finite-dimensional noise, the so-called spectral stochastic methods [7,13,20,9] have received a growing attention in the last two decades. These methods rely on a functional representation of random outputs, considered as second order random variables, by using truncated expansions on suitable Hilbertian basis. Classical basis consist in polynomial functions (finite-dimensional Polynomial Chaos [19,2,7]), piecewise polynomial functions [3,8,18] or more general orthogonal basis [16]. Of course, the accuracy of predictions depends on the quality of the input probabilistic model. Some works have been recently devoted to the identification of random variables (or processes), from a collection of samples, using Polynomial Chaos (PC) representations. Classical inference techniques have been used to identified the coefficients of functional expansions, such as maximum likelihood estimation [4,17] or Bayesian inference [6,1]. Polynomial Chaos a priori allows for the representation of second order random variables with arbitrary probability laws. However, for some classes of random variables, classical PC expansions may exhibit very slow convergence rates, thus requiring very high order expansions for an accurate representation. When introducing such representations for random input parameters of a physical model, very high order expansions are also required for an accurate approximation of random outputs. Classical spectral stochastic methods, such as Galerkin-type methods, then require to deal with high-dimensional approximation spaces, which leads to prohibitive computational costs. Although the use of efficient solvers or model reduction techniques based on separated representations [11,12,14] may help to reduce computational costs, a convenient alternative consists in identifying more suitable representations of random inputs.

The aim of the present article is to propose a PC-based numerical methodology for the identification of real-valued multi-modal random variables. In section 2, we briefly recall the basics of PC expansions of uni-variate random variables and introduce an empirical projection technique in order to identify these expansions from samples. This projection technique is an efficient alternative to classical inference techniques. We then illustrate the limitations of classical PC expansions when trying to represent multi-modal random variables. In section 3, we introduce a methodology for representing multi-modal real-valued random variables. From a theoretical point of view, it consists in introducing a complete set of events allowing a separation of modes. The probability density function of the random variable to be identified appears as a mixture of

Email address: anthony.nouy@ec-nantes.fr (Anthony Nouy).
probability density functions of random variables (finite mixture model \[10\]). We then propose a natural representation of the random variable on a generalized Polynomial Chaos, which can be interpreted as a “mixture of chaos expansions”, estimated from samples using an efficient empirical projection. Section 4 will illustrate the efficiency of the proposed methodology.

2 Polynomial chaos decomposition of a second order random variable

2.1 Polynomial chaos representation

Let \( X \) denote a real-valued random variable defined on an abstract probability space \( (\Theta, \mathcal{B}_\Theta, P) \). Let \( F_X \) denote its cumulative density function (CDF) and \( p_X \) its probability density function (PDF). We introduce a random variable \( \xi \) defined on \( (\Theta, \mathcal{B}_\Theta, P) \), with known support \( \Xi \subset \mathbb{R} \) and probability law \( P_\xi \), thus defining a new probability space \( (\Xi, \mathcal{B}_\Xi, P_\xi) \). The random variable \( g(\xi) := F_X^{-1} \circ F_\xi(\xi) \) have the same probability law as \( X \). We then make the hypothesis that \( g \) is a \( P_\xi \)-square integrable function from \( \Xi \) to \( \mathbb{R} \), i.e. \( g \in L^2(\Xi, dP_\xi) \).

Introducing an Hilbertian basis \( \{h_i\}_{i \in \mathbb{N}} \) of \( L^2(\Xi, dP_\xi) \), the random variable \( X \) then admits the following representation:

\[
X = \sum_{i \in \mathbb{N}} X_i h_i(\xi),
\]

with coefficients \( X_i \) being defined by

\[
X_i = \langle g, h_i \rangle_{L^2(\Xi, dP_\xi)} := E(g(\xi)h_i(\xi)),
\]

where \( E \) denotes the mathematical expectation. An approximate representation of \( X \) can be obtained by truncating the decomposition:

\[
X \approx \sum_{i=0}^p X_i h_i(\xi).
\]

A classical choice for the \( h_i \) consists in polynomial functions orthonormal with respect to scalar product \( \langle \cdot, \cdot \rangle_{L^2(\Xi, dP_\xi)} \), thus leading to the so-called uni-dimensional Polynomial Chaos (PC) expansion of degree \( p \) of \( X \) \([7,21,16]\).

2.2 Identification of the decomposition

Classical inference techniques have been applied for the identification of coefficients \( X_i \) from a collection of independent samples \( \{X^{(k)}\}_{k=1}^Q \) of \( X \): maximum likelihood estimation \([4]\), Bayesian inference \([6]\). Here, in order to estimate the coefficients, we use a simple and efficient numerical technique, named “empirical projection”. It is based on the estimation of mapping \( g \) from samples and the introduction of a quadrature scheme to compute its projection on the PC basis. We denote by \( \tilde{F}_X \) the empirical CDF of \( X \), estimated from samples:

\[
\tilde{F}_X(x) = \frac{1}{Q} \sum_{k=1}^Q I(X^{(k)} \leq x),
\]

where \( I(A) \) is the indicator function of event \( A \). We then introduce the following approximation

\[
g(\xi) \approx \tilde{F}_X^{-1} \circ F_\xi(\xi),
\]

where \( \tilde{F}_X^{-1} : [0, 1] \to \mathbb{R} \) is uniquely defined as

\[
\tilde{F}_X^{-1}(y) = \min \{x \in \{X^{(k)}\}_{k=1}^Q; \tilde{F}_X(x) \geq y\}.
\]

Then, the coefficients of the
Chaos expansion can be approximated using a numerical integration:

\[ X_i = \int_{\Xi} F^{-1}_X(F_\xi(y))h_i(y)dP_\xi(y) \approx \sum_{k=1}^{N} \omega_k \tilde{F}^{-1}_X(F_\xi(y_k))h_i(y_k) \tag{1} \]

where the \( \{\omega_k, y_k\}_{k=1}^{N} \) are integration weights and points. In practice, an accurate Gauss-quadrature associated with measure \( P_\xi \) can be used.

### 2.3 Limitations of classical polynomial chaos representations

Classical polynomial chaos decompositions allow for an accurate representation of a wide class of probability laws. The accuracy can be simply improved by choosing a suitable germ \( \xi \) (Gaussian, Uniform,...), associated with classical orthogonal polynomial basis (Hermite, Legendre,...). However, these classical polynomial decompositions may not be adapted for some classes of random variables, particularly for multi-modal random variables. Figure 1 illustrates the convergence of a Hermite polynomial chaos expansion\(^1\) of a bi-modal random variable \( X(\theta) \), defined by

\[ X(\theta) = \begin{cases} 
  a(\theta) - \delta & \text{if } b(\theta) < 1/3 \\
  a(\theta) + \delta & \text{if } b(\theta) > 1/3 
\end{cases} 
\]

where \( a \) and \( b \) are independent standard Gaussian random variables and where \( \delta \) is a parameter controlling the separation of modes. We observe that when increasing \( \delta \), the convergence of a classical PC expansion drastically deteriorates, thus needing for a high polynomial degree for an accurate representation of the PDF.

![Figure 1. PC expansion of a bi-modal random variable: convergence with the expansion’s degree \( p \)](image)

Figure 1. PC expansion of a bi-modal random variable: convergence with the expansion’s degree \( p \)

---

1. Expansions are identified with the empirical projection technique, using a highly accurate (and converged) Gauss-Hermite quadrature.
3 Identification of a multi-modal random variable

3.1 Mixture of probability laws

Let us denote by \( m \) the number of modes of the random variable \( X \) (which can be defined as the number of local maxima of \( p_X \)). We introduce a complete set \(^2\) of \( m \) events \( \{ \Theta_i \}_{i=1}^m \) of \( \mathcal{B}_\Theta \) and we define associated real-valued random variables \( Y_i \) with probability law defined for all \( B \in \mathcal{B}_\mathbb{R} \) by

\[
P_{Y_i}(B) = P(X \in B|\Theta_i) = P(X^{-1}(B)|\Theta_i) = P(X^{-1}(B) \cap \Theta_i)/P(\Theta_i).
\]

We admit that the events \( \Theta_i \) are such that the \( Y_i \) are uni-modal random variables. The probability law of \( X \) can then be defined by: \( \forall B \in \mathcal{B}_\mathbb{R}, \)

\[
P_X(B) = \sum_{i=1}^m P(X^{-1}(B) \cap \Theta_i) = \sum_{i=1}^m P_{Y_i}(B)P(\Theta_i).
\]

Its probability density function then appears as a mixture \(^{[10]}\) (i.e. convex combination) of probability density functions of uni-modal random variables:

\[
p_X(x) = \sum_{i=1}^m p_{Y_i}(x)P(\Theta_i).
\]

(2)

The identification of \( X \) is then replaced by the identification of random variables \( Y_i \), which are expected to admit accurate low-order chaos representations. The questions are now: how to define the partition \( \{ \Theta_i \}_{i=1}^m \), what kind of chaos representation can be used for \( X \) and how to identify this representation from samples?

3.2 Definition of \( \Theta_i \) by separation of samples

We introduce an artificial separation of samples \( \{ X^{(k)} \}_{k=1}^Q \) into \( m \) sets of uni-modal samples, which allow for the construction of the desired partition of \( \Theta \). We suppose that the empirical PDF allows estimating a set of points \( \{ x_i \}_{i=1}^{m-1} \) that separate samples into \( m \) sets of uni-modal samples, defined as follows:

\[
\mathcal{X}_i = \{ X^{(k)}, k \in \{ 1 \ldots Q \}; X^{(k)} \in [x_{i-1}, x_i), \quad i \in \{ 1 \ldots m \}, \quad \text{where by convention} \quad x_0 = -\infty \quad \text{and} \quad x_m = +\infty. \quad \text{Then,} \quad \Theta_i \quad \text{is defined as the abstract event associated with samples in} \quad \mathcal{X}_i \quad (i.e. \quad \mathcal{X}_i \subset X(\Theta_i)). \quad \text{The probability of event} \quad \Theta_i \quad \text{is then defined by} \quad P(\Theta_i) = \text{Card}(\mathcal{X}_i)/Q. \quad \text{In order to define}
\]

\[^{2}\cup_{i=1}^m \Theta_i = \Theta, \quad \Theta_i \cap \Theta_j = \emptyset \quad \text{for} \quad i \neq j\]
Let us now denote $\xi_2$ another random variable, independent on $\xi_1$. Letting $\xi = (\xi_1, \xi_2)$, we define a 2-dimensional probability space $\mathbb{P} \mathcal{B}_\mathbb{P} = (\Xi, \mathcal{B}_\Xi, P_\xi)$, with $\Xi = \Xi_1 \times \Xi_2$ and $P_\xi = P_{\xi_1} \otimes P_{\xi_2}$. Random variable $X$ is then seen as a function of $\xi$, defined by $X(\xi) = \sum_{i=1}^{m} I_{B_i}(\xi_1)Y_i(\xi_2)$, where $I_{B_i}$ denotes the indicator function of $B_i$. We next introduce a chaos representation of each random variable $Y_i = \sum_{j \in \mathbb{N}} X_{i,j} h_j(\xi_2)$, where the $h_j$ are classical orthonormal polynomials in $L^2(\Xi_2, dP_{\xi_2})$. A generalized chaos representation of $X$ is then sought as:

$$X(\xi) = \sum_{i=1}^{m} I_{B_i}(\xi_1)\left( \sum_{j \in \mathbb{N}} X_{i,j} h_j(\xi_2) \right) = \sum_{i=1}^{m} \sum_{j \in \mathbb{N}} X_{i,j} I_{B_i}(\xi_1) h_j(\xi_2), \quad (4)$$

which can be interpreted as a mixture of polynomial chaos expansions. Functions $\{I_{B_i} h_j\}$ form an orthogonal set of functions in $L^2(\Xi, dP_\xi) = L^2(\Xi_1, dP_{\xi_1}) \otimes L^2(\Xi_2, dP_{\xi_2})$, composed by piecewise polynomial functions (polynomial with respect to $\xi_2$ and piecewise constant with respect to $\xi_1$). The $L^2$-norm of a basis function is $E((I_{B_i}(\xi_1) h_j(\xi_2))^2)^{1/2} = P(\Theta_i)^{1/2}$. Coefficients $\{X_{i,j}\}$ of the decomposition of $X$ are defined as the orthogonal projections of $X$ on these basis functions:

$$X_{i,j} = P(\Theta_i)^{-1} \int_{\Xi_1 \times \Xi_2} X(y_1, y_2) I_{B_i}(y_1) h_j(y_2) p_{\xi_2}(y_2) dy_1 dy_2 = \int_{\Xi_2} Y_i(y_2) h_j(y_2) p_{\xi_2}(y_2) dy_2. \quad (5)$$

### 3.4 Identification of the decomposition from samples

Classical inference techniques [15,5] could be used in order to identify from samples the set of $m(p + 1)$ coefficients \(^3\) of the mixture of chaos expansions (4). However, the number of parameters is such that these classical techniques lead to high computational costs. With a maximum likelihood estimation, the identification requires the resolution of a hard optimization

\(^3\) Note that samples separation values $x_i, i = 1\ldots m - 1$, could also be added to the set of parameters to be identified.
problem (high dimension, objective function with many local maxima, possibly non-smooth function) for which classical algorithms may lack robustness. Here, we propose to apply the empirical projection technique introduced in section 2.2. The random variable $Y_i$ is written in terms of $\xi_2$ in the following way:

$$Y_i(\xi_2) = F_{Y_i}^{-1}(F_{\xi_2}(\xi_2)),$$

where $F_{Y_i} = F_{X|\Theta_i}$ is the conditional CDF of $X$ knowing $\Theta_i$. The subset of samples $X_i$ corresponds to independent samples of $Y_i$. Therefore, an approximation $\tilde{F}_{Y_i}$ of $F_{Y_i}$ can be simply estimated by

$$\tilde{F}_{Y_i}(y) = \frac{1}{\text{Card}(X_i)} \sum_{z \in X_i} I(z \leq y).$$

Random variable $X$ is then approximated by truncating polynomial chaos expansions to a degree $p$, the coefficients being estimated from samples in the following way: $\forall i \in \{1, \ldots, m\}$, $\forall j \in \{0, \ldots, p\}$,

$$X_{i,j} \approx \sum_{k=1}^{N} \omega_k F_{Y_i}^{-1}(F_{\xi_2}(y_k))h_j(y_k),$$

where the $\{\omega_k, y_k\}_{k=1}^{N}$ are integration weights and points of a classical quadrature rule (e.g. Gauss-quadrature) associated with measure $P_{\xi_2}$.

4 Numerical illustration

**Exemple 1:** We generate an artificial collection of $Q = 1,000$ samples from the bi-modal random variable defined in section 2.3. We consider three cases corresponding to the following three values of the mode-separation parameter: $\delta = 1.5$, $\delta = 2$ and $\delta = 3$. The corresponding empirical PDFs of samples are shown on Figures 2(a-c). On the same figures, also plotted are the PDFs associated with a mixture of Hermite polynomial chaos expansions ($\xi_2$ is a Gaussian random variable). The coefficients of the expansion have been obtained with the empirical projection technique (see section 3.4), using a 15-points Gauss-Hermite quadrature for the numerical integration. For the three cases, samples have been separated into two sets of uni-modal samples by choosing separation values $x_1 = 0$, $0.2$ and $0$ respectively. Theses values have been determined by simply locating local minima of the empirical PDFs. Whatever the separation of modes, we observe a very good agreement between the empirical PDFs of samples and the identified mixture of polynomial chaos expansions (4), even with a low degree of expansion ($p = 2$ or $3$).

**Exemple 2:** We consider a collection of $Q = 1,000$ samples corresponding to a 3-modal distribution, represented on Figure 3(a). On Figure 3(b), we illustrate the bad convergence of a classical Hermite PC expansion $X \approx \sum_{i=0}^{p} X_i h_i(\xi)$ ($\xi$ being a standard Gaussian random variable). The coefficients are computed using the empirical projection technique (see section 2.2) with a high order Gauss-Hermite quadrature (100 quadrature points). Figure 3(c)
Figure 2. Mixture of PC expansions for a bi-modal random variable: convergence with the expansion’s degree $p$.

illustrates the PDF obtained with a 3-modal mixture of Hermite PC expansions of degree $p$: $X(\xi_1, \xi_2) = \sum_{i=1}^{3} \sum_{j=0}^{p} X_{i,j} I_{B_i}(\xi_1) h_j(\xi_2)$. The samples have been separated into three sets of uni-modal samples by choosing separation values $x_1 = 4$ and $x_2 = 9$. These values have been chosen by approximatively locating the two local minima of the empirical PDF. For the computation of expansion coefficients, we have used the empirical projection technique introduced in section 3.4, with a 15-points Gauss-Hermite quadrature. We observe that a very good representation of the random variable is obtained with a mixture of PC expansions of low degree ($p \approx 3$). Figure 4 shows the weighted PDFs $y \mapsto P(\Theta_i)p_{Y_i}(y)$ of random variables $Y_i$. It also shows the resulting PDF of $X$, which appears as the mixture of the weighted conditional PDFs.

Figure 3. Probability density functions: samples (a), Hermite PC expansion (b), Mixture of Hermite PC expansions (c).

5 Conclusion

In this paper, we have introduced an efficient numerical technique for the identification of real-valued multi-modal random variables. A mixture of chaos representations is used, which can be interpreted as a 2-dimensional generalized polynomial chaos expansion. The expansion basis is defined by the product of polynomial functions of a first random variable and piecewise constant
functions of a second random variable. The coefficients of the expansion are estimated from samples by using an efficient empirical projection technique. Classical inference techniques such as maximum likelihood or Bayesian inference could also be used for estimating the coefficients of the expansion, although leading to much higher computational costs. The proposed mixture of polynomial chaos expansions and the empirical projection technique can be extended to vector-valued random variables. The empirical projection technique however requires more and more samples as the dimension increases.

References


