Use of polynomial chaos expansions and maximum likelihood estimation for probabilistic inverse problems

Frédéric Perrin [1,2], Bruno Sudret [3], Géraud Blatman [1,3] & Maurice Pendola [2]

[1] LaMI, Institut Français de Mécanique Avancée et Université Blaise Pascal, Campus des Cézeaux, 63175 Aubière Cedex, FRANCE
[2] Phimeca Engineering S.A., 1 Allée Alan Turing, 63170 Aubière, FRANCE - *perrin@phimeca.com*
[3] Electricité de France, R&D Division, Site des Renardières, 77818 Moret-sur-Loing, FRANCE

Abstract :

The present paper deals with the identification of probabilistic models of input variables using response measurements. The input random variables, whose probability density function has to be identified, are represented by their polynomial chaos expansion (PCE). The proposed method allows to solve the probabilistic inverse problem using an efficient maximum likelihood approach. An advanced optimization algorithm is used to maximize this likelihood and get the optimal values of unknown PCE coefficients. The approach is illustrated by determining the variability of the loading applied to a series of similar simply supported beams when a database of measured maximum deflection is at hand.

Key-words :

Polynomial chaos expansion - Maximum likelihood method - Probabilistic inverse problem

1 Introduction

In probabilistic engineering mechanics, the physical system under consideration is represented by a numerical model whose input parameters are random variables. In practice, the accurate description of these input (or basic) random variables can be done using classical statistical tools when a sufficient amount of data is available. If not, the domain of Bayesian statistics may help combine prior information on the variability of the parameters (*e.g.* expert judgment) and few experimental values.

However, in many situations, the data that can be easily collected concerns *response quantities* (*e.g.* displacements, strain, etc.) instead of input parameters (*e.g.* material properties, loading, etc.). In these situations, so-called *inverse probabilistic methods* have to be devised in order to properly identify the probabilistic model for the input random variables. Most of the available literature in this domain adresses the problem in a Bayesian context (Yuen and Katafygiotis, 2002; Tarantola, 2005). Recently, polynomial chaos expansion techniques (PCE), originally used for uncertainty propagation in the context of stochastic finite element analysis (Ghanem and Spanos, 1991) have been used. Desceliers et al. (2006) have proposed a method based on the maximum likelihood concept to identify the coefficients of a PCE representation of the spatially varying Young's modulus of a structure by solving *deterministic* inverse three-dimensional elasticity problems and a statistical post-processing of the latter. Attempts to applying this kind of approach within a Bayesian framework to stochastic models have also been carried out by Ghanem and Doostan (2006).

In the present paper, the probabilistic model consists of both random variables with *prescribed* distribution and random variables with *unknown* distribution. The aim of the paper is to develop an indirect method for identifying the probability density function (PDF) of the latter.

These variables of unknown PDF are represented by their polynomial chaos expansion (Section 2). Then the likelihood of the observations (of response quantities possibly perturbed by measurement or model error) is formulated conditionally to these expansion coefficients. An advanced optimization algorithm is finally used to maximize this likelihood and get the PCE coefficients of the unknown variables (Section 3).

To illustrate the approach, the problem of determining the variability of the loading applied to a series of similar simply supported beams is solved, introducing the measurements of the maximal deflection of each beam as the experimental data (Section 4).

2 Polynomial chaos expansion of random variables

Let $(\Omega, \mathcal{F}, P)$ be a probability space, where Ω is the sample space, $\mathcal{F}$ is the σ-algebra of subsets of Ω and P is a probability measure. Let $X(\omega)$ be a real random variable (ω represents randomness), whose probability density function (PDF) is denoted by $f_X(x)$. The space of real random variables with finite variance equipped with the inner product $\langle X, Y \rangle \stackrel{\text{def}}{=} \mathrm{E}\left[XY\right]$ is an Hilbert space denoted by $L_2\left(\Omega, \mathcal{F}, P\right)$ ($\mathrm{E}\left[.\right]$ denotes the mathematical expectation. It can be shown that any random variable in this space can be cast as a polynomial series expansion in a standard normal variable ξ:

$$X(\omega) = \sum_{i=0}^{\infty} a_i H_i\left(\xi\left(\omega\right)\right) \tag{1}$$

In this expression, $\{H_i(x),\, i \in \mathbb{N}\}$ are the Hermite polynomials, which form an orthogonal family with respect to the Gaussian probability measure $\varphi(x) = \exp(-x^2/2)/\sqrt{2\pi}$:

$$\langle H_i, H_j \rangle = \int_{-\infty}^{+\infty} H_i\left(x\right) H_j\left(x\right) \varphi\left(x\right) dx = \delta_{ij}\; i! \quad (\delta_{ij} \text{ Kronecker symbol}) \tag{2}$$

In other words, the coefficients $\{a_i,\, i \in \mathbb{N}\}$ in Eq.(1) are the “coordinates” of X in the Hermite polynomial basis.

When the PDF of X is prescribed, the PCE coefficients can be computed by projection or regression (Sudret et al., 2006). In practice, unimodal random variables can be accurately approximated using a truncated series expansion $X = \sum_{i=0}^{p} a_i H_i\left(\xi\left(\omega\right)\right)$ where $p = 3$ is usually sufficient (Berveiller, 2005). When several random variables are to be simultaneously represented by PCE expansions, a one-to-one correspondance with as many components of a standard normal Gaussian vector is used.

3 Inverse probabilistic identification

3.1 Formulation

Let $\tilde{y}$ be the true value of the response quantity y of a mechanical system. Suppose that this quantity can be predicted by a mathematical model $\mathcal{M}$, which depends on a vector $\boldsymbol{x}$ of input parameters.

If the mechanical model $\mathcal{M}$ was “perfect” and if the true value $\tilde{\boldsymbol{x}}$ of the input parameters was known for the system under consideration, one could write:

$$\tilde{y} = \mathcal{M}\left(\tilde{\boldsymbol{x}}\right) \tag{3}$$

In practice, none of these assumptions hold. Indeed, the input parameters are usually not well known, leading to the introduction of a random vector $\boldsymbol{X}(\omega)$ for their modeling. In some cases, the random response $Y(\omega)$ can be measured by an analyst through experimental investigations. The measurement value y_{mes} may differ from the true value $\tilde{y}$:

$$y_{mes} = \tilde{y} + e = \mathcal{M}(\tilde{\boldsymbol{x}}) + e \tag{4}$$

where e is a realization of ϵ, which takes into account the deviation between predicted and measured values of y, *i.e.* it encompasses both measurement and model errors. ϵ is supposed to be a zero-mean Gaussian random variable with known variance σ_e^2.

Provided $\tilde{\boldsymbol{x}}$ is known, Eq.(4) can be interpreted as the fact that y_{mes} is a realization of a Gaussian random variable whose mean value is $\mathcal{M}(\tilde{\boldsymbol{x}})$ and whose standard deviation is σ_e. Thus, the corresponding likelihood of y_{mes} is:

$$f_{Y|\boldsymbol{X}}(y_{mes}|\tilde{\boldsymbol{x}}) = \frac{1}{\sigma_e}\varphi\left(\frac{y_{mes} - \mathcal{M}(\tilde{\boldsymbol{x}})}{\sigma_e}\right) \tag{5}$$

where φ is the standard normal probability density function (PDF).

In the sequel, it is supposed that a single random variable, say X^u, has an unknown PDF to be identified. This random variable will be approximated using a truncated Hermite polynomial chaos expansion in a standrad normal variable ξ^u:

$$X^u \approx \sum_{i=0}^{p} a_i^u H_i(\xi^u) \tag{6}$$

The remaining input variables of the problem are gathered into a random vector $\boldsymbol{X}^{-u} = \left(X^1, \dots, X^{u-1}, X^{u+1}, \dots, X^M\right)$, which can be represented through an isoprobabilistic transform T as a function of a standard normal vector $\boldsymbol{\xi}^{-u} = \left(\xi^1, \dots, \xi^{u-1}, \xi^{u+1}, \dots, \xi^M\right)$:

$$\boldsymbol{X}^{-u}(\omega) = T\left(\boldsymbol{\xi}^{-u}\right) \tag{7}$$

where $\boldsymbol{\xi}^{-u} = \left(\xi^1, \dots, \xi^{u-1}, \xi^{u+1}, \dots, \xi^M\right)$. In the case when the components of $\boldsymbol{X}^{-u}$ are independent, this transform reduces to a one-to-one mapping $X^i = F_{X^i}^{-1} \circ \Phi(\xi^i)$, where F_{X^i} is the CDF of X^i and Φ is the standard normal CDF. In case of correlated components, the Nataf transform may be used. The above assumptions allow to approximate Eq.(5) using an explicit mapping from $\boldsymbol{x}$ to $\boldsymbol{\xi}$:

$$f_{Y|\boldsymbol{\xi}}(y_{mes}|\boldsymbol{\xi}, \boldsymbol{a}^u) = \frac{1}{\sigma_e}\varphi\left(\frac{y_{mes} - \mathcal{M}\left(\sum_{i=0}^{p} a_i^u H_i(\xi^u), F_{X^i}^{-1}\left(\boldsymbol{\xi}^{-u}\right)\right)}{\sigma_e}\right) \tag{8}$$

where $\boldsymbol{\xi} = \left\{\boldsymbol{\xi}^{-u}, \xi^u\right\}$ and $\boldsymbol{a}^u = \{a_i^u\}_{i=0}^p$ and the approximation is due to the truncated expansion of X^u.

3.2 Maximum likelihood estimation

The PDF of variable Y can be cast as follows:

$$f_Y(y_{mes}) = \int f_{Y|\boldsymbol{X}}(y_{mes}|\boldsymbol{x})\, f_{\boldsymbol{X}}(\boldsymbol{x})\, d\boldsymbol{x} \tag{9}$$

where $f_{Y|\boldsymbol{X}}$ is the conditional PDF given in Eq.(5) and $f_{\boldsymbol{X}}$ refers to the known joint PDF of input parameters, in the case when all the random variables have been identified. Nevertheless, in the present case the joint PDF $f_{\boldsymbol{X}}$ is not known since a random variable has to be identified. Thus, by replacing the likelihood function $f_{Y|\boldsymbol{X}}$ by Eq.(8), Eq.(9) becomes:

$$\begin{aligned} f_Y\left(y_{mes}|\boldsymbol{a}^u\right) &= \int_{\mathbb{R}^M} f_{Y|\boldsymbol{\xi}}\left(y_{mes}|\boldsymbol{\xi},\boldsymbol{a}^u\right)\varphi\left(\boldsymbol{\xi}\right)d\boldsymbol{\xi} \\ &= \int_{\mathbb{R}^M} \frac{1}{\sigma_e}\varphi\left(\frac{y_{mes}-\mathcal{M}\left(\sum_{i=0}^{p} a_i^u H_i\left(\xi^u\right), F_{X^i}^{-1}\left(\boldsymbol{\xi}^{-u}\right)\right)}{\sigma_e}\right)\varphi_M(\boldsymbol{\xi})d\boldsymbol{\xi} \end{aligned} \quad (10)$$

Let $y_1,\ldots,y_Q$ be an experimental set of independent observations of Y. The likelihood function of the samples $\{y_q\}_{q=1}^{Q}$ can be approximated as:

$$L_Y\left(y_1,\ldots,y_Q|a_0^u,\ldots,a_p^u\right) = \prod_{q=1}^{Q} f_Y\left(y_q|\boldsymbol{a}^u\right) \quad (11)$$

In order to estimate the coefficients $\left\{a_0^u,\ldots,a_p^u\right\}$, one has to solve the following maximum likelihood optimization problem:

$$\boldsymbol{a}^{u^*} = \arg\min_{\boldsymbol{a}^u\in\mathbb{R}^{p+1}}\left(-\ln f_Y\left(y_1,\ldots,y_q|\boldsymbol{a}^u\right)\right) = \arg\min_{\boldsymbol{a}^{\boldsymbol{u}}\in\mathbb{R}^{p+1}}\left(-\sum_{q=1}^{Q}\ln f_Y\left(y_q|\boldsymbol{a}^u\right)\right) \quad (12)$$

3.3 Computational aspects

The above relation refers to an unconstrained optimization problem which has to be solved with a well-suited algorithm. Classical quasi-Newton methods need explicit gradient information, usually obtained by the use of finite differences. In the present case, choosing an appropriate step size for approximating the gradient function by finite differences is quite delicate. Indeed, the function to be minimized in this study (related to Eq.(12)) may be very sensitive to the step size chosen for the polynomial chaos coefficients $\boldsymbol{a}^{\boldsymbol{u}}$ that have to estimated.

Practically speaking,the usual gradient-based algorithm fail to converge due to the fact that the function to optimize present flat regions. The CONDOR (*COnstrained, Non-linear, Direct, parallel Optimization using trust Region method for high-computing load function*), developed by Vanden Berghen and Bersini (2004), appears to be useful in the present case. This optimization technique refers to a special class of optimization algorithms, named Trust-Region methods. These methods construct local models (linear or quadratical approximations in CONDOR built using multivariate Lagrange interpolation) in a ball of predefined radius. A local model approximates the function of interest, using the least number of function evaluations. The solution is obtained by reducing the trust region radius of the sampled space in each iteration of the algorithm.

Another computational challenge leads in the evaluation of the integral defined in Eq.(10). One can use Monte-Carlo simulation but it is not attractive since it requires a large number of samples of the mechanical model $\mathcal{M}$, which may be a problem when using high-time-computing model (*e.g.* finite element codes). As an alternative, *Gaussian quadrature schemes*

(Abramowitz and Stegun, 1970) may be used to evaluate the integral. Thus, Eq.(10) may be computed by:

$$f_Y \left(y_{mes} | \boldsymbol{a}^u \right) \approx \sum_{k_1=1}^{K} \ldots \sum_{k_N=1}^{K} w_{k_1} \ldots w_{k_N} f_{Y|\boldsymbol{\xi}} \left(y_{mes} | \xi_{k_1}, \ldots, \xi_{k_N}, \boldsymbol{a}^u \right) \quad (13)$$

Both Monte-Carlo simulation (MCS) method and the Gauss-Hermite are applied and compared in the following application example.

4 Application example: simply supported beam submitted to midspan load

4.1 Mechanical problem statement

This example application deals with the probabilistic identification of the midspan load F applied to a series of identical simply supported bending beams (Fig.1).

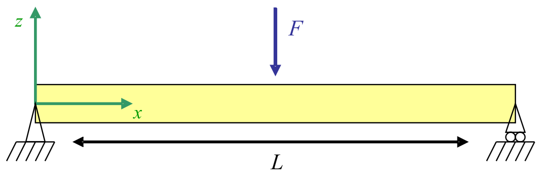


Figure 1: Bending beam submitted to midspan load

In the case of a simply supported beam, the maximal deflection is given by:

$$z_{max} = \mathcal{M} \left(F, L, E, I \right) = \frac{F L^3}{48 E I} \quad (14)$$

where L is the beam length, E is the Young's modulus and $I = bh^3/12$ is the moment of inertia which, in case of rectangular cross-section, depends on the width b and height h of the beam.

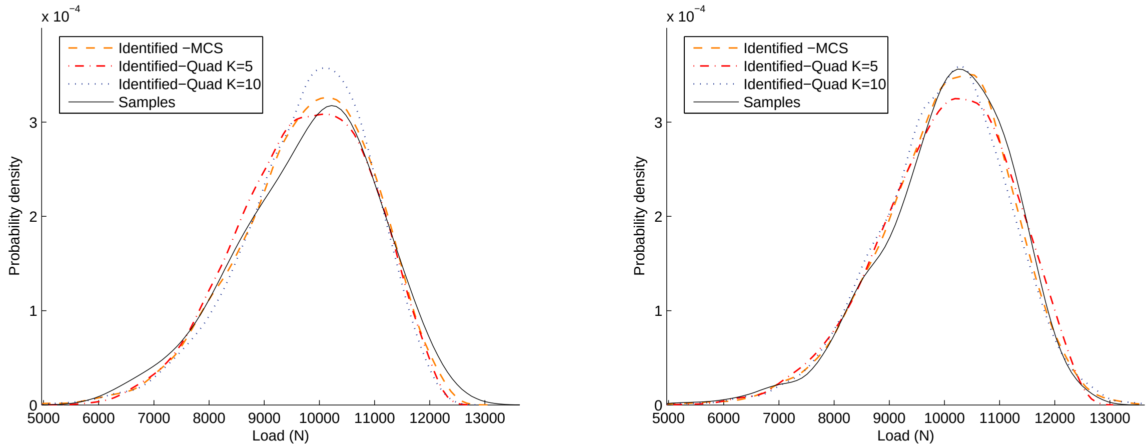
4.2 Probabilistic model

It is assumed that the only random parameter of known PDF is the Young's modulus E. Parameters related to the geometry of the beam are supposed to be deterministic (L =1 m, $b = h$ =0.1 m). The point of this application example is to identify the PDF of random variable F, *i.e.* to compute its PC coefficients. The proposed maximum likelihood estimation method is applied using an experimental database made of Q samples and associated measured deflections.

In the real world, such a database would be obtained by a comprehensive experimental program, *e.g.* inspection of a series of identical beams in a complex structure (bridges, packing of cooling tower, etc.). For the sake of illustration, the database is built here by Monte-Carlo simulation of the beam model $\mathcal{M}$ and by adding a measurement noise sampled from a zero-mean Gaussian random variable with standard deviation $\sigma_e = 0.005$ m. In these simulations, it is assumed that the load follows a Weibull distribution with mean $m_F = 10,000$ N and standard deviation $\sigma_F = 1,200$ N. The aim of the application example is to compute the optimal PC coefficients of F and show that the associated PDF is closed to the one used to set up the database.

4.3 Results

The proposed maximum likelihood method has been applied to two databases of measured deflections, comprising 100 and 1,000 samples respectively. In each case, the integral defined in Eq.(10) has been computed using 10,000 Monte-Carlo samplings and the Gauss-Hermite quadrature approximation using 5 (resp. 10) integration points.



(a) $Q = 100$ (b) $Q = 1000$

Figure 2: Identified probability density functions for the two databases

Plots of the three identified PDFs (approximated by a kernel PDF representation (Wand and Jones, 1995)) are shown in Figure 2, where the kernel PDF of the database is also plotted. It appears that identified PDFs of the load F are close to the histogram of the deflection database in each case.

As expected, the identified PDFs are improved when using a larger experimental database ($Q = 1,000$), as seen from Figure 2(b). The quadrature method (with 10 integration points) requires $10^2 = 100$ model calls and is rather accurate on this example at a cost which is 100 times less expensive than MCS.

Table 1: Estimates of four first statistical moments

Moments	Samples	Identification ($Q = 100$)		
		MC ($K = 10,000$)	Quad ($K = 5$)	Quad ($K = 10$)
Mean $\hat{m}$	9785.8	9761.7	9713.7	9774.8
Standard deviation $\hat{\sigma}$	1202.5	1205.7	1168.1	1151.7
Skewness $\hat{\delta}$	-0.486	-0.492	-0.317	-0.615
Kurtosis $\hat{\kappa}$	2.832	3.137	2.664	3.453

The four first statistical moments of the load F are further computed from the identified PDF. Estimates obtained from the PCE coefficients computed by one of the three integration methods (MCS with 10,000 samples, quadrature method with $K = 5$ (resp. $K = 10$) integration points) are reported in Table 1. The empirical moments of the database of F (which is in a real problem unknown, and represents in this application example the target to recover) are also given in Table 1, column #1 for the sake of comparison.

As far as the approximated mean is concerned, the MCS approach yields a relative error $\epsilon_{\hat{m}} = 0.4\ \%$ with respect to the database mean value whereas relative errors of 0.7 % and 0.1 %

are associated to the 5-point and 10-point quadrature schemes respectively. The MCS estimate of the standard deviation is much closer ($\epsilon_{\hat{\sigma}} = 0.3$ %) to that of the database than those provided by both quadrature schemes ($\epsilon_{\hat{\sigma}} = 3.0$ % and 4.5 % respectively). A similar trend is observed for the higher order moments.

5 Conclusion

A method for solving a probabilistic inverse problem using polynomial chaos decomposition of unknown input parameters is proposed. The method is based on the use of the maximum likelihood estimation to identify unknown polynomial chaos coefficients. The optimization problem is solved using appropriate computational methods: the maximization of the likelihood is performed using a trust-region optimization algorithm called CONDOR. The evaluation of the conditional PDF of an observation is carried out using the Monte Carlo simulation method or a Gauss-Hermite quadrature scheme.

The proposed method is applied in order to determine the variability of the loading applied to a series of similar simply supported beams. Both Monte Carlo simulation and quadrature methods give accurate results on estimating the PDF of the applied loading. Consequently, this method seems to be an efficient alternative to a Bayesian framework when identifying unknown input random model parameter from response measurements, in the case when the database at hand is sufficiently large.

References

Abramowitz, M., Stegun, I. A. (Eds.), 1970. Handbook of mathematical functions. Dover Publications, Inc.

Berveiller, M., 2005. Stochastic finite elements : intrusive and non intrusive methods for reliability analysis. Ph.D. thesis, Université Blaise Pascal, Clermont-Ferrand.

Desceliers, C., Ghanem, R., Soize, C., 2006. Maximum likelihood estimation of stochastic chaos representations from experimental data. Int. J. Numer. Meth. Engng. 66, 978–1001.

Ghanem, R., Doostan, A., 2006. Characterization of stochastic system parameters from experimental data: a Bayesian inference approach. J. Comp. Phys. (217), 63–81.

Ghanem, R., Spanos, P., 1991. Stochastic finite elements - A spectral approach. Springer Verlag.

Sudret, B., Berveiller, M., Lemaire, M., 2006. A stochastic finite element procedure for moment and reliability analysis. Revue Européenne de Mécanique Numérique 15 (7-8), 1819–1835.

Tarantola, A., 2005. Inverse problem theory and methods for model parameter estimation. Society for Industrial and Applied Mathematics (SIAM).

Vanden Berghen, F., Bersini, H., 2004. CONDOR, a new parallel, constrained extension of powell's UOBYQA algorithm: experimental results and comparison with the DFO algorithm. J. of Comp. Appl. Math. 181, 157–175.

Wand, M., Jones, M., 1995. Kernel smoothing. Chapman and Hall.

Yuen, K.-V., Katafygiotis, L.-S., 2002. Bayesian modal updating using input complete and incomplete response measurements. J. of Eng. Mech. 128 (3), 340–350.