

# Proposition de communication

Nabil Rachdi, doctorant CIFRE EADS.

2 décembre 2009

## Contexte général de la thèse.

Ma thèse intitulée "*Modélisation des incertitudes par simulation*" a débuté en Septembre 2008 sous la direction de Jean-Claude Fort (Université Paris V), de Thierry Klein (Université Toulouse III) et également encadrée par Fabien Mangeant et Régis Lebrun (EADS IW Suresnes). Le problème général est la prise en compte de toute l'information disponible pour modéliser les incertitudes. En particulier, nous nous plaçons dans le cadre où l'on dispose d'un nombre limité de données expérimentales et d'un (ou plusieurs) modèle sensé représenter le phénomène qui nous intéresse. Ce cadre est très représentatif des Avant-Projets chez Airbus, notamment par l'utilisation de plateformes de simulation multi-échelles pour la conception virtuelle d'avions. Cette situation est présente dans l'ingénierie en général, où plusieurs modèles numériques co-existent pour une même simulation, ainsi que des données expérimentales sur les entrées et sorties de ces modèles, pas nécessairement appariées, coûteuses à obtenir.

On se pose la question de savoir comment la simulation peut contribuer à améliorer notre connaissance (données expérimentales) sur l'incertitude d'un phénomène donné.

## Résumé.

On modélise une grandeur physique, résultat d'un phénomène complexe, par une variable aléatoire  $Y \in \mathbb{R}$  de densité  $f$  inconnue. On dispose de résultats expérimentaux de ce phénomène,  $Y_1, \dots, Y_n$ , considérés comme des réalisations i.i.d. de la variable aléatoire  $Y$ . L'information disponible sur  $Y$  est insuffisante pour réaliser des tests statistiques ou des calculs de quantités d'intérêt significatifs.

L'approche retenue pour pallier ce problème consiste à prendre en compte un ou plusieurs modèles numériques de ce phénomène complexe. Pour cela, on considère des fonctions  $h^{(i)} : \mathbb{R}^{d_i} \times \mathbb{R}^{k_i} \rightarrow \mathbb{R}$  telles que  $Y \simeq h^{(i)}(\mathbf{X}_i, \boldsymbol{\theta}_i)$  où  $\mathbf{X}_i = (X_{i1}, \dots, X_{id_i})^T$  est un vecteur aléatoire de loi  $\mathbb{P}_{\mathbf{X}_i}$  et  $\boldsymbol{\theta}_i = (\theta_{i1}, \dots, \theta_{ik_i})^T$  est un vecteur de paramètres qu'on pourra supposer dans un compact  $\Theta_i \subset \mathbb{R}^{k_i}$ . Le vecteur  $\mathbf{X}_i$  représente les paramètres incertains du modèle  $h^{(i)}$ .

On notera  $f_{\boldsymbol{\theta}_i}$  la densité de probabilité de la variable aléatoire  $h^{(i)}(\mathbf{X}_i, \boldsymbol{\theta}_i)$ .

Par ailleurs, on suppose la donnée de  $N$  configurations  $\mathbf{X}_i^1, \dots, \mathbf{X}_i^N$  où  $\mathbf{X}_i^m, m = 1, \dots, N$ , seront supposées issues de la loi  $\mathbb{P}_{\mathbf{X}_i}$  (on prendra  $N$  grand devant  $n$  car la simulation est moins coûteuse que l'acquisition de données).

Pour simplifier, on omettra l'indexation "i" spécifiant le modèle et on note  $h(\mathbf{X}, \boldsymbol{\theta}) := h_{\boldsymbol{\theta}}(\mathbf{X})$ , puis  $f_{\boldsymbol{\theta}}$  la densité de probabilité de la variable  $h_{\boldsymbol{\theta}}(\mathbf{X})$ .

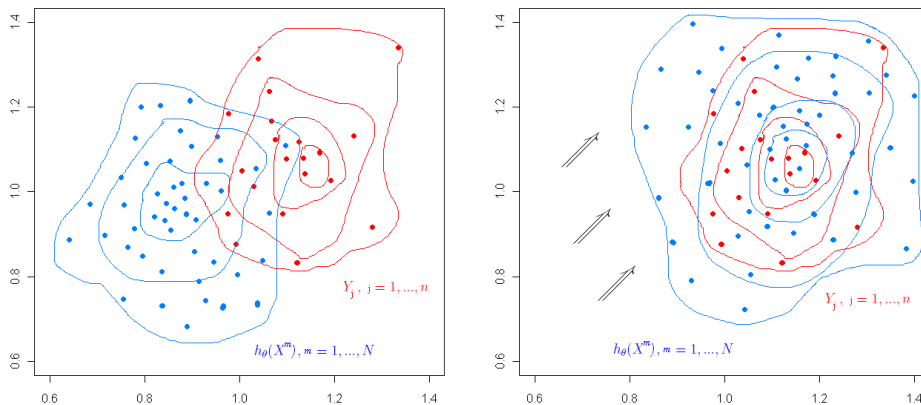


FIGURE 1 – Données expérimentales (en rouge) et données simulées (en bleu) : avant optimisation (à gauche), après optimisation (à droite)

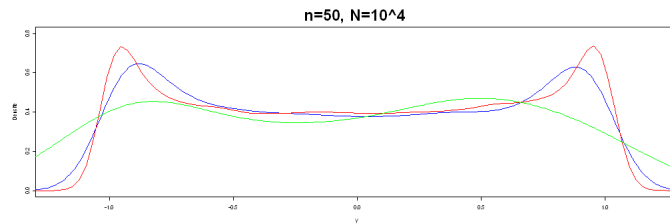


FIGURE 2 – Comparaison des densités de probabilité

On dispose ainsi de  $N$  points  $h_{\theta}(\mathbf{X}^1), \dots, h_{\theta}(\mathbf{X}^N)$  (données simulées sous  $\theta$ ).

Notre premier objectif est d'identifier un paramètre  $\theta \in \Theta$  tel que le modèle  $h_{\theta}$  sous l'incertitude  $\mathbf{X}$  ait un "comportement" proche de celui de  $Y$  au sens d'une distance sur les densités de  $Y$  et de  $h_{\theta}(\mathbf{X})$ . Nous discuterons également des autres choix permettant de mesurer une distance entre deux variables aléatoires, et justifierons le nôtre.

On espère pouvoir simuler des données qui soient fidèles aux données expérimentales. La difficulté réside dans le fait que ni  $f$  (densité de  $Y$ ) ni  $f_{\theta}$  (densité de  $h_{\theta}(\mathbf{X})$ ) ne sont connues analytiquement : on n'y a accès qu'à travers les échantillons  $(Y_j)_{j \in \{1, \dots, n\}}$  et  $(h_{\theta}(\mathbf{X}^m))_{m \in \{1, \dots, N\}}$ . On remplace donc  $f$  et  $f_{\theta}$  par des estimateurs non paramétriques (nous verrons qu'il suffira de ne remplacer que  $f_{\theta}$ ).

On établit alors un critère à minimiser où interviennent les données expérimentales  $(Y_j)_{j \in \{1, \dots, n\}}$  et les données  $(h_{\theta}(\mathbf{X}^m))_{m \in \{1, \dots, N\}}$ .

On construit une densité de probabilité  $f_{\hat{\theta}_{n, N}}^N$  sous laquelle on va pouvoir simuler un grand nombre de données qui vont permettre d'enrichir l'échantillon  $(Y_j)_{j \in \{1, \dots, n\}}$  d'un second échantillon  $(\tilde{Y}_s)_{s \in \{1, \dots, S\}}$ . La quantification de la loi de l'échantillon augmenté est alors beaucoup plus accessible.

La figure (2) compare les trois densités suivantes : celle de  $Y$  (inconnue, en rouge), celle reconstruite par noyaux uniquement à partir des données expérimentales (en vert), et celle reconstruite à noyaux à partir de l'échantillon augmenté (en bleu). On constate clairement la puissance de cette approche.

Nos premiers travaux ont tout d'abord consisté à formaliser le problème, puis à établir la consistance de notre procédure d'estimation à l'aide de la théorie des processus empiriques.

### Bibliographie.

L. Devroye, L. Györfi (1985). *Nonparametric Density Estimation*.

David W. Scott (1992). *Multivariate Density Estimation*.

Jun S. Liu (2001). *Monte Carlo Strategies in Scientific Computing*.

Mattei, P. et Druot, T. (2005). *General description of odip platform*. Rapport technique, Airbus. Ref FM0200016.

Sara van de Geer (2000). *Empirical Processes in M-Estimation*.

A.W van der Vaart (1998). *Asymptotics Statistics*.

A.W van der Vaart, Jon A. Wellner (1996). *Weak Convergence and Empirical Processes*.