Échantillonnage préférentiel pour estimer la probabilité d'événements rares.

Yves AUFFRAY*, Pierre BARBILLON†, Jean-Michel MARIN‡

2 janvier 2010

Je (Pierre Barbillon) suis étudiant en troisième année de thèse à l'université Paris-sud 11 sous la direction de Jean-Michel Marin. Ma thèse concerne les méthodes d'interpolation par noyaux pour les fonctions boîte noire coûteuses. Nous nous intéressons aux aspects théoriques de l'interpolation par noyaux et à des applications dans le cadre de problèmes industriels faisant intervenir des fonctions boîte noire coûteuses. Ici nous présentons avec Yves Auffray une méthodologie pour traiter un problème d'événement rare dépendant d'une fonction boîte noire coûteuse.

Soit

- X, une variable aléatoire à valeurs dans $\mathbb{R}^d$ de loi connue, simulable
- g sa densité par rapport à la mesure de Lebesgue.
- $f : \mathbb{R}^d \to \mathbb{R}_+$ une fonction "boîte noire"
- $R = \{x \in \mathbb{R}^d : f(x) \leq \rho\}$ où $\rho > 0$ est fixé.

$\{X \in R\}$ est un événement redouté et rare, dont on souhaite borner supérieurement la probabilité p avec un bon niveau de confiance.
Compte tenu de son coût d'évaluation, on 'dispose d'un "budget" maximal de N appels à f.

Pour traiter ce problème une méthode Monte Carlo standard consiste à produire un N-échantillon $(X_1, \ldots, X_N)$ de X et à proposer l'estimateur :

$$\hat{p}^{MC} = \frac{1}{N} \sum_{i=1}^{N} \mathbb{I}_R(X_i)\,.$$

Si $p << 1/N$, $\hat{p}^{MC}$ sera nul, très probablement et l'intervalle de confiance associé à l'estimateur, de forme $[0, p^+]$ pourra ne pas être assez précis.
Par exemple, si $p = 10^{-5}$ et $N = 100$, l'estimateur $\hat{p}^{MC}$ vaut 0 avec probabilité $(1-p)^N = 0.999$.
L'intervalle de confiance à 99% est $[0; 1 - (1/100))^{1/N}] = [0; 0.045]$.

Pour remédier à cela nous proposons de modifier le schéma précédent en y incluant une méthode d'échantillonnage préférentiel (Kahn and Marshall, 1953).
Cela consiste :

- à considérer la loi de probabilité sur $\mathbb{R}^d$ définie par le biais d'une "densité instrumentale" h
- produire un n échantillon $(Z_1, \ldots, Z_n)$ selon cette loi
- à proposer l'estimateur :

$$\hat{p}^{IS} = \sum_{i=1}^{n} \mathbb{I}_R(Z_i) \frac{g(Z_i)}{h(Z_i)}\,.$$

La densité instrumentale idéale serait

$$x \mapsto \mathbb{I}_R(x) g(x) \mathbb{P}(X \in R)^{-1}\,,$$

puisqu'elle permettrait d'obtenir un estimateur de variance nulle. Elle est évidemment inaccessible car elle dépend de $\mathbb{P}(X \in R)$ inconnue.
Cependant, on peut viser une densité "proche" en se servant, à la place de f, d'une approximation

*Dassault Aviation, Université Paris
†Université Paris-Sud 11, INRIA Saclay
‡Université Montpellier II

par krigeage (Cressie, 1990) $\hat{f}$ d'évaluation quasi instantanée, à laquelle on consent à consacrer la moitié du budget d'appels N.
Cela étant fait, définissant $\widehat{R}_\lambda = \{x : \hat{f}(x) < \lambda\}$ pour un λ réel quelconque, nous proposons la loi instrumentale :

$$h_\lambda : x \mapsto \mathbb{1}_{\widehat{R}_\lambda}(x) g(x) \mathbb{P}(X \in \widehat{R}_\lambda)^{-1} \,.$$

qui est simulable, puisque, du fait du faible coût de calcul demandé par $\hat{f}$, $\mathbb{P}(X \in \widehat{R}_\lambda)$ est estimable très finement.
On peut donc simuler selon la loi associée à h_λ pour produire un $\frac{N}{2}$-échantillon $(Z_1, \ldots, Z_n)$ (on note $n = N/2$). L'estimateur s'écrit alors :

$$\hat{p}^{IS} = \frac{1}{n} \sum_{i=1}^{n} \frac{g(Z_i)}{h_\lambda(Z_i)} \mathbb{1}_R(Z_i) = \frac{\mathbb{P}(X \in \widehat{R}_\lambda)}{n} \sum_{i=1}^{n} \mathbb{1}_R(Z_i) \,.$$

Il a

- pour espérance $\mathbb{E}(\hat{p}^{IS}) = \mathbb{P}(X \in R \cap \widehat{R}_\lambda)$, donc pour biais $-P(X \in R - \widehat{R}_\lambda)$, nul si $R \subset \widehat{R}_\lambda$
- et pour variance $\mathrm{Var}(\hat{p}^{IS}) = \frac{2}{N} \mathbb{P}(X \in R \cap \widehat{R}_\lambda) \Big(\mathbb{P}(X \in \widehat{R}_\lambda) - \mathbb{P}(X \in R \cap \widehat{R}_\lambda) \Big)$, qui, comparée à celle de $\hat{p}^{MC}$

$$\mathrm{Var}(\hat{p}^{MC}) = \frac{1}{N} \mathbb{P}(X \in R) \, (1 - \mathbb{P}(X \in R))$$

peut être bien plus faible si $\mathbb{P}(X \in \widehat{R}_\lambda)$ est proche de $\mathbb{P}(X \in R)$.

Par ailleurs, l'inégalite de Hoeffding (Hoeffding, 1963) fournit la borne suivante avec une probabilité $1 - \delta$:

$$p - \hat{p}^{IS} \leq P(X \in R - \widehat{R}_\lambda) + \mathbb{P}(X \in \widehat{R}_\lambda) \sqrt{\frac{\ln(1/\delta)}{2n}} \,.$$

En cas de biais nul, l'écart entre p et $\hat{p}^{IS}$ peut être d'un ordre comparable à $\mathbb{P}(X \in \widehat{R}_\lambda)$: ainsi, à supposer que λ a été choisi de façon que $R \subset \widehat{R}_\lambda$ et $\mathbb{P}(X \in \widehat{R}_\lambda) \leq 10 \cdot p$, la configuration de l'exemple précédent (i.e $p = 10^{-5}$, $N = 100$ donc $n = 50$) donne, avec probabilité 99%,

$$p - \hat{p}^{IS} \leq 2.15 \cdot 10^{-5} \,.$$

Références

Cressie, N. (1990). The origins of kriging. *Mathematical Geology*, **22(2)** : 239-252.

Fang, K.-T., Li, R. and Sudjianto, A. (2005). Design and Modeling for Computer Experiments (Computer Science & Data Analysis). *Chapman & Hall/CRC.*

Fishman, G.S. (1996). Monte Carlo Concepts, Algorithms, and Applications. *New York : Springer-Verlag.*

Hoeffding, W. (1963). Probability inequalities for sums of bounded random variables. *Journal of the American Statistical Association*, **58**, 13-30.

Kahn, H., and Marshall, A. (1953). Methods of Reducing Sample Size in Monte Carlo Computations. *Journal of the Operations Research Society of America*, **1**, 263-278.