

Estimation du volume des ensembles d'excursion d'un processus Gaussien par Krigeage intrinsèque

Emmanuel Vazquez, Miguel Piera-Martinez

► **To cite this version:**

Emmanuel Vazquez, Miguel Piera-Martinez. Estimation du volume des ensembles d'excursion d'un processus Gaussien par Krigeage intrinsèque. Conférence Journée de Statistiques, Jun 2007, Angers, France. pp.CD-ROM Proceedings. hal-00256148

HAL Id: hal-00256148

<https://hal-supelec.archives-ouvertes.fr/hal-00256148>

Submitted on 15 Feb 2008

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.

ESTIMATION DU VOLUME DES ENSEMBLES D'EXCURSION D'UN PROCESSUS GAUSSIEN PAR KRIGEAGE INTRINSÈQUE

Emmanuel Vazquez & Miguel Piera-Martinez

Supélec, Département Signaux et Systèmes Électroniques, 91192 Gif-sur-Yvette, France

Résumé — Estimer une probabilité de défaillance d'un système entrée-sortie à partir d'observations ponctuelles de celui-ci est un problème important pour le monde industriel. Une approche possible consiste à modéliser le système comme la réalisation d'un processus aléatoire et de s'intéresser au volume d'excursion de ce processus. Soit ξ un processus aléatoire gaussien et $|A_u(\xi)|$ le volume de l'ensemble d'excursion de ξ au dessus d'un seuil u sous une mesure de probabilité μ . L'objectif de cet article est d'estimer le volume d'excursion à partir de n observations ponctuelles du processus. Nous proposons d'approximer le volume d'excursion de ξ par celui d'un prédicteur ξ_n obtenu par krigeage intrinsèque à partir des n observations. Nous étudions les propriétés de convergence de cette approximation. Dans un deuxième temps, nous proposons un algorithme visant à choisir les points d'observation pour accélérer la convergence de l'estimateur.

Abstract — The estimation of the volume of an excursion set is a highly relevant problem for the industrial world since it corresponds to the estimation of the failure probability of a system that is known only through sampled observations. Assume that a Gaussian process ξ is predicted from n pointwise observations by intrinsic Kriging and that the volume of the excursion set of ξ above a given threshold u is approximated by the volume of the predictor. The first part of this paper gives a bound on the convergence rate of the approximated volume. The second part describes an algorithm that constructs a sequence of points to yield a fast convergence of the approximation.

Mots-clés principaux : Plans d'expérience, Statistique des processus

Mots-clés secondaires : Quantiles, Qualité - fiabilité

1 Introduction

Considérons un processus aléatoire gaussien ξ à valeurs réelles défini sur un ensemble \mathbb{X} muni d'une probabilité μ . Définissons l'ensemble d'excursion de ξ au dessus d'un niveau u par

$$A_u(\xi) := \{x \in \mathbb{X} : \xi(x) \geq u\}. \quad (1)$$

Le volume $\mu(A_u(\xi))$ sera par la suite noté $|A_u(\xi)|$. En ingénierie, $|A_u(\xi)|$ peut correspondre à la probabilité de défaillance d'un système (la probabilité que la sortie du système dépasse une valeur critique lorsqu'il est soumis à des entrées aléatoires).

Notons ξ_n l'estimateur de ξ par krigeage intrinsèque (Matheron, 1973) à partir d'un ensemble de n observations discrètes $\xi(x_i)$, $1 \leq i \leq n$. La section 2 étudie la convergence de $|A_u(\xi_n)|$ vers $|A_u(\xi)|$. Dans la section 3, nous proposons un algorithme de construction itérative d'une suite

de points (x_i) qui minimise à chaque étape la moyenne quadratique de l'erreur d'approximation $|A_u(\xi)| - |A_u(\xi_n)|$ conditionnée par les variables aléatoires $\xi(x_i)$, $i \leq n$. L'objectif de cet algorithme est d'accélérer la convergence de $|A_u(\xi_n)|$ vers $|A_u(\xi)|$. La section 4 illustre cet algorithme.

2 Estimation du volume d'excursion par krigeage intrinsèque

Une première façon d'estimer le volume d'excursion $|A_u(\xi)|$ est d'utiliser un estimateur de type Monte Carlo, tel que par exemple

$$|A_u(\xi)|_l := \frac{1}{l} \sum_{i=1}^l \mathbf{1}_{\{\xi(X_i) \geq u\}} \xrightarrow{l} |A_u(\xi)| \quad \text{p.s.} \quad (2)$$

où les X_i sont des variables aléatoires indépendantes de loi μ (en pratique, il est préférable d'utiliser des estimateurs avec échantillonnage d'importance).

Dans cet article, nous proposons d'estimer $|A_u(\xi)|$ en utilisant un prédicteur ξ_n de ξ et en approximant $|A_u(\xi)|$ par $|A_u(\xi_n)|_l$, avec l grand. Ceci a un intérêt pratique lorsque l'observation du processus ξ est coûteuse. On s'attend en effet à ce que ξ_n converge suffisamment vite vers ξ (quand $n \rightarrow \infty$) pour que l'approximation proposée soit satisfaisante pour des n relativement petits. Dans les paragraphes suivants, nous nous intéresserons à la convergence de $|A_u(\xi_n)|$ vers $|A_u(\xi)|$.

Nous utilisons le krigeage intrinsèque (Matheron, 1973) pour obtenir un prédicteur linéaire de ξ à partir d'un ensemble fini d'observations $\xi(x_i)$, $1 \leq i \leq n$. Rappelons que le krigeage intrinsèque étend le concept de prédiction linéaire non biaisée d'un processus aléatoire au cas où la moyenne du processus est inconnue.

La théorie de l'approximation de fonctions (voir par exemple Wu and Schaback, 1993 ; Light and Wayne, 1998 ; Narcowich et al., 2003 ; Wendland, 2005) permet d'établir la vitesse de décroissance de la variance $\sigma_n^2(x)$ de l'erreur de prédiction par krigeage intrinsèque lorsque la densité des observations du processus au voisinage de x augmente. Plus précisément, si ξ est un processus stationnaire de covariance $k(h)$, $h \in \mathbb{R}^d$, \mathbb{X} un domaine borné de \mathbb{R}^d , et si la transformée de Fourier de $k(h)$ satisfait

$$c_1(1 + \|\omega\|_2^2)^{-\nu} \leq \tilde{k}(\omega) \leq c_2(1 + \|\omega\|_2^2)^{-\nu}.$$

avec $\nu > d/2$, alors

$$\|\sigma_n(\cdot)\|_\infty \leq Ch_n^{\nu-d/2}, \quad (3)$$

où $h_n = \sup_{y \in \mathbb{X}} \min_i \|y - x_i\|_2$ mesure la densité de (x_1, \dots, x_n) dans \mathbb{X} .

La proposition suivante permet d'établir un résultat similaire pour le processus seuillé au niveau u .

Proposition 1. *Soit $\xi_n(x)$ le prédicteur de ξ par krigeage intrinsèque construit à partir des observations $\xi(x_i)$, $i = 1, \dots, n$. Posons $\sigma_n(x) := \text{var}[\xi(x) - \xi_n(x)]^{1/2}$. Alors,*

$$\mathbb{E} \left[(\mathbf{1}_{\xi(x) \geq u} - \mathbf{1}_{\xi_n(x) \geq u})^2 \right] = O(\sigma_n(x) |\log(\sigma_n(x))|^{1/2}) \quad \text{quand } \sigma_n(x) \rightarrow 0.$$

Démonstration. Voir Vazquez and Piera-Martinez (2006). □

Par conséquent, pour un domaine \mathbb{X} borné :

$$\begin{aligned} \mathbb{E} \left[(|A_u(\xi)| - |A_u(\xi_n)|)^2 \right] &= \mathbb{E} \left[\left(\int_{\mathbb{X}} \mathbf{1}_{\xi(x) \geq u} - \mathbf{1}_{\xi_n(x) \geq u} d\mu \right)^2 \right] \\ &\leq \int_{\mathbb{X}} \mathbb{E} [(\mathbf{1}_{\xi(x) \geq u} - \mathbf{1}_{\xi_n(x) \geq u})^2] d\mu \\ &\leq C \|\sigma_n(\cdot)\|_{\infty} |\log \|\sigma_n(\cdot)\|_{\infty}|^{1/2} \end{aligned} \quad (4)$$

quand $n \rightarrow \infty$ et $\|\sigma_n(\cdot)\|_{\infty} \rightarrow 0$.

Ce résultat simple montre que la vitesse de convergence de $|A_u(\xi_n)|$ vers $|A_u(\xi)|$ est liée à la vitesse de convergence de ξ_n vers ξ , et donc à la régularité de la covariance, cf. (3). En choisissant les points x_i sur une grille régulière, on obtient $h_n = O(n^{-1/d})$, où d est la dimension de \mathbb{X} . Supposons que les points x_i remplissent \mathbb{X} selon une telle grille. Dans ce cas, la vitesse de convergence de $|A_u(\xi_n)|$ vers $|A_u(\xi)|$ sera plus rapide que celle de l'estimateur (2) dès que $\nu > 3d/2$ (voir Vazquez and Piera-Martinez, 2006).

3 Accélération de convergence

3.1 Contrôle de la convergence

Soit $f : \mathbb{X} \rightarrow \mathbb{R}$ une trajectoire de ξ , observée en une suite de points $(x_n)_{n \in \mathbb{N}}$. Cette section propose un algorithme de construction de la suite (x_n) en vue d'optimiser la décroissance de l'erreur d'approximation $|A_u(\xi)| - |A_u(\xi_n)|$ conditionnée par les observations $\xi(x_i) = f(x_i)$, $i = 1, \dots, n$. Nous proposons une stratégie de contrôle à un pas consistant à choisir itérativement les x_n tels que

$$x_n = \operatorname{argmin}_{x_n \in \mathbb{X}} \Upsilon_n(x_n) := \mathbb{E} [(|A_u(\xi)| - |A_u(\xi_n)|)^2 | Z_{n-1}], \quad (5)$$

où pour tout n , $Z_n = (\xi(x_1), \dots, \xi(x_n))$. Notons que $\Upsilon_n(x_n)$ peut se réécrire

$$\Upsilon_n(x_n) = \mathbb{E} [\mathbb{E} [(|A_u(\xi)| - |A_u(\xi_n)|)^2 | Z_n] | Z_{n-1}]. \quad (6)$$

La loi de $|A_u(\xi)|$ conditionnée par les observations est généralement inconnue (voir Adler, 2000, section 4.4) et par conséquent, $\mathbb{E} [(|A_u(\xi)| - |A_u(\xi_n)|)^2 | Z_n]$ ne peut pas être déterminé analytiquement. Une première solution serait d'utiliser une approximation Monte Carlo de (5), en simulant le processus ξ conditionnellement aux observations. Formellement, il s'agirait de résoudre

$$\begin{aligned} x_n = \operatorname{argmin}_{x_n \in \mathbb{X}} \Upsilon_{n,m}(x_n) := \\ \mathbb{E} \left[m^{-1} \sum_{i=1}^m (|A_u(\xi_n + \zeta_n^i)| - |A_u(\xi_n)|)^2 | Z_{n-1}, \{\zeta_n^i, i \leq m\} \right], \end{aligned} \quad (7)$$

où les processus aléatoires ζ_n^i sont m copies indépendantes de ξ conditionné par $Z_n = (0, \dots, 0)$. La mise en œuvre numérique du programme (7) nécessite par ailleurs d'estimer $|A_u(\cdot)|$ par

$|A_u(\cdot)|_l$. S'il est en principe possible de simuler les processus conditionnés ζ_n^i (voir Chilès and Delfiner, 1999, chap. 7), ces simulations sont néanmoins coûteuses (le nombre d'opérations est de l'ordre de $O(l^3)$ pour simuler ξ aux points x_1, \dots, x_l). Comme l doit être grand afin d'obtenir une estimation suffisamment précise du volume d'excursion, l'utilisation de simulations est à éviter.

Nous proposons l'approche alternative suivante. Tout d'abord, approchons $\mathbb{E} [(|A_u(\xi)| - |A_u(\xi_n)|)^2 \mid Z_n]$ par $\mathbb{E} [(|A_u(\xi)|_l - |A_u(\xi_n)|_l)^2 \mid Z_n, \{X_i, i \leq l\}]$, avec l suffisamment grand. D'après l'inégalité de Minkowski, nous obtenons la majoration

$$\begin{aligned} & \mathbb{E} [(|A_u(\xi)|_l - |A_u(\xi_n)|_l)^2 \mid Z_n, \{X_i, i \leq l\}]^{1/2} \\ & \leq \frac{1}{l} \sum_{i=1}^l \mathbb{E} [(\mathbf{1}_{\xi(X_i) > u} - \mathbf{1}_{\xi_n(X_i) > u})^2 \mid Z_n, \{X_i, i \leq l\}]^{1/2}. \end{aligned} \quad (8)$$

3.2 Algorithme proposé

Soit $S = \{y_1, \dots, y_l\}$ un ensemble de l points de \mathbb{X} générés indépendamment selon la loi μ . Étant donné une suite finie $(x_i)_{1 \leq i \leq n-1}$ de points d'observation, cherchons le point d'observation x_n qui minimise la borne de l'erreur d'approximation du volume obtenue dans (8) :

$$x_n = \operatorname{argmin}_{x_n \in S} \Upsilon'_n(x_n) := \frac{1}{l} \sum_{i=1}^l \mathbb{E} [(\mathbf{1}_{\xi(y_i) > u} - \mathbf{1}_{\xi_n(y_i) > u})^2 \mid B_{n-1}]^{1/2}, \quad (9)$$

où B_n est l'évènement $\{\xi(x_1) = f(x_1), \dots, \xi(x_n) = f(x_n)\}$, $n > 0$.

Quelques étapes sont encore nécessaires pour mettre en œuvre numériquement le programme (9). Notons d'abord que

$$\begin{aligned} & \mathbb{E} [(\mathbf{1}_{\xi(y_i) > u} - \mathbf{1}_{\xi_n(y_i) > u})^2 \mid B_{n-1}] \\ & = \int_{z \in \mathbb{R}} \mathbb{E} [(\mathbf{1}_{\xi(y_i) > u} - \mathbf{1}_{\xi_n(y_i) > u})^2 \mid \xi(x_n) = z, B_{n-1}] \\ & \quad \times p_{\xi(x_n) \mid B_{n-1}}(z) dz, \quad \forall i \in \{1, \dots, l\}, \end{aligned} \quad (10)$$

où $p_{\xi(x) \mid B_{n-1}}$ est la densité de $\xi(x)$ conditionné par B_{n-1} . Le krigeage intrinsèque suppose que la moyenne de ξ est inconnue et par conséquent, pour $x \in \mathbb{X}$, $\mathbb{E} [(\mathbf{1}_{\xi(x) > u} - \mathbf{1}_{\xi_n(x) > u})^2 \mid \xi(x_n) = z, B_{n-1}]$ ne peut pas être déterminé exactement. Cependant, on peut écrire l'approximation suivante (voir Vazquez and Piera-Martinez, 2006) :

$$\mathbb{E} [(\mathbf{1}_{\xi(x) > u} - \mathbf{1}_{\xi_n(x) > u})^2 \mid \xi_n(x)] \approx v_n(x) := \Psi \left(\left| \frac{u - \xi_n(x)}{\sigma_n(x)} \right| \right), \quad (11)$$

où Ψ est la queue de distribution gaussienne.

Ensuite, soit l'opérateur de quantification défini par

$$\forall h \in \mathbb{R}, \quad \Delta_Q h = z_1 + \sum_{i=2}^Q (z_i - z_{i-1}) \mathbf{1}_{]z_i, +\infty[}(h)$$

avec $z_1 < z_2 < \dots < z_Q$. On peut alors réécrire (9) sous la forme :

$$x_n = \operatorname{argmin}_{x_n \in S} \Upsilon_n''(x_n) := \frac{1}{l} \sum_{i=1}^l \left(\sum_{j=1}^Q \mathbb{P}\{\Delta_Q \xi(x_n) = z_j \mid B_{n-1}\} \mathbb{E} [v_n(y_i) \mid \xi(x_n) = z_j, B_{n-1}] \right)^{1/2}. \quad (12)$$

Le programme (12) est directement transposable en langage informatique.

4 Exemple

Cette section illustre l'algorithme proposé (voir la figure 1). Considérons une fonction f définie sur \mathbb{R} et prenons $\mu = \mathcal{N}(0, \sigma^2)$. Modélisons f comme une trajectoire de ξ et cherchons à déterminer le volume d'excursion $|A_u(f)|$. Dans cet exemple, la covariance de ξ est estimée à chaque étape par maximum de vraisemblance (voir par exemple Stein, 1999). Après quelques itérations, on constate que la fonction f a été échantillonnée de façon à déterminer avec une bonne précision l'ensemble d'excursion de f au dessus du seuil choisi, dans la région où la densité de probabilité est élevée.

Références

- R. J. Adler. On excursion sets, tube formulas and maxima of random fields. *Ann. Appl. Probab.*, 10(1) :1–74, 2000.
- J.-P. Chilès and P. Delfiner. *Geostatistics : Modeling Spatial Uncertainty*. Wiley, New York, 1999.
- W. Light and H. Wayne. On power functions and error estimates for radial basis functions interpolation. *J. Approx. Theory*, 92(2) :245–266, 1998.
- G. Matheron. The intrinsic random functions, and their applications. *Adv. Appl. Prob.*, 5 : 439–468, 1973.
- F. J. Narcowich, J. D. Ward, and H. Wendland. Refined error estimates for radial basis function interpolation. *Constr. Approx.*, 19(4) :541–564, 2003.
- M. L. Stein. *Interpolation of Spatial Data : Some Theory for Kriging*. Springer, New York, 1999.
- E. Vazquez and M. Piera-Martinez. Estimation of the volume of an excursion set of a Gaussian process using intrinsic kriging, 2006. URL <http://arxiv.org/abs/math.ST/0611273>.
- H. Wendland. *Scattered Data Approximation*. Monographs on Applied and Computational Mathematics. Cambridge Univ. Press, Cambridge, 2005.
- Z. Wu and R. Schaback. Local error estimates for radial basis function interpolation of scattered data. *IMA J. Numer. Anal.*, 13 :13–27, 1993.

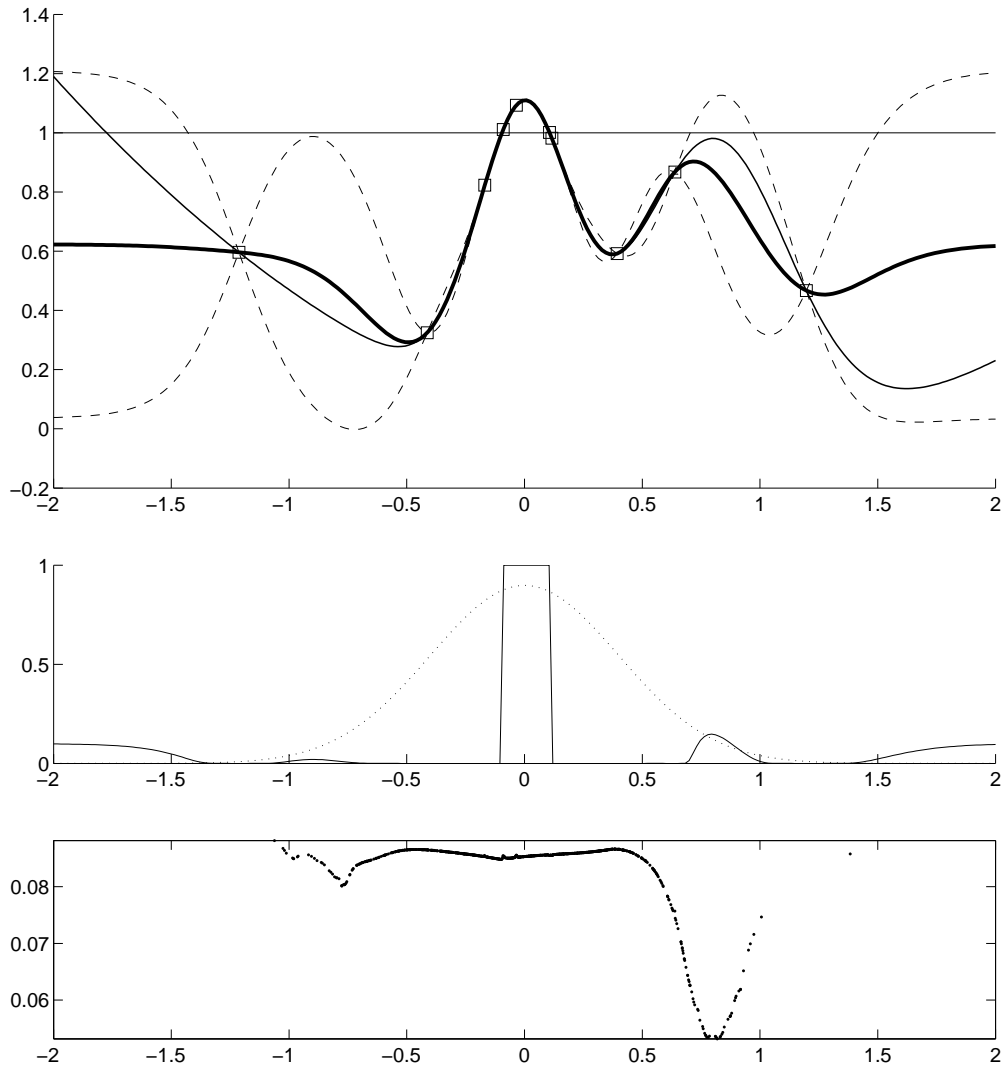


FIG. 1. Haut : Seuil u (ligne horizontale), fonction f (en trait fin), $n=10$ évaluations de f construites d'après l'algorithme proposé en utilisant $l = 800$ et $Q = 20$ (carrés), approximation f_n par krigeage intrinsèque (en trait épais), intervalles de confiance à 95% calculés en utilisant la variance de l'erreur de prédiction du krigeage intrinsèque (en trait interrompu). Milieu : Probabilité d'excursion $E[\mathbf{1}_{\xi(x) > u} | B_n]$ (en trait plein), densité de probabilité de μ (en trait pointillé). Bas : Graphe de $\Upsilon_n''(y_i)$, $i = 1, \dots, l = 800$. Le minimum de ce graphe indique la position de la prochaine évaluation de f (à environ 0.75).