Évaluation d'un risque d'inondation fluviale par planification séquentielle d'expériences

Aurélie Arnaud, Julien Bect, Mathieu Couplet, Alberto Pasanisi, Emmanuel Vazquez

▶ **To cite this version:**

Aurélie Arnaud, Julien Bect, Mathieu Couplet, Alberto Pasanisi, Emmanuel Vazquez. Évaluation d'un risque d'inondation fluviale par planification séquentielle d'expériences. 42èmes Journées de Statistique, 2010, Marseille, France, France. inria-00494767

HAL Id: inria-00494767
https://inria.hal.science/inria-00494767

Submitted on 24 Jun 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.

Évaluation d'un risque d'inondation fluviale par planification séquentielle d'expériences

Aurélie Arnaud[1], Julien Bect[2], Mathieu Couplet[1], Alberto Pasanisi[1]
et Emmanuel Vazquez[2,⋆]

1. *EDF R&D, Dépt. Management des Risques Industriels, 78401, Chatou, France*
2. *SUPELEC, Dépt. Signaux & Systèmes Électroniques, 91192 Gif-sur-Yvette, France*

Résumé : Nous nous intéressons au risque d'inondation d'une zone habitable ou industrielle, située à proximité d'un fleuve. Le risque est évalué à partir d'un modèle de la ligne d'eau du fleuve en présence d'incertitudes sur le débit et les caractéristiques du lit fluvial. Comme l'évaluation du modèle de la hauteur d'eau, pour un débit et des caractéristiques du lit fixés, est potentiellement coûteux en temps de calcul, l'estimation d'une probabilité de dépassement de seuil ou d'un quantile de la hauteur d'eau doit en pratique être conduite avec un budget réduit de simulations. Dans cet article, nous nous intéressons spécifiquement à l'estimation d'un quantile et nous proposons une méthode de planification d'expériences séquentielle qui construit une approximation du modèle par krigeage en choisissant les points d'évaluation du modèle de manière à réduire la variance d'estimation du quantile.

Abstract : The risk of river flooding in an inhabitable or industrial area is usually assessed by modeling the water-surface profile of the river, subject to uncertainties on the river discharge and the features of the riverbed. Because a single evaluation of such a model for known discharge and riverbed features is potentially time-consuming, the estimation of a probability of flooding must be achieved with a small budget of simulations. In this paper, we focus on the estimation of a water-level quantile. We propose a sequential Bayesian algorithm that selects relevant simulations to reduce the variance of estimation of the quantile.

1 Introduction

Cette étude concerne l'estimation du risque d'inondation d'une zone habitable ou industrielle, située à proximité d'un fleuve. Soit f la fonction à valeurs réelles, représentant la hauteur de l'eau du fleuve en un point donné, et dont l'argument est un vecteur de facteurs à valeurs dans $\mathbb{X} \subseteq \mathbb{R}^d$. Ces facteurs sont les grandeurs (physiques, morphologiques, etc.) susceptibles d'avoir une influence sur la hauteur d'eau observée. L'ensemble $\mathbb{X}$ est supposé muni d'une mesure de probabilité $\mathsf{P}_{\mathbb{X}}$, qui modélise le fait que les facteurs varient au cours du temps (on peut penser par exemple au débit du fleuve) ou qu'ils sont mal connus (par exemple, les caractéristiques du lit du fleuve). Nous nous intéressons à l'estimation du quantile $q_\alpha(f) = \inf\{u \in \mathbb{R};\, \mathsf{P}_{\mathbb{X}}\{f \leq u\} \geq \alpha\}$, pour une probabilité α donnée et proche de 1. En pratique, la connaissance d'un tel quantile permet de dimensionner la hauteur d'un ouvrage de protection.

La méthode standard pour estimer $q_\alpha(f)$ consiste à simuler un m-échantillon $X_1, \ldots, X_m$ selon la loi $\mathsf{P}_{\mathbb{X}}$, puis à considérer l'estimateur empirique

$$\widehat{q}_{\alpha,m}(f) = \min\left\{y; \frac{1}{m}\sum_{i=1}^{m} \mathbf{1}_{Y_i \leq y} \geq \alpha\right\} = Y_{\left(\lceil \alpha m \rceil\right)} \tag{1}$$

où $Y_i = f(X_i)$, $i = 1, \ldots, m$, et $Y_{(i)}$ désigne la statistique d'ordre de rang i de l'échantillon $Y_1, \ldots, Y_m$. Il est bien connu que $\sqrt{m}\big(\widehat{q}_{\alpha,m}(f) - q_\alpha(f)\big) \to_m \mathcal{N}\big(0, \sigma^2\big)$, avec $\sigma^2 = \frac{\alpha(1-\alpha)}{p_Y(q_\alpha(f))^2}$,

où p_Y est la densité de $Y = f(X)$, $X \sim \mathsf{P}_{\mathbb{X}}$ (voir par exemple [11]). Il est donc généralement nécessaire de simuler un échantillon de grande taille afin d'obtenir une estimation satisfaisante. Cependant, dans le cas où l'évaluation de la fonction f est coûteuse (par exemple, lorsqu'une évaluation du modèle en un point de l'espace des facteurs nécessite plusieurs heures de calcul), le budget d'évaluations de f sera très limité. La recherche d'estimateurs de quantile avec une faible variance constitue donc un enjeu important pour l'analyse de risque à partir de modèles coûteux. L'échantillonnage d'importance est l'idée la plus naturelle pour atteindre cet objectif [2, 4]. De plus, les techniques d'échantillonnage d'importance peuvent être sensiblement améliorées si l'on est capable de simuler facilement une variable aléatoire auxiliaire, disons Z, fortement corrélée avec Y [2,5]. Une telle variable aléatoire peut être obtenue en construisant une approximation $\widehat{f}$ de f et en posant $Z = \widehat{f}(X)$, $X \sim \mathsf{P}_{\mathbb{X}}$.

Dans cet article, nous proposons une approche fondée sur un algorithme bayésien de planification séquentielle d'expériences, inspiré des algorithmes bayésiens pour l'optimisation globale (voir par exemple [6,7,10]) et d'un algorithme pour estimer des probabilités de défaillance proposé dans [9]. On notera aussi que l'algorithme proposé possède des points communs avec [8]. Supposons que les points $X_i \overset{\text{i.i.d}}{\sim} \mathsf{P}_{\mathbb{X}}$, $i = 1, \ldots, m$ aient été générés mais qu'il n'est pas possible de calculer (1) en raison du coût d'évaluation de f. Notre objectif est de choisir séquentiellement des points d'évaluations $x_1, \ldots, x_n \in \{X_1, \ldots, X_m\}$ de f afin de construire un (méta-)estimateur $\tilde{q}_{\alpha,n}$ de $\widehat{q}_{\alpha,m}(f)$, consistant et rapidement convergent, de telle sorte que l'on puisse avoir $\tilde{q}_{\alpha,n}$ très proche de $\widehat{q}_{\alpha,m}(f)$ avec $n << m$. L'algorithme proposé est exposé dans la section 2. La section 3 fournit une évaluation partielle et empirique des performances de l'algorithme. Enfin la section 4 détaille le contexte applicatif et présente les résultats obtenus.

2 Algorithme séquentiel bayésien pour l'estimation de quantile

Dans [9], l'estimation d'une probabilité de défaillance est formulée comme un problème de planification séquentielle d'expériences dans un cadre bayésien, où l'information provenant des expériences effectuées à un instant est combinée à un a priori sur la fonction f, afin de choisir les expériences futures. Nous adoptons ici le même point de vue pour l'estimation d'un quantile. L'information a priori sur f est spécifiée sous la forme d'un processus aléatoire ξ dont la loi est choisie (ou estimée) par l'utilisateur. En général, on se restreint au cas des processus gaussiens, car il est possible dans ce cas d'écrire la loi a posteriori du processus après n évaluations de f en utilisant le krigeage (voir par exemple [3, 9, 10]).

Dans ce cadre, considérons l'estimateur $\tilde{q}_{\alpha,n} = \mathsf{E}[\,\widehat{q}_{\alpha,m}(\xi) \,|\, \mathcal{F}_n\,]$, où $\mathcal{F}_n$ désigne la σ-algèbre engendrée par les variables aléatoires $\xi(x_1), \ldots, \xi(x_n)$ et les points $X_1, \ldots, X_m$. En pratique, $\tilde{q}_{\alpha,n}$ peut être approché par l'estimateur $\tilde{q}'_{\alpha,n}$ construit de la manière suivante.

A-1 Pour $i = 1, \ldots, N$:

(a) Générer une trajectoire $f^{(n,i)}$ selon la loi de ξ conditionnée par $\xi(x_1), \ldots, \xi(x_n)$, évaluée aux points X_j.

(b) Calculer $q_\alpha^{(n,i)} = \widehat{q}_{\alpha,m}(f^{(n,i)})$, en utilisant l'échantillon $\left\{f^{(n,i)}(X_j)\right\}_{j=1,\ldots,m}$.

A-2 On obtient ainsi un échantillon $q_\alpha^{(n,1)}, \ldots, q_\alpha^{(n,N)}$ distribué selon la loi a posteriori de $\widehat{q}_{\alpha,m}(\xi)$. Définir $\tilde{q}'_{\alpha,n} = \frac{1}{N}\sum_{i=1}^{N} q_\alpha^{(n,i)}$.

Pour l'étape A-1.(a), la technique usuelle est celle du conditionnement par krigeage (pour plus de détails, voir par exemple [3,10]). Notons qu'évaluer $f^{(n,i)}$ en un grand nombre de points est généralement coûteux en temps de calcul. Ceci limite donc la valeur de m que l'on peut considérer en pratique.

Nous cherchons ensuite à réduire l'erreur d'estimation a posteriori de $\widehat{q}_{\alpha,m}(\xi)$ par $\tilde{q}_{\alpha,n}$ en choisissant les points d'évaluation de f (voir [9] pour des explications plus détaillées). Pour ce faire, nous adoptons une stratégie de planification à un pas, consistant à construire la suite $(x_n)_{n\geq 1}$ définie itérativement par

$$x_n = \underset{x\in\{X_1,\ldots,X_m\}}{\operatorname{argmin}} \Upsilon_n(x) := \mathsf{E}\left\{(\widehat{q}_{\alpha,m}(\xi) - \tilde{q}_{\alpha,n})^2 \mid \mathcal{F}_{n-1}\right\}, \tag{2}$$

où $\tilde{q}_{\alpha,n}$ est calculé à partir des observations $\xi(x_i)$, $i = 1, \ldots, n-1$ et de $\xi(x)$ qui n'a pas été observée. Notons que pour tout n, x_n est une fonction de $\xi(x_1), \ldots, \xi(x_{n-1})$. En pratique, le calcul du critère Υ_n en un point x se fait en deux étapes en remarquant que

$$\Upsilon_n(x) = \mathsf{E}\left\{ \mathsf{E}\left\{(\widehat{q}_{\alpha,m}(\xi) - \tilde{q}_{\alpha,n})^2 \mid \mathcal{F}_n\right\} \,\middle|\, \mathcal{F}_{n-1}\right\}.$$

Plus précisément, le calcul numérique de l'espérance conditionnelle intérieure peut se faire d'après la procédure suivante.

B-1 Faire l'étape A-1 ci-dessus en conditionnant les trajectoires par $\xi(x_1), \ldots, \xi(x_{n-1})$, et $\xi(x) = y$

B-2 Définir $\tilde{q}'_{\alpha,n}(x,y) = \frac{1}{N}\sum_{i=1}^{N} q_\alpha^{(n,i)}$ et $\Gamma_n(x,y) = \frac{1}{N-1}\sum_{i=1}^{N}(q_\alpha^{(n,i)} - \tilde{q}'_{\alpha,n}(x,y))^2$.

Le calcul numérique de l'espérance conditionnelle extérieure consiste à approcher l'intégrale $\int_{\mathbb{R}} \Gamma_n(x,y)\, p_{\xi(x)|\mathcal{F}_{n-1}}(y)dy$, où $p_{\xi(x)|\mathcal{F}_{n-1}}$ désigne la densité conditionnelle de $\xi(x)$ par rapport à $\mathcal{F}_{n-1}$. Ceci ne pose pas de problème en pratique.

3 Exemple illustratif

Cette section illustre le comportement de l'algorithme proposé lorsque f est une fonction d'une seule variable scalaire. Nous considérons l'expérience suivante. Nous simulons une trajectoire f d'un processus gaussien sur $\mathbb{X} = \mathbb{R}$, de moyenne nulle et avec une fonction de covariance stationnaire de Matérn (voir par exemple [10]) écrite sous la forme $k(h) = \frac{\sigma^2}{2^{\nu-1}\Gamma(\nu)} s\, \mathcal{K}_\nu(s)$, avec $s = 2\nu^{1/2}h/\rho$, $\sigma^2 = 1$, $\nu = 3$ et $\rho = 1/3$. Nous cherchons à estimer le quantile $q_\alpha(f)$, avec $\alpha = 0.97$, lorsque $\mathbb{X}$ est muni d'une probabilité uniforme sur $[-1, 1]$. Nous choisissons $m = 500$ points et $N = 200$. La figure 1 présente le comportement de l'algorithme de planification séquentielle après $n = 10$ itérations. Nous constatons que les points d'évaluation se concentrent dans les régions où les valeurs de f sont proches de q_α.

Pour compléter cet exemple, nous répétons cette expérience $K = 2000$ fois, et nous calculons les quantiles à 0.005 et 0.995 de l'erreur relative de l'estimateur proposé en fonction du nombre d'itérations. Les résultats sont reportés dans la table 1. On constate empiriquement que l'algorithme a un comportement satisfaisant.

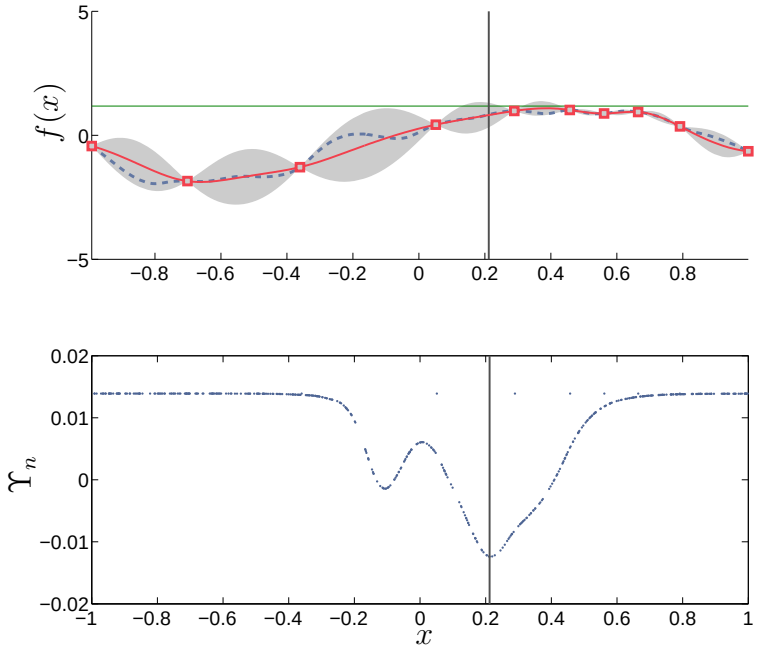

FIGURE 1: Haut : Fonction f (en trait interrompu), n=10 évaluations de f construites d'après l'algorithme proposé (carrés), approximation $\widehat{f}_n$ par krigeage (en trait continu), intervalles de confiance à 95% calculés en utilisant la variance de krigeage (grisé), estimation du quantile à 97% (ligne horizontale). Bas : Graphe de $\Upsilon_n(x_i)$, $i = 1, \ldots, m = 500$. Le minimum de ce graphe indique la position de la prochaine évaluation de f (à environ $x = 0.2$).

n	4	10	15	18	20
$[i^-, i^+]$	$[-12.9, 9.1]$	$[-0.92, 1.31]$	$[-0.12, 0.072]$	$[-2.210^{-2}, 2.3\,10^{-2}]$	$[-4.5\,10^{-3}, 7.0\,10^{-3}]$

TABLE 1: Intervalles empiriques $[i^-, i^+]$ à 99% de l'erreur relative $e = (\tilde{q}_{\alpha,n} - \widehat{q}_{\alpha,m}(\xi))/\widehat{q}_{\alpha,m}(\xi)$ en fonction de n.

4 Application industrielle

On s'intéresse à une portion de la Garonne, d'environ 50 km, comprise entre Tonneins (ville du Lot-et-Garonne, située en aval de la confluence avec le Lot) et La Réole (ville de Gironde, située à la limite de la zone d'influence hydrodynamique de la marée). Bien que cette portion ne présente pas d'installations industrielles importantes, elle se rapproche des configurations fluviales modélisées dans le cadre d'études à plus forts enjeux (notamment de protection d'installations nucléaires), et constitue donc un bon cas-test à la fois pour les études hydrauliques [1] et les analyses de risque.

Nous supposons ici que le problème est unidimensionnel. Plus particulièrement, la grandeur d'intérêt est la ligne d'eau, c'est-à-dire la relation, dépendante du temps, entre la hauteur d'eau et une abscisse curviligne. Le phénomène physique est régi par les équations de Saint-Venant qui lient la hauteur d'eau au débit, à la section mouillée, aux apports de débits latéraux, à la pente du tronçon et aux pertes de charges par frottement entre l'eau

et le lit fluvial. Nous considérons une modélisation dite « en lit composé », avec un débit de crue constant (régime permanent) et des apports latéraux nuls. Pour des raisons de simplicité, nous ne modélisons pas les zones d'expansion de la crue, une fois que la hauteur d'eau a dépassé la côte de la berge. Par conséquent, le modèle aura tendance à surestimer la hauteur d'eau. Les résultats obtenus ont uniquement une valeur d'exemple. Une section fluviale comporte ainsi deux zones : un lit mineur (zone principale d'écoulement) et un lit majeur (zone élargie qui est investie en présence de crues importantes). Ces deux zones sont caractérisées par des sols de nature différente et, par conséquent, par des rugosités différentes, exprimées classiquement par des coefficients de frottement de Strickler. Les coefficients de Strickler permettent d'évaluer, en fonction du débit et de la morphologie de la section fluviale, les pertes de charges dans les équations de Saint Venant. Ils ont la particularité de fournir une mesure décroissante du frottement : plus la valeur du coefficient de Strickler est faible, plus les pertes de charge seront élevées. Le calcul hydraulique est réalisé à l'aide du logiciel Mascaret développé par EDF-R&D et le CETMEF (Centre d'Etudes Techniques Maritimes et Fluviales) et disponible gratuitement.

Dans le cadre de cette étude, les grandeurs physiques supposées incertaines sont le débit, et les coefficients de frottement, affectés par une incertitude de type épistémique, due à un manque de connaissance. La morphologie du cours d'eau et les conditions limites sont considérées connues. La modélisation probabiliste du débit s'avère très facile en pratique car l'historique du débit du fleuve est bien connu. Le maximum annuel du débit peut être modélisé de manière très satisfaisante par une loi de Gumbel. La modélisation de l'incertitude des coefficients de Strickler est en revanche délicate. Ces derniers sont en effet des paramètres du modèle, mais ils ne sont pas directement observables. Il est possible de les estimer indirectement, à partir de couples hauteur-débit, relevés à différents endroits du fleuve. Comme ces estimations sont conduites dans un cadre bayésien, nous avons accès à la loi a posteriori des coefficients de Strickler. Nous choisissons ici de modéliser l'incertitude sur ces facteurs par leur loi a posteriori. Enfin, pour des raisons tenant à l'identifiabilité des coefficients de Strickler à partir des données disponibles, seuls deux coefficients seront considérés (un coefficient global pour le lit mineur du tronçon, et un autre pour le lit majeur). Notons alors que l'espace des facteurs $\mathbb{X}$ est de dimension 3. Nous testons le comportement de l'algorithme proposé pour $\alpha = 0.99$ et $m = 2000$ points. Les résultats sont présentés dans la figure 2 et paraissent satisfaisants.

En conclusion, la technique que nous proposons ici nous semble être intéressante dans le domaine de l'analyse de risque lorsque cette analyse est fondée sur l'évaluation de modèles informatiques coûteux et qui ne permettent pas l'utilisation d'estimateurs empiriques classiques.

Remerciements. Cette étude a été en partie financée par l'Agence Nationale de la Recherche française (ANR) dans le contexte du projet OPUS (réf. ANR-07-CIS7-010 et ANR-07-TLOG-015).

Références

[1] A. Besnard and N. Goutal. Comparison between 1D and 2D models for hydraulic modeling on a flood plain : Case of Garonne river. In *International Conference River Flow*, Cesme-Izmir (Turkey), 2008.

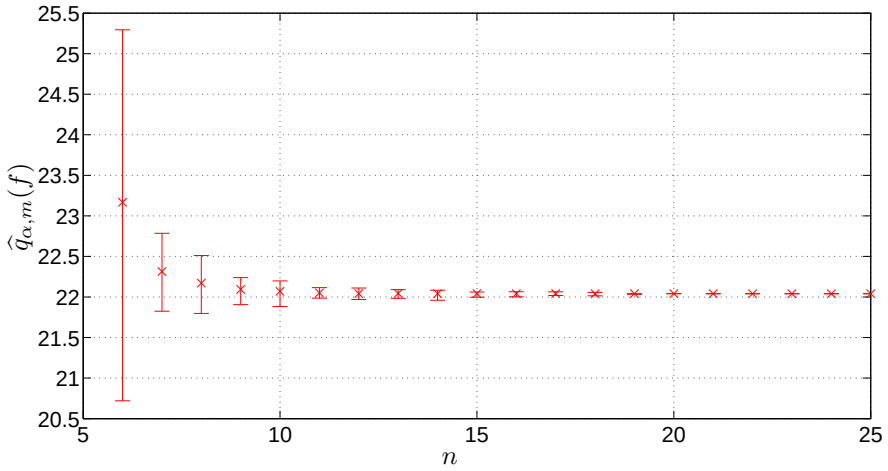

FIGURE 2: Estimation par $\tilde{q}_{\alpha,n}$ de $\widehat{q}_{\alpha,m}(f)$, avec $\alpha = 0.99$ et $m = 2000$, en fonction du nombre n d'évaluations de f. Les barres d'erreurs sont des intervalles a posteriori à 95%. On constate qu'après $n = 25$ évaluations, l'incertitude a posteriori sur $\widehat{q}_{\alpha,m}(f)$ est négligeable. On trouve ainsi $\tilde{q}_{\alpha,n=25} \approx 22.04\,\mathrm{m}$ au dessus du niveau de la mer. À titre indicatif, une estimation par bootstrap d'une région de confiance (fréquentiste) à 95% pour l'estimateur $\widehat{q}_{\alpha,m}(f)$ donne l'intervalle $[21.77, 22.40]$ (calculée à partir de l'approximation par krigeage après $n = 25$ évaluations).

[2] C. Cannamela, J. Garnier, and B. Iooss. Controlled stratification for quantile estimation. *Ann. Appl. Stat.*, 2(4) :1554–1580, 2008.

[3] J.-P. Chilès and P. Delfiner. *Geostatistics : Modeling Spatial Uncertainty*. Wiley, New York, 1999.

[4] P. W. Glynn. Importance sampling for Monte Carlo estimation of quantiles. In *Proceedings of the Second International Workshop on Mathematical Methods in Stochastic Simulation and Experimental Design*, pages 180–185, St. Petersburg, 1996.

[5] T. C. Hesterberg and B. L. Nelson. Control variates for probability and quantile estimation. *Management Science*, 44 :1295–1312, 1998.

[6] D. R. Jones, M. Schonlau, and W. J. Welch. Efficient global optimization of expensive black-box functions. *J. Global Optim.*, 13 :455–492, 1998.

[7] J. Mockus. *Bayesian Approach to Global Optimization : Theory and Applications*. Kluwer Acad. Publ., Dordrecht-Boston-London, 1989.

[8] J. Oakley. Estimating percentiles of uncertain computer code outputs. *J. Roy. Statist. Soc. Ser. C*, 53(1) :83–93, 2004.

[9] E. Vazquez and J. Bect. Sequential bayesian algorithm to estimate a probability of failure. In *15th IFAC Symposium on System Identification, SYSID09*, pages 546–550, Saint-Malo, France, July 6-8 2009.

[10] J. Villemonteix, E. Vazquez, and E. Walter. An informational approach to the global optimization of expensive-to-evaluate functions. *J. Global Optim.*, 44(4) :509–534, 2009.

[11] L. Wasserman. *All of nonparametric statistics*. Springer, 2006.