HAL
open science

Utilisation d'un estimateur non paramétrique de la densité conditionnelle pour l'étude de la dégradation spatiale du fil de contact caténaire

Jeanne Casaert, Denis Bosq, Rachid Ziani

To cite this version:

Jeanne Casaert, Denis Bosq, Rachid Ziani. Utilisation d'un estimateur non paramétrique de la densité conditionnelle pour l'étude de la dégradation spatiale du fil de contact caténaire. 41èmes Journées de Statistique, SFdS, Bordeaux, 2009, Bordeaux, France, France. inria-00386684

HAL Id: inria-00386684

https://inria.hal.science/inria-00386684v1

Submitted on 22 May 2009

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.

Utilisation d'un estimateur non paramétrique de la densité conditionnelle pour l'étude de la dégradation spatiale du fil de contact caténaire

Jeanne Casaert[1,2] & Denis Bosq[1] & Rachid Ziani[2]

[1] *Laboratoire de Statistique Théorique et Appliquée, Université Paris VI, 175 rue du Chevaleret, 75013 Paris*

[2] *SNCF, Direction de l'Innovation et de la Recherche, 45 rue de Londres, 75008 Paris*

Résumé : Dans le cadre de la maintenance préventive du fil de contact caténaire, nous cherchons à étudier la dégradation spatiale du fil de contact. En effet, des systèmes de mesure automatique du fil de contact permettent de connaître l'épaisseur du fil en plusieurs points espacés de façon régulière. Le pas de mesure correspondant à l'espacement entre les points de mesure est lié à la performance des systèmes de mesure. Plus les systèmes mesurent l'épaisseur avec un pas faible, plus la mise en oeuvre du système est coûteuse. Ainsi, afin d'optimiser l'efficacité de la maintenance préventive des fils, la SNCF souhaiterait optimiser ce pas de mesure et, si cela est possible, l'augmenter.
Pour répondre à cette problématique, nous cherchons, par des méthodes statistiques, à mieux connaître la dégradation spatiale du fil de contact et, en particulier, à être en mesure d'estimer un intervalle de confiance de l'épaisseur du fil entre deux points de mesures.
La méthode proposée ici pour estimer cet intervalle de confiance repose sur l'estimation non paramétrique de la densité conditionnelle d'un point non mesuré sachant un point connu proche.

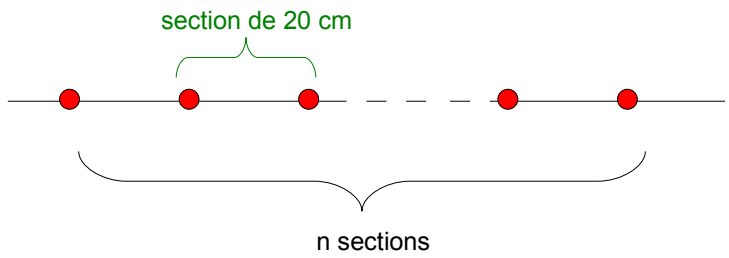
Mots clés : densité conditionelle, dégradation spatiale, maintenance

Abstract: The study deals with the catenary contact wire maintenance. We try to have a better knowledge of its spatial degradation. Thanks to regular inspections of the contact wire, its degradation can be observed. In order to improve the maintenance, the SNCF try to know how frequent the measures along the wire must be.
To answer to this problematic, we propose statistical methods to better know the spatial degradation of the contact wire and, more particularly, to estimate a confidence interval around the thickness of wire between two measures.
The approach developped here to estimate this confidence interval is based on the non parametric estimation of conditional density.

Keywords: contional density, spatial degradation, maintenance

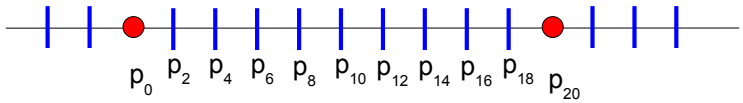
1 Introduction

Considérons un fil de contact tendu, de longueur approximative 1000 m.
Nous supposons que l'épaisseur de ce fil est observée tous les 20 cm et est inconnue ailleurs.
Nous pouvons donc représenter le fil découpé en sections de 20 cm :



Nous souhaitons prédire l'épaisseur du point central d'une section en connaissant l'épaisseur des deux extrêmités, ou encore, prédire l'épaisseur minimale d'une section (toujours en connaissant l'épaisseur des deux extrêmités).

Soit X_p l'épaisseur du point à la position p. Soit Y la variable à prédire, Y correspondra à l'épaisseur du milieu de la section ($X_{p_{10}}$) ou au minimum de la section ($\min_p X_p$) suivant les cas.



Nous cherchons donc à connaître Y sachant X_{p_0} et $X_{p_{20}}$.
La méthode repose sur l'estimation non paramétrique de la densité conditionnelle de Y sachant le couple X=(X_{p_0},$X_{p_{20}}$).

Supposons qu'il existe une relation entre Y et X :

$$Y = r(X)$$

Un article de Bosq (1989) permet de dire que dans ce cas (c'est à dire si Y est très corrélée à X) la densité conditionnelle $f^{Y|X}(.)$ 'explosera' au voisinage du support de r (où r est la fonction décrite ci-dessus) et s'évanouira ailleurs. Ainsi, une très forte corrélation entre les points est repérable à l'observation de la densité conditionnelle.
En observant le comportement de la densité conditionnelle, nous serons donc en mesure de dire si une relation entre Y et X existe et même d'estimer la fonction r par :

$$\hat{r}_n(x) = inf\left\{z : f_n(x,z) = max_{y\in\mathbb{R}} f_n^{X,Y}(x,y)\right\}$$

Nous proposons une amélioration des résultats de cet article et les appliquons à notre cas d'étude.

2 Estimateur de la densité et intervalle de confiance

On considère les processus réels (X_t) et (Y_t), $t \in \mathbb{Z}$ liés par la relation :

$$Y_t = r(X_t), t \in \mathbb{Z}$$

où r est borélienne

L'estimation de r sera basée sur la densité jointe du couple (X, Y).
Nous choisissons d'estimer la densité jointe par l'estimateur de Parzen-Rosenblatt :

$$f_n^{X,Y}(x,y) = \frac{1}{nh_n^2}\sum_{t=1}^{n} K_1\left(\frac{X_t - x}{h_n}\right) K_1\left(\frac{Y_t - y}{h_n}\right)$$

$$\text{avec } h_n \to 0 \text{ quand } n \to \infty$$

K_1 est un noyau sur $\mathbb{R}$ et nous définissons K sur $\mathbb{R}^2$ par :

$$\text{pour tout } u = (u_1, u_2) \in \mathbb{R}^2 \text{ , } K(u_1, u_2) = K_1(u_1)K_1(u_2)$$

On dira que K est régulier s'il vérifie :

1. K est une densité continue presque partout et bornée
2. $\exists\, m \geq 2 : \lim_{||u||\to\infty} ||u||^m K(u) = 0$
   où $||.||$ désigne la norme euclidienne sur $\mathbb{R}^2$
3. $\exists\, a$ et $b > 0 :\ K(u) > b$ pour $||u|| < a$
4. $\sup_{z\in\mathbb{R}} K(., z)$ est intégrable sur $\mathbb{R}$
5. K est symétrique : $K(u) = K(-u)$

Pour la suite de cette partie théorique, nous supposons que le processus $(X_t, t \in \mathbb{Z})$ est géométriquement fortement mélangeant.
Nous supposons également que X_t admet une densité f deux fois dérivable, bornée, strictement positive dans un voisinage de x.

On dira que $\hat{r}_n$ est un estimateur de r si $\hat{r}_n$ est une application mesurable vérifiant :

$$f_n^{X,Y}(x, \hat{r}_n(x)) = \max_{y \in \mathbb{R}} f_n^{X,Y}(x, y) \ , \ x \in \mathbb{R} \tag{1}$$

Le théorème suivant donne la convergence en probabilité de $\hat{r}_n$ vers r accompagnée d'un intervalle de confiance autour de $\hat{r}_n$:

Théorème. *Si r est lipschitzienne au voisinage de x et si (X_t) est géométriquement fortement mélangeant alors*

$$\frac{\sqrt{n}h_n}{\log(n)} \to \infty \implies \frac{1}{h_n^{1-\frac{2}{m}}} |\hat{r}_n(x) - r(x)| \overset{p.s.}{\to} 0$$

et plus précisément :

$$\forall \varepsilon > 0, \exists c > 0, \exists N_{\varepsilon,v} : \forall n \geq N_{\varepsilon,v}$$
$$\forall q \in [1, \frac{n}{2}]$$
$$\mathbb{P}\left[\frac{1}{h_n^{1-\frac{2}{m}}} |\hat{r}_n(x) - r(x)| \geq \varepsilon\right] \leq 2\exp\left\{-2c^2 h_n^2 q\right\} + 22\sqrt{1 + \frac{1}{ch_n}} q\alpha\left[\frac{n}{2q}\right]$$

c peut être explicité.

3 Application au fil de contact caténaire

Nous disposons des mesures de plusieurs fils de contact espacées tous les 2 cm. En considérant chaque fil séparément, ces données servent d'observations pour l'estimation de Y.
Notons :
n le nombre de sections de 20 cm sur le fil observé et j le numéro de la section, $j \in \{1...n\}$
$X_{p_0,j}$ l'épaisseur du point connu 'gauche' de la section j
$X_{p_{20},j}$ l' épaisseur du point connu 'droit' de la section j
$Y_j = X_{p_{10},j}$ l'épaisseur du point milieu de la section j

La démarche que nous employons est la suivante :

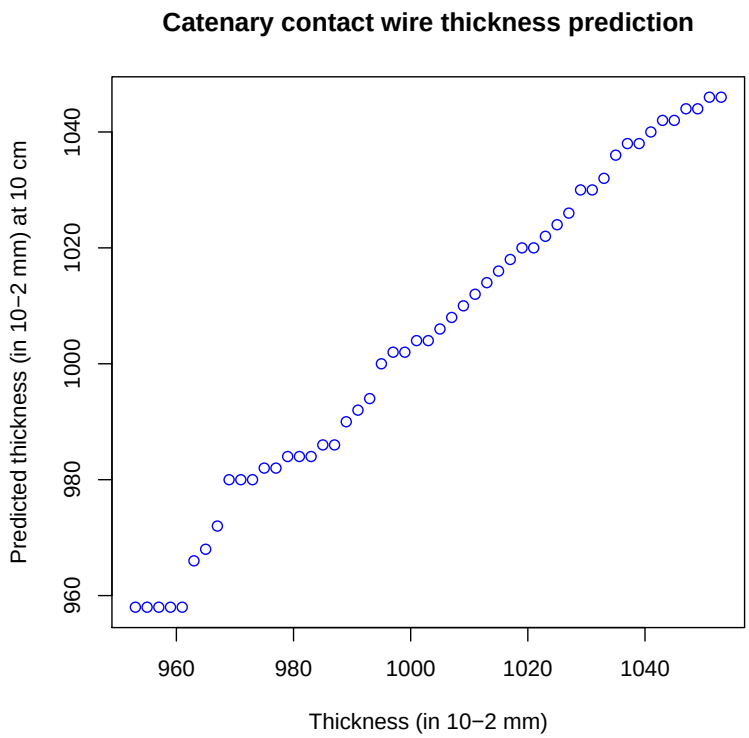
1. Observation de la densité estimée de la variable $X_{p_0,j}$ (de sorte à en connaître sa forme et surtout son intervalle de définition)

2. Estimation de $\hat{r}_n(x)$ pour toute valeur x réalisable

3. Recherche et optimisation numérique des constantes utilisées implicitement pour estimer l'intervalle de confiance

4. Application de l'intervalle de confiance tenant compte des paramètres propres au fil autour de l'estimation

Les résultats numériques sont en cours d'acquisition. Nous pouvons toutefois dors et déjà en présenter quelques-uns.
Le nombre d'observations varie entre 1200 et 2000 sur les différents fils (pour lesquels des mesures sont connues).
Le graphe ci-dessous représente $\hat{r}_n(x)$ en fonction de x (soit l'estimation de Y_j sachant $X_{p_0,j}$:



Cette illustration nous montre d'une part que l'estimation se fait correctement et que, d'autre part, deux points distants de 10 cm l'un de l'autre semblent avoir quasiment la même épaisseur.
Néanmoins, seul un intervalle de confiance pourra confirmer ce résultat : en effet, son amplitude indiquera la précision de cette estimation.
L'étape importante qu'il reste à mettre en place est donc le calcul numérique de cet intervalle de confiance. En particulier, nous cherchons à optimiser la valeur numérique

des constantes (qui ne sont pas données numériquement au cours de la démonstration mais sont exprimées en fonction de différentes contraintes à respecter).

Bibliographie

[1] Bosq, D (1989) Non parametric estimation of a non linear filter using a density estimator with a zero-one explosive behaviour in $\mathbb{R}^d$, *Statistics and Decisions*, 7, 229–241.
[2] Bosq, D (1998) Nonparametric Statitics for Stochastic Processes, *Lecture Notes in Statistics, Springer-Verlag*
[3] Doukhan, P. (1994) Mixing : Properties and Examples, *Lecture Notes in Statistics, Springer-Verlag*
[4] Parzen, E. (1962) On the estimation of probability density function and mode, *Ann. Math. Statist.*, 33, 1065–1076
[5] Rio, E. (2000) Théorie asymptotique des processus aléatoires faiblement dépendants, *Mathématiques et Applications, Springer*
[6] Rosenblatt, M. (1956) Remarks on some non parametric estimates of a density function, *Ann. Math. Statist.*, 27, 832–837