Available online at www.sciencedirect.com
ScienceDirect
ELSEVIER
C. R. Acad. Sci. Paris, Ser. I 344 (2007) 211–214
COMPTES RENDUS
MATHEMATIQUE
http://france.elsevier.com/direct/CRASS1/

Statistique/Probabilités

Estimation des quantiles d'une probabilité de franchissement de seuils avec observations dépendantes

Claire Pinçon [a], Odile Pons [b]

[a] *Université de Lille, biomathématiques, 3, rue du Professeur Laguesse, B.P. 83, 59006 Lille cedex, France*
[b] *INRA, mathématiques et informatique appliquée, 78352 Jouy-en-Josas cedex, France*

Reçu le 2 octobre 2006 ; accepté après révision le 13 décembre 2006

Présenté par Paul Deheuvels

Résumé

Dans un modèle de régression non paramétrique avec fonction de régression monotone et erreur, la probabilité de franchissement d'un seuil pour la variable réponse est définie en fonction d'un seuil sur la variable explicative. Des estimateurs de cette fonction de probabilité et de ses quantiles sont définis et leurs propriétés asymptotiques sont établies pour des observations dépendantes. ***Pour citer cet article : C. Pinçon, O. Pons, C. R. Acad. Sci. Paris, Ser. I 344 (2007).***

Abstract

Nonparametric estimation of the quantiles for a probability of threshold crossing with dependent data. In a nonparametric regression model with a monotone regression function and an error, we define a conditional probability of threshold crossing and its quantiles, and present the asymptotic properties of their estimators for dependent observations. ***To cite this article: C. Pinçon, O. Pons, C. R. Acad. Sci. Paris, Ser. I 344 (2007).***

1. Modèle et conditions

Soit $(X,Y)=(X_t,Y_t)_{t\in\mathbb{R}_+}$ un processus sur un espace de probabilité (Ω, A, P), à valeurs dans un espace métrique séparable complet $(\mathbb{X}\times\mathbb{Y}, B)$, dont la loi est définie par $Y=m(X)+\varepsilon$, $E\varepsilon=0$ où $\varepsilon=(\varepsilon_t)_{t\geqslant 0}$ est indépendente de X, de fonction de répartition homogène $\varepsilon_t\sim F$ et à densité symétrique ; sur un intervalle fini $I_X=[a,b]$ du support de X, m est une fonction continue monotone, supposée croissante. Seuls X et une indicatrice $\delta=I\{Y\leqslant\mu_Y\}$ correspondant au franchissement d'un seuil inconnu μ_Y du processus Y sont observés, on note μ_X le seuil qui lui correspond pour X. Pour $i=1,\dots,n$, un échantillonnage de n trajectoires indépendantes est effectué en des instants aléatoires $t_{i1},\dots,t_{iJ_i}$ d'observation de (X,δ), où J_i est aléatoire, les observations $X_{i1},\dots,X_{iJ_i},Y_{i1},\dots,Y_{iJ_i}$ sont dépendantes pour tout i et

$$p(x)=\Pr(\delta_t=1|X_t=x)=\Pr\bigl(\varepsilon_t\leqslant\mu_Y-m(x)\bigr)=F\bigl(\mu_Y-m(x)\bigr),\qquad \mu_X=p^{-1}(1/2):=q(1/2).$$

Adresses e-mail : claire.pincon@univ-lille2.fr (C. Pinçon), odile.pons@jouy.inra.fr, pons.odile@free.fr (O. Pons).

1631-073X/$ – see front matter
doi:10.1016/j.crma.2006.12.006

Soient

$$\hat{p}_{N,h}(x) = \frac{\sum_{i=1}^{n}\sum_{j=1}^{J_i} K_h(x - X_{ij})\delta_{ij}}{\sum_{i=1}^{n}\sum_{j=1}^{J_i} K_h(x - X_{ij})}, \qquad \hat{q}_{N,h}(u) = \sup\{x \in I_X\colon\ \hat{p}_{N,h}(x) < u\},$$

où $N = N_n = \sum_{i=1}^{n} J_i$ et $K_h(x) = h^{-1}K(h^{-1}x)$ avec $h = h_{n,x} = h_n(x)$ une fonction bornée de norme uniforme $\|h\|$. Les convergences sont étudiées avec une fenêtre de vitesse optimale pour l'erreur quadratique moyenne des estimateurs (voir Hall [2], Härdle [3], et Sheather and Marron [5]).

Les conditions sont les suivantes :

(K_1) K est positive, symétrique, à support $[-1; 1]$;
(K_2) $\int K(v)\,\mathrm{d}v = 1$, $\kappa = \int v^2 K(v)\,\mathrm{d}v < \infty$, $\kappa_r = \int K^r(v)\,\mathrm{d}v < \infty$ et $\int |K'(v)|^r\,\mathrm{d}v < \infty$ pour $r > 0$; pour tout n, la fonction $h = h_n$ admet une dérivée bornée sur $I_{X,h} = [a + \|h\|, b - \|h\|]$;
(K_3) pour $i = 1, \ldots, n$, $J_i < \infty$ p.s. ; quand $n \to \infty$, N et $J^{(k)}(n) = \sum_{i=1}^{n} J_i(J_i - 1)\cdots(J_i - k + 1)$ satisfont p.s. : $N \to \infty$ avec $N/n < \infty$ et $J^{(k)}(n)/N < \infty$, $\|h\| \to 0$, $N\|h\|$ et $N\|h\|^3 \to \infty$, $\|Nh^5 - N_0\| \to 0$, avec $\|N_0\| < \infty$.

Soient $f_{X_{ij}}$ la densité de X_{ij}, $f_{(X_{ij_1},\ldots,X_{ij_k})}$ celle de $X_{ij_1}, \ldots, X_{ij_k}$, $I_{X,h} = \{x; [a + \|h\|; b - \|h\|] \subset I_X\}$.

(D_1) La fonction p est $C^3(I_X)$ et strictement monotone sur I_X ;
(D_2) $EN^{-1}\sum_{i=1}^{n}\sum_{j=1}^{J_i} f_{X_{ij}} \to s^{(0)} > 0$ fonction de $C^3(I_X)$;
(D_3) Les fonctions $EN^{-1}\sum_{i=1}^{n}\sum_{j=1}^{J_i} f_{X_{ij}}(x_1)f_{X_{ij}}(x_2)$ et, pour $\gamma > 0$, $EN^{-1}\sum_{i=1}^{n}\sum_{j=1}^{J_i}(f_{X_{ij}})^{\gamma}$ convergent sur I_X^2 et I_X respectivement ;
(D_4) Pour $k \leqslant \max_i J_i$, $EJ^{(k)}(n)^{-1}\sum_{i=1}^{n}\sum_{j_1=1}^{J_i}\cdots\sum_{j_k \neq j_1\ldots j_{k-1}} f_{X_{ij_1},\ldots,X_{ij_k}}$ et $EJ^{(k)}(n)^{-1}\sum_{i=1}^{n}\sum_{j_1=1}^{J_i}\cdots\sum_{j_k=1}^{J_i} f_{X_{ij_1}}\cdots f_{X_{ij_k}}$ convergent sur $I_X^{\otimes k}$.

2. Convergence des estimateurs

Soient $\hat{n}_{N,h} = N^{-1}\sum_{i=1}^{n}\sum_{j=1}^{J_i} K_h(x - X_{ij})\delta_{ij}$, $\hat{d}_{N,h} = N^{-1}\sum_{i=1}^{n}\sum_{j=1}^{J_i} K_h(x - X_{ij})$ et, avec E_N l'espérance conditionnelle à N,

$$\begin{aligned}
&n_{N,h} = E_N\{\hat{n}_{N,h}\}, \qquad d_{N,h} = E_N\{\hat{d}_{N,h}\},\\
&\hat{n}^*_{N,h} = \hat{n}_{N,h} - n_{N,h}, \qquad \hat{d}^*_{N,h} = \hat{d}_{N,h} - d_{N,h},\\
&p_{N,h} = E_N\{\hat{p}_{N,h}\}, \qquad \hat{p}^*_{N,h} = \bigl(f^{(0)}\bigr)^{-1}(\hat{n}^*_{N,h} - p\hat{d}^*_{N,h}).
\end{aligned}$$

Les o et O des convergences sont uniformes.

En généralisant Nadaraya [4] :

Lemme 2.1. *Sur* $S_{X,h} = [a + \|h\|; b - \|h\|]$,

(a) $n_{N,h} = pf^{(0)} + \frac{h^2}{2}\kappa\{pf^{(0)}\}^{(2)} + \mathrm{o}(h^2)$, $\mathrm{var}_N\{\hat{n}_{N,h}\} = N^{-1}h^{-1}\kappa_2 pf^{(0)} + \mathrm{o}(N^{-1}h^{-1})$, $E_N|\hat{n}_h - n_{N,h}|^{\gamma} = \mathrm{O}((Nh)^{-(\gamma-1)})$, $\gamma > 2$.

(b) $d_{N,h} = f^{(0)} + \frac{h^2}{2}\kappa f^{(2)} + \mathrm{o}(h^2)$, $\mathrm{var}_N\{\hat{d}_h\} = N^{-1}h^{-1}\kappa_2 f^{(0)} + \mathrm{o}(N^{-1}h^{-1})$, $E_N|\hat{d}_h - d_{N,h}|^{\gamma} = \mathrm{O}((Nh)^{-(\gamma-1)})$, $\gamma > 2$.

(c) $\mathrm{cov}_N\{\hat{d}_h, \hat{n}_h\} = N^{-1}h^{-1}\kappa_2 pf^{(0)} + \mathrm{o}(N^{-1}h^{-1})$.

Proposition 2.2. *Sur* $S_{X,h}$,

(a) $E_N|\hat{p}_{N,h}(x) - p(x)| \to 0$, $p_{N,h} = p + h^2B + \mathrm{o}(h^2)$, $B = \kappa\{p^{(1)}f^{(1)}\{f^{(0)}\}^{-1} + \frac{1}{2}p^{(2)}\}$, $\mathrm{var}_N\{\hat{p}_{N,h}\} = N^{-1}h^{-1}\sigma^2 + \mathrm{o}(N^{-1}h^{-1})$, $\sigma^2 = p(1-p)\kappa_2\{f^{(0)}\}^{-1}$, $E_N|\hat{p}_{N,h} - p|^{\gamma} = \mathrm{O}(h^{2\gamma})$, $\gamma > 2$;

(b) $(Nh)^{1/2}\{\hat{p}_{N,h} - p_{N,h}\} = (Nh)^{1/2}\{f^{(0)}\}^{-1}[\{\hat{n}_{N,h} - n_{N,h}\} - p\{\hat{d}_{N,h} - d_{N,h}\}] + r_{N,h}$, $E_N|r_{N,h}|^{\gamma} = \mathrm{O}(N^{-(\gamma-1)} \times h^{\gamma+1})$, $\gamma \geqslant 1$.

Lemme 2.3. *Pour $\gamma \geqslant 2$, il existe c_γ, C_γ t.q. pour tout n, pour tous x et y de $S_{X,h}$*

$$E_N\left|\hat{p}^*_{N,h}(x) - \hat{p}^*_{N,h}(y)\right|^\gamma \leqslant c_\gamma |x-y|^\gamma \bigl(N\|h\|\bigr)^{-(\gamma-1)}, \quad E_N\left|\hat{r}_{N,h}(x) - \hat{r}_{N,h}(y)\right|^\gamma \leqslant C_\gamma |x-y|^\gamma \|h\|^{2\gamma}.$$

En utilisant Billingsley [1] :

Théorème 2.4. $\|\hat{p}_{N,h} - p\|_{S_{X,h}} \xrightarrow{P} 0$ *et* $U_{N,h} = (Nh)^{1/2}\{\hat{p}_{N,h} - p\}I\{S_{X,h}\}$ *converge en loi vers* $W_\sigma + N_0^{1/2}B$, W_σ *processus gaussien sur* S_X *t.q.* $EW_\sigma = 0$, $EW_\sigma(x)W_\sigma(x') = \sigma(x)^2 1\{x = x'\}$, $\sup_{x\in S_{X,h}} \sigma^{-1}(x)|U_{N,h}(x) - N_0^{1/2}(x)B(x)|$ *converge en loi vers* $\sup_{S_X} W_1$, *pour tout* $\varepsilon > 0$ *il existe* $c_\varepsilon > 0$ *t.q.* $\Pr\{\sup_{S_{X,h}} |\sigma^{-1}(U_{N,h} - N_0^{1/2}B) - W_1| > c_\varepsilon\} \xrightarrow{P} 0$.

Comme p est une fonction décroissante, l'estimateur des quantiles est défini par

$$\hat{q}_{N,h}(u) = \sup\bigl\{x \in S_{X,h}\colon\ \hat{p}_{N,h}(x) < u\bigr\} \quad \text{sur } \widehat{S}_{U,h} = \hat{p}_{N,h}(S_{X,h}).$$

Lemme 2.5. *Sur $S_{X,h}$, $p_{N,h}$ est monotone pour n assez grand et pour tous $x_1 < x_2$, $\zeta > 0$, il existe $C > 0$ t.q.* $\Pr\{\hat{p}_{N,h}(x_1) - \hat{p}_{N,h}(x_2) > C\} \geqslant 1 - \zeta$.

Proposition 2.6. *Sur* $\widehat{S}_{U,h}$, $E_N\hat{q}_{N,h}(u) = q(u) - \{h^2 \frac{B}{p^{(1)}}\} \circ q(u) + \mathrm{o}(\|h\|^2)$,

$$\mathrm{var}_N\bigl\{\hat{q}_{N,h}(u)\bigr\} = \left\{\frac{\sigma^2}{Nh\{p^{(1)}\}^2}\right\} \circ q(u) + \mathrm{o}\bigl(\bigl(N\|h\|\bigr)^{-1}\bigr), \qquad E_N\bigl|\hat{q}_{N,h}(u) - q(u)\bigr|^\gamma = \mathrm{O}\bigl(\|h\|^{2\gamma}\bigr), \quad \gamma > 2.$$

Théorème 2.7. (a) $\|\hat{q}_{N,h} - q\|_{\widehat{S}_{U,h}} \xrightarrow{P} 0$.

(b) $V_{N,h} = (Nh)^{1/2}\{\hat{q}_{N,h} - q\}I\{\widehat{S}_{U,h}\}$ *converge en loi vers* $\frac{W_\sigma - N_0^{1/2}B}{p^{(1)}} \circ q$,

(c) *pour tout* $\varepsilon > 0$ *il existe* $c_\varepsilon > 0$ *t.q.* $\Pr\{\sup_{\widehat{S}_{U,h}} |(\frac{p^{(1)}}{\sigma} \circ q\{V_{N,h} + (N_0^{1/2}B) \circ q\} - W_1| > c_\varepsilon\} \xrightarrow{P} 0$.

3. Choix de la fenêtre

L'erreur quadratique asymptotique de $\hat{q}_{N,h}$ est

$$\mathrm{AMSE}_q(u,h) = \frac{\sigma^2}{Nh\{p^{(1)}\}^2} \circ q(u) + h^4\left\{\frac{B}{p^{(1)}} \circ q(u)\right\}^2,$$

elle est minimale pour la fenêtre locale

$$h_{\mathrm{opt,loc}}(u) = N^{-1/5}\left\{\frac{\sigma^2}{4B^2} \circ q(u)\right\}^{1/5}$$

qui est égale à la fenêtre locale optimale pour l'erreur quadratique asymptotique de $\hat{p}_{N,h}$. D'après le Théorème 2.7 et sous les conditions D_1, D_2, la fonction $h_{\mathrm{opt,loc}}$ satisfait K_2, K_3 et, d'après le Théorème 2.7, $(Nh)^{1/2}\{\hat{q}_{N,h_{\mathrm{opt,loc}}} - q\}$ converge en loi vers G un processus gaussien de moyenne et variance

$$\mu(u) = -\frac{\sigma^2}{4p^{(1)}B} \circ q(u), \qquad \Sigma(u,u') = \frac{\sigma^2}{(p^{(1)})^2} \circ q(u) 1\{u = u'\}.$$

L'erreur quadratique moyenne asymptotique pour le quantile est

$$\int_{\widehat{S}_{U,h}} \mathrm{AMSE}_q(u,h)\,\mathrm{d}u = \int_{q(\widehat{S}_{U,h})} \frac{\mathrm{AMSE}_p(x,h)}{p^{(1)}(x)}\,\mathrm{d}x,$$

elle est approchée par

$$\mathrm{AMASE}_q(h) = N^{-1}\sum_{i=1}^n \sum_{j=1}^{J_i} \frac{\mathrm{AMSE}_p(X_{ij},h)}{p^{(1)}(X_{ij})}$$

et est minimale pour la fenêtre optimale globale

$$h_{\mathrm{opt,glob}} = N^{-1/5}\left\{\frac{\sum_{i=1}^{n}\sum_{j=1}^{J_i}\sigma^2(X_{ij})/p^{(1)}(X_{ij})}{4\sum_{i=1}^{n}\sum_{j=1}^{J_i}B^2(X_{ij})/p^{(1)}(X_{ij})}\right\}^{1/5}.$$

4. Généralisation

Lorsque (X, Y) est observé en des instants aléatoires $t_1, \ldots, t_J \subset [0, T]$ sous les mêmes conditions, sans l'hypothèse de symétrie de F, la loi de Y conditionnellement à X et les fonctions inverses q_1 et q_2 de p, à y ou x fixés, sont définies et estimées par

$$P(Y_t \leqslant y | X_t = x) := p(y; x) = F(y - m(x)), \qquad p(y; x) = u \Leftrightarrow \begin{cases} x = m^{-1}(y - F^{-1}(u)) := q_1(u; y), \\ y = m(x) + F^{-1}(u) := q_2(u; x), \end{cases}$$

$$\hat{p}_{N,h}(y; x) = \frac{\sum_{i=1}^{n}\sum_{j=1}^{J_i} K_h(x - X_{ij}) I_{\{Y_{ij} \leqslant y\}}}{\sum_{i=1}^{n}\sum_{j=1}^{J_i} K_h(x - X_{ij})},$$

$$\hat{q}_{1,N,h}(u; y) = \sup\bigl\{x \in I_{X,Y}\colon \hat{p}_{N,h}(y; x) < u\bigr\} = \hat{m}_{N,h}^{-1}\bigl(y - \widehat{F}_{N,h}^{-1}(u)\bigr),$$

$$\hat{q}_{2,N,h}(u; x) = \inf\bigl\{y \in I_{X,Y}\colon \hat{p}_{N,h}(y; x) > u\bigr\} = \widehat{F}_{N,h}^{-1}(u) + \hat{m}_{N,h}(x),$$

$$\widehat{F}_{N,h}(s) = \frac{1}{N}\sum_{i=1}^{n}\sum_{j=1}^{J_i}\hat{p}_{N,h}\bigl(s + \hat{m}_{N,h}(X_{ij}); X_{ij}\bigr),$$

où $\hat{m}_{N,h}$ est l'estimateur non paramétrique à noyau de m, des propriétés asymptotiques analogues sont établies sous les mêmes conditions.

Références

[1] P. Billingsley, Convergence of Probability Measures, Wiley, New York, 1968.
[2] P. Hall, Integrated square error properties of kernel estimators of regression functions, Ann. Statist. 12 (1984) 241–260.
[3] W. Härdle, Applied Nonparametric Regression, Cambridge University Press, Cambridge, 1990.
[4] E.A. Nadaraya, Nonparametric Estimation of Probability Densities and Regression Curves, Kluwer Academic Publishers, Boston, 1989.
[5] S.J. Sheather, J.S. Marron, Kernel quantile estimators, J. Amer. Statist. Assoc. 85 (1990) 410–416.