

ANNALES SCIENTIFIQUES  
DE L'UNIVERSITÉ DE CLERMONT-FERRAND 2  
*Série Mathématiques*

D. PICARD

J. DESHAYES

**Rupture de modèles : loi asymptotique des statistiques de tests et  
des estimateurs du maximum de vraisemblance**

*Annales scientifiques de l'Université de Clermont-Ferrand 2*, tome 71, série *Mathématiques*, n° 20 (1982), p. 115-118

[http://www.numdam.org/item?id=ASCFM\\_1982\\_\\_71\\_20\\_115\\_0](http://www.numdam.org/item?id=ASCFM_1982__71_20_115_0)

© Université de Clermont-Ferrand 2, 1982, tous droits réservés.

L'accès aux archives de la revue « Annales scientifiques de l'Université de Clermont-Ferrand 2 » implique l'accord avec les conditions générales d'utilisation (<http://www.numdam.org/conditions>). Toute utilisation commerciale ou impression systématique est constitutive d'une infraction pénale. Toute copie ou impression de ce fichier doit contenir la présente mention de copyright.

NUMDAM

Article numérisé dans le cadre du programme  
Numérisation de documents anciens mathématiques  
<http://www.numdam.org/>

RUPTURE DE MODELES : LOI ASYMPTOTIQUE DES STATISTIQUES DE TESTS ET DES ESTIMATEURS DU MAXIMUM DE VRAISEMBLANCE.

D. PICARD - J. DESHAYES.

Université de PARIS-SUD

Soient  $X_1, \dots, X_n$  des observations indépendantes. Le modèle paramétrique le plus classique est de supposer les variables  $X_i$  identiquement distribuées de loi  $G(., \theta)$  avec  $\theta \in \Theta$  : ouvert de  $\mathbb{R}^P$  et d'en déduire des renseignements sur  $\theta$ .

Ici, nous supposons que les indices des variables correspondent à un ordre (chronologique ou autre...) et nous voulons savoir si à partir d'un certain indice  $k$  inconnu, il a pu se produire une rupture dans le modèle. Nous traiterons donc les deux problèmes : (1) tester l'hypothèse  $H_0$  : "pas de rupture".  $X_1, \dots, X_n \sim G(., \theta)$  contre l'hypothèse  $H_1$  : "existence d'une rupture" :  $\exists k \in \{1, \dots, n-1\}$  tel que

$$X_1, \dots, X_k \sim G(., \theta^*)$$

$$X_{k+1}, \dots, X_n \sim G(., \theta^{**}) \text{ avec } \theta^{**} \neq \theta^*$$

(2) dans ce dernier modèle, estimer les paramètres  $\theta^*$ ,  $\theta^{**}$  et  $k$ .

Le paramètre  $k$  joue un rôle particulier et l'étude asymptotique conduit à des vitesses non standard.

L'étude de la convergence de la suite des processus de sommes partielles associée au processus de vraisemblance classique :

$$[0,1] \times \mathbb{R}^P \rightarrow \mathbb{R}$$

$$(t, \lambda) \quad \sum_{i=1}^{[nt]} \log \frac{g(X_i, \theta_0 + \frac{\lambda}{\sqrt{n}})}{g(X_i, \theta_0)}$$

fournit la loi asymptotique de la statistique de test et celles des estimateurs.

Théorème 1 :

Sous des hypothèses de régularité classiques sur le modèle [4], la suite de processus  $Z_n^\theta(t, \lambda^*, \lambda^{**})$  de  $\mathcal{C}([0,1] \times \mathbb{R}^p \times \mathbb{R}^p)$  définie par la ligne polygonale

$$\text{joignant les sommets } (0,0) \text{ et } \left(\frac{k}{n}, \sum_{i=1}^k \text{Log} \frac{g(X_i, \theta_0 + \frac{\lambda^*}{\sqrt{n}})}{g(X_i, \theta_0)} + \sum_{i=k+1}^n \text{Log} \frac{g(X_i, \theta_0 + \frac{\lambda^{**}}{\sqrt{n}})}{g(X_i, \theta_0)}\right)$$

converge étroitement sous  $H_0(\theta_0)$  dans  $\mathcal{C}([0,1] \times \mathbb{R}^p \times \mathbb{R}^p)$  muni de la topologie de la convergence uniforme sur les compacts vers le processus :

$$Z(t, \lambda^*, \lambda^{**}) = t_{\lambda^*} [I(\theta_0)]^{1/2} W_p(t) - \frac{1}{2} t_{\lambda^*} I(\theta_0) \lambda^* t + t_{\lambda^{**}} [I(\theta_0)]^{1/2} [W_p(1) - W_p(t)] - \frac{1}{2} t_{\lambda^{**}} I(\theta_0) \lambda^{**} (1-t)$$

avec  $I(\theta_0)$  : matrice d'information de Fisher de la loi  $G$  au point  $\theta_0$ .

$W_p(\cdot)$  : processus gaussien sur  $[0,1]$  à valeurs dans  $\mathbb{R}^p$  formé de  $p$  composantes browniennes indépendantes.

Pour obtenir une loi asymptotique de statistique de test, nous sommes amenés à renormaliser la vraisemblance afin de pénaliser les bords.

Corollaire 1 : Sous les conditions de régularité, la suite de statistiques

$$\text{Sup}_{k=1, \dots, n} 2 \frac{k}{n} \cdot \frac{n-k}{n} \cdot \text{Sup}_{\lambda^* \in \mathbb{R}^p} \text{Sup}_{\lambda^{**} \in \mathbb{R}^p} \text{Inf}_{\lambda \in \mathbb{R}^p} \sum_{i=1}^k \log \frac{g(X_i, \theta_0 + \frac{\lambda^*}{\sqrt{n}})}{g(X_i, \theta_0 + \frac{\lambda}{\sqrt{n}})} + \sum_{i=k+1}^n \log \frac{g(X_i, \theta_0 + \frac{\lambda^{**}}{\sqrt{n}})}{g(X_i, \theta_0 + \frac{\lambda}{\sqrt{n}})}$$

converge sous  $H_0(\theta_0)$  vers

$$\text{Sup}_{t \in [0,1]} t [W_p(t) - t W_p(1)] [W_p(t) - t W_p(1)] .$$

Les démonstrations de ces deux résultats figurent dans [4]. L'estimation des paramètres  $(k, \theta^*, \theta^{**})$  repose fondamentalement sur le comportement asymptotique de l'amplitude de la rupture :  $\theta^{**} - \theta^* = d$  et nous distinguons trois cas, l'instant de

rupture sera toujours supposé éloigné des bords :  $k = n\tau$  avec  $\tau \in ]0,1[$ ,

i)  $d$  constant (lorsque  $n$  tend vers l'infini). HINKLEY [1] a calculé la loi asymptotique de l'estimateur  $\hat{k}$  par des méthodes de fluctuations, mais la limite dépend de  $d$  (inconnu).

ii)  $d_n$  tend vers 0 avec  $\|d_n\| \sqrt{\frac{\text{Log Log } n}{n}} \rightarrow \infty$ . Le nouveau paramétrage défini par :

$$\begin{cases} T_n(k) = (k-n\tau) t_{d_n} \cdot I(\bar{\theta}) d_n \\ \bar{\theta} = \tau \theta^* + (1-\tau) \theta^{**} \\ \delta_n = \sqrt{\tau(1-\tau)} d_n \end{cases}$$

fournit des lois asymptotiques simples.

Théorème 2 :

Sous les hypothèses de régularité [6], la suite de processus  $Y_n(T, \bar{\lambda}, \delta)$  de  $\mathcal{C}(\mathbb{R}^{2p+1})$  définie par la ligne polygonale joignant les sommets (0,0) et

$$(T_n(k), \text{Log} \frac{dP_{H_{1,n}}(k, \bar{\theta} + \frac{\bar{\lambda}}{\sqrt{n}}, \delta_n + \frac{\delta}{\sqrt{n}})}{dP_{H_{1,n}}(n\tau, \bar{\theta}, \delta_n)}(X_1, \dots, X_n))$$

converge étroitement sous  $H_{1,n}(n\tau, \bar{\theta}, \delta_n)$  vers

$$Y(T, \bar{\lambda}, \delta) = t_{\bar{\lambda}} [I(\bar{\theta})]^{1/2} \xi_1 + t_{\delta} [I(\bar{\theta})]^{1/2} \xi_2 + W(T) \\ - \frac{1}{2} \left\{ t_{\bar{\lambda}} I(\bar{\theta}) \bar{\lambda} + t_{\delta} I(\bar{\theta}) \delta + \frac{|T|}{2} \right\}$$

où  $\xi_1$  et  $\xi_2$  sont des vecteurs gaussiens indépendants  $\mathcal{N}(0, I_{p \times p})$  et

$$W(T) = W_1(T) \cdot 1_{[0, \infty[}(T) + W_2(T) \cdot 1_{]-\infty, 0]}(T),$$

$W_1$  et  $W_2$  étant des mouvements browniens indépendants entre eux et indépendants de  $(\xi_1, \xi_2)$ .

Corollaire 2 :

Sous les conditions de régularité, et sous  $H_{1,n}(\pi, \bar{\theta}, d_n)$ , les estimateurs du maximum de vraisemblance convergent en loi respectivement

$$\sqrt{n}(\hat{\theta}_n - \bar{\theta}) \rightarrow \mathcal{N}(0, I(\bar{\theta})^{-1})$$

$$\sqrt{n}(\hat{\delta}_n - \delta_n) \rightarrow \mathcal{N}(0, I(\bar{\theta})^{-1})$$

$t_{d_n} I(\bar{\theta}) d_n (\hat{k} - \pi) \rightarrow \Psi$  et sont asymptotiquement indépendants . La loi  $\Psi$  admet pour densité :

$$\psi(T) = \psi(-T) = \frac{3}{2} e^T \int_{\frac{3}{2}\sqrt{T}}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{u^2}{2}} du - \frac{1}{2} \int_{\frac{\sqrt{T}}{2}}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{u^2}{2}} du, \quad T \geq 0$$

Les démonstrations de ces derniers résultats figurent dans [6] ainsi qu'une extension au cas de rupture dans un modèle de régression du type :  $X_i - t_{\theta} f\left(\frac{i}{n}\right)$  suit une loi  $G$  connue, avec  $f$  fonction continue sur  $[0,1]$  à valeurs dans  $\mathbb{R}^p$ .

REFERENCES :

- [1] HINKLEY D.V. : "Inference about the change-point in a sequence of random variables" *Biometrika* (1970) n° 57, p. 1-17
- [2] IBRAGIMOV I.A., KHASHMINSKII R.Z. : "Asymptotic behaviour of statistical estimators in the smooth case. I- Study of the likelihood ratio", T.P.A. (1972), vol. 17, p. 445-462
- [3] IBRAGIMOV I.A., KHASHMINSKII R.Z. : "Local asymptotic normality for non identically distributed observations", T.P.A. (1975) n° 20, p. 246-260
- [4] DESHAYES J., PICARD D. : "Testing for a change-point in statistical models" *Prépublications d'Orsay*, 1980, t. 52
- [5] DESHAYES J., PICARD D. : "Convergence de processus à double indice : application aux tests de rupture dans un modèle". *Note C.R.A.S. t.292* (1981), série I, p. 449-452.
- [6] PICARD D., DESHAYES J. : "Rupture dans les modèles de régression : loi asymptotique des tests et estimateurs du maximum de vraisemblance" 1981 - *Prépublication d'Orsay*.