ANNALES SCIENTIFIQUES
DE L'UNIVERSITÉ DE CLERMONT-FERRAND 2
Série Probabilités et applications

J. DESHAYES

Rupture de modèles pour des processus de Poisson

Annales scientifiques de l'Université de Clermont-Ferrand 2, tome 78, série *Probabilités et applications*, n° 2 (1984), p. 1-7

<http://www.numdam.org/item?id=ASCFPA_1984__78_2_1_0>

RUPTURE DE MODELES POUR DES PROCESSUS DE POISSON

J. DESHAYES

E.N.S.T. PARIS

Nous nous intéressons à des situations où les observations indicées chronologiquement présentent des changements brusques (ruptures) de modèles. Nous étudions principalement les modèles paramétriques à temps continu (processus à accroissements indépendants) et à temps discret (variables indépendantes, pas forcément de même loi) avec saut de paramètres à des instants inconnus. Les tests d'existence de rupture et les estimateurs des paramètres d'une rupture sont développés dans un cadre asymptotique permettant d'aboutir à des lois limites non classiques.

Citons quelques exemples classiques de modèles où les résultats s'appliquent :

. les translations du mouvement brownien (diffusions à coefficients déterministes) sur $[0, T]$

  $dX_t = f(\eta,t)\ dt + \sigma \,.\, dW_t$

  avec W : brownien standard et f : fonction connue

  $\theta = (\eta,\sigma)$ étant les paramètres à estimer et susceptibles de sauter.

. le processus de Poisson d'intensité paramétrique sur $[0, T]$: X_t est le processus à accroissement indépendant tel que $X_{s+h} - X_s$ suit une loi de Poisson de paramètre $\int_s^{s+h} f(\theta, t)dt$.

. les modèles statistiques classiques du n-échantillon : $(X_1, \ldots, X_n)$ où les X_i sont de densité $p(.,\theta)$ par rapport à une mesure dominante fixe.

. les modèles de régression paramétrique (linéaire ou non) où l'on peut écrire que $X_i - f(\eta, \frac{i}{n})$ est de densité $p(.,\sigma)$.

Les développements correspondant à ces différents modèles figurent dans /1/ ; afin d'énoncer rapidement des résultats, nous allons nous placer dans le cas simple du processus de Poisson homogène sur R et renvoyons à /2/ pour des démonstrations détaillées.

I - TEST D'EXISTENCE D'UNE RUPTURE D'INTENSITE

On considère un processus de Poisson X(.) homogène sur un intervalle de temps $[0, T]$ et on teste l'hypothèse nulle

H_o : X(.) est d'intensité constante θ sur $[0, T]$

contre l'hypothèse d'existence d'un changement d'intensité à un instant τ inconnu

H_1 : X(.) est l'intensité constante θ_1 sur $[0,\tau]$ et θ_2 sur $[\tau, T]$
les intensités θ, θ_1, θ_2 étant supposées inconnues.

Dans ce modèle particulier, la statistique X(T) est exhaustive complète pour le paramètre θ sous H_o ; alors la restriction aux tests sans biais de niveau α conduit immédiatement à la recherche de tests ayant une structure de Neyman, c'est-à-dire aux tests de niveau α conditionnellement en X(T).

1° Définition des tests

Conditionnellement en X(T) = n, le modèle du processus X(.) est équivalent :

. sous H_o, à un n-échantillon de densité uniforme sur $[0, T]$.

. sous H_1, à un n-échantillon de densité

$$\frac{1}{\tau+\rho(T-\tau)} \quad 1_{[0,\tau[} + \quad \frac{\rho}{\tau+\rho(T-\tau)} \quad 1_{[\tau, T[}$$

en posant $\dfrac{\theta_2}{\theta_1} = \rho$.

Ici, nous sommes ramenés au problème statistique classique de tester une hypothèse simple contre une hypothèse composée où le paramètre τ d'instant de rupture joue le même rôle que le paramètre plus classique ρ du rapport des intensités.

En l'absence de monotonie du rapport de vraisemblance, deux tests s'imposent ici :

a) Le test de Kolmogorov-Smirnov

Sa région de rejet est définie par :

$$\sup_{t \in [0, T[} \left| \frac{X(t)}{n} - \frac{t}{T} \right| > c_n(\alpha)$$

La statistique du test n'est autre que la différence renormalisée des intensités moyennes :

$$\frac{X(t)}{X(\ )} - \frac{t}{T} = \left[\frac{X(t)}{t} - \frac{X(\_)-X(t)}{T-t}\right] \cdot \frac{t(T-t)}{T.X(T)}$$

b) Le test du rapport de vraisemblance

Sa région de rejet se simplifie en :

$$\sup_{t \in [0, T[} n.K \left[\frac{X(t)}{n}, \frac{t}{T}\right] > d_n(\alpha)$$

où K désigne l'information de Kullback : $K(p,q) = p \log \frac{p}{q} + (1-p)\log \frac{1-p}{1-q}$.

Ces deux tests sont en fait d'utilisation non paramétrique puisqu'ils s'écrivent respectivement avec les statistiques $\sup_s |\hat{F}_n(s)-F(s)|$ et $\sup_s K[\hat{F}_n(s), F(s)]$ où $\hat{F}_n$ désigne la fonction de répartition empirique d'un n-échantillon de fonction de répartition F sous H_o.

2° Comparaisons asymptotiques des deux tests

Le seul cadre où une propriété d'optimalité peut être mise en évidence est celui des grandes déviations et nous avons la

PROPOSITION 1 :

Le test du rapport de vraisemblance est asymptotiquement optimal au sens non local pour tester H_o contre H_1 parmi les tests conditionnels (en X(T)) de même niveau exponentiel.

Malheureusement, l'approximation du niveau du test de vraisemblance conduit à renormaliser la statistique près des bords de l'intervalle d'observation (0 et T), ce qui lui enlève la propriété d'optimalité asymptotique au sens non local et pour la renormalisation particulière par t(T-t), on obtient l'équivalence asymptotique au sens local des deux tests :

PROPOSITION 2 :

Sous $H_o(\theta)$, $\sup\limits_{t\in[0,\,T]} \left| \frac{2t(T-t)}{T} \cdot X(T).K\left[\frac{X(t)}{X(T)}, \frac{t}{T}\right] - X(T) \cdot \left[\frac{X(t)}{X(T)} - \frac{t}{T}\right]^2 \right|$

tend vers 0 en probabilité.
En conséquence, les deux tests ont même puissance sous hypothèse contigüe à H_o.

Remarque : la suite de modèles définissant l'asymptotique n'a pas été précisée volontairement car nous avons le choix entre faire tendre l'intensité totale vers l'infini sur un intervalle fixe ou faire tendre l'intervalle d'observation vers l'infini avec une intensité fixe : cela revient au même par un changement d'échelle de temps à cause de l'homogénéité.

II - LOIS ASYMPTOTIQUES DES ESTIMATEURS DU MAXIMUM DE VRAISEMBLANCE D'UNE RUPTURE

Toujours sur l'exemple du processus de Poisson homogène nous allons envisager différents cadres asymptotiques pour étudier les estimateurs des paramètres d'une rupture. Pour l'étude de modèles plus généraux, on se reportera à /1/ - /2/ - /3/ - /4/.
Sous l'hypothèse H_1 d'existence d'un changement d'intensité, le processus X(.) est d'intensité :

$$\theta \left\{\frac{1}{\tau + \rho\,(1-\tau)}\, 1_{[0,\tau[}(t) + \frac{\rho}{\tau + \rho(1-\tau)}\, 1_{[\tau,\,1[}(t)\right\}$$

où θ désigne l'intensité totale et ρ le rapport des intensités avant et après rupture ; on s'est ramené à un intervalle d'observation unité par changement d'échelle et l'étude asymptotique consistera à faire tendre l'intensité totale θ vers l'infini.
Les estimateurs du maximum de vraisemblance sont respectivement

$$\hat{\theta} = X(1) \quad , \hat{\rho} = \frac{X(1)-X(\hat{\tau})}{1-\hat{\tau}} \cdot \frac{\hat{\tau}}{X(\hat{\tau})} \quad \text{et} \quad \hat{\tau} = \operatorname{Arg} \sup_{t \in]0,1[} K \left[\frac{X(t)}{X(1)} , t\right]$$

La loi exacte ne peut être calculée que pour le premier d'entre eux et nous allons définir des lois approchées pour une suite d'hypothèses $H_{1,\nu}$ de paramètres $(\theta_\nu , \tau_\nu, \rho_\nu)$ avec θ_ν tendant vers l'infini.
Comme dans /5/, les lois asymptotiques des estimateurs sont déduites de la convergence d'un processus de vraisemblance (voir également /3/).

<u>1er cas : τ et ρ restent fixes.</u>

<u>PROPOSITION 3</u> :

Sous la suite d'hypothèses $H_{1,\nu}$ $(\theta_\nu , \tau ,\rho)$, les estimateurs de chacun des paramètres sont asymptotiquement indépendants et de lois limites respectives :

$$\sqrt{\theta_\nu} \left(\frac{\hat{\theta}_\nu}{\theta_\nu} - 1\right) \xrightarrow{\mathcal{L}} \mathcal{N}(0 \; ; \; 1)$$

$$\sqrt{\theta_\nu} \left(\hat{\rho}_\nu - \rho\right) \xrightarrow{\mathcal{L}} \mathcal{N}\left(0 \; ; \; \frac{\tau(1-\tau)}{\rho\left[\tau+(1-\tau)\rho\right]^2}\right)$$

$$\theta_\nu \frac{(\hat{\tau}_\nu - \tau)}{\tau+(1-\tau)\rho} \xrightarrow{\mathcal{L}} \operatorname{Arg} \sup_{t \in R} \left\{\log \frac{1}{\rho} \cdot \overline{X}_\rho(t) - (1-\rho)t\right\}$$

avec $\overline{X}_\rho$ désignant le processus de Poisson sur R d'intensité :
$1_{]-\infty,0[}(t) + \rho . 1_{[0,\infty[}(t)$.
L'inconvénient de ce cadre asymptotique est que la loi limite de $\hat{\tau}$ dépend trop fortement du paramètre ρ inconnu car la rupture est trop brutale et ne permet pas le calcul d'intervalles de confiance

approchés. Des ruptures moins brutales vont fournir une loi limite universelle :

<u>2ème cas : τ reste fixe mais ρ_ν tend vers 1.</u>

Toutefois pour que les paramètres ρ et τ puissent être estimés de façon consistante, il ne faut pas que la suite d'hypothèses $H_{1,\nu}$ soit contigüe à H_o : cela nécessite la convergence vers l'infini de la suite $\sqrt{\theta_\nu}|\rho_\nu - 1|$.

<u>PROPOSITION 4</u> :

Sous la suite d'hypothèses $H_{1,\nu}$ $(\theta_\nu, \tau, \rho_\nu)$, les estimateurs du maximum de vraisemblance sont asymptotiquement indépendants et de lois respectives :

$$\sqrt{\theta_\nu}\left(\frac{\hat{\theta}_\nu}{\theta_\nu} - 1\right) \xrightarrow{\mathcal{L}} \mathcal{N}(0,1)$$

$$\sqrt{\theta_\nu}(\hat{\rho}_\nu - \rho_\nu) \xrightarrow{\mathcal{L}} \mathcal{N}(0\ ;\tau(1-\tau))$$

$$\theta_\nu(\rho_\nu-1)^2(\hat{\tau}_\nu-\tau) \xrightarrow{\mathcal{L}} \operatorname{Arg}\sup_{t\in R} \{W(t) - \frac{|t|}{2}\}$$

où W est un mouvement brownien standard sur R.
Le dernier cas à envisager est celui où la rupture (ou l'une des ruptures) se produit en début ou en fin d'observation.

<u>3ème cas : τ_ν tend vers 0</u>

Pour être dans un cadre asymptotique où les hypothèses $H_{1,\nu}$ ne sont pas contigües à H_o, on suppose donc que $\theta_\nu \tau_\nu$ $(\rho_\nu - 1)$ tend vers l'infini et c'est alors $\theta_\nu \tau_\nu$ qui donnent les vitesses d'estimations de ρ et τ au lieu de θ_ν, les lois limites étant les mêmes que précédemment.

REFERENCES :

/1/ J. DESHAYES, D. PICARD : "Ruptures de modèles en statistique" Thèses d'état - Orsay (1983).

/2/ J. DESHAYES : "Rupture de modèle pour des processus ponctuels" (à paraître).

/3/ J. DESHAYES, D. PICARD : "Principe d'invariance sur le processus de vraisemblance" (à paraître dans les Annales de l'I.H.P.).

/4/ J. DESHAYES, D. PICARD : "Ruptures dans les modèles de régression : "Lois asymptotiques des test et estimateurs du maximum de vraisemblance". (à paraître).

/5/ I.A. IBRAGIMOV, R.Z. HAS'MINSKII : "Asymptotic behaviour of statistical estimators in the smooth case". Theory of probability and applications (1972) p. 445-462.

J. DESHAYES
Ecole Nationale Supérieure
des Télécommunications
Département SYC
46, rue Barrault

75634 PARIS Cédex 13

Reçu en Septembre 1983