

REVUE DE STATISTIQUE APPLIQUÉE

DARIUSH GHORBANZADEH

RACHID LOUNES

Deux procédures de détection de rupture pour des observations poissonniennes groupées

Revue de statistique appliquée, tome 44, n° 3 (1996), p. 47-61

http://www.numdam.org/item?id=RSA_1996__44_3_47_0

© Société française de statistique, 1996, tous droits réservés.

L'accès aux archives de la revue « *Revue de statistique appliquée* » (<http://www.sfds.asso.fr/publicat/rsa.htm>) 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/>

DEUX PROCÉDURES DE DÉTECTION DE RUPTURE POUR DES OBSERVATIONS POISSONNIENNES GROUPÉES

Dariush Ghorbanzadeh (1), Rachid Lounes (2)

(1) CNAM-Département Mathématiques 292, Rue Saint-Martin,
75141 Paris cedex 03, France

(2) Université Paris V, Laboratoire de Statistique Médicale, UA CNRS 1323
45, rue des Saints-Pères, 75006 Paris, France

RÉSUMÉ

On considère deux procédures de détections de ruptures pour des observations poissonniennes groupées. La première procédure est basée sur la statistique du rapport de vraisemblance adaptée au problème de rupture. La loi asymptotique de la statistique de test, sous l'hypothèse nulle est celle du sup d'une fonctionnelle carrée d'un pont brownien. La deuxième procédure, bayésienne, est basée sur la loi a posteriori du paramètre de l'instant de rupture. Nous appliquons les deux procédures aux données concernant l'hypospadias, recueillies dans les registres des naissances à Liverpool durant l'année 1960 et 1982 et nous faisons une comparaison entre les deux méthodes.

Mots-clés : Rupture, Test du rapport de vraisemblance pénalisé, pont brownien, loi a priori, loi a posteriori.

ABSTRACT

We consider two tests with regard to detection of change of parameter in a sequence of independent Poisson random variables. The first test considered is based on the likelihood ratio appropriately adapted for a change point problem. The limiting distribution of the test statistics, under the null hypothesis is evaluated by asymptotical techniques. Under the null hypothesis the distribution of the test statistic obtained by maximization of the likelihood ratio converges to the supremum of an square function of the Brownian bridge process. The second test use Bayes technics, and is based on the marginal posterior density of the change. Finally, we apply the results for a set of data from the Liverpool Registry during 1960 to 1982, and compare the two procedures.

Keywords : Change point, Penalized likelihood ratio test, brownian bridge, prior distribution, posterior distribution.

1. Introduction et modèle statistique

Soit X_1, \dots, X_n une suite de variables aléatoires; dans ce travail nous nous proposons d'étudier le problème de test suivant :

$$H_0 : \mathcal{P}(X_i = x_i) = e^{-n_i \lambda_0} \frac{(n_i \lambda_0)^{x_i}}{x_i!} \quad \text{pour } i = 1, \dots, n.$$

Contre

$$H_1^{(N_n)} = \bigcup_{k=1}^{n-1} H_1 \left(\frac{N_k}{N_n} \right)$$

avec $H_1 \left(\frac{N_k}{N_n} \right) : \mathcal{P}(X_i = x_i) = \begin{cases} e^{-n_i \lambda_1} \frac{(n_i \lambda_1)^{x_i}}{x_i!} & \text{si } i = 1, \dots, k \\ e^{-n_i \lambda_2} \frac{(n_i \lambda_2)^{x_i}}{x_i!} & \text{si } i = k + 1, \dots, n. \end{cases}$

où

$$\sum_{i=1}^n n_i = N_n, \quad \sum_{i=1}^k n_i = N_k, \quad \sum_{i=k+1}^n n_i = N_k^*$$

et les paramètres $\lambda_0, \lambda_1, \lambda_2$ sont inconnus.

Remarque. Chaque X_i peut être vu comme une somme de n_i variables aléatoires indépendantes et équidistribuées de loi de Poisson de paramètre λ .

Dans la littérature, plusieurs auteurs comme Chernoff et Zacks [3], Sen et Srivastava [11] ont proposé des tests de détection de rupture dans la moyenne de variables gaussiennes. Hinkley et Hinkley [8] proposent des tests pour des variables binomiales. Worsley [15] propose un test de détection de rupture pour des observations poissonniennes qui fait intervenir le rapport des paramètres avant et après la rupture; il approxime la loi de la statistique du rapport de vraisemblances, conditionnellement à la somme totale des observations, par celle d'une statistique appelée «Somme Cumulée» qui suit, sous l'hypothèse nulle, une loi du χ^2 . D'autres auteurs comme Bhattacharya et Frierson [1] ont proposé des méthodes non paramétriques basées sur les rangs ou la fonction de répartition empirique associée aux observations. Simpkin et Downham (1988) ont proposé un test de détection pour des observations poissonniennes basé sur la statistique du rapport des vraisemblances maximales; en supposant l'instant de rupture fixé, ils approximent la loi de cette statistique par une loi du χ^2 et ils appliquent leurs résultats à un fichier de données concernant l'hypospadias, enregistrées durant les années 1960 à 1982 à Liverpool.

Dans cet article un test de détection de rupture du paramètre de variables de Poisson indépendantes est proposé; le test est basé sur la statistique Log-vraisemblance maximale pénalisée. Nous montrons que la loi asymptotique de cette statistique se comporte comme celle d'une fonctionnelle d'un pont brownien; ceci permet l'évaluation asymptotique du niveau du test étudié.

Nous proposons une application pratique aux données utilisées par Simpkin et Downham [13] et donnons un intervalle de confiance pour l'instant de rupture.

2. Test du rapport de vraisemblance

La statistique Log-vraisemblance est définie par :

$$\Lambda_{N_n} = \max_{1 < k < n} \Lambda_{N_n}^* \left(\frac{N_k}{N_n} \right) \quad (1)$$

avec

$$\Lambda_{N_n}^* \left(\frac{N_k}{N_n} \right) = N_k \bar{X}_{N_k} \log \left(\frac{\bar{X}_{N_k}}{\bar{X}_{N_n}} \right) + N_k^* \bar{X}_{N_k}^* \log \left(\frac{\bar{X}_{N_k}^*}{\bar{X}_{N_n}} \right) \quad (2)$$

où

$$\frac{1}{N_n} \sum_{i=1}^n X_i = \bar{X}_{N_n}, \quad \frac{1}{N_k} \sum_{i=1}^k X_i = \bar{X}_{N_k} \quad \text{et} \quad \frac{1}{N_k^*} \sum_{i=k+1}^n X_i = \bar{X}_{N_k}^*.$$

La valeur \hat{k} de k réalisant le maximum de $\Lambda_{N_n}^* \left(\frac{N_k}{N_n} \right)$ est l'estimateur du maximum de vraisemblance de k sous l'hypothèse alternative $H_1^{(N_n)}$. \square

Pour l'étude du comportement asymptotique de $\Lambda_{N_n}^*$, sous l'hypothèse H_0 , faisons quelques remarques préliminaires.

Dans cette étude, nous fixons n et supposons que les n_i sont suffisamment grands, de sorte que N_n soit grand.

1. Il est facile de vérifier que

$$\begin{cases} \bar{X}_{N_n} = \frac{N_k}{N_n} \bar{X}_{N_k} + \left(1 - \frac{N_k}{N_n}\right) \bar{X}_{N_n}^* \\ \bar{X}_{N_n} - \bar{X}_{N_k} = \left(1 - \frac{N_k}{N_n}\right) (\bar{X}_{N_k}^* - \bar{X}_{N_k}) \\ \bar{X}_{N_n} - \bar{X}_{N_k}^* = \frac{N_k}{N_n} (\bar{X}_{N_k} - \bar{X}_{N_k}^*) \end{cases} \quad (3)$$

De (3) on déduit

$$\begin{cases} \frac{\bar{X}_{N_n}}{\bar{X}_{N_k}} = 1 - \left(1 - \frac{N_k}{N_n}\right) \left(1 - \frac{\bar{X}_{N_k}^*}{\bar{X}_{N_k}}\right) \\ \frac{\bar{X}_{N_n}}{\bar{X}_{N_k}^*} = 1 - \frac{N_k}{N_n} \left(1 - \frac{\bar{X}_{N_k}}{\bar{X}_{N_k}^*}\right) \end{cases} \quad (4)$$

2. Pour tout $N_s \leq N_u < N_n$, on a :

$$\text{Cov}_{H_0}(\bar{X}_{N_s} - \bar{X}_{N_s}^*, \bar{X}_{N_u} - \bar{X}_{N_u}^*) = \frac{\lambda_0}{N_n} \left(\frac{N_u}{N_n} \left(1 - \frac{N_s}{N_n} \right) \right)^{-1} \quad (5)$$

3. Pour tout $t \in [0, 1]$, soit $V(t)$ un *pont Brownien*; c'est-à-dire un processus gaussien sur $[0, 1]$ conditionné par : $V(0) = V(1) = 0$, on sait alors que pour $t_1 \in [0, 1]$ et $t_2 \in [0, 1]$: $\text{Cov}(V(t_1), V(t_2)) = \min(t_1, t_2) - t_1 t_2$.

Alors pour tout $0 < t_1 < t_2 < 1$, on a :

$$\text{Cov} \left(\frac{V(t_1)}{\sqrt{t_1(1-t_1)}}, \frac{V(t_2)}{\sqrt{t_2(1-t_2)}} \right) = \sqrt{\frac{t_1(1-t_2)}{t_2(1-t_1)}} \quad (6)$$

DÉFINITION 1. Soit $g_n : [0, 1] \rightarrow [0, 1]$ définie par $g_n(t) = \frac{1}{N_n} \sum_{i=1}^{[nt]} n_i$ ($[A]$ désigne la partie entière du réel A).

Il est facile de vérifier que l'application g_n est croissante, que $g_n(0) = 0$ et que $g_n(1) = 1$. De plus, la restriction de g_n à $]0, 1[$ est l'interpolée linéaire de la fonction en escalier définie sur

$$\left\{ \frac{1}{n}, \dots, 1 - \frac{1}{n} \right\} \quad \text{par} \quad g_n \left(\frac{k}{n} \right) = \frac{1}{N_n} \sum_{i=1}^k n_i = \frac{N_k}{N_n}.$$

DÉFINITION 2. On définit $\{\Lambda_{N_n}^*(t), t \in [0, 1]\}$ comme le processus interpolé linéaire du processus $\left\{ \Lambda_{N_n}^* \left(\frac{N_k}{N_n} \right), 1 < k < n \right\}$.

THÉORÈME 1. Sous l'hypothèse H_0 , $\Lambda_{N_n}^*$ converge au sens de la convergence en loi des marginales de dimension finie vers le processus Λ_0 défini par :

$$\lambda_0(t) = \frac{V^2(g_n(t))}{2g_n(t)(1-g_n(t))} \quad (7)$$

où V est un pont Brownien.

Preuve. Le développement de Taylor de (2) donne

$$\begin{aligned} \Lambda_{N_n}^*(t) &= \frac{N_n}{2} \frac{[N_n g_n(t)]}{N_n} \left(1 - \frac{[N_n g_n(t)]}{N_n} \right) \\ &\quad \left(\bar{X}_{[N_n g_n(t)]} - \bar{X}_{[N_n g_n(t)]}^* \right)^2 \frac{\bar{X}_{N_n}}{\bar{X}_{[N_n g_n(t)]} \bar{X}_{[N_n g_n(t)]}^*} + o(R_{N_n}) \quad (8) \end{aligned}$$

où R_{N_n} converge vers 0 avec probabilité 1.

Notons

$$U_{N_n}(t) = \sqrt{\frac{N_n [N_n g_n(t)]}{\lambda_0 N_n} \left(1 - \frac{[N_n g_n(t)]}{N_n}\right)} \left(\bar{X}_{[N_n g_n(t)]} - \bar{X}_{[N_n g_n(t)]}^*\right).$$

Il est facile de voir que le processus $\{U_{N_n}(t), t \in [0, 1]\}$ est asymptotiquement gaussien (comme la différence de deux processus asymptotiquement gaussiens) et que sous l'hypothèse H_0 , pour tout $s \leq t$ et compte tenu de (5) il vérifie :

$$\text{Cov}(U_{N_n}(s), U_{N_n}(t)) = \sqrt{\frac{[N_n g_n(s)](1 - \frac{[N_n g_n(t)]}{N_n})}{[N_n g_n(t)](1 - \frac{[N_n g_n(s)]}{N_n})}}.$$

$$\text{D'où } \lim_{N_n \rightarrow \infty} \text{Cov}(U_{N_n}(s), U_{N_n}(t)) = \sqrt{\frac{g_n(s)(1 - g_n(t))}{g_n(t)(1 - g_n(s))}}.$$

$$\text{D'autre part, } \frac{\bar{X}_n}{\bar{X}_{[N_n g_n(t)]} \bar{X}_{[N_n g_n(t)]}^*} \text{ converge en probabilité vers } \frac{1}{\lambda_0}.$$

De (6) et (8) on déduit la convergence de $\Lambda_{N_n}^*$ vers λ_0 . \square

L'un des problèmes cruciaux de détection de rupture est posé par les effets de bords (les ruptures intervenant aux bords de l'intervalle d'observations). En effet, pour les t proches de 0 ou 1, Λ_0 n'est pas continue.

Pour pénaliser ces effets de bords, nous adoptons la méthode proposée par Shorack & Wellner [12] (inégalité de Birnbaum and Marshall en prenant $h(t) = \frac{\psi(t)}{t}$, page 129), en considérant la suite de processus $\Gamma_{N_n}(t) = \psi(t)\psi(1-t)\Lambda_{N_n}^*(t)$, où ψ est une fonction continue par morceaux sur $[0, 1]$ vérifiant $\int_0^1 \left(\frac{\psi(t)}{t}\right)^2 dt < \infty$.

Posons $\Gamma(t) = \psi(t)\psi(1-t)\Lambda_0(t)$, c'est un processus continu pour $t \in [0, 1]$; en considérant les moments d'ordre 4 de $\{\Gamma_{N_n}(t), t \in [0, 1]\}$ on montre facilement que $\{\Gamma_{N_n}(t), t \in [0, 1]\}$ converge en loi vers $\{\Gamma(t), t \in [0, 1]\}$ lorsqu'on munit l'espace $\mathcal{C}([0, 1])$ des fonctions continues sur $[0, 1]$ de la topologie de la convergence uniforme (voir Billingsley [2]).

En utilisant la continuité de la fonction «sup», (cf. Billingsley [2]), sur l'espace $\mathcal{C}([0, 1])$, on a alors la proposition suivante.

PROPOSITION. Sous l'hypothèse H_0 , pour tout $\eta \geq 0$, on a :

$$\lim_{N_n \rightarrow \infty} P \left(\sup_{t \in [0, 1]} \Gamma_{N_n}(t) \geq \eta \right) = P \left(\sup_{t \in [0, 1]} \Gamma(t) \geq \eta \right) \square \quad (9)$$

Construction du test

Pour la construction du test, nous prenons comme région de rejet

$$\left\{ \sup_{t \in [0,1]} \Gamma_{N_n}(t) \geq \eta \right\}$$

où pour un α donné dans $[0, 1]$, le seuil η_α est déterminé par :

$$P \left(\sup_{t \in [0,1]} \Gamma(t) \geq \eta_\alpha \right) \leq \alpha. \quad (10)$$

Le tableau suivant présente la répartition de la loi de $\sup_{t \in [0,1]} \Gamma(t)$ pour les fonctions $\psi(t) = t^\rho$ avec $\rho = 1, \frac{3}{2}, \frac{5}{2}, \frac{7}{2}$ et la fonction $g_n(t) = t^{0.75}$.

Pour α donné, on lit η_α tel que $P \left(\sup_{t \in [0,1]} \Gamma(t) \geq \eta_\alpha \right) = \alpha$.

$g_n(t) = t^{0.75}$	η_α			
$\psi(t) = t^\rho$	$\alpha = 1\%$	$\alpha = 5\%$	$\alpha = 10\%$	$\alpha = 15\%$
$\rho = 1$	1.22952	0.89089	0.70427	0.61690
$\rho = 3/2$	0.57135	0.40679	0.33460	0.28987
$\rho = 5/2$	0.14116	0.09638	0.07810	0.06513
$\rho = 7/2$	0.03507	0.02338	0.01872	0.01565

3. Application et Discussion

Dans cette partie nous proposons une application du test étudié ci-dessus aux données concernant le nombre de naissances enregistrées à Liverpool entre l'année 1960 et l'année 1982 (Simpkin et Downham [13]). Au travers de ces naissances, on s'intéresse à l'hypospadias, maladie rare issue d'une malformation congénitale. Il est admis (Simpkin et Downham [13]) que le nombre de malformations pour une année donnée i , suit une loi de Poisson. En effet, si la proportion d'une telle malformation durant l'année i (soit λ) est supposée constante, et si le nombre total de naissances durant la même année est n_i , alors le nombre de telles malformations pour l'année i est une variable aléatoire X_i de loi approximativement de Poisson de paramètre $n_i \lambda$.

Nous avons appliqué le test décrit au paragraphe 2, pour différentes fonctions de pénalisation.

La figure 1 montre que la fonction g_n la plus adaptée à ces données est la fonction décrite dans le paragraphe 2 et au sous paragraphe construction du test, et qui a servi à construire η_α , à savoir, la fonction $t^{0.75}$.

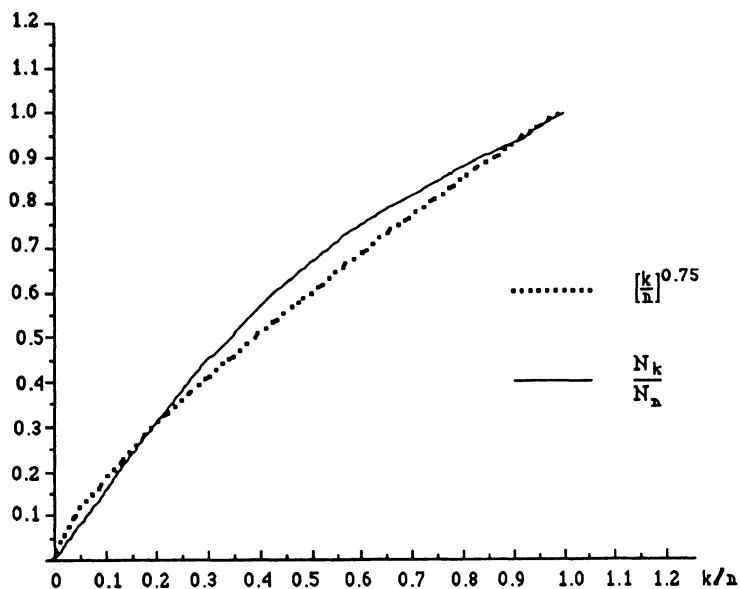


FIGURE 1
Choix de $g_n(t)$

La figure 2 représente les trajectoires des processus du rapport de vraisemblance pénalisée pour les différentes fonctions de pénalisation $\psi(t) = t^\rho$ avec $\rho = 1, \frac{3}{2}, \frac{5}{2}, \frac{7}{2}$.

Le tableau 1 représente les résultats obtenus dans lesquels $\Gamma_{N_n} \left(\frac{N_k}{N_n} \right)_\rho$ désignent (pour différentes valeurs de ρ) les rapports des vraisemblances maximales, et les valeurs soulignées sont évaluées au point \hat{k} donné par : $\hat{k} = \arg \max_{1 < k < n} \Gamma_{N_n} \left(\frac{N_k}{N_n} \right)$.

Dans la figure 2 comme dans le tableau 1, on remarque une zone de rupture $([9, 11])$ instable.

Simpkin et Downham [13] montrent que l'estimateur \hat{k} de l'instant de rupture k obtenu par la méthode du maximum de vraisemblance sous l'hypothèse alternative est obtenu pour $\hat{k} = 11$. Ils considèrent que le logarithme du rapport des vraisemblances maximales non pénalisées (à une constante près) se comporte asymptotiquement comme une variable du χ^2 à 2 degrés de liberté.

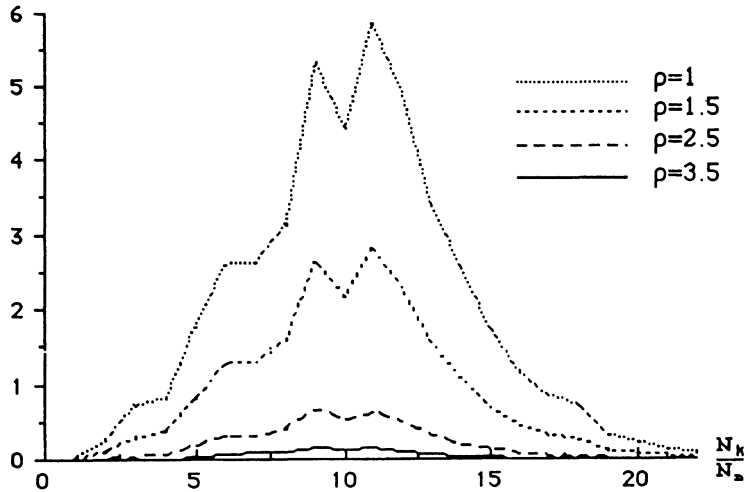


FIGURE 2
 $\Gamma_{N_n}(\frac{N_k}{N_n})$ pour différentes valeurs de ρ

La figure 3 donne le taux de prévalence (i.e. la proportion $\frac{x_i}{n_i}$ de nouveaux nés ayant la malformation) en fonction de l'année et montre clairement une zone de rupture se situant dans l'intervalle [8, 12].

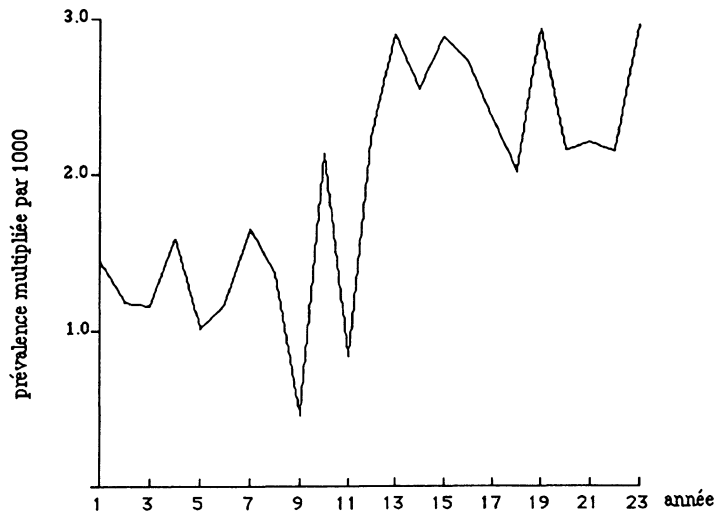


FIGURE 3
 Taux de prévalence observés pour l'hypospadias

TABLEAU 1

i	année	n_i	x_i	$\Gamma_{N_n} \left(\frac{N_i}{N_n} \right)_{\rho=1}$	$\Gamma_{N_n} \left(\frac{N_i}{N_n} \right)_{\rho=3/2}$	$\Gamma_{N_n} \left(\frac{N_i}{N_n} \right)_{\rho=5/2}$	$\Gamma_{N_n} \left(\frac{N_i}{N_n} \right)_{\rho=7/2}$
1	1960	17943	26				
2	1961	18663	22	0.244320	0.084040	0.009944	0.001177
3	1962	18165	21	0.700719	0.282966	0.046144	0.007525
4	1963	17626	28	0.804799	0.357793	0.070717	0.013977
5	1964	17866	18	1.741452	0.823825	0.184367	0.041260
6	1965	16217	19	2.575014	1.260893	0.302326	0.072489
7	1966	15099	25	2.808473	1.299042	0.322179	0.079905
8	1967	13970	19	3.110927	1.555277	0.388726	0.097158
9	1968	13269	6	5.287970	2.626470	<u>0.647947</u>	<u>0.159848</u>
10	1969	12662	27	4.412562	2.157191	0.515566	0.123220
11	1970	12055	10	<u>5.841753</u>	<u>2.786259</u>	0.633837	0.144190
12	1971	10791	24	4.910265	2.270024	0.485156	0.103689
13	1972	9696	28	3.384643	1.508231	0.299487	0.059469
14	1973	8652	22	2.563952	1.096274	0.200418	0.036640
15	1974	7987	23	1.694556	0.691273	0.115037	0.019144
16	1975	7394	20	1.136998	0.439815	0.065810	0.009847
17	1976	7241	17	0.838731	0.304390	0.040091	0.005280
18	1977	6962	14	0.709352	0.238547	0.026977	0.003051
19	1978	7161	21	0.324268	0.098633	0.009125	0.000844
20	1979	6946	15	0.217749	0.058099	0.004136	0.000294
21	1980	6821	15	0.127708	0.028249	0.001382	0.000068
22	1981	6989	15	0.064456	0.010144	0.000251	0.000006
23	1982	6784	20				

En effet, on constate dans cette zone des écarts considérables par rapport aux écarts situés en dehors de cette zone. L'estimation ponctuelle proposée par Simpkin et Downham [13], n'envisage pas une éventuelle rupture avant ou après l'instant 11. En effet, la méthode proposée est basée sur la fixation de l'instant de rupture, estimé en 11, c'est-à-dire, que leur statistique de test est conditionnée par l'instant k .

Nous avons appliqué le test étudié au paragraphe 2, à diverses fonctions de pondération, et le test détecte systématiquement une rupture significative située entre les instants 9 et 11. Ceci permet d'envisager un intervalle de confiance [9, 11] pour l'instant k de rupture.

4. Modèle Bayésien

Dans cette partie, pour l'approche bayésienne, nous proposons le modèle suivant :

$$\begin{aligned} X_i &\sim \mathcal{P}(n_i \lambda_1) & \text{si } i = 1, \dots, k \\ X_i &\sim \mathcal{P}(n_i \lambda_2) & \text{si } i = k + 1, \dots, n \end{aligned}$$

Le test étudié ici est le suivant :

$$H_0 : k = n \quad (\text{pas de rupture } \lambda_1 = \lambda_2)$$

Contre

$$H_1 : 1 \leq k \leq n - 1$$

On suppose ici k , λ_1 et λ_2 indépendants.

Le test proposé utilise l'analyse bayésienne informelle, basée sur la loi a postériori du paramètre k de l'instant de rupture (Diaz [4]). L'avantage de cette procédure est qu'elle ne requiert pas d'asymptotique.

4.1 Procédure de détection

La fonction de vraisemblance est donnée par :

$$L(x_1, \dots, x_n | \lambda_1, k, \lambda_2) \propto \begin{cases} e^{-\lambda_1 N_n} \lambda_1^{s_n} & \text{si } k = n \\ e^{-\lambda_1 N_k - \lambda_2 N_k^*} \lambda_1^{s_k} \lambda_2^{s_k^*} & \text{si } k \neq n \end{cases}$$

où

$$s_n = \sum_{i=1}^n x_i, \quad s_k = \sum_{i=1}^k x_i, \quad s_k^* = \sum_{i=k+1}^n x_i$$

Pour les paramètres λ_1 et λ_2 , nous avons choisi les lois a priori (impropres) : $f(\lambda_j) = \frac{1}{\lambda_j}$, $j = 1, 2$. Ce choix est d'une part motivé par des raisons techniques, et

d'autre part, conforme à ce qui est utilisé couramment dans la littérature (Diaz [4], Raftery et Akman [10], Smith [14]).

Pour le paramètre k , nous avons choisi les lois a priori suivantes :

$$\begin{aligned}\nu_0(k) &= p\mathbf{1}_{\{k=n\}} + \frac{1-p}{n-1}\mathbf{1}_{\{1 \leq k \leq n-1\}} \\ \nu_1(k) &= p\mathbf{1}_{\{k=n\}} + \frac{p(1-p)^k}{1-(1-p)^{n-1}}\mathbf{1}_{\{1 \leq k \leq n-1\}} \\ \nu_2(k) &= p\mathbf{1}_{\{k=n\}} + \frac{C_{n-1}^k p^k (1-p)^{n-k}}{1-(1-p)^{n-1}}\mathbf{1}_{\{1 \leq k \leq n-1\}}\end{aligned}$$

La loi ν_0 a été utilisée par Diaz [4] et les lois ν_1 et ν_2 sont obtenues à partir des lois (modifiées) géométrique et binomiale respectivement.

Les lois a posteriori pour le paramètre k sont (pour $j = 0, 1, 2$) :

$$\pi_j(k | x_1, \dots, x_n) \propto \begin{cases} \nu_j(k) \frac{\gamma(s_n)}{N_n^{s_n}} & \text{si } k = n \\ \nu_j(k) \frac{\gamma(s_k)\gamma(s_k^*)}{N_k^{s_k}(N_k^*)^{s_k^*}} & \text{si } k \neq n \end{cases}$$

où $\gamma(\cdot)$ désigne la fonction *Gamma*, $s_k = \sum_{i=1}^k x_i$, $s_k^* = \sum_{i=k+1}^n x_i$.

En effet, les lois a posteriori précédentes s'obtiennent de la formule classique de Bayes, et de l'indépendance de λ_1 , k et λ_2 . A savoir,

$$\pi_j(k | x_1, \dots, x_n) = \frac{\nu_j(k) \int_0^\infty \int_0^\infty L(x_1, \dots, x_n | \lambda_1, k, \lambda_2) f(\lambda_1) f(\lambda_2) d\lambda_1 d\lambda_2}{f_j(x_1, \dots, x_n)} \quad 1$$

avec

$$f_j(x_1, \dots, x_n) = \sum_{k=1}^n \nu_j(k) \int_0^\infty \int_0^\infty L(x_1, \dots, x_n | \lambda_1, k, \lambda_2) f(\lambda_1) f(\lambda_2) d\lambda_1 d\lambda_2.$$

Nous avons alors,

1. si $k = n$,

$$\int_0^\infty L(x_1, \dots, x_n | \lambda_1, n) f(\lambda_1) d\lambda_1 = \left(\prod_{i=1}^n \frac{n_i^{x_i}}{x_i!} \right) \left(\int_0^\infty \lambda_1^{s_n-1} e^{-\lambda_1 N_n} d\lambda_1 \right)$$

¹ en fait pour $k = n$, L ne dépend que de λ_1 , ce que nous n'avons pas fait apparaître dans ces formules pour ne pas alourdir l'écriture.

2. si $k \neq n$,

$$\int_0^\infty \int_0^\infty L(x_1, \dots, x_n | \lambda_1, k, \lambda_2) f(\lambda_1) f(\lambda_2) d\lambda_1 d\lambda_2$$

$$= \left(\prod_{i=1}^n \frac{n_i^{x_i}}{x_i!} \right) \left(\int_0^\infty \lambda_1^{s_k-1} e^{-\lambda_1 N_k} d\lambda_1 \right) \left(\int_0^\infty \lambda_2^{s_k^*-1} e^{-\lambda_2 N_k^*} d\lambda_2 \right)$$

4.2 Résultats et commentaires

TABLEAU 2

i	$\pi_0(i)$ $p = 0.1$	$\pi_0(i)$ $p = 0.5$	$\pi_0(i)$ $p = 0.9$	i	$\pi_0(i)$ $p = 0.1$	$\pi_0(i)$ $p = 0.5$	$\pi_0(i)$ $p = 0.9$
1	$2.06 \cdot 10^{-11}$	$2.06 \cdot 10^{-11}$	$2.06 \cdot 10^{-11}$	13	$1.68 \cdot 10^{-4}$	$1.68 \cdot 10^{-4}$	$1.68 \cdot 10^{-4}$
2	$8.35 \cdot 10^{-11}$	$8.35 \cdot 10^{-11}$	$8.35 \cdot 10^{-11}$	14	$8.39 \cdot 10^{-6}$	$8.39 \cdot 10^{-6}$	$8.39 \cdot 10^{-6}$
3	$6.66 \cdot 10^{-10}$	$6.66 \cdot 10^{-10}$	$6.66 \cdot 10^{-10}$	15	$1.87 \cdot 10^{-7}$	$1.87 \cdot 10^{-7}$	$1.87 \cdot 10^{-7}$
4	$4.66 \cdot 10^{-10}$	$4.66 \cdot 10^{-10}$	$4.66 \cdot 10^{-10}$	16	$1.48 \cdot 10^{-8}$	$1.48 \cdot 10^{-8}$	$1.48 \cdot 10^{-8}$
5	$1.79 \cdot 10^{-8}$	$1.79 \cdot 10^{-8}$	$1.79 \cdot 10^{-8}$	17	$4.58 \cdot 10^{-9}$	$4.58 \cdot 10^{-9}$	$4.58 \cdot 10^{-9}$
6	$3.29 \cdot 10^{-7}$	$3.29 \cdot 10^{-7}$	$3.29 \cdot 10^{-7}$	18	$4.40 \cdot 10^{-9}$	$4.40 \cdot 10^{-9}$	$4.40 \cdot 10^{-9}$
7	$2.52 \cdot 10^{-7}$	$2.52 \cdot 10^{-7}$	$2.52 \cdot 10^{-7}$	19	$3.09 \cdot 10^{-10}$	$3.09 \cdot 10^{-10}$	$3.09 \cdot 10^{-10}$
8	$1.69 \cdot 10^{-6}$	$1.69 \cdot 10^{-6}$	$1.69 \cdot 10^{-6}$	20	$2.21 \cdot 10^{-10}$	$2.21 \cdot 10^{-10}$	$2.21 \cdot 10^{-10}$
9	$1.35 \cdot 10^{-2}$	$1.35 \cdot 10^{-2}$	$1.35 \cdot 10^{-2}$	21	$1.66 \cdot 10^{-10}$	$1.66 \cdot 10^{-10}$	$1.66 \cdot 10^{-10}$
10	$6.80 \cdot 10^{-4}$	$6.80 \cdot 10^{-4}$	$6.80 \cdot 10^{-4}$	22	$2.15 \cdot 10^{-10}$	$2.15 \cdot 10^{-10}$	$2.15 \cdot 10^{-10}$
11	<u>$9.24 \cdot 10^{-1}$</u>	<u>$9.24 \cdot 10^{-1}$</u>	<u>$9.24 \cdot 10^{-1}$</u>	23	$6.76 \cdot 10^{-11}$	$6.08 \cdot 10^{-10}$	$5.47 \cdot 10^{-9}$
12	$6.19 \cdot 10^{-2}$	$6.19 \cdot 10^{-2}$	$6.19 \cdot 10^{-2}$				

On constate que la probabilité a postérieure est maximale pour $i = 11$. Pour $i \in \{9, \dots, 12\}$, la somme des probabilités a postérieure est nettement supérieure à 0.95, ce qui permet d'interpréter [9, 12] comme un intervalle de confiance a postérieure de niveau supérieur à 95 % pour l'instant de rupture.

Comme dans le cas précédent et pour $p \leq 0.5$ les probabilités a postérieure sont maximales pour $i = 11$ et de même, on peut interpréter [9, 12] comme un intervalle

TABLEAU 3

i	$\pi_1(i)$ $p = 0.1$	$\pi_1(i)$ $p = 0.5$	$\pi_1(i)$ $p = 0.9$	i	$\pi_1(i)$ $p = 0.1$	$\pi_1(i)$ $p = 0.5$	$\pi_1(i)$ $p = 0.9$
1	$5.94 \cdot 10^{-11}$	$2.09 \cdot 10^{-8}$	$3.78 \cdot 10^{-2}$	13	$1.36 \cdot 10^{-4}$	$4.15 \cdot 10^{-5}$	$3.07 \cdot 10^{-7}$
2	$2.16 \cdot 10^{-10}$	$4.23 \cdot 10^{-8}$	$1.53 \cdot 10^{-2}$	14	$6.13 \cdot 10^{-6}$	$1.04 \cdot 10^{-6}$	$1.53 \cdot 10^{-9}$
3	$1.55 \cdot 10^{-9}$	$1.69 \cdot 10^{-7}$	$1.22 \cdot 10^{-2}$	15	$1.23 \cdot 10^{-7}$	$1.16 \cdot 10^{-8}$	$3.43 \cdot 10^{-12}$
4	$9.77 \cdot 10^{-10}$	$5.90 \cdot 10^{-8}$	$8.52 \cdot 10^{-4}$	16	$8.79 \cdot 10^{-9}$	$4.59 \cdot 10^{-10}$	$2.71 \cdot 10^{-14}$
5	$3.38 \cdot 10^{-8}$	$1.13 \cdot 10^{-6}$	$3.28 \cdot 10^{-3}$	17	$2.44 \cdot 10^{-9}$	$7.08 \cdot 10^{-11}$	$8.37 \cdot 10^{-16}$
6	$5.59 \cdot 10^{-7}$	$1.04 \cdot 10^{-5}$	$6.02 \cdot 10^{-3}$	18	$2.11 \cdot 10^{-9}$	$3.40 \cdot 10^{-11}$	$8.05 \cdot 10^{-17}$
7	$3.85 \cdot 10^{-7}$	$3.99 \cdot 10^{-6}$	$4.61 \cdot 10^{-4}$	19	$1.33 \cdot 10^{-10}$	$1.20 \cdot 10^{-12}$	$5.66 \cdot 10^{-19}$
8	$2.33 \cdot 10^{-6}$	$1.34 \cdot 10^{-5}$	$3.10 \cdot 10^{-4}$	20	$8.61 \cdot 10^{-11}$	$4.28 \cdot 10^{-13}$	$4.05 \cdot 10^{-20}$
9	$1.67 \cdot 10^{-2}$	$5.34 \cdot 10^{-2}$	$2.47 \cdot 10^{-1}$	21	$5.81 \cdot 10^{-11}$	$1.61 \cdot 10^{-13}$	$3.04 \cdot 10^{-21}$
10	$7.57 \cdot 10^{-4}$	$1.34 \cdot 10^{-3}$	$1.24 \cdot 10^{-3}$	22	$6.76 \cdot 10^{-11}$	$1.04 \cdot 10^{-13}$	$3.93 \cdot 10^{-22}$
11	<u>$9.26 \cdot 10^{-1}$</u>	<u>$9.14 \cdot 10^{-1}$</u>	$1.69 \cdot 10^{-1}$	23	$7.96 \cdot 10^{-11}$	$5.60 \cdot 10^{-8}$	<u>$5.06 \cdot 10^{-1}$</u>
12	$5.59 \cdot 10^{-2}$	$3.06 \cdot 10^{-2}$	$1.13 \cdot 10^{-3}$				

de confiance a postérieur de niveau supérieur à 95 % pour l’instant de rupture. Pour $p > 0.5$, la probabilité a postérieur de non rupture est supérieure à 0.50; ceci montre que pour cette famille de lois, la procédure ne détecte pas systématiquement une rupture.

L’examen du tableau 4 montre qu’on retrouve pour cette famille de lois, les mêmes résultats que ceux obtenus pour la première famille de lois. Les résultats obtenus sur les données relatives à l’hypospadias montrent que la première et la dernière famille de lois sont en accord avec la première procédure; alors que ceux obtenus pour la deuxième famille de lois indiquent l’importance du choix de la loi a priori pour le paramètre de l’instant de rupture.

5. Conclusion

Chacune des procédures possède ses avantages et ses inconvénients. La première procédure paraît plus appropriée ici car les données sont nombreuses et par conséquent, l’asymptotique semble naturelle. Malheureusement, celle-ci nécessite le calcul (très long en temps machine) des seuils asymptotiques en plus du choix des fonctions de pénalisation. La méthode bayésienne, d’un point de vue pratique,

TABLEAU 4

i	$\pi_2(i)$ $p = 0.1$	$\pi_2(i)$ $p = 0.5$	$\pi_2(i)$ $p = 0.9$	i	$\pi_2(i)$ $p = 0.1$	$\pi_2(i)$ $p = 0.5$	$\pi_2(i)$ $p = 0.9$
1	$1.31 \cdot 10^{-6}$	$6.50 \cdot 10^{-16}$	$1.27 \cdot 10^{-25}$	13	$8.54 \cdot 10^{-7}$	$1.19 \cdot 10^{-4}$	$6.61 \cdot 10^{-3}$
2	$6.20 \cdot 10^{-6}$	$2.76 \cdot 10^{-14}$	$4.87 \cdot 10^{-23}$	14	$3.05 \cdot 10^{-9}$	$3.83 \cdot 10^{-6}$	$1.91 \cdot 10^{-3}$
3	$3.66 \cdot 10^{-5}$	$1.47 \cdot 10^{-12}$	$2.33 \cdot 10^{-20}$	15	$4.03 \cdot 10^{-12}$	$4.57 \cdot 10^{-8}$	$2.05 \cdot 10^{-4}$
4	$1.35 \cdot 10^{-5}$	$4.87 \cdot 10^{-12}$	$6.97 \cdot 10^{-19}$	16	$1.55 \cdot 10^{-14}$	$1.58 \cdot 10^{-9}$	$6.40 \cdot 10^{-5}$
5	$2.08 \cdot 10^{-4}$	$6.75 \cdot 10^{-10}$	$8.69 \cdot 10^{-16}$	17	$1.88 \cdot 10^{-16}$	$1.72 \cdot 10^{-10}$	$6.26 \cdot 10^{-5}$
6	$1.20 \cdot 10^{-3}$	$3.51 \cdot 10^{-8}$	$4.07 \cdot 10^{-13}$	18	$5.57 \cdot 10^{-18}$	$4.60 \cdot 10^{-11}$	$1.50 \cdot 10^{-4}$
7	$2.34 \cdot 10^{-4}$	$6.15 \cdot 10^{-8}$	$6.41 \cdot 10^{-12}$	19	$9.17 \cdot 10^{-21}$	$6.81 \cdot 10^{-13}$	$2.00 \cdot 10^{-5}$
8	$3.27 \cdot 10^{-4}$	$7.75 \cdot 10^{-7}$	$7.27 \cdot 10^{-10}$	20	$1.09 \cdot 10^{-22}$	$7.32 \cdot 10^{-14}$	$1.94 \cdot 10^{-5}$
9	$4.51 \cdot 10^{-1}$	$9.60 \cdot 10^{-3}$	$8.11 \cdot 10^{-5}$	21	$8.69 \cdot 10^{-25}$	$5.23 \cdot 10^{-15}$	$1.24 \cdot 10^{-5}$
10	$3.28 \cdot 10^{-3}$	$6.29 \cdot 10^{-4}$	$4.78 \cdot 10^{-5}$	22	$5.67 \cdot 10^{-27}$	$3.07 \cdot 10^{-16}$	$6.59 \cdot 10^{-6}$
11	<u>$5.40 \cdot 10^{-1}$</u>	<u>$9.32 \cdot 10^{-1}$</u>	<u>$6.37 \cdot 10^{-1}$</u>	23	$7.31 \cdot 10^{-7}$	$1.65 \cdot 10^{-10}$	$7.75 \cdot 10^{-5}$
12	$3.69 \cdot 10^{-3}$	$5.73 \cdot 10^{-2}$	$3.52 \cdot 10^{-1}$				

est plus simple à mettre en œuvre. L'inconvénient de cette procédure, comme on l'a remarqué, c'est que les résultats dépendent très fortement du choix de la loi a priori du paramètre de rupture.

Remerciements

Nous tenons à remercier les référés pour leurs remarques très judicieuses qui nous ont permis d'améliorer la rédaction, ainsi que Monsieur le Professeur Bernard BRU (Université PARIS V) pour ses précieux conseils.

Références

- [1] Bhattacharya P.K. and Frierson F.JR. (1981). A nonparametric control chart for detecting small disorders. *Ann.Statist.* **9** 544-554.
- [2] Billingsley P. (1968). *Convergence of Probability Measures*. Wiley, New York.

- [3] Chernoff H. and Zacks S.(1964). Estimating the current mean of a normal distribution which is subjected to change in time. *Ann. Math. Statist.* **35** 999-1018.
- [4] Diaz J. (1982). Bayesian detection of a change of scale parameter in sequences of independent Gamma random variables. *Journal of Econometrics* **19** 23-29.
- [5] Ghorbanzadeh D. (1995). Un test de détection de rupture de la moyenne dans un modèle gaussien. *Revue de la Statistique Appliquée* **XLIII(2)** 67-76.
- [6] Ghorbanzadeh D. et Picard D. (1992). Etude Asymptotique et Pratique du comportement de deux tests de détection de Rupture. *Statistique et Analyse des Données*. **Vol 16**. 63-84.
- [7] Hawkins D.M. (1977). Testing a sequence of observations for a shift in location. *J. Amer. Statist. Assoc.* **72** 180-186.
- [8] Hinkley D.V. and Hinkley E.A. (1970). Inference about the change point in a sequence of binomial random variables. *Biometrika*. **57** 477-488.
- [9] Hinkley D.V. (1971). Inference about the change point from cumulative sum test. *Biometrika*. **58** 509-523.
- [10] Raftery A.E. and Akman V.E. (1986). Bayesian analysis of a Poisson process with a change-point. *Biometrika*. **73** 85-89.
- [11] Sen A. and Srivastava M.S. (1975). On tests for detecting change in mean. *Ann. Statist.* **3** 90-108.
- [12] Shorack G.R. and Wellner J.A. (1986). *Empirical Processes With Applications to Statistics*. J. Wiley, New York.
- [13] Simpkin J.M. and Downham D.Y. (1988). Testing for a change-point in registry data with an example on Hypospadias. *Statistics in Medicine*. **7** 387-393.
- [14] Smith A.F.M. (1975). A bayesian approach to inference about a change-point in a sequence of random variables. *Biometrika*. **62** 407-416.
- [15] Worsley K.J. (1986). Confidence regions and tests for a change-point in a sequence of exponential family random variables. *Biometrika*. **73** 455-464.