Une Procedure asymptotique de détection de rupture pour des observations à incréments indépendants.

D.GHORBANZADEH
http ://www.cnam.fr/maths/Membres/ghorbanzadeh/

CNAM

17 juin 2008

Plan

Introduction

Modèle statistique

- Modèle de détection
- Comportements des observations sous l'hypothèse $\mathbf{H_0}$
- Statistique du test pénalisée aux bords
- Comportements des observations sous l'hypothèse $\mathbf{H_0}$
- Construction du test

Applications

- cas HSU (haemolytic uraemic syndrome)
- cas SIDA

Plan

Modèle statistique

On considère une suite de variables aléatoires $X_1, \ldots, X_n$ avec des incréments indépendants qui sont susceptibles de changer de loi après les k premières observations.

Modèle statistique

On considère une suite de variables aléatoires $X_1, \ldots, X_n$ avec des incréments indépendants qui sont susceptibles de changer de loi après les k premières observations.

avant la rupture

les incréments $X_1, \ldots, X_k - X_{k-1}$ sont indépendants et suivent une loi de Poisson $\mathcal{P}(\lambda)$

Modèle statistique

On considère une suite de variables aléatoires $X_1, \ldots, X_n$ avec des incréments indépendants qui sont susceptibles de changer de loi après les k premières observations.

avant la rupture

les incréments $X_1, \ldots, X_k - X_{k-1}$ sont indépendants et suivent une loi de Poisson $\mathcal{P}(\lambda)$

après la rupture

les incréments $X_{k+1} - X_k, \ldots, X_n - X_{n-1}$ sont indépendants et suivent une loi de Poisson $\mathcal{P}(\lambda + \frac{\delta}{\sqrt{n}})$

Modèle de détection

On propose le problème de test suivant :

$$\begin{cases} \mathbf{H_0} : & \delta = 0 \\ \mathbf{H_1} : & \delta \neq 0 \end{cases}$$

Statistique du test

Pour tester l'existance le point de rupture, on propose la statistique du test :

statistique du test

$$R_n(k) = \sqrt{n}\left(\frac{X_n - X_k}{n-k} - \frac{X_k}{k}\right) \qquad (1)$$

◂ $R_n(k)$

sous l'hypothèse $\mathbf{H_0}$

$X_i \sim \mathcal{P}(i\,\lambda)$

$\mathtt{Cov}(X_i, X_j) = \min(i, j)\,\lambda$

sous l'hypothèse $\mathbf{H_0}$

$X_i \sim \mathcal{P}(i\,\lambda)$

$\mathtt{Cov}(X_i, X_j) = \min(i,j)\,\lambda$

comportement de la statistique du test sous $\mathbf{H_0}$

$\mathbb{E}[R_n(k)] = 0$

$$\forall k_1 \leq k_2 \;\; \mathtt{Cov}\left(R_n(k_1)\,,\, R_n(k_2)\right) = \frac{\lambda}{\frac{k_2}{n}\,(1-\frac{k_1}{n})}$$

$R_n(k)$

$$\mathtt{Var}[R_n(k)] = \frac{\lambda}{\frac{k}{n}\,(1-\frac{k}{n})}$$

statistique du test pénalisée aux bords

Soit φ une fonction continue par morceaux sur $[0,1]$ satisfaisant :

$$\int_0^1 \left(\frac{\varphi(t)}{\sqrt{t}}\right)^2 dt < \infty$$

statistique du test pénalisée

$$R^{\star}_{n,\varphi}(k) = \varphi(\frac{k}{n})\,\varphi(1-\frac{k}{n})\,R_n(k) \tag{2}$$

◂ statistique pénalisée

comportement de la statistique du test pénalisée

$\forall\, k_1 \leq k_2$

$$\texttt{Cov}(R^\star_{n,\varphi}(k_1)\,,\, R^\star_{n,\varphi}(k_2)) = \frac{\varphi(\frac{k_1}{n})\,\varphi(1-\frac{k_1}{n})\,\varphi(\frac{k_2}{n})\,\varphi(1-\frac{k_2}{n})}{\dfrac{k_2}{n}\,(1-\dfrac{k_1}{n})}\lambda \quad (3)$$

$$\texttt{Var}(R^\star_{n,\varphi}(k)) = \left(\frac{\varphi(\frac{k}{n})\,\varphi(1-\frac{k}{n})}{\sqrt{\dfrac{k}{n}\,(1-\dfrac{k}{n})}}\right)^2 \lambda$$

fonction de pénalisation

Si l'on prend $\varphi(t) = t$, l'équation (2) devient :

statistique pénalisée pour la fonction $\varphi(t) = t$

$$\mathbb{K}_n(k) = n^{-3/2}(k\,X_n - n\,X_k) \tag{4}$$

équation (2)

fonction de pénalisation

Si l'on prend $\varphi(t) = t$, l'équation (2) devient :

statistique pénalisée pour la fonction $\varphi(t) = t$

$$\mathbb{K}_n(k) = n^{-3/2}(k\,X_n - n\,X_k) \tag{4}$$

▸ équation (2)

Définition

Pour tout estimateur consistant $\hat{\lambda}_n$ de λ, soit $\mathbb{K}^\star_n(k) = \frac{1}{\sqrt{\hat{\lambda}_n}}\,\mathbb{K}_n(k)$.

On définit $\{\mathbb{K}^\star_n(t)\,;\ t \in [0,1]\}$ comme le processus interpolé linéaire du processus $\{\mathbb{K}^\star_n(k)\,;\ 1 < k < n\}$.

comportement de $\mathbb{K}_n^\star$ sous l'hypothèse $\mathbf{H_0}$

Théorème

sous l'hypothèse $\mathbf{H_0}$, $\mathbb{K}_n^\star(t)$ converge au sens de la convergence en loi des marginales de dimension finie vers le pont Brownien $B(t) = W(t) - t\,W(1)$ où $\{W(t)\,;\, t \in [0,1]\}$ représente le mouvement Brownien.

Preuve

Par l'équation (3)on a :

$$\forall t_1 \leq t_2 \;\; \mathrm{cov}(\mathbb{K}_n(t_1)\,,\,\mathbb{K}_n(t_2)) = \frac{[nt_1]}{n}\,(1 - \frac{[nt_2]}{n})\,\lambda \qquad (5)$$

Or,

$$\lim_{n\to\infty} \mathtt{Cov}(\mathbb{K}_n^\star(t_1)\,,\,\mathbb{K}_n^\star(t_2)) = t_1\,(1 - t_2)$$

D'autre part, on a : $t_1 \leq t_2, \;\; \mathtt{Cov}(B(t_1), B(t_2)) = t_1\,(1 - t_2)$

Proposition

Sous $\mathbf{H_0}$, pour tout $\eta > 0$, on a :

$$\lim_{n \to \infty} \mathbb{P}_{H_0}\left(\sup_{t \in [0,1]} |\mathbb{K}_n^\star(t)| \geq \eta \right) = \mathbb{P}\left(\sup_{t \in [0,1]} |B(t)| \geq \eta \right) \tag{6}$$

Proposition

Sous $\mathbf{H_0}$, pour tout $\eta > 0$, on a :

$$\lim_{n\to\infty} \mathbb{P}_{H_0}\left(\sup_{t\in[0,1]} |\mathbb{K}_n^\star(t)| \geq \eta\right) = \mathbb{P}\left(\sup_{t\in[0,1]} |B(t)| \geq \eta\right) \qquad (6)$$

Conséquence

Soit $\hat{\tau}_n = \mathrm{Arg} \sup_{t\in[0,1]} |\mathbb{K}_n^\star(t)|$ et $\tau = \mathrm{Arg} \sup_{t\in[0,1]} |B(t)|$. Alors,

$$\hat{\tau}_n \xrightarrow[n\to\infty]{\mathrm{P}} \tau \qquad (7)$$

sous l'hypothèse $\mathbf{H_1}$

$$\mathbb{E}[X_i] = \begin{cases} i\,\lambda & \text{si } 1 \leq i \leq k \\ k\,\lambda + (i-k)\,(\lambda + \frac{\delta}{\sqrt{n}}) & \text{si } k+1 \leq i \leq n \end{cases}$$

pour tout k_1 et k_2 ,

$$\mathtt{Cov}(X_{k_1}, X_{k_2}) = \begin{cases} k\lambda + (\min(k_1,k_2) - k)(\lambda + \frac{\delta}{\sqrt{n}}) & \mathrm{s}i \min(k_1,k_2) \geq k \\ \min(k_1,k_2)\,\lambda & \mathrm{s}inon \end{cases}$$

comportement de $\mathbb{K}_n^\star$ sous l'hypothèse $\mathbf{H_1}$

Théorème

sous l'hypothèse $\mathbf{H_1}$, $\mathbb{K}_n^\star$ converge au sens de la convergence en loi des marginales de dimension finie vers le processus process $B_\tau^\star$ défini :

$$B_\tau^\star(t) = (t \wedge \tau)\,(1 - t \vee \tau)\,\frac{\delta}{\sqrt{\lambda}} + B(t) \tag{8}$$

où $t \wedge \tau = \inf\{t, \tau\}$ et $t \vee \tau = \sup\{t, \tau\}$

Zone de rejet

Nous prenons comme région de rejet : où pour α donné dans $[0,1]$, le seuil η_α est déterminé par :

$$P\,(\sup_{t\in[0,1]} |B(t)| \geq \eta_\alpha) = 2\sum_{j=1}^{\infty}(-1)^{j+1}\, e^{-2\,j^2\,\eta_\alpha^2} \tag{9}$$

Kolmogorov

On note que le test étudié a le même niveau asymptotique que le test classic de **Kolmogorov**

Puissance du test

puissance du test

$$\mathrm{p}(\tau) = P\left(\sup_{t\in[0,1]} |B^{\star}_{\tau}(t)| \geq \eta_{\alpha} \right) \tag{10}$$

Plan

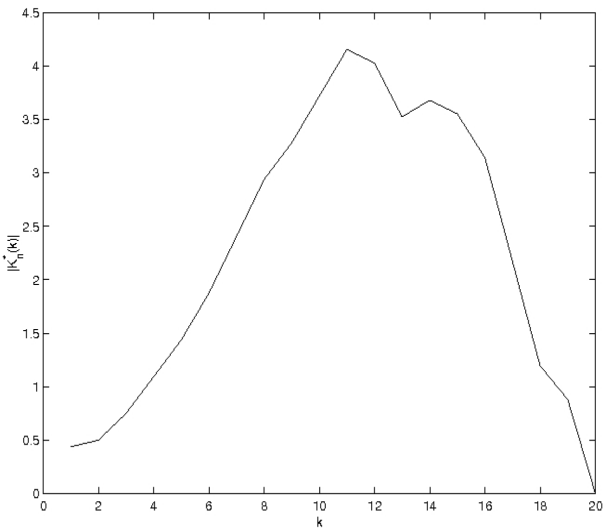
HSU (haemolytic uraemic syndrome)

Les données représentent le nombre de 'HUS' à Birmingham et Newcastle de 1970 à 1989.

référence

R. W. West & R. T. Ogden. (1997). Continuous-time estimation of a change-point in a Poisson process. Journ. Statist. Comput. Simul. 56, pp 293-302.

HSU (Birmingham)



graphe de $|\mathbb{K}_n^{\star}(k)|$ pour les données de Birmingham .

estimation (Birmingham)

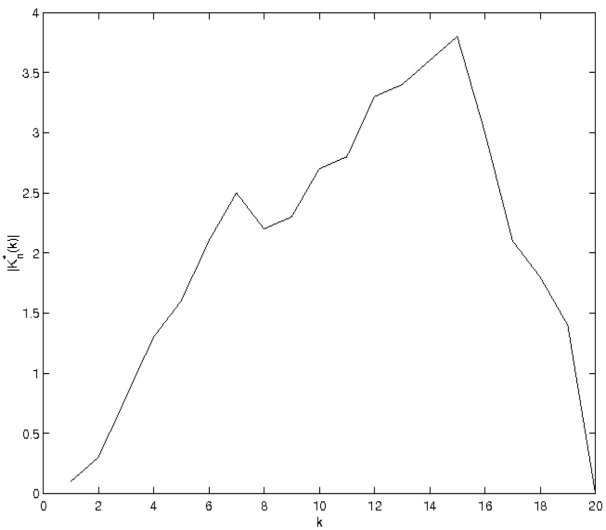
$\hat{\lambda}_n = \dfrac{x_n}{n} = 5.65$

le maximum est atteint au point : $k = 11$, ce qui correspond à la valeur de statistique : $|\mathbb{K}_n^\star(11)| = 4.15$.

Pour $\alpha = 5\%$, la valeur critique $\eta_\alpha = 1.36$, on rejette donc l'hypothèse reject H_0.

$\delta = |\mathbb{K}_n(11)| = 9.87$, ce qui donne une puissance ègale à 41%.

HSU (Newcastle)



graphe de $|\mathbb{K}_n^\star(k)|$ pour les données de Newcastle .

estimation (Newcastle)

$\hat{\lambda}_n = \frac{x_n}{n} = 5.$
le maximum est atteint au point : $k = 15$, ce qui correspond à la valeur de statistique : $|\mathbb{K}_n^{\star}(11)| = 1.6994$.
Pour $\alpha = 5\%$, la valeur critique $\eta_\alpha = 1.36$, on rejette donc l'hypothèse reject H_0.
$\delta = |\mathbb{K}_n(15)| = 8.50$, ce qui donne une puissance ègale à 21%.

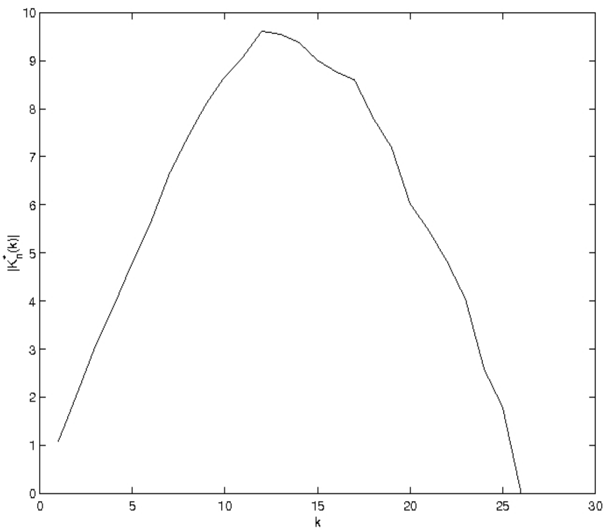
SIDA (cuba)

Les données représentent le nombre de 'SIDA' au Cuba de 1986 à 1992.

référence

The HIV-AIDS Epidemic in Cuba. Internal Report. Departamento de Epidemiologia, Sanatorio Santiago de Las Vegas. Cuba 1999.

SIDA (cuba)



graphe de $|\mathbb{K}_n^\star(k)|$ pour les données SIDA au Cuba .

estimation (SIDA)

$\hat{\lambda}_n = \frac{x_n}{n} = 32.15$
le maximum est atteint au point : $k = 12$, ce qui correspond à la valeur de statistique : $|\mathbb{K}_n^\star(11)| = 9.61$.
Pour $\alpha = 5\%$, la valeur critique $\eta_\alpha = 1.36$, on rejette donc l'hypothèse reject H_0.
$\delta = |\mathbb{K}_n(15)| = 54.49$, ce qui donne une puissance ègale à 98%.