

Astérisque

JEAN DESHAYES

DOMINIQUE PICARD

Application aux tests de rupture de régression

Astérisque, tome 68 (1979), p. 73-98

<http://www.numdam.org/item?id=AST_1979__68__73_0>

© Société mathématique de France, 1979, tous droits réservés.

L'accès aux archives de la collection « Astérisque » (<http://smf4.emath.fr/Publications/Asterisque/>) 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/>

APPLICATION AUX TESTS DE RUPTURE DE RÉGRESSION

Jean DESHAYES, Dominique PICARD.

I - VITESSE EXPONENTIELLE DES TESTS utilisés pour détecter une rupture de régression .

Nous reprenons brièvement l'exposé du problème et les tests proposés dans [1].
Soit le modèle :

$$y_k = x_k' \beta_k + u_k \quad k = 1, 2, \dots, n$$

où x_k désigne le vecteur sur lequel on régresse.

β_k est le vecteur des paramètres de régression au temps k .

u_k est l'erreur : les u_k sont supposées indépendantes et de loi normale centrée de variance σ_k^2 .

L'hypothèse à tester est qu'il n'y a pas de rupture :

$$H_0 = \{\beta_1 = \beta_2 = \dots = \beta_n \text{ et } \sigma_1 = \sigma_2 = \dots = \sigma_n\} .$$

En fait, nous n'envisagerons ici les propriétés des tests que dans le cas d'une contrehypothèse du type $\{\beta_1 = \beta_2 = \dots = \beta_r \neq \beta_{r+1} = \dots = \beta_n \text{ et } \sigma_1 = \sigma_2 = \dots = \sigma_n\}$ et du point de vue asymptotique ; nous supposons observer le processus Y_t sur $[0, 1]$ aux instants $\frac{k}{n}$ ($k = 1, 2, \dots, n$) donc en des instants de plus en plus rapprochés. Sous l'alternative, nous supposons par conséquent qu'il existe $t_0 \in]0, 1[$ tel que pour :

$$\begin{aligned} \frac{k}{n} < t_0, \quad y_{k,n} \text{ suit } \mathcal{N}(x_{k,n}' \beta_0 ; \sigma^2) \\ \frac{k}{n} \geq t_0, \quad y_{k,n} \text{ suit } \mathcal{N}(x_{k,n}' \beta_1 ; \sigma^2), \beta_0 \neq \beta_1 \end{aligned}$$

Aucun caractère d'optimalité ne se dégage de chacun des tests proposés, ceci étant dû, d'une part à l'étendue de l'alternative, d'autre part au fait que t_0 est mal estimable, même si on restreint l'alternative comme ci-dessous.

Les tests proposés dans [1] sont construits à partir des résidus récursifs (ceux-ci sont obtenus à partir des résidus habituels par orthogonalisation)

EXPOSÉ 5

$$\omega_k = \frac{y_k - x'_k b_{k-1}}{\sqrt{1 + x'_k (X'_{k-1} X_{k-1})^{-1} x_k}}$$

avec $b_k = (X'_k X_k)^{-1} X'_k Y_k$ et $X_k = \begin{pmatrix} x'_1 \\ x'_2 \\ \vdots \\ x'_k \end{pmatrix}$, $Y_k = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_k \end{pmatrix}$

Nous allons étudier du point de vue asymptotique, le test basé sur les sommes cumulées de résidus puis celui basé sur les sommes cumulées des carrés de résidus. Pour les deux tests, nous distinguerons les cas σ connu et σ inconnu et nous nous limiterons au cas unidimensionnel (rupture de moyenne).

I - 1) Test construit sur les sommes cumulées de résidus

I - 1. : 1) σ connu

Les variables $y_{k,n}$ suivent une loi normale $\mathcal{N}(m_1, \sigma^2)$ pour $k < nt_0$ et une loi normale $\mathcal{N}(m_2, \sigma^2)$ pour $k \geq nt_0$.

Sous H_0 , on a $t_0 = 0$ ou $m_1 = m_2$

sous H_1 , on a $t_0 \in]0, 1[$ et $m_1 - m_2 = d \neq 0$.

Les résidus récurrents valent : $\omega_k = \frac{y_k - \bar{y}^{k-1}}{\sqrt{1 + \frac{1}{k-1}}}$, $k = 2, \dots, n$ où

$$\bar{y}^{k-1} = \frac{\sum_{i=1}^{k-1} y_i}{k-1}.$$

Le test basé sur les sommes cumulées de résidus consiste à rejeter l'hypothèse H_0 s'il existe t dans $]0, 1[$ tel que $|s_{n-1}(t)| > g(t)$ avec $g(t) = at + \frac{a}{2}$ pour un $a > 0$.

$s_{n-1}(\cdot)$ désigne la ligne polygonale partant de 0 et joignant les points $(\frac{k-1}{n-1}, \frac{\sum_{i=2}^k \omega_i}{n-1})$, $k = 2, 3, \dots, n$.

I.1.1. a) Calcul du niveau

Sous H_0 , les ω_k sont des variables aléatoires indépendantes $\mathcal{N}(0, \sigma^2)$, leur transformée de Cramer est $h(u) = \frac{u^2}{2\sigma^2}$.

D'après le chapitre n° IV ([6]), on a :

$$\forall a > 0, \frac{1}{n-1} \text{Log } P_{H_0}(|s_{n-1}(t)| > at + \frac{a}{2}) \longrightarrow - \inf_t t \cdot h\left(\frac{at + \frac{a}{2}}{t}\right) = - \frac{a^2}{2\sigma^2}$$

ce qui signifie en particulier que ce test a toujours un niveau exponentiel.

I.1.1. b) Calcul de l'erreur de 2ème espèce

Sous H_1 , les $y_{k,n}$ suivent une loi normale de variance σ^2 et de moyenne égale à m_1 pour $k < nt_0$ et égale à m_2 pour $k \geq nt_0$. Alors les résidus récurrents sont indépendants et de loi $\mathcal{N}(E\omega_k, \sigma^2)$ où :

$$E \omega_k = 0 \quad \text{pour } k < nt_0$$

$$E \omega_k = \frac{\frac{[nt_0]}{k-1} \cdot (m_2 - m_1)}{\sqrt{1 + \frac{1}{k-1}}} = \frac{\frac{[nt_0]}{n} \cdot d}{\sqrt{\frac{k}{n} \left(\frac{k-1}{n} - \frac{1}{n}\right)}} \sim \frac{t_0 \cdot d}{\frac{k}{n}} \quad \text{pour } k \geq nt_0.$$

D'après le chapitre n° IV ([6]), l'erreur de 2ème espèce se calcule ainsi :

$$\begin{aligned} & \lim_{n \rightarrow \infty} \frac{1}{n} \text{Log } P_{H_1}(|s_{n-1}(t)| < at + \frac{a}{2}, \forall t) \\ &= \lim_{n \rightarrow \infty} \frac{1}{n} \text{Log } P_{H_0}(s_{n-1}(t) < at + \frac{a}{2} - t_0 |d| \cdot \text{Log } \frac{t}{t_0} \cdot 1_{t \geq t_0}, \forall t) \\ &= -\omega(g^*) \quad \text{en utilisant les notations du chapitre n° IV ([6], §III.3°).} \end{aligned}$$

Pour avoir une décroissance exponentielle, il faut déjà que :

$$\inf_t g(t) \text{ soit négatif. En particulier } g'(t) = a - \frac{t_0 |d|}{t} \cdot 1_{t \geq t_0}$$

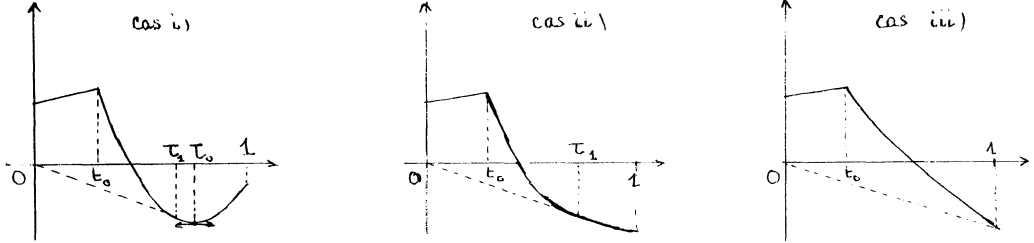
doit être négatif en t_0 , ce qui nécessite : $a < |d|$.

• Si $t_0 < \frac{a}{|d|} < 1$, alors g' s'annule au point $\tau_0 = \frac{t_0 |d|}{a} \in [t_0, 1]$,

EXPOSÉ 5

ce qui donne : $\inf_t g(t) = g(\tau_0) = a \left(t_0 \frac{|d|}{a} + \frac{1}{2} - t_0 \frac{|d|}{a} \text{Log} \frac{|d|}{a} \right)$

◦ Si $\frac{a}{|d|} < t_0$, alors $\inf_t g(t) = g(1) = a \left(\frac{3}{2} + t_0 \frac{|d|}{a} \text{Log} t_0 \right)$. Trois cas seront à distinguer pour le calcul explicite :



◦ Si $|d| \leq a$, pas de décroissance exponentielle de l'erreur.

◦ Si $t_0 \leq \frac{a}{|d|} < 1$ et $\frac{a}{|d|} \cdot \frac{1}{2t_0} + 1 + \text{Log} \frac{a}{|d|} \geq 0$, pas de décroissance exponentielle de l'erreur.

◦ Cas iii) $t_0 \leq \frac{a}{|d|} < 1$ et $\frac{a}{|d|} \cdot \frac{1}{2t_0} + 1 + \text{Log} \frac{a}{|d|} < 0$, la deuxième inégalité peut s'écrire $\frac{a}{|d|} < \mu(t_0)$ avec $\mu(t_0)$ solution de $\frac{\mu}{2t_0} + 1 + \text{Log} \mu = 0$.

Cette valeur $\mu(t_0)$ est supérieure à t_0 si et seulement si t_0 est inférieur à $e^{-3/2}$. Par conséquent, le cas i) ne se produit pas si t_0 est supérieur à $e^{-3/2}$. Notons τ_1 la racine comprise entre t_0 et $t_0 \frac{|d|}{a}$ de l'équation :

$g'(t) = \frac{g(t)}{t}$; τ_1 est nécessairement inférieur à t_0 dans ce cas puisque la fonction g est convexe sur $[t_0, 1]$. On obtient alors :

$$\begin{aligned} \omega(g^*) &= \int_0^1 h(g^*(t)) dt = \frac{g(\tau_1)^2}{2\sigma^2 \tau_1} + \int_{\tau_1}^{t_0} \frac{g'(t)^2}{2\sigma^2} dt \\ &= t_0 \frac{d^2}{\sigma^2} \left\{ e^{-(1 + \frac{a}{2t_0|d|})} + \frac{a}{|d|} \text{Log} \frac{a}{|d|} + \frac{1}{2t_0} \cdot \frac{a^2}{d^2} \right\} . \end{aligned}$$

◦ Si $t_0 \geq e^{-3/2}$ et $t_0 \leq \frac{a}{|d|} \leq \frac{2}{3} t_0 \text{Log} \frac{1}{t_0}$, alors $g(1)$ est positif et par conséquent, pas de décroissance exponentielle.

◦ si $t_0 < e^{-3/2}$ et $\frac{a}{|d|} < t_0$ ou bien si $t_0 \geq e^{-3/2}$ et $\frac{a}{|d|} < \frac{2}{3} t_0 \text{Log} \frac{1}{t_0}$, on est amené à distinguer deux cas suivant que τ_1 est inférieur à 1 ou non ;

TESTS DE RUPTURE

le cas ii) correspond à $\tau_1 \leq 1$, ce qui se produit si $t_o < e^{-1}$ et $\frac{a}{|d|} < 2t_o (\text{Log } \frac{1}{t_o} - 1)$, cette dernière borne étant plus petite que $\frac{2}{3} t_o \text{Log } \frac{1}{t_o}$ si $t_o \geq e^{-3/2}$ et plus grande que t_o dans le cas contraire. Le cas iii) correspond à $\tau_1 > 1$, ce qui se produit pour $t_o > e^{-1}$ ou bien $t_o < e^{-1}$ et $\frac{a}{|d|} > 2t_o (\text{Log } \frac{1}{t_o} - 1)$.

o Cas ii) $t_o \leq e^{-3/2}$ et $\frac{a}{|d|} < t_o$

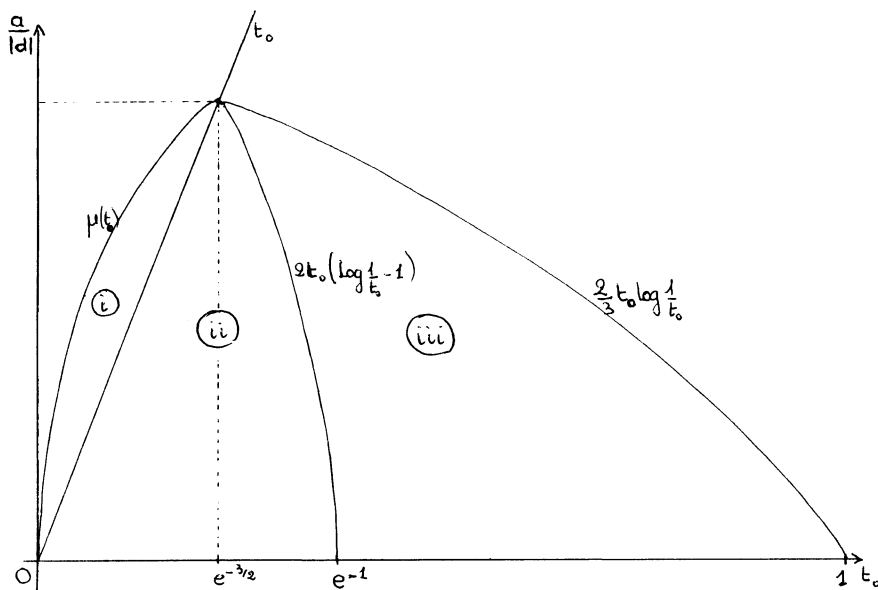
ou $e^{-3/2} \leq t_o \leq e^{-1}$ et $\frac{a}{|d|} \leq 2t_o (\text{Log } \frac{1}{t_o} - 1)$

alors $\omega(g^*) = \frac{g(\tau_1)^2}{2\sigma^2 \tau_1} + \int_{\tau_1}^1 \frac{g'(t)^2}{2\sigma^2} dt = \frac{t_o d^2}{\sigma^2} \left\{ \frac{a^2}{d^2} \cdot \frac{1}{t_o} + e^{-(1 + \frac{a}{2t_o |d|})} + \frac{a}{|d|} \cdot \text{Log } t_o - t_o \right\}$

o Cas iii) $t_o > e^{-1}$ et $\frac{a}{|d|} < \frac{2}{3} t_o \text{Log } \frac{1}{t_o}$

ou $e^{-3/2} \leq t_o \leq e^{-1}$ et $2t_o (\text{Log } \frac{1}{t_o} - 1) < \frac{a}{|d|} < \frac{2}{3} t_o \text{Log } \frac{1}{t_o}$

alors $\omega(g^*) = \frac{g(1)^2}{2\sigma^2} = \frac{t_o d^2}{\sigma^2} \left\{ \frac{9}{8t_o} \cdot \frac{a^2}{d^2} + \frac{3a}{2|d|} \text{Log } t_o + \frac{t_o}{2} \cdot (\text{Log } t_o)^2 \right\}$.



I.1.2) σ inconnu : niveau

La région de rejet est la même que la précédente si ce n'est que σ^2 est ici estimé par l'estimateur classique des moindres carrés

$$\hat{\sigma}^2 = \frac{1}{n-1} \sum_{k=1}^n (y_k - \bar{y}^n)^2 = \frac{1}{n-1} \sum_{k=2}^n \omega_k^2 \quad (\text{voir [1]}) :$$

$$D = \left\{ \int_0^1 t \in [0,1] \text{ tel que } \frac{|s_{n-1}(t)|}{\hat{\sigma}} > bt + \frac{b}{2} \right\} .$$

Du point de vue fonctionnel, D peut s'écrire :

$$D = \left\{ f \text{ telle que } \exists t \in [0,1] , \frac{|f(t)|}{\left(\int_0^1 f'(t)^2 dt \right)^{1/2}} > bt + \frac{b}{2} \right\} .$$

C'est un exemple d'ensemble fonctionnel où la théorie exposée au chapitre n°IV ([6]) ne donne aucune précision : en effet, l'intérieur de D est vide et sa fermeture contient 0 de sorte que $W(\overset{\circ}{D}) = +\infty$ et $W(\bar{D}) = 0$.

o. étant donné l'aspect homogène de la contrainte, il est évident que si D n'est pas vide, alors $\bar{D}$ contient la fonction identiquement nulle.

- si $b < \frac{\sqrt{2}}{2}$, alors il existe τ_2 compris entre 0 et 1 tel que $\sqrt{\tau_2} > b\tau_2 + \frac{b}{2}$ alors toute fonction f linéaire sur $[0, \tau_2]$ et constante ensuite appartient à D $f(t) = C \cdot t \cdot 1_{t < \tau_2} + C \cdot \tau_2 \cdot 1_{t > \tau_2}$ donc D ne sera pas vide.

- si $b \geq \frac{\sqrt{2}}{2}$, alors D est vide : il est clair ^{effet} en $\sqrt{}$ qu'il suffit de s'intéresser aux fonctions linéaires sur $[0, \tau_2]$ car si une fonction f appartient à D , il existe τ tel que $\frac{f(\tau)}{\left(\int_0^1 f'(t)^2 dt \right)^{1/2}} > b\tau + \frac{b}{2}$, alors la fonction

linéaire : $\frac{f(\tau)}{\tau} \cdot t \cdot 1_{t \leq \tau} + f(\tau) \cdot 1_{t > \tau}$ est à fortiori dans D . D étant vide,

$\overset{\circ}{D}$ et $\bar{D}$ le sont aussi.

o. Même dans le cas où $b < \frac{\sqrt{2}}{2}$, on constate immédiatement que $\overset{\circ}{D}$ est vide car dans tout ε -tube construit autour de f , il existe des fonctions qui "oscillent" trop pour appartenir à D .

Néanmoins, $P_{H_0}(D)$ décroît exponentiellement comme le montre l'encadrement suivant dans le cas où $b < \frac{1}{\sqrt{2}}$:

$$\bullet \forall \varepsilon > 0, \quad P_{H_0}(D) < P_{H_0}\left\{\exists t, \frac{|S_{n-1}(t)|}{\sigma(1-\varepsilon)} > bt + \frac{b}{2}\right\} + P_{H_0}\{\hat{\sigma} < \sigma(1-\varepsilon)\}$$

$$\text{or : } \frac{1}{n} \log P_{H_0}\left\{\exists t, \frac{|S_{n-1}(t)|}{\sigma} > (1-\varepsilon)bt + \frac{(1-\varepsilon)b}{2}\right\} \longrightarrow -b^2(1-\varepsilon)^2$$

$$\frac{1}{n} \log P_{H_0}\{\hat{\sigma} < \sigma(1-\varepsilon)\} \longrightarrow -\frac{(1-\varepsilon)^2 - 1 - \log(1-\varepsilon)^2}{2}$$

$$\lim_n \frac{1}{n} \log P_{H_0}(D) < -\inf\{b^2(1-\varepsilon)^2, \frac{(1-\varepsilon)^2 - 1 - \log(1-\varepsilon)^2}{2}\}$$

$$\bullet \forall \varepsilon > 0, \quad P_{H_0}(D) > P_{H_0}\left\{\exists t, \frac{|S_{n-1}(t)|}{\sigma(1+\varepsilon)} > bt + \frac{b}{2}\right\} - P_{H_0}\{\hat{\sigma} > \sigma(1+\varepsilon)\}$$

$$\frac{1}{n} \log P_{H_0}\left\{\exists t, \frac{|S_{n-1}(t)|}{\sigma} > (1+\varepsilon)bt + \frac{(1+\varepsilon)b}{2}\right\} \longrightarrow -b^2(1+\varepsilon)^2$$

$$\frac{1}{n} \log P_{H_0}\{\hat{\sigma} > \sigma(1+\varepsilon)\} \longrightarrow -\frac{(1+\varepsilon)^2 - 1 - \log(1+\varepsilon)^2}{2}$$

$$\sup_{\varepsilon} \frac{(1+\varepsilon)^2 - 1 - \log(1+\varepsilon)^2}{2(1+\varepsilon)^2} = \frac{1}{2} > b^2$$

$$\text{donc } \exists \varepsilon_0 > 0 \text{ tel que } b^2(1+\varepsilon_0)^2 < \frac{(1+\varepsilon_0)^2 - 1 - \log(1+\varepsilon_0)^2}{2}$$

$$\text{on en déduit que : } \lim_n \frac{1}{n} \log P_{H_0}(D) > -b^2(1+\varepsilon_0)^2.$$

I.2. Test construit sur les sommes des carrés de résidus - σ connu

I.2.1) Niveau

La région de rejet s'écrit : $\{\exists k \in \{2, 3, \dots, n\} \text{ tel que}$

$$\frac{1}{(n-1)\sigma^2} \sum_{i=2}^k \omega_i^2 > \frac{k-1}{n-1} + c\}$$
 soit finalement : $\{\exists t \in [0, 1] \text{ tel que } ss_{n-1}(t) > t+c\}$

avec $c > 0$, où $ss_{n-1}(\cdot)$ désigne la ligne polygonale joignant l'origine 0 et

$$\text{les points } \left\{\frac{k-1}{n-1}; \frac{1}{(n-1)\sigma^2} \sum_{i=2}^k \omega_i^2\right\} \text{ pour } k = 2, \dots, n.$$

EXPOSÉ 5

On a vu dans le chapitre n° IV([6]) que :

$$\frac{1}{n} \text{Log } P_{H_0}(\exists t, ss_{n-1}(t) > t+c) \longrightarrow - \inf_t t \cdot h_0\left(\frac{t+c}{t}\right)$$

où h_0 est la transformée de Cramer d'une variable suivant une loi χ_1^2 soit :

$$h_0(x) = \frac{x-1 - \text{Log } x}{2}, \quad x > 1.$$

Donc : $\forall c > 0, \frac{1}{n} \text{Log } P_{H_0}(\exists t, ss_{n-1}(t) > t+c) \longrightarrow - \frac{c - \text{Log}(1+c)}{2} = - \mathcal{H}_0(c)$.

On remarque, c'est la même vitesse que pour le test admettant pour région de rejet seulement $\{\frac{1}{(n-1)\sigma^2} \sum_{i=2}^n \omega_i^2 > 1+c\}$.

I.2.2) Erreur de 2ème espèce

Sous H_1 , les variables ω_k sont indépendantes et suivant des lois normales de variance σ^2 et d'espérance :

$$E \omega_k = \frac{t_o d}{\sqrt{\frac{k}{n}(\frac{k}{n} - 1)}} \mathbf{1}_{k \geq nt_o}.$$

On a donc à évaluer : $\frac{1}{n} \text{Log } P_{H_1}(\forall k, \frac{1}{(n-1)\sigma^2} \sum_{i=2}^k \omega_i^2 \leq \frac{k-1}{n-1} + c)$. Ce

problème n'a pas été traité explicitement dans le chapitre n° IV([6]), car la suite (ω_k) , $k = 2, \dots, n$ n'est plus stationnaire. Toutefois, on a de même (cf.[2]) :

$$\frac{1}{n} \text{Log } P_{H_1}(\forall k, \frac{1}{(n-1)\sigma^2} \sum_{i=2}^k \omega_i^2 \leq \frac{k-1}{n-1} + c) \rightarrow - \inf_{f \in D} \int_0^1 h_t(f'(t)) dt$$

où $D = \{f \text{ telle que } \forall t \in [0,1], f(t) \leq t+c \text{ et } f(0) = 0\}$

et $h_t(.)$ désigne la transformée de Cramer de la variable $\frac{1}{\sigma^2} \omega_{\lfloor nt \rfloor}^2$, soit ici :

$$h_t(x) = \sup_u (xu + \frac{1}{2} \text{Log}(1-2u) - \frac{t_o^2 d^2}{\sigma^2 t^2} \cdot \frac{u}{1-2u} \cdot \mathbf{1}_{t \geq t_o})$$

On est amené à résoudre le problème suivant :

TESTS DE RUPTURE

Trouver $\inf_f \int_0^1 h_1(t, f'(t)) dt$ sous la contrainte :

$f(0) = 0$ et $f(t) \leq C$, $\forall t \in [0, 1]$ si on note $h_1(t, v) = h_t(v+1)$. On utilisera le fait que h_1 est décroissante en t et convexe en v . Ce problème est équivalent au suivant :

"Trouver $\inf_g \int_0^1 \int_0^{\gamma(t)} u(t, v) dv$ sous la contrainte : $\int_0^x g(t) dt \leq c$,

$\forall x \in [0, 1]$, avec :

$$u(t, v) = - \frac{\partial h_1}{\partial v}(t, v)$$

$$\gamma(t) \text{ étant défini par } u(t, \gamma(t)) = 0 \iff \gamma(t) = \frac{t_0 d^2}{\sigma^2 t^2} \cdot 1_{t > t_0}.$$

Remarque 1 : Il est évident que la décroissance de l'erreur ne sera pas exponentielle si la fonction $\gamma(\cdot)$ respecte la contrainte :

$$\int_{t_0}^1 \gamma(t) dt \leq c \text{ soit si } t_0(1-t_0) \frac{d^2}{\sigma^2} < c.$$

Pour $c \in]0, t_0(1-t_0) \frac{d^2}{\sigma^2} [$, on va exhiber la fonction g réalisant l'infimum.

Remarque 2 : A t fixé, la convexité de h_1 par rapport à v implique que $u(t, v)$ est une fonction décroissante de v et que les courbes de niveau $g_\lambda(t)$ définies par : $v = g_\lambda(t) \iff u(t, v) = \lambda > 0$ sont des fonctions décroissantes de t sur $[t_0, 1]$. On a :

$$g_\lambda(t) = \frac{-2\lambda}{1+2\lambda} + \frac{t_0^2 d^2}{\sigma^2 t^2} \cdot \frac{1}{(1+2\lambda)^2} \cdot 1_{t \geq t_0}.$$

Remarque 3 : Il suffit de considérer les fonctions g inférieures à γ car pour toute fonction g réalisant la contrainte, la fonction $\inf(g, \gamma)$ réalise aussi la contrainte (car $\gamma \geq 0$) et donne une intégrale plus petite.

EXPOSÉ 5

Nous serons amenés à distinguer deux cas selon que la courbe de niveau g_λ

réalisant $\int_0^1 g_\lambda(t) dt = c$ vérifie : $g_\lambda(t) > 0$ pour $t \in [t_0, 1]$ ou non .

Cas i) la courbe de niveau $g_\lambda(t) = \frac{-2\lambda}{1+2\lambda} + \frac{t_0^2}{\sigma^2} \frac{d^2}{t^2} \cdot \frac{1}{(1+2\lambda)^2} \cdot 1_{t > t_0}$ telle que $\int_0^1 g_\lambda(t) dt = c$, vérifie $g_\lambda \geq 0$ sur $[t_0, 1]$.

Alors montrons que g_λ est meilleure que toute fonction g inférieure à γ et réalisant $\int_0^1 g(t) dt \leq c$.

$$\begin{aligned} \int_0^1 dt \int_{g(t)}^{\gamma(t)} u(t, v) dv &= \int_0^1 dt \int_{g \vee g_\lambda(t)}^{\gamma(t)} u(t, v) dv + \int_0^1 dt \int_{g(t)}^{g_\lambda(t) \vee g} u(t, v) dv \\ &\geq \int_0^1 dt \int_{g \vee g_\lambda(t)}^{\gamma(t)} u(t, v) dv + \lambda \int_0^1 dt \int_{g(t)}^{g_\lambda(t) \vee g} dv \\ &= \int_0^1 dt \int_{g \vee g_\lambda(t)}^{\gamma(t)} u(t, v) dv + \lambda \left[\int_0^1 g \vee g_\lambda(t) dt - \int_0^1 g(t) dt \right] \\ &\geq \int_0^1 dt \int_{g \vee g_\lambda(t)}^{\gamma(t)} u(t, v) dv + \lambda \left[\int_0^1 g_\lambda \vee g(t) dt - \int_0^1 g_\lambda(t) dt \right] \\ &\geq \int_0^1 dt \int_{g \vee g_\lambda(t)}^{\gamma(t)} u(t, v) dv + \int_0^1 \int_{g_\lambda(t)}^{g(t) \vee g} u(t, v) dv \\ &= \int_0^1 dt \int_{g_\lambda(t)}^{\gamma(t)} u(t, v) dv . \end{aligned}$$

On est dans ce cas si $t_0 \geq \frac{1}{2}$

ou si $t_0 < \frac{1}{2}$ et $\frac{4t_0(1-2t_0) \frac{d^2}{\sigma^2}}{(1 + \sqrt{1 + \frac{4t_0^2 d^2}{\sigma^2}})^2} \leq c \leq t_0(1-t_0) \frac{d^2}{\sigma^2}$

on obtient : $\inf_{f \in D} \int_0^1 h_t(f'(t)) dt = \frac{1}{2} \{ \text{Log } \Lambda - \frac{\Lambda-1}{\Lambda} + \frac{t_0(1-t_0)d^2}{\sigma^2} \cdot (\frac{\Lambda-1}{\Lambda})^2 \}$

avec $\Lambda = 1 + 2\lambda = \frac{1 + \sqrt{1 + 4t_0(1-t_0) \frac{d^2}{\sigma^2}}}{2(1+c)}$

Cas ii) la courbe de niveau g_λ réalisant $\int_0^1 g_\lambda(t)dt = c$ est telle que $g_\lambda(1) < 0$. Nous allons montrer que la courbe réalisant l'infimum est $\tilde{g}_\lambda$ définie par :

$$\tilde{g}_\lambda(t) = g_\lambda(t)(1_{t < t_0} + 1_{t > t_0 \cap g_\lambda(t) > 0}), \lambda \text{ étant tel que } \int_0^1 \tilde{g}_\lambda(t)dt = c$$

Si g est une fonction réalisant l'infimum, nécessairement g réalise l'égalité dans la contrainte : $\int_0^x g(t)dt \leq c$ pour tout x ; posons alors :

$$\tau_g = \inf \left\{ \tau, \int_0^\tau g(t)dt = c \right\}.$$

Nécessairement g est nulle pour t supérieur à τ_g . En effet, sur $[\tau_g, 1]$, toute "arche de type $g \geq 0$ " sera précédée d'une ou plusieurs "arches de type $g \leq 0$ " de surface au moins égale car $\int_0^{\tau_g} g(t)dt = c$ et $\int_0^x g(t)dt \leq c$ pour tout x . Ces deux arches sont alors séparées par une courbe de niveau $g_{\lambda'}, \lambda' > 0$; par un raisonnement analogue au cas i), on obtient :

$$\int_{\tau_g}^1 dt \int_{g(t)}^{\gamma(t)} u(t,v)dv \geq \int_{\tau_g}^1 dt \int_0^{\gamma(t)} u(t,v)dv.$$

La fonction g étant optimale, elle est donc nulle sur $[\tau_g, 1]$.

Un argument analogue permet d'affirmer que g est nécessairement positive sur $[t_0, \tau_g]$.

Pour achever la démonstration, il suffit de faire la même démonstration que dans le cas i)

$$\begin{aligned} \int_0^1 dt \int_{g(t)}^{\gamma(t)} u(t,v)dv &= \int_0^1 dt \int_{g \vee \tilde{g}_\lambda}^{\gamma(t)} u(t,v)dv + \int_0^1 dt \int_g^{\tilde{g}_\lambda} u(t,v)dv \\ &\geq \int_0^1 dt \int_{g \vee \tilde{g}_\lambda}^{\gamma(t)} u(t,v)dv + \lambda \left(\int_0^1 g \vee \tilde{g}_\lambda(t)dt - \int_0^1 g(t)dt \right) \\ &= \int_0^1 dt \int_{g \vee \tilde{g}_\lambda}^{\gamma(t)} u(t,v)dv + \lambda \left(\int_0^1 g \vee \tilde{g}_\lambda(t)dt - \int_0^1 \tilde{g}_\lambda(t)dt \right) \end{aligned}$$

D'où

$$\int_0^1 dt \int_{g(t)}^{\gamma(t)} u(t, v) dv \geq \int_0^1 dt \int_{g_\lambda(t)}^{\gamma(t)} u(t, v) dv.$$

On est dans ce cas si $t_0 < \frac{1}{2}$ et $0 < c < \frac{4t_0(1-2t_0)\frac{d^2}{\sigma^2}}{(1+\sqrt{1+4t_0^2\frac{d^2}{\sigma^2}})^2}$

on obtient ainsi : $\inf_{f \in D} \int_0^1 h_t(f'(t)) dt = \frac{1}{2} \text{Log} \frac{1+\sqrt{1+4t_0^2\frac{d^2}{\sigma^2}}}{2} + 1 - \frac{t_0^2 d^2}{2\sigma^2}$

$$+ t_0 \frac{|d|}{\sigma} \cdot \frac{\Lambda^2 - 1}{\sqrt{\Lambda(\Lambda - 1)}} - \sqrt{1 + 4t_0^2 \frac{d^2}{\sigma^2}} + \frac{t_0 d^2}{2\sigma^2} \cdot \left(\frac{\Lambda - 1}{\Lambda}\right)^2 + \frac{t_0 |d|}{2\sigma} \cdot \text{Log} \frac{2t_0 \frac{|d|}{\sigma} + \sqrt{1 + 4t_0^2 \frac{d^2}{\sigma^2}}}{2\sqrt{\Lambda(\Lambda - 1)} + (2\Lambda - 1)}$$

avec $\Lambda = 1 + 2\lambda$ solution de : $\left[\Lambda^2 \frac{c}{t_0} - \frac{d^2}{2}\right]^2 - 4 \frac{d}{\sigma} \Lambda(\Lambda - 1) = 0$.

II - ÉTUDE D'UN AUTRE TEST POUR DÉCELER UNE RUPTURE DANS UN MODÈLE

Nous avons supposé que les $y_{k,n}$, $k = 1, \dots, n$ sont indépendantes et suivent une loi F_1 pour $k < nt_0$ et une loi F_2 pour $k \geq nt_0$. Notre problème est toujours de déterminer s'il y a une rupture effective ($t_0 \neq 0$, et $F_1 \neq F_2$) ou non. On a vu dans le paragraphe précédent qu'une des difficultés sous-jacentes était la mauvaise estimabilité de t_0 pour de telles observations. On va essayer de contourner cette difficulté en "randomisant" les instants d'observations.

Au lieu d'observer le phénomène à des instants $\frac{k}{n}$ fixes, on va tirer au sort les instants d'observation, par exemple avec une loi $\Pi \otimes^n$ sur $[0, 1]^n$, de sorte qu'on obtient un n -échantillon $(X_1, X_2, \dots, X_n)$ avec $X_j = Y_{U_j}$ où $(U_1, U_2, \dots, U_n)$ est un n -échantillon de la loi Π sur $[0, 1]$ et les variables $Y_{u_j, n}$ suivent la loi F_1 si $u_j < t_0$ et la loi F_2 si $u_j \geq t_0$. Si Π est une loi absolument continue par rapport à la mesure de Lebesgue sur $[0, 1]$ et si on note $\pi_0 = \Pi(U < t_0)$, alors $(X_1, X_2, \dots, X_n)$ est un n -échantillon de la loi mélange : $\pi_0 F_1 + (1 - \pi_0)F_2$ dont les deux premiers moments sont : $\pi_0 m_1 + (1 - \pi_0)m_2$ et $\sigma^2 + \pi_0(1 - \pi_0)d^2$ (on a noté m_1 et σ^2 les moments de F_1 , m_2 et σ^2 les moments de F_2 et $d = m_1 - m_2$).

TESTS DE RUPTURE

Sous cette forme, l'hypothèse naturelle à tester est :

$$H_0 = \{ \pi_0 (1 - \pi_0) d^2 = 0 \}$$

et la méthode des moments amène à comparer l'estimateur classique de la variance à σ^2 .

On va étudier ce test d'un point de vue asymptotique dans le cas très simple où $F_1 = \mathcal{N}(m_1, \sigma^2)$ et $F_2 = \mathcal{N}(m_2, \sigma^2)$. Pour faciliter l'exposition, on conservera la randomisation, sous la forme $\Pi^{\otimes n}$, des temps d'observation. En réalité, aussi bien au niveau théorique que pratique, on procédera un peu différemment (l'équivalence des deux procédures étant montrée en appendice) : n étant fixé, nous supposons disposer de $N(n)$ (avec $\frac{N(n)}{n} \xrightarrow{n \rightarrow \infty} \infty$) temps possibles d'observations disposés sur $[0, 1]$: $U_1, U_2, \dots, U_N$ et leurs observations associées $Y_{U_1}, \dots, Y_{U_N}$ avec Y_{U_i} suit la loi F_1 si $U_i < t_0$ et la loi F_2 si $U_i > t_0$; et on va effectuer un tirage uniforme et sans remise des n instants parmi les $N(n)$. L'arbitraire dans la loi Π se retrouvant ici dans le choix arbitraire des instants $U_1, U_2, \dots, U_N$ possibles.

On verra que la loi Π peut être choisie de façon plus ou moins appropriée et qu'on aura intérêt à utiliser l'éventuelle information a priori sur l'instant de rupture t_0 .

II.1. Définition du test

II.1.1) σ connu

La région de rejet sera du type : $D = \left\{ \frac{1}{(n-1)\sigma^2} \sum_{i=1}^n (X_i - \bar{X})^2 > 1+c \right\}$ avec $c > 0$ déterminé par le niveau α du test : $P_{H_0}(D) = \alpha$.

II.1.1) σ inconnu

Dans ce cas, on conserve le même principe de test mais on remplace σ^2 par une estimation faite à l'aide d'un deuxième échantillon de taille n : $V_1, V_2, \dots, V_n$ réalisé suivant la même procédure que $X_1, X_2, \dots, X_n$ mais les instants d'observation étant cette fois tiré selon une loi $\Pi_n^{\otimes n}$, Π_n' ayant

EXPOSÉ 5

pour support $[0, a_n]$ avec a_n : suite décroissante vers 0. Ce second échantillon est supposé indépendant du premier et forme un n -échantillon de la loi $\pi'_{n,0} F_1 + (1 - \pi'_{n,0}) F_2$: il est clair que pour $t_0 > 0$, $\pi'_{n,0}$ vaut 1 pour n assez grand. D'un point de vue plus pratique, si on présume que la rupture se produit après un instant t_1 ou avant un instant t_2 , on prend alors $\Pi'_n = \Pi'$ loi chargeant uniquement $[0, t_1]$ ou $[t_2, 1]$ respectivement.

On pourra donc tester l'hypothèse H_0 à l'aide de la région de rejet

$$\frac{\sum_{i=1}^n (X_i - \bar{X})^2}{\sum_{i=1}^n (V_i - \bar{V})^2} > 1 + c$$

avec $c > 0$ déterminé par le niveau du test.

II.2. Propriétés asymptotiques

Soit maintenant un n -échantillon $(X_1, X_2, \dots, X_n)$ de loi $F = F_X$, nous noterons :

$$\phi_n(t) = \int \dots \int e^{t \sum_{i=1}^n (x_i - \bar{x})^2} dF(x_1) \dots dF(x_n)$$

On suppose que : $\frac{1}{n} \text{Log } \phi_n(t)$ converge vers une fonction notée $\text{Log } \phi(t)$ en tout point sauf éventuellement à la frontière du domaine de définition de $\text{Log } \phi$. On suppose de plus que 0 est un point intérieur à ce domaine. Si on note $h(\cdot)$ La duale de Young de $\text{Log } \phi(\cdot)$, d'après le chapitre n° I ([3]), on a :

$$\lim_{n \rightarrow \infty} \frac{1}{n} \text{Log } P\left(\frac{\sum_{i=1}^n (X_i - \bar{X})^2}{(n-1)\sigma^2} > 1+c\right) = -h((1+c)\sigma^2), \quad \forall c > 0$$

$$\lim_{n \rightarrow \infty} \frac{1}{n} \text{Log } P\left(\frac{\sum_{i=1}^n (X_i - \bar{X})^2}{(n-1)\sigma^2} < 1+c\right) = -h((1+c)\sigma^2), \quad \forall c < 0$$

II.2.1) Application à des lois particulières

a) F est la loi $\mathcal{N}(m, \sigma^2)$

On a dans ce cas $\phi_n(t) = \left(\frac{1}{1 - 2\sigma^2 t}\right)^{n-1}$, ce qui donne :

TESTS DE RUPTURE

$$\lim_{n \rightarrow \infty} \frac{1}{n} \log P \left(\sum_{i=1}^n \frac{(X_i - \bar{X})^2}{(n-1)\sigma^2} > 1 + c \right) = -H_0(c), \quad c > 0$$

$$\text{avec} \quad H_0(c) = \frac{c - \log(1+c)}{2}.$$

b) F est la loi de Bernouilli de paramètre π_0

Cette loi F a l'intérêt de montrer que $\sum_{i=1}^n (X_i - \bar{X})^2$ ne se comporte pas forcément comme $\sum_{i=1}^n [X_i - E(X)]^2$; elle introduira d'autre part plus facilement la loi mélange de deux gaussiennes.

$$\phi_n(t) = \sum_{k=0}^n C_n^k \pi_0^k (1 - \pi_0)^{n-k} e^{nt \frac{k}{n} (1 - \frac{k}{n})}$$

Il est facile de montrer que tout t réel, la limite de $\frac{1}{n} \log \phi_{F,n}(t)$ existe et vaut :

$$\log \phi(t) = \alpha \log \frac{\pi_0}{\alpha} + (1-\alpha) \log \frac{1-\pi_0}{1-\alpha} + t \alpha(1-\alpha)$$

avec $\alpha = \alpha(t)$ égal à l'unique solution comprise entre 0 et 1 de l'équation :

$$\log \frac{1-\alpha}{\alpha} + \frac{\pi_0}{1-\pi_0} + t(1-2\alpha) = 0.$$

Toutefois, plutôt que de chercher la duale de Young de $\log \phi_F(\cdot)$, il est plus simple de calculer directement :

$$\begin{aligned} P \left(\sum_{i=1}^n (X_i - \bar{X})^2 < (n-1)\sigma^2(1+c) \right) & \quad (-1 < c < 0) \\ = P \left[\bar{X}(1-\bar{X}) < \frac{n-1}{n} \cdot \pi_0(1-\pi_0)(1+c) \right] \end{aligned}$$

Pour $0 < \pi_0(1-\pi_0)(1+c) < \frac{1}{4}$, notons u la plus petite racine de l'équation :

$$u(1-u) = \pi_0(1-\pi_0)(1+c).$$

On obtient alors, par application simple des théorèmes de grandes déviations sur les loi de Bernouilli (cf. par exemple [5]), en désignant par

$$\bar{\pi}_0 = \inf(\pi_0, 1 - \pi_0)$$

$$P_{\pi_0}(\bar{X}(1-\bar{X}) < (1-u)) = P_{\pi_0}(\bar{X} < u) + P_{\pi_0}(\bar{X} > 1-u) \\ \leq 2 P_{\pi_0}(\bar{X} < u)$$

$$\lim_{n \rightarrow \infty} \frac{1}{n} \text{Log } P_{\pi_0}(\bar{X}(1-\bar{X}) < u(1-u)) = \lim_{n \rightarrow \infty} \frac{1}{n} \text{Log } P_{\pi_0}(\bar{X} < u) \\ = u \text{Log } \frac{\bar{\pi}_0}{u} + (1-u) \text{Log } \frac{1-\bar{\pi}_0}{1-u}$$

Ce résultat appelle les deux remarques suivantes :

i) pour $c > 0$, $\sum_{i=1}^n (X_i - \bar{X})^2 > (n-1)\sigma^2(1+c)$ n'est pas une queue exponentielle

ii) pour $c < 0$

$$\lim_{n \rightarrow \infty} \frac{1}{n} \text{Log } P_{\pi_0} \left(\sum_{i=1}^n (X_i - \pi_0)^2 < n \pi_0 (1-\pi_0) (1+c) \right) \\ = \lim_{n \rightarrow \infty} \frac{1}{n} \text{Log } P_{\pi_0} \left(\bar{X} < \bar{\pi}_0 + \frac{c \bar{\pi}_0 (1-\bar{\pi}_0)}{1-2\bar{\pi}_0} \right) \\ = \left(\bar{\pi}_0 + c \frac{\bar{\pi}_0 (1-\bar{\pi}_0)}{1-2\bar{\pi}_0} \right) \text{Log } \frac{1}{1+c \frac{\bar{\pi}_0 (1-\bar{\pi}_0)}{(1-2\bar{\pi}_0)}} + \left(1-\bar{\pi}_0 - c \frac{\bar{\pi}_0 (1-\bar{\pi}_0)}{1-2\bar{\pi}_0} \right) \text{Log } \frac{1}{1-c \frac{\bar{\pi}_0 (1-\bar{\pi}_0)}{1-2\bar{\pi}_0}}$$

Les deux vitesses exponentielles sont strictement différentes pour

$\sum_{i=1}^n (X_i - \bar{X})^2$ et $\sum_{i=1}^n (X_i - EX)^2$ dans le cas de variables de Bernouilli :

pour $c < 0$, $\lim_{n \rightarrow \infty} \frac{1}{n} \text{Log } P(\sum_{i=1}^n (X_i - \bar{X})^2 < (n-1)\pi_0(1-\pi_0)(1+c)) >$

$$\lim_{n \rightarrow \infty} \frac{1}{n} \text{Log } P \left(\sum_{i=1}^n (X_i - \pi_0)^2 < (n-1)\pi_0(1-\pi_0)(1+c) \right)$$

c) F est le mélange de deux gaussiennes

$\mathcal{N}(m_1, \sigma^2)$ avec probabilité π_0 et $\mathcal{N}(m_2, \sigma^2)$ avec probabilité $(1-\pi_0)$.

On note $d = m_1 - m_2$ et $u(t) = \frac{t \cdot d^2}{1-2\sigma^2 t}$,

TESTS DE RUPTURE

$$\phi_n(t) = (1-2\sigma^2 t)^{-\frac{n-1}{2}} \sum_{k=0}^n C_n^k \pi_o^k (1-\pi_o)^{n-k} e^{n \cdot u(t) \cdot \frac{k}{n} (1 - \frac{k}{n})}$$

On déduit du paragraphe précédent que pour tout t inférieur à $\frac{1}{2\sigma^2}$, la fonction $\frac{1}{n} \text{Log } \phi_n(t)$ converge vers $\text{Log } \phi(t)$ définie par :

$$\log \phi(t) = -\frac{1}{2} \text{Log}(1-2\sigma^2 t) + \alpha \text{Log} \frac{\pi_o}{\alpha} + (1-\alpha) \text{Log} \frac{1-\pi_o}{1-\alpha} + u(t) \alpha(1-\alpha)$$

avec $\alpha = \alpha(t)$ solution de : $\text{Log} \frac{1-\alpha}{\alpha} + \frac{\pi_o}{1-\pi_o} + u(t)(1-2\alpha) = 0$.

On peut écrire sa duale de Young :

$$h(x) = x \tau + \frac{1}{2} \text{Log}(1-2\sigma^2 \tau) - \alpha \text{Log} \frac{\pi_o}{\alpha} - (1-\alpha) \text{Log} \frac{1-\pi_o}{1-\alpha} - u(\tau) \alpha(1-\alpha)$$

$$\text{avec } \tau = \frac{2x - \sigma^2 - \sqrt{\sigma^4 + 4\alpha(1-\alpha)d^2 x}}{4x \sigma^2}$$

et $\alpha(\tau)$ solution de : $\text{Log} \frac{\pi_o}{\alpha} + \frac{1-\alpha}{1-\pi_o} + (1-2\alpha) \frac{d^2}{2\sigma^2} \cdot \frac{2x - \sigma^2 - \sqrt{\sigma^4 + 4\alpha(1-\alpha)d^2 x}}{\sigma^2 + \sqrt{\sigma^4 + 4\alpha(1-\alpha)d^2 x}} = 0$.

II.3. Vitesse exponentielle des erreurs pour ce test

II.2.1. σ connu

On déduit des paragraphes précédents que pour : $0 < c < \pi_o(1-\pi_o) \frac{d^2}{\sigma^2}$,

$$\lim_{n \rightarrow \infty} \frac{1}{n} \text{Log } P_{H_0} \left(\sum_{i=1}^n \frac{(X_i - \bar{X})^2}{(n-1)\sigma^2} > 1+c \right) = -H_0(c)$$

$$H_0(c) = \frac{c - \text{Log}(1+c)}{2}$$

$$\lim_{n \rightarrow \infty} \frac{1}{n} \text{Log } P_{H_1} \left(\sum_{i=1}^n \frac{(X_i - \bar{X})^2}{(n-1)\sigma^2} < 1+c \right) = -H_1(c)$$

$$\text{avec } H_1(c) = \frac{c+1}{2} \left[1 - \frac{1 + \sqrt{1 + 4\alpha(1-\alpha) \frac{d^2}{\sigma^2} (c+1)}}{2(c+1)} \right] - \frac{\alpha d^2}{2\sigma^2} \left[\frac{2(c+1)}{1 + \sqrt{1 + 4\alpha(1-\alpha) \frac{d^2}{\sigma^2} (c+1)}} - 1 \right] +$$

$$\frac{1}{2} \text{Log} \frac{1 + \sqrt{1 + 4\alpha(1-\alpha) \frac{d^2}{\sigma^2} (c+1)}}{2(c+1)} - \text{Log} \frac{1-\pi_o}{1-\alpha}$$

EXPOSÉ 5

où $\alpha = \alpha(c)$ est la solution de :

$$\text{Log } \frac{\pi_0}{\alpha} \cdot \frac{1-\alpha}{1-\pi_0} + (1-2\alpha) \frac{d^2}{2\sigma^2} \cdot \left[\frac{2(c+1)}{1+\sqrt{1+4\alpha(1-\alpha)\frac{d^2}{\sigma^2}(c+1)}} - 1 \right] = 0 \quad .$$

On remarque que H_1 dépend de la loi Π choisie pour échantillonner les instants d'observations seulement par l'intermédiaire de $\pi_0 = \Pi(t < t_0)$; et il est facile de montrer que :

$$H_1(c)(\pi_0) \leq H_1(c)(\pi) \leq H_1(c)\left(\frac{1}{2}\right)$$

pour toute valeur de π située entre π_0 et $\frac{1}{2}$ et pour toute valeur de c positive et inférieure à $\pi_0(1-\pi_0)\frac{d^2}{\sigma^2}$. On constate ainsi qu'on aura intérêt à choisir la loi Π la plus simple possible mais en essayant de se rapprocher de la valeur $\pi_0 = \frac{1}{2}$: dans la mesure où l'on peut pressentir l'instant de la rupture éventuelle, il faudra faire autant d'observation avant qu'après. Un choix optimal mais théorique de Π_n serait une loi chargeant équitablement les intervalles $[0, a_n]$ et $[1-a_n, 1]$ avec $a_n \xrightarrow{n \rightarrow \infty} 0$. Plus pratiquement, une rupture présumée dans une plage $[t_1, t_2]$ conduira à prendre une loi Π chargeant équitablement les intervalles $[0, t_1]$ et $[t_2, 1]$. Avec le meilleur échantillonnage, on obtiendrait :

$$\begin{aligned} H_1(c)\left(\frac{1}{2}\right) &= \frac{c+1}{2} \left(1 - \frac{1+\sqrt{1+\frac{d^2}{\sigma^2}(c+1)}}{2(c+1)}\right) - \frac{d^2}{8\sigma^2} \left(\frac{2(c+1)}{1+\sqrt{1+\frac{d^2}{\sigma^2}(c+1)}} - 1\right) \\ &+ \frac{1}{2} \text{Log } \frac{1+\sqrt{1+\frac{d^2}{\sigma^2}(c+1)}}{2(c+1)} \end{aligned}$$

II.3.2. σ inconnu

Par une simple application, dans R^2 cette fois, du chapitre n° I ([3]), on montre que :

$$\lim_{n \rightarrow \infty} \frac{1}{n} \text{Log } P_{H_0} \left(\frac{\sum_{i=1}^n (X_i - \bar{X})^2}{\sum_{i=1}^n (V_i - \bar{V})^2} > c+1 \right) = - \inf_{\substack{(x,y) \in R^2 \\ x \geq (c+1)y}} (\bar{H}_0(x) + \bar{H}_0(y)) \quad \text{si } c > 0$$

$$\text{où } \bar{H}_0(x) = H_0(x-1) = \frac{x-1 - \text{Log } x}{2}$$

$$\lim_{n \rightarrow \infty} \frac{1}{n} \log P_{H_1} \left(\frac{\sum_{i=1}^n (X_i - \bar{X})^2}{\sum_{i=1}^n (V_i - \bar{V})^2} < c+1 \right) = - \inf_{\substack{(x,y) \in R^2 \\ x \leq (c+1)y}} (\bar{H}_0(x) + \bar{H}_1(y)) \quad \text{si } c < \pi_0 (1 - \pi_0) \frac{d^2}{\sigma^2}$$

$$\begin{aligned} \text{où } \bar{H}_1(y) = H_1(y-1) &= \frac{y}{2} \left[1 - \frac{1 + \sqrt{1 + 4\alpha(1-\alpha) \frac{d^2}{\sigma^2} y}}{2y} \right] - \frac{\alpha^2 d^2}{2\sigma^2} \left[\frac{2y}{1 + \sqrt{1 + 4\alpha(1-\alpha) \frac{d^2}{\sigma^2} y}} - 1 \right] \\ &+ \frac{1}{2} \log \frac{1 + \sqrt{1 + 4\alpha(1-\alpha) \frac{d^2}{\sigma^2} y}}{2y} - \log \frac{1 - \pi_0}{1 - \alpha} \end{aligned}$$

avec $\alpha(y)$ solution de :

$$\log \frac{\pi_0}{\alpha} \cdot \frac{1 - \alpha}{1 - \pi_0} + (1 - 2\alpha) \frac{d^2}{\sigma^2} \left[\frac{2y}{1 + \sqrt{1 + 4\alpha(1-\alpha) \frac{d^2}{\sigma^2} y}} - 1 \right] = 0.$$

Les calculs effectués donnent les résultats suivants :

$$\lim_{n \rightarrow \infty} \frac{1}{n} \log P_{H_0} \left(\frac{\sum_{i=1}^n (X_i - \bar{X})^2}{\sum_{i=1}^n (V_i - \bar{V})^2} > 1+c \right) = - \log \frac{1 + \frac{c}{2}}{\sqrt{1+c}}, \quad c > 0$$

$$\begin{aligned} \lim_{n \rightarrow \infty} \frac{1}{n} \log P_{H_1} \left(\frac{\sum_{i=1}^n (X_i - \bar{X})^2}{\sum_{i=1}^n (V_i - \bar{V})^2} < 1+c \right) = \\ - \frac{1}{2} \{ \log(1-T) - 2\alpha \log \frac{\pi_0}{\alpha} - 2(1-\alpha) \log \frac{1-\pi_0}{1-\alpha} - \alpha(1-\alpha) \frac{\frac{d^2}{\sigma^2} T}{1-T} \\ + \log(1 + (c+1)T) \} \end{aligned}$$

$$\text{avec } T = 2\sigma^2 \tau = \frac{(3c+2) + \alpha(1-\alpha) \frac{d^2}{\sigma^2} (c+1) + \sqrt{(c+2)^2 + 6\alpha(1-\alpha) \frac{d^2}{\sigma^2} (c+1) (c+2) + (\alpha(1-\alpha) \frac{d^2}{\sigma^2} (c+1))^2}}{4(c+1)}$$

$$\text{et } \alpha(T) \text{ solution de : } \log \frac{\pi_0}{\alpha} \cdot \frac{1 - \alpha}{1 - \pi_0} + (1 - 2\alpha) \frac{d^2}{2\sigma^2} \left[\frac{T}{1-T} \right] = 0.$$

Remarque 1 : Comme dans le cas où σ est connu, le résultat sera asymptotiquement meilleur si on se rapproche de la proposition $\pi_0 = \frac{1}{2}$.

Remarque 2 : Pour un seuil donné $(1+c)$, on peut améliorer le niveau asymptotique du test en prenant un nombre différent d'observation au numérateur et au déno-

EXPOSÉ 5

minateur . En faisant comme ci-dessus $2n$ observations au total, nous pouvons en faire $2n \cdot \theta$ pour la statistique du test et $2n(1-\theta)$ pour estimer la variance σ^2 , $\theta \in]0,1[$. On obtient alors :

$$\lim_{n \rightarrow \infty} \frac{1}{n} \log P_{H_0} \left(\frac{\frac{1}{\theta} \sum_{i=1}^{2n \cdot \theta} (X_i - \bar{X})^2}{\frac{1}{1-\theta} \sum_{i=1}^{2n(1-\theta)} (V_i - \bar{V})^2} > 1+c \right) = - \inf_{x \geq (1+c)y} 2\theta \cdot \bar{H}_0(x) + 2(1-\theta) \bar{H}_0(y)$$

Le niveau le plus petit est atteint avec $\theta(c) = \frac{1}{\log(1+c)} - \frac{1}{c}$, ce qui donne :

$$\lim_{n \rightarrow \infty} \frac{1}{n} \log P_{H_0} \left(\frac{1-\theta(c)}{\theta(c)} \cdot \frac{\sum_{i=1}^{2n \cdot \theta(c)} (X_i - \bar{X})^2}{\sum_{i=1}^{2n(1-\theta(c))} (V_i - \bar{V})^2} > 1+c \right) = - \left\{ \frac{\log 1+c}{c} - 1 - \log \frac{\log c+1}{c} \right\}$$

Il faut d'autant mieux estimer la variance (θ petit) qu'on se fixe un seuil grand ($\theta(c) \xrightarrow{c \rightarrow \infty} 0$).

III - COMPARAISON DES 2 MÉTHODES DE TESTS

III.1. Comparaison du test basé sur les sommes partielles des carrés de résidus avec le test global introduit par échantillonnage du temps (σ connu).

La comparaison de ces deux tests se justifie à de nombreux points de vue :

o d'abord dans leur conception même : le premier consiste en une investigation de toutes les sommes $\sum_{i=2}^k \omega_i^2 = \sum_{i=1}^k (y_i - \bar{y}^k)^2$, $k = 2, \dots, n$ alors que le second est fondé uniquement sur la dernière somme : $\sum_{i=1}^n (X_i - \bar{X})^2$.

o sous H_0 , les variables X_i et Y_i ont même loi et nous remarquons, que pour un seuil donné, les deux tests ont même niveau asymptotique.

o sous H_1 , les variables X_i et Y_i n'ont plus même loi mais les erreurs de deuxième espèce asymptotiques ont des expression semblables :

$$- \text{ si } t_o \geq \frac{1}{2} \text{ ou si } t_o < \frac{1}{2} \text{ et } \frac{4t_o(1-2t_o)\frac{d^2}{\sigma^2}}{(1+\sqrt{1+4t_o^2\frac{d^2}{\sigma^2}})^2} \leq c < t_o(1-t_o)\frac{d^2}{\sigma^2}, \text{ alors}$$

$$\lim_{n \rightarrow \infty} \frac{1}{n} \text{Log } P_{H_1} \left(\bigvee_k, \frac{1}{(n-1)\sigma^2} \sum_{i=1}^k \omega_i^2 \leq \frac{k-1}{n-1} + c \right) = -h_1(c)$$

$$\text{avec } h_1(c) = \frac{1}{2} \left\{ \text{Log } \Lambda - \frac{\Lambda-1}{\Lambda} + t_o(1-t_o) \frac{d^2}{\sigma^2} \left(\frac{\Lambda-1}{\Lambda} \right)^2 \right\}$$

$$\text{et } \Lambda_{t_o} = \frac{1 + \sqrt{1 + 4t_o(1-t_o) \frac{d^2}{\sigma^2} (1+c)}}{2(1+c)}$$

$$- \text{ pour } 0 < c < \pi_o(1-\pi_o) \frac{d^2}{\sigma^2},$$

$$\lim_{n \rightarrow \infty} \frac{1}{n} \text{Log } P_{H_1} \left(\sum_{i=1}^n (X_i - \bar{X})^2 < (n-1)\sigma^2 (1+c) \right) = -H_1(c)$$

$$\text{avec } H_1(c) = \frac{1}{2} \left\{ (c+1)(1-\Lambda) - \frac{\alpha^2 d^2}{\sigma^2} \left(\frac{1}{\Lambda} - 1 \right) + \text{Log } \Lambda - 2 \text{Log } \frac{1-\pi_o}{1-\alpha} \right\}$$

$$\text{et } \Lambda_\alpha = \frac{1 + \sqrt{1 + 4\alpha(1-\alpha) \frac{d^2}{\sigma^2} (1+c)}}{2(1+c)}, \quad \alpha \text{ étant toujours la solution}$$

$$\text{Log } \frac{\pi_o}{\alpha} + \frac{1-\alpha}{1-\pi_o} + (1-2\alpha) \frac{d^2}{2\sigma^2} \left(\frac{1}{\Lambda} - 1 \right) = 0.$$

Pour t_o (inconnu) et c fixés, parmi toutes les lois Π possibles, on peut regarder ce qu'on obtient avec celles qui donnent : $\alpha(c) = t_o$

$$(\text{cela revient à prendre } \pi_{o,c} = \frac{t_o}{(1-2t_o) \left(\frac{2(c+1)}{1 + \sqrt{1 + 4t_o(1-t_o) \frac{d^2}{\sigma^2} (1+c)}} - 1 \right) t_o + (1-t_o)e}) :$$

$$H_1(c)(\pi_{o,c}) - h_1(c) = \frac{1}{2} \cdot \frac{\Lambda_{t_o}^{-1}}{\Lambda_{t_o}} \cdot \frac{d^2}{\sigma^2} t_o(2t_o-1) - \text{Log } \frac{1-\pi_{o,c}}{1-t_o}.$$

EXPOSE 5

$$\text{Or } \Lambda_{t_o} > 1 \iff c < t_o(1-t_o)\frac{d^2}{\sigma^2}$$

• α était toujours compris entre π_o et $\frac{1}{2}$, donc ici t_o sera toujours compris entre $\pi_{o,c}$ et $\frac{1}{2}$

de sorte que pour $t_o > \frac{1}{2}$, $\pi_{o,c} > t_o$ donc $H_1(c)(\pi_{o,c}) > h_1(c)$

pour $t_o < \frac{1}{2}$, $\pi_{o,c} < t_o$ donc $H_1(c)(\pi_{o,c}) < h_1(c)$,

il suffit dans ce dernier cas de prendre Π de sorte que $\alpha(c) = 1 - t_o$, cela revient à prendre $1 - \pi_{o,c}$ au lieu de $\pi_{o,c}$:

$$H_1(c)(1-\pi_{o,c}) - h_1(c) = \frac{1}{2} \cdot \frac{\Lambda_{t_o}^{-1}}{\Lambda_{t_o}} \cdot \frac{d^2}{\sigma^2} \cdot (1-t_o)(1-2t_o) - \text{Log} \frac{\pi_{o,c}}{t_o}$$

donc $H_1(c)(1-\pi_{o,c}) > h_1(c)$.

On déduit des inégalités précédentes et du fait que $H_1(c)(\pi) = H_1(c)(1-\pi)$, que pour toute loi Π telle que π_o soit compris entre $\pi_{o,c}$ et $1-\pi_{o,c}$, le test global est meilleur que celui basé sur la somme des carrés des résidus ; c'est vrai en particulier pour la loi Π uniforme sur $[0,1]$. On améliorera encore le résultat pourvu qu'on ait assez d'information a priori sur l'instant t_o de rupture pour "dilater" ou "contracter" l'intervalle $[0, t_o]$ pour se rapprocher de $[0, \frac{1}{2}]$: situation optimale. Dans les autres cas, le test basé sur la somme des carrés des résidus est meilleur.

Cette comparaison n'est valable que pour les grandes valeurs de c si $t_o < \frac{1}{2}$: en effet si :

$$0 < c < \frac{4t_o(1-2t_o)}{(1+\sqrt{1+4t_o^2\frac{d^2}{\sigma^2}})^2},$$

les expressions des erreurs de 2ème espèce rendent la comparaison difficilement praticable.

III. 2. Comparaison du test basé sur les sommes partielles de résidus avec le test global introduit par échantillonnage du temps (σ connu)

Le test basé sur les sommes partielles de résidus a pour niveau :

$$\lim_{n \rightarrow \infty} \frac{1}{n} \log P_{H_0} (\exists t \in [0,1] , |s_n(t)| > at + \frac{a}{2}) = - \frac{a^2}{\sigma^2} , \quad a > 0 .$$

Le test basé sur la variance empirique de l'échantillon a pour niveau :

$$\lim_{n \rightarrow \infty} \frac{1}{n} \log P_{H_0} \left(\frac{\sum_{i=1}^n (X_i - \bar{X})^2}{(n-1)\sigma^2} > 1+c \right) = - \frac{c - \log(1+c)}{2} , \quad c > 0 .$$

L'ajustement des niveaux se fera donc en prenant $\frac{a}{\sigma} = \sqrt{\frac{c - \log(c+1)}{2}} .$

La comparaison des expressions des erreurs de 2ème espèce asymptotiques s'avérant difficile, nous tirerons notre information de la comparaison des régions où la décroissance des erreurs de 2ème espèce est effectivement exponentielle.

Pour le test basé sur les sommes partielles de résidus, il faut que :

- si $t_0 \leq e^{-3/2}$, on ait : $\frac{a}{|d|} < \mu(t_0)$ avec $\mu(t_0)$ solution de $\frac{\mu}{2t_0} + 1 + \log \mu = 0$
- si $t_0 \geq e^{-3/2}$, on ait : $\frac{a}{|d|} < \frac{2}{3} t_0 \log \frac{1}{t_0}$

Pour le test basé sur la variance empirique de l'échantillon, il faut que :

$$c < \pi_0(1-\pi_0) \frac{d^2}{\sigma^2} .$$

Pour le premier test, les valeurs de a donnant une décroissance exponentielle vérifient :

$$\frac{a}{\sigma} \leq \frac{|d|}{\sigma} \times \{ \mu(t_0) \cdot 1_{t_0 \leq e^{-3/2}} + \frac{2}{3} t_0 \log \frac{1}{t_0} \cdot 1_{t_0 \geq e^{-3/2}} \} .$$

Pour le second test, les valeurs de a donnant une décroissance exponentielle vérifient :

$$\circ \text{ si } \pi_0(1-\pi_0) \frac{d^2}{\sigma^2} \text{ est petit , } \frac{a}{\sigma} \leq \frac{1}{2} \pi_0(1-\pi_0) \frac{d^2}{\sigma^2}$$

$$\circ \text{ si } \pi_0(1-\pi_0) \frac{d^2}{\sigma^2} \text{ est grand , } \frac{a}{\sigma} \lesssim \sqrt{\frac{\pi_0(1-\pi_0)}{2} \frac{|d|}{\sigma}} .$$

Afin de faciliter la comparaison, prenons une loi Π telle que $\pi_0 = t_0$ par exemple la loi uniforme sur $[0,1]$. Pour les petites valeurs de $\frac{|d|}{\sigma}$, on constate que la région de décroissance exponentielle de l'erreur de 2ème espèce est moins étendue pour le test basé sur la variance asymptotique que pour le test basé sur les sommes de résidus : ce dernier serait donc plus puissant. Pour les grandes valeurs de $\frac{|d|}{\sigma}$, on obtient :

$$\begin{aligned} \circ \mu(t_0) < \frac{t_0(1-t_0)}{2} &\iff \frac{\sqrt{\frac{t_0(1-t_0)}{2}}}{2t_0} + 1 + \text{Log} \sqrt{\frac{t_0(1-t_0)}{2}} > 0 \\ &\iff \sqrt{\frac{1-t_0}{2t_0}} + 2 + \text{Log} \frac{t_0(1-t_0)}{2} > 0 . \end{aligned}$$

Cette dernière inégalité est réalisée pour t_0 inférieur à $e^{-3/2}$ car la fonction qui intervient est décroissante sur $[0, \frac{1}{2}]$ et est positive en $\frac{1}{2}$.

$$\circ \frac{2}{3} t_0 \text{Log} \frac{1}{t_0} < \frac{t_0(1-t_0)}{2} \iff \sqrt{\frac{1-t_0}{t_0}} + \frac{2\sqrt{2}}{3} \text{Log} t_0 > 0$$

Cette dernière inégalité est réalisée pour t_0 supérieur à $e^{-3/2}$ car la fonction qui intervient est décroissante sur $[0,1]$ et est nulle en 1 .

En résumé pour les grande valeurs de $\frac{|d|}{\sigma}$, la région de décroissance exponentielle est plus étendue pour le test basé sur la variance empirique : celui-ci serait donc plus puissant que le test basé sur les sommes de résidus.

APPENDICE

On va montrer qu'asymptotiquement, pour le problème qui nous concerne il est équivalent de faire un échantillonnage de type $\Pi^{\otimes n}$ ou de faire un tirage sans remise de n instants parmi $N(n)$ possibles. La solution de ce problème classique en échantillonnage repose sur la convergence de la loi hypergéométrique vers la loi binomiale.

Notons $U_1, U_2, \dots, U_N$ les $N(n)$ instants possibles sur $[0,1]$ et $Y_{U_1}, \dots, Y_{U_N}$ les observations associées, les variables Y_{U_i} étant indépendantes et de loi $\mathcal{N}(m_1, \sigma^2)$ si $U_i < t_0$ et $\mathcal{N}(m_2, \sigma^2)$ si $U_i \geq t_0$. On fait un tirage de n instants parmi les $N(n)$ de façon uniforme et sans remise : on obtient les observations $(X_1, \dots, X_n) = (Y_{U_1}, \dots, Y_{U_n})$. Il nous suffit alors de montrer que $\frac{1}{n} \text{Log } \phi_n(t)$ converge vers la même fonction limite qu'avec l'échantillonnage de type $\Pi^{\otimes n}$:

$$\begin{aligned} \phi_n(t) &= E e^{t \sum_{i=1}^n (X_i - \bar{X})^2} = \sum_{k=0}^n \frac{C_{\pi_0}^k N^{n-k} (1-\pi_0)^N}{C_N^n} \cdot (1-2\sigma^2 t)^{\frac{n-k}{2}} e^{\frac{nt}{1-2\sigma^2 t} \frac{d^2}{2}} \cdot \frac{k}{n} (1 - \frac{k}{n}) \\ &= \sum_{k=0}^n A_k \end{aligned}$$

Pour n grand, $\frac{1}{n} \text{Log } \phi_n(t) \sim \frac{1}{n} \text{Log}(\max_k A_k) \sim \frac{1}{n} \text{Log } A_{\alpha n}$ avec :

$$\alpha = \alpha(t) \text{ solution de : } \text{Log } \frac{\pi_0}{\alpha} + \frac{1-\alpha}{1-\pi_0} + (1-2\alpha) \frac{t d^2}{1-2\sigma^2 t} = 0$$

ce qui donne, en utilisant la formule de Stirling :

$$\frac{1}{n} \text{Log } \phi_n(t) \longrightarrow -\frac{1}{2} \text{Log}(1-2\sigma^2 t) + \alpha \text{Log } \frac{\pi_0}{\alpha} + (1-\alpha) \text{Log } \frac{1-\pi_0}{1-\alpha} + \alpha(1-\alpha) \frac{t d^2}{1-2\sigma^2 t}.$$

EXPOSÉ 5

Bibliographie :

- [1] BROWN-DURBIN-EVANS : "Techniques for testing the consistency of regression relationships over time" (1975). JRSS (B), v. 149-199.

- [2] R. AZENCOTT : "Grandes déviations : théorèmes à la Cramer-Chernoff et petites perturbations de systèmes dynamiques", Cours de l'Ecole d'Eté de Saint-Flour, 1978 (à paraître)

- [3] D. DACUNHA-CASTELLE : Exposé n° I au Séminaire de Statistique d'Orsay 1977-1978

- [4] J. BRETAGNOLLE : Exposé n° III au Séminaire de Statistique d'Orsay de 1977-1978

- [5] P. ERDÖS- A.RENYI : "On a new law of large numbers", Journal d'analyse Mathématique n° 22 (1970)

- [6] J. DESHAYES-D. PICARD : Exposé n° IV au Séminaire de Statistique d'Orsay de 1977-1978.

Jean DESHAYES
Dominique PICARD
Mathématique - Bât. 425
ERA CNRS 532 "Statistique Appliquée"
Université Paris-Sud
91405 ORSAY