

# Etude de la stationnarité des séries hydrométéorologiques

**Jacques Bernier**

Laboratoire National d'Hydraulique

Electricité de France - Direction des Études et Recherches

## Introduction

L'étude dont le présent article expose les résultats a été entreprise dans le cadre d'un contrat passé en 1972 avec le Ministère de l'Équipement et du Logement. Elle concernait la détermination de méthodologies efficaces pour détecter et estimer les ruptures de tendances et les non-stationnarités dans les séries chronologiques. Pour remettre cette étude de 1972 en perspective, nous citerons les résultats récents qu'a donné P. Bois dans sa thèse "Contribution à la critique et à la prévision des variables hydrométéorologiques ; applications à la prévision des débits du Niger et des avalanches à Davos" présentée en septembre 1976 à l'Université de Grenoble.

Bien que faisant référence aux exemples hydrologiques et météorologiques, les méthodes sont d'une portée très générale.

## Historique

La détection des "cassures" dans les séries hydrologiques est une des premières tâches des hydrologues. Diverses méthodes sont utilisées. On connaît la vogue de la technique dite "doubles cumuls" (ou doubles masses). Cette technique graphique consiste à porter en ordonnées le cumul dans l'ordre chronologique des valeurs de la série à vérifier et en abscisse le cumul des valeurs concomitantes d'une série de contrôle. On détecte usuellement les écarts systématiques pouvant survenir à certaines époques par les cassures de cette courbe "double

cumuls". Le gros défaut de cette méthode est l'absence de tests statistiques valables pour préciser la signification des cassures apparentes. On peut démontrer par des raisonnements probabilistes que la courbe résultant de séries homogènes doit présenter nécessairement de telles cassures par sa nature même et il reste à distinguer les cassures "naturelles" de celles qui sont imputables aux ruptures de tendances.

Hinkley [1] a étudié les performances de tests précis dans le cas de séries unidimensionnelles, c'est-à-dire en ne tenant pas compte de variables de contrôle :

Considérons une séquence  $y_1, y_2, \dots, y_N$  dont on connaît l'espérance mathématique  $m_0$  et la variance  $\sigma^2$ , on construit les cumuls suivants :

$$S_0 = 0 ; S_n = \sum_{j=1}^n (y_j - m_0 + \delta\sigma)$$

ou  $\delta$  est un nombre convenablement choisi.

L'époque  $L$  éventuelle, où un changement de moyenne (écart à  $m_0$ ) apparaît, peut être estimée par le premier indice  $n$  où  $S_n$  atteint sa valeur maximale, mathématiquement définie par :

$$\hat{L} = \text{minimum } (n \text{ tel que } S_n \geq S_u) \\ \text{pour } u = 1, 2, \dots, N$$

Le test consiste à rejeter l'hypothèse d'absence de changement si :

$$S_N - \max_{n < N} S_n < -h$$

Hinkley définit la méthode de choix de  $h$  et  $\delta$  pour donner des performances désirées à cette procédure. Ce choix est assez délicat et les résultats du test peuvent être complètement faussés par un mauvais choix.

L'efficacité de la méthode est également fonction de

la variance  $\sigma^2$  de la série. Dans le cas de séries hydro-métriques et pluviométriques, cette variance peut être telle que le test unidimensionnel perd toute efficacité. C'est la raison de l'introduction de variables de contrôle comme le fait la méthode des "doubles cumuls".

La Division Technique Générale (D.T.G.) d'E.D.F. a proposé et utilise systématiquement la méthode suivante :

Soit :  $y_1, y_2, \dots, y_N$ , la série à vérifier et  $x_1, x_2, \dots, x_N$ , la série concomitante de contrôle :

$\bar{x}, \bar{y}, S_x^2, S_y^2, r$  les moyennes, variances et coefficients de corrélation empiriques de ces séries.

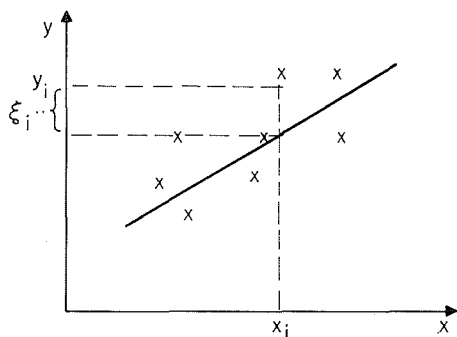
Le résidu  $\bar{\xi}_i$  correspondant à la valeur  $y_i$  dans la régression linéaire de  $y$  en  $x$  est défini par :

$$\bar{\xi}_i = y_i - \bar{y} - r \frac{S_y}{S_x} (x_i - \bar{x})$$

La D.T.G. considère les cumuls :

$$Z_k = \sum_{i=1}^k \bar{\xi}_i \quad (k \pm 1, 2, \dots, N)$$

La courbe de  $Z_k$  en fonction de  $k$  et notamment les maxima de cette courbe peuvent permettre la détection des ruptures de séquences.



On pourrait penser appliquer à ces résidus la procédure de Hinkley avec :

- $m_0 = 0$  (la moyenne des résidus est nulle)
- $\sigma^2 = S_y^2 (1 - r^2)$

On constate que la variance des résidus est d'autant plus faible que la corrélation avec la série de contrôle est forte. La corrélation accroît donc l'efficacité de la méthode. Cependant la méthode de Hinkley devrait être modifiée pour tenir compte de la contrainte imposée aux résidus, à savoir :

$$\sum_{i=1}^N \bar{\xi}_i = 0$$

### Les travaux de P. Bois

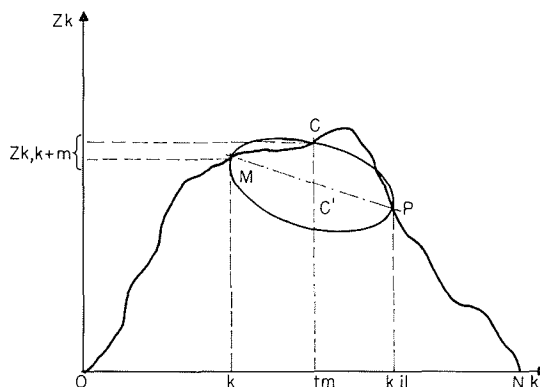
Dans une précédente publication P. Bois [2] avait proposé un test statistique basé sur les principes suivants :

Dans l'hypothèse de distribution gaussienne des résidus, ceux-ci sont distribués selon une loi normale de moyenne nulle et de variance égale à  $S_y^2 (1 - r^2)$  (résultat valable pour  $N$  grand). Un cumul partiel (cf. figure ci-après) :

$$Z_{k,k+m} = \sum_{i=k}^{k+m} \bar{\xi}_i$$

représente la distance verticale séparant deux points figuratifs de deux cumuls distants de  $m$  dans la série de résidus. C'est une réalisation d'une variable aléatoire normale d'espérance nulle et de variance donnée par Bois sous la forme :

$$\text{Var} [Z_{k,k+m}] = m S_y^2 (1 - r^2) \quad (1)$$



Une rupture peut exister entre les époques  $k$  et  $k + m$  si  $Z_{k,k+m}$  est trop grand. Dans son premier travail Bois utilisait la formule (1) pour définir un seuil de test en négligeant la dépendance des résidus et la contrainte essentielle :

$$\sum_{i=1}^N \bar{\xi}_i = 0$$

Dans sa thèse Bois tient compte de cette contrainte et plus généralement donne une approximation de la loi de l'ordonnée  $Z_k$ , de la courbe de cumul des résidus contrainte à passer par deux points  $M$  et  $P$  fixés, d'abscisses  $k$  et  $k + l$  (voir graphique joint). De façon spécifique Bois montre que la distance verticale  $CC'$  d'un point  $C$  de la courbe des cumuls d'abscisse  $k + m$  au segment  $MP$ , dans l'hypothèse de stationnarité, est une variable aléatoire normale d'espérance nulle et de variance approximative (pour  $N$  grand)

$$\text{Var} (CC') = \frac{Nm(l - m)}{(N - 1)l} S_y^2 (1 - r^2) \quad (2)$$

avec nos notations.

Pour  $l = N$ , et  $k = 0$

$$\text{Var} (CC') = \text{Var} (Z_m) = \frac{m(N - m)}{(N - 1)} S_y^2 (1 - r^2) \quad (3)$$

où  $Z_m$  est le cumul total à partir de l'origine.

La méthode de test proposé par Bois est alors la suivante :

Il existe une rupture significative entre  $M$  et  $C$  si :

$$CC' > t_{\frac{\alpha}{2}} \sqrt{\text{Var}(CC')} \quad (4)$$

où  $\text{Var}(CC')$  est calculée par (2) ou (3) et  $t_{\frac{\alpha}{2}}$  est la valeur de la variable normale centrée réduite dont la probabilité de dépassement est  $\frac{\alpha}{2}$  ( $1 - \alpha$  étant le seuil de confiance choisi)

$$. 1 - \alpha = 95 \% \quad t_{\frac{\alpha}{2}} = 1,96$$

$$. 1 - \alpha = 99 \% \quad t_{\frac{\alpha}{2}} = 2,57$$

Graphiquement la courbe représentative de la limite donnée par (4) est une ellipse passant par  $M$  et  $P$  si on se reporte au graphique précédent. Une difficulté de cette procédure est l'imprécision sur la performance du test où  $\alpha$  n'est pas liée directement à un risque d'erreur.

*En résumé*, il serait possible d'utiliser les techniques de Hinkley appliquées aux résidus  $\xi_i$ . La mise en œuvre de ces techniques n'est pas simple mais elle pourrait être utile dans certains cas d'espèce.

Les techniques pratiques que nous donnons dans le paragraphe suivant prennent pour point de départ les principes posés par P. Bois mais elles ont été développées à partir d'une étude plus poussée du processus des résidus, étude explicitée en annexe. La première méthode est essentiellement celle de notre étude de 1972, la deuxième méthode est nouvelle et basée sur l'approximation par le mouvement brownien contraint.

### Les méthodes proposées

#### 1<sup>ère</sup> méthode

1/ Pour un seuil de confiance  $1 - \alpha$  choisi (on pourra prendre  $1 - \alpha = 0,90$ ), on calcule  $t_{\frac{\alpha}{N(N-1)}}$  la valeur de la variable normale centrée réduite dont la probabilité de dépassement est  $\frac{\alpha}{N(N-1)}$ .

2/ On considère l'ellipse de contrôle d'équation

$$Z(m) = \pm t_{\frac{\alpha}{N(N-1)}} \sqrt{m(N-m)(N-1)} \cdot \frac{S_y \sqrt{1-r^2}}{N} \quad (4)$$

En déplaçant cette ellipse de façon que l'un de ses

sommets coïncide avec les différents points de la courbe "cumuls des résidus" on détecte les écarts  $Z_{k,k+m}$  qui sortent des limites de l'ellipse.

3/ La première époque  $L$  ( $0 \leq L \leq N$ ) de rupture est estimée par l'indice  $k$  correspondant au premier maximum de  $Z_k$  jugé significatif par la procédure ci-dessus.

La détection des époques de ruptures postérieures est effectuée en prenant comme origine des temps la précédente époque de rupture estimée et en effectuant un nouveau calcul des résidus.

4/ Estimation de  $\theta$  : écart moyen des espérances avant et après rupture :

$$\theta = E(\xi_i) - E(\xi_i) \quad \begin{matrix} \uparrow & \uparrow \\ i \leq L & i > L \end{matrix}$$

$$\hat{\theta} = \frac{N^2}{L(N-L)(N-1)S_y \sqrt{1-r^2}} \quad (5)$$

#### 2<sup>e</sup> méthode plus expéditive

Déplacer l'origine le long de la courbe des cumuls de résidus revient à considérer les sommes partielles  $Z_{k,k+m}$ . Considérons alors le maximum  $U$  de la valeur absolue des sommes partielles pour  $0 \leq k \leq N$  et  $0 < m \leq N - k$ . Ce maximum est en fait égal à l'étendue : écart entre le maximum et le minimum de la courbe des résidus cumulés. L'hypothèse d'absence de rupture dans la série des  $y$  est rejetée si :

$$U \geq \sqrt{\frac{(N-1)(a\gamma)}{2}} S_y \sqrt{1-r^2} \quad (6)$$

où  $\gamma$  est relié au seuil de confiance  $1 - \alpha$  choisi par :

$$\gamma = \frac{\alpha}{N}$$

et  $a\gamma$  est la valeur de la variable exponentielle dont la probabilité de dépassement est  $\gamma$  (voir annexe).

#### Remarques importantes :

1/ Les procédures ci-dessus supposent stationnaires les séries naturelles non perturbées. On ne peut pas les appliquer à des séries où apparaissent des variations saisonnières.

2/ Les couples  $(x_i, y_i)$  sont supposés indépendants. En toute rigueur, les procédures ne s'appliquent pas lorsque les séries testées présentent des auto-corrélations importantes. L'effet de ces auto-corrélations serait de sous-estimer la variance des résidus, ce qui entraînerait un risque plus grand de détecter des ruptures non réelles. Cependant la non-vérification de l'hypothèse d'indépendance ne semble pas cruciale.

3/ S'il y a rupture et donc un  $\theta$  significativement différent de 0, la statistique  $S_y^2(1-r^2)$  surestime la "vraie" variance des résidus. On pourrait dans ce cas effectuer une correction, cependant assez complexe. Pratiquement cette correction ne semble pas essentielle.

4/ Il est loisible d'utiliser plusieurs variables hydrologiques de contrôle. Dans ce cas, les résidus sont ceux apparaissant dans une régression à plusieurs dimensions et  $r$  est remplacé par le coefficient de corrélation multiple.

### Application

Le Service Central Hydrologique du Ministère de l'Équipement et du Logement nous a fourni les données de hauteurs d'eau relatives à 10 stations du bassin de l'Oise. Nous avons appliqué la technique exposée dans ce rapport aux séquences des hauteurs moyennes annuelles en utilisant la station d'Origny Sainte-Benoite comme contrôle. Le tableau ci-après donne les éléments des calculs par la première méthode.

Stations	Données existantes	$r$	$S_y \sqrt{1-r^2}$ (cm)	Signification (S - NS)	Date de rupture (maximal)
Condren	1876-1968	0,86	13,6	S	1921-1945
Moncornet	1876-1968	0,69	19,6	NS	(1915-1917)
Hirson	1885-1968	0,13	18,7	S	(1914-1939)
Origny-en-Thiérache	1873-1968 (sauf 1941)	0,78	16,7	S (seuil 70 %)	1942
Biermes	1885-1968	0,52	23,8	S	1928
Sainte-Menehould	1876-1968 (sauf 1935-1936)	0,56	22,1	S	1937
Vouziers	1884-1968	0,55	22,6	S	1920
Vraincourt	1876-1968 (sauf 1935-1936)	0,67	20,4	NS	(1939)
Etreux	1885-1968 (sauf 1924 à 1931)	0,68	20,2	S	1938
Origny-Ste-Benoite	1881-1968	-	-	-	-

- L'ensemble de ces stations présente des lacunes de 1915 à 1919.  
 - Le coefficient de corrélation  $r$  est calculé à partir de la station d'Origny-Ste-Benoite choisie comme contrôle.  
 Dans le cas de cassure non significative, la date du maximum de  $Z$  est donnée entre parenthèses à titre indicatif.

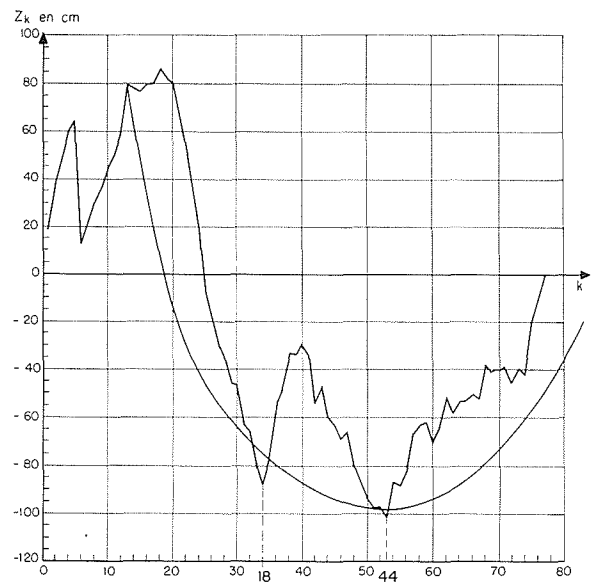
L'application de la deuxième méthode pour un seuil  $\alpha$  comparable ( $\alpha = 10\%$ ) a donné des résultats identiques sauf pour Condren, Vouziers et Etreux où la première méthode donnait des résultats très proches des limites de l'ellipse de contrôle.

Hinkley, dans son étude de la distribution d'échantillonnage de l'estimateur de  $L$ , a constaté la grande dispersion de cet estimateur. Il en résulterait donc une imprécision notable de l'estimation qui devrait être assortie d'un intervalle de confiance assez large. Il n'est pas douteux que ceci s'applique également au cas de l'analyse des résidus comme nous le faisons. Bien que cette méthode nous paraisse être la plus efficace possible, elle permet certes de tester avec une sécurité notable et calculable effectivement l'existence de rupture mais l'estimation de la date reste relativement

imprécise. Ceci ne doit pas être oublié dans l'interprétation des résultats du tableau (1). Une procédure approximative permettant d'apprécier la plage d'incertitude concernant  $L$  consiste à définir l'intervalle des valeurs de  $k$  correspondant aux  $Z_k$  qui sortent de l'ellipse de confiance.

Globalement, il apparaît donc que les séries vérifiées sont hétérogènes et les dates de rupture semblent coïncider nettement avec les deux périodes de guerre 1914-1918 d'une part, 1940-1945 d'autre part.

Notons que la station d'Origny-Ste-Benoite, choisie comme station de référence pour illustrer la méthode, a dû être soumise aux mêmes avatars que les autres stations ; elle n'est certainement pas exempte de toute rupture aux époques de guerre. Dans ce cas, le calcul de  $\theta$  n'a pas grande signification. Mais la méthode indiquée dans ce rapport peut s'appliquer dans d'autres contextes.



Résidus cumulés Condren-Origny Sainte-Benoite

### Annexe Étude des processus des résidus

Reprenons l'expression du résidu :

$$\bar{\xi}_i = y_i - \bar{y} - r \frac{S_y}{S_x} (x_i - \bar{x})$$

En utilisant l'identité :

$$\frac{r S_y}{S_x} = \frac{\sum_{j=1}^N (x_j - \bar{x}) y_j}{N S_x^2}$$

on peut mettre le résidu sous la forme :

$$\bar{\xi}_i = \left[ \frac{N-1}{N} - \frac{(x_i - \bar{x})^2}{N S_x^2} \right] y_i - \sum_{j \neq i} \left[ \frac{1}{N} + \frac{(x_i - \bar{x})(x_j - \bar{x})}{N S_x^2} \right] y_j \quad (A.1)$$

qui, les  $x_i$  fixés, exprime  $\bar{\xi}_i$  comme combinaison linéaire des  $y_j$ .

Les  $\bar{\xi}_i$  ne sont pas indépendants, ils sont fonctionnellement reliés par les deux équations :

$$\sum_{i=1}^N \bar{\xi}_i = 0 \tag{A.2}$$

$$\sum_{i=1}^N (x_i - \bar{x}) \bar{\xi}_i = 0 \tag{A.3}$$

La prise en compte de cette dépendance est essentielle pour une bonne représentation des fluctuations des résidus et de leurs cumulés.

La formule (A.1) permet l'expression de  $\bar{\xi}_i$  comme fonction linéaire de variables indépendantes des  $y_j$ . C'est la formule de base de notre calcul.

Rappelons qu'avec l'hypothèse de régression linéaire des  $y$  en fonction des  $x$ , on suppose que, conditionnellement aux  $x$  fixés :

$$- E_x (y_i) = a + b (x_i - \bar{x})$$

$$- \text{Var}_x (y_i) = \sigma^2 (1 - \rho^2)$$

$-\sigma^2$  étant la variance théorique de  $y$  et de  $\rho$  le coefficient de corrélation théorique des  $y$  avec les  $x$ .

(Le conditionnement par les  $x_i$  est indiqué par l'indice  $x$ ).

Il est possible alors de calculer les espérances et variances et covariances conditionnelles des  $\bar{\xi}_i$  par l'intermédiaire de (A.1) ; on trouve :

$$E_x (\bar{\xi}_i) = 0 \tag{A.4}$$

$$\frac{\text{Var}_x (\bar{\xi}_i)}{\sigma^2 (1 - \rho^2)} = \frac{N - 1}{N} - \frac{(x_i - \bar{x})^2}{N S_x^2} \tag{A.5}$$

$$\frac{\text{Cov}_x (\bar{\xi}_i, \bar{\xi}_j)}{\sigma^2 (1 - \rho^2)} = -\frac{1}{N} - \frac{(x_i - \bar{x})(x_j - \bar{x})}{N S_x^2} \tag{A.6}$$

Ces formules permettent de passer aux cumulés :

$$Z_k = \sum_{i=1}^k \bar{\xi}_i$$

$$E_x (Z_k) = 0 \tag{A.7}$$

$$\frac{\text{Var}_x (Z_k)}{\sigma^2 (1 - \rho^2)} = \frac{k(N - k)}{N} - \frac{k^2 (\bar{x}_k - \bar{x})^2}{N S_x^2} \tag{A.8}$$

avec

$$\bar{x}_k = \frac{1}{k} \sum_{i=1}^k x_i$$

et si  $k' > k$

$$\frac{\text{Cov}_x (Z_k, Z_{k'})}{\sigma^2 (1 - \rho^2)} = \frac{k(N - k')}{N} - \frac{k k' (\bar{x}_k - \bar{x})(\bar{x}_{k'} - \bar{x})}{N S_x^2} \tag{A.9}$$

### Distribution conditionnelle et non conditionnelle

Commençons par supposer les variables aléatoires  $y_i$

comme conditionnellement normales ; les moments (A.7) à (A.9) définissent donc les paramètres de la distribution normale des  $Z_k$ . On voit aussi que la distribution des  $Z_k$  dépend des valeurs  $x$  de la série de contrôle. En fait ici l'hypothèse de distribution normale n'est pas essentielle et au moins pour  $k$  grand, la distribution des  $Z_k$  sera approximativement normale si même la distribution des  $y_j$  ne l'est pas.

La mise en œuvre pratique des formules conditionnelles peut nécessiter des calculs assez lourds. Pour des raisons de commodité il est préférable d'utiliser des distributions et moments non conditionnels (quels que soient les  $x$ ) au prix d'une légère perte d'efficacité.

Les formules ci-dessous résultent cependant de l'approximation consistant à remplacer la variance empirique  $S_x^2$  figurant au dénominateur des formules (A.7) à (A.9) par la valeur théorique  $\sigma_x^2$  vers laquelle elle converge lorsque  $N \rightarrow \infty$  ; cette approximation est justifiée lorsque  $N$  est grand ( $N \geq 40$  à  $50$ ). On obtient alors :

$$E (Z_k) = 0 \tag{A.10}$$

$$\frac{\text{Var} (Z_k)}{\sigma^2 (1 - \rho^2)} = \frac{k(N - k)(N - 1)}{N^2} \tag{A.11}$$

$$\frac{\text{Cov} (Z_k, Z_{k'})}{\sigma^2 (1 - \rho^2)} = \frac{k(N - k')(N - 1)}{N^2} \tag{A.12}$$

De (A.12) on tire l'expression du coefficient de corrélation entre  $Z_k$  et  $Z_{k'}$  :

$$\rho(Z_k, Z_{k'}) = \frac{k(N - k')}{\sqrt{k'(N - k)}} \tag{A.13}$$

$Z_k$  et  $Z_{k'}$  sont donc très liés lorsque  $k' - k$  est petit.

### Détermination de l'ellipse de contrôle

Le problème est de rechercher une courbe de contrôle telle que tout dépassement amène à repousser l'hypothèse d'homogénéité de la série avec un seuil de confiance contrôlé.

Considérons alors l'ensemble de tous les intervalles :

$$- t_{\frac{\gamma}{2}} \sqrt{\text{Var} (Z_k)} \leq Z_k \leq + t_{\frac{\gamma}{2}} \sqrt{\text{Var} (Z_k)}$$

pour  $k = 1, 2, \dots, N$

avec une valeur  $t_{\frac{\gamma}{2}}$  de la variable normale centrée réduite dont la probabilité de dépassement est  $\frac{\gamma}{2}$ .

Soit  $E_k$  l'évènement : " $Z_k$  compris dans l'intervalle"

$$\text{Prob} [E_k] = 1 - \gamma$$

On peut écrire :

$$\begin{aligned} \text{Prob} [\text{tous les } E_k \text{ simultanément}] &= \\ &= \text{Prob} [E_1 \cap E_2 \dots \cap E_N] \geq 1 - \sum_{k=1}^N P(\bar{E}_k) \end{aligned}$$

où  $\bar{E}_k$  est l'évènement contraire de  $E_k : Z_k$  situé hors de l'intervalle.

La probabilité de  $\bar{E}_k$  est égale à  $\gamma$  donc :

$$\text{Prob [tous les } E_k \text{ simultanément]} \geq 1 - N\gamma$$

On voit ainsi qu'il ne suffit pas de contrôler la probabilité correspondant à un  $Z_k$  individuel comme le fait P. Bois mais il faut contrôler la probabilité de l'occurrence simultanée des  $E_k$  dont on connaît une borne inférieure. La méthode consiste donc à contrôler cette borne inférieure, par exemple :

$$1 - N\gamma = 1 - a$$

et d'en tirer la valeur de  $\gamma$  correspondante.

Si le sommet de l'ellipse est déplacé de façon à coïncider successivement avec les différents points de la courbe des cumuls, on aura en fait  $\frac{N(N-1)}{2}$  intervalles à prendre en compte de telle sorte que dans ce cas :

$$1 - \frac{N(N-1)}{2} \gamma = 1 - a$$

L'expression de  $\text{Var}(Z_k)$  en fonction de  $k$  (formule A.11) montre bien que les limites de la courbe de contrôle sont définies par l'équation d'une ellipse donnée par la formule (4).

**Approximation asymptotique (complément à l'étude de 1972)**

Considérons un changement d'échelle des temps en posant :

$$t = \frac{K}{N}$$

et le processus des  $Z_k$  transformé sous la forme :

$$B(t) = \frac{Z_{Nt}}{S_y \sqrt{(1-r^2)(N-1)}} \tag{A.14}$$

On peut alors montrer aisément que si  $N$  tend vers l'infini le processus  $B(t)$  converge vers un processus de Wiener-Levy ou mouvement brownien défini sur l'intervalle de temps  $0 \leq t \leq 1$  et soumis à la contrainte :

$$B(t) = 0 \text{ pour } t = 1$$

Bois avait déjà remarqué dans sa thèse l'analogie avec le processus de Wiener-Levy.

On peut cependant utiliser plus complètement les propriétés du mouvement brownien. On trouvera notamment dans l'ouvrage de S. Karlin [3], l'expression de la loi conjointe du maximum de  $B(t)$  sur l'intervalle  $0 \leq t \leq u$ , soit  $B_M$  et de  $B(u)$ . Connaissant la loi de  $B(u)$  on peut en déduire la loi du maximum  $B_M$  conditionnée par la contrainte  $B(u) = x$ . On obtient le résultat remarquable suivant :

La variable aléatoire  $A = \frac{2 B_M (B_M - x)}{u}$  est distri-

buée selon une loi exponentielle :

$$\text{Prob}[A \geq a] = e^{-a} \tag{A.15}$$

Pour un seuil de test fixé  $\gamma$  tel que :

$$\text{Prob}[A \geq a_\gamma] = \gamma$$

on peut alors bâtir un test à partir du maximum de  $Z_k$  en supposant valable l'approximation par le mouvement brownien ( $N > 50$ ).

Si on considère le maximum des  $Z_k$  contraintes à la seule condition  $Z_N = 0$ , d'où  $x = 0$  pour  $u = 1$  dans la formule de la loi de  $A$ , on définit un seul  $B_0$  vérifiant :

$$2 B_0^2 = \alpha_\gamma \text{ soit } B_0 = \sqrt{\frac{\alpha_\gamma}{2}} \tag{A.16}$$

On est donc amené à prendre en compte la limite de rejet de l'hypothèse d'absence de rupture sous la forme :

$$\frac{\max Z_k}{S_y \sqrt{1-r^2}} \geq \sqrt{(N-1) B_0} \tag{A.17}$$

Si de façon comparable à la méthode basée sur l'ellipse de contrôle, on envisage de déplacer l'origine de la courbe des  $Z_k$  en la faisant coïncider avec tous les points de cette courbe, un raisonnement analogue au précédent montre que  $\gamma$  doit être relié à un risque  $\alpha$  global contrôlé par la formule :

$$1 - N\gamma = 1 - a$$

On notera de plus que les propriétés des lois du mouvement brownien étant symétriques autour de  $B = 0$ , la loi conditionnelle pour  $Z_N = 0$  s'applique aussi au minimum de  $Z_k$  et la formule (A.17) est donc valable pour le minimum à condition de changer le sens de l'inégalité.

**Estimation de  $\theta$**

Admettons un écart moyen  $\theta$  intervenant dans la série à l'époque  $L$ , tel que :

$$E(y_i) = \mu + \theta \text{ pour } 0 \leq i \leq L$$

$$E(y_i) = \mu \text{ pour } L < i < N$$

On peut montrer que  $Z_k$  n'a pas une espérance mathématique nulle mais :

$$\text{si } k \leq L : \frac{E(Z_k)}{S_y \sqrt{1-r^2}} = \frac{\theta k (N-1) (N-L)}{N^2}$$

$$\text{si } k > L : \frac{E(Z_k)}{S_y \sqrt{1-r^2}} = \frac{\theta L (N-1) (N-k)}{N^2}$$

L'espérance de  $Z_k$  atteint donc un extrêmu pour  $k = L$  ce qui justifie la méthode de test et d'estimation de  $L$ .

On constate également que  $Z_k$  est en moyenne négative si l'écart  $\theta$  est négatif, c'est-à-dire si les  $y_i$  précédant l'époque  $L$  sont en moyenne inférieurs aux  $y_i$  suivant  $L$ .

Ces formules peuvent permettre une estimation de  $\theta$ .  
En utilisant la déviation maximale, on peut égaler  
l'espérance mathématique de  $Z_L$  à la valeur observée :

$$\frac{\hat{\theta} (N - 1) (N - L) L}{N^2} = \frac{Z_L}{S_y \sqrt{1 - r^2}}$$

On estime donc  $\theta$  par :

$$\hat{\theta} = \frac{N^2 Z_L}{L(N - 1) (N - L) S_y \sqrt{1 - r^2}} \quad (\text{A.19})$$

## Bibliographie

---

- [1] HINKLEY (D.V.) – Inference about the change-point from cumulative sum tests – *Biometrika* (1971) n° 58-3.
- [2] BOIS (P.) – Une méthode de contrôle de séries chronologiques utilisées en climatologie et en hydrologie. *Publication du laboratoire de Mécanique des Fluides* – Section Hydrologie – Université de Grenoble (mai 1971).
- [3] KARLIN (S.) – *A first course in Stochastic Processes* – Academic Press – 1966 – p. 281 – Problème 1.