

Record Number:
Author, Monographic: Cluis, D.//Laberge, C.
Author Role:
Title, Monographic: Étude de l'interpolation linéaire et de l'interpolation homoscédastique
Translated Title:
Reprint Status:
Edition:
Author, Subsidiary:
Author Role:
Place of Publication: Québec
Publisher Name: INRS-Eau
Date of Publication: 1985
Original Publication Date: Janvier 1985
Volume Identification:
Extent of Work: vi, 31
Packaging Method: pages
Series Editor:
Series Editor Role:
Series Title: INRS-Eau, Rapport de recherche
Series Volume ID: 173
Location/URL:
ISBN: 2-89146-171-1
Notes: Rapport annuel 1984-1985
Abstract: 10.00\$
Call Number: R000173
Keywords: rapport/ ok/ dl

Etude de l'interpolation linéaire et
de l'interpolation homoscédastique

par

D. CLUIS et C. LABERGE

Rapport scientifique no 173

Janvier 1985

RÉSUMÉ

L'interpolation est une technique rétrospective pouvant être utilisée dans plusieurs cas différents. Ainsi, son utilisation permet de compléter des valeurs absentes ou encore de réestimer des valeurs apparemment aberrantes. Le cas que nous traiterons dans ce rapport présente l'interpolation comme répartition, à l'intérieur des intervalles de mesures, de l'information contenue dans une série originale.

Une telle utilisation amène évidemment des changements dans les caractéristiques de la série. Nous tenterons alors de mettre en évidence les changements théoriques prévisibles ainsi que les changements empiriques observables.

Deux types d'interpolation seront étudiés: premièrement, l'interpolation linéaire qui répartit l'information linéairement à l'intérieur de l'intervalle formé par deux points de la série originale et, finalement, l'interpolation homoscedastique qui garde la variance constante d'une valeur interpolée à l'autre.

Dans la première partie du rapport, chaque type d'interpolation sera présenté tandis que dans la seconde partie nous comparerons les résultats obtenus pour chaque cas.

Mots-clé / Descripteurs:

Interpolation linéaire / interpolation homoscedastique / séries chronologiques /

ABSTRACT

Interpolation is a retrospective technique that can be used in different cases. It allows the replacement of missing data or the reestimation of outliers values. In this report, interpolation will be presented as a method for the repartition of time series information, at smaller intervals.

This kind of use brings, of course, changes in the series characteristics. We will try to show clearly the theoretical changes and illustrate them with empirical observations.

Two kinds of interpolation will be studied: first, linear interpolation which spreads the information linearly inside intervals formed by two successive points of the original series and, secondly, homoscedastic interpolation which keeps variance constant from one rank of interpolation to another.

In the first part of this report, each method of interpolation will be presented, while in the second part, we will compare the results obtained in each case.

Keywords / Descriptors:

Linear interpolation / homoscedastic interpolation / time series /

Table des matières

| | <u>Page</u> |
|---|-------------|
| Résumé | i |
| Abstract | ii |
| Table des matières | iii |
| Liste des figures | iv |
| Liste des tableaux | v |
| Notations | vi |
| | |
| 1. Interpolation linéaire | 1 |
| 1.1 Présentation | 1 |
| 1.2 Résultats théoriques | 3 |
| 1.1.1 Opérateur de différence ∇ | 3 |
| 1.2.2 Stationnarité de la moyenne | 3 |
| 1.2.3 Variance | 6 |
| 1.2.4 Autocovariance et fonctions d'autocorrélation . | 8 |
| 1.2.5 Résultats spéciaux | 10 |
| 1.3 Résultats empiriques | 12 |
| 2. Interpolation homoscédastique | 16 |
| 2.1 Présentation | 16 |
| 2.2 Résultats théoriques | 16 |
| 2.2.1 Moyenne | 16 |
| 2.2.2 Variance | 18 |
| 2.2.3 Autocovariance et fonctions d'autocorrélation . | 19 |
| 2.3 Résultats empiriques | 20 |
| 2.3.1 Moyenne | 20 |
| 2.3.2 Variance | 21 |
| 3. Comparaison entre les deux types d'interpolation | 25 |
| 3.1 Moyenne | 25 |
| 3.2 Variance | 25 |
| 3.3 Autocovariance et fonction d'autocorrélation | 25 |
| | |
| Conclusion | 30 |
| Bibliographie | 31 |

Liste des figures

| | <u>Page</u> |
|---|-------------|
| Figure 1.1: Interpolation linéaire ($p=10, k=16, n=151$) | 2 |
| Figure 1.2: Première différenciation (∇y_t) de la série de la figure 1.1 | 4 |
| Figure 1.3: Deuxième différenciation ($\nabla^2 y_t$) de la série de la figure 1.1 | 5 |
| Figure 2.1: Interpolation homoscédastique formée à partir des mêmes points noeuds que la série de la figure 1.1 | 22 |
| Figure 2.2: Interpolation homoscédastique non centrée | 23 |
| Figure 2.3: Interpolation homoscédastique centrée | 24 |

Liste des tableaux

| | <u>Page</u> |
|---|-------------|
| Tableau 1: Moyennes et variances d'une partition des différentes interpolations formées à partir d'une série de points noeuds sans persistance... | 13 |
| Tableau 2: Moyennes et variances d'une partition des différentes interpolations formées à partir d'une série de points noeuds ayant une persistance de l'ordre de 0.7 | 14 |
| Tableau 3: Écarts relatifs des moyennes des séries interpolées homoscédastiquement versus l'interpolation linéaire | 26 |
| Tableau 4a: Fonction d'autocorrélation de l'interpolation linéaire | 29 |
| Tableau 4b: Fonction d'autocorrélation de l'interpolation homoscédastique | 29 |

Notations

Les notations suivantes seront utilisées dans le présent rapport:

- $E(X)$: Espérance de la variable aléatoire X
- $Var(X)$: Variance de la variable aléatoire X
- $Cov(X, Y)$: Covariance entre les variables aléatoires X et Y
- $ACV(k)$: Autocovariance d'ordre k
- \bar{Y} : Moyenne de la variable aléatoire Y .
- X_{IL} : Point de l'interpolation linéaire.
- X_{IHC} : Point de l'interpolation homoscédastique centrée.
- X_{IHNC} : Point de l'interpolation homoscédastique non-centrée.
- \bar{X}_{IL} : Moyenne de l'interpolation linéaire.
- \bar{X}_{IHC} : Moyenne de l'interpolation homoscédastique centrée.
- \bar{X}_{IHNC} : Moyenne de l'interpolation homoscédastique non-centrée.
- $Y_{p'}$: Écriture abrégée de $Y_{(L-1)p + p'}$

1. INTERPOLATION LINÉAIRE

1.1 Présentation

Soit $X_1 \dots X_k$, des réalisations qui formeront notre série originale. Ces réalisations représenteront également les "points noeuds" de la série interpolée. C'est donc à partir de ces points que l'on construira l'interpolation.

Soit "p", le pas d'interpolation choisi par l'utilisateur: "p" représente le nombre d'intervalles de grandeurs équivalentes que l'on forme entre deux points noeuds consécutifs.

Il est bon de noter que plus "p" sera grand, moins les points interpolés contiendront d'information puisque l'information contenue dans les points noeuds est distribuée aux points interpolés.

Étant donnés $X_1 \dots X_k$ et p, l'interpolation linéaire est formée de points $Y_{(L-1)p+p'}$, définis par:

$$Y_{(L-1)p+p'} = \frac{(p'-1)X_{L+1} + (p-p'+1) X_L}{p} \quad (1)$$

où p' varie de 1 à p et représente la position de la valeur interpolée dans l'intervalle $[X_L, X_{L+1}[$.

Ainsi, la série $Y_{(L-1)p+1}$ pour $L=1, \dots, k$ représente la série des points noeuds, la série $Y_{(L-1)p+2}$ pour $L=1, \dots, k-1$ représente la série des points suivant un point noeud, etc.

La figure 1.1 donne un exemple d'interpolation linéaire. On y retrouve 16 points noeuds (la série originale, X_1, \dots, X_{16}), avec lesquels nous avons construit une série de 151 valeurs à raison d'un pas d'interpolation de 10.

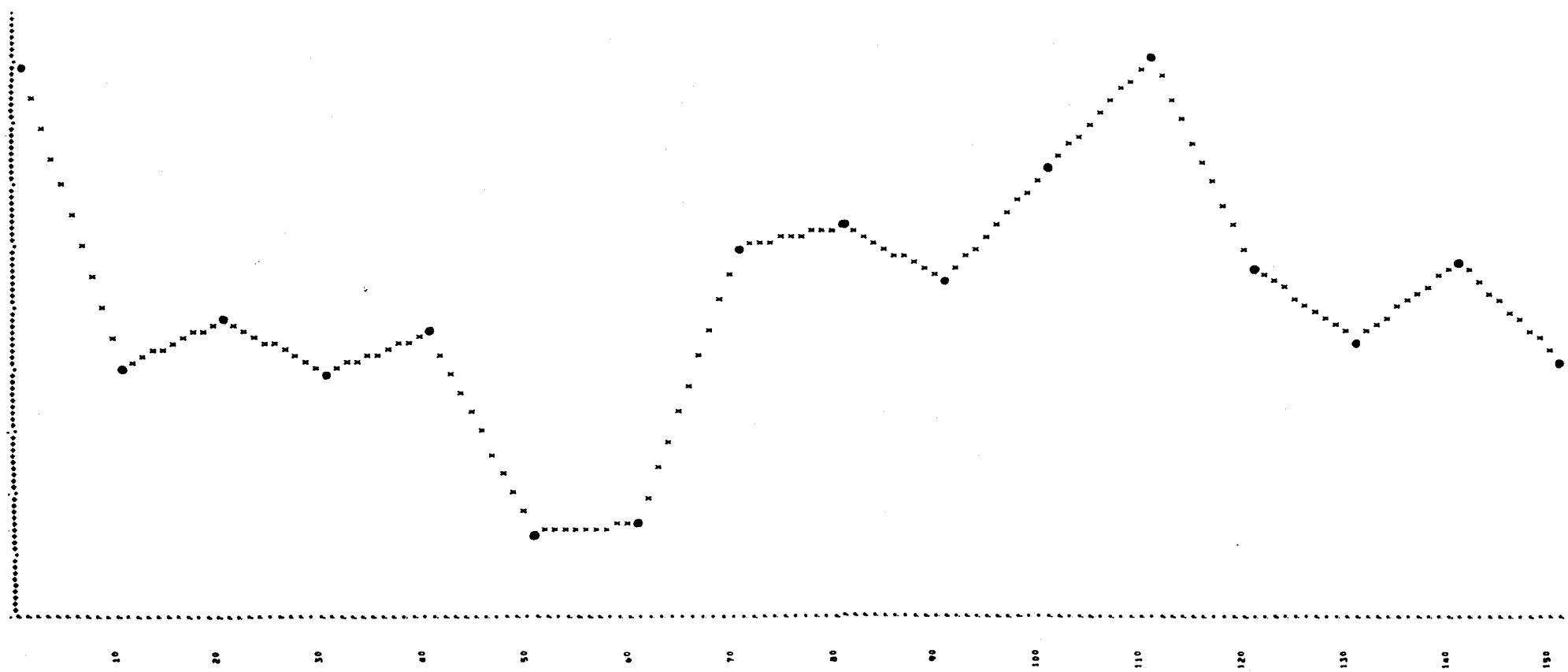


Figure 1.1: Interpolation linéaire ($p=10, k=16, n=151$).

Il faut cependant noter que nous ne créons pas de nouvelle information avec une telle interpolation mais, comme nous l'avons dit précédemment, nous répartissons l'information à un pas plus fin.

1.2 Résultats théoriques

1.2.1 Opérateur de différence ∇^*

Cet opérateur, souvent utilisé dans la méthode d'analyse de Box et Jenkins afin de stationnariser la série en moyenne, amène des résultats très spéciaux lorsqu'il est appliqué à une série interpolée linéairement.

Ainsi, la série ∇Y_t (où Y_t est la série interpolée linéairement) aura la forme illustrée sur la figure 1.2. On s'aperçoit que cette série reste constante dans les intervalles $]X_L, X_{L+1}]$, ce résultat provenant directement de la forme que nous imputons à la série interpolée.

La série de deuxième différence ($\nabla^2 Y_t$) aura, elle aussi, une forme très particulière (figure 1.3). Ainsi, seulement les points $Y_{(L-1)p+2}$ sont non-nuls; les autres points valent zéro, car ∇Y_t et ∇Y_{t-1} sont égaux dans les intervalles $]X_L, X_{L+1}]$.

La présentation ici, des effets de l'interpolation linéaire sur les séries ∇Y_t et $\nabla^2 Y_t$ sert à mettre en garde, tout utilisateur, devant les limitations de cette interpolation.

1.2.2 Stationnarité de la moyenne

Étant donné la forme d'une interpolation linéaire, il est simple de vérifier que si la série originale (points noeuds) est stationnaire en moyenne, alors la série Y_t (points interpolés) sera également stationnaire en moyenne.

* $\nabla = (1-B)$ i.e.: $\nabla Y_t = (1-B) Y_t = Y_t - Y_{t-1}$

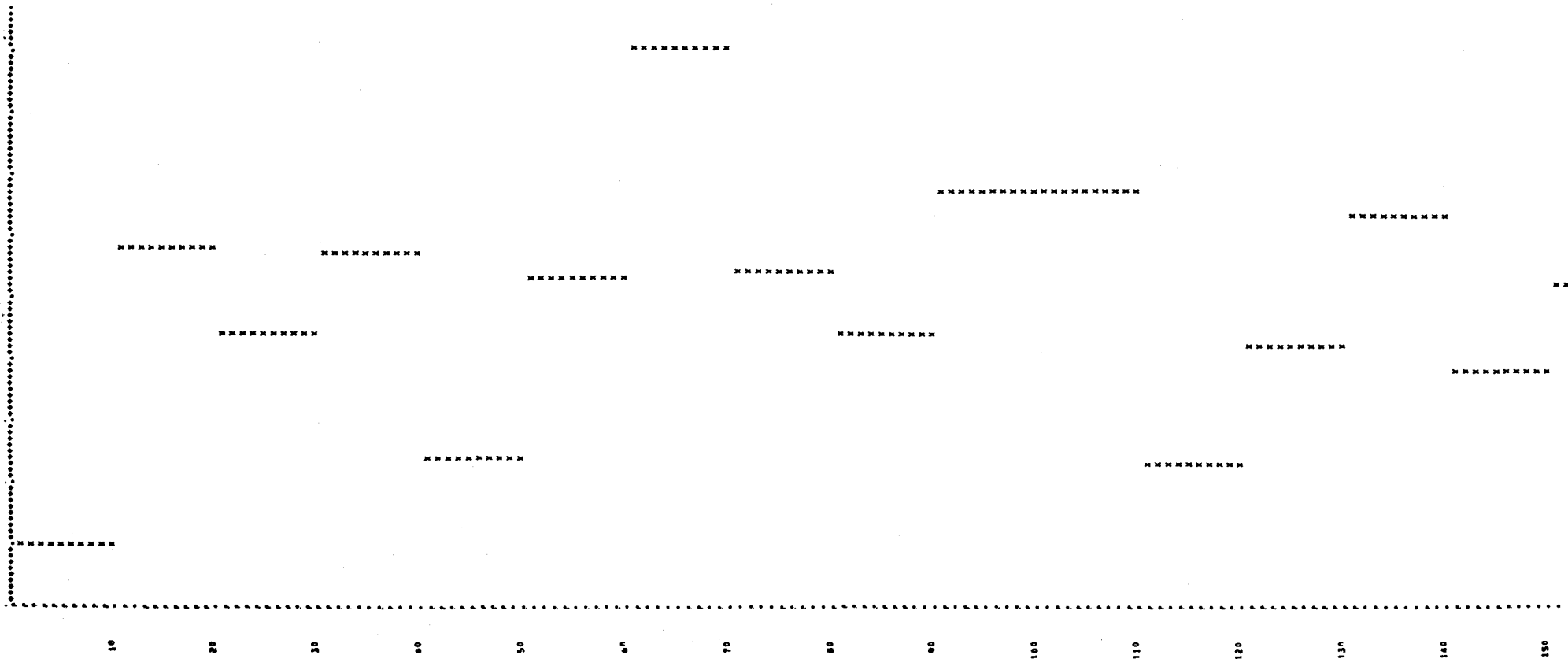


Figure 1.2: Première différenciation (∇Y_t) de la série de la figure 1.1.

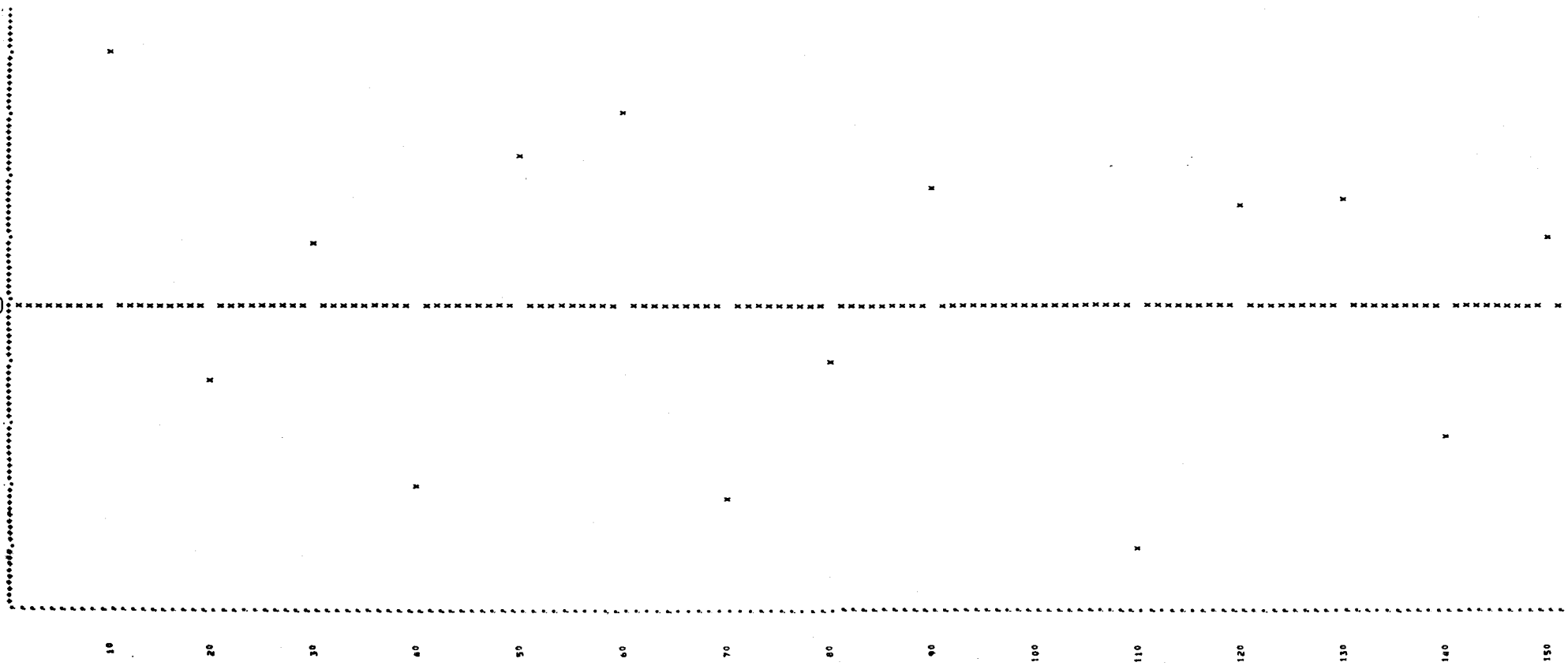


Figure 1.3: Deuxième différenciation ($\nabla^2 Y_t$) de la série de la figure 1.1

Preuve:

D'après l'équation (1), on a:

$$E(Y_{(L-1)p+p'}) = \frac{(p'-1)E(X_{L+1}) + (p-p'+1)E(X_L)}{p}$$

Comme X_t est stationnaire, on a $E(X_L) = E(X_{L+1}) = E(X_K)$ pour tout K

Ainsi:

$$E(Y_{(L-1)p+p'}) = \frac{(p'-1 + p - p' + 1) E(X_L)}{p} = E(X_L) \text{ pour tout } L, p, p'$$

1.2.3 Variance

Cependant, même si la série originale est stationnaire en variance, il n'en résulte pas une série interpolée linéairement qui est aussi stationnaire en variance.

Preuve:

D'après l'équation (1), on a:

$$\text{Var}(Y_{p'}) = \frac{(p'-1)^2 \text{Var}(X_{L+1}) + (p-p'+1)^2 \text{Var}(X_L) + 2(p'-1)(p-p'+1) \text{Cov}(X_L, X_{L+1})}{p^2}$$

Comme X_t est stationnaire, on a $\text{Var}(X_L) = \text{Var}(X_{L+1}) = \text{Var}(X_K)$ pour tout K .

De plus, généralement on a $\text{Cov}(X_L, X_{L+1}) \neq 0$

Ainsi:

$$\begin{aligned} \text{Var}(Y_{p'}) &= \left[\frac{(p-1)^2 + (p-p'+1)^2}{p^2} \right] \text{Var}(X_L) + \frac{2(p'-1)(p-p'+1)}{p^2} \text{Cov}(X_L, X_{L+1}) \\ &= \text{Var}(X_L) \left[\left(1 - \frac{2(p'-1)(p-p'+1)}{p^2} \right) + \left(\frac{2r_1(p'-1)(p-p'+1)}{p^2} \right) \right] \end{aligned}$$

$$\text{où: } r_1 = \frac{\text{Cov}(X_L, X_{L+1})}{\text{Var}(X_L)}$$

Finalement on obtient:

$$\text{Var}(Y_{(L-1)p+p'}) = \text{Var}(X_L) \left[1 - \frac{2(1-r_1)(p'-1)(p-p'+1)}{p^2} \right] \quad (2)$$

La présence de p' dans le terme de droite de l'équation (2) montre la non-stationnarité de la variance de la série interpolée linéairement.

Cependant, il est bon de remarquer, que la non-stationnarité de la variance apparaît à l'intérieur des intervalles d'interpolation (entre deux points noeuds). Si nous prenons une vue plus macroscopique de la série, il semble possible de penser que la variance est stationnaire (aux effets de bout près) d'un sous-ensemble assez grand de la série à un autre car la variance est constante d'un intervalle d'interpolation à l'autre.

Ainsi:

$$\begin{aligned} E(\text{Var}(Y_{p'})) &= \frac{p}{\sum_{p'=1}^p} \frac{\text{Var}(Y_{(L-1)p+p'})}{p} = \frac{\text{Var}(X_L)}{p} \frac{p}{\sum_{p'=1}^p} \left[1 - \frac{2(1-r_1)(p'-1)(p-p'+1)}{p^2} \right] \\ &= \frac{\text{Var}(X_L)}{p} \left[p - \sum_{p'=1}^p \frac{2(1-r_1)(p'-1)(p-p'+1)}{p^2} \right] = \frac{\text{Var}(X_L)}{p} \left[p - \frac{2(1-r_1)(p^2-1)}{6p} \right] \\ &= \frac{\text{Var}(X_L)}{p} \left[\frac{6p^2 - 2p^2 + 2 + 2r_1 p^2 - 2r_1}{6p} \right] = \text{Var}(X_L) \left[\frac{2p^2 + 1}{3p^2} + \frac{r_1(p^2 - 1)}{3p^2} \right] \quad (3) \end{aligned}$$

Donc pour une même interpolation (mêmes points noeuds et même pas d'interpolation), on voit que la variance moyenne des points interpolés ne varie pas d'un intervalle d'interpolation à l'autre car le terme de droite ne dépend pas de la variable p' .

On voit donc la particularité de la non-stationnarité de la variance dans le cas de l'interpolation linéaire. De plus, d'après Pankratz (1983)*: "If a data series is stationary then the variance of any major subset of the series will differ from the variance of any other major subset only by chance". Avec la précédente définition, il est donc légitime de parler de stationnarité du deuxième ordre d'une série interpolée linéairement. Cependant, si on demeure dans le contexte plus stricte, où la stationnarité d'une variable aléatoire est basée sur l'espérance mathématique indépendante du rang, il peut être utile de corriger l'interpolation linéaire par un facteur variable dépendant de p' et p qui fait disparaître artificiellement, la non-stationnarité de variance: c'est ce que nous appellerons l'interpolation homoscedastique.

1.2.4 Autocovariances et fonctions d'autocorrelation

Tout comme ce fut le cas avec les variances, on s'aperçoit que les autocovariances sont non-stationnaires, même si la série des points noeuds est stationnaire.

Pour illustrer ce problème, calculons la valeur de:

$$\begin{aligned} \text{ACV}(1) &= \text{Cov}(Y_{(L-1)p+p'}, Y_{(L-1)p+p'+1}) \\ &= \text{Cov} \left(\frac{(p'-1)X_{L+1} + (p-p'+1)X_L}{p}, \frac{(p')X_{L+1} + (p-p')X_L}{p} \right) \end{aligned}$$

* p. 16.

$$\begin{aligned}
&= \frac{p'(p'-1)\text{Var}(X_{L+1}) + (p-p'+1)(p-p')\text{Var}(X_L)}{p^2} \\
&\quad + \left[\frac{(p'-1)(p-p') + (p-p'+1)p'}{p^2} \right] \text{Cov}(X_L, X_{L+1}) \\
&= \left[\frac{p'(p'-1) + (p-p'+1)(p-p')}{p^2} \right] \text{Var}(X_L) + \left[\frac{(p'-1)(p-p') + (p-p'+1)p'}{p^2} \right] \text{Cov}(X_L, X_{L+1}) \quad (4)
\end{aligned}$$

Encore une fois le terme de droite dépend de p' , donc la position où l'on se trouve entre deux points noeuds fera varier la valeur de l'autocovariance d'ordre 1.

On remarquera un tel problème pour toutes les autres autocovariances et leurs calculs est un peu plus complexe étant donné la possibilité de chevauchement d'un point noeud. Cependant, comme pour la variance, la non-stationnarité n'apparaît plus lorsque l'on prend une vue plus macroscopique de la série. (voir Cluis et Bobée (1980)).

Quant aux coefficients d'autocorrélation, on a, encore une fois, l'apparition d'une non-stationnarité.

Soit:

$\rho_k =$ coefficient d'autocorrélation d'ordre k

$$\rho_k = \begin{cases} \frac{\text{ACV}(k)}{\text{Var}(Y_t)} & \text{si } Y_t \text{ est stationnaire en variance} \\ \frac{\text{ACV}(k)}{[\text{Var}(Y_t) \cdot \text{Var}(Y_{t+k})]^{1/2}} & \text{dans le cas contraire} \end{cases}$$

On peut facilement voir la non-stationnarité en remplaçant le numérateur et le dénominateur par les valeurs obtenues dans les paragraphes précédents (équations (2) et (3)). Mais, encore une fois, cette non-stationnarité disparaît lorsque l'on prend comme unité de mesure, un pas complet d'interpolation.

1.2.5 Résultats spéciaux

Il existe une exception à la non-stationnarité présentée dans le dernier paragraphe. En effet, si $p=2$, les coefficients d'autocorrélation d'ordre impair sont stationnaires d'un p' à l'autre. Voyons la démonstration pour le coefficient d'ordre 1 de la série interpolée (noté ρ_{1y}), une démonstration identique peut être faite pour ρ_{3y} , ρ_{5y} , etc.

Démonstration:

Soit:

ρ_{iy} = coefficient d'autocorrélation d'ordre "i" de la série interpolée

ρ_{ix} = coefficient d'autocorrélation d'ordre "i" de la série des points noeuds

X_L = valeur "L" de la série des points noeuds.

$$\rho_{1y} = \frac{\text{Cov}(Y_{p'}, Y_{p'+1})}{[\text{Var}(Y_t) \cdot \text{Var}(Y_{t+1})]^{1/2}} \quad \text{et} \quad Y_{p'} = Y_{(L-1)2+p'} = \frac{(p'-1)X_{L+1} + (2-p'+1)X_L}{2}$$

si $p'=1$

$$\rho_{1y} = \frac{\text{Cov}(X_L, \frac{X_{L+1} + X_L}{2})}{\{\text{Var}(X_L) \cdot \text{Var}(\frac{X_{L+1} + X_L}{2})\}^{1/2}} = \frac{\frac{1}{2}[\text{Var}(X_L) + \text{Cov}(X_L, X_{L+1})]}{\{\text{Var}(X_L) \cdot \frac{1}{4}[\text{Var}(X_L) + \text{Var}(X_{L+1}) + 2\text{Cov}(X_L, X_{L+1})]\}^{1/2}}$$

$$= \frac{\frac{1}{2}[\text{Var}(X_L) + \text{Cov}(X_L, X_{L+1})]}{\{\text{Var}(X_L) \cdot \frac{1}{4}[2\text{Var}(X_L) + 2\text{Cov}(X_L, X_{L+1})]\}^{\frac{1}{2}}} \text{ car } \{X_L\}_{L=1}^k, \text{ est stationnaire}$$

$$\rho_{1y} = \frac{1 + \rho_{1x}}{[2 + 2\rho_{1x}]^{\frac{1}{2}}} \quad (5a)$$

Si $p'=2$

$$\rho_{1y} = \frac{\text{Cov}\left(\frac{X_{L+1} + X_L}{2}, X_{L+1}\right)}{\{\text{Var}\left(\frac{X_{L+1} + X_L}{2}\right) \cdot \text{Var}(X_{L+1})\}^{\frac{1}{2}}} = \frac{\frac{1}{2}(\text{Var}(X_{L+1}) + \text{Cov}(X_L, X_{L+1}))}{\{\text{Var}(X_{L+1}) \cdot \frac{1}{4}[2\text{Var}(X_{L+1}) + 2\text{Cov}(X_L, X_{L+1})]\}^{\frac{1}{2}}}$$

car $\{X_L\}_1^k$ est stationnaire.

$$\rho_{1y} = \frac{1 + \rho_{1x}}{[2 + 2\rho_{1x}]^{\frac{1}{2}}} \quad (5b)$$

Ce résultat nous dit également que si l'on prend une série sans persistance (dont les données successives sont indépendantes) à partir de laquelle nous créons une autre série par interpolation linéaire de pas $p=2$, cette série aura alors asymptotiquement une persistance de 0.7071.

De plus, étant donné que plus le pas d'interpolation augmente et plus la persistance augmente, on a donc asymptotiquement 0.7071 comme valeur minimum du coefficient d'autocorrélation d'ordre 1 d'une série interpolée linéairement. On crée donc, et il fallait s'y attendre, une série plus persistante que la série originale lorsque l'on construit une série interpolée linéairement. (Ces résultats peuvent être généralisés à l'interpolation homoscédastique; voir partie 3.3).

1.3 Résultats empiriques

Étant donné que les résultats suivant ne serviront qu'à illustrer les résultats théoriques discutés dans les parties précédentes, il nous apparaissait inutile de procéder à un grand nombre de simulations.

Ainsi, nous avons créé, par simulation deux séries markoviennes d'ordre 1, de 22 valeurs. La première série est sans persistance alors que la seconde a une persistance de l'ordre de 0,7. (N.B: on a 22 points noeuds, mais seulement 21 points pour $p'=2,3,4,\dots,p$)*.

Les différents résultats quant aux moyennes et aux variances des séries interpolées sont donnés dans les tableaux 1 et 2. Les résultats sont donnés pour une partition de dix sous-ensembles formée de la façon suivante: le premier regroupe tous les points noeuds ($p'=1$), le second regroupe tous les points suivant un point noeud ($p'=2$) et ainsi de suite jusqu'au dixième sous-ensemble qui contient tous les points précédant immédiatement un point noeud ($p'=10$).

Pour le cas qui nous intéresse présentement, les moyennes d'une série interpolée linéairement, les résultats sont présentés dans les deuxièmes colonnes des tableaux 1 et 2 (immédiatement à la droite de la colonne des p'). On voit que la moyenne n'est pas égale pour chaque p' , ce qui semble, à prime abord, en désaccord avec notre résultat théorique disant que la moyenne est stationnaire.

Cependant, cette variation est essentiellement due aux effets de bout comme le démontre le résultat suivant. D'après l'équation (1) on obtient d'abord la moyenne des noeuds, \bar{Y}_N en posant $p'=1$.

$$\bar{Y}_N = \sum_{L=1}^k Y_{p'} = \sum_{L=1}^k \left[\frac{0 \cdot X_{L+1} + p X_L}{p k} \right] = \frac{\sum_{L=1}^k X_L}{k} = \bar{X}_N$$

* Les figures 1.1 à 1.3 et 2.1 à 2.3 sont toutes tirées de la série de 22 points de persistance 0.7. Cependant, ces figures ne montrent que les 161 premières valeurs pour fin de concision.

Tableau 1: Moyennes et variances d'une partition des différentes interpolations formées à partir d'une série de points noeuds sans persistance.

| p' | MOYENNE | | | VARIANCE | |
|----|------------------------|--------------------------------------|--|------------------------|------------------------------|
| | Interpolation linéaire | Interpolation homoscéastique centrée | Interpolation homoscéastique non-centrée | Interpolation linéaire | Interpolation homoscéastique |
| 1 | -.3315 | -.3315 | -.3315 | .6514 | .6514 |
| 2 | -.3599 | -.3640 | -.4111 | .5113 | .6672 |
| 3 | -.3643 | -.3744 | -.4764 | .3862 | .6606 |
| 4 | -.3687 | -.3866 | -.5467 | .2943 | .6471 |
| 5 | -.3731 | -.3902 | -.6076 | .2356 | .6251 |
| 6 | -.3774 | -.4091 | -.6371 | .2102 | .5988 |
| 7 | -.3818 | -.4135 | -.6281 | .2179 | .5781 |
| 8 | -.3862 | -.4126 | -.5726 | .2589 | .5691 |
| 9 | -.3906 | -.4088 | -.5108 | .3330 | .5696 |
| 10 | -.3950 | -.4040 | -.4512 | .4404 | .5747 |

Tableau 2: Moyennes et variances d'une partition des différentes interpolations formées à partir d'une série de points noeuds ayant une persistance de l'ordre de 0,7.

| p' | MOYENNE | | | VARIANCE | |
|----|------------------------|---------------------------------------|---|------------------------|-------------------------------|
| | Interpolation linéaire | Interpolation homoscédastique centrée | Interpolation homoscédastique non-centrée | Interpolation linéaire | Interpolation homoscédastique |
| 1 | -1.2628 | -1.2628 | -1.2628 | 1.2033 | 1.2033 |
| 2 | -1.3450 | -1.3499 | -1.4250 | 1.0031 | 1.1260 |
| 3 | -1.3463 | -1.3558 | -1.4996 | 0.9137 | 1.1337 |
| 4 | -1.3475 | -1.3610 | -1.5609 | 0.8484 | 1.1383 |
| 5 | -1.3488 | -1.3650 | -1.6020 | 0.8073 | 1.1388 |
| 6 | -1.3501 | -1.3674 | -1.6174 | 0.7933 | 1.1342 |
| 7 | -1.3514 | -1.3680 | -1.6050 | 0.7975 | 1.1249 |
| 8 | -1.3527 | -1.3669 | -1.5668 | 0.8288 | 1.1120 |
| 9 | -1.3539 | -1.3643 | -1.5081 | 0.8842 | 1.0971 |
| 10 | -1.3552 | -1.3607 | -1.4358 | 0.9638 | 1.0819 |

Si $p' \neq 1$

$$\sum_{L=1}^{k'} Y_{p'} = \sum_{L=1}^{k'} \left[\frac{(p'-1) X_{L+1} + (p-p'+1) X_L}{p k'} \right] \quad \text{où } k' = k-1$$

$$= \frac{\sum_{L=2}^{k'} X_L + \frac{(p-p'+1)}{p} X_1 + \frac{(p'-1)}{p} X_k}{k'}$$

$$\sum_{L=1}^k Y_{p'} = \frac{k}{k'} \bar{Y}_N - \underbrace{\left[\frac{(p'-1)}{p k'} X_1 + \frac{(p-p'+1)}{p k'} X_k \right]}_{\text{effet de bout}} \quad (6)$$

On voit bien que la variation est due aux valeurs de bout X_1 et X_k dont la présence plus ou moins grande selon le p' fait varier la moyenne de façon assez visible à cause, surtout, du faible nombre de valeurs représentant un p' ($k=22$ dans ce cas précis). Il est bon de noter que l'effet de bout tendra vers 0 à mesure que l'on augmentera la longueur de la série:

$$\lim_{N \rightarrow \infty} Y_{p'} = \lim_{k \rightarrow \infty} Y_{p'} = \lim_{k \rightarrow \infty} \left(\left[\frac{k}{k'} \bar{Y}_N \right] - \left[\frac{(p'-1)}{p k'} X_1 + \frac{(p-p'+1)}{p k'} X_k \right] \right)$$

On voit que $Y_{p'}$ tendra vers \bar{Y}_N par le fait même.

En ce qui concerne les variances, les troisièmes colonnes des tableaux 1 et 2 mettent en évidence la variation de la variance d'un p' à l'autre. Il fallait s'attendre à une telle variation car, comme on l'a vu dans la partie théorique on a une non-stationnarité d'un p' à l'autre. On s'aperçoit également que, comme prévu, la variance à l'intérieur de l'intervalle compris entre deux points noeuds est plus petite que la variance aux points noeuds. C'est pour contrer de tels effets que nous introduisons maintenant l'interpolation homoscedastique.

2. INTERPOLATION HOMOSCEDASTIQUE

2.1 Présentation

La forme de l'interpolation homoscedastique est sensiblement la même que pour l'interpolation linéaire. Son but est cependant de garder la variance constante d'un point interpolé à l'autre (d'un p' à l'autre), on utilise le résultat obtenu dans la partie 1.2.3, (équation 2):

$$\text{Var } (Y_{(L-1)p+p'}) = \text{Var } (X) \left[1 - \frac{2(1-r_1) (p'-1) (p-p'+1)}{p^2} \right]$$

pour former la série suivante:

$$\begin{aligned} Y_{(L-1)p+p'} &= \frac{(p'-1) X_{L+1} + (p-p'+1) X_L}{p} \cdot \left\{ \frac{p^2}{p^2 - 2(1-r_1) (p'-1) (p-p'+1)} \right\}^{\frac{1}{2}} \\ &= \frac{(p'-1) X_{L+1} + (p-p'+1) X_L}{[p^2 - 2(1-r_1) (p'-1)(p-p'+1)]^{\frac{1}{2}}} \end{aligned} \quad (7)$$

qui représente la série interpolée homoscedastiquement.

2.2 Résultats théoriques

2.2.1 Moyenne

Avec l'interpolation homoscedastique telle que définie par l'équation (7), on obtient une moyenne qui n'est pas stationnaire d'un p' à l'autre.

N.B. Pour fin de simplification $D(p')$ sera utilisé dans la suite de ce rapport pour remplacer le dénominateur de l'expression (7).

Ainsi la moyenne des points noeuds (\bar{X}_N) est:

$$\bar{X}_N = \sum_{L=1}^k \frac{X_L}{k+1} = \bar{Y}_N \text{ pour } p' = 1$$

si $p' \neq 1$ on aura:

$$\bar{Y}_{p'} = \frac{\sum_{L=1}^{k'} Y_{p'}}{k'} = \frac{1}{k'} \sum_{L=1}^{k'} \frac{(p'-1) X_{L+1} + (p-p'+1) X_L}{D(p')} \quad \text{où } k' = k-1$$

$D(p')$ est constant ici, car on somme sur une sous-série où chaque élément a le même p' ; on a alors:

$$\begin{aligned} \bar{Y}_{p'} &= \frac{p}{k' D(p')} \left[\sum_{L=2}^{k'} X_L + \frac{(p-p'+1) X_1 + (p'-1) X_k}{p} \right] \\ &= \underbrace{\frac{p}{k' D(p')} \left[\sum_{L=2}^{k'} X_L \right]}_{\text{effet sur la moyenne}} + \underbrace{\frac{1}{k' D(p')} [(p'-p'+1)X_1 + (p'-1)X_k]}_{\text{effet de bout}} \end{aligned}$$

$$\bar{Y}_{p'} = \frac{p}{D(p')} \bar{X}_N - \left[\frac{(p'-1) X_1 + (p-p'+1) X_k}{k' D(p')} \right] \quad (8)$$

On s'aperçoit alors que lorsque l'on prendra n et k suffisamment grand, l'effet de bout tendra vers 0, mais la moyenne de $\bar{Y}_{p'}$ ne tendra pas vers \bar{X}_N . De plus, chaque sous-série $\bar{Y}_{p'}$ tendra vers une moyenne différente et la moyenne globale de la série interpolée tendra vers:

$$\frac{p}{E(D(p'))} \cdot \bar{X}_N$$

Ces résultats sont peu intéressants car s'il faut sacrifier la stationnarité de la moyenne pour avoir la stationnarité en variance: ce n'est que déplacer le problème de place.

Il semble donc plus intéressant d'utiliser une interpolation homocédastique centrée:

$$Y_{p(L-1)+p'} = \frac{(p'-1)(X_{L+1} - \bar{X}_N) + (p-p'+1)(X_L - \bar{X}_N)}{D(p')} + \bar{X}_N$$

On obtient alors les résultats suivants:

$$\bar{X}_N = \frac{\sum_{L=1}^k X_L}{k+1} = \bar{Y}_N \text{ pour } p' = 1$$

$$\text{et } \bar{Y}_{p'} = \frac{p + k' D(p')}{k D(p')} \bar{X}_N - \frac{p}{n D(p')} \left[\frac{(p'-1) X_1 + (p-p'+1) X_k}{p} \right] \quad (9)$$

On s'aperçoit qu'alors, toutes les sous-séries Y_p , tendront en moyenne vers \bar{X}_N , (les effets de bouts tendant vers 0 et la fonction devant \bar{X}_N tendant vers 1), lorsque n (k par le fait même) tend vers l'infini.

Ces résultats apparaissent beaucoup plus intéressants.

2.2.2 Variance

Prenons la sous série $\{Y_{p(L-1)+p'}\}_{L=1}^k$ avec un p' constant, on a alors:

$$\text{Var}(Y_{p'}) = \text{Var}\left(\frac{(p'-1) X_{L+1} + (p-p'+1) X_L}{D(p')}\right)$$

$$= \frac{1}{[D(p')]^2} \text{Var} ((p'-1) X_{L+1} + (p-p'+1) X_L)$$

car $D(p')$ est constant si p' est constant

$$= \frac{1}{[D(p')]^2} [(p'-1)^2 \text{Var}(X_{L+1}) + (p-p'+1)^2 \text{Var}(X_L) + 2(p'-1)(p-p'+1) \text{Cov}(X_L, X_{L+1})]$$

$$= \frac{[(p'-1)^2 + (p-p'+1)^2] \text{Var}(X_L) + 2(p'-1)(p-p'+1) \text{Cov}(X_L, X_{L+1})}{[D(p')]^2}$$

$$= \frac{\text{Var}(X_L) [p^2 - 2(p-1)(p-p'+1) + 2r_1(p'-1)(p-p'+1)]}{[D(p')]^2} = \frac{\text{Var}(X_L) \cdot [D(p')]^2}{[D(p')]^2}$$

$$= \text{Var}(X_L) \text{ pour tout } p'$$

Ainsi, on s'aperçoit que, comme on le voulait, la variance est maintenant stationnaire d'un p' à l'autre. Cependant, la démonstration précédente ne tient pas compte des effets de bout.

De plus on a utilisé l'interpolation homoscédastique non-centrée, mais des résultats identiques sont obtenus avec une interpolation homoscédastique centrée.

2.2.3 Autocovariance et fonctions d'autocorrélation

Le dernier résultat théorique présenté, n'implique cependant pas une stationnarité du deuxième ordre, car encore une fois, les autocovariances ne sont pas stationnaires d'un p' à l'autre, cette dernière condition étant nécessaire afin d'avoir la stationnarité recherchée (BOX et JENKINS, 1976))*.

* p. 30.

Pourtant encore une fois, la non-stationnarité disparaît lorsque l'on prend une vue plus macroscopique de la série.

Pour ce qui est des fonctions d'autocorrélations, il sera montré dans la partie 3.3 qu'elles sont identiques aux fonctions d'autocorrélation des séries interpolées linéairement.

2.3 Résultats empiriques

2.3.1 Moyenne

Les 3e et 4e colonnes des tableaux 1 et 2 permettent de bien illustrer les points soulevés dans la partie théorique. Ainsi, il est facile de s'apercevoir que la série homoscedastique centrée à une moyenne beaucoup plus stable (d'un p' à l'autre) que la série non-centrée. De plus, on voit également que la moyenne globale sera beaucoup plus près de la moyenne des points noeuds dans le cas de la série centrée.

La figure 2.1 nous montre l'allure générale d'une interpolation homoscedastique par rapport à une interpolation linéaire. On s'aperçoit qu'au dessus d'un certain axe, les points interpolés homoscedastiquement sont plus grands que les points interpolés linéairement, tandis qu'au dessous de cet axe, les points interpolés homoscedastiquement sont plus petits que les points interpolés linéairement. C'est la position de cet axe, qui peut amener des problèmes de non-stationnarité de la moyenne.

La figure 2.2 illustre l'interpolation homoscedastique non-centrée. Dans ce cas, l'axe susmentionné est représenté par l'équation $Y_{p'}=0$. Donc tout point interpolé linéairement qui est plus grand que zéro sera plus petit que le point interpolé homoscedastiquement qui lui est relié par l'équation (7). Le phénomène inverse se produisant si le point interpolé linéairement est plus petit que zéro. Ces résultats découlent directement du fait que $D(p')$ est toujours dans l'intervalle $]0,1]$.

Ainsi, lorsque la moyenne des points noeuds \bar{X}_N sera plus grande que zéro, il faudra s'attendre à avoir plus de points positifs que de points négatifs, c'est ce qui amène une moyenne de la série interpolée homoscédastiquement (non-centrée) plus grande que la moyenne des points noeuds. De même si \bar{X}_N est plus petit que zéro, la moyenne de la série interpolée homoscédastiquement (non-centrée) sera plus petite que \bar{X}_N .

Dans le cas de l'interpolation homoscédastique centrée (figure 2.3), l'axe devient $Y_p = \bar{X}_N$. On aura alors:

- $X_{p'IHC} > X_{p'IL}$ si $X_{p'IL} > \bar{X}_N$
- $X_{p'IHC} < X_{p'IL}$ si $X_{p'IL} < \bar{X}_N$

où

| | |
|-------------|--|
| $X_{p'IHC}$ | = point de l'interpolation homoscédastique centrée |
| $X_{p'IL}$ | = point de l'interpolation linéaire |
| \bar{X}_N | = moyenne des points noeuds. |

Ces résultats amène directement le résultat asymptotique $\bar{X}_{IHC} = \bar{X}_N$.

2.3.2 Variance

Les dernières colonnes des tableaux 1 et 2 regroupent les interpolations homoscédastiques centrée et non-centrée car les résultats sont exactement semblables à la quatrième décimale. Ceci est en complet accord avec nos résultats théoriques.

Il est aussi intéressant de remarquer que nous avons considérablement stabilisé la variance comparativement à l'interpolation linéaire.

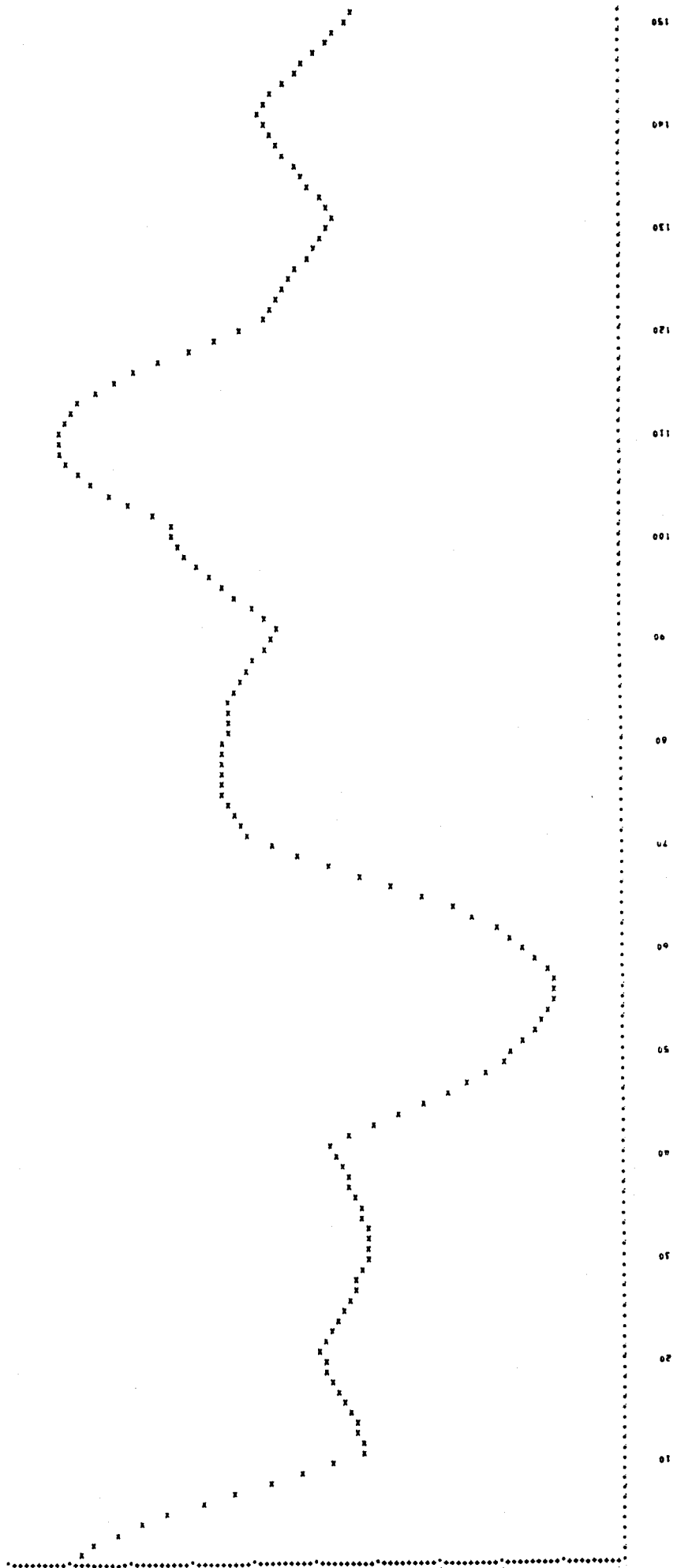


Figure 2.1: Interpolation homoscédastique formée à partir des mêmes points noeuds que la série de la figure 1.1.

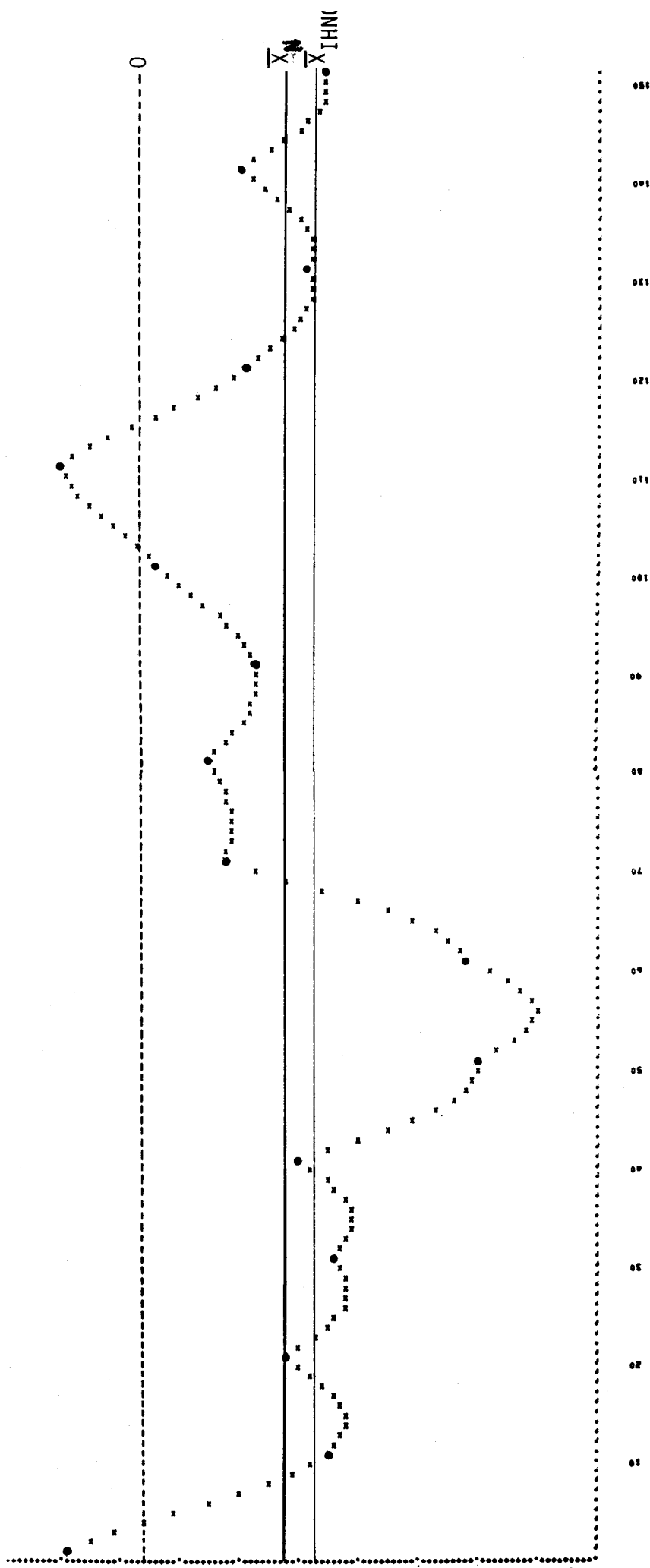


Figure 2.2: Interpolation homogénéistique non-centrée.

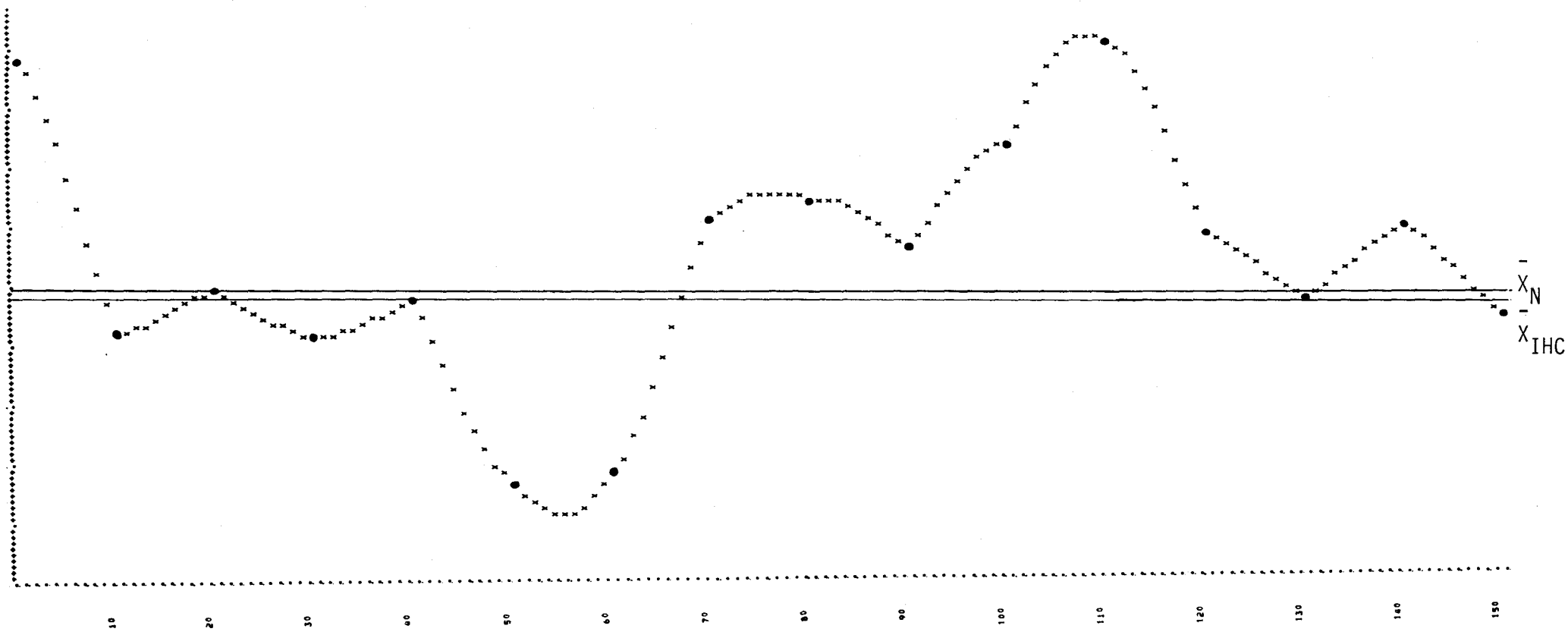


Figure 2.3: Interpolation homoscedastique centrée.

3. COMPARAISON ENTRE LES DEUX TYPES D'INTERPOLATION

3.1 Moyenne

Nous comparerons maintenant, de façon relative les différentes interpolations au niveau de la moyenne. L'interpolation linéaire étant stationnaire de par sa définition, nous l'utiliserons comme "standard". Le tableau 3 nous montre les écarts relatifs des deux autres interpolations vis-à-vis celle-ci. Il est alors facile de voir que l'interpolation homoscédastique centrée est beaucoup plus près de la stationnarité que l'interpolation homoscédastique non-centrée.

On s'aperçoit également que les séries homoscédastiques formées à partir d'une série de points noeuds persistante est beaucoup plus près, en moyenne, de l'interpolation linéaire. Cela est dû à la moins grande importance de $D(p')$ à mesure que l'autocorrélation d'ordre 1 augmente (r_1).

3.2 Variance

Il est facile de s'apercevoir, avec les tableaux 1 et 2 que la variance est beaucoup plus stable pour l'interpolation homoscédastique (centrée ou non) que pour l'interpolation linéaire. De fait, l'écart entre la plus grande variance ($p'=1$, 0.6514) et la plus petite ($p'=6$, 0.2102) pour l'interpolation linéaire est de l'ordre de 200% comparativement à 14.46% pour l'interpolation homoscédastique ($p'=1$, 0.6514 et $p'=8$, 0.5691 étant les valeurs extrêmes).

3.3 Autocovariance et fonction d'autocorrélation

Il nous apparaît intéressant de citer ici que, pour un p' donné, on aura la même fonction d'autocorrélation pour l'interpolation linéaire et l'interpolation homoscédastique.

Regardons d'où vient ce résultat:

Tableau 3: Ecart relatif des moyennes des séries interpolées homoscédastiquement versus l'interpolation linéaire

| p' | Série de points noeuds sans persistance | | Série des points noeuds avec persistance | |
|----|---|--|---|--|
| | Ecart entre \bar{X}_{IHC} et \bar{X}_{IL} pour chaque sous-ensemble | Ecart entre \bar{X}_{IHNC} et \bar{X}_{IL} pour chaque sous-ensemble | Ecart entre \bar{X}_{IHC} et \bar{X}_{IL} pour chaque sous-ensemble | Ecart entre \bar{X}_{IHNC} et \bar{X}_{IL} pour chaque sous-ensemble |
| 1 | 0.0% | 0.0% | 0.0% | 0.0% |
| 2 | 1.1% | 14.2% | 0.4% | 5.9% |
| 3 | 2.8% | 30.8% | 0.7% | 11.4% |
| 4 | 4.9% | 48.3% | 1.0% | 15.8% |
| 5 | 7.0% | 62.9% | 1.2% | 18.8% |
| 6 | 8.4% | 68.8% | 1.3% | 19.8% |
| 7 | 8.3% | 62.9% | 1.2% | 18.8% |
| 8 | 6.8% | 48.3% | 1.0% | 15.8% |
| 9 | 4.7% | 30.8% | 0.8% | 11.4% |
| 10 | 2.3% | 14.2% | 0.4% | 5.9% |

Soit $\{X_L\}_{L=1}^k$: la série des points noeuds et $\{Y_t\}_{t=1}^{(n-1)p+1}$: la série interpolée linéairement; on a vu que:

$$\text{Var}(Y_{p'}) = \text{Var}(X_L) [F(p')]$$

$$\text{posons Cov}(Y_{p'}, Y_{p'+k}) = \text{ACV}(k)$$

notre interpolation homoscedastique a, par définition, la forme:

$$w_t = \frac{Y_t}{F(p')}$$

afin de garder la variance constante.

Voyons alors l'allure des fonctions d'autocorrelation.

LINÉAIREMENT

$$\begin{aligned} \rho_k &= \frac{\text{ACV}(k)}{\text{Var}(Y_{p'}) \cdot \text{Var}(Y_{p'+1})} \\ &= \frac{\text{ACV}(k)}{\text{Var}(X) [F(p')] \cdot \text{Var}(X) [F(p'+k)]} \\ &= \frac{\text{ACV}(k)}{\text{Var}(X) [F(p') \cdot F(p'+k)]} \end{aligned}$$

HOMOSCEDASTIQUEMENT

$$\begin{aligned} \text{Cov}(w_{p'}, w_{p'+k}) &= \text{Cov}\left(\frac{Y_{p'}}{F(p')}, \frac{Y_{p'+k}}{F(p'+k)}\right) \\ &= \frac{1}{F(p') \cdot F(p'+k)} \cdot \text{Cov}(Y_{p'}, Y_{p'+k}) \\ &= \frac{\text{ACV}(k)}{F(p') \cdot F(p'+k)} \end{aligned}$$

et

$$\rho_k = \frac{\text{Cov}(w_{p'}, w_{p'+k})}{\text{Var}(w_{p'}) \cdot \text{Var}(w_{p'+k})}$$

on a:

$$\rho_k = \frac{\text{ACV}(k)}{\text{Var}(X) [F(p') \cdot F(p'+k)]}$$

Si l'on prend les deux séries complètes un tel résultat n'est pas totalement vérifié. Des expériences ont cependant montré que pour une série de 211 valeurs, la différence absolue entre un coefficient d'autocorrélation d'une série interpolée linéairement et d'une autre interpolée homoscédastiquement, ne dépassait pas 0.02 (voir tableaux 4a et 4b).

Il faut noter que de tels résultats ne tiennent que pour les séries qui n'ont pas été transformées. En effet lorsque l'on effectue une différenciation, (opérateur ∇) ou toute autre transformation, l'égalité des fonctions d'autocorrélation, pour un même p' , n'existe plus.

MEAN OF THE SERIES = $-.13410E+01$
 ST. DEV. OF SERIES = $.95194E+00$
 NUMBER OF OBSERVATIONS = 211

| | | | | | | |
|--------|------|------|------|------|------|------|
| 1- 6 | .97 | .94 | .90 | .85 | .80 | .75 |
| ST.E. | .07 | .12 | .15 | .17 | .19 | .21 |
| 7- 12 | .69 | .64 | .58 | .53 | .48 | .44 |
| ST.E. | .22 | .23 | .24 | .24 | .25 | .25 |
| 13- 18 | .39 | .35 | .31 | .27 | .23 | .20 |
| ST.E. | .26 | .26 | .26 | .26 | .27 | .27 |
| 19- 24 | .16 | .14 | .11 | .08 | .06 | .04 |
| ST.E. | .27 | .27 | .27 | .27 | .27 | .27 |
| 25- 30 | .02 | .01 | -.01 | -.02 | -.03 | -.04 |
| ST.E. | .27 | .27 | .27 | .27 | .27 | .27 |
| 31- 36 | -.06 | -.07 | -.08 | -.09 | -.10 | -.12 |
| ST.E. | .27 | .27 | .27 | .27 | .27 | .27 |

Tableau 4a: Fonction d'autocorrélation de l'interpolation linéaire.

MEAN OF THE SERIES = $-.13517E+01$
 ST. DEV. OF SERIES = $.10632E+01$
 NUMBER OF OBSERVATIONS = 211

| | | | | | | |
|--------|------|------|------|------|------|------|
| 1- 6 | .98 | .95 | .91 | .86 | .81 | .75 |
| ST.E. | .07 | .12 | .15 | .17 | .19 | .21 |
| 7- 12 | .70 | .64 | .59 | .54 | .49 | .44 |
| ST.E. | .22 | .23 | .24 | .25 | .25 | .26 |
| 13- 18 | .40 | .35 | .31 | .27 | .24 | .20 |
| ST.E. | .20 | .26 | .26 | .27 | .27 | .27 |
| 19- 24 | .17 | .14 | .11 | .09 | .06 | .04 |
| ST.E. | .27 | .27 | .27 | .27 | .27 | .27 |
| 25- 30 | .03 | .01 | .00 | -.02 | -.03 | -.04 |
| ST.E. | .27 | .27 | .27 | .27 | .27 | .27 |
| 31- 36 | -.06 | -.07 | -.08 | -.09 | -.10 | -.12 |
| ST.E. | .27 | .27 | .27 | .27 | .27 | .27 |

Tableau 4b: Fonction d'autocorrélation de l'interpolation homoscedastique.

Conclusion

Dans ce rapport, deux modes de génération de données à un pas plus fin que celui des mesures sont étudiés: l'interpolation linéaire classique et l'interpolation homoscédastique centrée.

L'interpolation linéaire est un estimateur local dépendant uniquement de deux points de mesures successifs et non des paramètres de l'échantillonnage; la série ainsi générée n'est pas stationnaire en variance.

L'interpolation homoscédastique centrée est, quant à elle, plus stable en variance, mais les valeurs qu'elle génère dépendent des caractéristiques globales de l'échantillon (moyenne, variance, coefficient d'autocorrélation de la série des points-noeuds); dans cette optique, elle est tout à fait inadaptée à des évaluations progressives.

Notons enfin que les corrélogrammes des séries interpolées par les deux techniques sont identiques; ces interpolations augmentent la persistance de la série des points noeuds, et cette augmentation est d'autant plus forte que la série originale était peu persistante.

Bibliographie

BOBÉE, B. et D. CLUIS.

Caractéristiques statistiques de séries générées par interpolation linéaire. INRS-Eau, Rapport scientifique no 121, 87 pages. 1980.

BOX, P.E. et G.M. JENKINS.

Time series analysis: forecasting and control. San Francisco, Holden-Day, 575 pages. 1976.

PANKRATZ, A.

Forecasting with univariate Box-Jenkins models: concept and cases. New York, John Wiley & Sons, 525 pages. 1983.