Record Number: 1220
Author, Monographic: Cluis, D.//Bobée, B.
Author Role:
Title, Monographic: Caractéristiques statistiques de séries générées par interpolation linéaire
Translated Title:
Reprint Status:
Edition:
Author, Subsidiary:
Author Role:
Place of Publication: Québec
Publisher Name: INRS-Eau
Date of Publication: 1980
Original Publication Date: Juillet 1980
Volume Identification:
Extent of Work: ix, 106
Packaging Method: pages incluant 3 annexes
Series Editor:
Series Editor Role:
Series Title: INRS-Eau, Rapport de recherche
Series Volume ID: 121
Location/URL:
ISBN: 2-89146-118-5
Notes: Rapport annuel 1980-1981
Abstract: 15.00$

Keywords: rapport/ ok/ dl

CARACTERISTIQUES STATISTIQUES
DE SERIES GENEREES
PAR INTERPOLATION LINEAIRE

RAPPORT SCIENTIFIQUE No 121

Juillet 1980

par

D. Cluis et B. Bobée

INRS-Eau
Université du Québec
C.P. 7500
Ste-Foy, Québec
G1V 4C7

ISBN 2-89146-118-5
DEPOT LEGAL 1980

CARACTERISTIQUES STATISTIQUES
DE SERIES GENEREES
PAR INTERPOLATION LINEAIRE

RAPPORT SCIENTIFIQUE No 121

Juillet 1980

par

D. Cluis et B. Bobée

INRS-Eau
Université du Québec
C.P. 7500
Ste-Foy, Québec
G1V 4C7

RESUME

L'interpolation linéaire est l'un des modes d'estimation des plus utilisés pour générer dans la pratique des valeurs à une fréquence supérieure à celle de la mesure d'une variable hydrologique; les caractéristiques statistiques de cet estimateur non optimal sont étudiées; on considère, en particulier, les coefficients d'autocorrélation et de la variance de la moyenne de la série interpolée en fonction des mêmes caractéristiques de la série de base; ces calculs sont effectués dans le cas général d'une série de base quelconque et appliqués à quelques cas particuliers simples tels que série de base indépendante, de faible persistance ou Markovienne d'ordre 1; quand cela est possible analytiquement, les termes de boût sont pris en compte et le biais introduit en les négligeant est estimé. Finalement, une évaluation quantitative globale est effectuée en utilisant la notion de contenu en information.

MOTS-CLES

Interpolation linéaire / estimation statistique / variance / autocovariance / coefficients d'autocorrélation / persistance

ABSTRACT

Iterated linear interpolation is one of the most common and practical estimating technique used to generate values at a frequency higher than that of the measured hydrological variate. The statistical properties of this non-optimal estimator are studied in particular the variance, autocovariances, coefficients of autocorrelation and variance of the mean of the interpolated series are calculated as functions of the same caracteristics of the original series. The calculations are developed for the case of a general generating series and applied to some simple cases as an independent series, a short-term persistent series and a Markovian series of lag one. When analytically feasible, end-effects are taken into account and the biases introduced by neglecting them are estimated. Finally a quantitative evaluation is made, using the concept of content of information.

KEYWORDS

Linear interpolation / statistical estimation / variance / autocovariance / autocorrelation / coefficient / persistence

TABLE DES MATIERES

PAGE

LISTE DES FIGURES

LISTE DES TABLEAUX

NOTATIONS

X_i	série de base
Y_j	série interpolée
i	indice variant de 1 à k
j	indice variant de 1 à (k - 1) P + 1
p	entier relatif au pas de l'interpolation, degré de la partition
p'	entier variant de 1 à p
s	somme des X_i
S	somme des Y_j
$\overline{X}$	moyenne des X_i
$\overline{Y}$	moyenne des Y_j
var X	variance des X_i
var Y	variance des Y_j
$ACV(Y)_t$	autocovariance d'ordre t de la série Y_j
r_1	coefficient d'autocorrélation d'ordre 1 de la série X_i
r_2	coefficient d'autocorrélation d'ordre 2 de la série X_i
R_t	coefficient d'autocorrélation d'ordre t de la série Y_j

(suite)

N	borne de variation supérieure de l'indice j
$(Cv)_X$	coefficient de variation de la série X_i
$(Cv)_Y$	coefficient de variation de la série Y_j

INTRODUCTION

La plupart des développements théoriques récents en hydrologie ont été axés sur les modèles de simulation de processus; il s'agit essentiellement d'une optique prospective où les résultats des recherches peuvent être utilisés pour de la prévision à plus ou moins long terme.

Beaucoup moins d'efforts ont été consacrés à des techniques anciennes et très largement utilisées comme l'interpolation linéaire, mais qui procèdent à une évaluation rétrospective; la technique est alors utilisée une fois les données acquises, soit pour compléter des valeurs absentes ou réestimer des valeurs apparemment aberrantes; elle peut être aussi utilisée pour fins d'interprétation ou pour le calcul d'autres quantités dérivées; par exemple, l'interpolation des concentrations est utilisée de façon classique pour le calcul de débits massiques ou de bilans, quand les débits liquides sont disponibles à une fréquence plus élevée que celle des échantillonnages de qualité.

Dans ce rapport, nous préciserons certaines règles d'utilisation de la technique et nous effectuerons les calculs complets et exacts de la variance et des coefficients d'autocorrélation d'une série interpolée linéairement en évaluant l'effet des termes de boût et l'influence de la persistance de la série originale.

1. Conditions d'utilisation de l'interpolation linéaire

Le but du présent chapitre n'est pas d'établir en toute rigueur mathématique la valeur de l'interpolation linéaire comme estimateur de réalisations non mesurées d'une série de temps selon la structure de cette série, mais plutôt d'évaluer globalement les caractéristiques des séries pour lesquelles cette opération est efficace en pratique, c'est-à-dire qu'elle réduit la perte d'information dans des opérations mathématiques nécessitant un pas de temps plus faible que le pas de mesure.

1.1 Définition de l'interpolation linéaire

Considérons un processus continu $\Phi(t)$; ce processus possède des réalisations à chaque instant, mais en pratique, les séries sont généralement obtenues à partir de mesures effectuées à des intervalles de temps équidistants p pendant une durée P. Il peut arriver, une fois les données acquises, que l'on soit intéressé à avoir une estimation de la valeur prise par Φ à des pas de temps plus petits que p, par exemple à une fraction de cet intervalle de mesure. Un des modes d'estimation des valeurs désirées est l'interpolation linéaire qui est une moyenne pondérée des deux réalisations aux temps t_1 et t_2 entourant la réalisation à estimer, on a, à un temps t':

$$\Phi_{t'} = \frac{(t_2 - t')\ \Phi_2 + (t' - t_1)\ \Phi_1}{t_2 - t_1} \qquad [\ 1\ ]$$

avec $t_2 - t_1 = P$

Φ_2 et Φ_1 étant les valeurs prises par la variable aux temps t_2 et t_1

et $\Phi_{t'}$ la valeur de Φ estimée au temps t'.

On doit noter que cette estimation ne fait appel qu'à 2 réalisations successives du processus et qu'il existe d'autres modes d'estimation plus complexes faisant appel à davantage d'information comme les polynomes orthogonaux ou les ajustements de fonctions (spline).

1.2 Décomposition d'une série de temps

Dans sa forme la plus générale, une série de temps de longueur indéfinie peut se décomposer en:

$$\Phi_{(t)} = R_{(t)} + P_{(t)} + \eta_{(t)} \qquad [2]$$

où $R_{(t)}$ représente la composante de tendance, $P_{(t)}$ la composante périodique et $\eta_{(t)}$ la composante stochastique ou purement aléatoire.

Ces trois éléments étant indépendants, la composition des variances est très simple:

$$\text{var } \Phi = \text{var } R + \text{var } P + \text{var } \eta \qquad [3]$$

On démontre par ailleurs que toute série périodique peut se décomposer de façon unique en une somme de séries trigonométriques construites sur une fréquence fondamentale ω:

$$P = \sum_{j=1}^{\infty} C_j \cos \frac{2\pi t}{\omega} j + \Theta_j \qquad [4]$$

Dans cette représentation $\text{var } P = \frac{1}{2} \sum_{j=1}^{\infty} C_j^2$

On peut noter ici une des difficultés pratiques inhérentes à l'estimation des coefficients C_j et Θ_j à partir d'une série de longueur finie: La série historique contient des nombres variables de cycles reliés aux fréquences contenues, ce qui conduit à des précisions variables dans l'identification des coefficients par la méthode des moindres carrés par exemple. Comme on procède généralement de façon soustractive et récurrente, à partir des basses fréquences en commençant par les tendances, on introduit dans la série résiduelle toutes les incertitudes d'ajustement des coefficients des séries de basse fréquence, courtes en nombre de cycles, ce qui rend très aléatoires les résultats obtenus sur les cycles de haute fréquence; on a même montré que cette méthode de déconvolution peut introduire les fréquences absolument absentes de la série originale.

1.3 Résolution en fréquence d'une série de mesures de longueur finie

On considère une série de longueur n, de mesures effectuées avec un pas de temps p, pour une durée totale $P = (n - 1)\,p$. Avec l'information ainsi recueillie, que peut-on dire de la population-mère $\phi(t)$? Du côté des basses fréquences, il faut de 8 à 10 cycles pour confirmer une périodicité, donc les fréquences plus basses que

$$\omega_1 = \frac{2\pi}{P/10} = \frac{20\pi}{(n-1)p}$$

ne peuvent être distinguées des termes de tendances.

Du côté des hautes fréquences, la fréquence la plus élevée pour laquelle l'échantillon contient de l'information est la fréquence dite de Nyquist $\omega_0 = \frac{\Pi}{p}$; les composantes harmoniques de fréquences supérieures ne peuvent être distinguées par les mesures des termes aléatoires. On peut donc réécrire l'équation [3] en fonction des fréquences de résolution de la série de mesures:

$$\text{var } \Phi = [\text{var } R + \text{var } P_1] + [\text{var } P_2] + [\text{var } P_3 + \text{var } \eta] \qquad [\ 5\ ]$$

où • P_1 représente les harmoniques de la population-mère de fréquences inférieures à ω_1,

• P_2 les harmoniques de fréquences comprises entre ω_1 et ω_2,

• P_3 les harmoniques de fréquences supérieures à ω_2,

ce qui s'écrit, avec les mêmes indices:

$$\text{var } \Phi = \text{var } \Phi_1 + \text{var } \Phi_2 + \text{var } \Phi_3 \qquad [\ 6\ ]$$

Ces résultats apparaissent au tableau 1.

1.4 Intérêt de l'interpolation linéaire

Une interpolation linéaire est toujours possible mais elle présente un intérêt particulier quand les valeurs interpolées représentent des meilleurs estimateurs de la série réelle (non mesurée aux pas de temps intermédiaires) que les valeurs les plus probables de la totalité ou

Termes de tendance	Basse fréquence	ω_1		ω_2	Haute fréquence	Termes aléatoires
var R	var P_1	$\frac{20\pi}{(n-1)\,p}$	var P_2	$\frac{\pi}{p}$	var P_3	var η
var Φ_1			var Φ_2		var Φ_3	

Tableau 1: Décomposition de la variance d'une série de temps.

d'une partie de la population-mère, c'est-à-dire des moyennes générales ou locales. On note donc que la persistance de la série est un élément favorable à l'usage de l'interpolation linéaire. Intuitivement, on peut dire que cet estimateur sera d'autant meilleur que le rapport

$\frac{\text{var } \Phi_1 + \text{var } \Phi_2}{\text{var } \Phi}$ sera proche de 1 ou que le rapport $\frac{\text{var } \Phi_3}{\text{var } \Phi}$ sera faible. Compte tenu des difficultés <u>pratiques</u> à obtenir la décomposition réelle des variances, l'évaluation théorique de la précision de l'estimateur "interpolation linéaire" en fonction des valeurs de ce rapport, ne sera pas poussée plus avant ici. La figure 1 met en évidence trois cas de répartition des variances et les conséquences qui en découlent pour l'interpolation linéaire.

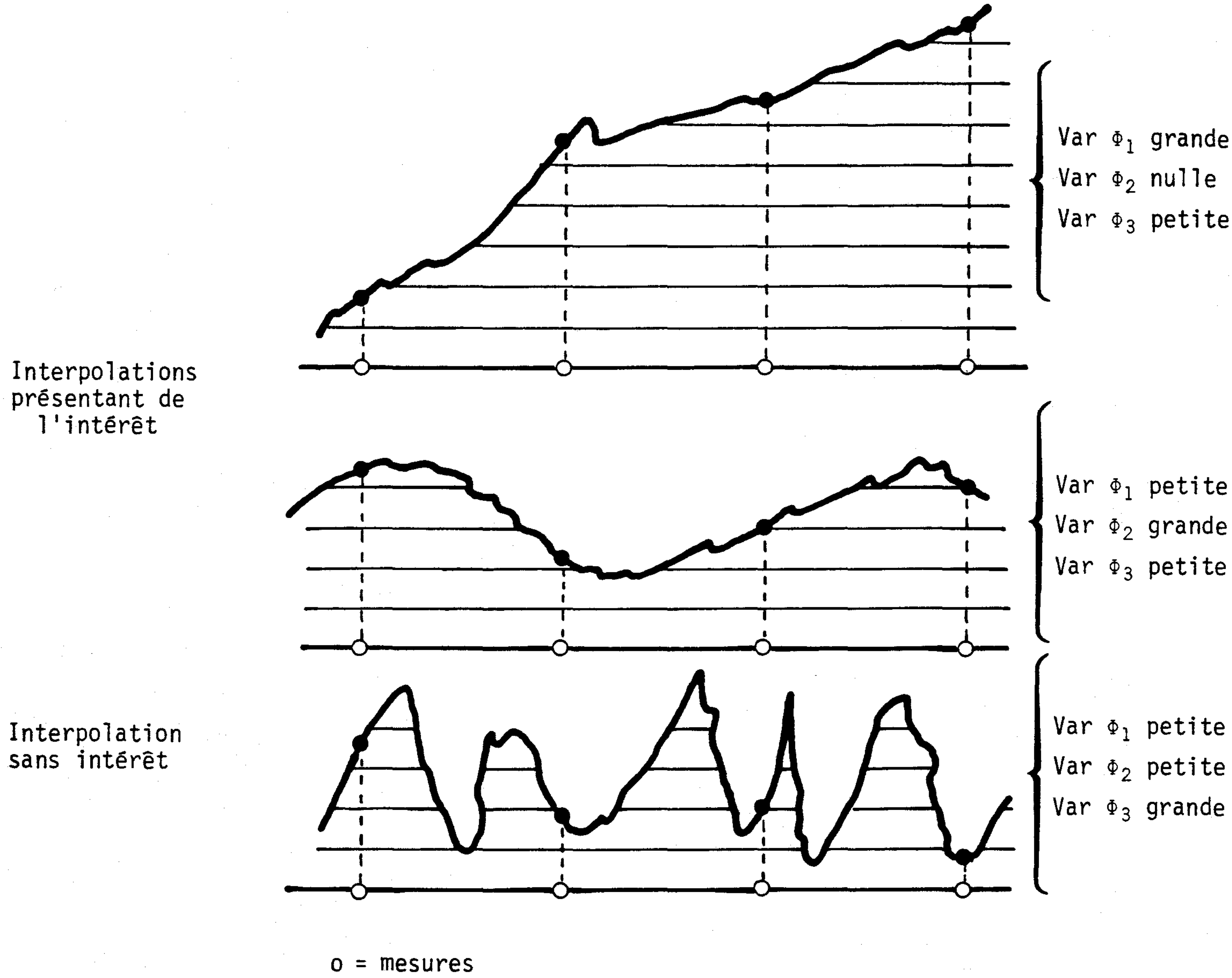


FIGURE 1. Exemple de séries de temps sujettes à interpolation.

2. Caractéristiques générales d'une série interpolée

2.1 Définitions de la série de base X_i et de la série interpolée Y_j

On considère une série X_i où i varie de 1 à k; l'intervalle entre deux valeurs consécutives de la série X_i est subdivisée en p intervalles égaux; on obtient ainsi par interpolation une série Y_j où j varie de 1 à (k - 1) p + 1 . Si p = 1, on retrouve la série originale X_i . D'où le schéma de définition:

$$\begin{array}{lllllll}
X_i = & 1 & 2 & 3 & 4 & 5 & i = k \\
Y_j = & \underbrace{1\ 2 \quad\quad}_{p} & \underbrace{p+1 \quad\quad}_{p} & \underbrace{2p+1 \quad\quad}_{p} & \underbrace{3p+1 \quad\quad}_{p} & 4p+1 & j = (k-1)\ p+1
\end{array}$$

Chaque élément de la série Y_j s'écrit:

$$Y_j = Y_{(i-1)\ p\ +\ p'} = \frac{(p' - 1)\ X_{i+1} + (p - p' + 1)\ X_i}{p} \qquad [\ 7\ ]$$

où i est un entier variant de 1 à k,

p' est un entier variant de 1 à p,

et p est un entier relatif au pas d'interpolation.

j varie donc de 1 à (k-1) p + 1 et l'on doit noter, pour la suite des sommations que, pour i = k, p' ne prend que la valeur 1, ce qui correspond au dernier élément $Y_{(k-1)\ p\ +\ 1} \equiv X_k$; dans cette indiciation, k et p sont des indices fixes, j, i et p' sont variables. Les relations [B - 1] à [B - 4], développées en Annexe 2, s'appliquent à la série X_i comme à la série Y_j .

2.2 Moyenne des Y_j en fonction des caractéristiques de la série originale

$$S = \sum_{j=1}^{(k-1)\,p+1} Y_j = [(k-1)\,p + 1]\,\overline{Y}$$

$$s = \sum_{i=1}^{k} X_i = k\overline{X}$$

$\overline{Y}$ et $\overline{X}$ représentent les moyennes de la série interpolée et de la série originale respectivement.

D'après [1] $$S = \frac{1}{p}\left\{\sum_{p'=1}^{p}\sum_{i=1}^{k-1} (p'-1)\,X_{i+1} + (p-p'+1)\,X_i\right\} + X_k$$

D'après l'équation [A-1] de l'Annexe 1, cette équation s'écrit:

$$S = \frac{1}{p}\left\{\frac{p\,(p-1)}{2}\,(s - X_1) + \frac{p\,(p+1)}{2}\,(s - X_k)\right\} + X_k$$

En développant, il vient:

$$S = ps - \frac{p-1}{2}\,(X_1 + X_k) \qquad [\ 8\ ]$$

et $$\overline{Y} = \frac{S}{(k-1)\,p + 1} = \frac{2ps - (p-1)\,(X_1 + X_k)}{2\,[(k-1)\,p + 1]}$$

$$\overline{Y} = \frac{2pk\overline{X} - (p-1)(X_1 + X_k)}{2\,[(k-1)\,p + 1]} \qquad [\ 9\ ]$$

Cas particulier

Si on ne connaît pas les valeurs de X_1 et X_k et que l'on les estime par la valeur la plus probable $\overline{X}$, on obtient des estimations biaisées: pour $X_1 = X_k = \overline{X}$ les équations [8] et [9] deviennent:

$$S = \overline{X} \; (p\,k - p + 1)$$

$$\overline{Y} = \overline{X} \; \frac{p\,k - p + 1}{(k-1)\,p + 1} = \overline{X}$$

2.3 Variance des Y_j en fonction des caractéristiques de la série originale

Notons var X et r_1 respectivement la variance et le coefficient d'autocorrélation d'ordre 1 de la série X_i, tels que définis à l'Annexe 2:

Par définition, on a:

$$\text{var } Y = \frac{1}{(k-1)\,p} \sum_{j=1}^{N} (Y_j - \overline{Y})^2 \qquad \text{avec } N = (k-1)\,p + 1$$

ce qui s'écrit:

$$\text{var } Y = \frac{\sum_{j=1}^{N} Y_j^2 - [(k-1)\,p + 1]\,\overline{Y}^2}{(k-1)\,p}$$

Nous allons évaluer successivement $\sum Y^2$ et $\overline{Y}^2$:

2.3.1 Calcul de $\sum_{j=1}^{N} Y_j^2$

D'après l'équation [7]:

$$\sum_{j=1}^{N} Y_j^2 = \sum_{p'=1}^{p} \frac{(p'-1)^2}{p^2} \sum_{1}^{k-1} X_{i+1}^2 + \sum_{p'=1}^{p} \frac{(p-p'+1)^2}{p^2} \sum_{1}^{k-1} X_i^2 + X_k^2$$

$$+ 2 \sum_{p'=1}^{p} \frac{(p'-1)\ (p-p'+1)}{p^2} \sum_{1}^{k-1} X_i\ X_{i+1}$$

On note que les deux premiers termes constituent des sommes complètes à l'exception du terme en X_1 et X_k; d'après [B-1], pour les termes carrés, [A-3] et [A-5] pour les sommations d'entiers et [B-2] pour les termes produits, on a:

$$\sum_{j=1}^{N} Y_j^2 = \frac{(p-1)\ p\ (2p-1)}{6\ p^2} \left[(k-1) \text{ var } X + k\ \overline{X}^2 - X_1^2\right] +$$

$$p\ \frac{(p+1)\ (2p+1)}{6\ p^2} \left[(k-1) \text{ var } X + k\ \overline{X}^2 - X_k^2\right] + X_k^2 +$$

$$\frac{2\ p\ (p^2-1)}{6\ p^2} \left[r_1\ (k-1) \text{ var } X + (k+1)\ \overline{X}^2 - \overline{X}\ (X_1 + X_k)\right]$$

Cette expression introduit le coefficient d'autocorrélation r_1 de la série X_i, défini à l'Annexe 2.

L'expression précédente se développe:

$$\sum_{j=1}^{N} Y^2 = \frac{2\ p^2 - 3\ p + 1}{6\ p} \left[(k-1)\ \text{var}\ X + k\ \overline{X}^2 - X_1{}^2\right] + \frac{2\ p^2 + 3\ p + 1}{6\ p}$$

$$\left[(k-1)\ \text{var}\ X + k\ \overline{X}^2 + \overline{X}_k{}^2\right] + X_k{}^2 + \frac{p^2-1}{3\ p} \left[r_1\ (k-1)\ \text{var}\ X\right.$$

$$\left. + (k+1)\ \overline{X}^2 - \ \overline{X}\ (X_1 + X_k)\right]$$

ou, si l'on sépare les termes centraux des termes de boût:

$$\sum_{j=1}^{N} Y^2 = \frac{2\ p^2 + 1}{3\ p} \left[(k-1)\ \text{var}\ X + k\ \overline{X}^2\right] + \frac{p^2-1}{3\ p} \left[r_1\ (k-1)\ \text{var}\ X\right.$$

$$\left. + (k+1)\ \overline{X}^2\right] - \overline{X}\ (X_1 + X_k) - \frac{(p-1)\ (2\ p-1)}{6\ p}\ (X_1{}^2 + X_k{}^2)$$

2.3.2 <u>Calcul de $\overline{Y}^2$</u>

D'après [2] $$\overline{Y}^2 = \frac{2\ p\ s - (p-1)\ (X_1 + X_k)}{[2 \cdot (k-1)\ p + 1]^2}$$

$$\overline{Y}^2 = \frac{4\ p^2\ k^2\ \overline{X}^2 - 4\ p\ k\ \overline{X}\ (p-1)\ (X_1 + X_k) + (p-1)^2\ (X_1 + X_k)^2}{4\ [(k-1)\ p + 1]^2}$$

Ces évaluations de $\sum_{j=1}^{N} Y_j^2$ et de $\overline{Y}^2$ permettent l'établissement de la formule générale donnant la variance de Y:

$$\text{var } Y = \overline{X}^2 \left[\frac{2\ p^2+1}{3\ p^2}\ \frac{k}{k-1} + \frac{p^2-1}{3\ p^2}\ \frac{k+1}{k-1} - \frac{p}{(k-1)\ p + 1}\ \frac{k^2}{k-1}\right] \qquad [\ 10a\ ]$$

$$+ \text{ var } X \left[\frac{2\ p^2+1}{3\ p^2} + r_1\ \frac{p^2-1}{3\ p^2}\right] \qquad [\ 10b\ ]$$

$$+ \overline{X}\ (X_1+X_k)\ \left[\frac{k}{(k-1)\ p+1}\ \frac{p-1}{k-1} - \frac{p^2-1}{3\ p^2 \cdot k-1}\right] \qquad [\ 10c\ ]$$

$$- \frac{(p-1)}{6\ p^2}\ \frac{(2\ p-1)}{k-1}\ [X_1^2 + X_k^2] - \frac{(p-1)^2}{4(k-1)\ p \cdot [(k-1)\ p+1]}\ (X_1+X_k)^2 \qquad [\ 10d\ ]$$

Exemple numérique

Pour mettre en évidence l'ordre de grandeur des termes de l'équation [10], les coefficients des 4 termes ont été calculés dans le cas où $k = 100$ $p = 5$ et $r_1 = 0.7$. On a alors:

$$\begin{aligned} \text{var } Y = & - 0.004914\ \overline{X}^2 + 0.90 \text{ var } X \\ & + 0.004913\ \overline{X}\ (X_1 + X_k) \\ & - 0.002424\ [X_1^2 + X_k^2] \\ & - 0.00001629\ [X_1 + X_k]^2 \end{aligned}$$

On note d'abord que, pour une série longue (k très grand), seule la partie [10b] de l'équation [10] subsiste quel que soit $\overline{X}$.

En effet, pour k très grand, [10a] s'écrit:

$$\overline{X}^2 \cdot \left[1 \left\{\frac{2\ p^2+1}{3\ p^2} + \frac{p^2-1}{3\ p^2} - 1\right\} \simeq \overline{X}^2 \cdot \left[0\ (k^{-1})\right]\right.$$

et [10c] et [10d] ne contiennent que des termes en k^{-1} et k^{-2}, qui tendent donc vers zéro quand k croît.

Si, d'autre part, on suppose la série X_i centrée ($\overline{X} = 0$), ce qui ne réduit pas la généralité des résultats, l'équation [10] s'écrit:

$$\text{var } Y = \text{var } X \left[\frac{2\ p^2+1}{3\ p^2} + r_1 \frac{p^2-1}{3\ p^2}\right] - \frac{(p-1)\ (2\ p-1)}{6\ p^2\ (k-1)} \left[X_1{}^2 + X_k{}^2\right]$$

$$- \frac{(p-1)^2}{4(k-1)\ p \cdot \left[(k-1)\ p + 1\right]} (X_1 + X_k)^2$$

Cette variance peut se décomposer en 3 parties:

$a = \frac{2p^2+1}{3\ p^2}$ var X variance provenant d'une série infinie de termes indépendants X_n .

$b = r_1 \frac{p^2-1}{3\ p^2}$ var X variance provenant de la persistance d'une série infinie X_n .

$$C = \frac{(p-1)\ (2\ p-1)}{6\ p^2\ (k-1)} \left[X_1^2 + X_k^2\right] - \frac{(p-1)^2}{4(k-1)\ p \cdot [(k-1)\ p + 1]^2} \left[X_1+X_2\right]^2$$

variance traduisant l'effet des termes de boût.

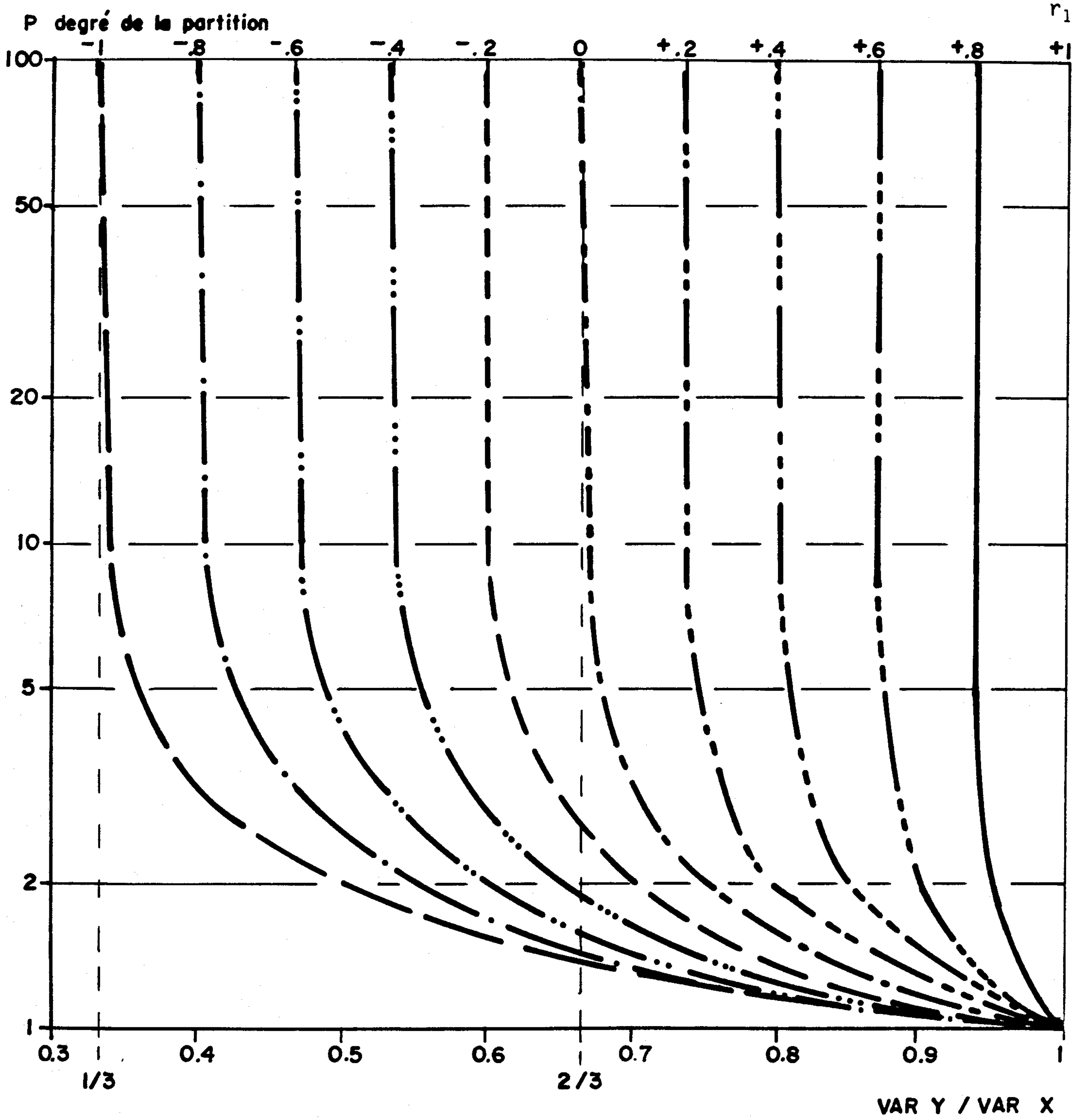
Nous allons examiner ces trois parties:

a: la fonction $\frac{2\ p^2+1}{3\ p^2}$ est une fonction monotone décroissante de 1 (quand P = 1) à 2/3 (quand P est infini): la variance d'une série interpolée à partir d'une série infinie (k infini) de termes indépendants (r_1 = 0) est réduite aux 2/3 de la variance de la série de base; la première condition (k infini) rend la partie c négligeable et la deuxième condition (série X_i indépendante, r_1 très faible) fait disparaître la partie b.

b: la fonction $\frac{p^2-1}{3\ p^2}$ est une fonction monotone croissante de 0 (quand p = 1) à 1/3 (quand p devient infini).

On voit que si r_1 est proche de 1 (forte persistance de la série de base) la perte de variance de la série interpolée à partir d'une série infinie est très faible, en effet a + b tend vers 1 quand k augmente. Par contre, si r_1 est négatif, la variance de la série interpolée peut être divisée jusqu'à un facteur 3 par rapport à la série de base (si celle-ci est infinie et que r_1 est proche de -1).

Ces résultats apparaissent à la figure 2.



r_1 = coefficient d'autocorrélation de X_i

Figure 2. Reduction de variance induite par l'interpolation d'une série infinie (k infini).

c: on note d'abord que l'effet des termes de boût est toujours négatif, c'est-à-dire que l'omission de ces termes surestime la variance de la série interpolée. Nous allons évaluer ce biais:

Si k est grand et si p devient négligeable devant k, on a $k \sim k-1 >> p$; si, d'autre part, on a $k\ p >> 1$, alors le dénominateur de la seconde fraction se simplifie:

$$4\ (k-1)\ p\ [(k-1)\ p + 1] \sim 4\ k^2\ p^2$$

$$\text{d'où} \quad c \sim - \frac{(p-1)\ (2\ p-1)}{6\ p^2\ k} [X_1{}^2 + X_k{}^2] - \frac{(p-1)^2}{4\ k^2\ p^2} [X_1+X_k]^2$$

Puisque k est grand, la deuxième fraction (en k^{-2}) est alors négligeable devant la première (en k^{-1}). Nous en évaluons maintenant la valeur la plus probable:

D'après l'équation [B-1] de l'Annexe 2 où $\overline{X} = 0$:

$$c = \frac{(p-1)\ (2\ p-1)}{6\ p^2 k}\ 2 \cdot E(X^2) \sim - \frac{(p-1)\ (2\ p-1)}{3\ p^2 k} \text{ var } X$$

On a démontré plus haut que, pour k grand, var Y se réduisait à [10b]:

$$\text{var } Y = \left[\frac{(2\ p^2+1)}{3\ p^2} + r_1\ \frac{p^2-1}{3\ p^2}\right] \text{ var } X \qquad [\ 11\ ]$$

D'où l'évaluation du pourcentage de biais par rapport à la variance totale introduite, en négligeant les effets de bout pour une série de longueur k grande, de persistance r_1 et partitionnée en p éléments équidistants:

$$B = - \frac{(p-1)\ (2\ p-1)}{3\ p^2\ k} \text{ var } X \cdot \frac{3\ p^2}{[(2\ p^2+1) + r\ \ (p^2-1)]\text{ var } X}$$

$$\text{d'où } B\ (\%) = \frac{100}{k} \frac{(p-1)\ (2\ p-1)}{[(2\ p^2+1) + r_1\ (p^2+1)} \qquad [\ 12\ ]$$

Cette équation [12] se simplifie si l'on considère que p est très grand devant 1:

$B_2(\%) \sim - \frac{100}{k} \frac{2}{2+r_1}$; cette expression varie de $- \frac{66}{k}$ à $\frac{200}{k}$ selon les valeurs de r_1, ce qui indique la longueur k de la série X_i nécessaire pour limiter le biais sur la variance totale quand on néglige les effets de bout dans l'équation [10].

On peut aussi se demander pour quelles valeurs de r_1 et P_1 la quantité B_2 constitue une borne supérieure du biais; pour cela, on effectue le rapport $\frac{B_2}{B_1}$ et l'on contrôle s'il reste supérieur à +1:

$$\frac{B_2}{B_1} = \frac{2}{2 + r_1} \frac{(2\ p^2+1) + r_1\ (p^2-1)}{(p-1)\ (2\ p-1)} > 1$$

ce qui conduit à $2\ p + r_1\ p - r_1 > 0$ ou $r_1 > \frac{-2\ p}{p-1}$; comme cette condition est toujours réalisée ($|r_1| < 1$), l'expression $B_2 = \frac{-100}{k} \frac{2}{2+r_1}$ constitue bien une borne supérieure du biais.

2.4 Coefficient de variation

Nous avons vu précédemment que la partition d'une série X_j induit une réduction de la variance de la série interpolée Y_j (équation [10]), la moyenne $\overline{Y}$ restant stable (équation [9]) aux effets de bout près; on sait aussi qu'une mesure de la variabilité d'une série peut être définie par le coefficient de variation:

$$(Cv) = \frac{\text{écart-type}}{\text{moyenne}} = \frac{\sqrt{\text{variance}}}{\text{moyenne}}$$

Ce coefficient de variation pour la série interpolée Y_j s'évalue en fonction de la série de base X dans le cas où k est grand (ce qui conduit à négliger les effets de bout dans l'équation [14]), mais sans hypothèse sur $\overline{X}$:

$$(Cv)_y = \frac{(\text{var } Y)^{\frac{1}{2}}}{\overline{Y}} = \left[\frac{2\ p^2+1\ +\ r_1\ (p^2-1}{3\ p^2}\right]^{\frac{1}{2}} \Big/ \frac{2\ k\ p\ \overline{X}}{2\ (k-1)\ p\ +\ 1}$$

d'où en tenant compte du fait que k est grand:

$$\frac{(Cv)_y}{(Cv)_x} = \frac{(2\ p^2+1)\ +\ r_1\ (p^2-1)}{p\sqrt{3}} \qquad [\ 13\ ];$$

selon les valeurs de p et de r_1 , cette fonction varie de 1 à $1/\sqrt{3}$ = 0.577 ; il y a donc toujours diminution du coefficient de variation (Tableau 2).

Tableau 2: Evolution du rapport $R = \frac{(Cv)_y}{(Cv)_x} = \frac{\sqrt{(2\ p^2+1) + r_1\ (p^2-1)}}{p\sqrt{3}}$

r_1 \ p	-1	0	+1	
2	$\sqrt{1/2}$	$\sqrt{3/4}$	1	$\sqrt{(3 + r_1)/4}$
3	$\sqrt{11/27}$	$\sqrt{19/27}$	1	$\sqrt{(19 + 8\ r_1)/27}$
4	$\sqrt{3/8}$	$\sqrt{11/16}$	1	$\sqrt{(11 + 15\ r_1)/16}$
	$\sqrt{\frac{p^2 + 2}{3\ p^2}}$	$\sqrt{\frac{2\ p^2 + 1}{3\ p^2}}$	1	R

3. Autocovariances d'ordre t

On écrit tout d'abord l'équation [B-4] pour la série Y_j de longueur $N = (k-1)\,p + 1$

$$ACV(Y)_t = \frac{1}{(k-1)\,p+1-t}\left\{ \overline{Y}\left(\sum_{j=1}^{t} Y_j + \sum_{j=(k-1)p+2-t}^{(k-1)p+1} Y_j\right) - [(k-1)\,p+1+t]\,\overline{Y}^2 + \sum_{j=1}^{j=(k-1)p+1-t} Y_j\,Y_{j+t}\right\} \qquad [\ 14\ ]$$

nous allons évaluer successivement les 3 parties de l'équation [14]:

$$A = \overline{Y}\left(\sum_{j=1}^{t} Y_j + \sum_{j=(k-1)p+2-t}^{(k-1)p+1} Y_j\right)$$

$$B = -[(k-1)p+1+t]\,\overline{Y}^2$$

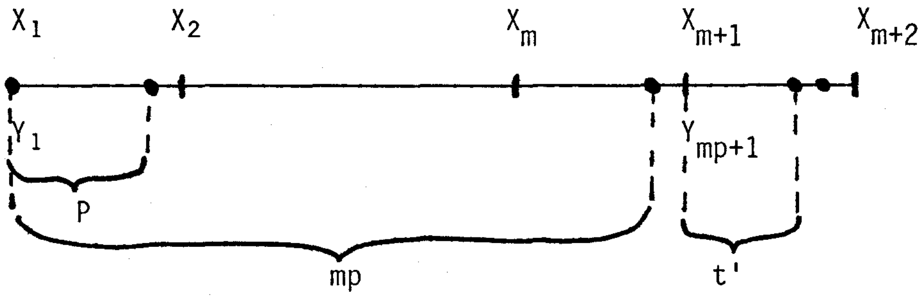
$$C = \sum_{j=1}^{(k-1)p+1-t} Y_j\,Y_{j+t}$$

on a:

$$ACV(Y)_t = \frac{1}{(k-1)\,p+1-t}\,[A+B+C]$$

3.1 Calcul des sommes partielles

D'après notre schéma de définition



Le calcul de l'expression $\sum_{j=1}^{t} Y_j$ s'effectue de la façon suivante:

on pose $t = mp+t'$ avec $1 \leqslant t' \leqslant p$ pour les valeurs de m telles que $0 \leqslant m \leqslant k-2$ et $t' = 1$ quand $m = k-1$. On a alors $Y_{mp+1} \equiv X_{m+1}$

$$\sum_{j=1}^{t} Y_j = \sum_{1}^{mp+t'} Y_j = \sum_{1}^{mp} Y_j + \sum_{mp+1}^{mp+t'} = \sum_{1}^{mp+1} Y_j + \sum_{mp+1}^{mp+t'} Y_j - Y_{mp+1}$$

Le premier terme correspond à la somme des mp+1 premiers Y_j, , soit d'après l'équation [8]:

$$\sum_{1}^{mp+1} Y_j = P \sum_{1}^{m+1} X_i - \frac{P-1}{2} (X_1 + X_{m+1})$$

Le deuxième terme correspond à une somme partielle:

$$\sum_{mp+1}^{mp+t'} Y_j = \frac{\sum_{1}^{t'} (t'-1)\, X_{m+2} - \sum_{1}^{t'} (p-t'+1)\, X_{m+1}}{P} = \frac{t' \cdot t'-1}{2p}\, X_{m+2}$$

$$+ \frac{2pt'-t' \cdot t'-1}{2p}\, X_{m+1}$$

Le troisième terme correspond au terme $X_{m+1} \equiv Y_{mp+1}$ qui a été compté dans chacune des sommes précédentes, donc une fois de trop.

d'où $$\sum_{1}^{t=mp+t'} Y_j = p \sum_{1}^{m+1} X_i - \frac{p-1}{2} (X_1 + X_{m+1}) + \frac{t' \cdot t'-1}{2p}\, X_{m+2}$$

$$+ \frac{2pt'-t' \cdot t'-1}{2p}\, X_{m+1} - X_{m+1}$$

ce qui s'écrit aussi:

$$\sum_{1}^{t=mp+t'} Y_j = p \sum_{1}^{m+1} X_i - \frac{p-1}{2} (X_1 + X_{m+1}) + \frac{t'-1}{2p} [(2p-t')\, X_{m+1}$$

$$+ t'\, X_{m+2}] \qquad [\ 15\ ]$$

La première partie de l'équation [15] correspond au cumul d'un nombre entier d'intervalles, borne supérieure comprise, et la deuxième partie correspond à la fraction d'intervalle restante qui disparaît si $t' = 1$.

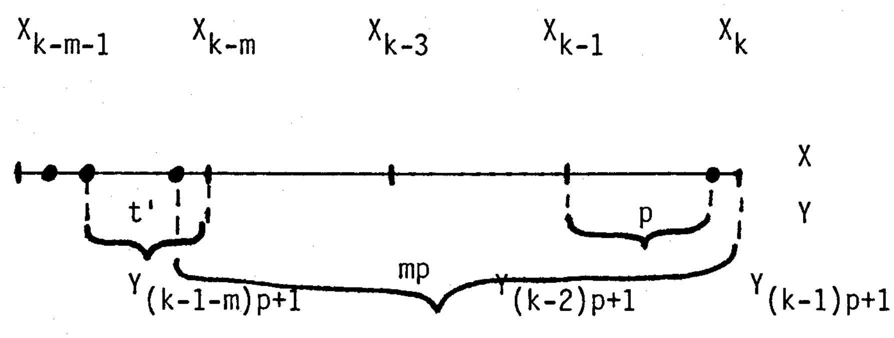
Cas particuliers

- si $t' = p$ les termes en X_{m+1} disparaissent.

- pour $m = 0$ l'expression se réduit à:

$$X_1 + \frac{t'-1}{2p}\left[(2p-t')\ X_1 + t'X_2\right] = \frac{t'\cdot t'-1}{2p}\ X_2 + \frac{2pt'-t'\cdot t'-1}{2p}\ X_1$$

L'évaluation de la somme partielle des $t = mp+t'$ derniers termes s'effectue de la même façon:



on a: $X_{k-m} \equiv Y_{(k-1-m)\ p+1}$

$$\sum_{(k-1)p+2-t}^{(k-1)p+1} Y_j = \sum_{(k-1-m)p+2-t'}^{(k-1)p+1} Y_j = \sum_{(k-1-m)p+2-t'}^{(k-1-m)p+1} Y_j + \sum_{(k-1-m)p+1}^{(k-1)p+1} Y_j - Y_{(k-1-m)p+1}$$

le premier terme s'écrit: $\dfrac{2pt' - t'(t'-1)}{2p} X_{k-m} + \dfrac{t'\cdot t'-1}{2p} X_{k-m-1}$

le second terme s'écrit: $p \sum_{k-m}^{k} X_i - \dfrac{p-1}{2} (X_{k-m} + X_k)$

le troisième terme s'écrit: $Y_{(k-1-m)p+1} = X_{k-m}$

d'oũ

$$\sum_{(k-1)p+2-t}^{(k-1)p+1} Y_j = p \sum_{k-m}^{k} X_i - \frac{p-1}{2} X_{k-m} + X_k)$$

$$\frac{t'-1}{2p} \left[(2p-t')\ X_{k-m} + t'\cdot X_{k-m-1}\right] \qquad [\ 16\ ]$$

d'où l'expression regroupée des sommes partielles A :

$$A = \overline{Y}\left[\sum_{j=1}^{t} Y_j + \sum_{j=(k-1)p+2-t}^{(k-1)p+1} Y_j\right] = \frac{2pk\overline{X} - (p-1)\,(X_1+X_k)}{2[(k-1)p+1]}$$

$$\{p\,\left(\sum_{1}^{m+1} X_i + \sum_{k-m}^{k} X_i\right) - \frac{p-1}{2}\,[X_1 + X_{m+1} + X_{k-m} + X_k]$$

$$+ \frac{t'-1}{2p}\,[(2p-t')\,(X_{m+1} + X_{k-m}) + t'\,(X_{m+2} + X_{k-m-1})]\} \qquad [\ 17\ ]$$

Cette équation est valable pour:

$t = mp + t'$ avec $0 \leqslant m \leqslant k-2$ $1 \leqslant t' \leqslant p$

ou pour $m = k-1$ on a $t' = 1$

<u>Cas particuliers:</u>

- pour $m = k-1$ et $t' = 1$, on vérifie que l'on obtient deux fois $\sum_{1}^{(k-1)p+1} Y_j$

- pour $m = 0$ on retrouve $X_1 + X_k$

3.2 Calcul du terme en $\overline{Y}^2$

D'après l'équation [9], on a :

$$\overline{Y} = \frac{2ps - (p-1)\ (X_1 + X_k)}{2\ [(k-1)p+1]} = \frac{2pk\overline{X} - (p-1)\ (X_1 + X_k)}{2\ [(k-1)p+1]}$$

d'où

$$\boxed{B = -\ [(k-1)p+1+t]\ \overline{Y}^2 = -\ [(k-1)p+1+t]\ \cdot \frac{4p^2k^2\overline{X}^2 - 4kp\overline{X}\ (p-1)\ (X_1 + X_k) + (p-1)^2\ (X_1 + X_k)^2}{4\ [(k-1)p+1]^2}} \qquad [\ 18\ ]$$

3.3 Calcul du terme $\sum_{j=1}^{(k-1)\ p+1-t} Y_j\ Y_{j+t} = C$

On cherche à évaluer la quantité $C = \sum_{j=1}^{j=(k-1)p+1-t} Y_j\ Y_{j+t}$. Pour calculer cette sommation, nous effectuons les 2 changements de variables $t = mp+t'$ et $j = lp + l'$, qui comparent les indices des termes au pas p de l'interpolation de la série originale X_i . Il faut maintenant calculer la borne supérieure de l'indice j; Pour cela, on évalue les termes réguliers et singuliers (de boût) de la série t, puis on identifie la série $t = mp+t'$ est constituée de termes réguliers où $0 \leqslant m \leqslant k-2$ et $1 \leqslant t' \leqslant p$; le terme singulier $m = k-1$ et $t' = 1$ ne peut être atteint par Y_{j+t} .

La borne supérieure de j donne alors:

$$j_{max} = l_{max}\, p + l'_{max} = (k-1)p+1-t = (k-2-m)p+p+1-t'$$

La série $j = lp + l'$ est donc constituée <u>de termes réguliers</u> où

$0 \leq l \leq k-3-m$ et $1 \leq l' \leq p$ et une <u>série de termes singuliers</u> où
$l = k-2-m$ et $1 \leq l' \leq p+1-t'$.

<u>Remarque importante:</u> D'après les gammes de variations précédentes, on note que la quantité $[t' + l']$ peut varier de 2 à 2p.

L'expression initiale C peut donc s'écrire:

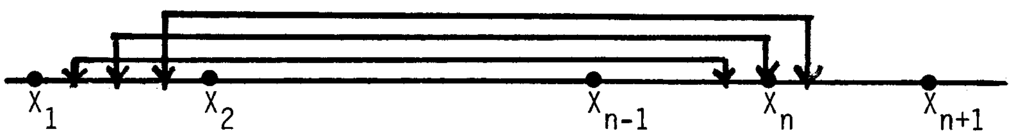
$$C = \sum_{l=0}^{l=k-2-m} \sum_{l'=1}^{l'=p-t'+1} Y_{lp+l'} \; Y_{(l+m)p+t'+l'}$$

$$+ \sum_{l=0}^{l=k-3-m} \sum_{l'=p-t'+2}^{p} Y_{lp+l'} \; Y_{(l+m+1)p+t'+l'-p}$$

Dans cette équation, k, p, m et t' sont fixes, l et l' varient; la première partie de l'équation contient les termes singuliers et les termes réguliers pour lesquels $2 \leq t'+l' \leq p+1$. La deuxième partie de l'équation correspond aux termes réguliers pour lequels

$$p+2 \leqslant t'+l' \leqslant 2p$$

Cette séparation, en deux parties, correspond au fait que, pour un décalage t fixé, les termes Y_{j+t} correspondant à des Y_j, situés entre deux X_i sont eux-mêmes situés entre trois X_i.



Calcul du premier terme C_1

Il faut évaluer $C_1 = \sum\limits_{l=0}^{l=k-2-m} \sum\limits_{l'=1}^{l'=p-t'+1} Y_{lp+l'}\, Y_{(l+m)p+t'+l'}$

Or, d'après [7]:

$$Y_{lp+l'} = \frac{(l'-1)\ X_{l+2} + (p-l'+1)\ X_{l+1}}{p}$$

$$Y_{(l+m)p+t'+l'} = \frac{(l'+t'-1)\ X_{l+m+2} + (p-l'-t'-1)\ X_{l+m+1}}{p}$$

D'où l'expression du produit:

$$C_l = \frac{1}{p^2} [\sum \sum (l'-1)\ (l'+t'-1)\ X_{l+2}\ X_{l+m+2}$$

$$+ (p-l'+1)\ (p-l'-t'+1)\ X_{l+1}\ X_{l+m+1}$$

$$+ (l'-1)\ (p-l'-t'+1)\ X_{l+2}\ X_{l+m+1}$$

$$+ (p-l'+1)\ (l'+t'-1)\ X_{l+1}\ X_{l+m+2}]$$

On remarque que ces sommes doubles sont indépendantes et peuvent donc être séparées:

$$C_l = \frac{1}{p^2} [\sum_{l'} (l'-1)\ (l'+t'-1) \sum_{l} X_{l+2}\ X_{l+m+2} + \cdots]$$

Nous devons donc évaluer 4 termes:

$$a_1 = \sum (l'-1)\ (l'+t'-1)$$

$$a_2 = \sum (p-l'+1)\ (p-l'-t'+1)$$

$$a_3 = \sum (l'-1)\ (p-l'-t'+1)$$

$$a_4 = \sum (p-l'+1)\ (l'+t'-1)$$

et ceci, pour l' variant de 1 à $p-t'+1$. On développe chacune des expressions a_1, a_2, a_3 et a_4 , puis on les ordonne en puissance de l': l'^2, l' et termes constants, et l'on effectue les sommations en utilisant les identités [A-1] et [A-3]:

$$a_1 = \sum \left[l'^2 + (t'-2)\, l'+1-t'\right]$$

$$a_2 = \sum \left[l'^2 + (t'-2p-2)\, l' + (p+1)\,(p+1-t')\right]$$

$$a_3 = \sum \left[-l'^2 + (-t'+p+2)\, l' - (p+1-t')\right]$$

$$a_4 = \sum \left[-l'^2 + (p+2-t')\, l' + (p+1)\,(t'-1)\right]$$

Après sommation et réarrangement, il vient:

$$a_1 = \frac{\overline{p-t'+1} \cdot \overline{p-t'} \cdot \overline{2p+t'+1}}{6}$$

$$a_2 = \frac{\overline{p-t'+1} \cdot \overline{p-t'} \cdot \overline{2p+t'+1}}{6}$$

$$a_3 = \frac{\overline{p-t'+1} \cdot \overline{p-t'} \cdot \overline{p-t'-1}}{6}$$

$$a_4 = \frac{p-t'+1}{6}\left[(p-t'+2)\,(p-t'+3) + 6\,(p+1)\,(t'-1)\right]$$

<u>N.B.</u>: Ce même résultat aurait pu être obtenu en effectuant le changement de variable $L' = p-t'+1$ et en sommant de 1 à L'.

<u>Calcul du second terme C_2:</u>

Il faut évaluer $$C_2 = \sum_{l=0}^{l=k-3-m} \sum_{l'=p+t'+2}^{p} Y_{lp+l'} \quad Y_{(l+m+1)p+t'+l'-p}$$

Or, d'après [7]:

$$Y_{lp+l'} = \frac{(l'-1)\ X_{l+2} + (p-l'+1)\ X_{l+1}}{p}$$

$$Y_{(l+m+1)\ p+t'+l'-p} = \frac{(t'+l'-p-1)\ X_{l+m+3} + (2p-t'-l'+1)\ X_{l+m+2}}{p}$$

d'où $$C_2 = \frac{1}{p^2}\ [\sum \sum (l'-1)(t'+l'-p-1)\ X_{l+2} \quad X_{l+m+3}$$

$$+ \sum \sum (p-l'+1)\ (2p-t'-l'+1)\ X_{l+1}\ X_{l+m+2}$$

$$+ \sum \sum (l'-1)\ (2p-t'-l'+1)\ X_{l+2}\ X_{l+m+2}$$

$$+ \sum \sum (p-l'+1)\ (t'+l'-p-1)\ X_{l+1}\ X_{l+m+3}]$$

On remarque que les sommes doubles sont indépendantes et peuvent donc être séparées:

$$C_2 = \frac{1}{p^2} \sum_{l'} (l'-1)\ (t'+l'-p-1) \quad \sum_{l} X_{l+2}\ X_{l+m+3} + \cdots$$

Nous devons donc évaluer 4 termes:

$$b_1 = \sum (l'-1)(t'+l'-p-1)$$

$$b_2 = \sum (p-l'+1)\ (2p-t'-l'+1)$$

$$b_3 = \sum (l'-1)\ (2p-t'-l'+1)$$

$$b_4 = \sum (p-l'+1)\ (t'+l'-p-1)$$

et ceci pour l' variant de $p-t'+2$ à p. Ce système se développe et s'ordonne en puissances de l' :

$$b_1 = \sum [l'^2-l'\ (p-t'+2) + (p-t'+1)]$$

$$b_2 = \sum [l'^2-l'\ (3p-t'+2) + (p+1)\ (2p-t'+1)]$$

$$b_3 = \sum [-l'^2 + l'\ (2p-t'+2) - (2p-t'+1)]$$

$$b_4 = \sum [-l'^2 + l'\ (2p-t'+2) - (p+1)\ (p-t'+1)]$$

Pour sommer ces expressions, on utilise les identités [A-10], [A-11] et [A-12] de l'Annexe 1 et après regroupement, on obtient:

$$b_1 = \frac{t' \cdot t'-1 \cdot [3p-(t'+1)]}{6}$$

$$b_2 = \frac{t' \cdot t'-1 \cdot [3p-(t'+1)]}{6}$$

$$b_3 = \frac{t'-1}{6} \cdot [6p(p-t') + t'(1+t')]$$

$$b_4 = \frac{t'-1}{6} \cdot t' \cdot t'+1$$

D'oū l'expression générale de $C = C_1 + C_2 = \sum Y_j\, Y_{j+2}$:

$$C = \sum_{j=1}^{j=(k-1)p+1-t} Y_j\, Y_{j+t} = \frac{1}{6p^2} \cdot \Big[$$

$$(p-t'+1)(p-t')(2p+t'+1) \sum_{l=0}^{k-2-m} X_{l+2}\, X_{l+m+2} \qquad [19a_1]$$

$$+ (p-t'+1)\ (p-t')\ (2p+t'+1) \sum_{l=0}^{k-2-m} X_{l+1}\, X_{l+m+1} \qquad [19a_2]$$

$$+ (p-t'+1)\ (p-t')\ (p-t'-1) \sum_{l=0}^{k-2-m} X_{l+2}\, X_{l+m+1} \qquad [19a_3]$$

$$+ (p-t'+1)\ \left[(p-t'+2)\ (p-t'+3) + 6(p+1)(t'-1)\right] \sum_{l=0}^{k-2-m} X_{l+1}\, X_{l+m+2} \qquad [19a_4]$$

$$+ t'\ (t'-1)\ (3p-(t'+1)) \sum_{l=0}^{k-3-m} X_{l+2}\ X_{l+m+3} \qquad [19b_1]$$

$$+ t'\ (t'-1)\ (3p-(t'+1)) \sum_{l=0}^{k-3-m} X_{l+1}\ X_{l+m+2} \qquad [19b_2]$$

$$+ (t'-1)\ \left[6p(p-t') + t'(1+t')\right] \sum_{l=0}^{k-3-m} X_{l+2}\ X_{l+m+2} \qquad [19b_3]$$

$$+ (t'-1)\ t'\ (t'+1) \sum_{l=0}^{k-3-m} X_{l+1}\ X_{l+m+3}\ \Big] \qquad [19b_4]$$

avec $t' = t-mp$

D'où l'expression complète d'autocovariance d'ordre t de la série Y :

$$ACV(Y)_t = \frac{1}{(k-1)\ p+1-t}\ [A+B+C] \qquad t = mp+t' \quad 0 \leqslant m \leqslant k-2 \rightarrow 1 \leqslant t' \leqslant p$$
$$m = k-1 \rightarrow t' = 1$$

Avec

$$A = \frac{2\ pk\ \overline{X} - (p-1)\ (X_1+X_k)}{2[(k-1)\ p+1]}\ [p\ [\sum_{1}^{m+1} X_i + \sum_{k-m}^{k} X_i] - \frac{p-1}{2}\ (X_1+X_{m+1}+X_{k-m}+X_k)$$
$$+ \frac{t'-1}{2p}\ [(2p-t')\ (X_{m+1}+X_{k-m}) + t'\ (X_{m+2}+X_{k-m-1})]] \qquad [\ 17\ ]$$

$$B = -[(k-1)p+1+t] \cdot \frac{4\ p^2k^2\overline{X}^2 - 4\ p\ \overline{X}\ (p-1)(X_1+X_k)+(p-1)^2\ (X_1+X_k)^2}{4\ [(k-1)\ p+1]^2} \qquad [\ 18\ ]$$

$$C = C_1 + C_2 \qquad [\ 19\ ]$$

$$C_1 = \frac{1}{p^2}\ [a_1 \sum_{l=0}^{k-2-m} X_{l+2}\ X_{l+m+2} + a_2 \sum_{l=0}^{k-2-m} X_{l+1}\ X_{l+m+1}$$
$$+ a_3 \sum_{l=0}^{k-2-m} X_{l+2}\ X_{l+m+1} + a_4 \sum_{l=0}^{k-2-m} X_{l+1}\ X_{l+m+2}]$$

$$C_2 = \frac{1}{p^2}\ [b_1 \sum_{l=0}^{k-3-m} X_{l+2}\ X_{l+m+3} + b_2 \sum_{l=0}^{k-3-m} X_{l+1}\ X_{l+m+2} +$$
$$b_3 \sum_{l=0}^{k-3-m} X_{l+2}\ X_{l+m+2} + b_4 \sum_{l=0}^{k-3-m} X_{l+1}\ X_{l+m+3}]$$

Avec $a_1 = a_2 = \frac{(p-t'+1)(p-t')(2p+t'+1)}{6}$ $a_3 = \frac{(p-t'+1)(p-t')(p-t'-1)}{6}$

$a_4 = \frac{p-t'+1}{6}\left[(p-t'+2)\ (p-t'+3) + 6(p+1)(t'-1)\right]$

$b_1 = b_2 = \frac{t'(t'-1)[3p-(t'+1)]}{6}$ $b_3 = \frac{t'-1}{6}\left[6p(p-t')+t'(1+t')\right]$

$b_4 = \frac{t'-1}{6}\, t'\ (t'+1)$

On peut vérifier que pour $p=1$, on a alors $X_i \equiv Y_j$ et $ACV(X)_t = ACV(Y_t)$ en faisant $p=1$, on a $t'=1$ $0 \leq m \leq k-1$.

$t=m+1$ $1 \leq t \leq k$

En reportant dans A, B, C, on retrouve l'expression de $ACV(X)_t$ donnée en Annexe 2.

L'équation [19] peut se spécialiser pour $m=0$, $m=1$ et $m \geqslant 2$, en regroupant les termes de même puissance; on note que les coefficients a_1 et a_2 sont égaux, de même que b_1 et b_2 ; on a donc, en notant par $\sum$ les sommes complètes:

• <u>m = 0</u>

$$C = \sum Y_j\, Y_{j+t} = \frac{1}{p^2} \cdot \{(2a_1+b_3)\sum_{i=1}^{k} X_i^2 + (2b_1+a_3+a_4)\sum_{i=1}^{k-1} X_i X_{i+1} + b_4 \sum_{i=1}^{k-2} X_i\, X_{i+2} \quad [20a]$$

$$- (a_1+b_3)\,[X_1^2+X_k^2] - b_1\,[X_1\, X_2 + X_{k-1}\, X_k]\} \quad [20b]$$

• <u>m = 1</u>

$$C = \sum Y_j\, Y_{j+t} = \frac{1}{p^2} \cdot \{a_3 \sum_{i=1}^{k} X_i^2 + (2a_1 + b_3)\sum_{i=1}^{k-1} X_i X_{i+1}$$

$$+ (2b_1 + a_4)\sum_{1}^{k-2} X_i\, X_{i+2} + b_4 \sum_{1}^{k-3} X_i\, X_{i+3} \quad [21a]$$

$$- a_3\,[X_1^2 + X_k^2] - (a_1 + b_3)\,[X_1 X_2 + X_{k-1} X_k]$$

$$- b_1\,[X_1 X_3 + X_{k-2} X_k]\} \quad [21b]$$

• <u>m ⩾ 2</u>

$$C = \sum Y_j\, Y_{j+t} = \frac{1}{p^2} \cdot \{a_3 \sum_{i=1}^{k-m+1} X_i X_{i+(m-1)} + (2a_1 + b_3)\sum_{i=1}^{k-m} X_i\, X_{i+m}$$

$$+ (2b_1 + a_4)\sum_{i=1}^{k-m-1} X_i\, X_{i+(m+1)} + b_4 \sum_{i=1}^{k-m-2} X_i X_{i+(m+2)} \quad [22a]$$

$$- a_3\,[X_1 X_m + X_{k-m+1}\, X_k] - (a_1 + b_3)\,[X_1\, X_{m+1} + X_{k-m}\, X_k]$$

$$- b_1\,[X_1 X_{m+2} + X_{k-(m+1)}\, X_k]\} \quad [22b]$$

3.4 Cas particuliers

Cas d'une série centrée

Si on considère la série d'origine X_i comme centrée ($\overline{X}$=0), ce qui ne réduit pas la généralité, on a toujours:

$$ACV(Y)_t = \frac{1}{(k-1)p+1-t} [A + B + C]$$

C reste inchangé (équation [19]), mais A et B (équations [17] et [18]) se simplifient et donnent, dans les notations habituelles:

$$A' = \frac{(p-1)\ (X_1+X_k)}{2[(k-1)p+1]} \cdot [p[\sum_{1}^{m+1} X_i + \sum_{k-m}^{k} X_i] - \frac{p-1}{2} [X_1+X_{m+1}$$

$$+ X_{k-m} + X_k] + \frac{t'-1}{2p} [(2p-t')\ (X_{m+1} + X_{k-m})$$

$$+ t'\ (X_{m+2} + X_{k-m-1})]]$$

et $$B' = -\frac{(k-1)\ p+1+t}{4[(k-1)\ p+1]^2} (p-1)^2\ (X_1 + X_k)^2$$

Cas d'une série centrée et très longue

Si, en plus d'être centrée ($\overline{X}$=0), la série X_i est très longue (k très grand), les deux premiers termes de l'autocovariance A' et B' , qui sont en k^{-1} , deviennent négligeables devant les termes de C ; les termes de bout de l'équation C disparaissent aussi, à cause de la présence du dénominateur en $1/[(k-1)\ p+1-t]$; il reste alors:

$$ACV(Y)_t \approx \frac{1}{(k-1)p+1-t} \cdot \frac{1}{p^2} \left[a_3 \sum_{i=1}^{k-m+1} X_i X_{i+m-1} + (2a_1 + b_3) \sum_{i=1}^{k-m} X_i X_{i+m}\right.$$

$$\left. + (a_4 + 2b_1) \sum_{i=1}^{k-m-1} X_i X_{i+m+1} + b_4 \sum_{i=1}^{k-m-2} X_i X_{i+m+2}\right]$$

Avec ces hypothèses, l'équation [10] et les équations:

$$\sum_{i=1}^{k} X_i^2 = (k-1) \text{ var } X + k \overline{X}^2 \qquad [B-1]$$

$$\sum_{i=1}^{k-t} X_i X_{i+t} = r_t \cdot (k-t) \cdot \text{var } X + (k-t) \overline{X}^2 - \overline{X} \left(\sum_{1}^{t} X_i + \sum_{k-t+1}^{k} X_i\right) \qquad [B-4]$$

(valable pour $t > 0$)

se simplifient et se réduisent à:

$$\sum_{i=1}^{k} X_i^2 = (k-1) \text{ var } X \qquad [23]$$

$$\sum_{i=1}^{k-t} X_i X_{i+t} = r_t \cdot (k-t) \text{ var } X \qquad \text{avec } t > 0 \qquad [24]$$

et $\text{Var } Y = \left[\frac{2p^2+1}{3p^2} + r_1 \frac{p^2-1}{3p^2}\right] \text{var } X$ (équation [11]) [25]

<u>Autocovariances à l'ordre t, lorsque $\overline{X}=0$ et quand k est grand</u>

Dans ces conditions, en substituant les équations [23] et [24] dans [20a], [21a] et [22a], on obtient:

• m = 0

$$ACV(Y)_t = \frac{1}{(k-1)\ p+1-t}\ \frac{1}{p^2}\ \text{var } X\ [(2a_1 + b_3)\ (k-1)$$

$$+ r_1\ (2b_1 + a_3 + a_4)\ (k-1) + r_2\ b_4\ (k-2)] \qquad [26]$$

• m = 1

$$ACV(Y)_t = \frac{1}{(k-1)p+1-t}\ \frac{1}{p^2}\ \text{var } X\ [a_3\ (k-1) + r_1\ (2a_1 + b_3)\ (k-1)$$

$$+ r_2\ (2b_1 + a_4)\ (k-2) + r_3\ b_4\ (k-3)] \qquad [27]$$

• m ⩾ 2

$$ACV(Y)_t = \frac{1}{(k-1)\ p+1-t}\ \frac{1}{p^2}\ \text{var } X\ [r_{m-1}\ a_3\ (k-(m-1))$$

$$+ r_m\ (2a_1 + b_3)\ (k-m) + r_{m+1}\ (2b_1 + a_4)$$

$$(k-(m+1)) + r_{m+2}\ \ b_4\ (k-(m+2))] \qquad [28]$$

Cette formule générale, lorsqu'elle est appliquée pour $m=1$ conduit à la formule précédente [27], sauf pour le coefficient de a_3 qui diffère d'un facteur $(k-1)$ à cause de l'utilisation de la formule non biaisée à la variance. On peut remarquer que, lorsque $m = k+2$, le coefficient de $r_{m+2} = r_k$ est nul, ce qui est cohérent avec la non-définition de r_k . Bien que r_k n'existe pas, on peut donc conserver la forme générale de l'équation [28] même pour $m = k-2$. D'une manière générale, on a $t = mp+t'$; à la limite m peut atteindre $(k-2)$ et m ne peut donc pas toujours être négligé devant k systématiquement; cependant, quand k est grand:

• <u>si $m = 0$</u> on a : $t = t' \leqslant p$

$(k-1)\ p+1-t \simeq kp$ et $k-1 \simeq k$

d'où:

$$ACV(Y)_t \simeq \frac{\text{var } X}{p^3} \left[(2a_1 + b_3) + r_1\ (2b_1 + a_3 + a_4) + r_2\ b_4\right] \quad [29]$$

• <u>si $m = 1$</u> on a : $t = p + t'$

$$ACV(Y)_t \simeq \frac{\text{var } X}{p^3} \left[a_3 + r_1\ (2a_1+b_3) + r_2\ (2b_1 + a_4) + r_3 b_4\right] \quad [30]$$

Le cas général <u>$m \geqslant 2$</u> peut poser plus de problèmes; en remplaçant t par $mp+t'$, on vérifie que l'on a bien $m \leqslant k-2$;

• <u>Dans le cas où</u> $\frac{k}{m} \gg 1$, alors $(k-1-m)\ p-1-t' \sim kp$, la formule [28] se simplifie et devient:

$$ACV(Y)_t = \frac{1}{p^3} \left[a_3\ r_{m-1} + (2a_1 + b_3)\ r_m + (2b_1 + a_4)\ r_{m+1} + b_4\ r_{m+2}\right] \quad [31]$$

• <u>Si la condition</u> $\frac{k}{m} \gg 1$ n'est pas remplie, alors on doit conserver la formule [28].

4. Coefficients d'autocorrélation d'ordre t , lorsque $\overline{X}$=0 et quand k est grand

4.1 Généralités

Par définition, on a $R_t = \frac{ACV(Y)_t}{var\ Y}$; on substitue [25] dans [26], [27] [28] ou [31], selon la valeur de $\frac{m}{k}$:

. Si m = 0

$$R_t = \frac{3}{p} \frac{[(2a_1 + b_3) + r_1 (2b_1 + a_3 + a_4) + r_2 b_4]}{(2p^2+1) + r_1 (p^2-1)} \qquad [32]$$

vérification: Calculons R pour t' = 0 et m = 0, donc pour t=0 :

$$2a_1 + b_3 = \frac{2p (p+1)(2p+1) - 6p^2}{6} = \frac{2p (2p^2+1)}{6}$$

$$b_1 = 0 \qquad a_3 + a_4 = \frac{2p (p^2-1)}{6}$$

$$b_4 = 0$$

On trouve donc $R_0 \equiv 1$ pour toute valeur de p, de r_1 et de r_2.

. Si m = 1

$$R_t = \frac{3}{p} \frac{[a_3 + r_1 (2a_1 + b_3) + r_2 (2b_1 + a_4) + r_3 b_4]}{2p^2+1 + r_1 (p^2-1)} \qquad [33]$$

• Si $m \geq 2$

Dans le cas où $\overline{X} = 0$, k est grand et $2 \leq m \leq k-2$, on a, d'après l'équation [31]:

$$R_t = \frac{3}{(k-1)p+1-t} \left[\frac{r_{m-1}\, a_3\, (k-m+1) + r_m\, (2a_1 + b_3)\, (k-m)}{[(2p^2+1) + r_1\, (p^2-1)]}\right.$$

$$\left. + \frac{r_{m+1}\, (2b_1 + a_4)\, (k-m-1) + r_{m+2}\, b_4\, (k-m-2)}{[(2p^2+1) + r_1\, (p^2-1)]}\right] \qquad [34]$$

Si, de plus, $\frac{k}{m} >> 1$, on a, d'après l'équation [28]:

$$R_t = \frac{3}{p} \; \frac{a_3\, r_{m-1} + (2a_1+b_3)\, r_m + (2b_1 + a_4)\, r_{m+1} + b_4\, r_{m+2}}{(2p^2+1) + r_1\, (p^2-1)} \qquad [35]$$

On note que cette expression de R_t peut être considérée comme générale pour tout m tel que $(\frac{k}{m} >> 1)$, si on utilise la convention habituelle de symétrie des autocorrélations:

$$r_{-m} = r_m \qquad \text{pour tout } m \neq 0$$

et $\quad r_0 = +1$

Cette convention permet de généraliser [34] pour m=0 et m=1.

<u>Vérification</u>: Si p=1 et t'=1, alors t=p+1 , on a $a_i = b_i = 0$, sauf $a_4 = 1$ et $R_t = a_4\, \frac{r_{m+1}}{3}\, \frac{3}{1} = r_{m+1}$

4.2 Coefficients d'autocorrélation de la série Y_j aux points de mesure de la série X_i

Le correlogramme de la série Y_j étant une fonction continue, il peut être intéressant d'en déterminer la forme aux points de mesure de la série X_i . Pour cela, on pose $t' = 1$ dans les équations [19], [29], [30] et [31]; dans tous ces cas $b_1 = b_2 = b_3 = b_4 = 0$ et

$$a_1=a_2=\frac{p \cdot p-1 \cdot p+1}{3} \qquad a_3=\frac{p \cdot p-1 \cdot p-2}{6} \qquad a_4=\frac{p \cdot p+1 \cdot p+2}{6}$$

Calculons R_t pour $t = mp+1$

. <u>Si m = 0</u> $(t = 1)$

$$R_t= R_1 = \frac{1}{2}\ \frac{4p^2-4+(2p^2+4)\ r_1}{2p^2+1 + r_1\ (p^2-1)} = \frac{2p^2-2 + (p^2+2)r_1}{2p^2+1 +(p^2-1)r_1}$$

Cas particuliers: si $p = 1$ $\qquad R_1 = r_1$

si $p \to \infty$ $\qquad R_1 \to \frac{4+2r_1}{4+2r_1} = 1$

. <u>Si m = 1</u> $(t = p + 1)$

$$R_t = R_{p+1} = \frac{1}{2}\ \frac{p^2-3p+2 + r_1\ (4p^2-4)+r^2\ (p^2+3p+2)}{2p^2+1 + r_1\ (p^2-1)}$$

Cas particuliers: si $p = 1$ $\qquad R_2 = r_2$

si $p \to \infty$ $\qquad R_{p+1} \to \frac{1+4r_1+r_2}{4+2r_1}$

• Si $m \geq 2$ $(t = mp + 1)$

Dans le cas général, on a d'après l'équation [34]:

$$R_t = R_{mp+1} = \frac{3}{(k-1-m)p} \left[\frac{r_{m-1}\, a_3\, (k-m+1) + r_m\, (2a_1 + b_3)\, (k-m)}{[(2p^2+1) + r_1\, (p^2-1)]} + \frac{r_{m+1}\, (2b_1 + a_4)\, (k-m-1) + r_{m+2}\, b_4\, (k-m-1)}{[(2p^2+1) + r_1\, (p^2-1)]}\right]$$

$$R_t = R_{mp+1} = \frac{3}{p[(2p^2+1) + r_1\, (p^2-1)]} \left[r_{m-1}\, a_3\, \frac{k-m+1}{k-m-1} + r_m\, (2a_1 + b_3)\, \frac{k-m}{k-m-1} + r_{m+1}\, (2b_1+a_4) + r_{m+2}\, b_4\, \frac{k-m-2}{k-m-1}\right]$$

En remplaçant les a_i et les b_i par leurs valeurs, il vient:

$$R_t = R_{mp+1} = \frac{1}{(2p^2+1) + r_1\, (p^2-1)} \left[r_{m-1}\, \frac{k-m+1}{k-m-1}\, \frac{(p-1)(p-2)}{2} + r_m\, \frac{k-m}{k-m-1}\, 2\, (p-1)\, (p-2) + r_{m+1}\, \frac{(p+1)(p+2)}{2}\right] \qquad [36]$$

Exemples:

Si $p = 1$ $\qquad R_{m+1} = \frac{1}{3}\, r_{m+1} \cdot \frac{2 \cdot 3}{2} = r_{m+1}$

Si $p \to \infty$ $\qquad R_{mp+1} \to \frac{1}{2+r_1} \left[\frac{1}{2}\, r_{m-1}\, \frac{k-m+1}{k-m-1} + 2\, r_m\, \frac{k-m}{k-m-1} + \frac{1}{2}\, r_m+1\right]$

- lorsque k est pair, si par exemple $\frac{k}{m} = 2$, on obtient:

$$R_{(\frac{k}{2}\ p+1)} = \frac{1}{(2p^2+1) + r_1\ (p^2-1)} \cdot \frac{1}{(m-1)} \left[(m+1)\ \frac{(p-1)(p-2)}{2}\ r_{m-1} + r_m \cdot m \cdot 2\ (p-1)(p+1) + r_{m+1}\ \frac{(p+1)(p+2)}{2}\right]$$

- à la limite, $m = k-2$ on a:

$$R_{(k-2)\ p+1} = \frac{1}{(2p^2+1)+r_1\ (p^2-1)} \left[\frac{3}{2}\ (p-1)(p-2)\ r_{m-1} + 4(p-1)(p+1)\ r_m + \frac{1}{2}\ (p+1)(p+2)\ r_{m+1}\right]$$

Dans les cas plus particuliers où $\frac{k}{m} >> 1$, on peut utiliser l'équation [35], ce qui correspond au cas assez fréquent où k est grand et où l'on s'intéresse aux valeurs de R pour les faibles valeurs de m (persistance à court terme); on a alors:

$$R_t = \frac{3}{p}\ \frac{a_3\ r_{m-1} + (a_1 + a_2)\ r_m + a_4\ r_{m+1}}{2p^2+1 + r_1\ (p^2-1)}$$

$$= \frac{1}{2}\ \frac{(p^2-3p+2)\ r_{m-1} + (4p^2-4)\ r_m + (p^2+3p+2)\ r_{m+1}}{2p^2+1 + r_1\ (p^2-1)} \qquad [37]$$

<u>Exemples</u>:

- si $p = 1$ $\qquad R_t = R_{mp+1} = r_{m+1}$
- si $p \to \infty$ $\qquad R_t \to \frac{r_{m-1} + 4r_m + 4_{m+1}}{4 + 2r_1}$

4.3 Cas d'une série X_i indépendante

Si la série X_i est indépendante $r_1 = r_2 = r_3 = r_m = 0$, le corrélogramme se simplifie:

Si $\underline{m = 0}$
alors $t = t'$ et l'on a $R_t = \frac{3}{p} \frac{2a_1 + b_3}{2p^2+1} = \frac{4p^3 + 2p(1-3t^2) + 3t\,(t^2-1)}{4p^3+2p}$ [38]

Si $\underline{m = 1}$
alors $t = t'+p$ et l'on a $R_t = \frac{3}{p} \frac{a_3}{2p^2+1} = \frac{(2p-t)\,(2p-t+1)(2p-t-1)}{4p^3+2p}$ [39]

Si $\underline{m \geq 2}$ $R_t = 0$

On note que pour $p = t$ les deux fonctions prennent la même valeur:

$R_{(t=p)} = \frac{p^2-1}{4p^2+1}$. si $m = 0$ $1 \leq t' \leq p$ $t = p$ peut être atteint

si $m = 1$ $1 \leq t' \leq p$ $t \geq p+1$; par continuité le résultat est logique et montre que cette expression algébrique serait restée inchangée même avec d'autres conventions d'indices.

Si on appelle "mémoire d'un processus" le lag le plus faible pour laquelle la fonction d'autocorrélation devient nulle, alors l'interpolation linéaire est un processus de mémoire $(2p-1)$ et non $2p$ comme plusieurs auteurs l'indiquent (MEIJA, RODRIGEZ-ITURBE et DAWDY, 1972).

A partir des équations [38] et [39], on effectue des développements limités en $\frac{1}{p}$, si l'on suppose p suffisamment grand pour que les puissances inférieures soient négligeables:

- pour <u>m = 0</u>, on a $1 \leqslant t \leqslant p$ et R_t peut s'écrire:

$$R_t = [1 + \frac{1}{2p^2}(1 - 3t^2) + \frac{3t}{4p^3}(t^2-1)]\ [1 + \frac{1}{2p^2}]^{-1}$$

$$[1 + \frac{1}{2p^2}]^{-1} = 1 - \frac{1}{2p^2} + \frac{1}{4p^4} + \varepsilon\ (\frac{1}{p^5})\ ;\ \text{d'où}$$

$$R_t \simeq 1 - \frac{3}{4}\frac{t^2}{p^2}(2 - \frac{t}{p} + \frac{1}{tp} - \frac{1}{p^2})$$

Si on se limite aux termes en $\frac{1}{p^3}$:

$$R_t \simeq 1 - \frac{3}{4}\ \frac{t^2}{p^2}\ [2 - \frac{t}{p}\ (1 - \frac{1}{t^2})] \qquad [\ 40\ ]$$

Dans le cas <u>particulier</u> où $\frac{1}{t^2}$ est négligeable devant 1 (t supérieur à 3 ou 4 par exemple), l'équation [40] se simplifie et devient

$$R_t \simeq 1 - \frac{3}{4}\ \frac{t^2}{p^2}\ [2 - \frac{t}{p}] \qquad [\ 41\ ]$$

- pour <u>m = 1</u>, on a $p < t \leqslant 2p$ et R_t peut s'écrire, de la même façon:

$$R_t = \frac{8p^3 - 12p^2t + 6pt^2 - 2p - t^3 + t}{4p^3\ (1 + \frac{1}{2p^2})}$$

$$R_t \simeq (2 - \frac{3t}{p} + \frac{3t^2-1}{2p^2} - \frac{t^3-t}{4p^3})\ (1 - \frac{1}{2p^2} + \frac{1}{4p^4} + \varepsilon\ (\frac{1}{p^5}))$$

$$= \quad 2 - \frac{3t}{p} + \frac{3t^2-3}{2p^2} + \frac{7t-t^3}{4p^3} + \frac{-3t^2+3}{4p^4} + \varepsilon\,(\frac{1}{p^5})$$

en se limitant aux termes en $\frac{1}{p^3}$

$$R_t \simeq \frac{1}{4}\,[8 - 3 \cdot 4 \cdot \frac{t}{p} + 3 \cdot 2 \cdot (\frac{t}{p})^2\,(1 - \frac{1}{t^2}) - (\frac{t}{p})^3\,(1 - \frac{7}{t^2})]$$

Si on considère les équations [40] et [41], c'est-à-dire le cas où p est grand et où t est supérieur à 3 ou 4, on retrouve les résultats obtenus par MEIJA et al (1972) pour le processus de lignes brisées [Broken Line Process].

- <u>pour $m = 0$</u> $\qquad 0 < t \leqslant p$

$$R_t \approx \frac{4p^3 - 6pt^2 + 3t^3}{4p^3} = 1 - \frac{3}{4}\,\frac{t^2}{p^2}\,[2 - \frac{t}{p}]$$

quand $t = p$ $\qquad R_t = R_p \rightarrow \frac{1}{4}$

- <u>pour $m = 1$</u> $\qquad 0 < t \leqslant 2p$

$$R_t \approx \frac{(2p-t)\,(4p^2-4pt+t^2)}{4p^3} = \frac{1}{4}\,[2 - \frac{t}{p}]^3$$

- <u>pour $m > 1$</u> $\qquad t \geqslant 2p$

$$R_t \equiv 0$$

On note qu'il s'agit d'une forme affine qui n'est fonction que du rapport $\frac{t}{p} = \zeta$:

$$R\,(t, p) = 1 - \frac{3}{4}\,\zeta^2 \cdot (2 - \zeta) \qquad\qquad \zeta \leqslant 1$$

$$R\ (t,\ p) = \frac{1}{4} \cdot (2 - \zeta)^3 \qquad\qquad 2 \geqslant \zeta > 1$$

$$R\ (t,\ p) = 0 \qquad\qquad \zeta > 2$$

Ce corrélogramme est calculé au tableau 3.

A la figure 3, la forme du corrélogramme a été tracée et comparée à celle de deux processus classiques qui ont même autocorrélation à l'ordre 1 $[R_{\zeta=1} = 0.25]$:

Le processus markovien d'ordre 1:

$$\rho = (\tfrac{1}{4})\zeta$$

La moyenne mobile de portée $\frac{4}{3}$ dont le corrélogramme s'écrit

$$\rho = 1 - \frac{3\zeta}{4} \quad \text{pour} \quad \zeta \leqslant \frac{4}{3} \quad \text{et} \quad \zeta = 0 \quad \text{pour} \quad \zeta > \frac{4}{3} .$$

R (ζ)	$\zeta = \frac{t}{p}$
0	1
.1	0.986
.2	0.946
.3	0.885
.4	0.808
.5	0.719
.6	0.622
.7	0.522
.8	0.422
.9	0.332
1.0	0.250
1.1	0.203
1.2	0.160
1.3	0.123
1.4	0.090
1.5	0.063
1.6	0.040
1.8	0.010
1.9	0.003
2	0

Tableau 3 : Autocorrélogramme d'une série interpolée linéairement.

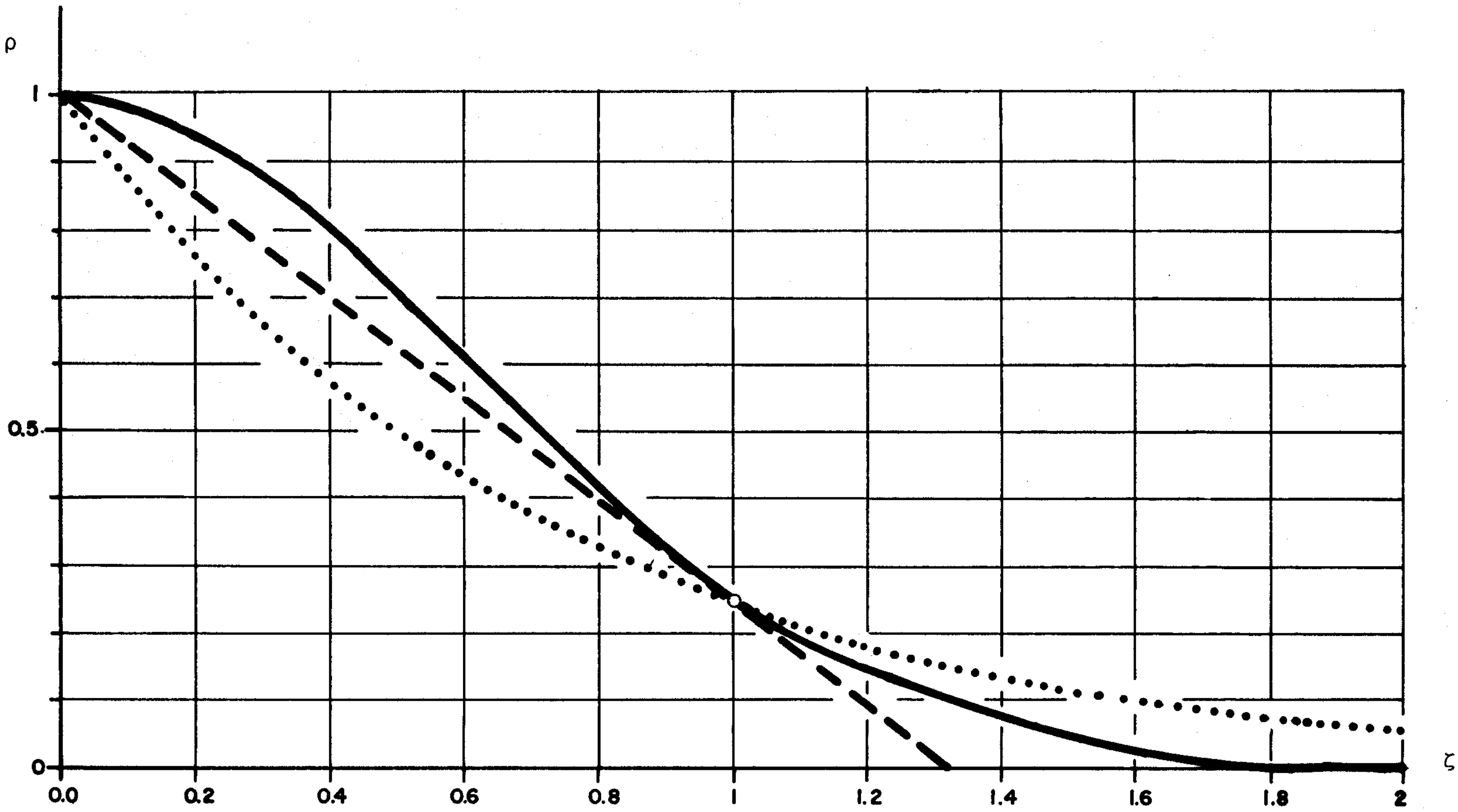


FIGURE 3 : Comparaison des formes des autocorrélogrammes

——— Série indépendante interpolée; $\rho_1 = 1 - \frac{3}{4}\zeta^2 (2-\zeta)$; $\rho_2 = \frac{1}{4}(2-\zeta)^3$; $\rho_3 = 0$

- - - - Moyenne mobile; $\rho = 1 - \frac{3}{4}\zeta$

······ Série markovienne d'ordre 1 ; $\rho = (\frac{1}{4})^{\zeta}$

5. Variance de la moyenne et contenu en information d'une série interpolée

5.1 Concept de contenu en information

Le contenu en information I d'une série Z est un indice global de la variabilité du phénomène représenté par une série de temps; Pour un échantillon constitué de réalisations indépendantes d'une variate normale de variance σ^2 , la variance de la moyenne vaut σ^2/N où N est le nombre d'éléments de l'échantillon. Cette variance est, en pratique, estimée par s^2/N où s^2 est la variance de l'échantillon, considérée comme un estimateur de σ^2 . Prenant une série normale formée d'éléments indépendants comme référence, MATALAS et LANGBEIN (1962) définissent le contenu en information I d'une série quelconque de taille N par le rapport:

$$I = (\sigma^2/N) \;/\; \mathrm{var}\, \overline{Z} \qquad [42]$$

où var $\overline{Z}$ représente la variance de la moyenne de la série étudiée. On peut en déduire un nombre effectif $N' = NI$ d'observations; Ainsi, les N observations sont équivalentes, en termes d'information, à N' observations indépendantes d'une variate normale de variance σ^2 .

BAYLEY et HAMMERSLEY (1946) ont exprimé la variance de la moyenne d'une série formée d'éléments non indépendants Z_j , en fonction de la variance var Z , de la longueur N et des coefficients d'autocorrélations R_t de la série étudiée:

$$\mathrm{var}\, \overline{Z} = \frac{\mathrm{var}\, Z}{N} \left[1 + \frac{2}{N} \sum_{t=1}^{N-1} (N-t)\, R_t\right] \qquad [43]$$

Pour une série indépendante ($R_t = 0$ pour $t \neq 0$), on retrouve bien σ^2/N .

Dans le cas qui nous intéresse d'une série Y_j interpolée à partir d'une série X_i , on a démontré [2.3] que pour une série X_i centrée ($\overline{X} = 0$), de taille k assez grande, les termes de boût sont négligeables et l'on a, d'après l'équation [10]:

$$\text{var } Y = \left[\frac{2p^2+1}{3p^2} + r_1 \frac{p^2-1}{3p^2}\right] \text{var } X \qquad [44]$$

Si l'on combine les équations [43] et [44], on obtient:

$$\text{var } \overline{Y} = \frac{\text{var } X}{N} \left[\frac{2p^2+1}{3p^2} + r_1 \frac{p^2-1}{3p^2}\right] \left[1 + \frac{2}{N} \sum_{t=1}^{N-1} (N-t) R_t\right] \qquad [45]$$

La taille de la série interpolée est $N = (k-1)\,p + 1$.

5.2 Calcul de var $\overline{Y}$ dans le cas général

Pour évaluer cette expression, on utilise l'équation générale [34] relative à une série centrée ($\overline{X} = 0$) et longue (k grand):

$$R_t = \frac{3}{(k-1)\,p+1-t} \cdot$$

$$\left[\frac{r_{m-1}\, a_3\, (k-m+1) + r_m\, (2a_1+b_3)(k-m) + r_{m+1}(2b_1+a_4)(k-m-1) + r_{m+2}\, b_3(k-m-2)}{(2p^2+1) + r_1\, (p^2-1)}\right]$$

Pour faciliter le calcul de var $\overline{Y}$ dans l'équation [45], on calcule d'abord:

$$U = \sum_{t=1}^{N-1} (N-t)\ R_t = \sum_{t=1}^{(k-1)p} [(k-1)\ p+1-t]\ R_t$$

En posant:

$$A_{(t',m)} = r_{m-1}\ a_3\ (k-m+1) + r_m\ (2a_1 + b_3)\ (k-m)$$

$$+ r_{m+1}\ (2b_1 + a_4)(k-m-1) + r_{m+2}\ b_3\ (k-m-2)$$

$$B = \frac{(2p^2+1) + r_1\ (p^2-1)}{3}$$

$$N = (k-1)\ p + 1 \quad \text{et} \quad t = mp + t'$$

on a alors: $$R_t = \frac{A_{(t',m)}}{[(k-1)\ p+1-t]\cdot B}$$

$$V_{(t',m)} = \sum A_{(t',m)} = BU = \sum [(k-1)\ p+1-t]\ \cdot B \cdot R_t$$

l'équation [45] devient:

$$\text{var}\ \overline{Y} = \frac{\text{var}\ X}{N} \left[\frac{2p^2+1}{3p^2} + r_1\ \frac{p^2-1}{3p^2}\right] \left[1 + \frac{2}{N}\ U\right]$$

$$= \frac{\text{var}\ X}{N\ p^2}\ B \left[1 + \frac{2}{N}\ \frac{V}{B}\right] = \frac{\text{var}\ X}{N\ p^2} \left[B + \frac{2V}{N}\right] \qquad [46]$$

Il faut donc évaluer: $$V_{(t',m)} = \sum_{m=0}^{k-2} \sum_{t'=1}^{p} A_{(t',m)}$$

On note que:

- les coefficients a_i et b_i sont indépendants de m et ne dépendent que de t' ,

- que les coefficients r_i sont indépendants de t' et ne dépendent que de m ;

Cette situation permet de séparer les sommations et d'écrire $V_{(t',m)}$ sous la forme:

$$V_{(t',m)} = \sum_{m=0}^{k-2} [r_{m-1} (k-m+1) \sum_{t'=1}^{p} a_3 + r_m (k-m) \sum_{t'=1}^{p} (2a_1 + b_3)$$

$$+ r_{m+1} (k-m-1) \sum_{t'=1}^{p} (2b_1 + a_4) + r_{m+2} (k-m-2) \sum_{t'=1}^{p} b_4] \quad [47]$$

Evaluation de $V_{(t',m)}$

Pour évaluer cette expression, on pose d'abord:

$$\alpha = \sum_{t'=1}^{t'=p} a_3 = \frac{1}{6} \sum_{1}^{p} [-t'^3 + 3p\, t'^2 + t' (1-3p^2) + p^3-p]$$

$$\beta = \sum_{t'=1}^{t'=p} (2a_1 + b_3) = \frac{1}{6} \sum_{1}^{p} [3t'^3 - 6pt'^2 - 3t' + 4p^3 + 2p]$$

$$\gamma = \sum_{t'=1}^{t'=p} (2b_1 + a_4) = \frac{1}{6} \sum_{1}^{p} [-3t'^3 + 3pt'^2 + 3t'(p^2+1) + p^3-p]$$

$$\delta = \sum_{t'=1}^{t'=p} b_4 = \frac{1}{6} \sum_{1}^{p} [t'^3-t']$$

Ces sommations sont effectuées en utilisant les identités [A-1], [A-3] et [A-5]:

$$\sum_{1}^{p} t' = \frac{p \cdot p+1}{2} \; ; \; \sum_{1}^{p} t'^2 = \frac{p \cdot p+1 \cdot 2p+1}{6} \; ; \; \sum_{1}^{p} t'^3 = \frac{p^2 \cdot (p+1)^2}{4}$$

ce qui donne:

$$\alpha = \frac{p(p+1)(p-1)(p-2)}{24} \; ; \qquad \beta = \frac{p(p-1)(11p^2+5p+6)}{24}$$

$$\gamma = \frac{p(p+1)(11p^2-5p+6)}{24} \; ; \qquad \delta = \frac{p(p+1)(p-)(p+2)}{24} \qquad [48]$$

L'équation [47] s'écrit donc:

$$V_{(t',m)} = \sum_{m=0}^{k-2} [r_{m-1} (k-m+1) \alpha + r_m (k-m) \beta$$

$$+ r_{m+1} (k-m-1) \gamma + r_{m+2} (k-m-2) \delta]$$

La sommation s'écrit, en alignant les coefficients d'autocorrélation r_i de même rang:

[m=0] $V_{(t'm)} = r_{-1}(k+1)\alpha + r_0 k \beta + r_1(k-1)\gamma + r_2(k-2)\delta$

[m=1] $+ r_0 k \alpha + r_1(k-1)\beta + r_2(k-2)\gamma + r_3(k-3)\delta \ldots$

[m=2] $+ r_1(k-1)\alpha + r_2(k-2)\beta + r_3(k-3)\gamma \ldots$

[m=3] -------- $+ r_2(k-2)\alpha + r_3(k-3)\beta \ldots$

[m=4] $+ r_3(k-3)\alpha \ldots$

....

....

....

....

$$[m=k-5] \quad + 6\, r_{k-6} + 5\, r_{k-5}\, \beta + 4\, r_{k-4}\, \gamma + 3\, r_{k-3}\, \delta \quad \text{--------} \quad \text{--------}$$

$$[m=k-4] \quad + 5\, r_{k-5}\, \alpha + 4\, r_{k-4}\, \beta + 3\, r_{k-3}\, \gamma + 2\, r_{k-2}\, \delta \quad \text{--------}$$

$$[m=k-3] \quad + 4\, r_{k-4}\, \alpha + 3\, r_{k-3}\, \beta + 2\, r_{k-2}\, \gamma + 1\, r_{k-1}\, \delta$$

$$[m=k-2] \quad + 3\, r_{k-3}\, \alpha + 2\, r_{k-2}\, \beta + 1\, r_{k-1}\, \gamma + 0.$$

La sommation peut donc s'écrire, en isolant les termes centraux

$C = \sum_{j=1}^{k-1} r_j\ (k-j)\ (\alpha + \beta + \gamma + \delta)$ et en complétant par des termes de début (D) et de fin (F):

$$V_{(t'm)} = D + C + F$$

où
$$C = \sum_{j=1}^{k-1} r_j\ (k-j)\ (\alpha + \beta + \gamma + \delta)$$

$$D = -r_1\ (k-1)\ \delta + r_0\ k\ (\alpha + \beta) + r_{-1}\ (k+1)\ \alpha$$

or $r_0 = 1$ et $r_{-1} = r_1$, d'où:

$$D = k\ (\alpha + \beta) + r_1\ [k\ (\alpha - \delta) + (\alpha + \delta)]$$

$$F = -[(\alpha+\beta)\ r_{k-1} + 2\ \alpha\ r_{k-2}]$$

D'où l'expression sommée de $V_{(t',m)}$:

$$V_{(t',m)} = k\,(\alpha+\beta) + r_1\,[k\,(\alpha-\delta) + (\alpha+\delta)]$$

$$- [(\alpha+\beta)\,r_{k-1} + 2\,\alpha\,r_{k-2}] \qquad [49]$$

$$+ (\alpha+\beta+\gamma+\delta) \sum_{j=1}^{k-1} r_j\,(k-j)$$

Cette expression [49] peut s'exprimer en fonction de p , à partir des équations [48], on a, en effet:

$$\alpha + \beta + \gamma + \delta = p^4 \; ; \qquad \alpha + \beta = \frac{p\cdot p-1}{6}\,[3p^2+p+1]$$

$$\alpha + \delta = \frac{p^2\cdot(p+1)\cdot(p-1)}{12} \; ; \qquad \alpha - \delta = -\,\frac{p(p+1)(p-1)}{6} \; ;$$

$$\alpha = \frac{p(p+1)(p-1)(p-2)}{24}$$

$$V_{(t',m)} = p^4 \sum_{j=1}^{k-1} r_j\,(k-j) + \frac{p\,(p-1)}{12}\,\{2k\,(3p^2+p+1) + r_1[p(p+1)$$

$$- 2\,k\,(p+1)] - [(p+1)(p-2)\,r_{k-2} + 2\,r_{k-1}\,(3p^2+p+1)]\}$$

D'où la valeur de V :

$$V_{(t',m)} = p^4 \sum_{j=1}^{k-1} r_j\,(k-j) + \frac{p(p-1)}{12}\,[2k\,(3p^2+p+1) - (p+1)\,(2k-p)\,r_1$$

$$- (p+1)\,(p-2)\,r_{k-2} - 2\,(3p^2+p+1)\,r_{k-1}] \qquad [50]$$

Si on revient à l'équation [46]:

$$\text{var } \overline{Y} = \frac{\text{var } X}{N\ p^2} \left[B + \frac{2V}{N}\right]$$

avec $N = (k-1)\ p+1$

et $B = \dfrac{(2p^2+1) + r_1\ (p^2-1)}{3}$

on obtient l'expression générale de la variance de la moyenne d'une série interpolée var Y en fonction de la variance de X , de la longueur k et des coefficients d'autocorrélation r_j de la série d'origine X , ainsi que du pas p de l'interpolation linéaire:

$$\begin{aligned}\text{var } \overline{Y} = \text{var } X \Big[&\frac{(2p^2+1) + r_1\ (p^2-1)}{3p^2\ [(k-1)p + 1]} + \frac{2p^2}{[(k-1)p+1]^2} \sum_{j=1}^{k-1} r_j\ (k-j) \\ &+ \frac{(p-1)}{6\ p\ [(k-1)p+1]^2} \{2k\ (3p^2+p+1)-(p+1)(2k-p)\ r_1 \\ &-(p+1)(p-2)\ r_{k-2} - 2\ (3p^2+p+1)\ r_{k-1}\}\Big]\end{aligned} \qquad [51]$$

Cette expression est valable dans le cas d'une série initiale X_i centrée ($\overline{X} = 0$) et longue (k grand) , ces deux conditions conduisant à négliger les effets de bout.

Si dans l'équation [51], on substitue

$$\frac{(2p^2+1) + r_1\ (p^2-1)}{3p^2} = \frac{\text{var } Y}{\text{var } X} \qquad \text{(équation [11])}$$

$$2 \sum_{j=1}^{k-1} (k-j)\ r_j = k^2 \frac{\text{var}\ \overline{X}}{\text{var}\ X} - k \qquad \text{(équation [43])}$$

On obtient après quelques manipulations:

$$\left[\text{var}\ \overline{Y} - \frac{\text{var}\ Y}{(k-1)p+1}\right] = \frac{p^2k^2}{[(k-1)p+1]^2} \left[\text{var}\ \overline{X} - \frac{\text{var}\ X}{k}\right]$$

$$+ \frac{\text{var}\ X}{[(k-1)p+1]^2}\ \frac{p-1}{6p} \{2k\ (3p^2+p+1)-(p+1)(2k-p)r_1$$

$$- (p+1)\ (p-2)\ r_{k-2} - 2\ (3p^2+p+1)\ r_{k-1}\} \qquad [52]$$

Cette équation définit la perte de précision sur la moyenne de la série interpolée par rapport à une série indépendante de même longueur, en fonction des mêmes caractéristiques de la série de base.

5.3 <u>Calcul de var $\overline{Y}$ dans quelques cas particuliers</u>

a) <u>$p = 1$</u>

Dans ce cas, la série interpolée coincide avec la série originale; l'effet des termes de boût compris dans l'accolade disparaît, de plus $B = 1$, on a donc:

$$\text{var}\ \overline{Y} = \text{var}\ X \left[\frac{1}{k} + \frac{2}{k^2} \sum_{j=1}^{k-1} r_j\ (k-j)\right]$$

$$= \frac{\text{var}\ X}{k} \left[1 + \frac{2}{k} \sum_{j=1}^{k-1} r_j\ (k-j)\right] = \text{var}\ \overline{X}$$

On retrouve bien la variance de la moyenne de la série d'origine.

b) cas où p est grand

(k est lui-même grand par hypothèse)

$$\text{var } \overline{Y} \approx \text{var } X \left[\frac{p^2 (2+r_1)}{3 k p^3} + \frac{2p^2}{k^2p^2} \sum_{j=1}^{k-1} r_j (k-j)\right.$$

$$\left. + \frac{1}{6 k^2p^2} \{6 k p^2 - (2k-p) p r_1 - p^2 r_{k-2} - 6 p^2 r_{k-1}\}\right]$$

Si on regroupe les termes en r_1, on tire:

$$\text{var } \overline{Y} \approx \text{var } X \left[\frac{2}{k^2} \sum_{j=1}^{k-1} r_j (k-j) + \frac{2}{3 k p}\right.$$

$$\left. + \frac{1}{k} + \frac{r_1}{6 k^2} - \frac{r_{k-2}}{6 k^2} - \frac{r_{k-1}}{k^2}\right]$$

$$\text{var } \overline{Y} \approx \text{var } X \left[\left(\frac{1}{k} + \frac{2}{k^2} \sum_{j=1}^{k-1} r_j (k-j)\right) +\right.$$

$$\left. + \left(\frac{2}{3 k p} + \frac{r_1 - r_{k-2} - 6 r_{k-1}}{6 k^2}\right)\right] \qquad [53]$$

Dans cette expression, on note que la première partie constitue la variance de la moyenne de la série d'origine et que la deuxième partie représente l'influence de la longueur k, du degré de partition p , et de r_1 , r_{k-1} et r_{k-2} sur var $\overline{Y}$; on note que cet effet tend vers zéro quand k augmente;

c) <u>Cas d'une persistance à court terme</u>

Physiquement, il arrive souvent (par exemple en hydrologie), que les éléments d'une série soient d'autant plus indépendants les uns des autres qu'ils sont distants dans le temps; La dépendance décroit alors avec l'ordre de l'autocorrélation et à partir d'un certain rang L , on peut admettre que les coefficients d'autocorrélation sont négligeables: $r_{L+j} \simeq 0$ pour toute valeur de $j \geqslant 1$.

La valeur de $\text{var}\,\overline{Y}$ dans l'équation [51] se trouve donc simplifiée puisqu'on peut admettre en première approximation que:

$$r_{L+1} = r_{L+2} = \cdots r_{k-2} = r_{k-1} = 0$$

En pratique, pour déterminer l'ordre L à partir duquel les autocorrélations sont négligeables, on peut tester si les autocorrélations sont significativement différentes de 0. Si on regroupe les termes constants et ceux en r_1 et r_j, il vient, si $L < (k-3)$:

$$\text{var}\,\overline{Y} = \frac{\text{var}\,X}{6\,p^2\,[(k-1)p+1]^2}\,[6\,k\,p^4 - 2(p-1)(2\,p^2+1)$$

$$+ r_1\,(p^2-1)\,(p^2 - 2pk + 2) + 12\,p^4 \sum_{j=1}^{L} r_j\,(k-j)] \quad [54]$$

Dans le cas où $L = 1$, c'est-à-dire où seul r_1 est différent de zéro (série de mémoire 1 lag), la formule [53] se simplifie encore:

$$\text{var } \overline{Y} = \frac{\text{var } X}{6\ p^2[(k-1)p+1]^2} [6\ k\ p^4 - 2(p-1)(2p^2+1)$$

$$+ r_1\ (p^2-1)\ (p^2-2kp+2) + 12\ r_1\ (k-1)\ p^4] \qquad [55]$$

Dans le cas où k est très grand, cette équation [55] tend asymptotiquement vers:

$$\text{var } \overline{Y} \approx (1 + 2r_1 - \frac{1}{3}\ \frac{r_1}{p})\ \frac{\text{var } X}{k}$$

Si p est grand, cette expression tend vers:

$$\text{var } \overline{Y} \approx (1 + 2r_1)\ \frac{\text{var } X}{k}$$

d) <u>Cas d'une série de base indépendante</u>

Si la série X_i est indépendante, $r_j = 0$ pour toute valeur de j différente de zéro, les équations [51] et [53] se simplifient et peuvent s'écrire:

$$\text{var } \overline{Y} = \frac{3\ k\ p^4 - (p-1)\ (2p^2+1)}{3\ p^2\ [(k-1)\ p + 1]^2}\ \text{var } X$$

Si on introduit la variance de la moyenne de la série X_i, $\text{var } \overline{X} = \frac{\text{var } X}{k}$, il vient:

$$\text{var } \overline{Y} = \frac{3\ k^2\ p^4 - k\ (p-1)\ (2p^2 + 1)}{3\ p^2\ [(k-1)\ p + 1]^2}\ \text{var } \overline{X} \qquad [56]$$

Si k est très grand, on peut écrire:

$$\operatorname{var} \overline{Y} = \frac{1 - \frac{1}{k} \frac{(p-1)(2p^2+1)}{3\,p^4}}{[1 - \frac{1}{k} \;(1 - \frac{1}{p})]^2} \cdot \operatorname{var} X$$

En posant $\alpha = \frac{(p-1)(2p^2+1)}{3\,p^4}$ et $\beta = (1 - \frac{1}{p})$, on développe cette expression en séries:

$$\operatorname{var} Y = (1 - \frac{\alpha}{k})\;(1 - \frac{\beta}{k})^{-2} \;\operatorname{var} X$$

$$(1 - \frac{\beta}{k})^{-2} = 1 + \frac{2\beta}{k} + \frac{3\beta^2}{k^2} + \varepsilon\;(\frac{1}{k^3})$$

$$\frac{\operatorname{var} \overline{Y}}{\operatorname{var} \overline{X}} \simeq (1 - \frac{\alpha}{k})\;(1 + \frac{2\beta}{k} + \frac{3\beta^2}{k^2}) = 1 + \frac{1}{k}\;(2\beta - \alpha)$$

$$+ \frac{1}{k^2}\;(3\beta - 2\alpha)\;\beta + \varepsilon\;(\frac{1}{k^3})$$

$$2\beta - \alpha = \frac{p-1}{3p^4}\;[6\;p^3 - 2\;p^2 - 1]$$

$$3\beta - 2\alpha = \frac{p-1}{3p^4}\;[9\;p^3 - 4\;p^2 - 2]$$

Finalement:

$$\operatorname{var} \overline{Y} = [1 + \frac{1}{k}\;\frac{p-1}{3\;p^4}\;(6\;p^3 - 2\;p^2 - 1)$$

$$+ \frac{1}{k^2}\;\frac{(p-1)^2}{3\;p^5}\;(9\;p^3 - 4\;p^2 - 2)]\;\operatorname{var} \overline{X} \qquad [57]$$

Si, de plus, p est grand $(p > 10)$

$$\frac{p-1}{3p^4}\;(6\;p^3 - 2\;p - 1) \simeq 2\;[1 - \frac{4}{3p} + \frac{1}{3p^2}] \quad \text{et}$$

$$\frac{(p-1)^2}{3\ p^5}\ (9\ p^3 - 4\ p^2 - 2) \simeq 3\ [1 - \frac{22}{9\ p} + \frac{17}{9\ p^2}]$$

Donc, en se limitant aux termes en k^{-2} et p^{-2}:

$$\text{var}\ \overline{Y} = [1 + \frac{2}{k}\ (1 - \frac{4}{3\ p} + \frac{1}{3\ p^2}) + \frac{3}{k^2}\ (1 - \frac{22}{9\ p} + \frac{11}{9\ p^2})]\ \text{var}\ \overline{X} \quad [58]$$

e) <u>Cas d'un processus de base Markovien</u>

Si la série de base X_j suit un processus Markovien d'ordre 1, alors on a $r_j = {r_1}^j$; BROOKS et CARRUTHERS (1953) ont montré que la variance de la moyenne donnée par la formule générale [43]

$$\text{var}\ \overline{X} = \frac{\text{var}\ X}{k}\ [1 + \frac{2}{k} \sum_{j=1}^{k-1} r_j\ (k-j)]$$

peut s'écrire pour $r_1 \neq 1$ (voir annexe 3, équations [C-3], [C-4] et [C-5]).

$$\text{var}\ \overline{X} = \frac{\text{var}\ X}{k}\ [\frac{1 + r_1}{1 - r_1} - \frac{2}{k}\ r_1\ \frac{1 - {r_1}^k}{(1-r_1)^2}] \quad [59]$$

on a alors:

$$\sum_{j=1}^{k-1} r_j\ (k-j) = \frac{k\ r_1}{1-r_1} - \frac{r_1\ (1-{r_1}^k)}{(1-r_1)^2}$$

Cette expression est introduite dans l'équation [51], en faisant apparaître $\frac{\text{var}\ X}{k}$, variance de la moyenne d'une série indépendante

$$\text{var } \overline{Y} = \frac{\text{var } X}{k} \{\frac{[(2p^2+1)+r_1(p^2-1)]\ k}{3p^2\ [(k-1)\ p + 1]} + \frac{2\ p^2\ k}{[(k-1)p+1]^2}$$

$$[\frac{kr_1}{1-r_1} - \frac{r_1\ (1-r_1{}^k)}{(1-r_1)^2}] + \frac{(p-1)\ k}{6p[(k-1)p+1]^2}$$

$$[2k\ (3p^2+p+1)-(p+1)(2k-p)r_1 - (p+1)(p-2)r_1{}^{k-2}$$

$$- 2\ (3p^2+p+1)\ r_1{}^{k-1}]\} \qquad [60]$$

Dans le cas où k est très grand, cette équation [60] tend asymptotiquement vers:

$$\text{var } \overline{Y} \simeq \frac{\text{var } X}{k} \{\frac{(2p^2+1) + r_1\ (p^2-1)}{3p^3} + \frac{2r_1}{1-r_1}$$

$$+ \frac{(p-1)\ [(3p^2+p+1)-(p-1)\ r_1]}{3p^3}\}$$

Si, de plus p devient très grand:

$$\text{var } \overline{Y} \simeq \frac{\text{var } X}{k} \cdot \frac{1 + r_1}{1 - r_1}$$

Dans le cas où $r_1 = 1$, la formule de BROOKS et CARRUTHERS n'est formellement plus valable:

$$\sum_{j=1}^{j-1} r_j\ (k-j) = \sum_{j=1}^{k-j} r_1{}^j\ (k-j) = \sum_{j=1}^{k-j} (k-j) =$$

$$(k-1) + \ldots + 1 = \frac{k\ (k-1)}{2}$$

En reportant ces résultats dans [51], on obtient:

$$\text{var } \overline{Y} = \text{var } X \left[\frac{(2p^2+1) + p^2 - 1}{3p^2 \left[(k-1)\ p + 1\right]} + \frac{2p^2}{\left[(k-1)\ p+1\right]^2} \cdot \frac{k(k-1)}{2}\right.$$

$$+ \frac{(p-1)}{6p\left[(k-1)p+1\right]^2} (2k\ (3p^2+p+1) - (p+1)\ (2k-p)$$

$$\left. -(p+1)(p-2) -2\ (3p^2 + p + 1)\right]$$

$$= \text{ var } X \cdot \frac{p^2\ k^2 - p^2 + 1}{\left[(k-1)\ p + 1\right]^2}$$

- Vérification pour $p = 1$

On a $\text{var } \overline{Y} = \text{var } \overline{X} = \text{var } X = 0$, dans ce cas, tous les éléments de la série originale sont égaux, il en est de même de ceux de la série interpolée.

5.4 Décomposition du contenu en information

Si on prend comme référence la variance de la série de base supposée indépendante (équation [42]), le contenu en information peut s'écrire:

$$I = \frac{\frac{\text{var } X}{k}}{\text{var } \overline{Y}} = \frac{\frac{\text{var } X}{k}}{\text{var } \overline{X}} \cdot \frac{\text{var } \overline{X}}{\text{var } \overline{Y}} = I_1 \cdot I_2 \qquad [61]$$

Cette équation [60] définit deux indices:

$$I_1 = \frac{\frac{\text{var } X}{k}}{\text{var } \overline{X}}$$: contenu en information de la série de base relié à la persistance de cette série

$I_2 = \dfrac{\text{var } \overline{X}}{\text{var } \overline{Y}}$: effet de l'interpolation sur les variances des moyennes des séries de base et interpolées

Dans le cas général, l'équation [51] ne permet pas la décomposition analytique de I en I_1 et I_2, c'est pourquoi nous allons effectuer cette décomposition uniquement dans les cas particuliers étudiés au paragraphe 5.3.

5.5 Décomposition du contenu en information dans quelques cas particuliers

Les différentes valeurs de I, I_1 et I_2 ont été calculées pour $p = 2, 3, 4, 5, 10, 20$ et 50 , $k = 5, 10, 20, 50, 100$ et 500 $r_1 = 0.1, 0.3, 0.5, 0.7$ et 0.9 .

Les valeurs de I_1 sont calculées en faisant $p = 1$ dans la formule pertinente (tableau 4).

a) Cas d'une série indépendante

Dans ce cas $I_1 = 1$, et d'après l'équation [56]

$$I_2 = \frac{3p^2 \left[(k-1)\ p + 1\right]^2}{3k^2p^4 - k(p-1)(2p^2+1)}$$

On a donc ici $I = I_2$. Les valeurs de I_2 déduites de l'équation [56] apparaissent aux tableaux 5 et 6, cas 1; les valeurs calculées par l'équation [56] sont pratiquement identiques à celles calculées par les équations [57] et [58] pour $k > 5$. On note que pour k fixe, I_2 décroit quand p croît et que, Cas 1 : Série de persistance à court terme ($r_1 \neq 0$).

r_1 \ k	5	10	20	50	100	500	Infini
0	1	1	1	1	1	1	1
0.1	0.8621	0.8475	0.8403	0.8361	0.8347	0.8336	0.8333
0.3	0.6757	0.6494	0.6369	0.6297	0.6274	0.6255	0.6250
0.5	0.5556	0.5263	0.5128	0.5051	0.5025	0.5005	0.5000
0.7	0.4717	0.4425	0.4292	0.4216	0.4191	0.4172	0.4167
0.9	0.4098	0.3817	0.369	0.3618	0.3595	0.3576	0.3571

Cas 2 : Série Markovienne d'ordre 1.

r_1 \ k	5	10	20	50	100	500	Infini
0	1	1	1	1	1	1	1
0.1	0.8526	0.8351	0.8265	0.8215	0.8198	0.8185	0.8182
0.3	0.62	0.5765	0.5568	0.5457	0.542	0.5392	0.5385
0.5	0.4494	0.3846	0.3571	0.3425	0.3378	0.3342	0.3333
0.7	0.3248	0.2407	0.2045	0.1867	0.1815	0.1774	0.1765
0.9	0.2349	0.1374	0.09014	0.06486	0.05814	0.05365	0.0526

Tableau 4 : Contenu en information I_1 de la série de base.

Cas 1

$r_1 = 0$

p \ k	5	10	20	50	100	500
2	0.8416	0.9197	0.9596	0.9838	0.9919	0.9984
3	0.7754	0.8849	0.9418	0.9766	0.9883	0.9976
4	0.7416	0.8668	0.9324	0.9727	0.9863	0.9973
5	0.7213	0.8557	0.9266	0.9704	0.9851	0.997
10	0.6806	0.8331	0.9148	0.9655	0.9827	0.9965
20	0.6603	0.8216	0.9087	0.963	0.9814	0.9963
50	0.6481	0.8147	0.905	0.9614	0.9806	0.9961

Cas 2

$r_1 = 0.1$

p \ k	5	10	20	50	100	500
2	0.8436	0.9255	0.9677	0.9933	1.002	1.009
3	0.7762	0.889	0.9478	0.9839	0.9961	1.006
4	0.7417	0.8696	0.9369	0.9784	0.9925	1.004
5	0.721	0.8577	0.9302	0.975	0.9901	1.002
10	0.6794	0.8335	0.9162	0.9677	0.9852	0.9992
20	0.6587	0.8212	0.9091	0.9639	0.9826	0.9976
50	0.6464	0.8138	0.9048	0.9617	0.981	0.9966

Cas 3

$r_1 = 0.3$

p \ k	5	10	20	50	100	500
2	0.8465	0.9332	0.9781	1.006	1.015	1.022
3	0.7774	0.8942	0.9556	0.9933	1.006	1.016
4	0.7419	0.8732	0.9428	0.9857	1	1.012
5	0.7205	0.8603	0.9347	0.9808	0.9964	1.009
10	0.6778	0.834	0.9181	0.9705	0.9883	1.003
20	0.6567	0.8207	0.9096	0.9652	0.9841	0.9993
50	0.6441	0.8127	0.9045	0.9619	0.9815	0.9973

Tableau 5.1: Valeurs de I_2 selon k et p pour une série de base ayant une persistance à court terme $r_1 \neq 0$.

Cas 4 $r_1 = 0.5$

p \ k	5	10	20	50	100	500
2	0.8483	0.938	0.9845	1.013	1.023	1.03
3	0.7782	0.8975	0.9604	0.9991	1.012	1.023
4	0.742	0.8755	0.9464	0.9902	1.005	1.017
5	0.7202	0.8619	0.9375	0.9844	1	1.013
10	0.6768	0.8343	0.9192	0.9722	0.9902	1.005
20	0.6554	0.8204	0.9099	0.9659	0.9849	1
50	0.6426	0.812	0.9043	0.9621	0.9818	0.9977

Cas 5 $r_1 = 0.7$

p \ k	5	10	20	50	100	500
2	0.8496	0.9413	0.989	1.018	1.028	1.036
3	0.7787	0.8998	0.9637	1.003	1.016	1.027
4	0.7421	0.8771	0.9488	0.9932	1.008	1.02
5	0.72	0.863	0.9394	0.9868	1.003	1.016
10	0.6761	0.8345	0.92	0.9733	0.9914	1.006
20	0.6544	0.8202	0.9101	0.9664	0.9855	1.001
50	0.6416	0.8116	0.9042	0.9622	0.982	0.9979

Cas 6 $r_1 = 0.9$

p \ k	5	10	20	50	100	500
2	0.8505	0.9437	0.9922	1.022	1.032	1.04
3	0.7791	0.9015	0.966	1.006	1.019	1.03
4	0.7421	0.8782	0.9506	0.9953	1.01	1.023
5	0.7198	0.8639	0.9408	0.9885	1.005	1.018
10	0.6756	0.8347	0.9206	0.9741	0.9923	1.007
20	0.6538	0.82	0.9103	0.9668	0.986	1.001
50	0.6408	0.8112	0.9041	0.9623	0.9821	0.9981

Tableau 5.2 : Valeurs de I_2 selon k et p pour une série de base ayant une persistance à court terme $r_1 \neq 0$.

Cas 1
$r = 0$

p \ k	5	10	20	50	100	500
2	0.8416	0.9197	0.9596	0.9838	0.9919	0.9984
3	0.7754	0.8849	0.9418	0.9766	0.9883	0.9976
4	0.7416	0.8668	0.9324	0.9727	0.9863	0.9973
5	0.7213	0.8557	0.9266	0.9704	0.9851	0.997
10	0.6806	0.8331	0.9148	0.9655	0.9827	0.9965
20	0.6603	0.8216	0.9087	0.963	0.9814	0.9963
50	0.6481	0.8147	0.905	0.9614	0.9806	0.9961

Cas 2
$r = .1$

p \ k	5	10	20	50	100	500
2	0.8358	0.9164	0.9578	0.983	0.9915	0.9983
3	0.7706	0.882	0.9402	0.9759	0.9879	0.9976
4	0.7375	0.8642	0.931	0.9721	0.986	0.9972
5	0.7176	0.8534	0.9253	0.9698	0.9849	0.997
10	0.6778	0.8313	0.9138	0.9651	0.9825	0.9965
20	0.658	0.8202	0.9079	0.9626	0.9812	0.9962
50	0.6461	0.8134	0.9043	0.9611	0.9805	0.9961

Cas 3
$r = .3$

p \ k	5	10	20	50	100	500
2	0.8278	0.9114	0.9551	0.9819	0.9909	0.9982
3	0.7643	0.8778	0.9379	0.9749	0.9874	0.9975
4	0.7323	0.8606	0.929	0.9713	0.9856	0.9971
5	0.7132	0.8501	0.9236	0.9691	0.9845	0.9969
10	0.675	0.829	0.9125	0.9645	0.9822	0.9964
20	0.656	0.8184	0.9069	0.9622	0.981	0.9962
50	0.6446	0.812	0.9035	0.9608	0.9803	0.996

Tableau 6.1 : Valeurs de I_2 selon k et p pour une série de base Markovienne d'ordre 1.

Cas 4

r = .5

p \ k	5	10	20	50	100	500
2	0.8245	0.908	0.9533	0.9812	0.9905	0.9981
3	0.7627	0.875	0.9364	0.9743	0.9871	0.9974
4	0.7318	0.8583	0.9277	0.9708	0.9853	0.9971
5	0.7133	0.8482	0.9225	0.9686	0.9842	0.9968
10	0.6765	0.8279	0.9119	0.9643	0.9821	0.9964
20	0.6582	0.8177	0.9066	0.9621	0.9809	0.9962
50	0.6473	0.8116	0.9033	0.9607	0.9803	0.996

Cas 5

r = .7

p \ k	5	10	20	50	100	500
2	0.8258	0.9062	0.952	0.9806	0.9903	0.9981
3	0.766	0.8739	0.9354	0.9739	0.9869	0.9974
4	0.7362	0.8577	0.927	0.9704	0.9852	0.997
5	0.7183	0.8479	0.9219	0.9684	0.9841	0.9968
10	0.6828	0.8283	0.9117	0.9642	0.982	0.9964
20	0.6652	0.8185	0.9066	0.9621	0.981	0.9962
50	0.6546	0.8126	0.9035	0.9608	0.9803	0.996

Cas 6

r = .9

p \ k	5	10	20	50	100	500
2	0.8311	0.9072	0.9515	0.9803	0.9901	0.998
3	0.7737	0.8759	0.9353	0.9736	0.9868	0.9973
4	0.7449	0.8603	0.9271	0.9703	0.9851	0.997
5	0.7277	0.8509	0.9222	0.9683	0.9841	0.9968
10	0.6935	0.8322	0.9124	0.9643	0.982	0.9964
20	0.6764	0.8228	0.9075	0.9622	0.981	0.9962
50	0.6662	0.8172	0.9045	0.961	0.9804	0.9961

Tableau 6.2 : Valeurs de I_2 selon k et p pour une série de base Markovienne d'ordre 1.

pour p fixe, I_2 croît quand k croit; on note aussi que le premier effet, c'est-à-dire l'influence du degré de partition p paraît le plus important sur les valeurs de I_2 .

b) Cas d'une persistance à court terme

Le tableau 4, cas 1, présente le contenu en information I_1 de la série de base pour diverses valeurs de k et r_1 ; on note que, pour r_1 fixe, I_1 décroît légèrement quand k augmente, alors que, pour k fixé, I_1 décroît rapidement quand r_1 croît de 0 à 0.9. Cette décroissance est mise en évidence à la figure 4 dans le cas où k vaut 20.

Les tableaux 5.1 et 5.2 donnent les valeurs de I_2 selon k et p , pour diverses valeurs de r_1; on note que, pour p et r_1 fixes, I_2 croît quand k croît; pour k et r_1 fixes, I_2 décroît quand p croît et pour k et p fixes, I_2 reste très stable quand r_1 croît.

On remarque que les sens de variation de I_1 et I_2 sont les mêmes selon p et r_1 , mais opposés par rapport à k ; ceci explique que, pour r_1 et p fixes, la valeur asymptotique $I = \frac{1}{1 + 2r_1}$ (équation [55]) de $I = I_1 \cdot I_2$ puisse être atteinte par valeurs supérieures ou inférieures.

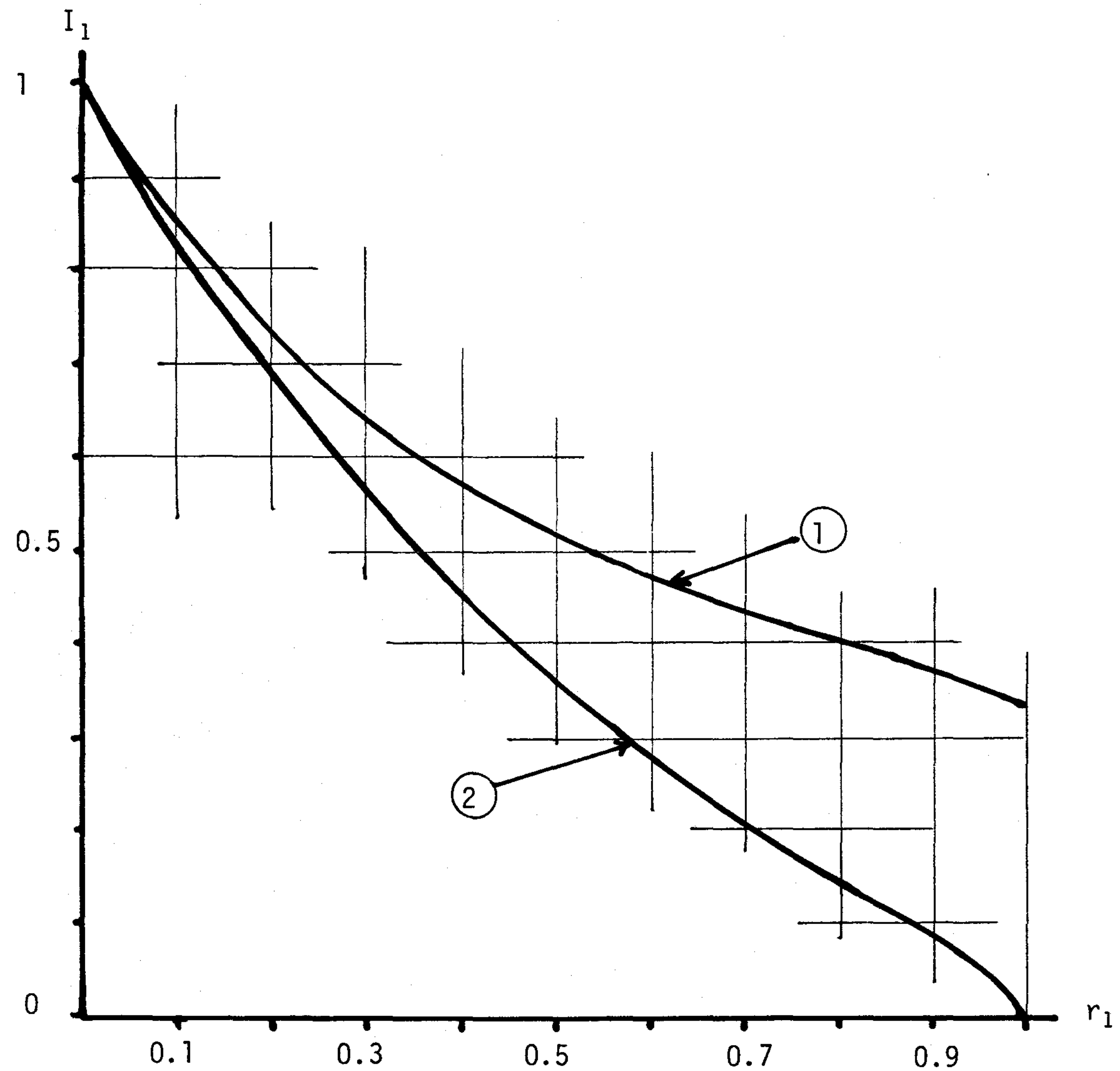


Figure 4 : Comparaison des contenus en information I_1 d'une série de faible persistance ① et d'une série Markovienne ② selon les valeurs de r_1 (k = 20).

c) Cas d'un processus de base Markovien

Le tableau 4, cas 2, présente le contenu en information I_1 de la série de base pour diverses valeurs de k et r ; on note que pour r fixe, I_1 décroît lentement quand k augmente, alors que, pour k fixe, I_1 décroit très rapidement quand r croît de 0 à 0.9; cette décroissance, mise en évidence à la figure 4 pour $k = 20$ est encore plus rapide que dans le cas d'une série ayant une persistance à court terme.

Les tableaux 6.1 et 6.2 donnent les valeurs de I_2 selon k et p pour diverses valeurs de r , on note que pour p et r fixes, I_2 croît quand k croît; pour k et r fixes, I_2 décroît quand p croît et pour k et p fixes, I_2 décroît légèrement quand r croît. On remarque que les sens de variation de I_1 et I_2 sont les mêmes selon p et r_1 , mais opposés par rapport à k ; ceci explique que, pour r et p fixes, la valeur asymptotique $I_{k,p\ \infty} = \frac{1-r_1}{1+r_1}$ (équation [60]) puisse être atteinte par valeurs supérieures ou inférieures.

5.6 Séparation approximative des deux effets

Les tableaux 5 et 6 montrent que pour k et p donnés, les valeurs de I_2 pour différentes valeurs de r_1 diffèrent très peu de celles d'une série indépendante; avec cette approximation, on peut donc séparer les deux effets: celui de la structure de la série de base et celui de l'interpolation:

$$I = I_1 \cdot I_2 \qquad \text{avec}$$

$$I_2 = \frac{3p^2\ [(k-1)\ p + 1]^2}{3k^2p^4 - k\ (p-1)\ (2p^2+1)}$$ expression indépendante de r_i

et $$I_1 = \frac{\frac{\text{var } X}{k}}{\text{var } \overline{X}} = [1 + \frac{2}{k} \sum_{j=1}^{k-1} r_j\ (k-j)]^{-1}$$

D'après l'équation [43], I_1 s'écrit, dans le cas d'une série de courte persistance:

$$I_1 = [1 + \frac{2}{k}\ (k-1)\ r_1]^{-1} = \frac{1}{1 + 2r_1 \frac{k-1}{k}}$$

expression indépendante de p et dont la valeur asymptotique pour k très grand est bien $\frac{1}{1+2r_1}$.

D'après l'équation [59], I_1 s'écrit, dans le cas d'une série markovienne d'ordre 1:

$$I_1 = [\frac{1 + r_1}{1 - r_1} - \frac{2}{k}\ r_1 \frac{1 - r_1^{\ k}}{(1 - r_1)^2}]^{-1}$$

expression indépendante de p et dont la valeur asymptotique pour k très grand est bien $\frac{1 - r_1}{1 + r_1}$.

On évalue maintenant dans quelles conditions une telle approximation est valable pour une série quelconque.

Pour une série indépendante, on a:

$$I_{2_i} = \frac{6p^2\ [(k-1)\ p + 1]^2}{6k^2p^4 - 2k\ (p-1)(2p^2+1)} \qquad [62]$$

Pour une série quelconque, on a, d'après [51]:

$$I_{2_q} = \frac{1}{k}\left[\frac{(2p^2+1)+r_1(p^2-1)}{3p^2[(k-1)\ p+1]} + \frac{2p^2}{[(k-1)p+1]^2}\sum_{j=1}^{k-1}(k-j)\ r_j\right.$$

$$+ \frac{(p-1)}{6p[(k-1)p+1]^2}\{2k(3p^2+p+1)-(p+1)(2k-p)r_1-(p+1)(p+2)r_{k-2}$$

$$\left. - 2\ (3p^2+p+1)\ r_{k-1}\}\right]^{-1} \cdot \left[1+\frac{2}{k}\sum_{j=1}^{k-1}(k-j)\ r_j\right] \qquad [63]$$

Après réduction au même dénominateur, on évalue:

$$\frac{I_{2_q}}{I_{2_i}} = \frac{N}{D}$$

$$N = \left[1+\frac{2}{k}\sum_{j=1}^{k-1}(k-j)r_j\right]\left[6k^2p^4-2k(p-1)(2p^2+1)\right]$$

$$D = 6k^2p^4-2k(p-1)(2p^2+1)+kr_1(p^3-2kp^2+2k-p+2)+12kp^4\sum_{j=1}^{k-1}(k-j)r_j$$

$$-k(p^2-1)(p-2)r_{k-2} - 2k(p-1)(3p^2+p+1)r_{k-1}$$

pour k et p élevés.

Définissons comme YULE (1945) une persistance globale de la série:

$$\overline{r} = \frac{2\sum_{j=1}^{k-1} r_j\ (k-j)}{k \cdot k-1}$$

On en tire: $\frac{2}{k} \sum_{j=1}^{k-1} r_j (k-j) = (k-1)\,\overline{r}$, le rapport $\frac{I_{2_q}}{I_{2_i}}$ peut s'écrire, en conservant les termes prépondérants en p et k :

$$\frac{I_{2_q}}{I_{2_i}} \simeq \frac{1 + k\overline{r}}{1 + 6k\overline{r} - \frac{r_1}{3k^2} + \frac{r_1 - r_{k-2} - 6r_{k-1}}{6kp}}$$

$$\simeq (1 + k\overline{r})\ (1 - 6k\overline{r} + \frac{r_1}{6k^2} - \frac{r_1 - r_{k-2} - 6r_{k-1}}{6kp} \qquad [64]$$

Cette équation [64] permet de définir les conditions pour lesquelles une séparation approximative des deux effets a lieu; $\frac{I_{2_q}}{I_{2_i}}$ doit être peu différent de 1; les termes en r_1, r_{k-2} et r_{k-1} , dont le numérateur est borné par 1 en valeur absolue, et dont les dénominateurs décroissent en k^2 ou en kp sont rapidement négligeables devant 1;
il reste $\frac{I_{2_q}}{I_{2_i}} \simeq (1+k\overline{r})\ (1-6k\overline{r}) = 1 - 5k\overline{r} - 6k^2\overline{r}^2$. Cette quantité est proche de 1 pour $k\overline{r} \simeq 0$, soit

$$\frac{2 \sum_{j=1}^{k-1} r_j (k-j)}{k-1} \simeq 0 \qquad [65]$$

Cette relation exclut donc les séries non-stationnaires et les séries dont l'autocorrélogramme r_j décroît moins vite que $r_1{}^j$ (série Markovienne d'ordre 1).

6. Synthèse et perspective d'applications

Il est évident que l'interpolation linéaire ne crée pas d'information nouvelle par rapport à la série d'origine, mais la répartit dans l'intervalle entre les mesures; Dans cette opération, les caractéristiques statistiques globales des deux séries sont modifiées, et dans ce rapport, nous avons déduit celles de la série interpolée de celles de la série de base dont elle dérive.

De l'étude effectuée sur la variance de la moyenne et sur la notion de contenu en information de la série, il ressort que l'interpolation elle-même introduit une faible perte de contenu en information I_2 , alors que l'effet de structure de la série de base I_2 peut être très important et prépondérant; On a aussi établi les conditions nécessaires au découplage approximatif de ces deux effets.

Considérons maintenant une expression fonctionnelle $z = f(x, y)$, les variates géophysiques x et y ayant été mesurées à des pas de temps différents, multiples entiers ou non. Dans ces conditions, il peut être très intéressant, compte tenu des variabilités propres de x et de y de les transformer par interpolation en des séries x' et y' contenant un peu moins d'information, mais permettant d'exploiter complètement la relation fonctionnelle $z = f(x', y')$ au pas de temps choisi. Cette perspective d'application sera développée dans une autre étude à partir des résultats acquis ici.

REFERENCES BIBLIOGRAPHIQUES

BAYLEY, G.V. et J.M. HAMMERSLEY (1946)
The effective number of independent observations in an autocorrelated time series. J. Roy. Stat. Soc., 8-18, p. 184-197.

BROOKS, C.E.P. et N. CARRUTHERS (1953)
Handbook of statistical methods in meteorology, M.O. 538. Meteorological office, London, 412 p.

MATALAS, N.C. et W.B. LANGBEIN (1962)
Information content of the mean. J. Geoph. Res., Vol. 67, No 9, p. 3441-3448.

MEJIA, J.M., I. RODRIGUEZ-ITURBE et D.R. DAWDY (1972)
Streamflow simulation. 2. The Broken Line Process as a potential model for hidrologic simulation. Water Res. Res., Vol. 8, No 4, p. 931-941.

YULE, G.U. (1945)
A method of studying time series based on their internal correlations. J. Roy Statis. Soc., 108, 208.

ANNEXE 1

Identités concernant les nombres entiers

1.1 Nombres entiers consécutifs: $P' = 1, 2, 3 \ldots n$

$$\sum_{p'=1}^{n} p' = \frac{n \cdot n+1}{2} \qquad [\text{ A-1 }]$$

$$\sum_{p'=1}^{n-1} p' = \frac{n-1 \cdot n}{2} \qquad [\text{ A-2 }]$$

$$\sum_{p'=1}^{n} p'^2 = \frac{n \cdot n+1 \cdot 2\ n+1}{6} \qquad [\text{ A-3 }]$$

$$\sum_{p'=1}^{n-1} p'^2 = \frac{n-1 \cdot n \cdot 2\ n-1}{6} \qquad [\text{ A-4 }]$$

$$\sum_{1}^{n} p'^3 = \frac{n^2 \cdot (n+1)^2}{4} \qquad [\text{ A-5 }]$$

$$\sum_{1}^{n} p'^4 = \frac{n}{30}\ (n+1)\ (2n+1)\ (3n^2+3n-1) \qquad [\text{ A-6}]$$

$$\sum_{p'=1}^{p} (p'-1)\ (p-p'+1) = \frac{p\ (p^2-1)}{6} \qquad [\text{ A-7 }]$$

1.2 Séries partielles de nombres entiers

On utilise la relation:

$$\sum_{(m+1)\,p-t+1}^{(m+1)\,p} p' = \sum_{1}^{(m+1)p} p' - \sum_{1}^{(m+1)\,p\,-\,t} p'$$

Cette relation est appliquée aux identités [A-1] et [A-3] et les résultats sont simplifiés en utilisant les identités:

$$a \bullet a+1 - b \bullet b+1 = (a-b)\ (a+b+1)$$

et $$\frac{a \bullet a+1 \bullet 2\ a+1 - b \bullet b+1 \bullet 2\ b+1}{6} = (a-b)\ \left[(a+b+1)\ (2a+2b+1)-2ab\right]$$

il vient:

$$\sum_{(m+1)\ p-t+1}^{(m+1)\ p} p' = \frac{2\ (m+1)\ p-t+1}{2} \bullet t \qquad [\ A\text{-}8\ ]$$

$$\sum_{(m+1)\ p-t+1}^{(m+1)\ p} p'^2 = \left[6(m+1)^2\ p^2 + 6(m+1)(1-t)\ p+(1-t)(1-2\ t)\right] \bullet \frac{t}{6} \qquad [\ A\text{-}9\ ]$$

$$\sum_{p-t'+2}^{p} K = (t'-1)\ K \qquad [\ A\text{-}10\ ]$$

$$\sum_{p-t'+2}^{p} p' = \frac{p \cdot p+1}{2} - \frac{(p-t'+1)(p-t'+2)}{2} = \frac{(t'-1)(2p-t'+2)}{2} \quad [\text{ A-11 }]$$

$$\sum_{p-t'+2}^{p} p'^{2} = \frac{p \cdot p+1 \cdot 2p+1}{6} - \frac{(p-t'+1)(p-t'+2)(2p-2t'+3)}{6} \quad [\text{ A-12 }]$$

$$= \frac{t'-1}{6} \left[(2p-t'+2)(4p-2t'+3) - 2p(p-t'+1) \right]$$

$$= \frac{t'-1}{6} \left[6p^2 + 6p(2-t') + (2-t')(3-2t') \right]$$

ANNEXE 2

Identités concernant toute série de temps X_i

2.1 Définitions

Dans cette partie, on considère une série X_i de longueur k d'éléments équidistants (c'est-à-dire avec un poids égal); aucune hypothèse n'est faite concernant la distribution statistique des éléments de cette série ou leur persistance; la seule hypothèse implicite concerne la définition des coefficients d'autocorrélation qui présuppose la stationnarité au niveau des moyennes et des écarts-types, soit:

s la somme des éléments X_i ,

$\overline{X}$ la moyenne de k éléments X_i ,

et $\text{var}\, X$ la variance de la série, on a, par définition:

$$s = \sum_1^k X_i = k\,\overline{X}$$

$$\text{var}\, X = \frac{1}{k-1} \sum_1^k (X_i - \overline{X})^2 = \frac{1}{k-1} \left[\sum X_i^2 - k\,\overline{X}^2\right]$$

on en déduit

$$\sum_1^k X_i^2 = (k-1)\,\text{var}\, X + k\,\overline{X}^2 \qquad [\text{ B-1 }]$$

2.2 Autocovariances et coefficients d'autocorrélation

Notons ACV_t et r_t , les auto-covariances et les coefficients d'autocorrélation d'ordre t; par définition, on a: $r_t = \frac{ACV_t}{\text{var } X} = \frac{ACV_t}{ACV_0}$

Calcul à l'ordre 1:

$$ACV_1 = \frac{1}{k-1} \sum_{1}^{k-1} (X_i - \overline{X})\ (X_{i+1} - \overline{X})$$

$$= \frac{1}{k-1} [\sum X_i X_{i+1} - \overline{X} \sum (X_i + X_{i+1}) + \sum \overline{X}^2]$$

or, $\sum_{i=1}^{k-1} (X_i + X_{i+1}) = 2s - (X_1 + X_k) = 2\ k\ \overline{X} - (X_1 + X_k)$, ce qui fait

apparaître les effets de boût X_1 et X_k . En substituant dans ACV_1 , il vient: $ACV_1 = \frac{1}{k-1} [\sum X_i\ X_{i+1} + \overline{X}\ (X_1 + X_k) - (k+1)\ \overline{X}^2]$

$$r_1 = \frac{ACV_1}{\text{var } X} = \frac{\sum X_i\ X_{i+1} + \overline{X}\ (X_1 + X_k) - (k+1)\ \overline{X}^2}{\sum X_i^2 - k\ \overline{X}^2}$$

d'où l'on tire:

$$\sum_{i=1}^{k-1} X_i \; X_{i+1} = r_1 \; (k-1) \; \text{var} \; X + (k+1) \; \overline{X}^2 - (X_1 + X_k) \; \overline{X} \qquad [\text{ B-2 }]$$

Calcul à l'ordre 2: De la même façon

$$ACV_2 = \frac{1}{k-2} \sum_{1}^{k-2} (X_i - \overline{X}) \; (X_{i+2} - \overline{X})$$

$$= \frac{1}{k-2} \; [\sum X_i \; X_{i+2} - \overline{X} \sum (X_i + X_{i+2}) + \sum \overline{X}^2]$$

or $$\sum_{i=1}^{k-2} = (X_i + X_{i+2}) = 2s - [X_1 + X_2 + X_{k-1} + X_k]$$

$$= 2 \; k \; \overline{X} - [X_1 + X_2 + X_{k-1} + X_k]$$

En substituant dans ACV_2 , il vient:

$$ACV_2 = \frac{1}{k-2} \; [\sum X_i \; X_{i+2} - 2 \; k \; \overline{X}^2 + \overline{X} \; (X_1 + X_2 + X_{k-1} + X_k) + (k-2) \; \overline{X}^2]$$

d'où $$r_2 = \frac{ACV_2}{\text{var} \; X} = \frac{k-1}{k-2} \cdot \frac{\sum X_i \; X_{i+2} + \overline{X} \; (X_1 + X_2 + X_{k-1} + X_k) - (k+2) \; \overline{X}^2}{\sum X_i^2 - k \; \overline{X}^2}$$

On en tire:

$$\sum_{i=1}^{k-2} X_i \, X_{i+2} = r_2 \, (k-2) \text{ var } X + (k+2) \, \overline{X}^2 - \overline{X} \, (X_1 + X_2 + X_{k-1} + X_k) \quad [\text{ B-3 }]$$

<u>Généralisation à l'ordre t:</u>

$$ACV_t = \frac{1}{k-t} \sum_{1}^{k-t} (X_i - \overline{X}) \, (X_{i+t} - \overline{X})$$

$$= \frac{1}{k-t} \, [\sum X_i \, X_{i+t} - \overline{X} \sum (X_i + X_{i+t}) + \sum \overline{X}^2]$$

$$\sum_{1}^{k-t} (X_i + X_{i+t}) = 2 \, s - \sum_{i=1}^{t} X_i - \sum_{i=k-t+1}^{k} X_i$$

d'où $$ACV_t = \frac{1}{k-t} \, [\sum X_i \, X_{i+t} - (k+t) \, \overline{X}^2 + \overline{X} \, (\sum_{1}^{t} X_i + \sum_{k-t+1}^{k} X_i)]$$

$$r_t = \frac{ACV_t}{\text{var } X} = \frac{k-1}{k-t} \; \frac{\sum X_i \, X_{i+t} - (k+t) \, \overline{X}^2 + X \, (\sum_{1}^{t} X_i + \sum_{k-t+1}^{k} X_i)}{\sum X_i^2 - k \, X^2}$$

et $$\sum_{1}^{k-t} X_i \, X_{i+t} = r_t \, (k-t) \text{ var } X + (k+t) \, \overline{X}^2 - \overline{X} \, (\sum_{1}^{t} X_i + \sum_{k-t+1}^{k} X_i) \quad [\text{ B-4 }]$$

Cette expression est valable pour $t \geqslant 1$. Pour $t = 0$, on ne retrouve pas l'équation [B-1] à cause de la correction (k-1) effectuée sur le biais de la variance.

ANNEXE 3

Variance de la moyenne d'une série de Markov
(formule de BROOKS et CARRUTHERS)

Dans cette annexe, nous allons démontrer la formule de BROOKS et CARRUTHERS; l'expression générale de la variance de la moyenne d'une série quelconque s'écrit:

$$\mathrm{var}\ \overline{X} = \frac{\mathrm{var}\ X}{k} \left[1 + \frac{2}{k} \sum_{j=1}^{k-1} r_j\ (k-j)\right] \qquad [C-1]$$

où X représente la série étudiée

var X sa variance

var $\overline{X}$ la variance de sa moyenne

k sa longueur et

r_j les coefficients d'autocorrélation successifs

Si la série étudiée est Markovienne, on a:

$$r_j = r_1^{\,j} \text{ ; si l'on pose } W = \sum_{j=1}^{k-1} (k-j)\ r_j = \sum_{j=1}^{k-1} r_1^{\,j}\ (k-j) \text{ et}$$

$$W' = \sum_{j=1}^{k-1} j\ r_1^{\,j} \text{ , on a } W + W' = \sum_{j=1}^{k-1} k\ r_1^{\,j} = k \sum_{j=1}^{k-1} r_1^{\,j}$$

si l'on considère maintenant la somme

$$S = \sum_{j=1}^{k-1} r_1^{\,j} = r_1 + r_1^2 + \ldots\ r_1^{\,k-1} = r_1\ \frac{(1-r_1^{\,k-1})}{1-r_1}$$

$$\frac{dS}{dr_1} = 1 + 2r_1 + \dots (k-1)\ r_1^{k-2}$$

d'oũ $$W' = r_1 \frac{dS}{dr_1}$$

Finalement:

$$W = kS - W' = kS - r_1 \frac{dS}{dr_1}$$

avec $$S = r_1 \frac{(1-r_1^{k-1})}{1-r_1} \qquad [C-2]$$

d'oũ $$\frac{dS}{dr_1} = \frac{1-kr_1^{k-1}}{1-r_1} + \frac{r_1\ (1-r_1^{k-1})}{(1-r_1)^2}$$

$$W = k\ r_1 \frac{(1-r_1^{k-1})}{1-r_1} - r_1 \left[\frac{1 - k\ r_1^{k-1}}{1-r_1} + \frac{r_1\ (1-r_1^{k-1})}{(1-r_1)^2}\right]$$

donc $$W = \frac{k\ r_1}{1-r_1} - r_1 \frac{1-r_1^{k}}{(1-r_1)^2} \qquad [C-3]$$

Si on introduit ce résultat dans l'expression [C-1], on trouve l'expression de la variance de la moyenne:

$$\text{var}\ \overline{X} = \frac{\text{var}\ X}{k} \left[1 + \frac{2}{k} W\right]$$

$$= \frac{\text{var}\ X}{k} \left[1 + 2 \frac{r_1 - r_1^{k}}{1-r_1} - \frac{2}{k} r_1 \left[\frac{(1-r_1)(1-k\ r_1^{k-1})+(r_1-r_1^{k})}{(1-r_1)^2}\right]\right]$$

$$\text{var}\ \overline{X} = \frac{\text{var}\ X}{k} \left[\frac{1+r_1}{1-r_1} - \frac{2}{k} r_1 \frac{1-r_1^{k}}{(1-r_1)^2}\right] \qquad [C-4]$$

ou

$$\text{var}\ \overline{X} = \frac{\text{var}\ X}{k} \left[1 + \frac{2r_1}{1-r_1} \left(1 - \frac{1}{k} \frac{1-r_1^{k}}{1-r_1}\right)\right] \qquad [C-5]$$