

IF505  
Simulation par  
événements discrets



Louis Granger



**ÉCOLE  
POLYTECHNIQUE  
MONTRÉAL**  
*Le génie  
sans frontières*

A-306.13  
340-4782  
granger@granger.gegi.polymtl.ca

## CHAPITRE 9

# ANALYSE DES RÉSULTATS

## ANALYSE DES RÉSULTATS INTRODUCTION

- ☞ Analyse statistique des résultats est souvent négligée.
  - Effectuer une passe de simulation unique d'une longueur déterminée de façon plutôt arbitraire, et de traiter les estimations ainsi obtenues de comme étant les valeurs vraies des paramètres du système.
- ☞ Indispensable d'en faire une analyse statistique appropriée.
- ☞ Une autre difficulté: l'hypothèse que les observations sont indépendantes et distribuées selon une même loi.
  - Les extrants: processus non stationnaire (la distribution des observations varie avec le temps) et autocorrélés (Les observations dépendent les unes des autres).

## ANALYSE DES RÉSULTATS Expérience statistique

- Définition du système

Une compagnie possède un ordinateur central auquel sont reliés en interactif  $n$  terminaux. Chaque usager d'un terminal réfléchit pendant une certaine période temps, puis envoie une requête (tâche) à l'ordinateur. Cette requête entre dans une file d'attente. Le processeur (UCT) traite les requêtes selon la méthode du tourniquet [round-robin] en allouant une tranche de temps à chaque tâche à exécuter. Cette méthode permet de satisfaire plus rapidement les plus petites tâches.

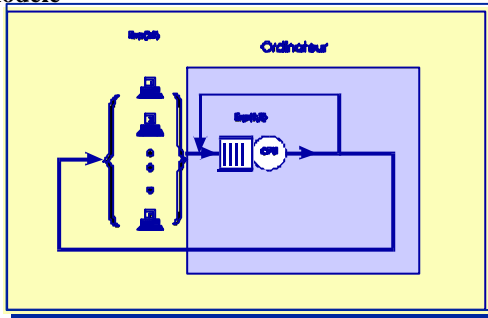
Le temps de réflexion des usagers est en moyenne 25 secondes et suit une distribution exponentielle. Le processeur (UCT) traite ces tâches en 0.8 seconde selon une distribution exponentielle. Le processeur alloue une tranche de 0.1 seconde à laquelle il faut ajouter un [overhead] de 0.015 seconde.

Si on définit le temps de réponse comme l'intervalle de temps entre le moment où la tâche a été soumise et le moment où elle se termine, la compagnie veut maintenir un objectif d'un temps de réponse inférieur à 30 secondes.

Pour un nombre de terminaux  $n = 30, 35, \dots, 60$ , on demande de simuler le système pour 1000 tâches, et de recueillir des statistiques le nombre moyen de tâches en attente, et l'utilisation du processeur.

## ANALYSE DES RÉSULTATS Expérience statistique

### ☞ Modèle



Les résultats obtenus de ce modèle pour un nombre fixe de terminaux, mais pour 30000 tâches avec impression périodique à toutes les 1000 tâches, sont:

**Modèle d'un ordinateur en temps partagé --- Paramètres d'entrée**

**Temps de réflexion est exponentiel de moyenne 25.00 secondes**

**Temps de service est exponentiel de moyenne .8 secondes**

**Quantum 2.000 secondes**

**Changement de contexte .015 secondes**

**Nombre de taches traitées 30000**

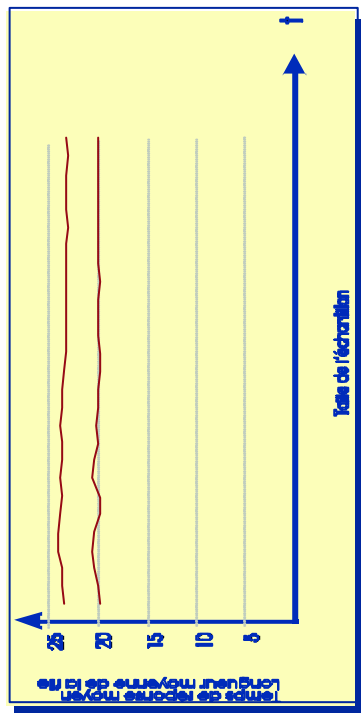
**Intervalle d'impression 1000 taches**

## ANALYSE DES RÉSULTATS Expérience statistique

| Moyenne temps réponse | Nombre moyen dans la file | Nombre de tâches |
|-----------------------|---------------------------|------------------|
| 19.5930               | 23.024                    | 1000             |
| 19.9346               | 23.400                    | 2000             |
| 20.2331               | 23.334                    | 3000             |
| 20.3596               | 23.464                    | 4000             |
| 20.2600               | 23.482                    | 5000             |
| 19.8986               | 23.264                    | 6000             |
| 19.8359               | 23.247                    | 7000             |
| 20.2582               | 23.529                    | 8000             |
| 20.1084               | 23.306                    | 9000             |
| 19.9869               | 23.298                    | 10000            |
| 20.1245               | 23.451                    | 11000            |
| 19.8920               | 23.334                    | 12000            |
| 19.7261               | 23.206                    | 13000            |
| 19.5366               | 23.070                    | 14000            |
| 19.4451               | 22.984                    | 15000            |
| 19.5380               | 23.066                    | 16000            |
| 19.5182               | 23.035                    | 17000            |
| 19.6135               | 23.117                    | 18000            |
| 19.5482               | 23.051                    | 19000            |
| 19.5528               | 23.080                    | 20000            |
| 19.5430               | 23.100                    | 21000            |
| 19.5203               | 23.063                    | 22000            |
| 19.5769               | 23.118                    | 23000            |
| 19.5428               | 23.086                    | 24000            |
| 19.5466               | 23.096                    | 25000            |
| 19.6009               | 23.175                    | 26000            |
| 19.5083               | 23.067                    | 27000            |
| 19.5163               | 23.069                    | 28000            |
| 19.5979               | 23.144                    | 29000            |
| 19.5384               | 23.093                    | 30000            |

## ANALYSE DES RÉSULTATS

### Expérience statistique



## ANALYSE DES RÉSULTATS

### Théorèmes limites

loi (forte) des grands nombres

**Théorème:** Soit  $\{x_1, x_2, \dots, x_n\}$  une suite de variables aléatoires indépendantes et de même loi (i.i.d : indépendantes et Identiquement distribuées) telle que  $E(x_i) = \mu$ . Alors, on a avec une probabilité égale à 1 :

$$\frac{\sum_{i=1}^n x_i}{n} \rightarrow \mu \text{ lorsque } n \rightarrow \infty$$

**Théorème:** Soit  $\{x_1, x_2, \dots, x_n\}$  une suite de variables aléatoires indépendantes et de même loi (i.i.d : indépendantes et identiquement distribuées) ayant une espérance égale à  $\mu$  et une variance égale à  $\sigma^2$ . Alors la variable aléatoire :

$$\frac{\sum_{i=1}^n x_i - n\mu}{\sigma\sqrt{n}}$$

tend vers la loi normale centrée réduite lorsque  $n \rightarrow \infty$  ou encore

$$\lim_{n \rightarrow \infty} P \left\{ \frac{\sum_{i=1}^n x_i - n\mu}{\sigma\sqrt{n}} \in [a, b] \right\} = \frac{1}{\sqrt{2\pi}} \int_a^b e^{-x^2/2} dx = \Phi(x)$$

tend vers la loi normale centrée réduite  $N(0, 1)$  lorsque

$$\bar{x} = \frac{\sum_{i=1}^n x_i}{n} \quad \text{et} \quad \frac{\sigma}{\sqrt{n}}$$

$n \rightarrow \infty$

## ANALYSE DES RÉSULTATS

### Estimation

☞ Introduction

- l'estimation ponctuelle
- l'estimation par intervalle.

☞ Propriétés des estimateurs.

- Estimateur sans biais
- Estimateur efficace
- Le meilleur estimateur linéaire et centré

## ANALYSE DES RÉSULTATS

### Estimation

☞ Estimateur sans biais

- variable aléatoire  $x$  dont la distribution est caractérisée, entre autres, par un paramètre  $\theta$

○  $\hat{q}$  : un estimateur sans biais du paramètre  $\theta$   
 $\hat{q}$  est un estimateur de  $\theta$  si  $E(\hat{q}) = \theta$

$$\bar{x} = \frac{\sum_{i=1}^n x_i}{n} : \text{estimateur sans biais de } \mu$$

$$s^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1} : \text{estimateur sans biais de } \sigma^2.$$

☞ Estimateur efficace

- Un estimateur est efficace si et seulement si il est sans biais et s'il possède une variance minimale.

- Mathématiquement

$\hat{q}$  un estimateur efficace de  $\theta$  si les conditions suivantes sont respectées:

- a)  $\hat{q}$  est sans biais,
- b)  $Var(\hat{q}) \leq Var(\tilde{q})$  où  $\tilde{q}$  un autre estimateur sans biais de  $\theta$ .

☞ Le meilleur estimateur linéaire et centré

L'efficacité se définit de la façon suivante:

$\hat{q}$  est un le meilleur estimateur linéaire et centré de  $\theta$  si les conditions suivantes sont respectées:

- a)  $\hat{q}$  est une fonction linéaire des observations de l'échantillon,
- b)  $\hat{q}$  est sans biais,
- c)  $Var(\hat{q}) \leq Var(\tilde{q})$  où  $\tilde{q}$  est un autre estimateur sans biais de  $\theta$ .

## ANALYSE DES RÉSULTATS

### Méthodes d'estimation

☞ **Méthode des moindres carrés**

∞ Le principe: minimiser l'erreur entre un estimateur et le paramètre estimé.

Exemple:

Echantillon aléatoire:  $\{x_1, x_2, \dots, x_n\}$  provenant d'une population de moyenne  $\mu$  inconnue et de variance  $\sigma^2$  connue.

Chaque observation peut s'exprimer comme  $x_i = \mu + \varepsilon_i$  où  $\mu$ : la moyenne

$\varepsilon_i$ : fluctuation aléatoire autour de  $\mu$

$$S = \sum_{i=1}^n e_i^2 = \sum_{i=1}^n (x_i - \mathbf{m})^2$$

Un estimateur qui minimise cette fonction doit satisfaire les conditions suivantes

$$\frac{dS}{d\mathbf{m}} = 0, \quad \frac{d^2S}{d\mathbf{m}^2} > 0.$$

## SIMULATION DISCRETE

### ANALYSE DES RÉSULTATS

Méthodes d'Estimation

Ce qui entraîne que  $\frac{dS}{d\mathbf{m}} = -2 \sum_{i=1}^n (x_i - \mathbf{m})$  et en utilisant  $\tilde{\mathbf{m}}$  comme estimateur de  $\mu$ , nous pouvons écrire:

$$\sum_{i=1}^n (x_i - \tilde{\mathbf{m}}) = 0$$

$$\sum_{i=1}^n x_i - \sum_{i=1}^n \tilde{\mathbf{m}} = 0$$

$$\sum_{i=1}^n x_i - n\tilde{\mathbf{m}} = 0$$

$$n\tilde{\mathbf{m}} = \sum_{i=1}^n x_i$$

$$\tilde{\mathbf{m}} = \frac{\sum_{i=1}^n x_i}{n}$$

On pourrait aussi vérifier que la dérivée seconde est positive.

## SIMULATION DISCRETE ANALYSE DES RÉSULTATS Méthodes d'Estimation

### ☞ Méthode du maximum de vraisemblance

- Principe du maximum de vraisemblance: maximiser par rapport aux paramètres la fonction de densité  $L=f(x_1, x_2, \dots, x_n / \mathbf{q}_1, \mathbf{q}_2, \dots, \mathbf{q}_k)$  où  $x_1, x_2, \dots, x_n$  est la suite échantillonnée et  $\theta_1, \theta_2, \dots, \theta_k$  représente la suite des paramètres à estimer.

$$\frac{\partial L}{\partial \mathbf{q}_i} = 0 \quad i = 1, \dots, k$$

- Exemple:

La distribution normale a deux paramètres  $\theta_1$  et  $\theta_2$  qui sont en fait  $\mu$  et  $\sigma$ . Pour un échantillon  $\{x_i\}$ , la fonction de vraisemblance est:

$$L = \frac{1}{\mathbf{s}^n (2\mathbf{p})^{1/2}} e^{\left( -\sum_{i=1}^n (x_i - \mathbf{m})^2 / 2\mathbf{s}^2 \right)}$$

- En prenant le logarithme naturel de  $L_n$  on obtient:  $\ln L = -n / 2 (1n2\mathbf{p} + 1n\mathbf{s}^2) - \sum_{i=1}^n (x_i - \mathbf{m}) / 2\mathbf{s}^2$
- Les dérivées partielles sont:

$$\frac{\partial \ln L}{\partial \mathbf{m}} = \sum_{i=1}^n (x_i - \mathbf{m}) / \mathbf{s}^2 = 0$$

$$\frac{\partial \ln L}{\partial \mathbf{s}^2} = -n / 2\mathbf{s}^2 - \sum_{i=1}^n (x_i - \mathbf{m})^2 / 2\mathbf{s}^4 = 0$$

## SIMULATION DISCRETE ANALYSE DES RÉSULTATS Méthodes d'Estimation

- En solutionnant pour des estimateurs de  $\mu$  et  $\sigma^2$ , on obtient:

$$\hat{\mathbf{m}} = \frac{\sum_{i=1}^n x_i}{n}$$

$$\hat{\mathbf{s}}^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n}$$

- La méthode des moments donnait:

$$\hat{\mathbf{m}} = \frac{\sum_{i=1}^n x_i}{n}$$

$$\hat{\mathbf{s}}^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1}$$

## SIMULATION DISCRETE ANALYSE DES RÉSULTATS Intervalle de confiance

☞ Définition

$P \{a \leq \text{vraie valeur du paramètre} \leq b\} = 100(1 - \alpha)\%$  où  $0 \leq \alpha \leq 1$ .

☞ Niveau de confiance:  $100(1 - \alpha)\%$ , par exemple pour  $\alpha = 0.05$ , le niveau de confiance est de 95% ou 95%.

☞ Il y a un risque de 5% que la vraie valeur soit en dehors des limites  $(a, b)$ , qu'on appelle l'intervalle de confiance.

## SIMULATION DISCRETE ANALYSE DES RÉSULTATS Intervalle de confiance

☞ **Intervalle de confiance pour  $m$  lorsque  $s$  est connue.**

Échantillon de taille  $n$  d'une population normale de variance  $\sigma^2$  connue et les observations indépendantes, la variable  $\bar{x}$  suit une loi normale  $N(m, s / \sqrt{n})$ .

Si l'hypothèse de normalité n'est pas satisfaite, d'après le théorème central limite, si la taille de l'échantillon est suffisamment grande ( $n \geq 30$ ), alors la variable  $y = \frac{\bar{x} - m}{s / \sqrt{n}}$  est distribuée selon une normale  $N(0,1)$ . On peut donc trouver un intervalle de confiance de niveau  $100(1 - \alpha)\%$  en suivant la procédure suivante:

- 1) calculer la moyenne de l'échantillon (l'estimateur ponctuel),
- 2) calculer l'écart-type  $s_{\bar{x}}$ , où  $s_{\bar{x}} = \frac{s}{\sqrt{n}}$ ,
- 3) choisir un risque  $\alpha$ ; la probabilité associée à l'intervalle de confiance est  $1 - \alpha$ ,
- 4) calculer les limites de confiance. L'intervalle de confiance peut être bilatéral ou unilatéral. Ainsi

$$P(a \leq \bar{x} \leq b) = 1 - \alpha$$

$$P(a \leq m \leq b) = 1 - \alpha$$

$$\text{et } a = \bar{x} - z_{\alpha/2} \frac{s}{\sqrt{n}}, \quad b = \bar{x} + z_{\alpha/2} \frac{s}{\sqrt{n}}$$

## SIMULATION DISCRETE ANALYSE DES RÉSULTATS Intervalle de confiance

☞ Ainsi, pour un risque  $\alpha = 5\%$ ,  $z_{0.05/2} = z_{0.25} = 1.96$  et l'intervalle de confiance est

$$\bar{x} - 1.96 \frac{S}{\sqrt{n}} \leq m \leq \bar{x} + 1.96 \frac{S}{\sqrt{n}}$$

☞ 5) définir l'intervalle de confiance et interpréter.

$$\bar{x} - z_{\alpha/2} \frac{S}{\sqrt{n}} \leq m \leq \bar{x} + z_{\alpha/2} \frac{S}{\sqrt{n}}$$

a une probabilité  $1 - \alpha$  d'inclure le paramètre  $\mu$ .

## SIMULATION DISCRETE ANALYSE DES RÉSULTATS Intervalle de confiance

☞ **Intervalle de confiance pour m lorsque s est inconnue et que la taille de l'échantillon  $n \leq 30$ .**

☐ La procédure à suivre dans cette situation est la même que celle décrite à la section précédente.

☐ L'écart-type de la population inconnu et estimé

☐ Taille de l'échantillon  $n \leq 30$ .

☐ Sous ces conditions, la quantité  $\frac{\bar{x} - m}{s / \sqrt{n}}$  est distribuée

☐ selon la loi *t* de **Student** avec  $(n - 1)$  degré de liberté. Alors les limites de l'intervalle de confiance seront:

$$\bar{x} \pm t_{\alpha/2, n-1} \frac{S}{\sqrt{n}}$$

## SIMULATION DISCRETE ANALYSE DES RÉSULTATS Ergodicité et stationnarité

### ☞ Stationnarité et ergodicité

□ Soit  $\{x_1, x_2, \dots, x_n\}$  une suite de variables aléatoires représentant  $n$  observations sur l'un des paramètres de simulation. Posons

$$\mu_i = E(x_i)$$

$$R_{ij} = E[(x_i - \mathbf{m})(x_j - \mathbf{m})],$$

$$R_{ii} = E[(x_i - \mathbf{m})^2] = \mathbf{s}_i^2$$

où  $\mu_i$  est la moyenne de  $x_i$ ,

$R_{ij}$  la covariance entre  $x_i$  et  $x_j$ ,

et  $R_{ii}$  la variance de  $x_i$ .

□ On dira que les variables aléatoires  $\{x_1, x_2, \dots, x_n\}$  forment un *processus stationnaire du deuxième ordre*, si

i)  $\mu_1 = \mu_2 = \dots = \mu_n = \mu$

ii)  $R_{ij} = R_h$ , pour tout  $i$  et  $j$  tels que  $|i - j| = h$ .

## SIMULATION DISCRETE ANALYSE DES RÉSULTATS Ergodicité et stationnarité

☞ **Théorème:** Si les variables aléatoires  $x_1, x_2, \dots, x_n$  forme un processus stationnaire de deuxième ordre, alors:

$$E(\bar{x}_n) = \mathbf{m}$$

$$Var(\bar{x}_n) = \frac{1}{n} \left( \mathbf{s}^2 + 2 \sum_{i=1}^{n-1} \left(1 - \frac{i}{n}\right) R_i \right)$$

☞ **Théorème:** Si est un  $x_1, x_2, \dots, x_n$  est un processus stationnaire, et

$$S^2 = \frac{1}{n-1} \sum_{i=1}^{n-1} (x_i - \bar{x}_n)^2$$

alors:

$$E(S^2) = \mathbf{s}^2 - \frac{2}{n-1} \sum_{i=1}^{n-1} \left(1 - \frac{i}{n}\right) R_i$$

Si les variables  $x_1, x_2, \dots, x_n$  sont indépendantes, alors  $R_i = 0$  pour  $i > 0$  et

$$Var(\bar{x}_n) = \frac{R_0}{n}$$

☞ Par conséquent, si les variables aléatoires  $x_1, x_2, \dots, x_n$  sont positivement corrélées, alors  $E(S^2) < \mathbf{s}^2$

## SIMULATION DISCRETE ANALYSE DES RÉSULTATS Ergodicité et stationnarité

De plus, d'après l'équation:

$$Var(\bar{x}_n) = \frac{1}{n} \left( s^2 + 2 \sum_{i=1}^{n-1} \left(1 - \frac{i}{n}\right) R_i \right) > \frac{s^2}{n}$$

- Nous avons donc la double inégalité:

$$Var(\bar{x}_n) \geq \frac{s^2}{n} > \frac{E(S^2)}{n}$$

- En estimant  $Var(\bar{x}_n)$  par  $S^2/n$ , on néglige le biais de  $S^2$  comme estimateur de  $\sigma^2$  et d'autre part l'effet de la covariance des  $x_i$  sur la variance de  $\bar{x}_n$ .

- Supposons que  $\lim_{i \rightarrow \infty} R_i = 0$

- Si  $R_i$  converge rapidement vers zéro, on peut montrer

$$\lim_{n \rightarrow \infty} nVar(\bar{x}_n) = V < \infty$$

$$V = \sum_{i=-\infty}^{\infty} R_i$$

- Si cette dernière condition est satisfaite, on dit que les variables  $\{x_i\}$  sont ergodiques pour la moyenne. C'est cette propriété qui permet d'obtenir un estimé par intervalle consistant de la moyenne  $\mu$  à partir de données dépendantes.

## SIMULATION DISCRETE ANALYSE DES RÉSULTATS systèmes se terminant

- Dans les systèmes discrets, on peut les classer en deux catégories, les systèmes se terminant et les systèmes ne se terminant pas. Dans un système se terminant, les événements engendrant la dynamique du système cessent de se produire à un moment dans le temps; tandis que ceci ne se produit pas dans les systèmes ne se terminant pas. On appelle une époque ou sous-intervalle (lot), la période où les événements se produisent, et point de régénération le moment où les événements cessent. Chaque époque recommence avec un système à vide.

- Exemple:

- **Systèmes se terminant**

- \* Banque: ouverte 9:00 à 17:00
- \* Ordinateur central:
- \* Inventaire de pièces de rechange:
- \* Système de base de données:

- **Systèmes ne se terminant pas**

- \* Atelier:
- \* Système d'inventaire
- \* Hôpital

- Il faudra faire la distinction lors de la cueillette des données entre les variables obtenues en régime transitoire, et les variables en régime permanent.

## SIMULATION DISCRÈTE ANALYSE DES RÉSULTATS systèmes se terminant

### ☞ Analyse des résultats pour les systèmes se terminant

- Les paramètres les plus usuels que l'on utilise sont la moyenne et l'écart-type. Par exemple, pour  $x_1, x_2, \dots, x_n$  une suite de variables indépendantes et de même loi (iid), les paramètres mesurés sont:

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

$$s = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1}}$$

- On utilisera la méthode des moyennes de lots pour évaluer les paramètres des systèmes se terminant.

### □ Détermination du nombre de lots

Étant donné la nature fluctuante dans le temps des paramètres de la simulation, on cherchera un estimateur par intervalle. Cependant, pour combien de lots  $K$  devra-t-on simuler pour obtenir un intervalle de confiance assez étroit? Rappelons que la demi-largeur d'un intervalle de confiance est:

$$I = t_{\alpha/2, K-1} s / \sqrt{K}$$

## SIMULATION DISCRÈTE ANALYSE DES RÉSULTATS systèmes se terminant

- **Variance connue**  
On peut fixer la largeur de l'intervalle de confiance; et pour un niveau de confiance  $100(1 - \alpha)\%$  donné, déterminer le nombre  $K$  de lots.
- Par exemple, dans un système M/M/1, pour un niveau de confiance de 95%, une demi-largeur de l'intervalle de confiance de 1.0, et un écart-type de 8.31, et  $I = t_{\alpha/2, K-1} s / \sqrt{K}$  on a alors:

$$K = (1.96 \sigma / I)^2$$

$$= 1.96(8.31/1.0)^2$$

$$= 256$$

## SIMULATION DISCRETE ANALYSE DES RÉSULTATS systèmes se terminant

□ **Variance inconnue**

Toutefois, en pratique la variance est habituellement inconnue. La variance est une fonction du nombre de lots  $K$  et on cherche à déterminer  $K$ .

□ **Théorème:** Lorsque le but de la simulation est d'estimer une moyenne, quand  $\sigma^2$  est inconnue, les étapes 1 à 7 permettent d'obtenir un estimé de  $\mu$ , avec un niveau de confiance  $100(1 - \alpha)\%$ .

1) Prendre  $k_0$  un entier positif  $\geq 2$ . Soit  $w = \frac{t_{\alpha/2, k_0 - 1}}{I}$ .

2) Observer  $x_1, x_2, \dots, x_{k_0}$ .

3) Calculer 
$$\bar{x}_{k_0} = \frac{\sum_{i=1}^{k_0} x_i}{k_0}$$

$$s^2 = \frac{\sum_{i=1}^{k_0} (x_i - \bar{x}_{k_0})^2}{k_0 - 1}$$

$$K = \max\{k_0 + 1, \lceil (ws)^2 \rceil\}$$

4) Prendre où  $\lceil x \rceil$  dénote le plus grand entier qui ne dépasse pas  $x$ .

5) Observer..  $x_{k_0+1}, \dots, x_K$

6) Calculer 
$$\bar{x}_K = \frac{\sum_{i=1}^K x_i}{K}$$

7) Calculer l'intervalle de confiance de niveau

$$\bar{x} - I < \mu < \bar{x} + I$$

## SIMULATION DISCRETE ANALYSE DES RÉSULTATS systèmes se terminant

□ **Exemple:**

Supposons une caserne de pompiers qui doit desservir plusieurs municipalités. De par l'expérience, on sait qu'il y a en moyenne 4 incendies par 24 heures selon un processus de Poisson. On sait aussi que 3/4 des incendies nécessite une alerte, et que 1/4 en nécessite deux. La durée des incendies pour:

1 alerte (unité)            N(3 heures, 0.5 heures)

2 alertes (unités)        N(4 heures, 1.0 heures)

On veut déterminer le total minutes par jour où les pompiers (1440 x 2 x 60) sont occupés avec un précision de 60 minutes.

**Solution:**

On veut estimer  $\mu$  avec  $I = 60$  pour  $\alpha=10\%$ . Supposons que l'on choisit  $k_0 = 10$  et on simule pour 10 jours. Les résultats du programme sont:

|         |      |     |     |     |      |      |      |      |     |     |
|---------|------|-----|-----|-----|------|------|------|------|-----|-----|
| Jours   | 1    | 2   | 3   | 4   | 5    | 6    | 7    | 8    | 9   | 10  |
| Moyenne | 1505 | 804 | 864 | 869 | 1314 | 1599 | 1215 | 1232 | 885 | 603 |

On calcule  $\bar{x}_{k_0} = 969, S = 425.07$

3). Puisque  $t_{5\%, 9} = 1.83$ , on a:

$$K = \max\{11, \lceil (1.83)^2 (425.07)^2 / (60)^2 \rceil\}$$

$$= \max\{11, \lceil 168.08 \rceil\}$$

$$= \max\{11, 169\}$$

On simule pour 159 jours de plus et on obtient pour un niveau de confiance de 90%, on a un intervalle de confiance pour entre les bornes

SIMULATION DISCRETE  
ANALYSE DES RÉSULTATS  
système se terminant

```

SIMULATE
RMULT      12345,12345,12345
EXPON      FUNCTION      RN3,C24 Exponential distribution
0,0/. 1., 104/.2.,222/.3.,355/.4.,509/.5.,69/.6.,915
.7,1.2/.75,1.38/.8,1.6/.84,1.83/.88,2.12/.9,2.3
.92,2.52/.94,2.81/.95,2.99/.96,3.2/.97,3.5/.98,3.9
.99,4.6/.995,5.3/.998,6.2/.999,7/.9998,8

SNORM      FUNCTION      RN2,C23 Standard normal distribution
0,-4/.00138,-3/.00621,-2.5/.02275,-2/.06681,-1.5
11507,-1.2/.15866,-1/.21186,-.8/.27425,-.6
.34458,-.4/.42074,-.2/.5,0/.57926,.2/.65542,.4
.72575,.6/.78814,.8/.84134,1/.88493,1.2/.93319,1.5
.97725,2/.99379,2.5/.99862,3/1,4

TIME1      FVARIABLE      30*FN$SNORM+180 Duration of 1-unit fires
TIME2      FVARIABLE      60*FN$SNORM+240 Duration of 2-unit fires

UNIT       STORAGE       10      10 units should suffice.

           GENERATE      360,FN$EXPON Create fires.
           TRANSFER      .25,ONE,TWO Branch to 1 - or 2-unit fires.

ONE        ENTER         UNIT,1  Get a unit.
           ASSIGN        1,V$TIME1 P1= Time to put out fire.
    
```

SIMULATION DISCRETE  
ANALYSE DES RÉSULTATS  
système se terminant

```

SAVEVALUE  DAYTOT+,P1    Accumulate times for the day.
ADVANCE    P1            Service the fire.
LEAVE      UNIT,1       Return the unit.
TERMINATE

TWO        ENTER        UNIT,2    Get 2 units.
           ASSIGN        1,V$TIME2 P1= Time to put out fire.
SAVEVALUE  DAYTOT+,P1    Accumulate times for the day,
SAVEVALUE  DAYTOT+,P1    - for each unit.
ADVANCE    P1            Service the fire.
LEAVE      UNIT,2       Return both units.
TERMINATE

GENERATE   1440          At the end of each day,
SAVEVALUE  TOTAL+,X$DAYTOT accumulate times,
PRINT      ,,X          obtain output, and
SAVEVALUE  DAYTOT,0     zero total for day.
TERMINATE  1

START      10,NP
END
    
```

SIMULATION DISCRETE  
ANALYSE DES RÉSULTATS  
système se terminant

PRINT BLOCK 18 AT CLOCK TIME 1440  
NON-ZERO FULLWORD SAVEVALUES  
SAVEX VALUE SAVEX VALUE  
DAYTOT1505 TOTAL 1505

PRINT BLOCK 18 AT CLOCK TIME 2880  
NON-ZERO FULLWORD SAVEVALUES  
SAVEX VALUE SAVEX VALUE  
DAYTOT804 TOTAL 2309

PRINT BLOCK 18 AT CLOCK TIME 14400  
NON-ZERO FULLWORD SAVEVALUES  
SAVEX VALUE SAVEX VALUE  
DAYTOT603 TOTAL 9890  
50 END

SIMULATION DISCRÈTE  
ANALYSE DES RÉSULTATS  
SYSTÈMES SE TERMINANT

Observations:

- 1) Une durée de 10 jours est insuffisante.
- 2) Si on croit qu'une estimation d'une journée suffit, on prendrait comme estimateur de  $\bar{x} = 1505$ , ce qui est très loin de la vraie utilisation des équipements par jour.
- 3) Si on augmente le niveau de confiance, ceci requiert d'augmenter l'échantillonnage. Pour un niveau de confiance de 95%,  $t_{25, 9} = 2.262$  et

$$K = \max\left\{11, \left[ (2.262)^2 (425.07)^2 / (60)^2 \right] \right\}$$

$$= \max\{11, 257\} = 257 \text{ jours de simulation.}$$

- 4) Si on peut se permettre seulement 100 jours de simulation et que l'on désire conserver un niveau de confiance de 90%, alors

$$100 = \max\left\{11, \left[ (1.83)^2 (425.07)^2 / I^2 \right] \right\}$$

$$100 = \frac{(1.83)^2 (425.07)^2}{I^2}$$

$$\Rightarrow I = 77.79$$

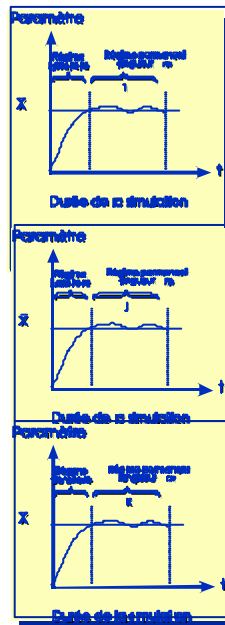
- 5) Pour ce problème, on peut trouver une solution théorique pour comparer aux résultats de la simulation. Il est clair que la demande moyenne journalière est:  $(3 \text{ incendies}) \times (180 \text{ min.}) + (1 \text{ incendie}) \times (2 \text{ alertes}) \times (240 \text{ min.}) = 1020 \text{ min.}$  qui se situe à l'intérieur de notre intervalle de confiance de niveau 90% pour  $\mu$ .
- 6) Pour un niveau de confiance de 90%, nos énoncés sont corrects 9 sur 10. Pour 95%, 19 sur 20. Pour un niveau de 99%, on fait une erreur 1 fois sur 100, et dans ce cas,  $t_{0,005, 9} = 3.25$ , ce qui entraîne  $K = \max\{11, 531\} = 531$  jours de simulation.

SIMULATION DISCRÈTE  
ANALYSE DES RÉSULTATS  
SYSTÈMES NE SE TERMINANT PAS

MÉTHODE DES MOYENNES  
DES LOTS

Répétitions indépendantes

- Chaque observation est obtenue d'une passe de simulation.
- Chaque passe est réalisée à partir des mêmes conditions initiales, mais avec des suites différentes de nombres aléatoires indépendants.
- La valeur du paramètre à estimer sera la moyenne des observations des  $K$  passes de simulation.
  - L'avantage principal: observations indépendantes, et on peut calculer un intervalle de confiance pour évaluer la précision de la valeur estimée.
  - L'inconvénient majeur est que chaque observation risque d'être biaisée par les condition initiales

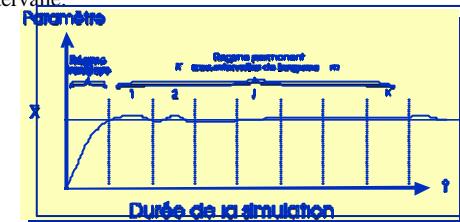


SIMULATION DISCRÈTE  
ANALYSE DES RÉSULTATS  
SYSTÈMES NE SE TERMINANT PAS

Division en sous-intervalles

But: réduire les effets du régime transitoire associés à la méthode des répétitions indépendantes.

Diviser une passe de simulation en  $K$  sous-intervalles.  
Sous-intervalle :  $m$  observations, et la valeur estimée du paramètre  $x_i$  = moyenne des  $K$  valeurs obtenues dans le  $i$ -ième sous-intervalle.



- Remarque : Covariance entre les résultats obtenus à la fin d'un sous-intervalle et au début de l'intervalle suivant n'est pas nulle. Cependant si les sous-intervalles sont suffisamment longs ( $m$  suffisamment grand), l'hypothèse des observations indépendantes est approximativement satisfaite.
- Avantage: seule passe de simulation, et les effets du régime transitoire disparaissent avec le temps, et les observations obtenues sont donc plus représentatives du système réel.
- Désavantages: la détermination de la longueur  $m$  des sous-intervalles et le nombre de sous-intervalles peut être complexe.

SIMULATION DISCRÈTE  
ANALYSE DES RÉSULTATS  
SYSTÈMES NE SE TERMINANT PAS

- Soit  $x_{ij}$  : ième observation du lot  $j$ .  
Avec  $m$  observations de chacune des  $K$  répétitions, on peut estimer la moyenne  $\mu$  par

$$\bar{x}_n = \frac{1}{K} \sum_{j=1}^K \bar{x}_{m,j}$$

où

$$\bar{x}_{m,j} = \frac{1}{m} \sum_{i=1}^m x_{ij}$$

- Un estimateur non-biaisé de la variance de  $\bar{x}_{m,j}$  est

$$S_K^2(\bar{x}_{m,j}) = \frac{1}{(K-1)} \sum_{j=1}^K (\bar{x}_{m,j} - \bar{x}_n)^2$$

On peut utiliser  $S_K^2(\bar{x}_{m,j})/K$  comme un estimateur non-biaisé de  $Var(\bar{x}_n)$ .  
Pour trouver un intervalle de confiance pour la moyenne  $\mu$ , on approxime la distribution de la variable  $T = (\bar{x}_n - \mu) / \sqrt{S_K^2 / K}$  par une distribution  $t$  de Student avec  $K-1$  degrés de liberté.  
Avec la méthode des répétitions indépendantes, on devrait s'assurer de la normalité des variables  $\bar{x}_{m,1}, \bar{x}_{m,2}, \bar{x}_{m,3}, \dots, \bar{x}_{m,K}$ , en appliquant un test d'hypothèse approprié, par exemple, le test de Shapiro-Wilk.

SIMULATION DISCRÈTE  
ANALYSE DES RÉSULTATS  
SYSTÈMES NE SE TERMINANT PAS

Détermination de la longueur d'un lot

En changeant de notation, on calcule la moyenne des lots

$$y_{i,m} = \frac{1}{m} \sum_{j=1}^m x_{m(i-1)+j} \quad i = 1, \dots, K = \langle n/m \rangle$$

pour un  $m$  donné. Soit  $R_{s,m} = Cov(y_{i,m}, y_{i+s,m})$  la covariance entre les variables  $y_{i,m}$  et  $y_{i+s,m}$ . Si les moyennes des échantillons sont indépendantes, alors le coefficient de corrélation entre les moyennes des lots voisins est

$$r_{1,m} \approx 0$$

$$r_{1,m} \equiv \frac{R_{1,m}}{R_{0,m}}$$

$$\text{Si } \bar{y}_k = \frac{1}{k} \sum_{i=1}^k y_{i,m}$$

$$\text{alors } Var(\bar{y}_k) = \frac{R_{0,m}}{k}$$

Mais  $\bar{y}_k \approx \bar{x}_k$ . Si on estime  $R_{s,m}$  par

$$\hat{R}_{s,m} = \frac{1}{k-s} \sum_{i=1}^{k-s} (y_{i,m} - \bar{y}_k)(y_{i+s,m} - \bar{y}_k) \quad 0 \leq s \leq k$$

on peut approximer la distribution de  $(\bar{x} - \mu) / \sqrt{R_{0,m} / k}$  par une distribution  $t$  de Student avec  $k-1$  degrés de liberté.

SIMULATION DISCRÈTE  
ANALYSE DES RÉSULTATS  
SYSTÈMES NE SE TERMINANT PAS

☞ Problème: déterminer un  $m^*$  tel que les variables  $y_{1,m^*}, \dots, y_{k^*,m^*}$  pour  $k^* = \lfloor n/m^* \rfloor$  sont indépendantes.

○ Plusieurs procédures: un test d'hypothèse sur le coefficient de corrélation.

Un estimé de  $\rho_{1,m}$  est:

$$\hat{r}_{1,m} = \frac{\sum_{i=1}^{k-1} (y_{i,m} - \bar{y})(y_{i+1,m} - \bar{y})}{\sum_{i=1}^k (y_{i,m} - \bar{y}_k)^2}$$

Le test statistique

$$C_k = 1 - \frac{\sum_{i=1}^{k-1} (y_{i,m} - y_{i+1,m})^2}{2 \sum_{i=1}^k (y_{i,m} - \bar{y}_k)^2}$$

peut être écrit comme

$$C_k = \hat{r}_{1,m} + \frac{(y_{1,m} - \bar{y}_k)^2 (y_{k,m} - \bar{y}_k)^2}{2(k-1)\hat{R}_{0,m}}$$

lorsque  $k$  est grand,  $C_k \approx \hat{r}_{1,m}$

SIMULATION DISCRÈTE  
ANALYSE DES RÉSULTATS  
SYSTÈMES NE SE TERMINANT PAS

☞ Hypothèse

$H_0$  :  $y_{1,m}, \dots, y_{k,m}$  sont indépendantes

Si les  $y_{1,m}, \dots, y_{k,m}$  : distribuées selon une normale, alors  $C_k$  a une moyenne de zéro  $E(C_k) = 0$ , une variance

$Var(C_k) = (k-2)/(k^2-1)$ , et tend vers une normale pour  $k$  aussi petit que 8.

Si les  $y_{1,m}, \dots, y_{k,m}$  ne sont pas distribuées selon une normale, alors  $C_k$  a une moyenne de zéro. Si  $k$  augmente, la variance tend vers  $Var(C_k) = (k-2)/(k^2-1)$ , et les moments d'ordre 3, le coefficient d'asymétrie et 4, le coefficient d'aplatissement convergent vers zéro.

Approximer  $C_k / \sqrt{(k-2)/(k^2-1)}$  par une normale  $N(0,1)$  et si les  $y_{1,m}, \dots, y_{k,m}$  sont distribuées selon une normale, et l'utiliser pour  $k \geq 8$ .

Si les  $\{x_i\}$  possède une fonction d'autocovariance monotone croissante, on utilise une test unilatéral de grandeur  $\beta$ . On accepte  $H_0$  si  $C_k \leq c_{2b} \sqrt{(k-2)/(k^2-1)}$  où  $c_{2b}$  est:

$$\frac{1}{\sqrt{2b}} \int_{-\infty}^{c_{2b}} e^{-x^2/2} dx = 1 - 2b$$

Si comme dans les systèmes de files d'attente, les  $\{x_i\}$  ont une fonction d'autocovariance harmonique amortie, alors on utilise un test bilatéral. Par exemple, si  $|C_k| \leq c_b \sqrt{(k-2)/(k^2-1)}$ , on accepte  $H_0$ .

SIMULATION DISCRÈTE  
ANALYSE DES RÉSULTATS  
SYSTÈMES NE SE TERMINANT PAS

Voici un algorithme qui implante un test bilatéral sur des lots dont la taille croît jusqu'à ce que  $H_0$  soit accepté.

1.  $m_0 \leftarrow -1$  et  $j \leftarrow 0$
2.  $m_{j+1} \leftarrow m_j^* 2$ ,  $k_j \leftarrow \lfloor n/m_j \rfloor$
3. Si  $k_j < 8$ , alors pas de solution.
4. Calculer les moyennes des lots  $y_{1,m}, \dots, y_{k_j,m}$
5. Calculer  $C_{k_j}$
6. Si  $C_{k_j} > c_{2b} \sqrt{(k-2)/(k^2-1)}$ ,  $j \leftarrow j+1$  ,et aller à 2,
7.  $m^* \leftarrow m_j$  et  $k^* \leftarrow k_j$ ,
8. Calculer  $\hat{R}_{0,m^*}$ ,
9. Retourner un intervalle de confiance  

$$\bar{x}_n \pm t_{\alpha/2, k^*} \sqrt{\hat{R}_{0,m^*} / k^*}$$
 où  $t_{\alpha/2, k^* - 1}$  est le  $1 - \alpha$  point de la distribution  $t$  de Student avec  $k^* - 1$  degrés de liberté

SIMULATION DISCRETE  
ANALYSE DES RÉSULTATS  
MÉTHODE MOYENNE DES LOTS

```

ROUTINE Accounting GIVEN Wait
" Ce sous-programme est appele a toutes les fois qu'un appel
" est complété
" Wait = temps d'attente d'un appel complété

DEFINE Wait AS a DOUBLE VARIABLE

DEFINE Epoch, i, N.Epoch,
    Num, Rep, Trun AS SAVED INTEGER VARIABLE
DEFINE X AS a REAL SAVED 1-DIM ARRAY
DEFINE j AS an INTEGER VARIABLE

IF N.Epoch = 0, " Au début
    READ N.Epoch, Trun, i, Rep
    " N.Epoch = nombre désiré d'appel complété
    " Trun = nombre d'appel a éliminer au début
    " i = nombre désiré de la première répétition
    " Rep = nombre de répétitions
    epoch = 1
    RESERVE X(*) AS N. Epoch
    Num = 0
    ADD i-1 TO Rep
    RETURN
ENDIF
ADD 1 TO Num

IF Num <= Trun, " Doit-t-on tronquer les observation
    RETURN
ENDIF
    
```

**SIMULATION DISCRETE  
ANALYSE DES RÉSULTATS  
MÉTHODE MOYENNE DES LOTS**

```
X(Epoch) = .EnMinute * Wait " convertir en minutes
IF Epoch = N.Epoch,
  CALL Title ( N.Epoch, U.Reservationist(1) + N.X.Reservationist(1) , i,
Trun)
  CALL Methode.Moyenne.Lots  GIVING N.Epoch, X(*)
  CALL Methode.Autoregressive  GIVING N.Epoch, X(*)
  CALL Methode.Spectrale  GIVING N.Epoch, X(*)
  IF i = Rep  "Est-on rendu au nombre de répétitions limites
    STOP
  ENDIF
  FOR j=1 TO Dim.F(S(*)), S(j) = SEED.V(j)
  epoch = 1
  Num = 0
  ADD 1 TO i
  ELSE
  ADD 1 TO Epoch
  ENDIF
  RETURN
END " Accounting
```

**SIMULATION DISCRETE  
ANALYSE DES RÉSULTATS  
MÉTHODE MOYENNE DES LOTS**

```
ROUTINE Methode.Moyenne.Lots GIVEN N, X
" Calcule un intervalle de confiance a 0.95
" par la méthode des moyennes de lots

DEFINE i, k, N, No, Ntilde      AS INTEGER VARIABLES
DEFINE C, Valeur.Critique,
  D, E, demi.Largeur, Na, Xbar  AS REAL VARIABLES
DEFINE X, y AS 1-DIM REAL ARRAYS

IF N < 8,
  PRINT 1 LINE AS FOLLOWS
  Dimension de l'échantillon trop petite pour le test d'indépendance
  RETURN
ENDIF

RESERVE Y(*) AS N
Ntilde = N
K = 1
for i = 1 to N, Y(i) = X(i)
for i = 1 to N, Compute Xbar as the mean of X(i)
```

**SIMULATION DISCRETE  
ANALYSE DES RÉSULTATS  
MÉTHODE MOYENNE DES LOTS**

```

WHILE (.Vrai )
  DO
  FOR i = 1 TO Ntilde,
  DO
  COMPUTE D AS THE VARIANCE OF Y(i)
  IF i < Ntilde,
  COMPUTE E AS THE SUM OF (Y(i) - Y(i+1)) * (Y(i) - Y(i+1))
  ENDIF
  LOOP

```

**SIMULATION DISCRETE  
ANALYSE DES RÉSULTATS  
MÉTHODE MOYENNE DES LOTS**

```

D = D * Ntilde
C = 1.0 - E / (2.0 * D)
Na = Ntilde * (Ntilde - 1)
Valeur.Critique = 1.645 * SQRT.F (( Ntilde - 2.0) / (Ntilde * Ntilde - 1.0))
demi.Largeur = Student.F ( REAL.F(Ntilde-1)) * SQRT.F ( D / Na )
PRINT 1 LINE WITH Ntilde, K, D / Na,
      Xbar-demi.Largeur, Xbar + demi.Largeur, C,
      Valeur.Critique
      THUS
***** ***** ***** ***** ***** * ***** * ****
SKIP 1 LINE
IF ABS.F (C) <= Valeur.Critique And No = 0,
  No = Ntilde
ENDIF

IF Ntilde < 16,
  SKIP 1 LINE
  IF No > 0,
    PRINT 1 LINE WITH No THUS
  ***** lots suffisent pour passer le test d'indépendance a 0.05
  ELSE
    PRINT 1 LINE THUS
  Rejet du test d'indépendance
  ENDIF
  RELEASE Y
  RETURN
ENDIF

Ntilde = TRUNC.F (Ntilde / 2.0)
K = 2 * K
FOR i = 1 TO Ntilde,
  Y(i) = (Y(2*i-1) + Y(2*i)) / 2.0
LOOP
END " REPORT

```

**SIMULATION DISCRETE  
ANALYSE DES RÉSULTATS  
MÉTHODE MOYENNE DES LOTS**

```

FUNCTION Studentf (Df)
  DEFINE Df AS A REAL VARIABLE
  DEFINE A AS A REAL VARIABLE
  DEFINE SS AS A SAVED INTEGER VARIABLE
  DEFINE i AS AN INTEGER VARIABLE
  DEFINE B, H AS SAVED, REAL 1-DIM ARRAYS

  IF SS = 0 ,
    RESERVE B(*) AS 9,
    H(*) AS 4
    LET SS = 1
    LET B(1) = 1.96
    FOR i = 2 TO 9,
      DO
        LET B(i) = B(i-1) * B(1)
      LOOP

    H(1) = ( B(3) + B(1) ) / 4.0
    H(2) = ( 5.0 * B(5) + 16.0 * B(3) + 3.0 * B(1) ) / 96.0
    H(3) = ( 3.0 * B(7) + 19.0 * B(5) + 17.0 * B(3) - 15.0 * B(1) ) / 384.0
    H(4) = ( 79.0 * B(9) + 776.0 * B(7) + 1482.0 * B(5) - 1920.0 * B(3) -
    945.0 * B(1) ) / 92160.0
  ENDIF

  FOR i BACK FROM 4 TO 1 BY 1
    DO
      LET A = ( A + H(i) ) / Df
    LOOP

  RETURN WITH A + B(1)
END " Student.f
    
```

**SIMULATION DISCRETE  
ANALYSE DES RÉSULTATS  
MÉTHODE MOYENNE DES LOTS**

- ☞ Pour réduire la demi-largeur d'un intervalle de confiance, on doit réduire la variance des observations ou augmenter la taille de l'échantillon.
- ☞ La méthode des sous-intervalles séquentiels tente de déterminer le plus petit sous-intervalle tel que les sous-intervalles demeurent indépendants.
- ☞ Une mesure de la dépendance de sous-intervalles adjacents est le coefficient de corrélation:

$$\hat{r} = \frac{\sum_{j=1}^{k-1} (\bar{x}_{m,j} - \bar{x}_n)(\bar{x}_{m,j+1} - \bar{x}_n)}{\sum_{j=1}^k (\bar{x}_{m,j} - \bar{x}_n)^2}$$

où  $\bar{x}_n$  représente la moyenne des moyennes des  $K$  sous-intervalles et  $\bar{x}_{m,j}$  la moyenne du  $j$ -ième lot de taille  $m$ , pour  $j = 1, \dots, K$ .

- ☞ Pour réduire le biais de l'estimateur de  $\rho$ , on suggère d'utiliser l'estimateur:

$$\hat{r}' = 2\hat{r}_1 - (\hat{r}_1 + \hat{r}_2) / 2$$

où  $\hat{r}_1$  est calculé selon l'équation précédente en utilisant les  $K/2$  premiers sous-intervalles et  $\hat{r}_2$  en utilisant les  $K/2$  derniers sous-intervalles.

## SIMULATION DISCRETE ANALYSE DES RÉSULTATS MÉTHODE MOYENNE DES LOTS

☞ Une implantation de cet algorithme est la suivante:

- 1) on simule le système pour  $K$  lots de  $m$  observations chacun,
- 2) on calcule  $\hat{r}$  et on teste si les lots peuvent être considérés indépendants,
- 3) si les lots sont indépendants, calculer un intervalle de confiance, et si la précision désirée est atteinte, on arrête. Sinon, on augmente la taille des lots, par exemple  $m \rightarrow 2*m$  et on retourne à l'étape 1).
- 4) si les lots sont dépendants, on augmente la taille des lots  $m$ , et on retourne à l'étape 1).

## SIMULATION DISCRETE ANALYSE DES RÉSULTATS MÉTHODE MOYENNE DES LOTS

☞ Law et Carson, les développeurs de cette méthode:

proposé une procédure où le nombre de sous-intervalles  $K$  demeure constant lorsque la taille de chaque lot augmente.

La taille des lots est  $m_i/K$  où  $m_i$  est le nombre d'observations total à la  $i$ -ième itération.

L'intervalle de confiance est établi en combinant les  $m_i/K$  lots en  $n$  lots plus grands.

## SIMULATION DISCRETE ANALYSE DES RÉSULTATS MÉTHODE MOYENNE DES LOTS

☞ Law et Carson:

- 1) on initialise  $i = 1, n = 40, m_0 = 600, m_1 = 800, K = 400, u = .04$ ; il faut définir  $nf=K$  où  $n=40, f = 10$ .
- 2) on simule pour  $m_i$  observations. On divise les  $m_i$  observations en  $K$  lots de taille  $k$ . On calcule  $\hat{r}'$
- 3) si,  $\hat{r}' > u$ , les lots sont indépendants et on va à l'étape 6). Sinon si  $\hat{r}' < 0$ , on va à l'étape 5) pour tenter de définir un intervalle de confiance. Dans le cas contraire, on va à l'étape 4).
- 4) on divise les  $m_i$  observations en  $K/2$  lots de taille  $2k$ . On recalcule  $\hat{r}'$ . Si la nouvelle valeur de  $\hat{r}'$  est inférieure à la précédente (étape 2), on va à l'étape 5), sinon à l'étape 6).

## SIMULATION DISCRÈTE ANALYSE DES RÉSULTATS MÉTHODE MOYENNE DES LOTS

- 5) on divise les  $m_i$  observations en  $n$  lots de taille  $m_i/n$ . On calcule

$$S_x^2 = \frac{1}{n-1} \sum_{i=1}^n (\bar{x}_{m,i} - \bar{x}_n)^2$$

et la demi-largeur de l'intervalle de confiance est

$$t_{\alpha, n-1} S_x / \sqrt{n}$$

Si l'intervalle de confiance est acceptable, on arrête, sinon, on va à l'étape 6).

- 6) Augmenter  $i$ ;  $i \rightarrow i+1, m_i = 2m_{i-2}$  et retourner à l'étape 2).

## SIMULATION DISCRETE ANALYSE DES RÉSULTATS MÉTHODE AUTORÉGRESSIVE

☞ Méthode des moyennes des lots:

- l'utilisation de l'approximation
- l'estimation de l'approximation.

☞ Si le nombre d'observations par lots augmente, les corrélations entre les paires d'échantillons diminuent.

Une fois  $n$  déterminé, on estime  $Var(\bar{x})$ .

Méthode: basée sur le fait que l'on prend une suite d'observations dépendantes  $x_p, \dots, x_n$  et on les transforme en une suite d'observations indépendantes.

Supposons qu'une suite non-stochastique  $b_0, \dots, b_p$  existe telle que

$$y_i = \sum_{s=0}^p b_s (x_{i-s} - m) \quad i = p+1, \dots, n$$

est une suite de variables aléatoires indépendantes avec moyenne zéro  $E(y_i) = 0$  et variance  $\sigma^2$ .

## SIMULATION DISCRETE ANALYSE DES RÉSULTATS MÉTHODE AUTORÉGRESSIVE

☞ Dans la littérature des séries temporelles, la dernière expression : la représentation autorégressive de  $\{x_i\}$ , et  $p$ , l'ordre autorégressif.

Il est à noter que  $\bar{y}_{n-p} = \frac{1}{n-p} \sum_{i=p+1}^n y_i$

possède une variance:  $Var(\bar{y}_{n-p}) = \sigma^2 / (n-p)$ .

Pour voir la relation avec  $Var(\bar{x}_n)$ , on réécrit l'expression précédente:

$$\begin{aligned} \bar{y}_{n-p} &= \frac{1}{n-p} \sum_{i=p+1}^n \left[ \sum_{s=0}^p b_s (x_{i-s} - m) \right] \\ &= \frac{nb}{n-p} (\bar{x}_n - m) - \frac{1}{n-p} \left[ b_0 \sum_{i=1}^n x_i + \sum_{s=1}^{p-1} b_s \left( \sum_{i=n-s+1}^n x_i + \sum_{i=1}^{p-s} x_i \right) \right. \\ &\quad \left. + b_p \sum_{i=p+1}^n x_i - pbm \right] \end{aligned}$$

$$b = \sum_{s=0}^p b_s, \quad p \geq 2$$

alors que  $\lim_{n \rightarrow \infty} [\bar{y}_{n-p} - b(\bar{x}_n - m)]$

Il en découle que  $Var(\bar{x}_n) \approx \frac{Var(\bar{y}_{n-p})}{b^2} \approx \frac{\sigma^2}{nb^2}$

## SIMULATION DISCRETE ANALYSE DES RÉSULTATS MÉTHODE AUTORÉGRESSIVE

- ☞ L'intérêt: trouver une approximation de  $Var(\bar{x}_n)$  par une fonction de  $p$ ,  $b_0, \dots, b_p$  et  $\sigma^2$  au lieu d'une moyenne d'autocovariances.
- ☞ Les valeurs  $b_0, b_1, \dots, b_p$  : obtenues par la solution du système d'équations

$$\sum_{s=1}^p b_s R_{r-s} = -R_r \quad r=1, \dots, p$$

en fixant  $b_0 \equiv 1$ .

Les valeurs des autocovariances sont estimées par

$$\hat{R}_s = \frac{1}{n-s} \sum_{i=1}^{n-s} [(x_i - \bar{x}_n)(x_{i+s} - \bar{x}_n)] \quad s=0, \dots, p$$

ce qui permet d'obtenir des estimés pour  $\hat{b}_0, \hat{b}_1, \dots, \hat{b}_p$  qui sont les estimés des coefficients autorégressifs des moindres carrés.

- ☞ Un estimé de la variance  $\sigma^2$  est

$$\hat{S}_i^2 = \sum_{s=0}^i \hat{b}_{i,s} \hat{R}_s \quad \hat{b}_{i,0} \equiv 1$$

$$y_i = \sum_{s=0}^p b_s (x_{i-s} - m) \quad i = p+1, \dots, n$$

## SIMULATION DISCRETÈ ANALYSE DES RÉSULTATS MÉTHODE AUTORÉGRESSIVE

- ☞ Détermination de l'ordre autorégressif

- ☞ Pour déterminer l'ordre autorégressif  $p$ :

calcule pour  $i = 0, \dots, q$  où  $q$  est beaucoup plus grand que l'ordre autorégressif anticipé.

Si l'ordre  $0 \leq j \leq q$ , alors la distribution

$$T_{q-j} = n \left( 1 - \frac{\hat{S}_q^2}{\hat{S}_j^2} \right)$$

converge vers une chi-carrée  $X^2$  avec  $q-j$  degrés de liberté lorsque  $n$  augmente.

- ☞ Test d'hypothèse pour  $j = 0, 1, \dots$  jusqu'à ce que le test soit satisfait pour un  $\alpha$  donné.

- ☞ L'ordre  $p$  étant déterminé, on estime la variance par

$$Var(\bar{x}_n) = \frac{\hat{S}_p^2}{n \left( \sum_{s=0}^p \hat{b}_{p,s} \right)^2}$$

### SIMULATION DISCRÈTE ANALYSE DES RÉSULTATS MÉTHODE AUTORÉGRESSIVE

☞ On calcule le degré de liberté  $n_l$  par l'expression

$$n_l = \frac{nb}{2 \sum_{s=0}^p (p-2s)b_s}$$

où  $b = \sum_{s=0}^p \hat{b}_{p,s}$  (En pratique, on substitue pour  $b_s$ ).

☞ La méthode autorégressive mesure:

- la distance de corrélation entre les observations,
- l'information additionnelle sur le biais dû aux conditions initiales.

### SIMULATION DISCRÈTE ANALYSE DES RÉSULTATS MÉTHODE AUTORÉGRESSIVE

☞ Fonction de corrélation

- Une fois les covariances calculées par

$$R_j = - \sum_{s=1}^p b_s R_{j-s} \quad j > 0$$

- Calculer récursivement la fonction d'autocovariance  $\{R_j; j=p, p+1, \dots\}$ .

La quantité  $r_j = R_j / R_0$  représente la corrélation entre les observations distancées de  $j$  unités et  $\{r_j; j = 0, 1, 2, \dots\}$  la fonction d'autocorrélation des  $\{x_t\}$  que l'on peut aisément calculée par

$$\hat{r}_j = - \sum_{s=1}^p b_s r_{j-s} \quad j > 0$$

- Le graphique de cette fonction donne une image de la corrélation entre les observations distancées de  $j$  unités, pour  $j = 1, 2, \dots$ . De cette fonction, on peut déterminer pour un  $0 < \alpha < 1$  donné, un  $j$  tel que

$$j^* = \min(j : |r_j| \leq \alpha)$$

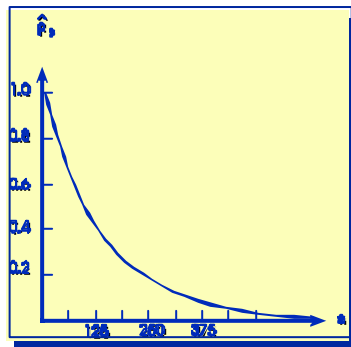
### SIMULATION DISCRÈTE ANALYSE DES RÉSULTATS MÉTHODE AUTORÉGRESSIVE

Les estimés des  $r_j$  sont basés sur les valeurs échantillonnées

$$\hat{R}_0, \hat{R}_1, \dots, \hat{R}_{p-1} \text{ et } \hat{b}_{p,1}, \hat{b}_{p,2}, \dots, \hat{b}_{p,p}$$

Ainsi,

$$r_j = \begin{cases} \frac{\hat{R}_j}{\hat{R}_0} & j = 0, 1, \dots, p-1 \\ -\sum_{s=1}^p \hat{b}_{p,s} \hat{r}_{j-s} & j \geq p \end{cases}$$



### SIMULATION DISCRÈTE ANALYSE DES RÉSULTATS MÉTHODE AUTORÉGRESSIVE

*Influence des conditions initiales*

Moyenne conditionnelle:

$$m_j = E(x_j | x_{j-p}, x_{j-2}, \dots, x_{j-p}) \text{ pour un ordre } p$$

Si on prend l'espérance mathématique de

$$y_j = \sum_{s=0}^p b_s (x_{j-s} - m) \quad j = p+1, \dots, n$$

on obtient

$$m_j = m - \sum_{s=1}^p b_s (m_{j-s} - m)$$

⇔ une suite de moyenne conditionnelle pour

$$\{m_j; j = p+1, p+2, \dots, x_1, \dots, x_p\}$$

Si on commence la simulation avec  $x_1 = x_2 = x_3 = \dots = x_p$ , alors la moyenne conditionnelle de  $x_{p+1}$  est

$$m_{p+1} = m \sum_{s=0}^p b_s \quad b_0 \equiv 0$$

### SIMULATION DISCRÈTE ANALYSE DES RÉSULTATS MÉTHODE AUTORÉGRESSIVE

Lorsque  $j$  croît, on pourrait montrer que

$$\lim_{n \rightarrow \infty} m_j = m$$

Le rapport  $(m_j - m)/m$  donne une mesure de l'influence des  $x_j$  sur les conditions initiales.

On peut ainsi calculer un  $j^*$  tel que

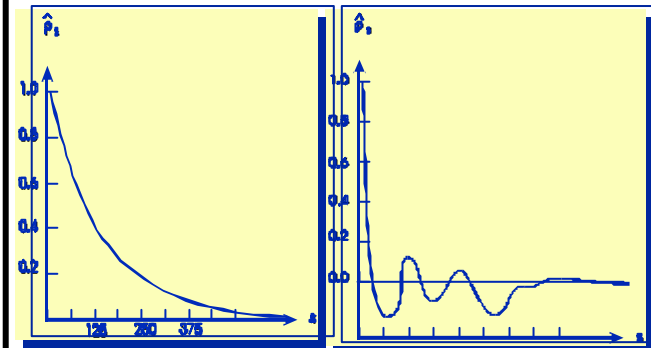
$$j^* = \min \left( j : \frac{|m_j - m|}{|m|} \leq b \right)$$

pour un  $\beta$  donné.

### SIMULATION DISCRÈTE ANALYSE DES RÉSULTATS MÉTHODE AUTORÉGRESSIVE

*Fonction de densité spectrale*

- La fonction d'autocorrélation  $\{r_j; j=0, 1, 2, \dots\}$  peut être monotone décroissante.
- Les fonctions d'autocorrélation peuvent osciller de façon cyclique.
- Si tel est le cas, on cherche à identifier la fréquence de ce cycle. Cette fonction d'autocorrélation ne permet pas toujours d'identifier ce cycle.



## SIMULATION DISCRÈTE ANALYSE DES RÉSULTATS MÉTHODE AUTORÉGRESSIVE

☞ Il existe une transformation de  $\{R_s\}$  qui permet d'identifier assez directement la fréquence de ce cycle.

☞ Puisque  $\{x_i\}$  est un processus stationnaire, on peut représenter  $R_s$  par

$$R_s = \int_{-p}^p g(I) e^{iIs} dI$$

où

$$g(I) = \frac{1}{2p} \sum_{s=-\infty}^{\infty} r_s e^{-is}$$

$$g(I) = g(-I) \quad |I| \leq p$$

☞ Ici,  $R_s$  et  $g(I)$  forme une paire de transformée de Fourier. La fonction  $g(I)$  est le spectre de  $\{x_i\}$  et la fonction de densité

spectrale est  $h(I) = \frac{g(I)}{R_0}$  où

$$R_0 = \int_{-p}^p g(I) dI = 2 \int_0^p g(I) dI$$

## SIMULATION DISCRÈTE ANALYSE DES RÉSULTATS MÉTHODE AUTORÉGRESSIVE

☞ Le spectre peut être exprimé en termes de  $\sigma^2$  et  $b_1, \dots, b_p$  par l'expression:

$$g(I) = \frac{\sigma^2}{2p \sum_{r,s=0}^p b_r b_s \cos I(r-s)} \quad |I| \leq p$$

où on substitue pour  $\sigma$  et  $b_1, \dots, b_p$  les estimés  $\hat{\sigma}^2$  et  $\hat{b}_1, \dots, \hat{b}_p$ .

☞ On peut alors déterminer la période  $P = \frac{2p}{I}$  pour une fréquence  $\pi$ .

## SIMULATION DISCRÈTE ANALYSE DES RÉSULTATS MÉTHODE AUTORÉGRESSIVE

```

ROUTINE Methode.Autoregressive GIVING N, X
" Calcule un intervalle a 0.95 end
" utilisant la méthode de régression
DEFINE i, j, k, N, No, p      AS INTEGER VARIABLE
DEFINE Critical.Value, demi.Largeur, T AS REAL VARIABLE
DEFINE Df, V, W, Xbar      AS REAL VARIABLE
DEFINE bsum, Lz, R, S, Vxbar, X      AS a 1-DIM REAL ARRAY
DEFINE b      AS a REAL 2-DIM ARRAY

IF N < 50,
  PRINT 1 LINE THUS
  Erreur ---- nombre d'observations insuffisant
  RETURN
  ENDIF

RESERVE S (*)          AS 51
RESERVE b(*,*)        AS 51 BY *,
  bsum(*), Lz(*), R(*), Vxbar(*) AS 51

FOR i = 1 TO N,
  DO
    COMPUTE Xbar AS THE MEAN  OF X(i)
    COMPUTE R(1) AS THE VARIANCE OF X(i)
  LOOP
  LET S(1) = R(1)
  SKIP 5 LINES
  PRINT 3 LINES WITH Xbar THUS
  Estimation d'un intervalle a 0.95 --- Approche autorégressive
  Moyenne échantillonnée = ***** *****
  SKIP 3 LINES
  BEGIN REPORT
  
```

## SIMULATION DISCRÈTE ANALYSE DES RÉSULTATS MÉTHODE AUTORÉGRESSIVE

```

" BEGIN HEADING
  PRINT 3 LINES THUS
  Variance échan. Intervalle (0.95) Degré Valeur
  Ordre de la moy.échan. Inf. Sup. liberté T Critique
  -----
" END HEADING
RESERVE b(1,*) AS 1
LET b(1,1) = 1.0
FOR i = 1 TO 50,
  DO
    LET k = i + 1
    RESERVE b(k,*) AS k
    FOR j = 1 TO N i,
      COMPUTE R(k) AS THE MEAN OF (X(j) - Xbar) *(X(i+j) - Xbar)
    FOR j = 1 TO i,
      DO
        COMPUTE V AS THE SUM OF b(i, j) * R(j)
        COMPUTE W AS THE SUM OF b(i, j) * R(i+j+2)
      LOOP
    LET b(k, k) = - W / V
    LET b(k, 1) = 1.0
    IF i > 1,
      FOR j = 2 TO i,
        LET b(k, j) = b(i, j) + b(k, k) * b(i, k - j + 1)
    ENDIF
  FOR j = 1 TO k,
    DO
      COMPUTE S(k) AS THE SUM OF b(k, j) * R(j)
      IF j < k,
        COMPUTE bsum(i) AS THE SUM OF b(i, j)
        COMPUTE Lz(i) AS THE SUM OF (i - 1.0 - 2.0 * (j - 1.0)) * b(i, j)
      ENDIF
    LOOP
  LET Vxbar(i) = S(i) / (N * bsum(i) * bsum(i))
  LOOP
  
```

## SIMULATION DISCRÈTE ANALYSE DES RÉSULTATS MÉTHODE AUTORÉGRESSIVE

```

LET Df = N - 1.0

FOR j = 1 TO 50
DO
LET T = N * (1.0 - S(51) / S(j))
LET Critical.Value = ChiCarree.f ( REAL.F ( 51 - j), 0.90)
IF j > 1,
LET Df = MAX.F ( 1.0, N * bsum(j) / (2.0 * Lz(j)))
ENDIF
LET demi.Largeur = Student.F ( Df) * Sqrt.F ( Vxbar(j))
PRINT 1 LINE WITH j-1, Vxbar(j), Xbar-demi.Largeur, Xbar +
demi.Largeur,
Df, T, Critical.Value THUS
***** ***** ***** ***** ***** ***** ***** ***** *****
***** **
SKIP 1 LINE
IF T <= Critical.Value AND No = 0,
LET p = j
LET No = 1
ENDIF
LOOP
    
```

## SIMULATION DISCRÈTE ANALYSE DES RÉSULTATS MÉTHODE AUTORÉGRESSIVE

```

IF p = 0,
PRINT 1 LINE THUS
Aucun ordre ne satisfait le test
ELSE
PRINT 1 LINE WITH p-1 THUS
Ordre = *** satisfait le test
ENDIF
END " REPORT
SKIP 5 LINES
PRINT 1 LINE THUS
Autocorrélation échantillonnée
SKIP 1 LINE
FOR i = 1 TO p, WRITE R(i) / R(1) AS (4) E(18,6)
SKIP 3 LINES
PRINT 1 LINE THUS
Echantillon coefficient de corrélation
FOR i = 1 TO p, WRITE b(p, i) AS (4) E(18,6)
IF p = 1,
RELEASE b(*, *), bsum(*), Lz(*), R(*), S(*), Vxbar(*)
RETURN
ENDIF
    
```

## SIMULATION DISCRÈTE ANALYSE DES RÉSULTATS MÉTHODE AUTORÉGRESSIVE

```

DEFINE Y AS a 1-DIM REAL ARRAY
DEFINE Theta, Z, ZZ AS REAL VARIABLE
RESERVE Y(*) AS p-1
FOR i = 1 TO p-1, LET Y(i) = R(i) / R(1)
LET j = 1
LET Theta = 0.05
LET ZZ = RINF.C
WHILE (ZZ > 0.01)
DO
  LET Z = 0
  FOR i = 1 TO p-1,
  DO
    SUBTRACT b(p, p+1-i) * Y(i) FROM Z
    IF i < p-1,
      LET y(i) = Y(i+1)
    ENDIF
  LOOP
  LET Y (p-1) = Z
  LET ZZ = ABS.F (Z)
  IF ZZ <= Theta,
    SKIP 2 LINES
    PRINT 1 LINE WITH j, theta THUS
    J minimum = ***** pour ABS.F(autocorrelation(j)) <= *,**
    IF ZZ > 0.01,
      LET Theta = 0.01
    ENDIF
  ENDIF
  ADD 1 TO j
LOOP
RELEASE bsum(*), Lz(*),R(*), S(*), Vxbar(*)

```

## SIMULATION DISCRÈTE ANALYSE DES RÉSULTATS MÉTHODE AUTORÉGRESSIVE

```

FOR i = 1 TO p-1,
  LET Y(i) = -1.0
LET j = 1
LET Theta = 0.05
LET ZZ = RINF.C
WHILE (ZZ > 0.01)
DO
  LET Z = 0
  FOR i = 1 TO p-1
  DO
    SUBTRACT b(p, p+1-i) * Y(i) FROM Z
    IF i < p-1,
      LET y(i) = Y(i+1)
    ENDIF
  LOOP

```

## SIMULATION DISCRÈTE ANALYSE DES RÉSULTATS MÉTHODE AUTORÉGRESSIVE

```

LET Y (p-1) = Z
LET ZZ = ABS.F (Z)
IF ZZ <= Theta,
  SKIP 2 LINES
  PRINT 1 LINE WITH j, theta THUS
  J minimum = ***** pour ABS MEAN(j) cond - MEAN) / MEAN <=
  * : ***
    IF ZZ > 0.01,
      LET Theta = 0.01
    ENDIF
  ENDIF
  ADD 1 TO j

LOOP
RELEASE b(*, *), Y(*)
RETURN
end "Methode. Autoregressive
    
```

## SIMULATION DISCRÈTE ANALYSE DES RÉSULTATS MÉTHODE SPECTRALE

☞ Rappelons que

$$Var(\bar{x}_n) = \frac{1}{n} \left( s^2 + 2 \sum_{i=1}^{n-1} \left(1 - \frac{i}{n}\right) R_i \right)$$

☞ Cette expression peut aussi s'écrire sous la forme

$$Var(\bar{x}_n) = \frac{1}{n} \left( R_0 + 2 \sum_{i=1}^{n-1} \left(1 - \frac{i}{n}\right) R_i \right)$$

☞ ou une autre forme est

$$Var(\bar{x}_n) = \frac{1}{n} \sum_{i=-n+1}^{n-1} \left(1 - \frac{|i|}{n}\right) R_i$$

☞ La convergence de  $nVar(\bar{x}_n)$  vers  $V = \sum_{i=-\infty}^{\infty} r_i$  dépend de la fonction de pondération

$$W_{n,i}^{(1)} = 1 - \frac{|i|}{n} \quad i = 0, 1, \dots, \pm n$$

## SIMULATION DISCRÈTE ANALYSE DES RÉSULTATS MÉTHODE SPECTRALE

- ☞ Si pour  $n$  suffisamment grand, l'utilisation de  $V/n$  engendre une imprécision minimale,
- ☞ Existe-t-il d'autres fonctions de pondération qui permettent de faire l'approximation de  $V$  par une combinaison linéaire d'un nombre moindre de covariances et avec une imprécision acceptable.
- ☞ Considérons l'expression

$$V_j = \sum_{i=k}^k W_{ni}^{(j)} \quad k \leq n$$

- ☞ où la suite  $\{W_{ni}^{(i)}\}$  est telle que

$$\lim_{k \rightarrow \infty} V_j = \sum_{i=-\infty}^{\infty} R_i = V$$

## SIMULATION DISCRÈTE ANALYSE DES RÉSULTATS MÉTHODE SPECTRALE

- ☞ Les covariances étant inconnues, on utilise un estimé.
- ☞ En manipulant l'expression, on obtient:

$$\begin{aligned} \hat{R}_i &= \frac{1}{n-1} \sum_{j=1}^{n-i} [(x_j - \bar{x}_i)(x_{j+i} - \bar{x}_i)] \quad i=0, \dots, p \\ &= \frac{1}{n-1} \sum_{j=1}^{n-i} [(x_j - \bar{m})(x_{j+i} - \bar{m}) - (\bar{x}_i - \bar{m})^2] \end{aligned}$$

- ☞ On peut montrer que:

$$\lim_{n \rightarrow \infty} nE(\hat{R}_i - R_i) = -V$$

$$\lim_{n \rightarrow \infty} nE(\hat{V}_j - V_j) = -V \sum_{i=k}^k W_{k,i}^{(j)}$$

- ☞ On en déduit que:  $E(\hat{V}_j) = V_j - \frac{1}{n} \sum_{i=k}^k W_{k,i}^{(j)}$

- ☞ Il est aussi clair que:  $\lim_{k \rightarrow \infty} \lim_{n \rightarrow \infty} E(\hat{V}_j) = V$

si  $k$  croît moins rapidement que  $n$ . Si  $k$  est suffisamment grand pour que  $V_j \approx V$ , alors

$$\tilde{V}_j = \frac{\hat{V}_j}{1 - \frac{1}{n} \sum_{i=k}^k W_{k,i}^{(j)}}$$

est un meilleur estimateur pour  $V$  car il élimine le biais introduit dans  $\hat{R}_i$  par l'utilisation de  $\bar{x}_n$  au lieu de  $\mu$ . En pratique, le diviseur  $n-i$  est remplacé par  $n$ , car l'effet est négligeable si  $n \gg k$ .

## SIMULATION DISCRÈTE ANALYSE DES RÉSULTATS MÉTHODE SPECTRALE

☞ Pour la variance de  $\tilde{V}_j$ , on pourrait montrer que

$$\lim_{n \rightarrow \infty} n \text{Var}(\tilde{V}_j) = \lim_{n \rightarrow \infty} n \text{Var}(\hat{V}_j) \approx V^2 \left\{ \sum_{-k}^k [W_{k,i}^{(j)}]^2 + k \right\}$$

où  $k$ : contribution de la variance due à la non-normalité de  $x_1, x_2, \dots, x_n$ .

☞ La dernière expression signifie que si  $k$  est déterminé tel que  $E(\tilde{V}_j) \approx V$  une plus grande valeur de  $k$  améliore l'approximation de  $E(\tilde{V}_j) \approx V$  au détriment d'une plus grande valeur de la variance  $\text{Var}(\tilde{V}_j)$ .

☞ Deux fonctions de pondération méritent considération. Il s'agit des fonctions de Tukey-Hanning

$$W_{k,i}^{(2)} = \begin{cases} 0.5 \left( 1 + \cos \frac{i}{k} \pi \right) & |i| \leq k \\ 0 & \text{ailleurs} \end{cases}$$

et de Parzen

$$W_{k,i}^{(3)} = \begin{cases} 1 - 6 \left| \frac{i}{k} \right|^2 + 6 \left| \frac{i}{k} \right|^3 & |i| \leq k/2 \\ 2 \left( 1 - \left| \frac{i}{k} \right| \right)^3 & k/2 \leq |i| \leq k \\ 0 & \text{ailleurs} \end{cases}$$

## SIMULATION DISCRÈTE ANALYSE DES RÉSULTATS MÉTHODE SPECTRALE

☞ Pour un même  $k_2$ , l'erreur d'approximation de la fonction  $W_{k,i}^{(2)}$  est inférieure à l'erreur de la fonction  $W_{k,i}^{(3)}$ . Toutefois, si on considère la variances, l'inégalité est renversée. En effet,

$$E(\tilde{V}_2) \approx V$$

$$E(\tilde{V}_3) \approx V$$

$$\text{Var}(\tilde{V}_2) \approx 0.75V^2 \frac{k}{n}$$

$$\text{Var}(\tilde{V}_3) \approx 0.539V^2 \frac{k}{n}$$

☞ De plus, on a

$$\sum_{i=-k}^k W_{k,i}^{(2)} = k$$

$$\sum_{i=-k}^k W_{k,i}^{(3)} = 0.75k$$

ce qui facilite les calculs de  $V_2$  et  $\tilde{V}_3$

## SIMULATION DISCRÈTE ANALYSE DES RÉSULTATS MÉTHODE SPECTRALE

☞ Le calcul des degrés de liberté s'effectue de la même façon que dans la méthode autorégressive. Ainsi  $l_1 = 2.667 \frac{n}{k}$  pour la fonction de pondération de Tukey-Hanning et pour  $l_2 = 3.711 \frac{n}{k}$  la fonction de Parzen.

☞ Un intervalle de confiance de niveau  $100(1-\alpha)\%$  pour la moyenne  $\mu$  est

$$\bar{x}_n \pm t_{\alpha/2,1} \sqrt{V_i / n}$$

☞ Il n'existe pas vraiment de méthode pour déterminer le nombre de sous-intervalles  $K$ .

☞ Mais si on applique une règle conservatrice d'au moins 100 degrés de liberté, alors pour un nombre d'observations  $n$  fixé, on peut déterminer  $k=c_1 n/l$  où  $c_1$  prendra les valeurs 2.667 ou 3.711 selon la fonction de pondération choisie.

## SIMULATION DISCRÈTE ANALYSE DES RÉSULTATS MÉTHODE SPECTRALE

ROUTINE Methode.Spectrale GIVING N, X

" Cette routine utilise méthode spectrale (Tukey-Hanning)

" pour calculer un intervalle de confiance a 0.95

DEFINE i, j, k, N AS INTEGER VARIABLE

DEFINE Df, demi.Largeur, V, Vxbar, Xbar AS REAL VARIABLE

DEFINE R, X AS 1-DIM REAL ARRAYS

IF N < 512,

PRINT 1 LINE THUS

Erreur -- Nombre insuffisant d'observations

RETURN

ENDIF

RESERVE R(\*) AS 513

FOR i = 1 TO N,

COMPUTE Xbar AS THE MEAN OF X(i)

SKIP 2 LINES

PRINT 3 LINES WITH Xbar THUS

Estimation d'un intervalle a 0.95 --- Approche spectrale

Moyenne échantillonnée = \*\*\*\*\*.\*\*\*\*\*

## SIMULATION DISCRÈTE ANALYSE DES RÉSULTATS MÉTHODE AUTORÉGRESSIVE

```
ROUTINE Methode.Autoregressive GIVING N, X
" Calcule un intervalle a 0.95 end
" utilisant la méthode de régression
DEFINE i, j, k, N, No, p AS INTEGER VARIABLE
DEFINE Critical.Value, Demie.Largeur, T AS REAL VARIABLE
DEFINE Df, V, W, Xbar AS REAL VARIABLE
DEFINE bsum, Lz, R, S, Vxbar, X AS a 1-DIM REAL ARRAY
DEFINE b AS a REAL 2-DIM ARRAY
```

```
IF N < 50,
  PRINT 1 LINE THUS
  Erreur ---- nombre d'observations insuffisant
  RETURN
ENDIF
```

```
RESERVE S (*) AS 51
RESERVE b(*,*) AS 51 BY *,
  bsum(*), Lz(*), R(*), Vxbar(*) AS 51
```

```
FOR i = 1 TO N,
  DO
    COMPUTE Xbar AS THE MEAN OF X(i)
    COMPUTE R(1) AS THE VARIANCE OF X(i)
  LOOP
```

```
LET S(1) = R(1)
SKIP 5 LINES
PRINT 3 LINES WITH Xbar THUS
  Estimation d'un intervalle a 0.95 --- Approche autorégressive
  Moyenne échantillonnée = *****
SKIP 3 LINES
BEGIN REPORT
```

$$\bar{x}_n = \frac{\sum_{i=1}^n x_i}{n}$$

$$R_0 = Var(\bar{x}_n)$$

## SIMULATION DISCRÈTE ANALYSE DES RÉSULTATS MÉTHODE AUTORÉGRESSIVE

```
BEGIN HEADING
PRINT 3 LINES THUS
  Variance échan. Intervalle (0.95) Degré Valeur
  Ordre de la moy. échan. Inf. Sup. liberté T Critique
-----
```

```
END HEADING
RESERVE b(1, *) AS 1
LET b(1, 1) = 1.0
FOR i = 1 TO 50,
  DO
    LET k = i + 1
    RESERVE b( k, *) AS k
    FOR j = 1 TO N-i,
      COMPUTE R(k) AS THE MEAN OF (X(j) - Xbar) * (X(i+j) - Xbar)
    FOR j = 1 TO i,
      DO
        COMPUTE V AS THE SUM OF b(i, j) * R(j)
        COMPUTE W AS THE SUM OF b(i, j) * R(i-j+2)
      LOOP
    LET b(k, k) = - W / V
    LET b(k, 1) = 1.0
    IF i > 1,
      FOR j = 2 TO i,
        LET b(k, j) = b(i, j) + b(k, k) * b(i, k-j+1)
    ENDIF
```

```
FOR j = 1 TO k,
  DO
    COMPUTE S(k) AS THE SUM OF b(k, j) * R(j)
```

```
IF j < k,
  COMPUTE bsum(i) AS THE SUM OF b(i, j)
```

## SIMULATION DISCRÈTE ANALYSE DES RÉSULTATS MÉTHODE AUTORÉGRESSIVE

```
" BEGIN HEADING
  PRINT 3 LINES THUS
    Variance échan. Intervalle (0.95) Degré      Valeur
  Ordre  de la moy. échan. Inf.  Sup.  liberté T  Critique
  -----
" END HEADING
RESERVE b(1, *) AS 1
LET b(1, 1) = 1.0
FOR i = 1 TO 50,
DO
  LET k = i + 1
  RESERVE b( k, *) AS k
  FOR j = 1 TO N-i,
    COMPUTE R(k) AS THE MEAN OF (X(j) - Xbar) *(X(i+j) -
Xbar)
  FOR j = 1 TO i,
DO
  COMPUTE V AS THE SUM OF b(i, j) * R(j)
  COMPUTE W AS THE SUM OF b(i, j) * R(i-j+2)
LOOP
LET b(k, k) = - W / V
LET b(k, 1) = 1.0
IF i > 1,
  FOR j = 2 TO i,
    LET b(k, j) = b(i, j) + b(k, k) * b(i, k-j+1)
  ENDIF
FOR j = 1 TO k,
DO
  COMPUTE S(k) AS THE SUM OF b(k, j) * R(j)
  IF j < k,
    COMPUTE hsum(i) AS THE SUM OF b(i, j)
  COMPUTE Lz(i) AS THE SUM OF (i - 1.0 - 2.0 * (j - 1.0)) *
b(i, j)
ENDIF
```

## SIMULATION DISCRÈTE ANALYSE DES RÉSULTATS MÉTHODE AUTORÉGRESSIVE

```
LET Df = N - 1.0

FOR j = 1 TO 50
DO
  LET T = N * (1.0 - S(51) / S(j))
  LET Critical.Value = ChiCarree.f ( REAL.F ( 51 - j), 0.90)
  IF j > 1,
    LET Df = MAX.F ( 1.0, N * bsum(j) / (2.0 * Lz(j)))
  ENDIF
  LET Demie.Largeur = Student.F ( Df) * Sqrt.F ( Vxbar(j))
  PRINT 1 LINE WITH j-1, Vxbar(j), Xbar-Demie.Largeur, Xbar +
Demie.Largeur,
    Df, T, Critical.Value THUS
  *****
  *****
  *****
SKIP 1 LINE
IF T <= Critical.Value AND No = 0,
  LET p = j
  LET No = 1
ENDIF
LOOP
```

## SIMULATION DISCRÈTE ANALYSE DES RÉSULTATS MÉTHODE AUTORÉGRESSIVE

```

IF p = 0,
    PRINT 1 LINE THUS
    Aucun ordre ne satisfait le test
ELSE
    PRINT 1 LINE WITH p-1 THUS
    Ordre = *** satisfait le test
ENDIF
END " REPORT
SKIP 5 LINES
PRINT 1 LINE THUS
    Autocorrélation échantillonnée
SKIP 1 LINE
FOR i = 1 TO p, WRITE R(i) / R(1) AS (4) E(18,6)
SKIP 3 LINES
PRINT 1 LINE THUS
    Echantillon coefficient de corrélation
FOR i = 1 TO p, WRITE b(p, i) AS (4) E(18,6)
IF p = 1,
    RELEASE b(*,*), bsum(*), Lz(*), R(*), S(*), Vxbar(*)
    RETURN
ENDIF
    
```

## SIMULATION DISCRÈTE ANALYSE DES RÉSULTATS MÉTHODE AUTORÉGRESSIVE

```

DEFINE Y AS a 1-DIM REAL ARRAY
DEFINE Theta, Z, ZZ AS REAL VARIABLE
RESERVE Y(*) AS p-1
FOR i = 1 TO p-1, LET Y(i) = R(i) / R(1)
LET j = 1
LET Theta = 0.05
LET ZZ = RINF.C
WHILE (ZZ > 0.01)
DO
    LET Z = 0
    FOR i = 1 TO p-1,
    DO
        SUBTRACT b(p, p+1-i) * Y(i) FROM Z
        IF i < p-1,
            LET y(i) = Y(i+1)
        ENDIF
    LOOP
    LET Y (p-1) = Z
    LET ZZ = ABS.F (Z)
    IF ZZ <= Theta,
        SKIP 2 LINES
        PRINT 1 LINE WITH j, theta THUS
        J minimum = ***** pour ABS.F(autocorrelation(j)) <= ***
        IF ZZ > 0.01,
            LET Theta = 0.01
        ENDIF
    ENDIF
    ADD 1 TO j
LOOP
RELEASE bsum(*), Lz(*),R(*), S(*), Vxbar(*)
    
```

## SIMULATION DISCRÈTE ANALYSE DES RÉSULTATS MÉTHODE AUTORÉGRESSIVE

```

FOR i = 1 TO p-1,
  LET Y(i) = -1.0
LET j = 1
LET Theta = 0.05
LET ZZ = RINF.C
WHILE (ZZ > 0.01)
DO
  LET Z = 0
  FOR i = 1 TO p-1
  DO
    SUBTRACT b(p, p+1 -i) * Y(i) FROM Z
    IF i < p-1,
      LET y(i) = Y(i+1)
    ENDIF
  LOOP

```

## SIMULATION DISCRÈTE ANALYSE DES RÉSULTATS MÉTHODE AUTORÉGRESSIVE

```

LET Y (p-1) = Z
LET ZZ = ABS.F (Z)
IF ZZ <= Theta,
  SKIP 2 LINES
  PRINT 1 LINE WITH j,theta THUS
  J minimum = ***** pour ABS MEAN(j) cond - MEAN) / MEAN<=
  * ***
  IF ZZ > 0.01,
    LET Theta = 0.01
  ENDIF
ENDIF
ADD 1 TO j

LOOP
RELEASE b(*,*), Y(*)
RETURN
end "Methode.Autoregressive

```

## SIMULATION DISCRÈTE ANALYSE DES RÉSULTATS MÉTHODE SPECTRALE

☞ Rappelons que

$$\text{Var}(\bar{x}_n) = \frac{1}{n} \left( s^2 + 2 \sum_{i=1}^{n-1} \left(1 - \frac{i}{n}\right) R_i \right)$$

☞ Cette expression peut aussi s'écrire sous la forme

$$\text{Var}(\bar{x}_n) = \frac{1}{n} \left( R_0 + 2 \sum_{i=1}^{n-1} \left(1 - \frac{i}{n}\right) R_i \right)$$

☞ ou une autre forme est

$$\text{Var}(\bar{x}_n) = \frac{1}{n} \sum_{i=-n+1}^{n-1} \left(1 - \frac{|i|}{n}\right) R_i$$

☞ La convergence de  $n\text{Var}(\bar{x}_n)$  vers  $V = \sum_{i=-\infty}^{\infty} r_i$  dépend de la fonction de pondération

$$W_{n,i}^{(1)} = 1 - \frac{|i|}{n} \quad i = 0, 1, \dots, \pm n$$

## SIMULATION DISCRÈTE ANALYSE DES RÉSULTATS MÉTHODE SPECTRALE

☞ Si pour  $n$  suffisamment grand, l'utilisation de  $V/n$  engendre une imprécision minimale,

☞ Existe-t-il d'autres fonctions de pondération qui permettent de faire l'approximation de  $V$  par une combinaison linéaire d'un nombre moindre de covariances et avec une imprécision acceptable.

☞ Considérons l'expression

$$V_j = \sum_{i=k}^k W_{n,i}^{(j)} \quad k \leq n$$

☞ où la suite  $\{W_{n,i}^{(j)}\}$  est telle que

$$\lim_{k \rightarrow \infty} V_j = \sum_{i=-\infty}^{\infty} R_i = V$$

## SIMULATION DISCRÈTE ANALYSE DES RÉSULTATS MÉTHODE SPECTRALE

- ☞ Les covariances étant inconnues, on utilise un estimé.
- ☞ En manipulant l'expression, on obtient:

$$\hat{R}_i = \frac{1}{n-1} \sum_{j=1}^{n-i} [(x_j - \bar{x}_n)(x_{j+i} - \bar{x}_n)] \quad i = 0, \dots, p$$

$$= \frac{1}{n-1} \sum_{j=1}^{n-i} [(x_j - m) - (x_n - m)][(x_{j+i} - m) - (x_n - m)]$$

- ☞ On peut montrer que:

$$\lim_{n \rightarrow \infty} nE(\hat{R}_i - R_i) = -V$$

- ☞ On en déduit que:  $\lim_{n \rightarrow \infty} nE(\hat{V}_j - V_j) = -V \sum_{i=-k}^k W_{k,i}^{(j)}$

- ☞ Il est aussi clair que:  $E(\hat{V}_j) \approx V_j - \frac{1}{n} \sum_{i=-k}^k W_{k,i}^{(j)}$

- ☞ si  $k$  croît moins rapidement que  $n$ ,  $\lim_{k \rightarrow \infty} \lim_{n \rightarrow \infty} E(\hat{V}_j) = V$  si  $k$  est suffisamment grand pour que  $\frac{1}{n} \sum_{i=-k}^k W_{k,i}^{(j)} \approx 0$ .

$$V_j \approx V,$$

$$\tilde{V}_j = \frac{\hat{V}_j}{\bar{x}_n}$$

- ☞ est un meilleur estimateur pour  $V_j$  car il élimine le biais introduit dans  $\hat{V}_j$  par l'utilisation de  $\bar{x}_n$  au lieu de  $\mu$ . En pratique, le diviseur  $n-i$  est remplacé par  $n$ , car l'effet est négligeable si  $n \gg k$ .

## SIMULATION DISCRÈTE ANALYSE DES RÉSULTATS MÉTHODE SPECTRALE

- ☞ Pour la variance de  $\tilde{V}_j$ , on pourrait montrer que

$$\lim_{n \rightarrow \infty} nVar(\tilde{V}_j) = \lim_{n \rightarrow \infty} nVar(\hat{V}_j) \approx V^2 \left\{ \sum_{i=-k}^k [W_{k,i}^{(j)}]^2 + k \right\}$$

- ☞ où  $k$ : contribution de la variance due à la non-normalité de  $x_1, x_2, \dots, x_n$ .

- ☞ La dernière expression signifie que si  $k$  est déterminé tel que  $E(\tilde{V}_j) \approx V$ , une plus grande valeur de  $k$  améliore l'approximation de  $E(\tilde{V}_j) \approx V$  au détriment d'une plus grande valeur de la variance  $Var(\tilde{V}_j)$ .

- ☞ Deux fonctions de pondération méritent considération. Il s'agit des fonctions de Tukey-Hanning

$$W_{k,i}^{(2)} = \begin{cases} 0.5 \left( 1 + \cos \frac{i\pi}{k} \right) & |i| \leq k \\ 0 & \text{ailleurs} \end{cases}$$

- ☞ et de Parzen

$$W_{k,i}^{(3)} = \begin{cases} 1 - 6 \left| \frac{i}{k} \right|^2 + 6 \left| \frac{i}{k} \right|^3 & |i| \leq k/2 \\ 2 \left( 1 - \left| \frac{i}{k} \right| \right)^3 & k/2 < |i| \leq k \\ 0 & \text{ailleurs} \end{cases}$$

## SIMULATION DISCRÈTE ANALYSE DES RÉSULTATS MÉTHODE SPECTRALE

☞ Pour un même  $k_2$ , l'erreur d'approximation de la fonction  $W_{k,i}^{(2)}$  est inférieure à l'erreur de la fonction  $W_{k,i}^{(3)}$ . Toutefois, si on considère la variances, l'inégalité est renversée. En effet,

$$E(\tilde{V}_2) \approx V$$

$$E(\tilde{V}_3) \approx V$$

$$Var(\tilde{V}_2) \approx 0.75V^2 \frac{k}{n}$$

$$Var(\tilde{V}_3) \approx 0.539V^2 \frac{k}{n}$$

☞ De plus, on a

$$\sum_{i=-k}^k W_{k,i}^{(2)} = k$$

$$\sum_{i=-k}^k W_{k,i}^{(3)} = 0.75k$$

☞ ce qui facilite les calculs de

$$V_2 \text{ et } \tilde{V}_3$$

## SIMULATION DISCRÈTE ANALYSE DES RÉSULTATS MÉTHODE SPECTRALE

☞ Le calcul des degrés de liberté s'effectue de la même façon que dans la méthode autorégressive. Ainsi  $l_1 = 2.667 \frac{n}{k}$  pour la fonction de pondération de Tukey-Hanning et pour  $l_2 = 3.711 \frac{n}{k}$  la fonction de Parzen.

☞ Un intervalle de confiance de niveau  $100(1-\alpha)\%$  pour la moyenne  $\mu$  est

$$\bar{x}_n \pm t_{\alpha/2, l} \sqrt{V_i/n}$$

☞ Il n'existe pas vraiment de méthode pour déterminer le nombre de sous-intervalles  $K$ .

☞ Mais si on applique une règle conservatrice d'au moins 100 degrés de liberté, alors pour un nombre d'observations  $n$  fixé, on peut déterminer  $k = c_i n/l$  où  $c_i$  prendra les valeurs 2.667 ou 3.711 selon la fonction de pondération choisie.

## SIMULATION DISCRÈTE ANALYSE DES RÉSULTATS MÉTHODE SPECTRALE

```

ROUTINE Methode.Spectrale GIVING N, X
" Cette routine utilise méthode spectrale (Tukey-Hanning)
" pour calculer un intervalle de confiance a 0.95
DEFINE i, j, k, N          AS INTEGER VARIABLE
DEFINE Df, Demie.Largeur, V, Vxbar, Xbar AS REAL VARIABLE
DEFINE R, X AS 1-DIM REAL ARRAYS
IF N < 512,
  PRINT 1 LINE THUS
  Erreur -- Nombre insuffisant d'observations
  RETURN
ENDIF
RESERVE R(*) AS 513
FOR i = 1 TO N,
  COMPUTE Xbar AS THE MEAN OF X(i)
SKIP 2 LINES
PRINT 3 LINES WITH Xbar THUS
  Estimation d'un intervalle a 0.95 --- Approche spectrale
  Moyenne échantillonnée = *****

```

## SIMULATION DISCRÈTE ANALYSE DES RÉSULTATS MÉTHODE SPECTRALE

```

SKIP 2 LINES
BEGIN REPORT
" BEGIN HEADING
  PRINT 3 LINES THUS
No.  Degré  Variance échan.  Intervalle (0.95)
interv.  liberté de la moy. échan.  Inf.  Sup.
-----
" END HEADING
FOR j = 1 TO 513,
  DO
  FOR i = 1 TO N-j+1,
    COMPUTE R(j) AS THE SUM OF (X(i) - Xbar)*(X(i+j-1)-Xbar)
  LET R(j) = R(j) / N
  LOOP
  LET k = 1
  FOR i = 1 TO 9,
  DO
  LET k = 2 * k
  FOR j = 1 TO k,
  COMPUTE V AS THE SUM OF ( 1.0 + COS.F(J * PI.C / k)) * R(j+1)
  LET Vxbar = ( V + R(1)) / ( N - k )
  LET Df = 2.667 * N / k
  LET Demie.Largeur = Student.F(Df) * SQRT.F ( Vxbar)
  SKIP 1 LINE
  PRINT 1 LINE WITH k, Df, Vxbar, Xbar - HalfWidth, Xbar +
HalfWidth
  THUS
  *****
  LOOP
  RELEASE R
  RETURN
END " REPORT
END " Methode.Spectrale

```

## SIMULATION DISCRÈTE ANALYSE DES RÉSULTATS AUTRES MÉTHODES

- ☞ *Méthode antithétique*
- ☞ Basée sur l'observation que si  $u$  est un nombre aléatoire dans 0, 1, il en est de même pour le nombre complémentaire  $1 - u$ .
- ☞ Le principe consiste à faire une première passe de simulation avec la suite de nombres aléatoires  $u_1, u_2, \dots, u_k$  suivie d'une autre passe de simulation avec la suite complémentaire  $1 - u_1, 1 - u_2, \dots, 1 - u_k$ .
- ☞ Soient  $x_i$  et  $x'_i$  les variables aléatoires associées avec la  $i$ -ième observation de chacune des passes; les suites  $\{x_i\}$  et  $\{x'_i\}$  sont corrélées négativement, car, en principe lorsqu'une faible valeur  $x_i$  est obtenue, la valeur obtenue de la suite  $x'_i$  complémentaire  $x'_i$  est plus grande, et vice versa.
- ☞ Rappelons que 
$$Var(\bar{x}_n) = \sum_{i=1}^n Var(x_i) + 2 \sum_{i < j} Cov(x_i, x_j)$$
- ☞ Une corrélation négative va réduire la variance de  $\bar{x}_n$  c'est-à-dire l'erreur échantillonnée. Dans cette méthode, on estime la moyenne de l'échantillon par:

$$\bar{y}_n = \frac{1}{n} \sum_{i=1}^n y_i = \frac{1}{n} \sum_{i=1}^n \frac{x_i + x'_i}{2}$$

- ☞ Équivalent à utiliser un échantillon de  $2n$  observations.

## SIMULATION DISCRÈTE ANALYSE DES RÉSULTATS AUTRES MÉTHODES

- ☞ On peut aussi montrer que la variable aléatoire  $y_i$  a une variance plus petite que celle obtenue à l'aide d'un échantillon de  $2n$  observations non-corrélées. En effet,

$$\begin{aligned} Var(\bar{y}_n) &= Var\left(\frac{\bar{x}_n + \bar{x}'_n}{2}\right) = \frac{1}{4} Var(\bar{x}_n + \bar{x}'_n) \\ &= \frac{1}{4} [Var(\bar{x}_n) + Var(\bar{x}'_n)] + \frac{1}{2} Cov(\bar{x}_n, \bar{x}'_n) \end{aligned}$$

- ☞ Puisqu'en principe  $Cov(\bar{x}_n, \bar{x}'_n) < 0$ , il s'ensuit que:

$$Var(\bar{y}_n) \leq \frac{1}{4} [Var(\bar{x}_n) + Var(\bar{x}'_n)]$$

- ☞ Il résulte de cette inégalité que même dans le cas le plus défavorable, la variance de  $\bar{y}_n$  est inférieure à la moitié de la plus grande des variances de  $\bar{x}_n$  et  $\bar{x}'_n$ .

## MÉTHODE ANTITHÉTIQUE

### ☞ Exemple

Soit un système M/M/1 avec arrivée à toutes les 5 minute ( $\lambda=1/5$ ) et service moyen de 4 minute ( $\mu=1/4$ ). Pour ce modèle, nous savons que:

$$L_Q = \frac{I^2}{m(m-1)} = 3.2 \text{ clients}$$

$$W_Q = L_Q / I = 16 \text{ minutes}$$

### ☞ Simulation:

- 10 répétitions indépendantes, chacune d'une durée de 6 heures et on recueille les observations pour les 3 dernières heures.
- Pour les arrivées et services, on utilise des suites indépendantes de nombres aléatoires  $\{u_i\}$  et  $\{u'_i\}$ .
- On répète pour 10 autres répétitions avec les variables antithétiques  $\{1-u_i\}$  et  $\{1-u'_i\}$
- Soient

$x_i$  : longueur moyenne de la queue pour la  $i^{\text{ème}}$  passe

$x'_i$  : longueur moyenne de la queue pour la  $i^{\text{ème}}$  passe antithétique

$w_i$  : temps moyen d'attente dans la queue pour la  $i^{\text{ème}}$  passe et,

$w'_i$  : temps moyen d'attente de la queue pour la  $i^{\text{ème}}$  passe antithétique

## MÉTHODE ANTHITÉTIQUE

### ☞ Définissons

$$y_i = (x_i + x'_i)/2 \text{ et } v_i = (w_i + w'_i)/2$$

| $i$      | $x_i$   | $x'_i$ | $y_i$  | $w_i$   | $w'_i$  | $v_i$   |
|----------|---------|--------|--------|---------|---------|---------|
| 1        | 12.1948 | 4.7130 | 8.4539 | 55.3378 | 19.4276 | 37.3827 |
| 2        | 1.6369  | 2.4363 | 2.0366 | 8.1092  | 12.7728 | 10.4410 |
| 3        | 1.3258  | 1.3960 | 1.3609 | 6.4496  | 6.3884  | 6.4190  |
| 4        | 3.5298  | 4.1751 | 3.8525 | 16.5749 | 20.1298 | 18.3524 |
| 5        | 0.6205  | 0.9716 | 0.7961 | 3.2529  | 4.9970  | 4.1250  |
| 6        | 5.5517  | 1.1496 | 3.3507 | 22.8847 | 6.2077  | 14.5462 |
| 7        | 2.0837  | 2.8093 | 2.4465 | 11.3658 | 14.1777 | 12.7718 |
| 8        | 1.7364  | 1.8446 | 1.7905 | 10.0824 | 9.4865  | 9.7845  |
| 9        | 4.9972  | 2.1074 | 3.5523 | 20.5993 | 12.2364 | 16.4178 |
| 10       | 2.8583  | 5.2744 | 4.0664 | 12.5487 | 24.1370 | 18.3429 |
| 11       | 15.4249 |        |        | 54.9799 |         |         |
| 12       | 3.4915  |        |        | 17.9562 |         |         |
| 13       | 0.9759  |        |        | 4.9717  |         |         |
| 14       | 3.0276  |        |        | 14.3414 |         |         |
| 15       | 7.8885  |        |        | 34.0781 |         |         |
| 16       | 2.3025  |        |        | 10.7187 |         |         |
| 17       | 1.7459  |        |        | 8.4178  |         |         |
| 18       | 7.6546  |        |        | 30.6185 |         |         |
| 19       | 1.9837  |        |        | 9.3146  |         |         |
| 20       | 1.4891  |        |        | 7.6480  |         |         |
| Moyenne  | 4.1260  |        | 3.1706 | 18.0130 |         | 14.8582 |
| Variance | 0.7695  |        | 0.4665 | 11.3355 |         | 8.5585  |

☞ Réduction de  $Var(L_Q)$  de 40%

☞ Réduction de  $Var(W_Q)$  de 25%

## SIMULATION DISCRÈTE ANALYSE DES RÉSULTATS AUTRES MÉTHODES

- ☞ *Méthode des suites communes de nombres aléatoires*  
Comparer plusieurs stratégies ou configuration d'un système. la taille de l'échantillon soit la même pour toutes les stratégies envisagées.
- ☞ La méthode exige que pour chaque source de variation aléatoire dans le modèle, on dispose d'une suite de nombres aléatoires que l'on peut reproduire d'une suite à l'autre.
- ☞ Cette procédure introduit une corrélation positive entre les deux passes de simulation. Soient  $\bar{x}_n$  la variable aléatoire correspondant au paramètre à estimer pour la première stratégie, et  $\bar{x}'_n$  la variable aléatoire associée avec le paramètre pour l'autre stratégie, alors la variance de la différence  $\bar{x}_n - \bar{x}'_n$ 

$$Var(\bar{x}_n - \bar{x}'_n) = Var(\bar{x}_n) + Var(\bar{x}'_n) - 2Cov(\bar{x}_n, \bar{x}'_n)$$
- ☞ La covariance est nulle si les suites de nombres aléatoires sont indépendantes.

Mais si on utilise une même suite pour chacune des stratégies, on doit s'attendre à une corrélation positive. La variance de  $\bar{x}_n - \bar{x}'_n$  utilisant une suite commune de nombres aléatoires est donc en principe inférieure à la variance utilisant des suites indépendantes.

## MÉTHODE RÉGÉNÉRATIVE

- ☞ Méthodes précédentes : une période de chauffe pour éliminer le biais initial des statistiques dû au régime transitoire du système en éliminant un certain nombre d'observations au début de la simulation.
- ☞ La méthode régénérative élimine la période de chauffe en identifiant le point de régénération. Un point de régénération est un état du système dans lequel le comportement futur est indépendant du comportement passé.

## MÉTHODE RÉGÉNÉRATIVE

☞ Considérons, sur la droite réelle, une suite de points  $(S_i)$  tels que :

$$S_i + T_1 + T_2 + \dots + T_i$$

où les  $T_k$  sont des variables indépendantes et identiquement distribuées (i.i.d.) de même loi  $F$  :

$$P(T_1 = t) = \dots = P(T_k = t) = F(t)$$

☞ La séquence  $\{S_i, i = 0\}$  est appelée un processus de renouvellement. La théorie du renouvellement est l'un des outils de base en matière de files d'attente et de fiabilité.

☞ Le processus  $\{Z_t, t = 0\}$  est un processus régénératif si il existe une suite  $(S_i)$ , qui soit un renouvellement (qu'on dira inclus dans  $Z$ ), tel que  $\{Z_t, S_k = t = S_{k+1}\}$  soit une réplique de  $\{Z_t, S_1 = t = S_2\}$ .

☞ Pour une file d'attente G/G/1, la suite des numéros de clients qui arrivent et trouvent le serveur libre est un renouvellement. Le temps d'attente du  $n$ -ième client est un processus régénératif discret, dans lequel le précédent est inclus.

☞ On notera  $\{\beta_k\}$  le renouvellement discret :

$$\beta_k = \beta_{k-1} + t_k, \quad t_k \in \mathbb{N}$$

## MÉTHODE RÉGÉNÉRATIVE

☞ On note  $Z_n, n = 0$  le processus régénératif.

☞ Intérêt de la méthode :

- . Soit une fonction  $f$  définie sur le processus, on cherche à calculer  $E[f(z)]$  : si  $f(x) = x$ , on calcule la moyenne, etc... Le théorème-clé affirme que :

$$E[f(z)] = \frac{E\left[\sum_{i=1}^{t_1} f(z_i)\right]}{E(t_1)}$$

☞ Le théorème affirme qu'on peut restreindre l'étude du phénomène à un seul intervalle de régénération : on somme les mesures sur un intervalle de régénération. L'espérance mathématique divisée par le nombre moyen de mesures dans l'intervalle, donne la moyenne cherchée.

## MÉTHODE RÉGÉNÉRATIVE

### ☞ Application à la simulation

- Soit le processus régénératif  $Z$  représentant le temps d'attente des clients dans une file et on le notera  $W$ .
- $Y_i$  représente les sommes sur chaque intervalle de renouvellement :

$$Y_i = \sum_{j=b_i+1}^{b_i+t_{i+1}} f(W_j)$$

- On peut réécrire :

$$\begin{aligned} E(W) &= \lim_{n \rightarrow \infty} \frac{W_1 + W_2 + \dots + W_n}{n} \\ &= \lim_{k \rightarrow \infty} \frac{Y_1 + Y_2 + \dots + Y_k}{t_1 + t_2 + \dots + t_k} \\ &= \lim_{k \rightarrow \infty} \frac{\frac{Y_1 + Y_2 + \dots + Y_k}{k}}{\frac{t_1 + t_2 + \dots + t_k}{k}} \end{aligned}$$

## MÉTHODE RÉGÉNÉRATIVE

### ☞ Application à la simulation

- Sous des conditions générales :

$$\lim_{k \rightarrow \infty} \frac{\sum_{i=1}^k Y_i}{k} = E(Y_1)$$

$$\lim_{k \rightarrow \infty} \frac{\sum_{i=1}^k t_i}{k} = E(t_1)$$

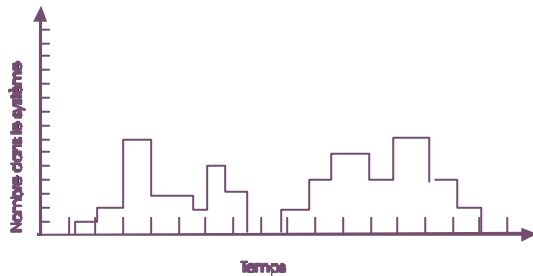
- Le théorème affirme simplement que la limite du quotient est égale au quotient des limites :

$$E(W) = \lim_{k \rightarrow \infty} \frac{\frac{\sum_{i=1}^k Y_i}{k}}{\frac{\sum_{i=1}^k t_i}{k}} = \frac{E(Y_1)}{E(t_1)}$$

## MÉTHODE RÉGÉNÉRATIVE

### ☞ Application à la simulation

- On détecte les points de régénération, et on collecte les  $Y_i$  successifs.
- L'expérience est conduite pour  $n$  intervalles de régénération successifs.
- À la différence de la méthode des lots, les  $Y_i$  sont indépendants les uns des autres – par construction, c'est l'intérêt de la méthode.
- Les cycles successifs n'ont pas la même longueur (ne servent pas le même nombre de clients)



## MÉTHODE RÉGÉNÉRATIVE

### ☞ Exemples

| Système                         | Z  | Y  | T  |
|---------------------------------|--|--|--|
| File d'attente                  | Temps moyen d'attente des clients $W_Q$    | Temps total d'attente de tous les clients durant le cycle<br>$\sum_{j=b_i+1}^{b_i+n_i} W_i^Q$      | Nombre de clients arrivant dans le cycle : $n$ |
| File d'attente                  | Nombre moyen de clients dans la file $L_Q$ | Intégral sur le temps du nombre de clients dans la file<br>$\sum_{j=b_i+1}^{b_i+n_i} t_j L_Q(t_j)$ | Durée du cycle $t_i$                           |
| Inventaire d'une banque de sang | Moyenne des stocks                         | Intégral sur le temps du niveau de stocks  | Durée du cycle $t_i$                           |

Rappel :  $t_i$  désigne le nombre de mesures effectuées dans le  $i$ -ième intervalle de renouvellement.

La séquence  $\{Y_i, t_i\}$  est une suite de vecteurs i.i.d. (indépendants et identiquement distribués).

## MÉTHODE RÉGÉNÉRATIVE

☞ Estimé par intervalle

Soit  $r$  la vraie valeur du quotient :

$$r = \frac{E(Y_1)}{E(t_1)}$$

La simulation nous donne de cette quantité un estimateur

$$\hat{r} = \frac{\sum_{k=1}^N Y_k}{\sum_{k=1}^N t_k}$$

Pour trouver un intervalle de confiance de  $r$ , on doit déterminer la variance de  $r$ . Pour trouver  $Var(r)$ , définissons  $V = Y - r t$ ,

alors la suite  $\{V_j = Y_j - r t_j\}$  est une suite de variable indépendantes i.i.d., de moyenne nulle  $E(V) = 0$ . On montre que :

$$\begin{aligned} s^2 &= Var(V) \\ &= Var(Y) - 2rCov(Y, t) + r^2 Var(t) \\ &= E(Y^2) - 2rE(Yt) + r^2 E(t^2) \end{aligned}$$

## MÉTHODE RÉGÉNÉRATIVE

☞ Estimé par intervalle

Considérons la moyenne empirique

$$\bar{V} = \frac{\sum_{k=1}^N V_k}{N} = \bar{Y} - r \bar{t}$$

Le théorème de la limite centrale nous apprend que  $\bar{V}$  tend vers une loi normale, dont la variance est  $s^2/N$  ( $N$  est le nombre total de mesures), ce qui se traduit en :

$$\lim_{N \rightarrow \infty} P \left( \frac{\sum_{k=1}^N Y_k - r \sum_{k=1}^N t_k}{N} \leq x \right) = \Phi(x)$$

En introduisant l'estimateur  $\hat{r}$ , et  $\bar{t} = \left( \sum_{k=1}^N t_k \right) / N$ , on obtient :

$$P \left( \left( \hat{r} - r \right) \frac{\bar{t} \sqrt{N}}{s_V} \leq x \right) = \Phi(x)$$

## MÉTHODE RÉGÉNÉRATIVE

☞ Estimé par intervalle

D'où on obtient un intervalle de confiance de niveau  $100(1 - \alpha) \%$  sur le quotient :

$$P\left(\hat{r} - \frac{Z_{\alpha/2} S_V}{t\sqrt{N}} \leq E(W) \leq \hat{r} + \frac{Z_{\alpha/2} S_V}{t\sqrt{N}}\right) = 100(1 - \alpha) \%$$

## MÉTHODE RÉGÉNÉRATIVE

☞ **Résumé**

- Soit  $N$  cycles de régénération
- Soit  $t_k$  : le nombre de mesures collectées au cours du cycle numéro  $k$  (le nombre de clients du cycle, dans l'exemple),
- Soit  $Y_k$  : le cumul des mesures du cycle  $k$  (c'est-à-dire la somme des temps - noter que le premier temps d'attente)

$$\bullet Y1 = \sum_{k=1}^N Y_k$$

$$\bullet Y2 = \sum_{k=1}^N (Y_k)^2$$

$$\bullet T1 = \sum_{k=1}^N t_k$$

$$\bullet T2 = \sum_{k=1}^N (t_k)^2$$

$$\bullet YT = \sum_{k=1}^N Y_k t_k$$

## MÉTHODE RÉGÉNÉRATIVE

### ☞ Résumé

- À la fin de la simulation,  $N$  cycles ayant été observés, on calcule :

$$\bar{W} = Y1/T1$$

$$V = \frac{(Y2 + T2 \times \bar{W}^2 - 2\bar{W} \times YT)}{N - 1}$$

- Quand la longueur de simulation augmente,  $V$  doit fluctuer autour d'une valeur constante. La demi-largeur ou l'imprécision ?  $W$  est donnée par :

$$\Delta W = z_{\alpha/2} \sqrt{\frac{NV}{T1}}$$

- et pour un risque  $\alpha$ , l'intervalle de confiance est :

$$\bar{W} - \Delta W \leq E(w) \leq \bar{W} + \Delta W$$

## MÉTHODE RÉGÉNÉRATIVE

### ☞ Exemple

| Méthode de lots                          | Méthode régénérative               |  |  |
|--|------------------------------------|--|--|
| pour $2 \times 10^6$ clients<br>200 lots | $10^4$ cycles<br>(100 860 clients) | $10^5$ cycles,<br>soit 991 000 clients | $2 \times 10^5$ cycles, soit environ $2 \times 10^6$ clients |
| Moyenne : 8.82<br>Écart-Type : 3.03      |                                    |  |  |
| Intervalle de confiance<br>(8.82 ± 0.42) | (9.56 ± 1.26)                      | (8.92 ± 0.37)                          | (8.9 ± 0.26 )  |

### ☞ Difficultés

- détecter des points de régénération, pas trop éloignés
- si le système comprend plusieurs files d'attente, il faut trouver un point commun ou un point de régénération à toutes les files

*INF4701 Simulation par événements discrets*

9- 109

IF505 Simulation par événements discrets-- L.Granger--SIM4-109--2000-11-23-10:46