

# Guide de survie en actuariat

Gabriel Crépeault-Cauchon et Nicholas Langevin

19 septembre 2019

# Table des matières

<b>I</b>	<b>Fondements mathématiques utiles</b>	<b>4</b>
<b>1</b>	<b>Calculs</b>	<b>5</b>
1.1	Règles de dérivation . . . . .	5
<b>2</b>	<b>Algèbre linéaire</b>	<b>6</b>
2.1	Définition d'un vecteur et une matrice . . . . .	6
2.2	Matrice transposée . . . . .	8
2.3	Opérations matricielles . . . . .	8
2.4	Trace, déterminant et matrice inverse . . . . .	9
2.4.1	Trace d'une matrice . . . . .	9
2.4.2	Déterminant d'une matrice . . . . .	9
2.4.3	Matrice inverse . . . . .	10
2.5	Décomposition LDU de Choleski . . . . .	10
2.6	Vecteurs et valeurs propres . . . . .	10
2.6.1	Définition . . . . .	10
2.6.2	Propriétés intéressantes . . . . .	10
2.6.3	Décomposition spectrale . . . . .	11
2.7	Dérivées de matrice ou vecteurs . . . . .	11
<b>II</b>	<b>Matière vue dans le baccalauréat en actuariat</b>	<b>12</b>
<b>3</b>	<b>Probabilités et statistiques</b>	<b>13</b>
3.1	Concepts de probabilité de base . . . . .	13
3.1.1	Probabilité conditionnelle . . . . .	13
3.1.2	Théorème de Bayes . . . . .	14
3.2	Définition d'une variable aléatoire . . . . .	14
3.3	Distribution d'une variable aléatoire . . . . .	14

3.4	Moments et quantités importantes . . . . .	14
3.5	Distribution de probabilité qui reviennent souvent . . . . .	15
<b>4</b>	<b>Mathématiques financières</b>	<b>16</b>
<b>5</b>	<b>Processus aléatoire</b>	<b>17</b>
5.1	Chaîne de Markov . . . . .	17
<b>III</b>	<b>Matière pour les examens professionnels</b>	<b>18</b>
<b>6</b>	<b>Probability Models</b>	<b>19</b>
6.1	Reliability Theory . . . . .	20
6.1.1	Struture Functions . . . . .	20
<b>7</b>	<b>Time Series</b>	<b>21</b>
7.1	Base Models . . . . .	22
7.1.1	Estimating Trends . . . . .	22
7.1.2	Estimating Seasonal Effect . . . . .	23
7.1.3	Smoothing Procedure . . . . .	23
7.2	Correlation . . . . .	24
7.2.1	The Correlogram . . . . .	26
7.2.2	Covariance of sums of random variables . . . . .	27
7.3	Forecasting Strategies . . . . .	28
7.3.1	Relationships of different time serie . . . . .	28
7.4	Basic Stochastic Models . . . . .	29
7.4.1	White Noise . . . . .	29
7.4.2	Random Walks . . . . .	30
7.4.3	Autoregressive Models . . . . .	31
7.5	Regression . . . . .	33
7.5.1	Linear Models . . . . .	33
7.5.2	Linear Models with Seasonal Variables . . . . .	34
7.5.3	Forecasting from regression . . . . .	34
7.6	Stationary Models . . . . .	36
7.6.1	Strictly Stationary Series . . . . .	36
7.6.2	Moving average models . . . . .	37
7.6.3	Mized Models : The ARMA process . . . . .	38
7.7	Non-stationary Models . . . . .	39

7.7.1	Differencing . . . . .	39
7.7.2	Non-Seasonal ARIMA Models . . . . .	40
7.7.3	Seasonal ARIMA models . . . . .	40
<b>8</b>	<b>Extended Linear Models</b>	<b>41</b>
8.1	Inference . . . . .	42
8.1.1	Sampling distribution for the score statistic . . . . .	42
8.2	Normal Linear Models . . . . .	43
8.2.1	Basis Result . . . . .	43
<b>A</b>	<b>Principales distribution de probabilité utilisées</b>	<b>44</b>
<b>B</b>	<b>Preuves</b>	<b>45</b>
B.1	Stop-Loss ( $\pi_X(d)$ ) . . . . .	45
B.2	TVaR . . . . .	47
B.2.1	Les 3 formes explicites de la <i>TVaR</i> . . . . .	47
B.3	Sous-additivité de la <i>TVaR</i> . . . . .	48
B.3.1	À l'aide de la fonction convexe $\varphi(x)$ . . . . .	49
B.3.2	Avec les fonctions indicatrices . . . . .	51
B.3.3	À l'aide des statistiques d'ordre . . . . .	53
B.4	Loi des grands nombres . . . . .	53
B.5	Somme de v.a. indépendantes d'une loi Poisson Composée . . . . .	54
B.6	Théorème d'Euler . . . . .	56
B.7	Dérivée de l'écart-type (générale) . . . . .	57
B.8	Distribution limite de $W_n$ . . . . .	60

Première partie

Fondements mathématiques  
utiles

# Chapitre 1

## Calculs

### 1.1 Règles de dérivation

Dans le tableau, on utilise  $k \in \mathbb{R}$  ou  $n$  pour parler d'une constante,  $u$ ,  $v$  ou  $w$  pour parler d'une fonction.

Fonction	Dérivée
$f(x) = k$	$f'(x) = 0$
$f(x) = kx$	$f'(x) = k$
$f(x) = x^n$	$f'(x) = nx^{n-1}$
$f(x) = kg(x)$	$f'(x) = kg'(x)$
$f(x) = g(x) \pm h(x)$	$f'(x) = g'(x) \pm h'(x)$
$f(x) = g(x) \cdot h(x)$	$f'(x) = g'(x) \cdot h(x) + g(x) \cdot h'(x)$
$f(x) = \frac{g(x)}{h(x)}$	$f'(x) = \frac{g'(x)h(x) - g(x)h'(x)}{h(x)^2}$
$f(x) = g(x)^n$	$f'(x) = n \cdot g(x)^{n-1} \cdot g'(x)$
$f(x) = k^{g(x)}$	$f'(x) = k^{g(x)} \ln k \cdot g'(x)$
$f(x) = e^{g(x)}$	$f'(x) = e^{g(x)} \cdot g'(x)$
$f(x) = \ln(g(x))$	$f'(x) = \frac{g'(x)}{g(x)}$

# Chapitre 2

## Algèbre linéaire

### 2.1 Définition d'un vecteur et une matrice

**Vecteur ligne** Un vecteur ligne  $\mathbf{x}$  est un vecteur de dimension  $p \times 1$ , tel que

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \dots \\ x_p \end{bmatrix}$$

**Matrice** Une matrice  $\mathbf{A} = [a_{ij}]_{m \times n}$  de dimension  $m$  lignes par  $n$  colonnes, définie telle que

$$\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \dots & \dots & \ddots & \dots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{bmatrix} \quad (2.1)$$

**Matrice carrée** Une matrice carrée  $\mathbf{A}$  de dimensions  $m \times m$  a autant de lignes que de colonnes.

**non-négative**  $\mathbf{A}$  est définie comme *non-négative* si  $\mathbf{x}^\top \mathbf{A} \mathbf{x} \geq 0, \forall \mathbf{x} \in \mathbb{R}^n$ .

**positive**  $\mathbf{A}$  est définie comme *positive* si  $\mathbf{x}^\top \mathbf{A} \mathbf{x} > 0, \forall \mathbf{x} \neq 0$ .

**semi-positive**  $\mathbf{A}$  est définie comme non-négative, mais elle n'est pas définie positive.

**Orthogonale**  $\mathbf{A}$  est *orthogonale* si elle est non-singulière et  $\mathbf{A}^{-1} = \mathbf{A}^\top$  (voir [sous-section 2.4.3](#) pour définition de  $\mathbf{A}^{-1}$ )

**Matrice symétrique** La matrice  $\mathbf{A}$  est symétrique si  $a_{ij} = a_{ji} \forall i, j$ , i.e

$$\mathbf{A} = \begin{bmatrix} 1 & 2 & 3 \\ 2 & 1 & 4 \\ 3 & 4 & 1 \end{bmatrix}$$

**Matrice triangulaire (inférieure ou supérieure)** Une matrice inférieure  $\mathbf{L}$  est constituée de 0 en dessous de la diagonale :

$$\mathbf{L} = \begin{bmatrix} 1 & 2 & 3 \\ 0 & 1 & 4 \\ 0 & 0 & 1 \end{bmatrix}$$

À l'inverse, on peut aussi avoir une matrice triangulaire supérieure  $\mathbf{U}$ , où les éléments en haut de la diagonale sont tous égaux à 0 :

$$\mathbf{U} = \begin{bmatrix} 1 & 0 & 0 \\ 8 & 3 & 0 \\ 2 & 3 & 9 \end{bmatrix}$$

**Matrice diagonale** Une matrice diagonale  $\mathbf{D}$  a des éléments  $d_{ii} > 0$  sur sa diagonale seulement. Cette matrice est à la fois triangulaire inférieure et supérieure. i.e.

$$\mathbf{D} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & 2 \end{bmatrix}$$

Un cas spécial de la matrice diagonale est la matrice identité  $\mathbf{I}$ , où  $\mathbf{I}_{ii} = 1, \forall i$ , i.e

$$\mathbf{I} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad (2.2)$$



**Matrice diagonalisable** Une matrice  $\mathbf{A}_{n \times n}$  est dite *diagonalisable* s'il existe une matrice carrée  $\mathbf{Q}_{n \times n}$  inversible (ou non-singulière) et une matrice  $\mathbf{D}$  diagonale telle que

$$\mathbf{Q}^{-1}\mathbf{A}\mathbf{Q} = \mathbf{D} \leftrightarrow \mathbf{A} = \mathbf{Q}\mathbf{D}\mathbf{Q}^{-1} \quad (2.3)$$

(Théorème sur les matrices symétriques) : Toute matrice carrée symétrique est diagonalisable par une matrice orthogonale  $\mathbf{Q}$ .

## 2.2 Matrice transposée

Soit la matrice  $\mathbf{A}$  définie en (2.1). On peut trouver la matrice transposée  $\mathbf{A}^\top$ , où  $[a_{ij}] = [a_{ji}]$ . **En d'autres mots, les lignes deviennent des colonnes.** Voici quelques propriétés intéressantes avec les matrices transposées :

- $(\mathbf{A}^\top)^\top = \mathbf{A}$
- $(\mathbf{A} + \mathbf{B})^\top = \mathbf{A}^\top + \mathbf{B}^\top$
- $(k\mathbf{A})^\top = k\mathbf{A}^\top$
- $(\mathbf{AB})^\top = \mathbf{B}^\top \mathbf{A}^\top$
- $\mathbf{A}^\top \mathbf{A}$  et  $\mathbf{AA}^\top$  sont symétriques.

## 2.3 Opérations matricielles

Voici une liste non-exhaustive des opérations matricielles possibles. Côté notation,  $A$  et  $B$  représente des matrices,  $c$  représente une constante

- $\mathbf{A} + \mathbf{B} = [a_{ij} + b_{ij}]$
- $\mathbf{A} - \mathbf{B} = [a_{ij} - b_{ij}]$
- $c\mathbf{A} = [ca_{ij}]$
- Produit matriciel :

$$\mathbf{AB} = \left[ \sum_{k=1}^p a_{ik} b_{kj} \right]_{i \times j} \quad (2.4)$$

, avec  $\mathbf{A} = [a_{ip}]$  et  $\mathbf{B} = [b_{pj}]$

- $\mathbf{A}(\mathbf{B} + \mathbf{C}) = \mathbf{AB} + \mathbf{AC}$
- $\mathbf{A}^{-1}\mathbf{A} = \mathbf{I} = \mathbf{AA}^{-1}$ , où  $\mathbf{I}$  est la matrice identité (voir [Équation 2.2](#)) et  $\mathbf{A}^{-1}$  est la matrice inverse de  $\mathbf{A}$  (voir [sous-section 2.4.3](#) au besoin)
- $(\mathbf{AB})^{-1} = \mathbf{B}^{-1}\mathbf{A}^{-1}$

## 2.4 Trace, déterminant et matrice inverse

### 2.4.1 Trace d'une matrice

Soit la matrice carrée  $\mathbf{A}$ . On peut trouver la trace de cette matrice en sommant les éléments de sa diagonale, i.e.

$$\text{Tr}(\mathbf{A}) = \sum_{i=1}^n a_{ii} \quad (2.5)$$

**Propriétés de la trace d'une matrice**

- $\text{Tr}(\mathbf{A} + \mathbf{B}) = \text{Tr}(\mathbf{A}) + \text{Tr}(\mathbf{B})$
- $\text{Tr}(\mathbf{AB}) = \text{Tr}(\mathbf{BA})$  et  $\text{Tr}(\mathbf{ABC}) = \text{Tr}(\mathbf{CAB}) = \text{Tr}(\mathbf{BCA})$

### 2.4.2 Déterminant d'une matrice

Soit la matrice carrée  $\mathbf{A}$ . On peut trouver le déterminant de  $\mathbf{A}$ , noté  $\det(\mathbf{A})$  ou  $|\mathbf{A}|$ , avec

$$\det(\mathbf{A}) = \begin{vmatrix} a & b \\ c & d \end{vmatrix} = ad - bc \quad (2.6)$$

De façon générale, lorsque les dimensions de la matrice carrée sont supérieures à 2, on a

$$\det(A) = \sum_{j=1}^n a_{ij} C_{ij} \quad (2.7)$$

avec  $1 \leq i \leq n$  où  $C_{ij} = (-1)^{i+j} M_{ij}$  et  $M_{ij}$  est le déterminant de la nouvelle matrice en enlevant la ligne  $i$  et la colonne  $j$ .

Si la matrice  $\mathbf{A}$  est inversible (ou non-singulière, voir la [sous-section 2.4.3](#)), alors le déterminant aura les propriétés suivantes :

- $\det(A^\top) = \det(A)$
- $\det(kA) = k^n \det(A)$
- $\det(A + B) \neq \det(A) + \det(B)$
- $\det(AB) = \det(A) \det(B)$
- $\det(A^{-1}) = \frac{1}{\det(AB)} = \det(A)^{-1}$

### 2.4.3 Matrice inverse

Soit la matrice carrée  $\mathbf{A}$ . On peut trouver la matrice inverse  $\mathbf{A}^{-1}$  telle que

$$\mathbf{A}^{-1} = \frac{1}{\det(\mathbf{A})} \text{Adj}(\mathbf{A}) \quad (2.8)$$

où  $\text{Adj}(\mathbf{A}) = [C_{ij}]_{m \times n}^T$  et  $C_{ij} = (-1)^{i+j} M_{ij}$ .

## 2.5 Décomposition LDU de Choleski

Soit  $\mathbf{A}$  une matrice carrée symétrique définie positive. Alors, il existe une décomposition unique telle que

$$\mathbf{A} = \mathbf{L}\mathbf{D}\mathbf{U} \quad (2.9)$$

où  $\mathbf{L}$ ,  $\mathbf{D}$ ,  $\mathbf{U}$  sont respectivement des matrices triangulaire inférieure, triangulaire supérieure et diagonale.

Cette décomposition peut être fortement utile en programmation lorsqu'on fait des opérations sur des matrices, afin de limiter le nombre d'opérations.

## 2.6 Vecteurs et valeurs propres

### 2.6.1 Définition

Soit  $\mathbf{A}$  une matrice carrée. On dit que  $\lambda$  est une *valeur propre* de  $\mathbf{A}$  s'il existe un vecteur  $\mathbf{x} \neq 0$  tel que

$$\mathbf{A}\mathbf{x} = \lambda\mathbf{x} \quad (2.10)$$

On appelle le vecteur  $\mathbf{x}$  un *vecteur propre* correspondant à la valeur propre  $\lambda$ . De plus, l'ensemble des nombres réels  $\lambda$  satisfaisant l'Équation 2.10 est appelé *spectre* de la matrice  $\mathbf{A}$ .

### 2.6.2 Propriétés intéressantes

Les vecteurs propres et valeurs propres permettent d'avoir plusieurs propriétés appréciables, notamment :

- Si  $\mathbf{x}$  est un vecteur propre de  $\mathbf{A}$  correspondant à la valeur propre  $\lambda$ , alors  $c\mathbf{x}$  sera également un vecteur propre de  $\mathbf{A}$  correspondant à  $\lambda$ .

- Si  $\mathbf{A}$  est symétrique et  $\mathbf{x}_1$  et  $\mathbf{x}_2$  sont des vecteurs propres correspondant à des valeurs propres différentes de  $\mathbf{A}$ , alors  $\mathbf{x}_1$  et  $\mathbf{x}_2$  sont des vecteurs orthogonaux, i.e.  $\mathbf{x}_1^\top \mathbf{x}_2 = 0$ .
- Si  $\mathbf{A}$  a les valeurs propres (pas nécessairement distinctes)  $\lambda_1, \dots, \lambda_n$ , alors  $\det(\mathbf{A}) = \prod_{i=1}^n \lambda_i$  et  $\text{Tr}(\mathbf{A}) = \sum_{i=1}^n \lambda_i$ .

### 2.6.3 Décomposition spectrale

Soit  $\mathbf{A}_{n \times n}$  une matrice symétrique avec les  $n$  valeurs propres  $\lambda_1, \dots, \lambda_n$ . Il existe une matrice orthogonale  $\mathbf{Q}$  telle que

$$\mathbf{A} = \mathbf{Q} \mathbf{\Lambda} \mathbf{Q}^\top \quad (2.11)$$

avec  $\mathbf{\Lambda} = \text{diag}(\lambda_1, \dots, \lambda_n)$ . Cette décomposition est fort utile lorsqu'on veut faire des produits matriciels successifs de la même matrice (appliqué directement dans les chaînes de Markov, voir [section 5.1](#)) :

$$\begin{aligned} \mathbf{A} \mathbf{A} &= \mathbf{Q} \mathbf{\Lambda} \underbrace{\mathbf{Q}^\top \mathbf{Q}}_{=\mathbf{I}} \mathbf{\Lambda} \mathbf{Q}^\top \\ &= \mathbf{Q} \mathbf{\Lambda}^2 \mathbf{Q}^\top \end{aligned}$$

## 2.7 Dérivées de matrice ou vecteurs

Voici quelques entités pratiques :

$$\frac{\partial}{\partial \mathbf{v}} \mathbf{w}^\top \mathbf{v} = w$$

$$\frac{\partial}{\partial \mathbf{v}} = \mathbf{v}^\top \mathbf{A} \mathbf{v} = (\mathbf{A} + \mathbf{A}^\top) \mathbf{v}$$

## Deuxième partie

Matière vue dans le  
baccalauréat en actuariat

# Chapitre 3

## Probabilités et statistiques

### 3.1 Concepts de probabilité de base

#### 3.1.1 Probabilité conditionnelle

Définition de base

$$\Pr(A|B) = \frac{\Pr(A \cap B)}{\Pr(B)} \quad (3.1)$$

**Loi des probabilités totales** Soit  $E_i$  le *outcome*  $i$  parmi l'ensemble des  $n$  *outcome* possibles de l'évènement  $E$ , alors, on peut représenter la probabilité que l'évènement  $A$  survienne comme

$$\Pr(A) = \sum_{i=1}^n \Pr(A|E_i) \Pr(E_i) \quad (3.2)$$

avec  $\sum_{i=1}^n \Pr(E_i) = 1$ .

**Relation importante** de l'Équation 3.1, on peut représenter  $\Pr(A|B)$  comme

$$\Pr(A|B) = \frac{\Pr(B|A) \Pr(A)}{\Pr(B)} \quad (3.3)$$

### 3.1.2 Théorème de Bayes

En combinant l'Équation 3.3 et la loi des probabilités totales (l'Équation 3.2), on obtient le théorème de Bayes :

$$\Pr(A|B) = \frac{\Pr(B|A) \Pr(A)}{\sum_{i=1}^n \Pr(B|A_i) \Pr(A_i)} \quad (3.4)$$

## 3.2 Définition d'une variable aléatoire

## 3.3 Distribution d'une variable aléatoire

Fonction de densité, répartition, survie, hazard rate, etc.

## 3.4 Moments et quantités importantes

Espérance, variance, covariance, coefficient de variation, corrélation

**Espérance** Soit une v.a.  $X$  (continue ou discrète). Son espérance est définie telle que

$$\mathbb{E}[X] = \mu = \sum_{x=0}^{\infty} x \Pr(X = x) = \int_0^{\infty} x f_X(x) dx \quad (3.5)$$

L'espérance d'une fonction de la v.a  $X$  est

$$\mathbb{E}[g(X)] = \sum_{x=0}^{\infty} g(x) \Pr(X = x) = \int_0^{\infty} g(x) f_X(x) dx \quad (3.6)$$

**Variance**

$$\text{Var}(X) = \sigma^2 = \mathbb{E}[(X - \mathbb{E}[X])^2] = \mathbb{E}[X^2] - \mathbb{E}[X]^2 \quad (3.7)$$

quelques propriétés à savoir :

$$\begin{aligned} \text{Var}(aX) &= a^2 \text{Var}(X) \\ \text{Var}(X + b) &= \text{Var}(X) \end{aligned}$$

### Covariance

$$\text{Cov}(X, Y) = \sigma_{X,Y} = E[(X - E[X])(Y - E[Y])] = E[XY] - E[X]E[Y] \quad (3.8)$$

## 3.5 Distribution de probabilité qui reviennent souvent

Un tableau récapitulatif des différentes distribution de probabilité est disponible à l'



## Chapitre 4

# Mathématiques financières

# Chapitre 5

## Processus aléatoire

### 5.1 Chaîne de Markov

## Troisième partie

### Matière pour les examens professionnels

# Chapitre 6

## Probability Models

This section resume the *Introduction to probability models*

## 6.1 Reliability Theory

### 6.1.1 Struture Functions

**State Vector** The state vector is difine as  $\mathbf{x} = (x_1, \dots, x_n)$ , where  $x_i = 1$  if the component is functioning, 0 otherwise.

**Structure function** The structure function  $\phi(\mathbf{x}) = 1$  if the system is fonctionning when the state vector is  $\mathbf{x}$ .

**Serie Structure** A series system functions if and only if all of its compo-nents are functioning.

$$\phi(\mathbf{x}) = \min(x_1, \dots, x_n) = \prod_{i=1}^n x_i$$

**Parallel Struture** A parallel system functions if and only if at least one of its components is functioning.

$$\phi(\mathbf{x}) = \max(x_1, \dots, x_n) = 1 - \prod_{i=1}^n (1 - x_i)$$

**k-out-of-n Struture** The series and parallel systems are both special cases of a k-out-of-n system. Such a system functions if and only if at least k of the n components are functioning.

$$\phi(\mathbf{x}) = \begin{cases} 1, & \text{if } \sum_{i=1}^n x_i \geq k \\ 0, & \text{if } \sum_{i=1}^n x_i < k \end{cases}$$

#### Note

(If and only if) mean  $\min(x_1, \dots, x_n) = \prod x_i$

(At least one) mean  $\max(x_1, \dots, x_n) = 1 - \prod(1 - x_i)$

# Chapitre 7

## Time Series

This section resume the Chapters 1-5 (excluding Sections 3.3 and 3.4), 6, 7 (Sections 7.1, 7.2 and 7.3) of *An Introductory Time Series with R*

A mettre dans la bio : Cowpertwait, P. and Metcalfe, A., *Introductory Time Series with R*, Springer, 2009.

**Trend** In general, a systematic change in a time series that does not appear to be periodic is known as a trend.

**Seasonal Variation** A repeating pattern within any fixed period.

**Notation** We represent a time series of length  $n$  by  $x_t : t = 1, \dots, n = x_1, x_2, \dots, x_n$

**Forecast** A forecast  $\hat{x}_{t+k|t}$  is a predicted future value, and the number of time steps into the future is the **lead time (k)**

## 7.1 Base Models

**Additive Decomposition** The additive decomposition model is given by

$$x_t = m_t + s_t + z_t$$

where  $t$  is the time,  $x_t$  the observed series,  $m_t$  the trend,  $s_t$  the seasonal effect and  $z_t$  the error term.

**Multiplication Model** If the seasonal effect tends to increase as the trend inscrease, we use a multiplication model define as

$$x_t = m_t \cdot s_t + z_t$$

where  $t$  is the time,  $x_t$  the observed series,  $m_t$  the trend,  $s_t$  the seasonal effect and  $z_t$  the error term.

### 7.1.1 Estimating Trends

**Centered Moving Average** A moving average is an average of a specified number of time series values around each value in the time series, with the

exception of the first few and last few terms. The length of the moving average is chosen to average out the seasonal effects, which can be estimated

$$\hat{m}_t = \frac{0.5m_{t-k} + m_{t-k-1} + \dots + m_t + \dots + m_{t+k-1} + 0.5m_{t+k}}{2k}$$

### 7.1.2 Estimating Seasonal Effect

**Additive Effect** An estimate of the monthly additive effect ( $s_t$ ) at time  $t$  is define by

$$\hat{s}_t = x_t - \hat{m}_t$$

It is usual to adjust those estimate in order that the sum of one period of the time serie equal zero. Let  $c$  be that adjustment in order to solve this expression.

$$\sum (s_t + c) = 0$$

**Multiplicative Effect** An estimate of the monthly multiplicative effect ( $s_t$ ) at time  $t$  is define by

$$\hat{s}_t = \frac{x_t}{\hat{m}_t}$$

And we found the ajustment  $c$  in order to solve this expression.

$$\sum \frac{(\hat{s}_t + c)}{n} = 1$$

### 7.1.3 Smoothing Procedure

Smoothing procedures can, and usually do, use points before and after the time at which the smoothed estimate is to be calculated. A consequence is that the smoothed series will have some points missing at the beginning and the end unless the smoothing algorithm is adapted for the end points.

Ex.1 The centering moving average is an exemple of a *smoothing* procedure.

Ex.2 The *loess* technique is also a *smoothing* that use a locally weighted regression.



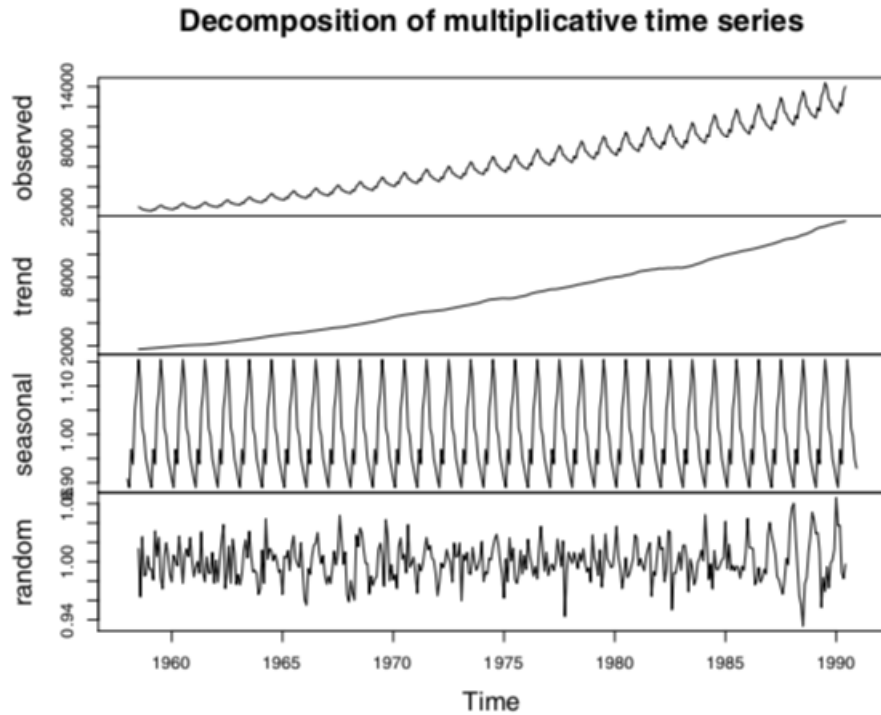


FIGURE 7.1 – In this example, the multiplicative model would seem more appropriate than the additive model because the variance of the original series and trend increase with time

## 7.2 Correlation

**Mean Function** The mean function is, in general, a function of  $t$  and it define as

$$\mu(t) = E[x_t]$$

**Sample Mean** The sample mains is define as

$$\bar{x} = \sum \frac{x_i}{n}$$

**Variance Function** The variance function of a time series model that is stationary in the mean is

$$\sigma^2(t) = E[(x_t - \mu^2)]$$

**Sample Variance** If the model is stationary in the variance, we can estimate the variance with the sample variance define as

$$\text{Var}(x) = \frac{\sum (x_t - \bar{x})^2}{n - 1}$$

**Stationarity** If the mean function is constant, we say that the time series model is stationary in the mean. The time serie can also be stationary in the variance, if the variance function is constant  $\sigma^2$ .

**Ergodic Serie** A time series model that is stationary in the mean is ergodic in the mean if the time average for a single time series tends to the ensemble mean as the length of the time series increases.

$$\lim_{n \rightarrow \infty} \bar{x} = \mu$$

**Covariance** The covariance measure the *linear association* between two random variables. The covariance is define as

$$\gamma(x, y) = \text{E}[(x - \mu_x)(y - \mu_y)]$$

**Sample Covariance** We can estimate the covariance with the sample covariance define as

$$\text{Cov}(x, y) = \sum \frac{(x_i - \bar{x})(y_i - \bar{y})}{n - 1}$$

**Correlation** Correlation is a dimensionless measure of the linear association between a pair of variables (x,y) and is obtained by standardising the covariance.

$$\rho(x, y) = \frac{\text{E}[(x - \mu_x)(y - \mu_y)]}{\sigma_x \sigma_y} = \frac{\gamma(x, y)}{\sigma_x \sigma_y}$$

**Sample Correlation** We can estimate the covariance with the sample covariance define as

$$\text{Cor}(x, y) = \frac{\text{Cov}(x, y)}{\text{sd}(x)\text{sd}(y)}$$

**Autocovariance (acvf)** If a time series model is second-order stationary, we can define an autocovariance function (acvf),  $\gamma_k$ , as a function of the lag  $k$ .

$$\gamma_k = E[(x_t - \mu)(x_{t+k} - \mu)]$$

**Sample Autocovariance** The acvf function can be estimated by

$$c_k = \frac{1}{n} \sum_{t=1}^{n-k} (x_t - \bar{x})(x_{t+k} - \bar{x})$$

#### Note

The sample autocovariance at lag 0,  $c_0$ , is the variance calculated with a denominator  $n$ .

**Autocorrelation (acf)** The lag  $k$  autocorrelation function (acf),  $\rho_k$ , is defined by

$$\rho_k = \frac{\gamma_k}{\sigma^2}$$

where  $\rho_0 = 1$

**Sample Autocorrelation** The acf function can be estimated by

$$r_k = \frac{c_k}{c_0}$$

**Second-order Stationary** Suppose a time series that is stationary in the mean and the variance. Then, the time series model is second-order stationary if the correlation between variable depends only on the number of time steps separating them.

### 7.2.1 The Correlogram

The correlogram is plot of  $r_k$  against  $k$ . The main use of the correlogram is to detect autocorrelations in the time series after we have removed an estimate of the trend and seasonal variation.

- The x-axis gives the lag ( $k$ ) and the y-axis gives the autocorrelation ( $\rho_k$ ) at each lag. The unit of lag is the sampling interval. Correlation is dimensionless, so there is no unit for the y-axis.
- The dotted lines on the correlogram are drawn at

$$-\frac{1}{n} \pm \frac{2}{\sqrt{n}}$$

If the sample  $\text{acf}(r_k)$  falls outside these line, we have evidence against the null hypothesis that  $p_k = 0$  at the 5% level.

- The lag at 0 is always.

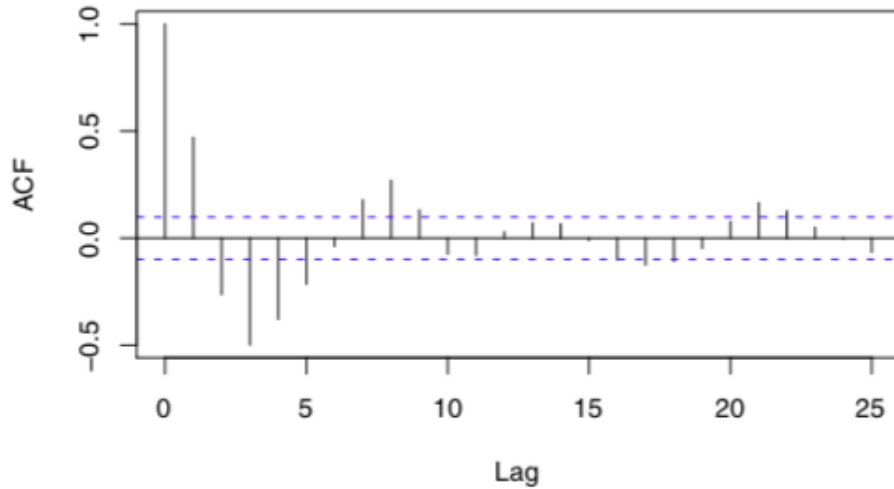


FIGURE 7.2 – Correlogram

### 7.2.2 Covariance of sums of random variables

Let  $x_1, x_2, \dots, x_n$  and  $y_1, y_2, \dots, y_m$  be random variables. Then

$$\text{Cov} \left( \sum_{i=1}^n x_i, \sum_{j=1}^m y_j \right) = \sum_{i=1}^n \sum_{j=1}^m \text{Cov}(x_i, y_j)$$

The result tells us that the covariance of two sums of variables is the sum of all possible covariance pairs of the variables.

## 7.3 Forecasting Strategies

### 7.3.1 Relationships of different time serie

**Cross-covariance (ccvf)** The cross-correlation (ccvf) between two time series is define as

$$\gamma_k(x, y) = E[(x_{t+k} - \mu_x)(y_t - \mu_y)]$$

#### Note

This is not a symmetric relationship, and the variable  $x$  is lagging variable  $y$  by  $k$ .

$$\gamma_k(x, y) = \gamma_{-k}(y, x)$$

**Sample Cross-covariance** The ccvf function can be estimed by

$$c_k(x, y) = \frac{1}{n} \sum_{t=1}^{n-k} (x_{t+k} - \bar{x})(y_t - \bar{y})$$

**Cross-correlation (ccf)** The cross-correlation (ccf) between two time serie is define as

$$\rho_k(x, y) = \frac{\gamma_k(x, y)}{\sigma_x \sigma_y}$$

#### Note

This is not a symmetric relationship, and the variable  $x$  is lagging variable  $y$  by  $k$ .

$$\rho_k(x, y) = \rho_{-k}(y, x)$$

**Sample Cross-correlation** The ccf function can be estimed by

$$r_k(x, y) = \frac{c_k(x, y)}{\sqrt{c_0(x, x)c_0(y, y)}}$$

## 7.4 Basic Stochastic Models

We may consider a **trend** to be **stochastic** when it shows inexplicable changes in direction. Those type of trend can be simulated by the models of this section.

### 7.4.1 White Noise

**Residual Error** A residual error is the difference between the observed value and the model predicted value at time  $t$ . Then the residual error,  $x_t$ , is defined by

$$x_t = y_t - \hat{y}_t$$

where  $y_t$  is the observed value and  $\hat{y}_t$  the predicted value.

As the residual errors occur in time, they form a time series :  $x_1, x_2, \dots, x_n$ .

**Definition** A white noise is a time series  $\{w_t : t = 1, 2, \dots, n\}$  where each term is independent and with a constant variance  $\sigma^2$ .

$$w_t \sim N(0, \sigma^2)$$

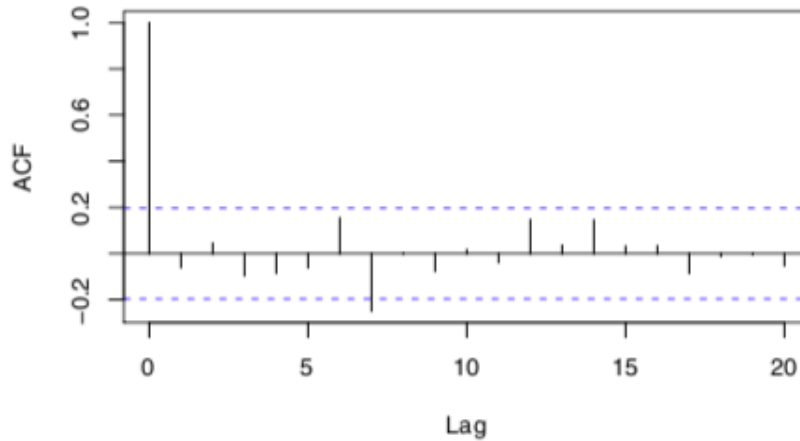


FIGURE 7.3 – Correlogram of a simulated white noise series. The underlying autocorrelations are all zero (except at lag 0); the statistically significant value at lag 7 is due to sampling variation.

### 7.4.2 Random Walks

**Definition** Let  $\{x_t\}$  be a time series. Then  $\{x_t\}$  is a random walk if

$$x_t = x_{t-1} + w_t$$

where  $\{w_t\}$  is a white noise.

**Backward Operator** The backward operator (or lag operator)  $\mathbf{B}$  is defined by

$$\mathbf{B}^n = x_{t-n}$$

**Second-order Properties** The covariance is in function of time, so a random walk is non-stationary. This model is only suitable for short term predictions.

$$\begin{aligned}\mu(t) &= 0 \\ \sigma^2(t) &= t\sigma_w^2 \\ \gamma_k(t) &= t\sigma_w^2 \\ \rho_k(t) &= \frac{1}{\sqrt{1 + \frac{k}{t}}}\end{aligned}$$

**The Difference Operator** The difference operator is defined by

$$\nabla x_t = x_t - x_{t-1}$$

We can also express it with the Backward operator

$$\nabla^n x_t = (1 - \mathbf{B})^n x_t$$

#### Note

Differencing adjacent terms of a series can transform a non-stationary series to a stationary series.

$$x_t - x_{t-1} = w_t$$

Here differencing two non-stationary random walk give a stationary white noise.

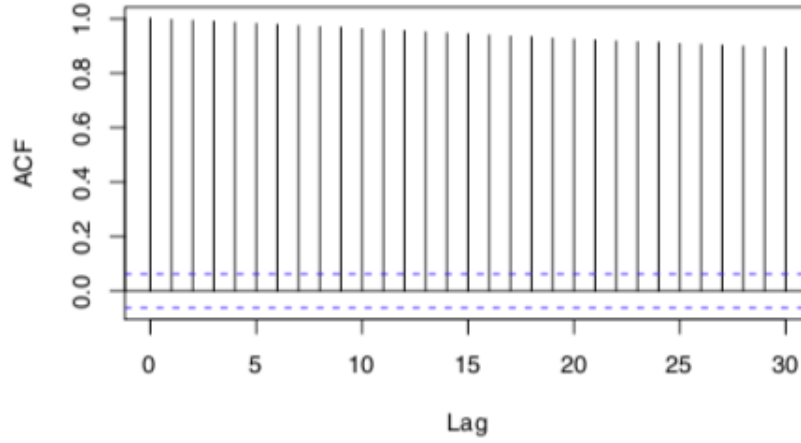


FIGURE 7.4 – The correlogram for the simulated random walk. A gradual decay from a high serial correlation is a notable feature of a random walk series.

**Random Walk with Drift** A Random walk with drift is a random walk with a mean. it defined as

$$x_t = \delta + x_{t-1} + w_t$$

### 7.4.3 Autoregressive Models

**Definition** The series  $\{x_t\}$  is an autoregressive process of order  $p$ , abbreviated to  $AR(p)$ , if

$$x_t = \alpha_1 x_{t-1} + \alpha_2 x_{t-2} + \dots + \alpha_p x_{t-p} + w_t$$

where  $\alpha_i$  are the model parameters with  $\alpha_p \neq 0$ .

The following points should be noted :

- The random walk is a special case of  $AR(1)$  with  $\alpha_1 = 1$ .
- The model is a regression of  $x_t$  on past terms from the same series, that where the name *autoregressive* come.
- A prediction at time  $t$  is given by

$$\hat{x}_t = x_t = \alpha_1 x_{t-1} + \alpha_2 x_{t-2} + \dots + \alpha_p x_{t-p}$$

- The model parameters can be estimated by minimising the sum of squared errors.



**Stationary and non-stationary AR processes** We can express a AR model as a polynomial of order  $p$  in terms of the backward shift operator

$$\theta_p(B)x_t = (1 - \alpha_1 B - \alpha_2 B^2 - \dots - \alpha_p B^p)x_t = w_t$$

Then a AR model is stationary if all roots of  $\theta_p(B)$  exceed one in absolute values. In other words if

if all  $|B_i| > 1$  then the model is stationary  
otherwise, the model is non-stationary

### Second-order properties of an AR(1) model

$$\begin{aligned}\mu_k &= 0 \\ \gamma_k &= \frac{\alpha^k \sigma_w^2}{1 - \alpha^2} \\ \rho &= \alpha^k\end{aligned}$$

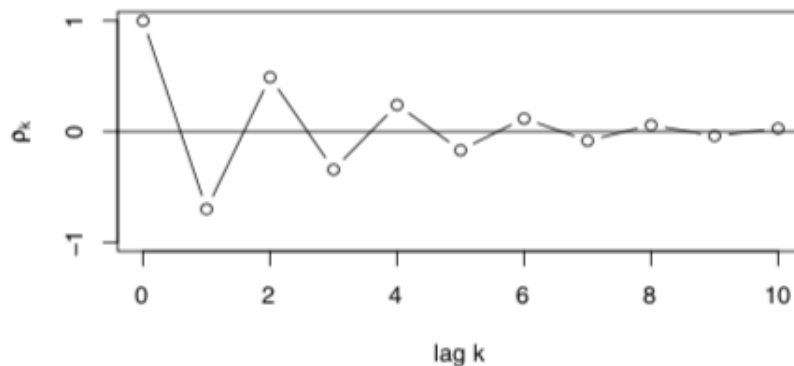


FIGURE 7.5 – For an AR model, the correlogram decays to zero. The correlogram decays to zero more rapidly for small  $\alpha$ .

**Partial Autocorrelation** An AR( $p$ ) process has a correlogram of partial autocorrelation  $\alpha_k$  that is zero after lag  $p$ .

## 7.5 Regression

When we have some plausible physical explanation for a **trend** we will usually wish to model it in some **deterministic** manner. Therefore, the model of this section can be used.

The difference with **deterministic trend** is the when we make short term forecast, we assume that the trend will change slowly.

Time series regression usually differs from a standard regression analysis because the residual form a time series and therefore tend to be serially correlated. When this correlation is positive, the estimated standard errors of the parameter estimates, read from the computer output of a standard regression analysis, will tend to be less than their true value.

### 7.5.1 Linear Models

**Definition** A model for a time series  $\{x_t : t = 1, \dots, n\}$  is linear if it can be expressed as

$$x_t = \alpha_0 + \alpha_1 u_{1,t} + \alpha_2 u_{2,t} + \dots + \alpha_m u_{m,t} + z_t$$

where  $u_{i,t}$  is the value of the  $i^{\text{th}}$  explanatory variable,  $\alpha_i$  the estimated model parameters,  $z_t$  the error at time  $t$ .

An example of a linear model is the  $p$ th-order polynomial function of  $t$  :

$$x_t = \alpha_0 + \alpha_1 t + \alpha_2 t^2 + \dots + \alpha_p t^p + z_t$$

#### Note

Note that the errors form a time series  $\{z_t\}$ , with mean 0, that does not have to be Gaussian or white noise.

**Stationarity** Linear models for time series are non-stationary when they include functions of time.

Differencing can remove both stochastic and deterministic trends from time series. Then for a polynomial of order  $m$ , the  $m$ th-order differencing is required to remove the trend.

**Autocorrelation variance estimation of the sample mean** Let  $\{x_t : t = 1, \dots, n\}$  be a stationary time series with mean  $\mu$ , variance  $\sigma^2$  and autocovariance  $\text{Cov}(x_t, x_{t+k})$ . Then the variance of the sample mean is given by

$$\text{Var}(\bar{x}) = \frac{\sigma^2}{n} \left[ 1 + 2 \sum_{k=1}^{n-1} \left(1 - \frac{k}{n}\right) \rho_k \right]$$

**Generalised least squares** A fitting procedure known as generalised least squares (GLS) can be used to provide better estimates of the standard errors of the regression parameters to account for the autocorrelation in the residual series.

## 7.5.2 Linear Models with Seasonal Variables

**Additive Seasonal Indicator Variables** To include seasonal effect, we change the constant  $\alpha_o$  depending on the season. Let  $s$  be the time series measured (Ex. for monthly time series,  $s = 12$ ). Then for each  $s$ , we fit a constant term.

$$x_t = s_t + m_t + z_t$$

where  $s_t$  is the seasonal constant when  $t$  falls in the  $i^{\text{th}}$  season,  $m_t$  a linear model for the trend and  $z_t$  the error term.

**Harmonic Seasonal models** The advantage of this model is that we can represent the seasonal effect with something that is smoothly. For a time series  $\{x_t\}$  with  $s$  season, there are  $\lfloor s/2 \rfloor$  possible cycles. The harmonic model is defined by

$$x_t = m_t + \sum_{i=1}^{\lfloor s/2 \rfloor} \left\{ s_i \sin\left(\frac{2\pi i t}{s}\right) + c_i \cos\left(\frac{2\pi i t}{s}\right) \right\} + z_t$$

where  $m_t$  is a trend model **that include a constant term** ( $\alpha_0$ ) and  $s_i$  and  $c_i$  are unknown parameters.

## 7.5.3 Forecasting from regression

When we predict a regression time series, we try to predict in the future. The problem is that the trend might change. Therefore, it is better to think of

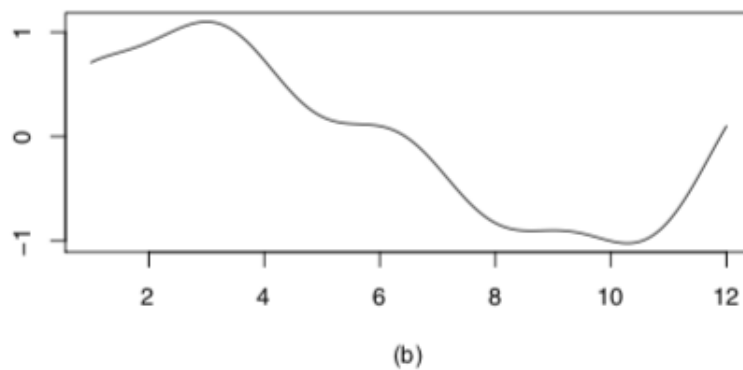
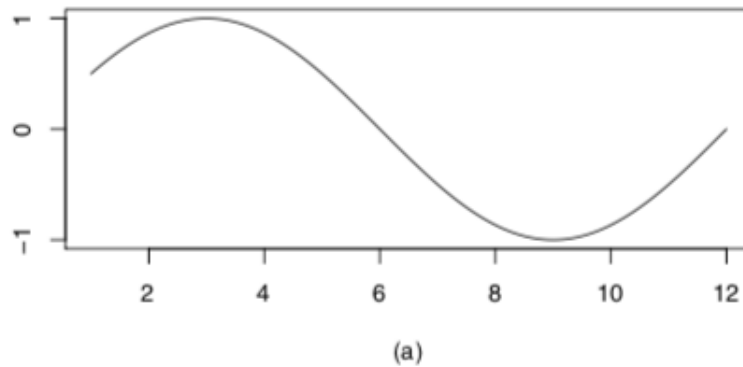


FIGURE 7.6 – Two possible underlying seasonal patterns for monthly series based on the harmonic model (Equation (5.10)). Plot (a) is of the first harmonic over a year and is usually too regular for most practical applications. Plot (b) is of the same wave but with a further two harmonics added. Plot (b) illustrates just one of many ways that an underlying sine wave can be perturbed to produce a less regular, but still dominant, seasonal pattern of period 12 months.

a forecast from a regression model as an expected value conditional on past trends continuing into the future.

**Bias Correction** The process of transforming the model introduce some bias in the mean. We need to apply a correction to the mean. Note that this

correction doesn't need to be apply in simulation.

$$\hat{x}'_t = \hat{x}_t * \text{Correction}$$

**Lognormal Correction**

$$e^{\sigma^2/2}$$

**Empirical Correction**

$$\frac{1}{n} \sum e^{z_t}$$

## 7.6 Stationary Models

Sometime, the residual will be correlated in time, as this is not accounted in the fitted regression model, we need other model.

### 7.6.1 Strictly Stationary Series

A time series model  $\{x_t\}$  is *strictly stationary* if the joint statistical distribution  $x_{t_1}, \dots, x_{t_n}$  is the same as the joint distribution of  $x_{t_1+m}, \dots, x_{t_n+m}$  for all  $t_1, \dots, t_n$  and  $m$ , so that the distribution is unchanged after an arbitrary time shift.

#### Note

Note that strict stationarity implies that the mean and variance are constant in time and that the autocovariance  $\text{Cov}(x_t, x_s) = \gamma_k$  (i.e. only depend on the lag  $k$ ). If a series is not strictly stationary but the mean and variance are constant in time and the autocovariance only depends on the lag, then the series is called *second-order stationary*.

We focus on the second-order properties in this chapter, but the stochastic processes discussed are strictly stationary.

Stationarity is an idealisation that is a property of models. If we fit a stationary model to data, we assume our data are a realisation of a stationary process. So our first step in an analysis should be to check whether there is any evidence of a trend or seasonal effects and, if there is, remove them.

Regression can break down a non-stationary series to a trend, seasonal components, and residual series. It is often reasonable to treat the time series of residuals as a realisation of a stationary error series. Therefore, the models in this chapter are often fitted to residual series arising from regression analyses.

### 7.6.2 Moving average models

**Definition** moving average (MA) process of order  $q$  is a linear combination of the current white noise term and the  $q$  most recent past white noise terms and is defined by

$$x_t = w_t + \beta_1 w_{t-1} + \dots + \beta_q w_{t-q} = \phi_q(B)w_t$$

where  $\phi_q$  is a polynomial of order  $q$ . Because MA processes consist of a finite sum of stationary white noise terms, they are stationary and hence have a time-invariant mean and autocovariance.

#### Second-order properties

$$\begin{aligned}\mu &= 0 \\ \sigma^2 &= \sigma_w^2(1 + \beta_1^2 + \dots + \beta_q^2) \\ \gamma_k &= \sigma_w^2 \sum_{i=0}^{q-k} \beta_i \beta_{i+k} \\ \rho_k &= \frac{\sum_{i=0}^{q-k} \beta_i \beta_{i+k}}{\sum_{i=0}^q \beta_i^2}\end{aligned}$$

where  $\beta_0 = 1$ .

**Invertible properties** An MA process is invertible if it can be expressed as a stationary  $AR(\infty)$  process of infinite order without an error term.

$$w_t = (1 - \beta B)^{-1}x_t = x_t + \beta x_{t-1} + \beta^2 x_{t-2} + \dots$$

provided  $|\beta| < 1$ , which is required for convergence.

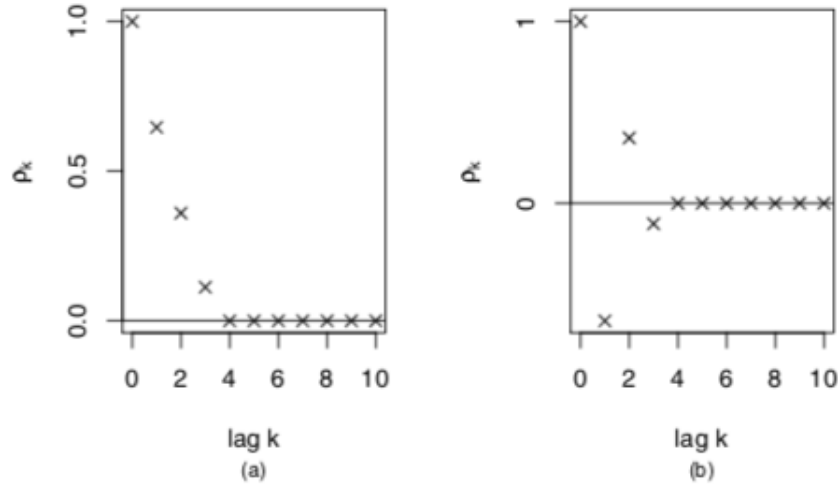


FIGURE 7.7 – Plots of the autocorrelation functions for two MA(3) processes. The autocorrelation for lag  $k > q$  are all zero. (a)  $\beta_1 = 0.7, \beta_2 = 0.5, \beta_3 = 0.2$ ; (b)  $\beta_1 = -0.7, \beta_2 = 0.5, \beta_3 = -0.2$ .

### 7.6.3 Mized Models : The ARMA process

**Definition** The ARMA model is a AR(p) + MA(q) models defined as

$$x_t = \alpha_1 x_{t-1} + \alpha_2 x_{t-2} + \dots + \alpha_p x_{t-p} + w_t + \beta_1 w_{t-1} + \beta_2 w_{t-2} + \dots + \beta_q w_{t-q}$$

We can also express it as

$$\theta_p(B)x_t = \phi_q(B)w_t$$

- The process is stationary when the roots of  $\theta$  are all exceed unity in absolute value.
- The process is invertible when the roots of  $\phi$  all exceed unity in absolute value.
- The AR(p) model is the special case ARMA(p, 0).
- The MA(q) model is the special case ARMA(0, q).
- **Parameter parsimony.** When fitting to data, an ARMA model will often be more parameter efficient (i.e., require fewer parameters) than a single MA or AR model.

- **Parameter redundancy.** When  $\theta$  and  $\phi$  share a common factor, a stationary model can be simplified. For example, the model :

$$\begin{aligned}(1 - \frac{1}{2}B)(1 - \frac{1}{3}B)x_t &= (1 - \frac{1}{6}B)w_t \\ (1 - \frac{1}{3}B)x_t &= w_t\end{aligned}$$

## Second-order Properties

$$\begin{aligned}\sigma^2 &= \sigma_w^2 \left( 1 + \frac{(\alpha + \beta)^2}{1 - \alpha^2} \right) \\ \gamma_0 &= \sigma_w^2 \left( \frac{1 + 2\alpha\beta + \beta^2}{1 - \alpha^2} \right) \\ \gamma_k &= \sigma_w^2 (\alpha + \beta) \alpha^{k-1} \left( \frac{1 + \alpha\beta}{1 - \alpha^2} \right) \\ \rho_k &= \frac{\alpha^{k-1}(\alpha + \beta)(1 + \alpha\beta)}{1 + \alpha\beta + \beta^2} = \alpha\rho_{k-1}\end{aligned}$$

## 7.7 Non-stationary Models

### 7.7.1 Differencing

Differencing a serie  $\{x_t\}$  can remove trends, whether these trend are stochastic, as in a random walk, or deterministic, as in the case of a linear trend.

#### Differencing random walk

$$\nabla x_t = x_t - x_{t-1} = w_t$$

which is a stationary white noise.

#### Differencing linear trend

$$\nabla x_t = x_t - x_{t-1} = b + w_t - w_{t-1}$$

which is a stationary moving average process rather than white noise.



**Integrated Model** A serie  $\{x_t\}$  is integrated of order  $d$ ,  $I(d)$ , if the  $d^{\text{th}}$  difference of  $\{x_t\}$  is a white noise

$$(1 - B)^d x_t = w_t$$

#### Note

The random walk is a special case  $I(1)$

### 7.7.2 Non-Seasonal ARIMA Models

**Definition** A time series  $\{x_t\}$  follows an  $\text{ARIMA}(p, d, q)$  process if the  $d^{\text{th}}$  differences of the  $\{x_t\}$  series are an  $\text{ARIMA}(p, q)$  process

$$\theta_p(B)(1 - B)^d x_t = \phi_q(B)w_t$$

In general :

- $\text{ARIMA}(0, d, q) \equiv \text{IMA}(d, q)$
- $\text{ARIMA}(p, d, 0) \equiv \text{ARI}(p, d)$

### 7.7.3 Seasonal ARIMA models

A seasonal ARIMA model uses differencing at a lag equal to the number of seasons ( $s$ ) to remove additive seasonal effects. As with lag 1 differencing to remove a trend, the lag  $s$  differencing introduces a moving average term.

The  $\text{ARIMA}(p, d, q)(P, D, Q)_s$  model can be defined as

$$\Theta_P(B^s)\theta_p(B)(1 - B^s)^D(1 - B)^d x_t = \Phi_Q(B^s)\phi_q(B)w_t$$

## Chapitre 8

### Extended Linear Models

This section resume the Chapters 5-9 of *An Introduction to Generalized Linear Models*.

Add to biblio :

## 8.1 Inference

The two tool to do inference are **confidence intervals** and **hypothesis tests**. For GLM, a hypothesis tests can be use to compare two models, but their need to have the same probability function and the same link. Also, the null hypothesis  $H_0$  is a sinpler model and must be a special case of the other.

### 8.1.1 Sampling distribution for the score statistic

**Score Function** Let  $\ell = \ln f(y)$  be the log-likelihood function, then the score function  $U$ , is define as

$$U_j = \frac{\partial \ell}{\partial \mu} = \sum_{i=1}^N \left[ \frac{(Y_i - \mu_i)}{\text{Var}(Y_i)} x_{i,j} \left( \frac{\partial \mu_i}{\partial \eta} \right) \right]$$

**Information matrix** The information matrix is define as the variance-covariance matrix of the score function. This information matrix is defined as

$$I(\theta) = \text{Var}(U)$$

**Score Statistic** If there is only one  $\beta$ , the score statistic has the asymptotic sampling distribution

$$\frac{U}{\sqrt{I(\theta)}} \sim N(0, 1), \frac{U^2}{I(\theta)} \sim \chi^2(1)$$

If there is a vector of  $\underline{\beta}$

$$U^T I(\theta)^{-1} U \sim \chi^2(p)$$

## 8.2 Normal Linear Models

### 8.2.1 Basis Result

**Maximum likelihood** The maximum likelihood estimation of  $\beta$  is given by

$$\mathbf{b} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{y}$$

The estimator is unbiased with the variance-covariance matrix

$$I(\theta)^{-1} = \sigma^2 (\mathbf{X}^T \mathbf{X})^{-1}$$

However, the unbiased estimator of  $\sigma^2$  is given by

$$\hat{\sigma}^2 = \frac{1}{N - p} (\mathbf{y} - \mathbf{X}\mathbf{b})^T (\mathbf{y} - \mathbf{X}\mathbf{b})$$

**Least Square Estimation** In the case of linear models, we obtain the same result as the maximum likelihood

$$\mathbf{b} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{y}$$

**Deviance** The deviance is defined by the square of the error  $\varepsilon$ .

$$\frac{1}{\sigma^2} \varepsilon^T \varepsilon = \frac{1}{\sigma^2} (\mathbf{y} - \mathbf{X}\mathbf{b})^T (\mathbf{y} - \mathbf{X}\mathbf{b})$$

or, in case of simple linear model,

$$\frac{1}{\sigma^2} \sum (Y_i - \hat{Y}_i)^2$$

# Annexe A

## Principales distribution de probabilité utilisées

introduction

# Annexe B

## Preuves

### B.1 Stop-Loss ( $\pi_X(d)$ )

Dans un contexte continu,

$$\pi_X(d) = \int_d^\infty \overline{F}(d) du$$

*Démonstration.*

$$\begin{aligned}\pi_X(d) &= E[\max(X - d, 0)] \\ &= \int_0^\infty \max(x - d, 0) F_X(x) dx \\ &= \int_0^\infty (x - d) 1_{\{X > d\}} f_X(x) dx \\ &= \int_d^\infty (x - d) f_X(x) dx \\ &= \int_d^\infty x f_X(x) dx - \int_d^\infty df_X(x) dx\end{aligned}$$

On doit alors faire une intégration par partie, en posant

$$\begin{aligned}u &= x & du &= dx \\ dv &= dF_X(x) & v &= -S(x)\end{aligned}$$

Note : si on fait tendre  $S(x)$  vers l'infini, ça va tendre plus rapidement vers 0 que  $x$  seul.

$$\begin{aligned}
\pi_X(d) &= -xS(x) \Big|_d^\infty - \int_d^\infty -S(x)dx - d(F(\infty) - F(d)) \\
&= 0 + \cancel{dS(d)} + \int_d^\infty S(x)dx - \cancel{dS(d)} \\
&= \int_d^\infty S(x)dx
\end{aligned}$$

□

Il existe aussi le contexte discret :

$$\pi_X(d) = \sum_{k=d}^{\infty} S(k)$$

*Démonstration.*

$$\begin{aligned}
\pi_X(k) &= E[\max(N - k, 0)] \\
&= \sum_{j=k}^{\infty} (j - k)P(N = j) \\
&= (k - k)P(N = k) + ((k + 1) - k)P(N = k + 1) + P((k + 2) - k)P(N = k + 2) + \dots \\
&= P(N = k + 1) + 2P(N = k + 2) + 3P(N = k + 3) + \dots \\
&= \underbrace{(P(N = k + 1) + P(N = k + 2) + P(N = k + 3) + \dots)}_{S(k)} \\
&\quad + \underbrace{(P(N = k + 2) + P(N = k + 3) + P(N = k + 4) + \dots)}_{S(k+1)} \\
&\quad + \underbrace{(P(N = k + 3) + P(N = k + 4) + P(N = k + 5) + \dots)}_{S(k+2)} \\
&\quad + \dots \\
&= \sum_{i=k}^{\infty} S(i)
\end{aligned}$$

□

## B.2 TVaR

### B.2.1 Les 3 formes explicites de la $TVaR$

Pour la  $TVaR$ , il y a 3 preuves à bien connaître :

$$TVaR_\kappa(X) = \frac{1}{1-\kappa} \pi_X(VaR_\kappa(X)) + VaR_\kappa(X)$$

*Démonstration.*

$$\begin{aligned} TVaR_\kappa(X) &= \frac{1}{1-\kappa} \int_\kappa^1 VaR_u(X) du \\ &= \frac{1}{1-\kappa} \int_\kappa^1 (VaR_u(X) - VaR_\kappa(X) + VaR_\kappa(X)) du \\ &= \frac{1}{1-\kappa} \int_\kappa^1 \underbrace{(VaR_u(X) - VaR_\kappa(X))}_{\text{fonction quantile}} du + \underbrace{\int_\kappa^1 VaR_\kappa(X) du}_{\text{intégration d'une constante}} \\ &= \frac{1}{1-\kappa} \int_\kappa^1 (F_X^{-1}(u) - VaR_\kappa(X)) \underbrace{f_U(u)}_{U \sim Unif(0,1)} du + \frac{1}{1-\kappa} VaR_\kappa(X)(1-\kappa) \\ &= \frac{1}{1-\kappa} E[\max(\underbrace{F_X^{-1}(U)}_{F_X^{-1} \sim X} - VaR_\kappa(X); 0)] + VaR_\kappa(X) \\ &= \frac{1}{1-\kappa} E[\max(X - VaR_\kappa(X); 0)] + VaR_\kappa(X) \\ &= \frac{1}{1-\kappa} \pi_X(VaR_\kappa(X)) + VaR_\kappa(X) \end{aligned}$$

□

à partir de la preuve ci-dessus, on peut démontrer celle-ci :

$$TVaR_\kappa(X) = \frac{E[X \times 1_{\{X > VaR_\kappa(X)\}}] + VaR_\kappa(X)(F_X(VaR_\kappa(X)) - \kappa)}{1-\kappa}$$



*Démonstration.*

$$\begin{aligned}
TVaR_\kappa(X) &= \frac{1}{1-\kappa} \pi_X(VaR_\kappa(X)) + VaR_\kappa(X) \\
&= \frac{1}{1-\kappa} E[\max(X - VaR_\kappa(X); 0)] + VaR_\kappa(X) \\
&= \frac{1}{1-\kappa} E[(X - VaR_\kappa(X)) \times 1_{\{X > VaR_\kappa(X)\}}] + VaR_\kappa(X) \\
&= \frac{1}{1-\kappa} E[X \times 1_{\{X > VaR_\kappa(X)\}}] - \frac{1}{1-\kappa} E[VaR_\kappa(X) \times \underbrace{1_{\{X > VaR_\kappa(X)\}}}_{=S_X(VaR_\kappa(X))}] + VaR_\kappa(X) \\
&= \frac{1}{1-\kappa} E[X \times 1_{\{X > VaR_\kappa(X)\}}] - \frac{1}{1-\kappa} VaR_\kappa(X)(1 - F_X(VaR_\kappa(X))) + \frac{1-\kappa}{1-\kappa} VaR_\kappa(X) \\
&= \frac{E[X \times 1_{\{X > VaR_\kappa(X)\}}] + VaR_\kappa(X)(-1 + F_X(VaR_\kappa(X)) + 1 - \kappa)}{1-\kappa} \\
&= \frac{E[X \times 1_{\{X > VaR_\kappa(X)\}}] + VaR_\kappa(X)(F_X(VaR_\kappa(X)) - \kappa)}{1-\kappa}
\end{aligned}$$

□

Une dernière preuve fortement utilisée pour la  $TVaR$ , qui découle directement de la dernière :

$$TVaR_\kappa(X) = \frac{E[X \times 1_{\{X > VaR_\kappa(X)\}}]}{1-\kappa}$$

*Démonstration.* Étant donné que cette formule ne fonctionne seulement que pour une v.a. continue, elle est très facile à prouver :

$$\text{si } X \text{ est continue, } \forall x, F_X(VaR_\kappa(X)) = \kappa$$

Alors, on peut enlever la partie de droite de l'équation.

□

### B.3 Sous-additivité de la $TVaR$

Il y a [plusieurs façons](#) de prouver la sous-additivité de la  $TVaR$ .

### B.3.1 À l'aide de la fonction convexe $\varphi(x)$

On sait que la fonction  $\varphi(x)$  est convexe :

$$\varphi(x) = x + \frac{1}{1 - \kappa} \pi_X(x)$$

Et on sait aussi que

$$TVaR_\kappa(X) = \inf \{ \varphi(x) \}$$

Il faut prouver que  $TVaR_\kappa(X + Y) \leq TVaR_\kappa(X) + TVaR_\kappa(Y)$

*Démonstration.* Puisque  $\varphi(x)$  est une fonction convexe, on peut dire que

$$\begin{aligned} TVaR_\kappa(X) &\leq \varphi(x) \\ &\leq x + \frac{1}{1 - \kappa} \pi_X(x) \end{aligned}$$

On pose le changement de variable  $\boxed{X^* = \alpha X + (1 - \alpha)Y}$

On peut donc remplacer  $x$  dans  $\varphi(x)$  par

$$\begin{aligned} x_0 &= VaR_\kappa(X^*) \\ &= VaR_\kappa(\alpha X + (1 - \alpha)Y) \\ &= \alpha VaR_\kappa(X) + (1 - \alpha) VaR_\kappa(Y) \end{aligned}$$

Alors,

$$\begin{aligned}
TVaR_\kappa(\alpha X + (1 - \alpha)Y) &\leq \alpha VaR_\kappa(X) + (1 - \alpha)VaR_\kappa(Y) \\
&\quad + \frac{1}{1 - \kappa} E[\max(\alpha X + (1 - \alpha)Y - \alpha VaR_\kappa(X) - (1 - \alpha)VaR_\kappa(Y); 0)] \\
&= \alpha VaR_\kappa(X) + (1 - \alpha)VaR_\kappa(Y) \\
&\quad + \frac{1}{1 - \kappa} E[\max(\alpha(X - VaR_\kappa(X)) + (1 - \alpha)(Y - VaR_\kappa(Y)); 0)] \\
&\leq \alpha VaR_\kappa(X) + (1 - \alpha)VaR_\kappa(Y) \\
&\quad + \alpha \left( \frac{1}{1 - \kappa} E[\max(X - VaR_\kappa(X); 0)] \right) \\
&\quad + (1 - \alpha) \left( \frac{1}{1 - \kappa} E[\max(X - VaR_\kappa(X); 0)] \right) \\
\text{Si on met en commun, on retrouve les expressions de la } TVaR \\
&= \alpha \left( \frac{1}{1 - \kappa} \pi_X(VaR_\kappa(X)) + VaR_\kappa(X) \right) \\
&\quad + (1 - \alpha) \left( \frac{1}{1 - \kappa} \pi_Y(VaR_\kappa(Y)) + VaR_\kappa(Y) \right) \\
TVaR_\kappa(\alpha X + (1 - \alpha)Y) &\leq \alpha TVaR_\kappa(X) + (1 - \alpha)TVaR_\kappa(Y)
\end{aligned}$$

La relation se vérifie très bien avec le cas où  $\alpha = 0,5$  :

$$\begin{aligned}
TVaR_\kappa(0,5X + (1 - 0,5)Y) &\leq 0,5TVaR_\kappa(X) + (1 - 0,5)TVaR_\kappa(Y) \\
0,5TVaR_\kappa(X + Y) &\leq 0,5TVaR_\kappa(X) + 0,5TVaR_\kappa(Y) \\
&\quad \text{on multiplie par 2 pour enlever les } 0,5 \\
\mathbf{TVaR}_\kappa(\mathbf{X} + \mathbf{Y}) &\leq \mathbf{TVaR}_\kappa(\mathbf{X}) + \mathbf{TVaR}_\kappa(\mathbf{Y})
\end{aligned}$$

□

### B.3.2 Avec les fonctions indicatrices

Si on a les v.a. continues  $X$  et  $Y$  (les espérances existent) avec les fonctions de répartition respectives  $F_X$  et  $F_Y$ , alors

$$TVaR_\kappa(X) = \frac{E[X \times 1_{\{X > VaR_\kappa(X)\}}]}{1 - \kappa}$$

$$(1 - \kappa)TVaR_\kappa(X) = E[X \times 1_{\{X > VaR_\kappa(X)\}}]$$

est valide pour toute v.a. continue  $X$ .

On veut alors démontrer que

$$TVaR_\kappa(X) + TVaR_\kappa(Y) - TVaR_\kappa(X + Y) \geq 0 \quad (\text{B.1})$$

*Démonstration.* .

(1) On peut écrire le membre de gauche de l'inégalité (B.1) comme

$$\begin{aligned} & \underbrace{(1 - \kappa)TVaR_\kappa(X)}_{E[X \times 1_{\{X > VaR_\kappa(X)\}}]} + \underbrace{(1 - \kappa)TVaR_\kappa(Y)}_{E[Y \times 1_{\{Y > VaR_\kappa(Y)\}}]} - \underbrace{(1 - \kappa)TVaR_\kappa(X + Y)}_{E[(X+Y) \times 1_{\{X+Y > VaR_\kappa(X+Y)\}}]} \\ &= E[X \times 1_{\{X > VaR_\kappa(X)\}}] + E[Y \times 1_{\{Y > VaR_\kappa(Y)\}}] - \underbrace{E[(X + Y) \times 1_{\{X+Y > VaR_\kappa(X+Y)\}}]}_{\text{On split cette espérance}} \\ &= \underbrace{E[X \times 1_{\{X > VaR_\kappa(X)\}}] - E[X \times 1_{\{X+Y > VaR_\kappa(X+Y)\}}]}_{\text{On peut rassembler les indicatrices}} \\ &+ \underbrace{E[Y \times 1_{\{Y > VaR_\kappa(Y)\}}] - E[Y \times 1_{\{X+Y > VaR_\kappa(X+Y)\}}]}_{\text{ici aussi}} \end{aligned}$$

$$= E[X \times (1_{\{X > VaR_\kappa(X)\}} - 1_{\{X+Y > VaR_\kappa(X+Y)\}})] + E[Y \times (1_{\{Y > VaR_\kappa(Y)\}} - 1_{\{X+Y > VaR_\kappa(X+Y)\}})] \quad (\text{B.2})$$

(2) Rendu ici, on veut prouver que chacun de ces espérance  $\geq 0$ , pour que la somme des 2 soit  $\geq 0$  aussi. Étant donné que les 2 parties du membre de gauche sont identiques, on va le prouver seulement pour un côté.

(2.1) Pour nous aider, on va créer un terme *auxiliaire*, i.e un terme qui est égal à zéro, mais qui va nous aider à faire la preuve, soit le terme suivant :

$$\begin{aligned}
& E[VaR_\kappa(X) \times (1_{\{X > VaR_\kappa(X)\}} - 1_{\{X+Y > VaR_\kappa(X+Y)\}})] \\
&= VaR_\kappa(X) E[(1_{\{X > VaR_\kappa(X)\}} - 1_{\{X+Y > VaR_\kappa(X+Y)\}})] \\
&= VaR_\kappa(X) ((1 - \kappa) - (1 - \kappa)) \\
&= 0
\end{aligned}$$

(2.2) Alors, l'équation (B.2) devient

$$E[(X - VaR_\kappa(X)) \times (1_{\{X > VaR_\kappa(X)\}} - 1_{\{X+Y > VaR_\kappa(X+Y)\}})]$$

(2.3) On va prouver que la quantité à l'intérieur de l'espérance ci-haut sera toujours  $\geq 0$ , de sorte que l'espérance sera toujours positive aussi :

$$\begin{aligned}
(X - VaR_\kappa(X))(1_{\{X > VaR_\kappa(X)\}} - 1_{\{X+Y > VaR_\kappa(X+Y)\}}) &\geq 0 \text{ si } X < VaR_\kappa(X) \\
(X - VaR_\kappa(X))(1_{\{X > VaR_\kappa(X)\}} - 1_{\{X+Y > VaR_\kappa(X+Y)\}}) &= 0 \text{ si } X = VaR_\kappa(X) \\
(X - VaR_\kappa(X))(1_{\{X > VaR_\kappa(X)\}} - 1_{\{X+Y > VaR_\kappa(X+Y)\}}) &\geq 0 \text{ si } X > VaR_\kappa(X)
\end{aligned}$$

(2.4) Alors, on déduit que

$$\begin{aligned}
& E[(X - VaR_\kappa(X)) \times (1_{\{X > VaR_\kappa(X)\}} - 1_{\{X+Y > VaR_\kappa(X+Y)\}})] \\
&= E[X \times (1_{\{X > VaR_\kappa(X)\}} - 1_{\{X+Y > VaR_\kappa(X+Y)\}})] \geq 0
\end{aligned}$$

(3) Par conséquent,

$$\begin{aligned}
(1 - \kappa)TVaR_\kappa(X) + (1 - \kappa)TVaR_\kappa(Y) - (1 - \kappa)TVaR_\kappa(X + Y) &\geq 0 \\
\mathbf{TVaR}_\kappa(\mathbf{X}) + \mathbf{TVaR}_\kappa(\mathbf{Y}) - \mathbf{TVaR}_\kappa(\mathbf{X} + \mathbf{Y}) &\geq 0
\end{aligned}$$

□

### B.3.3 À l'aide des statistiques d'ordre

Pour pouvoir prouver la sous-additivité de la  $TVaR$ , on peut aussi utiliser la relation des statistiques d'ordre<sup>1</sup> :

$$TVaR_{\kappa}(X) = \lim_{n \rightarrow \infty} \frac{\sum_{j=[n\kappa]+1}^n X_{j:n}}{[n(1-\kappa)]}$$

à compléter plus tard

## B.4 Loi des grands nombres

Cette preuve était demandée à l'examen Intra traditionnel H2017 du cours ACT-2001.

#### Théorème

Soit les v.a. *iid*  $X_1, \dots, X_n$  avec  $E[X^m] < \infty$ ,  $m = 1, 2, \dots$  et  $Var(X) < \infty$ . Alors,

$$\lim_{n \rightarrow \infty} F_{W_n}(x) \longrightarrow F_Z(x) \quad (\text{B.3})$$

où  $Z$  est une v.a. tel que  $P(Z = E[X]) = 1$ .

*Démonstration.* (1) Première étape, on va démontrer que  $\lim_{n \rightarrow \infty} \mathcal{L}_{w_n}(t) \rightarrow \mathcal{L}_Z(t)$

(1.1) On sait que  $\mathcal{L}_{w_n}(t) = \mathcal{L}_X\left(\frac{t}{n}\right)^n$   $n = 1, 2, \dots$

(1.2) Soit une v.a.  $Y$  positive. On fixe  $t$  tout petit

(1.3) Alors

$$\begin{aligned} \mathcal{L}_Y(t) &= E[e^{-tY}] \\ &\approx E[1 - tY] \quad \text{par dév. de Taylor} \quad = E[1] - tE[Y] \end{aligned}$$

---

1. Relation qui d'ailleurs est utilisée dans le contexte de simulation Monte Carlo pour estimer la TVaR d'une variable aléatoire.

(1.4)

$$\begin{aligned}\mathcal{L}_{w_n}(t) &= \mathcal{L}_X \left( \frac{t}{n} \right)^n \\ &\simeq \left( 1 - \frac{t}{n} E[X] \right)^n\end{aligned}$$

(1.5) On prends la limite de part et d'autre de l'égalité en (1.3)

$$\begin{aligned}\lim_{n \rightarrow \infty} \mathcal{L}_{w_n}(t) &= \lim_{n \rightarrow \infty} \left( \mathcal{L}_X \left( \frac{t}{n} \right) \right)^n \\ &\simeq \lim_{n \rightarrow \infty} \left( 1 - \frac{t}{n} E[X] \right)^n \\ &= e^{-tE[X]} \\ &= \mathcal{L}_Z(t)\end{aligned}$$

Ce qui correspond à la Transformée de la v.a.  $Z$  où  $P(Z = E[X]) = 1$

(2) On applique le résultat de (1.4)

$$\lim_{n \rightarrow \infty} F_{w_n}(x) = F_Z(x), \quad \forall x$$

□

## B.5 Somme de v.a. indépendantes d'une loi Poisson Composée

*Démonstration.* Soit les v.a. indépendantes  $X_1, \dots, X_n$  où  $X_i \sim \text{PoisComp}(\lambda_i; F_{B_i})$ ,  $i = 1, \dots, n$   
Ainsi,

$$\begin{aligned}\mathcal{L}_{X_1}(t) &= \mathcal{P}_{M_i}(\mathcal{L}_{B_i}(t)) \\ &= e^{\lambda(\mathcal{L}_{B_i}(t)-1)}, \quad i = 1, 2, \dots, n\end{aligned}$$

On peut trouver la transformée de  $S$ ,

$$\mathcal{L}_S(t) = \prod_{i=1}^n \mathcal{L}_{X_i}(t) = \prod_{i=1}^n e^{\lambda(\mathcal{L}_{B_i}(t)-1)} \quad (\text{B.4})$$

Le passage de l'équation (B.4) aux étapes suivantes résulte d'une propriété de la loi de Poisson, i.e.

$$\begin{aligned} \mathcal{L}_S(t) &= e^{\sum_{i=1}^n \lambda_i(\mathcal{L}_{B_i}(t)-1)} \\ &= e^{\sum_{i=1}^n \lambda_i \mathcal{L}_{B_i}(t) - \sum_{i=1}^n \lambda_i} \\ &= e^{\sum_{i=1}^n \lambda_i \mathcal{L}_{B_i}(t) - \lambda_S} \\ &= e^{\sum_{i=1}^n \lambda_i \mathcal{L}_{B_i}(t) - \lambda_S} \end{aligned}$$

Si on met en évidence le  $\lambda_S$  ...

$$\mathcal{L}_S(t) = e^{\lambda_S \left( \sum_{i=1}^n \frac{\lambda_i}{\lambda_S} \mathcal{L}_{B_i}(t) - 1 \right)} \quad (\text{B.5})$$

Si on pose  $c_i = \frac{\lambda_i}{\lambda_S}$ , on observe que  $0 < c_i < 1$  et que  $\sum_{i=1}^n c_i = 1$ .

On se définit une nouvelle v.a.,  $D$ , où

$$\mathcal{L}_D(t) = \sum_{i=1}^n c_i \mathcal{L}_{B_i}(t) \quad (\text{B.6})$$

Ce qui implique que  $D$  obéit à une loi mélange :

$$F_D(x) = \sum_{i=1}^n c_i F_{B_i}(x) \quad , x \geq 0$$

en combinant (B.5) et (B.6), on obtient

$$\mathcal{L}_S(t) = e^{\lambda_S(\mathcal{L}_D(t)-1)} \quad (\text{B.7})$$

On introduit une nouvelle v.a.,  $N_S \sim \text{Pois}(\lambda_S)$  et  $P_N(s) = e^{\lambda_S(s-1)}$ . Alors, (B.7) devient

$$\mathcal{L}_S(t) = \mathcal{P}_{N_S}(\mathcal{L}_D(t))$$



On peut donc représenter  $S$  comme

$$S = \begin{cases} \sum_{k=1}^{N_s} D_k & N_s > 0 \\ 0 & N_s = 0 \end{cases}$$

où  $D_k$ ,  $k = 1, 2, \dots$  forme une suite de v.a *iid*, et  $D$  et  $N_s$  sont indépendants.  $\square$

## B.6 Théorème d'Euler

**Définition** Soit une fonction  $\phi : \mathbb{R}^n \rightarrow \mathbb{R}$  homogène d'ordre  $n$ . Alors, pour toute fonction  $\phi$  dérivable partout, on a

$$n\phi(x_1, \dots, x_n) = \sum_{i=1}^n x_i \frac{\partial \phi(x_1, \dots, x_n)}{\partial x_i}$$

*Démonstration.* .

(1) Puisque  $\phi$  est homogène d'ordre  $n$ , on a

$$\phi(x_1, \dots, x_n) = \lambda^n \phi(x_1, \dots, x_n) \quad (\text{B.8})$$

(2) On dérive le terme de gauche de l'équation dans l'équation (B.8) par rapport à  $\lambda$  et on pose  $\lambda = 1$ .

$$\begin{aligned} \left. \frac{\partial}{\partial \lambda} \lambda^n \phi(x_1, \dots, x_n) \right|_{\lambda=1} &= n \lambda^{n-1} \phi(x_1, \dots, x_n) \Big|_{\lambda=1} \\ &= n \phi(x_1, \dots, x_n) \end{aligned}$$

(3) On dérive le terme de droite de l'équation dans l'équation (B.8) par rapport à  $\lambda$  et on pose  $\lambda = 1$ .

$$\begin{aligned} \left. \frac{\partial}{\partial \lambda} \lambda^n \phi(x_1, \dots, x_n) \right|_{\lambda=1} &= \sum_{i=1}^n \frac{\partial \phi(x_1, \dots, x_n)}{\partial (\lambda x_i)} \frac{\partial (\lambda x_i)}{\partial \lambda} \Big|_{\lambda=1} \\ &= \sum_{i=1}^n \frac{\partial (\lambda x_i, \dots, \lambda x_n)}{\partial \lambda x_i} x_i \Big|_{\lambda=1} \\ &= \sum_{i=1}^n x_i \frac{\partial \phi(x_1, \dots, x_n)}{\partial x_i} \end{aligned}$$

(4) On pose  $(4) = (3)$ , et on obtient le résultat souhaité.

□

## B.7 Dérivée de l'écart-type (générale)

Lorsqu'on prouve la contribution  $C(X_i)$  pour  $\rho(X) = \sqrt{\text{Var}(\sum_{i=1}^n X_i)}$ , on doit dériver l'écart-type... voici le développement complet, avec un exemple où  $n = 3$ . Ce qui est important de suivre, c'est qu'on cherche ici la contribution de la v.a.  $X_i$  : alors, lorsqu'on dérive par rapport à  $\lambda_i$ , ça peut être n'importe quoi le  $i : 1, 2, \dots, n$ .

**Rappel d'ACT-1002** Pour les propriétés de la covariance, voir la sous-section ??.

$$\frac{\partial}{\partial \lambda_i} \sqrt{\text{Var} \left( \sum_{i=1}^n \lambda_i X_i \right)} = \frac{1}{2} \left( \frac{1}{\sqrt{\text{Var}(\sum_{i=1}^n \lambda_i X_i)}} \right) \times$$

$$\frac{\partial}{\partial \lambda_i} \left( \sum_{i=1}^n \lambda_i^2 \text{Var}(X_i) + \sum_{i=1}^n \sum_{k=1, k \neq i}^n \lambda_i \lambda_k \text{Cov}(X_i, X_k) \right)$$

### Explication de la forme générale de la variance

Avec un exemple  $n = 3$ , il est très facile de comprendre d'où vient la formule générale de la variance (qui est universelle si les  $X_i$  sont indépendants ou non).

$$\begin{aligned} \text{Var}(X_1 + X_2 + X_3) &= \text{Cov}(X_1 + X_2 + X_3, X_1 + X_2 + X_3) \\ &= \text{Cov}(X_1, X_1) + \text{Cov}(X_1, X_2) + \text{Cov}(X_1, X_3) \\ &\quad + \text{Cov}(X_2, X_1) + \text{Cov}(X_2, X_2) + \text{Cov}(X_2, X_3) \\ &\quad + \text{Cov}(X_3, X_1) + \text{Cov}(X_3, X_2) + \text{Cov}(X_3, X_3) \end{aligned}$$

On remarque **en bleu** les variances séparées pour chacun de nos  $X_i$  de notre exemple, qu'on va pouvoir rassembler ensemble dans une même somme. On remarque aussi que les covariances sont similaires. On remarque **en orange** les covariances reliées à  $X_1$ , **en vert** les covariances reliées à  $X_2$  et finalement **en violet** les covariances qui sont reliées à  $X_3$ .

En étant attentif, on remarque qu'on peut sommer ensemble chaque *paquet* de covariance sur tout le support ( $n = 3$ ), sauf la combinaison  $\text{Cov}(X_i, X_i)$ , car celle-ci a été prise pour rassembler les variances ensemble ( $\text{Var}(X_i) = \text{Cov}(X_i, X_i)$ ).

Alors, on obtient (pour le cas  $n = 3$ ) :

$$\text{Var}\left(\sum_{i=1}^3 X_i\right) = \sum_{i=1}^3 \text{Var}(X_i) + \sum_{i=1}^3 \sum_{k=1, k \neq i}^3 \text{Cov}(X_i, X_k)$$

Si on développe le **la dérivée en rouge** seule du reste (en prenant l'exemple

du cas  $n = 3$  et qu'on dérive par rapport à  $\lambda_1$ ), on obtient :

$$\begin{aligned}
\frac{\partial}{\partial \lambda_1} \rho(X_1 + X_2 + X_3) &= \frac{\partial}{\partial \lambda_1} \left[ \lambda_1^2 \text{Var}(X_1) + \lambda_2^2 \text{Var}(X_2) + \lambda_3^2 \text{Var}(X_3) \right. \\
&\quad + \lambda_1 \lambda_2 \text{Cov}(X_1, X_2) + \lambda_1 \lambda_3 \text{Cov}(X_1, X_3) \\
&\quad + \lambda_2 \lambda_1 \text{Cov}(X_2, X_1) + \lambda_2 \lambda_3 \text{Cov}(X_2, X_3) \\
&\quad \left. + \lambda_3 \lambda_1 \text{Cov}(X_3, X_1) + \lambda_3 \lambda_2 \text{Cov}(X_3, X_2) \right] \\
&= 2\lambda_1 \text{Var}(X_1) \\
&\quad + \lambda_2 \text{Cov}(X_1, X_2) + \lambda_3 \text{Cov}(X_1, X_3) \\
&\quad + \lambda_2 \text{Cov}(X_2, X_1) + \lambda_3 \text{Cov}(X_3, X_1) \\
&= 2\lambda_1 \text{Var}(X_1) + \sum_{k=1, k \neq 1}^3 \lambda_k \text{Cov}(X_1, X_k) + \sum_{k=1, k \neq 1}^3 \lambda_k \text{Cov}(X_k, X_1) \\
&= 2\lambda_1 \text{Var}(X_1) + 2 \sum_{k=1, k \neq 1}^3 \lambda_k \text{Cov}(X_1, X_k)
\end{aligned}$$

Il ne reste plus qu'à remettre toute l'équation ensemble :

$$\frac{\partial}{\partial \lambda_i} = \frac{1}{2} \frac{2\lambda_i \text{Var}(X_i) + 2 \sum_{k=1, k \neq i}^3 \lambda_k \text{Cov}(X_i, X_k)}{\sqrt{\text{Var}(\sum_{i=1}^n \lambda_i X_i)}}$$

Si on pose  $\lambda_1 = \dots = \lambda_i = \dots = \lambda_n = 1$  et qu'on utilise les définitions des covariances pour rentrer les sommes dans la covariance, tel que

$$\begin{aligned}
\text{Var}(X_i) + \sum_{k=1, k \neq i}^n \text{Cov}(X_i, X_k) &= \sum_{k=1}^n \text{Cov}(X_i, X_k) \\
&= \text{Cov}\left(X_i, \sum_{k=1}^n X_k\right)
\end{aligned}$$

Alors,

$$C(X_i) = \frac{\text{Cov}(X_i, \sum_{k=1}^n X_k)}{\sqrt{\text{Var}(\sum_{k=1}^n X_k)}} = \frac{\text{Cov}(X_i, S)}{\sqrt{\text{Var}(S)}}$$

## B.8 Distribution limite de $W_n$

Soit la v.a.  $Z$  où  $\Pr(Z = E[X]) = 1$ . On veut démontrer (à l'aide des transformées de Laplace) que

$$F_{W_n}(x) \longleftarrow F_Z(x) \quad x > 0$$

où  $\Pr(Z = \gamma_j) = \Pr(\Theta = \theta_j)$  et  $\gamma_j = E[X|\Theta = \theta_j]$ .

*Démonstration.* Pour faire la preuve, il faut savoir les 2 résultats suivants :

$$e^{-x} \approx 1 - x \tag{B.9}$$

$$\lim_{n \rightarrow \infty} \left(1 + \frac{x}{n}\right)^n = e^x \tag{B.10}$$

Si on développe la transformée de Laplace :

$$\begin{aligned}
\mathcal{L}_{W_n}(t) &= E[e^{-tW_n}] \\
&= E_{\Theta}[E[e^{-tW_n}|\Theta = \theta]] \\
&= \int_0^{\infty} E[e^{-tW_n}|\Theta = \theta] f_{\Theta}(\theta) d\theta \\
&= \int_0^{\infty} E[e^{\frac{t}{n}(X_1 + \dots + X_n)}|\Theta = \theta] f_{\Theta}(\theta) d\theta \\
&= \int_0^{\infty} \prod_{i=1}^n E[e^{-\frac{t}{n}X_i}|\Theta = \theta] f_{\Theta}(\theta) d\theta \quad (\text{Car les risques sont cond. indép.}) \\
&= \int_0^{\infty} E[e^{-\frac{t}{n}X}|\Theta = \theta]^n f_{\Theta}(\theta) d\theta \quad (\text{car les v.a. sont id}) \\
&\approx \int_0^{\infty} E\left[\left(1 - \frac{t}{n}X\right)|\Theta = \theta\right]^n f_{\Theta}(\theta) d\theta \quad (\text{par l'équation (B.9)}) \\
&= \int_0^{\infty} \left(E[1|\Theta] - \frac{t}{n}E[X|\Theta]\right)^n f_{\Theta}(\theta) d\theta \\
&= \int_0^{\infty} \left(1 - \frac{t}{n}E[X|\Theta]\right)^n f_{\Theta}(\theta) d\theta
\end{aligned}$$

Si on pose la limite  $n \rightarrow \infty$ ,

$$\begin{aligned}
\lim_{n \rightarrow \infty} \mathcal{L}_{W_n}(t) &= \int_0^{\infty} \lim_{n \rightarrow \infty} \left(1 - \frac{t}{n}E[X|\Theta]\right)^n f_{\Theta}(\theta) d\theta \\
&= \int_0^{\infty} e^{-tE[X|\Theta]} f_{\Theta}(\theta) d\theta \quad (\text{par l'équation (B.10)}) \\
&= \int_0^{\infty} e^{-t\gamma} f_{\Theta}(\theta) d\theta \quad , \text{ où } \gamma = E[X|\Theta] \\
&= \mathcal{L}_Z(t)
\end{aligned}$$

□