Guide de survie en actuariat

Gabriel Crépeault-Cauchon et Nicholas Langevin

16 septembre 2019

Table des matières

Première partie

Fondements mathématiques utiles

Chapitre 1

Algèbre linéaire

1.1 Définition d'un vecteur et une matrice

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

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

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

$$\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} & ... & a_{1n} \\ a_{21} & a_{22} & ... & a_{2n} \\ ... & ... & \ddots & ... \\ a_{m1} & a_{m2} & ... & a_{mn} \end{bmatrix} \tag{1.1}$$

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 **A** est définie comme non-négative, mais elle n'est pas définie positive.

Orthogonale **A** est *orthogonale* si elle est non-singulière et $\mathbf{A}^{-1} = \mathbf{A}^{\top}$ (voir sous-section 1.4.3 pour définition de $\mathbf{A}^{-1}$)

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

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

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

$$\mathbf{I} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \tag{1.2}$$

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

$$\mathbf{Q}^{-1}\mathbf{A}\mathbf{Q} = \mathbf{D} \leftrightarrow \mathbf{A} = \mathbf{Q}\mathbf{D}\mathbf{Q}^{-1} \tag{1.3}$$

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

1.2 Matrice transposée

Soit la matrice $\mathbf{A}$ définie en (1.1). 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 :

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

1.3 Opérations matricielles

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

$$\mathbf{A}\mathbf{B} = \left[\sum_{k=1}^{p} a_{ik}b_{kj}\right]_{i \times j} \tag{1.4}$$

  , avec $\mathbf{A} = [a_{ip}]$ et $B = [b_{pj}]$
- $\mathbf{A}(\mathbf{B} + \mathbf{C}) = \mathbf{A}\mathbf{B} + \mathbf{A}\mathbf{C}$
- $\mathbf{A}^{-1}\mathbf{A} = \mathbf{I} = \mathbf{A}\mathbf{A}^{-1}$, où I est la matrice identité (voir Équation 1.2) et A^{-1} est la matrice inverse de A (voir sous-section 1.4.3 au besoin)
- $(AB)^{-1} = B^{-1}A^{-1}$

1.4 Trace, déterminant et matrice inverse

1.4.1 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.

$$\text{Tr}(\mathbf{A}) = \sum_{i=1}^{n} a_{ii} \tag{1.5}$$

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 $\text{Tr}(\mathbf{ABC}) = \text{Tr}(\mathbf{CAB}) = \text{Tr}(\mathbf{BCA})$

1.4.2 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

$$\det(\mathbf{A}) = \begin{vmatrix} a & b \\ c & d \end{vmatrix} = ad - bc \tag{1.6}$$

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

$$\det(A) = \sum_{j=1}^{n} a_{ij} C_{ij} \tag{1.7}$$

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), alors le déterminant aura les propriétés suivantes :

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

1.4.3 Matrice inverse

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

$$\mathbf{A}^{-1} = \frac{1}{\det(\mathbf{A})} \operatorname{Adj}(\mathbf{A}) \tag{1.8}$$

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

1.5 Décomposition LDU de Choleski

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

$$\mathbf{A} = \mathbf{LDU} \tag{1.9}$$

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.

1.6 Vecteurs et valeurs propres

1.6.1 Définition

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

$$\mathbf{Ax} = \lambda \mathbf{x} \tag{1.10}$$

On appelle le vecteur $\mathbf{x}$ un *vecteur propre* correspondant à la valeur propre λ. De plus, l'ensemble des nombres réels λ satisfaisant l'Équation 1.10 est appelé *spectre* de la matrice $\mathbf{A}$.

1.6.2 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 λ, alors $c\mathbf{x}$ sera également un vecteur propre de $\mathbf{A}$ correspondant à λ.

- 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 ortogonaux, i.e. $\mathbf{x_1}^\top \mathbf{x_2} = 0$.

- Si $\mathbf{A}$ a les valeurs propres (pas nécessairement distinctes) $\lambda_1, \ldots, \lambda_n$, alors $\det(\mathbf{A}) = \prod_{i=1}^n \lambda_i$ et $\text{Tr}(\mathbf{A}) = \sum_{i=1}^n \lambda_i$.

1.6.3   Décomposition spectrale

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

$$\mathbf{A} = \mathbf{Q}\boldsymbol{\Lambda}\mathbf{Q}^\top \tag{1.11}$$

avec $\boldsymbol{\Lambda} = diag(\lambda_1, \ldots, \lambda_n)$. 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) :

$$\mathbf{A}\mathbf{A} = \mathbf{Q}\boldsymbol{\Lambda} \underbrace{Q^\top Q}_{=\mathbf{I}} \boldsymbol{\Lambda}\mathbf{Q}^\top$$
$$= \mathbf{Q}\boldsymbol{\Lambda}^\mathbf{2}\mathbf{Q}^\top$$

1.7   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 2

Probabilités et statistiques

2.1 Concepts de probabilité de base

2.1.1 Probabilité conditionnelle

Définition de base

$$\Pr(A|B) = \frac{\Pr(A \cap B)}{\Pr(B)} \tag{2.1}$$

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

$$\Pr(A) = \sum_{i=1}^{n} \Pr(A|E_i) \Pr(E_i) \tag{2.2}$$

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

Relation importante de l'Équation 2.1, on peut représenter $\Pr(A|B)$ comme

$$\Pr(A|B) = \frac{\Pr(B|A) \Pr(A)}{\Pr(B)} \tag{2.3}$$

2.1.2 Théorème de Bayes

En combinant l'Équation 2.3 et la loi des probabilités totales (l'Équation 2.2), on obtient le théorème de Bayes :

$$\Pr\left(A|B\right) = \frac{\Pr\left(B|A\right)\Pr\left(A\right)}{\sum_{i=1}^{n}\Pr\left(B|A_i\right)\Pr\left(A_i\right)} \tag{2.4}$$

2.2 Définition d'une variable aléatoire

2.3 Distribution d'une variable aléatoire

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

2.4 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

$$\mathrm{E}\left[X\right] = \mu = \sum_{x=0}^{\infty} x\Pr\left(X = x\right) = \int_0^{\infty} x f_X(x)dx \tag{2.5}$$

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

$$\mathrm{E}\left[g(X)\right] = \sum_{x=0}^{\infty} g(x)\Pr\left(X = x\right) = \int_0^{\infty} g(x) f_X(x)dx \tag{2.6}$$

Variance

$$\mathrm{Var}\left(X\right) = \sigma^2 = \mathrm{E}\left[(X - \mathrm{E}\left[X\right])^2\right] = \mathrm{E}\left[X^2\right] - \mathrm{E}\left[X\right]^2 \tag{2.7}$$

quelques propriétés à savoir :

$$\mathrm{Var}\left(aX\right) = a^2\mathrm{Var}\left(X\right)$$
$$\mathrm{Var}\left(X + b\right) = \mathrm{Var}\left(X\right)$$

Covariance

$$\mathrm{Cov}\left(X,Y\right) = \sigma_{X,Y} = \mathrm{E}\left[(X - \mathrm{E}\left[X\right])(Y - \mathrm{E}\left[Y\right])\right] = \mathrm{E}\left[XY\right] - \mathrm{E}\left[X\right]\mathrm{E}\left[Y\right]$$
$$(2.8)$$

2.5 Distribution de probabilité qui reviennent souvent

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

Chapitre 3

Mathématiques financières

Chapitre 4

Processus aléatoire

4.1   Chaîne de Markov

Troisième partie

Matière pour les examens professionnels

Chapitre 5

Time Series

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 $x_t : t = 1, ..., n = x_1, x_2, ..., x_n$

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)**

5.1   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.

5.1.1   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

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

5.1.2   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.

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

5.1.3   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.

Decomposition of multiplicative time series
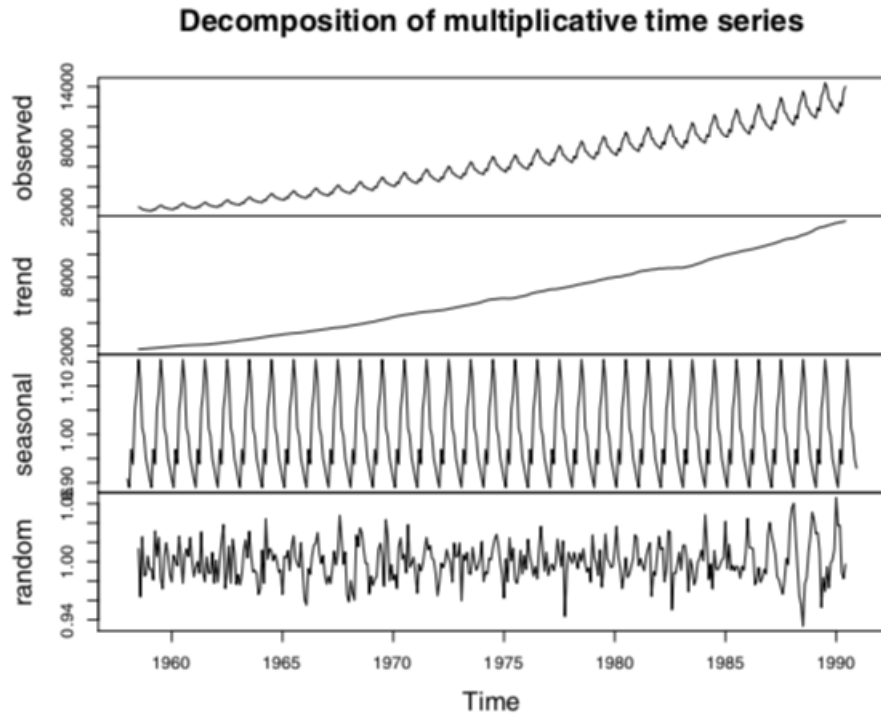
FIGURE 5.1 – In this example, the multiplicative model would seem more appropriate than the additive model because the variance of the original series and trend increase with time

5.2 Correlation

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

$$\mu(t) = \mathrm{E}\left[x_t\right]$$

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

$$\sigma^2(t) = \mathrm{E}\left[(x_t - \mu^2)\right]$$

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

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

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 σ^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.

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

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

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

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.

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

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), γ_k , as a function of the lag k.

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

Sample Autocorrelation   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 autocorrelaton function (acf), ρ_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 the time serie that is stationary in the mean and the variance. Then, the time serie model is second-order stationary if the correlation between variable depends only on the number of time steps separating them.

5.2.1   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 (ρ_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 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.

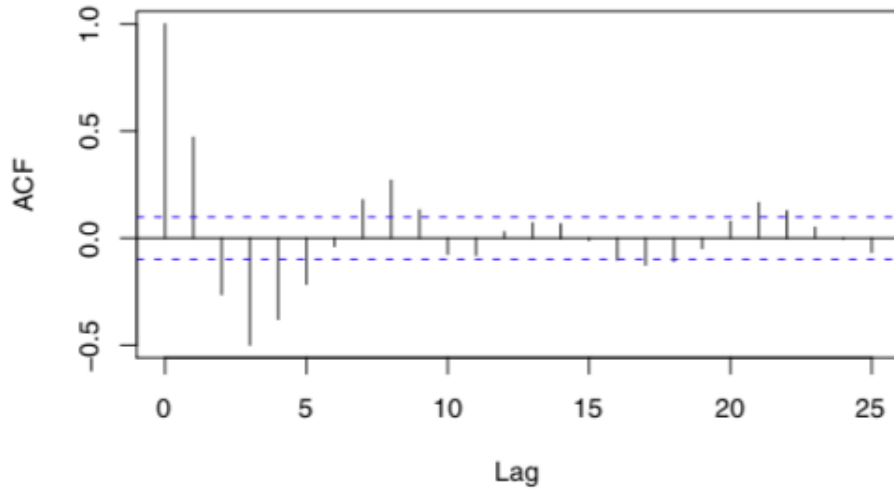FIGURE 5.2 – Correlogram

5.2.2 Covariance of sums of random variables

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

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

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

23

5.3 Forecasting Strategies

5.3.1 Relationships of different time serie

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

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

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

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

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.

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

Sample Cross-correlation The ccf function can be estimed by

$$r_k(x, y) = \frac{c_k(x, y)}{\sqrt{c_0(x, x) x_0(y, y)}}$$

5.4 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.

5.4.1 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 : $x_1, x_2, ..., x_n$.

Definition A white noise is a time serie $\{w_t : t = 1, 2, ..., n\}$ where each term is independent and with a constant variance σ^2.
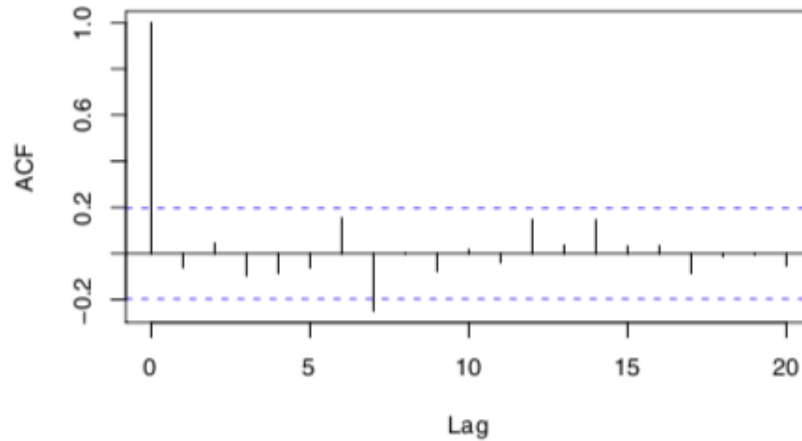
$$w_t \sim \mathrm{N}(0, \sigma^2)$$

FIGURE 5.3 – Correlogram of a simulated white noise series. The underlying autocorre- lations are all zero (except at lag 0) ; the statistically significant value at lag 7 is due to sampling variation.

5.4.2 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.

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

The Difference Operator The defference operator is defined by

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

We can also expresse 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.
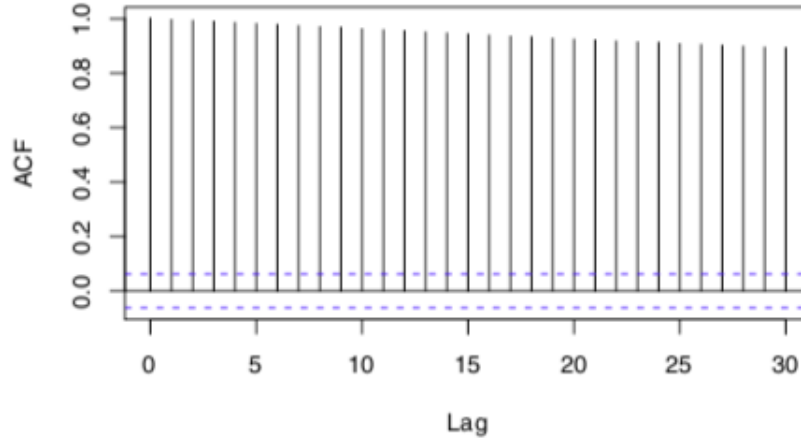
FIGURE 5.4 – The correlogram for the simulated random walk. A gradual decay from a high serial correlation is a notable feature of a random walk series.

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

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

5.4.3   Autoregressive Models

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

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

where α_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

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

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

27

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

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

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

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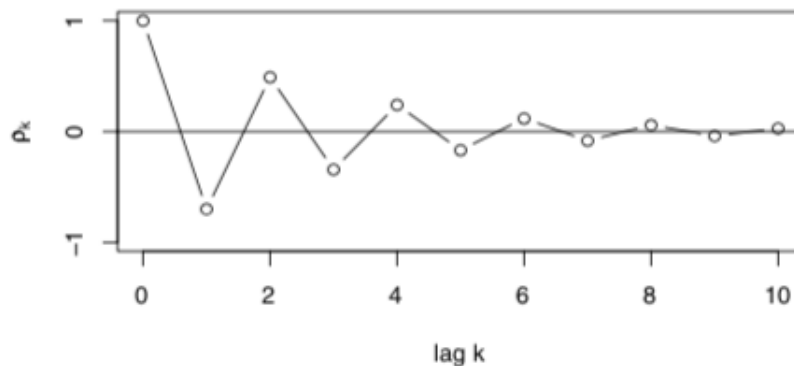$$\rho = \alpha^k$$

FIGURE 5.5 – For an AR model, the correlograms decays to zero. The correlogram decays to zero more rapidly for small α.

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

5.5 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 **deterministics trend** is the when we make short term forecast, we assume that the trend will change slowly.

Time series regression usually differs from a standart regression analysis because the residual form a time serie and therefore tend to be serially correlated. When this correlation is possitive, 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.

5.5.1 Linear Models

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

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

where $u_{i,t}$ is the value of the i$^{\text{th}}$ explanatory variable, α_i the estimated model parameters, z_t the error at time t.
An example of a linear model is the pth-order polynomial function of t :

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

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 roder m, the mth-order differencing is required to remove the trend.

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

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

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.

5.5.2 Linear Models with Seasonal Variables

Additive Seasonal Indicator Variables To include seasonal effect, we change the constant α_o depending on the season. Let s be the time serie measured (Ex. for monthly time serie, $s = 12$). 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 represente 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

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

where m_t is a trend model **that include a constant term** (α_0) and s_i and c_i are unknown parameters.

5.5.3 Forecasting from regression

When we predic a regression time series, we try to predict in the future. The problem is that the trend might change. Therefore, it is better ot think of
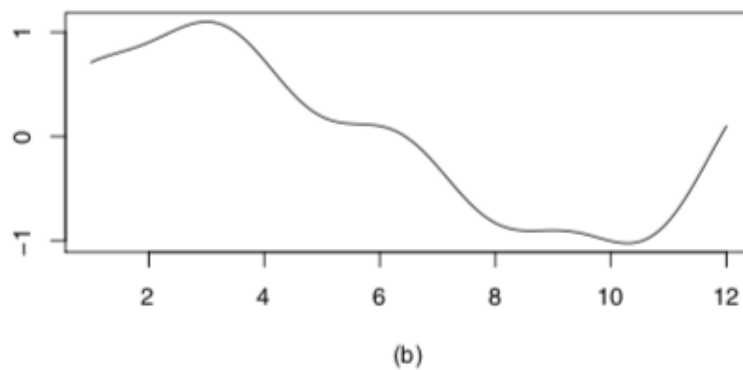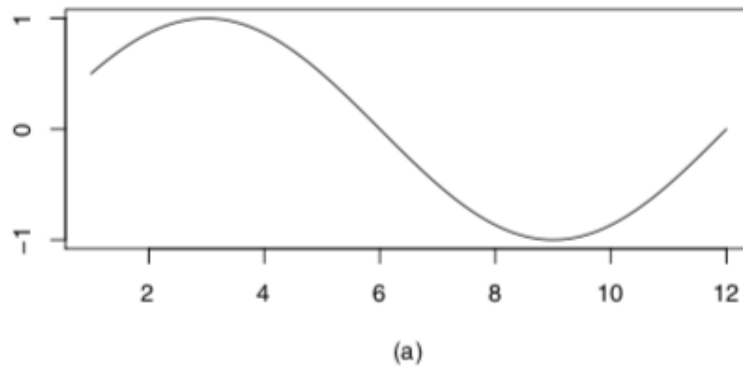
FIGURE 5.6 – Two possible underlying seasonal patterns for monthly series based on the harmonic model (Equation (5.10)). Plot (a) is of the first harmonic over a year and is usually too regular for most practical applications. Plot (b) is of the same wave but with a further two harmonics added. Plot (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

31

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

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

5.6 Stationary Models

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

5.6.1 Strictly Stationary Series

A time series model $\{x_t\}$ is *strictly stationry* if the joint statistical distribution $x_{t_1}, ..., x_{t_n}$ is the same as the joint distribution of $x_{t_1+m}, ..., x_{t_n+m}$ for all $t_1, ..., t_n$ 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.

5.6.2   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

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

where ϕ_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

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

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

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

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

5.6.3   Mized Models : The ARMA process

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

$$x_t = \alpha_1 x_{t-1} + \alpha_2 x_{t-2} + ... + \alpha_p x_{t-p} + w_t + \beta_1 w_{t-1} + \beta_2 w_{t-2} + ... + \beta_q w_{t-q}$$
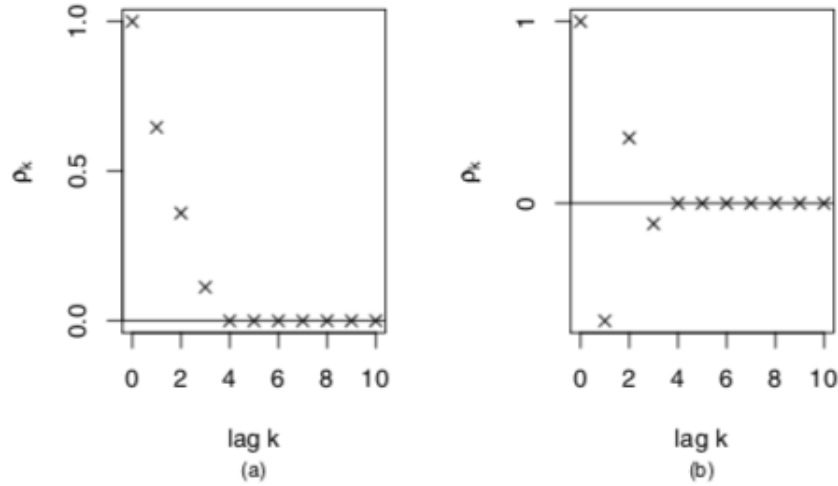
FIGURE 5.7 – Plots of the autocorrelation functions for two MA(3) processes. The autocorrelation for lag $k > q$ are all zero. $(a)\beta_1 = 0.7, \beta_2 = 0.5, \beta_3 = 0.2; (b)\beta_1 = -0.7, \beta_2 = 0.5, \beta_3 = -0.2$.

We can also express it as

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

- The process is stationary when the roots of θ are all exceed unity in absolute value.
- The process is invertible when the roots of ϕ 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 θ and ϕ share a common factor, a stationary model can be simplified. For example, the model :

$$(1 - \frac{1}{2}B)(1 - \frac{1}{3}B)x_t = (1 - frac12B)w_t$$
$$(1 - \frac{1}{3}B)x_t = w_t$$

34

Second-order Properties

$$\sigma^2 = \sigma_w^2 \left(1 + \frac{(\alpha + \beta)^2}{1 - \alpha^2} \right)$$

$$\gamma_0 = \sigma_w^2 \left(\frac{1 + 2\alpha\beta + \beta^2}{1 - \alpha^2} \right)$$

$$\gamma_k = \sigma_w^2 (\alpha + \beta) \alpha^{k-1} \left(\frac{1 + \alpha\beta}{1 - \alpha^2} \right)$$

$$\rho_k = \frac{\alpha^{k-1}(\alpha + \beta)(1 + \alpha\beta)}{1 + \alpha\beta + \beta^2} = \alpha\rho_{k-1}$$

5.7   Non-stationary Models

5.7.1   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 $I(1)$

5.7.2   Non-Seasonal ARIMA Models

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

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

In general :

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

5.7.3   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 ARIMA$(p, d, q)(P, D, Q)_s$ model can be defined as

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

Annexe A

Principales distribution de probabilité utilisées

introduction