Avd. Matematisk statistik

KTH Matematik

Formulas and survey
Time series analysis

Jan Grandell

Innehåll

1 Some notation

$\mathbb{R} = (-\infty, \infty)$
$\mathbb{Z} = \{0, \pm 1, \pm 2, \dots\}$
$\mathbb{C} =$ The complex numbers $= \{x + iy;\ x \in \mathbb{R},\ y \in \mathbb{R}\}$
$\stackrel{\text{def}}{=}$ means "equality by definition".

2 General probabilistic formulas

$(\Omega, \mathcal{F}, P)$ is a *probability space*, where:

Ω is the *sample space*, i.e. the set of all possible outcomes of an experiment.

$\mathcal{F}$ is a σ-*field* (or a σ-algebra), i.e.

(a) $\emptyset \in \mathcal{F}$;

(b) if $A_1, A_2, \dots \in \mathcal{F}$ then $\bigcup_1^\infty A_i \in \mathcal{F}$;

(c) if $A \in \mathcal{F}$ then $A^c \in \mathcal{F}$.

P is a *probability measure*, i.e. a function $\mathcal{F} \to [0, 1]$ satisfying

(a) $P(\Omega) = 1$;

(b) $P(A) = 1 - P(A^c)$;

(c) if $A_1, A_2, \dots \in \mathcal{F}$ are disjoint, then $P\left(\bigcup_1^\infty A_i\right) = \sum_1^\infty P(A_i)$.

Definition 2.1 *A* random variable X *defined on* $(\Omega, \mathcal{F}, P)$ *is a function* $\Omega \to \mathbb{R}$ *such that* $\{\omega \in \Omega :\ X(\omega) \le x\} \in \mathcal{F}$ *for all* $x \in \mathbb{R}$.

Let X be a random variable.

$F_X(x) = P\{X \le x\}$ is the *distribution function* (fördelningsfunktionen).

$f_X(\cdot)$, given by $F_X(x) = \int_{-\infty}^x f_X(y)\, dy$, is the *density function* (täthetsfunktionen).

$p_X(k) = P\{X = k\}$ is the *probability function* (sannolikhetsfunktionen).

$\phi_X(u) = E\big(e^{iuX}\big)$ is the *characteristic function* (karakteristiska funktionen).

Definition 2.2 *Let* $X_1, X_2, \dots$ *be a sequence of random variables. We say that* X_n converges in probability to *the real number* a*, written* $X_n \xrightarrow{P} a$*, if for every* $\varepsilon > 0$*,*

$$\lim_{n\to\infty} P(|X_n - a| > \varepsilon) = 0.$$

Definition 2.3 *Let $X_1, X_2, \ldots$ be a sequence of random variables with finite second moment. We say that X_n* converges in mean-square to *the random variable X, written $X_n \xrightarrow{\text{m.s.}} X$, if*

$$E[(X_n - X)^2] \to 0 \text{ as } n \to \infty.$$

An important property of mean-square convergence is that Cauchy-sequences do converge. More precisely this means that if $X_1, X_2, \ldots$ have finite second moment and if

$$E[(X_n - X_k)^2] \to 0 \text{ as } n, k \to \infty,$$

then there exists a random variable X with finite second moment such that $X_n \xrightarrow{\text{m.s.}} X$.

The space of square integrable random variables is complete *under mean-square convergence.*

2.1 Some distributions

The Binomial Distribution $X \sim \text{Bin}(n, p)$ if

$$p_X(k) = \binom{n}{k} p^k (1-p)^{n-k}, \quad k = 0, 1, \ldots, n \text{ and } \ 0 < p < 1.$$

$E(X) = np, \ \text{Var}(X) = np(1-p).$

The Poisson Distribution $X \sim \text{Po}(\lambda)$ if

$$p_X(k) = \frac{\lambda^k}{k!} e^{-\lambda}, \quad k = 0, 1, \ldots \text{ and } \lambda > 0.$$

$E(X) = \lambda, \ \text{Var}(X) = \lambda, \ \phi_X(u) = e^{-\lambda(1-e^{iu})}.$

The Exponential Distribution $X \sim \text{Exp}(\lambda)$ if

$$f_X(x) = \begin{cases} \frac{1}{\lambda} e^{-x/\lambda} & \text{if } x \geq 0, \\ 0 & \text{if } x < 0 \end{cases} \quad \lambda > 0.$$

$E(X) = \lambda, \ \text{Var}(X) = \lambda^2.$

The Standard Normal Distribution $X \sim N(0, 1)$ if

$$f_X(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}, \quad x \in \mathbb{R}.$$

$E(X) = 0, \ \text{Var}(X) = 1, \ \phi_X(u) = e^{-u^2/2}$.
The density function is often denoted by $\varphi(\cdot)$ and the distribution function by $\Phi(\cdot)$.

The Normal Distribution $X \sim N(\mu, \sigma^2)$ if

$$\frac{X-\mu}{\sigma} \sim N(0,1), \quad \mu \in \mathbb{R},\ \sigma > 0.$$

$E(X) = \mu,\ \mathrm{Var}(X) = \sigma^2,\ \phi_X(u) = e^{i\mu u} e^{-u^2\sigma^2/2}$.

The (multivariate) Normal Distribution
$\boldsymbol{Y} = (Y_1, \ldots, Y_m)' \sim N(\boldsymbol{\mu}, \Sigma)$, if there exists

$$\text{a vector } \boldsymbol{\mu} = \begin{pmatrix} \mu_1 \\ \vdots \\ \mu_m \end{pmatrix}, \text{ a matrix } B = \begin{pmatrix} b_{11} & \ldots & b_{1n} \\ \vdots & & \\ b_{m1} & \ldots & b_{mn} \end{pmatrix} \text{ with } \Sigma = BB'$$

and a random vector $\boldsymbol{X} = (X_1, \ldots, X_n)'$ with independent and $N(0,1)$-distributed components, such that $\boldsymbol{Y} = \boldsymbol{\mu} + B\boldsymbol{X}$. If

$$\begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix} \sim N\left(\begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix} \right) ,$$

then

$$Y_1 \text{ conditional on } Y_2 = y_2 \sim N\left(\mu_1 + \frac{\rho\sigma_1}{\sigma_2}(y_2 - \mu_2), (1-\rho^2)\sigma_1^2 \right).$$

More generally, if

$$\begin{pmatrix} \boldsymbol{Y}_1 \\ \boldsymbol{Y}_2 \end{pmatrix} \sim N\left(\begin{pmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{pmatrix}, \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix} \right) ,$$

then $\boldsymbol{Y}_1$ conditional on $\boldsymbol{Y}_2 = \boldsymbol{y}_2$

$$\sim N\big(\boldsymbol{\mu}_1 + \Sigma_{12}\Sigma_{22}^{-1}(\boldsymbol{y}_2 - \boldsymbol{\mu}_2), \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}\big) .$$

Asymptotic normality

Definition 2.4 *Let $Y_1, Y_2, \ldots$ be a sequence of random variables.*
$Y_n \sim \mathrm{AN}(\mu_n, \sigma_n^2)$ means that

$$\lim_{n\to\infty} P\left(\frac{Y_n - \mu_n}{\sigma_n} \le x \right) = \Phi(x).$$

Definition 2.5 *Let $\boldsymbol{Y}_1, \boldsymbol{Y}_2, \ldots$ be a sequence of random k-vectors.*
$\boldsymbol{Y}_n \sim \mathrm{AN}(\boldsymbol{\mu}_n, \Sigma_n)$ means that

(a) *$\Sigma_1, \Sigma_2, \ldots$ have no zero diagonal elements;*

(b) *$\boldsymbol{\lambda}'\boldsymbol{Y}_n \sim \mathrm{AN}(\boldsymbol{\lambda}'\boldsymbol{\mu}_n, \boldsymbol{\lambda}'\Sigma_n\boldsymbol{\lambda})$ for every $\boldsymbol{\lambda} \in \mathbb{R}^k$ such that $\boldsymbol{\lambda}'\Sigma_n\boldsymbol{\lambda} > 0$ for all sufficiently large n.*

2.2 Estimation

Let $x_1, \ldots, x_n$ be observations of random variables $X_1, \ldots, X_n$ with a (known) distribution depending on the unknown parameter θ. A point estimate (punktskattning) of θ is then the value $\widehat{\theta}(x_1, \ldots, x_n)$. In order to analyze the estimate we consider the estimator (stickprovsvariabeln) $\widehat{\theta}(X_1, \ldots, X_n)$. Some nice properties of an estimate are the following:

- An estimate $\widehat{\theta}$ of θ is *unbiased* (väntevärdesriktig) if $E(\widehat{\theta}(X_1, \ldots, X_n)) = \theta$ for all θ.
- An estimate $\widehat{\theta}$ of θ is *consistent* if $P(|\widehat{\theta}(X_1, \ldots, X_n) - \theta| > \varepsilon) \to 0$ for $n \to \infty$.
- If $\widehat{\theta}$ and θ^* are unbiased estimates of θ we say that $\widehat{\theta}$ is *more effective* than θ^* if $\mathrm{Var}(\widehat{\theta}(X_1, \ldots, X_n)) \leq \mathrm{Var}(\theta^*(X_1, \ldots, X_n))$ for all θ.

3 Stochastic processes

Definition 3.1 (Stochastic process) *A* stochastic process *is a family of random variables* $\{X_t,\, t \in T\}$ *defined on a probability space* $(\Omega, \mathcal{F}, P)$.

A stochastic process with $T \subset \mathbb{Z}$ is often called a *time series.*

Definition 3.2 (The distribution of a stochastic process) *Put*

$$\mathcal{T} = \{\boldsymbol{t} \in T^n : t_1 < t_2 < \cdots < t_n,\, n = 1, 2, \ldots\}.$$

The (finite-dimensional) distribution functions are the family $\{F_{\boldsymbol{t}}(\cdot),\, \boldsymbol{t} \in \mathcal{T}\}$ *defined by*

$$F_{\boldsymbol{t}}(\boldsymbol{x}) = P(X_{t_1} \leq x_1, \ldots, X_{t_n} \leq x_n), \quad \boldsymbol{t} \in T^n,\, \boldsymbol{x} \in \mathbb{R}^n.$$

With "the distribution of $\{X_t,\, t \in T \subset \mathbb{R}\}$åe mean the family $\{F_{\boldsymbol{t}}(\cdot),\, \boldsymbol{t} \in \mathcal{T}\}$.

Definition 3.3 *Let* $\{X_t,\, t \in T\}$ *be a stochastic process with* $\mathrm{Var}(X_t) < \infty$. *The* mean function *of* $\{X_t\}$ *is*

$$\mu_X(t) \stackrel{\text{def}}{=} E(X_t), \quad t \in T.$$

The covariance function *of* $\{X_t\}$ *is*

$$\gamma_X(r, s) = \mathrm{Cov}(X_r, X_s), \quad r, s \in T.$$

Definition 3.4 (Standard Brownian motion) *A* standard Brownian motion*, or a standard Wiener process* $\{B(t),\, t \geq 0\}$ *is a stochastic process satisfying*

(a) $B(0) = 0$*;*

(b) *for every* $\boldsymbol{t} = (t_0, t_1, \ldots, t_n)$ *with* $0 = t_0 < t_1 < \cdots < t_n$ *the random variables* $\Delta_1 = B(t_1) - B(t_0), \ldots, \Delta_n = B(t_n) - B(t_{n-1})$ *are independent;*

(c) $B(t) - B(s) \sim N(0, t-s)$ *for* $t \geq s$.

Definition 3.5 (Poisson process) *A* Poisson process $\{N(t),\, t \geq 0\}$ *with mean rate (or intensity)* λ *is a stochastic process satisfying*

(a) $N(0) = 0$*;*

(b) *for every* $\boldsymbol{t} = (t_0, t_1, \ldots, t_n)$ *with* $0 = t_0 < t_1 < \cdots < t_n$ *the random variables* $\Delta_1 = N(t_1) - N(t_0), \ldots, \Delta_n = N(t_n) - N(t_{n-1})$ *are independent;*

(c) $N(t) - N(s) \sim \mathrm{Po}(\lambda(t-s))$ *for* $t \geq s$.

Definition 3.6 (Gaussian time series) *The time series* $\{X_t,\, t \in \mathbb{Z}\}$ *is said to be a Gaussian time series if all finite-dimensional distributions are normal.*

4 Stationarity

Definition 4.1 *The time series* $\{X_t,\, t \in \mathbb{Z}\}$ *is said to be* strictly stationary *if the distributions of*

$$(X_{t_1}, \ldots, X_{t_k}) \quad \textit{and} \quad (X_{t_1+h}, \ldots, X_{t_k+h})$$

are the same for all k*, and all* $t_1, \ldots, t_k, h \in \mathbb{Z}$.

Definition 4.2 *The time series* $\{X_t,\, t \in \mathbb{Z}\}$ *is said to be* (weakly) stationary *if, see Definition 3.3 on the preceding page for notation,*

(i) $\mathrm{Var}(X_t) < \infty \quad$ *for all* $t \in \mathbb{Z}$,

(ii) $\mu_X(t) = \mu \quad$ *for all* $t \in \mathbb{Z}$,

(iii) $\gamma_X(r, s) = \gamma_X(r + t, s + t) \quad$ *for all* $r, s, t \in \mathbb{Z}$.

(iii) implies that $\gamma_X(r, s)$ is a function of $r - s$, and it is convenient to define

$$\gamma_X(h) \stackrel{\text{def}}{=} \gamma_X(h, 0).$$

The value "h"is referred to as the "lag".

Definition 4.3 *Let* $\{X_t,\, t \in \mathbb{Z}\}$ *be a stationary time series. The* autocovariance function (ACVF) *of* $\{X_t\}$ *is*

$$\gamma_X(h) = \mathrm{Cov}(X_{t+h}, X_t).$$

The autocorrelation function (ACF) *is*

$$\rho_X(h) \stackrel{\text{def}}{=} \frac{\gamma_X(h)}{\gamma_X(0)}.$$

5 Spectral theory

Definition 5.1 *The complex-valued time series* $\{X_t,\, t \in \mathbb{Z}\}$ *is said to be stationary if*

(i) $E|X_t|^2 < \infty \quad$ *for all* $t \in \mathbb{Z}$,

(ii) EX_t *is independent of* t *for all* $t \in \mathbb{Z}$,

(iii) $E[X_{t+h}\overline{X_t}]$ *is independent of* t *for all* $t \in \mathbb{Z}$.

Definition 5.2 *The autocovariance function* $\gamma(\cdot)$ *of a complex-valued stationary time series* $\{X_t\}$ *is*

$$\gamma(h) = E[X_{t+h}\overline{X_t}] - EX_{t+h}E\overline{X_t}.$$

Suppose that $\sum_{h=-\infty}^{\infty} |\gamma(h)| < \infty$. The function

$$f(\lambda) = \frac{1}{2\pi} \sum_{h=-\infty}^{\infty} e^{-ih\lambda}\gamma(h), \quad -\pi \le \lambda \le \pi, \tag{1}$$

is called the *spectral density* of the time series $\{X_t,\, t \in \mathbb{Z}\}$. We have the *spectral representation* of the ACVF

$$\gamma(h) = \int_{-\pi}^{\pi} e^{ih\lambda} f(\lambda)\, d\lambda.$$

For a real-valued time series f is symmetric, i.e. $f(\lambda) = f(-\lambda)$.
For any stationary time series the ACVF has the representation

$$\gamma(h) = \int_{(-\pi,\pi]} e^{ih\nu}\, dF(\nu) \quad \text{for all } h \in \mathbb{Z},$$

where *the spectral distribution function* $F(\cdot)$ is a right-continuous, non-decreasing, bounded function on $[-\pi, \pi]$ and $F(-\pi) = 0$.
The time series itself has a spectral representation

$$X_t = \int_{(-\pi,\pi]} e^{it\nu}\, dZ(\nu)$$

where $\{Z(\lambda),\, \lambda \in [-\pi, \pi]\}$ is an orthogonal-increment process.

Definition 5.3 (Orthogonal-increment process) *An orthogonal-increment process on* $[-\pi, \pi]$ *is a complex-valued process* $\{Z(\lambda)\}$ *such that*

$$\langle Z(\lambda), Z(\lambda)\rangle < \infty, \quad -\pi \le \lambda \le \pi,$$

$$\langle Z(\lambda), 1\rangle = 0, \quad -\pi \le \lambda \le \pi,$$

and

$$\langle Z(\lambda_4) - Z(\lambda_3), Z(\lambda_2) - Z(\lambda_1)\rangle = 0, \quad \text{if } (\lambda_1, \lambda_2] \cap (\lambda_3, \lambda_4] = \emptyset$$

where $\langle X, Y\rangle = EX\overline{Y}$.

6 Time series models

Definition 6.1 (White noise) *A process* $\{X_t,\, t \in \mathbb{Z}\}$ *is said to be a white noise with mean* μ *and variance* σ^2*, written*

$$\{X_t\} \sim \mathrm{WN}(\mu, \sigma^2),$$

if $EX_t = \mu$ *and* $\gamma(h) = \begin{cases} \sigma^2 & \text{if } h = 0, \\ 0 & \text{if } h \neq 0. \end{cases}$

A $\mathrm{WN}(\mu, \sigma^2)$ has spectral density

$$f(\lambda) = \frac{\sigma^2}{2\pi}, \quad -\pi \leq \lambda \leq \pi.$$

Definition 6.2 (Linear processes) *The process* $\{X_t,\, t \in \mathbb{Z}\}$ *is said to be a* linear process *if it has the representation*

$$X_t = \sum_{j=-\infty}^{\infty} \psi_j Z_{t-j}, \quad \{Z_t\} \sim \mathrm{WN}(0, \sigma^2),$$

where $\sum_{j=-\infty}^{\infty} |\psi_j| < \infty$.

A linear process is stationary with mean 0, autocovariance function

$$\gamma(h) = \sum_{j=-\infty}^{\infty} \psi_j \psi_{j+h} \sigma^2,$$

and spectral density

$$f(\lambda) = \frac{\sigma^2}{2\pi} |\psi(e^{-i\lambda})|^2, \quad -\pi \leq \lambda \leq \pi,$$

where $\psi(z) = \sum_{j=-\infty}^{\infty} \psi_j z^j$.

Definition 6.3 (IID noise) *A process* $\{X_t,\, t \in \mathbb{Z}\}$ *is said to be an IID noise with mean 0 and variance* σ^2*, written*

$$\{X_t\} \sim \mathrm{IID}(0, \sigma^2),$$

if the random variables X_t *are independent and identically distributed with* $EX_t = 0$ *and* $\mathrm{Var}(X_t) = \sigma^2$.

6.1 ARMA processes

Definition 6.4 (The ARMA(p, q) process) *The process* $\{X_t, t \in \mathbb{Z}\}$ *is said to be an* ARMA(p, q) process *if it is stationary and if*

$$X_t - \phi_1 X_{t-1} - \ldots - \phi_p X_{t-p} = Z_t + \theta_1 Z_{t-1} + \ldots + \theta_q Z_{t-q}, \tag{2}$$

where $\{Z_t\} \sim \text{WN}(0, \sigma^2)$. *We say that* $\{X_t\}$ *is an* ARMA(p, q) *process with mean* μ *if* $\{X_t - \mu\}$ *is an* ARMA(p, q) *process.*

Equations (2) can be written as

$$\phi(B)X_t = \theta(B)Z_t, \quad t \in \mathbb{Z},$$

where

$$\phi(z) = 1 - \phi_1 z - \ldots - \phi_p z^p,$$
$$\theta(z) = 1 + \theta_1 z + \ldots + \theta_q z^q,$$

and B is the backward *shift operator*, i.e. $(B^j X)_t = X_{t-j}$. The polynomials $\phi(\cdot)$ and $\theta(\cdot)$ are called *generating polynomials.*

Definition 6.5 *An* ARMA(p, q) *process defined by the equations*

$$\phi(B)X_t = \theta(B)Z_t \quad \{Z_t\} \sim \text{WN}(0, \sigma^2),$$

is said to be causal *if there exists constants* $\{\psi_j\}$ *such that* $\sum_{j=0}^{\infty} |\psi_j| < \infty$ *and*

$$X_t = \sum_{j=0}^{\infty} \psi_j Z_{t-j}, \quad t \in \mathbb{Z}. \tag{3}$$

Theorem 6.1 *Let* $\{X_t\}$ *be an* ARMA(p, q) *for which* $\phi(\cdot)$ *and* $\theta(\cdot)$ *have no common zeros. Then* $\{X_t\}$ *is causal if and only if* $\phi(z) \neq 0$ *for all* $|z| \leq 1$. *The coefficients* $\{\psi_j\}$ *in* (3) *are determined by the relation*

$$\psi(z) = \sum_{j=0}^{\infty} \psi_j z^j = \frac{\theta(z)}{\phi(z)}, \quad |z| \leq 1.$$

Definition 6.6 *An* ARMA(p, q) *process defined by the equations*

$$\phi(B)X_t = \theta(B)Z_t, \quad \{Z_t\} \sim \text{WN}(0, \sigma^2),$$

is said to be invertible *if there exists constants* $\{\pi_j\}$ *such that* $\sum_{j=0}^{\infty} |\pi_j| < \infty$ *and*

$$Z_t = \sum_{j=0}^{\infty} \pi_j X_{t-j}, \quad t \in \mathbb{Z}. \tag{4}$$

Theorem 6.2 *Let $\{X_t\}$ be an* ARMA(p,q) *for which $\phi(\cdot)$ and $\theta(\cdot)$ have no common zeros. Then $\{X_t\}$ is invertible if and only if $\theta(z) \neq 0$ for all $|z| \leq 1$. The coefficients $\{\pi_j\}$ in* (4) *are determined by the relation*

$$\pi(z) = \sum_{j=0}^{\infty} \pi_j z^j = \frac{\phi(z)}{\theta(z)}, \quad |z| \leq 1.$$

A causal and invertible ARMA(p,q) process has spectral density

$$f(\lambda) = \frac{\sigma^2}{2\pi} \frac{|\theta(e^{-i\lambda})|^2}{|\phi(e^{-i\lambda})|^2}, \quad -\pi \leq \lambda \leq \pi.$$

Definition 6.7 (The AR(p) process) *The process $\{X_t, t \in \mathbb{Z}\}$ is said to be an* AR(p) autoregressive process of order p *if it is stationary and if*

$$X_t - \phi_1 X_{t-1} - \ldots - \phi_p X_{t-p} = Z_t, \quad \{Z_t\} \sim \text{WN}(0, \sigma^2).$$

We say that $\{X_t\}$ is an AR(p) *process with mean μ if $\{X_t - \mu\}$ is an* AR(p) *process.*

A causal AR(p) process has spectral density

$$f(\lambda) = \frac{\sigma^2}{2\pi} \frac{1}{|\phi(e^{-i\lambda})|^2} \quad -\pi \leq \lambda \leq \pi.$$

Its ACVF is determined be the the *Yule-Walker equations*:

$$\gamma(k) - \phi_1 \gamma(k-1) - \ldots - \phi_p \gamma(k-p) = \begin{cases} 0, & k = 1, \ldots, p, \\ \sigma^2, & k = 0. \end{cases} \tag{5}$$

A causal AR(1) process defined by

$$X_t - \phi X_{t-1} = Z_t, \quad \{Z_t\} \sim \text{WN}(0, \sigma^2),$$

has ACVF

$$\gamma(h) = \frac{\sigma^2 \phi^{|h|}}{1 - \phi^2}$$

and spectral density

$$f(\lambda) = \frac{\sigma^2}{2\pi} \frac{1}{1 + \phi^2 - 2\phi\cos(\lambda)}, \quad -\pi \leq \lambda \leq \pi.$$

Definition 6.8 (The MA(q) process) *The process $\{X_t, t \in \mathbb{Z}\}$ is said to be a* moving average of order q *if*

$$X_t = Z_t + \theta_1 Z_{t-1} + \ldots + \theta_q Z_{t-q}, \quad \{Z_t\} \sim \text{WN}(0, \sigma^2),$$

where $\theta_1, \ldots, \theta_q$ are constants.

An invertible MA(1) process defined by

$$X_t = Z_t + \theta Z_{t-1}, \quad \{Z_t\} \sim \mathrm{WN}(0, \sigma^2),$$

has ACVF

$$\gamma(h) = \begin{cases} (1+\theta^2)\sigma^2 & \text{if } h = 0, \\ \theta\sigma^2 & \text{if } |h| = 1, \\ 0 & \text{if } |h| > 1. \end{cases}$$

and spectral density

$$f(\lambda) = \frac{\sigma^2}{2\pi}\left(1 + \theta^2 + 2\theta\cos(\lambda)\right), \quad -\pi \le \lambda \le \pi.$$

6.2 ARIMA and FARIMA processes

Definition 6.9 (The ARIMA(p, d, q) process) *Let d be a non-negative integer. The process $\{X_t, t \in \mathbb{Z}\}$ is said to be an* ARIMA(p, d, q) process *if $(1-B)^d X_t$ is a causal* ARMA(p, q) *process.*

Definition 6.10 (The FARIMA(p, d, q) process) *Let $0 < |d| < 0.5$. The process $\{X_t, t \in \mathbb{Z}\}$ is said to be a* fractionally integrated ARMA process *or a* FARIMA(p, d, q) process *if $\{X_t\}$ is stationary and satisfies*

$$\phi(B)(1-B)^d X_t = \theta(B) Z_t, \quad \{Z_t\} \sim \mathrm{WN}(0, \sigma^2).$$

6.3 Financial time series

Definition 6.11 (The ARCH(p) process) *The process $\{X_t, t \in \mathbb{Z}\}$ is said to be an* ARCH(p) process *if it is stationary and if*

$$X_t = \sigma_t Z_t, \qquad \{Z_t\} \sim \mathrm{IID}\, N(0,1),$$

where

$$\sigma_t^2 = \alpha_0 + \alpha_1 X_{t-1}^2 + \ldots + \alpha_p X_{t-p}^2$$

and $\alpha_0 > 0$, $\alpha_j \ge 0$ for $j = 1, \ldots, p$, and if Z_t and $X_{t-1}, X_{t-2}, \ldots$ are independent for all t.

Definition 6.12 (The GARCH(p, q) process) *The process $\{X_t, t \in \mathbb{Z}\}$ is said to be an* GARCH(p, q) process *if it is stationary and if*

$$X_t = \sigma_t Z_t, \qquad \{Z_t\} \sim \mathrm{IID}\, N(0,1),$$

where

$$\sigma_t^2 = \alpha_0 + \alpha_1 X_{t-1}^2 + \ldots + \alpha_p X_{t-p}^2 + \beta_1 \sigma_{t-1}^2 + \ldots + \beta_q \sigma_{t-q}^2$$

and $\alpha_0 > 0$, $\alpha_j \ge 0$ for $j = 1, \ldots, p$, $\beta_k \ge 0$ for $k = 1, \ldots, q$, and if Z_t and $X_{t-1}, X_{t-2}, \ldots$ are independent for all t.

7 Prediction

Let $X_1, X_2, \ldots, X_n$ and Y be any random variables with finite means and variances. Put $\mu_i = E(X_i),\ \mu = E(Y)$,

$$\Gamma_n = \begin{pmatrix} \gamma_{1,1} & \ldots & \gamma_{1,n} \\ \vdots & & \\ \gamma_{n,1} & \ldots & \gamma_{n,n} \end{pmatrix} = \begin{pmatrix} \mathrm{Cov}(X_1, X_1) & \ldots & \mathrm{Cov}(X_1, X_n) \\ \vdots & & \\ \mathrm{Cov}(X_n, X_1) & \ldots & \mathrm{Cov}(X_n, X_n) \end{pmatrix}$$

and

$$\boldsymbol{\gamma}_n = \begin{pmatrix} \gamma_1 \\ \vdots \\ \gamma_n \end{pmatrix} = \begin{pmatrix} \mathrm{Cov}(X_1, Y) \\ \vdots \\ \mathrm{Cov}(X_n, Y) \end{pmatrix}.$$

Definition 7.1 *The* best linear predictor $\widehat{Y}$ *of* Y *in terms of* $X_1, X_2, \ldots, X_n$ *is a random variable of the form*

$$\widehat{Y} = a_0 + a_1 X_1 + \ldots + a_n X_n$$

such that $E\big[(Y - \widehat{Y})^2\big]$ *is minimized with respect to* $a_0, \ldots, a_n$. $E\big[(Y - \widehat{Y})^2\big]$ *is called the* mean-squared error.

It is often convenient to use the notation

$$P_{\overline{\mathrm{sp}}\{1,\, X_1, \ldots, X_n\}} Y \stackrel{\mathrm{def}}{=} \widehat{Y}.$$

The predictor is given by

$$\widehat{Y} = \mu + a_1(X_1 - \mu_1) + \ldots + a_n(X_n - \mu_n)$$

where

$$\boldsymbol{a}_n = \begin{pmatrix} a_1 \\ \vdots \\ a_n \end{pmatrix}$$

satisfies $\boldsymbol{\gamma}_n = \Gamma_n \boldsymbol{a}_n$. If Γ_n is non-singular we have $\boldsymbol{a}_n = \Gamma_n^{-1} \boldsymbol{\gamma}_n$.

- *There is no restriction to assume all means to be* 0.
- *The predictor* $\widehat{Y}$ *of* Y *is determined by*

$$\mathrm{Cov}(\widehat{Y} - Y, X_i) = 0, \quad \text{for } i = 1, \ldots, n.$$

7.1 Prediction for stationary time series

Theorem 7.1 *If* $\{X_t\}$ *is a zero-mean stationary time series such that* $\gamma(0) > 0$ *and* $\gamma(h) \to 0$ *as* $h \to \infty$, *the best linear predictor* $\widehat{X}_{n+1}$ *of* X_{n+1} *in terms of* $X_1, X_2, \ldots, X_n$ *is*

$$\widehat{X}_{n+1} = \sum_{i=1}^{n} \phi_{n,i} X_{n+1-i}, \quad n = 1, 2, \ldots,$$

where

$$\boldsymbol{\phi}_n = \begin{pmatrix} \phi_{n,1} \\ \vdots \\ \phi_{n,n} \end{pmatrix} = \Gamma_n^{-1}\boldsymbol{\gamma}_n, \quad \boldsymbol{\gamma}_n = \begin{pmatrix} \gamma(1) \\ \vdots \\ \gamma(n) \end{pmatrix} \quad \textit{and}$$

$$\Gamma_n = \begin{pmatrix} \gamma(1-1) & \dots & \gamma(1-n) \\ \vdots & & \\ \gamma(n-1) & \dots & \gamma(n-n) \end{pmatrix}.$$

The mean-squared error is $v_n = \gamma(0) - \boldsymbol{\gamma}_n'\Gamma_n^{-1}\boldsymbol{\gamma}_n$.

Theorem 7.2 (The Durbin–Levinson Algorithm) *If* $\{X_t\}$ *is a zero-mean stationary time series such that* $\gamma(0) > 0$ *and* $\gamma(h) \to 0$ *as* $h \to \infty$, *then* $\phi_{1,1} = \gamma(1)/\gamma(0)$, $v_0 = \gamma(0)$,

$$\phi_{n,n} = \left[\gamma(n) - \sum_{j=1}^{n-1} \phi_{n-1,j}\gamma(n-j)\right] v_{n-1}^{-1}$$

$$\begin{pmatrix} \phi_{n,1} \\ \vdots \\ \phi_{n,n-1} \end{pmatrix} = \begin{pmatrix} \phi_{n-1,1} \\ \vdots \\ \phi_{n-1,n-1} \end{pmatrix} - \phi_{n,n} \begin{pmatrix} \phi_{n-1,n-1} \\ \vdots \\ \phi_{n-1,1} \end{pmatrix}$$

and

$$v_n = v_{n-1}[1 - \phi_{n,n}^2].$$

Theorem 7.3 (The Innovations Algorithm) *If* $\{X_t\}$ *has zero-mean and* $E(X_iX_j) = \kappa(i,j)$, *where the matrix* $\begin{pmatrix} \kappa(1,1) & \dots & \kappa(1,n) \\ \vdots & & \\ \kappa(n,1) & \dots & \kappa(n,n) \end{pmatrix}$ *is non-singular, we have*

$$\widehat{X}_{n+1} = \begin{cases} 0 & \textit{if } n = 0, \\ \sum_{j=1}^{n} \theta_{n,j}(X_{n+1-j} - \widehat{X}_{n+1-j}) & \textit{if } n \geq 1, \end{cases} \tag{6}$$

and

$$\begin{aligned} v_0 &= \kappa(1,1), \\ \theta_{n,n-k} &= v_k^{-1}\left(\kappa(n+1,k+1) - \sum_{j=0}^{k-1} \theta_{k,k-j}\theta_{n,n-j}v_j\right), \ k = 0,\dots,n-1, \\ v_n &= \kappa(n+1,n+1) - \sum_{j=0}^{n-1} \theta_{n,n-j}^2 v_j. \end{aligned}$$

7.2 Prediction of an ARMA Process

Let $\{X_t\}$ be a causal ARMA(p,q) process defined by $\phi(B)X_t = \theta(B)Z_t$. Then

$$\widehat{X}_{n+1} = \begin{cases} \sum_{j=1}^{n} \theta_{n,j}(X_{n+1-j} - \widehat{X}_{n+1-j}) & \text{if } 1 \le n < m, \\ \phi_1 X_n + \cdots + \phi_p X_{n+1-p} & \\ \quad + \sum_{j=1}^{q} \theta_{n,j}(X_{n+1-j} - \widehat{X}_{n+1-j}) & \text{if } n \ge m, \end{cases}$$

where $m = \max(p,q)$. The θ_{nj}:s are obtained by the innovations algorithm applied to

$$\begin{cases} W_t = \sigma^{-1} X_t, & \text{if } t = 1, \ldots, m, \\ W_t = \sigma^{-1}\phi(B)X_t, & \text{if } t > m. \end{cases}$$

8 Partial correlation

Definition 8.1 *Let Y_1, Y_2 and $W_1, \ldots, W_k$ be random variables. The* partial correlation coefficient *of Y_1 and Y_2 with respect to $W_1, \ldots, W_k$ is defined by*

$$\alpha(Y_1, Y_2) \stackrel{\text{def}}{=} \rho(Y_1 - \widehat{Y}_1, Y_2 - \widehat{Y}_2),$$

where $\widehat{Y}_1 = P_{\overline{\text{sp}}\{1, W_1, \ldots, W_k\}} Y_1$ and $\widehat{Y}_2 = P_{\overline{\text{sp}}\{1, W_1, \ldots, W_k\}} Y_2$.

8.1 Partial autocorrelation

Definition 8.2 *Let $\{X_t,\, t \in \mathbb{Z}\}$ be a zero-mean stationary time series. The* partial autocorrelation function (PACF) *of $\{X_t\}$ is defined by*

$$\alpha(0) = 1,$$

$$\alpha(1) = \rho(1),$$

$$\alpha(h) = \rho(X_{h+1} - P_{\overline{\text{sp}}\{X_2, \ldots, X_h\}} X_{h+1}, X_1 - P_{\overline{\text{sp}}\{X_2, \ldots, X_h\}} X_1), \quad h \ge 2.$$

Theorem 8.1 *Under the assumptions of Theorem 7.2 $\alpha(h) = \phi_{h,h}$ for $h \ge 1$.*

9 Linear filters

A *filter* is an operation on a time series $\{X_t\}$ in order to obtain a new time series $\{Y_t\}$. $\{X_t\}$ is called the *input* and $\{Y_t\}$ the *output*. The following operation

$$Y_t = \sum_{k=-\infty}^{\infty} c_{t,k} X_k$$

defines a *linear filter*. A filter is called *time-invariant* if $c_{t,k}$ depends only on $t-k$, i.e. if

$$c_{t,k} = h_{t-k}.$$

A time-invariant linear filter (TLF) is said to by *causal* if

$$h_j = 0 \quad \text{for } j < 0,$$

A TLF is called *stable* if $\sum_{-\infty}^{\infty} |h_k| < \infty$.

Put $h(z) = \sum_{-\infty}^{\infty} h_k z^k$. Then $Y = h(B)X$. The function $h(e^{-i\lambda})$ is called the *transfer function* (överföringsfunktion eller frekvenssvarsfunktion). The function $|h(e^{-i\lambda})|^2$ is called the *power transfer function.*

Theorem 9.1 *Let $\{X_t\}$ be a possibly complex-valued stationary input in a stable TLF $h(B)$ and let $\{Y_t\}$ be the output, i.e. $Y = h(B)X$. Then*

(a) *$EY_t = h(1)EX_t$;*

(b) *Y_t is stationary;*

(c) *$F_Y(\lambda) = \int_{(-\pi,\lambda]} |h(e^{-i\nu})|^2 \, dF_X(\nu), \quad$ for $\lambda \in [-\pi, \pi]$.*

10 Estimation in time series

Definition 10.1 (Strictly linear time series) *A stationary time series $\{X_t\}$ is called* strictly linear *if it has the representation*

$$X_t = \mu + \sum_{j=-\infty}^{\infty} \psi_j Z_{t-j}, \qquad \{Z_t\} \sim \text{IID}(0, \sigma^2).$$

10.1 Estimation of μ

Consider $\overline{X}_n = \frac{1}{n} \sum_{j=1}^{n} X_j$, which is a natural unbiased estimate of μ.

Theorem 10.1 *If $\{X_t\}$ is a stationary time series with mean μ and autocovariance function $\gamma(\cdot)$, then as $n \to \infty$,*

$$\text{Var}(\overline{X}_n) = E[(\overline{X}_n - \mu)^2] \to 0 \quad \textit{if } \gamma(n) \to 0,$$

and

$$n \, \text{Var}(\overline{X}_n) \to \sum_{h=-\infty}^{\infty} \gamma(h) = 2\pi f(0) \quad \textit{if} \sum_{h=-\infty}^{\infty} |\gamma(h)| < \infty.$$

Theorem 10.2 *If $\{X_t\}$ is a strictly linear time series where $\sum_{j=-\infty}^{\infty} |\psi_j| < \infty$ and $\sum_{j=-\infty}^{\infty} \psi_j \neq 0$, then*

$$\overline{X}_n \sim \text{AN}\Big(\mu, \frac{v}{n}\Big),$$

where $v = \sum_{h=-\infty}^{\infty} \gamma(h) = \sigma^2 \left(\sum_{j=-\infty}^{\infty} \psi_j\right)^2$.

The notion AN is found in Definitions 2.4 and 2.5 on page 4.

10.2 Estimation of $\gamma(\cdot)$ and $\rho(\cdot)$

Consider

$$\widehat{\gamma}(h) = \frac{1}{n}\sum_{t=1}^{n-h}(X_t - \overline{X}_n)(X_{t+h} - \overline{X}_n), \quad 0 \le h \le n-1,$$

and

$$\widehat{\rho}(h) = \frac{\widehat{\gamma}(h)}{\widehat{\gamma}(0)},$$

respectively.

Theorem 10.3 *If $\{X_t\}$ is a strictly linear time series where $\sum_{j=-\infty}^{\infty}|\psi_j| < \infty$ and $EZ_t^4 = \eta\sigma^4 < \infty$, then*

$$\begin{pmatrix}\widehat{\gamma}(0)\\ \vdots \\ \widehat{\gamma}(h)\end{pmatrix} \sim \text{AN}\left(\begin{pmatrix}\gamma(0)\\ \vdots \\ \gamma(h)\end{pmatrix}, n^{-1}V\right),$$

where $V = (v_{ij})_{i,j=0,\ldots,h}$ is the covariance matrix and

$$v_{ij} = (\eta-3)\gamma(i)\gamma(j) + \sum_{k=-\infty}^{\infty}\{\gamma(k)\gamma(k-i+j) + \gamma(k+j)\gamma(k-i)\}.$$

Note: If $\{Z_t,\, t \in \mathbb{Z}\}$ is Gaussian, then $\eta = 3$. □

Theorem 10.4 *If $\{X_t\}$ is a strictly linear time series where $\sum_{j=-\infty}^{\infty}|\psi_j| < \infty$ and $EZ_t^4 < \infty$, then*

$$\begin{pmatrix}\widehat{\rho}(1)\\ \vdots \\ \widehat{\rho}(h)\end{pmatrix} \sim \text{AN}\left(\begin{pmatrix}\rho(1)\\ \vdots \\ \rho(h)\end{pmatrix}, n^{-1}W\right),$$

where $W = (w_{ij})_{i,j=1,\ldots,h}$ is the covariance matrix and

$$w_{ij} = \sum_{k=-\infty}^{\infty}\{\rho(k+i)\rho(k+j) + \rho(k-i)\rho(k+j) + 2\rho(i)\rho(j)\rho^2(k) - 2\rho(i)\rho(k)\rho(k+j) - 2\rho(j)\rho(k)\rho(k+i)\}. \quad (7)$$

In the following theorem, the assumption $EZ_t^4 < \infty$ is relaxed at the expense of a slightly stronger assumption on the sequence $\{\psi_j\}$.

Theorem 10.5 *If $\{X_t\}$ is a strictly linear time series where $\sum_{j=-\infty}^{\infty}|\psi_j| < \infty$ and $\sum_{j=-\infty}^{\infty}\psi_j^2|j| < \infty$, then*

$$\begin{pmatrix}\widehat{\rho}(1)\\ \vdots \\ \widehat{\rho}(h)\end{pmatrix} \sim \text{AN}\left(\begin{pmatrix}\rho(1)\\ \vdots \\ \rho(h)\end{pmatrix}, n^{-1}W\right),$$

where W is given by the previous theorem.

10.3 Estimation of the spectral density

The *Fourier frequencies* are given by $\omega_j = \frac{2\pi j}{n},\ -\pi < \omega_j \le \pi$. Put

$$F_n \stackrel{\text{def}}{=} \{j \in \mathbb{Z},\ -\pi < \omega_j \le \pi\} = \left\{ -\left[\frac{n-1}{2}\right], \ldots, \left[\frac{n}{2}\right] \right\},$$

where $[x]$ denotes the integer part of x.

10.3.1 The periodogram

Definition 10.2 *The* periodogram $I_n(\cdot)$ *of* $\{X_1, \ldots, X_n\}$ *is defined by*

$$I_n(\omega_j) = \frac{1}{n}\left|\sum_{t=1}^{n} X_t e^{-it\omega_j}\right|^2, \quad j \in F_n.$$

Definition 10.3 (Extension of the periodogram) *For any* $\omega \in [-\pi, \pi]$ *we define*

$$I_n(\omega) = \begin{cases} I_n(\omega_k) & \text{if } \omega_k - \pi/n < \omega \le \omega_k + \pi/n \text{ and } 0 \le \omega \le \pi, \\ I_n(-\omega) & \text{if } \omega \in [-\pi, 0). \end{cases}$$

Theorem 10.6 *We have*

$$EI_n(0) - n\mu^2 \to 2\pi f(0) \quad \text{as } n \to \infty$$

and

$$EI_n(\omega) \to 2\pi f(\omega) \quad \text{as } n \to \infty \text{ if } \omega \neq 0.$$

(If $\mu = 0$ *then* $I_n(\omega)$ *converges uniformly to* $2\pi f(\omega)$ *on* $[-\pi, \pi)$*.)*

Theorem 10.7 *Let* $\{X_t\}$ *be a strictly linear time series with*

$$\mu = 0, \quad \sum_{j=-\infty}^{\infty} |\psi_j||j|^{1/2} < \infty \ \text{ and } \ EZ^4 < \infty.$$

Then

$$\mathrm{Cov}(I_n(\omega_j), I_n(\omega_k)) = \begin{cases} 2(2\pi)^2 f^2(\omega_j) + O(n^{-1/2}) & \text{if } \omega_j = \omega_k = 0 \text{ or } \pi, \\ (2\pi)^2 f^2(\omega_j) + O(n^{-1/2}) & \text{if } 0 < \omega_j = \omega_k < \pi, \\ O(n^{-1}) & \text{if } \omega_j \neq \omega_k. \end{cases}$$

10.3.2 Smoothing the periodogram

Definition 10.4 *The estimator* $\widehat{f}(\omega) = \widehat{f}(g(n, \omega))$ *with*

$$\widehat{f}(\omega_j) = \frac{1}{2\pi} \sum_{|k| \le m_n} W_n(k) I_n(\omega_{j+k}),$$

where

$$m_n \to \infty \quad \textit{and} \quad m_n/n \to 0 \quad \textit{as} \quad n \to \infty,$$
$$W_n(k) = W_n(-k), \quad W_n(k) \geq 0, \quad \textit{for all } k,$$
$$\sum_{|k| \leq m_n} W_n(k) = 1,$$

and

$$\sum_{|k| \leq m_n} W_n^2(k) \to 0 \quad \textit{as} \quad n \to \infty,$$

is called a discrete spectral average estimator *of* $f(\omega)$.
(*If* $\omega_{j+k} \notin [-\pi, \pi]$ *the term* $I_n(\omega_{j+k})$ *is evaluated by defining* I_n *to have period* 2π.)

Theorem 10.8 *Let* $\{X_t\}$ *be a strictly linear time series with*

$$\mu = 0, \quad \sum_{j=-\infty}^{\infty} |\psi_j||j|^{1/2} < \infty \quad \textit{and} \quad EZ^4 < \infty.$$

Then

$$\lim_{n\to\infty} E\widehat{f}(\omega) = f(\omega)$$

and

$$\lim_{n\to\infty} \frac{1}{\sum_{|k|\leq m_n} W_n^2(k)} \operatorname{Cov}(\widehat{f}(\omega), \widehat{f}(\lambda)) = \begin{cases} 2f^2(\omega) & \textit{if } \omega = \lambda = 0 \textit{ or } \pi, \\ f^2(\omega) & \textit{if } 0 < \omega = \lambda < \pi, \\ 0 & \textit{if } \omega \neq \lambda. \end{cases}$$

Remark 10.1 If $\mu \neq 0$ we ignore $I_n(0)$. Thus we can use

$$\widehat{f}(0) = \frac{1}{2\pi}\left(W_n(0)I_n(\omega_1) + 2\sum_{k=1}^{m_n} W_n(k)I_n(\omega_{k+1}) \right).$$

Moreover, whenever $I_n(0)$ appears in $\widehat{f}(\omega_j)$ we replace it with $\widehat{f}(0)$. □

Example 10.1 (The simple moving average estimate) For this estimate we have

$$W_n(k) = \begin{cases} 1/(2m_n + 1) & \text{if } |k| \leq m_n, \\ 0 & \text{if } |k| > m_n, \end{cases}$$

and

$$\operatorname{Var}(\widehat{f}(\omega)) \sim \begin{cases} \frac{1}{m_n} f^2(\omega) & \text{if } \omega = 0 \text{ or } \pi, \\ \frac{1}{m_n} \frac{f^2(\omega)}{2} & \text{if } 0 < \omega < \pi. \end{cases}$$

□

11 Estimation for ARMA models

11.1 Yule-Walker estimation

Consider a causal zero-mean AR(p) process $\{X_t\}$:

$$X_t - \phi_1 X_{t-1} - \ldots - \phi_p X_{t-p} = Z_t, \quad \{Z_t\} \sim \text{IID}(0, \sigma^2).$$

The Yule-Walker equations (5) on page 10 can be written on the form

$$\Gamma_p \boldsymbol{\phi} = \boldsymbol{\gamma}_p \quad \text{and} \quad \sigma^2 = \gamma(0) - \boldsymbol{\phi}' \boldsymbol{\gamma}_p,$$

where

$$\Gamma_p = \begin{pmatrix} \gamma(0) & \ldots & \gamma(p-1) \\ \vdots & & \\ \gamma(p-1) & \ldots & \gamma(0) \end{pmatrix} \quad \text{and} \quad \boldsymbol{\gamma}_p = \begin{pmatrix} \gamma(1) \\ \vdots \\ \gamma(p) \end{pmatrix}.$$

If we replace Γ_p and $\boldsymbol{\gamma}_p$ with the estimates $\widehat{\Gamma}_p$ and $\widehat{\boldsymbol{\gamma}}_p$ we obtain the following equations for the *Yule-Walker estimates*

$$\widehat{\Gamma}_p \widehat{\boldsymbol{\phi}} = \widehat{\boldsymbol{\gamma}}_p \quad \text{and} \quad \widehat{\sigma}^2 = \widehat{\gamma}(0) - \widehat{\boldsymbol{\phi}}' \widehat{\boldsymbol{\gamma}}_p,$$

where

$$\widehat{\Gamma}_p = \begin{pmatrix} \widehat{\gamma}(0) & \ldots & \widehat{\gamma}(p-1) \\ \vdots & & \\ \widehat{\gamma}(p-1) & \ldots & \widehat{\gamma}(0) \end{pmatrix} \quad \text{and} \quad \widehat{\boldsymbol{\gamma}}_p = \begin{pmatrix} \widehat{\gamma}(1) \\ \vdots \\ \widehat{\gamma}(p) \end{pmatrix}.$$

Theorem 11.1 *If $\{X_t\}$ is a causal* AR(p) *process with $\{Z_t\} \sim \text{IID}(0, \sigma^2)$, and $\widehat{\boldsymbol{\phi}}$ is the Yule-Walker estimate of $\boldsymbol{\phi}$, then*

$$\widehat{\boldsymbol{\phi}} \sim \text{AN}\left(\boldsymbol{\phi}, \frac{\sigma^2 \Gamma_p^{-1}}{n}\right), \quad \textit{for large values of } n.$$

Moreover,

$$\widehat{\sigma}^2 \xrightarrow{P} \sigma^2.$$

A usual way to proceed is as if $\{X_t\}$ were an AR(m) process for $m = 1, 2, \ldots$ until we believe that $m \geq p$. In that case we can use the Durbin-Levinson algorithm, see Theorem 7.2 on page 13, with $\gamma(\cdot)$ replaced by $\widehat{\gamma}(\cdot)$.

11.2 Burg's algorithm

Assume as usual that $x_1, \ldots, x_n$ are the observations. The idea is to consider one observation after the other and to "predict"it both by forward and backward data. The forward and backward prediction errors $\{u_i(t)\}$ and $\{v_i(t)\}$ satisfy the recursions

$$u_0(t) = v_0(t) = x_{n+1-t},$$
$$u_i(t) = u_{i-1}(t-1) - \phi_{ii} v_{i-1}(t),$$

and

$$v_i(t) = v_{i-1}(t) - \phi_{ii} u_{i-1}(t-1).$$

Suppose now that we know $\phi_{i-1,k}$ for $k = 1, \ldots, i-1$ and ϕ_{ii}. Then $\phi_{i,k}$ for $k = 1, \ldots, i-1$ may be obtained by the Durbin-Levinson algorithm. Thus the main problem is to obtain an algorithm for calculating ϕ_{ii} for $i = 1, 2, \ldots$

Burg's algorithm:

$$d(1) = \tfrac{1}{2}x_1^2 + x_2^2 + \ldots + x_{n-1}^2 + \tfrac{1}{2}x_n^2 \tag{8}$$

$$\phi_{ii}^{(B)} = \frac{1}{d(i)} \sum_{t=i+1}^{n} v_{i-1}(t) u_{i-1}(t-1) \tag{9}$$

$$\sigma_i^{(B)2} = \frac{d(i)\big(1 - \phi_{ii}^{(B)2}\big)}{n-i} \tag{10}$$

$$d(i+1) = d(i)\big(1 - \phi_{ii}^{(B)2}\big) - \tfrac{1}{2}v_i^2(i+1) - \tfrac{1}{2}u_i^2(n). \tag{11}$$

The Burg estimates for an AR(p) have the same statistical properties for large values of n as the Yule-Walker estimate, i.e. Theorem 11.1 on the preceding page holds.

11.3 The innovations algorithm

Since an MA(q) process

$$X_t = Z_t + \theta_1 Z_{t-1} + \ldots + \theta_q Z_{t-q}, \quad \{Z_t\} \sim \mathrm{IID}(0, \sigma^2),$$

has, by definition, an innovation representation, it is natural to use the innovations algorithm for prediction in a similar way as the Durbin-Levinson algorithm was used. Since, generally, q is unknown, we can try to fit MA models

$$X_t = Z_t + \widehat{\theta}_{m1} Z_{t-1} + \ldots + \widehat{\theta}_{mm} Z_{t-m}, \quad \{Z_t\} \sim \mathrm{IID}(0, \widehat{v}_m),$$

of orders $m = 1, 2, \ldots$, by means of the innovations algorithm.

Definition 11.1 (Innovations estimates of MA parameters)
If $\widehat{\gamma}(0) > 0$ we define the innovations estimates

$$\widehat{\boldsymbol{\theta}}_m = \begin{pmatrix} \widehat{\theta}_{m1} \\ \vdots \\ \widehat{\theta}_{mm} \end{pmatrix} \quad \textit{and} \quad \widehat{v}_m, \quad m = 1, 2, \ldots, n-1,$$

by the recursion relations

$$\begin{cases} \widehat{v}_0 & = \widehat{\gamma}(0), \\ \widehat{\theta}_{m,m-k} & = \widehat{v}_k^{-1}\left(\widehat{\gamma}(m-k) - \sum_{j=0}^{k-1} \widehat{\theta}_{m,m-j}\widehat{\theta}_{k,k-j}\widehat{v}_j\right), \ k = 0, \dots, m-1, \\ \widehat{v}_m & = \widehat{\gamma}(0) - \sum_{j=0}^{m-1} \widehat{\theta}_{m,m-j}^2 \widehat{v}_j. \end{cases}$$

This method works also for causal invertible ARMA processes. The following theorem gives asymptotic statistical properties of the innovations estimates.

Theorem 11.2 *Let* $\{X_t\}$ *be the causal invertible* ARMA *process* $\phi(B)X_t = \theta(B)Z_t, \{Z_t\} \sim \text{IID}(0, \sigma^2)$, $EZ_t^4 < \infty$, *and let* $\psi(z) = \sum_{j=0}^{\infty} \psi_j z^j = \frac{\theta(z)}{\phi(z)}, |z| \leq 1$ *(with* $\psi_0 = 1$ *and* $\psi_j = 0$ *for* $j < 0$*). Then for any sequence of positive integers* $\{m(n),\ n = 1, 2, \dots\}$ *such that* $m \to \infty$ *and* $m = o(n^{1/3})$ *as* $n \to \infty$, *we have for each fixed* k,

$$\begin{pmatrix} \widehat{\theta}_{m1} \\ \vdots \\ \widehat{\theta}_{mk} \end{pmatrix} \sim \text{AN}\left(\begin{pmatrix} \psi_1 \\ \vdots \\ \psi_k \end{pmatrix}, n^{-1}A\right),$$

where $A = (a_{ij})_{i,j=1,\dots,k}$ *and*

$$a_{ij} = \sum_{r=1}^{\min(i,j)} \psi_{i-r}\psi_{j-r}.$$

Moreover,

$$\widehat{v}_m \xrightarrow{P} \sigma^2.$$

11.4 The Hannan–Rissanen algorithm

Let $\{X_t\}$ be an ARMA(p, q) process:

$$X_t - \phi_1 X_{t-1} - \dots - \phi_p X_{t-p} = Z_t + \theta_1 Z_{t-1} + \dots + \theta_q Z_{t-q}, \quad \{Z_t\} \sim \text{IID}(0, \sigma^2).$$

The *Hannan–Rissanen algorithm* consists of the following two steps:

Step 1

A high order AR(m) model (with $m > \max(p, q)$) is fitted to the data by Yule-Walker estimation. If $\widehat{\phi}_{m1}, \dots, \widehat{\phi}_{mm}$ are the estimated coefficients, then Z_t is estimated by

$$\widehat{Z}_t = X_t - \widehat{\phi}_{m1}X_{t-1} - \dots - \widehat{\phi}_{mm}X_{t-m}, \quad t = m+1, \dots, n.$$

Step 2

The vector $\boldsymbol{\beta} = (\boldsymbol{\phi}, \boldsymbol{\theta})$ is estimated by least square regression of X_t onto

$$X_{t-1}, \dots, X_{t-p}, \widehat{Z}_{t-1}, \dots, \widehat{Z}_{t-q},$$

i.e. by minimizing

$$S(\boldsymbol{\beta}) = \sum_{t=m+1}^{n} (X_t - \phi_1 X_{t-1} - \ldots - \phi_p X_{t-p} - \theta_1 \widehat{Z}_{t-1} - \ldots - \theta_q \widehat{Z}_{t-q})^2$$

with respect to $\boldsymbol{\beta}$. This gives the Hannan–Rissanen estimator

$$\widehat{\boldsymbol{\beta}} = (Z'Z)^{-1} Z' \boldsymbol{X}_n \text{ provided } Z'Z \text{ is non-singular,}$$

where

$$\boldsymbol{X}_n = \begin{pmatrix} X_{m+1} \\ \vdots \\ X_n \end{pmatrix}$$

and

$$Z = \begin{pmatrix} X_m & X_{m-1} & \ldots & X_{m-p+1} & \widehat{Z}_m & \widehat{Z}_{m-1} & \ldots & \widehat{Z}_{m-q+1} \\ \vdots & \vdots & & & & & & \\ X_{n-1} & X_{n-2} & \ldots & X_{n-p} & \widehat{Z}_{n-1} & \widehat{Z}_{n-2} & \ldots & \widehat{Z}_{n-q} \end{pmatrix}.$$

The Hannan–Rissanen estimate of the white noise variance σ^2 is

$$\widehat{\sigma}^2_{\mathrm{HR}} = \frac{S(\widehat{\boldsymbol{\beta}})}{n-m}.$$

11.5 Maximum Likelihood and Least Square estimation

It is possible to obtain better estimates by the maximum likelihood method (under the assumption of Gaussian processes) or by the least square method. In the least square method we minimize

$$S(\boldsymbol{\phi}, \boldsymbol{\theta}) = \sum_{j=1}^{n} \frac{(X_j - \widehat{X}_j)^2}{r_{j-1}},$$

where $r_{j-1} = v_{j-1}/\sigma^2$, with respect to $\boldsymbol{\phi}$ and $\boldsymbol{\theta}$. The estimates has to be obtained by recursive methods, and the estimates discussed are natural starting values. The least square estimate of σ^2 is

$$\widehat{\sigma}^2_{\mathrm{LS}} = \frac{S(\widehat{\boldsymbol{\phi}}_{\mathrm{LS}}, \widehat{\boldsymbol{\theta}}_{\mathrm{LS}})}{n-p-q},$$

where $(\widehat{\boldsymbol{\phi}}_{\mathrm{LS}}, \widehat{\boldsymbol{\theta}}_{\mathrm{LS}})$ is the estimate obtained by minimizing $S(\boldsymbol{\phi}, \boldsymbol{\theta})$.
Let us assume, or at least act as if, the process is Gaussian. Then, for any fixed values of $\boldsymbol{\phi}$, $\boldsymbol{\theta}$, and σ^2, the innovations $X_1 - \widehat{X}_1, \ldots, X_n - \widehat{X}_n$ are independent and normally distributed with zero means and variances $v_0 = \sigma^2 r_0, \ldots, v_{n-1} = \sigma^2 r_{n-1}$. The likelihood function is then

$$L(\boldsymbol{\phi}, \boldsymbol{\theta}, \sigma^2) = \prod_{j=1}^{n} f_{X_j - \widehat{X}_j}(X_j - \widehat{X}_j) = \prod_{j=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2 r_{j-1}}} \exp\left\{ -\frac{(X_j - \widehat{X}_j)^2}{2\sigma^2 r_{j-1}} \right\}.$$

Proceeding "in the usual wayåe get

$$\ln L(\boldsymbol{\phi}, \boldsymbol{\theta}, \sigma^2) = -\frac{1}{2}\ln((2\pi\sigma^2)^n r_0 \cdots r_{n-1}) - \frac{S(\boldsymbol{\phi}, \boldsymbol{\theta})}{2\sigma^2}.$$

Obviously $r_0, \ldots, r_{n-1}$ depend on $\boldsymbol{\phi}$ and $\boldsymbol{\theta}$ but they do not depend on σ^2. To maximize $\ln L(\boldsymbol{\phi}, \boldsymbol{\theta}, \sigma^2)$ is the same as to minimize

$$\ell(\boldsymbol{\phi}, \boldsymbol{\theta}) = \ln(n^{-1}S(\boldsymbol{\phi}, \boldsymbol{\theta})) + n^{-1}\sum_{j=1}^{n} \ln r_{j-1},$$

which has to be done numerically.
In the causal and invertible case $r_n \to 1$ and therefore $n^{-1}\sum_{j=1}^{n} \ln r_{j-1}$ is asymptotically negligible compared with $\ln S(\boldsymbol{\phi}, \boldsymbol{\theta})$. Thus both methods – least square and maximum likelihood – give asymptotically the same result in that case.

11.6 Order selection

Assume now that we want to fit an ARMA(p, q) process to real data, i.e. we want to estimate p, q, $(\boldsymbol{\phi}, \boldsymbol{\theta})$, and σ^2. We restrict ourselves to maximum likelihood estimation. Then we maximize $L(\boldsymbol{\phi}, \boldsymbol{\theta}, \sigma^2)$, or – which is the same – minimize $-2\ln L(\boldsymbol{\phi}, \boldsymbol{\theta}, \sigma^2)$, where L is regarded as a function also of p and q. Most probably we will get very high values of p and q. Such a model will probably fit the given data very well, but it is more or less useless as a mathematical model, since it will probably not be lead to reasonable predictors nor describe a different data set well. It is therefore natural to introduce a "penalty factorto discourage the fitting of models with too many parameters. Instead of maximum likelihood estimation we may apply the *AICC Criterion*:

Choose p, q, and $(\boldsymbol{\phi}_p, \boldsymbol{\theta}_q)$, to minimize

$$\text{AICC} = -2\ln L(\boldsymbol{\phi}_p, \boldsymbol{\theta}_q, S(\boldsymbol{\phi}_p, \boldsymbol{\theta}_q)/n) + 2(p+q+1)n/(n-p-q-2).$$

(The letters AIC stand for "Akaike's Information Criterionänd the last C for "biased-Corrected".)
The AICC Criterion has certain nice properties, but also its drawbacks. In general one may say the order selection is genuinely difficult.

12 Multivariate time series

Let

$$\boldsymbol{X}_t \stackrel{\text{def}}{=} \begin{pmatrix} X_{t1} \\ \vdots \\ X_{tm} \end{pmatrix}, \quad t \in \mathbb{Z},$$

where each component is a time series. In that case we talk about *multivariate time series*.

The second-order properties of $\{\boldsymbol{X}_t\}$ are specified by the mean vector

$$\boldsymbol{\mu}_t \stackrel{\text{def}}{=} E\boldsymbol{X}_t = \begin{pmatrix} \mu_{t1} \\ \vdots \\ \mu_{tm} \end{pmatrix} = \begin{pmatrix} EX_{t1} \\ \vdots \\ EX_{tm} \end{pmatrix}, \quad t \in \mathbb{Z},$$

and the covariance matrices

$$\Gamma(t+h,t) \stackrel{\text{def}}{=} E[(\boldsymbol{X}_{t+h}-\boldsymbol{\mu}_{t+h})(\boldsymbol{X}_t-\boldsymbol{\mu}_t)'] = \begin{pmatrix} \gamma_{11}(t+h,t) & \dots & \gamma_{1m}(t+h,t) \\ \vdots & & \\ \gamma_{m1}(t+h,t) & \dots & \gamma_{mm}(t+h,t) \end{pmatrix}$$

where $\gamma_{ij}(t+h,t) \stackrel{\text{def}}{=} \text{Cov}(X_{t+h,i}, X_{t,j})$.

Definition 12.1 *The m-variate time series $\{\boldsymbol{X}_t,\, t \in \mathbb{Z}\}$ is said to be* (weakly) stationary *if*

(i) $\boldsymbol{\mu}_t = \boldsymbol{\mu}$ *for all $t \in \mathbb{Z}$,*

(ii) $\Gamma(r,s) = \Gamma(r+t, s+t)$ *for all $r, s, t \in \mathbb{Z}$.*

Item (ii) implies that $\Gamma(r,s)$ is a function of $r-s$, and it is convenient to define

$$\Gamma(h) \stackrel{\text{def}}{=} \Gamma(h,0).$$

Definition 12.2 (Multivariate white noise) *An m-variate process*

$$\{\boldsymbol{Z}_t,\, t \in \mathbb{Z}\}$$

is said to be a white noise with mean $\boldsymbol{\mu}$ and covariance matrix $\mathbb{\Sigma}$, written

$$\{\boldsymbol{Z}_t\} \sim \text{WN}(\boldsymbol{\mu}, \mathbb{\Sigma}),$$

if $E\boldsymbol{Z}_t = \boldsymbol{\mu}$ and $\Gamma(h) = \begin{cases} \mathbb{\Sigma} & \text{if } h = 0, \\ 0 & \text{if } h \neq 0. \end{cases}$

Definition 12.3 (The ARMA(p,q) process) *The process $\{\boldsymbol{X}_t,\, t \in \mathbb{Z}\}$ is said to be an* ARMA(p,q) process *if it is stationary and if*

$$\boldsymbol{X}_t - \Phi_1\boldsymbol{X}_{t-1} - \ldots - \Phi_p\boldsymbol{X}_{t-p} = \boldsymbol{Z}_t + \Theta_1\boldsymbol{Z}_{t-1} + \ldots + \Theta_q\boldsymbol{Z}_{t-q}, \tag{12}$$

where $\{\boldsymbol{Z}_t\} \sim \text{WN}(\boldsymbol{0}, \mathbb{\Sigma})$. We say that $\{\boldsymbol{X}_t\}$ is an ARMA(p,q) *process with mean $\boldsymbol{\mu}$ if $\{\boldsymbol{X}_t - \boldsymbol{\mu}\}$ is an* ARMA(p,q) *process.*

Equations (12) can be written as

$$\Phi(B)\boldsymbol{X}_t = \Theta(B)\boldsymbol{Z}_t, \quad t \in \mathbb{Z},$$

where

$$\Phi(z) = I - \Phi_1 z - \ldots - \Phi_p z^p,$$

$$\Theta(z) = I + \Theta_1 z + \ldots + \Theta_q z^q,$$

are matrix-valued polynomials.
Causality and invertibility are characterized in terms of the generating polynomials:

Causality: $\boldsymbol{X}_t$ is causal if $\det \Phi(z) \neq 0$ for all $|z| \leq 1$;

Invertibility: $\boldsymbol{X}_t$ is invertible if $\det \Theta(z) \neq 0$ for all $|z| \leq 1$.

Assume that

$$\sum_{h=-\infty}^{\infty} |\gamma_{ij}(h)| < \infty, \quad i, j = 1, \ldots, m. \tag{13}$$

Definition 12.4 (The cross spectrum) *Let* $\{\boldsymbol{X}_t, t \in \mathbb{Z}\}$ *be an* m*-variate stationary time series whose ACVF satisfies* (13)*. The function*

$$f_{jk}(\lambda) = \frac{1}{2\pi} \sum_{h=-\infty}^{\infty} e^{-ih\lambda} \gamma_{jk}(h), \quad -\pi \leq \lambda \leq \pi, \;\; j \neq k,$$

is called the cross spectrum *or* cross spectral density *of* $\{X_{tj}\}$ *and* $\{X_{tk}\}$*. The matrix*

$$f(\lambda) = \begin{pmatrix} f_{11}(\lambda) & \ldots & f_{1m}(\lambda) \\ \vdots & & \\ f_{m1}(\lambda) & \ldots & f_{mm}(\lambda) \end{pmatrix}$$

is called the spectrum *or* spectral density matrix *of* $\{\boldsymbol{X}_t\}$*.*

The spectral density matrix $f(\lambda)$ is non-negative definite for all $\lambda \in [-\pi, \pi]$.

13 Kalman filtering

We will use the notation

$$\{\boldsymbol{Z}_t\} \sim \mathrm{WN}(\boldsymbol{0}, \{\Sigma_t\}),$$

to indicate that the process $\{\boldsymbol{Z}_t\}$ has mean $\boldsymbol{0}$ and that

$$E\boldsymbol{Z}_s\boldsymbol{Z}_t' = \begin{cases} \Sigma_t & \text{if } s = t, \\ 0 & \text{otherwise.} \end{cases}$$

Notice this definition is an extension of Definition 12.2 on the page before in order to allow for non-stationarity.
A *state-space model* is defined by

the *state equation*

$$\boldsymbol{X}_{t+1} = F_t \boldsymbol{X}_t + \boldsymbol{V}_t, \quad t = 1, 2, \ldots, \tag{14}$$

where
$\{\boldsymbol{X}_t\}$ is a v-variate process describing the state of some system,
$\{\boldsymbol{V}_t\} \sim \mathrm{WN}(\mathbf{0}, \{Q_t\})$, and
$\{F_t\}$ is a sequence of $v \times v$ matrices

and

the *observation equation*

$$\boldsymbol{Y}_t = G_t \boldsymbol{X}_t + \boldsymbol{W}_t, \quad t = 1, 2, \ldots, \tag{15}$$

where
$\{\boldsymbol{Y}_t\}$ is a w-variate process describing the observed state of some system,
$\{\boldsymbol{W}_t\} \sim \mathrm{WN}(\mathbf{0}, \{R_t\})$, and
$\{G_t\}$ is a sequence of $w \times v$ matrices.
Further $\{\boldsymbol{W}_t\}$ and $\{\boldsymbol{V}_t\}$ are uncorrelated. To complete the specification it is assumed that the initial state $\boldsymbol{X}_1$ is uncorrelated with $\{\boldsymbol{W}_t\}$ and $\{\boldsymbol{V}_t\}$.

Definition 13.1 (State-space representation) *A time series $\{\boldsymbol{Y}_t\}$ has a* state-space representation *if there exists a state-space model for $\{\boldsymbol{Y}_t\}$ as specified by equations* (14) *and* (15).

Put

$$P_t(\boldsymbol{X}) \stackrel{\text{def}}{=} P(\boldsymbol{X} \mid \boldsymbol{Y}_0, \ldots, \boldsymbol{Y}_t),$$

i.e. the vector of best linear predictors of $X_1, \ldots, X_v$ in terms of all components of $\boldsymbol{Y}_0, \ldots, \boldsymbol{Y}_t$.
Linear estimation of $\boldsymbol{X}_t$ in terms of

- $\boldsymbol{Y}_0, \ldots, \boldsymbol{Y}_{t-1}$ defines the *prediction problem*;
- $\boldsymbol{Y}_0, \ldots, \boldsymbol{Y}_t$ defines the *filtering problem*;
- $\boldsymbol{Y}_0, \ldots, \boldsymbol{Y}_n, \ n > t,$ defines the *smoothing problem*.

Theorem 13.1 (Kalman Prediction) *The predictors $\widehat{\boldsymbol{X}}_t \stackrel{\text{def}}{=} P_{t-1}(\boldsymbol{X}_t)$ and the error covariance matrices*

$$\Omega_t \stackrel{\text{def}}{=} E[(\boldsymbol{X}_t - \widehat{\boldsymbol{X}}_t)(\boldsymbol{X}_t - \widehat{\boldsymbol{X}}_t)']$$

are uniquely determined by the initial conditions

$$\widehat{\boldsymbol{X}}_1 = P(\boldsymbol{X}_1 \mid \boldsymbol{Y}_0), \quad \Omega_1 \stackrel{\text{def}}{=} E[(\boldsymbol{X}_1 - \widehat{\boldsymbol{X}}_1)(\boldsymbol{X}_1 - \widehat{\boldsymbol{X}}_1)']$$

and the recursions, for $t = 1, \ldots,$

$$\widehat{\boldsymbol{X}}_{t+1} = F_t \widehat{\boldsymbol{X}}_t + \Theta_t \Delta_t^{-1} (\boldsymbol{Y}_t - G_t \widehat{\boldsymbol{X}}_t) \tag{16}$$

$$\Omega_{t+1} = F_t \Omega_t F_t' + Q_t - \Theta_t \Delta_t^{-1} \Theta_t', \tag{17}$$

where

$$\Delta_t = G_t \Omega_t G_t' + R_t,$$
$$\Theta_t = F_t \Omega_t G_t'.$$

The matrix $\Theta_t \Delta_t^{-1}$ is called the Kalman gain.

Theorem 13.2 (Kalman Filtering) *The filtered estimates* $\boldsymbol{X}_{t|t} \stackrel{\text{def}}{=} P_t(\boldsymbol{X}_t)$ *and the error covariance matrices*

$$\Omega_{t|t} \stackrel{\text{def}}{=} E[(\boldsymbol{X}_t - \boldsymbol{X}_{t|t})(\boldsymbol{X}_t - \boldsymbol{X}_{t|t})']$$

are determined by the relations

$$X_{t|t} = P_{t-1}(\boldsymbol{X}_t) + \Omega_t G_t' \Delta_t^{-1}(\boldsymbol{Y}_t - G_t \widehat{\boldsymbol{X}}_t)$$

and

$$\Omega_{t|t+1} = \Omega_t - \Omega_t G_t' \Delta_t^{-1} G_t \Omega_t'.$$

Theorem 13.3 (Kalman Fixed Point Smoothing) *The smoothed estimates* $\boldsymbol{X}_{t|n} \stackrel{\text{def}}{=} P_n(\boldsymbol{X}_t)$ *and the error covariance matrices*

$$\Omega_{t|n} \stackrel{\text{def}}{=} E[(\boldsymbol{X}_t - \boldsymbol{X}_{t|n})(\boldsymbol{X}_t - \boldsymbol{X}_{t|n})']$$

are determined for fixed t *by the recursions, which can be solved successively for* $n = t, t+1, \ldots$:

$$\begin{aligned} P_n(\boldsymbol{X}_t) &= P_{n-1}(\boldsymbol{X}_t) + \Omega_{t.n} G_n' \Delta_n^{-1}(\boldsymbol{Y}_n - G_n \widehat{\boldsymbol{X}}_n), \\ \Omega_{t.n+1} &= \Omega_{t.n}[F_n - \Theta_n \Delta_n^{-1} G_n]', \\ \Omega_{t|n} &= \Omega_{t|n-1} - \Omega_{t.n} G_n' \Delta_n^{-1} G_n \Omega_{t.n}', \end{aligned}$$

with initial conditions $P_{t-1}(\boldsymbol{X}_t) = \widehat{\boldsymbol{X}}_t$ *and* $\Omega_{t.t} = \Omega_{t|t-1} = \Omega_t$ *found from Kalman prediction.*

Sakregister