743- Regression and Time Series

Mamikon S. Ginovyan

Estimation in Stationary ARMA(p,q) Models

The Sample Autocorrelation Function

- Although we know how to compute the autocorrelation function for some time series models,
- **in practical problems we do not start with a model**, but with observed data $\{x_1, x_2, \ldots, x_n\}$.

- To assess the degree of dependence in the data set, and to select an appropriate model, one of the important tools we use is **the sample autocorrelation function** (**sample ACF**) of the data.

- If we believe that the data are realized values of a stationary time series $\{X_t, t = 0, \pm 1, \ldots\}$, then the sample ACF will provide an estimate of the ACF of X_t.

The Sample Autocorrelation Function

- This estimate may suggest which of the many possible stationary time series models is a **suitable candidate** for representing the dependence in the data.

- **For example,** a sample ACF that is **close to zero for all nonzero lags** suggests that an appropriate model for the data might be **white noise**.

- The following definitions are natural sample analogues of those for the **ACVF** and **ACF** functions given earlier for stationary time series models.

The Sample Autocorrelation Function

Definition 1.

Let $\{\boldsymbol{x_1}, \boldsymbol{x_2}, \ldots, \boldsymbol{x_n}\}$ be an observation of a time series.

1. The **sample mean** of $\{\boldsymbol{x_1}, \boldsymbol{x_2}, \ldots, \boldsymbol{x_n}\}$ is

$$\bar{x} = \frac{1}{n}\sum_{t=1}^{n} x_t .$$

2. The **sample autocovariance function** is

$$\hat{g}(h) := \frac{1}{n}\sum_{t=1}^{n-|h|}(x_{t+|h|} - \bar{x})(x_t - \bar{x}), \quad -n < h < n.$$

3. The **sample autocorrelation function** is

$$\hat{r}(h) = \frac{\hat{g}(h)}{\hat{g}(0)}, \quad -n < h < n.$$

Properties of Sample Mean, ACVF and ACF

- A second-order stationary process $\{\boldsymbol{X_t}, \boldsymbol{t} = 0, \pm 1, \ldots\}$ is characterized by its **mean** $\mu = \mathrm{E}[\boldsymbol{X_t}]$ and **ACVF**

$$g(h) = Cov(X_h, X_0) = E[(X_h - m)(X_0 - m)].$$

- The estimators of $\boldsymbol{\mu}$, $\gamma(\boldsymbol{h})$ and **ACF** $\rho(\boldsymbol{h}) = \gamma(\boldsymbol{h}) / \gamma(0)$ computed from the sample $\{X_1, X_2, \ldots, X_T\}$ therefore plays a crucial role in problems of inference and construction an appropriate model for the data.

- We examine some of the properties of the estimators

$$\overline{X}_T, \ \hat{g}_T(h) \text{ and } \hat{r}_T(h).$$

Estimation of mean μ = E[X_t]

As an estimator of μ we consider the **sample mean:**

$$\bar{X}_T = \frac{1}{T}\sum_{t=1}^{T} X_t .$$

1. $\bar{X}_T$ is an **unbiased** estimator of $m : E[\bar{X}_T] = m .$

2. The **mean squared error (MSE)** of $\bar{X}_T$ is

$$MSE[\bar{X}_T] = E[\bar{X}_T - m]^2$$

$$= Var(\bar{X}_T) = \frac{1}{T}\sum_{h=-T}^{T}\left(1 - \frac{|h|}{T}\right) g(h). \qquad (1)$$

Estimation of mean $\mu = \mathrm{E}[\, X_t \,]$

Observe that:

1.If $\gamma(h) \to 0$ as $h \to \infty$, then the right hand side of (1) tends to 0, so that we have

$$MSE[\bar{X}_T] \to 0 \quad \text{as} \quad T \to \infty.$$

2.If $\sum_{h=-\infty}^{\infty} |g(h)| < \infty$, then (1) gives

$$\lim_{T \to \infty} T \cdot Var(\bar{X}_T) = \sum_{|h|<\infty} g(h).$$

•Thus, we have the following result.

Estimation of mean μ = E[X_t]

Theorem 1.

If $\{\boldsymbol{X_t}, \boldsymbol{t} = 0, \pm 1, \ldots\}$ is a second-order stationary process with mean $\boldsymbol{\mu}$ and covariance function $\gamma(\boldsymbol{h})$. Then as $\boldsymbol{T} \to \infty$

$$(a)\quad MSE[\overline{X}_T] = Var(\overline{X}_T) \to 0 \quad \text{if} \quad g(T) \to 0;$$

$$(b)\quad T \cdot MSE[\overline{X}_T] = T \cdot Var(\overline{X}_T) \to \sum_{|h|<\infty} g(h) \quad \text{if} \quad \sum_{h=-\infty}^{\infty} |g(h)| < \infty.$$

(c) If $\{\boldsymbol{X_t}, \boldsymbol{t} = 0, \pm 1, \ldots\}$ is Gaussian, then

$$T^{1/2}(\overline{X}_T - m) : AN\left(0, \sum_{h=-T}^{T}\left(1 - \frac{|h|}{T}\right) g(h)\right).$$

Estimation of mean $\mu = E[X_t]$

Remark

- For many time-series, in particular for Linear and ARMA models, for large $\boldsymbol{n}$,

$$\bar{X}_T : AN\left(m, \frac{1}{T} \sum_{|h|<\infty} g(h) \right).$$

- An **approximate 95% CI** for $\boldsymbol{\mu}$ is then

$$\bar{X}_T \pm 1.96\sqrt{v/T}, \text{ where } v = \sum_{|h|<\infty} g(h).$$

- If $\boldsymbol{v}$ is **unknown**, as an estimator of $\boldsymbol{v}$ we consider the statistic:

$$\hat{v} = \sum_{h=-\sqrt{T}}^{\sqrt{T}} \left(1 - |h|/T\right)\hat{g}(h).$$

Estimation of mean μ = E[X_t]

Remark-Example .

Let $\boldsymbol{X_t}$ be an AR(1) process with mean $\boldsymbol{\mu}$, defined by

$$\boldsymbol{X_t} - \boldsymbol{\mu} = \boldsymbol{\Phi}\,(\,\boldsymbol{X_t} - \boldsymbol{\mu}\,) + \boldsymbol{Z_t}\,,$$

where $|\, f \,| < 1$ and $Z_t \sim WN(0, s^{\,2})$.

We have $\quad g\,(h) = \dfrac{f^{\,|h|} s^{\,2}}{1 - f^{\,2}}, \quad$ and hence

$$v = \sum_{|h|<\infty} g\,(h) = \frac{s^{\,2}}{1 - f^{\,2}}\left(1 + 2\sum_{h=1}^{\infty} f^{\,h}\right) = \frac{s^{\,2}}{(1 - f\,)^2}.$$

Hence an approximate 95% CI for $\boldsymbol{\mu}$ is

$$\bar{X}_T \pm 1.96\, s\, T^{-1/2} \,/\, (1 - f\,),$$

provided that **σ** and $\boldsymbol{\phi}$ are known.

Estimation of $\gamma(h)$ and $\rho(h)$

As estimators for unknown $\boldsymbol{\gamma(h)}$ and $\boldsymbol{\rho(h)}$ we consider the

sample autocovariance function:

$$\hat{g}(h) := \frac{1}{n}\sum_{t=1}^{n-|h|}(x_{t+|h|} - \overline{x})(x_t - \overline{x}), \quad -n < h < n.$$

and the **sample autocorrelation function:**

$$\hat{r}(h) = \frac{\hat{g}(h)}{\hat{g}(0)}, \quad -n < h < n.$$

Examples

Example 1.

Suppose that in a sample of size 100 from an AR(1) process with mean μ , $\phi = 0.6$ and $\sigma^2 = 2$ we obtain $\overline{x}_{100} = 0.271$.

a) Construct an approximate 95% confidence interval for μ.

b) Are the data compatible with the hypothesis that $\mu = 0$?

Examples

Solution:

(a) We have $\boldsymbol{X_t}$ is an AR(1) with mean $\boldsymbol{\mu}$, so $\boldsymbol{X_t}$ satisfies

$$X_t - m = f(X_{t-1} - m) + Z_t, \{Z_t : t \in Z\} \sim WN\left(0, s^2\right),$$

with $\phi = 0.6$ and $\sigma^2 = 2$. For AR(1) process we have

$$g_X(h) = \frac{f^{|h|} s^2}{1 - f^2}.$$

We estimate $\boldsymbol{\mu}$ by $\bar{X}_n = \frac{1}{n} \sum_{k=1}^{n} X_k$.

For large sample sizes $\bar{X}_n$ is approximately normally distributed with mean $\boldsymbol{\mu}$ and variance $\frac{1}{n} \sum_{|h|<\infty} g(h)$

(see Section 2.4 in Brockwell and Davis).

Examples

In our case, the variance is

$$\frac{1}{n}\left(1+2\sum_{h=1}^{\infty} f^{h}\right)\frac{s^{2}}{1-f^{2}}=\frac{1}{n}\left(1+2\left(\frac{1}{1-f}-1\right)\right)\frac{s^{2}}{1-f^{2}}$$

$$=\frac{1}{n}\left(\frac{2}{1-f}-1\right)\frac{s^{2}}{1-f^{2}}=\frac{1}{n}\left(\frac{1+f}{1-f}\right)\frac{s^{2}}{1-f^{2}}=\frac{s^{2}}{n(1-f)^{2}}.$$

Hence, $\overline{X}_n$ is approximately $N\left(m,\frac{s^{2}}{n(1-f)^{2}}\right)$.

Examples

Thus, a 95% confidence interval is given by

$$I = \left(\overline{x}_n - l_{0.025} \frac{s}{\sqrt{n}(1-f)}, \overline{x}_n + l_{0.025} \frac{s}{\sqrt{n}(1-f)} \right).$$

Putting in the numeric values gives $\boldsymbol{I} = \mathbf{0.271 \pm 0.69.}$

(b) Since $0 \in I$, so the hypothesis $\boldsymbol{H_0 : \mu = 0}$ **cannot be rejected.**

Estimation of Parameters of ARMA(p,q) Models

- We briefly discuss **estimation of the parameters** in **ARMA(p,q)** models.
- The method of estimation used is the **least square technique**, with some minor modifications.
- **Recall** that Least squares was previously employed in Regression Analysis.
- The **only significant difference** here is that the observations are assumed to be **dependent** as opposed to the Regression Analysis, where they were always assumed to be **independent**.

- We consider some special cases of **ARMA(p,q)** model.

Estimation of AR(1) Model

AR(1) Model:

$$y_t - m - f_1(y_{t-1} - m) = e_t .$$

- The parameters to be estimated are $\boldsymbol{\mu}$ and ϕ_1.
- We minimize the sum of squares of errors denoted by $\boldsymbol{S}$:

$$S = \sum_{t=1}^{T} e_t^2 = \sum_{t=1}^{T} [y_t - m - f_1(y_{t-1} - m)]^2 , \qquad (1)$$

provided $\boldsymbol{y_0}$ is available.

- Obtain $\partial \boldsymbol{S} / \partial \boldsymbol{\mu}$ and $\partial \boldsymbol{S} / \partial \phi_1$, and write the normal equations by setting them equal to 0.

Estimation of AR(1) Model

Remark.

Note that, in the sum $\boldsymbol{S}$, we assume that $\boldsymbol{y_0}$ is available.
If we assume that only the time series data $\boldsymbol{y_1 , y_2 ,\ldots, y_T}$ are available, then we should minimize the sum $\boldsymbol{S_1}$:

$$S_1 = \sum_{t=2}^{T} e_t^2 = \sum_{t=2}^{T} [y_y - m - f_1(y_{t-1} - m)]^2, \; e_1 = 0. \qquad (2)$$

This is referred to as :
condition $\boldsymbol{\varepsilon_1 = 0}$. **Conditional Least Squares Estimation**

- In this case we have the following **normal equations** to estimate $\boldsymbol{\mu}$ and ϕ_1 :

Estimation of AR(1) Model

$$\frac{\partial S_1}{\partial m} = 2\sum_{t=2}^{T}[(y_t - m - f_1(y_{t-1} - m)] = 0,$$

$$\frac{\partial S_1}{\partial f_1} = 2\sum_{t=2}^{T}[(y_t - m - f_1(y_{t-1} - m)][-(y_{t-1} - m)] = 0.$$

- The solution of these equations gives the **least squares estimates** of $\boldsymbol{\mu}$ and ϕ_1 as

$$\hat{m} = \frac{\sum_{t=2}^{T}(y_t - \hat{f}_1 y_{t-1})}{(T-1)(1-\hat{f}_1)}, \quad \hat{f}_1 = \frac{\sum_{t=2}^{T}(y_t - \hat{m})(y_{t-1} - \hat{m})}{\sum_{t=2}^{T}(y_{t-1} - \hat{m})^2}.$$

Estimation of AR(1) Model

Remark. When the sample size, $\boldsymbol{T}$ is **sufficiently large** we can use the following approximations:

$$\hat{m} = \overline{y} = \frac{1}{T}\sum_{t=1}^{T} y_t \ , \qquad (3)$$

$$\hat{f}_1 = \frac{\sum_{t=2}^{T}(y_t - \overline{y})(y_{t-1} - \overline{y})}{\sum_{t=2}^{T}(y_{t-1} - \overline{y})^2} \approx \hat{r}_1, \qquad (4)$$

where $\hat{r}_1$ is the sample autocorrelation coefficient at lag 1.

- Using these estimates for $\boldsymbol{\mu}$ and ϕ_1, we obtain the **fitted values** $\hat{y}_t$ for $t = 2, 3, \ldots, T$:

$$\hat{y}_t = \hat{m} + \hat{f}_1(y_{t-1} - \hat{m}).$$

Estimation of AR(1) Model

Consequently, the **residuals** are

$$\hat{e}_t = y_t - \hat{y}_t \quad \text{for} \quad t = 2, 3, \ldots, T,$$

and the **Residual Sum of Squares** is

$$\hat{S}_1 = \sum_{t=2}^{T} \hat{e}_t^2 = \sum_{t=2}^{T} [y_t - \hat{m} - \hat{f}_1(y_{t-1} - \hat{m})]^2.$$

We can use $\hat{S}_1$ to obtain an **unbiased** estimator of σ_ε^2.

Thus, in the AR(1) model an unbiased estimator of σ_ε^2 is

$$\hat{s}_e^2 = \hat{S}_1 / (T - 3).$$

Estimation of AR(1) Model

Example 1. Use the model

$$y_t - m = f_1(y_{t-1} - m) + e_t$$

to represent a time series, for which the first 10 values are

t	1	2	3	4	5	6	7	8	9	10
y_t	8	10	7	6	9	8	6	5	7	4

Conditional on the initial value $\boldsymbol{y_1 = 8}$ and assuming $\hat{m} = \overline{y}$, calculate the least squares estimate of ϕ_1.

a) .29 **b) .19** **c) .17** **d) .05** **e) .03**

Estimation of AR(1) Model. Example 1

Solution.

Observe that $T = 10$, and we are asked to use

$$\hat{m} = \overline{y} = \frac{1}{10}\sum_{t=1}^{10} y_t = \frac{70}{10} = 7.0.$$

Hence the least squares estimate of ϕ_1 is given by $(t\text{-}1 \to t)$

$$\hat{f}_1 = \frac{\sum_{t=2}^{10}(y_t - \overline{y})(y_{t-1} - \overline{y})}{\sum_{t=2}^{10}(y_{t-1} - \overline{y})^2} = \frac{\sum_{t=1}^{9}(y_{t+1} - \overline{y})(y_t - \overline{y})}{\sum_{t=1}^{9}(y_t - \overline{y})^2} = \frac{4}{21} \approx .19,$$

where the numerator and the denominator are calculated from the following **table**:

Estimation of AR(1) Model. Example 1

t	1	2	3	4	5	6	7	8	9	10
y_t	8	10	7	6	9	8	6	5	7	4
$y_t - \bar{y}$	1	3	0	-1	2	1	-1	-2	0	-3
$y_{t+1} - \bar{y}$	3	0	-1	2	1	-1	-2	0	-3	

We have:

the numerator = 3 – 2 + 2 – 1 + 2 = 4,

the denominator = 1+9+1+ 4 +1+1 + 4 = 21.

- **Thus the correct choice is b).**

Estimation of AR(1) Model. Example 2

Example 2.

Given the ten observation as in **Example 1** and the model assumed to be

$$Y_t = f_1 Y_{t-1} + e_t \ ,$$

obtain the least squares estimate of $\boldsymbol{\phi_1}$.

a) .90 b) .91 c) .92 d) .93 e) .94.

Solution.

Observe that minimizing the sum of the squares of the errors in this model leads to the following estimate of $\boldsymbol{\phi_1}$.

Estimation of AR(1) Model. Example 2

$$\hat{f}_1 = \frac{\sum_{t=2}^{10} y_t y_{t-1}}{\sum_{t=2}^{10} y^2{}_{t-1}} = \frac{459}{504} \approx .91,$$

where

$$\sum_{t=2}^{10} y_t y_{t-1} = 8(10) + 10(7) + 7(6) + \ldots + 7(4) = 459,$$

$$\sum_{t=2}^{10} y_{t-1}^2 = 8^2 + 10^2 + \ldots + 7^2 = 504.$$

- **Thus the correct choice is b).**

Estimation of MA(1) Model

For the **MA(1)** model: $y_t - m = e_t + q_1 e_{t-1}$,

the sum of squares of the error terms can be written as:

$$S = \sum_{t=1}^{T} e_t^2 = \sum_{t=1}^{T} [y_t - m - q_1 e_{t-1}]^2,$$

where at time $\boldsymbol{t} = 1$, the value of $\boldsymbol{\varepsilon_0}$ is involved in $\boldsymbol{S}$.

- If $\boldsymbol{\varepsilon_0}$ is **known**, then by taking some initial values of $\boldsymbol{\mu}$ and $\boldsymbol{\theta_1}$ we can calculate:

$$e_1 = y_1 - m - q_1 e_0$$

$$e_2 = y_2 - m - q_1 e_1$$

$$\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots$$

$$e_T = y_T - m - q_1 e_{T-1}.$$

Estimation of MA(1) Model

- Squaring and adding the resulting expressions we obtain $\boldsymbol{S}$.
- If the value of $\boldsymbol{\varepsilon_0}$ is **not known** then it might be reasonable to assume that $\boldsymbol{\varepsilon_0} = \mathbf{0}$.
- An initial value of $\boldsymbol{\mu}$ is usually taken as $\overline{y}$, whereas
- the initial value of $\boldsymbol{\theta_1}$ based on sample ACF.

Remark.

It is not straightforward to minimize $\boldsymbol{S}$ by setting $\partial \boldsymbol{S} / \partial \boldsymbol{\mu}$ and $\partial \boldsymbol{S} / \partial \boldsymbol{\theta_1}$ equal to $\mathbf{0}$ because the **resulting normal equations** are **nonlinear** in the parameters $\boldsymbol{\mu}$ and $\boldsymbol{\theta_1}$.

Diagnostic Checking of the Model

Description of the Procedure.

Having specified a **tentative model**, we consider procedure for **checking if the model indeed fits given time series data.**

- This procedure is based on the **residuals** obtained by fitting the tentative model to the given data.

Remark.

For simplicity we discuss the case where the tentative model selected is **AR(1)**. The described procedure applies equally well to **any ARMA(p,q)** model with appropriate modifications.

Diagnostic Checking of the Model

So our model is **AR(1)**:

$$(Y_t - m) - f_1(Y_{t-1} - m) = e_t .$$

Having estimated the parameters $\boldsymbol{\mu}$ and $\boldsymbol{\phi_1}$, we obtain the fitted values $\hat{y}_t$ as

$$\hat{y}_t = \hat{m} + \hat{f}_1(y_{t-1} - \hat{m}) \quad \text{for} \quad t = 2, 3, 4, \ldots, T, \qquad (1)$$

and residuals $\hat{e}_t$ defined as

$$\hat{e}_t = y_t - \hat{y}_t \quad \text{for} \quad t = 2, 3, 4, \ldots, T .$$

Note that there are only $\boldsymbol{T} - 1$ residuals available.

Diagnostic Checking of the Model

An Approach.

Since $\boldsymbol{\varepsilon_t}$'s are assumed to be $\boldsymbol{WN}(0\,,\sigma_\varepsilon^2)$ = uncorrelated RV's with mean 0 and a constant variance σ_ε^2,

if the assumed AR(1) model is indeed the underlying model for the given time series, then it is reasonable to expect that

i) $$\bar{e} = \frac{1}{T-1}\sum_{t=2}^{T} \hat{e}_t \approx 0.$$

ii) The sample ACF $\hat{r}_k$, at various lags, calculated using $\hat{e}_t$'s, should also be **close to 0**.

- In fact it can be shown that if the underlying model is **AR(1),**

  then $E[\hat{r}_k] \approx 0$ for all $\boldsymbol{k} > 0$.

Diagnostic Checking of the Model

A test of the hypothesis $\boldsymbol{H_0 : \rho_k = 0}$ for a given value of $\boldsymbol{k}$ against the alternative $\boldsymbol{H_1 : \rho_k \neq 0}$ is suggested as follows:

STEP-1: Calculate $$\hat{r}_k = \frac{\sum_{t=2}^{T-k} \hat{e}_t \hat{e}_{t+k}}{\sum_{t=2}^{T} \hat{e}_t^2}, \quad k = 1, 2, \ldots, K.$$

(We are assuming that only $\boldsymbol{T}-1$ residuals are available.)

STEP-2 (Decision Rule):

Do not reject $\boldsymbol{H_0}$ if $|\hat{r}_k| \leq 2\sqrt{\hat{V}[\hat{r}_k]}$, otherwise reject $\boldsymbol{H_0}$.

Diagnostic Checking of the Model

Another Test of Adequacy

- If T is large and K is small compared to T, then it can be shown that the statistic

$$Q^* = T \cdot \sum_{k=1}^{K} \hat{r}_k^2 \sim \chi^2_{K-m}$$

is distributed **approximately as a chi-square** RV with $df = K - m$, where m is the total number of parameters in a model to be estimated by the least squares technique.

- **For example**, in **AR(1)** model $m = 2$, since μ and ϕ_1 are the parameters to be estimated.
  Similarly, for **AR(2)** model, $m = 3$.

Diagnostic Checking of the Model

- A **large observed value** of Q^* should be viewed as **evidence that the tentative model selected is inadequate** for describing the time series data at hand.

- For small values of T, Ljung & Box suggest using the statistic Q given by:

$$Q = T(T+2) \cdot \sum_{k=1}^{K} \frac{\hat{r}_k^{\,2}}{T-k},$$

because it provides a **better approximation to the chi-square distribution.**

Forecasting Stationary Time Series

The MSE – Best Linear Prediction.

Let $\{\boldsymbol{X_t}, \boldsymbol{t} = 0, \pm 1, \ldots\}$ be a stationary time series with

- mean $\boldsymbol{\mu}$,
- autocovariance function (ACVF) $\boldsymbol{\gamma_X}(h)$,
- variance $\sigma^2 = \boldsymbol{\gamma_X}(0)$,
- autocorrelation (ACF) function $r_X(h) = \dfrac{g_X(h)}{g_X(0)}$.
- The ACVF and ACF provide a useful measure of the degree of dependence among the values of the time series $\boldsymbol{X_t}$, and play an important role in **prediction** of the **future values of the series in terms of the past and present values.**

The MSE – Best Linear Prediction

- The **role of the ACVF in prediction** is illustrated by the following **observations**.

- Essentially, now I am going to show that for **MSE – Best Linear Prediction** of X_{n+h} by means of

$$X_n, X_{n-1}, X_{n-2}, \ldots,$$

we need to know **only the mean and ACVF** $\gamma_X(h)$ of the process X_t, but not the complete specification of X_t.

The MSE – Best Linear Prediction

Observation 1.

Suppose that $\{X_t, t = 0, \pm 1, \ldots\}$ is a Gaussian stationary time series and that we have observed the value X_n .

We would like to find the function of X_n that gives

- the **best MSE – predictor** of the value X_{n+h} for some $h > 0$, and
- the **MSE – prediction error**.

We have solved this problem, and the solution is:

- Since the RV's $Y = X_{n+h}$ and $X = X_n$ have **joint bivariate normal distribution:**

$$N(m_1 = m_2 = m,\ s_1^{\ 2} = s_2^{\ 2} = s^{\ 2},\ r = r(h)),$$

The MSE – Best Linear Prediction

the **conditional distribution** of $\boldsymbol{Y = X_{n+h}}$ given that $\boldsymbol{X = X_n = x_n}$ is:

$$N(m + r(h)(x_n - m),\ \ s^2(1 - r(h))^2).$$

Hence

a) The **best MSE – predictor** $\hat{X}_{n+h}$ of the value $\boldsymbol{X_{n+h}}$ in terms of $\boldsymbol{X_n}$ is

$$\hat{X}_{n+h} = g(X_n) = E[X_{n+h} \mid X_n] = m + r(h)(X_n - m) \qquad (1)$$

b) The **MSE – prediction error** is

$$E(X_{n+h} - \hat{X}_{n+h})^2 = s^2(1 - r(h))^2.$$

The MSE – Best Linear Prediction

Observation 2. Observation 1 shows that

- at least for Gaussian process, the best MSE – prediction of $\boldsymbol{X_{n+h}}$ in terms of $\boldsymbol{X_n}$ is **more accurate** as $|\rho(h)|$ becomes closer to 1, and
- in the limit as $\rho(h) \to \pm 1$, the best MSE – predictor $\hat{X}_{n+h}$ approaches to $\mu \pm (X_n - \mu)$, (for $\mu = 0$ we have $\boldsymbol{X_n}$)
- the MSE – prediction error $\sigma^2(1 - \rho(h))^2$ approaches to 0.

Note that in Observation 1 the assumption that the RV's $\boldsymbol{Y} = \boldsymbol{X_{n+h}}$ and $\boldsymbol{X} = \boldsymbol{X_n}$ have **joint normal distribution** played a crucial role.

The MSE – Best Linear Prediction

Observation 3.

- For time series with **non-normal joint distributions** the corresponding calculations are in general much more complicated.
- However, if instead of looking for the best MSE – predictor for $\boldsymbol{X_{n+h}}$, we look for the best **linear** MSE – predictor, that is, the best MSE – predictor of the form

$$L(\boldsymbol{X_n}) = a\boldsymbol{X_n} + b ,$$

then our problem becomes of finding constants $\boldsymbol{a}$ and $\boldsymbol{b}$ to minimize the MSE –error

$$E(X_{n+h} - aX_n - b)^2 .$$

The MSE – Best Linear Prediction

- **Below in Observation 4 we will show that the best Linear MSE – predictor** for $\boldsymbol{X_{n+h}}$ is given by

$$\hat{X}_{n+h,L} = L(X_n) = \mu + \rho(h)(X_n - \mu) = \rho(h)X_n + \mu(1 - \rho(h)), \qquad (2)$$

that is, $a = \rho(h)\,\boldsymbol{X_n}$ and $b = \mu\,(1 - \rho(h))$.

- The **MSE – prediction error** is given by

$$E(X_{n+h} - L(X_n))^2 = \sigma^2(1 - \rho(h))^2.$$

The MSE – Best Linear Prediction

- **Comparing (1) and (2) we find that** for **Gaussian processes** $g(\boldsymbol{X}_n)$ and $L(\boldsymbol{X}_n)$ are the same.

- **In general,** of course, $g(\boldsymbol{X}_n)$ will give smaller MSE-error than $L(\boldsymbol{X}_n)$, since it is the best of the larger class of predictors.

- **However,** the fact that the **best Linear MSE-predictor depends only on the mean and ACF** of the process X_t means that it can be calculated without more detailed knowledge of the joint distributions, that is, we **need not to have complete specification of the underlying process** X_t .

The MSE – Best Linear Prediction

- This fact is extremely important in practice because of the difficulty of complete specification of the underlying process X_t and because of the difficulty of computing the required conditional expectations even if the distributions are known.

- As we will see later, similar conclusions apply when we consider the more general problem of predicting X_{n+h} by means of $X_n, X_{n-1}, X_{n-2}, \ldots,$.

Conclusion.

- **For best Linear** MSE – prediction of X_{n+h} by means of X_n, $X_{n-1}, X_{n-2}, \ldots,$ .we need to know **only the mean and ACF** of the underlying process X_t .

The MSE – Best Linear Prediction

Observation 4.

Problem 2.1: Let $\{\boldsymbol{X_t}, \boldsymbol{t} = 0, \pm 1, \ldots\}$ be stationary time-series with mean $\boldsymbol{\mu}$ and ACF $\boldsymbol{\rho}(h)$.

Show that the best predictor $\hat{X}_{n+h}$ of $\boldsymbol{X_{n+h}}$ of the form $\hat{X}_{n+h} = aX_n + b$ is obtained by choosing $a = \rho(h)$ and $b = \mu\,(1 - \rho(h))$, that is,

$$\hat{X}_{n+h} = r\,(h)X_n + m[1 - r\,(h)].$$

Solution.

We find the best linear predictor $\hat{X}_{n+h} = aX_n + b$ of $\boldsymbol{X_{n+h}}$ by finding constants $\boldsymbol{a}$ and $\boldsymbol{b}$ such that

$$E\left[X_{n+h} - \hat{X}_{n+h}\right] = 0 \quad \text{and} \quad E\left[\left(X_{n+h} - \hat{X}_{n+h}\right)X_n\right] = 0.$$

The MSE – Best Linear Prediction

We have $E\left[X_{n+h}-\hat{X}_{n+h}\right]=E\left[X_{n+h}-aX_n-b\right]$

$$=E\left[X_{n+h}\right]-aE\left[X_n\right]-b=m(1-a)-b,$$

and $E\left[\left(X_{n+h}-\hat{X}_{n+h}\right)X_n\right]=E\left[\left(X_{n+h}-aX_n-b\right)X_n\right]$

$$=E\left[X_{n+h}X_n\right]-aE\left[X_n^2\right]-bE\left[X_n\right]$$

$$=E\left[X_{n+h}X_n\right]-E\left[X_{n+h}\right]E\left[X_n\right]+E\left[X_{n+h}\right]E\left[X_n\right]$$

$$-a\left(E\left[X_n^2\right]-E\left[X_n\right]^2+E\left[X_n\right]^2\right)-bE\left[X_n\right]$$

$$=Cov\left(X_{n+h},X_n\right)+m^2-a\left(Cov\left(X_n,X_n\right)+m^2\right)-bm$$

$$=g(h)+m^2-a\left(g(0)+m^2\right)-bm.$$

The MSE – Best Linear Prediction

Thus, we have

$$b = m(1-a),\ a = \frac{g(h)+m^2-bm}{g(0)+m^2}.$$

Solving the obtained system of equations for $\boldsymbol{a}$ and $\boldsymbol{b}$, we get

$$a = \gamma(h)/\gamma(0) = \rho(h) \quad \text{and} \quad b = \mu\,(1-\rho(h)),$$

$$\hat{X}_{n+h} = r(h)X_n + m\big(1-r(h)\big).$$

Thus, we have shown that:

For best Linear MSE – prediction of $\boldsymbol{X_{n+h}}$ by means of RV's $\boldsymbol{X_n}$, $\boldsymbol{X_{n-1}}$, $\boldsymbol{X_{n-2}}$, …, we need to know **only the mean and ACF** of the underlying process $\boldsymbol{X_t}$.

Forecasting Stationary Time Series

- Let $\{\boldsymbol{X_t}, \boldsymbol{t} = 0, \pm 1, \dots\}$ be a second-order stationary process with mean $\boldsymbol{\mu}$ and ACVF $\gamma(h)$.
- We now consider the problem of **linear MSE-predicting** the values $\boldsymbol{X_{n+h}}$, $\boldsymbol{h} > 0$, of the process $\boldsymbol{X_t}$ in terms of the values $\{\boldsymbol{X_1}, \dots, \boldsymbol{X_n}\}$.
- That is, our goal is to find the **linear combination** of $1, \boldsymbol{X_n}, \boldsymbol{X_{n-1}}, \boldsymbol{X_{n-2}}, \dots, \boldsymbol{X_1}$, that forecasts $\boldsymbol{X_{n+h}}$ with **minimum mean squared error**.
- The best linear predictor in terms of $1, \boldsymbol{X_n}, \boldsymbol{X_{n-2}}, \dots, \boldsymbol{X_1}$, denoted by $\hat{X}_{n+h} = P_n X_{n+h}$, has the form

$$\hat{X}_{n+h} = P_n X_{n+h} = a_0 + a_1 X_n + \dots + a_n X_1. \qquad (1)$$

Forecasting Stationary Time Series

- It remains only to determine the coefficients $\boldsymbol{a_0}$, $\boldsymbol{a_1}$, … , $\boldsymbol{a_n}$, by finding the values that **minimize the MSE**

$$S(a_0,\dots,a_n) = E(X_{n+h} - P_n X_{n+h})^2$$
$$= E(X_{n+h} - a_0 - a_1 X_n - \cdots - a_n X_1)^2. \qquad (2)$$

- Since $\boldsymbol{S}$ is a quadratic function of $\boldsymbol{a_0}$, $\boldsymbol{a_1}$, … , $\boldsymbol{a_n}$, there is at least one value of ($\boldsymbol{a_0}$,…,$\boldsymbol{a_n}$) that minimizes $\boldsymbol{S}$ and that the minimum satisfies the equations

$$\frac{\partial S(a_0,\dots,a_n)}{\partial a_j} = 0, \quad j = 0,\dots,n. \qquad (3)$$

- Evaluation of the derivatives in equations (3) gives the equivalent equations:

Forecasting Stationary Time Series

$$E\left[X_{n+h} - a_0 - \sum_{i=1}^{n} a_i X_{n+1-i}\right] = 0 \qquad (4)$$

$$E\left[\left(X_{n+h} - a_0 - \sum_{i=1}^{n} a_i X_{n+1-i}\right) X_{n+1-j}\right] = 0,\ j = 1,\dots,n. \qquad (5)$$

These equations can be written more neatly in vector notation as

$$a_0 = m\left(1 - \sum_{i=1}^{n} a_i\right) \qquad (6)$$

and $\quad \Gamma_n \vec{a}_n = \vec{g}_n(h) \qquad (7)$

where $\vec{a}_n = (a_0,\dots,a_n)', \quad \Gamma_n = [g(i-j)]_{i,j=1}^{n},$

and $\vec{g}_n(h) = \left(g(h), g(h+1),\dots, g(h+n-1)\right)'.$

Forecasting Stationary Time Series

Hence

- the best linear predictor is given by

$$\hat{X}_{n+h} = P_n X_{n+h} = \mu + \sum_{i=1}^{n} a_i \left(X_{n+1-i} - \mu \right), \qquad (8)$$

- the mean square prediction error given by

$$\begin{aligned} E(X_{n+h} - P_n X_{n+h})^2 &= \gamma(0) - 2\sum_{i=1}^{n} a_i \gamma\left(h+i-1\right) + \sum_{i=1}^{n}\sum_{j=1}^{n} a_i \gamma\left(i-j\right) a_j \\ &= \gamma(0) - \vec{a}_n' \gamma_n(h), \qquad (9) \end{aligned}$$

Forecasting Stationary Time Series

Summary: Properties of $\hat{X}_{n+h} = P_n X_{n+h}$:

1. $P_n X_{n+h} = \mu + \sum_{i=1}^{n} a_i (X_{n+1-i} - \mu),$

   where $\vec{a}_n = (a_0, \ldots, a_n)'$ satisfies (7)

2. $E(X_{n+h} - P_n X_{n+h})^2 = \gamma(0) - \vec{a}_n' \vec{\gamma}_n(h),$

   where $\vec{\gamma}_n(h) = \left(\gamma(h), \ldots, \gamma(h+n-1)\right)'$.

3. $E(X_{n+h} - P_n X_{n+h}) = 0.$

4. $E\left[(X_{n+h} - P_n X_{n+h}) X_j\right] = 0$ for all $j = 1, \ldots, n.$

Forecasting Stationary Time Series

Example 1 (One – step prediction of an AR(1) series).
Consider the **AR(1)** stationary time series:

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

From (7) and (8), the best linear predictor of $\boldsymbol{X_{n+1}}$ in terms of $\{1, \boldsymbol{X_n}, \ldots, \boldsymbol{X_1}\}$ is $P_n X_{n+1} = \vec{a}_n' \vec{X}_n$, where $\vec{X}_n = (X_n, \ldots, X_1)'$, and

$$\begin{bmatrix} 1 & \phi & \phi^2 & \cdots & \phi^{n-1} \\ \phi & 1 & \phi & \cdots & \phi^{n-2} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \phi^{n-1} & \phi^{n-2} & \phi^{n-3} & \cdots & 1 \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix} = \begin{bmatrix} \phi \\ \phi^2 \\ \vdots \\ \phi^n \end{bmatrix}$$

from which we find $\vec{a}_n = (\phi, 0, \ldots, 0)'$.

Forecasting Stationary Time Series

Thus, the best linear predictor of $\boldsymbol{X_{n+1}}$ in terms of $\{\boldsymbol{X_1}, \ldots, \boldsymbol{X_n}\}$ is

$$P_n X_{n+1} = a_n' \vec{X}_n = f X_n,$$

with mean squared error

$$E\left(X_{n+1} - P_n X_{n+1}\right)^2 = g(0) - a_n' g_n(1) = \frac{s^2}{1-f^2} - fg(1) = s^2.$$

Remark.

Thus, the best linear predictor of $\boldsymbol{X_{n+1}}$ in terms of $\{\boldsymbol{X_1}, \ldots, \boldsymbol{X_n}\}$ **depends only on last observation $\boldsymbol{X_n}$**. This is not surprising because **AR(1)** process possess **Markovian property**:

$$P\left\{X_{n+1} \le x_{n+1} \mid X_n = x_n, \ldots, X_1 = x_1\right\} = P\left\{X_{n+1} \le x_{n+1} \mid X_n = x_n\right\}.$$

Forecasting Stationary Time Series

Example 2 (AR(1) model with nonzero mean).

The time series $\{Y_t\}$ is said to be an **AR(1)** process with mean μ if $X_t = Y_t - \mu$ is a zero-mean **AR(1)** process.
Defining X_t as in Example 1: $X_t = \phi X_{t-1} + Z_t$, and letting $Y_t = X_t + \mu$, we see that Y_t satisfies the equation

$$Y_t - \mu = \Phi(Y_t - \mu) + Z_t \qquad (1)$$

If $P_n Y_{n+h}$ is the best linear predictor of Y_{n+h} in terms of $\{1, Y_n, \ldots, Y_1\}$, then application of P_n to (1) with $t = n+1, n+2, \ldots$ gives the recursions

$$P_n Y_{n+h} - \mu = \phi (P_n Y_{n+h-1} - \mu), \quad h = 1, 2, \ldots$$

Forecasting Stationary Time Series

Noting that $P_n Y_n = Y_n$, we can solve these equations recursively for $P_n Y_{n+h}$, h = 1, 2, … , to obtain

$$P_n Y_{n+h} = \mu + \phi^h (Y_n - \mu).$$

For the corresponding mean squared error we have

$$E(Y_{n+h} - P_n Y_{n+h})^2 = \sigma^2 (1 - \phi^{2h})/(1 - \phi^2) .$$

The Durbin-Levinson Algorithm

- Let $\{X_t\}$ be a zero-mean stationary series with ACVF $\gamma(\cdot)$, then we have obtained formulas for determining the best linear predictor $P_n X_{n+h}$ of X_{n+h} in terms of $\{X_n, \ldots, X_1\}$.
- However, the direct approach requires the determination of solution of a system of n equations, which for large n may be difficult and time-consuming.
- **An Idea:** It would be helpful to use the one-step predictor $P_n X_{n+1}$ to simplify the calculation of $P_n X_{n+2}$.
- Prediction algorithms that utilize this idea are said to be **recursive**.
- We discuss here the **Durbin-Levinson algorithm.**
- It is of interest also the **innovations algorithm**.

The Durbin-Levinson Algorithm

Assume that the covariance matrix $\Gamma_n = [g(i-j)]_{i,j=1}^{n}$ is nonsingular, then the best linear predictor is given by

$$\hat{X}_{n+1} = P_n X_{n+1} = f'_n X_n = f_{n1} X_n + \ldots + f_{nn} X_1,$$

where $f_n = \Gamma_n^{-1} \Upsilon_n, \quad \Upsilon_n = \left(g(1), \ldots, g(n)\right)',$

and the corresponding mean squared error is

$$s_n^2 := E(X_{n+1} - P_n X_{n+1})^2 = g(0) - f'_n \Upsilon_n.$$

The Durbin-Levinson Algorithm

The Durbin-Levinson Algorithm: The coefficients $\phi_{n1}, \ldots, \phi_{nn}$ can be computed recursively from the equations

$$\phi_{nn} = \left[\gamma(n) - \sum_{j=1}^{n-1} \phi_{n-1,j}\gamma(n-j)\right]\sigma_{n-1}^{-2},$$

$$\begin{bmatrix} \phi_{n1} \\ \vdots \\ \phi_{n,n-1} \end{bmatrix} = \begin{bmatrix} \phi_{n-1,1} \\ \vdots \\ \phi_{n-1,n-1} \end{bmatrix} - \phi_{nn} \begin{bmatrix} \phi_{n-1,n-1} \\ \vdots \\ \phi_{n-1,1} \end{bmatrix}$$

$$\text{and} \quad \sigma_n^2 = \sigma_{n-1}^2\left[1 - \phi_{nn}^2\right],$$

where $\phi_{11} = \gamma(1)/\gamma(0)$ and $\sigma_0^2 = \gamma(0)$.

Remark.

The coefficients $\phi_{n1}, \ldots, \phi_{nn}$ determine the **partial ACF** of $\{X_t\}$.