4. Estimation of the Expectation and the Autocovariance function

Up to now:

- Theoretical moments of the stationary process \( \{X_t\} \)
  - expectation \( \mathbb{E}(X_t) = \mu \) (1st moment)
  - autocovariance function \( \text{Cov}(X_t, X_{t+h}) = \gamma_X(h) \) (2nd moments)

- Link between expectation/autocovariance function and process parameters
Remarks: (I)

- For the expectation of the stationary ARMA\((p, q)\) process we have

\[
E(X_t) = E\left[c + \phi_1 X_{t-1} + \ldots + \phi_p X_{t-p} + \epsilon_t + \theta_1 \epsilon_{t-1} + \ldots + \theta_q \epsilon_{t-q}\right]
\]

\[
= c + \phi_1 E(X_{t-1}) + \ldots + \phi_p E(X_{t-p}) + 0 + 0 + \ldots + 0
\]

\[
= c + \phi_1 E(X_t) + \ldots + \phi_p E(X_t)
\]

\[
= c + E(X_t) \cdot (\phi_1 + \phi_2 + \ldots + \phi_p),
\]

and thus

\[
E(X_t) = \frac{c}{1 - \phi_1 - \phi_2 - \ldots - \phi_p} \equiv \mu
\]
Remarks: (II)

- If the stationary ARMA\((p, q)\) process \(\{X_t\}\) has a non-zero expectation (mean) \(\mu \neq 0\), we often consider the process \(\{X_t - \mu\}\) instead of \(\{X_t\}\)

- The process \(\{X_t - \mu\}\) characterizes the process deviations from its mean and has the form

\[
X_t - \mu = c - \mu + \phi_1 X_{t-1} + \ldots + \phi_p X_{t-p} \\
\quad + \epsilon_t + \theta_1 \epsilon_{t-1} + \ldots + \theta_q \epsilon_{t-q} \\
= \mu(1 - \phi_1 - \ldots - \phi_p) - \mu + \phi_1 X_{t-1} + \ldots + \phi_p X_{t-p} \\
\quad + \epsilon_t + \theta_1 \epsilon_{t-1} + \ldots + \theta_q \epsilon_{t-q} \\
= \phi_1(X_{t-1} - \mu) + \ldots + \phi_p(X_{t-p} - \mu) \\
\quad + \epsilon_t + \theta_1 \epsilon_{t-1} + \ldots + \theta_q \epsilon_{t-q}
\]
Remarks: (III)

• The process $\{Y_t\} = \{X_t - \mu\}$ is an ARMA($p, q$) process with mean 0.

• The processes $\{X_t\}$ and $\{Y_t\}$ have the same autocovariance function:

$$\gamma_X(h) = \gamma_Y(h) \quad \text{for all } h \in \mathbb{Z}$$
Now:

• Estimation of the mean and the autocovariance function using a sample of $T$ observations (process realizations)

• We denote the observations (trajectory) by $x_1, \ldots, x_T$ and the random sample by $X_1, \ldots, X_T$ (process variables)

Difference to "Classical Estimation Theory":

• In time series analysis, the sample $X_1, \ldots, X_T$ in general constitutes a sequence of dependent random variables that are not identically distributed (violation of the i.i.d. assumption)
Question:

- Under what conditions can we obtain reliable estimators of the parameters of the random variable $X_t$ and the vector $(X_t, X_{t+h})'$ for fixed values of $t$ and $h$ on the basis of a single trajectory?

$\rightarrow$ Ergodicity
4.1 Ergodicity

Definition 4.1: (Ergodicity)

Let \( \{X_t\} \) be a stationary process with expectation \( \mu \) and auto-
covariance function \( \gamma_X(h) \). We call \( \{X_t\} \) mean-ergodic, if

\[
\lim_{T \to \infty} E \left[ \left( \frac{1}{T} \sum_{t=1}^{T} X_t - \mu \right)^2 \right] = 0.
\]

We call \( \{X_t\} \) covariance-ergodic, if for all \( h \in \mathbb{Z} \) we have

\[
\lim_{T \to \infty} E \left\{ \left[ \frac{1}{T} \sum_{t=1}^{T} (X_t - \mu)(X_{t+h} - \mu) - \gamma_X(h) \right]^2 \right\} = 0. \]
Remark:

- The concept of "Ergodicity" may be interpreted as quadratic-mean consistency of the estimators

\[
\bar{X}_T \equiv \frac{1}{T} \sum_{t=1}^{T} X_t
\]

and

\[
\tilde{\gamma}_X(h) = \frac{1}{T} \sum_{t=1}^{T} (X_t - \mu)(X_{t+h} - \mu)
\]

of the unknown parameters \(\mu\) and \(\gamma_X(h)\) in the case of dependent random variables
Recall:

- Consistency of an estimator means that "the estimator become more and more accurate" when the sampling size increases
  (distinct definitions of consistency)

Now:

- Sufficient ergodicity-conditions for a stochastic process
**Theorem 4.2:** (Condition for mean-ergodicity)

Let \( \{X_t\}_{t \in \mathbb{Z}} \) be a stationary process with mean \( \mu \) and autocovariance function \( \gamma_X(h) \). If \( \{\gamma_X(h)\}_{h \in \mathbb{Z}} \) is absolutely summable, i.e. if

\[
\sum_{h = -\infty}^{\infty} |\gamma_X(h)| < \infty,
\]

then \( \{X_t\} \) is mean-ergodic. \( \blacksquare \)

(Proof: Fuller, 1976)
Remarks:

- Mean-ergodicity requires that the autocovariance function $\gamma_X(h)$ converges to zero sufficiently fast for $h \to \infty$

- For practical purposes Theorem 4.2 implies that we can estimate the process mean $\mu$ consistently by $\overline{X} = \frac{1}{T} \sum_{t=1}^{T} X_t$ when the dependency of two process variables lying far away from each other becomes sufficiently small (in practice, this often is the case)

Conditions for covariance ergodicity:

- Restriction to normal processes
Definition 4.3: (Normal process)

We call a stationary process \( \{X_t\}_{t \in \mathbb{Z}} \) a normal process if for every selection of \( n \) points in time \( t_1, t_2, \ldots, t_n \) the corresponding random variables \( X_{t_1}, X_{t_2}, \ldots, X_{t_n} \) follow a multivariate normal distribution.

Theorem 4.4: (Condition for covariance-ergodicity)

A stationary normal process \( \{X_t\}_{t \in \mathbb{Z}} \) with absolutely summable autocovariance function is covariance-ergodic.
4.2 Estimation of the Expectation

Point of departure:

• We consider a stationary process \( \{X_t\} \) with mean \( \mu \equiv E(X_t) \)
  and covariance function \( \gamma_X(h) = \text{Cov}(X_t, X_{t+h}) \)

Now:

• Consider

\[
\overline{X}_T = \frac{1}{T} \sum_{t=1}^{T} X_t
\]

as an estimator of \( \mu \)
We have:

\[ E(\overline{X}_T) = \frac{1}{T} \sum_{t=1}^{T} E(X_t) = \mu \]

Furthermore, Theorem 4.2 states:

- If

\[ \sum_{h=-\infty}^{\infty} |\gamma_X(h)| < \infty, \]

then \( \overline{X}_T \) is a consistent estimator of \( \mu \)

(cf. Slide 137)
Summary:

• Provided that \( \sum_{h=-\infty}^{\infty} |\gamma_X(h)| < \infty \) (absolutely summable covariance function) \( \bar{X}_T \) is an unbiased and consistent estimator of the mean \( \mu \)

Now:

• Results on the variance of the estimator \( \bar{X}_T \)
Theorem 4.5: (Variance of $\bar{X}_T$)

Let $\{X_t\}_{t \in \mathbb{Z}}$ be a stationary process with mean $\mu$ and auto-
covariance function $\gamma_X(h)$. We consider the estimator $\bar{X}_T = 1/T \sum_{t=1}^{T} X_t$ of $\mu$.

1. For the variance of $\bar{X}_T$ we have

$$\text{Var} \left( \bar{X}_T \right) = \frac{1}{T} \left[ \gamma_X(0) + 2 \sum_{h=1}^{T-1} \frac{T - h}{T} \gamma_X(h) \right].$$
2. If \( \{ \gamma_X(h) \}_{h \in \mathbb{Z}} \) is absolutely summable, then

\[
\lim_{T \to \infty} \text{Var}(\overline{X}_T) = 0 
\]

and

\[
\lim_{T \to \infty} T \cdot \text{Var}(\overline{X}_T) = \gamma_X(0) + 2 \sum_{h=1}^{\infty} \gamma_X(h) = \sum_{h=-\infty}^{\infty} \gamma_X(h). 
\]

(Proof: Schlittgen & Streitberg, 2001)
Remark:

- Theorems 4.2 and 4.5 hold for arbitrary stationary processes and, in particular, for stationary ARMA($p, q$) processes

Now:

- Results on the statistical distribution of $\bar{X}_T$
  (relevant to constructing confidence intervals, hypothesis tests)

To this end:

- Some claryfing notation
Current notation:

- For a white noise process \( \{X_t\} \sim \text{WN}(\mu, \sigma^2) \) we have
  - \( E(X_t) = \mu \) and \( \text{Var}(X_t) = \sigma^2 \) for all \( t \in \mathbb{Z} \)
  - \( \text{Cov}(X_t, X_s) = 0 \) for all \( t \neq s \)

  (cf. Slide 27)

- For the Gaussian white noise process \( \{X_t\} \sim \text{GWN}(\mu, \sigma^2) \) we have
  - \( X_t \sim N(\mu, \sigma^2) \) for all \( t \in \mathbb{Z} \)
  - \( \text{Cov}(X_t, X_s) = 0 \) for all \( t \neq s \)

  \( \rightarrow \) due to normality \( X_t \) and \( X_s \) are independent

  (cf. Definition 2.3 on Slide 25)
Now:

- Special case of a white noise process

**Definition 4.6: (iid process)**

Let \( \{X_t\}_{t \in \mathbb{Z}} \) be a sequence of stochastically independent random variables that are all identically distributed with \( E(X_t) = \mu \) and \( \text{Var}(X_t) = \sigma^2 \) for all \( t \in \mathbb{Z} \). We then call the process an independent, identically distributed process and write \( \{X_t\} \sim \text{iid}(\mu, \sigma^2) \).
Remarks:

- An iid process is always a white noise process, i.e. if \( \{X_t\} \sim \text{iid}(\mu, \sigma^2) \), then it is also true that \( \{X_t\} \sim \text{WN}(\mu, \sigma^2) \) (the reverse does not hold)

- A Gaussian white noise process is always an iid process, i.e. if \( \{X_t\} \sim \text{GWN}(\mu, \sigma^2) \), then it is also true that \( \{X_t\} \sim \text{iid}(\mu, \sigma^2) \) (the reverse does not hold)

Now:

- Limiting distribution of \( \overline{X}_T \)
Theorem 4.7: (Limiting distribution of $X_T$)

We consider the general linear process

$$X_t = \mu + \sum_{j=-\infty}^{\infty} \psi_j \epsilon_{t-j}$$

with \(\{\epsilon_t\} \sim \text{iid}(0, \sigma^2)\) and the absolutely summable sequence of coefficients \(\{\psi_j\}_{j \in \mathbb{Z}}\). Then the cdf of $\sqrt{T} \left( \bar{X}_T - \mu \right)$ converges towards the cdf of a normal distribution with zero-mean and variance $\sigma^2 \sum_{j=-\infty}^{\infty} \psi_j^2$:

$$\sqrt{T} \left( \bar{X}_T - \mu \right) \xrightarrow{d} N \left( 0, \sigma^2 \sum_{j=-\infty}^{\infty} \psi_j^2 \right).$$

(Proof: Anderson, 1971)
Remarks:

- The limiting distribution of $\bar{X}_T$ holds for causal ARMA($p$, $q$) processes according to Definition 3.7 (Slide 82), which satisfy the causality condition stated in Theorem 3.8 (Slide 84)

- Theorem 4.7 can be viewed as a generalization of the central limit theorem applied to dependent random variables

- For many processes Theorem 4.7 justifies the use of normality theory in inference about $\mu$ as long as we consider a sufficiently large time series
4.3 Estimation of the Autocovariance and the Autocorrelation Function

Now:

- Estimation of $\gamma_X(h)$ and $\rho_X(h)$ of the stationary process $\{X_t\}$

Intuitive estimators: (I)

- Estimators of $\gamma_X(h)$:
  - for known $\mu$: $\hat{\gamma}_X(h) = \frac{1}{T} \sum_{t=1}^{T} (X_t - \mu)(X_{t+h} - \mu)$
  - for unknown $\mu$: $\hat{\gamma}_X(h) = \frac{1}{T} \sum_{t=1}^{T} (X_t - \bar{X}_T)(X_{t+h} - \bar{X}_T)$
Intuitive estimators: (II)

- Estimators of $\rho_X(h)$:
  - for known $\mu$: $\tilde{\rho}_X(h) = \frac{\tilde{\gamma}_X(h)}{\tilde{\gamma}_X(0)}$
  - for unknown $\mu$: $\hat{\rho}_X(h) = \frac{\hat{\gamma}_X(h)}{\hat{\gamma}_X(0)}$
Remarks: (I)

- Statements on the distributions of $\tilde{\gamma}_X(h), \hat{\gamma}_X(h)$ and $\tilde{\rho}_X(h), \hat{\rho}_X(h)$ are more complex than for the estimator $\bar{X}_T$ of $\mu$.

- Theorem 4.4 on Slide 139 states that $\tilde{\gamma}_X(h)$ is a consistent estimator of $\gamma_X(h)$ if $\{X_t\}$ is a normal process with absolutely summable autocovariance function.
Remarks: (II)

- We focus on the estimators

\[
\hat{\gamma}_X(h) = \frac{1}{T} \sum_{t=1}^{T} (X_t - \bar{X}_T)(X_{t+h} - \bar{X}_T)
\]

and

\[
\hat{\rho}_X(h) = \frac{\hat{\gamma}_X(h)}{\hat{\gamma}_X(0)},
\]

since \(\mu\) is generally unknown and has to be estimated by \(\bar{X}_T\)

- Under some conditions \(\hat{\gamma}_X(h)\) and \(\hat{\rho}_X(h)\) provide accurate estimates of \(\gamma_X(h)\) and \(\rho_X(h)\) for sampling sizes \(T > 50\) (cf. Box and Jenkins, 1976)
Now:

- Asymptotic distribution of $\hat{\rho}_X(h)$

**Theorem 4.8: (Asymptotic distribution of $\hat{\rho}_X(h)$)**

We consider the general linear process

$$X_t = \mu + \sum_{j=-\infty}^{\infty} \psi_j \epsilon_{t-j}$$

with $\{\epsilon_t\} \sim \text{iid}(0, \sigma^2)$, $\sum_{j=-\infty}^{\infty} |\psi_j| < \infty$ (absolutely summable sequence of coefficients) and $\sum_{j=-\infty}^{\infty} j |\psi_j|^2 < \infty$. 
Then, for any $h = 1, 2, \ldots$ and for large $T$, the $(h \times 1)$ random vector $[\hat{\rho}_X(1), \ldots, \hat{\rho}_X(h)]'$ approximately follows a multivariate normal distribution:

$$
\begin{bmatrix}
\hat{\rho}_X(1) \\
n \\
\hat{\rho}_X(h)
\end{bmatrix}
\approx
N
\left(
\begin{bmatrix}
\rho_X(1) \\
n \\
\rho_X(h)
\end{bmatrix},
\frac{1}{T}W
\right),
$$

where the elements of the $(h \times h)$ matrix $W = (w_{ij})$ are given by

$$w_{ij} = \sum_{k=1}^{\infty} \left[ \rho_X(k + i) + \rho_X(k - i) - 2\rho_X(i)\rho_X(k) \right] \times \left[ \rho_X(k + j) + \rho_X(k - j) - 2\rho_X(j)\rho_X(k) \right].$$

(Proof: Brockwell and Davis, 1991)
Remark:

- The concept of the "asymptotic distribution" from Theorem 4.8 is different from the concept of the "limiting distribution" used in Theorem 4.7 (Slide 149)

Example #1: (I)

- Let \( \{X_t\} \sim \text{iid}(0, \sigma^2) \), i.e. using the notation from Theorem 4.8

\[
X_t = \epsilon_t = \mu + \sum_{j=-\infty}^{\infty} \psi_j \epsilon_{t-j}
\]

with \( \mu = 0, \psi_0 = 1 \) and \( \psi_j = 0 \) for all \( j \neq 0 \)
Example #1: (II)

- It follows that

\[
\sum_{j=-\infty}^{\infty} |\psi_j| = \ldots + |0| + |\psi_0| + |0| + \ldots = 1 < \infty
\]

and

\[
\sum_{j=-\infty}^{\infty} j|\psi_j|^2 = \ldots + (-1)|0|^2 + 0|\psi_0|^2 + 1|0|^2 + \ldots = 0 < \infty
\]

i.e. the conditions stated in Theorem 4.8 are satisfied
Example #1: (III)

- Owing to the stochastic independence of the $\{X_t\}$, $\gamma_X(h)$ and $\rho_X(h)$ have the form

$$\gamma_X(h) = \begin{cases} 
\sigma^2, & \text{for } h = 0 \\
0, & \text{for } h \neq 0
\end{cases}$$

and

$$\rho_X(h) = \begin{cases} 
1, & \text{for } h = 0 \\
0, & \text{for } h \neq 0
\end{cases}$$
Example #1: (IV)

- Thus, the matrix $\mathbf{W}$ from Theorem 4.8 is given by

$$w_{ij} = \begin{cases} 
1 & \text{, für } i = j \\
0 & \text{, für } i \neq j 
\end{cases}$$

i.e. $\mathbf{W}$ is the $(h \times h)$ identity matrix

$$\mathbf{W} = \begin{bmatrix}
1 & 0 & \cdots & 0 \\
0 & 1 & \cdots & 0 \\
\vdots & \vdots & \ddots & \vdots \\
0 & 0 & \cdots & 1
\end{bmatrix}$$
Example #1: (V)

For large $T$ the estimators $\hat{\rho}_X(1), \ldots, \hat{\rho}_X(h)$ are approximately independent, normally distributed random variables each with mean

$$E[\hat{\rho}_X(1)] = \rho_X(1) = 0, \ldots, E[\hat{\rho}_X(h)] = \rho_X(h) = 0$$

and variance

$$\text{Var}[\hat{\rho}_X(1)] = \ldots = \text{Var}[\hat{\rho}_X(h)] = \frac{1}{T}$$

Thus, a 95% confidence interval for each $\hat{\rho}_X(h)$ ($h \geq 1$) is given by

$$[-1.96 \cdot \sqrt{1/T}, +1.96 \cdot \sqrt{1/T}]$$

(test for an iid process)
Estimated autocorrelation function of the iid process $X_t = \epsilon_t$
with $\epsilon_t \sim GWN(0, 1)$
Example #2: (I)

- Again, let \( \{X_t\} \sim \text{iid}(0, \sigma^2) \) so that the conditions from Theorem 4.8 are satisfied.

- For \( h \in \mathbb{N} \) we now consider the multiple testing problem
  \[
  H_0 : \rho_X(1) = \rho_X(2) = \ldots = \rho_X(h) = 0
  \]
  versus
  \[
  H_1 : \text{For at least one } j \in \{1, \ldots, h\} \text{ we have } \rho_X(j) \neq 0
  \]
  (interpretation of \( H_1 \): the time series is autocorrelated)
Example #2: (II)

- From Theorem 4.8 we have for every $j = 1, 2, \ldots$

  \[
  \hat{\rho}_X(j) \overset{\text{approx.}}{\sim} N(\underbrace{\rho_X(j)}_{=0}, 1/T)
  \text{ under } H_0
  \]

  or, under $H_0$,

  \[
  \sqrt{T} \cdot \hat{\rho}_X(j) \overset{\text{approx.}}{\sim} N(0, 1)
  \]

- Furthermore, owing to the diagonal structure of the matrix $\mathbf{W}$ from Theorem 4.8 the estimators

  \[
  \hat{\rho}_X(1), \ldots, \hat{\rho}_X(h)
  \]

  are (approximately) stochastically independent for large $T$
Example #2: (III)

• Thus, under $H_0$, the following result obtains:

$$\tilde{Q} \equiv \sum_{j=1}^{h} \left[ \sqrt{T} \cdot \hat{\rho}_X(j) \right]^2 = T \sum_{j=1}^{h} [\hat{\rho}_X(j)]^2 \approx \chi^2_h$$

(chi-square distribution with $h$ degrees-of-freedom)

$\rightarrow$ **Box-Pierce-test:**

reject $H_0$ at the significance level $\alpha$, if

$$\tilde{Q} > \chi^2_{h;1-\alpha}$$

$((1 - \alpha)$ quantile of the $\chi^2_h$ distribution)
Remarks:

• Refinement of the Box-Pierce test statistic:

\[ Q \equiv T(T + 2) \sum_{j=1}^{h} \frac{[\hat{\rho}_X(j)]^2}{T - j} \]

• Under \( H_0 \) we have \( Q \approx \chi^2_h \)

\[ Q > \chi^2_{h;1-\alpha} \]

→ Ljung-Box test (also Q test):

Reject \( H_0 \) at the significance level \( \alpha \), if

• The Ljung-Box test has better properties for small and moderate sampling sizes