Estimating mean, acf, and pacf of weakly stationary processes

In the following, we consider a weakly stationary process \((X_t)_{t \in \mathbb{Z}}\) with mean \(\mu := E(X_t)\), autocovariance function \(\gamma_X\), autocorrelation function \(\rho_X\), and partial autocorrelation function \(\alpha_X\). Our task is to estimate these quantities upon knowledge of \(X_1, \ldots, X_n\), i.e. to find estimators

- \(\hat{\mu}\) for the mean \(\mu\),
- \(\hat{\gamma}_X(h)\) for the autocovariance function \(\gamma_X(h)\),
- \(\hat{\rho}_X(h)\) for the acf \(\rho_X(h)\), and
- \(\hat{\alpha}_X(h)\) for the pacf \(\alpha_X(h)\).
Estimating the mean $\mu$ I

to estimate $\mu$, we use $\hat{\mu} := \bar{X}^n := \frac{1}{n}(X_1 + \ldots + X_n)$, which has the following properties:

1. $\hat{\mu}$ is unbiased, i.e. $E\hat{\mu} = \mu$,

2. $\text{Var}(\hat{\mu}) = E(\hat{\mu} - \mu)^2 = \frac{1}{n} \sum_{h=-n}^{n} (1 - \frac{|h|}{n}) \gamma_X(h)$,

3. if $X$ is mean-ergodic, $\hat{\mu}$ is mse-consistent for $\mu$: 
   $E(\hat{\mu} - \mu)^2 \xrightarrow{n \to \infty} 0$,

4. if $\sum_{h=-\infty}^{\infty} |\gamma_X(h)| < \infty$, we have $n \text{Var}(\hat{\mu}) \xrightarrow{n \to \infty} \sum_{h=-\infty}^{\infty} \gamma_X(h)$,

5. in many cases, $\hat{\mu}$ is asymptotically $N(\mu, \frac{\nu}{n})$ distributed with $\nu := \sum_{h=-\infty}^{\infty} \gamma_X(h)$, i.e. $\sqrt{n}(\hat{\mu} - \mu) \xrightarrow{d} N(0, \nu)$. 

Estimating $\mu$ II, estimating $\gamma_X(h), \rho_X(h)$ I

Remarks:
If $\nu$ is known, approximative confidence regions and tests for $\mu$ can be constructed using the asymptotic distribution of $\hat{\mu}$. If $\nu$ is unknown, $\nu$ and thereby also $\gamma_X$ have to be estimated from the (same!) data. For ARMA models, one chooses

$\hat{\nu} := \sum_{|h|<\sqrt{n}} (1 - \frac{|h|}{n})\hat{\gamma}_X(h)$, with $\hat{\gamma}_X(h)$ an estimator for $\gamma_X(h)$.

Estimating $\gamma_X(h), \rho_X(h)$ I

To estimate $\gamma_X(h)$, we use

$$\hat{\gamma}_X(h) := \frac{1}{n} \sum_{t=1}^{n-|h|} (X_{t+|h|} - \hat{\mu})(X_t - \hat{\mu})$$

for $|h| < n$.

$\rho_X(h)$ is then estimated by $\hat{\rho}_X(h) := \frac{\hat{\gamma}_X(h)}{\hat{\gamma}_X(0)}$. 
Estimating $\gamma_X(h)$, $\rho_X(h)$ II

Properties of $\hat{\gamma}$ and $\hat{\rho}$:

1. In general, both $\hat{\gamma}_X(h)$ and $\hat{\rho}_X(h)$ are not unbiased for $\gamma_X(h)$ and $\rho_X(h)$, resp.; the bias remains even if we replace the factor $\frac{1}{n}$ by $\frac{1}{n-|h|}$. However, both estimators are asymptotically unbiased.

2. The matrices $\hat{\Gamma}_n := (\hat{\gamma}_X(i-j))_{i,j=1,...,n}$ and $\hat{R}_n := (\hat{\rho}_X(i-j))_{i,j=1,...,n}$ are positive semidefinite.

3. For values of $h$ close to $n$, the sum in the definition of $\hat{\gamma}_X(h)$ consists of only a few terms, so that the estimator is not very reliable. A rule of thumb by Box and Jenkins suggests $n > 50$ and $h \leq \frac{n}{4}$. 
In many cases, $\hat{\rho}_k := (\hat{\rho}_X(1), \ldots, \hat{\rho}_X(k))'$ is asymptotically $N(\rho_k, \frac{1}{n} W)$ distributed, with $\rho_k := (\rho_X(1), \ldots, \rho_X(k))'$ and $W = (w_{i,j})_{i,j=1 \ldots k}$, where $w_{i,j}$ is given by Bartlett's formula

$$w_{i,j} = \sum_{l=1}^{\infty} \left( \rho_X(l+i) + \rho_X(l-i) - 2\rho_X(i)\rho_X(l) \right) \cdot$$

$$(\rho_X(l+j) + \rho_X(l-j) - 2\rho_X(j)\rho_X(l)) .$$
Example of empirical acf, MA(1) process

Empirical ACF, MA(1) process, n=200, $\theta = -0.9$
Example of empirical acf, MA(1) process II

Empirical ACF, MA(1) process, n=200, $\theta = -0.9$

$\hat{\rho}_X(h)$

WN 95% range of acf
Example of empirical acf, MA(1) process III

Empirical ACF, MA(1) process, n=200, $\theta = -0.9$

WN 95% range of acf
MA(1), $\theta = -0.9$, 95% range of acf
Empirical vs. theoretical acf, MA(1) with $\theta = -0.9$

Empirical acf
95% confidence region

MA(1) acf with $\theta = -0.9$

Empirical acf
95% confidence region
Remark

Using the asymptotic distribution of $\hat{\rho}_k$, it is possible to construct approximative tests for checking whether given data may have been generated by a certain stochastic process.

Estimating $\alpha_X(h)$

To estimate $\alpha_X(h)$, one determines $\hat{\alpha}_X(h)$ as the last component of the solution $\hat{a}_h$ of the system $\hat{R}_h\hat{a}_h = \hat{\rho}_h$.

For computational reasons, this is usually done by applying the Levinson-Durbin recursion.
2. Univariate lineare time series models: ARIMA processes

2.1 Moving Average processes

2.1.1 MA(1) processes

Definition (MA(1) process)

1. A stochastic process \((X_t)_{t \in \mathbb{Z}}\) is called Moving Average process of order 1 or MA(1) process, if there exist a white noise \((\varepsilon_t)_{t \in \mathbb{Z}}\) and a real number \(\theta \in \mathbb{R}\) such that we have for all \(t \in \mathbb{Z}\):

   \[X_t = \varepsilon_t + \theta \varepsilon_{t-1}\.

2. A stochastic process \((X_t)_{t \in \mathbb{Z}}\) is called Moving Average process of order 1 with mean \(\mu\) or MA(1) process with mean \(\mu\), if \(X_t - \mu\) is an MA(1) process according to 1.
In the literature, sometimes the words ‘with mean $\mu$’ are omitted when talking about an MA(1) process with non-zero mean.

In the same vein, a stochastic process $X$ with mean $\mu$ is often called causal or invertible w.r.t. a white noise if $X - \mu$ has the corresponding property.
Properties of MA(1) processes

An MA(1) process $X$ (with mean $\mu$) with $X_t - \mu = \varepsilon_t + \theta \varepsilon_{t-1}$ for a white noise $\varepsilon$ has the following properties:

1. $X$ is weakly stationary and causal w.r.t. $\varepsilon$.

2. $\gamma_X(h) = \begin{cases} (1 + \theta^2)\sigma_\varepsilon^2, & h = 0 \\ \theta \sigma_\varepsilon^2, & |h| = 1 \\ 0, & \text{otherwise} \end{cases}$

3. $\rho_X(h) = \begin{cases} 1, & h = 0 \\ \frac{\theta}{1 + \theta^2}, & |h| = 1 \\ 0, & \text{otherwise} \end{cases}$
Properties of MA(1) processes II

4 For the pacf $\alpha_X$, we have: $\alpha_X(h) = (-1)^{h-1} \frac{\theta^h}{1 + \theta^2 + \ldots + \theta^{2h}}$.

5 W.r.t. invertibility, we have:

1 For $|\theta| < 1$, $X$ is invertible w.r.t. $\varepsilon$:

$$
\varepsilon_t = (X_t - \mu) - \theta(X_{t-1} - \mu) + \theta^2(X_{t-2} - \mu) \pm \ldots
$$

2 For $|\theta| \geq 1$, $X$ is not invertible w.r.t. $\varepsilon$. For $|\theta| > 1$, $X$ is invertible w.r.t. the white noise

$$
Z_t := X_t - \mu - \frac{1}{\theta}(X_{t-1} - \mu) + \frac{1}{\theta^2}(X_{t-2} - \mu) \pm \ldots
$$

and we have: $X_t - \mu = Z_t + \frac{1}{\theta}Z_{t-1}$
Remarks

1. For an MA(1) process $X$ (with mean $\mu$) it follows that the acf $\rho_X$ vanishes from lag 2 on, whereas the pacf $\alpha_X$ doesn’t (except for the trivial case $\theta = 0$), but decays gradually.

2. The acf $\rho_X(1)$ of lag 1 for an MA(1) process $X$ (with mean $\mu$) always stays in the interval $[-\frac{1}{2}, \frac{1}{2}]$.

3. If $X_t$ is allowed to depend not only on $\varepsilon_t$ and $\varepsilon_{t-1}$, but also on further lags of the white noise $\varepsilon$, we arrive at Moving Average processes of higher order.
2.1.2 MA($q$) processes

Definition (MA($q$) process)

1. A stochastic process $(X_t)_{t \in \mathbb{Z}}$ is called Moving Average process of order $q \in \mathbb{N}$ or MA($q$) process, if there exist a white noise $(\varepsilon_t)_{t \in \mathbb{Z}}$ and real numbers $\theta_1, \ldots, \theta_q$ such that we have for all $t \in \mathbb{Z}$:

$$X_t = \varepsilon_t + \theta_1 \varepsilon_{t-1} + \cdots + \theta_q \varepsilon_{t-q}.$$ 

2. A stochastic process $(X_t)_{t \in \mathbb{Z}}$ is called Moving Average process of order $q$ with mean $\mu$ or MA($q$) process with mean $\mu$, if $X_t - \mu$ is an MA($q$) process according to 1.
Properties of MA(\(q\)) processes I

An MA(\(q\)) process \(X\) (with mean \(\mu\)) with

\[
X_t - \mu = \varepsilon_t + \theta_1 \varepsilon_{t-1} + \cdots + \theta_q \varepsilon_{t-q}
\]

for a white noise \(\varepsilon\) has the following properties:

1. \(X\) is weakly stationary and causal w.r.t. \(\varepsilon\).
2. The autocovariance function of \(X\) is given by (\(\theta_0 := 1\)):

\[
\gamma_X(h) = \begin{cases} 
\sigma^2 \sum_{l=0}^{q-|h|} \theta_l \theta_{l+|h|}, & |h| \leq q \\
0 & , \text{otherwise}
\end{cases}
\]
Properties of MA($q$) processes II

3 The acf of $X$ is then given by:

$$\rho_X(h) = \begin{cases} 
q - |h| & \text{if } |h| \leq q \\
\sum_{l=0}^{q-|h|} \theta_l \theta_{l+|h|} & \text{otherwise}
\end{cases}$$

4 W.r.t. the invertibility of $X$, we have:

1. $X$ is invertible w.r.t. $\varepsilon$ if and only if all (possibly complex) roots of the so-called MA-polynomial $1 + \theta_1 z + \theta_2 z^2 + \ldots + \theta_q z^q$ lie outside of the unit circle.

2. If $1 + \theta_1 z + \theta_2 z^2 + \ldots + \theta_q z^q \neq 0$ for all complex numbers $z$ with $|z| = 1$, then there exist $\tilde{\theta}_1, \ldots, \tilde{\theta}_q \in \mathbb{R}$ and a white noise $Z$ such that $X_t = Z_t + \tilde{\theta}_1 Z_{t-1} + \ldots + \tilde{\theta}_q Z_{t-q}$ and $X$ is invertible w.r.t. $Z$. 
Remarks

1. For an MA($q$) process (with mean $\mu$) it follows that the acf $\rho_X$ vanishes from lag $q + 1$ on, whereas the pacf $\alpha_X$ in general doesn’t, but decays gradually.

2. MA($q$) processes are related to $q$-correlated processes which we will define below.

**Definition (q-correlated)**

A weakly stationary process $X$ is called $q$-correlated for a natural number $q \in \mathbb{N}$, if we have:

$$\gamma_X(h) = 0 \text{ for } |h| > q \text{ and } \gamma_X(q) \neq 0.$$
Theorem

1. If $X$ is an MA($q$) process (with mean $\mu$) with $\theta_q \neq 0$, then $X$ is $q$-correlated.

2. If $X$ is a $q$-correlated process, there exist a white noise $(\varepsilon_t)_{t \in \mathbb{Z}}$ and real numbers $\theta_1, \ldots, \theta_q$ such that we have with $\mu := E(X_t)$:

$$X_t - \mu = \varepsilon_t + \theta_1 \varepsilon_{t-1} + \cdots + \theta_q \varepsilon_{t-q},$$

i.e. $X$ is an MA($q$) process (with mean $\mu$).
The preceding theorem tells us that the class of all MA($q$) processes equals the class of all weakly stationary processes whose autocorrelation function vanishes from lag $q + 1$ on.

As for instance in the Wold decomposition, one often encounters representations of a stochastic process as an infinite sum of lagged values of a white noise. This leads to the notion of MA($\infty$) processes.
2.1.3 MA(∞) processes

Definition (MA(∞) process)

1. A stochastic process \((X_t)_{t \in \mathbb{Z}}\) is called Moving Average process of infinite order or MA(∞) process, if there exist a white noise \((\varepsilon_t)_{t \in \mathbb{Z}}\) and real numbers \(\theta_1, \theta_2, \ldots\) with \(\sum_{j=1}^{\infty} |\theta_j| < \infty\) such that for all \(t \in \mathbb{Z}\), we have:

\[
X_t = \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \cdots.
\]

2. A stochastic process \((X_t)_{t \in \mathbb{Z}}\) is called Moving Average process of infinite order with mean \(\mu\) or MA(∞) process with mean \(\mu\), if \(X_t - \mu\) is an MA(∞) process according to 1.
The definitions of MA(1)-, MA(q)-, and MA(∞) processes (with mean \( \mu \)) are chosen in such a way that every MA(1) process (with mean \( \mu \)) is an MA(q)-Prozess (with mean \( \mu \)) and every MA(q) process (with mean \( \mu \)) is an MA(∞) process (with mean \( \mu \)).

It is easy to see that every process \( X \) of the form
\[
X_t = \mu + \sum_{j=0}^{\infty} \psi_j \varepsilon_{t-j},
\]
with \( \varepsilon \) a white noise and \( \sum_{j=0}^{\infty} |\psi_j| < \infty \), is an MA(∞) process with mean \( \mu \).

The class of all MA(∞) processes (with mean \( \mu \)) contains more processes than one might think after a first glance. For instance, all AR(1) processes of the form \( X_t = \phi X_{t-1} + \varepsilon_t \) for a white noise \( \varepsilon \) with \( |\phi| \neq 1 \) are also MA(∞) processes.
Properties of MA($\infty$) processes

An MA($\infty$) process $X$ (with mean $\mu$) with $X_t - \mu = \varepsilon_t + \sum_{j=1}^{\infty} \theta_j \varepsilon_{t-j}$

for a white noise $\varepsilon$ and $\sum_{j=0}^{\infty} |\theta_j| < \infty$ has the following properties:

1. $X$ is weakly stationary and causal w.r.t. $\varepsilon$.

2. For $\gamma_X(h)$, we have with $\theta_0 := 1$: $\gamma_X(h) = \sigma_\varepsilon^2 \sum_{l=0}^{\infty} \theta_l \theta_{l+|h|}$.

3. For the acf of $X$, we have: $\rho_X(h) = \frac{\sum_{l=0}^{\infty} \theta_l \theta_{l+|h|}}{\sum_{l=0}^{\infty} \theta_l^2}$.