Parameter estimation: ACVF of AR processes

Yule-Walker’s for AR processes: a method of moments, i.e. $\mu = \bar{x}$

and choose parameters so that $\gamma(h) = \hat{\gamma}(h)$ (for $h$ “small”).
Parameter estimation: ACVF of AR processes

Yule-Walker’s for AR processes: a method of moments, i.e. $\mu = \bar{x}$ and choose parameters so that $\gamma(h) = \hat{\gamma}(h)$ (for $h$ “small”).

Multiplying $X_t = \varphi_1 X_{t-1} + \cdots + \varphi_p X_{t-p} + Z_t$ by $X_{t-h}$ ($h = 1 \ldots p$) and taking expectations, one has $\Gamma_p \phi = \gamma_p$ with $(\Gamma_p)_{i,j} = \gamma(i-j)$, i.e.

$$\Gamma_p = \begin{pmatrix} \gamma(0) & \gamma(1) & \gamma(2) & \cdots & \gamma(p-1) \\ \gamma(1) & \gamma(0) & \gamma(1) & \cdots & \vdots \\ \gamma(2) & \gamma(1) & \gamma(0) & \cdots & \vdots \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \gamma(p-1) & \cdots & \gamma(1) & \gamma(0) \end{pmatrix}, \quad \gamma_p = \begin{pmatrix} \gamma(1) \\ \gamma(2) \\ \vdots \\ \gamma(p) \end{pmatrix},$$

and multiplying by $X_t$ one obtains $\sigma^2 = \gamma(0) - \phi^t \gamma_p$. 
Parameter estimation: Yule-Walker’s method

Hence estimate $\hat{\phi} = \begin{pmatrix} \hat{\phi}_1 \\ \hat{\phi}_2 \\ \vdots \\ \hat{\phi}_p \end{pmatrix}$ by solving $\hat{\Gamma}_p \hat{\phi} = \hat{\gamma}_p$. \hspace{1cm} (1)

Moreover $\hat{\sigma}^2 = \hat{\gamma}(0) - \hat{\phi}^t \hat{\gamma}_p$. \hspace{1cm} (2)
Parameter estimation: Yule-Walker’s method

Hence estimate $\hat{\phi} = \begin{pmatrix} \hat{\phi}_1 \\ \hat{\phi}_2 \\ \vdots \\ \hat{\phi}_p \end{pmatrix}$ by solving $\hat{\Gamma}_p \hat{\phi} = \hat{\gamma}_p$. \hspace{1cm} (1)

Moreover $\sigma^2 = \hat{\gamma}(0) - \hat{\phi}^t \hat{\gamma}_p$. \hspace{1cm} (2)

Remarks
Parameter estimation: Yule-Walker’s method

Hence estimate \( \hat{\phi} = \begin{pmatrix} \hat{\phi}_1 \\ \hat{\phi}_2 \\ \vdots \\ \hat{\phi}_p \end{pmatrix} \) by solving \( \hat{\Gamma}_p \hat{\phi} = \hat{\gamma}_p \). \hspace{1cm} (1)

Moreover \( \hat{\sigma}^2 = \hat{\gamma}(0) - \hat{\phi}^t \hat{\gamma}_p \). \hspace{1cm} (2)

Remarks

1. We proved \( \hat{\Gamma}_p \) non-negative definite.
Parameter estimation: Yule-Walker’s method

Hence estimate \( \hat{\phi} = \begin{pmatrix} \hat{\phi}_1 \\ \hat{\phi}_2 \\ \vdots \\ \hat{\phi}_p \end{pmatrix} \) by solving \( \hat{\Gamma}_p \hat{\phi} = \hat{\gamma}_p \). \( (1) \)

Moreover \( \sigma^2 = \hat{\gamma}(0) - \hat{\phi}^t \hat{\gamma}_p \). \( (2) \)

Remarks

1. We proved \( \hat{\Gamma}_p \) non-negative definite. If \( \hat{\gamma}(0) > 0 \) (i.e. data are not identical), is positive definite, hence invertible \( \implies \hat{\phi} \) well-defined.
Parameter estimation: Yule-Walker’s method

Hence estimate \( \hat{\phi} = \begin{pmatrix} \hat{\phi}_1 \\ \hat{\phi}_2 \\ \vdots \\ \hat{\phi}_p \end{pmatrix} \) by solving \( \hat{\Gamma}_p \hat{\phi} = \hat{\gamma}_p \). (1)

Moreover \( \hat{\sigma}^2 = \hat{\gamma}(0) - \hat{\phi}^t \hat{\gamma}_p \). (2)

Remarks

1. We proved \( \hat{\Gamma}_p \) non-negative definite. If \( \hat{\gamma}(0) > 0 \) (i.e. data are not identical), is positive definite, hence invertible \( \implies \hat{\phi} \) well-defined.

2. One can divide (1) by \( \gamma(0) \) to obtain \( \hat{R}_p \hat{\phi} = \hat{\rho}_p \) with \( (R_p)_{i,j} = \rho(i-j), \rho_i = \rho(i) \).
Parameter estimation: Yule-Walker’s method

Hence estimate \( \hat{\phi} = \begin{pmatrix} \hat{\phi}_1 \\ \hat{\phi}_2 \\ \vdots \\ \hat{\phi}_p \end{pmatrix} \) by solving \( \hat{\Gamma}_p \hat{\phi} = \hat{\gamma}_p \). \hspace{1cm} (1)

Moreover \( \hat{\sigma}^2 = \hat{\gamma}(0) - \hat{\phi}^T \hat{\gamma}_p \). \hspace{1cm} (2)

Remarks

1. We proved \( \hat{\Gamma}_p \) non-negative definite. If \( \hat{\gamma}(0) > 0 \) \textit{(i.e. data are not identical)}, is positive definite, hence invertible \( \Rightarrow \) \( \hat{\phi} \) well-defined.

2. One can divide (1) by \( \gamma(0) \) to obtain \( \hat{R}_p \hat{\phi} = \hat{\rho}_p \) with \( (R_p)_{i,j} = \rho(i-j), \rho_i = \rho(i) \).

3. One can compute \( \phi \) iteratively, applying Durbin-Levinson to \( \hat{\gamma} \).
Asymptotic properties of Yule-Walker

Theorem

If \( \{X_t\} \) is the causal AR process satisfying

\[
X_t = \varphi_1 X_{t-1} + \cdots + \varphi_p X_{t-p} + Z_t \quad \{Z_t\} \sim \text{IID}(0, \sigma^2)
\]

then

\[
n^{1/2}(\hat{\phi} - \phi) \overset{d}{\rightarrow} N(0, \sigma^2 \Gamma_p^{-1}) \quad \text{and} \quad \hat{\sigma}^2 \overset{P}{\rightarrow} \sigma^2.
\]
Asymptotic properties of Yule-Walker

**Theorem**

If \( \{X_t\} \) is the causal AR process satisfying

\[
X_t = \varphi_1 X_{t-1} + \cdots + \varphi_p X_{t-p} + Z_t \quad \{Z_t\} \sim \text{IID}(0, \sigma^2)
\]

then

\[
n^{1/2}(\hat{\phi} - \phi) \xrightarrow{d} N(0, \sigma^2 \Gamma_p^{-1}) \quad \text{and} \quad \hat{\sigma}^2 \xrightarrow{P} \sigma^2.
\]

An approximate \( \gamma \)-confidence interval for \( \varphi_i, \ i = 1 \ldots p \) is then

\[
(\hat{\phi}_i - \hat{\sigma} \left[ (\hat{\Gamma}_p)^{-1} \right]_{ii}^{1/2} \frac{Z_{\gamma}}{\sqrt{n}}, \hat{\phi}_i + \hat{\sigma} \left[ (\hat{\Gamma}_p)^{-1} \right]_{ii}^{1/2} \frac{Z_{\gamma}}{\sqrt{n}})
\]

where \( P(|N(0, 1)| \leq z_\gamma) = \gamma \).
Comparison with linear regression

One could write the problem as \( Y = X\phi + Z \) with

\[
Y = \begin{pmatrix} X_n \\ X_{n-1} \\ \vdots \\ X_{p+1} \end{pmatrix} \quad X = \begin{pmatrix} X_{n-1} & \cdots & X_{n-p} \\ X_{n-2} & \cdots & X_{n-p-1} \\ \vdots & \ddots & \vdots \\ X_p & \cdots & X_1 \end{pmatrix}
\]

and \( Z \sim \text{IID}(0, \sigma^2) \).

Least square estimation asks for minimizing \( \| Y - X\phi \|^2 \) and results into

\[
\hat{\phi} = (X^t X)^{-1} X^t Y.
\]

\[
X^t Y = \begin{pmatrix} \sum_{t=p+1}^{n} X_t X_{t-1} \\ \vdots \\ \sum_{t=p+1}^{n} X_t X_{t-p} \end{pmatrix} \approx n\hat{\gamma}_p \quad (X^t X)_{ij} = \sum_{t=p-i+1}^{p-i} X_t X_{t+i-j} \approx n(\hat{\Gamma})_{ij}.
\]
Yule-Walker’s method finds $\hat{\phi}$ as the coefficients of the best linear predictor of $X_{p+1}$ using $X_1, \ldots, X_p$, assuming that the ACF coincides with the sample ACF for $i = 1, \ldots p$. 
Variations on Yule-Walker: Burg’s algorithm

Yule-Walker’s method finds $\hat{\phi}$ as the coefficients of the best linear predictor of $X_{p+1}$ using $X_1, \ldots, X_p$, assuming that the ACF coincides with the sample ACF for $i = 1, \ldots p$.

Burg’s algorithm considers actual forward and backward linear predictors. Define $u_p(t) \ (p \leq t < n)$ as the difference between $X_{n-t+p}$ and the best linear prediction using the previous $p$ values, i.e.

$$u_p(t) = X_{n-t+p} - \mathbf{P}_L(X_{n-t}, \ldots, X_{n-t+p-1})X_{n-t+p}$$

and $v_p(t) \ (1 \leq t < n)$ be the difference between the value of $X_{n-t}$ and the best linear prediction using the following $p$ values, i.e.

$$v_p(t) = X_{n-t} - \mathbf{P}_L(X_{n-t+1}, \ldots, X_{n-t+p})X_{n-t}.$$
Variations on Yule-Walker: Burg’s algorithm

Yule-Walker’s method finds $\hat{\phi}$ as the coefficients of the best linear predictor of $X_{p+1}$ using $X_1, \ldots, X_p$, assuming that the ACF coincides with the sample ACF for $i = 1, \ldots, p$.

Burg’s algorithm considers actual forward and backward linear predictors. Define $u_p(t)$ ($p \leq t < n$) as the difference between $X_{n-t+p}$ and the best linear prediction using the previous $p$ values, i.e.

$$u_p(t) = X_{n-t+p} - \mathbf{P}_L(X_{n-t}, \ldots, X_{n-t+p-1})X_{n-t+p}$$

and $v_p(t)$ ($1 \leq t < n$) be the difference between the value of $X_{n-t}$ and the best linear prediction using the following $p$ values, i.e.

$$v_p(t) = X_{n-t} - \mathbf{P}_L(X_{n-t+1}, \ldots, X_{n-t+p})X_{n-t}.$$ 

The coefficients are found by minimizing $\sum_{t=p}^{n-1}(u_p^2(t) + v_p^2(t))$. Details on the algorithm in an exercise.
Order of the auto-regressive process

Assume the true model is AR(p) and we fit a model of order $m$. If $m < p$ (i.e., the model is wrong), we should see this from the residuals.

$$X_t - \hat{X}_t = X_t - p \sum_{j=1}^{p} \phi_j X_{t-j} = Z_t \sim WN(0, \sigma^2) \text{ for } t > p.$$  

$\hat{W}_t = X_t - p \sum_{j=1}^{m} \hat{\phi}_j X_{t-j}$ are not exactly WN, but close for $n$ large.

If $m < p$ and we compute $(\hat{W}_m)_t = X_t - m \sum_{j=1}^{m} (\hat{\phi}_m)_j X_{t-j}$, we expect to find them correlated. Significant ACF of $\hat{W}_m$ suggests $m < p$.

If $m \geq p$, $n^{-1/2} (\hat{\phi}_m - \phi) = \Rightarrow N(0, \sigma^2 \Gamma^{-1} m)$ (Theorem).

In particular for $m > p$, $\hat{\phi}_m \sim N(0, 1/n)$. Finding $\hat{\phi}_m$ within $\pm 1.96 n^{-1/2}$ suggests that the true value of $\phi_m$ may be 0, and to fit a model of lower order.
Order of the auto-regressive process

Assume the true model is AR(p) and we fit a model of order $m$. If $m < p$ (i.e. the model is wrong), we should see this from the residuals.

$$X_t - \hat{X}_t = X_t - \sum_{j=1}^{p} \varphi_j X_{t-j} = Z_t \sim WN(0, \sigma^2) \text{ for } t > p.$$
Assume the true model is AR($p$) and we fit a model of order $m$. If $m < p$ (i.e. the model is wrong), we should see this from the residuals.

$$X_t - \hat{X}_t = X_t - \sum_{j=1}^{p} \varphi_j X_{t-j} = Z_t \sim WN(0, \sigma^2) \text{ for } t > p.$$  

$\hat{W}_t = X_t - \sum_{j=1}^{p} \hat{\varphi}_j X_{t-j}$ are not exactly WN, but close for $n$ large.
Order of the auto-regressive process

Assume the true model is AR(p) and we fit a model of order $m$. If $m < p$ (i.e. the model is wrong), we should see this from the residuals.

$$X_t - \hat{X}_t = X_t - \sum_{j=1}^{p} \varphi_j X_{t-j} = Z_t \sim WN(0, \sigma^2) \text{ for } t > p.$$  

$\hat{W}_t = X_t - \sum_{j=1}^{p} \hat{\varphi}_j X_{t-j}$ are not exactly WN, but close for $n$ large.

If $m < p$ and we compute $(\hat{W}_m)_t = X_t - \sum_{j=1}^{m} (\hat{\varphi}_m)_j X_{t-j}$, we expect to find them correlated.
Order of the auto-regressive process

Assume the true model is AR(p) and we fit a model of order \( m \). If \( m < p \) (i.e. the model is wrong), we should see this from the residuals.

\[
X_t - \hat{X}_t = X_t - \sum_{j=1}^{p} \varphi_j X_{t-j} = Z_t \sim WN(0, \sigma^2) \quad \text{for } t > p.
\]

\[\hat{W}_t = X_t - \sum_{j=1}^{p} \hat{\varphi}_j X_{t-j}\] are not exactly WN, but close for \( n \) large.

If \( m < p \) and we compute \((\hat{W}_m)_t = X_t - \sum_{j=1}^{m} (\hat{\varphi}_m)_j X_{t-j}\), we expect to find them correlated. Significant ACF of \( \hat{W}_m \) suggests \( m < p \).
Order of the auto-regressive process

Assume the true model is AR(p) and we fit a model of order $m$. If $m < p$ (i.e. the model is wrong), we should see this from the residuals.

$$X_t - \hat{X}_t = X_t - \sum_{j=1}^{p} \varphi_j X_{t-j} = Z_t \sim WN(0, \sigma^2) \quad \text{for } t > p.$$ 

$\hat{W}_t = X_t - \sum_{j=1}^{p} \hat{\varphi}_j X_{t-j}$ are not exactly WN, but close for $n$ large.

If $m < p$ and we compute $(\hat{W}_m)_t = X_t - \sum_{j=1}^{m} (\hat{\varphi}_m)_j X_{t-j}$, we expect to find them correlated. Significant ACF of $\hat{W}_m$ suggests $m < p$.

If $m \geq p$, $n^{1/2}(\hat{\varphi}_m - \phi) \implies N(0, \sigma^2 \Gamma_m^{-1})$ (Theorem).

In particular for $m > p$, $\hat{\varphi}_m \sim N(0, 1/n)$. 

Significant ACF of $\hat{W}_m$ suggests $m < p$. 

If $m \geq p$, $n^{1/2}(\hat{\varphi}_m - \phi) \implies N(0, \sigma^2 \Gamma_m^{-1})$ (Theorem).

In particular for $m > p$, $\hat{\varphi}_m \sim N(0, 1/n)$. 

Finding $\hat{\varphi}_m$ within $\pm 1.96 n^{-1/2}$ suggests that the true value of $\varphi_m$ may be 0, and to fit a model of lower order.
Estimation in ARMA processes

Order of the auto-regressive process

Assume the true model is AR(p) and we fit a model of order \( m \). If \( m < p \) (i.e. the model is wrong), we should see this from the residuals.

\[
X_t - \hat{X}_t = X_t - \sum_{j=1}^{p} \varphi_j X_{t-j} = Z_t \sim WN(0, \sigma^2) \quad \text{for } t > p.
\]

\( \hat{W}_t = X_t - \sum_{j=1}^{p} \hat{\varphi}_j X_{t-j} \) are not exactly WN, but close for \( n \) large.

If \( m < p \) and we compute \((\hat{W}_m)_t = X_t - \sum_{j=1}^{m} (\hat{\varphi}_m)_j X_{t-j}\), we expect to find them correlated. Significant ACF of \( \hat{W}_m \) suggests \( m < p \).

If \( m \geq p \), \( n^{1/2}(\hat{\varphi}_m - \varphi) \Rightarrow N(0, \sigma^2 \Gamma_m^{-1}) \) (Theorem).

In particular for \( m > p \), \( \hat{\varphi}_m \sim N(0, 1/n) \).

Finding \( \hat{\varphi}_m \) within \( \pm 1.96 n^{-1/2} \) suggests that the true value of \( \varphi_m \) may be 0, and to fit a model of lower order.
A standard method of parameter estimate is maximum likelihood estimation (MLE).
A standard method of parameter estimate is maximum likelihood estimation (MLE).
Likelihood of ARMA processes is computed assuming \( \{Z_t\} \sim IIDN(0, \sigma^2) \).
Then \( \{X_t\} \) is also normally distributed.
A standard method of parameter estimate is maximum likelihood estimation (MLE).
Likelihood of ARMA processes is computed assuming \( \{Z_t\} \sim IIDN(0, \sigma^2) \).
Then \( \{X_t\} \) is also normally distributed.

**Example of AR(1):** \( X_t = \mu + \varphi(X_{t-1} - \mu) + Z_t, \ |\varphi| < 1 \).
Then \( X_t \sim N(\mu, \frac{\sigma^2}{1-\varphi^2}) \).
Maximum likelihood estimation

A standard method of parameter estimate is maximum likelihood estimation (MLE).
Likelihood of ARMA processes is computed assuming \( \{Z_t\} \sim IIDN(0, \sigma^2) \). Then \( \{X_t\} \) is also normally distributed.

**Example of AR(1):** \( X_t = \mu + \varphi(X_{t-1} - \mu) + Z_t, |\varphi| < 1 \).
Then \( X_t \sim N(\mu, \frac{\sigma^2}{1-\varphi^2}) \).
Joint distributions can be obtained by conditional distributions: conditional on the value of \( X_{t-1} \), \( X_t \sim N(\mu + \varphi(X_{t-1} - \mu), \sigma^2) \).
A standard method of parameter estimate is maximum likelihood estimation (MLE).
Likelihood of ARMA processes is computed assuming \( \{Z_t\} \sim \text{IIDN}(0, \sigma^2) \).
Then \( \{X_t\} \) is also normally distributed.

**Example of AR(1):** \( X_t = \mu + \varphi(X_{t-1} - \mu) + Z_t, \ |\varphi| < 1 \).
Then \( X_t \sim \mathcal{N}(\mu, \frac{\sigma^2}{1-\varphi^2}) \).
Joint distributions can be obtained by conditional distributions:
conditional on the value of \( X_{t-1} \), \( X_t \sim \mathcal{N}(\mu + \varphi(X_{t-1} - \mu), \sigma^2) \).
Hence likelihood of data \( (x_1, x_2, \ldots, x_n) \) can be obtained as

\[
f(x_1)f(x_2|x_1) \cdots f(x_n|x_{n-1})
\]

where \( f(\cdot) \) is the (unconditional) density of \( X_t \),
while \( f(\cdot|x) \) is the density of \( X_t \) conditional on \( X_{t-1} = x \).
Computing likelihood for AR(1)

\[ f(x) = \left(2\pi \frac{\sigma^2}{1 - \varphi^2}\right)^{-1/2} \exp\left\{ -\frac{(x - \mu)^2}{2\sigma^2/(1 - \varphi^2)} \right\} \]

\[ = \left(\frac{1 - \varphi^2}{2\pi \sigma^2}\right)^{1/2} \exp\left\{ -\frac{(1 - \varphi^2)(x - \mu)^2}{2\sigma^2} \right\} \]

\[ f(y, x) = (2\pi \sigma^2)^{-1/2} \exp\left\{ -\frac{(y - \mu - \varphi(x - \mu))^2}{2\sigma^2} \right\}. \]
Computing likelihood for AR(1)

\[ f(x) = (2\pi \frac{\sigma^2}{1 - \varphi^2})^{-1/2} \exp\left\{ -\frac{(x - \mu)^2}{2\sigma^2/(1 - \varphi^2)} \right\} \]

\[ = \left( \frac{1 - \varphi^2}{2\pi\sigma^2} \right)^{1/2} \exp\left\{ -\frac{(1 - \varphi^2)(x - \mu)^2}{2\sigma^2} \right\} \]

\[ f(y, x) = (2\pi\sigma^2)^{-1/2} \exp\left\{ -\frac{(y - \mu - \varphi(x - \mu))^2}{2\sigma^2} \right\}. \]

Parameters of the model are \( \mu, \varphi, \sigma^2 \):

\[ L(\mu, \varphi, \sigma^2) = (1 - \varphi^2)^{1/2}(2\pi\sigma^2)^{-n/2} \exp\left\{ -\frac{(1 - \varphi^2)(x_1 - \mu)^2}{2\sigma^2} \right\} \]

\[ \times \prod_{j=2}^{n} \exp\left\{ -\frac{(x_j - \mu - \varphi(x_{j-1} - \mu))^2}{2\sigma^2} \right\} \]
Computing likelihood for AR(1)

\[ f(x) = \left(2\pi \frac{\sigma^2}{1 - \varphi^2}\right)^{-1/2} \exp\left\{ -\frac{(x - \mu)^2}{2\sigma^2/(1 - \varphi^2)} \right\} \]

\[ = \left( \frac{1 - \varphi^2}{2\pi \sigma^2} \right)^{1/2} \exp\left\{ -\frac{(1 - \varphi^2)(x - \mu)^2}{2\sigma^2} \right\} \]

\[ f(y, x) = (2\pi \sigma^2)^{-1/2} \exp\left\{ -\frac{(y - \mu - \varphi(x - \mu))^2}{2\sigma^2} \right\}. \]

Parameters of the model are \(\mu, \varphi, \sigma^2:\)

\[ L(\mu, \varphi, \sigma^2) = (1 - \varphi^2)^{1/2} (2\pi \sigma^2)^{-n/2} \exp\left\{ -\frac{(1 - \varphi^2)(x_1 - \mu)^2}{2\sigma^2} \right\} \]

\[ \times \prod_{j=2}^{n} \exp\left\{ -\frac{(x_j - \mu - \varphi(x_{j-1} - \mu))^2}{2\sigma^2} \right\} \]

\[ = \frac{(1 - \varphi^2)^{1/2}}{(2\pi \sigma^2)^{n/2}} \exp\left\{ -\frac{1}{2\sigma^2} \left[ (1 - \varphi^2)(x_1 - \mu)^2 + \sum_{t=2}^{n} (x_t - \mu - \varphi(x_{t-1} - \mu))^2 \right] \right\} \]