Introduction to Time Series Analysis. Lecture 9.

Peter Bartlett

Last lecture:

1. Linear prediction.
2. Projection in Hilbert space.
3. Forecasting and backcasting.

Introduction to Time Series Analysis. Lecture 9.

Peter Bartlett

1. Review: Linear prediction.
2. Prediction operator.
3. Partial autocorrelation function.
4. Recursive methods: Durbin-Levinson.

Review: Linear prediction

Given $X_1, X_2, \ldots, X_n$, the best linear predictor

$$X_{n+m}^n = \alpha_0 + \sum_{i=1}^n \alpha_i X_i$$

of X_{n+m} satisfies the **prediction equations**

$$\mathrm{E}\left(X_{n+m} - X_{n+m}^n\right) = 0$$

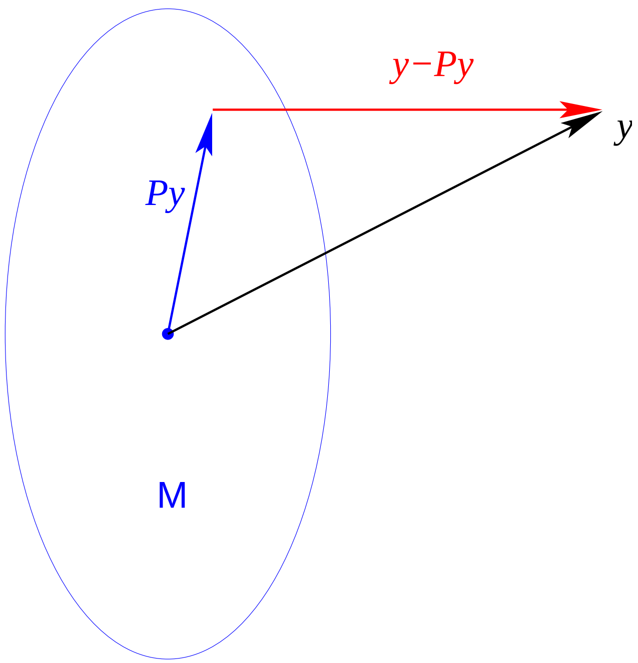
$$\mathrm{E}\left[\left(X_{n+m} - X_{n+m}^n\right) X_i\right] = 0 \qquad \text{for } i = 1, \ldots, n.$$

That is, the *prediction errors* ($X_{n+m}^n - X_{n+m}$) *are uncorrelated with the prediction variables* ($1, X_1, \ldots, X_n$).

Review: Projection Theorem

If $\mathcal{H}$ is a Hilbert space,
$\mathcal{M}$ is a closed linear subspace of $\mathcal{H}$,
and $y \in \mathcal{H}$,
then there is a point $Py \in \mathcal{M}$
(the **projection of** y **on** $\mathcal{M}$)
satisfying

1. $\|Py - y\| \leq \|w - y\|$ for $w \in \mathcal{M}$,
2. $\|Py - y\| < \|w - y\|$ for $w \in \mathcal{M}$, $w \neq y$
3. $\langle y - Py, w \rangle = 0$ for $w \in \mathcal{M}$.



Review: One-step-ahead linear prediction

$$X_{n+1}^{n} = \phi_{n1}X_n + \phi_{n2}X_{n-1} + \cdots + \phi_{nn}X_1$$

$$\Gamma_n \phi_n = \gamma_n,$$

$$P_{n+1}^{n} = \mathrm{E}\left(X_{n+1} - X_{n+1}^{n}\right)^2 = \gamma(0) - \gamma_n' \Gamma_n^{-1} \gamma_n,$$

$$\Gamma_n = \begin{bmatrix} \gamma(0) & \gamma(1) & \cdots & \gamma(n-1) \\ \gamma(1) & \gamma(0) & & \gamma(n-2) \\ \vdots & & \ddots & \vdots \\ \gamma(n-1) & \gamma(n-2) & \cdots & \gamma(0) \end{bmatrix},$$

$$\phi_n = (\phi_{n1}, \phi_{n2}, \ldots, \phi_{nn})', \quad \gamma_n = (\gamma(1), \gamma(2), \ldots, \gamma(n))'.$$

Introduction to Time Series Analysis. Lecture 9.

1. Review: Linear prediction.
2. Prediction operator.
3. Partial autocorrelation function.
4. Recursive methods: Durbin-Levinson.

The prediction operator

For random variables $Y, Z_1, \ldots, Z_n$, define the **best linear prediction of** Y **given** $Z = (Z_1, \ldots, Z_n)'$ as the operator $P(\cdot|Z)$ applied to Y:

$$P(Y|Z) = \mu_Y + \phi'(Z - \mu_Z)$$

with

$$\Gamma\phi = \gamma,$$

where

$$\gamma = \text{Cov}(Y, Z)$$

$$\Gamma = \text{Cov}(Z, Z).$$

Properties of the prediction operator

1. $\mathrm{E}(Y - P(Y|Z)) = 0$, $\mathrm{E}((Y - P(Y|Z))Z) = 0$.

2. $\mathrm{E}((Y - P(Y|Z))^2) = \mathrm{Var}(Y) - \phi'\gamma$.

3. $P(\alpha_1 Y_1 + \alpha_2 Y_2 + \alpha_0 | Z) = \alpha_0 + \alpha_1 P(Y_1|Z) + \alpha_2 P(Y_2|Z)$.

4. $P(Z_i|Z) = Z_i$.

5. $P(Y|Z) = \mathrm{E}Y$ if $\gamma = 0$.

Example: predicting m steps ahead

Write
$$X^n_{n+m} = \phi^{(m)}_{n1} X_n + \phi^{(m)}_{n2} X_{n-1} + \cdots + \phi^{(m)}_{nn} X_1$$
$$\Gamma_n \phi^{(m)}_n = \gamma^{(m)}_n,$$

with
$$\Gamma_n = \mathrm{Cov}(X, X),$$
$$\gamma^{(m)}_n = \mathrm{Cov}(X_{n+m}, X) = (\gamma(m), \gamma(m+1), \ldots, \gamma(m+n-1))'.$$

Also,
$$\mathrm{E}((X_{n+m} - X^n_{n+m})^2) = \gamma(0) - \phi^{(m)\prime} \gamma^{(m)}_n.$$

Introduction to Time Series Analysis. Lecture 9.

1. Review: Linear prediction.
2. Prediction operator.
3. Partial autocorrelation function.
4. Recursive methods: Durbin-Levinson.

Partial autocovariance function

AR(1) model: $X_t = \phi_1 X_{t-1} + W_t$

$$\begin{aligned}
\gamma(1) &= \text{Cov}(X_0, X_1) = \phi_1 \gamma(0) \\
\gamma(2) &= \text{Cov}(X_0, X_2) \\
&= \text{Cov}(X_0, \phi_1 X_1 + W_2) \\
&= \text{Cov}(X_0, \phi_1^2 X_0 + \phi_1 W_1 + W_2) \\
&= \phi_1^2 \gamma(0).
\end{aligned}$$

Clearly, X_0 and X_2 are correlated through X_1.

In the PACF, we remove this dependence by considering the covariance of the *prediction errors* of X_2^1 and X_0^1.

Partial autocovariance function

For AR(1) model:

$$X_2^1 = \phi_1 X_1,$$

$$X_0^1 = \phi_1 X_1,$$

so

$$\begin{aligned} \mathrm{Cov}(X_2^1 - X_2, X_0^1 - X_0) &= \mathrm{Cov}(\phi_1 X_1 - X_2, \phi_1 X_1 - X_0) \\ &= \mathrm{Cov}(W_2, \phi_1 X_1 - X_0) \\ &= 0. \end{aligned}$$

Partial autocorrelation function

The Partial AutoCorrelation Function (PACF) of a stationary time series $\{X_t\}$ is

$$\phi_{11} = \text{Corr}(X_1, X_0) = \rho(1)$$

$$\phi_{hh} = \text{Corr}(X_h - X_h^{h-1}, X_0 - X_0^{h-1}) \quad \text{for } h = 2, 3, \ldots$$

This removes the linear effects of $X_1, \ldots, X_{h-1}$:

$$\ldots, X_{-1}, \underline{X_0}, \underbrace{X_1, X_2, \ldots, X_{h-1}}_{\text{partial out}}, \underline{X_h}, X_{h+1}, \ldots$$

Partial autocorrelation function

The PACF ϕ_{hh} is also the last coefficient in the best linear prediction of X_{h+1} given $X_1, \ldots, X_h$:

$$\Gamma_h \phi_h = \gamma_h \qquad X_{h+1}^h = \phi_h' X$$

$$\phi_h = (\phi_{h1}, \phi_{h2}, \ldots, \phi_{hh}).$$

Example: Forecasting an AR(p)

For $\quad X_t = \sum_{i=1}^{p} \phi_i X_{t-i} + W_t,$

$$
\begin{aligned}
X_{n+1}^n &= P(X_{n+1}|X_1, \ldots, X_n) \\
&= P\left(\sum_{i=1}^{p} \phi_i X_{n+1-i} + W_{n+1}|X_1, \ldots, X_n\right) \\
&= \sum_{i=1}^{p} \phi_i P\left(X_{n+1-i}|X_1, \ldots, X_n\right) \\
&= \sum_{i=1}^{p} \phi_i X_{n+1-i} \qquad \text{for } n \geq p.
\end{aligned}
$$

Example: PACF of an AR(p)

$$\text{For} \quad X_t = \sum_{i=1}^{p} \phi_i X_{t-i} + W_t,$$

$$X_{n+1}^n = \sum_{i=1}^{p} \phi_i X_{n+1-i}.$$

$$\text{Thus, } \phi_{hh} = \begin{cases} \phi_h & \text{if } 1 \leq h \leq p \\ 0 & \text{otherwise.} \end{cases}$$

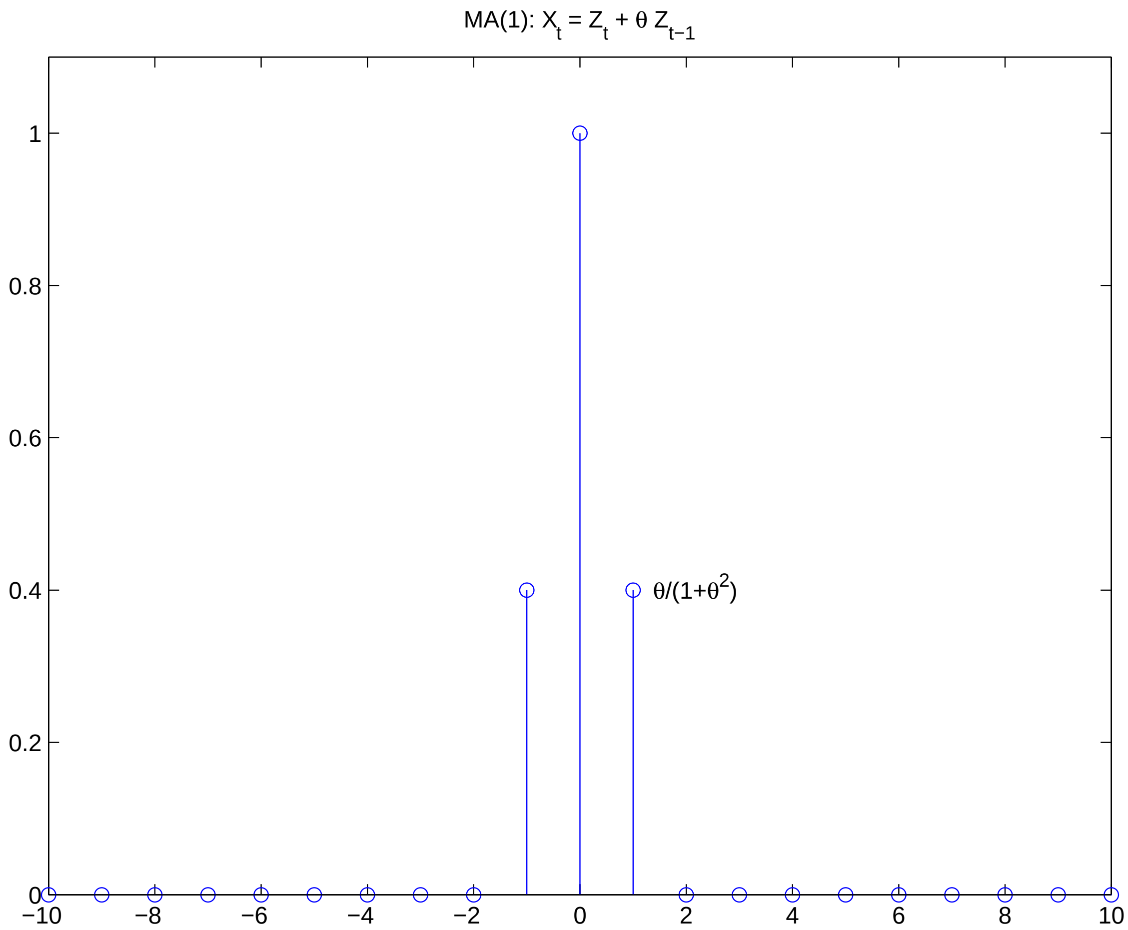
Example: PACF of an invertible MA(q)

For $\quad X_t = \sum_{i=1}^{q} \theta_i W_{t-i} + W_t, \qquad X_t = -\sum_{i=1}^{\infty} \pi_i X_{t-i} + W_t,$

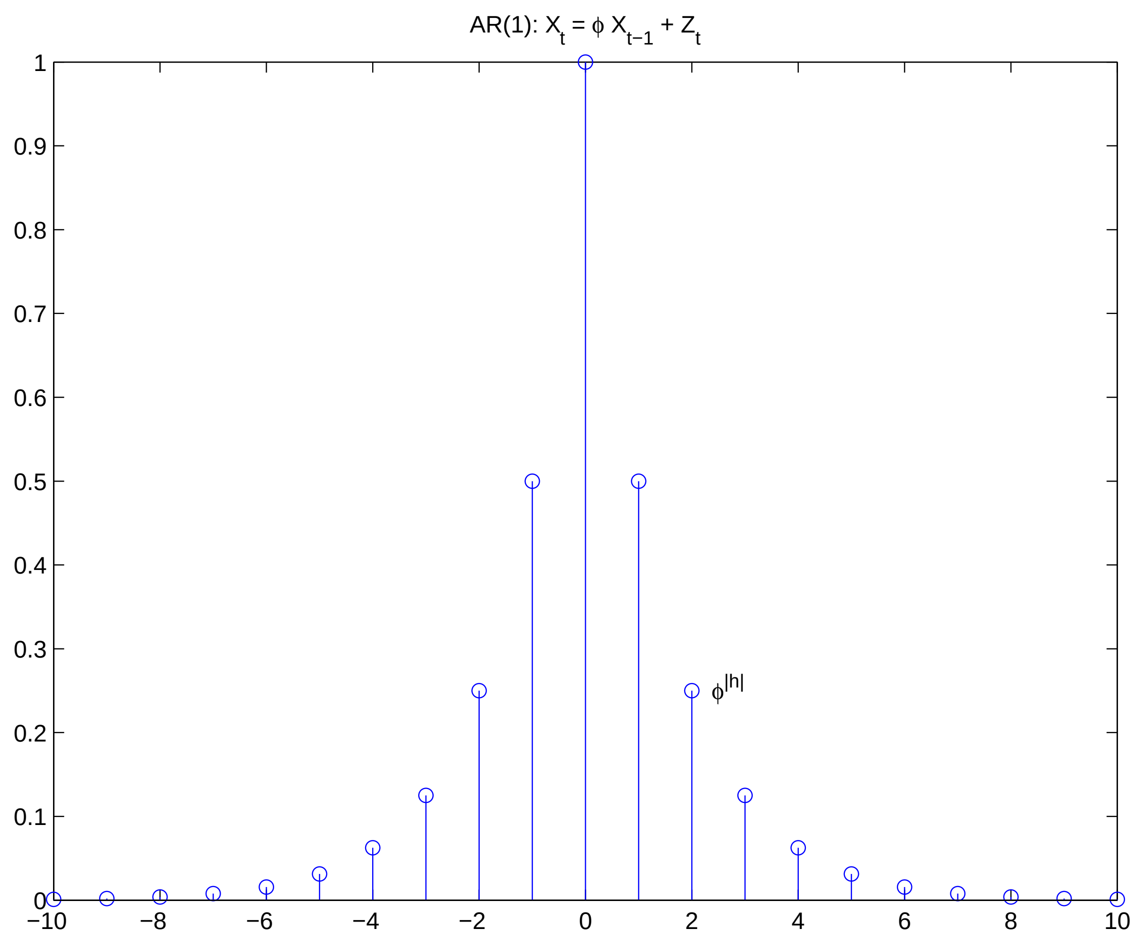
$$
\begin{aligned}
X_{n+1}^n &= P(X_{n+1} | X_1, \dots, X_n) \\
&= P\left(-\sum_{i=1}^{\infty} \pi_i X_{n+1-i} + W_{n+1} | X_1, \dots, X_n \right) \\
&= -\sum_{i=1}^{\infty} \pi_i P\left(X_{n+1-i} | X_1, \dots, X_n \right) \\
&= -\sum_{i=1}^{n} \pi_i X_{n+1-i} - \sum_{i=n+1}^{\infty} \pi_i P\left(X_{n+1-i} | X_1, \dots, X_n \right).
\end{aligned}
$$

In general, $\phi_{hh} \neq 0$.

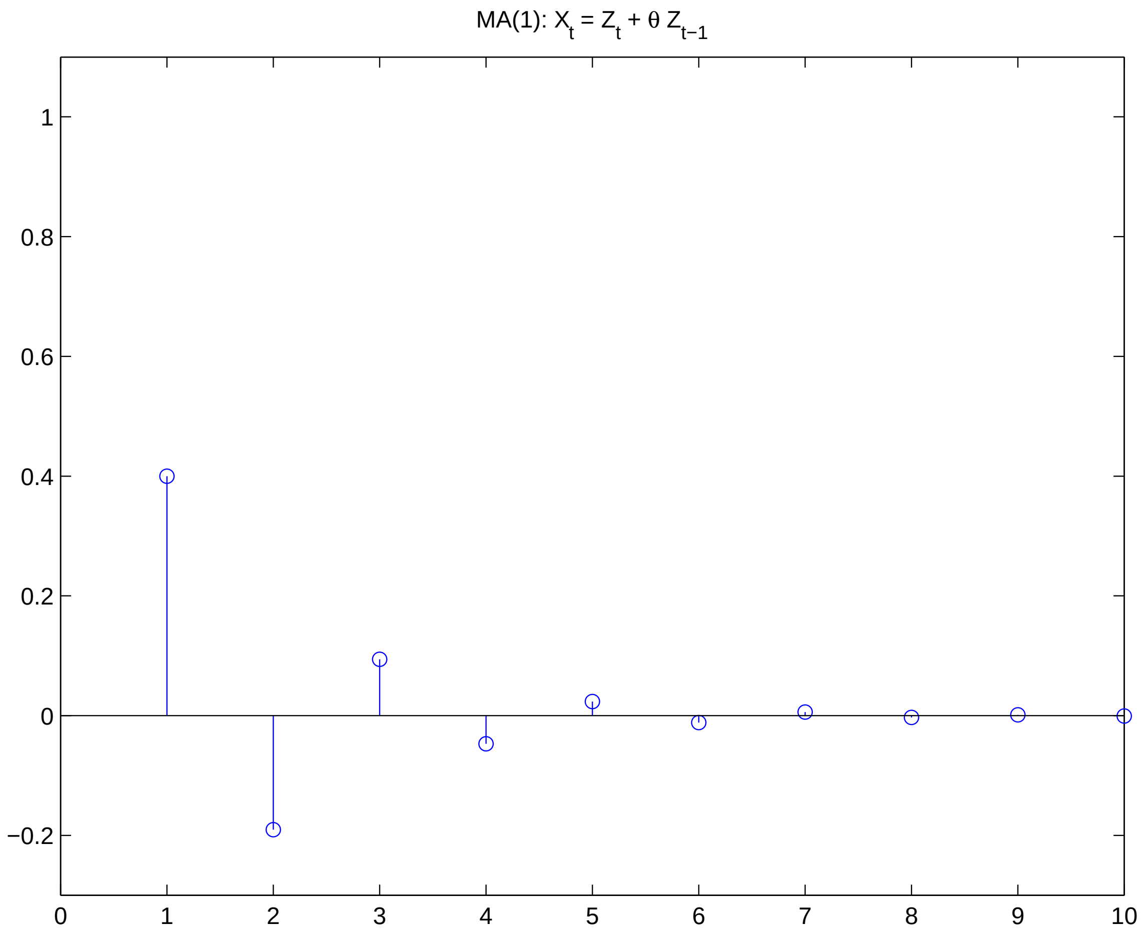
ACF of the MA(1) process



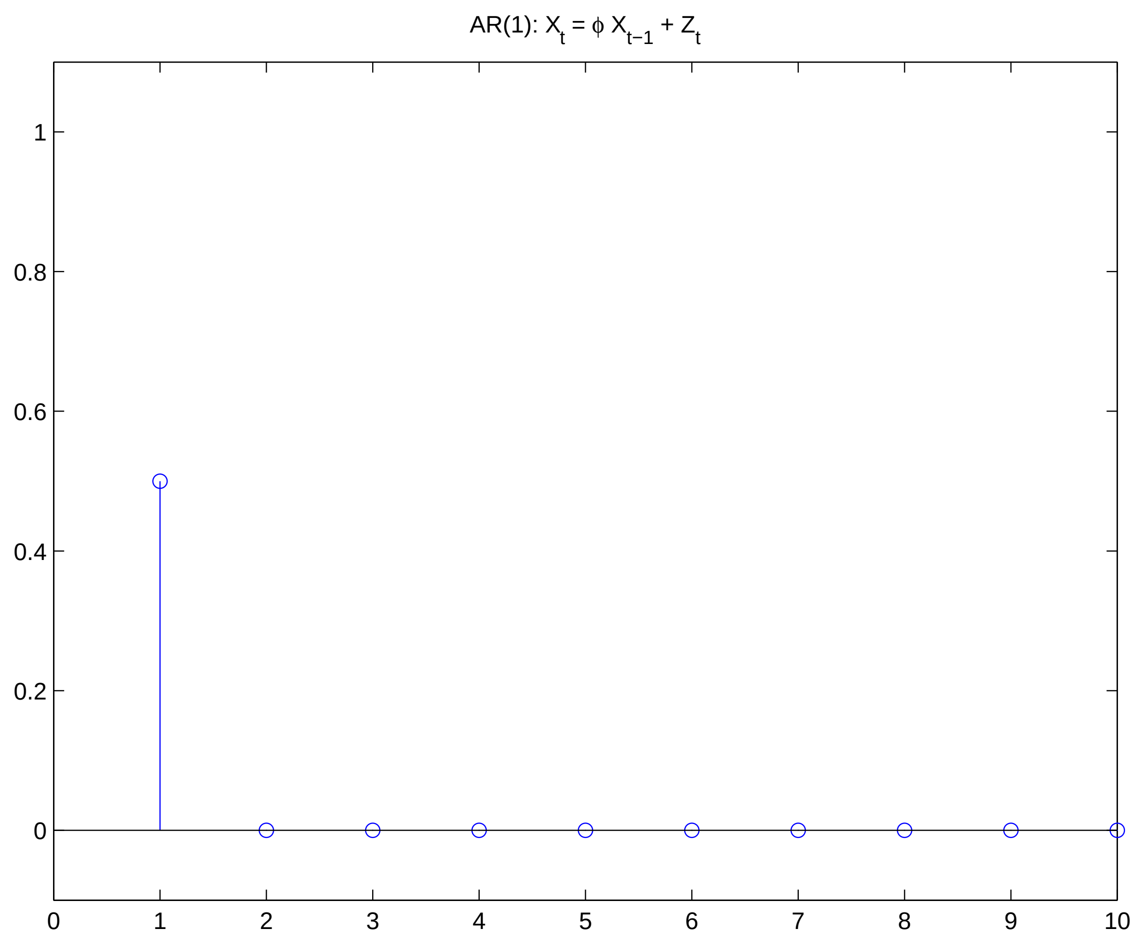
ACF of the AR(1) process



PACF of the MA(1) process



PACF of the AR(1) process



PACF and ACF

Model:	ACF:	PACF:
AR(p)	decays	zero for $h > p$
MA(q)	zero for $h > q$	decays
ARMA(p,q)	decays	decays

Sample PACF

For a realization $x_1, \ldots, x_n$ of a time series,
the **sample PACF** is defined by

$$\hat{\phi}_{00} = 1$$

$$\hat{\phi}_{hh} = \text{ last component of } \hat{\phi}_h,$$

where $\hat{\phi}_h = \hat{\Gamma}_h^{-1}\hat{\gamma}_h$.

Introduction to Time Series Analysis. Lecture 9.

1. Review: Linear prediction.
2. Prediction operator.
3. Partial autocorrelation function.
4. Recursive methods: Durbin-Levinson.

The importance of P_{n+1}^n: Prediction intervals

$$X_{n+1}^n = \phi_{n1} X_n + \phi_{n2} X_{n-1} + \cdots + \phi_{nn} X_1$$

$$\Gamma_n \phi_n = \gamma_n, \qquad P_{n+1}^n = \mathrm{E}\left(X_{n+1} - X_{n+1}^n\right)^2 = \gamma(0) - \gamma_n' \Gamma_n^{-1} \gamma_n.$$

After seeing $X_1, \ldots, X_n$, we forecast X_{n+1}^n. The expected squared error of our forecast is P_{n+1}^n. We can construct a prediction interval:

$$X_{n+1}^n \pm c_{\alpha/2} \sqrt{P_{n+1}^n}.$$

For a Gaussian process, the prediction error has distribution $\mathcal{N}(0, P_{n+1}^n)$, so $c_{0.05/2} = 1.96$ gives a 95% prediction interval. For any process with finite second moments, we can apply Chebyshev's inequality:

$$\Pr\left(|X - \mathrm{E}X| \geq t\sqrt{\mathrm{Var}(X)}\right) \leq \frac{1}{t^2}.$$

Computing linear prediction coefficients

$$X_{n+1}^n = \phi_{n1} X_n + \phi_{n2} X_{n-1} + \cdots + \phi_{nn} X_1$$

$$\Gamma_n \phi_n = \gamma_n,$$

$$P_{n+1}^n = \mathrm{E}\left(X_{n+1} - X_{n+1}^n\right)^2 = \gamma(0) - \gamma_n' \Gamma_n^{-1} \gamma_n.$$

How can we compute these quantities recursively?
i.e., given the coefficients ϕ_{n-1} of X_n^{n-1}, how can we compute the coefficients ϕ_n of X_{n+1}^n, without solving another linear system $\Gamma_n \phi_n = \gamma_n$?

Durbin-Levinson

$$\phi_0 = 0, \qquad \phi_{00} = 0;$$

$$\phi_1 = \phi_{11}, \qquad \phi_{11} = \frac{\gamma(1)}{\gamma(0)};$$

$$\phi_n = \begin{pmatrix} \phi_{n-1} - \phi_{nn}\tilde{\phi}_{n-1} \\ \phi_{nn} \end{pmatrix}, \quad \phi_{nn} = \frac{\gamma(n) - \phi'_{n-1}\tilde{\gamma}_{n-1}}{\gamma(0) - \phi'_{n-1}\gamma_{n-1}}.$$

$$\phi_n = (\phi_{n1}, \ldots, \phi_{nn})' \qquad \tilde{\phi}_n = (\phi_{nn}, \ldots, \phi_{n1})',$$

$$\gamma_n = (\gamma(1), \ldots, \gamma(n))' \qquad \tilde{\gamma}_n = (\gamma(n), \ldots, \gamma(1))'.$$

Durbin-Levinson: Example

$$\phi_0 = 0, \qquad \phi_{00} = 0;$$

$$\phi_1 = \phi_{11}, \qquad \phi_{11} = \frac{\gamma(1)}{\gamma(0)};$$

$$\phi_n = \begin{pmatrix} \phi_{n-1} - \phi_{nn}\tilde{\phi}_{n-1} \\ \phi_{nn} \end{pmatrix}, \quad \phi_{nn} = \frac{\gamma(n) - \phi'_{n-1}\tilde{\gamma}_{n-1}}{\gamma(0) - \phi'_{n-1}\gamma_{n-1}}.$$

This algorithm computes $\phi_1, \phi_2, \phi_3, \ldots$, where

$$X_2^1 = X_1\phi_1, \quad X_3^2 = (X_2, X_1)\phi_2, \quad X_4^3 = (X_3, X_2, X_1)\phi_3, \ldots$$

Durbin-Levinson: Example

$$\phi_1 = \phi_{11}, \qquad \phi_{11} = \frac{\gamma(1)}{\gamma(0)};$$

$$\phi_n = \begin{pmatrix} \phi_{n-1} - \phi_{nn}\tilde{\phi}_{n-1} \\ \phi_{nn} \end{pmatrix}, \quad \phi_{nn} = \frac{\gamma(n) - \phi'_{n-1}\tilde{\gamma}_{n-1}}{\gamma(0) - \phi'_{n-1}\gamma_{n-1}}.$$

$$\phi_1 = \gamma(1)/\gamma(0),$$

$$\phi_2 = \begin{pmatrix} \phi_1 - \phi_{22}\phi_{11} \\ \phi_{22} \end{pmatrix} = \begin{pmatrix} \frac{\gamma(1)}{\gamma(0)}\left(1 - \frac{\gamma(2)-\gamma(1)}{\gamma(0)-\gamma(1)}\right) \\ \frac{\gamma(2)-\gamma(1)}{\gamma(0)-\gamma(1)} \end{pmatrix}, \text{ etc.}$$

Durbin-Levinson: Why it works (Details)

Clearly, $\Gamma_1 \phi_1 = \gamma_1$.

Suppose $\Gamma_{n-1}\phi_{n-1} = \gamma_{n-1}$. Then $\Gamma_{n-1}\tilde{\phi}_{n-1} = \tilde{\gamma}_{n-1}$, and so

$$\begin{aligned}
\Gamma_n \phi_n &= \begin{pmatrix} \Gamma_{n-1} & \tilde{\gamma}_{n-1} \\ \tilde{\gamma}'_{n-1} & \gamma(0) \end{pmatrix} \begin{pmatrix} \phi_{n-1} - \phi_{nn}\tilde{\phi}_{n-1} \\ \phi_{nn} \end{pmatrix} \\
&= \begin{pmatrix} \gamma_{n-1} \\ \tilde{\gamma}'_{n-1}\phi_{n-1} + \phi_{nn}\left(\gamma(0) - \gamma'_{n-1}\phi_{n-1}\right) \end{pmatrix} \\
&= \gamma_n.
\end{aligned}$$

Durbin-Levinson: Evolution of mean square error

$$
\begin{aligned}
P_{n+1}^{n} &= \gamma(0) - \phi_n' \gamma_n \\
&= \gamma(0) - \begin{pmatrix} \phi_{n-1} - \phi_{nn}\tilde{\phi}_{n-1} \\ \phi_{nn} \end{pmatrix}' \begin{pmatrix} \gamma_{n-1} \\ \gamma(n) \end{pmatrix} \\
&= P_n^{n-1} - \phi_{nn}\left(\gamma(n) - \tilde{\phi}_{n-1}' \gamma_{n-1}\right) \\
&= P_n^{n-1} - \phi_{nn}^2\left(\gamma(0) - \phi_{n-1}' \gamma_{n-1}\right) \qquad \text{(From expression for } \phi_{nn}\text{)} \\
&= P_n^{n-1}\left(1 - \phi_{nn}^2\right).
\end{aligned}
$$

i.e., variance reduces by a factor $1 - \phi_{nn}^2$.

Introduction to Time Series Analysis. Lecture 9.

1. Review: Linear prediction.
2. Prediction operator.
3. Partial autocorrelation function.
4. Recursive methods: Durbin-Levinson.