Review: least squares linear prediction

Consider a **linear predictor** of X_{n+h} given $X_n = x_n$:

$$f(x_n) = \alpha_0 + \alpha_1 x_n.$$

For a stationary time series $\{X_t\}$, the best linear predictor is $f^*(x_n) = (1 - \rho(h))\mu + \rho(h)x_n$:

$$\begin{aligned} \mathrm{E}\left(X_{n+h} - (\alpha_0 + \alpha_1 X_n)\right)^2 &\geq \mathrm{E}\left(X_{n+h} - f^*(X_n)\right)^2 \\ &= \sigma^2(1 - \rho(h)^2). \end{aligned}$$

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**

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

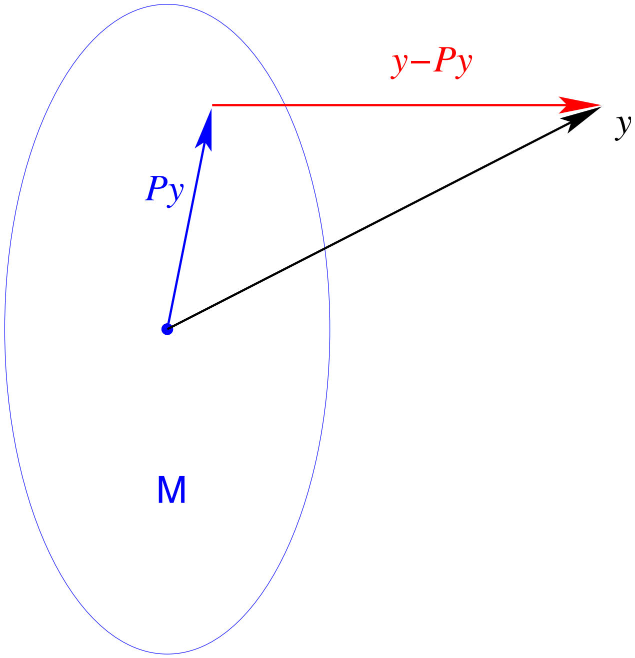
$$\mathrm{E}\left[\left(X_{n+m} - X_{n+m}^n\right) X_i\right] = 0 \qquad \text{for } i = 1, \ldots, n.$$

This is a special case of the *projection theorem*.

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 $w \in \mathcal{M}$.



Projection theorem: Linear prediction

Let X^n_{n+m} denote the best linear predictor:

$$\|X^n_{n+m} - X_{n+m}\|^2 \le \|Z - X_{n+m}\|^2 \quad \text{for all } Z \in \mathcal{M}.$$

The projection theorem implies the orthogonality

$$\langle X^n_{n+m} - X_{n+m}, Z\rangle = 0 \quad \text{for all } Z \in \mathcal{M}$$

$$\Leftrightarrow \quad \langle X^n_{n+m} - X_{n+m}, Z\rangle = 0 \quad \text{for all } Z \in \{1, X_1, \ldots, X_n\}$$

$$\Leftrightarrow \quad \begin{aligned} \mathrm{E}\left(X^n_{n+m} - X_{n+m}\right) &= 0 \\ \mathrm{E}\left[\left(X^n_{n+m} - X_{n+m}\right) X_i\right] &= 0 \end{aligned}$$

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

Linear prediction

Error $(X_{n+m}^n - X_{n+m})$ is uncorrelated with the prediction variable 1:

$$
\begin{aligned}
& \mathrm{E}\left(X_{n+m}^n - X_{n+m}\right) = 0 \\
\Leftrightarrow \quad & \mathrm{E}\left(\alpha_0 + \sum_i \alpha_i X_i - X_{n+m}\right) = 0 \\
\Leftrightarrow \quad & \mu\left(1 - \sum_i \alpha_i\right) = \alpha_0 .
\end{aligned}
$$

So $\quad X_{n+m}^n = \alpha_0 + \sum_i \alpha_i X_i \quad \Leftrightarrow \quad X_{n+m}^n - \mu = \sum_i \alpha_i \left(X_i - \mu\right).$

Thus, for forecasting, we can assume $\mu = 0$. So we'll ignore α_0.

One-step-ahead linear prediction

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

Prediction equations: $\mathrm{E}\left((X_{n+1}^n - X_{n+1}) X_i\right) = 0, \text{ for } i = 1, \ldots, n$

$$\Leftrightarrow \quad \sum_{j=1}^{n} \phi_{nj} \mathrm{E}\left(X_{n+1-j} X_i\right) = \mathrm{E}(X_{n+1} X_i)$$

$$\Leftrightarrow \quad \sum_{j=1}^{n} \phi_{nj} \gamma(i-j) = \gamma(i)$$

$$\Leftrightarrow \quad \Gamma_n \phi_n = \gamma_n,$$

One-step-ahead linear prediction

Prediction equations: $\Gamma_n \phi_n = \gamma_n$.

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

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

Mean squared error of one-step-ahead linear prediction

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

where $X = (X_n, X_{n-1}, \ldots, X_1)'$.

Backcasting: Predicting m steps in the past

Given $X_1, \ldots, X_n$, we wish to predict X_{1-m} for $m > 0$.
That is, we choose $Z \in \mathcal{M} = \bar{\text{sp}}\{X_1, \ldots, X_n\}$ to minimize $\|Z - X_{1-m}\|^2$.

The prediction equations are

$$\langle X_{1-m}^n - X_{1-m}, Z \rangle = 0 \quad \text{for all } Z \in \mathcal{M}$$

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

One-step backcasting

Write the least squares prediction of X_0 given $X_1, \ldots, X_n$ as

$$X_0^n = \phi_{n1} X_1 + \phi_{n2} X_2 + \cdots + \phi_{nn} X_n = \phi_n' X,$$

where the predictor vector is reversed: now $X = (X_1, \ldots, X_n)'$.
The prediction equations are

$$\begin{aligned}
& \mathrm{E}\left(\left(X_0^n - X_0\right) X_i\right) = 0 \quad \text{for } i = 1, \ldots, n \\
\Leftrightarrow \quad & \mathrm{E}\left(\left(\sum_{j=1}^n \phi_{nj} X_j - X_0\right) X_i\right) = 0 \\
\Leftrightarrow \quad & \sum_{j=1}^n \phi_{nj} \gamma(j-i) = \gamma(i) \\
\Leftrightarrow \quad & \Gamma_n \phi_n = \gamma_n.
\end{aligned}$$

One-step backcasting

The prediction equations are

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

which is exactly the same as for forecasting, but with the indices of the predictor vector reversed: $X = (X_1, \ldots, X_n)'$ versus $X = (X_n, \ldots, X_1)'$.

Example: Forecasting AR(1)

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

linear prediction of X_2: $$X_2^1 = \phi_{11} X_1$$

Prediction equation:
$$\begin{aligned} \gamma(0)\phi_{11} &= \gamma(1) \\ &= \text{Cov}(X_0, X_1) \\ &= \phi_1 \gamma(0) \end{aligned}$$

$\Leftrightarrow$ $$\phi_{11} = \phi_1.$$

Example: Backcasting AR(1)

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

linear prediction of X_0: $$X_0^1 = \phi_{11} X_1$$

Prediction equation:
$$\begin{aligned}\gamma(0)\phi_{11} &= \gamma(1) \\ &= \text{Cov}(X_0, X_1) \\ &= \phi_1 \gamma(0)\end{aligned}$$

$\Leftrightarrow$ $$\phi_{11} = \phi_1.$$

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

$$\begin{aligned}
\text{Write} \qquad X_{n+m}^n &= \phi_{n1}^{(m)} X_n + \phi_{n2}^{(m)} X_{n-1} + \cdots + \phi_{nn}^{(m)} X_1 \\
\Gamma_n \phi_n^{(m)} &= \gamma_n^{(m)}, \\
\text{with} \qquad \Gamma_n &= \mathrm{Cov}(X, X), \\
\gamma_n^{(m)} &= \mathrm{Cov}(X_{n+m}, X) \\
&= (\gamma(m), \gamma(m+1), \ldots, \gamma(m+n-1))'. \\
\text{Also,} \quad \mathrm{E}((X_{n+m} - X_{n+m}^n)^2) &= \gamma(0) - \phi^{(m)'} \gamma_n^{(m)}.
\end{aligned}$$

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 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)

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

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

Example: PACF of an AR(p)

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

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

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

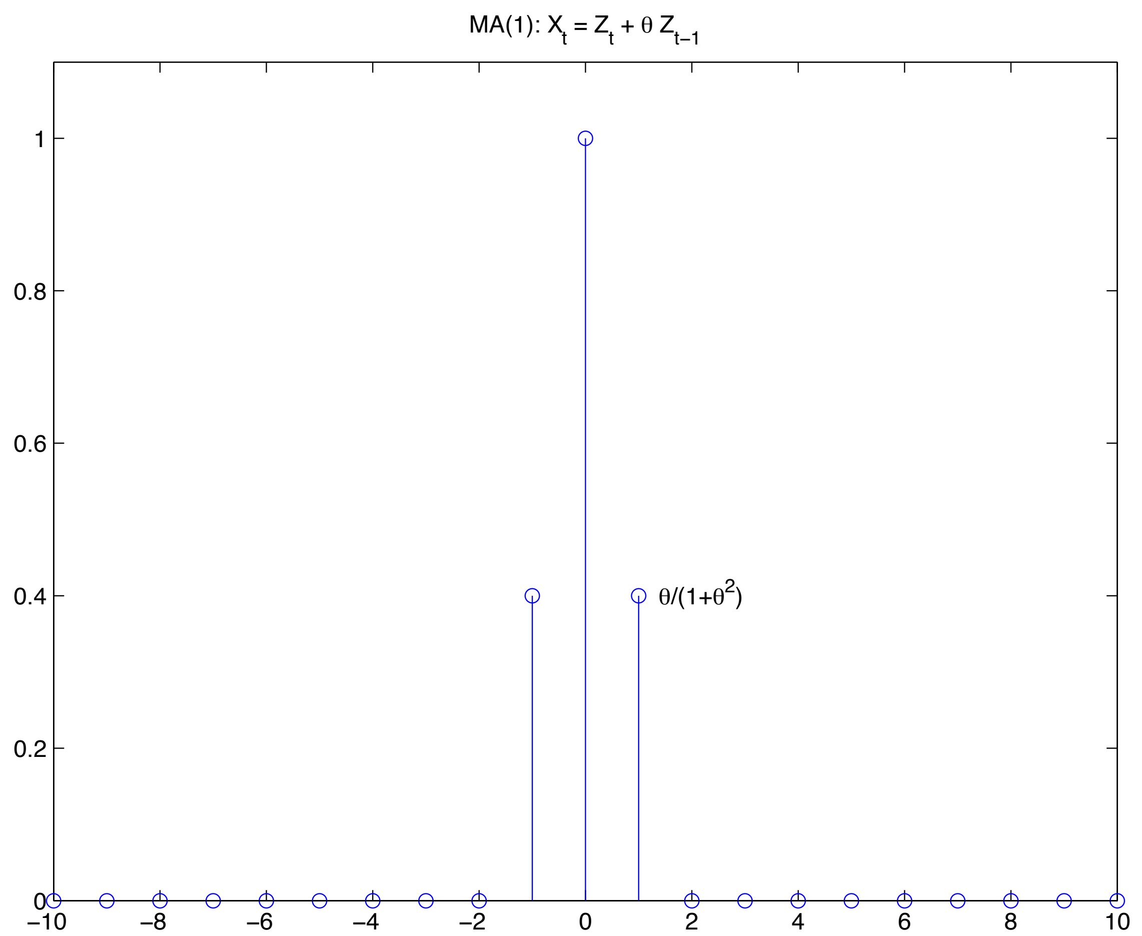
Example: PACF of an invertible MA(q)

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

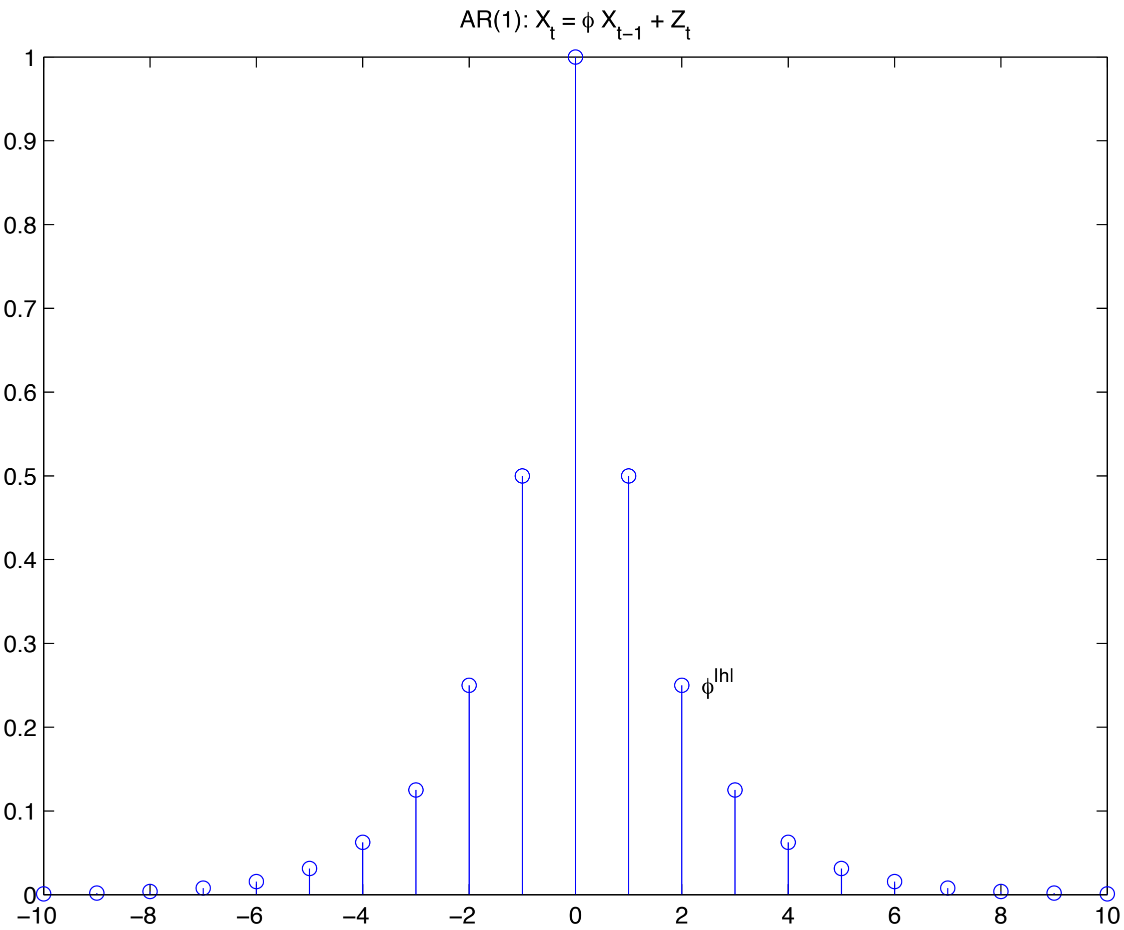
$$\begin{aligned} X_{n+1}^{n} &= P(X_{n+1} | X_1, \ldots, X_n) \\ &= P\left(\sum_{i=1}^{\infty} \pi_i X_{n+1-i} + W_t | X_1, \ldots, X_n \right) \\ &= \sum_{i=1}^{\infty} \pi_i P\left(X_{n+1-i} | X_1, \ldots, X_n \right) \\ &= \sum_{i=1}^{n} \pi_i X_{n+1-i} + \sum_{i=n+1}^{\infty} \pi_i P\left(X_{n+1-i}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

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

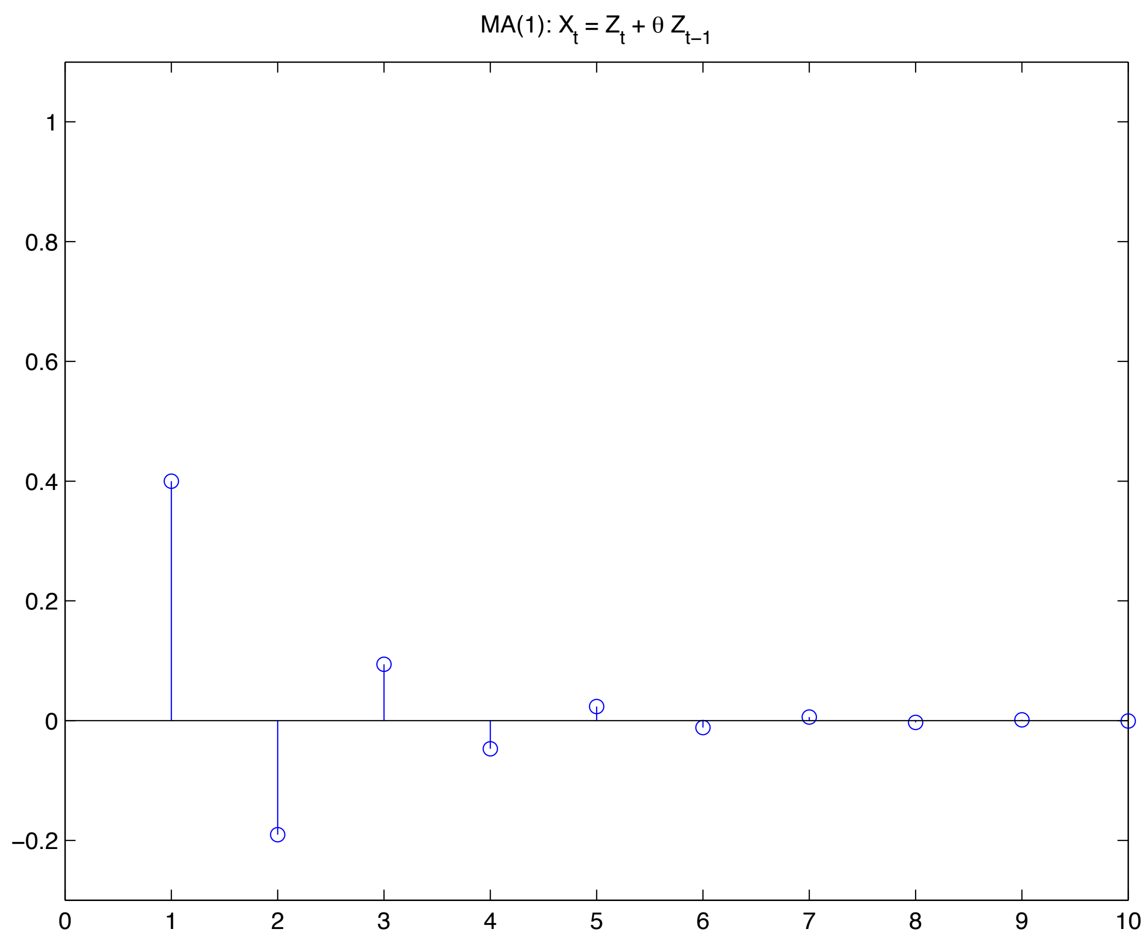
ACF of the MA(1) process



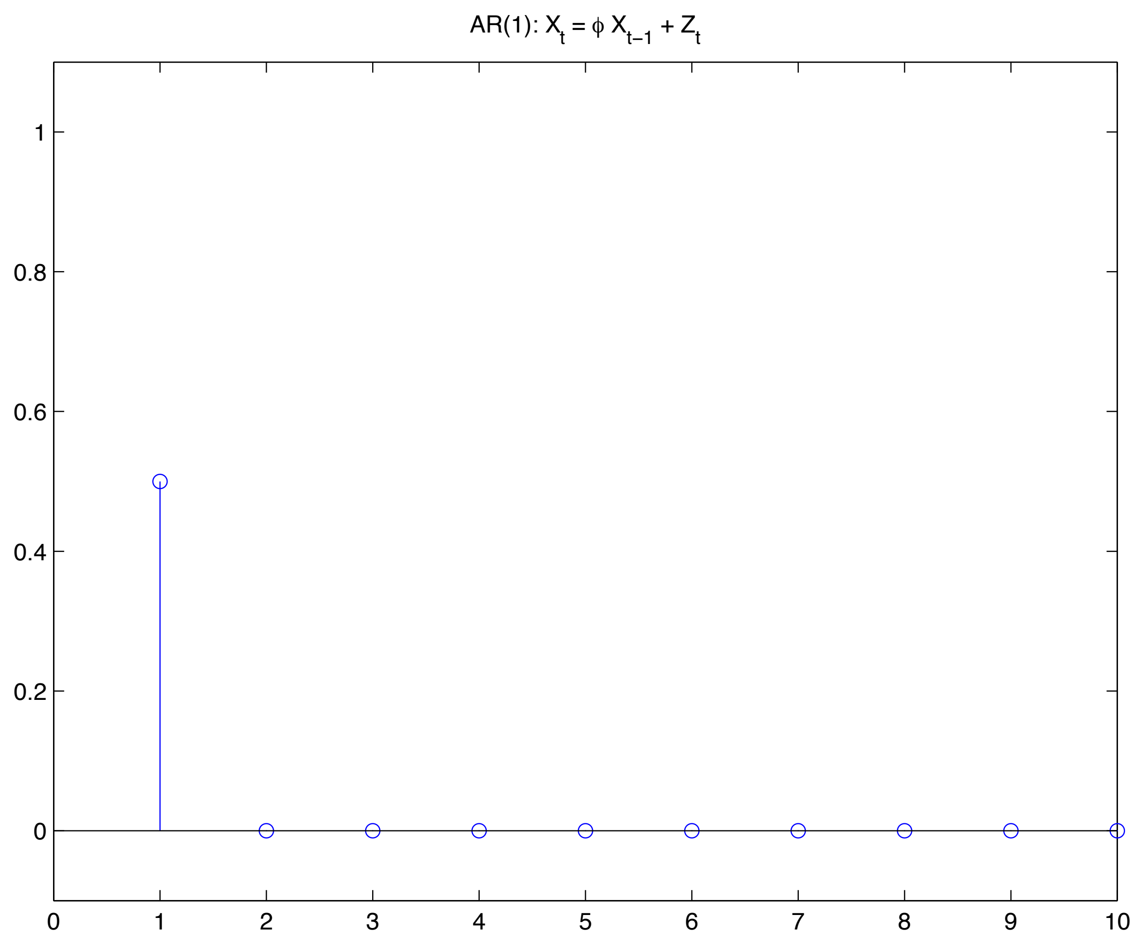
ACF of the AR(1) process



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

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 $\Gamma_n \phi_n = \gamma_n$?