The innovations algorithm. Basis

Another recursive algorithm ('innovations algorithm') works better in some cases. It will be important in the estimation of ARMA processes.

The innovations algorithm. Basis

Another recursive algorithm ('innovations algorithm') works better in some cases. It will be important in the estimation of ARMA processes.
Let $\hat{X}_{n+1} = \mathbf{P}_{\mathcal{L}(X_1,\dots,X_n)} X_{n+1} \in \mathcal{L}(X_1, \dots, X_n)$. We wish to write

$$\hat{X}_{n+1} = \sum_{j=1}^{n} \vartheta_{n,j}(X_{n+1-j} - \hat{X}_{n+1-j}).$$

The innovations algorithm. Basis

Another recursive algorithm ('innovations algorithm') works better in some cases. It will be important in the estimation of ARMA processes.
Let $\hat{X}_{n+1} = \mathbf{P}_{\mathcal{L}(X_1,\ldots,X_n)} X_{n+1} \in \mathcal{L}(X_1, \ldots, X_n)$. We wish to write

$$\hat{X}_{n+1} = \sum_{j=1}^{n} \vartheta_{n,j}(X_{n+1-j} - \hat{X}_{n+1-j}).$$

$\{X_{n+1-j} - \hat{X}_{n+1-j}\}_{j=1\ldots n}$ is an orthogonal basis of $\mathcal{L}(X_1, \ldots, X_n)$.
In fact $X_{k+1} - \hat{X}_{k+1}$ by definition is orthogonal to $\mathcal{L}(X_1, \ldots, X_k)$,
hence to $X_j - \hat{X}_j$ for all $j = 1 \ldots k$.

($X_{k+1} - \hat{X}_{k+1}$ is named *innovation*, as it could not be predicted before)

The innovations algorithm. Steps

The orthogonality condition reads: for $j = 1 \dots n$

$$\begin{aligned}\langle X_{n+1}, X_{n+1-j} - \hat{X}_{n+1-j} \rangle &= \langle \hat{X}_{n+1}, X_{n+1-j} - \hat{X}_{n+1-j} \rangle \\ &= \vartheta_{n,j} \| X_{n+1-j} - \hat{X}_{n+1-j} \|^2 = \vartheta_{n,j} v_{n-j}. \end{aligned} \tag{1}$$

The innovations algorithm. Steps

The orthogonality condition reads: for $j = 1 \dots n$

$$\begin{aligned} \langle X_{n+1}, X_{n+1-j} - \hat{X}_{n+1-j} \rangle &= \langle \hat{X}_{n+1}, X_{n+1-j} - \hat{X}_{n+1-j} \rangle \\ &= \vartheta_{n,j} \| X_{n+1-j} - \hat{X}_{n+1-j} \|^2 = \vartheta_{n,j} v_{n-j}. \end{aligned} \tag{1}$$

Take $j = n$. Then

$$\vartheta_{n,n} v_0 = \langle X_{n+1}, X_1 - \hat{X}_1 \rangle = \langle X_{n+1}, X_1 \rangle = \gamma(n).$$

The innovations algorithm. Steps

The orthogonality condition reads: for $j = 1 \dots n$

$$\begin{aligned}\langle X_{n+1}, X_{n+1-j} - \hat{X}_{n+1-j}\rangle &= \langle \hat{X}_{n+1}, X_{n+1-j} - \hat{X}_{n+1-j}\rangle \\ &= \vartheta_{n,j} \|X_{n+1-j} - \hat{X}_{n+1-j}\|^2 = \vartheta_{n,j} v_{n-j}.\end{aligned} \tag{1}$$

Take $j = n$. Then

$$\vartheta_{n,n} v_0 = \langle X_{n+1}, X_1 - \hat{X}_1\rangle = \langle X_{n+1}, X_1\rangle = \gamma(n).$$

For $j < n$, $\quad \vartheta_{n,j} v_{n-j} = \langle X_{n+1}, X_{n+1-j} - \hat{X}_{n+1-j}\rangle$

$$= \gamma(j) - \sum_{k=1}^{n-j} \vartheta_{n-j,k} \langle X_{n+1}, X_{n+1-j-k} - \hat{X}_{n+1-j-k}\rangle.$$

The innovations algorithm. Steps

The orthogonality condition reads: for $j = 1 \dots n$

$$\begin{aligned}\langle X_{n+1}, X_{n+1-j} - \hat{X}_{n+1-j} \rangle &= \langle \hat{X}_{n+1}, X_{n+1-j} - \hat{X}_{n+1-j} \rangle \\ &= \vartheta_{n,j} \| X_{n+1-j} - \hat{X}_{n+1-j} \|^2 = \vartheta_{n,j} v_{n-j}.\end{aligned} \tag{1}$$

Take $j = n$. Then

$$\vartheta_{n,n} v_0 = \langle X_{n+1}, X_1 - \hat{X}_1 \rangle = \langle X_{n+1}, X_1 \rangle = \gamma(n).$$

For $j < n$, $\quad \vartheta_{n,j} v_{n-j} = \langle X_{n+1}, X_{n+1-j} - \hat{X}_{n+1-j} \rangle$

$$= \gamma(j) - \sum_{k=1}^{n-j} \vartheta_{n-j,k} \langle X_{n+1}, X_{n+1-j-k} - \hat{X}_{n+1-j-k} \rangle.$$

Now insert (1) in the rightmost term.

The innovations algorithm. Steps (cont.)

$$\langle X_{n+1}, X_{n+1-j} - \hat{X}_{n+1-j} \rangle = \vartheta_{n,j} v_{n-j}. \qquad (1)$$

Hence $\vartheta_{n,j} v_{n-j} = \gamma(j) - \sum_{k=1}^{n-j} \vartheta_{n-j,k} \langle X_{n+1}, X_{n+1-j-k} - \hat{X}_{n+1-j-k} \rangle$

$$= \gamma(j) - \sum_{k=1}^{n-j} \vartheta_{n-j,k} \vartheta_{n,j+k} v_{n-j-k}.$$

The innovations algorithm. Steps (cont.)

$$\langle X_{n+1}, X_{n+1-j} - \hat{X}_{n+1-j} \rangle = \vartheta_{n,j} v_{n-j}. \qquad (1)$$

Hence $\vartheta_{n,j} v_{n-j} = \gamma(j) - \sum_{k=1}^{n-j} \vartheta_{n-j,k} \langle X_{n+1}, X_{n+1-j-k} - \hat{X}_{n+1-j-k} \rangle$

$$= \gamma(j) - \sum_{k=1}^{n-j} \vartheta_{n-j,k} \vartheta_{n,j+k} v_{n-j-k}.$$

In order to compute $\vartheta_{n,j}$ we need $\vartheta_{n-j,k}$ (as $j \geq 1$ this value has already been obtained) and $\vartheta_{n,j+k}$, i.e. $\vartheta_{n,l}$ with $l > j$. At step n, one can then compute $\vartheta_{n,n}$ (first formula), then $\vartheta_{n,n-1}$ down to $\vartheta_{n,1}$.

The innovations algorithm. Steps (cont.)

$$\langle X_{n+1}, X_{n+1-j} - \hat{X}_{n+1-j} \rangle = \vartheta_{n,j} v_{n-j}. \qquad (1)$$

Hence $$\vartheta_{n,j} v_{n-j} = \gamma(j) - \sum_{k=1}^{n-j} \vartheta_{n-j,k} \langle X_{n+1}, X_{n+1-j-k} - \hat{X}_{n+1-j-k} \rangle$$
$$= \gamma(j) - \sum_{k=1}^{n-j} \vartheta_{n-j,k} \vartheta_{n,j+k} v_{n-j-k}.$$

In order to compute $\vartheta_{n,j}$ we need $\vartheta_{n-j,k}$ (as $j \geq 1$ this value has already been obtained) and $\vartheta_{n,j+k}$, i.e. $\vartheta_{n,l}$ with $l > j$. At step n, one can then compute $\vartheta_{n,n}$ (first formula), then $\vartheta_{n,n-1}$ down to $\vartheta_{n,1}$.

One needs still a recursive formula for v_n.

The innovations algorithm. Summary

$$\begin{aligned} v_n = \|X_{n+1} - \hat{X}_{n+1}\|^2 &= \|X_{n+1}\|^2 + \|\hat{X}_{n+1}\|^2 - 2\langle X_{n+1}, \hat{X}_{n+1}\rangle \\ &= \|X_{n+1}\|^2 + \|\hat{X}_{n+1}\|^2 - 2\langle X_{n+1} - \hat{X}_{n+1}, \hat{X}_{n+1}\rangle - 2\langle \hat{X}_{n+1}, \hat{X}_{n+1}\rangle \\ &= \|X_{n+1}\|^2 - \|\hat{X}_{n+1}\|^2 \end{aligned}$$

as $X_{n+1} - \hat{X}_{n+1}$ is orthogonal to $\mathcal{L}(X_1, \ldots, X_k)$, hence to $\hat{X}_{n+1}$.

The innovations algorithm. Summary

$$
\begin{aligned}
v_n = \|X_{n+1} - \hat{X}_{n+1}\|^2 = \|X_{n+1}\|^2 + \|\hat{X}_{n+1}\|^2 - 2\langle X_{n+1}, \hat{X}_{n+1}\rangle \\
= \|X_{n+1}\|^2 + \|\hat{X}_{n+1}\|^2 - 2\langle X_{n+1} - \hat{X}_{n+1}, \hat{X}_{n+1}\rangle - 2\langle \hat{X}_{n+1}, \hat{X}_{n+1}\rangle \\
= \|X_{n+1}\|^2 - \|\hat{X}_{n+1}\|^2
\end{aligned}
$$

as $X_{n+1} - \hat{X}_{n+1}$ is orthogonal to $\mathcal{L}(X_1, \dots, X_k)$, hence to $\hat{X}_{n+1}$.

$\|X_{n+1}\|^2 = \gamma(0)$, $\|\hat{X}_{n+1}\|^2 = \sum_{j=1}^{n} \vartheta_{n,j}^2 v_{n-j}$.

The innovations algorithm. Summary

$$
\begin{aligned}
v_n = \|X_{n+1} - \hat{X}_{n+1}\|^2 &= \|X_{n+1}\|^2 + \|\hat{X}_{n+1}\|^2 - 2\langle X_{n+1}, \hat{X}_{n+1}\rangle \\
&= \|X_{n+1}\|^2 + \|\hat{X}_{n+1}\|^2 - 2\langle X_{n+1} - \hat{X}_{n+1}, \hat{X}_{n+1}\rangle - 2\langle \hat{X}_{n+1}, \hat{X}_{n+1}\rangle \\
&= \|X_{n+1}\|^2 - \|\hat{X}_{n+1}\|^2
\end{aligned}
$$

as $X_{n+1} - \hat{X}_{n+1}$ is orthogonal to $\mathcal{L}(X_1, \dots, X_k)$, hence to $\hat{X}_{n+1}$.

$\|X_{n+1}\|^2 = \gamma(0)$, $\|\hat{X}_{n+1}\|^2 = \sum_{j=1}^{n} \vartheta_{n,j}^2 v_{n-j}$.

The algorithm starts with $v_0 = \gamma(0)$.

Then for each n, $\quad \vartheta_{n,n} = \gamma(n)/v_0$,

The innovations algorithm. Summary

$$v_n = \|X_{n+1} - \hat{X}_{n+1}\|^2 = \|X_{n+1}\|^2 + \|\hat{X}_{n+1}\|^2 - 2\langle X_{n+1}, \hat{X}_{n+1}\rangle$$
$$= \|X_{n+1}\|^2 + \|\hat{X}_{n+1}\|^2 - 2\langle X_{n+1} - \hat{X}_{n+1}, \hat{X}_{n+1}\rangle - 2\langle \hat{X}_{n+1}, \hat{X}_{n+1}\rangle$$
$$= \|X_{n+1}\|^2 - \|\hat{X}_{n+1}\|^2$$

as $X_{n+1} - \hat{X}_{n+1}$ is orthogonal to $\mathcal{L}(X_1, \ldots, X_k)$, hence to $\hat{X}_{n+1}$.

$\|X_{n+1}\|^2 = \gamma(0)$, $\|\hat{X}_{n+1}\|^2 = \sum_{j=1}^{n} \vartheta_{n,j}^2 v_{n-j}$.

The algorithm starts with $v_0 = \gamma(0)$.

Then for each n, $\quad \vartheta_{n,n} = \gamma(n)/v_0$,

$$\vartheta_{n,j} = [\gamma(j) - \sum_{k=1}^{n-j} \vartheta_{n-j,k}\vartheta_{n,j+k} v_{n-j-k}]/v_{n-j}, \quad j = n-1, \ldots, 1.$$

The innovations algorithm. Summary

$$v_n = \|X_{n+1} - \hat{X}_{n+1}\|^2 = \|X_{n+1}\|^2 + \|\hat{X}_{n+1}\|^2 - 2\langle X_{n+1}, \hat{X}_{n+1}\rangle$$
$$= \|X_{n+1}\|^2 + \|\hat{X}_{n+1}\|^2 - 2\langle X_{n+1} - \hat{X}_{n+1}, \hat{X}_{n+1}\rangle - 2\langle \hat{X}_{n+1}, \hat{X}_{n+1}\rangle$$
$$= \|X_{n+1}\|^2 - \|\hat{X}_{n+1}\|^2$$

as $X_{n+1} - \hat{X}_{n+1}$ is orthogonal to $\mathcal{L}(X_1, \dots, X_k)$, hence to $\hat{X}_{n+1}$.
$\|X_{n+1}\|^2 = \gamma(0)$, $\|\hat{X}_{n+1}\|^2 = \sum_{j=1}^{n} \vartheta_{n,j}^2 v_{n-j}$.

The algorithm starts with $v_0 = \gamma(0)$.
Then for each n, $\quad \vartheta_{n,n} = \gamma(n)/v_0$,

$$\vartheta_{n,j} = [\gamma(j) - \sum_{k=1}^{n-j} \vartheta_{n-j,k}\vartheta_{n,j+k} v_{n-j-k}]/v_{n-j}, \quad j = n-1, \dots, 1.$$

$$v_n = \gamma(0) - \sum_{j=1}^{n} \vartheta_{n,j}^2 v_{n-j}.$$

Innovations algorithm applied to MA(1)

It is easy to see that $\vartheta_{n,j} = 0$ for $n > 1$ and $j > 1$. In fact

$$\vartheta_{n,j} = [\gamma(j) - \sum_{k=1}^{n-j} \vartheta_{n-j,k} \vartheta_{n,j+k} v_{n-j-k}]/v_{n-j}.$$

Innovations algorithm applied to MA(1)

It is easy to see that $\vartheta_{n,j} = 0$ for $n > 1$ and $j > 1$. In fact

$$\vartheta_{n,j} = [\gamma(j) - \sum_{k=1}^{n-j} \vartheta_{n-j,k} \vartheta_{n,j+k} v_{n-j-k}]/v_{n-j}.$$

Then

$$\vartheta_{n,1} = \frac{\gamma(1)}{v_{n-1}} \quad \text{and} \quad v_n = \gamma(0) - \vartheta_{n,1}^2 v_{n-1} = \gamma(0) - \frac{\gamma^2(1)}{v_{n-1}}.$$

Projection on infinite past

We can consider projections based on knowledge of all the past:

$$\mathcal{M}_t = \overline{\text{sp}}(X_s)_{s \le t}$$

i.e. the smallest closed subset containing all the finite linear combinations of X_s, $s \le t$, i.e. the limits (in L^2) of finite linear combinations of X_s.

Projection on infinite past

We can consider projections based on knowledge of all the past:

$$\mathcal{M}_t = \overline{\mathrm{sp}}(X_s)_{s \le t}$$

i.e. the smallest closed subset containing all the finite linear combinations of X_s, $s \le t$, i.e. the limits (in L^2) of finite linear combinations of X_s.

An example. MA(1): $X_t = Z_t - \vartheta Z_{t-1}$. Show that, if $|\vartheta| < 1$,

$$-\sum_{j=1}^{\infty} \vartheta^j X_{t+1-j} = \mathbf{P}_{\mathcal{M}_t} X_{t+1}.$$

Projection on infinite past

We can consider projections based on knowledge of all the past:

$$\mathcal{M}_t = \overline{\mathrm{sp}}(X_s)_{s \le t}$$

i.e. the smallest closed subset containing all the finite linear combinations of X_s, $s \le t$, i.e. the limits (in L^2) of finite linear combinations of X_s.

An example. MA(1): $X_t = Z_t - \vartheta Z_{t-1}$. Show that, if $|\vartheta| < 1$,

$$-\sum_{j=1}^{\infty} \vartheta^j X_{t+1-j} = \mathbf{P}_{\mathcal{M}_t} X_{t+1}.$$

1. the series converges.

Projection on infinite past

We can consider projections based on knowledge of all the past:

$$\mathcal{M}_t = \overline{\mathrm{sp}}(X_s)_{s \le t}$$

i.e. the smallest closed subset containing all the finite linear combinations of X_s, $s \le t$, i.e. the limits (in L^2) of finite linear combinations of X_s.

An example. MA(1): $X_t = Z_t - \vartheta Z_{t-1}$. Show that, if $|\vartheta| < 1$,

$$-\sum_{j=1}^{\infty} \vartheta^j X_{t+1-j} = \mathbf{P}_{\mathcal{M}_t} X_{t+1}.$$

1. the series converges.
2. $X_{t+1} + \sum_{j=1}^{\infty} \vartheta^j X_{t+1-j}$ is orthogonal to X_{t-i}, $i \ge 0$.

Projection on infinite past

We can consider projections based on knowledge of all the past:

$$\mathcal{M}_t = \overline{\mathrm{sp}}(X_s)_{s \le t}$$

i.e. the smallest closed subset containing all the finite linear combinations of X_s, $s \le t$, i.e. the limits (in L^2) of finite linear combinations of X_s.

An example. MA(1): $X_t = Z_t - \vartheta Z_{t-1}$. Show that, if $|\vartheta| < 1$,

$$-\sum_{j=1}^{\infty} \vartheta^j X_{t+1-j} = \mathbf{P}_{\mathcal{M}_t} X_{t+1}.$$

1. the series converges.
2. $X_{t+1} + \sum_{j=1}^{\infty} \vartheta^j X_{t+1-j}$ is orthogonal to X_{t-i}, $i \ge 0$.

What could be $\mathbf{P}_{\mathcal{M}_t} X_{t+1}$ if $|\vartheta| > 1$?

Wold's theorem. General statement

Let $\sigma^2 = \mathbb{E}\left(|X_{t+1} - \mathbf{P}_{\mathcal{M}_t} X_{t+1}|^2\right)$ (*does not depend on t because of stationarity of X_t*).

Definition X_t is said to be *deterministic* if $\sigma^2 = 0$. Example: $X_t = A\cos(\omega t) + B\sin(\omega t)$.

Wold's theorem. General statement

Let $\sigma^2 = \mathbb{E}\left(|X_{t+1} - \mathbf{P}_{\mathcal{M}_t} X_{t+1}|^2\right)$ (*does not depend on* t *because of stationarity of* X_t).

Definition X_t is said to be *deterministic* if $\sigma^2 = 0$. Example: $X_t = A\cos(\omega t) + B\sin(\omega t)$.

Wold's theorem Every stationary process can be written as the sum of an $MA(\infty)$ process and of a deterministic process.

Wold's theorem. Precise statement

Theorem

Let X_t *be a non–deterministic stationary process, i.e.* $\sigma^2 > 0$. *Then there exist unique*

1. $\{\psi_j\}_{j\geq 0}$ *with* $\psi_0 = 1$ *and* $\sum_{j=0}^{\infty} \psi_j^2 < \infty$;
2. $\{Z_t\} \sim WN(0, \sigma^2)$

such that

$$X_t = \sum_{j=0}^{\infty} \psi_j Z_{t-j} + V_t$$

and

1. $Cov(Z_s, V_t) = 0 \quad \forall\ s, t \in \mathbb{Z}$;
2. $\{V_t\}$ *is deterministic.*

Wold's theorem. Sketch of proof

Let $\mathcal{M}_t = \overline{\text{sp}}(X_s)_{s \leq t}$ and $\mathcal{M}_{-\infty} = \cap_{t \in \mathbb{Z}} \mathcal{M}_t$.

Define $Z_t = X_t - \mathbf{P}_{\mathcal{M}_{t-1}} X_t$ and $\psi_j = \dfrac{\langle X_t, Z_{t-j} \rangle}{\sigma^2}$.

Wold's theorem. Sketch of proof

Let $\mathcal{M}_t = \overline{\text{sp}}(X_s)_{s \leq t}$ and $\mathcal{M}_{-\infty} = \cap_{t \in \mathbb{Z}} \mathcal{M}_t$.

Define $Z_t = X_t - \mathbf{P}_{\mathcal{M}_{t-1}} X_t$ and $\psi_j = \dfrac{\langle X_t, Z_{t-j} \rangle}{\sigma^2}$.

$Z_t \in \mathcal{M}_t$ and orthogonal to $\mathcal{M}_{t-1}$ hence to Z_s for $s < t$, proving $\{Z_t\} \sim WN(0, \sigma^2)$.

Wold's theorem. Sketch of proof

Let $\mathcal{M}_t = \overline{\text{sp}}(X_s)_{s \leq t}$ and $\mathcal{M}_{-\infty} = \cap_{t \in \mathbb{Z}} \mathcal{M}_t$.

Define $Z_t = X_t - \mathbf{P}_{\mathcal{M}_{t-1}} X_t$ and $\psi_j = \dfrac{\langle X_t, Z_{t-j} \rangle}{\sigma^2}$.

$Z_t \in \mathcal{M}_t$ and orthogonal to $\mathcal{M}_{t-1}$ hence to Z_s for $s < t$, proving $\{Z_t\} \sim WN(0, \sigma^2)$.
As $\{Z_{t-j}\}_{j \geq 0}$ is an orthogonal sequence,

$$\mathbf{P}_{\overline{\text{sp}}\{Z_s,\ s \leq t\}} X_t = \sum_{j=0}^{\infty} \frac{\langle X_t, Z_{t-j} \rangle}{\|Z_{t-j}\|^2} Z_{t-j} = \sum_{j=0}^{\infty} \psi_j Z_{t-j}$$

with $\sum_j \psi_j^2 < \infty$.

Wold's theorem. Sketch of proof

Let $\mathcal{M}_t = \overline{\text{sp}}(X_s)_{s \le t}$ and $\mathcal{M}_{-\infty} = \cap_{t \in \mathbb{Z}} \mathcal{M}_t$.

Define $Z_t = X_t - \mathbf{P}_{\mathcal{M}_{t-1}} X_t$ and $\psi_j = \dfrac{\langle X_t, Z_{t-j} \rangle}{\sigma^2}$.

$Z_t \in \mathcal{M}_t$ and orthogonal to $\mathcal{M}_{t-1}$ hence to Z_s for $s < t$, proving $\{Z_t\} \sim WN(0, \sigma^2)$.
As $\{Z_{t-j}\}_{j \ge 0}$ is an orthogonal sequence,

$$\mathbf{P}_{\overline{\text{sp}}\{Z_s,\ s \le t\}} X_t = \sum_{j=0}^{\infty} \frac{\langle X_t, Z_{t-j} \rangle}{\|Z_{t-j}\|^2} Z_{t-j} = \sum_{j=0}^{\infty} \psi_j Z_{t-j}$$

with $\sum_j \psi_j^2 < \infty$.
Define $V_t = X_t - \mathbf{P}_{\overline{\text{sp}}\{Z_s,\ s \le t\}} X_t$. By definition $\langle V_t, Z_s \rangle = 0$ for $t \ge s$.
On the other hand $V_t \in \mathcal{M}_t$; for $s > t$, Z_s is orthogonal to $\mathcal{M}_{s-1}$, hence to V_t.

Wold's theorem. Sketch of proof

Let $\mathcal{M}_t = \overline{\text{sp}}(X_s)_{s \le t}$ and $\mathcal{M}_{-\infty} = \cap_{t \in \mathbb{Z}} \mathcal{M}_t$.

Define $Z_t = X_t - \mathbf{P}_{\mathcal{M}_{t-1}} X_t$ and $\psi_j = \dfrac{\langle X_t, Z_{t-j} \rangle}{\sigma^2}$.

$Z_t \in \mathcal{M}_t$ and orthogonal to $\mathcal{M}_{t-1}$ hence to Z_s for $s < t$, proving $\{Z_t\} \sim WN(0, \sigma^2)$.
As $\{Z_{t-j}\}_{j \ge 0}$ is an orthogonal sequence,

$$\mathbf{P}_{\overline{\text{sp}}\{Z_s,\ s \le t\}} X_t = \sum_{j=0}^{\infty} \frac{\langle X_t, Z_{t-j} \rangle}{\|Z_{t-j}\|^2} Z_{t-j} = \sum_{j=0}^{\infty} \psi_j Z_{t-j}$$

with $\sum_j \psi_j^2 < \infty$.
Define $V_t = X_t - \mathbf{P}_{\overline{\text{sp}}\{Z_s,\ s \le t\}} X_t$. By definition $\langle V_t, Z_s \rangle = 0$ for $t \ge s$.
On the other hand $V_t \in \mathcal{M}_t$; for $s > t$, Z_s is orthogonal to $\mathcal{M}_{s-1}$, hence to V_t.
Need only to prove that $\{V_t\}$ is deterministic (*a bit involved*).

Processes with ACF of finite length

We saw: if $\{X_t\} \sim MA(q)$, i.e. $X_t = Z_t + \vartheta_1 Z_{t-1} + \cdots + \vartheta_q Z_{t-q}$ $\{Z_t\} \sim WN(0, \sigma^2)$, then $\gamma(h) = 0$ for $|h| > q$, while $\gamma(q) = \vartheta_q \sigma^2$.

Processes with ACF of finite length

We saw: if $\{X_t\} \sim MA(q)$, i.e. $X_t = Z_t + \vartheta_1 Z_{t-1} + \cdots + \vartheta_q Z_{t-q}$
$\{Z_t\} \sim WN(0, \sigma^2)$, then $\gamma(h) = 0$ for $|h| > q$, while $\gamma(q) = \vartheta_q \sigma^2$.
Vice versa

Theorem

If $\{X_t\}$ is a (0 mean) stationary process s.t. $\gamma(h) = 0$ for $|h| > q$, while $\gamma(q) \neq 0$, then there exist unique $\{Z_t\} \sim WN(0, \sigma^2)$, $\vartheta_1, \ldots, \vartheta_q$ s.t. $X_t = Z_t + \vartheta_1 Z_{t-1} + \cdots + \vartheta_q Z_{t-q}$.

Processes with ACF of finite length are MA(q).Proof

Proof (sketch): As in Wold's, $Z_t = X_t - \mathbf{P}_{\mathcal{M}_{t-1}} X_t$.

Processes with ACF of finite length are MA(q).Proof

Proof (sketch): As in Wold's, $Z_t = X_t - \mathbf{P}_{\mathcal{M}_{t-1}} X_t$.
From $X_{t-1} = Z_{t-1} - \mathbf{P}_{\mathcal{M}_{t-2}} X_{t-1}$, one sees $\mathcal{M}_{t-1} = \overline{\text{sp}}(\mathcal{M}_{t-2}, Z_{t-1})$.
Iterating, one arrives at

$$\mathcal{M}_{t-1} = \overline{\text{sp}}(\mathcal{M}_{t-q-1}, Z_{t-q}, \ldots, Z_{t-1}).$$

$$\text{Hence } \mathbf{P}_{\mathcal{M}_{t-1}} X_t = \mathbf{P}_{\mathcal{M}_{t-q-1}} X_t + \mathbf{P}_{\mathcal{L}(Z_{t-q},\ldots,Z_{t-1})} X_t.$$

Processes with ACF of finite length are MA(q).Proof

Proof (sketch): As in Wold's, $Z_t = X_t - \mathbf{P}_{\mathcal{M}_{t-1}} X_t$.
From $X_{t-1} = Z_{t-1} - \mathbf{P}_{\mathcal{M}_{t-2}} X_{t-1}$, one sees $\mathcal{M}_{t-1} = \overline{\text{sp}}(\mathcal{M}_{t-2}, Z_{t-1})$.
Iterating, one arrives at

$$\mathcal{M}_{t-1} = \overline{\text{sp}}(\mathcal{M}_{t-q-1}, Z_{t-q}, \dots, Z_{t-1}).$$

Hence $\mathbf{P}_{\mathcal{M}_{t-1}} X_t = \mathbf{P}_{\mathcal{M}_{t-q-1}} X_t + \mathbf{P}_{\mathcal{L}(Z_{t-q},\dots,Z_{t-1})} X_t$.

For $h > q$, by assumption $\gamma(h) = 0$, hence

$$\langle X_t, X_{t-h} \rangle = 0 \Longrightarrow \mathbf{P}_{\mathcal{M}_{t-q-1}} X_t = 0.$$

Processes with ACF of finite length are MA(q).Proof

Proof (sketch): As in Wold's, $Z_t = X_t - \mathbf{P}_{\mathcal{M}_{t-1}} X_t$.
From $X_{t-1} = Z_{t-1} - \mathbf{P}_{\mathcal{M}_{t-2}} X_{t-1}$, one sees $\mathcal{M}_{t-1} = \overline{\text{sp}}(\mathcal{M}_{t-2}, Z_{t-1})$.
Iterating, one arrives at

$$\mathcal{M}_{t-1} = \overline{\text{sp}}(\mathcal{M}_{t-q-1}, Z_{t-q}, \ldots, Z_{t-1}).$$

$$\text{Hence } \mathbf{P}_{\mathcal{M}_{t-1}} X_t = \mathbf{P}_{\mathcal{M}_{t-q-1}} X_t + \mathbf{P}_{\mathcal{L}(Z_{t-q},\ldots,Z_{t-1})} X_t.$$

For $h > q$, by assumption $\gamma(h) = 0$, hence

$$\langle X_t, X_{t-h} \rangle = 0 \Longrightarrow \mathbf{P}_{\mathcal{M}_{t-q-1}} X_t = 0.$$

Then $\mathbf{P}_{\mathcal{M}_{t-1}} X_t = \sum_{j=1}^{q} \frac{\langle X_t, Z_{t-j} \rangle}{\|Z_{t-j}\|^2} Z_{t-j}$.

$$\text{This is the thesis with } \vartheta_j = \frac{\mathbb{E}(X_t Z_{t-j})}{\sigma^2}.$$

Periodogram of data. Quick reminder

Let $\omega_k = 2\pi \frac{k}{n}$ and $\mathbf{e}_k = \frac{1}{\sqrt{n}} \begin{pmatrix} e^{i\omega_k} \\ \vdots \\ e^{in\omega_k} \end{pmatrix}$.

$\{\mathbf{e}_k\}_{k \in F_n}$ is an **orthonormal basis** of $\mathbb{C}^n$ where $F_n = \{-\left[\frac{n-1}{2}\right], \dots, \left[\frac{n}{2}\right]\}$.
Note: F_n has n elements. Alternatively, one can use $\{\mathbf{e}_k\}_{k=1,\dots,n}$

Periodogram of data. Quick reminder

Let $\omega_k = 2\pi\frac{k}{n}$ and $\mathbf{e}_k = \frac{1}{\sqrt{n}}\begin{pmatrix} e^{i\omega_k} \\ \vdots \\ e^{in\omega_k} \end{pmatrix}$.

$\{\mathbf{e}_k\}_{k\in F_n}$ is an **orthonormal basis** of $\mathbb{C}^n$ where $F_n = \{-\left[\frac{n-1}{2}\right], \ldots, \left[\frac{n}{2}\right]\}$.

Note: F_n has n elements. Alternatively, one can use $\{\mathbf{e}_k\}_{k=1,\ldots,n}$

Given data $x_1, \ldots, x_n$, we write $\mathbf{x} = (x_1, \ldots, x_n) = \sum_{k\in F_n} a_k \mathbf{e}_k$.

The periodogram (ess. a discrete Fourier transform) of the data is given by

$$I_n(\omega_k) = |a_k|^2 = |\langle x, \mathbf{e}_k\rangle|^2 = \frac{1}{n}\left|\sum_{t=1}^{n} x_t e^{-it\omega_k}\right|^2 \qquad k \in F_n.$$

Periodogram of data. Quick reminder

Let $\omega_k = 2\pi \frac{k}{n}$ and $\mathbf{e}_k = \frac{1}{\sqrt{n}} \begin{pmatrix} e^{i\omega_k} \\ \vdots \\ e^{in\omega_k} \end{pmatrix}$.

$\{\mathbf{e}_k\}_{k \in F_n}$ is an **orthonormal basis** of $\mathbb{C}^n$ where $F_n = \{-\left[\frac{n-1}{2}\right], \dots, \left[\frac{n}{2}\right]\}$.
Note: F_n has n elements. Alternatively, one can use $\{\mathbf{e}_k\}_{k=1,\dots,n}$
Given data $x_1, \dots, x_n$, we write $\mathbf{x} = (x_1, \dots, x_n) = \sum_{k \in F_n} a_k \mathbf{e}_k$.
The periodogram (ess. a discrete Fourier transform) of the data is given by

$$I_n(\omega_k) = |a_k|^2 = |\langle x, \mathbf{e}_k \rangle|^2 = \frac{1}{n} \left| \sum_{t=1}^{n} x_t e^{-it\omega_k} \right|^2 \qquad k \in F_n.$$

If $x_t = A\cos(\omega_k(t+\varphi))$, $I_n(\omega_k) = nA^2$, $I_n(\omega_j) = 0$ for $j \neq k$.

From the periodogram to the spectral density. Motivation

$$\text{If } x_t = \sum_{k=1}^{q} A_k \cos(\omega_k(t + \varphi_k)), \text{ then } I_n(\omega_k) = nA_k^2$$

The height of the periodogram at ω_k represents the strength of the component.

From the periodogram to the spectral density. Motivation

$$\text{If } x_t = \sum_{k=1}^{q} A_k \cos(\omega_k(t+\varphi_k)), \text{ then } I_n(\omega_k) = nA_k^2$$

The height of the periodogram at ω_k represents the strength of the component.

Need a tool analogous to periodogram for a stationary process, assuming that it can decomposed similarly.

From the periodogram to the spectral density. Motivation

$$\text{If } x_t = \sum_{k=1}^{q} A_k \cos(\omega_k(t+\varphi_k)), \text{ then } I_n(\omega_k) = nA_k^2$$

The height of the periodogram at ω_k represents the strength of the component.
Need a tool analogous to periodogram for a stationary process, assuming that it can decomposed similarly. If

$$X_t = \sum_{k=1}^{q} [B_k \cos(\omega_k t) + C_k \sin(\omega_k t)]$$

$$\text{with } \mathbb{E}(B_k) = \mathbb{E}(C_k) = \mathbb{E}(B_k C_j) = 0 \qquad \mathbb{V}(B_k) = \mathbb{V}(C_k) = \sigma_k^2$$

$\gamma(h) = \sum_{k=1}^{q} \sigma_k^2 \cos(\omega_k h)$.

From the periodogram to the spectral density. Motivation

$$\text{If } x_t = \sum_{k=1}^{q} A_k \cos(\omega_k(t+\varphi_k)), \text{ then } I_n(\omega_k) = nA_k^2$$

The height of the periodogram at ω_k represents the strength of the component.

Need a tool analogous to periodogram for a stationary process, assuming that it can decomposed similarly. If

$$X_t = \sum_{k=1}^{q} [B_k \cos(\omega_k t) + C_k \sin(\omega_k t)]$$

$$\text{with } \mathbb{E}(B_k) = \mathbb{E}(C_k) = \mathbb{E}(B_k C_j) = 0 \qquad \mathbb{V}(B_k) = \mathbb{V}(C_k) = \sigma_k^2$$

$\gamma(h) = \sum_{k=1}^{q} \sigma_k^2 \cos(\omega_k h)$. More generally, we will obtain

$$\gamma(h) = \int_{-\pi}^{\pi} e^{ih\lambda} dF(\lambda).$$

The spectral density

Assume $\gamma(h)$, $h \in \mathbb{Z}$ is the ACVF of a process, with $\sum_{h=-\infty}^{\infty} |\gamma(h)| < \infty$.

The **spectral density**: $f(\lambda) = \frac{1}{2\pi} \sum_{h=-\infty}^{\infty} \gamma(h) e^{-ih\lambda}, \quad \lambda \in (-\pi, \pi]$.

The series converges because of the assumption on $\gamma(\cdot)$.

The spectral density

Assume $\gamma(h)$, $h \in \mathbb{Z}$ is the ACVF of a process, with $\sum_{h=-\infty}^{\infty} |\gamma(h)| < \infty$.

The **spectral density**: $f(\lambda) = \frac{1}{2\pi} \sum_{h=-\infty}^{\infty} \gamma(h) e^{-ih\lambda}, \quad \lambda \in (-\pi, \pi]$.

The series converges because of the assumption on $\gamma(\cdot)$.

Three properties of f:

1. f is even: $f(\lambda) = f(-\lambda)$;
2. $f(\lambda) \geq 0$;
3. $\gamma(h) = \int_{-\pi}^{\pi} e^{ih\lambda} f(\lambda)\, d\lambda$.

The spectral density

Assume $\gamma(h)$, $h \in \mathbb{Z}$ is the ACVF of a process, with $\sum_{h=-\infty}^{\infty} |\gamma(h)| < \infty$.

The **spectral density**: $f(\lambda) = \frac{1}{2\pi} \sum_{h=-\infty}^{\infty} \gamma(h) e^{-ih\lambda}, \quad \lambda \in (-\pi, \pi]$.

The series converges because of the assumption on $\gamma(\cdot)$.

Three properties of f:

1. f is even: $f(\lambda) = f(-\lambda)$;
2. $f(\lambda) \geq 0$;
3. $\gamma(h) = \int_{-\pi}^{\pi} e^{ih\lambda} f(\lambda)\, d\lambda$.

The third is the property we looked for. Note that using 1.,

$$\int_{-\pi}^{\pi} e^{ih\lambda} f(\lambda)\, d\lambda = \int_{0}^{\pi} e^{ih\lambda} f(\lambda)\, d\lambda + \int_{0}^{\pi} e^{-ih\lambda} f(\lambda)\, d\lambda = 2 \int_{0}^{\pi} \cos(h\lambda) f(\lambda)\, d\lambda.$$