AR(p) Models

An autoregression of order p, or AR(p), has the form

$$y_t = \sum_{j=1}^{p} \phi_j y_{t-j} + \epsilon_t,$$

ϵ_t : sequence of uncorrelated error terms; typically $\epsilon_t \sim N(0, v)$.

Under Gaussianity, if $\mathbf{y} = (y_T, y_{T-1}, \ldots, y_{p+1})'$, we have

$$p(\mathbf{y}|y_{1:p}) = \prod_{t=p+1}^{T} p(y_t|y_{(t-p):(t-1)}) = \prod_{t=p+1}^{T} N(y_t|\mathbf{f}_t'\phi, v) = N(\mathbf{y}|\mathbf{F}'\phi, v\mathbf{I}_n)$$

with $\phi = (\phi_1, \ldots, \phi_p)'$, $\mathbf{f}_t = (y_{t-1}, \ldots, y_{t-p})'$, $\mathbf{F} = [\mathbf{f}_T, \ldots, \mathbf{f}_{p+1}]$.

AR Models: Causality and Stationarity

Definition

An AR(p) process y_t is *causal* if it can be written as

$$y_t = \Psi(B)\epsilon_t = \sum_{j=0}^{\infty} \psi_j \epsilon_{t-j},$$

with B the backshift operator $B\epsilon_t = \epsilon_{t-1}$, $\psi_0 = 1$ and $\sum_{j=0}^{\infty} |\psi_j| < \infty$.

Definition

The *AR characteristic polynomial* is defined as:

$$\Phi(u) = 1 - \sum_{j=1}^{p} \phi_j u^j.$$

AR Models: Causality and Stationarity

- y_t is causal only when $\Phi(u)$ has all its roots outside the unit circle (or the reciprocal roots inside the unit circle). In other words, y_t is causal if $\Phi(u) = 0$ only when $|u| > 1$.
- Causality $\Rightarrow$ Stationarity. The reverse is not necessarily true, i.e., we can have stationary processes that are not causal.

AR Models: State-space representation

$y_t \sim \mathrm{AR}(p)$ can be written as

$$\begin{aligned} y_t &= \mathbf{F}'\mathbf{x}_t \\ \mathbf{x}_t &= \mathbf{G}\mathbf{x}_{t-1} + \omega_t, \end{aligned}$$

with $\mathbf{x}_t = (y_t, y_{t-1}, \ldots, y_{t-p+1})'$, $\omega_t = (\epsilon_t, 0, \ldots, 0)'$, $\mathbf{F} = (1, 0, \ldots, 0)'$ and

$$\mathbf{G} = \begin{pmatrix} \phi_1 & \phi_2 & \phi_3 & \cdots & \phi_{p-1} & \phi_p \\ 1 & 0 & 0 & \cdots & 0 & 0 \\ 0 & 1 & 0 & \cdots & 0 & 0 \\ \vdots & & & \ddots & 0 & \vdots \\ 0 & 0 & \cdots & \cdots & 1 & 0 \end{pmatrix}.$$

AR Models: State-space representation

- The eigenvalues of the matrix $\mathbf{G}$, denoted as $\alpha_1, \ldots, \alpha_p$, are the reciprocal roots of the AR characteristic polynomial $\Rightarrow$

$$\phi(u) = 1 - \sum_{j=1}^{p} \phi_j u^j = \prod_{j=1}^{p} (1 - \alpha_j u).$$

- The expected behavior of the process in the future is given by

$$f_t(h) = E(y_{t+h} | y_{1:t}) = \mathbf{F}' \mathbf{G}^h \mathbf{x}_t = \sum_{j=1}^{p} c_{t,j} \alpha_j^h,$$

with $c_{t,j} = d_j e_{t,j}$, and d_j, $e_{t,j}$ elements of $\mathbf{d} = \mathbf{E}'\mathbf{F}$, and $\mathbf{e}_t = \mathbf{E}^{-1} \mathbf{x}_t$, where $\mathbf{E}$ is an eigenmatrix of $\mathbf{G}$.

AR Models: Forecast function

- If y_t is such that $|\alpha_j| < 1$ for all j, $f_t(h) \to 0$ as h increases.
- If α_j is real, its contribution to the forecast function is $c_{t,j}\alpha_j^h$.
- If α_j and α_{j+1} are complex conjugates, $c_{t,j}$ and $c_{t,j+1}$ are also complex conjugates that can be written as $a_t \exp(\pm ib_t)$. The contribution of this pair of complex reciprocal roots to $f_t(h)$ is $2a_t r^h \cos(\omega h + b_t)$, with r and ω the modulus and frequency of α_j and α_{j+1}.

AR Models: ACF

The autocorrelation structure of an AR(p) process is given in terms of the solution of the homogeneous difference equation

$$\rho(h) - \phi_1 \rho(h-1) - \ldots - \phi_p \rho(h-p) = 0, \quad h \geq p.$$

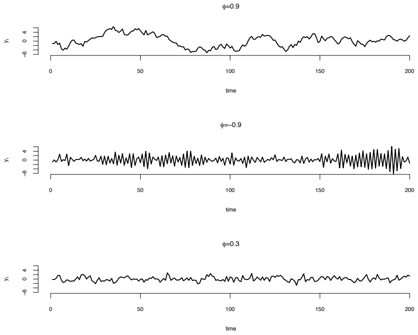
If $\alpha_1, \ldots, \alpha_r$ are the reciprocal characteristic roots with multiplicities $m_1, \ldots, m_r$ and $\sum_{i=1}^r m_i = p$, the general solution of the equation is

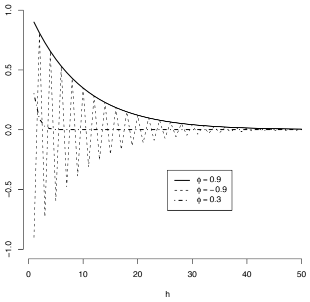
$$\rho(h) = \alpha_1^h p_1(h) + \ldots + \alpha_r^h p_r(h), \quad h \geq p,$$

where $p_i(h)$ is a polynomial of degree $m_i - 1$.

AR(1): $y_t = \phi y_{t-1} + \epsilon_t, \;\; \epsilon_t \sim N(0, v).$

- *Characteristic polynomial:* $\Phi(u) = 1 - \phi u$. If $|\phi| < 1$ the process is stationary.
- *Forecast function:* $f_t(h) = \phi^h y_t$. *Autocorrelation function (ACF):* $\rho(h) = Corr[y_t, y_{t-h}] = \phi^h$, $h \geq 0$.





AR Models: Forecast function, AR(2)

Let $y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \epsilon_t$ and α_1, α_2 be the reciprocal roots of $\Phi(u) = 1 - \phi_1 u - \phi_2 u^2$.

- α_1, α_2 **real and distinct.**

$$f_t(h) = c_{t,1}\alpha_1^h + c_{t,2}\alpha_2^h.$$

- $\alpha_1 = \alpha_2 = \alpha$ **real.**

$$f_t(h) = p(h)\alpha^h,$$

  with $p(h)$ polynomial of degree one, i.e., $p(h) = d + eh$.

- α_1, α_2 **complex conjugates.**

$$f_t(h) = 2a_t r^h \cos(\omega h + b_t).$$

AR Models: ACF of an AR(2)

Autocorrelation Function

- α_1, α_2 **real and distinct.**

$$\rho(h) = a\alpha_1^h + b\alpha_2^h, \quad h \geq 2$$

- $\alpha_1 = \alpha_2 = \alpha$ **real.**

$$\rho(h) = (a + bh)\alpha^h, \quad h \geq 2$$

- α_1, α_2 **complex conjugates.**

$$\rho(h) = ar^h \cos(h\omega + b), \quad h \geq 2.$$

AR Models: PACF

Let $\phi(h,h)$ be the *partial autocorrelation coefficient at lag h*, given by

$$\phi(h,h) = \begin{cases} \rho(y_1, y_0) = \rho(1) & h = 1 \\ \rho(y_h - y_h^{h-1}, y_0 - y_0^{h-1}) & h > 1, \end{cases}$$

with y_h^{h-1} the minimum mean square linear predictor of y_h given $y_{h-1}, \ldots, y_1$, and y_0^{h-1} the minimum mean square linear predictor of y_0 given $y_1, \ldots, y_{h-1}$.

Result: If $y_t \sim \mathrm{AR}(p)$, $\phi(h,h) = 0$ for all $h > p$.

AR Models: Computing the PACF

- $\Gamma_n \phi_n = \gamma_n$, with Γ_n an $n \times n$ matrix with elements $\{\gamma(h-j)\}_{j,h=1}^{n}$, $\gamma_n = (\gamma(1), \ldots, \gamma(n))'$, and $\phi_n = (\phi(n,1), \ldots, \phi(n,n))'$.
- **Durbin-Levinson recursion.** For $n = 0$ $\phi(0,0) = 0$, and for $n \geq 1$

$$\phi(n,n) = \frac{\rho(n) - \sum_{h=1}^{n-1} \phi(n-1,h)\rho(n-h)}{1 - \sum_{h=1}^{n-1} \phi(n-1,h)\rho(h)},$$

with

$$\phi(n,h) = \phi(n-1,h) - \phi(n,n)\phi(n-1,n-h),$$

for $n \geq 2$ and $h = 1 : (n-1)$.

Sample PACF can also be computed using these algorithms.

AR Models: Yule-Walker Estimation

$$\hat{\Gamma}_p \hat{\phi} = \hat{\gamma}_p, \quad \hat{v} = \hat{\gamma}(0) - \hat{\gamma}_p' \hat{\Gamma}_p^{-1} \hat{\gamma}_p.$$

It can be shown that

$$\sqrt{T}(\hat{\phi} - \phi) \approx N(\mathbf{0}, v\Gamma_p^{-1}),$$

and that $\hat{v}$ is close to v when T is large.

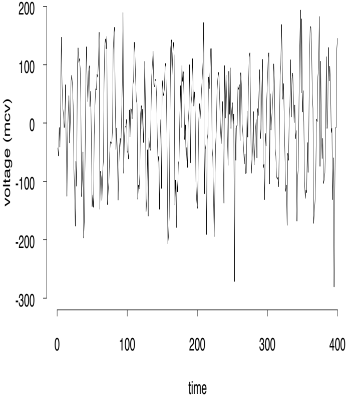
AR Models: MLE and Bayesian estimation

- **MLE.** Find $\hat{\phi}$ that maximizes

$$
\begin{aligned}
p(\mathbf{y}|\phi, v, y_{1:p}) &= \prod_{t=p+1}^{T} p(y_t|\phi, v, y_{(t-p):(t-1)}) \\
&= \prod_{t=p+1}^{T} N(y_t|\mathbf{f}_t'\phi, v) = N(\mathbf{y}|\mathbf{F}'\phi, v\mathbf{I}_n).
\end{aligned}
$$

- **Bayesian.** Combine $p(\mathbf{y}|\phi, v, y_{1:p})$ with prior $p(\phi, v)$.
  - Reference prior $p(\phi, v) \propto 1/v$.
  - Conjugate prior $p(\phi|v) = N(\phi|\mathbf{m}_0, v\mathbf{C}_0)$ and $p(v) = IG(n_0/2, d_0/2)$.
  - Non-conjugate.

AR Models: EEG data analysis



Posterior mean from AR(8) reference analysis ($n = 392$):

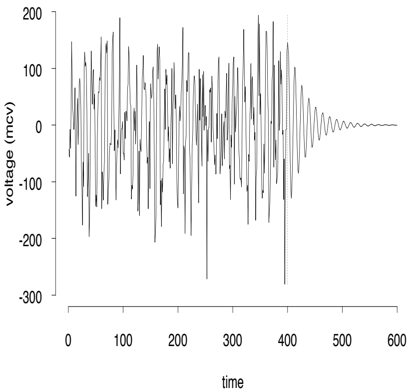
$$\hat{\phi} = (0.27, 0.07, -0.13, -0.15, -0.11, -0.15, -0.23, -0.14)'$$

and $s = 61.52$. These estimates lead to the following estimates of the reciprocal characteristic roots:

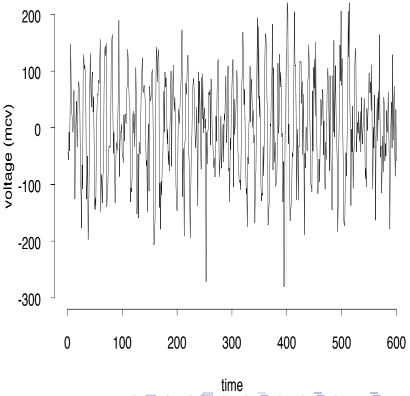
$$(0.97, 12.73); \quad (0.81, 5.10); \quad (0.72, 2.99); \quad (0.66, 2.23).$$

AR Models: EEG data analysis

Forecast function



Future sample



AR Models: Model Order Assessment

Choose a value p^* and for all $p \leq p^*$ compute

- **Akaike's Information Criterion (AIC):**

$$2p + n\log(s_p^2).$$

- **Bayesian Information Criterion (BIC):**

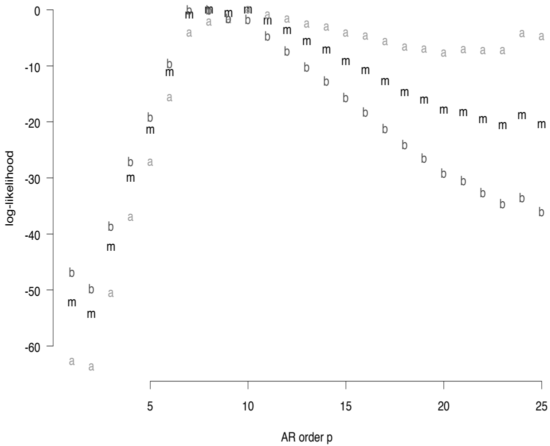
$$\log(n)p + n\log(s_p^2).$$

- **Marginal:**

$$p(y_{(p^*+1):T}|y_{1:p^*}, p) = \int p(y_{(p^*+1):T}|\phi_p, v, y_{1:p^*})p(\phi_p, v)d\phi_p dv.$$

Here $n = T - p^*$.

Order Assessment in EEG Example: Take $p^* = 25$ and $n = 400 - p^*$.



AR Models: Initial Observations

Full likelihood:

$$\begin{aligned} p(y_{1:T}|\phi, v) &= p(y_{(p+1):T}|\phi, v, y_{1:p})p(y_{1:p}|\phi, v) \\ &= p(\mathbf{y}|\phi, v, \mathbf{x}_p)p(\mathbf{x}_p|\phi, v). \end{aligned}$$

What about $p(\mathbf{x}_p|\phi, v)$?

- $N(\mathbf{x}_p|\mathbf{0}, \mathbf{A})$ with $\mathbf{A}$ known.
- $N(\mathbf{x}_p|\mathbf{0}, v\mathbf{A}(\phi))$ with $\mathbf{A}(\phi)$ depending on ϕ through the autocorrelation function and

  $$p(y_{1:T}|\phi, v) \propto v^{-T/2}|\mathbf{A}(\phi)|^{-1/2}\exp(-Q(y_{1:T}, \phi)/2v),$$

  where

  $$Q(y_{1:T}, \phi) = \sum_{t=p+1}^{T} (y_t - \mathbf{f}_t'\phi)^2 + \mathbf{x}_p'\mathbf{A}(\phi)^{-1}\mathbf{x}_p.$$

AR Models: Initial Observations

It can be shown (e.g., see Box, Jenkins, and Reinsel, 2008) that

$$Q(y_{1:T}, \phi) = a - 2\mathbf{b}'\phi + \phi'\mathbf{C}\phi,$$

with $a, \mathbf{b}$, and $\mathbf{C}$ obtained from

$$\mathbf{D} = \begin{pmatrix} a & -\mathbf{b}' \\ -\mathbf{b} & \mathbf{C} \end{pmatrix},$$

and $\mathbf{D}$ a $(p+1) \times (p+1)$ with $D_{ij} = \sum_{r=0}^{T+1-j-i} y_{i+r} y_{j+r}$.

- If $|\mathbf{A}(\phi)|^{-1/2}$ is ignored when computing $p(y_{1:T}|\phi, v)$, the likelihood function is that of a standard linear model form and so, if $p(\phi, v) \propto 1/v$ we have a normal/inverse-gamma posterior with $\hat{\phi}^* = \mathbf{C}^{-1}\mathbf{b}$.
- Jeffreys' prior is approximately $p(\phi, v) \propto |\mathbf{A}(\phi)|^{1/2} v^{-1/2}$.

AR Models: Structured non-conjugate priors

If $y_t \sim \text{AR}(p)$, y_t is causal and stationary if all the AR reciprocal roots have moduli less than one.
Huerta and West (1999) proposed priors on the reciprocal characteristic roots as follows:

- Let C be the maximum number of pairs of complex roots and R the maximum number of real roots with $p = 2C + R$.
- Denote the complex roots as (r_j, λ_j), for $j = 1 : C$ and the real roots as r_j, for $j = (C+1) : (R+C)$.

Then...

AR Models: Structured non-conjugate priors

- **Prior on the real reciprocal roots.**

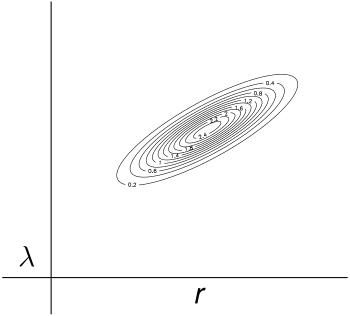
$$\begin{aligned} r_j &\sim \pi_{r,-1} I_{(-1)}(r_j) + \pi_{c,0} I_0(r_j) + \pi_{r,1} I_1(r_j) + \\ &\quad + (1 - \pi_{r,0} - \pi_{r,-1} - \pi_{r,1}) g_r(r_j), \end{aligned}$$

  with $g_r(\cdot)$ a continuous distribution on $(-1, 1)$, e.g., $g_r(\cdot) = U(\cdot | -1, 1)$.
- **Prior on the complex reciprocal roots.**

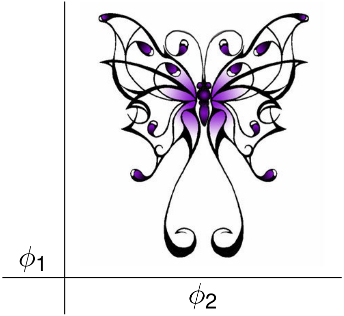
$$\begin{aligned} r_j &\sim \pi_{c,0} I_0(r_j) + \pi_{c,1} I_1(r_j) + (1 - \pi_{c,1} - \pi_{c,0}) g_c(r_j), \\ \lambda_j &\sim h(\lambda_j), \end{aligned}$$

  with $g_c(r_j)$ and $h(\lambda_j)$ continuous distributions on $0 < r_j < 1$ and $2 < \lambda_j < \lambda_u$, with $\lambda_u \leq T/2$.
- $Dir(\pi_{r,-1}, \pi_{r,0}, \pi_{r,1} | 1, 1, 1)$ and $Dir(\pi_{c,0}, \pi_{c,1} | 1, 1)$.
- $IG(v | a, b)$.

AR Models:Structured non-conjugate priors



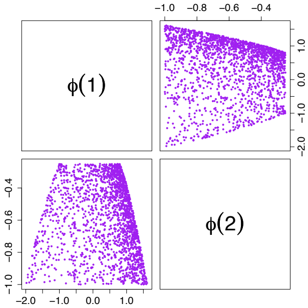
$? \Rightarrow ?$



AR Models: Structured non-conjugate priors

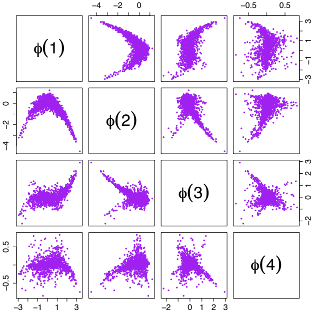
One pair of complex roots

$r \sim U(0.5, 1), \lambda \sim U(2, 10)$



One pair of complex roots and two real roots

$r_1 \sim U(0, 1), \lambda_1 \sim U(2, 10),$
$r_2, r_3 \sim U(-1, 1)$



AR Models: Structured non-conjugate priors

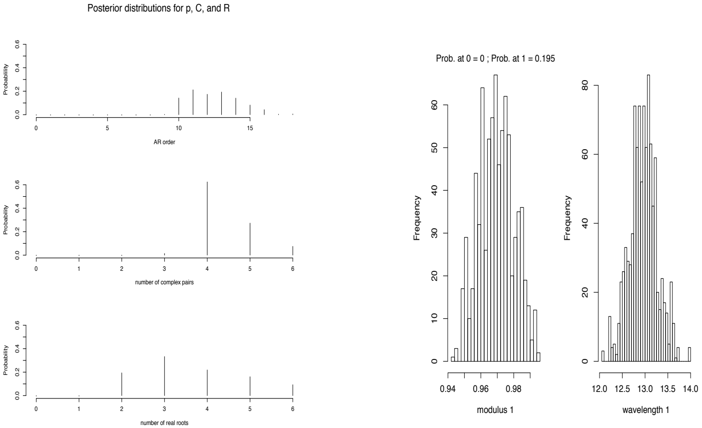
Posterior inference under these priors can be achieved using a reversible jump MCMC algorithm as detailed in Huerta and West (1999).

`ARcomp` code available at

`www.stat.duke.edu/software/research`

AR Models: Structured non-conjugate priors

Analysis of EEG data using $C = R = 6$ (i.e., maximum model order of 18).



ARMA Models

y_t follows an autoregressive moving average model, ARMA(p, q), if

$$y_t = \sum_{i=1}^{p} \phi_i y_{t-i} + \sum_{j=1}^{q} \theta_j \epsilon_{t-j} + \epsilon_t,$$

We can also write

$$\underbrace{(1 - \phi_1 B - \ldots - \phi_p B^p)}_{\Phi(B)} y_t = \underbrace{(1 + \theta_1 B + \ldots + \theta_q B^q)}_{\Theta(B)} \epsilon_t$$

We typically assume $\epsilon_t \sim N(0, v)$. If $q = 0$ $y_t \sim$ AR(p) and if $p = 0$ $y_t \sim$ MA(q).

ARMA Models

Definition

A MA(q) process is *identifiable or invertible* if the roots of the MA characteristic polynomial $\Theta(u)$ lie outside the unit circle. In this case is possible to write the process as an infinite order AR.

Example

Let $y_t \sim$ MA(1) with MA coefficient θ. The process is stationary for all θ and

$$\rho(h) = \begin{cases} 1 & h = 0 \\ \frac{\theta}{(1+\theta^2)} & h = 1 \\ 0 & \text{otherwise.} \end{cases}$$

Note that a MA process with coefficient $1/\theta$ has the same ACF $\Rightarrow$ the identifiability condition is $1/\theta > 1$.

An ARMA(p, q) process is *causal* if the roots of $\Phi(u)$ lie outside the unit circle. In this case:

$$y_t = \Phi^{-1}(B)\Theta(B)\epsilon_t = \Psi(B)\epsilon_t = \sum_{j=0}^{\infty} \psi_j \epsilon_{t-j},$$

with $\Phi(B)\Psi(B) = \Theta(B)$. The ψ_js can be found by solving the homogeneous difference equations

$$\psi_j - \sum_{h=1}^{p} \phi_h \psi_{j-h} = 0, \quad j \geq \max(p, q+1),$$

with initial conditions

$$\psi_j - \sum_{h=1}^{j} \phi_h \psi_{j-h} = \theta_j, \quad 0 \leq j < \max(p, q+1),$$

and $\theta_0 = 1$.

ARMA Models

The general solution is given by

$$\psi_j = \alpha_1^j p_1(j) + \ldots + \alpha_r^j p_r(j),$$

where $\alpha_1, \ldots, \alpha_r$ are the reciprocal roots of $\Phi(u)$, with multiplicities $m_1, \ldots, m_r$, respectively, and each $p_i(j)$ is a polynomial of degree $m_i - 1$.

ARMA Models: ACF of MA(q)

If $y_t \sim$ MA(q), its ACF is

$$\rho(h) = \begin{cases} 1 & h = 0 \\ \frac{\sum_{j=0}^{q-h} \theta_j \theta_{j+h}}{1+\sum_{j=1}^{q} \theta_j^2} & h = 1 : q \\ 0 & h > q, \end{cases}$$

ARMA Models: ACF of ARMA(p, q)

The autocovariance function can be written in terms of the general homogeneous equations

$$\gamma(h) - \phi_1\gamma(h-1) - \ldots - \phi_p\gamma(h-p) = 0, \quad h \geq \max(p, q+1),$$

with initial conditions

$$\gamma(h) - \sum_{j=1}^{p} \phi_j\gamma(h-j) = v\sum_{j=h}^{q} \theta_j\psi_{j-h}, \quad 0 \leq h < \max(p, q+1).$$

PACF can also be computed. The PACF coefficients of MA and ARMA processes will never drop to zero.

ARMA Models: Inverting AR Components

Assume that $y_t \sim \text{AR}(p)$, i.e., $\Phi(B)y_t = \prod_{i=1}^{p}(1-\alpha_i B)y_t = \epsilon_t$. Then

$$\prod_{i=1}^{r}(1-\alpha_i B)y_t = \prod_{i=r+1}^{p}(1-\alpha_i B)^{-1}\epsilon_t = \Psi^*(B)\epsilon_t,$$

with $\Psi^*(u) = 1 + \sum_{j=1}^{\infty}\psi_j^* u^j$, such that
$1 = \Psi^*(u)\prod_{i=r+1}^{p}(1-\alpha_i u)$.

$\Rightarrow$

$$\begin{aligned} y_t &= \sum_{j=1}^{r}\phi_j^* y_{t-j} + \epsilon_t + \sum_{j=1}^{\infty}\psi_j^*\epsilon_{t-j}, \\ y_t &\approx \sum_{j=1}^{r}\phi_j^* y_{t-j} + \epsilon_t + \sum_{j=1}^{q}\psi_j^*\epsilon_{t-j}, \end{aligned}$$

where $\Phi^*(u) = \prod_{i=1}^{r}(1-\alpha_i u) = 0$.

ARMA Models: Inverting AR Components

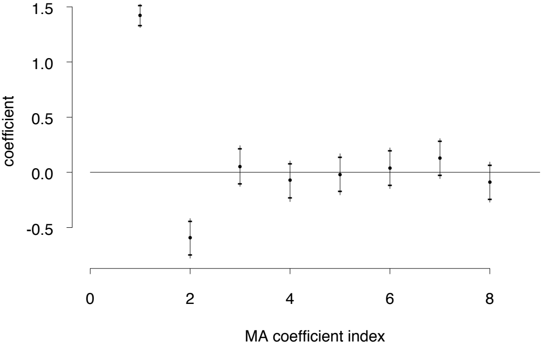
Algorithm

1. Initialize the algorithm by setting $\psi_i^* = 0$ for all $i = 1 : q$.
2. For $i = (r+1) : p$, update $\psi_1^* = \psi_1^* + \alpha_i$, and then,
   - for $j = 2 : q$, update $\psi_j^* = \psi_j^* + \alpha_i \psi_{j-1}^*$.

EEG data, AR(8)**:** Bayesian reference analysis: we obtained estimates of reciprocal characteristic roots given by $(0.97, 12.73)$, $(0.81, 5.10)$, $(0.72, 2.99)$, and $(0.66, 2.23)$ $\Rightarrow$

$$y_t \approx \phi_1^* y_{t-1} + \phi_2^* y_{t-2} + \epsilon_t + \sum_{j=1}^{q} \psi_j^* \epsilon_{t-j},$$

where $\phi_1^* = 2r_1 \cos(2\pi/\lambda_1)$ and $\phi_2^* = -r_1^2$. Taking $q = 8$ we have...



Note: the optimal ARMA(p, q) model for these data, among all the models with $p, q \leq 8$, is an ARMA$(2, 2)$. The MLEs for the MA coefficients are $\hat{\theta}_1 = 1.37$ and $\hat{\theta}_2 = -0.51$.