Chapter 3: Parametric spectral estimation

Motivation

Non-parametric methods

- quite general: applicable to any kind of signals
- neglect the specific properties of the signals under analysis

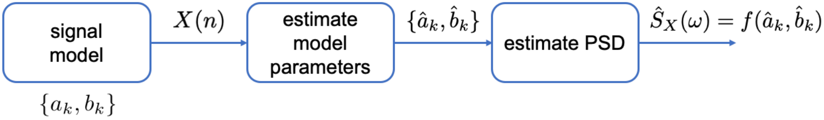
Parametric methods

- take into account the signal properties $\rightarrow$ signal model, defined by model parameters
- estimate the model parameters
- estimate the PSD from the estimated model parameters

Benefits

- typically reduced number of model parameters
  $\rightarrow$ allows sample size reduction for equivalent performance
- improved performance (resolution, variance)...
- ... if the assumed model is correct.

General concept



1. The **signal model** is built from prior information about the process $\rightarrow$ model parameters
2. The **model parameters** are estimated from the observed process
3. The **PSD** is constructed from the estimated model parameters.

Properties of the PSD

P1) The PSD is defined as the Fourier transform of the ACS

$$S(\omega) = \mathcal{F}\{r(k)\} = \sum_{k=-\infty}^{+\infty} r(k) e^{-\jmath\omega k}$$

We can also define the **PSD in the z-domain** as:

$$S(z) = \mathcal{Z}\{r(k)\} = \sum_{k=-\infty}^{+\infty} r(k) z^{-k} \qquad \text{with} \qquad S(\omega) = S(z)\big|_{z=e^{\jmath\omega}}$$

P2) Since the ACS is conjugate symmetric, i.e., $r(k) = r^*(-k)$, the PSD is **real valued**:

$$S(\omega) = S^*(\omega) \qquad \text{and} \qquad S(z) = S^*(1/z^*)$$

P3) If the process is real valued, the PSD is an **even function**:

$$S(\omega) = S(-\omega) \qquad \text{and} \qquad S(z) = S^*(z^*)$$

P4) The PSD is **nonnegative**:

$$S(\omega) \geq 0, \qquad \forall \omega \in \mathbb{R}$$

P5) **Total power**:

$$P_X = \mathrm{E}\{|x(n)|^2\} = r(0) = \frac{1}{2\pi} \int_{-\pi}^{\pi} S(\omega) d\omega$$

$\rightarrow$ $S(\omega)$ represents the power density of the process in the frequency domain (= PSD).

Filtering random processes

WSS random process $X(n)$ with mean μ_X, ACS $r_X(k)$ and PSD $S_X(\omega)$ is filtered with a linear time-invariant (LTI) filter with transfer function $H(z)$:

Q: What is the mean, ACS and PSD of the output process $Y(n)$?
We can prove (see tutorial) that

$$\boxed{\begin{aligned} \mu_Y &= \mu_X H(e^{\jmath 0}) \\ r_Y(k) &= r_X(k) * h(k) * h^*(-k) \\ S_Y(\omega) &= S_X(\omega)|H(\omega)|^2 \\ S_Y(z) &= S_X(z)H(z)H^*(1/z^*) \end{aligned}}$$

Special case: white noise input with variance σ_X^2

$$\boxed{\begin{aligned} S_Y(\omega) &= \sigma_X^2 |H(\omega)|^2 \\ S_Y(z) &= \sigma_X^2 H(z)H^*(1/z^*) \end{aligned}}$$

Spectral factorization theorem

Let $S(\omega)$ be the PSD of a WSS random process $X(n)$.

If $S(\omega)$ is a continuous function of ω, then it can be factored as

$$S(z) = \sigma_\varepsilon^2 H(z) H^*(1/z^*)$$

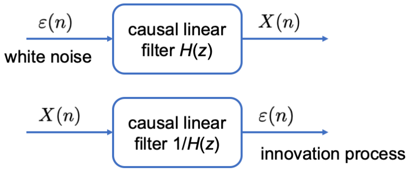
where $H(z)$ is a causal, stable, minimum phase LTI filter, i.e., with no poles or zeros outside the unit circle.

Any process that admits this factorization is called a **regular process**.

Properties of regular processes

- They can be realized as the output of a causal, stable filter driven by white noise with variance $\sigma_\varepsilon^2 \rightarrow$ **innovations representation** of the process.

- If filtered with $1/H(z)$, the output is a white noise with variance $\sigma_\varepsilon^2 \rightarrow$ **whitening**
  - $1/H(z)$: **whitening filter**
  - white noise output $\varepsilon(n)$: **innovation process**
  - σ_ε^2: **innovation variance** or **modeling error**.



Spectral factorization (cont'd)

Wold decomposition theorem

Any WSS random process may be decomposed into the sum of two orthogonal processes:

$$X(n) = X_r(n) + X_p(n)$$

with $\mathrm{E}\{X_r(n)X_p^*(n-k)\} = 0$, $\forall k \in \mathbb{Z}$, where

- $X_r(n)$: **regular** process
- $X_p(n)$: **predictable** process.

A **predictable process** can be predicted without error from a linear combination of its previous values:

$$X_p(n) = \sum_{k=1}^{\infty} a(k)X_p(n-k).$$

Its PSD consists of impulses:

$$S_{X_p}(\omega) = \sum_{k=1}^{K} \alpha_k \delta(\omega - \omega_k).$$

Corollary: the general form of the PSD of a WSS process $X(n)$ is given by

$$S_X(\omega) = S_{X_r}(\omega) + \sum_{k=1}^{K} \alpha_k \delta(\omega - \omega_k).$$

Stochastic process models for rational spectra

Important special case: $S(z)$ is a rational function

$$S(z) = \frac{N(z)}{D(z)} \underset{\substack{\uparrow \\ \text{spectral factorization}}}{=} \sigma_\varepsilon^2 H(z) H^*(1/z^*) = \sigma_\varepsilon^2 \frac{B(z)B^*(1/z^*)}{A(z)A^*(1/z^*)} \qquad R_1 < |z| < R_2$$

with

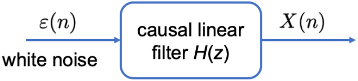
$$H(z) \stackrel{\text{def}}{=} \frac{B(z)}{A(z)} = \frac{\sum_{k=0}^{q} b_k z^{-k}}{1 + \sum_{k=1}^{p} a_k z^{-k}} \qquad |z| > R_1.$$

By the spectral factorization theorem, $H(z)$ and $1/H(z)$ are causal, stable, minimum phase $\rightarrow B(z)$ and $A(z)$ have all their roots inside the unit circle.

Power spectral density

$$\boxed{S(\omega) = \sigma_\varepsilon^2 \frac{|B(\omega)|^2}{|A(\omega)|^2} = \sigma_\varepsilon^2 \frac{\left|\sum_{k=0}^{q} b_k e^{-\jmath\omega k}\right|^2}{\left|1 + \sum_{k=1}^{p} a_k e^{-\jmath\omega k}\right|^2}}$$

Equivalent time series representation



$$X(z) = H(z)\varepsilon(z) = \frac{B(z)}{A(z)}\varepsilon(z)$$

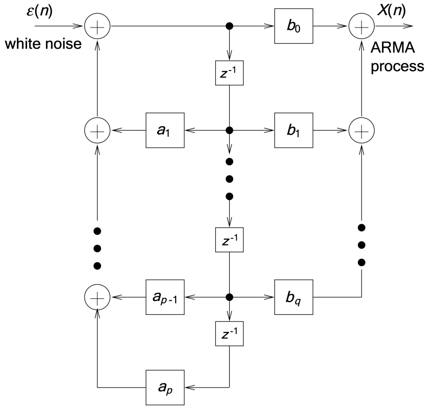
$$A(z)X(z) = B(z)\varepsilon(z) \qquad \Rightarrow \qquad \mathcal{Z}^{-1}\{A(z)X(z)\} = \mathcal{Z}^{-1}\{B(z)\varepsilon(z)\}$$

$$\boxed{X(n)+\sum_{k=1}^{p} a_k X(n-k) = \sum_{k=0}^{q} b_k \varepsilon(n-k) \quad (12)}$$

Autoregressive moving average (ARMA) processes

If $p > 0$ and $q > 0$
$\rightarrow$ **autoregressive moving average (ARMA) process** of order (p, q): ARMA(p, q)

$$X(n) + \sum_{k=1}^{p} a_k X(n-k) = \sum_{k=0}^{q} b_k \varepsilon(n-k) \quad \Rightarrow \quad S(\omega) = \sigma_\varepsilon^2 \frac{\left|\sum_{k=0}^{q} b_k e^{-\jmath\omega k}\right|^2}{\left|1 + \sum_{k=1}^{p} a_k e^{-\jmath\omega k}\right|^2}$$



ARMA processes — examples

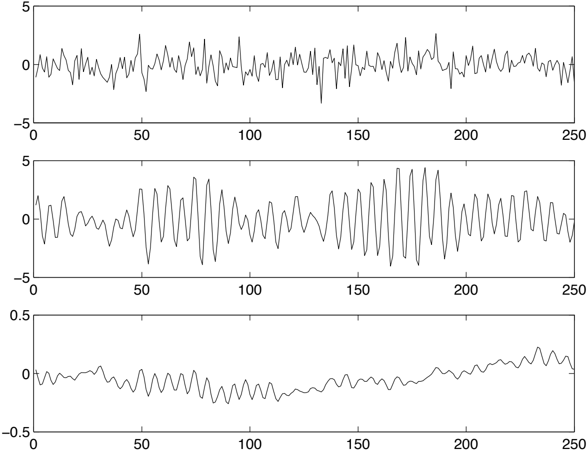


Figure: (Top) White noise input.
(Middle) $\mathbf{a} = [1, -0.975, 0.95]$, $\mathbf{b} = [1, 1, 1, 1]/4$.
(Bottom) $\mathbf{a} = [1, -0.975, 0.95]$, $\mathbf{b} = \underbrace{[1, 1, \ldots, 1]}_{\text{100 entries}}/100$.

Link between ARMA parameters and ACS

Establishing this link is useful for estimating the model parameters.

Multiplying eqn. (12) by $X^*(n-k)$ and taking expectations:

$$X(n)X^*(n-k) + \sum_{m=1}^{p} a_m X(n-m)X^*(n-k) = \sum_{m=0}^{q} b_m \varepsilon(n-m)X^*(n-k)$$

$$\mathrm{E}\{X(n)X^*(n-k)\} + \sum_{m=1}^{p} a_m \mathrm{E}\{X(n-m)X^*(n-k)\} = \sum_{m=0}^{q} b_m \mathrm{E}\{\varepsilon(n-m)X^*(n-k)\}.$$

Hence:

$$r(k) + \sum_{m=1}^{p} a_m r(k-m) = \sum_{m=0}^{q} b_m r_{\varepsilon X}(k-m).$$

To compute the **cross-correlation** $r_{\varepsilon X}(k)$, we note that:

- The innovation processe is white: $r_{\varepsilon}(k) = \sigma_{\varepsilon}^2 \delta(k)$.
- The model output is given by the convolution of the filter impulse response $h(n) \stackrel{\text{def}}{=} \mathcal{Z}^{-1}\{H(z)\}$ and the filter input:

$$X(n) = h(n) * \varepsilon(n) = \sum_{m=-\infty}^{+\infty} h(m)\varepsilon(n-m)$$

- $H(z)$ is causal (spectral factorization theorem), i.e., $h(n) = 0$, $n < 0$, and then:

$$\sum_{m=-\infty}^{+\infty} h(m)\varepsilon(n-m) = \sum_{m=0}^{+\infty} h(m)\varepsilon(n-m).$$

Accordingly:

$$r_{\varepsilon X}(k) = \mathrm{E}\{\varepsilon(n)X^*(n-k)\} = \mathrm{E}\Big\{\sum_{m=0}^{\infty} h^*(m)\varepsilon^*(n-k-m)\varepsilon(n)\Big\}$$

$$= \sum_{m=0}^{\infty} h^*(m)\underbrace{\mathrm{E}\{\varepsilon^*(n-k-m)\varepsilon(n)\}}_{r_{\varepsilon}(k+m)=\sigma_{\varepsilon}^2\delta(k+m)} = \sigma_{\varepsilon}^2 \sum_{m=0}^{\infty} h^*(m)\delta(k+m) = \sigma_{\varepsilon}^2 h^*(-k).$$

Link between ARMA parameters and ACS (cont'd)

Again, because $h(n)$ is causal (spectral factorization theorem), then

$$\sum_{m=0}^{q} b_m r_{\varepsilon X}(k-m) \underset{r_{\varepsilon X}(k)=\sigma_\varepsilon^2 h^*(-k)}{\overset{\uparrow}{=}} \sigma_\varepsilon^2 \sum_{m=0}^{q} b_m h^*(m-k) = \sigma_\varepsilon^2 c(k)$$

with

$$c(k) \stackrel{\text{def}}{=} \sum_{m=0}^{q-k} b_{m+k} h^*(m).$$

Hence, for $k \geq 0$:

$$\boxed{r(k) + \sum_{m=1}^{p} a_m r(k-m) = \begin{cases} \sigma_\varepsilon^2 c(k) & 0 \leq k \leq q \\ 0 & k > q. \end{cases} \qquad (13)}$$

For $k < 0$, we just enforce the conjugate symmetry of the ACS: $r(k) = r^*(-k)$.

These are the **Yule-Walker equations**, linking the ARMA parameters with the ACS.

Yule-Walker equations — matrix form

Casting the equations for $k = 0, 1, \ldots, p+q$ in matrix form, we have:

$$\left[\begin{array}{llll} r(0) & r(-1) & \cdots & r(-p) \\ r(1) & r(0) & \cdots & r(-p+1) \\ \vdots & \vdots & \ddots & \vdots \\ r(q) & r(q-1) & \cdots & r(q-p) \\ \hline r(q+1) & r(q) & \cdots & r(q-p+1) \\ \vdots & \vdots & \ddots & \vdots \\ r(q+p) & r(q+p-1) & \cdots & r(q) \end{array}\right] \left[\begin{array}{c} 1 \\ a_1 \\ \vdots \\ a_p \end{array}\right] = \sigma_\varepsilon^2 \left[\begin{array}{c} c(0) \\ c(1) \\ \vdots \\ c(q) \\ \hline 0 \\ \vdots \\ 0 \end{array}\right] \qquad (14)$$

- Difficult solution for model coefficients of a general ARMA(p, q) process, because $c(k)$ depends nonlinearly on them (see previous slide) $\rightarrow$ nonlinear system of equations.

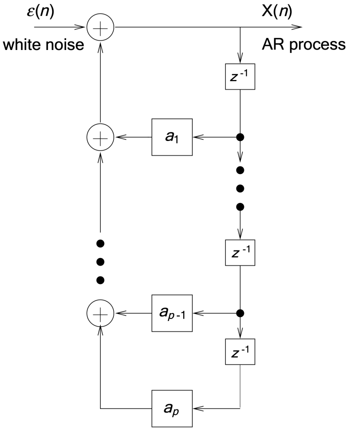
- The system of equations becomes linear for $q = 0$, since in that case $c(k) = |b_0|^2 \delta(k)$.

Autoregressive (AR) processes

If $q = 0 \rightarrow$ ARMA$(p, 0) \rightarrow$ **autoregressive (AR) process** of order p: AR(p)

$$X(n) + \sum_{k=1}^{p} a_k X(n-k) = \varepsilon(n) \qquad \Rightarrow \qquad S(\omega) = \frac{\sigma_\varepsilon^2}{\left|1 + \sum_{k=1}^{p} a_k \mathrm{e}^{-\jmath\omega k}\right|^2} \tag{15}$$

Also known as **all-pole model**.



AR processes — examples

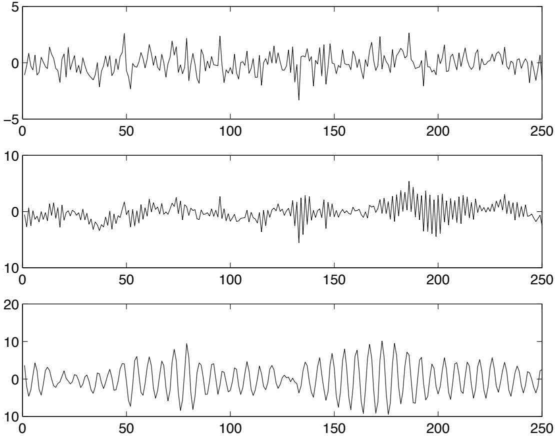


Figure: (Top) White noise input.
(Middle) AR process with $\mathbf{a} = [1, 0.1, -0.8]$.
(Bottom) AR process with $\mathbf{a} = [1, -0.975, 0.95]$.

AR processes — Yule-Walker equations

For AR(p) processes, assuming that $b_0 = 1$, the Yule-Walker equations (13) simplify to:

$$r(k) + \sum_{m=1}^{p} a_m r(k-m) = \sigma_\varepsilon^2 \delta(k) \qquad k \geq 0$$

and their matrix form (14) becomes

$$\boxed{\begin{bmatrix} r(0) & r(-1) & \cdots & r(-p) \\ r(1) & r(0) & \cdots & r(-p+1) \\ \vdots & \vdots & \ddots & \vdots \\ r(p) & r(p-1) & \cdots & r(0) \end{bmatrix} \begin{bmatrix} 1 \\ a_1 \\ \vdots \\ a_p \end{bmatrix} = \begin{bmatrix} \sigma_\varepsilon^2 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \quad (16)}$$

AR processes — Yule-Walker equations (cont'd)

The **last p equations** of (16) yield:

$$\underbrace{\begin{bmatrix} r(1) \\ \vdots \\ r(p) \end{bmatrix}}_{\mathbf{r}_p} + \underbrace{\begin{bmatrix} r(0) & r(-1) & \cdots & r(-p+1) \\ \vdots & \vdots & \ddots & \vdots \\ r(p-1) & r(p-2) & \cdots & r(0) \end{bmatrix}}_{\mathbf{R}_p} \underbrace{\begin{bmatrix} a_1 \\ \vdots \\ a_p \end{bmatrix}}_{\boldsymbol{\theta}} = \begin{bmatrix} 0 \\ \vdots \\ 0 \end{bmatrix}$$

that is:

$$\boxed{\mathbf{R}_p\boldsymbol{\theta} = -\mathbf{r}_p \qquad (17)}$$

The **first equation** of (16) yields:

$$\boxed{\sigma_\varepsilon^2 = r(0) + \mathbf{r}_p^{\mathrm{H}}\boldsymbol{\theta} \qquad (18)}$$

where $(\cdot)^{\mathrm{H}}$ denotes the Hermitian (conjugate-transpose) operator.

Yule-Walker method for AR models

Yule-Walker (or autocorrelation) method

1. Select the **model order** p.
2. Compute suitable **ACS estimates** $\{\hat{r}(k)\}_{k=0}^{p}$ from the available samples $\{x(n)\}_{n=0}^{N-1}$.
3. Build **autocorrelation matrix** $\mathbf{R}_p$.
4. Solve **Yule-Walker linear system** (17) to estimate the AR model coefficients $\hat{\boldsymbol{\theta}}$.
5. Estimate the **innovation variance** (or modeling error) $\hat{\sigma}_\varepsilon^2$ through eqn. (18).
6. Compute the AR **PSD estimate** (15).

Remarks

- Choosing the model order is a difficult problem $\rightarrow$ seen later in the chapter
- Biased ACS estimates are preferred: $\mathbf{R}_p$ positive definite $\rightarrow$ lower variance in PSD estimate
- Inverting matrix $\mathbf{R}_p \rightarrow O(p^3)$ products-divisions (Gaussian elimination)...
- ... but $\mathbf{R}_p$ is a Hermitian Toeplitz matrix
  $\rightarrow$ computationally efficient algorithm with $O(p^2)$ prod.-div.: **Levinson-Durbin recursion**
- ACS estimates $\hat{r}(k)$: $O(Np)$ operations
  $\rightarrow$ cost reduction of Levinson-Durbin algorithm may be negligible if $N \gg p$.

Link with linear prediction

Linear prediction: estimate or predict $x(n)$ from a linear combination of previous samples

$$\hat{x}(n) = -\sum_{k=1}^{p} a_k x(n-k)$$

Criterion: minimize the mean square error (MSE)

$$J_{\text{MSE}} = \mathrm{E}\{|e(n)|^2\} \tag{19}$$

- $e(n) \stackrel{\text{def}}{=} \hat{x}(n) - x(n)$: **linear prediction error**
- $\{a_k\}_{k=1}^{p}$: **linear prediction coefficients**
- $A(z) = 1 + \sum_{k=1}^{p} a_k z^{-k}$: **linear prediction filter**

Linear prediction and AR modeling

The linear prediction coefficients minimizing the MSE criterion (19) are given by the solution of Yule-Walker equations (17) for AR modeling.

The linear prediction error is the innovation process: $e(n) = \varepsilon(n)$.

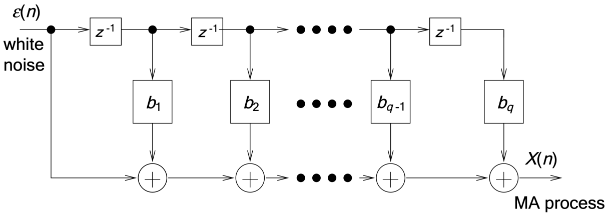
The minimum MSE is given by the innovation variance (18): $\min J_{\text{MSE}} = \sigma_\varepsilon^2$.

Moving average (MA) processes

If $p = 0 \rightarrow$ ARMA$(0, q) \rightarrow$ **moving average (MA) process** of order q: MA(q)

$$x(n) = \sum_{k=0}^{q} b_k \varepsilon(n-k) \qquad \Rightarrow \qquad S(\omega) = \sigma_\varepsilon^2 \left| \sum_{k=0}^{q} b_k e^{-\jmath \omega k} \right|^2$$

Also known as **all-zero model**.



MA processes — examples

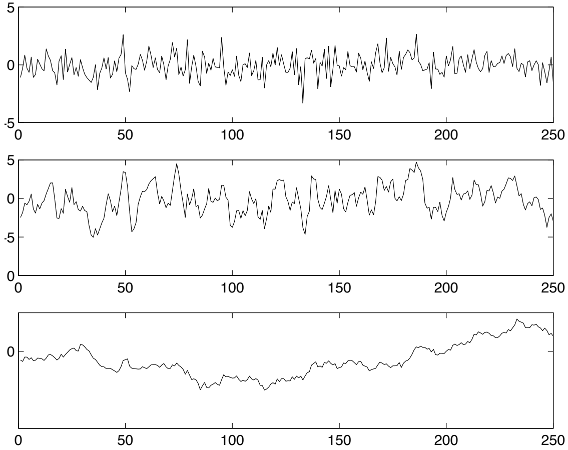


Figure: (Top) White noise input.
(Middle) MA process with $\mathbf{b} = [1, 1, 1, 1]/4$.
(Bottom) MA process with $\mathbf{b} = \underbrace{[1, 1, \ldots, 1]}_{\text{100 entries}}/100$.

MA processes — Yulker-Walker equations

For MA(q) processes:

$$h(k) = \begin{cases} b_k & 0 \leq k \leq q \\ 0 & \text{otherwise} \end{cases}$$

Hence, the ACS (13) becomes

$$r(k) = \begin{cases} \sigma_\varepsilon^2 \sum_{m=0}^{q-k} b_{m+k} b_m^* & 0 \leq k \leq q \\ r^*(-k) & k < 0 \\ 0 & |k| > q. \end{cases}$$

Because $r(k) = 0$ for $|k| > q$, a **natural PSD estimate** is:

$$\hat{S}(\omega) = \sum_{k=-q}^{q} \hat{r}(k) \mathrm{e}^{-\jmath\omega k}$$

where $\hat{r}(k)$ is a suitable ACS estimate.

- Equivalent to **Blackman-Tuckey** with rectangular window of length $2q+1$ [Chap. 2]
- Biased PSD estimate if process $X(n)$ is not actually governed by an MA(q) model.

Model order selection

Important step in rational spectrum modeling: estimate model order (p, q) from observed data.

Selecting model order p in AR modeling

- if too small $\rightarrow$ smoothed spectrum, poor resolution
- if too large $\rightarrow$ spurious peaks (spectral line splitting).

General approach

- increase model order until modeling error is minimized
- *problem:* error is a monotonically nonincreasing function of p
- *idea:* incorporate a penalty term that increases with model order p
  $\rightarrow$ select p minimizing the criterion:

$$\boxed{C(p) = N \log \sigma_\varepsilon^2(p) + f(N)p}$$

N : data record length

$\sigma_\varepsilon^2(p)$: modeling error for model order p

$f(N)$: constant that may depend on N.

Model order selection (cont'd)

Akaike information criterion (AIC)

$$\mathrm{AIC}(p) = N \log \sigma_\varepsilon^2(p) + 2p$$

- estimated p typically too small for non-AR processes
- it tends to overestimate p as N increases

Minimum description length (MDL)

$$\mathrm{MDL}(p) = N \log \sigma_\varepsilon^2(p) + (\log N)p$$

- consistent model order estimator: $\hat{p}_{\mathrm{MDL}} \underset{N\to\infty}{\longrightarrow} p$

Akaike's final prediction error (FPE)

$$\mathrm{FPE}(p) = \sigma_\varepsilon^2(p) \frac{N+p+1}{N-p-1}$$

Remarks

- no criterion works particularly well for short data sequences
- generally they should just be used as 'indicators' of the model order
- prediction error $\sigma_\varepsilon^2(p)$ depends on modeling technique.

Summary

- Exploit prior knowledge about the random process to be analyzed
- **General approach**
  1. Select an appropriate model for process under analysis
  2. Estimate the model parameters from the available data
  3. Estimate the PSD by incorporating the estimated parameters into the parametric form for the PSD
- **PSD with rational function structure** — 3 models:
  - autoregressive moving average (ARMA): spectrum with poles and zeros
  - autoregressive (AR): all-pole model
  - moving average (MA): all-zero model
- Model parameters can be estimated using **Yule-Walker equations**
  - estimate ACS $\rightarrow$ build autocorrelation matrix $\rightarrow$ solve a linear system of equations
- Several criteria for **model order selection**
  - AIC, MDL, FPE, ...
- Performance of parametric approach depends on fitness of model to the process being analyzed.