Lecture 26

Modeling (1): Time Series Prediction

Outline

- Time series models
- Linear Time Series Models
  - Moving Average Model
  - Auto-regressive Model
  - ARMA model
- Nonlinear Time Series Estimation
- Applications

Time Series

- What is a time series?
  - A scalar or vector-valued function of time indices
- Examples:
  - Stock prices
  - Temperature readings
  - Measured signals of all kinds
- What is the use of a time series?
  - Prediction of future time series values based on past observations

Modeling of a time series

Values of a time series at successive time indices are often correlated. Otherwise, prediction is impossible.

Most time series can be modeled mathematically as a wide-sense stationary (WSS) random process. The statistical properties do not change with respect to time.

Some time series exhibits chaotic nature. A chaotic time series can be described by a deterministic model but behaves as if it is random, and highly un-predictable.

Time Series Models

- Most time series are sampled from continuous time physical quantities at regular sampling intervals. One may label each such interval with an integer index. E.g.

  {y(t); t = 0, 1, 2, …}.

- A time series may have a starting time, say t = 0. If so, it will have an initial value.
- In other applications, a time series may have been run for a while, and its past value can be traced back to t= –∞.
- Notations
  - y(t): time series value at present time index t.
  - y(t–1): time series value one unit sample interval before t.
  - y(t+1): the next value in the future.
- Basic assumption
  - y(t) can be predicted with certainly degree of accuracy by its past values {y(t–k); k > 0} and/or the <u>present</u> and past values of <u>other</u> time series such as {x(t–m); m ≥ 0}

Time Series Prediction

- Problem Statement

  Given {y(i); i = t–1, …} estimate $y(t+t_o)$, $t_o \geq 0$ such that
  $$C = E\left[\left\|e(t+t_o)\right|^2\right]$$
  $$= E\left[\left\|y(t+t_o) - \hat{y}(t+t_o)\right|^2\right]$$
  is minimized.

  when $t_o = 0$, it is called a 1-step prediction.

  Sometimes, additional time series {u(i); i = **t,** t–1, …} may be available to aid the prediction of y(i)

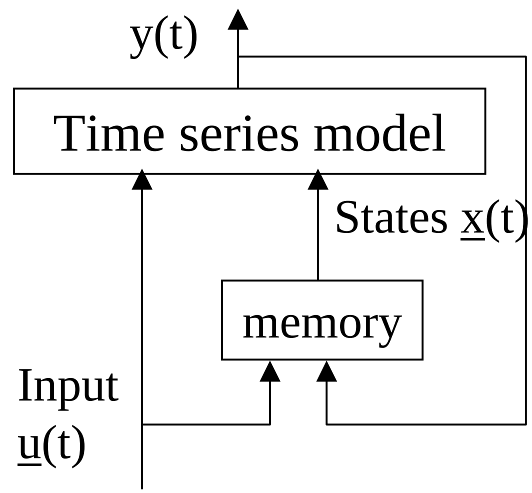
- The estimate of $y(t+t_o)$ that minimizes C is the conditional expectation given past value and other relevant time series.
  $$\hat{y}(t+t_o)$$
  $$= E\left[y(t+t_o) \mid \{y(i)\}_{i=-\infty}^{t-1}, \{u(i)\}_{i=-\infty}^{t}\right]$$
- This conditional expectation can be modeled by a linear function (linear time series model) or a nonlinear function.

A Dynamic Time Series Model

- State {x(t)}
  - Past values of a time series can be summarized by a finite-dimensional state vector.
- Input {u(t)}
  - Time series that is not dependent on {y(t)}

$$y(t) = f(\underline{x}(t), \underline{u}(t))$$

- The mapping is a dynamic system as y(t) depends on both present time inputs as well as past values.



$\underline{x}(t)$ = [x(t) x(t–1) … x(t –p)]
consists past values of {y(t)}
$\underline{u}(t)$ = [u(t) u(t–1) … u(t –q)]

Linear Time Series Models

- y(t) is a linear combination of x(t) and/or u(t).
- White noise random process model of input {u(t)}:
  - E(u(t)) = 0
  - E{u(t)u(s)} = 0 if t $\neq$ s; = σ^2 if t = s.
- Three popular linear time series models:

1. Moving Average (MA) Model:

$$y(t) = \sum_{m=0}^{M} b(m)u(t-m)$$

2. Auto-Regressive (AR) Model:

$$y(t) = \sum_{n=1}^{N} a(n)y(t-n) + u(t)$$

3. Moving Average, Auto-regressive (ARMA) Model:

$$y(t) = \sum_{n=1}^{N} a(n)y(t-n) + \sum_{m=0}^{M} b(m)u(t-m)$$

Moving Average Model

$$y(t)=\sum_{m=0}^{M} b(m)u(t-m)$$

- Cross correlation function

$$R_{yu}(i)=E\big(y(t)u(t-i)\big)$$
$$=\sum_{m=0}^{M} b(m)E\big(u(t-m)u(t-i)\big)$$
$$=\begin{cases}\sigma^2 b(i) & \text{if } 0\le i\le M,\\ 0 & \text{otherwise.}\end{cases}$$

- Auto-correlation function:

$$R_y(k)=E\big(y(t)y(t-k)\big)$$
$$=\begin{cases}\sigma^2\sum_{m=0}^{M-|k|} b(m)b(m+k) & |k|\le M;\\ 0 & |k|>M.\end{cases}$$

- An MA model is recognized by the finite number of non-zero auto-correlation lags.
- {b(m)} can be solved from $\{R_y(k)\}$ using optimization procedure.
- Example: If {u(t)} is the stock price, then

$$y(t)=\frac{u(t)+u(t-1)+u(t-2)+u(t-3)}{4}$$

is a moving average model – An average that moves with respect to time!

Finding MA Coefficients

- Problem: Given a MA time series {y(t)}, how to find {b(m)} withtout knowing {u(t)}, except the knowledge of σ^2?
- One way to find the MA model coefficients {b(m)} is *spectral factorization*.
- Consider an example:
  $$y(t) = b(0)u(t) + b(1)u(t-1)$$
- Given {y(t); t = 1, …, T}, estimate auto-correlation lag
  $$\hat{R}(k) = \frac{1}{T-k}\sum_{t=1}^{T-k} y(t)y(t+k)$$
- For this MA(1) model, R(k)≈0 for k > 1.

- Power spectrum
  $$S_{YY}(z) = B(z)B(1/z^*) = \sum_{k=-\infty}^{\infty} R(k)z^{-1}$$
  $$= \frac{b(0)z + b(1)}{z} \cdot \frac{b(0) + b(1)z^*}{1}$$

	B(z)	*B(1/z*)*
zeros	–b(1)/b(0)	–(b(1)/b(0))*
poles	0	∞

- Spectral factorization:
  - Compute S(z) from {R(k)} and factorize its zeros and poles to construct B(z)
- Or comparing the coefficients of polynomial and solve a set of nonlinear equations.

Auto-Regressive Model

$$y(t)=\sum_{n=1}^{N}a(n)y(t-n)+u(t)$$

$$E\big(y(t)\big)=0,\quad E\big(y(t)u(t)\big)=\sigma^2$$

Hence,

$$R_y(k)=E\big(y(t)y(t-k)\big)$$

$$=E\big(u(t)y(t-k)\big)+\sum_{n=1}^{N}a(n)E\big(y(t-n)y(t-k)\big)$$

$$=\sigma^2\cdot\delta(t-k)+\sum_{n=1}^{N}a(n)R_y(k-n)$$

This leads to the Yule-Walker equation:

$$\mathbf{Ra}=\sigma^2[1\quad 0\quad\cdots\quad 0]^T=\sigma^2\mathbf{e}$$

or in matrix form:

$$\underbrace{\begin{bmatrix} R(0) & R(-1) & \cdots & R(-N)\\ R(1) & R(0) & \ddots & R(-N+1)\\ \vdots & \ddots & \ddots & \vdots\\ R(N) & R(N-1) & \cdots & R(0)\end{bmatrix}}_{\text{R}}\underbrace{\begin{bmatrix}1\\ -a(1)\\ \vdots\\ -a(N)\end{bmatrix}}_{\text{a}}=\underbrace{\begin{bmatrix}\sigma^2\\ 0\\ 0\\ 0\end{bmatrix}}_{\sigma^2\mathbf{e}}$$

- $R_y(m)$ is replaced with R(m) to simplify notations.
- **R** is a Töeplitz matrix and is positive definite. Fast Cholesky factorization algorithms such as the Levinson algorithm can be devised to solve the Y-W equation effectively.

Auto-Regressive, Moving Average (ARMA) Model

$$y(t) = \sum_{n=1}^{N} a(n) y(t-n) + \sum_{m=0}^{M} b(m) u(t-m)$$

- A combination of AR and MA model.
- Denote $v(t) = \sum_{m=0}^{M} b(m) u(t-m)$

  Then

  $$E\big(y(t-s)v(t)\big)$$
  $$= \sum_{m=0}^{M} b(m) E\big(y(t-s)u(t-m)\big)$$
  $$= 0 \quad \text{for } s > M$$

Thus,

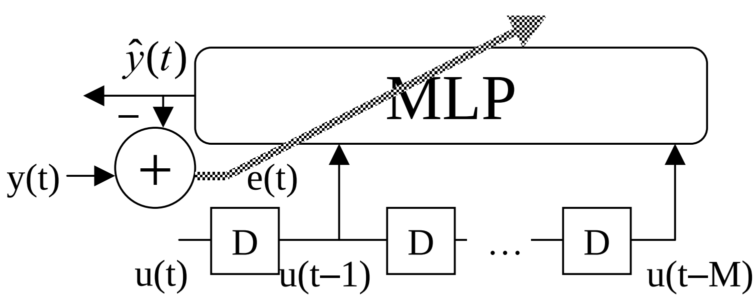
$$R_y(k) = E\big(y(t)\cdot y(t-k)\big)$$
$$= E\left(\left(\sum_{n=1}^{N} a(n) y(t-n) + \sum_{m=0}^{M} b(m) u(t-m)\right)\cdot y(t-k)\right)$$
$$= \sum_{n=1}^{N} a(n) R_y(k-n) + \sum_{m=0}^{M} b(m) E\big(y(t-k)u(t-m)\big)$$
$$= \sum_{n=1}^{N} a(n) R_y(k-n) \quad \text{if } k > M.$$

Higher order Y-W equation

$$\begin{bmatrix} R(k-1) & R(k-2) & \cdots & R(k-N) \\ R(k) & R(k-1) & \cdots & R(k-N+1) \\ \vdots & \vdots & \ddots & \vdots \\ R(k+L-1) & R(k+L-2) & \cdots & R(k+L-N) \end{bmatrix} \begin{bmatrix} a(1) \\ \vdots \\ a(N) \end{bmatrix} = \begin{bmatrix} R(k) \\ R(k+1) \\ \vdots \\ R(k+L) \end{bmatrix}$$

Nonlinear Time Series Model

- *f(x(t), u(t))* is a nonlinear function or mapping:
  - *MLP*
  - *RBF*
- *Time Lagged Neural Net (TLNN)*
  - The input of a MLP network is formed by a time-delayed segment of a time series.



- A neuronal filter