Lecture 14 (March 7, 2019)

Correlations amongst prediction errors

Covariance between prediction errors exists, even though modeling errors are uncorrelated.

$$E[\hat{e}_t(l)\,\hat{e}_{t+j}(l)] = E[(G_0 a_{t+l} + G_1 a_{t+l-1} + \dots + G_{l-1} a_{t+1}) \cdot (G_0 a_{t+j+l} + G_1 a_{t+j+l-1} + \dots + G_{l-1} a_{t+j+1})]$$

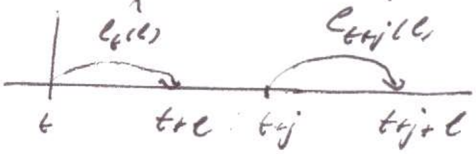
For $j < l$, we have "overlap" between these 2 prediction error terms

Then

$$E[\hat{e}_t(l)\, \hat{e}_{t+j}(l)] = \sigma_a^2 [G_0 G_j + G_1 G_{j+1} + \dots + G_{l-j-1} G_{l-1}]$$

Otherwise, if $j \geq l$, $E[\hat{e}_t(l)\, \hat{e}_{t+j}(l)] = 0$

Furthermore

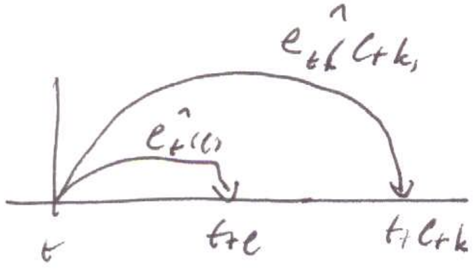


$$E[\hat{e}_t(l)\, \hat{e}_t(l+k)] =$$

$$= E[(G_0 a_{t+l} + G_1 a_{t+l-1} + \dots + G_{l-1} a_{t+1})(G_0 a_{t+l+k} + \dots$$

$$+ G_k a_{t+l} + G_{k+1} a_{t+l-1} + \dots + G_{k+l-1} a_{t+1}) =$$

$$= \sigma_a^2 (G_0 G_k + G_1 G_{k+1} + \dots + G_{l-1} G_{k+l-1})$$



Exponentially Weighted Moving Average (EWMA) Models

Parallelly with theoretical models, empirical models were pursued in business & industry

* We just learnt that ARMA models capture dynamics of the system & used that to predict what will happen in the future.

* What if I just want a quick way to predict next sample, without getting into details of the underlying dynamic model?

- $X_{t+1} \approx \frac{1}{N} \sum_{j=0}^{N-1} X_{t-j}$ ← one intuitive approximation

- Perhaps, as I go farther back in the past, I can taper off the significance of each sample

$$\hat{X}_t(1) = \sum_{j=0}^{\infty} w_j X_{t-j} \qquad w_j = c \cdot \theta^j$$

for some $|\theta| < 1$

$$\sum w_j = 1 \Rightarrow c \cdot \frac{1}{1-\theta} = 1 \Rightarrow$$

$$\Rightarrow c = (1-\theta) \Rightarrow w_j = (1-\theta)\theta^j$$

Then

$$\hat{X}_t(1) = \sum_{j=0}^{\infty} (1-\theta)\theta^j X_{t-j}$$ ← This is EWMA

Let $\lambda = 1-\theta$. Then

$$\hat{X}_t(1) = \sum_{j=0}^{\infty} \lambda(1-\lambda)^j X_{t-j}$$

λ - exponentially smoothing constant (parameter)

λ small => weights change slowly and die off slowly

λ large => —— " —— rapidly (taper off quickly)

Reasons for EWMA

(a) Intuitively appealing & simple

(b) No need to store the entire past to make forecasts!

$$\hat{X}_t(1) = \sum_{j=0}^{\infty} \lambda(1-\lambda)^j X_{t-j} = \lambda X_t + \sum_{j=1}^{\infty} \lambda(1-\lambda)^j X_{t-j}$$

$$\hat{X}_{t-1}(1) = \sum_{j=0}^{\infty} \lambda(1-\lambda)^j X_{t-j-1} = \frac{1}{1-\lambda} \sum_{j=0}^{\infty} \lambda(1-\lambda)^{j+1} X_{t-(j+1)}$$

$$= \frac{1}{1-\lambda} \sum_{j=1}^{\infty} \lambda(1-\lambda)^j X_{t-j} .$$ (see now $\hat{X}_t(1)$)

$$\hat{X}_t(1) = \lambda X_t + (1-\lambda) \hat{X}_{t-1}(1) \qquad (*)$$

=> all I need is the new observation of X_t and the previous forecast of $\hat{X}_{t-1}(1)$ (it's forecast of X_t at time $t-1$).

Relation of EWMA models with ARMA Models

Let's observe an ARMA(1,1) model with $\phi_1 = 1$

$$X_t - X_{t-1} = a_t - \theta_1 a_{t-1} \Rightarrow$$

$$X_{t+1} - X_t = a_{t+1} - \theta_1 a_t \qquad / E[\cdot | t]$$

$$\hat{X}_t(1) = X_t - \theta_1 a_t = X_t - \theta_1 (-\hat{X}_{t-1}(1) + X_t)$$

$$= (1-\theta_1) X_t + \theta_1 \hat{X}_{t-1}(1)$$

It's the same as what you get with EWMA in (*), using $\lambda = 1 - \theta_1$ and (or) $\theta_1 = 1 - \lambda$.

Another way to demonstrate equivalence between EWMA and ARMA(1,1) models is

$$\hat{X}_t(1) = \sum_{j=0}^{\infty} \lambda (1-\lambda)^j X_{t-j} \Rightarrow$$

$$X_{t+1} = \hat{X}_t(1) + a_{t+1} = a_{t+1} + \sum_{j=0}^{\infty} \lambda (1-\lambda)^j X_{t-j} \Rightarrow$$

$$a_{t+1} = X_{t+1} - \sum_{j=0}^{\infty} \lambda(1-\lambda)^j X_{t-j} \Rightarrow$$

$$a_t = X_t - \sum_{j=0}^{\infty} \lambda(1-\lambda)^j X_{t-j-1} = X_t - \frac{\lambda}{1-\lambda} \sum_{j=0}^{\infty} (1-\lambda)^{j+1} X_{t-j-1}$$

$$= \left(1 - \frac{\lambda}{1-\lambda} \sum_{j=0}^{\infty} (1-\lambda)^{j+1} B^{j+1}\right) X_t$$

$$= \left(1 - \frac{\lambda}{\cancel{1-\lambda}} \frac{\cancel{(1-\lambda)} B}{1-(1-\lambda)B}\right) X_t = a_t \Rightarrow$$

$$\Rightarrow \frac{1-B}{1-(1-\lambda)B} X_t = a_t$$

$$\Rightarrow X_t - X_{t-1} = a_t - \underset{\substack{\uparrow \\ \theta_1}}{(1-\lambda)} a_{t-1}.$$

In essence, we see that a special case of ARMA is EWMA. I can now use EWMA to get ARMA!

Remember, I used "dynamical" reasoning (based in dynamics) to get different orders of ARMA models. Now, I will use EWMA & its "prognostic" connotations to get different ARMA models!

Obviously, EWMA leads to

$$X_t = a_t + \sum_{j=0}^{\infty} \lambda(1-\lambda)^j X_{t-j-1}$$

which corresponds to an ARMA(1,1) model

$$X_t - X_{t-1} = a_t - \theta_1 a_{t-1} \quad \text{with } \theta_1 = 1-\lambda.$$

$$X_t = a_t + \sum_{j=0}^{\infty} (1-\theta_1)\theta_1^j X_{t-j-1}$$

ii) Let's relax the first coefficient and say that

$$X_t = a_t + \sum_{j=0}^{\infty} (\phi_1 - \theta_1)\theta_1^j X_{t-j-1}$$

That leads to

$$X_t - \phi_1 X_{t-1} = a_t - \theta_1 a_{t-1}$$

In essence, I relaxed the need for all weights in EWMA model to sum up to 1!

(iii) $$I_1 = \phi_1 - \theta_1$$

$$I_j = (\phi_2 + \phi_1\theta_1 - \theta_1^2)\,\theta_1^{j-2} \quad j \geq 2$$

will lead to

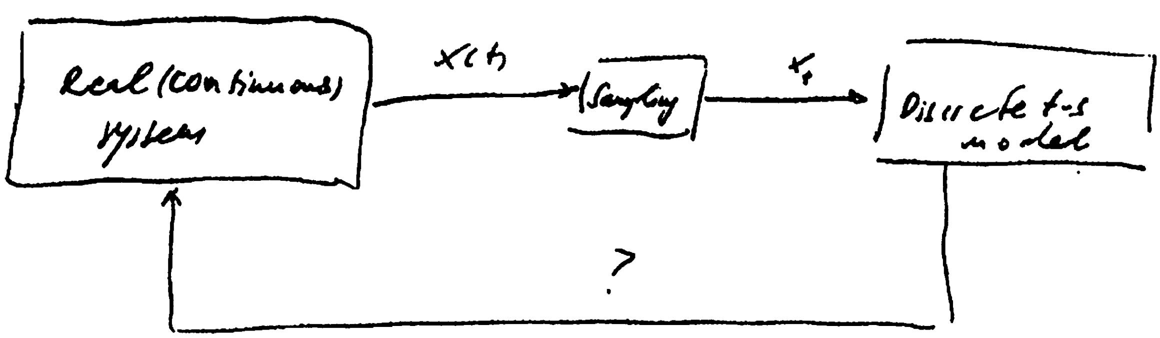
$$X_t - \phi_1 X_{t-1} - \phi_2 X_{t-2} = a_t - \theta_1 a_{t-1}$$

For more complex ARMA models, I must be more elaborate with initial conditions on the inverse function.

Uniform Sampling of Continuous Time-Series
(Chapter 6)

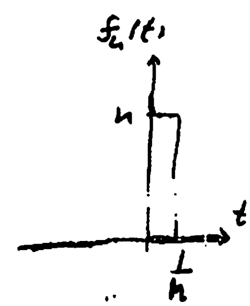
Actual physical phenomena are continuous in nature & discrete time-series are sampled versions of what actually happens in continuous time

Can we observe the discrete-time time-series & make judgement about the continuous-time (actual, physical) system?



To find this connection, let's review continuous-time Dirac's delta function

Def. $f_n(t) = \begin{cases} n, & |t|<1/n \\ 0, & \text{otherwise} \end{cases}$ $\qquad \delta(t) = \lim_{n\to\infty} f_n(t)$



Properties of $\delta(t)$

$$\delta(t) = \begin{cases} \infty, & t=0 \\ 0, & t \neq 0 \end{cases} \qquad \int_{-\varepsilon}^{\varepsilon} \delta(t)\,dt = 1$$

$$\int_{-\infty}^{t} \delta(u)\,du = \begin{cases} 1 & t>0 \\ 0 & t\leq 0 \end{cases} \qquad \int_{-\infty}^{\infty} f(u)\,\delta(t-u)\,du = f(t)$$

$\Rightarrow \delta(t) = \frac{d}{dt} S(t) \leftarrow$ step function

Integration by parts can give $\int_{-\infty}^{\infty} f(t-u)\,\delta^{(k)}(u)\,du = (-1)^k f^{(k)}(t)$

Review of Linear Ordinary Differential Equations with Constant Coefficients

i) Linear Homogeneous Differential Eqns with Constant Coefficients

$$X^{(n)}(t) + \alpha_{n-1} X^{(n-1)}(t) + \dots + \alpha_0 X(t) = 0$$

$$(D^n + \alpha_{n-1} D^{n-1} + \dots + \alpha_0)\, X(t) = 0$$

Solution must be of the shape

$$X(t) = C_1 e^{\lambda_1 t} + C_2 e^{\lambda_2 t} + \dots + C_n e^{\lambda_n t}$$

where λ_i, $i=1,2,\dots,n$ are roots of the characteristic polynomial

$$s^n + \alpha_{n-1} s^{n-1} + \dots + \alpha_0 = 0$$

and constants $C_1, C_2, \dots, C_n$ are obtained from the initial values of the function $x(0), x'(0), \dots, x^{(n-1)}(0)$

First Order H.O.D.E.

$$x'(t) + \alpha_0 x(t) = 0$$

Physically: changes of signals are proportional to signal values!

$$x(t) = C_1 e^{-\alpha_0 t} = C_1 e^{-\frac{t}{\tau}} \quad \text{where } \tau = \frac{1}{\alpha_0} \text{ - time constant}$$

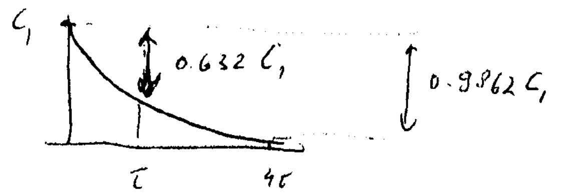
$\alpha_0 > 0 \Rightarrow x(t) \to 0$ (stable system)

$\alpha_0 = 0 \Rightarrow x(t) = C_1$ (marginally stable system)

$\alpha_0 < 0 \Rightarrow x(t) \to \infty$ (unstable system)

Large α_0 (small τ) $\Rightarrow$ ~~fa~~ rapid decay (quick system)

Small α_0 (large τ) $\Rightarrow$ slow decay (slow system)



General Linear ODE-s with Constant Coefficients

$$X^{(n)}(t) + \alpha_{n-1} X^{(n-1)}(t) + \dots + \alpha_0 X(t) = \beta_m U^{(m)}(t) + \beta_{m-1} U^{(m-1)}(t) + \dots + \beta_0 U(t)$$

where $m < n$ for causal systems

Solutions of these eqns can be found as

$$X(t) = \int_{-\infty}^{\infty} G(\tau) U(t-\tau)\, d\tau \underset{\uparrow}{=} \int_{0}^{\infty} G(\tau) U(t-\tau)\, d\tau$$

for a causal system

Where G is the impulse response of this system (this formula holds only when all initial conditions are 0 - if not true, one needs some modifications here, but that becomes control theory already).

Hence

$$G^{(n)}(t) + \alpha_{n-1} G^{(n-1)}(t) + \dots + \alpha_0 G(t) = \beta_m \delta^{(m)}(t) + \beta_{m-1} \delta^{(m-1)}(t) + \dots + \beta_0 \delta(t)$$

How to find $G(t)$?

$\boxed{\text{Thm}}$ For a Lin. Differential Equ with constant coeff.

$$X^{(n)}(t) + \alpha_{n-1} X^{(n-1)}(t) + \dots + \alpha_0 X(t) = V(t)$$

$G(t)$ can be found by solving homogeneous ODE

$$G^{(n)}(t) + \alpha_{n-1} G^{(n-1)}(t) + \dots + \alpha_0 G(t) = 0$$

with initial conditions

$$G(0) = 0; \quad G'(0) = 0; \; \dots; \; G^{(n-1)}(0) = 1$$

(this is for $t \geq 0$; for $t<0$, $G(t)=0$, since it's a causal system).

Proof: Easiest using Laplace transforms - please feel free to discuss with me this proof, so we can avoid turning this into a controls class

Ex. For a first order system:

$\dot{X}(t) + \alpha_0 (t) = u(t)$ $\qquad$ $\dot{G}(t) + \alpha_0 G(t) = 0$

$G(0) = 1 \Rightarrow$

$\Rightarrow G(t) = e^{-\alpha_0 t}$

Let's now swith gears & observe discrete version of differential equs $\rightarrow$ difference equations.

AR(1) model can be seen as a 1st order ~~[illegible]~~

Homogeneous Difference Eqn.

$$X_t - \phi_1 X_{t-1} = 0$$

Driven by white noise a_t as

$$X_t - \phi_1 X_{t-1} = a_t$$

Now, before introducing "stochastic continuous ODE and systems, I need to define continuous-time white noise.

Def) $Z(t)$, $t \in R$ is a white noise stochastic process iff.

i) $E[Z(t)] = 0$ ii) $E[Z(t) Z(t-s)] = cov(Z(t), Z(t-s))$
$= \sigma_z^2 \delta(s)$

Hence, it's a stochastic process where each sample is independent of samples infinitely close to it - does NOT really exist in nature, but very convenient for analysis because of the orthogonality property, i.e. because of property ii).

$\boxed{\text{Def}}$ Stochastic linear ODE with constant coefficients is an equation of the form

$$X^{(n)}(t)+\alpha_{n-1}X^{(n-1)}(t)+\dots+\alpha_0 X(t)=$$
$$=\beta_m Z^{(m)}(t)+\beta_{m-1}Z^{(m-1)}(t)+\dots+\beta_0 Z(t)$$

where $Z(t)$ is a Gaussian white noise process.

Solution of this equation for any trace (realization) of the random process $Z(t)$ is

$$X(t)=\int_0^\infty G(\tau)Z(t-\tau)\,d\tau \qquad (**)$$

where $G(\tau)$ is the impulse response of the system above, i.e. $G(\tau)$ is solution of the equation

$$G^{(n)}(t)+\alpha_{n-1}G^{(n-1)}(t)+\dots+\alpha_0 G(t)=$$
$$=\beta_m \delta^{(m)}(t)+\beta_{m-1}\delta^{(m-1)}(t)+\dots+\beta_0\delta(t)$$

where $\delta(t)$ is the impulse function.

Obviously, (**) denotes a stochastic process.

* $E[X(t)] = 0$ since $E[Z(t)] = 0$

* $\gamma(s) = E[X(t)\, X(t-s)] = E\left[\int_0^\infty\int_0^\infty G(s_1)\, Z(t-s_1)\, G(s_2)\, Z(t-s-s_2)\, ds_1\, ds_2\right]$

$$= \int_0^\infty\int_0^\infty G(s_1)\, G(s_2)\, E[Z(t-s_1)\, Z(t-s-s_2)]\, ds_1\, ds_2 =$$

$$= \int_0^\infty\int_0^\infty G(s_1)\, G(s_2)\, \delta(t-s_1-t+s+s_2)\, \sigma_z^2\, ds_1\, ds_2 =$$

$$= \int_0^\infty\int_0^\infty G(s_1)\, G(s_2)\, \delta(s+s_2-s_1)\, \sigma_z^2\, ds_1\, ds_2 =$$

$$= \sigma_z^2 \int_0^\infty G(s_2)\, ds_2 \int_0^\infty ds_1\, \underset{\underset{u}{\uparrow}}{G(s_1)}\, \delta(\underbrace{s+s_2}_{t} - \underset{\underset{u}{\uparrow}}{s_1}) =$$

$$= \sigma_z^2 \int_0^\infty G(s_2)\, G(s+s_2)\, ds_2$$

For a first order system, $X'(t) + \alpha_0 X(t) = u(t)$, impulse response is $G(s) = e^{-\alpha_0 s}$ and hence:

$$\gamma(s) = \sigma_z^2 \int_0^\infty e^{-\alpha_0 s_2}\, e^{-\alpha_0 (s+s_2)}\, ds_2 =$$

$$= \sigma_z^2 \int_0^\infty e^{-\alpha_0 s}\, e^{-2\alpha_0 s_2}\, ds_2 = \sigma_z^2\, \frac{e^{-\alpha_0 s}}{2\alpha_0}$$ //

Can we connect this to some discrete system?

Note: $X'(t) + \alpha_0 X(t) = Z(t)$ is often referred to as a stochastic autoregressive model of order 1 (label as $A(1)$)