Time Series Forecasting

Armine Hayrapetyan

Stochastic processes

A **discrete stochastic process** is a family of random variables structured as a sequence (finite or infinite) and has a discrete time index, denoted X_t. Discrete stochastic processes may model, for instance, the recorded daily high temperatures in Melbourne, Australia.

A **continuous stochastic process** is also a family of random variables, but is indexed by a continuous variable and denoted as $X(t)$. A commonly encountered continuous process is the Weiner Process, having nothing to do with hot dogs, but instead describing a particle's position as a function of time as it floats on the surface of a liquid (Brownian Motion). Another commonly encountered continuous process is the Poisson Process.

Ensembles and Realizations

When we acquire time series data in the field, we don't usually have the luxury of observing multiple trajectories.

We usually just have one sequentially observed data set and must infer the properties of the generating process from this single trajectory.

The usual terminology here is that an individual trajectory corresponds to a realization of a stochastic process. This is what we have been calling a time series. The set of all possible trajectories is called the **ensemble**.

A stochastic process is a rather complicated thing. To fully specify its structure we would need to know the joint distribution of the full set of random variables.

Strict stationarity

The problem here is that, for a given stochastic process we only have one observed time series. This is a bit like asking you whether a coin is fair by letting you toss it just once. It's just a silly question.

To get some traction, we often assume a property called stationarity. This will help us to essentially "pool" our data to in order to say something useful about the process.

We say a stochastic process is strictly stationary if

$$\begin{array}{c} \textit{Joint Distribution of} \\ X(t_1),\ X(t_2), \dots,\ X(t_k) \end{array} \quad \textit{is the same as} \quad \begin{array}{c} \textit{Joint Distribution of} \\ X(t_1+\tau),\ X(t_2+\tau), \dots,\ X(t_k+\tau) \end{array}$$

ACF

In this setting, if we shift our attention by a distance τ we will still have the same joint distribution. In particular, the joint distribution only depends upon our lag spacings. This is actually quite profound and yields beautifully simplifying results.

First, consider the simple case of $k = 1$. Then immediately

$$\begin{array}{ccc} \textit{The distribution of} & & \textit{The distribution of} \\ X(t_1) & \text{is the same as} & X(t_1 + \tau) \end{array}$$

The joint distribution depends only on the lag spacing, so

$$\textit{Autocovariance Function}:\ \gamma(t_1, t_2) = \gamma(t_2 - t_1) = \gamma(\tau)$$

Strict stationarity

It doesn't matter where you are sitting along the process, we just care about the distance between the random variables. Strict stationarity is a very strong condition, but as you can see it gives us some extremely simplifying results. Just stating the obvious, we can also scale the covariance to obtain the ***autocorrelation function***

$$\rho(\tau) \equiv \frac{\gamma(\tau)}{\gamma(0)}$$

Weak (also called second-order) Stationarity

A process is weakly stationary if

$$\text{Mean Function:} \quad \mu(t) = \mu$$

$$\text{Autocovarince Function:} \ \gamma(t_1, t_2) = \gamma(t_2 - t_1) = \gamma(\tau)$$

We are being efficient in our definition and are including the variance en passant. For an easy result, note that if $\tau = 0$ we also immediately have constant variance:

$$\gamma(0) = constant = E[(X(t) - \mu)(X(t) - \mu)]$$

Backward shift operator

Backward shift operator is defined as $BX_t = X_{t-1}$

- $B^2X_t = BBX_t = BX_{t-1} = X_{t-2}$
- $B^kX_t = X_{t-k}$

Random walk example:

$X_t = X_{t-1} + Z_t$; $X_t = BX_t + Z_t$; $(1 - B)X_t = Z_t$; $\phi(B)\ X_t = Z_t$ where $\phi(B) = 1 - B$

MA(q) process with a drift

$X_t = \mu + \beta_0 Z_t + \beta_1 Z_{t-1} + \cdots + \beta_q Z_{t-q},$

$X_t = \mu + \beta_0 Z_t + \beta_1 B^1 Z_t + \cdots + \beta_q B^q Z_t,$

$X_t - \mu = \beta(B) Z_\mathrm{t},$

Where

$\beta(B) = \beta_0 + \beta_1 (B) + \cdots + \beta_q B^\mathrm{q}$

AR(p) process

$X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + \cdots + \phi_p X_{t-p} + Z_\mathrm{t}$,

$X_t - \phi_1 X_{t-1} - \phi_2 X_{t-2} - \cdots - \phi_p X_{t-p} = Z_\mathrm{t}$,

$X_t - \phi_1 B X_t - \phi_2 B^2 X_t - \cdots - \phi_p B^p X_t = Z_\mathrm{t}$,

$\phi(B) X_t = Z_t$,

Where

$\phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \cdots - \phi_p B^p$.

Invertibility of a stochastic process

Two MA(1) models

- Model 1 - $X_t = Z_t + 2Z_{t-1}$,
- Model 2 - $X_t = Z_t + (½)\, Z_{t-1}$

Theoretical Auto Covariance Function of Model 1

γ(k) = $Cov(X_{t+k}, X_t) = Cov(Z_{t+k} + 2Z_{t+k-1}, Z_t + 2Z_{t-1})$

If $k > 1$, then $t + k - 1 > t$, so all Z's are uncorrelated, thus $\gamma(k) = 0$

If $k = 0$, then

$\gamma(0) = Cov(Z_t + 2Z_{t-1}, Z_t + 2Z_{t-1}) = Cov(Z_t, Z_t) + 4Cov(Z_{t-1}, Z_{t-1}) = \sigma_Z^{\ 2} + 4\sigma_Z^{\ 2} = 5\sigma_Z^{\ 2}$.

If k = 1, then

$\gamma(1) = Cov(Z_{t+1} + 2Z_t , Z_t + 2Z_{t-1}) = Cov(2Z_t , Z_t) = 2\sigma_Z^{\ 2}$

If k < 0, then

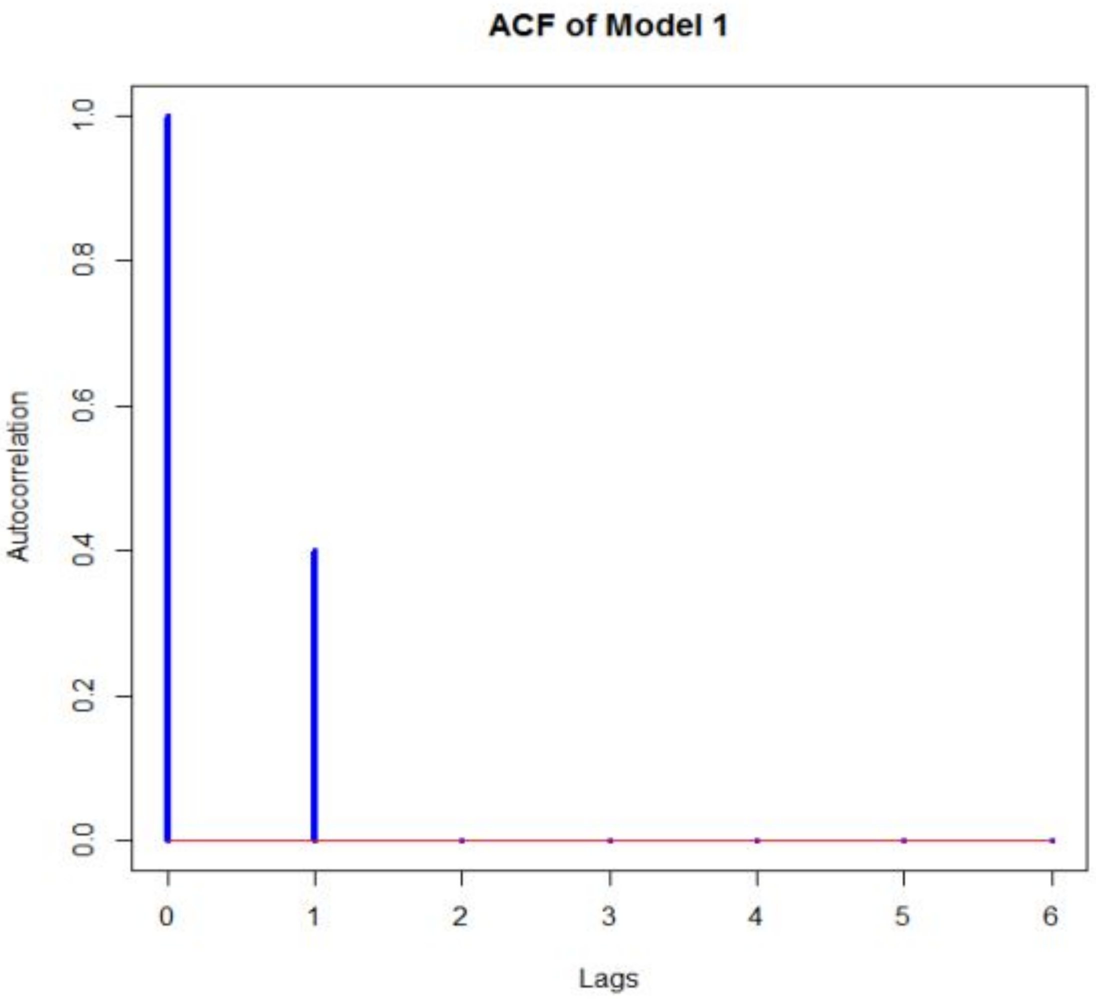
$\gamma(k) = \gamma(-k)$

Auto Covariance Function and ACF of Model 1

$$\gamma(k) = \begin{cases} 0, & k > 1 \\ 2\sigma_Z^2, & k = 1 \\ 5\sigma_Z^2, & k = 0 \\ \gamma(-k), & k < 0 \end{cases}$$

Then, since $\rho(k) = \frac{\gamma(k)}{\gamma(0)}$,

$$\rho(k) = \begin{cases} 0, & k > 1 \\ \frac{2}{5}, & k = 1 \\ 1, & k = 0 \\ \rho(-k), & k < 0 \end{cases}$$


ACF of Model 1
Autocorrelation
0.0
0.2
0.4
0.6
0.8
1.0
0
1
2
3
4
5
6
Lags

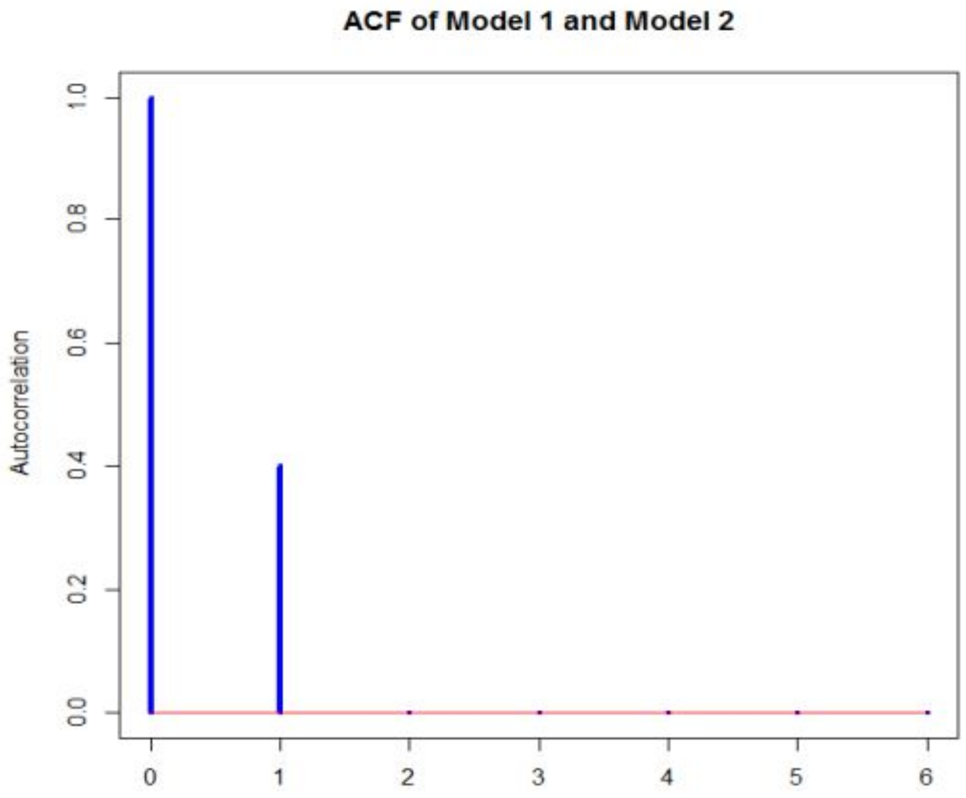
ACF of Model 2

$$\rho(1) = \frac{\gamma(1)}{\gamma(0)} = \frac{Cov\left[Z_{t+1} + \frac{1}{2}Z_t, Z_t + \frac{1}{2}Z_{t-1}\right]}{Cov[Z_t + \frac{1}{2}Z_{t-1},\ Z_t + \frac{1}{2}Z_{t-1}]} = \frac{\frac{1}{2}}{1 + \frac{1}{4}} = \frac{2}{5}.$$

Thus we obtain the same ACF:

$$\rho(k) = \begin{cases} 0, & k > 1 \\ \frac{2}{5}, & k = 1 \\ 1, & k = 0 \\ \rho(-k), & k < 0 \end{cases}$$

ACFs are same!



Invertibility - definition

X_t is a stochastic process.

Z_t is innovations, i.e., random disturbances or white noise.

$\{X_t\}$ is called **invertible**, if $Z_t = \sum_{k=0}^{\infty} \pi_k X_{t-k}$ where $\sum_{k=0}^{\infty} |\pi_k|$ is convergent.

Model 1 vs Model 2

- Model 1 is **not invertible** since

$$\sum_{k=0}^{\infty} |\pi_k| = \sum_{k=0}^{\infty} 2^k, \qquad Divergent$$

- Model 2 is **invertible** since

$$\sum_{k=0}^{\infty} |\pi_k| = \sum_{k=0}^{\infty} \frac{1}{2^k}, \qquad Geometric\ Series, \qquad Convergent$$

Model choice

- For ‘invertibility’ to hold, we choose Model 2, since $|½| < 1$
- This way, ACF uniquely determines the MA process.

So invertibility condition guarantees unique MA process corresponding to observed ACF.

MA(q) process

$$X_t = \beta_0 Z_t + \beta_1 Z_{t-1} + \cdots \beta_q Z_{t-q}$$

Using Backward shift operator,

$$X_t = \left(\beta_0 + \beta_1 B + \cdots + \beta_q B^q\right) Z_t = \beta(B) Z_t$$

We obtain innovations Z_t in terms of present and past values of X_t,

$$Z_t = \beta(B)^{-1} X_t = (\alpha_0 + \alpha_1 B + \alpha_2 B^2 + \cdots) X_t$$

For this to hold, "complex roots of the polynomial $\beta(B)$ must lie outside of the unit circle where B is regarded as complex variable".

Invertibility condition for MA(q)

MA(q) process is invertible if the roots of the polynomial

$$\beta(B) = \beta_0 + \beta_1 B + \cdots + \beta_q B^q$$

all lie outside the unit circle, where we regard B as a complex variable (not an operator).

(Proof is done using mean-square convergence, see optional reading)

Stationarity condition for AR(p)

AR(p) process

$$X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + \cdots + \phi_p X_{t-p} + Z_t$$

is (weakly) stationary if the roots of the polynomial

$$\phi\,(B) = 1 - \phi_1 B - \phi_2 B^2 - \cdots - \phi_p B^p.$$

all lie outside the unit circle, where we regard B as a complex variable (not an operator).

Duality between AR and MA processes

Under invertibility condition of MA(q),

$$\mathrm{MA}(q) \Longrightarrow \mathrm{AR}(\infty)$$

Under stationarity condition of AR(p)

$$\mathrm{AR}(p) \Longrightarrow \mathrm{MA}(\infty)$$

Mean square convergence

Let

$$X_1, X_2, X_3, \dots$$

be a sequence of random variables (i.e. a stochastic process).

We say X_n converge to a random variable X in the mean-square sense
if

$$E[(X_n - X)^2] \to 0 \; as \; n \to \infty$$

Difference equations

- General term of a sequence is given, ex: $a_n = 2n + 1$. So,

$$3, 5, 7, \dots$$

- General term not given, but a relation is given, ex:

$$a_n = 5a_{n-1} - 6a_{n-2}$$

- This is a difference equation (recursive relation)

- We look for a solution in the format

$$a_n = \lambda^n$$

- For the previous problem,

$$\lambda^n = 5\lambda^{n-1} - 6\lambda^{n-2}$$

We simplify

$$\lambda^2 - 5\lambda + 6 = 0$$

- Auxiliary equation or characteristic equation.

- $\lambda = 2, \lambda = 3$
- $a_n = c_1 2^n + c_2 3^n$
- With some initial conditions, say $a_0 = 3, a_1 = 8$.

We get

$$\begin{cases} c_1 + c_2 = 3 \\ 2c_1 + 3c_2 = 8 \end{cases}$$

Thus,

$$c_1 = 1, c_2 = 2.$$

Solution

$$a_n = 2^n + 2 \cdot 3^n$$

Is the solution of 2nd order difference equation

$$a_n = 5a_{n-1} - 6a_{n-2}$$

k-th order difference equation

$$a_n = \beta_1 a_{n-1} + \beta_2 a_{n-2} + \cdots + \beta_k a_{n-k}$$

Its characteristic equation

$$\lambda^k - \beta_1 \lambda^{k-1} - \cdots - \beta_{k-1}\lambda - \beta_k = 0$$

Then we look for the solutions of the characteristic equation. Say, all k solutions are distinct real numbers, $\lambda_1, \lambda_2, \dots, \lambda_k$, then

$$a_n = c_1 \lambda_1^n + c_2 \lambda_2^n + \cdots + c_n \lambda_k^n$$

Coefficients $c_j' s$ are determined using initial values.

Yule-Walker Equations

Procedure:

- We assume stationarity in advance (a priori assumption)
- Take product of the AR model with $X_{n-\mathrm{k}}$
- Take expectation of both sides
- Use the definition of covariance, and divide by γ(0_ = ${\sigma_X}^2$
- Get difference equation for $\varrho(k)$, ACF of the process
- This set of equations is called Yule-Walker equations
- Solve the difference equation

Example

We have an AR(2) process

$$X_t = \frac{1}{3}X_{t-1} + \frac{1}{2}X_{t-2} + Z_t \dots (*)$$

Polynomial

$$\phi(B) = 1 - \frac{1}{3}B - \frac{1}{2}B^2$$

has real roots $\frac{-2 \pm \sqrt{76}}{6}$ both of which has magnitude greater than 1, so roots are outside of the unit circle in $\mathbb{R}^2$. Thus, this AR(2) process is a stationary process.

Example cont.

Note that if $E(X_t) = \mu$, then

$$E(X_t) = \frac{1}{3}E(X_{t-1}) + \frac{1}{2}E(X_{t-2}) + E(Z_t)$$
$$\mu = \frac{1}{3}\mu + \frac{1}{2}\mu$$
$$\mu = 0$$

Multiply both side of $(*)$ with X_{t-k}, and take expectation

$$E(X_{t-k}X_t) = \frac{1}{3}E(X_{t-k}X_{t-1}) + \frac{1}{2}E(X_{t-k}X_{t-2}) + E(X_{t-k}Z_t)$$

Yule-Walker equations

Since $\mu = 0$, and assume $E(X_{t-k}Z_t) = 0$,

$$\gamma(-k) = \frac{1}{3}\gamma(-k+1) + \frac{1}{2}\gamma(-k+2)$$

Since $\gamma(k) = \gamma(-k)$ for any k,

$$\gamma(k) = \frac{1}{3}\gamma(k-1) + \frac{1}{2}\gamma(k-2)$$

Divide by $\gamma(0) = \sigma_X^2$

$$\rho(k) = \frac{1}{3}\rho(k-1) + \frac{1}{2}\rho(k-2)$$

This set of equations is called Yule-Walker equations.

ACF of the AR(2) model

$For\, any\, k \geq 0,$

$$\rho(k) = \frac{4+\sqrt{6}}{8}\left(\frac{2+\sqrt{76}}{12}\right)^k + \frac{4-\sqrt{6}}{8}\left(\frac{2-\sqrt{76}}{12}\right)^k$$

And

$$\rho(k) = \rho(-k)$$

References

- https://d3c33hcgiwev3.cloudfront.net/_9ea06758bd41cded9788592bdac148a3_Stationarity---Intuition-and-Definition.pdf?Expires=1582675200&Signature=L31kWUMJcPLPopEYXa9feSw1bmCqnTxUL30xUVBfaAvqlsB6LK1FeROWB4VXilH4nfZI9s0Ryc2dfUQwX~J4phSIlFEjPXy4SRZgRxr0lGvfm~nu5hCN0af9oGZhap03boQK-3d1fGXXihE3zuMsYNqAWP5NJewy1c8EMtv4zHY_&Key-Pair-Id=APKAJLTNE6QMUY6HBC5A
- https://d3c33hcgiwev3.cloudfront.net/_6db95799d385f60a655c818d699a697c_Week-2---slides---White-Background.pdf?Expires=1582675200&Signature=EKR5szAXCeYhfBnTMHsyyYWchn0Pmy5NjjIrhYQsBuwvkpI-EV36XHYZko-5SHUtPKdJjYB6pLjUKWqhh93hKjJRsZKGO9N1FbzLLZEcaEEDd87ZEKxBIzK743xLGHscWDVCEN5VyrlCx6wBg2a26QFg22ByWoUjx4uANMg~jBk_&Key-Pair-Id=APKAJLTNE6QMUY6HBC5A