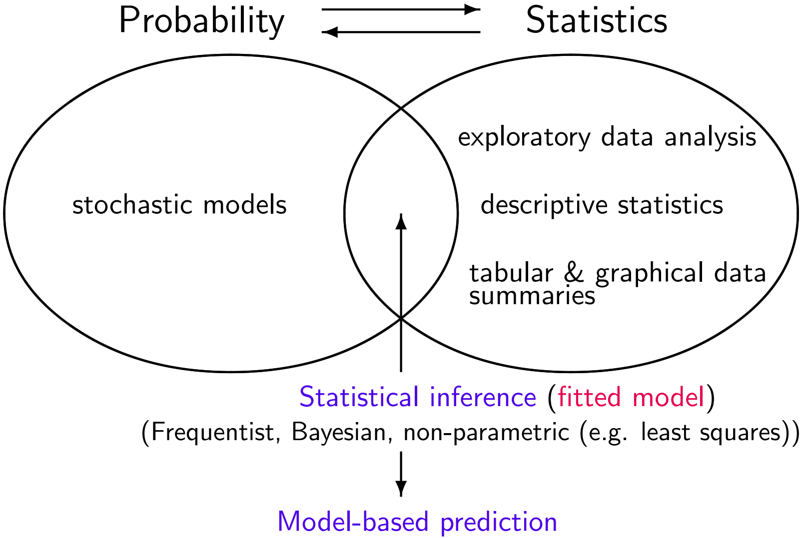
STAT 443: Time Series and Forecasting

Chapter 2

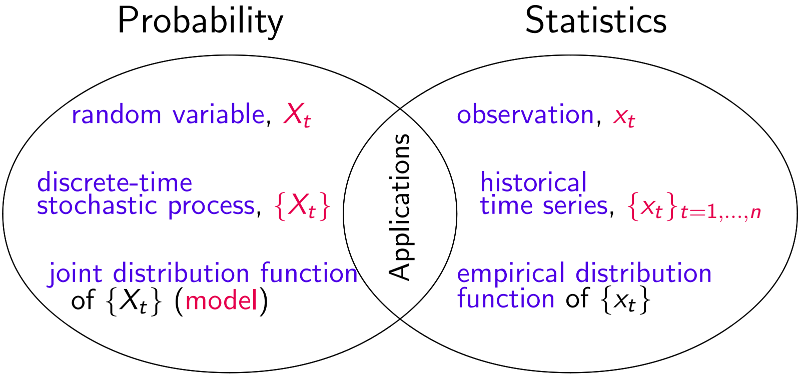
Stochastic models for time series

Essentially, all models are wrong, but some are useful.
/George E.P. Box, 1987/

Interlude on Probability and Statistics



Interlude on Probability and Statistics (cont'd)



A little bit of probability theory

- Random variables and their probability distributions are building blocks for stochastic models

- Let X and Y be two random variables

- Suppose X has distribution function F: $F(x) := \mathbb{P}\{X \leq x\}$

- Define:

  - Expected value or mean (first moment) of X: $\mathbb{E}(X) := \int x \, dF(x)$

    - **Interpretation**: weighted average of all possible values of X

    - **Linearity**: $\mathbb{E}(aX + bY) = a\,\mathbb{E}(X) + b\,\mathbb{E}(Y) \qquad (a, b \text{ constants})$

  - Variance (second moment) of X: $Var(X) := \mathbb{E}[(X - \mathbb{E}(X))^2]$

    - Interpretation: measure of how "spread out" the distribution of X is in relation to its mean

    - Square root of the variance is known as the standard deviation

    - If X and Y are independent then $Var(X + Y) = Var(X) + Var(Y)$

What is $Var(X + Y)$ if X and Y are not independent?

A little bit of probability theory (cont'd)

- If X and Y are not independent, then from the definition

$$\begin{aligned} Var(X+Y) &= \mathbb{E}[((X+Y)-\mathbb{E}(X+Y))^2] \\ &= \mathbb{E}[((X-\mathbb{E}(X))+(Y-\mathbb{E}(Y)))^2] \quad \text{(rearranging terms)} \\ &= \mathbb{E}[(X-\mathbb{E}(X))^2+(Y-\mathbb{E}(Y))^2+2(X-\mathbb{E}(X)(Y-\mathbb{E}(Y))] \\ &= Var(X)+Var(Y)+2Cov(X,Y), \end{aligned}$$

where $Cov(X, Y)$ is called the covariance between X and Y, and is defined as

$$Cov(X,Y) := \mathbb{E}[(X-\mathbb{E}(X))(Y-\mathbb{E}(Y))]$$

- Note: $Cov(X,X) = Var(X)$
- After simplifying, we see: $Cov(X,Y) = \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y)$
- If X and Y are independent then $Cov(X,Y) = 0$

A little bit of probability theory (cont'd)

- Note that the units of covariance are determined by the units of X and Y, which makes it hard to interprete

- It is hence common to standardize covariance by the product of standard deviations of X and Y, giving the correlation (coefficient) between X and Y:

$$\rho_{X,Y} := \frac{Cov(X,Y)}{\sqrt{Var(X)Var(Y)}}$$

- Some properties:
  - Correlation has no units
  - $-1 \leq \rho_{X,Y} \leq 1$
  - If $\rho_{X,Y} = \pm 1$, then X and Y are exactly linearly related: $Y = aX + b$ for some constants a and b

Describing models for time series

- Models for time series are examples of discrete-time stochastic processes

- A discrete-time stochastic process can be thought of as a sequence of random variables $\{X_1, X_2, \ldots\}$, also denoted $\{X_t\}$

- Naturally, certain properties of the distribution of X_t, such as its expected value $\mathbb{E}(X_t)$ and variance $Var(X_t)$, may vary with time t

- To measure serial dependence, it is natural to define an autocovariance function, the covariance between X_t at two different time points say t_1 and t_2:

$$\gamma(t_1, t_2) := Cov(X_{t_1}, X_{t_2})$$

- The difference $|t_2 - t_1|$ is referred to as the lag

Describing models for time series (cont'd)

In this module we will mostly deal with stochastic models defined below

Definition: *A stochastic process $\{X_t\}$ is called (weakly or second order) stationary if its mean is constant, i.e.*

$$\mathbb{E}(X_t) = \mu \qquad \text{for all } t,$$

variance is finite

$$\mathrm{Var}(X_t) < \infty,$$

and its autocovariance function depends only on the lag, i.e.

$$Cov(X_t, X_{t+h}) =: \gamma(h) \qquad \text{for all } t \text{ and } h$$

Stationarity for model versus data

- Given a stochastic process (time series model), it is generally possible to prove whether it is stationary or non-stationary

  (this is a probability exercise!)

- It is also common to refer to data as being stationary or not

  "Is this time series stationary?"

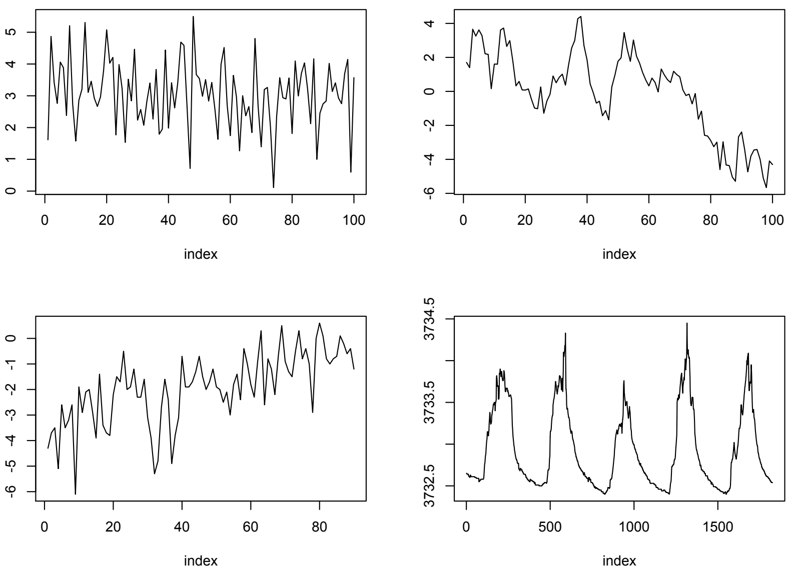
  What one really means in this case is whether the time series at hand can be well described or modelled by a stationary stochastic process (model)

Stationarity for model versus data (cont'd)

- Remember: for real data, there is no "true model"
- Sometimes certain features of a time series (e.g., trend, seasonality) make it clearly non-stationary
- Sometimes a good argument can be made for data to be stationary
- And in some cases an argument can be made for either.

  The ultimate choice can be based on data context and performance (e.g., model fit or predictive accuracy)

Quiz: Which of the following time series appear to be stationary?



What is the significance of stationarity for modelling?

- Stationary time series are easier to model
- Stationarity ensures that past is informative about the future, which makes prediction possible

Describing models for time series (cont'd)

- The mean and autocovariance contain considerable information about a stochastic process
- However, they do not completely describe its evolution over time
- A stronger condition is that of strong stationarity requiring the same joint distribution of the sequences $\{X_{t_1}, \ldots, X_{t_n}\}$ and $\{X_{t_1+h}, \ldots, X_{t_n+h}\}$ for all $t_1, \ldots, t_n$ and all h
- Strong stationarity is a very difficult condition to check in practice for a realization of a stochastic process
- For us stationarity will be in the weak sense

Describing models for time series (cont'd)

- For a stationary stochastic process $\{X_t\}$, the autocorrelation function (acf) at lag h is defined as

$$\rho(h) := \frac{Cov(X_t, X_{t+h})}{Var(X_t)} = \frac{\gamma(h)}{\gamma(0)} = \frac{\gamma(h)}{\sigma_X^2}$$

- Properties:
  - $\rho(0) = 1$
  - $\rho(-h) = \rho(h)$ (due to stationarity)
  - $-1 \leq \rho(h) \leq 1$
  - The acf does not uniquely specify a stochastic process; two different processes could have the same acf

Survey of popular stochastic models for time series

- We have already seen one example of a stochastic process which is a basic "building block" for other processes – the white noise $\{Z_t\}$, a sequence of i.i.d. random variables with mean zero and variance σ^2, denoted $WN(0, \sigma^2)$

- The autocovariance and autocorrelation functions are easy to compute

$$\gamma(h) = \begin{cases} 0 & \text{for } h \neq 0, \\ \sigma^2 & \text{for } h = 0 \end{cases} \qquad \text{and} \qquad \rho(h) = \begin{cases} 0 & \text{for } h \neq 0, \\ 1 & \text{for } h = 0 \end{cases}$$

- A white noise process is (strongly) stationary

- If $Z_t \sim \mathcal{N}(0, \sigma^2)$ for all $t = 1, 2, \ldots$ and are independent, then $\{Z_t\}$ is an example of a white noise process

Moving average processes

- Let $\{Z_t\}$ be a $WN(0, \sigma^2)$

Definition: *The process $\{X_t\}$ is said to be a moving average process of order q, denoted MA(q), if*

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

for some constants $\beta_0, \beta_1, \ldots, \beta_q$ (usually $\beta_0 = 1$ and this will be our convention)

- In words: The value of an MA(q) process at time t is a weighted sum of the last q values of the white noise process $\{Z_t\}$ plus the new value from $\{Z_t\}$

- Interpretation: the effect of some random event (sometimes called an innovation) can have an immediate impact and also a "shock-wave" effect on later time periods

Activity: Moving Average Processes

Properties of an MA(q) process

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

- $\mathbb{E}(X_t) = 0$
- $Var(X_t) = \sigma^2(1 + \beta_1^2 + \cdots + \beta_q^2) = \sigma^2 \sum_{i=0}^{q} \beta_i^2 \qquad (\beta_0 = 1)$
- The autocovariance function (acvf) at lag h:

$$\begin{aligned}
\gamma(h) &= Cov(X_t, X_{t+h}) \\
&= Cov(Z_t + \beta_1 Z_{t-1} + \cdots + \beta_q Z_{t-q}, Z_{t+h} + \beta_1 Z_{t-1+h} + \cdots + \beta_q Z_{t-q+h}) \\
&= \begin{cases} \sigma^2 \sum_{i=0}^{q-h} \beta_i \beta_{i+h} & \text{for } h = 0, 1, \ldots, q \\ 0 & \text{for } h > q \\ \gamma(-h) & \text{for } h < 0 \end{cases}
\end{aligned}$$

- Any MA(q) process is stationary since both the mean and the acvf do not depend on time t

Properties of an MA(q) process (cont'd)

- Combining expressions for the variance and acvf, the autocorrelation function (acf) is given by

$$\rho(h) = \begin{cases} \sum_{i=0}^{q-h} \beta_i \beta_{i+h} / \sum_{i=0}^{q} \beta_i^2 & \text{for } h = 0, 1, \ldots, q \\ 0 & \text{for } h > q \\ \rho(-h) & \text{for } h < 0 \end{cases}$$

  that is, it "cuts off" at lag q

- This property of the acf partly characterizes MA processes

Invertibility of moving average processes

- Note: both MA(1) processes $X_t = Z_t + \beta Z_{t-1}$ and $X_t = Z_t + \frac{1}{\beta} Z_{t-1}$ have the same acf

$$\rho(\pm 1) = \frac{\beta}{1+\beta^2} = \frac{1/\beta}{1+(1/\beta)^2} \qquad \text{and} \qquad \rho(h) = 0 \text{ for } |h| > 1$$

- For $X_t = Z_t + \beta Z_{t-1}$, we have

$$\begin{aligned} Z_t &= X_t - \beta Z_{t-1} \\ &= X_t - \beta(X_{t-1} - \beta Z_{t-2}) = \dots \\ &= X_t - \beta X_{t-1} + \beta^2 X_{t-2} - \beta^3 X_{t-3} + \cdots \end{aligned}$$

- Similarly, "inverting" $X_t = Z_t + \frac{1}{\beta} Z_{t-1}$ gives

$$Z_t = X_t - \frac{1}{\beta} X_{t-1} + \frac{1}{\beta}^2 X_{t-2} - \frac{1}{\beta}^3 X_{t-3} + \cdots$$

Invertibility of moving average processes (cont'd)

- In these representations, we have two sequences of coefficients in front of X terms

$$1, -\beta, \beta^2, -\beta^3, \ldots \qquad \text{and} \qquad 1, -1/\beta, 1/\beta^2, -1/\beta^3, \ldots$$

- If $|\beta| < 1$, the first sequence forms a convergent sum making the first process well-defined whereas the sum of the second sequence of coefficients diverges

- Converse is true if $|\beta| > 1$

- Hence, the two processes cannot be both sensibly defined

Invertibility of moving average processes (cont'd)

Definition:

A process is said to be invertible if it can be expressed in the form:

$$Z_t = \sum_{j=0}^{\infty} \pi_j X_{t-j}, \qquad \text{with } \sum_j |\pi_j| < \infty.$$

Remark: The condition of invertibility ensures that an MA process is uniquely specified by its acf (model identifiability, a desirable property in estimation)

Example: An MA(1) process $X_t = Z_t + \beta Z_{t-1}$ is invertible provided $|\beta| < 1$

Invertibility of moving average processes (cont'd)

Definition:

The *backward shift operator B* *is defined as*

$$B^j X_t := X_{t-j} \qquad (j = 0, 1, 2, \dots)$$

- *i.e., B^j transforms a time series into another time series by shifting it back j time units*

- We can then re-write an MA(q) process in the following form:

$$X_t = \theta(B) Z_t,$$

where $\theta(B) = 1 + \beta_1 B + \cdots + \beta_q B^q$ is referred to as the characteristic polynomial

Invertibility of moving average processes (cont'd)

Theorem:

An MA(q) process $X_t = \theta(B)Z_t$ is invertible if all roots of the characteristic polynomial $\theta(B)$ lie outside the unit circle in complex plane; i.e., the roots have modulus greater than unity.

- (here B is regarded as a complex variable, not as an operator)

Remarks on the proof:

- From the definition of invertibility, $\frac{1}{\theta(B)}$ must have a convergent series expansion in powers of B:

$$\frac{1}{\theta(B)} = \pi(B)$$

- A key step is based on the Fundamental Theorem of Algebra (saying that any homogeneous polynomial of order q can be factorized into a product of first order polynomials involving its roots)
- O.D. Anderson (1978). On the invertibility conditions for moving average processes. *Series Statistics* 9: 525-529. (and references therein)

Exercise:

Let

$$X_t = Z_t - 1.3Z_{t-1} + 0.4Z_{t-2}, \qquad \{Z_t\} \sim WN(0, \sigma^2)$$

Is the process $\{X_t\}$ invertible? Is it stationary?

Activity: Autoregressive Processes

Autoregressive processes

- As usual, let $\{Z_t\}$ be a $WN(0, \sigma^2)$

Definition: *The process $\{X_t\}$ is said to be an autoregressive process of order p, denoted AR(p), if*

$$X_t = \alpha_1 X_{t-1} + \alpha_2 X_{t-2} + \cdots + \alpha_p X_{t-p} + Z_t$$

for some constants $\alpha_1, \ldots, \alpha_p$

- That is, the value of the process at time t is obtained by regressing on the previous p values plus a random error

AR(p) processes

- Using the backward shift operator, an $AR(p)$ process can be written as

$$\phi(B)X_t = Z_t, \qquad \underbrace{\phi(B) = 1 - \alpha_1 B - \alpha_2 B^2 - \cdots - \alpha_p B^p}_{\textit{characteristic polynomial}}$$

- Alternatively, an $AR(p)$ process can be expressed as an $MA(\infty)$ process for any p

$$X_t = \phi(B)^{-1} Z_t$$

as it is possible to express $\phi(B)^{-1}$ as an infinite sum

$$\phi(B)^{-1} = 1 + \beta_1 B + \beta_2 B^2 + \cdots \quad \text{for some constants } \beta_i$$

AR(p) processes (cont'd)

$$\boxed{X_t = \phi(B)^{-1} Z_t = \sum_{j=0}^{\infty} \beta_j Z_{t-j}}$$

- Properties:
  - $\mathbb{E}(X_t) = 0$
  - $\gamma(h) = \sigma^2 \sum_{j=0}^{\infty} \beta_j \beta_{j+h}$

- Proposition:

  The process $X_t = \phi(B)^{-1} Z_t$ is stationary if and only if $\sum_{j=0}^{\infty} \beta_j^2 < \infty$.

  Proof: ...

AR(p) processes (cont'd)

- Similarly to the invertibility condition for $MA(q)$ processes, stationarity of an $AR(p)$ process can be determined using the characteristic polynomial

- Theorem:

  An $AR(p)$ process $\phi(B)X_t = Z_t$ is stationary if and only if the roots of the characteristic polynomial $\phi(B)$ lie outside the unit circle in the complex plane.

Exercise

Consider

$$\boxed{X_t = X_{t-1} - 1.25X_{t-2} + Z_t, \quad \{Z_t\} \sim WN(0, \sigma^2)}$$

Is this $AR(2)$ process stationary? Is it invertible?

Computation of the ACF for an $AR(p)$ process

$$\boxed{AR(p): \quad X_t = \alpha_1 X_{t-1} + \alpha_2 X_{t-2} + \cdots + \alpha_p X_{t-p} + Z_t}$$

Possible approaches:

- Use the $MA(\infty)$ representation $X_t = \sum_{j=0}^{\infty} \beta_j Z_{t-j}$
  - Then $\gamma(h) = \sigma^2 \sum_{j=0}^{\infty} \beta_j \beta_{j+h}$ and

$$\rho(h) = \frac{\gamma(h)}{\gamma(0)} = \frac{\sum_{j=0}^{\infty} \beta_j \beta_{j+h}}{\sum_{j=0}^{\infty} \beta_j^2}$$

  - Requires computation of β's from α's (e.g., by equating coefficients) $\Rightarrow$ tedious!!

- By solving the Yule-Walker Equations (Y-W eq'ns)

Yule-Walker Equations

Derivation:

1. Start with the defining equation

$$X_t = \alpha_1 X_{t-1} + \alpha_2 X_{t-2} + \cdots + \alpha_p X_{t-p} + Z_t \qquad (*)$$

   Assume $\{X_t\}$ is stationary.

2. Multiply both sides of $(*)$ by X_{t-k} for $k = 1, \ldots, p$.
   This gives p equations.

3. Take expectation on both sides of <u>each</u> of these p equations.

4. Divide each equation by $\sigma_X^2 = Var(X_t)$.
   (Note $\sigma_X^2 < \infty$ and independent of t due to stationarity.)

$\Rightarrow$ **Activity: Yule-Walker Equations** Questions 1-3

Yule-Walker Equations - Derivation

1. Assume $\{X_t\}$ is stationary.

$$\boxed{X_t = \alpha_1 X_{t-1} + \alpha_2 X_{t-2} + \cdots \alpha_p X_{t-p} + Z_t \qquad (*)}$$

2. Multiply both sides of $(*)$ by X_{t-k} for $k = 1, \ldots, p$.
   This gives p equations:

$$X_t X_{t-k} = \alpha_1 X_{t-1} X_{t-k} + \alpha_2 X_{t-2} X_{t-k} + \cdots + \alpha_p X_{t-p} X_{t-k} + Z_t X_{t-k}, \ (k = 1, \ldots, p)$$

3. Take expectation on both sides of each of these p equations.

4. Divide each equation by $\sigma_X^2 = Var(X_t)$.
   (Note $\sigma_X^2 < \infty$ and independent of t due to stationarity.)

Note: $\mathbb{E}(X_t X_{t-k})/\sigma_X^2 = \rho(k)$, $k = 1, \ldots, p$

Hence, the k^{th} equation above is $\rho(k) = \alpha_1 \rho(k-1) + \alpha_2 \rho_{k-2} + \ldots + \alpha_p \rho_{k-p}$

Since $\rho(-k) = \rho(k)$ and $\rho(0) = 1$, the p equations become ...

The Yule-Walker Equations and General Solution

$$\boxed{\begin{cases} \rho(1) = \alpha_1 + \alpha_2\rho(1) + \cdots + \alpha_p\rho(p-1), \\ \rho(2) = \alpha_1\rho(1) + \alpha_2 + \cdots + \alpha_p\rho(p-2), \\ \vdots \\ \rho(p) = \alpha_1\rho(p-1) + \alpha_2\rho(p-2) + \cdots + \alpha_p \end{cases}} \qquad (YWE)$$

- The acf can be obtained by solving the Y-W equations (for which some knowledge of difference equations is useful)
- The general solution to (YWE) is of the form

$$\rho(h) = A_1 d_1^{|h|} + \cdots + A_p d_p^{|h|},$$

where $d_1, \ldots, d_p$ are the roots of polynomial

$$D^p - \alpha_1 D^{p-1} - \cdots - \alpha_p D^0 = 0$$

in D, and constants A_i are subject to the constraint $\sum_{i=1}^p A_i = 1$ (Why?)

Activity: Yule-Walker Equations (cont'd)

- We will next introduce 3 model classes related to MA and AR processes studied earlier
  - ARMA
  - ARIMA ("I" for "integrated")
  - SARIMA ("S" for "seasonal")
- They were proposed in pioneering work of G. Box and G. Jenkins

ARMA models

Combining an AR(p) and an MA(q) process gives a mixed model, denoted ARMA(p,q) and defined as

$$X_t = \alpha_1 X_{t-1} + \cdots + \alpha_p X_{t-p} + Z_t + \beta_1 Z_{t-1} + \cdots + \beta_q Z_{t-q}$$

for some constants $\alpha_1, \ldots, \alpha_p, \beta_1, \ldots, \beta_q$ and a $WN(0, \sigma^2)$ process $\{Z_t\}$

What is a potential advantage of an ARMA model versus a pure MA or AR process?

- ARMA models are of importance in view of the "Principle of Parsimony", a quest for the simplest model:

  it might be possible to fit a mixed model to a time series with fewer parameters than either a pure AR or a pure MA model

An $ARMA(p,q)$ process can be written as

$$\boxed{\phi(B)X_t = \theta(B)Z_t}$$

where:

$$\begin{aligned}\phi(B) &= 1 - \alpha_1 B - \alpha_2 B^2 - \cdots - \alpha_p B^p \\ \theta(B) &= 1 + \beta_1 B + \beta_2 B^2 + \cdots + \beta_q B^q\end{aligned}$$

Properties:

- An $ARMA(p,q)$ process is stationary if the roots of $\phi(B)$ have modulus greater than unity.
- An $ARMA(p,q)$ process is invertible if the roots of $\theta(B)$ have modulus greater than unity.
- Computation of the acvf is generally quite tedious, no simple formula exists

- An ARMA model can be written as a pure MA process:

$$\boxed{X_t = \psi(B) Z_t \quad \text{for some polynomial } \psi(B) = \sum_{i=0}^{\infty} \psi_i B^i}$$

  - note: $\psi(B) = \theta(B)/\phi(B)$
  - useful representation for creating confidence intervals in forecasting

- An ARMA model can also be written as a pure AR process:

$$\boxed{\pi(B) X_t = Z_t \quad \textit{where } \pi(B) = 1 - \sum_{i=1}^{\infty} \pi_i B^i}$$

  - obviously: $\pi(B)\psi(B) = 1$

- The weights $\{\psi_i\}$ and $\{\pi_i\}$ can be found via division or by equating coefficients in powers of B using either

$$\psi(B)\phi(B) = \theta(B) \qquad \text{or} \qquad \theta(B)\pi(B) = \phi(B)$$

ARIMA models

- Most observed time series are non-stationary
- Suppose $X_t = \mu t + Z_t$. Define a process

$$Y_t = \nabla X_t := X_t - X_{t-1},$$

where ∇ is called the difference operator.

Question: Is the process $\{Y_t\}$ stationary?

- Differencing can be used to remove many types of non-stationary effects
- This gives rise to a general class of models where initially the process is differenced, say d times, before an ARMA(p,q) is appropriate

Definition: *Given a stochastic process* $\{X_t\}$*, create a new process* $\{Y_t\}$ *by applying the difference operator d times to* X_t*:*

$$Y_t := \nabla^d X_t.$$

If $\{Y_t\}$ *is an ARMA(p,q) process, i.e.,*

$$Y_t = \alpha_1 Y_{t-1} + \cdots + \alpha_p Y_{t-p} + Z_t + \beta_1 Z_{t-1} + \cdots + \beta_q Z_{t-q}$$

for some constants $\alpha_1, \ldots, \alpha_p, \beta_1, \ldots, \beta_q$ *and a* $WN(0, \sigma^2)$ *process* $\{Z_t\}$*, then* $\{X_t\}$ *is said to be an ARIMA(p,d,q) process.*

- "I" stands for "integrated"
- An ARIMA(p,d,q) process is non-stationary (Why?)
- In practice, first order differencing $d = 1$ usually suffices
- Differencing is widely used for econometric data

SARIMA models

- For time series with seasonal variation, differencing at the seasonal frequency can be used to remove seasonal effects
  - E.g.: For a process $X_t = s_t + Z_t$ with a seasonal effect s_t of period 4 (e.g., quarterly) satisfying $s_t = s_{t-4}$ for all t, the differenced process

    $$\nabla_4 X_t := X_t - X_{t-4} = s_t + Z_t - s_{t-4} - Z_{t-4} = Z_t - Z_{t-4}$$

    is stationary

- Applying seasonal differencing to an ARIMA process gives a class of seasonal ARIMA models, abbreviated as SARIMA

Definition: $\{X_t\}$ *is a SARIMA process of order* $(p, d, q) \times (P, D, Q)_s$ *if it is of the form:*

$$\phi(B)\ \Phi(B^s)\ W_t = \theta(B)\ \Theta(B^s)\ Z_t$$

where

$$W_t = \nabla^d \nabla_s^D X_t, \qquad \nabla_s X_t := X_t - X_{t-s}$$

and ϕ, Φ, θ *and* Θ *are polynomials of order* p, P, q *and* Q, *respectively*

Remarks:

- Typical values for d and D are $\{0, 1, 2\}$
- When both d and D are nonzero, apply the difference operators starting from leftmost

Summary

- Describing stochastic models for time series
  - mean function, autocovariance function ($\gamma(h)$) and autocorrelation function ($\rho(h)$) at lag h
  - Stationarity
  - Invertibility
- Popular stochastic models for time series
  - White noise process, $WN(0, \sigma^2)$
  - Moving average processes, MA(q)
  - Autoregressive processes, AR(p)
  - Mixed models, ARMA(p,q)
  - Non-stationary models, ARIMA(p,d,q) and SARIMA$(p, d, q) \times (P, D, Q)_s$

Concluding remarks

- We have discussed a broad class of stochastic models which proved to be useful in describing time series data
- All the models involve parameters (the α's, β's and σ^2) which in reality are not known and hence need to be sensibly estimated from the data
- The chosen model has to be assessed on whether it is appropriate for the data
- If the fitted model appears suitable, it can then be used in, for instance, predicting future values of the process

The last three points are topics for the next two chapters