IDS 702: Module 7.2

Stationarity and autocorrelation

Dr. Olanrewaju Michael Akande

IDS 702

STATIONARITY

- The most common time series models usually assume stationarity.
- Stationarity of a time series process means that the marginal distribution of any part of the series does not depend on time.
- Basically, the locations in time themselves does not matter in the marginal and joint distributions; however, the differences in locations, that is, the lags, do matter!
- Put a different way, a stationary time series is one whose properties do not depend on the particular time at which the series is observed.
- Examples of non stationary series:
  - Steadily increasing trend (like the melanoma example).
  - Known seasonal trends, like increase in sales before Christmas.
  - Break in trend due to some external event.

IDS 702

Stationarity

- Denote the times series for the outcome by y_t.
- Stationarity $\Rightarrow$
  - $\mathbb{P}\mathrm{r}(y_t)$ is the same for all t,
  - $\mathbb{P}\mathrm{r}(y_t, y_{t+1})$ is the same for all t,
  and so on...
- Weak stationarity requires that only marginal moments, that is, means, variances and covariances are the same.
- Stationarity $\Rightarrow$ weak stationarity, but the converse need not hold.
- For a normal distribution, the mean and variance completely characterizes the distribution, so that stationarity and weak stationarity will be equivalent.
- Why does that matter?
- When dealing with linear models, what distribution do we assume??

IDS 702

Popular stationary models

- We will mainly focus on two types of stationary time series models:
  - Autoregressive models (AR models)
    - Value of outcome at time t is correlated with value at previous times.
  - Moving average models (MA models)
    - Value of outcome at time t is correlated with value of prediction errors at previous times.
- Note: autoregressive moving average models (ARMA models)
  - Combination of AR and MA.
- There are many more types of time series models (see STA 642/942).

IDS 702

Autocorrelation

- Autocorrelation (serial correlation) measures the strength of the linear relationship between y_t and its lagged values.

- The lag k autocorrelation ρ_k measures the correlation in outcomes at time t and at time $t-k$, where k indicates how far back to go; k is called a lag.

- The sample lag k autocorrelation r_k can be calculated using

$$r_k = \frac{\sum_{t=k+1}^{T}(y_t - \bar{y})(y_{t-k} - \bar{y})}{\sum_{t=1}^{T}(y_t - \bar{y})^2}.$$

IDS 702

Partial autocorrelation

- Autocorrelation ρ_k (and the sample version r_k) between y_t and y_{t-k} will also include the linear relationships between y_t and each of $y_{t-1}, y_{t-2}, \ldots, y_{t-k+1}$.
- As you will see, we will need to be able to assess the correlation between y_t and y_{t-k} without interference from the other lags.
- Partial autocorrelation lets us do just that.
- It is the autocorrelation between y_t and y_{t-k}, with all the linear relationships between y_t and each of $y_{t-1}, y_{t-2}, \ldots, y_{t-k+1}$ removed.
- In `R`, use `acf` to compute and plot autocorrelations and `pacf` to compute and plot partial autocorrelations.

IDS 702

What's next?

Move on to the readings for the next module!

IDS 702