**ARMA Models**

ARMA models combine two models:

* The first is an autoregressive (AR) model. Autoregressive models anticipate series’ dependence on **its own past values**.
* The second is a moving average (MA) model. Moving average models anticipate series’ dependence on **past forecast errors**.
* The combination (ARMA) is also known as the **Box - Jenkins approach**.

ARMA models are often expressed using orders ***p*** and ***q*** for the ***AR*** and ***MA*** components.

**You can determine if seasonality is present by using :**

1. **autocorrelation** **and** **partial autocorrelation plots** 2. **seasonal subseries plots**, and **intuition**

**Autocorrelation Plot-**It summarizes total (2-way) correlation between the variable and its past values.

**Partial Autocorrelation Plot** also summarizes dependence on past observations(it measures partial results)

1. **Seasonal Subseries Plot** is one approach for measuring seasonality. This chart shows the **average level for each seasonal period** and illustrates how individual observations relate to this level.

### **Estimation**

Once we have a stationary series, we can estimate AR and MA models. We need to determine ***p*** and ***q***, the order of the AR and MA models.

1. One approach here is to look at autocorrelation and partial autocorrelation plots.
2. Another approach is to treat ***p*** and ***q*** as hyperparameters and apply standard approaches (grid search, cross validation, etc.)

**ARIMA models have three components:**

* AR Model
* Integrated Component
* MA Model

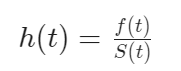
## ****ARIMA and SARIMA Estimation****

These are the steps to estimate p, d, q and P, D, Q?

* Visually inspect a run sequence plot for trend and seasonality.
* Generate an ACF Plot.
* Generate a PACF Plot.
* Treat as hyperparameters (cross validate).
* Examine ***information criteria*** (AIC, BIC) which penalize the number of parameters the model uses.

**Survival Function :** Survival Analysis is used **to estimate the lifespan** of a particular **event** under study.

**Hazard Rate** is defined as:



It represents the instantaneous **rate at which events occur**, given that it has not occurred already.

**Kaplan-Meier Curve** allows us to visually inspect differences in survival rates by category.

### The **Cox Proportional Hazard (CPH) model**

This is one of the most common survival models. It assumes features have a **constant proportional impact** on the hazard rate.

### **Accelerated Failure Time (AFT) models** (several variants including the **Weibull AFT model**)

These models differ with respect to assumptions they make about the hazard rate function, and the impact of features.

**Moving average smoothing** is useful for estimating trend and seasonality of past data.

**Moving-average model** is conceptually a linear regression of the current value of the series against current and previous (unobserved) white noise error terms or random shocks.

AR models propagate shocks infinitely

MA models do not propagate shocks infinitely; they die after q lags.

Some rules to highlight from the Duke ARIMA Guide:

1. If the series has positive autocorrelations out to a high number of lags, then it probably needs a higher order of differencing
2. If the lag-1 autocorrelation is zero or negative, or the autocorrelations are all small and patternless, then the series does not need a higher order of differencing. If the lag-1 autocorrelation is -0.5 or more negative, the series may be overdifferenced. BEWARE OF OVERDIFFERENCING!!
3. A model with no orders of differencing assumes that the original series is stationary (mean-reverting). A model with one order of differencing assumes that the original series has a constant average trend (e.g. a random walk or SES-type model, with or without growth). A model with two orders of total differencing assumes that the original series has a time-varying trend (e.g. a random trend or LES-type model)