Time Series

Sudipta Das

Assistant Professor,

Department of Computer Science,
Ramakrishna Mission Vivekananda Educational & Research Institute

Outline I

- Testing the Noise Sequence
 - Testing i.i.d Sequence
 - Testing Normality



Testing the Noise Sequence I

- If there is no dependence among the residuals, then
 - we can regard them as observations of independent random variables,
 - and there is no further modeling to be done except to estimate their mean and variance.
- However, if there is significant dependence among the residuals, then
 - we need to look for a more complex model for the noise that accounts for the dependence.
 - as a result, the past observations of the noise sequence can assist in predicting future values.

Testing the Noise Sequence II

 Therefore, we examine some simple tests for checking the hypothesis that the residuals are observed values of independent and identically distributed random variables.

The sample autocorrelation function I

• For large n, the sample auto-correlations of an *i.i.d.* sequence X_1, \ldots, X_n is defined as

$$\hat{
ho}(h) = rac{\hat{\gamma}(h)}{\hat{\gamma}(0)} ext{ for } -n < h < n,$$

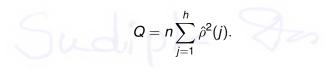
where
$$\hat{\gamma}(h) = n^{-1} \sum_{t=1}^{n-|h|} (x_{t+h} - \bar{x})(x_t - \bar{x})$$
 and $\bar{x} = n^{-1} \sum_{t=1}^{n} x_t$.

The sample autocorrelation function II

- The auto-correlations of an **i.i.d. sequence** X_1, \ldots, X_n , with **finite variance** are approximately *i.i.d.* with distribution N(0, 1/n)
- Hence, if x_1, \ldots, x_n is a realization of such an *i.i.d.* sequence, about 95% of the sample auto-correlations should fall between the bounds $\pm 1.96/\sqrt{n}$.

The portmanteau test I

• Instead of checking to see whether each sample auto-correlation $\hat{\rho}(j)$ falls inside the bounds defined above, it is also possible to consider the single statistic



• For an **i.i.d. sequence** X_1, \ldots, X_n , with finite variance Q is approximately distributed as the sum of squares of the independent N(0,1) random variables, $\sqrt{n}\hat{\rho}(j), j=1,\ldots,h$, i.e., as chi-squared with h degrees of freedom.

The portmanteau test II

- A large value of Q suggests that the sample autocorrelations of the data are too large for the data to be a sample from an i.i.d. sequence.
 - We therefore reject the *i.i.d.* hypothesis at level α if $Q > \chi^2_{1-\alpha}(h)$, where $\chi^2_{1-\alpha}(h)$ is the $1-\alpha$ quantile of the chi-squared distribution with h degrees of freedom.

Ljung Box test I

A better and modified estimator

$$Q_{LB} = n(n+2) \sum_{j=1}^{h} \hat{\rho}^2(j)/(n-j).$$

• For an **i.i.d. sequence** X_1, \dots, X_n , with finite variance Q is approximately distributed as a chi-squared with h degrees of freedom.

Ljung Box test II

- A large value of Q_{LB} suggests that the sample autocorrelations of the data are too large for the data to be a sample from an i.i.d. sequence.
 - We therefore reject the *i.i.d.* hypothesis at level α if $Q_{LB} > \chi^2_{1-\alpha}(h)$, where $\chi^2_{1-\alpha}(h)$ is the $1-\alpha$ quantile of the chi-squared distribution with h degrees of freedom.

The turning point test I

- If x_1, \ldots, x_n is a sequence of observations, we say that there is a turning point at time i, 1 < i < n,
 - if $x_{i-1} < x_i$ and $x_i > x_{i+1}$ or if $x_{i-1} > x_i$ and $x_i < x_{i+1}$.
- If T is the number of turning points of an i.i.d. sequence of length n, then,
 - the probability that a point at time *i* is a turning point is $\frac{2}{3}$
 - $\mu_T = E[T] = 2(n-2)/3$
 - $\sigma_T^2 = Var[T] = (16n 29)/90$

The turning point test II

- A large value of $T \mu_T$ indicates that the series is fluctuating more rapidly than expected for an *i.i.d.* sequence.
- On the other hand, a value of $T \mu_T$ much smaller than zero indicates a positive correlation between neighboring observations.

The turning point test III

- For an *i.i.d.* sequence with n large, it can be shown that T is approximately $N(\mu_T, \sigma_T^2)$.
- Therefore, we can carry out a test of the **i.i.d.** hypothesis and reject it at level α if $|T \mu_T|/\sigma_T > z_{1-\alpha/2}$,
 - where $z_{1-\alpha/2}$ is the 1 $-\alpha/2$ quantile of the standard normal distribution

The test for Normality. I

Q-Q Plot: Graphical check for normality Steps:

- Given an *i.i.d.* sequence $\{x_1, x_2, \dots, x_n\}$ turn it to a standardized form $\{z_1, z_2, \dots, z_n\}$
- Sort the standardized sequence to $\{z_{(1)}, z_{(2)}, \dots, z_{(n)}\}$

The test for Normality. II

• Corresponding to each $z_{(i)}$ calculate the associated quantile from standard normal, i.e.,

$$N_{q_{z_{(i)}}} = \frac{i - 0.5}{n}$$

• Note: Empirical distribution of $Z_{(i)}$:

$$P(Z_{(i)} \le Z_{(i)}) = \frac{\text{Number of points less than equal } z_i}{\text{Total Points}} = \frac{i}{n}$$

- ullet Plot the pair points $(N_{q_{Z_{(i)}}}, Z_i)$ for $i=1,\ldots,n$
- If the sequence $\{x_1, x_2, \dots, x_n\}$ is coming from normal, the plot described above will be straight line passing through origin with slop 1.

The test for Normality. III

Shapiro R2 test

• Test Statistics, under normality assumption:

$$R^{2} = \left[\frac{\sum_{i=1}^{n} N_{q_{Z_{(i)}}} Z_{i}}{\sum_{i=1}^{n} N_{q_{Z_{(i)}}} \sum_{i=1}^{n} Z_{i}}\right]^{2}$$

- p values can be found from standard tables
- Decision on Acceptance/Rejection of normality can be made by seeing the p value