

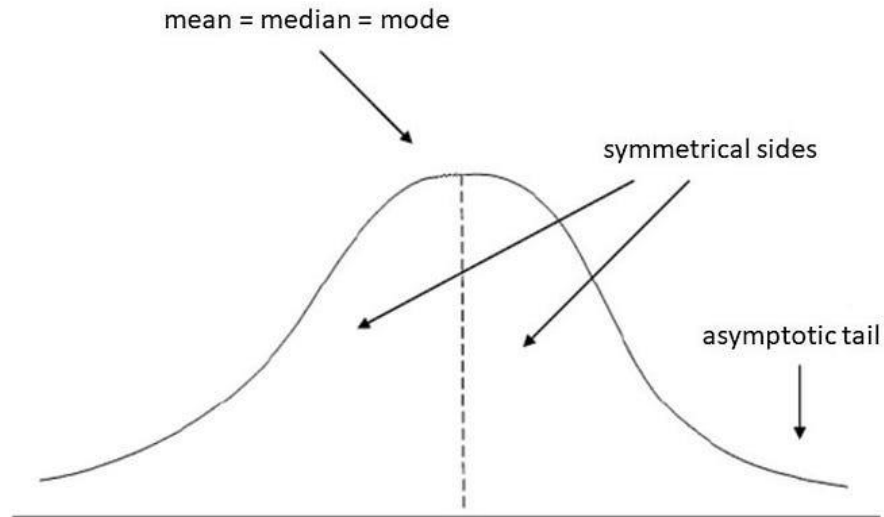
Time series analysis – basic data introduction

Wengang Mao (Marine Technology)
Department of Mechanics and Maritime Sciences,
Chalmers University of Technology,
Goteborg, Sweden

Outline

- Why to transfer to Gaussian
- How to check if a RV is Gaussian
- Techniques for the transformation
 - ✓ Log Transformation
 - ✓ Square root Transformation
 - ✓ Reciprocal Transformation
 - ✓ Exponential Transformation
 - ✓ Box-Cox Transformation

Why Gaussian



Importance of Data Distribution Transformation

- Machine learning algorithms provide better results(predictions) when numerical data present are Normally distributed.
- To make the cost function minimize better the error of the predictions.

Why Gaussian

- **Importance of Normality in Machine Learning!**

1. Gaussian distribution is found everywhere because a dataset with finite variance turns into Gaussian as long as the dataset with independent feature-probabilities is allowed to grow in size.
2. Datasets with Gaussian distributions make applicable to a variety of methods that fall under parametric statistics.
The methods such as propagation of uncertainty and least squares parameter fitting that make a data-scientist life easy apply only to datasets with normal or normal-like distributions.
3. Since Gaussian(Normal) Distribution is easy to explain, the intuition behind the conclusion and summary of the test can be easily conveyed to people with little statistical knowledge.
4. The entire distribution is described by two numbers, the mean and variance.
5. Unlike much other distribution that changes their nature on transformation, a Gaussian tends to remain a Gaussian.
 - * The product of two Gaussian is a Gaussian
 - * The Sum of two independent Gaussian random variables is a Gaussian
 - * Convolution of Gaussian with another Gaussian is a Gaussian
 - * Fourier transform of Gaussian is a Gaussian

How to check Gaussian

- Histogram
- Q-Q plot
- KDE plot
- Skewness and Kurtosis
- Five Number Summary (This one just gives the glimpse and not the entire detail. Check out my blog if you want to know more about this concept)

Need for Data Transformation!!

- To more closely approximate a theoretical distribution that has nice statistical properties
- To spread data out more evenly -to make data distributions more symmetrical
- To make relationships between variables more linear
- To make data have more constant variance (homoscedastic)

Transformation (1) -- Log Transformation

- Coming soon ...



CHALMERS
UNIVERSITY OF TECHNOLOGY