План курса

1. Введение во врем. ряды;
2. Фильтрация, сглаживание
3. Линейные модели: AR, MA, ARIMA
4. Нейросетевые подходы: П. и А. В.Р.
5. А. и П. хаотические временные ряды
6. Фрактальный анализ.

Оценка: 4 лабы $\times 0.6 + 2$ экзамена $\times 0.4$

Time series is a sequence of measurements of certain quantity ordered in time. $y = \{y_t\}_{t=1}^{T}, y_t \in R^n$.

Continuous time series examples:

1. Economic: GDP, consumer price index;
2. Financial time series;
3. Biological time series: ECG, heart rate.

Time Series Decomposition (TSD)

Typical decomposiotion looks like:

$$y_t = T_t + S_t + R_t$$

T_t — trend, S_t — seasonality component, R_t — random fluctuations.

The decomposition can also take on the following forms:

$$y_t = T_t S_t R_t,$$

$$y_t = (T_t + S_t) R_t.$$

Decomposition algorythms:

I. Classical TSD (using moving averages).

$$\mathrm{ma}(y_t;m)=\frac{1}{m}\sum_{j=-k}^{k}y_t, m=2k+1$$

The window size has to be odd. Backward formula:

$$\mathrm{ma}(y_t;m)=\frac{1}{m}\sum_{j=-m}^{0}y_t,$$

Forward formula:

$$\mathrm{ma}(y_t;m)=\frac{1}{m}\sum_{j=0}^{m}y_t.$$

For $m=4$:

$$\mathrm{ma}(y_t;4)=\frac{1}{4}(y_{t-1},y_t,y_{t+1},y_{t+2})$$

MA over MA:

$$\begin{aligned}\mathrm{ma}(\mathrm{ma}(y_t,4);2)&=\frac{1}{2}[\mathrm{ma}(y_{t-1};4),\mathrm{ma}(y_t;4)]=\\&=\frac{1}{2}\left[\frac{1}{4}(y_{t-2},y_{t-1},y_t,y_{t+1})+\frac{1}{4}(y_{t-1},y_t,y_{t+1},y_{t+2})\right]=\\&=\frac{1}{8}y_{t-2}+\frac{1}{4}y_{t-1}+\frac{1}{4}y_t+\frac{1}{4}y_{t+1}+\frac{1}{8}y_{t+2}.\end{aligned}$$

This is used to: 1) smoothen the data; 2) extranct the trend.

Weighted moving average (WMA):

$$\mathrm{wma}(y_t;m)=\sum_{j=-k}^{k}y_{t+j}\cdot w_j, w_j\geq 0,\sum w_j=1.$$

So, the classical TSD algorithm is given as follows:

1. Compute trend component using either MA over MA if m is even or WMA if m is odd.

$$\hat{T}_t = \mathrm{ma}(y_t; m) \text{ or } \hat{T}_t = \mathrm{ma}(\mathrm{ma}(y_t; m); m).$$

2. Detrend the TS: $y_t - \hat{T}_t = S_t + R_t$.
3. Compute $\hat{S}_t$ by averaging detrended time series for a season.
4. $\hat{R}_t = y_t - \hat{S}_t - \hat{T}_t$ assuming that S_t the same of each season.

II. STL Decomposition (seasonal trend decomposition)

This algorithm realies on a technique called LOECS — a type of local regression for modeling and smoothing data $(x_i, y_i)_{i=1}^{m}$. Its key components are:

1. Kernel funciton. For example, Gaussian kernel $w_i = \exp\left(-\frac{(x_i - x)^2}{2\tau^2}\right)$.
2. Smoothing parameter τ. Smaller τ leads to narrower windows, larger τ — to wider windows and $\tau \to +\infty$ means that $w_i = 1$, hence leads to model becoming a simple linear regression.

Given data $(x_i, y_i)_{t=1}^{m}$ or $(t, y_t)_{t=1}^{T}$, the LOECS algorithm step-by-step:

1. Choose a kernel function kernel_fn and τ.
2. For all x_i:

2.1. Calculate $w_i = \mathrm{kernel\_fn}(x_i, x, \tau)$

2.2. Build weighted regression model. For example, weighted least squares: $L = \sum_{i=1}^{n} w_i (y_i - \Theta^T x_i)^2$, where $\Theta = (X^T W X)^{-1} X^T W y$.

2.3. Make predictions $\hat{y}(X)$ for X only.

2.4. "Forget" the model.

STL algorithm.

Input: $Y = \{y_1, ..., y_\tau\}$.

Parameters: n_p — # of outer iterations (1-2)

n_i — # of innter iterations (1-2)

n_l — trend smoothing parameter (smoothing parameter for LOECS)

n_s — seasonality smoothing parameter

n_o — residual smoothing parameter (optional, for residues R_t).

Outer loop: repeat n_p times.

1. Initialization:

   1) set trend $T^{(0)} = 0$ (initialize the approximation using MA for example);

   2) set weights $w = \{1, 1, ...1\}$ (optional, for residues).

2. Inner loop: repeat n_i times

2.1. Detrend time series: $D = Y - T$.

2.2. Compute seasonal component:

2.2.1. Split D subseries by seasons;

2.2.2. For each subseries apply the LOECS technique with $\tau = n_l$ and weights W.

2.2.3. Assemble the smoothed subseries into a seasonal component C.

2.2.4. Compute this C.

2.3. Update seasonal component $S = C$.

2.4. Deaseasonalize the data: $Y_{\text{deseasonalized}} = Y - S$

2.5. Update the trend: apply LOECS for $Y_{\text{deseasonalized}}$ with $\tau = n_l$ and “robust” weights w (obtain T).

3. Compute the residuals $R = Y - T - S$.
4. Update weights: recompute weights based on residues R to reduce the influence of outliers. Usually we sue Tuikey's biweight function.

Post-processing:

1) Normalize seasonality;

2) Smoothen the trend.

Result: T, S, R

STL is:

robust to outliers,

can model non-linear trends,

work with any seasonality.

How to update weights using Tuikey's biweight function?

1. Obtain the residuals $R = Y - S - T$
2. Compute median absolute deviation (MAD)

$$\mathrm{MAD} = \mathrm{median}(|R - \mathrm{median}(R)|).$$

Normalize: $s \approx 1.4826$, s — standard deviation(??????????)

3. Compute the normalized residuals: $u_i = \frac{R_i}{C \cdot S}$, where C is a tuning constant ($C = 4.685$).
4. Bisquare function $w_i = \begin{cases} (1-u_i)^2, |u_i| <1, \\ 0, |u_i| \geq 1. \end{cases}$.
5. If $S = 0$, then $w_i = 0$ (all residuals are the same). If $\mathrm{MAD} = 0$, but the residuals are not the same, we use STD instead of MAD.

For example, if $R = [0.1, -0.2, 3.0, -0.1, 10.0]$:

1. MAD: $\mathrm{median}(R) = 0.1$, hence $\mathrm{MAD} = \mathrm{median}(|R - 0.1|) = 0.3$

whatever yada-yada...

Stationarity and Ergoticity

Stationarity is a key feature of time series. There are several kinds of stationarity:

Strict stationarity: joint distribution of any segment of time series $\left(y_{t_1}, y_{t_2}, ..., y_{t_k}\right)$ is equivalent to $\left(y_{t_1+\tau}, y_{t_2+\tau}, ..., y_{t_k+\tau}\right)$ $\forall \tau$.

Weak stationarity: (erased)

Non-stationary time series:

1. Time servies with determinitstic trend:

$$y_t = \alpha + \beta t + \varepsilon_t, \varepsilon_t \sim N(0, \sigma^2).$$

Here, $\mathbb{F}[y_t^T] = \alpha + \beta t$.

2. (erased)

3. Random Walk:

$$\begin{aligned}
y_t &= y_{t-1} + \varepsilon_t,\ \varepsilon_t \sim N(0, \sigma^2), \quad \operatorname{cov}(\varepsilon_t, \varepsilon_s) = 0,\ t \neq s \\
y_1 &= y_0 + \varepsilon_1, \\
y_2 &= y_1 + \varepsilon_2 = y_0 + \varepsilon_1 + \varepsilon_2, \\
&\dots \\
y_t &= y_0 + \sum_{i=1}^{t} \varepsilon_i
\end{aligned}$$

So, $\mathbb{E}[y_t] = y_0,\ \mathbb{D}[y_t] = t\sigma^2$.

Some examples:

1. $y_t = S_t, \varepsilon_t \sim \text{iid } N(0, \sigma^2)$ – white noise. In this case,

$$\mathbb{E}[y_t] = 0,\ \mathbb{D}[y_t] = \varepsilon^2 < \infty \to \text{stationary}, \quad \operatorname{cov}(\varepsilon_t, \varepsilon_s) = 0$$

2. $$y_t = \beta y_{t-1} + \varepsilon_t,\ \beta \in (-1, 1),\ \varepsilon_t \sim \text{iid } N(0, \sigma^2)$$

(erased)

1.
$$\mathbb{E}[y_t] = \beta_1^t \mathbb{E}[y_0] + \hat{\beta}^{t-1}\mathbb{E}[\varepsilon_1] + ... + \mathbb{E}[\varepsilon_t]$$
$$= \beta_1^t y_0 \quad \text{if} \quad t \to \infty, \;\; \beta_1^t \to 0.$$

2.
$$\mathbb{D}[\beta^t y_0 + \beta^{t-1}\varepsilon_1 + ... + \varepsilon_t] =$$
$$= \beta^{2t-2}\mathbb{D}(\varepsilon_1) + \beta^{2t-4}\mathbb{D}(\varepsilon_2) + ... + \mathbb{D}[\varepsilon_t] =$$
$$= (\beta^{2t-2} + \beta^{2t-4} + ... + 1)\sigma^2$$
(erased)

3. $\operatorname{cov}\;(y_t, y_{t+1}) =$
$$= \operatorname{cov}(\beta^t y_0 + \beta^{t-1}\varepsilon_1 + ... + \varepsilon_t, \beta^{t+1}y_0 + \beta^t\varepsilon_1 + ... + \varepsilon_{t+1}) =$$
$$= \beta \operatorname{cov}(\varepsilon_t, \varepsilon_t) + \beta^3 \operatorname{cov}(\varepsilon_{t-1}, \varepsilon)... \text{ (erased)}$$

Unit root

$$y_t = \varphi \cdot y_{t-1} + \varepsilon_t, \;\; \varepsilon_i \overset{\text{iid}}{\sim} N(0, \sigma^2), \;\; \varphi \text{ is constant.}$$

1. $|\varphi| < 1$ means that the process is stationary;
2. $|\varphi| = 1$ is the unit root case, not stationary;
3. $|\varphi| > 1$ is a non-stationary or explosive time series.

Why unit root?

$$Ly_t = y_{t-1}, \;\; y_t = \varphi L y_t + S_t \to (1 - \varphi L) \text{ (erased)}$$

If $(1 - \varphi z) = 0$, $z = \frac{1}{\varphi} = 1 \to \varphi = 1$

2. Dickey-Fuller test

1)
$$y_t = \varphi y_{t-1} + \varepsilon_t$$
$$y_t - y_{t-1} = \varphi y_{t-1} - y_{t-1} + \varepsilon_t$$
$$\Delta y_t = (\varphi - 1) y_{t-1} + \varepsilon_t = \gamma y_{t-1} + \varepsilon_t.$$

2)

$H_0 : \gamma = 0 \ (\varphi = 1) \to \text{unit root} \to \text{non-stationary time series.}$

$H_1 : \gamma > 0 \ (\varphi < 1) \to \text{no unit root} \to \text{stationary process.}$

3) Evaluate γ by fitting regression:

$$\Delta y_t = \gamma y_{t-1} + \varepsilon_t t_{\text{stat}} = \frac{\hat{\gamma}}{\text{SE}(\hat{\gamma})}$$

4) Distributed Dickey-Fuller:

S.L. Crit.Val.

1% −3.43

5% −2.86

10% −2.57

5) If $t_{\text{stat}} < \text{crit. val.} \to H_0$ is rejected,

If $t_{\text{stat}} > \text{crit. val.} \to H_0$ is not rejected.

Modification

(erased)

1) $p \approx \sqrt[3]{T}, \ p \neq (?)\sqrt{T}.$

2) Test different p, choose p which gives you the “best” regression: BIC, AIC, MQIC.

4. KPSStat

1) KPSS assumes that the time series is dependant on $y_t = \xi_t + r_t + \varepsilon_t$, where ξ_t… (FINISH LATER!!!!!!)