Spring 2015 Statistics 153 (Time Series) : Lecture Two

Aditya Guntuboyina

22 January 2015

1 Last Class

- Data Examples
- White noise model
- Sample Autocorrelation Function and Correlogram

2 Trend Models

Many time series datasets show an increasing or decreasing trend. A simple model for such datasets is obtained by adding a deterministic trend function of time to white noise:

$$X_t = m_t + Z_t$$

Here m_t is a deterministic trend function and Z_t is white noise. There exist two main techniques for fitting this model to the data.

2.1 Parametric form for m_t and linear regression

Assume a simple parametric form for m_t, say linear or quadratic, and fit it via linear regression.

2.2 Smoothing

Here we estimate m_t without making any parametric assumptions about its form.

The idea is that to get m_t from $X_t = m_t + Z_t$, we need to eliminate Z_t. It is well-known that noise is eliminated by averaging. Consider

$$\hat{m}_t = \frac{1}{2q+1} \sum_{j=-q}^{q} X_{t+j}. \tag{1}$$

If m_t is linear on the interval $[t-q, t+q]$, then check that

$$\frac{1}{2q+1} \sum_{j=-q}^{q} m_{t+j} = m_t.$$

Thus if m_t is approximately linear over $[t-q, t+q]$, then

$$\hat{m}_t \approx m_t + \frac{1}{2q+1} \sum_{j=-q}^{q} Z_{t+j} \approx m_t.$$

The defining equation for $\hat{m}_t$ will have trouble when calculating averages near end-points. To counter, just define X_t to be X_1 for $t < 1$ and X_n for $t > n$.

$\hat{m}_t$ is also called the Simple Moving Average of X_t.

Key Question: How to choose the smoothing parameter q? Observe that:

$$\hat{m}_t = \frac{1}{2q+1} \sum_{j=-q}^{q} m_{t+j} + \frac{1}{2q+1} \sum_{j=-q}^{q} Z_{t+j}.$$

If q is very small, then the second term above is not quite small and so the trend estimate will also involve some noise component and therefore $\hat{m}_t$ will be very noisy. On the other hand, if q is large, then the assumption that m_t is linear on $[t-q, t+q]$ may not be quite true and thus, $\hat{m}_t$ may not be close to m_t. This is often referred to as the Bias-Variance tradeoff. Therefore q should be neither too small nor too large.

2.2.1 Parametric Curve Fitting versus Smoothing

Suppose there is good reason to believe that there is an underlying linear or quadratic trend function. In this case, is it still okay to use smoothing?

No, when there is really is an underlying linear trend, fitting a line gives a more precise estimate of the trend. On the other hand, the estimation of trend by smoothing only uses a few observations for each time point and the resulting estimate is not as precise. This is the price one has to pay for giving up the assumption of linearity. Of course, when there is no reason to believe in an underlying linear trend, it might not make sense at all to fit a line. Smoothing is the way to go in such cases.

2.3 More General Filtering

The smoothing estimate (1) of the trend function m_t is a special case of *linear filtering*. A linear filter converts the observed time series X_t into an estimate of the trend $\hat{m}_t$ via the linear operation:

$$\hat{m}_t = \sum_{j=-q}^{s} a_j X_{t+j}.$$

The numbers $a_{-q}, a_{-q+1}, \ldots, a_{-1}, a_0, a_1, \ldots, a_s$ are called the weights of the filter. The Smoothing method is clearly a special instance of filtering with $s = q$ and $a_j = 1/(2q+1)$ for $|j| \le q$ and 0 otherwise.

One can think of the filter as a (linear) *system* which takes the observed series X_t as input and produces the estimate of trend, $\hat{m}_t$ as output.

In addition to the choice $a_j = 1/(2q+1)$ for $|j| \le q$, there are other choice of filters that people commonly use.

(1) **Binomial Weights**: Based on the following idea. When we are estimating the value of the trend m_t at t, it makes sense to give a higher weight to X_t compared to $X_{t\pm1}$ and a higher weight to $X_{t\pm1}$ compared to $X_{t\pm2}$ and so on. An example of such weights are:

$$a_j = 2^{-q} \binom{q}{q/2+j} \qquad \text{for } j = -q/2, -q/2+1, \ldots, -1, 0, 1, \ldots, q/2.$$

As in usual smoothing, choice of q is an issue here.

(2) **Spencer's 15 point moving average**: We have seen that simple moving average filter leaves linear functions untouched. Is it possible to design a filter which leaves higher order polynomials untouched? For example, can we come up with a filter which leaves all quadratic polynomials untouched. Yes!

For a filter with weights a_j to leave all quadratic polynomials untouched, we need the following to be satisfied for every quadratic polynomial m_t:

$$\sum_j a_j m_{t+j} = m_t \qquad \text{for all } t$$

In other words, if $m_t = \alpha t^2 + \beta t + \gamma$, we need

$$\sum_j a_j \left(\alpha(t+j)^2 + \beta(t+j) + \gamma\right) = \alpha t^2 + \beta t + \gamma \qquad \text{for all } t.$$

Simplify to get

$$\alpha t^2 + \beta t + \gamma = (\alpha t^2 + \beta t + \gamma)\sum_j a_j + (2\alpha t + \beta)\sum_j j a_j + \alpha \sum_j j^2 a_j \qquad \text{for all } t.$$

This will clearly be satisfied if

$$\sum_j a_j = 1 \qquad \sum_j j a_j = 0 \qquad \sum_j j^2 a_j = 0. \tag{2}$$

An example of such a filter is Spencer's 15 point moving average defined by

$$a_0 = \frac{74}{320}, a_1 = \frac{67}{320}, a_2 = \frac{46}{320}, a_3 = \frac{21}{320}, a_4 = \frac{3}{320}, a_5 = \frac{-5}{320}, a_6 = \frac{-6}{320}, a_7 = \frac{-3}{320}$$

and $a_j = 0$ for $j > 7$. Also the filter is symmetric in the sense that $a_{-1} = a_1, a_{-2} = a_2$ and so on. Check that this filter satisfies the condition (2).

Because this is a symmetric filter, it can be checked that it allows all cubic polynomials to pass unscathed as well.

(3) **Exponential Smoothing**: Quite a popular method of smoothing (wikipedia has a big page on this). It is also used as a forecasting technique.

To obtain $\hat{m}_t$ in this method, one uses only the *previous* observations $X_{t-1}, X_{t-2}, X_{t-3}, \ldots$. The weights assigned to these observations *exponentially decrease* the further one goes back in time. Specifically,

$$\hat{m}_t := \alpha X_{t-1} + \alpha(1-\alpha)X_{t-2} + \alpha(1-\alpha)^2 X_{t-3} + \cdots + \alpha(1-\alpha)^{t-2}X_1 + (1-\alpha)^{t-1}X_0.$$

Check that the weights add up to 1. α is a parameter that determines the amount of smoothing (α here is analogous to q in smoothing). If α is close to 1, there is very little smoothing and vice versa.

A more succinct way of writing out the formula for $\hat{m}_t$ is via the recursion:

$$\hat{m}_1 := X_0 \quad \text{and} \quad \hat{m}_t := \alpha X_{t-1} + (1-\alpha)\hat{m}_{t-1} \text{ for } t > 1.$$

2.4 Isotonic Trend Estimation

Isotonic estimation is a very elegant way of estimating **monotone** trends.

Suppose we want to estimate m_t from the model $X_t = m_t + Z_t$. We have reasons to believe that m_t is non-decreasing. This is quite often the case. The Isotonic estimator $\hat{m}_t$ for m_t is the solution to the following minimization problem: Minimize $\sum_{t=1}^{n}(X_t - a_t)^2$ under the constraint $a_1 \leq \cdots \leq a_n$.

This minimization problem is a convex optimization problem. It can be solved highly efficiently (in $O(n \log n)$ time using the Pool Adjacent Violators algorithm). This often produces very good estimates $\hat{m}_t$. Isotonic estimator is one of the early examples of estimation procedures in statistics that are based on convex optimization.

Isotonic estimates have the following advantages over smoothing methods:

1. One gets $\hat{m}_t$ for all t unlike moving average smoothing which does yield $\hat{m}_t$ at end-points.
2. More importantly, there is no need to select a smoothing parameter. Smoothing parameter selection is a very tricky issue in general and isotonic methods completely avoid this issue.

The problem however is that this method only works in situations when we know that trend to be either non-decreasing or non-increasing.

3 Differencing for Trend Elimination

So far, we looked at trend models: $X_t = m_t + Z_t$ where m_t is a deterministic trend function and $\{Z_t\}$ is white noise. The residuals obtained after fitting the trend function m_t in the model $X_t = m_t + Z_t$ are studied to see if they are white noise or have some dependence structure that can be exploited for prediction.

Suppose that the goal is just to produce such detrended residuals. Differencing is a simple technique which produces such de-trended residuals.

One just looks at $Y_t = X_t - X_{t-1}, t = 2, \dots, n$. If the trend m_t in $X_t = m_t + Z_t$ is linear, then this operation simply removes it because if $m_t = \alpha t + b$, then $m_t - m_{t-1} = \alpha$ so that $Y_t = \alpha + Z_t - Z_{t-1}$.

Suppose that the first differenced series Y_t appears like white noise. What then would be a reasonable forecast for the original series: X_{n+1}? Because Y_t is like white noise, we forecast Y_{n+1} by the sample mean $\bar{Y} := (Y_2 + \cdots + Y_n)/(n-1)$. But since $Y_{n+1} = X_{n+1} - X_n$, this results in the forecast $X_n + \bar{Y}$ for X_{n+1}.

Sometimes, even after differencing, one can notice a trend in the data. In that case, just difference again. It is useful to follow the notation ∇ for differencing:

$$\nabla X_t = X_t - X_{t-1} \qquad \text{for } t = 2, \dots, n$$

and second differencing corresponds to

$$\nabla^2 X_t = \nabla(\nabla X_t) = \nabla X_t - \nabla X_{t-1} = X_t - 2X_{t-1} + X_{t-2} \qquad \text{for } t = 3, \dots, n.$$

It can be shown that quadratic trends simply disappear with the operation ∇^2. Suppose the data $\nabla^2 X_t$ appear like white noise, how would you obtain a forecast for X_{n+1}?

Differencing is a quick and easy way to produce detrended residuals and is a key component in the ARIMA forecasting models (later). A problem however is that it does not result in any estimate for the trend function m_t.