STAT 153 - Introduction to Time Series Lecture Two

Fall 2022, UC Berkeley

Aditya Guntuboyina

August 30, 2022

1 Last Class

In the last lecture, we looked at the Gaussian White Noise model for modeling observed time series. This model simply says that the observations are i.i.d $N(0, \sigma^2)$ for some unknown parameter $\sigma > 0$. This model is not directly applicable for observed time series mainly because of two reasons:

1. Observed time series often have different kinds of trends which are not explained by the Gaussian white noise model.
2. Even when there are no trends, observed time series tend to be autocorrelated which is also not present in data generated according to the Gaussian White Noise model. The presence of autocorrelation in an observed time series dataset can be detected by plotting the correlogram.

Today we shall discuss simple models for modeling the trend in observed time series.

2 A note on notation

From now on, we shall use one of the following two kinds of notation:

1. **Regularly sampled times**: In the case where the observed time series is sampled at regular or equally spaced time points, we shall denote the observed time series by $y_0, \ldots, y_T$.
2. **Irregularly sampled times**: In the case where the time series is sampled at possibly irregular time points $t_1, \ldots, t_n$, we shall denote the observed data by $y_1, \ldots, y_n$. In other words, y_i is the observed value of the time series at time t_i for $i = 1, \ldots, n$.

While describing models for observed time series, we shall work with random variables for which we use the similar notation. In the regularly sampled case, the random variables will be denoted by $Y_0, \ldots, Y_T$ and in the irregularly sampled case, the random variables with be denoted by $Y_1, \ldots, Y_n$.

3 Trend Models

The simplest models for trend are obtained by simply taking a parametric function of time (such as linear or quadratic) and then adding it to white noise. More specifically, a linear trend model is given by

$$Y_i = \beta_0 + \beta_1 t_i + Z_i$$

for $i = 1, \ldots, n$ (we are considering the irregular sampled times setting here) where

$$Z_1, \ldots, Z_n \overset{\text{i.i.d}}{\sim} N(0, \sigma^2).$$

There are three parameters in this model β_0, β_1 and σ^2.

How to fit this model to the observed data $(t_1, y_1), \ldots, (t_n, y_n)$? We shall do this via the Bayesian approach which generalizes very naturally to more complicated models.

3.1 Usual Least Squares Analysis of the Linear Trend Model

The parameters β_0 and β_1 are estimated by minimizing the least squares criterion:

$$S(\beta_0, \beta_1) := \sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 t_i)^2 .$$

This minimization can be carried out by differentiating $S(\beta_0, \beta_1)$ with respect to β_0 and β_1 and then setting the derivatives to zero. It is notationally quite convenient to do this analysis in matrix notation. Let

$$Y = \begin{pmatrix} y_1 \\ \cdot \\ \cdot \\ \cdot \\ y_n \end{pmatrix} \quad \text{and} \quad X = \begin{pmatrix} 1 & t_1 \\ \cdot & \cdot \\ \cdot & \cdot \\ \cdot & \cdot \\ 1 & t_n \end{pmatrix} \quad \text{and} \quad \beta = \begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix}.$$

Note that Y is a $n \times 1$ vector, X is a $n \times 2$ matrix and β is a 2×1 vector. With this notation, we can write

$$\begin{aligned} S(\beta_0, \beta_1) &= \sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 t_i)^2 \\ &= \|Y - X\beta\|^2 \\ &= (Y - X\beta)^T (Y - X\beta) = Y^T Y - Y^T X\beta - \beta^T X^T Y + \beta^T X^T X\beta \end{aligned}$$

The partial derivatives $\frac{\partial S}{\partial \beta_0}$ and $\frac{\partial S}{\partial \beta_1}$ can be computed in vector form as

$$\begin{pmatrix} \frac{\partial S}{\partial \beta_0} \\ \frac{\partial S}{\partial \beta_1} \end{pmatrix} = 2X^T X\beta - 2X^T Y.$$

Equating the above to zero, we obtain the so-called "normal equations" of linear regression:

$$X^T X\beta = X^T Y.$$

If we now assume that $X^T X$ is invertible (this will usually be the case), we obtain the least squares estimates:

$$\hat{\beta} = \begin{pmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \end{pmatrix} = (X^T X)^{-1} X^T Y$$

The smallest possible value of $S(\beta_0, \beta_1)$ is then:

$$S(\hat{\beta}_0, \hat{\beta}_1) = \sum_{i=1}^{n} (y_i - \hat{\beta}_0 - \hat{\beta}_1 t_i)^2 = \|Y - X\hat{\beta}\|^2.$$

This quantity is known as the Residual Sum of Squares (RSS).

The parameter σ is estimated via

$$\hat{\sigma} := \sqrt{\frac{RSS}{n - \text{rank}(X)}}$$

where $\text{rank}(X)$ denotes the rank of the X matrix. Under the assumption that $X^T X$ is invertible, the rank of X is simply equal to 2.

The quantities $\hat{\beta}_0, \hat{\beta}_1, \hat{\sigma}$ provide point estimates of the unknown parameters β_0, β_1 and σ. More work is needed for uncertainty quantification. Specifically, one proves that

$$\hat{\beta} \sim N(\beta, \sigma^2 (X^T X)^{-1}) \quad \text{and} \quad (n - \text{rank}(X)) \frac{\hat{\sigma}^2}{\sigma^2} = RSS \sim \chi^2_{n-\text{rank}(X)}$$

and that $\hat{\beta}$ and $\hat{\sigma}$ are independent. In the above, $N(\beta, \sigma^2 (X^T X)^{-1})$ denotes the multivariate normal distribution (with mean β and covariance $\sigma^2 (X^T X)^{-1}$). These two properties together imply that

$$\frac{\hat{\beta}_0 - \beta_0}{\hat{\sigma}\sqrt{(X^T X)^{11}}} \sim t_{n-\text{rank}(X)} \quad \text{and} \quad \frac{\hat{\beta}_1 - \beta_1}{\hat{\sigma}\sqrt{(X^T X)^{22}}} \sim t_{n-\text{rank}(X)}$$

where $(X^T X)^{11}$ and $(X^T X)^{22}$ denote the $(1,1)$ and $(2,2)$ entries of the matrix $(X^T X)^{-1}$. These results allow for uncertainty quantification for the parameters β_0 and β_1.

3.2 Bayesian Analysis of the Linear Trend Model

A natural Bayesian analysis of this model leads to exactly the same answers as usual least squares. This is demonstrated below.

The first step is to select a prior for the unknown parameters β_0, β_1, σ. A reasonable prior reflecting ignorance is

$$\beta_0, \beta_1, \log \sigma \overset{\text{i.i.d}}{\sim} \text{Unif}(-C, C)$$

for a large number C (the exact value of C will not matter in the following calculations). Note that as σ is always positive, we have made the uniform assumption on $\log \sigma$ (by the change of variable formula, the density of σ would be given by $f_\sigma(x) = f_{\log \sigma}(\log x)\frac{1}{x} = \frac{I\{-C < \log x < C\}}{2Cx} = \frac{I\{e^{-C} < x < e^C\}}{2Cx}$.

The joint posterior for all the unknown parameters β_0, β_1, σ is then given by (below we write the term "data" for $Y_1 = y_1, \ldots, Y_n = y_n$):

$$f_{\beta_0, \beta_1, \sigma | \text{data}}(\beta_0, \beta_1, \sigma) \propto f_{Y_1, \ldots, Y_n | \beta_0, \beta_1, \sigma}(y_1, \ldots, y_n) f_{\beta_0, \beta_1, \sigma}(\beta_0, \beta_1, \sigma).$$

The two terms on the right hand side above are

$$\begin{aligned} f_{Y_1,\ldots,Y_n|\beta_0,\beta_1,\sigma}(y_1,\ldots,y_n) &\propto \prod_{i=1}^{n} f_{Y_i|\beta_0,\beta_1,\sigma}(y_i) \\ &= \prod_{i=1}^{n} f_{Z_i|\beta_0,\beta_1,\sigma}(y_i-\beta_0-\beta_1 t_i) \\ &= \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}\sigma}\exp\left(-\frac{(y_i-\beta_0-\beta_1 t_i)^2}{2\sigma^2}\right) \\ &\propto \sigma^{-n}\exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i-\beta_0-\beta_1 t_i)^2\right), \end{aligned}$$

and

$$\begin{aligned} f_{\beta_0,\beta_1,\sigma}(\beta_0,\beta_1,\sigma) &= f_{\beta_0}(\beta_0) f_{\beta_1}(\beta_1) f_{\sigma}(\sigma) \\ &\propto \frac{I\{-C<\beta_0<C\}}{2C}\frac{I\{-C<\beta_1<C\}}{2C}\frac{I\{e^{-C}<\sigma<e^{C}\}}{2C\sigma} \\ &\propto \frac{1}{\sigma} I\left\{-C<\beta_0,\beta_1,\log\sigma<C\right\}. \end{aligned}$$

We thus obtain

$$\begin{aligned} &f_{\beta_0,\beta_1,\sigma|\text{data}}(\beta_0,\beta_1,\sigma) \\ &\propto \sigma^{-n-1}\exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i-\beta_0-\beta_1 t_i)^2\right) I\left\{-C<\beta_0,\beta_1,\log\sigma<C\right\}. \end{aligned}$$

The above is the joint posterior over β_0,β_1,σ. The posterior over only the main parameters β_0,β_1 can be obtained by integrating the parameter σ as follows:

$$\begin{aligned} f_{\beta_0,\beta_1|\text{data}}(\beta_0,\beta_1) &= \int f_{\beta_0,\beta_1,\sigma|\text{data}}(\beta_0,\beta_1,\sigma)d\sigma \\ &\propto I\{-C<\beta_0,\beta_1<C\}\int_{e^{-C}}^{e^{C}} \sigma^{-n-1}\exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i-\beta_0-\beta_1 t_i)^2\right)d\sigma. \end{aligned}$$

When C is large, the above integral can be evaluated from 0 to ∞ which gives

$$f_{\beta_0,\beta_1|\text{data}}(\beta_0,\beta_1) \propto I\{-C<\beta_0,\beta_1<C\}\int_{0}^{\infty} \sigma^{-n-1}\exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i-\beta_0-\beta_1 t_i)^2\right)d\sigma.$$

The change of variable

$$s = \frac{\sigma}{\sqrt{\sum_{i=1}^{n}(y_i-\beta_0-\beta_1 t_i)^2}}$$

allows us to write the integral as

$$\begin{aligned} &\int_{0}^{\infty} \sigma^{-n-1}\exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i-\beta_0-\beta_1 t_i)^2\right)d\sigma \\ &= \left(\sum_{i=1}^{n}(y_i-\beta_0-\beta_1 t_i)^2\right)^{-n/2}\int_{0}^{\infty} s^{-n-1}\exp\left(-\frac{1}{2s^2}\right)ds \propto \left(\sum_{i=1}^{n}(y_i-\beta_0-\beta_1 t_i)^2\right)^{-n/2}. \end{aligned}$$

The posterior density of (β_0,β_1) is thus

$$f_{\beta_0,\beta_1|\text{data}}(\beta_0,\beta_1) \propto I\{-C<\beta_0,\beta_1<C\}\left(\sum_{i=1}^{n}(y_i-\beta_0-\beta_1 t_i)^2\right)^{-n/2}. \tag{1}$$

We shall show in the next class that this density is related to the multivariate t-distribution (`https://en.wikipedia.org/wiki/Multivariate_t-distribution`).