Time Series Data

A time series is a sequence of observations taken sequentially in time. A discrete time series is usually sampled or accumulated from a continuous series. Examples in the text include

- Chemical process concentration reading: every 2 hours.
- IBM common stock closing prices: daily, May 17, 1961 – November 2, 1962.
- International airline passengers: monthly totals, January 1949 – December 1960.

A time series can be deterministic or stochastic. We will be concerned with the analysis of stochastic time series.

Time Series in R

R is an open-source programming environment for data analysis and graphics. R is packaged nicely for all major platforms including Unix, Windows, and Mac. Check `http://cran.r-project.org`.

Among the most useful R commands are `help.start()`, which loads the on-line documentations, and `q()`, which quits R.

A numeric vector can be made into an `ts` object by the `ts` function. The series can be plotted by `plot`. The time stamps can be moved by `lag`. The difference at lag 1 can be obtained through `diff`.

```
air<-ts(scan("airline-pass"),start=1949,frequency=12)
plot(air); plot.default(air)
lag(air); lag(air,12)
diff(air)
```

Times Series Analysis

Successive observations in a time series tend to be correlated. The modeling of time dependency and the forecasting of future are important topics of time series analysis.

```
chem<-scan("chem-conc"); plot(ts(chem))
plot(chem[-197],chem[-1])
```

A dynamic system transfers a input series into an output series. The understanding of the system dynamics is also an important topic of time series analysis.

```
chem2<-3*chem[-1]-2*chem[-197]
chem3<-chem2+rnorm(chem2)
plot(ts(chem2)); plot(ts(chem3))
plot(cbind(ts(chem),ts(chem3)))
```

Probability: Basic Concepts

The behavior of a r.v. X is characterized by its distribution function $F(x) = P(X \leq x)$. The first two moments are

$$\mu_x = \int_{-\infty}^{\infty} x\, dF(x), \quad \sigma_x^2 = \int_{-\infty}^{\infty} (x - \mu_x)^2 dF(x),$$

where the integral is in the Riemann-Stieltjes sense. For continuous r.v., $dF(x) = f(x)dx$, where $f(x)$ is the density.

The joint distribution of two r.v.'s (X, Y) is given by

$$F(x, y) = P(X \leq x, Y \leq y).$$

When the joint density $f(x, y) = \partial^2 F(x, y)/\partial x\, \partial y$ exists, the conditional density of $(Y|X = x)$ is given by $f(y|x) = f(x, y)/f_x(x)$, where $f_x(x) = \int_{-\infty}^{\infty} f(x, y)dy$ is the marginal density of X. X and Y are independent if $F(x, y) = F_x(x)F_y(y)$, or $f(x, y) = f_x(x)f_y(y)$.

Matrix Algebra: Block Matrix

Block matrices add and multiply just as ordinary matrices, so long as the dimensions match. For example,

$$\begin{pmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{C} & \mathbf{D} \end{pmatrix} \begin{pmatrix} \mathbf{a} \\ \mathbf{b} \end{pmatrix} = \begin{pmatrix} \mathbf{Aa} + \mathbf{Bb} \\ \mathbf{Ca} + \mathbf{Db} \end{pmatrix}$$

Let $\boldsymbol{\Sigma} = \left(\begin{smallmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{smallmatrix}\right)$ be positive definite, where $\boldsymbol{\Sigma}_{11}$ and $\boldsymbol{\Sigma}_{22}$ are square, and write $\mathbf{E} = \boldsymbol{\Sigma}_{22} - \boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12}$. One has

$$\begin{pmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{pmatrix}^{-1} = \begin{pmatrix} \boldsymbol{\Sigma}_{11}^{-1} + \boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12}\mathbf{E}^{-1}\boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1} & -\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12}\mathbf{E}^{-1} \\ -\mathbf{E}^{-1}\boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1} & \mathbf{E}^{-1} \end{pmatrix}$$

$$= \begin{pmatrix} \mathbf{I} & -\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12} \\ \mathbf{O} & \mathbf{I} \end{pmatrix} \begin{pmatrix} \boldsymbol{\Sigma}_{11}^{-1} & \mathbf{O} \\ \mathbf{O} & \mathbf{E}^{-1} \end{pmatrix} \begin{pmatrix} \mathbf{I} & \mathbf{O} \\ -\boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1} & \mathbf{I} \end{pmatrix}$$

Multivariate Normal Distribution

A n-dimensional normal distribution $N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ has a density

$$f(\mathbf{x}|\boldsymbol{\mu}, \boldsymbol{\Sigma}) = \frac{1}{(2\pi)^{n/2}|\boldsymbol{\Sigma}|^{1/2}} \exp\{-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})\}.$$

One has $E[\mathbf{X}] = \boldsymbol{\mu}$ and $\text{cov}[\mathbf{X}] = \text{cov}[\mathbf{X}, \mathbf{X}^T] = \boldsymbol{\Sigma}$.

Partition $\mathbf{x} - \boldsymbol{\mu} = \begin{pmatrix} \mathbf{x}_1 - \boldsymbol{\mu}_1 \\ \mathbf{x}_2 - \boldsymbol{\mu}_2 \end{pmatrix} = \begin{pmatrix} \mathbf{z}_1 \\ \mathbf{z}_2 \end{pmatrix}$ and $\boldsymbol{\Sigma} = \begin{pmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{pmatrix}$. Using results on the previous slide, one has $|\boldsymbol{\Sigma}| = |\boldsymbol{\Sigma}_{11}|\,|E|$ and

$$\begin{aligned}(\mathbf{x}-\boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu}) = {} & \mathbf{z}_1^T\boldsymbol{\Sigma}_{11}^{-1}\mathbf{z}_1 \\ & + (\mathbf{z}_2 - \boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}\mathbf{z}_1)^T\mathbf{E}^{-1}(\mathbf{z}_2 - \boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}\mathbf{z}_1).\end{aligned}$$

It follows that the conditional density $f(\mathbf{x}_2|\mathbf{x}_1)$ has mean $\boldsymbol{\mu}_2 + \boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}(\mathbf{x}_1 - \boldsymbol{\mu}_1)$ and covariance $\mathbf{E} = \boldsymbol{\Sigma}_{22} - \boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12}$. $\mathbf{X}_1$ and $\mathbf{X}_2$ are independent if and only if $\boldsymbol{\Sigma}_{12} = \mathbf{O}$.

Linear Models

Linear regression and ANOVA are collectively known as linear models, which have the standard form

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}.$$

where $E[\boldsymbol{\epsilon}] = \mathbf{0}$. Assuming $\text{cov}[\mathbf{Y}] = \sigma^2\mathbf{I}$, one minimizes the least squares criterion $(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})^T(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})$ to obtain

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{Y}; \quad \hat{\mathbf{Y}} = \mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{Y} = \mathbf{H}\mathbf{Y}.$$

The hat matrix $\mathbf{H}$ is idempotent, $\mathbf{H}^2 = \mathbf{H}$.

Assuming $\text{cov}[\mathbf{Y}] = \sigma^2\mathbf{W}^{-1}$, one minimizes the weighted least squares criterion $(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})^T\mathbf{W}(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})$ to obtain

$$\hat{\boldsymbol{\beta}}_W = (\mathbf{X}^T\mathbf{W}\mathbf{X})^{-1}\mathbf{X}^T\mathbf{W}\mathbf{Y}.$$

Maximum Likelihood Estimates

Consider a parametric density function $f(\mathbf{x}|\boldsymbol{\theta})$. Plugging in the observations $\mathbf{X}$, $L(\boldsymbol{\theta}|\mathbf{X}) = f(\mathbf{X}|\boldsymbol{\theta})$ as a function of $\boldsymbol{\theta}$ is called the likelihood function. The $\hat{\boldsymbol{\theta}}$ that maximizes $L(\boldsymbol{\theta}|\mathbf{X})$ is called a maximum likelihood estimate (MLE) of $\boldsymbol{\theta}$.

Write $l(\boldsymbol{\theta}|\mathbf{X}) = \log L(\boldsymbol{\theta}|\mathbf{X})$ the log likelihood. $I(\boldsymbol{\theta}) = -\partial^2 l/\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^T$ is the (observed) information matrix. For large samples, it usually holds that $\hat{\boldsymbol{\theta}} \overset{asy.}{\sim} N(\boldsymbol{\theta}_0, I^{-1}(\hat{\boldsymbol{\theta}}))$, where $\boldsymbol{\theta}_0$ generated the data.

Let $\hat{\boldsymbol{\theta}}$ be the MLE in the parameter space Θ and $\hat{\boldsymbol{\theta}}_0$ be the MLE in $\Theta_0 \subset \Theta$. If $\boldsymbol{\theta}_0 \in \Theta_0$, It usually holds that

$$2\big\{l(\hat{\boldsymbol{\theta}}|\mathbf{X}) - l(\hat{\boldsymbol{\theta}}_0|\mathbf{X})\big\} \overset{asy.}{\sim} \chi^2_\nu,$$

where $\nu = \dim(\Theta) - \dim(\Theta_0)$. This yiels the likelihood ratio tests.