Handout 7

[Chapter 2.4 in Brockwell and Davis.]
Estimation of mean, autocovariances and autocorrelations

Recall that a stationary process with mean μ can be described by two properties

a) $E(X_t) = \mu$, for all t,

b) $Cov(X_t, X_{t+j})$ is the same for all t (for any given positive integer j).

[Recall that properties (b) and (c) of a stationary series given on page 1 in Handout 6 (under the heading "Mathematical set up") can be combined as one property: "$Cov(X_t, X_{t+j})$ is the same for all t (for any given positive integer j)"].

The common value of $Cov(X_t, X_{t+j})$ is denoted by $\gamma(j)$ (and is called autocovariance of lag j). Also recall that $Corr(X_t, X_{t+j})$ is the same for all t and this common value is called the autocorrelation and is denoted by $\rho(j)$. This note is concerned with the issue of estimating μ, $\gamma(j)$ and $\rho(j)$.

Consider the series "annual precipitation in lake Michigan". This observed series $\{X_t\}$ looks stationary and it seems to fluctuate around a constant value μ. Clearly, it is of interest to know the avearge annual precipitation in lake Michigan during the period 1900-1986.

Often the observed series $\{Y_t\}$ has a trend (and/or seasonal components), as in the temperature data, then the practice is to first estimate the trend and then subtract it from the observed series in order to get an estimate of the stationary component X_t. Mathematically, if $Y_t = m_t + X_t$, we first estimate m_t, call the estimate $\hat{m}_t$ and then obtain $\hat{X}_t = Y_t - \hat{m}_t$. Then $\hat{X}_t$ is an estimate of X_t. Now, if the trend has been modeled and estimated by using a polynomial, i.e., $m_t = \beta_0 + \beta_1 t + \cdots + \beta_p t^p$, then mean of $\{\hat{X}_t\}$ is zero, and hence there is no need to estimate its mean. But if the trend has been estimated by the loess method, then the mean of $\{\hat{X}_t\}$ is close to zero, but not necessarily equal to zero. In such a case, we may still need to estimate the mean.

Why do we need to estimate the autocovariances and autocorrelations?

Recall that we are often concerned with forecasting, i.e. we want to forecast X_{n+h} when the given the available data $\{X_1, ..., X_n\}$. If the sequence $\{X_t\}$ is white noise, then knowing the past does not help us in forecasting the future. If the future X_{n+h} depends on the past, then in order to forecast it we need to have an idea of the relationships of X_t's to one another. Autocorrelations (and autocovariances) can provide useful clue to the nature of the relationships. We will explore this issue in detail later.

Estimation of μ.

If $\{X_t\}$ is stationary with mean μ, then an estimate of μ is given by the sample average $\bar{X} = (1/n)\sum_{1\le t\le n} X_t$. The following result is useful

$$E(\bar{X}) = \mu,\ Var(\bar{X}) = \tau_n^2/n, where$$
$$\tau_n^2 = \sum_{h=-(n-1)}^{n-1}(1-|h|/n)\gamma(h) = \gamma(0) + 2\sum_{h=1}^{n-1}(1-h/n)\gamma(h).$$

If $\gamma(h)$ is negligible for large h, then it can be shown that

$$\tau_n^2 \approx \gamma(0) + 2\sum_{h=1}^{\infty}\gamma(h) = \gamma(0)[1+2\sum_{h=1}^{\infty}\rho(h)].$$

Here and elsewhere, the notation " $\approx$ " is used to denote the phrase "approximately equal to". Though the last expression looks formidable, it can have simple expressions in some cases (Examples 1 and 2 below).

The following result holds if $\{X_t\}$ are normally distributed

$$\bar{X} \sim N(\mu, \tau_n^2/n).$$

Even if $\{X_t\}$ are not normally distributed, the above result still is true approximately when n is large, under some technical mathematical assumptions. A consequence of this result is that we can construct approximate confidence intervals for μ. For instance if we have an estimate $\hat{\tau}_n$ of τ_n, then an approximate 95% confidence interval for μ is given by $\bar{X} \pm 1.96\hat{\tau}_n/\sqrt{n}$.

Estimation of τ_n :

A general method of estimating τ_n can be done by using estimates of the autocovariances. If $\hat{\gamma}(j)$, $j = 0, 1, ...$ are the autocovariances, then it is reasonable to estimate τ_n^2 by

$$\hat{\tau}_n^2 = \hat{\gamma}(0) + 2\sum_{h=1}^{L}(1-h/n)\hat{\gamma}(h) = \hat{\gamma}(0)[1+2\sum_{h=1}^{L}(1-h/n)\hat{\rho}(h)], \tag{1}$$

where L is some (large) integer, e.g., $L = \sqrt{n}$ or the integer part of $\sqrt{n}$. Sometimes the value of L can be guessed by looking at the ACF plot. Choose L so that all the autocorrelations of order $L+1$ or higher seem to be negligible. [Note that we have used the relation $\hat{\rho}(h) = \hat{\gamma}(h)/\hat{\gamma}(0)$, and hence $\hat{\gamma}(h) = \hat{\gamma}(0)\hat{\rho}(h)$.]

For the series "annual precipitation in Lake Michigan, 1900-1986", we have

$$\bar{X} = 31.509,\ n = 87,\ L = 10,\ \hat{\tau}_n^2 = \hat{\gamma}(0) + 2\sum_{h=1}^{L}(1-h/n)\hat{\gamma}(h) = 13.2690,\ \hat{\tau}_n = 3.6427.$$

So a 95% confidence interval for μ is approximately

$$\bar{X} \pm 1.96\hat{\tau}_n/\sqrt{n}, i.e., 31.509 \pm (1.96)(3.6427)/\sqrt{87}, i.e., 31.509 \pm 0.765, i.e., (30.74, 32.27).$$

Special cases:

In some cases it may be possible to find simple expressions for τ_n if there is some reasonably simple model for the sequence $\{X_t\}$.

Example 1. If $\{X_t\}$ can be described by a moving average model, then estimation of τ_n is particularly simple. If we have an MA(1) model, then all the autocovarianes (and autocorrelations) of order 2 or higher are zero. For an $MA(2)$ model, then all the autocovarianes (and autocorrelations) of order 3 or higher are zero. Similarly, for an $MA(3)$ model, then all the autocovarianes (and autocorrelations) of order 4 or higher are zero. In such cases,

$$\begin{aligned}
&MA(1) : \tau_n^2 = \gamma(0) + 2(1-1/n)\gamma(1), \hat{\tau}_n^2 = \hat{\gamma}(0) + 2(1-1/n)\hat{\gamma}(1) \\
&MA(2) : \tau_n^2 = \gamma(0) + 2(1-1/n)\gamma(1) + 2(1-2/n)\gamma(2), \\
&\qquad \hat{\tau}_n^2 = \hat{\gamma}(0) + 2(1-1/n)\hat{\gamma}(1) + 2(1-2/n)\hat{\gamma}(2), \\
&MA(3) : \tau_n^2 = \gamma(0) + 2(1-1/n)\gamma(1) + 2(1-2/n)\gamma(2) + 2(1-3/n)\gamma(3), \\
&\qquad \hat{\tau}_n^2 = \hat{\gamma}(0) + 2(1-1/n)\hat{\gamma}(1) + 2(1-2/n)\hat{\gamma}(2) + 2(1-3/n)\hat{\gamma}(3).
\end{aligned}$$

For the Lake Michigan data if we fit an $MA(1)$ model: $X_t - \mu = \varepsilon_t + \theta\varepsilon_{t-1}$, then using R (some R commandsare given on the on the last page) we get

$$\begin{aligned}
&\hat{\mu} = \bar{X} = 31.509,\ \hat{\theta} = -0.0673, s(\hat{\theta}) = 0.1041, \hat{\sigma}^2 = 9.907, \\
&\hat{\gamma}(0) = 9.9554, \hat{\gamma}(1) = -0.7085, \\
&\hat{\tau}_n^2 = \hat{\gamma}(0) + 2(1-1/n)\hat{\gamma}(1) = 9.9554 + 2(1-1/87)(-0.7085) = 8.5533, \hat{\tau}_n = 2.925.
\end{aligned}$$

So if it is assumed that an $MA(1)$ model is reasonable for the lake Michigan data, then an approximate 95% confidence interval for μ is given by

$$\bar{X} \pm 1.96\hat{\tau}_n/\sqrt{n}, i.e., 31.509 \pm (1.96)(2.925/\sqrt{87}, i.e., 31.509 \pm 0615, i.e., (30.89, 32.12).$$

Example 2. If $\{X_t\}$ is $AR(1)$, i.e.,

$$X_t - \mu = \phi(X_{t-1} - \mu) + \varepsilon_t,$$

where $\{\varepsilon_t\}$ is white noise with variance σ^2, then we know that

$$\gamma(0) = Var(X_t) = (1-\phi^2)^{-1}\sigma^2, \rho(h) = \phi^h.$$

In such a case

$$\tau_n^2 \approx \gamma(0)[1 + 2\sum_{h=1}^{\infty}\rho(h)] = (1-\phi^2)^{-1}\sigma^2[1 + 2\sum_{h=1}^{\infty}\phi^h] = \sigma^2/(1-\phi)^2.$$

Technical* details are given below. So if we have estimates of ϕ and σ, then an estimate of τ_n^2 is given by

$$\hat{\tau}_n^2 = \hat{\sigma}^2/(1-\hat{\phi})^2.$$

[Technical details: Note that

$$\sum_{h=1}^{\infty} \phi^h = \phi + \phi^2 + \cdots = \phi(1 + \phi + \phi^2 + \cdots) = \phi(1 - \phi)^{-1}.$$

Hence using some algebra we can find that $\tau_n^2 \approx (1 - \phi^2)^{-1}\sigma^2[1 + 2\phi(1 - \phi)^{-1}] = \sigma^2/(1 - \phi)^2$.]

Remark: Typically, models are not known in advance. For this reason, it is often reasonable to estimate τ_n^2 using the formula given in (1) above.

Estimation of autocovariance and autocorrelation

We have already discussed this issue in Handout 6. Estimates of autocovariance $\gamma(h)$ and $\rho(h)$ are given by

$$\hat{\gamma}(h) = (1/n)\sum_{t=1}^{n-h}(X_t - \bar{X})(X_{t+h} - \bar{X}), \hat{\rho}(h) = \hat{\gamma}(h)/\hat{\gamma}(0).$$

It is known that, for large n, the distribution of $\hat{\rho}(h), h \geq 1$, is approximately normal with mean $\rho(h)$ and variance w_{hh}/n, where

$$w_{hh} = \sum_{k=1}^{\infty}[\rho(k + h) + \rho(k - h) - 2\rho(h)\rho(k)]^2.$$

So if we have an estimate $\hat{w}_{hh}$ of w_{hh}, then an approximate 95% confidence interval of $\rho(h)$ is given by

$$\hat{\rho}(h) \pm 1.96 s(\hat{\rho}(h)), \text{ where } s(\hat{\rho}(h))^2 = \hat{w}_{hh}/n.$$

We have already discussed in Handout 6 that in order to check if $\rho(h) = 0$, we look at the ACF plot to find out if $\hat{\rho}(h)$ is inside the interval $\pm 1.96/\sqrt{n}$. If $\hat{\rho}(h)$ is inside this interval, then it indicates that $\rho(h)$ may be close to zero. ACF plots are commonly used in the analysis of time series data.

In general obtaining an estimate of w_{hh} using the estimated autocorrelations requires a bit of care. However, for a few simple cases given below, it is possible to find nice expressions for w_{hh} and these espressions can be used to obtain estimate for w_{hh}.

Example 3. Suppose that the sequence $\{X_t\}$ is i.i.d. with mean μ and variance σ^2. In this case, $\rho(h) = 0$, $h = 1, 2, ...$ The quantity $w_{hh} = 1$ for any integer h. As a matter of fact, $\hat{\rho}(1), \hat{\rho}(2), ...$ are all independent and each is approximately normally distributed with mean zero and variance $1/n$. This result is behind Portmanteu and Box-Ljeung tests.

Example 4. Suppose that the sequence $\{X_t\}$ has mean μ and follows an $MA(1)$ sequence with the model $X_t - \mu = \varepsilon_t + \theta\varepsilon_{t-1}$, where $\{\varepsilon_t\}$ is white noise with variance σ^2. Then $\rho(1) = \theta/(1 + \theta^2)$ and $\rho(2)$, $\rho(3), \rho(4), ...$ are all zero. In this case, simple mathematical calculations (by using (1)) will show

$$w_{hh} = \begin{cases} 1 - 3\rho(1)^2 + 4\rho(1)^4 & h = 1 \\ 1 + 2\rho(1)^2 & h = 2, 3, ... \end{cases} \tag{1}$$

So in this case, $\hat{\rho}(1)$ is approximately normally distributed with mean $\rho(1)$ and variance w_{11}/n. However, for any $h = 2, 3, ...,$ $\hat{\rho}(h)$ is approximately normally distributed with **mean zero** and variance $(1 + 2\rho(1)^2)/n$.

If we fit an MA(1) model for the Lake Michigan data using R, we find that

$$\begin{aligned}
\hat{\gamma}(0) &= 9.9554,\ \hat{\gamma}(1) = -0.7085,\ \hat{\rho}(1) = -0.0712,\\
\hat{w}_{11} &= 1 - 3\hat{\rho}(1)^2 + 4\hat{\rho}(1)^4 = 0.9849,\\
\hat{w}_{hh} &= 1 + 2\hat{\rho}(1)^2 = 1.0101,\ h = 2, 3, ...,\\
s^2(\hat{\rho}(1)) &= \hat{w}_{11}/n = 0.9849/87 = 0.0113,\ s(\hat{\rho}(1)) = 0.1064,\\
s^2(\hat{\rho}(h)) &= \hat{w}_{hh}/n = 1.0101/87 = 0.0116,\ s(\hat{\rho}(h)) = 0.1078, h = 2, 3, ...
\end{aligned}$$

Example 5. If the sequence $\{X_t\}$ with mean μ follows an $AR(1)$ model, i.e., $X_t - \mu = \phi(X_{t-1} - \mu) + \varepsilon_t$, where $\{\varepsilon_t\}$ is white noise with variance σ^2, then $\gamma(h) = \gamma(0)\phi^h$ and $\rho(h) = \phi^h$, $h = 0, 1, ...$ In this case, $\hat{\rho}(h)$ is approximately normally distributed with mean $\rho(h) = \phi^h$ and variance

$$w_{hh} = (1 - \phi^{2h})(1 + \phi^2)(1 - \phi^2)^{-1} - 2h\phi^{2h}.$$

The expression for w_{hh} is obtained from (1) using the expressions for $\{\rho(j)\}$ in the AR(1) case. If we fit an $AR(1)$ model for the detrended temperature data $\{\hat{X}_t = Y_t - \hat{m}_t\}$ (where $\hat{m}_t$ has been estimated by loess in Handout 6), then the estimated coefficients are (using R)

$$\hat{\mu} = 0.0004,\ \hat{\phi} = 0.2145,\ s(\hat{\phi}) = 0.0766,\ \hat{\sigma}^2 = 0.01085.$$

You will find below the estimated standard errors for autocorrelation estimates for $h = 1, 2, 3:$

$$\begin{aligned}
\hat{w}_{11} &= (1 - \hat{\phi}^2)(1 + \hat{\phi}^2)(1 - \hat{\phi}^2)^{-1} - 2\hat{\phi}^2 = 0.9540,\ s(\hat{\rho}(1)) = \sqrt{\hat{w}_{11}/n} = 0.1047,\\
\hat{w}_{22} &= (1 - \hat{\phi}^4)(1 + \hat{\phi}^2)(1 - \hat{\phi}^2)^{-1} - 4\phi^4 = 1.0857,\ s(\hat{\rho}(2)) = \sqrt{\hat{w}_{22}/n} = 0.1117,\\
\hat{w}_{33} &= (1 - \hat{\phi}^6)(1 + \hat{\phi}^2)(1 - \hat{\phi}^2)^{-1} - 6\phi^6 = 1.0958,\ s(\hat{\rho}(3)) = \sqrt{\hat{w}_{33}/n} = 0.1122.
\end{aligned}$$

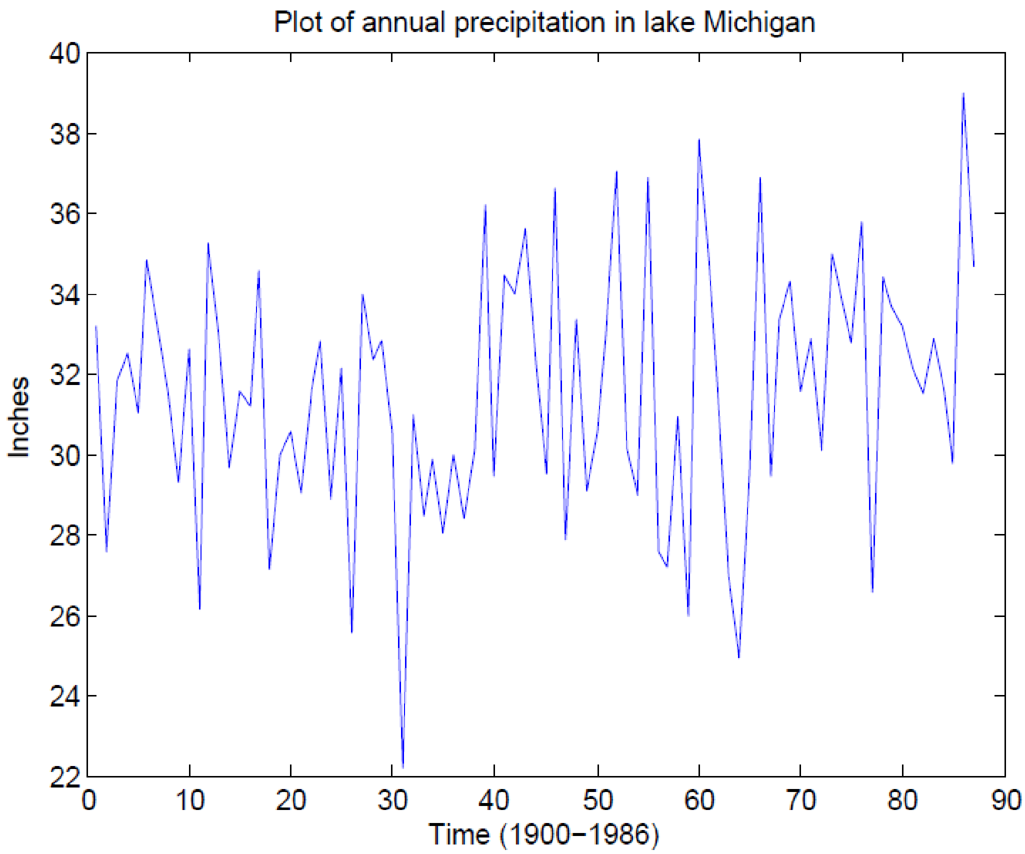
Remark: In Examples 4 and 5, estimated standrd errors $s(\hat{\rho}(h))$ are obtained for some simple models. Before one applies the formulas for w_{hh} given in these examples, it is important to be sure that these models are reasonable for the series $\{X_t\}$.

Appendix: Some R commands

If the given series is stored in x, then commands for fitting AR and MA models and getting the parameter estimates are given below.

arima(x, order=c(0,0,1)) [fits an MA(1) model]

arima(x, order=c(0,0,2)) [fits an MA(2) model],



arima(x, order=c(1,0,0)) [fits an AR(1) model],

arima(x, order=c(2,0,0)) [fits an AR(2) model].

Suppose you want to get the autocorrelation function upto lag 15 for an MA(2) sequence with $\theta_1 = 0.3$ and $\theta_2 = -0.5$, then the R command is

ARMAacf(ma=c(0.3,-0.5), lag.max=15).