# Stat 804

## Lecture 14 Notes

Our goal in this lecture is to develop asymptotic distribution theory for the sample autocorrelation function. We let $\rho_k$ and $\hat\rho_k$ be the ACF and estimated ACF respectively.

We begin by reducing the behaviour of $\hat\rho_k$ to the behaviour of $\hat C$, the sample autocovariance. Our approach is standard Talyor expansion.

### Large sample theory for ratio estimates

Suppose you have pairs $(X_n, Y_n)$ of random variables with

$$n^{1/2}(X_n - \mu) \Rightarrow N(0, \sigma^2)$$

and

$$n^{1/2}(Y_n - \nu) \Rightarrow N(0, \tau^2)$$

We study the large sample behaviour of $X_n/Y_n$ under the assumption that $\nu$ is not 0. We will see that the case $\mu = 0$ results in some simplifications. Begin by writing

$$\frac{X_n}{Y_n} = \frac{X_n}{\nu + (Y_n - \nu)} = \frac{X_n}{\nu}\frac{1}{1+\epsilon_n}$$

where

$$\epsilon_n = \frac{Y_n - \nu}{\nu}$$

Notice that $\epsilon_n \to 0$ in probability. We may expand

$$\frac{1}{1+\epsilon_n} = \sum_{k=0}^{\infty}(-\epsilon_n)^k$$

and then write

$$\frac{X_n}{Y_n} = \frac{X_n}{\nu}\sum_{k=0}^{\infty}(-\epsilon_n)^k$$

We want to compute the mean of this expression term by term and the variance by using the formula for the variance of the sum and so on. However, what we really do is truncate the infinite sum at some finite number of terms and compute moments of the finite sum. I want to be clear about the distinction; to do so I give an example. Imagine that $(X_n, Y_n)$ has a bivariate normal distribution with means $\mu, \nu$, variances $\sigma^2/n$, $\tau^2/n$ and correlation $\rho$ between $X_n$ and $Y_n$. The quantity $X_n/Y_n$ does not have a well defined mean because $\mathrm{E}(|X_n/Y_n|) = \infty$. Our expansion is still valid, however. Stopping the sum at $k = 1$ leads to the approximation

$$\begin{aligned}\frac{X_n}{Y_n} &\approx \frac{X_n}{\nu} + \frac{X_n(Y_n - \nu)}{\nu^2} \\ &= \frac{X_n}{\nu} + \frac{\mu(Y_n - \nu)}{\nu^2} + \frac{(X - n - \mu)(Y_n - \nu)}{\nu^2}\end{aligned}$$

I now want to look at these terms to decide which are big and which are small. To do so I introduce big O notation:
**Definition**: : If $U_n$ is a sequence of random variables and $a_n > 0$ a sequence of constants then we write

$$U_n = O_P(a_n)$$

if, for each $\epsilon > 0$ there is an $M$ (depending on $\epsilon$ but not $n$) such that

$$P(|U_n| > M|a_n|) < \epsilon$$

The idea is that $U_n = O_P(a_n)$ means that $U_n$ is proportional in size to $a_n$ with the "constant of proportionality" being a random variable which is not likely to be too large. We also often have use for notation indicating that $U_n$ is actually small compared to $a_n$.
**Definition**: : We say $U_n = o_P(a_n)$ if $U_n/a_n \to 0$ in probability: for each $\epsilon > 0$

$$\lim_{n\to\infty} P(|U_n/a_n| > \epsilon) = 0$$

You can manipulate $O_P$ and $o_P$ notation algebraically with a few rules:

1. If $b_n$ is a sequence of constants such that $b_n = ca_n$ with $c > 0$ then

$$U_n = O_P(b_n) \Leftrightarrow U_n = O_P(a_n)$$

   We write

$$cO_P(a_n) = O_P(a_n)$$

2. If $U_n = O_P(a_n)$ and $V_n = O_P(b_n)$ for two sequences $a_n$ and $b_n$ then

$$U_nV_n = O_P(a_nb_n)$$

   We express this as

$$O_P(a_n)O_P(b_n) = O_P(a_nb_n)$$

3. In particular

$$b_nO_P(a_n) = O_P(b_na_n)$$

4. 

$$O_P(a_n) + O_P(b_n) = O_P(\max(a_n, b_n))$$

5. $co_P(a_n) = o_P(a_n)$
6. $o_P(a_n)O_P(b_n) = o_P(a_n)o_P(b_n) = o_P(a_nb_n)$
7. In particular $b_no_P(a_n) = o_P(b_na_n)$
8. $o_P(a_n) + o_P(b_n) = o_P(\max(a_n, b_n))$

These notions extend Landau's $o$ and $O$ notation to random quantities.
**Example**: : In our ratio example we have

$$X_n = \mu + O_P(n^{-1/2})$$

and

$$Y_n = \nu + O_P(n^{-1/2})$$

In our geometric expansion

$$\epsilon_n^k = O_P(n^{-k/2})$$

Look first at the expansion stopped at $k = 1$. We have

$$\begin{aligned}\frac{X_n}{Y_n} - \frac{\mu}{\nu} &\approx \frac{X_n - \mu}{\nu} - \frac{\mu(Y_n - \nu)}{\nu^2} - \frac{(X - n - \mu)(Y_n - \nu)}{\nu^2} \\ &= O_P(n^{-1/2}) + O_P(n^{-1/2}) + O_P(n^{-1})\end{aligned}$$

(The three terms on the RHS of the first line are being described in terms of roughly how big each is.) If we stop at $k = 2$ we get

$$\begin{aligned}\frac{X_n}{Y_n} - \frac{\mu}{\nu} &\approx \frac{X_n - \mu}{\nu} - \frac{\mu(Y_n - \nu)}{\nu^2} - \frac{(X - n - \mu)(Y_n - \nu)}{\nu^2} \\ &\quad + \frac{\mu(Y_n - \nu)^2}{\nu^3} + \frac{(X - n - \mu)(Y_n - \nu)^2}{\nu^3} \\ &= O_P(n^{-1/2}) + O_P(n^{-1/2}) + O_P(n^{-1}) + O_P(n^{-1}) + O_P(n^{-3/2})\end{aligned}$$

Keeping only terms of order $O_P(n^{-1/2})$ we find

$$\frac{X_n}{Y_n} - \frac{\mu}{\nu} = \frac{X_n - \mu}{\nu} + \frac{\mu(Y_n - \nu)}{\nu^2} + O_P(n^{-1})$$

We now take expected values and discover that up to an error of order $n^{-1}$

$$\mathrm{E}(X_n/Y_n) = \mu/\nu$$

BUT you are warned that what is really meant is simply that there is a random variable which is approximately (neglecting something which is probably proportional in size to $n^{-1}$)

$$\frac{X_n}{Y_n} - \frac{\mu}{\nu}$$

whose expected value is 0. For the normal example the remainder term in this expansions, that is, the term $O_P(n^{-1})$), is probably small but its expected value is not defined.

To keep terms up to order $O_P(n^{-1})$ we have to keep terms out to $k = 2$ (In general

$$X_n \epsilon_n^k = (\mu + O_P(n^{-1/2}) O_P(n^{-k/2})$$

For $k > 2$ this is $o_P(n^{-1})$ but for $k = 2$ the $\mu O_P(n^{-1})$ term is not negligible. If we retain terms out to $k = 2$ then we get

$$\frac{X_n}{Y_n} - \frac{\mu}{\nu} = \frac{X_n - \mu}{\nu} + \frac{\mu(Y_n - \nu)}{\nu^2} + \frac{(X - n - \mu)(Y_n - \nu)}{\nu^2} + \frac{\mu(Y_n - \nu)^2}{\nu^3} + O_P(n^{-3/2})$$

Taking expected values here we get

$$\mathrm{E}\left[\frac{X_n}{Y_n}-\frac{\mu}{\nu}\right] \approx \mathrm{E}\left[(X-n-\mu)(Y_n-\nu)\right]/\nu^2$$

up to terms of order $n^{-1}$. In the normal case we get

$$\mathrm{E}\left[\frac{X_n}{Y_n}-\frac{\mu}{\nu}\right] \approx \rho\sigma\tau/n$$

In order to compute the approximate variance we ought to compute the second moment of $X_n/Y_n - \mu/\nu$ and subtract the square of the first moment. Imagine you had a random variable of the form

$$\sum_{k=1} \frac{W_k}{n^{k/2}}$$

where I assume that the $W_k$ do not depend on $n$. The mean, taken term by term would be of the form

$$\sum_{k=1} \frac{\eta_k}{n^{k/2}}$$

and the second moment of the form

$$\sum_{j=1}\sum_{k=1} \frac{\mathrm{E}(W_j W_k)}{n^{(j+k)/2}}$$

This leads to a variance of the form

$$\frac{\mathrm{Var}(W_1)}{n} + \frac{\mathrm{Cov}(W_1, W_2)}{n^{3/2}} + O(n^{-2})$$

Our expansion above gave

$$W_1 = \frac{n^{1/2}(X_n-\mu)}{\nu} - \frac{\mu n^{1/2}(Y_n-\nu)}{\nu^2}$$

and

$$W_2 = -\frac{n^{1/2}(X_n-\mu)n^{1/2}(Y_n-\nu)}{\nu^2} + \frac{\mu[n^{1/2}(Y_n-\nu)]^2}{\nu^3}$$

from which we get the approximate variance

$$\left(\frac{\sigma^2}{\nu^2} + \frac{\mu^2\tau^2}{\nu^4} + \frac{\rho\sigma\tau\mu}{\nu^3}\right)/n + O(n^{-3/2})$$

Now I want to apply these ideas to estimation of $\rho_k$. We make $X_n$ be $\hat{C}(k)$ and $Y_n$ be $\hat{C}(0)$ (and replace $n$ by $T$). Our first order approximation to $\hat{\rho}-\rho$ is

$$A_1 = T^{1/2}(\hat{C}(k) - C(k))/C(0) - T^{1/2}C(k)(\hat{C}(0) - C(0))/C^2(0)$$

Our second order approximation would be

$$A_1 + [T^{1/2}(\hat{C}(k) - C(k))]^2 C(k)/C^3(0) - T^{1/2}(\hat{C}(k) - C(k))T^{1/2}(\hat{C}(0) - C(0))/C^2(0)$$

I now evaluate means and variances in the special case where $\hat{C}$ has been calculated using a known mean of 0. That is

$$\hat{C}(k) = \frac{1}{T}\sum_0^{T-1-k} X_t X_{t+k}$$

Then

$$\mathrm{E}(\hat{C}(k) - C(k)) = -kC(k)/T$$

so

$$\mathrm{E}(A_1) = -k\rho_k/T^{1/2}$$

To compute the variance we begin with the second moment which is

$$\frac{1}{T^2}\sum_s\sum_t \mathrm{E}(X_s X_{s+k} X_t X_{t+k})$$

The expectations in question involve the fourth order product moments of $X$ and depend on the distribution of the $X$'s and not just on $C_X$. However, for the interesting case of white noise, we can compute the expected value. For $k > 0$ you may assume that $s < t$ or $s = t$ since the $s > t$ cases can be figured out by swapping $s$ and $t$ in the $s < t$ case. For $s < t$ the variable $X_s$ is independent of all 3 of $X_{s+k}$, $X_t$ and $X_{t+k}$. Thus the expectation factors into something containing the factor $\mathrm{E}(X_s) = 0$. For $s = t$, we get $\mathrm{E}(X_s^2)\mathrm{E}(X_{s+k}^2) = \sigma^4$. and so the second moment is

$$\frac{T-k}{T^2}\sigma^4$$

This is also the variance since, for $k > 0$ and for white noise, $C_X(k) = 0$.

For $k = 0$ and $s < t$ or $s > t$ the expectation is simply $\sigma^4$ while for $s = t$ we get $\mathrm{E}(X_t^4) \equiv \mu_4$. Thus the variance of the sample variance (when the mean is known to be 0) is

$$\frac{T-1}{T}\sigma^4 + \mu_4/T - \sigma^4 = (\mu_4 - \sigma^4)/T\,.$$

For the normal distribution the fourth moment $\mu_4$ is given simply by $3\sigma^4$.

Having computed the variance it is usual to look at the large sample distribution theory. For $k = 0$ the usual central limit theorem applies to $\sum X_t^2/T$ (in the case of white noise) to prove that

$$\sqrt{T}(\hat{C}_X(0) - \sigma^2)/\sqrt{\mu_4} \to N(0,1)\,.$$

The presence of $\mu_4$ in the formula shows that the approximation is quite sensitive to the assumption of normality.

For $k > 0$ the theorem needed is called the $m$-dependent central limit theorem; it shows that

$$\sqrt{T}\hat{C}_X(k)/\sigma^2 \to N(0,1)\,.$$

In each of these cases the assertion is simply that the statistic in question divided by its standard deviation has an approximate normal distribution.

The sample autocorrelation at lag $k$ is

$$\hat{C}_X(k)/\hat{C}_X(0).$$

For $k > 0$ we can apply Slutsky's theorem to conclude that

$$\sqrt{T}\hat{C}_X(k)/\hat{C}_X(0) \to N(0,1).$$

This justifies drawing lines at $\pm 2/\sqrt{T}$ to carry out a 95% test of the hypothesis that the $X$ series is white noise based on the $k$th sample autocorrelation.

It is possible to verify that subtraction of $\bar{X}$ from the observations before computing the sample covariances does not change the large sample approximations, although it does affect the exact formulas for moments.

When the $X$ series is actually not white noise the situation is more complicated. Consider as an example the model

$$X_t = \phi X_{t-1} + \epsilon_t$$

with $\epsilon$ being white noise. Taking

$$\hat{C}_X(k) = \frac{1}{T} X_t X_{t+k}$$

we find that

$$T^2 \mathrm{E}(\hat{C}_X(k)^2) = \sum_s \sum_t \sum_{u_1} \sum_{u_2} \sum_{v_1} \sum_{v_2} \phi^{u_1+u_2+v_1+v_2} \mathrm{E}(\epsilon_{s-u_1}\epsilon_{s+k-u_2}\epsilon_{t-v_1}\epsilon_{t+k-v_2})$$

The expectation is 0 unless either all 4 indices on the $\epsilon$'s are the same or the indices come in two pairs of equal values. The first case requires $u_1 = u_2 - k$ and $v_1 = v_2 - k$ and then $s - u_1 = t - v_1$. The second case requires one of three pairs of equalities: $s - u_1 = t - v_1$ and $s - u_2 = t - v_2$ or $s - u_1 = t + k - v_2$ and $s + k - u_2 = t - v_1$ or $s - u_1 = s + k - u_2$ and $t - v_1 = t - +k - v_2$ along with the restriction that the four indices not all be equal. The actual moment is then $\mu_4$ when all four indices are equal and $\sigma^4$ when there are two pairs. It is now possible to do the sum using geometric series identities and compute the variance of $\hat{C}_X(k)$. It is not particularly enlightening to finish the calculation in detail.

There are versions of the central limit theorem called mixing central limit theorems which can be used for ARMA$(p,q)$ processes in order to conclude that

$$\sqrt{T}(\hat{C}_X(k) - C_X(k))/\sqrt{\mathrm{Var}(\hat{C}_X(k))}$$

has asymptotically a standard normal distribution and that the same is true when the standard deviation in the denominator is replaced by an estimate. To get from this to distribution theory for the sample autocorrelation is easiest when the true autocorrelation is 0.

The general tactic is the $\delta$ method or Taylor expansion. In this case for each sample size $T$ you have two estimates, say $N_T$ and $D_T$ of two parameters. You want distribution theory for the ratio $R_T = N_T/D_T$. The idea is to write $R_T = f(N_T, D_T)$ where $f(x,y) = x/y$ and then make use of the fact that $N_T$ and $D_T$ are close to the parameters they are estimates of. In our case $N_T$ is the sample autocovariance at lag $k$ which is close to the true autocovariance $C_X(k)$ while the denominator $D_T$ is the sample autocovariance at lag 0, a consistent estimator of $C_X(0)$.

Write

$$\begin{aligned} f(N_T, D_T) &= f(C_X(k), C_X(0)) \\ &\quad +(N_T - C_X(k))D_1 f(C_X(k), C_X(0)) \\ &\quad +(D_T - C_X(0))D_2 f(C_X(k), C_X(0)) + \text{remainder} \end{aligned}$$

If we can use a central limit theorem to conclude that

$$(\sqrt{T}(N_T - C_X(k)), \sqrt{T}(D_T - C_X(0)))$$

has an approximately bivariate normal distribution and if we can neglect the remainder term then

$$\sqrt{T}(f(N_T, D_T) - f(C_X(k), C_X(0))) = \sqrt{T}(\hat{\rho}(k) - \rho(k))$$

has approximately a normal distribution. The notation here is that $D_j$ denotes differentiation with respect to the $j$th argument of $f$. For $f(x,y) = x/y$ we have $D_1 f = 1/y$ and $D_2 f = -x/y^2$. When $C_X(k) = 0$ the term involving $D_2 f$ vanishes and we simply get the assertion that

$$\sqrt{T}(\hat{\rho}(k) - \rho(k))$$

has the same asymptotic normal distribution as $\hat{C}_X(k)/C_X(0)$.

Similar ideas can be used for the estimated sample partial ACF.

## Portmanteau tests

In order to test the hypothesis that a series is white noise using the distribution theory just given, you have to produce a single statistic to base youre test on. Rather than pick a single value of $k$ the suggestion has been made to consider a sum of squares or a weighted sum of squares of the $\hat{\rho}(k)$.

A typical statistic is

$$T \sum_{k=1}^{K} \hat{\rho}^2(k)$$

which, for white noise, has approximately a $\chi^2_K$ distribution. (This fact relies on an extension of the previous computations to conclude that

$$\sqrt{T}(\hat{\rho}(1), \ldots, \hat{\rho}(K))$$

has approximately a standard multivariate distribution. This, in turn, relies on computation of the covariance between $\hat{C}(j)$ and $\hat{C}(k)$.)

When the parameters in an ARMA$(p,q)$ have been estimated by maximum likelihood the degrees of freedom must be adjusted to $K-p-q$. The resulting test is the Box-Pierce test; a refined version which takes better account of finite sample properties is the Box-Pierce-Ljung test. S-Plus plots the $P$-values from these tests for 1 through 10 degrees of freedom as part of the output of *arima.diag*.