Asymptotic Distributions in Time Series

Overview

Standard proofs that establish the asymptotic normality of estimators constructed from random samples (*i.e.*, independent observations) no longer apply in time series analysis. The usual version of the central limit theorem (CLT) presumes independence of the summed components, and that's not the case with time series. This lecture shows that normality still rules for asymptotic distributions, but the arguments have to be modified to allow for correlated data.

1. Types of convergence
2. Distributions in regression (Th A.2, section B.1)
3. Central limit theorems (Th A.3, Th A.4)
4. Normality of covariances, correlations (Th A.6, Th A.7)
5. Normality of parameter estimates in ARMA models (Th B.4)

Types of convergence

Convergence The relevant types of convergence most often seen in statistics and probability are

- Almost surely, almost everywhere, with probability one, w.p. 1:

$$X_n \stackrel{\text{a.s.}}{\to} X: \quad \mathbb{P}\{\omega : \lim X_n = X\} = 1.$$

- In probability, in measure:

$$X_n \stackrel{\text{p}}{\to} X: \quad \lim_n \mathbb{P}\{\omega : |X_n - X| > \epsilon\} = 0 \; .$$

- In distribution, weak convergence, convergence in law:

$$X_x \stackrel{\text{d}}{\to} X: \quad \lim_n \mathbb{P}(X_n \le x) = \mathbb{P}(X \le x) \; .$$

- In mean square or ℓ_2, the variance goes to zero:

$$X_n \stackrel{\text{m.s.}}{\to} X : \quad \lim_n \mathbb{E}\,(X_n - X)^2 = 0\ .$$

Connections Convergence almost surely (which is much like good old fashioned convergence of a sequence) implies covergence almost surely which implies covergence in distribution:

$$\stackrel{\text{a.s.}}{\to} \quad \Rightarrow \quad \stackrel{\text{p}}{\to} \quad \Rightarrow \quad \stackrel{\text{d}}{\to}$$

Convergence in distribution only implies convergence in probability if the distribution is a point mass (*i.e.*, the r.v. converges to a constant). The various types of converence "commute" with sums, products, and smooth functions.

Mean square convergence is a bit different from the others; it implies convergence in probabiity,

$$\stackrel{\text{m.s.}}{\to} \quad \Rightarrow \quad \stackrel{\text{p}}{\to}$$

which holds by Cheybyshev's inequality.

Asymptotically normal (Defn A.5)

$$\frac{X_n - \mu_n}{\sigma_n} \stackrel{\text{d}}{\to} N(0,1) \qquad \Leftrightarrow \qquad X_n \sim AN(\mu_n, \sigma_n^2)$$

Distributions in regression

Not i.i.d. The asymptotic normality of the slope estimates in regression is not so obvious if the errors are not normal. Normality requires that we can handle sums of independent, but not identically distributed r.v.s.

Scalar This example and those that follow only do scalar estimators to avoid matrix manipulations that conceal the underlying ideas. The text illustrates the use of the "Cramer-Wold device" for handling vector-valued estimators.

Model In scalar form, we observe a sample of independent observations that follow $Y_i = \alpha + \beta X_i + \epsilon_i$. Assume Y and ϵ denote random variables, the x_i are fixed, and the deviations ϵ_i have mean 0 and variance σ_ϵ^2.

Least squares estimators are $\hat\alpha = \overline{Y} - \hat\beta\overline{X}$ and

$$\hat\beta = \frac{\sum_i (x_i - \overline{x})(y_i - \overline{y})}{\sum_i (x_i - \overline{x})^2} = \frac{\text{cov}_n(x,y)}{\text{var}_n(x)}$$

For future reference, write $SS_{xn} = \sum_{i=1}^n (x_i - \overline{x})^2$.

Distribution of $\hat\beta$ We can avoid centering y in the numerator and write

$$\begin{aligned}
\hat\beta &= \frac{\sum_i (x_i - \overline{x}) y_i}{SS_{xn}} \\
&= \frac{\sum_i (x_i - \overline{x})(x_i\beta + \epsilon_i)}{SS_{xn}} \\
&= \beta + \sum w_{ni}\epsilon_i \qquad (1)
\end{aligned}$$

where the key weights depend on n and have the form

$$w_{ni} = \frac{x_i - \overline{x}}{SS_{xn}} \ . \qquad (2)$$

Hence, we have written $\hat\beta - \beta$ as a weighted sum of random variables, and it follows that

$$\mathbb{E}\,\hat\beta = \beta, \quad \text{Var}(\hat\beta) = \frac{\sigma_\epsilon^2}{\sum (x_i - \overline{x})^2} \ .$$

To get the limiting distribution, we need a version of the CLT that allows for unequal variances (Lindeberg CLT) and weights that change with n.

Bounded leverage Asymptotic normality of $\hat\beta$ requires, in essence, that the all of the observations have roughly equal impact on the response. In order that one point not dominate $\hat\beta$ we have to require that

$$\max \frac{w_{ni}}{SS_{xn}} \to 0 \ .$$

Lindeberg CLT This theorem allows a triangular array of random variables. Think of sequences as in a lower triangular array, $\{X_{11}\}$, $\{X_{21}, X_{22}\}$, $\{X_{31}, X_{32}, X_{33}\}$ with mean zero and variances $\text{Var}(X_{ni}) = \sigma_{ni}^2$. Now let $T_n = \sum_{i=1}^n X_{ni}$ with variance $\text{Var}(T_n) = s_n^2 = \sum_{i=1}^n \sigma_{ni}^2$. Hence,

$$\mathbb{E}\,\frac{T_n}{s_n} = 0, \qquad \text{Var}\,\frac{T_n}{s_n} = 1 \ .$$

If the so-called *Lindeberg condition* holds, (I(set) is the indicator function of the set)

$$\forall \delta > 0, \quad \frac{1}{s_n^2} \sum_{i=1}^{n} \mathbb{E}\left(X_{ni}^2 \, I(\delta s_n < |X_{ni}|\right) \to 0 \text{ as } n \to \infty \,. \tag{3}$$

then

$$\frac{T_n}{s_n} \xrightarrow{\text{d}} N(0,1), \qquad T_n \sim AN(0, s_n^2) \,.$$

Regression application In this case, define the random variables $X_{ni} = w_{ni}\epsilon_i$ so that $\hat\beta = \beta + \sum X_{ni}$, where the model error ϵ_i has mean 0 with variance σ^2 and (as in 2)

$$w_{ni} = \frac{(x_i - \overline{x}_n)}{\sum (x_i - \overline{x}_n)^2} \,, \quad s_n^2 = \sigma^2 \sum_i w_{ni}^2 \,. \tag{4}$$

Asymptotic normality requires that no one observation have too much effect on the slope, which we specify by the condition

$$\frac{\max_i w_{ni}^2}{\sum_{i=1}^n w_{ni}^2} \to 0 \text{ as } n \to \infty \tag{5}$$

To see that the condition (5) implies the Lindeberg condition, let's keep things a bit simpler by assuming that the ϵ_i share a common distribution, differing only in scale.

Now define the largest weight $W_n = \max |w_{ni}|$ and observe that

$$\frac{1}{s_n^2} \sum_i w_{ni}^2 \mathbb{E}\left(\epsilon_i^2 I(\delta s_n < |w_{ni}\epsilon_i|)\right) \le \frac{1}{s_n^2} \sum_i w_{ni}^2 \mathbb{E}\left(\epsilon_i^2 I(\tfrac{\delta s_n}{W_n} < |\epsilon_i|)\right) \tag{6}$$

The condition (5) implies that $s_n/W_n \to \infty$. Since $\operatorname{Var} \epsilon_i = \sigma^2$ is finite, choose N so that the summands have a common bound (here's where we use the common distribution of the ϵ_i)

$$\mathbb{E}\left[\epsilon_1^2 I(\frac{\delta s_n}{W_n} < |\epsilon_1|)\right] < \eta \quad \forall i.$$

with $\eta \to 0$ as $n \to \infty$. Then the sum (6) is bounded by $\eta/\sigma^2 \to 0$.

Central limit theorems for dependent processes

M-dependent process A sequence of random variables $\{X_t\}$ (stochastic process) is said to be *M-dependent* if $|t-s| > M$ implies that X_t is independent of X_s,

$$\{X_t, t \le T\} \qquad \text{ind} \qquad \{X_s : s > T+M\} \tag{7}$$

Variables with time indices t and s that are more than M apart, $|t-s| > M$, are independent.

Basic theorem (A.2) This is the favorite method of proof in S&S and other common textbooks. An up-over-down argument. The idea is to construct an approximating r.v. Y_{mn} that is similar to X_n (enforced by condition 3 below), but simpler to analyze in the sense that its asymptotic distribution which is controlled by the "tuning parameter" m is relatively easy to obtain.

Theorem A.2 If

$$\begin{array}{lll} (1) & \forall m \quad Y_{mn} \xrightarrow{\text{d}} Y_m & \text{as } n \to \infty, \\ (2) & Y_m \xrightarrow{\text{d}} Y & \text{as } m \to \infty, \\ (3) & \mathbb{E}\,(X_n - Y_{mn})^2 \to 0 & \text{as } m, n \to \infty, \end{array}$$

then $X_n \xrightarrow{\text{d}} Y$.

CLT for M-dependence (A.4) Suppose $\{X_t\}$ is M-dependent with covariances γ_j. The variance of the mean of n observations is then

$$\text{Var}\left(\sqrt{n}\,\overline{X}_n\right) = n\,\text{Var}\overline{X}_n = \sum_{h=-M}^{M} \frac{n-|h|}{n}\gamma_h \to \sum_{-M}^{M} \gamma_h := V_M \tag{8}$$

Theorem A.4 If $\{X_t\}$ is an M-dependent stationary process with mean μ and $V_M > 0$ then

$$\sqrt{n}(\overline{X}_n - \mu) \xrightarrow{\text{d}} N(0, V_M) \qquad (\overline{X}_n \sim AN(\mu, V_M/n))$$

Proof Assume w.l.o.g. that $\mu = 0$ (recall our prior discussion in Lecture 4 that we can replace the sample mean by μ) so that the statistic of interest is $\sqrt{n}\,\overline{X}_n$. Define the similar statistic Y_{mn} as follows by forming the data into blocks of length m and tossing enough so that the blocks are independent of each other. Choose $m > 2M$ and let $r = \lfloor n/m \rfloor$ (greatest integer less than) count the blocks.

Now arrange the data to fill a table with r rows and m columns (pretend for the moment that $n = r\,m$. The approximation Y_{mn} sums the first $m - M$ columns of this table,

$$Y_{mn} = \frac{1}{\sqrt{n}} \left(\begin{array}{ccccccc} & X_1 & + & X_2 & + \cdots + & X_{m-M} \\ + & X_{m+1} & + & X_{m+2} & + \cdots + & X_{2m-M} \\ + & X_{2m+1} & + & X_{2m+2} & + \cdots + & X_{3m-M} \\ & \cdots & & & & \\ + & X_{(r-1)m+1} & + & X_{(r-1)m+2} & + \cdots + & X_{(r-1)m-M} \end{array} \right) \tag{9}$$

and omits the intervening blocks,

$$U_{mn} = \frac{1}{\sqrt{n}} \left(\begin{array}{cccc} & X_{m-M+1} & + \cdots + & X_m \\ + & X_{2m-M+1} & + \cdots + & X_{2m} \\ + & X_{3m-M+1} & + \cdots + & X_{3m} \\ & \cdots & & \\ + & X_{(r-1)m-M+1} & + \cdots + & X_n \end{array} \right) \tag{10}$$

The idea is that we will get the *shape* of the distribution from the variation among the blocks.

Label the row sums that define Y_{mn} in (9) as Z_i so that we can write

$$Y_{mn} = \frac{1}{\sqrt{n}} \left(Z_1 + Z_2 + \ldots + Z_r \right).$$

Clearly, because the X_t are centered, $\mathbb{E}\, Z_i = 0$ and

$$\operatorname{Var} Z_i = \sum_h (m - M - |h|)\gamma_h := S_{m-M}$$

Now let's fill in the 3 requirements of Theorem A.2:

(1)

$$Y_{mn} = \frac{1}{\sqrt{n}} \sum_{i=1}^{r} Z_i = (\underbrace{n/r}_{\to m})^{-1/2}\, \underbrace{r^{-1/2} \sum Z_i}_{\xrightarrow{\mathrm{d}} N}$$

Now observe that $n/r \to m$, and since the Z_i are independent by construction, the usual CLT shows that

$$r^{-1/2} \sum Z_i \xrightarrow{\text{d}} N(0, S_{m-M})$$

Hence, we have $Y_{mn} \xrightarrow{\text{d}} Y_m := N(0, S_{m-M}/m)$ as $n \to \infty$.

(2) Now let $m \to \infty$. As the number of columns in the table (9) increases, the variance expression S_{m-M} approaches the limiting variance V_M, so $Y_m \xrightarrow{\text{d}} N(0, V_M)$.

(3) Last, we have to check that we did not make too rough an approximation when forming Y_{mn}. The approximation omits the last M columns (10), so the error is

$$\sqrt{n}\,\overline{X}_n - Y_{mn} = \tfrac{1}{\sqrt{n}}(U_1 + \cdots + U_r)$$

where $U_i = X_{im-M+1} + \cdots + X_{im}$ and $\text{Var}(U_i) = S_M$. Thus,

$$\mathbb{E}\,(\sqrt{n}\,\overline{X}_n - Y_{mn})^2 = \tfrac{r}{n} S_M \to \tfrac{1}{m}\, S_M$$

as $n \to \infty$, and $(S_M)/m \to 0$ as $m \to \infty$.

CLT for stationary processes, Th A.5 In particular, for linear processes of the form $X_t - \mu = \sum \psi_j w_{t-j}$ where the white noise terms are independent with mean 0 and variance σ^2. The coefficients, as usual, are absolutely summable, $\sum |\psi_j| < \infty$.

Before stating the theorem, note that the variance of the mean of such a linear process has a particularly convenient form. Following our familiar manipulations of these sums (and Fatou's lemma!), we define V as the limiting variance of the mean as $n \to \infty$,

$$\begin{aligned} n\ \text{Var}(\overline{X}_n) &= \sum_{h=-n+1}^{n-1} \gamma_h \\ &\to \sum_{h=-\infty}^{\infty} \gamma_h \\ &= \sigma^2 \sum_{h=-\infty}^{\infty} \sum_{j=0}^{\infty} \psi_j \psi_{j+h} \qquad (\psi_j = 0, j < 0) \end{aligned}$$

$$\begin{aligned} &= \sigma^2 \left(\sum_j \psi_j \right)^2 \\ &:= V \end{aligned} \tag{11}$$

Theorem A.5 If $\{X_t\}$ is a linear process of this form with $\sum \psi_j \neq 0$ then

$$\overline{X}_n \sim AN(\mu, V/n)$$

Proof. Again use the basic three-part theorem A.2. This time, form the key approximation to $\sqrt{n}(\overline{X}_n - \mu)$ from the M-dependent process matched up to resemble X_t,

$$X_t^M = \sum_{j=0}^{M} \psi_j w_{t-j}$$

(1) Form the approximation based on averaging the M-dependent process rather than the original, infinite order stationary process X_t,

$$Y_{Mn} = \sqrt{n}(\overline{X}_{n,M} - \mu), \quad \overline{X}_{n,M} = \tfrac{1}{n} \sum_{t=1}^{n} X_t^M \ .$$

Since X_t^M is M-dependent, Theorem A.4 implies that as $n \to \infty$

$$Y_{Mn} \xrightarrow{\mathrm{d}} Y_M \sim N(0, V_M), \quad V_M = \sum_h \gamma_h = \sigma^2 (\sum_{j=0}^{M} \psi_j)^2 \ .$$

(2) As the order M increases, $V_M \to V$ as defined in (11).

(3) The error of the approximation is the tail sum of the moving average approximation, which is easily bounded since the weights ψ_j are absolutely summable

$$\mathbb{E} \left[\sqrt{n}(\overline{X}_n - \overline{X}_n^M) \right]^2 = \mathbb{E} \left(\frac{1}{n} \sum_{t=1}^{n} \sum_{j=M+1}^{\infty} \psi_j w_{t-j} \right)^2 = \frac{\sigma^2}{n} (\sum_{j=M+1}^{\infty} \psi_j)^2 \to 0$$

Normality of covariances, correlations

Covariances Th A.6 The text shows this for linear processes. We did the messy part previously.

Correlations Th A.7 This is an exercise in the use of the delta method. For scalar r.v.s, we have for differentiable functions $g()$ that

$$X \sim AN(\mu, \sigma_n^2) \qquad \Rightarrow \qquad g(X) \sim AN(g(\mu), g'(\mu)^2\, \sigma_n^2) \tag{12}$$

The argument is based on the observation that if $g(x) = a + bx$, then $g(X) \sim N(g(\mu), b^2\sigma^2)$. The trick is to show that the remainder in the linear Taylor series expansion of $g(X)$ around μ is small,

$$g(X) = g(\mu) + g'(\mu)(X - \mu) + \underbrace{\frac{g''(m)}{2}(X-\mu)^2}_{remainder}\ . \tag{13}$$

Normality of estimates in ARMA models

Conditional least squares in the AR(1) model (with mean zero) gives the estimator

$$\begin{aligned} \hat\phi &= \frac{\sum_{t=1}^{n-1} X_{t+1}X_t}{\sum_{t=1}^{n-1} X_t^2} \\ &= \frac{\sum(\phi X_t + w_{t+1})X_t}{\sum_{t=1}^{n-1} X_t^2} \\ &= \phi + \frac{\sum w_{t+1}X_t}{\sum X_t^2} \end{aligned} \tag{14}$$

Hence,

$$\sqrt{n}(\hat\phi - \phi) = \frac{\frac{1}{\sqrt{n}}\sum X_t W_{t+1}}{\frac{1}{n}\sum X_t^2}$$

Asymptotic distribution Results for the covariances of stationary processes (Theorem A.6) shows that the denominator converges to the variance γ_0. For the numerator, the expected value is zero (since the terms are independent) and variances of the components are

$$\mathrm{Var}(X_t w_{t+1}) = \mathbb{E}\, X_t^2\, \mathbb{E}\, w_t^2 = \gamma_0\, \sigma^2$$

These have equal variance, unlike the counterparts in the analysis of a regression model with fixed xs. The covariances are zero (owing again to the independence of the w_t),

$$\mathrm{Cov}(X_t w_{t+1}, X_s w_{s+1}) = 0, \quad s \neq t.$$

The summands in $\frac{1}{\sqrt{n}} \sum X_t W_{t+1}$ are thus uncorrelated (though not independent) and are a "trivial" stationary process. Though this is not a linear process as in A.6, comparable results hold owing to the lack of lingering dependence. (One has to build an approximating M-dependent version of the products, say $w_{t+1} X_t^M$; the approximation works with A.2 again since X_t is a linear process – see S&S.) This version of the prior theorem gives (Theorem B.4)

$$\sqrt{n}(\hat{\phi} - \phi) \xrightarrow{\mathrm{d}} N(0, 1 - \phi^2)$$

where the variance expression comes from

$$\frac{\sigma^2 \, \gamma_0}{\gamma_0^2} = \frac{\sigma^2}{\gamma_0} = 1 - \phi^2.$$

(In the general case, the asymptotic variance is $\sigma^2 \Gamma_p^{-1}$, but this is a nice way to see that the expression is invariant of σ^2.)