Properties of Descriptive Estimators

Overview

1. Properties of $\overline{X}$
2. Simulation of estimator compared to $\hat{\mu}$
3. Properties of $\hat{\gamma}(h)$
4. Simulation of pointwise and "sequence-wide" properties

See S&S, Appendix A, for further details on the properties of these estimators that we'll cover in the next class. These notes consider the means and covariances of the estimators. The text shows that the estimators are also approximately normally distributed.

Moment estimators

Estimators

Mean:
$$\overline{X} = \frac{\sum_{t=1}^{n} X_t}{n} \left(= \frac{\mathbf{1}' X_1^n}{n} \right)$$

Covariances:
$$\hat{\gamma}(h) = \frac{\sum_{t=1}^{n-|h|} (X_t - \overline{X})(X_{t+|h|} - \overline{X})}{n} \;, \quad h = \ldots, -2, -1, 0, 1, 2, \ldots$$

Correlations:
$$\hat{\rho}(h) = \frac{\hat{\gamma}(h)}{\hat{\gamma}(0)}$$

Estimating the mean

Kronecker lemma If the real-valued sequence c_i is absolutely summable, then
$$\lim_{n\to\infty} \sum_{i=0}^{n} \tfrac{i}{n} |c_i| = 0 \;. \tag{1}$$

The proof works by taking some N for which $\sum_{N+1}^{\infty} |a_j| < \epsilon$ for arbitrary $\epsilon > 0$. Then divide the sum as $\sum_{i=0}^{n} = \sum_0^N + \sum_{N+1}^n$ and observe that the first summand goes to zero as $n \to \infty$ and the second is bounded by ϵ. Presto.

Unbiased Easy to see that $\overline{X}$ is unbiased. What about $\hat\mu$ if Γ is known (unknown)?

Variance Were $\{X_t\}$ uncorrelated, then $\text{Var}(\overline{X}) = \sigma_x^2/n$. What happens in the case of dependence?

Let $\sigma_x^2 = \gamma(0)$ so that the final expression looks more familiar. The variance of $\overline{X}$ is the sum of the elements in the $n \times n$ covariance matrix $\Gamma = [\gamma(i-j)]$, $i, j = 1, \ldots, n$ divided by n^2:

$$\begin{aligned}
\text{Var}(\overline{X}) &= \frac{1}{n^2} \sum_{t,s} \gamma(t,s) \\
&= \frac{1}{n^2} \sum_{t,s} \gamma(t-s) \\
&= \frac{\sigma^2}{n^2} \sum_{t,s} \rho(t-s) \\
&= \frac{\sigma^2}{n} \sum_{d=-(n-1)}^{n-1} \left(1 - \frac{|d|}{n}\right) \rho(d)
\end{aligned}$$

Comparison to GLS Compare

$$\text{Var}(\overline{X}) = \mathbf{1}'\Gamma\mathbf{1}/n^2 \quad \text{to } \text{Var}(\hat\mu) = \frac{1}{\mathbf{1}'\Gamma^{-1}\mathbf{1}} .$$

Start by forming a unit vector $u = \mathbf{1}/\sqrt{n}$ and recognize the comparison of eigenvalues. When are these the same?

Limiting variance In the limit, then (assuming that the correlation function is summable and using (1))

$$\lim_n n \ \text{Var}(\overline{X}) = \sigma^2 \sum_{h=-\infty}^{\infty} \rho(h)$$

Example Consider the case in which $\rho(h) = \alpha^h$ for some $|\alpha| < 1$. (Which process is this?) Then

$$\begin{aligned}\mathrm{Var}(\overline{X}) &= \frac{\sigma^2}{n} \sum_{h=-\infty}^{\infty} \rho(h) = \frac{\sigma^2}{n} \frac{2}{1-\rho} - 1 \\ &= \frac{\sigma^2}{n} \frac{1+\rho}{1-\rho} = \frac{\sigma^2}{n(1-\rho)/(1+\rho)} \ .\end{aligned}$$

Equivalent uncorrelated observations What's the "cost" of working with dependent data in terms of learning the mean of a stationary process? If we use $\overline{X}$ to estimate μ, then n dependent observations with correlation function $\rho(h)$ are only worth (in the sense of the length of confidence intervals, say)

$$n \, \frac{1-\rho}{1+\rho}$$

uncorrelated observations. (Of course, negative dependence is a whole different experience!)

ρ	$n \, \frac{1-\rho}{1+\rho}$
0	n
0.5	$n/3$
0.8	$n/9$
0.95	$n/39$

Simulation evidence

Model Simulate realizations of the autoregression

$$(X_t - \mu) = \alpha(X_{t-1} - \mu) + w_t, \qquad w_t \overset{\text{iid}}{\sim} N(0,1)$$

for $|\alpha| < 1$. The mean of the process is μ. The process is often written as

$$X_t = \mu(1-\alpha) + \alpha X_{t-1} + W_t \tag{2}$$

Notice that the "intercept" in this equation is not the mean.

Covariance function By back-substitution, it is easy to show that

$$\mathrm{Corr}(X_{t+h}, X_t) = \rho(h) = \alpha^{|h|} \ .$$

Convenient matrix inverse For this process, we can simulate the GLS estimator of $\mu = 0$ using the sample correlation. That allows us to compare three estimators ($X'_{1:n} = (X_1, \ldots, X_n)$):

- The OLS estimator, ignoring the dependence: $\overline{X}$
- The GLS estimator, using α known: $\mathbf{1}'\Gamma^{-1}X_{1:n}/D$
- The approximate GLS estimator: $\mathbf{1}'\hat{\Gamma}^{-1}X_{1:n}/\hat{D}$

where $D = \mathbf{1}'\Gamma^{-1}\mathbf{1}$. It is easy to do this simulation because we can approximately factor Γ^{-1}. Notice that $X_t - \alpha X_{t-1} = W_t$. Hence the product

$$\underbrace{\begin{pmatrix} 1 & 0 & 0 & 0 & 0 & \ldots \\ -\alpha & 1 & 0 & 0 & 0 & \ldots \\ 0 & -\alpha & 1 & 0 & 0 & \ldots \\ 0 & 0 & -\alpha & 1 & 0 & \ldots \\ 0 & 0 & 0 & \ddots & & \ldots \end{pmatrix}}_{A} \begin{pmatrix} X_1 - \mu \\ X_2 - \mu \\ X_3 - \mu \\ X_4 - \mu \\ \vdots \end{pmatrix} = \begin{pmatrix} X_1 - \mu \\ w_2 \\ w_3 \\ w_4 \\ \vdots \end{pmatrix} \approx w_{1:n}$$

but for the end effects. It follows that $A \approx \Gamma^{-1/2}$ since A converts $X_{1:n}$ into white noise (but for the first value). To make the expression work perfectly, standardize the leading random variable (which has variance $\sigma^2/(1-\alpha^2)$ to have the variance σ^2 as W_t. Do this by replacing the leading diagonal 1 in A by $\sqrt{1-\alpha^2}$.

Mean estimator If we apply A directly to $X_{1:n}$, then the average of the result is approximately (look back at (2)) $\mu(1-\alpha)$, so to estimate μ, we have to then divide by $1 - \hat{\alpha}$.

Effects of mean estimation

Goal Approximate the covariance of the covariances within a certain order of error: could you recognize the process, such as an AR(1) shown in (2) from seeing estimates of $\gamma(h)$ or $\rho(h)$?

Note: Treat the lag h as positive throughout to avoid having to add absolute values in the expressions that follow.

Simplify calculations by replacing the sample mean $\overline{X}$ by μ. To see the effect of this substitution, add and subtract μ in each of the terms in the numerator of $\hat{\gamma}(h)$,

$$n\hat{\gamma}(h) = \sum_{t=1}^{n-h}(X_{t+h}-\mu)(X_t-\mu)-(X_{t+h}-\mu)(\overline{X}-\mu)-(X_t-\mu)(\overline{X}-\mu)+(\overline{X}_n-\mu)^2$$

Only the first term on the r.h.s. is ultimately needed since the other terms are of smaller magnitude. Basically, the sum of the deviations from the mean is zero (but for end effects). For example,

$$\sum_{t=1}^{n-h}(X_t-\mu)(\overline{X}-\mu) = (n-h)(\overline{X}-\mu)(\overline{X}_h-\mu) \approx (n-h)(\overline{X}_n-\mu)^2 \ ,$$

where $\overline{X}_h$ is the average of $X_1, X_2, \ldots, X_{n-h}$. When we divide by n, the expected value of this term is approximately

$$\mathbb{E}\,(1-\frac{h}{n})(\overline{X}_n-\mu)^2 \approx \mathrm{Var}(\overline{X}_n) \to 0 \text{ as } n \to \infty.$$

Simplified estimator. With $\overline{X}$ replaced by μ, assume "without loss of generality" that $\mu = 0$ and consider

$$c(h) = \frac{\sum_{t=1}^{n-h} X_{t+h}X_t}{n}, \quad h \geq 0,$$

with $c(-h) = c(h)$.

Covariances of covariances

Bias The p.s.d. estimator of $\gamma(h)$ is biased, having been pulled toward zero (particularly for h large),

$$\mathbb{E}\,c(h) = \left(1-\frac{h}{n}\right)\gamma(h)\ . \tag{3}$$

By definition

$$\mathrm{Cov}(c(h), c(h+v)) = \mathbb{E}\ (c(h)c(h+v)) - \mathbb{E}\,c(h)\ \mathbb{E}\,c(h+v).$$

Expand the terms on the right, noting from (3) that the estimates are biased,

$$\mathrm{Cov}(c(h), c(h+v)) = \frac{1}{n^2} \sum_{t=1}^{n-h} \sum_{s=1}^{n-(h+v)} \mathbb{E}\, X_t X_{t+h} X_s X_{s+h+v} - (1-\frac{h}{n})(1-\frac{h+v}{n})\gamma(h)\gamma(h+v). \tag{4}$$

Isserlis' theorem The expected value of the product of four r.v.'s with mean zero is given by the old (*Biometrika*, 1918) result:

$$\mathbb{E}\, Y_1 Y_2 Y_3 Y_4 = \mathbb{E}\,(Y_1Y_2)\mathbb{E}\,(Y_3Y_4) + \mathbb{E}\,(Y_1Y_3)\mathbb{E}\,(Y_2Y_4) + \mathbb{E}\,(Y_1Y_4)\mathbb{E}\,(Y_2Y_3) + \kappa_4\ ,$$

where κ_4 is the fourth-order cumulant of the joint distribution of Y_1, Y_2, Y_3, and Y_4. The fourth-order cumulant measures the kurtosis of the joint distribution. In the univariate case,

$$\kappa_4 = \mathbb{E}\,(Y - \mathbb{E}\,Y)^4 - 3\mathbb{E}\,(Y - \mathbb{E}\,Y)^2.$$

You can check that $\kappa_4 = 0$ for normal random variables. The key step in the proof is "regression":

$$\mathbb{E}\,(X_1X_2X_3X_4) = \mathbb{E}\,(\mathbb{E}\,(x_1X_2X_3X_4 | X_1 = x_1))$$

and noting that (for normal r.v.) the conditional expectations have the form $\mathbb{E}\,(X_2|X_1) = (\rho\sigma_2/\sigma_1)x_1$.

Covariances, cntd. Using Isserlis' expression in (4)

$$\begin{aligned} \mathrm{Cov} \quad &= \quad \frac{1}{n^2} \sum_{t=1}^{n-h} \sum_{s=1}^{n-h-v} \mathbb{E}\, X_{t+h}X_t\, \mathbb{E}\, X_{s+h+v}X_s + \mathbb{E}\, X_tX_s\, \mathbb{E}\, X_{t+h}X_{s+h+v} \\ &\qquad + \mathbb{E}\, X_tX_{s+h+v}\, \mathbb{E}\, X_{t+h}X_s - (1-\frac{h}{n})(1-\frac{h+v}{n})\gamma(h)\gamma(h+v). \end{aligned}$$

Now replace the expectations of products by the covariances, assuming the process is Gaussian, (otherwise have a cumulant term)

$$\begin{aligned} \mathrm{Cov} \quad &= \quad \frac{1}{n^2} \sum_{t=1}^{n-h} \sum_{s=1}^{n-h-v} \gamma(h)\gamma(h+v) + \gamma(t-s)\gamma(t-s-v) + \gamma(t-s-h-v)\gamma(t+h-s) \\ &\quad -(1-\frac{h}{n})(1-\frac{h+v}{n})\gamma(h)\gamma(h+v). \end{aligned}$$

The first term on the right cancels with the product of the covariances since it has neither of the variables being summed (t nor s). Thus,

$$\text{Cov} = \frac{1}{n^2}\sum_{t=1}^{n-h}\sum_{s=1}^{n-h-v} \gamma(t-s)\gamma(t-s-v) + \gamma(t-s-h-v)\gamma(t-s+h)\ .$$

Since the indices in the sum always appear in the stationary form $t-s$, we can collapse this sum just as we did in the expression for the variance of $\overline{X}$. Replacing $t-s$ by m, we obtain the useful approximation first derived by Bartlett (1946),

$$\text{Cov}(\hat\gamma(h), \hat\gamma(h+v)) \approx \frac{1}{n}\sum_{|m|} \gamma(m)\gamma(m-v) + \gamma(m-h-v)\gamma(m+h) \quad (5)$$

Example. Consider again an AR(1) process for which $\gamma(h) = \gamma(0)\alpha^h$. From (5) with $v = h = 0$,

$$\text{Var}(\hat\gamma(0)) \approx \frac{2}{n}\sum_{|m|}\gamma(m)^2 = \frac{2\gamma(0)^2}{n}\sum_{|m|}\alpha^{2|m|}.$$

Since this is a geometric series,

$$\text{Var}(c_0) \approx \frac{2\gamma(0)^2}{n}\frac{1+\alpha^2}{1-\alpha^2}.$$

Compare this to the iid normal case where for $Z_1, \ldots, Z_n \sim N(0, \sigma_z^2)$ for which the variance of the usual variance estimator s_z^2 is

$$\text{Var}(s_z^2) \approx \frac{2\sigma_z^4}{n}\ .$$

Thus, the effect of the dependence upon the estimated variance is less than its effect upon the estimate of the mean ($\text{Var}\,\overline{X}_n = \frac{\sigma_x^2}{n}\frac{1+\alpha}{1-\alpha}$). (Humm... Is this true in general???)

More general expressions. In general, using (5)

$$\text{Var}(\hat\gamma_h) \approx \frac{1}{n}\sum_{|m|}\gamma(m)^2 + \gamma(m+h)\gamma(m-h).$$

This formula reduces to the usual expression for the variance of the familiar estimator s^2 when $\alpha = 0$. Hence, the variance of covariance estimates depends on

1. the lag h we are considering and
2. the properties of the underlying process.

In the case of a non-Gaussian process, one has to assume more than second-order stationarity in order to derive these properties. In particular, we have to assume that the process is stationary to fourth order, so that moments like $\mathbb{E}\, X_{t+h} X_t X_{s+v} X_s$ do not depend on the time origin.

Correlations

Similar expressions exist for the covariances of the estimated correlation function formed by the ratio $\hat\rho(h) = \frac{\hat\gamma_h}{\hat\gamma_0}$. The calculations are, in general, more difficult since we must consider properties of a ratio of estimators. Like most classical results in statistics, one uses a Taylor series expansion.

Delta method. Think of the estimated correlation $\hat\rho(h) = g(\hat\gamma(0), \hat\gamma(h)) = \hat\gamma(h)/\hat\gamma(0)$. Now expand the function g about the point $(\gamma(0), \gamma(h))$ and use the linear expansion in place of g.

Result. The approximation, also derived by Bartlett, resembles the previous expression for the covariance of the covariance estimates:

$$\begin{aligned}
\mathrm{Cov}(\hat\rho(h), \hat\rho(h+v)) &\approx \frac{1}{n}\sum_{|m|} \rho(m)\rho(m+v) + \rho(m+h+v)\rho(m-h) \\
&\approx \frac{1}{n}\sum_{|m|} \rho(m)\rho(m+v) \; .
\end{aligned}$$

The last approximation drops the second summand and *only applies for large sample size n and large lags h.*

Stationarity Notice that, in the most approximate approximation, the covariances of the autocorrelations *do not* depend on h, but only on the separation v.

In this sense, the estimated autocorrelations of a stationary process at substantial lags themselves behave like a *stationary process.* Their covariance function only depends on the separation, not the lags being considered. Indeed, the estimated correlations are more strongly correlated than the process itself.

No cumulant The cumulant κ_4 does not appear in this approximation for the covariance of the estimated correlations even if the process is not Gaussian. Its absence only requires that process under consideration is a moving average of an independent white noise sequence rather than an uncorrelated sequence — normality is not needed.

Examples. If the process is white noise, the covariances of the autocorrelation estimates are approximately zero. The variance of the estimates is about $1/n$. These properties are the origin of the +/- limits that appear in many autocorrelation plots (noted in the text, equation 1.38).

For an AR(1) process with coefficient $|\alpha| < 1$

$$\text{Var}(\hat{\rho}(h)) \approx \frac{1}{n} \sum_{|m|} \alpha^{2|m|} = \frac{1}{n} \frac{1+\alpha^2}{1-\alpha^2}$$

and

$$\text{Cov}(\hat{\rho}(h), \hat{\rho}(h+v)) \approx \frac{1}{n} \alpha^{|v|} \left(\frac{1+\alpha^2}{1-\alpha^2} + |v|\right).$$