

14.384 Time Series Analysis, Fall 2008
 Recitation by Paul Schrimpf
 Supplementary to lectures given by Anna Mikusheva
 October 24, 2008

Recitation 8

More Empirical Process Theory

This section of notes, especially the subsection on stochastic integrals are based on <http://cscs.umich.edu/~crshalizi/weblog/472.html>.

Convergence of Random Walks

We never did formally show the weak convergence of $\frac{1}{\sqrt{T}} \sum_{t=1}^{\lceil \tau T \rceil}$ to a Brownian motion. Let's fill that hole now. Remember the functional central limit theorem.

Theorem 1. Functional Central Limit Theorem: *If*

1. *there exists a finite-dimensional distribution convergence of ξ_T to ξ (as in (??))*
2. $\forall \varepsilon, \eta > 0$ *there exists a partition of Θ into finitely many sets, $\Theta_1, \dots, \Theta_k$ such that*

$$\limsup_{T \rightarrow \infty} P(\max_i \sup_{\tau_1, \tau_2 \in \Theta_i} |\xi_T(\tau_1) - \xi_T(\tau_2)| > \eta) < \varepsilon$$

then $\xi_T \Rightarrow \xi$

Finite dimensional convergence follows from a standard central limit theorem. We just need to verify the second condition, stochastic equicontinuity. It is not very hard to do this directly. $\Theta = [0, 1]$ is compact, so we can choose our partition to be a collection of intervals of length δ . This gives

$$\begin{aligned} P(\max_i \sup_{\tau_1, \tau_2 \in \Theta_i} |\xi_T(\tau_1) - \xi_T(\tau_2)| > \eta) &\leq P(\sup_{|\tau_1 - \tau_2| < \delta} |\xi_T(\tau_1) - \xi_T(\tau_2)| > \eta) \\ &\leq P(\sup_{\tau \in [0, 1]} |\frac{1}{\sqrt{T}} \sum_{t=\lceil \tau T \rceil}^{\lceil (\tau + \delta) T \rceil} \epsilon_t| > \eta) \end{aligned}$$

Chebyshev's inequality says that if $E[x] = \mu$ and $Var(x) = \sigma^2$, then $P(|x - \mu| > \eta) \leq \frac{\sigma^2}{\eta^2}$. Here, $E[\frac{1}{\sqrt{T}} \sum_{t=\lceil \tau T \rceil}^{\lceil (\tau + \delta) T \rceil} \epsilon_t] = 0$ and $Var(\frac{1}{\sqrt{T}} \sum_{t=\lceil \tau T \rceil}^{\lceil (\tau + \delta) T \rceil} \epsilon_t) = \frac{[\delta T] \sigma^2}{T}$ and these things do not depend on τ , so

$$P(\sup_{\tau \in [0, 1]} |\frac{1}{\sqrt{T}} \sum_{t=\lceil \tau T \rceil}^{\lceil (\tau + \delta) T \rceil} \epsilon_t| > \eta) \leq \frac{[\delta T] \sigma^2}{T \eta^2}$$

Hence, setting $\delta \leq \frac{\eta^2 \varepsilon}{\sigma^2}$ we have the desired result.

Stochastic Integrals

Our interest in stochastic integrals comes from the fact that we would like to say something about the convergence of

$$\frac{1}{T} \sum y_t \epsilon_t = \frac{1}{T} \sum \xi_T(t/T) \sqrt{T} (\xi_T(t/T) - \xi_T((t-1)/T))$$

We know that $\xi_T \Rightarrow W$, a Brownian motion. If W were differentiable, we'd also know that $\lim_{T \rightarrow \infty} \sqrt{T}(W(\tau T/T) - W(([\tau T] - 1)/T)) = W'(\tau)$. Finally, a sum like the above would usually converge to an integral

$$\frac{1}{T} \sum_{t=1}^T f(t/T) g'(t/T) \rightarrow \int_0^1 f(x) dg(x)$$

We would like to generalize the idea of an integral to apply to random functions such as ξ . There are a few problems. First, W is not differentiable. Nonetheless, we can still look at

$$\lim_{T \rightarrow \infty} \frac{1}{T} \sum_{t=1}^T \xi(t/T) (W(t/T) - W((t-1)/T))$$

We must verify that this sum converges to something. We will call its limit a stochastic integral, and write it as $\int_0^1 \xi(t) dW(t)$. We then want to define stochastic integrals more generally and make them have some of the properties of Riemann integrals. In particular, we want to make sure that the limit of the sum does not depend on how we partition $[0, 1]$, i.e. for any increasing $t_i^T \in [0, 1]$ such that $\lim_T \max_i t_i^T - t_{i+1}^T = 0$ we should have

$$\lim_{T \rightarrow \infty} \frac{1}{T} \sum_{t=1}^T \xi(t/T) * (W(t/T) - W((t-1)/T)) = \lim_{T \rightarrow \infty} \frac{1}{T} \sum_{t=1}^T \xi(t_i^T) * (W(t_{i+1}^T) - W(t_i^T))$$

Furthermore, we want to define

$$\int x(t) dW(t)$$

for a suitable broad class of stochastic functions, $x(t)$.

Here's a sketch of how to proceed:

1. Define stochastic integrals for *elementary processes*. An elementary process is a step function, like ξ_T . More precisely, $X(t)$ is elementary if there exist $\{t_i\}$ such that $X(t) = X(t_i)$ is $t \in [t_i, t_{i+1})$. Also, assume that $E \int X(t)^2 dt < \infty$, then we define our integral as:

$$\int X(t) dW = \sum X(t_i) (W(t_{i+1}) - W(t_i))$$

we have no limits or choice of partitions here, so this is a well defined object.

2. Note that any $X(t)$ that is square integrable, $E \int X(t)^2 dt < \infty$, can be approximated by a sequence of elementary processes, X_n , in the sense that $\lim_n E \left[\int (X(t) - X_n(t))^2 dt \right] \rightarrow 0$
3. Define $\int X(t) dW(t) = \lim_n \int X_n(t) dW(t)$ where X_n is any sequence of elementary processes converging to $X(t)$. Show that this definition does not depend on the choice of sequence.

Ito's Lemma

An Ito process is defined as

$$X(t) = X_0 \int_0^1 A(s)ds + \int_0^t B(s)dW$$

where X_0 is random variable independent of W and A and B are stochastic processes. This is also written as

$$dX = Adt + BdW$$

Ito's lemma tells us how write $f(t, X(t))$ in terms of integrals instead of a composition of functions. Ito's lemma says that

$$dF = \frac{\partial f}{\partial t}(t, X(t))dt + \frac{\partial f}{\partial x}(t, X(t))dX + \frac{1}{2}B(t)^2 \frac{\partial^2 f}{\partial x^2}(t, X(t))dt$$

or in terms of integrals,

$$F(t) - F(0) = \int_0^t \left[\frac{\partial f}{\partial t}(s, X(s)) + A(s) \frac{\partial f}{\partial x}(s, X(s)) + \frac{1}{2}B(s)^2 \frac{\partial^2 f}{\partial x^2}(s, X(s)) \right] dt + \int_0^t B(s) \frac{\partial f}{\partial x}(s, X(s))dW$$

Remark 2. Let's compare this lemma to what we would have with nonstochastic integrals and functions. Suppose $x(t) = x_0 + \int_0^t A(s)ds + \int_0^t B(s)dW$ where all integrals and functions are nonstochastic. Let $F(t) = f(t, x(t))$. Differentiating gives:

$$F'(t) = \frac{\partial f}{\partial t} + \frac{\partial f}{\partial x} (A(t) + B(t)W'(t))$$

which is the same as Ito's lemma, but without the second order term. The reason the second order term appears in Ito's lemma is that $dWdW$ is not "small", unlike dt .

Example 3. Consider $\int WdW$. In lecture we saw that $\int WdW = \frac{W(t)^2 - t}{2}$. We "showed" this by working with ξ_T and rearranging various sums. Using the definition of stochastic integrals and the exact same argument with ξ_T playing the role of elementary processes you can formally verify this equality. Alternatively, Ito's lemma can be used. Take $X(t) = W(t)$. This is an Ito process with $A = 0$ and $B = 1$. Consider $f(t, X(t)) = X(t)^2/2$, then

$$\begin{aligned} dF &= \frac{\partial f}{\partial t}(t, X(t))dt + \frac{\partial f}{\partial x}(t, X(t))dX + \frac{1}{2}B(t)^2 \frac{\partial^2 f}{\partial x^2}(t, X(t))dt \\ &= X(t)dX + \frac{1}{2}dt \\ &= WdW + \frac{1}{2}dt \end{aligned}$$

and integrating tells us:

$$\begin{aligned} \frac{1}{2} \int dW^2 &= \int WdW + \frac{t}{2} \\ \int_0^t WdW &= \frac{1}{2}(W(t)^2 - t) \end{aligned}$$

Convergence to Stochastic Integrals

De Jong and Davidson (2000) give general conditions for sums of the type

$$\sum_{t=1}^T \left(\sum_{s=1}^t u_s \right) w_{t+1}$$

to converge to $\int UdW$. I would say something more about this if I had time.

Local to Unity Asymptotics

An important general point to get from the lecture on local to unity asymptotics is that uniform convergence is often important for asymptotic results to give good approximations in practice. Recall that uniform convergence means that:

$$\sup_{\theta \in \Theta} \sup_x |P(t(\theta, T) \leq x) - \Phi(x)| \not\rightarrow 0$$

What this means is that no matter how large your sample size, there is some parameter value, θ , such that the error in the asymptotic approximation remains large. Whether this is merely a theoretical curiosity or a real problem largely depends on the situation. Over the last fifteen years, lots of simulation and empirical evidence has shown that lack of uniform convergence does create practical problems in econometrics. Anna talked about how badly sized unit root tests can be. The problems with weak instruments can also be thought of as due to the lack of uniformity of convergence for parameters with the first stage coefficients near 0.

Testing for Breaks

I've never found testing for breaks to a particularly useful endeavor in applied macro. There just is not enough data to reach strong conclusions. However, testing for breaks raises many interesting econometric issues, which I suspect is why it has generated so many papers. One thing that makes breaks interesting is that their limit theory is nonstandard. Another is that testing involves nuisance parameters that are only present under the alternative. Let's take a moment to review the standard approach to hypothesis testing.

Suppose we want to test $H_0 : \theta \in \Theta_0$ against $H_a : \theta \in \Theta_1$. Moreover, being clever econometricians, we want to do the "best" test possible. What do we mean by best? Usually, we mean the most powerful test with correct size. A basic result is the Neyman-Pearson lemma.

Lemma 4. Neyman-Pearson *the most powerful test for testing a point hypothesis $H_0 : \theta = \theta_0$ against a point alternative is $H_a : \theta = \theta_1$ is the likelihood ratio test, i.e. reject if $\frac{L(\theta_0|x)}{L(\theta_1|x)} \leq \eta$, where η is such that $\alpha = P(\frac{L(\theta_0|x)}{L(\theta_1|x)} \leq \eta | H_0)$*

This is an important result, and it is the starting point for much of our thinking about testing problems. Unfortunately there are some difficulties in extending it to more complicated situations. For one thing, it is for exact finite sample tests, but we usually work with asymptotic tests.

To analyze asymptotic tests, we simply think about tests that have the correct size asymptotically, and try to maximize asymptotic power against local alternatives. That is, we seek to maximize:

$$\lim_{T \rightarrow \infty} P(\text{reject} | \theta = \theta_{1,T} = \theta_0 + T^{-1/2}\delta)$$

Among tests such that

$$\lim P(\text{accept} | \theta = \theta_0) = \alpha$$

An important result says that every sequence of tests of asymptotic size α converges to a test of size α in the "limit experiment", i.e. the world where the distribution of $\hat{\theta}$ is its asymptotic distribution. The Neyman-Pearson lemma tells us that in the limit experiment, the likelihood ratio test is the most powerful test for point hypotheses. Furthermore, we usually know the distribution of the likelihood ratio in the limit experiment, so this test is easy to implement.

Another limitation of the Neyman-Pearson lemma is that it only applies to point hypotheses. For composite hypotheses, a good definition of the "best" test is one that is uniformly most powerful against all alternatives. That is a test of correct size that also maximizes power for each $\theta_1 \in \Theta_1$, i.e. it maximizes $P(\text{reject} | \theta = \theta_1 \in \Theta_1)$ for every θ_1 . Unfortunately, in general, uniformly most powerful tests do not exist. A simple example is testing if y has a two dimensional normal distribution with mean 0. We want to test

$H_0 : \mu = 0$ vs $H_a : \mu \neq 0$. We can draw many different ellipsoidal (or any other shape) acceptance regions around 0 that have correct size, but it is clear that no single acceptance rule will be uniformly most powerful. For example, a thin, long acceptance regions will have high power in some directions, but low power in others.

The lack of uniformly most powerful tests is disappointing because it introduces some arbitrariness into defining what we mean by the “best” test. One way to proceed is to require some sort invariance in testing procedures. For example, we might require that the gives the same answer when we shift our data by adding a constant, or when we apply a rotation, or more generally when we take any linear combination. For testing whether the mean of a multivariate normal is 0, the likelihood ratio test is the uniformly most powerful rotationally invariant test.

Another common testing problem is that we often only want to test a subset of our parameters, say θ^1 , while we do not care about the value of the rest of the “nuisance” parameters, θ^2 . If we can consistently estimate the nuisance parameters under H_0 , then we test conditional on our estimated nuisance parameters, and conditional likelihood ratio tests will be optimal in the sense described above.

Nuisance parameters become even more of a nuisance if they cannot be consistently estimated under H_0 . One way to proceed is to try to find a test that is invariant to any possible value of the nuisance parameters, so that we ignore them. Unfortunately this type of invariance is often too stringent of a requirement and invariant tests might not exist. In this situation, it is still easy enough to think of tests with correct size. An obvious test statistic to consider is the maximum over possible values of the nuisance parameters of the likelihood ratio statistic. For example, when testing for breaks, the nuisance parameter is the time of the break, $t_0 = [\delta T]$. The Quant statistic is the maximum of the likelihood ratio.

$$Q = \sup_{[\delta T] \leq t_0 \leq [(1-\delta)T]} F_T\left(\frac{t_0}{T}\right)$$

This is a fine test statistic, and given the Neyman-Pearson lemma, we would expect it to have good power properties. But how good? One parsimonious way of quantifying the test’s power properties is to place a distribution on possible values of the nuisance parameters and parameters of interest, say $F(\delta, \theta)$, and then look at weighted average power,

$$WAP = \int P(\text{reject} | \theta = (\theta_1^1, \delta)) dF(\delta, \theta)$$

This idea goes back to Wald. He showed that likelihood ratio tests for $H_0 : \mu = 0$ for normal distributions maximize WAP over alternatives with uniform weight on $\{\theta_1 : \|\Sigma^{-1}\theta_1\| = c\}$. Andrews and Ploberger (1994) considered the more general problem of maximizing weighted average power in the presence of nuisance parameters. They showed that the test statistic that maximizes WAP is of the form:

$$Exp - LR = (1 + c)^{-p/2} \int \exp\left(\frac{1}{2} \frac{1}{1 + c} LR(\delta)\right) dF(\delta)$$

where $LR(\delta)$ is the usual likelihood ratio statistic for test $H_0 : \theta = 0$ vs $H_a : \theta \neq 0$, $\delta_0 = \delta$, $F(\delta)$ is the marginal weight δ and c is a constant that depends on the joint weight function, $F(\delta, \theta)$. Furthermore, since LR, Lagrange multiple, and Wald tests are asymptotically equivalent, the LR statistic can be replaced with the LM or Wald statistic.

MIT OpenCourseWare
<http://ocw.mit.edu>

14.384 Time Series Analysis
Fall 2013

For information about citing these materials or our Terms of Use, visit: <http://ocw.mit.edu/terms>.