Unit Roots

Now consider the unstable (unit root) AR(1) process

$$x_t = x_{t-1} + \varepsilon_t \quad t = 1, 2, 3, \ldots \tag{*}$$

with $\varepsilon_t \sim iid(0, \sigma^2)$ and $x_0 = 0$.

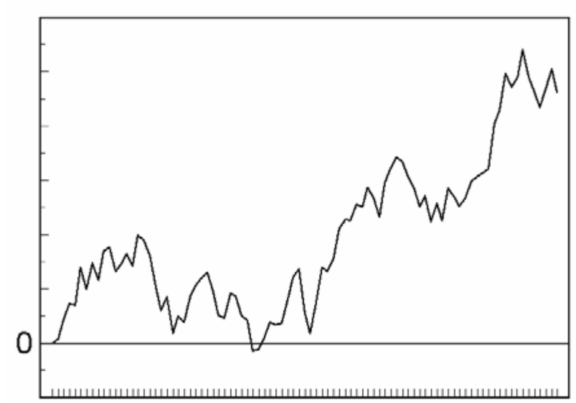
The solution is $x_t = \sum_{s=1}^{t} \varepsilon_s$, so that $E(x_t) = 0$, but

$$\text{Var}(x_t) = t\sigma^2$$

$$\text{Cov}(x_t x_{t+m}) = t\sigma^2 \text{ for all } m.$$

Therefore, the process is nonstationary, and has long memory. This is called a stochastic trend process.

Here is a realization of length 100 from (*), where $\varepsilon_t \sim N(0, 1)$:



Non-Standard Asymptotics

The usual limit theorems don't apply here.

If

$$\bar{x} = T^{-1} \sum_{t=1}^{T} x_t$$

note that

$$\bar{x}_T = T^{-1} \sum_{t=1}^{T} \sum_{s=1}^{t} \varepsilon_s = T^{-1} \sum_{t=1}^{T} (T - t + 1)\varepsilon_t$$

so that

$$\mathrm{Var}(\bar{x}_T) = \sigma^2 \frac{1}{T^2} \sum_{t=1}^{T} (T - t + 1)^2 = \sigma^2 \frac{1}{T^2} \sum_{t=1}^{T} t^2 = \sigma^2 \frac{T(T+1)(2T+1)}{6T^2} \to \infty$$

Thus, the sample mean diverges to infinity.

However,also note that

$$\mathrm{Var}\left(\frac{\bar{x}_T}{\sqrt{T}}\right) = \sigma^2 \frac{T(T+1)(2T+1)}{6T^3} \to \frac{\sigma^2}{3}$$

Further the CLT for nonstationary (trending variance) processes yields
$\frac{\bar{x}_T}{\sqrt{T}} \xrightarrow{d} N\left(0, \frac{\sigma^2}{3}\right)$.

Functional Central Limit Theorem

We also know that

$$\frac{x_T}{\sqrt{T}} = \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \varepsilon_t \xrightarrow{\mathrm{d}} N(0,\sigma^2).$$

Note that this result extends to

$$X_T(r) = \frac{x_{[Tr]}}{\sigma\sqrt{T}} = \frac{1}{\sigma\sqrt{T}} \sum_{t=1}^{[Tr]} \varepsilon_t \xrightarrow{\mathrm{d}} N(0,r), \quad 0 < r \le 1$$

where $[Tr]$ denotes the largest integer not exceeding Tr (the "floor" function).

X_T is a function of a real variable (step function) defined on the unit interval.

The CLT can be extended to a result of the form

$$X_T \xrightarrow{\mathrm{d}} B$$

where the limit function B is called *Brownian motion*.

This is the *FCLT* (the i.i.d. case is Donsker's Theorem).

This is a new kind of convergence problem, where the objects in question are not random variables or vectors, but random functions.

We consider some mathematical background to this convergence problem...

Random Functions

Let R denote the collection of all possible real-valued functions, $x : [0,1] \mapsto \mathbb{R}$.

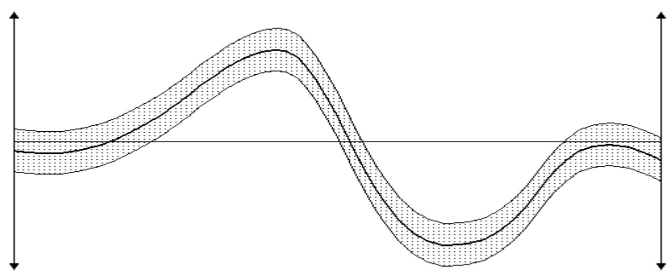
- Recall: a function is a correspondence that associates *every* element of the domain (t) with *one unique* element of the range ($x(t)$).
- If the domain represents time, we call x a process. If the x we observe arises randomly, it is a *stochastic process*.

Can we make R into a measurable space by defining a suitable Borel field?

We must define a metric (distance measure) on R: for example, the uniform metric:

$$d_U(x,y) = \sup_t |x(t) - y(t)|.$$

Definition: The Borel field $\mathcal{B}_R$ is the smallest σ-field of the metric space (R, d_U) containing the open spheres $B(x,r) = \{y : d_U(x,y) < r\}$ all $x \in R$ and $r > 0$.



A Note on Metrics

In the space $\mathbb{R}$ of real numbers, we take the existence of a measure of the distance between points x and y for granted - it's $|x - y|$.

However, in arbitrary sets of objects, the concept of the distance between objects may be non-obvious.

Definition A *metric* is a distance measure $d(x,y)$ with the following properties:

1. $d(x,y) = d(y,x)$
2. $d(x,y) = 0$ if and only if $x = y$.
3. $d(x,y) \leq d(x,z) + d(z,y)$ (triangle inequality)

A *metric space* (Ω, d) is any set of objects Ω having a metric d defined for each pair of set members.

Examples:

- In $\mathbb{R}^k$, a family of metrics is defined by $d_p(x,y) = \sqrt[p]{\sum_{j=1}^{k}|x_j - y_j|^p}$. Note that $d_\infty = \max_{1\leq j\leq k}|x_j - y_j|$.
- In $\mathbb{R}^2$, the distance d_1 between two points in New York City is the distance by taxicab, while d_2 is the distance as the crow flies.
- A *bounded* metric for $\mathbb{R}$ is $d(x,y) = |x - y|/(1 + |x - y|)$

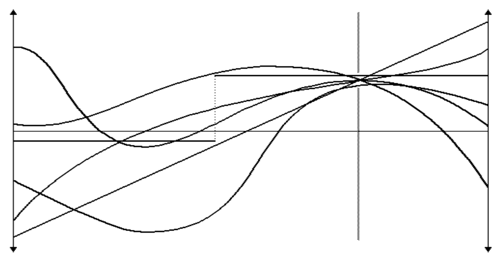
Assigning probabilities

How to assign probabilities to sets of functions? It's not obvious!

Start with the *finite dimensional* sets (cylinder sets).

Definition: The f.d. sets of R are the sets of functions whose co-ordinates $x(t)$ are restricted at only a finite number of points, $t_1, \ldots, t_k$.

e.g. ($k = 1$):



- The f.d. sets can be paired with k –dimensional random vectors, and we know how to assign probabilities to these.
- After defining measures on f.d. sets, we attempt to extend to general classes of sets.
- Letting $\mathcal{H}$ denote the collection of all f.d. sets, the *projection σ-field* is $\mathcal{P} = \sigma(\mathcal{H})$.
- However - $\mathcal{P} \subset \mathcal{B}_R$!! Borel sets exist that are not countable unions of $\mathcal{H}$-sets.

Constructing distributions for functions is accordingly more difficult than for sequences.

The Measurability Problem

Suppose a $\mathcal{F}/\mathcal{P}$-measurable function $X : \ \Omega \mapsto R$ associates each outcome ω with a unique element of R.

A p.m. defined on $(\Omega, \mathcal{F})$ can generate a p.m. on $(R, \mathcal{P})$ by this mapping.

- The finite-dimensional distributions (fidis) of this p.m. are the joint distributions of finite sets of co-ordinates, $x_{t_1}, \ldots, x_{t_k}$.

- Subject to a further technical condition, the Consistency Theorem can be extended to identify this p.m. uniquely with a family of fidis.

- There is no way to represent this p.m. by a simple device like the distribution function or characteristic function.

- However: the expected values $E(f(x))$ for all bounded, continuous functionals $f : R \mapsto \mathbb{R}$, are always defined. These do define the distribution *uniquely*.

(NB a functional is a function whose argument is a function.)

So far, so good...

The problem is that $\mathcal{P}$ may not contain all the sets we might be interested in. We would like to be able to assign probabilities to all the Borel sets.

This problem is approached in a combination in two ways.

1. Make the sample space a subset of *R* containing all the cases of interest.

2. Use a different concept of "closeness" for functions (i.e., define a different metric.)

The Space C

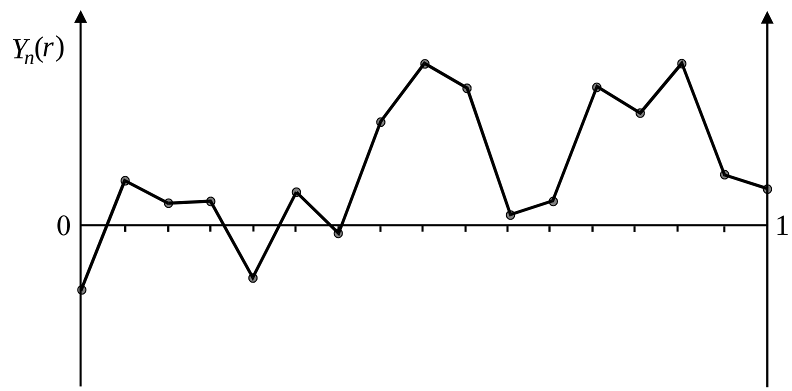
C is the space of continuous functions on [0,1] equipped with the uniform metric d_U.

Some important stochastic processes are elements of C.

Properties of C:

1. C is *separable*. In other words, it contains a countable, dense subset.

- *Countable* means that the set can be put in correspondence with the natural numbers
- A *dense* subset is one that has members arbitrarily close to every member of the space. (i.e. every open sphere contains an element.)
- Example: $\mathbb{R}$ is a separable space. The rational numbers are both dense in $\mathbb{R}$ and countable.
- The set of *piecewise linear* functions (composed of straight lines connecting a finite number of rational vertices) is a countable set which is also dense in C. Every continuous function is *uniformly* close to a piecewise linear function.



2. *C* is *complete*.

This means that if we have a sequence of elements of *C* converging to a limit, the limit is also a member of *C*.

- This generalises the well-known completeness of $\mathbb{R}$ - the limit of a Cauchy sequence of reals must also be real.
- Note: the rational numbers are not complete! The limit of a Cauchy sequence of rationals can be a real.
- Completeness is an important property, because it means that arguments involving limits (which often arise in probability) will not take us outside the space.

A Cauchy sequence $\{x_j\}$ in *C* has the property $d_U(x_j, x_{j+1}) \to 0$ as $j \to \infty$.

Separability and completeness together give a space that resembles $\mathbb{R}$ in important respects.

Although the properties of the "space of functions" may be hard to visualise, our basic intuitions about real numbers carry over to a great extent. This is especially useful when it comes to constructing distributions.

The Borel field of C

Let $\mathcal{B}_C$ denotes the Borel field of C, the σ-field generated by the open sets of C where an open set is defined with respect to the uniform metric.

Separability implies that $\mathcal{P}_C = \mathcal{B}_C$ where $\mathcal{P}_C$ denotes the restriction of $\mathcal{P}$ to C.

Hence, $(C, \mathcal{B}_C)$ *is* a measurable space, by the Consistency Theorem.

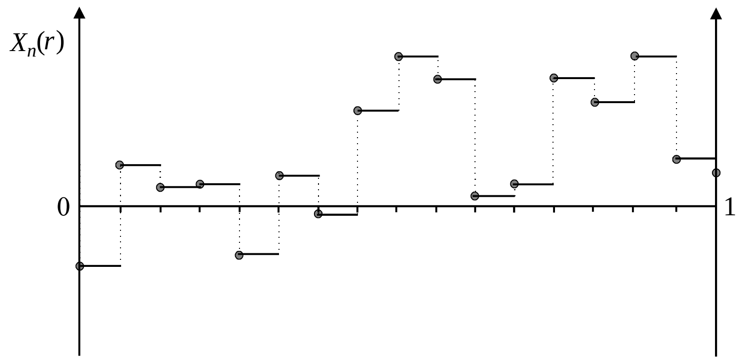
The Space D

D is the space of *cadlag* functions: right-continuous, and having left limits at every point.

(Also sometimes called: functions with discontinuities of the first kind.)

$C \subset D$, since every continuous function is cadlag. However, it contains some important stochastic processes that are not in C.

Example: a step function on [0,1].



The partial sum process

$$X_T(r) = T^{-1/2} \sum_{t=1}^{[Tr]} \xi_t$$

is in D but *not* in C.

- X_T is constant on the ranges $t/T \le r < (t+1)/T$ for $t = 0, \dots, T-1$,
- $T^{-1/2}\xi_t$ represents the height of the jump at time t.

A technical problem:

If we endow D with the uniform metric, $\mathcal{P}_D \subset \mathcal{B}_D$ and the space $(D, \mathcal{B}_D)$ is not a measurable space. $\mathcal{B}_D$ contains some "non-measurable" sets.

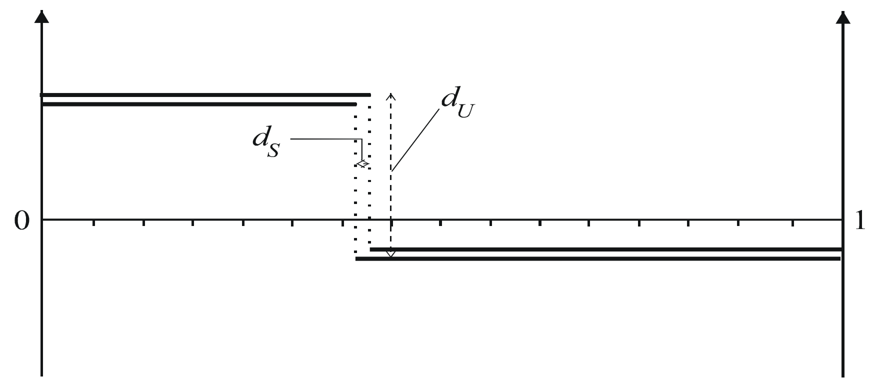
- This problem can be overcome by adopting a different metric for D.
- For example, the Skorokhod metric,

$$d_S = \inf_{\lambda \in \Lambda} \left\{ \varepsilon > 0 : \sup_t |\lambda(t) - t| \leq \varepsilon, \ \sup_t |x(t) - y(\lambda(t))| \leq \varepsilon \right\}$$

where Λ is the set of all increasing continuous functions

$$\lambda : [0,1] \mapsto [0,1].$$

- When functions can have discontinuities, d_S is a more "natural" distance measure.
- In the diagram, the *uniform* distance between the two functions is the height of the jump (d_U) but the Skorokhod distance d_S is the merely the horizontal separation.



- Any metric equivalent to d_S, such that it endows D with the *Skorokhod topology* has the same property.

- Example: when D is equipped with Billingsley's metric (d_B, equivalent to d_S) it is separable and complete. $(D, \mathcal{B}_D)$ can be a measurable space.

Probability Measures on C and D

p.m.s can be assigned to the spaces $(C, \mathcal{B}_C)$ and $(D, \mathcal{B}_D)$, endowed with suitable metrics. (d_U for C but d_B for D).

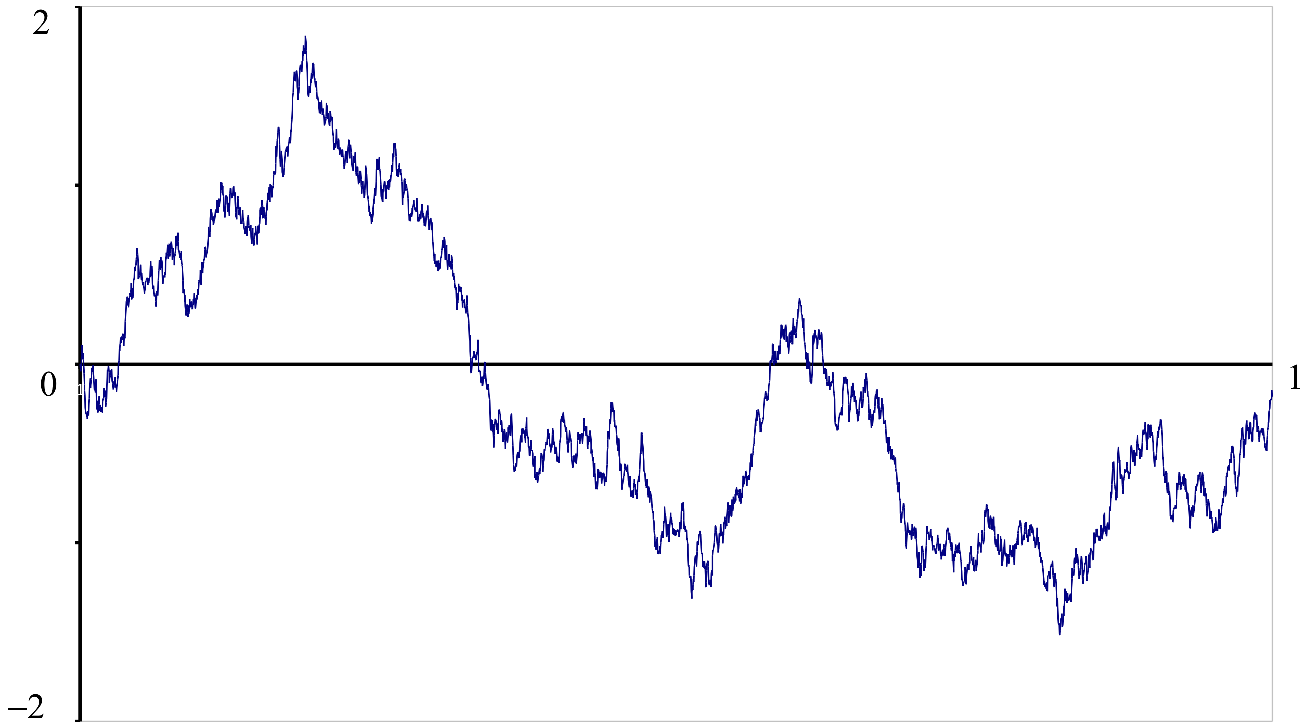
These measures can be constructed by extending the finite-dimensional distributions – the distributions of random vectors $(x(t_1), \dots, x(t_k))$ for all finite collections $t_1, \dots, t_k$ and all $k > 0$.

Example: Wiener measure

Better known as Brownian motion B. Has the following properties:

1. $P(B \in C) = 1$
2. $E(B(t)) = 0$ and $E(B(t)^2) = t$ for $t \geq 0$.
3. For any finite partition $t_0 < t_1 < t_2 <,\dots,< t_k$ of the line, where $t_0 = 0$, the increments $B(t_j) - B(t_{j-1})$, $j = 1,\dots,k$ are totally independent.

A typical sample path of BM:



Properties of Brownian motion

Consider the probability space $(C, \mathcal{B}_C, W)$, W the unique p.m. that assigns the above properties to the sample paths.

- We can also define $(D, \mathcal{B}_D, W)$, having the property that $W(C) = 1$ (the sample paths of a BM lie in C with probability 1.)

The sample paths have some remarkable properties:

1. The process is Gaussian (has normal attributes)! It can be proved that

$$B(t+h) - B(t) \sim \mathrm{N}(0, h) \qquad (*)$$

for $0 \leq t \leq 1 - h$ and all $h > 0$. This follows from the continuity of the sample paths, and the Central Limit Theorem.

The increment $B(t+h) - B(t)$ for $h > 0$ behaves like the sum of many small independent shocks.

2. The paths are continuous with probability.1, but nowhere smooth: (non-differentiable at every point, w.p.1). Note that (*) is equivalent to

$$\frac{B(t+h)-B(t)}{h} \sim \mathrm{N}(0,h^{-1}).$$

Letting $h \to 0$, the "derivative" of a BM is found to have infinite variance.

3. BMs have unbounded variation:

$$\sum_{j=0}^{n-1} |B((j+1)/n) - B(j/n)| \to \infty$$

as $n \to \infty$ w. p.1

Construction of BM as the limit of a discrete process:

Let $\xi_j \sim \text{iid}(0,1)$, and consider the partial sum process

$$X_T(t) = T^{-1/2} \sum_{j=1}^{[Tt]} \xi_j$$

Note that $E(X_T(t)) = 0$ and $E(X_T(t)^2) = T^{-1}[Tt]$.

Example. Let

$$\xi_j = \begin{cases} +1 & \text{with probability } \frac{1}{2} \\ -1 & \text{with probability } \frac{1}{2} \end{cases}$$

Notice that $\text{Var}(\xi_j) = 1$ and the $X_T(t)$ have standardised Binomial distributions.

$X_T \in D$, since it jumps by an amount $T^{-1/2}\xi_j$ at the points $t = j/T$, for $j = 1,2,3,\ldots$

However, the heights of the jumps are tending to 0 as T increases.

- The FCLT shows that $X_T \xrightarrow{\text{d}} B$.
- equivalently, $\mu_T \Rightarrow W$ where μ_T is the p.m. of X_T on $(D, \mathcal{B}_D)$.

In words: the sequence of p.m.s μ_T, defined on D equipped with the Skorokhod topology, converges weakly to Wiener measure.

Issues with Convergence

- Convergence in a function space implies more than the pointwise convergence of $X_T(r)$ for each $r \in [0,1]$.

- Functional convergence allows us the derive the limit distributions of quantities such as $\sup_r X_T(r)$.

- Functional convergence $X_T \xrightarrow{\text{d}} X$ requires showing:

1. The fidis converge;

2. The sequence of p.m.s is *tight*, and the limit process lies in C w.p.1.

Tightness and Stochastic Equicontinuity

A probability measure μ is *tight* if for every $\varepsilon > 0$ there exists a compact set $A_\varepsilon \in \mathcal{F}$ such $\mu(A_\varepsilon) \geq 1 - \varepsilon$.

Tightness of a sequence of distributions on a space means that the limit is a well-defined p.m. on the same space.

- Counter-example: $U[-r, r]$ is not uniformly tight on $\mathbb{R}$.
- In C and D, a non-tight distribution could exhibit arbitrary discontinuities with positive probability.
- To show that a sequence of p.m.s μ_T on $(D, \mathcal{B}_D)$ is uniformly tight as $T \to \infty$ requires establishing that the random elements are *uniformly strongly stochastically equicontinuous* functions.
- Fortunately, this property generally follows when the increment distribution has finite variance, as required in any case for CLT.

Continuous Mapping Theorem

If $X_T \xrightarrow{d} X$, and h is a measurable functional such that $h(X)$ is continuous with probability 1, then

$$h(X_T) \xrightarrow{d} h(X)$$

Analogous to Slutsky's Theorem, but for weak convergence, not convergence in prob.

Using the continuous mapping theorem we can also derive results of the following form:

$$\frac{1}{\sigma T^{3/2}} \sum_{t=1}^{T} x_t \xrightarrow{d} \int_0^1 B(r)dr \tag{1}$$

$$\frac{1}{\sigma^2 T^2} \sum_{t=1}^{T} x_t^2 \xrightarrow{d} \int_0^1 B(r)^2 dr \tag{2}$$

- Note that these are the sample means of the quantities $\dfrac{x_t}{\sigma\sqrt{T}}$ and $\dfrac{x_t^2}{T\sigma^2}$.
- They converge to random variables, not to constants!

Stochastic Integral with respect to B

Another important result:

$$\frac{1}{\sigma^2 T}\sum_{t=1}^{T} x_{t-1}\varepsilon_t \xrightarrow{\mathrm{d}} \int_0^1 BdB \tag{3}$$

The limit random variable here is called an Itô integral.

- Since $x_t^2 = x_{t-1}^2 + 2x_{t-1}\varepsilon_t + \varepsilon_t^2$, note that

$$\begin{aligned}\frac{1}{\sigma^2 T}\sum_{t=1}^{T} x_{t-1}\varepsilon_t &= \frac{1}{2\sigma^2 T}\sum_{t=1}^{T}(x_t^2 - x_{t-1}^2) - \frac{1}{2\sigma^2 T}\sum_{t=1}^{T}\varepsilon_t^2 \\ &= \frac{1}{2\sigma^2}\left(\frac{x_T^2}{T} - \frac{1}{T}\sum_{t=1}^{T}\varepsilon_t^2\right) \xrightarrow{\mathrm{d}} \frac{1}{2}(\chi^2(1) - 1).\end{aligned}$$

However, cases of the form

$$\frac{1}{\sigma_1\sigma_2 T}\sum_{t=1}^{T} x_{1,t-1}\varepsilon_{2t} \xrightarrow{\mathrm{d}} \int_0^1 B_1 dB_2$$

require more advanced arguments, based on the Itô calculus.

Tests of the Unit Root Hypothesis

Consider the regression

$$x_t = \alpha + \lambda x_{t-1} + \varepsilon_t, \quad \varepsilon_t \sim iid(0, \sigma^2).$$

To test $H_0 : \lambda = 1$, we cannot use the t test, since the standard asymptotics don't apply.

Note, this regression is equivalent to

$$\Delta x_t = \alpha + \phi x_{t-1} + \varepsilon_t$$

where $\phi = \lambda - 1$.

The null hypothesis is $\phi = 0$, and the alternative is $\phi < 0$.

Dickey-Fuller Test

The "t statistic" for this test is

$$\tau = \frac{\hat{\phi}}{\text{s.e.}(\hat{\phi})} = \frac{\sum_{t=2}^{T}(x_{t-1}-\bar{x})\varepsilon_t}{\hat{\sigma}\sqrt{\sum_{t=2}^{T}(x_{t-1}-\bar{x})^2}}$$

where $\hat{\sigma}^2 = T^{-1}\sum_{t=2}^{T}\hat{\varepsilon}_t^2$ and $\hat{\varepsilon}_t$ are the usual regression residuals, and $\hat{\sigma}^2 \to_{pr} \sigma^2$ under both null and alternative.

If H_0 is true and $\alpha = 0$, then using above results,

$$\tau = \frac{T^{-1}\sum_{t=2}^{T}x_{t-1}\varepsilon_t - (T^{-1/2}\bar{x})T^{-1/2}(x_T - x_1)}{\hat{\sigma}\sqrt{T^{-2}\sum_{t=2}^{T}x_{t-1}^2 - T^{-1}\bar{x}^2}}$$

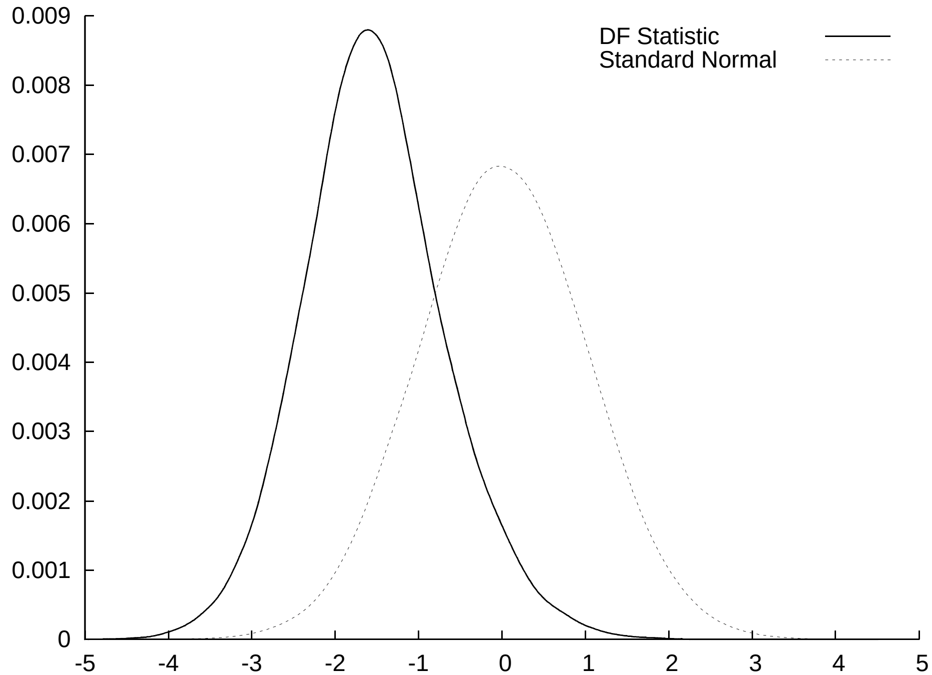
$$\xrightarrow{\mathrm{d}} \frac{B(1)^2 - 1 - 2B(1)\int_0^1 Bdr}{2\sqrt{\int_0^1 B^2 dr - \left(\int_0^1 Bdr\right)^2}}$$

This is (one variant of) the Dickey-Fuller distribution, first tabulated by Dickey and Fuller (by simulation) in 1976.

Points

- Note that this distribution does not depend on σ^2, and therefore can be tabulated.

- The Dickey-Fuller distribution is different from the *t* or normal - shifted, and skewed.



- The asymptotic 5% critical value for the *one-tailed* test of a unit root is -2.86 (compare -1.64 for the normal).

- We assumed $\alpha = 0$, but still included an intercept in the regression. The test can also be done with intercept suppressed, but then the D-F distribution is different. The formula is

$$\frac{B(1)^2 - 1}{2\sqrt{\int_0^1 B^2 dr}}$$

and the 5% critical value is -1.95.

- Tests can also be based on the statistic $\hat{\phi}$ " itself - Once again, the distributions are different – but do not depend on σ^2, and have also been tabulated by Dickey and Fuller.

Deterministic Trend

Suppose $\lambda = 1$ and $\alpha \neq 0$. Solving the model yields

$$x_t = \alpha t + \sum_{s=1}^{t} \varepsilon_s.$$

That is: the sum of a deterministic trend and a stochastic trend.

To test this null hypothesis, the following regression is run (even though $\gamma = 0$!):

$$\Delta x_t = \alpha + \phi x_{t-1} + \gamma t + \varepsilon_t.$$

The statistic $\tau = \dfrac{\hat{\phi}}{\mathrm{s.e.}(\hat{\phi})}$ in this regression has a distribution not depending on or σ^2.

- It has a different distribution, also tabulated by Dickey and Fuller.
- This test is valid whether or not $\alpha = 0$, but probably less powerful than the other test when the restriction is true.

Weak Dependence

So far, we have assumed implicitly that $\{\varepsilon_t\}$ is an i.i.d. process. In practice, the increments are allowed to be weakly dependent, just as for the ordinary CLT.

- They can be martingale differences.
- They can also be autocorrelated, provided the autocorrelation sequence is summable.
- Some additional restrictions on the dependence are generally required.
- *Many* different conditions can be cited: for example:
  - stationary and mixing, with $E|\varepsilon_t|^r < \infty$ and mixing coefficients $\alpha_m = O(m^{-r/(r-2)})$.
  - "Strong mixing" condition, allows nonstationarity...
  - "Near-epoch dependence".
- See Palgrave chapter for further details.

Autocorrelation

In the case of autocorrelation an important model parameter is the *long run variance*

$$\omega^2 = \sigma^2 + 2\gamma_1 + 2\gamma_2 + \cdots.$$

where $\gamma_j = E(\varepsilon_t \varepsilon_{t+j})$.

Note,

$$\begin{aligned} E(x_T^2) &= E\Big(\sum_{t=1}^{T} \varepsilon_t\Big)^2 \\ &= E\Big(\sum_{t=1}^{T} \varepsilon_t^2 + 2\sum_{t=1}^{T-1} \varepsilon_t \varepsilon_{t+1} + \cdots\Big) \\ &\approx T(\sigma^2 + 2\gamma_1 + 2\gamma_2 + \cdots) \\ &= T\omega^2. \end{aligned}$$

- ω^2 is proportional to the spectral density at 0.
- Defining

$$\omega^2 = \lim_{T\to\infty} T^{-1} E(x_T^2)$$

  allows the increments to be nonstationary, provided the limit exists.
- The FCLT requires $0 < \omega^2 < \infty$.

The ADF Test

Consider the practical problem: is a process ARIMA($p, 1, q$) or ARIMA($p + 1, 0, q$)?

The case $q = 0$.

If H0 is true, the model has the form

$$\Delta x_t = \alpha + \phi x_{t-1} + \beta_1 \Delta x_{t-1} + \cdots + \beta_p \Delta x_{t-p} + \varepsilon_t$$

- If $\alpha = 0$, the usual t statistic for $\hat{\phi}$ in this regression has the *standard Dickey-Fuller* distribution in large samples.
- If $\alpha \neq 0$, the t statistic for $\hat{\phi}$ in the regression

$$\Delta x_t = \alpha + \phi x_{t-1} + \beta_1 \Delta x_{t-1} + \cdots + \beta_p \Delta x_{t-p} + \gamma t + \varepsilon_t$$

has the standard Dickey-Fuller trend-adjusted distribution in large samples.

This is the *Augmented Dickey-Fuller* (ADF) test.

The case $q > 0$.

This is trickier. An invertible ARIMA($p, 1, q$), after writing in ARIMA($\infty, 1, 0$) form, can be approximated by ARIMA($p^*, 1, 0$), for some $p^* > p$.

The problem is to choose p^*. To get the correct asymptotic distribution, this needs to increase with T. $p^* = O(T^{1/3})$ has been proposed (Said and Dickey, 1984)

Nonparametric Autocorrelation Correction

When ε_t is autocorrelated, another way to see the problem with the "simple" Dickey-Fuller regression is that the variance estimate is wrong.

Another approach is to estimate ω^2 consistently from the sample autocovariances of the increments.

The *Newey-West* (1987) estimator is

$$\hat{\omega}^2 = \hat{\gamma}_0 + 2\sum_{j=1}^{M} w_j \hat{\gamma}_j$$

where $w_j = 1 - j/M$ for $j = 1,\ldots,M$. (Bartlett kernel) and $M = O(T^{1/3})$.

Phillips and Perron (1988) show that if $\alpha = 0$, the modified statistic

$$\hat{Z}_\tau = \frac{\hat{\sigma}}{\hat{\omega}}\tau + \frac{\hat{\sigma}^2 - \hat{\omega}^2}{2\hat{\omega}\sqrt{\sum_{t=2}^{T}(x_{t-1} - \bar{x})^2}}$$

has the standard Dickey-Fuller distribution in large samples, where $\hat{\omega}$ is estimated by Newey-West's method using the residual autocovariances.

- Note the two-part correction of τ - a re-scaling, plus a mean correction (second term).
- There is a similar modification for the trend-adjusted case.

Tests of I(0)

In the ADF and PP tests, I(1) represents the null hypothesis, and a rejection on the (one-sided) test is taken as evidence of I(0).

However, when making a decision to difference the data, or not, it's also desirable to test the null hypothesis of I(0).

Kwiatkowski-Phillips-Schmidt-Shin (1992) test

This is explicitly a test of I(0), with I(d) for $d > 0$ as the one-sided alternative. Let

$$S_t = \sum_{s=1}^{t} (x_s - \bar{x})$$

- the partial sum of mean deviations.

- Note, $S_T = 0$ identically.

The statistic is computed as

$$\hat{\eta}_T = \frac{1}{T^2\hat{\omega}^2} \sum_{t=1}^{T} S_t^2$$

where $\hat{\omega}^2$ is the Newey-West-type kernel variance estimator.

Asymptotic Distribution

$$\hat{\eta}_T \xrightarrow{d} \int_0^1 V(r)^2 dr$$

where $V(r) = B(r) - rB(1)$.

V is a modification of Brownian motion called a *Brownian Bridge* – similar type of process, but tied down at both ends, so that $V(1) = V(0) = 0$.

The distribution of $\hat{\eta}$ estimated by simulation by KPSS give critical values for the test, similar to PP and ADF.

Also, an extension to the deviations-from trend case. Different tabulations, also computed.

Problem:

The KPSS test effectively compares two estimates of the long-run variance:

Numerator: unconstrained estimate;

Denominator: number of non-zero autocovariances constrained by bandwidth choice

- There are no natural constraints on bandwidth choice.
- The requirement $M = O(T^{1/3})$ is compatible with any choice of $M < T$ whatever when T is finite.
- Hence you can get *any result you want* by choice of bandwidth!
- When $M = T - 1$, $\hat{\eta}_T = \frac{1}{2}$ identically. (This is in the rejection region of the KPSS limiting distribution.)

Problems with Testing I(0)

I(0) is basically the property we have also called ‘weak dependence’.

A critical property, since it legitimizes the application of the CLT to the partial sums of the process.

Essentially, this is the condition that the autocovariance sequence is summable (has finite absolute sum). Equivalently, the spectral density at zero is finite.

However: inference on $f(0)$ has been called an "ill-posed estimation problem".

- Consider the class $\mathcal{A}$ of stationary linear models, $x_t = \sum_{j=0}^{\infty} a_j u_{t-j}$ where $u_t \sim \text{iid}(0,1)$, hence defined by infinite sequences $\{a_0, a_1, a_2, \ldots\}$ satisfying $\sum_{j=0}^{\infty} a_j^2 < \infty$.
  - The sub-class $\mathcal{A}_0$ of I(0) processes (having $f(0) < \infty$) satisfies $\sum_{j=0}^{\infty} |a_j| < \infty$
  - Can show that both $\mathcal{A}_0$ and $\mathcal{A} - \mathcal{A}_0$ are dense in $\mathcal{A}$!

- Consider the IMA model $\Delta x_t = u_t - \theta_1 u_{t-1}$. This is I(1) for every $\theta_1 < 1$, but with $\theta_1 = 1$ it is i.i.d.!

Better to think of the I(0) property in relation to sample size – how close is the Gaussian approximation for the normalized sum?

- See Davidson (2009) "When is a Time series I(0)?"