Lecture Notes on Dependence[1]

Econ 624

February 1, 2021

[1] These notes are for instructional purposes only and are not to be distributed outside of the classroom.

Challenges Posed by Time Series Data

- **Question:** What are some of the challenges of working with dependent observations as opposed to independent observations?
- Recall that, in the most basic statistical framework, a random sample $X_1, X_2, ...X_n$ is assumed to be $i.i.d.\left(\mu, \sigma^2\right)$, where $0 < \sigma^2 < \infty$. We get a lot of mileage out of this simplified but restrictive setup.

(i) **Law of Large Numbers:**

$$\overline{X}_n = \frac{1}{n}\sum_{t=1}^{n} X_t \xrightarrow{p} \mu = E\left[X_t\right] \text{ as } n \rightarrow \infty.$$

Note from a statistical perspective what the law of large numbers says is that the sample mean is a consistent estimator of the population mean.

(ii) **Central Limit Theorem:**

$$\sqrt{n}\left(\overline{X}_n - \mu\right) \xrightarrow{d} N\left(0, \sigma^2\right) \text{ as } n \rightarrow \infty.$$

Challenges Posed by Time Series Data

- **Remark:** The $i.i.d.$ framework, however, is clearly not suitable for time series data since it assumes mutual independence of $X_1, X_2, ...X_n$.
- **Example:** Consider an extreme example of a dependent process

$$X_t = X \quad \text{for all } t$$

where

$$X = \begin{cases} 1 & \text{with probability } 1/2 \\ 0 & \text{with probability } 1/2 \end{cases}$$

It follows that for this process

$$\overline{X}_n = \frac{1}{n} \sum_{t=1}^{n} X_t = X$$

so we are not going to estimate $\mu = E[X_t]$ consistently. The problem is that we have too much memory or dependence in this case.

Challenges Posed by Time Series Data

- **Other challenges posed by time series:** Suppose we consider the classical (random regressor) linear regression model

$$\underset{n\times 1}{y} = \underset{n\times k}{X}\underset{k\times 1}{\beta} + \underset{n\times 1}{u}$$

where we assume that

(i) $Rank\left(X\right) = k \ (\leq n) \ a.s.$

(ii) $E\left[u|X\right] = 0 \ a.s.$

(iii) $E\left[uu'|X\right] = \sigma^2 I_n \ a.s.$

Challenges Posed by Time Series Data

- Note that, for this model, it is well known that the OLS estimator $\widehat{\beta}$ is an unbiased estimator of β since, by the usual regression algebra,

$$\widehat{\beta} = \beta + \left(X'X\right)^{-1} X'u$$

so that

$$\begin{aligned} E\left[\widehat{\beta}\right] &= \beta + E\left[\left(X'X\right)^{-1} X'u\right] \\ &= \beta + E_X\left[\left(X'X\right)^{-1} X'E\left[u|X\right]\right] \\ &= \beta. \end{aligned}$$

- **Remark:** Note also that the unbiasedness of $\widehat{\beta}$ as shown above is a finite sample property which does not involve taking n to infinity.

Challenges Posed by Time Series Data

- On the other hand, consider a dynamic time series analogue of the linear regression model

$$y = \beta y_{-1} + u$$

where

$$y = \begin{pmatrix} y_2 \\ y_3 \\ \vdots \\ y_n \end{pmatrix}, \ y_{-1} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_{n-1} \end{pmatrix}, \text{ and } u = \begin{pmatrix} u_2 \\ u_3 \\ \vdots \\ u_n \end{pmatrix}$$

Challenges Posed by Time Series Data

- Now, under the assumptions that $|\beta| < 1$ and $\{y_t\}$ is stationary so that $E\left[y_t^2\right]$ is a constant not depending t, we have

$$y_t = \sum_{j=0}^{\infty} \beta^j u_{t-j}$$

  Note that y_t here depends on current and past errors, so that the unbiasedness argument given previously for the classical linear regression does not work here. For this and other reasons, analyzing the performance of estimators for time series models is a challenging task and relies heavily on asymptotic, or large sample, analysis and approximation.

Basic Setup for Time Series and Stochastic Processes

- Let $(\Omega, \mathfrak{F}, P)$ be a probability space, and we consider a setup where each realization of a random sequence is modeled as a point in a probability space so that for $\omega \in \Omega$, we have the (infinite) sequence

$$(....X_{t-1}(\omega), X_t(\omega), X_{t+1}(\omega), ...)$$

  and for another element $\omega' \in \Omega$, we have a (possibly) different sequence

$$(....X_{t-1}(\omega'), X_t(\omega'), X_{t+1}(\omega'), ...)$$

  More formally, we have the following definition.

- **Defintion (Stochastic Process):** A stochastic process is a family (or collection) of random variables

$$x(\omega) = \{X_t(\omega) : t \in \mathbb{T}, \omega \in \Omega\}$$

  defined on a probability space $(\Omega, \mathfrak{F}, P)$. Here, $\mathbb{T}$ is called the index set.

Basic Setup for Time Series and Stochastic Processes

- Oftentimes, we take $\mathbb{T} = \mathbb{Z}$ (i.e., the set of integers) or $\mathbb{T} = \mathbb{R}$ (i.e., the set of real numbers). When $\mathbb{T}$ is countable, $\{X_t(\omega) : t \in \mathbb{T}, \omega \in \Omega\}$ is said to be a stochastic sequence. The random variable $X_t(\omega)$ at a particular point in time, say $t = \tau$, is called the τ^{th} coordinate of the process.
- **Remark:** Within this framework, we think of the available time series sample as being just a single realization of a random sequence, i.e.,

$$(....X_{t-1}(\omega), X_t(\omega), X_{t+1}(\omega), ...)$$

for a particular $\omega \in \Omega$. In fact, we only observe a finite subset of this sequence given that, in reality, the sample size of our time series is always finite.

- **Food for Thought:** Suppose in an idealized world, we are able to do repeated sampling, and suppose we do this by making *i.i.d.* draws of ω (i.e., $\omega_1, \dots, \omega_N$) at time $t = t_0$. In this setting, we would be able to construct the ensemble average

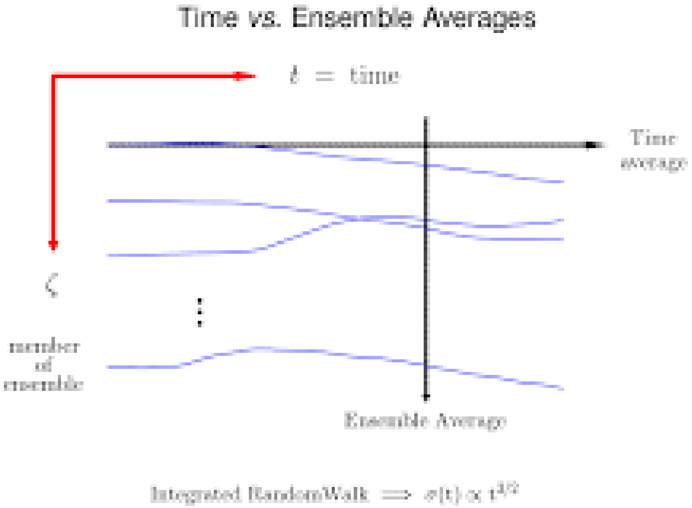
$$\overline{X}_{N,t_0} = \frac{1}{N} \sum_{j=1}^{N} X_{t_0}(\omega_j)$$

which is of course a consistent estimator of μ. On the other hand, in the case where we cannot do this and must be content with analyzing data that come from a single draw $\omega_i \in \Omega$; then, we must make do with the following time average

$$\overline{X}_n = \frac{1}{n} \sum_{t=1}^{n} X_t(\omega_i).$$

- In general, these two averages are not going to converge to the same thing with the former giving you the right answer and the latter giving you the wrong answer. However, the interesting question is whether there are circumstances under which these two averages would converge to the same limit, so that the time average also gives you the right answer. Under what conditions would this occur.

Basic Setup for Time Series and Stochastic Processes



Stationarity

- **Some Intuition:** Given that typically we only observe a finite subset $\{X_t(\omega)\}_{t=1}^n$ of a single realization of a random sequence, this would seem to suggest that, for us to be able to extract useful information from time series data, the probability distribution of X_t cannot vary too much over time. Motivated by this, the classical approach to time series modeling allows for dependence but imposes some rather strong homogeneity assumptions which have come to be known collectively as stationarity conditions.
- **Definition (Finite Dimensional Distributions):** Let $\mathcal{T}$ be the set of vectors of the form

$$\left\{\tau = (t_1, .., t_p)' : t_1 < t_2 < \cdots < t_p,\ p = 1, 2, ...\right\}$$

The finite-dimensional distribution functions of $\{X_t\}$ are given by

$$F_\tau(b) = P\left(X_{t_1} \leq b_1, X_{t_2} \leq b_2, ..., X_{t_p} \leq b_p\right)$$

for $b = (b_1, ..., b_p)' \in \mathbb{R}^p$.

- **Definition (Strict Stationarity):** A stochastic sequence $\{X_t\}_{t=-\infty-}^{\infty}$ is said to be strictly stationary if the finite-dimensional distributons are all translation invariant, i.e.,

$$\left(X_{t_1}, X_{t_2}, ..., X_{t_p}\right) \overset{d}{=} \left(X_{t_1+h}, X_{t_2+h}, ..., X_{t_p+h}\right)$$

  or, alternatively,

$$\begin{aligned} & P\left(X_{t_1} \leq b_1, X_{t_2} \leq b_2, ..., X_{t_p} \leq b_p\right) \\ = \; & P\left(X_{t_1+h} \leq b_1, X_{t_2+h} \leq b_2, ..., X_{t_p+h} \leq b_p\right) \end{aligned}$$

  for all $h \in \mathbb{Z}$, $\forall p \in \mathbb{N}$, $\forall t_1, ..., t_p \in \mathbb{Z}$, and for all $b = \left(b_1, ..., b_p\right)' \in \mathbb{R}^p$.

- Note that the assumption of strict stationarity is stronger than the assumption of identical distribution so that, in particular, the latter does not imply the former. This is illustrated in the following example.

Stationarity

- **Example:** Consider the stochastic sequence $\{X_t\}_{t=1}^{\infty}$ where

$$
\begin{aligned}
Z_1 &= (X_1, X_2)' \sim N(0, \Sigma), \\
Z_2 &= (X_3, X_4)' \sim N(0, \Sigma), \\
Z_3 &= (X_5, X_6)' \sim N(0, \Sigma), \\
&\vdots
\end{aligned}
$$

Suppose that $\{Z_s\}_{s=1}^{\infty}$ is a mutually independent sequence of 2×1 random vectors but suppose that

$$\Sigma = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}, \text{ where } \rho \neq 0.$$

Note that under these assumptions

$$X_t \sim N(0, 1) \text{ for every } t$$

so this process is identically distributed.

Stationarity

- On the other hand,

$$\begin{pmatrix} X_1 \\ X_2 \end{pmatrix} \sim N(0, \Sigma)$$

whereas

$$\begin{pmatrix} X_2 \\ X_3 \end{pmatrix} \sim N(0, I_2)$$

so $\{X_t\}_{t=1}^{\infty}$ is not strictly stationary.

- Note also that $\{X_t\}_{t=1}^{\infty}$ is not an independent sequence. An $i.i.d.$sequence would have been strictly stationary.
- **Definition (Autocovariance Function):** Let $\{X_t\}$ be a stochastic sequence such that $\sup_t E\left[X_t^2\right] < \infty$. Then, the autocovariance function is given by

$$\begin{aligned} \gamma_X(t,s) &= Cov(X_t, X_s) \\ &= E\left\{(X_t - E[X_t])(X_s - E[X_s])\right\} \end{aligned}$$

for $t, s \in \mathbb{Z}$ (or $t, s \in \mathbb{N}$).

Stationarity

- **Remark:** Note that by the Cauchy-Schwarz inequality, we have that

$$\begin{aligned} |\gamma_X(t,s)| &= |E\{(X_t - E[X_t])(X_s - E[X_s])\}| \\ &\leq \sqrt{E\left\{(X_t - E[X_t])^2\right\}}\sqrt{E\left\{(X_s - E[X_s])^2\right\}} \\ &\leq \sqrt{Var(X_t)}\sqrt{Var(X_s)} \end{aligned}$$

Now, by Jensen's inequality and Liapunov's inequality

$$|E[X_t]| \leq E|X_t| \leq \sqrt{E[X_t^2]}.$$

Moreover,

$$Var(X_t) = E\left[X_t^2\right] - (E[X_t])^2 \leq E\left[X_t^2\right].$$

Hence, if $\sup_t E\left[X_t^2\right] < \infty$; then, the mean, the variance, and the autocovariance are all well-defined for every t and s.

Stationarity

- **Defintion (Covariance Stationarity):** A stochastic process $\{X_t\}$ is said to be covariance stationary, weakly stationary, or stationary in the wide sense if

(a) $E\left[X_t^2\right] < \infty$ for all t;
(b) $E\left[X_t\right] = \mu$ for all t;
(c) $\gamma_X\left(t, s\right) = \gamma_X\left(t+k, s+k\right)$ for all s, t, and k.

- **Remark:** If $\{X_t\}$ is covariance stationary; then, setting $k = -s$, we have

$$\begin{aligned}\gamma_X\left(t, s\right) &= \gamma_X\left(t-s, 0\right) \\ &= \gamma_X\left(h, 0\right) \\ &= \gamma_X\left(h\right)\end{aligned}$$

where $h = t - s$. Hence, the autocovariance in this case depends only on "how far apart" in time X_t and X_s are and not on their locations in time.

- **Remark (con't):** In addition, note that by setting $k = -t$, we have

$$\begin{aligned}\gamma_X(h) &= \gamma_X(t, s) \\ &= \gamma_X(0, s-t) \\ &= \gamma_X(0, -h) \\ &= \gamma_X(-h)\end{aligned}$$

where $-h = s - t$, so that

$$\gamma_X(h) = \gamma_X(-h)$$

with

$$\gamma_X(0) = Var(X_t) = E\left\{(X_t - E[X_t])^2\right\}.$$

Stationarity

- Similarly, under covariance stationarity, the **autocorrelation coefficient** is

$$Corr\left(X_t, X_s\right) = \frac{Cov\left(X_t, X_s\right)}{\sqrt{Var\left(X_t\right)}\sqrt{Var\left(X_s\right)}} = \frac{\gamma_X\left(h\right)}{\sqrt{\gamma_X\left(0\right)}\sqrt{\gamma_X\left(0\right)}} = \frac{\gamma_X\left(h\right)}{\gamma_X\left(0\right)}$$

  where again $h = t - s$.

- **Definition (White Noise):** A process $\{u_t : t \in \mathbb{Z}\}$ is said to be **white noise** with mean zero and variance σ_u^2 and denoted

$$\{u_t\} \sim WN\left(0, \sigma_u^2\right),$$

  if and only if

$$\begin{aligned} E\left[u_t\right] &= 0 \text{ for all } t \in \mathbb{Z} \\ \gamma_X\left(t, s\right) &= E\left[u_t u_s\right] = \begin{cases} \sigma_u^2 & \text{for } t = s \\ 0 & \text{for } t \neq s \end{cases} \end{aligned}$$

Stationarity

- **Remarks:**

(i) Note that a white noise process is not necessarily $i.i.d.$ since we are only making assumptions about the first two moments. It is of course covariance stationary.

(ii) A strictly stationary stochastic sequence $\{X_t\}$ such that $E\left[X_t^2\right] < \infty$ is also covariance stationary, i.e., subject to a second moment condition, strict stationary implies covariance stationarity.

- **Remarks (con't):**

(iii) In general, covariance stationarity does not imply strict stationarity. An exception is a Gaussian stochastic sequence. A stochastic sequence $\{X_t\}$ is Gaussian if all finite-dimensional distributions are multivariate Gaussian (i.e., multivariate normal). A Gaussian, covariance stationary stochastic sequence is also strictly stationary since in this case $\left(X_{t_1}, X_{t_2}, ..., X_{t_p}\right)$ and $\left(X_{t_1+h}, X_{t_2+h}, ..., X_{t_p+h}\right)$ would have the same mean vector and the same covariance matrix $\forall h \in \mathbb{Z}$, $\forall p \in \mathbb{N}$, and $\forall t_1, ..., t_p \in \mathbb{Z}$. Since a multivariate normal distribution is completely determined by its mean vector and covariance matrix (i.e., it is completely determined by the first two moments), it follows that $\left(X_{t_1}, X_{t_2}, ..., X_{t_p}\right)$ and $\left(X_{t_1+h}, X_{t_2+h}, ..., X_{t_p+h}\right)$ would have the same distribution $\forall h \in \mathbb{Z}$, $\forall p \in \mathbb{N}$, and $\forall t_1, ..., t_p \in \mathbb{Z}$.

Shift Transformation

- Consider a $1-1$ measurable mapping

$$T : \Omega \rightarrow \Omega \text{ (onto)}$$

This a rule for pairing (or mapping) each outcome ω with another outcome $\omega' = T(\omega)$ of the space Ω, but since each $\omega \in \Omega$ maps into an infinite sequence, T induces a mapping from one sequence to another.

- **Definition:** T is called **measure-preserving** if

$$P(TE) = P(E) \text{ for all } E \in \mathfrak{F}.$$

- **Shift Transformation:** T is a (back) shift operator if

$$X_t(T\omega) = X_{t+1}(\omega)$$

for a sequence $\{X_t(\omega)\}_{t=-\infty}^{\infty}$. Note that a shift operator T takes each outcome ω and maps it into another outcome whereby the realized value of X which had previously occurred in period $t+1$ under outcome ω now occurs in period t, for every t.

Shift Transformation

- **Illustration:** Suppose that

$$\begin{aligned} \omega &= (\dots, x_1, x_2, x_3, \dots), \\ T\omega &= (\dots, x_2, x_3, x_4, \dots), \\ T^2\omega &= (\dots, x_3, x_4, x_5, \dots). \end{aligned}$$

Let

$$X_1(\omega) = x_1, X_2(\omega) = x_2, X_3(\omega) = x_3.$$

In this case,

$$\begin{aligned} X_1(T\omega) &= X_2(\omega) = x_2, \\ X_2(T\omega) &= X_3(\omega) = x_3, \\ X_3(T\omega) &= X_4(\omega) = x_4 \end{aligned}$$

- **Illustration (con't):**
  and

$$\begin{aligned} X_1\left(T^2\omega\right) &= X_3\left(\omega\right) = x_3, \\ X_2\left(T^2\omega\right) &= X_4\left(\omega\right) = x_4, \\ X_3\left(T^2\omega\right) &= X_5\left(\omega\right) = x_5. \end{aligned}$$

- Note that our previous definition of strict stationarity corresponds to the case where the shift transformation is measure preserving since strict stationarity means that

$$\begin{aligned} & P\left(X_{t_1}\left(E\right) \leq x_1, X_{t_2}\left(E\right) \leq x_2, ..., X_{t_p}\left(E\right) \leq x_p\right) \\ = \; & P\left(X_{t_1}\left(T^h E\right) \leq x_1, X_{t_2}\left(T^h E\right) \leq x_2, ..., X_{t_p}\left(T^h E\right) \leq x_p\right) \\ = \; & P\left(X_{t_1+h}\left(E\right) \leq x_1, X_{t_2+h}\left(E\right) \leq x_2, ..., X_{t_p+h}\left(E\right) \leq x_p\right) \end{aligned}$$

for all $E \in \mathfrak{F}$, $h \in \mathbb{Z}$, $p \in \mathbb{N}$ as well as $\forall t_1, ..., t_p \in \mathbb{Z}$, and also for all $(x_1, ..., x_p)' \in \mathbb{R}^p$.

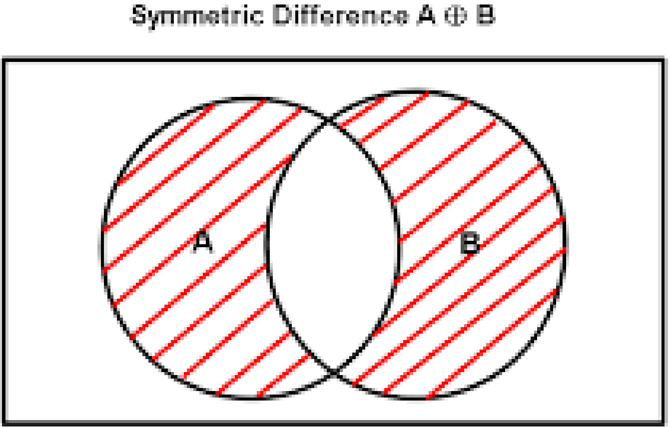
Shift Transformation

- **Definition of Invariant Event:** $E \in \mathfrak{F}$ is said to be **invariant** under a transformation T if $TE = E$.
- **Alternative Definition of Invariant Event:** $E \in \mathfrak{F}$ is said to be **invariant** under a transformation T if

$$P\left(TE \oplus E\right) = 0$$

  where $\oplus$ denotes symmetric difference, i.e., $A \oplus B = A \cup B \backslash \left(A \cap B\right)$.

- **Symmetric Difference:**



Shift Transformation

- Note that if $\omega \in \mathfrak{F}$ is invariant under the shift transformation, then $T\omega = \omega$ and

$$\begin{aligned} X_t(\omega) &= X_t(T\omega) = X_{t+1}(\omega), \\ X_{t+1}(\omega) &= X_{t+1}(T\omega) = X_{t+2}(\omega), \\ &\vdots \end{aligned}$$

Ergodicity

- **Definition (Ergodicity):** A strictly stationary sequence $\{X_t(\omega)\}$ is ergodic if the probability of every invariant event is either 0 or 1.
- **Remarks:**
  - (i) Note that events that are invariant under ergodic transformation must either occur almost surely or do not occur almost surely.
  - (ii) Absence of ergodicity means that there exists at least one invariant event E for which
  $$0 < P(E) < 1.$$
  If this is the event which actually occurs, then the realized time series will not be informative about the other possible realized values of X_t.
- **Ergodic Theorem:** Let $\{X_t(\omega)\}_{t=-\infty}^{\infty}$ be a strictly stationary and ergodic stochastic sequence such that $E|X_t| < \infty$. Then,
$$\lim_{n\to\infty} \frac{1}{n}\sum_{t=1}^{n} X_t(\omega) = E[X_t] \ \ a.s.$$
- **Remark:** Note that the ergodic theorem is in essence a strong law of large numbers for dependent processes.

Ergodicity

- **Example 1:** Consider a simple case where Ω only has two sequences. More specifically,

$$\Omega = \{\omega_1, \omega_2\},$$

where

$$\begin{aligned} \omega_1 &= (\ldots, 1, 1, 1, \ldots) \\ \omega_2 &= (\ldots, 0, 0, 0, \ldots) \end{aligned}$$

Clearly, both of these are invariant events under the shift operator T. Suppose further that

$$\Pr(\omega = \omega_1) = p \text{ where } 0 < p < 1$$

from which it follows that

$$\Pr(\omega = \omega_2) = 1 - p = q \in (0, 1)$$

Ergodicity

- **Example 1 (con't):** In this case, define the stochastic sequence $\{X_t(\omega)\}_{t=-\infty}^{\infty}$ to be

$$\begin{aligned} X_t(\omega_1) &= 1 \text{ for all } t \in \mathbb{Z}, \\ X_t(\omega_2) &= 0 \text{ for all } t \in \mathbb{Z}. \end{aligned}$$

Note that this stochastic sequence is non-ergodic since there exists invariant events whose probability is not 0 or 1.

- **Example 2:** Consider instead the case where Ω only has one sequence. More specifically, let

$$\Omega = \{\omega_*\}$$

where

$$\omega_* = (..., 1, 1, 1, ...)$$

so that, again, ω_* is an invariant event under the shift operator T. Suppose that

$$\Pr(\omega = \omega_*) = 1$$

and define $\{X_t(\omega)\}_{t=-\infty}^{\infty}$ to be the "trivial" process

$$X_t = X_t(\omega_*) = 1 \text{ for all } t \in \mathbb{Z}.$$

Note that $\{X_t(\omega)\}_{t=-\infty}^{\infty}$ is ergodic in this case since there is only one event, and it is an invariant event, but its probability is 1.

Ergodicity

- **Example 3:** Let X_t be generated by the equation

$$X_t = U_t + Z \quad \text{for } t \in \mathbb{Z},$$

where

$$\{U_t\}_{t=-\infty}^{\infty} \equiv i.i.d.\text{Uniform } [0,1] \text{ and } Z \sim N(0,1)$$

and U_t and Z are independent for all $t \in \mathbb{Z}$. Now, take

$$\begin{aligned} E &= \bigcap_{t=-\infty}^{\infty} \{\omega \in \Omega : X_t(\omega) < 0\} \\ &= \{\omega \in \Omega : ..., X_{t-1}(\omega) < 0, X_t(\omega) < 0, X_{t+1}(\omega) < 0, ...\} \end{aligned}$$

Clearly, E is invariant since

$$\begin{aligned} & TE \\ = \; & \{\omega \in \Omega : ..., X_{t-1}(T\omega) < 0, X_t(T\omega) < 0, X_{t+1}(T\omega) < 0, ...\} \\ = \; & \{\omega \in \Omega : ..., X_t(\omega) < 0, X_{t+1}(\omega) < 0, X_{t+2}(\omega) < 0, ...\} \end{aligned}$$

- **Example 3 (con't):** However, E occurs if and only if $\{Z < -1\}$; hence,

$$P(E) = \Phi(-1) = P(Z < -1)$$

so that, in this case, $0 < P(E) < 1$, and $\{X_t\}$ is not ergodic. One can, in fact, show that

$$\overline{X}_n = \frac{1}{n}\sum_{t=1}^{n} X_t = \overline{U}_n + Z \overset{a.s.}{\rightarrow} E[U_t] + Z = \frac{1}{2} + Z.$$

Another way to understand this problem is to calculate the autocovariance

$$\begin{aligned} Cov(X_t, X_{t+h}) &= Cov(U_t + Z, U_{t+h} + Z) \\ &= Var(Z) \\ &= 1 \end{aligned}$$

for all $h \in \mathbb{Z} \setminus \{0\}$, so there is simply too much dependence in this case.