An Imprecise Introduction to Stochastic Processes

Stelios Arvanitis

2018

Acknowledgement. I am grateful to Anastasia Chatzilena that helped with the typesetting of the present notes.

Part II

1 Stationarity and Birkhoff's LLN

The two stationarity notions do not suffice for the establishment of LLN's with degenerate limits. Further restrictions are needed on the dependence between the elements of the process as structured in the fidis. There are several ways to proceed in order to handle this structure. Given the non-rigorous nature of the notes the most convenient way to proceed (w.r.t. the simplicity of verification in what follows) is by the utilization of the notion of ergodicity.

In order to proceed in this fashion, we briefly examine the asymptotic behavior of arithmetic means in the framework of strict stationarity. Suppose thus that $x = (x_t)_{t\in\mathbb{Z}}$ is a stationary process. The following definition concerns an information set constructed by the measurable subsets of the set of sample paths of the process which have a particular connection to the action of the $L_\star$ operator.

Definition 1. The algebra of invariant events $\mathcal{J}_x$ of x is the one generated by every measurable subset of the set of sample paths of x that remains invariant w.r.t. $L_\star$, i.e. every such A, for which $L_\star(A) \subseteq A$.

Intuitively $\mathcal{J}_x$ can be perceived as equivalent to the informational content of the set of random processes *essentially* constructed by the elements of the sets of invariant events. It can be chosen to be "large enough" so as not to depend on the particular process x but merely on the σ-algebra with which its range is endowed with. $\mathcal{J}_x$ is trivial iff every event in it has either zero or one probability (w.r.t. $\mathbb{P}$). This is equivalent to that every one of the aforementioned random processes is essentially constant. By the definition of the conditional expectation if $\mathbb{E}(|x_0|) < +\infty$, then $\mathbb{E}(x_0/\mathcal{J}_x)$ is a well defined random variable, with expectation equal to $\mathbb{E}(x_0)$ due to the L.I.E., and $\mathbb{E}(x_0/\mathcal{J}_x)$ is constant iff

$\mathcal{J}_x$ is trivial. The following LLN may therefore specify convergence to a non-constant (i.e. stochastic) limit.

Lemma 1. *[Birkhoff's LLN-Stationarity] If $x = (x_t)_{t\in\mathbb{Z}}$ is stationary and $\mathbb{E}(|x_0| < +\infty)$ then $\frac{1}{T}\sum_{t=1}^{T} x_t \to \mathbb{E}(x_0/\mathcal{J}_x)$, $\mathbb{P}$ a.s.*

Proof. Out of the scope of the course. □

Hence the arithmetic mean may converge to a random variable, and this is precisely the case when (and only then) $\mathcal{J}_x$ is non trivial.

Example 1. Suppose that $\varepsilon = (\varepsilon_t)_{t\in\mathbb{Z}}$ is a stationary white noise process (WN(σ^2)) and let the causal $x = (x_t)_{t\in\mathbb{Z}}$ be defined by $x_t = \sum_{j=0}^{\infty} \beta_j \varepsilon_{t-j}$ for $(\beta_j)_{j\in\mathbb{N}}$ absolutely summable. Suppose that for some $b_0 \in \mathbb{R}$, $y = (y_t)_{t\in\mathbb{Z}}$ be defined by $y_t = b_0 x_{t-1} + \varepsilon_t$. Given the results in the paragraph Transformations and Hereditarity in Part I, and the ones about Causal Linear Processes we have that $(x_{t-1}^2)_{t\in\mathbb{Z}}$, $(\varepsilon_t x_{t-1})_{t\in\mathbb{Z}}$ are stationary, and that (why?) $\mathbb{E}(\varepsilon_0 x_{-1}) = 0$. Given a sample $(y_t, x_{t-1})_{t=1,\ldots,T}$ the OLSE for b_0 is $b_T = \frac{\sum_{t=1}^{T} y_t x_{t-1}}{\sum_{t=1}^{T} x_{t-1}^2} = b_0 + \frac{\frac{1}{T}\sum_{t=1}^{T} \varepsilon_t x_{t-1}}{\frac{1}{T}\sum_{t=1}^{T} x_{t-1}^2}$. As noted above when $\mathcal{J}_\varepsilon$ is chosen to be "large enough" then it can be shown to coincide with the algebra of invariant events of $(\varepsilon_t x_{t-1})_{t\in\mathbb{Z}}$, and the one of $(x_{t-1})_{t\in\mathbb{Z}}$. If $\mathbb{E}(x_{-1}^2/\mathcal{J}_\varepsilon) > 0$ a.s., due to the previous LLN and the CMT (continuous mapping theorem),

$$b_T \to b_0 + \frac{\mathbb{E}(\varepsilon_0 x_{-1}/\mathcal{J}_\varepsilon)}{\mathbb{E}(x_{-1}^2/\mathcal{J}_\varepsilon)},\ \mathbb{P} \text{ a.s.}$$

Then b_T would be (strongly) consistent iff $\mathcal{J}_\varepsilon$ is trivial.

2 Ergodicity

Given the heuristic description of $\mathcal{J}_x$ the precise understanding of the concept of ergodicity requires notions from dynamical systems and measure theory that lie beyond the scope of our lectures. However, the following definition is precise.

Definition 2. A stationary process $x = (x_t)_{t\in\mathbb{Z}}$ is ergodic (w.r.t. $L_\star$) iff $\mathcal{J}_x$ is trivial.

Example 2. If x is i.i.d. then it is stationary, ergodic. This is essentially due to Kolmogorov's zero-one law.

Example 3. When independence is retained but heterogeneity is introduced then x can not be ergodic, since it can not be stationary.

Example 4. There exist stationary processes, that are not ergodic. For an example consider $\varepsilon = (\varepsilon_t)_{t\in\mathbb{Z}}$ be i.i.d. and $\varepsilon_0 \sim$ Unif$_{[-1,1]}$. Suppose that $z \sim N(0,1)$ and z is independent of ε_t for all $t \in \mathbb{Z}$. Consider $x = (x_t)_{t\in\mathbb{Z}}$ with $x_t = \varepsilon_t + z$. The latter is stationary since for any $\kappa \in \mathbb{N}$ and any $m \in \mathbb{Z}$, due to the L.I.E. (Law of Iterated Expectations),

$$\mathbb{P}\left(x_{t_1} \in \cdot, x_{t_2} \in \cdot, \dots, x_{t_{1\kappa}} \in \cdot\right) \overset{\text{L.I.E.}}{=} \mathbb{E}\left(\mathbb{P}\left(x_{t_1} \in \cdot, x_{t_2} \in \cdot, \dots, x_{t_{1\kappa}} \in \cdot\right)/\sigma(z)\right)$$

$$\overset{\text{ind.}}{=} \mathbb{E}\left(\prod_{i=1}^{\kappa}\mathbb{P}\left(x_{t_i} \in \cdot\right)/\sigma(z)\right) \overset{\text{ident.}}{=} \mathbb{E}\left(\prod_{i=1}^{\kappa}\mathbb{P}\left(x_{t_i+m} \in \cdot\right)/\sigma(z)\right)$$

$$\overset{\text{ind.}}{=} \mathbb{E}\left(\mathbb{P}\left(x_{t_1+m} \in \cdot, x_{t_2+m} \in \cdot, \dots, x_{t_{1\kappa}+m} \in \cdot\right)/\sigma(z)\right)$$

$$\overset{\text{L.I.E.}}{=} \mathbb{P}\left(x_{t_1+m} \in \cdot, x_{t_2+m} \in \cdot, \dots, x_{t_{1\kappa}+m} \in \cdot\right).$$

For ergodicity consider $A \triangleq \{x_t \leq -1, t \in \mathbb{Z}\}$. Then

$$L_\star(A) = \{Lx_t \leq -1, t \in \mathbb{Z}\} = \{x_{t-1} \leq -1, t \in \mathbb{Z}\} = A.$$

Hence A is invariant. But,

$$\mathbb{P}(A) = \mathbb{P}(\{\varepsilon_t + z \leq -1, t \in \mathbb{Z}\}) = \mathbb{P}\left(\max_{t\in\mathbb{Z}}[\varepsilon_t + z] \leq -1\right)$$

$$\geq P(z \leq -2) = \Phi(-2) > 0,$$

where Φ denotes the cdf of the standard normal distribution. Furthermore,

$$\mathbb{P}(A) = \mathbb{P}(\{\varepsilon_t + z \leq -1, t \in \mathbb{Z}\}) \leq \mathbb{P}\left(\min_{t\in\mathbb{Z}}[\varepsilon_t + z] \leq -1\right)$$

$$\leq P(z \leq 0) = \Phi(0) = \frac{1}{2},$$

thus we obtain that for the invariant event A,

$$0 < \mathbb{P}(A) \leq \frac{1}{2},$$

hence x cannot be ergodic.

Remark 1. It can be proven that ergodicity is somewhat asymptotic independence on average between the elements of the process, i.e. $\lim_{k\to\infty}\frac{1}{k}\sum_{j=1}^{k}\text{Cov}\left(f(x_0), g(x_{-j})\right) = 0$, for any measurable real functions f, g for which the covariance exists.

3 Ergodicity: Transformations and Hereditarity

As with the concepts of stationarity we examine the hereditarity of ergodicity w.r.t. analogous transformations. It is possible to prove the following result.

Lemma 2. *If in Lemmata 1-3 (with 1 appropriately modified) in Part I the initial processes are assumed not only strictly stationary but also ergodic then the resulting processes are also strictly stationary and ergodic.*

The following result concerns the existence of "unique" stationary and ergodic solutions of systems of recurrence relations, termed stochastic recurrence equations-**SRE**s (or stochastic difference equations). Those systems define new processes as solutions of those recursions, given initial ones that structure the relations. They constitute the second major method of obtaining a stochastic process that we will use in the course, apart from the linear transformations that resulted in the category of causal linear processes that we have already begun to examine. In some cases the two methods may coincide.

Lemma 3. *Suppose that* $\varepsilon = (\varepsilon_t)_{t\in\mathbb{Z}}$ *is stationary, ergodic. Suppose that* $f : \mathbb{R}^2 \to \mathbb{R}$ *is appropriately measurable and for any* $z \in \mathbb{R}$ $f(\cdot, z)$ *is continuously differentiable w.r.t. the first argument. Suppose that certain conditions hold (2) and that* $\mathbb{E}\left(\ln\left(\sup_{x\in\mathbb{R}}\left|\frac{\partial f}{\partial x}(x,\varepsilon_0)\right|\right)\right) < 0$ *(or* $-\infty$*). Then there exists a "unique" process* $x = (x_t)_{t\in\mathbb{Z}}$ *satisfying*

$$x_t = f(x_{t-1}, \varepsilon_t), \ t \in \mathbb{Z},$$

defined as

$$x_t = \lim_{m\to\infty} \underbrace{f \circ f \circ f \circ \ldots \circ f}_{m}(y, \varepsilon_t, \ldots, \varepsilon_{t-m}), \ \mathbb{P} \text{ a.s.},$$

for some $y \in \mathbb{R}$*, that is strictly stationary and ergodic.*

Remark 2. Those conditions involve the existence of positive logarithmic moments for ε_0 and the existence of $y \in \mathbb{R}$ such that the positive logarithmic moment of the distance between y and $f(y, \varepsilon_0)$ exists. We do not further specify them since they anyhow hold in all cases that we will be using the result and for all relevant y. Notice that this lemma can be further generalized w.r.t. the form and properties of f. It is essentially based on the Banach Fixed Point Theorem. A generalization based on a more general fixed point theorem, which is obviously quite outside the scope of the course, can be found in https://doi.org/10.1080/23311835.2017.1380392

Example 5. AR(1) recursion. Suppose that $\varepsilon = (\varepsilon_t)_{t\in\mathbb{Z}}$ is a stationary and ergodic white noise ($\sim$ WN (σ^2)). Let $f(x_1, x_2) = \beta x_1 + x_2$. Consider the recursion $x_t = f(x_{t-1}, \varepsilon_t) = b_0 x_{t-1} + \varepsilon_t$, $t \in \mathbb{Z}$. Obviously $f(\cdot, z)$ is continuously differentiable w.r.t. $\cdot$ for all z. Then $\frac{\partial f(x,z)}{\partial x} = b_0$ and $\ln\left(\sup_x\left|\frac{\partial f(x,\varepsilon_0)}{\partial x}\right|\right) = \ln|b_0| = E\left(\ln\left(\sup_{x\in\mathbb{R}}\left|\frac{\partial f(x,e_0)}{\partial x}\right|\right)\right) < 0$ (or $-\infty$) iff

$|b_0| < 1$. Hence, the previous condition implies along with Lemma 3 that the recursion defined above admits a "unique" stationary and ergodic solution $x = (x_t)_{t\in\mathbb{Z}}$ defined by the limiting arguments $x_t = \lim_{m\to\infty} \underbrace{f \circ f \circ f \circ \ldots \circ f}_{m} (y, \varepsilon_t, \ldots, \varepsilon_{t-m}) = \lim_{m\to\infty} b_0^m y + \lim_{m\to\infty} \sum_{i=0}^{m} b_0^i \varepsilon_{t-i} = \sum_{i=0}^{\infty} b_0^i \varepsilon_{t-i}$ $\mathbb{P}$ a.s. Thereby, when $|\beta| < 1$ the linear process x defined on $(\varepsilon_t)_{t\in\mathbb{Z}}$, $(\beta^i)_{i\in\mathbb{N}}$ is the unique stationary ergodic solution of the recursion $x_t = b_0 x_{t-1} + \varepsilon_t$ termed as the AR(1) recursion.

4 Birkhoff's Law of Large Numbers

Ergodicity implies a simple yet powerful corollary of Doob's LLN, termed as Birkhoff's LLN.

Lemma 4. *(Birkhoff's LLN) Suppose that $x = (x_t)_{t\in\mathbb{Z}}$ is stationary ergodic, and that $\mathbb{E}(|x_0|) < +\infty$ exists as a real number. Then,*

$$\lim_{T\to\infty} \frac{1}{T} \sum_{t=1}^{T} x_t \to \mathbb{E}(x_0),\ \mathbb{P} \textit{ a.s.}$$

Obviously this extends Kolmogorov's LLN and suffices for consistency of the OLSE in frameworks such as the one in Example 1 as long as ergodicity is added.

Example 6. Consider the framework of Example 1, suppose that ε is also ergodic, and let $x_t = y_{t-1}$, for all $t \in \mathbb{Z}$, and $\beta_i = b_0^i$, $\forall i \in \mathbb{Z}$. Hence we obtain a regression model that corresponds to the linear AR(1) recursion in Example 5. Since now $b_T = \frac{\sum_{t=1}^{T} y_t y_{t-1}}{\sum_{t=1}^{T} y_{t-1}^2}$, due to Lemmata 2, 3, 4, and the results on the causal linear processes, we have that

$$\frac{1}{T} \sum_{t=1}^{T} y_t y_{t-1} \to \mathbb{E}\left(y_0 y_{-1}\right) = b_0 \frac{\sigma^2}{1 - b_0^2},\ \mathbb{P} \text{ a.s.,}$$

$$\frac{1}{T} \sum_{t=1}^{T} y_{t-1}^2 \to \mathbb{E}\left(y_{-1}^2\right) = \frac{\sigma^2}{1 - b_0^2} > 0,\ \mathbb{P} \text{ a.s.,}$$

and thereby due to the CMT (why is it applicable?) $b_n \to b_0$, $\mathbb{P}$ a.s., hence, the OLSE is strongly consistent. Question, is the OLSE unbiased? (hint: if $X = (y_0, y_1, \ldots, y_{T-1})^{\prime}$ and $\epsilon = (\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_T)^{\prime}$, does $\mathbb{E}\left(\epsilon / \sigma\left(X\right)\right) = 0_{T\times 1}$ hold in this case?).

Remark 3. Stationarity and ergodicity are not sufficient for the establishment of the asymptotic behavior of $\frac{1}{\sqrt{T}} \sum_{t=1}^{T} (x_t - \mathbb{E}(x_0))$ given the on average asymptotic independence interpretation of the latter. Conditions that ensure that Cov $\left(f\left(x_t\right), g\left(x_{t-k}\right)\right)$ converges to zero appropriately fast as $k \to \infty$, along with moment existence conditions for the elements of the process can suffice. The former are related to relevant notions of *mixing* (with appropriate rates) between the elements of the process.

The notes are in state of "perpetual" correction. They do not substitute the lectures. Please report any typos to stelios@aueb.gr on the course's e-class.