1 Stochastic Processes

A brief treatment of stochastic processes is provided, focusing on time series frameworks and stationary environments. A stochastic process is a random element having a distribution describable via a collection of joint distributions of random vectors. The formulation of a temporal dependence framework between random variables is briefly investigated by exploring the concepts of strict stationarity, ergodicity, and strong mixing.

It should be noted that this treatment is far from exhaustive; instead, it serves as a very introductory paragraph for interested readers who wish to explore the rich literature in this field. Some indicative references include [3], and [5].

2 Definition

Stochastic processes are collections of random variables indexed by elements of a parameter set, satisfying specific compatibility conditions. Alternatively, they can be seen as random elements that take values in spaces of functions.

Definition 1. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, Θ a non-empty set, and $(\mathbb{R}, \mathcal{B})$ the Borel space of real numbers. A stochastic process with values in $\mathbb{R}$ indexed by Θ is a collection of random variables $\{X_\theta : \Omega \to \mathbb{R}, \theta \in \Theta\}$ such that a unique probability measure is well-defined on the set of functions

$$\mathbb{R}^\Theta = \{f : \Theta \to \mathbb{R}\},$$

equipped with the smallest σ-algebra containing the elements of the product $\underbrace{\mathcal{B} \times \mathcal{B} \times \cdots}_{\text{indexed by } \Theta}$.

It is important to note that some authors adopt a broader approach and do not require the process to be a random element. Instead, they define it simply as a collection of random variables indexed uniquely by the parameter θ (see, for instance, [8]). In this generalized perspective, X can be thought of as a "bivariate" function $X : \Omega \times \Theta \to \mathbb{R}$ that satisfies the requirement that $X(\cdot, \theta)$ is measurable for each $\theta \in \Theta$. This alone, however, is insufficient to ensure the second condition of the definition, namely, the transfer of $\mathbb{P}$ via X onto $\mathbb{R}^\Theta$.

The latter depends on how the collection of functions can be constructed as a measurable space. A general method is through the topology of pointwise convergence: $f_n \to f$ if and only if $f_n(\theta) \to f(\theta)$ for all $\theta \in \Theta$, where the limits on the right-hand side are standard limits of real sequences. Using this construction, one immediately defines closed sets[1] and subsequently open sets[2] in the space. Consequently, the associated Borel σ-algebra is derived, allowing the space to become measurable. It is shown that this Borel σ-algebra coincides with the one referenced in the definition.

3 The Daniell-Kolmogorov Theorem

The definition of stochastic processes can be generalized so that the process takes values in a more general metric space S. In the general case, a

[1] $A \subset \mathbb{R}^\Theta$ is closed under pointwise convergence if and only if for every $f_n \in A$, $f_n \to f$ implies $f \in A$, i.e., the set contains all limits of converging sequences.

[2] $A \subset \mathbb{R}^\Theta$ is open under pointwise convergence if and only if its complement is closed.

stochastic process is a random element with values in the set of functions $S^{\Theta} \triangleq \{f : \Theta \to S\}$, equipped with an appropriate σ-algebra. In certain cases, it is useful to study processes taking values in a suitable "subspace" of S^{Θ}, such as the space of continuous functions from Θ to S when S and Θ have suitable topological structures. Note that this generalization is relatively straightforward when $S = \mathbb{R}^n$, $n \geq 1$, but the technical details become more complex for general S.

The definition implies that for all $\omega \in \Omega$, $X(\omega) = f_\omega : \Theta \to \mathbb{R}$. Such functions f_ω are called "sample paths" of the process.

The definition highlights the mapping of $\mathbb{P}$ into a probability distribution over $\mathbb{R}^{\Theta}$, which is inherently a complex object. Part of stochastic process theory involves representing such objects through more "familiar" notions. In this context, the Daniell-Kolmogorov theorem-see, for instance, [1] and [7]-implies that under specific weak consistency conditions regarding the behavior of finite subsets of the process, the distribution over $\mathbb{R}^{\Theta}$ is fully characterized by the collection of "finite-dimensional distributions (FiDis)" of the process.

Definition 2. The collection of FiDis of X is:

$$\left\{\mathbb{P}_{X_{\theta_1}, X_{\theta_2}, \ldots, X_{\theta_\kappa}},\ \theta_1, \theta_2, \ldots, \theta_\kappa \in \Theta^\star,\ \Theta^\star \subseteq \Theta,\ \sharp\Theta^\star = \kappa,\ \kappa \in \mathbb{N}^\star\right\},$$

where

$$\mathbb{P}_{X_{\theta_1}, X_{\theta_2}, \ldots, X_{\theta_\kappa}} \triangleq \mathbb{P}(X_{\theta_1} \in \cdot, X_{\theta_2} \in \cdot, \ldots, X_{\theta_\kappa} \in \cdot)$$

is the joint distribution of the random vector $(X_{\theta_1}, X_{\theta_2}, \ldots, X_{\theta_\kappa})$.

Given the above definition, the theorem states:

Theorem 1 (Daniell-Kolmogorov)**.** *If the collection of FiDis of X satisfies the following two consistency conditions:*

For every $\kappa \in \mathbb{N}^\star$, every $\Theta^\star \subseteq \Theta$ with $\sharp\Theta^\star = \kappa$, and any Borel sets $A_1, A_2, \ldots, A_\kappa \subseteq \mathbb{R}$,

1. *For every permutation π of the elements of $\Theta^\star$:*

$$\mathbb{P}_{X_{\pi(\theta_1)},X_{\pi(\theta_2)},\ldots,X_{\pi(\theta_\kappa)}}(A_{\pi(1)} \times \cdots \times A_{\pi(\kappa)}) = \mathbb{P}_{X_{\theta_1},X_{\theta_2},\ldots,X_{\theta_\kappa}}(A_1 \times \cdots \times A_\kappa), \quad (1)$$

2. *For every $0 < m \leq \kappa$:*

$$\mathbb{P}_{X_{\theta_1},\ldots,X_{\theta_m}}(A_1 \times \cdots \times A_m) = \mathbb{P}_{X_{\theta_1},\ldots,X_{\theta_\kappa}}(A_1 \times \cdots \times A_m \times \underbrace{\mathbb{R} \times \cdots \times \mathbb{R}}_{\kappa - m \text{ times}}), \quad (2)$$

then there exists a unique Borel distribution on $\mathbb{R}^\Theta$ consistent with the given FiDis.

The first consistency condition requires that rearranging the components of any random vector from the process correspondingly changes the joint distribution. The second condition requires that joint distributions of lower-dimensional vectors can be obtained by integrating out surplus components from higher-dimensional ones. The theorem indicates that these consistency conditions (1)-(2) are sufficient to define a valid distribution over the complex object X in $\mathbb{R}^\Theta$ and that the collection of FiDis fully characterizes the distribution of such processes over $\mathbb{R}^\Theta$. Consequently, the theorem allows for describing probability distributions in function spaces using the joint distributions of finite-dimensional random vectors. The proof is beyond the scope of this book, and the interested reader is referred

to [7]. The theorem's validity extends to processes taking values in metric spaces.

Later we will see that under weak conditions, objective functions in the theory of M-estimation for (semi-)parametric statistical models—e.g., sum of squares criteria, likelihood functions, quadratic criteria of the GMM type, etc.—are instances of stochastic processes defined over the corresponding parameter spaces.

Example 1. When $\Theta = \{1\}$, we recover the notion of a random variable, while when $\Theta = \{1, 2, \ldots, n\}$ (as an ordered set), we recover the concept of an n-dimensional random vector. (Exercise: Describe the set of FiDis!)

A "direct" application of the theorem essentially extends the definition of the normal distribution to function spaces:

Definition 3. A "Gaussian process" on Θ with values in the real numbers is a stochastic process such that every FiDi follows a normal distribution.

It can be shown that every well-defined Gaussian process X is fully characterized, in terms of its distribution, by the following pair of functions:

1. The "mean function" $\mu : \Theta \to \mathbb{R}$, where $\mu(\theta) := \mathbb{E}(X_\theta)$, and,
2. the "covariance kernel" $\Gamma : \Theta \times \Theta \to \mathbb{R}$, where

$$\Gamma(\theta, \theta^\star) := \mathbb{E}(X_\theta X_{\theta^\star}) - \mu(\theta)\mu(\theta^\star) = \text{Cov}(X_\theta, X_{\theta^\star}).$$

Example 2. When $\Theta = \mathbb{R}$, $z \sim N(0, 1)$, and $x_\theta = \theta z$, the process is Gaussian with $\mu(\theta) = \mathbb{E}(X_\theta) = \theta\mathbb{E}(z) = 0$ for all $\theta \in \mathbb{R}$, and

$$\Gamma(\theta, \theta^\star) = \mathbb{E}(x_\theta x_{\theta^\star}) = \theta\theta^\star\mathbb{E}(z^2) = \theta\theta^\star.$$

4 Time Series

In the following, we will briefly discuss a category of processes where the elements of Θ represent points in time. These are called time series' processes (time series) and are widely used in fields such as economics as models for phenomena evolving over time under uncertainty:

Definition 4. A time series with values in $\mathbb{R}$ is defined as any process for which Θ is a totally ordered set, thus representing time.

Usually, Θ is a subset of $\mathbb{R}$ with the usual ordering, and we will denote its typical element by t. When Θ is an interval, the time series evolves in continuous time (Continuous-Time Process). A characteristic example of a Gaussian process in continuous time is the following:

Definition 5. Let $\Theta = [0, 1]$, and consider the process W, defined by the following properties:

- $W_0 = 0$ almost surely,
- For every $t, u \in [0, 1]$ such that $t + u \leq 1$, the random variable $W_{t+u} - W_t$ is independent of the σ-algebra $\sigma(W_s, 0 \leq s \leq t)$.
- For every $t, u \in [0, 1]$ such that $t + u \leq 1$, the random variable $W_{t+u} - W_t$ follows the distribution $N(0, u)$.
- The sample paths of the process are almost surely continuous functions $[0, 1] \to \mathbb{R}$,

is called the (one-dimensional) Wiener process on $[0, 1]$.

It can be shown that the process is well-defined via the Daniell-Kolmogorov theorem. Apart from the above definition, the Wiener process has other equivalent representations, e.g., through suitable series of iid random variables following the standard normal distribution. This illustrates how processes can be constructed via transformations of existing processes. Finally, the Wiener process serves as a prototypical example of a weak limit of sequences of discrete-time processes.

When $\Theta = \mathbb{N}$ or $\mathbb{Z}$ (or more generally, Θ is countably infinite), X is called a discrete-time time series (Discrete-Time Process) or a stochastic sequence (double stochastic sequence when Θ has no initial element). In the following, we will mainly assume $\Theta = \mathbb{Z}$-the concepts we discuss will also hold for $\mathbb{N}$ with appropriate modifications. In this context, $X = (X_t)_{t\in\mathbb{Z}} \triangleq (\ldots, X_{-1}, X_0, X_1, \ldots)$, where X_t is a random variable for each $t \in \mathbb{Z}$.

Example 3. The process X is iid, i.e., it consists of jointly independent random variables (see the related definitions in the previous chapter).

Example 4. The process X consists of jointly independent random variables with $X_t \sim N(t, t^2)$, $t \in \mathbb{Z}$.

4.1 Strong Stationarity

A particularly appealing property that a time series can have is the invariance of its FiDis with respect to time shifts. This is expected to hold for processes representing dynamic phenomena in some form of stochastic dynamic stability. We describe this with the property of strong stationarity.

Specifically, given a process $X = (x_t)_{t\in\mathbb{Z}}$, consider the finite and ordered subset of $\mathbb{Z}$, $\{t_1, t_2, \ldots, t_\kappa\}$, where $\kappa \in \mathbb{N}$. This defines the FiDi $\mathbb{P}_{x_{t_1}, x_{t_2}, \ldots, x_{t_\kappa}}$.

Since $\Theta = \mathbb{Z}$, we can shift the time points in the previous set by adding the same integer constant to each element, thus obtaining a new subset of Θ. Let $m \in \mathbb{Z}$ and consider the set $\{t_1 + m, t_2 + m, \ldots, t_\kappa + m\}$, referred to as the m-shift of $\{t_1, t_2, \ldots, t_\kappa\}$, and the corresponding FiDi is $\mathbb{P}_{x_{t_1+m}, x_{t_2+m}, \ldots, x_{t_\kappa+m}}$. For a general process, the two FiDis are not identical:

Definition 6. If for every $m \in \mathbb{Z}$, $\mathbb{P}_{X_{t_1}, X_{t_2}, \ldots, X_{t_\kappa}} = \mathbb{P}_{X_{t_1+m}, X_{t_2+m}, \ldots, X_{t_\kappa+m}}$ then we say that the specific FiDi remains invariant under time shifts. The process X is called strongly stationary if and only if every FiDi remains invariant under time shifts.

Thus, strong stationarity imposes a collection of strong restrictions on the behavior of the process. Automatically, we obtain the following necessary condition for strong stationarity.

Lemma 1. *If x is strongly stationary, then $\mathbb{P}_{x_t} = \mathbb{P}_{x_{t^\star}}$ for $t, t^\star \in \mathbb{Z}$, i.e., all marginal distributions in the process are identical.*

Proof. Consider $\{t\}$ as a singleton subset of $\mathbb{Z}$ and apply the definition. □

The process in Example 3 is stationary because

$$\mathbb{P}_{X_{t_1}, X_{t_2}, \ldots, X_{t_\kappa}} \overset{\text{ind.}}{=} \mathbb{P}_{X_{t_1}} \cdot \ldots \cdot \mathbb{P}_{X_{t_\kappa}} \overset{\text{homogeneity.}}{=} \mathbb{P}_{X_{t_1+m}} \cdot \ldots \cdot \mathbb{P}_{X_{t_\kappa+m}}$$

$$\overset{\text{ind.}}{=} \mathbb{P}_{X_{t_1+m}, X_{t_2+m}, \ldots, x_{t_\kappa+m}},$$

for arbitrary κ, m.

Example 5. The process in Example 4 is not stationary because it violates Lemma 1 (explain!).

Example 6. The Wiener process in Example 4 is not stationary because it also violates Lemma 1 (explain!).

4.1.1 Transformations and Stationarity

The above background suffices to study the issue of stationarity preservation under transformations. This will be useful, among other things, to determine whether such properties are inherited by processes that form solutions to systems of stochastic recurrence equations involving stationary processes. We first need a minor extension of the definition of stationarity:

Definition 7. The processes $X_1 = (X_{1_t})_{t\in\mathbb{Z}}$, $X_2 = (X_{2_t})_{t\in\mathbb{Z}}$, ..., $X_\rho = (x_{\rho_t})_{t\in\mathbb{Z}}$ are called jointly strongly stationary if and only if the joint distribution of every finite collection of random variables from the processes $X_1, X_2, \ldots, X_\rho$ is invariant under time shifts.

Thus, stationarity is a special case of joint stationarity for $\rho = 1$. The two concepts are not generally equivalent.

Lemma 2. *Let $X_1 = (X_{1_t})_{t\in\mathbb{Z}}$, $X_2 = (X_{2_t})_{t\in\mathbb{Z}}$, ..., $X_\rho = (X_{\rho_t})_{t\in\mathbb{Z}}$ be jointly strongly stationary. Let $f : \mathbb{R}^\rho \to \mathbb{R}$ be a suitably measurable function. Then $Y = \left(f\left(X_{1_t}, \ldots, X_{\rho_t}\right)\right)_{t\in\mathbb{Z}}$ is strongly stationary.*

Proof. Exercise! □

Example 7. Let $\rho = 1$, $f(z) = z^4$. The process $y = (X_t^4)_{t\in\mathbb{Z}}$ is strongly stationary when X is strongly stationary.

Example 8. Let $\rho = 2$, $f(z_1, z_2) := z_1 + z_2 + z_1 z_2$. Then the process $Y = \left(X_{1_t}^4 + X_{2_t} + X_{1_t} X_{2_t}^4\right)_{t\in\mathbb{Z}}$ is strongly stationary when X_1, X_2 are jointly stationary.

The lag operator L is defined by $L((z_t)_{t\in\mathbb{Z}}) := (z_{t-1})_{t\in\mathbb{Z}}$, which shifts each component of any doubly indexed sequence to the previous position. It is easily shown to be linear and suitably measurable. It is also easy to prove that it preserves stationarity:

Lemma 3. *If X is strongly stationary, then $L^m X := \underbrace{L \circ L \circ \cdots \circ L}_{m \text{ times}} X$ is strongly stationary.*

Proof. Exercise! □

Example 9. If X is strongly stationary, then $Y = \left(X_t^3 X_{t-m}^2\right)$ is strongly stationary. This follows directly from combining Lemmas 2-3.

Finally, it can be proven that suitable limits of stationary transformations are also stationary. The following result is given without proof. The interested reader is referred to Chapter 2 of [6] for the proof.

Lemma 4. *If X is strongly stationary and for every m, f_m:$\mathbb{R}^{\mathbb{N}} \to \mathbb{R}$ is suitably measurable and $f_m(y_1, y_2, \ldots) \to f(y_1, y_2, \ldots)$ as $m \to \infty$, then $Y = (f(X_t, X_{t-1}, \ldots))_{t\in\mathbb{Z}}$ exists as a stochastic process and is strongly stationary.*

Example 10. Let $f_m(y_1, y_2, \ldots) = f(y_1, y_2, \ldots) = \max_{0\le i\le m} y_i$. Then, if X is stationary and consists of random variables taking values in a compact subset of $\mathbb{R}$, the process $Y = (\max_{i\in\mathbb{N}} X_{t-i})_{t\in\mathbb{Z}}$ is also stationary, due to Lemma 4.

4.1.2 Stationarity and the Law of Large Numbers

It can be shown that strong stationarity alone is insufficient to guarantee a Law of Large Numbers, in which the almost sure limit is degenerate. One

way to achieve such a result is through the concept of ergodicity. This requires additional restrictions on the dependence structure among the random variables comprising the process, as encoded in their FiDis.

To proceed, we briefly examine the asymptotic behavior of arithmetic means of a strongly stationary process and outline a general version of Birkhoff's Law of Large Numbers. Let $X = (X_t)_{t\in\mathbb{Z}}$ be a stationary process. The following definition concerns a sub-σ-algebra that captures information about the effect of the lag operator L on the sample paths of the process.

Definition 8. The invariant σ-algebra $\mathcal{J}_X$ of X is the σ-algebra generated by all measurable subsets of $\mathbb{R}^{\mathbb{Z}}$ consisting of values of X that remain unchanged under the transformation induced by L. In other words, it is generated by any such set A for which $L(A) = A$.

The $\mathcal{J}_X$ is trivial if and only if every invariant event has either zero or unit probability (with respect to $\mathbb{P}$). From the definition of 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 Law of Iterated Expectations (LIE). Moreover, $\mathbb{E}(X_0/\mathcal{J}_X)$ is a constant if and only if $\mathcal{J}_X$ is trivial. The following LLN describes convergence to a limit that may not follow a degenerate distribution, i.e., it may converge to a non-constant limit. For the concept of almost sure convergence, see the part of the notes discussing stochastic convergence.

Lemma 5 (General Version of Birkhoff's LLN)**.** *If $X = (X_t)_{t\in\mathbb{Z}}$ is strongly stationary and $\mathbb{E}(|X_0|) < +\infty$, then*

$$\frac{1}{T}\sum_{t=1}^{T} X_t \to \mathbb{E}(X_0/\mathcal{J}_X),\ \mathbb{P}\text{ a.s.}.$$

Proof. Out of the scope of this text. Interested readers may refer to [9]. □

Thus, the arithmetic mean may converge to a non-degenerate random variable, and this occurs precisely when $\mathcal{J}_X$ is not trivial.

4.2 Ergodicity

Given the heuristic description of $\mathcal{J}_X$, a deeper understanding of the concept of ergodicity requires ideas from the theories of dynamical systems and measure theory, which go beyond the scope of this text. Nevertheless, the following definition is precise:

Definition 9. A stationary process $X = (X_t)_{t\in\mathbb{Z}}$ is called ergodic (with respect to L) if and only if $\mathcal{J}_X$ is trivial.

Example 11. It can be shown that if X is iid, then it is ergodic. The proof uses the Borel-Cantelli Lemma; see, for instance, [4] to show that the corresponding σ-algebra is trivial.

Example 12. If independence holds but there is heterogeneity, the process cannot be ergodic since it is not stationary.

Example 13. There are stationary processes that are not ergodic. For example, let $\varepsilon = (\varepsilon_t)_{t\in\mathbb{Z}}$ be iid with $\varepsilon_0 \sim \text{Unif}_{[-1,1]}$. Let $z \sim N(0,1)$, independent of ε_t for every $t \in \mathbb{Z}$. Construct the process $X = (X_t)_{t\in\mathbb{Z}}$ with $X_t = \varepsilon_t + z$. This process is stationary because for every $\kappa \in \mathbb{N}$ and every $m \in \mathbb{Z}$, using the Law of Iterated Expectations (LIE-see the Addendum below),

$$\mathbb{P}\left(X_{t_1} \in \cdot, X_{t_2} \in \cdot, \ldots, X_{t_{1\kappa}} \in \cdot\right) \overset{\text{LIE}}{=} \mathbb{E}\left(\mathbb{P}\left(X_{t_1} \in \cdot, X_{t_2} \in \cdot, \ldots, X_{t_{1\kappa}} \in \cdot\right)/\sigma\left(z\right)\right)$$

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

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

$$\overset{\text{LIE}}{=} \mathbb{P}\left(X_{t_1+m} \in \cdot, X_{t_2+m} \in \cdot, \ldots, X_{t_{1\kappa}+m} \in \cdot\right).$$

For ergodicity, consider $A := \{x_t \leq -1, t \in \mathbb{Z}\}$. We have

$$L_\star\left(A\right) = \{LX_t \leq -1, t \in \mathbb{Z}\} = \{X_{t-1} \leq -1, t \in \mathbb{Z}\} = A,$$

so this is invariant. However,

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

$$\geq \mathbb{P}\left(z \leq -2\right) = \Phi\left(-2\right) > 0,$$

where Φ is the cumulative distribution function of the standard normal distribution. Also,

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

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

so for the invariant A, we have

$$0 < \mathbb{P}\left(A\right) \leq \frac{1}{2}.$$

Hence, X cannot be ergodic.

It can be shown that ergodicity necessarily implies the property of weak

mixing, which describes the asymptotic independence (on average) between elements of the process. Specifically, it holds that

$$\lim_{k\to\infty} \frac{1}{k} \sum_{j=1}^{k} \text{Cov}\left(f\left(x_0\right), g\left(x_{-j}\right)\right) = 0,$$

for any measurable f, g for which the covariance exists. It is shown that ergodicity is equivalent to the above property, but it does not necessarily imply the stronger property of strong mixing, $\lim_{j\to\infty} \text{Cov}\left(f\left(x_0\right), g\left(x_{-j}\right)\right) = 0$. For the concept of mixing and its relationship with ergodicity, see, e.g., [2].

4.2.1 Ergodicity and Transformations

As with stationarity, we examine the issue of whether ergodicity is preserved under transformations. It can be shown that the following holds:

Lemma 6. *Lemmas 2-3-4 also hold for ergodicity.*

The following concerns the existence and "uniqueness" of stationary and ergodic solutions in systems of stochastic recurrence equations (SREs) that depend on past processes. The solutions of these are stochastic processes constructed as potentially complex transformations of the preceding processes, with the forms of these transformations depending on the structure of the recurrence relations. The result is presented without proof and can be easily generalized to cover a significant class of such systems. Note that the omitted proof is an application of Banach's Fixed Point Theorem (see, e.g., [6]).

Lemma 7. *Let $\varepsilon = (\varepsilon_t)_{t\in\mathbb{Z}}$ be stationary and ergodic. Let $f : \mathbb{R}^2 \to \mathbb{R}$ be suitably measurable and assume that for every $z \in \mathbb{R}$, $f(\cdot, z)$ is continuously differentiable with respect to its first argument. Suppose that:*

1. *There exists $y \in \mathbb{R}$ such that $\mathbb{E}\left(\ln^+\left(|f(y,\varepsilon_0) - y|\right)\right) < +\infty$, where $\ln^+$ denotes the positive part of the natural logarithm.*
2. *$\mathbb{E}\left(\ln^+\left(\sup_{x\in\mathbb{R}}\left|\frac{\partial f}{\partial x}(x,\varepsilon_0)\right|\right)\right) < +\infty$.*
3. *$\mathbb{E}\left(\ln\left(\sup_{x\in\mathbb{R}}\left|\frac{\partial f}{\partial x}(x,\varepsilon_0)\right|\right)\right) < 0$ (or equals $-\infty$).*

Then there exists a "unique" stationary and ergodic process $X = (X_t)_{t\in\mathbb{Z}}$ that satisfies

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

and is defined by

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

for any $y \in \mathbb{R}$ satisfying condition a).

The following constitutes the standard example of a recursion commonly referred to as an autoregressive scheme of first order (AR(1)):

Example 14. Let $\varepsilon = (\varepsilon_t)_{t\in\mathbb{Z}}$ be as in Lemma 7, with $\mathbb{E}\left(\ln^+\left(|\varepsilon_0|\right)\right) < +\infty$ and $f(x_1, x_2) = b_0 x_1 + x_2$. Consider the recursion $X_t = f(X_{t-1}, \varepsilon_t) = b_0 X_{t-1} + \varepsilon_t$, $t \in \mathbb{Z}$. Clearly, $f(\cdot, z)$ is continuously differentiable with respect to $\cdot$ for every z. Also, $\frac{\partial f(x,z)}{\partial x} = b_0$ and $\ln\left(\sup_{x\in\mathbb{R}}\left|\frac{\partial f(x,\varepsilon_0)}{\partial x}\right|\right) = \ln|b_0| = \mathbb{E}\left(\ln\left(\sup_{x\in\mathbb{R}}\left|\frac{\partial f(x,\varepsilon_0)}{\partial x}\right|\right)\right) < 0$ (or $-\infty$) if and only if $|b_0| < 1$. Thus, this recursion has a "unique" stationary

and ergodic solution $X = (x_t)_{t\in\mathbb{Z}}$ given by the limit

$$X_t = \lim_{m\to\infty} \underbrace{f \circ f \circ f \circ \ldots \circ f}_{m}(y, \varepsilon_t, \ldots, \varepsilon_{t-m}) = \lim_{m\to\infty} b_0^m y + \lim_{m\to\infty} \sum_{i=0}^{m} b_0^i \varepsilon_{t-i} = \sum_{i=0}^{\infty} b_0^i \varepsilon_{t-i},$$

almost surely, for any $y \in \mathbb{R}$. Consequently, when $|\beta| < 1$, the linear process

$$\left(\sum_{i=0}^{\infty} b_0^i \varepsilon_{t-i} \right)_{t\in\mathbb{Z}},$$

is the corresponding stationary and ergodic process of type AR(1).

The following example concerns a nonlinear process widely used in Financial Econometrics to model, among other things, various empirical characteristics of financial return time series.

Example 15. Let $Z = (z_t)_{t\in\mathbb{Z}}$ be iid with $\mathbb{E}(z_0) = 0$, $\mathbb{E}(z_0^2) = 1$, and let the function $f(x_1, x_2) = \omega_0 + (a_0 x_1^2 + b_0)x_2$, with parameters $\omega_0, a_0, b_0 > 0$. Consider the recursion $h_t = f(z_{t-1}, h_{t-1})$, $t \in \mathbb{Z}$. Clearly, $f(z, \cdot)$ is continuously differentiable with respect to $\cdot$ for every z. Also, $\frac{\partial f(x,z)}{\partial x} = a_0 x^2 + b_0$ and since $\mathbb{E}(\ln\left(\sup_{x\in\mathbb{R}} \left|\frac{\partial f(x,\varepsilon_0)}{\partial x}\right|\right)) = \mathbb{E}(\ln(a_0 x^2 + b_0)) < 0$ (or $-\infty$), this recursion has a "unique" stationary and ergodic solution $H = (h_t)_{t\in\mathbb{Z}}$ which, due to Lemma 5, is given by

$$h_t = \lim_{m\to\infty} \underbrace{f \circ f \circ f \circ \ldots \circ f}_{m}(y, z_t, \ldots, z_{t-m}) = \ldots = \omega_0[1 + \sum_{i=1}^{\infty} \prod_{j=1}^{i} (a_0 z_{t-j}^2 + b_0)], \quad (3)$$

almost surely. Due to Lemma 6, the process

$$Z \cdot (\sqrt{H}) := (z_t \sqrt{\omega_0[1 + \sum_{i=1}^{\infty} \prod_{j=1}^{i} (a_0 z_{t-j}^2 + b_0)]})_{t \in \mathbb{Z}},$$

is stationary and ergodic and is called a GARCH(1,1) process. In the above framework, the square root on the left-hand side is considered to hold pointwise, while the omission of details in deriving the limit involves, among other things, the asymptotic behavior as $m \to \infty$ of the term $\prod_{j=1}^{m}(a_0 z_{t-j}^2 + b_0)]$.

> *Exersice*: Show that for the above example, $\prod_{j=1}^{m}(a_0 z_{t-j}^2 + b_0)] \overset{\text{a.s.}}{\to} 0$ using the Birkhoff Ergodic Theorem. Consequently, derive (3).

Addendum

The Law of Iterated Expectations

The **Law of Iterated Expectations**, also known as the *Tower Property*, is a fundamental theorem in probability theory. It states that for any two information sets (algebras) $\mathcal{F}_1 \subseteq \mathcal{F}_\in$ and a random variables X:

$$\mathbb{E}(X \mid \mathcal{F}_1) = \mathbb{E}\left(\mathbb{E}(X \mid \mathcal{F}_2) \mid \mathcal{F}_1\right).$$

This means that projections on poorer information can be recursively constructed via projections on richer information. When $\mathcal{F}_1$ is non-informative, the outer expectation is the unconditional one.

Applications

- Sequential Decision Making: In finance and economics, the law is used to evaluate decisions that unfold over time.
-Stochastic Processes: It helps in analyzing processes where future states depend on current states.
-Simplifying Complex Expectations: By breaking down expectations into conditional components, complex problems become more manageable.

Stochastic Recurrence Equations

A **Stochastic Recurrence Equation** (SRE) is a recursive equation that models the evolution of a stochastic process over time. It is typically expressed as:

$$X_t = f(X_{t-1}, \varepsilon_t)$$

where:

- X_t is the state of the process at time t.
- f is a function describing the dynamics of the system.
- $\{\varepsilon_n\}$ is a sequence of random variables representing random shocks or noise.

A common linear form of an SRE is:

$$X_t = A_t X_{t-1} + B_n$$

with:

- $\{A_t\}$ and $\{B_t\}$ being sequences of random variables.
- X_0 as the initial state-when the underlying process is defined on $\mathbb{N}$.

Example

Stochastic Logistic Map

A nonlinear SRE example:

$$X_t = rX_{t-1}(1 - X_{t-1}) + \varepsilon_n$$

where:

- r is a growth rate parameter.
- ε_n introduces randomness into the system.

Solution Techniques

- **Analytical Methods**: For linear SREs, solutions can sometimes be derived explicitly. For some non-linear ones they can be approximated as limits oof backward substitutions.
- **Numerical Simulation**: Monte Carlo methods are used when analytical solutions are intractable.