1 Stationary sequences and Birkhoff's Ergodic Theorem

A stochastic process $\mathbf{X} = \{X_n : n \geq 0\}$ is called *stationary* if, for each $j \geq 0$, the shifted sequence $\theta_j \mathbf{X} = \{X_{j+n} : n \geq 0\}$ has the same distribution, that is, the same distribution as $\mathbf{X}$. In particular, this implies that X_n has the same distribution for all $n \geq 0$, but stationarity is much stronger than only that: It means that all joint distributions are the same: for any $n_1 < n_2 < \cdots < n_l$, any $l \geq 1$, the joint distribution of the vector $(X_{j+n_1}, X_{j+n_2}, \cdots, X_{j+n_l})$ is the same for all $j \geq 0$, that is, the same as $(X_{n_1}, X_{n_2}, \cdots, X_{n_l})$.

$$P(X_{j+n_1} \leq x_1,\ X_{j+n_2} \leq x_2, \ldots, X_{j+n_2} \leq x_l) = P(X_{n_1} \leq x_1,\ X_{n_2} \leq x_2, \ldots, X_{n_2} \leq x_l),$$

for all $x_1, \ldots x_l$.

A very special case of a stationary sequence is an independent identically distributed sequence (iid), but much more complex examples exist in applications. A large class of examples is illustrated by positive recurrent (discrete-valued with state space $\mathcal{S}$) Markov chains. Such processes have a unique limiting/stationary distribution $\pi = (\pi_i),\ i \in \mathcal{S}$, and it is well known that if the chain is started off initially with distribution π (independent of all else), $P(X_0 = i) = \pi_i,\ i \in \mathcal{S}$, then the chain is a stationary process, called a *stationary version* of the chain.

It turns out that any stationary sequence $\mathbf{X}$ is either *ergodic* or not. Non-ergodic means that the distribution of $\mathbf{X}$ can be expressed as the *mixture* of 2 distinct *stationary* sequences $\mathbf{X}(1)$ and $\mathbf{X}(2)$ (say) distributions: For some $p \in (0,1)$, $P(\mathbf{X} \in \cdot) = pP(\mathbf{X}(1) \in \cdot) + (1-p)P(\mathbf{X}(2) \in \cdot)$. The idea is that you initially flip a coin (once) that lands H wp p and lands T wp $1-p$. If it lands H then set $\mathbf{X} = \mathbf{X}(1)$; if it lands T then set $\mathbf{X} = \mathbf{X}(2)$. The point is that if no such mixture can be found to exist, then $\mathbf{X}$ is ergodic.

Ergodic sequences (can be shown) to include all iid sequences, positive recurrent regenerative sequences (such as positive recurrent Markov chains) and many others.

Non-ergodic sequences can of course easily be constructed from any 2 distinct stationary (and ergodic) sequences. Here is a simple example: Let $\mathbf{X}(1)$ be an iid sequence of exponential rate 1 rvs, and let $\mathbf{X}(2)$ be an iid sequence of exponential rate 2 rvs. In essence, each sequence represents a Poisson process, but with different rates, 1 and 2 respectively. Now flip a p coin and create the mixture $\mathbf{X}$ mentioned above. Then with probability p all sample paths are that of a Poisson process at rate 1 and with probability $1-p$ all all sample paths are that of a Poisson process at rate 2. $\mathbf{X} = \{X_n : n \geq 0\}$ is stationary with common mean $E(X) = (1/2)(1 + 0.5) = 0.75,\ n \geq 0$.

Note how now, the elementary renewal theorem does not hold: wp1, $\lim_{t\to\infty} N(t)/t = \Lambda$, where Λ is not a constant, but a random variable instead; $P(\Lambda = 1) = P(\Lambda = 2) = 0.5$. Also note that it does not hold that $E(\Lambda) = 1/E(X)$: $E(\Lambda) = 3/2$ while $1/E(X) = 4/3$.

Similarly (equivalently) the SLLN does not hold: wp1

$$\lim_{n\to\infty} \frac{1}{n} \sum_{j=1}^{n} X_j = Y,$$

where Y is a not a constant but a random variable: $P(Y = 1) = P(Y = 0.5) = 0.5$. The sequence $\mathbf{X}$ is stationary, but it is not iid. It is, however, *conditionally iid*: Conditional on the coin landing H, it is iid with exponential 1 interarrival times, while conditional on the coin landing T it is iid with exponential 2 interarrival times. (The point process constructed using $\mathbf{X}$ for interarrival times is not a Poisson process: Its counting process $\{N(t) : t \geq 0\}$ does not have independent increments: Choosing t huge, and looking at the value of $N(t)/t$ would allow us to estimate which way the coin landed, via what the value of λ is (1 or 2) , thus biasing any future increment's distribution to have that value of λ.

It turns out that when ergodic, a stationary sequence exhibits limiting behavior as if it were iid: If $\mathbf{X}$ is ergodic, and $E|X| < \infty$, then wp1

$$\lim_{n\to\infty} \frac{1}{n} \sum_{j=1}^{n} X_j = E(X).$$

The formal statement of this result is much stronger than the classic SLLN, and includes it as a very special case. It is called *Birkhoff's Ergodic Theorem.* We deal with this next.

1.1 Birkhoff's Ergodic Theorem

To prepare for it, we need to consider real-valued functions $f = f(\mathbf{x})$ where $\mathbf{x} = \{x_0, x_1, \ldots\}$ is an infinite sequence. Examples include

1. $f(\mathbf{x}) = x_0$ or $f(\mathbf{x}) = x_5$.
2. Indicator functions such as $f(\mathbf{x}) = I\{x_0 \leq y_0\}$ or more generally $f(\mathbf{x}) = I\{x_{n_1} \leq y_1, \ldots, x_{n_l} \leq y_l\}$.
3. $f(\mathbf{x}) = x_2 + x_7 x_{100}$
4. $$f(\mathbf{x}) = \sum_{n=0}^{\infty} a_n x_n,$$
   where $\{a_n\}$ is a sequence of real numbers such that the sum converges.

In a stochastic setting, when $\mathbf{X}$ is a stochastic process, we get (for example) $f(\mathbf{X}) = X_0$, or $f(\mathbf{X}) = I\{X_{n_1} \leq y_1, \ldots, X_{n_l} \leq y_l\}$. Thus we can take expected values of such functionals; $E(f(\mathbf{X})) = E(X_0)$ or $E(f(\mathbf{X})) = P(X_{n_1} \leq y_1, \ldots, X_{n_l} \leq y_l)$.

The proof of this theorem is beyond the scope of these notes. But we state it here:

Theorem 1.1 (Birkhoff's Ergodic Theorem) *If $\mathbf{X}$ is stationary and ergodic, and f is such that $E|f(\mathbf{X})| < \infty$, then wp1*

$$\lim_{n\to\infty} \frac{1}{n} \sum_{j=1}^{n} f(\theta_j \mathbf{X}) = E(f(\mathbf{X})). \tag{1}$$

Note how this includes the SLLN (extended to stationary ergodic sequences): Choosing $f(\mathbf{x}) = x_0$ yields $f(\theta_j \mathbf{X}) = X_j$ and thus we get wp1

$$\lim_{n\to\infty} \frac{1}{n} \sum_{j=1}^{n} X_j = E(X),$$

where we are letting X denote a generic X_n since they all have the same distribution by definition of stationarity.

We finally point out that in fact a stationary sequence $\mathbf{X}$ is ergodic if and only if (1) holds wp1 for all f such that $E|f(\mathbf{X})| < \infty$.