EE477 Universal Schemes in Information Theory	Lecture 9 - 10/20/2009

Ergodic Processes

Lecturer: Tsachy Weissman — *Scribe: Matt Kraning*

1 Ergodic Processes

Note: In this class, as in many scientific communities, we define ergodicity of a process in the context of a stationary process.

Definition 1. *A finite-alphabet stationary process* $\mathbf{X}$ *is ergodic if* $\forall k$ *and every* $f : \mathcal{X}^{2k+1} \to \mathbb{R}$,

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} f(X_{i-k}^{i+k}) \to \mathbb{E} f(X_{-k}^{k}) \qquad w.p.1$$

Intuitivly, ergodic processes are "reliable" in that one process realization (L.H.S. of above equation) reveals all process statistics (R.H.S. of above equation).

Exercise 1. Ergodic Examples [1, 2, 3]

(a) Show X_i i.i.d. $\Rightarrow \mathbf{X}$ is ergodic

(b) Show that $\mathbf{X} = \left\{ \begin{array}{ll} \ldots 000000000 \ldots & w.p. \frac{1}{2} \\ \ldots 111111111 \ldots & w.p. \frac{1}{2} \end{array} \right\}$ is not ergodic

(c) Show that $\mathbf{X} = \left\{ \begin{array}{ll} \ldots 010101010 \ldots & w.p. \frac{1}{2} \\ \ldots 101010101 \ldots & w.p. \frac{1}{2} \end{array} \right\}$ is ergodic

Theorem 2. *(The "Ergodic Decomposition" Theorem) Every stationary process is a mixture of ergodic processes. That is,* $\forall \mathbf{P}$ *stationary,* $\exists$ *a family of ergodic processes* $\{\mathbf{P}_\theta\}_{\theta \in \Theta}$ *and a probability measure* $\mu(\theta)$ *on* Θ *such that* $\mathbf{P} = \int_\Theta \mathbf{P}_\theta d\mu(\theta)$.

Conversely, if $\mathbf{P}$ *is not ergodic, then it can be expressed as a non-trivial mixture of different stationary processes.*

For $\mathbf{X}$, let $\mathbf{X}^{(k)}$ denote the k-th order super source (e.g. $X_1^{(k)} = X_1^k$, $X_2^{(k)} = X_{k+1}^{2k}$)

Definition 3. $\mathbf{X}$ *is "totally ergodic" if* $\mathbf{X}^{(k)}$ *is ergodic* $\forall k$

Example 1 Look at $\mathbf{X} = \left\{ \begin{array}{ll} \ldots 01010101 \ldots & w.p. \frac{1}{2} \\ \ldots 10101010 \ldots & w.p. \frac{1}{2} \end{array} \right\}$ and consider $\mathbf{X}^{(2)} = \left\{ \begin{array}{ll} \ldots (01)(01)(01)(01) \ldots & w.p. \frac{1}{2} \\ \ldots (10)(10)(10)(10) \ldots & w.p. \frac{1}{2} \end{array} \right\}$.

$\mathbf{X}^{(2)}$ is not ergodic as it is equivalent to the single-letter process $\mathbf{Y} = \left\{ \begin{array}{ll} \ldots 000000000 \ldots & w.p. \frac{1}{2} \\ \ldots 111111111 \ldots & w.p. \frac{1}{2} \end{array} \right\}$, which is not ergodic. Thus, $\mathbf{X}$ is not totally ergodic.

Exercise 2. Let $\mathbf{X}$ be a "B-process" defined as $X_i = f(Y_{i-k}^{i+k})$ for i.i.d. process $\mathbf{Y}$. Show that $\mathbf{X}$ is totally ergodic.

2 Ergodicity: General Setup

We consider a more general setup for ergodicity on the probability space $(\Omega, \mathfrak{F}, \Pr)$. A mapping $\varphi : \Omega \to \Omega$ is *measure-preserving* if $\Pr(\varphi(A)) = \Pr(A)$ for all $A \in \mathcal{F}$. A random variable $X = X(\omega)$ is a measurable mapping from Ω to $\mathbb{R}$. An event $A \in \mathcal{F}$ is said to be *invariant* if $\Pr(\varphi^{-1}(A) \triangle A) = 0$, where the notation $A \triangle B = (A \setminus B) \cup (B \setminus A)$ is the symmetric difference of sets A and B.

Exercise 3. Show that the class of invariant events $\mathcal{I}$ is a σ-field.

Definition 4. *The mapping φ is "ergodic" if $\mathcal{I}$ is trivial, that is,* $\Pr(A) \in \{0,1\}$ *for all $A \in \mathcal{I}$.*

Theorem 5 (The ergodic theorem/Birkhoff's ergodic theorem/Pointwise ergodic theorem [4])**.** *For any $X \in L_1$,*

$$\frac{1}{n}\sum_{m=0}^{n-1} X(\varphi^m \omega) \overset{n\to\infty}{\longrightarrow} \mathsf{E}[X|\mathcal{I}] \text{ a.s. and in } L_1.$$

Since $\mathsf{E}[X|\mathcal{F}] = \mathsf{E}[X]$ a.s. for all X if and only if $\mathcal{F}$ is trivial, by the ergodic theorem, the following statements are equivalent.

(a) For any $X \in L_1$, $(1/n)\sum_{m=0}^{n-1} X(\varphi^m\omega) \overset{n\to\infty}{\longrightarrow} \mathsf{E}[X]$ w.p.1 and in L_1.

(b) $\mathcal{I}$ is trivial.

(c) φ is ergodic.

To define ergodicity for random processes, we consider the probability space $(\Omega, \mathcal{F}, \Pr)$, where $\Omega = \mathcal{X}^\infty$, the probability measure Pr is the distribution of the stationary random process, and $\varphi : \Omega \to \Omega$ is defined as a shift in time, namely $\varphi(\omega)_i = \omega_{i+1}$. Note that the stationarity of the process implies that φ is measure preserving. Then, the above equivalence justifies the following definition of ergodicity for process:

Definition 6. *A stationary process $\mathbf{X}$ is "ergodic" if for all k and every $f : \mathcal{X}^{2k+1} \to \mathbb{R}_X$ s.t. $f(X_{-k}^k) \in L_1$:*

$$\frac{1}{n}\sum_{i=1}^{n} f(X_{i-k}^{i+k}) \overset{n\to\infty}{\longrightarrow} \mathsf{E}[f(X_{-k}^k)]\ , \text{ w.p.1 and in } L_1.$$

Exercise 4. For the examples given in Exercise 1, verify ergodicity or lack thereof using Definition 4.

Definition 4 brings some insight into the ergodic decomposition theorem which states that a process $\mathbf{X}$ is ergodic if and only if it cannot be represented as a mixture of different stationary processes. By definition, if $\mathbf{X}$ is not ergodic then $\mathcal{I}$ defined by φ is not trivial. It follows that there exists an event A such that $\Pr(A) > 0$, $\Pr(A^c) > 0$, $\varphi(A) = A$, and $\varphi(A^c) = A^c$. Essentially, we can divide the sample space Ω into A and A^c, and interpret the random process $\mathcal{X}$ as a mixture of two processes in each subset with probability $\Pr(A)$ and $\Pr(A^c)$. For details we refer to Chapter 6 of [4].

The examples of ergodic processes we looked at have infinite memory. Although there are multiple ways to formally define decaying memory, when a process is "mixing" or has decaying memory, even the weakest definition of a "mixing" process implies ergodicity.

References

[1] T. Berger, "Rate-Distortion Theory", Prentice-Hall, 1971

[2] K. Petersen, "Ergodic theory", Cambridge University Press, 1983

[3] R. M. Gray, "Entropy and Information Theory", Springer-Verlag, 1990

[4] R. Durrett, *Probability: Theory and Examples,* 3rd ed., Duxbury Press,