Northwestern University
Department of Electrical and Computer Engineering

ECE 428: Information Theory Spring 2004

Lecture 5 Supplemental Notes:

In this lecture (and the end of the previous one) we began looking at the entropy rate of a (discrete-time) stochastic process. We showed that stationary processes always have a finite entropy rate and that it can be calculated by either

$$\lim_{n\to\infty} \frac{1}{n} H(X_1, \ldots, X_n),$$

or

$$\lim_{n\to\infty} H(X_n | X_{n-1}, \ldots, X_1).$$

The main reason that entropy rate is the right quantity to look at is because the A.E.P. can be generalized to stochasitic processes using this quantity. This more general form of the A.E.P. is also known as the "Shannon-McMillan-Breiman Theorem". It says that for any finite-valued stationary ergodic process $\{X_n\}$,

$$-\frac{1}{n} \log p(X_1, \ldots, X_n) \to H(\mathcal{X}) \quad \text{with prob. 1.}$$

A statement of this and a proof can be found in Sect. 15.7 of the text. Based on this, we can then use entropy rate to define the set of "typical sequences" from any stationary ergodic source and extend the analysis from the previous lecture to this setting.

In introductory stochastic process courses, you are often introduced the concept of ergodicity as meaning that time-average are equal to stochastic averages, i.e.,

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i}^{n} f(X_i) = E(X_1),$$

for some function f. In mathematical probability, ergodic processes are defined soemwhat differently and the above result is a consequence of the *ergodic theorem*. You can find some discussion of this in Sect. 15.7 on the text.