Northwestern University
Department of Electrical and Computer Engineering

ECE 428: Information Theory — Spring 2004

Lecture 3 Supplemental Notes:

In this lecture, we discussed the data processing inequality and Fano's inequality from Chapter 2; we will use both of these when we discuss channel capacity later in the quarter. We then began discussing the Asymptotic Equipartition Property (see Chapter 3 in text). This results requires an understanding of the convergence of random variables and the law of large numbers. The following notes will review these concepts.

Recall a random variable X is a function defined on a probability space, Ω, also called the sample space. Today, we will be interested in real-valued random variables, *i.e.,* $X : \Omega \mapsto \mathbb{R}$.

A sequence of random variables $\{X_n\}$ can be viewed as a sequence of functions all defined on the same probability space, i.e., $X_n : \Omega \mapsto \mathbb{R}$. Each sample point in Ω then corresponds to a different sequence of real numbers; this sequence may or may not converge.

Several different types of convergence are useful when dealing with random variables, these include:

- **Convergence in Probability:** $X_n \to X$ in probability if given any $\epsilon > 0$,

$$\lim_{n\to\infty} \Pr(|X_n - X| > \epsilon) = 0.$$

- **Almost sure convergence:** $X_n \to X$ almost surely (abbreviated (a.s.)) if $\Pr(X_n \to X) = 1$. This is also called convergence with probability one (w.p.1).
- **Convergence in Distribution:** $X_n \to X$ in distribution if $F_n(x) \to F(x)$ for all x at which $F(x)$ is continuous. Here $F_n(x) = \Pr(X_n \leq x)$.
- **Mean Square Convergence:** $X_1 \to X$ in the m.s. sense if

$$\lim_{n\to\infty} \mathbb{E}_{X_n}[|X_n - \mathbb{E}_X[X]|^2] \to 0.$$

The relationship between these is shown in fig. 1

In ECE 422, you deal primarily with convergence in distribution (e.g. the central limit theorem) and mean square convergence (e.g. filtering and estimation). Our main interest here will be in convergence in probability and a.s. convergence. These two types of convergence are used to state the weak and the strong laws of large numbers. Let $\{X_n\}_{n=1}^{\infty}$ be an i.i.d. sequence of random variables with finite mean $\bar{X} = \mathbb{E}(X_n)$. For $n = 1, 2, \ldots$, let $S_n = X_1 + \cdots + X_n$ be the nth partial sum.

Weak law of large numbers: $\frac{1}{n}S_n \to \bar{X}$ in probability.

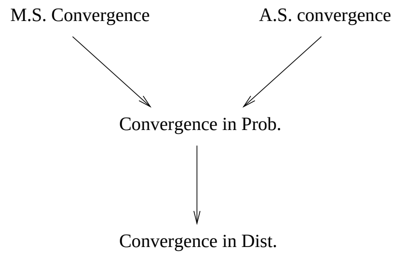

Figure 1: Relationship between different types of convergence for RV's.

When the X_n's have finite variance, the weak LLN's is a simple consequence of Chebyshev's inequality - see problem 3.1 on your homework. The finite variance assumption is not necessary, but a different proof is required.

Given the stronger assumption that $\mathbf{E}[|X_n|] < \infty$, then the following holds:

Strong law of large numbers: $\frac{1}{n}S_n \to \bar{X}$ a.s.

The AEP as given in Chapter 3 of the book is a direct consequence of the weak law of large numbers. Under various conditions, there are also versions of the AEP that hold with probability one.