EE 121 : Introduction to Digital Communication Systems

Problem Set for Discussion Section 1

Mon 1/28/2008 and Wed 1/30/2008

1. (based on 2.3 Gallager's book) (a) For a nonnegative integer-valued r.v. N, show that $E[N] = \sum_{n=0}^{\infty} \Pr(N > n)$.

(b) Show, with whatever mathematical care you feel comfortable with, that for an arbitrary nonnegative r.v. X that $E[X] = \int_0^{\infty} \Pr(X \geq a) da$.

(c) Derive the *Markov inequality*, which says that for any nonnegative r.v. $\Pr(X \geq a) \leq E[X]/a$.

(d) Derive the *Chebyshev inequatliy*, which says that $\Pr(|Y - E[Y]| \geq b) \leq \sigma_Y^2/b^2$ for any r.v. Y with finite mean $E[Y]$ and finite variance σ_Y^2. Hint: Use part (c) with $(Y - E[Y])^2 = X$.

(e) Derive the *(Weak) Law of Large Numbers* for i.i.d. random variables, which says that if $X_1, X_2, \ldots, X_n$ is a sequence of i.i.d random variables with finite variance, $\frac{1}{n}\sum_{i=1}^{n} X_i \rightarrow E[X_i]$ in probability, i.e. for any $\epsilon > 0$ as $n \rightarrow \infty$

$$\Pr\left(\left|\frac{1}{n}\sum_{i=1}^{n} X_i - E[X_i]\right| \geq \epsilon\right) \rightarrow 0.$$

(f) Can you come up with an example of a sequence of *dependent* random variables where the Law of Large Numbers still holds? no longer holds?

2. Craig has just finished sequencing his own DNA but only has a single compact disc on which to store it. His DNA sequence consists of 3 billion nucleotides. Each nucleotide is a character drawn from the 4-letter alphabet {*Adenine, Cytosine, Guanine, Thymine*} or $\{A, C, G, T\}$ for short.

(a) Can Craig use a naive, fixed-length source code, where each nucleotide is represented by a 2-bit string? Assume the compact disc can hold up to 700 megabytes of data.

(b) After browsing the internet Craig finds a website that claims Adenine, Cytosine, Guanine and Thymine nucleotides occur with relative frequencies 3:1:4:2, in human DNA. If this information were true, describe a variable-length source code that enables Craig to store his DNA sequence on a single compact disc? Assume nucleotides are independent.

(c) After a more throughout investigation, Craig learns that these frequencies are incorrect, and that the true relative frequencies are 5:2:2:1. If he uses your source code without modification, will the data still fit on the compact disc?

(d) Can you suggest a way of creating a variable-length source code that uses even less storage space than your code from part (b)?