

**Math 186, Winter 2009, Prof. Tesler**  
**Homework #1, Due Monday January 12, 2009**

Larsen & Marx **Fourth Edition:**

2.2# 2, 3, 4, 6, 12, 26, 28, 31, 32

2.3# 1, 5, 6, 10, 18

2.4# 1, 9, 22, 29, 34, 45

and the problems below: H-1 and H-2.

**Problem H-1.** DNA sequences can be specified as strings in the alphabet of nucleotides, A, C, G, T; for example, AACGTC is a DNA sequence of length 6. (DNA is double-stranded; note that we are only specifying the sequence on one strand.)

- (a) What is the number of DNA sequences of length 10?
- (b) What is the number of DNA sequences of length 10 that contain at least one G? (This restriction only refers to the one strand, not to the complementary strand.)
- (c) What is the number of DNA sequences of length 10 that contain at least one G but do not contain any A's? (This restriction only refers to the one strand, not to the complementary strand.)

**Problem H-2.** The SARS coronavirus was recently sequenced and is available at

[http://www.ncbi.nlm.nih.gov/entrez/viewer.fcgi?val=NC\\_004718](http://www.ncbi.nlm.nih.gov/entrez/viewer.fcgi?val=NC_004718)

<http://www.ncbi.nlm.nih.gov:80/genomes/SARS/SARS.html>

The sequence consists of 8481 A's, 5940 C's, 6187 G's and 9143 T's, in a particular order.

- (a) Mathematically define the sample space consisting of all positions in this sequence.
- (b) What are the empirical probabilities of each nucleotide?
- (c) What is the empirical probability that a nucleotide is a purine (A or G)?
- (d) What is the empirical probability that a nucleotide is a pyrimidine (C or T)?