

Stat 536 HW 1
August 28,2008

1. In a population of a 100 individuals, 25 have genotype **AA** and 75 have genotype **aa** at a locus. 5% of the **AA** type are susceptible to anemia while only 2% of the **aa** type suffer from this condition. If an individual is picked at random from the population and she is susceptible to anemia, what is the probability that she has genotype **AA**?
2. A 500 bp long sequence has 150 purines and 350 pyrimidines. A point mutation at any position can cause a purine to pyrimidine change with probability $1/4$ and pyrimidine to purine change with probability $1/3$.
 - (a) If a purine is observed at a particular location, what is the probability that it was a purine before the last mutation event?
 - (b) If a dinucleotide purine-purine is observed, what is the probability distribution of the four possible dinucleotide sequences that could have resulted in the observed dinucleotide. (*Note: dinucleotide refers to two successive nucleotides in the sequence of interest*)
3. Mendel performed very simple early genetics experiments by crossing varieties of peas. The characteristics he selected (e.g. wrinkled vs. smooth peas) were obvious to the eye and turned out to be what we now call Mendelian traits (i.e. controlled by a single locus) with just two alleles, labeled A and a . In the first season, Mendel selected an AA plant and an aa plant. (In fact, he did not know the plants' genotypes, but selected plants based on phenotypes that had been selected by breeders for many generations.) He meticulously crossed AA with aa and collected the seeds. The next season he planted only the sampled seeds from the previous generation and crossed them with each other. Again he collected the resulting seeds.
 - (a) What is the sample space of possible genotypes for the seeds collected from the second generation?
 - (b) Applying the *Law of Segregation*, derive the probability mass function for the second generation seed genotypes.
 - (c) For some traits, Mendel observed a 3:1 ratio of phenotypes in the second generation. Explain how the phenotype pmf derives from the genotype pmf.

- (d) To show that the larger class actually consisted of two distinct types of plants, Mendel sampled 100 plants with the dominant phenotype, selfed them (crossed these plants with themselves) and typed 10 offspring. What is the probability that all 10 offspring have the dominant phenotype?
- (e) If all 10 offspring shared the dominant phenotype, Mendel called the parent type I, else type II. He theorized (hypothesized) that the ratio of type I to type II was 2:1. He was actually slightly wrong. Why?
- (f) Test whether the following data shows evidence against the correct ratio (not exactly 2:1).

Type I	Type II
72	28

4. You have discovered a new gene, 768 base pairs long, that encodes for a protein you think interacts with the HIV virus and makes people more or less susceptible to the disease. You sample 11 individuals from an African population where the virus has been prevalent and active for a long time. Your excitement about a potential discovery increases when you detect variation, i.e. multiple alleles, in your sample. You hypothesize that though HIV is a relatively new human pathogen, that it has exerted extreme selection on the population you study. If mutants of this gene can make people less susceptible to the disease, then you theorize that these mutations will have been highly selected in the population you study. Looking more closely at your data, you observe that 14 of the 768 sites in the gene are *segregating* two nucleotides and 10 of these mutations are *nonsynonymous*. Use the genetic code to formulate a simple model that assumes there has been no selection (assume selection acts at the protein level only). Use your model to test whether there is evidence of selection acting on this gene. Critique your model.