Bi/GE105: Evolution
Homework 3
Due Date: Tuesday, February 4, 2014

"the theory of probabilities is basically just common sense reduced to calculus; it makes one appreciate with exactness that which accurate minds feel with a sort of instinct, often without being able to account for it." - Pierre Simon de Laplace

1. The Great Dying and the Siberian Traps.

(a) Read the paper by Saunders and Reichow and provide a two-paragraph summary of the paper that describes the geological event and the hypotheses being considered for its biological significance. Submit your response in pdf form to both professors and all of the TAs. No word documents will be accepted.

(b) In the paper, the authors talk about current estimates of the total volume of eruptives associated with the Siberian Traps and offer a number of 3×10^6 km^3. How thick a layer could one cover the United States with with this volume of Lava? Estimate how this compares to the volume it took to create the big island of Hawaii? (NOTE: remember that Hawaii has a large underwater component.) The goal here is to get a sense of the volume scale associated with these massive events.

2. The Poisson distribution and the number of mutations per genome.

(a) In this first part of the problem, we derive the Poisson distribution. This distribution is one of the great probability distributions in all of science and is relevant to our thinking about mutations and genomes (and much more). The question we ask ourselves is simple: what is the probability that a genome will have n mutations? What we imagine is that in going from one generation to the next, replication at each base is like flipping a highly biased coin where if we get heads (with probability $q = 1 - p$), the base is copied correctly and if we get tails (with probability p) the base is copied incorrectly. However, the coin is so biased that the probability of a tail is very small (i.e. $p << 1$). Begin by arguing why the probability that in a genome with N base pairs, the probability of getting n mutations is given as the binomial distribution

$$p(n,N) = \frac{N!}{(N-n)!n!} p^n q^{N-n}. \tag{1}$$

Note that for our case, we have the proviso that $p << q$. If we exploit the fact that $N >> n$, we can use the approximation that

$$\frac{N!}{(N-n)!} \approx N^n, \tag{2}$$

resulting in

$$p(n,N) \approx \frac{N^n}{n!} p^n (1-p)^N, \tag{3}$$

where we have also exploited the fact that $p + q = 1$. Make sure to demonstrate that all of these steps make sense and give some argument about the approximation of eqn. 2. For example, what happens when $N = 10^9$ and $n = 2$ - this should help you demonstrate why the approximation works. Maybe you can even say something about how big an error we make with this approximation in such cases.

For the binomial distribution, recall that the average value of n is given by $\langle n \rangle = Np$. Use this to show that we can rearrange the expression given above as

$$p(n,N) \approx \frac{(\langle n \rangle)^n}{n!} (1 - \frac{\langle n \rangle}{N})^N. \tag{4}$$

Finally, by invoking the definition of the exponential function embodied in the equation

$$e^{-x} = lim_{N \to \infty} (1 - \frac{x}{N})^N, \tag{5}$$

we see that our distribution adopts the Poisson form, namely,

$$p(n) = \frac{(\langle n \rangle)^n e^{-\langle n \rangle}}{n!}. \tag{6}$$

Make sure you perform the relevant steps in the derivation to show that everything works out as claimed here.

(b) Work out the mean and the variance of the Poisson distribution. In particular, use the fact that

$$\langle n \rangle = \sum_{n=0}^{\infty} n p(n) = \sum_{n=0}^{\infty} n \frac{(\langle n \rangle)^n e^{-\langle n \rangle}}{n!} \tag{7}$$

and rewrite our average as

$$\langle n \rangle = e^{-\langle n \rangle} \langle n \rangle \frac{d}{d\langle n \rangle} \sum_{n=0}^{\infty} \frac{(\langle n \rangle)^n}{n!}. \tag{8}$$

Make sure you demonstrate that this way of rewriting the average is correct and then evaluate the sum on the right and carry out the operations to find the result of interest. Use exactly the same method of differentiation to find $\langle n^2 \rangle$ and then the variance and compute the variance over the mean for a Poisson distribution.

(c) Given the Poisson distribution for the number of mutations in a genome of length N, figure out the probability of getting 1 mutation, 2 mutations, etc. for the case of the human genome. Work out the average number of mutations by explicitly calculating it from the distribution function using the result of part (b). Note that to get a numerical answer, you will need to choose a value for p from part (a) based upon your knowledge of mutation rates.

(d) Another Frances Arnold problem! In Prof. Arnold's lab, they do directed evolution. Their goal is to screen through various mutant proteins to see if they can find those that do a "better" job at some task than the naturally occurring protein. In a qualifying exam that she and Prof. Phillips gave, one question centered on questions such as how one might create new proteins by creating libraries of mutants in the active site of the enzyme. Specifically, in a paper by Oliphant and Struhl, they focused on β-lactamase proteins and considered a 17 amino acid section of the protein in the active site. What Prof. Arnold was interested in was based upon the mutation strategy used by Oliphant and Struhl. They prepared a library of roughly 500,000 mutant proteins based upon mutated oligonucleotides for the 17-amino acid active site. On average, 20% of the nucleotides in these oligos were mutated. Using the Poisson distribution worked out in the earlier part of the problem, provide the distribution which tells how many mutations these oligos will have. Make a histogram of the distribution showing the probability of having m mutations in the N-base pair fragment (what is N?). Roughly 2000 of the proteins obtained from this library were able to confer ampicillin resistance. Out of the 500,000 members of the library, how many of them does the Poisson distribution tell us to expect had *zero* mutations?

3. Your Galapagos Project - part 1.

Provide a one-paragraph summary of the topic you have chosen/been assigned for our trip to the Galapagos. In addition, submit the name of at least one reference from the primary literature which is focused on your topic. Submit your response in pdf form to both professors and all of the TAs. No word documents will be accepted.