EEB 600A, Problem Set One Solutions

1 : Suppose you are following two species (A and B) in the fossil record, and observe that 20 percent of all your samples contain both species, while species A is found in 35 percent of the samples and species B is found in 50 percent of the samples.

(a) What is the conditional probability of finding species A when species B is present?

Pr(A | B) = Pr(A,B)/ Pr(B) = 0.20/0.50 = 0.40

(b) What is the conditional probability of finding species B when species A is present?

Pr(B | A) = Pr(A,B)/ Pr(A) = 0.20/0.35 = 0.57

2 : Assume that substitutions follow a Poisson process (i.e., the number follows a Poisson distribution), and that there is site-to-site variation in the sequence of interest, with fast sites having a rate of 5×10^{-7}/yr, immediately sites having a rate of 10^{-8}/yr, and slow sites having a rate of 10^{-9}/yr. The frequency of fast, medium, and slow sites are 0.05, 0.65, 0.30. If we pick a random site, what is the probability that it has not mutated after 10^7 years? *Hint:* Condition over all the possible sites.

Pr(No mutation) = Pr(no mut | slow site)*Pr(slow) + Pr(no mut | medium)*Pr(medium) + Pr(no mut | fast)*Pr(fast)

$$= \exp(-10^{-9} * 10^7) * 0.30 + \exp(-10^{-8} * 10^7) * 0.65 + \exp(-5 \times 10^{-7} * 10^7) * 0.05 = 0.885$$

3 : Consider the same parameters as in Problem 3, and suppose we pick a random site showing no mutation. What is the probability that this is a fast site? A medium site? A slow sight? (Hint: Bayes)

$$\Pr(\textit{fast} \mid \textit{no mutation}) = \frac{\Pr(\textit{no mutation} \mid \textit{fast}) * \Pr(\textit{fast})}{\Pr(\textit{no mutation})} = \frac{e^{-5} * 0.05}{0.885} = 0.0003805$$

Likewise, $\Pr(\textit{medium} \mid \textit{no mutation}) = 0.6642$ ***and*** $\Pr(\textit{slow} \mid \textit{no mutation}) = 0.3354$***.***

4 : Suppose the mutation rate is 10^{-6} per generation. How many generations to we have to wait to have a 50 percent chance that (at least) one mutation has occured? How many generations for a 90 percent chance?

$$\Pr(\textit{at least one mutation} \mid t \textit{ gens}) = 1 - \Pr(\textit{no mutations} \mid t \textit{ gens})$$

$$= \left(1 - 10^{-6}\right)^t \simeq \exp(-10^{-6} \cdot t)$$

$$\Pr(\textit{at least one mutation} = \exp(-10^{-6} \cdot t) \geq 0.5$$

Taking logs gives, $t \geq -ln(0.5)/10^{-6} = 6.9 \times 10^5$ ***generations. Likewise, for 90%,*** $t \geq -ln(0.1)/10^{-6} = 2.3 \times 10^6$ ***generations.***

5 : As a comparison of the binomal and Poisson distributions, suppose the mutation rate is 10^3 per site per year and suppose we look at 500 sites.

(a) Using the binomal, what is the probability of two or fewer mutations?

Here $p = 10^{-3}$ ***and*** $n = 500$***, and hence***

$$P(x) = \frac{500!}{(500 - x)!\, x!}\, 0.001^x\, (0.999)^{500-x}$$

so that Pr(2 or less) = Pr(0) + Pr(1) + Pr(2),

$$= (0.999)^{500} + 500 \cdot 0.001\, (0.999)^{499} + \frac{500 \cdot 499}{2} \cdot 0.001^2\, (0.999)^{498}$$

$$= 0.6064 + 0.3035 + 0.0758 = 0.98577$$

(b) What is this same probability using the Possion?

The expected number of mutations is $p \cdot n = 0.5$***, and***

$$\begin{aligned}\Pr(0) + \Pr(1) + \Pr(2) &= e^{-0.5} + 0.5 \cdot e^{-0.5} + \frac{0.5^2}{2} \cdot e^{-0.5} \\ &= 0.6065 + 0.3033 + 0.0758 = 0.9856\end{aligned}$$

6 : Consider the following discrete random variable X, which takes on three values:

$$\Pr(X) = \begin{cases} -1 & \text{with probability } 0.1 \\ 0 & \text{with probability } 0.3 \\ 1 & \text{with probability } 0.6 \end{cases}$$

Compute the following

(a) $E[\,X\,]$

$$E[\,X\,] = \mu_x = -1 \cdot 0.1 + 0 \cdot 0.3 + 1 \cdot 0.6 = 0.5$$

(b) $E[\,X^2\,]$

$$E[\,X^2\,] = 1 \cdot 0.1 + 0 \cdot 0.3 + 1 \cdot 0.6 = 0.7$$

(c) $\sigma^2(X)$

$$\sigma^2 = E[\,X^2\,] - \mu_X^2 = 0.45$$

(d) the Skew of X, $E[\,(X-\mu_x)^3\,]$

$$E[\,(X-\mu_x)^3\,] = (-1-0.5)^3 \cdot 0.1 + (0-0.5)^3 \cdot 0.3 + (1-0.5)^3 \cdot 0.6 = -0.3$$

(e) the fourth moment of X, $E[\,(X-\mu_x)^4\,]$

$$E[\,(X-\mu_x)^4\,] = (-1-0.5)^4 \cdot 0.1 + (0-0.5)^4 \cdot 0.3 + (1-0.5)^4 \cdot 0.6 = 0.5625$$

(f) the kurotsis of X

From the Lecture 2 notes, the kurtosis is

$$\frac{E[\,(X-\mu_x)^4\,] - 3\sigma^4}{\sigma^4} = \frac{0.5625 - 3 \cdot 0.45^2}{0.45^2} = -2/9 = 0.222$$

(g) $E[\,\exp(X)\,]$

$$E[\,\exp(X)\,] = e^{-1} \cdot 0.1 + e^0 \cdot 0.3 + e^1 \cdot 0.6 = 1.968$$