Probability Models, solutions to Test 1

1(a) The only probabilistic element in Orr's model is that an interaction may or may not lead to an incompatibility, and the probability it does is p. In our model, we also introduce randomness in the number of interactions, described by a random variable X with range $\{0, 1, ..., \binom{K}{2}\}$ (not all values in the range are necessarily possible). In Orr's model, the number of interactions is constant $\binom{K}{2}$.

(b) Recall that selecting a node corresponds to a substitution being fixed at a gene. If difference nodes have different probabilities of being selected, it means that different genes have different mutation and fixation rates.

2(a) We have $K = 20, p = 0.01$, and $\alpha = 0.0087$ which gives speciation probability

$$E[S] \approx 1 - (1 - 0.01)^{0.0087 \cdot \binom{20}{2}} = 0.0165$$

Consider the yeast network with $K = 20$ and $p = 0.01$.

(b) Doubling K to $K = 40$ gives

$$E[S] \approx 1 - (1 - 0.01)^{0.0087 \cdot \binom{40}{2}} = 0.0659$$

which is a 300% increase.

(c) To double the speciation probability to 0.033, trial and error shows that K must be about 28 which is a 40% increase.

(d) Here $\alpha = 1$ which gives speciation probability

$$S = 1 - (1 - 0.01)^{\binom{20}{2}} = 0.85$$

(e) We must have

$$\alpha \binom{K}{2} = \binom{20}{2}$$

where $\alpha = 0.0087$ which gives $K \approx 209$, that is, 10 times as many substitutions are required.

(f) Snowball!

3. We have $p = 0.5$ and estimate the speciation probability to be $14/30 \approx 0.47$. Over 100 generations we expect $K = 5$ substitutions and estimate α by solving

$$1 - (1-p)^{\alpha\binom{5}{2}} = 1 - 0.5^{10\alpha} = \frac{14}{30}$$

which gives $\alpha = 0.09$.

4(a) The total number of choices of 3 nodes is $\binom{5}{3} = 10$ and by direct counting we obtain the probabilities

$$P(X=0) = 0.2,\ P(X=1) = 0.5,\ P(X=2) = 0.2, P(X=3) = 0.1$$

(b) By definition of expected value:

$$E[X] = 0 \cdot 0.2 + 1 \cdot 0.5 + 2 \cdot 0.2 + 3 \cdot 0.1 = 1.2$$

Further,

$$E[X^2] = 0^2 \cdot 0.2 + 1^2 \cdot 0.5 + 2^2 \cdot 0.2 + 3^2 \cdot 0.1 = 3.8$$

which gives

$$\text{Var}[X] = E[X^2] - (E[X])^2 = 0.76$$

Alternatively, use the formulas for $E[X]$ and $\text{Var}[X]$ from the article.

(c) We have $N_E = 4$ and $N = 5$ which gives

$$\alpha = \frac{N_E}{\binom{N}{2}} = 0.4$$

and, just to check, we get $E[X] = \alpha\binom{K}{2} = 0.4\binom{3}{2} = 1.2$ as we should.

(d) Exactly:

$$E[S] = 1 - \sum_{j=0}^{3}(1-p)^j P(X=j) = 0.9^0 \cdot 0.2 + 0.9^1 \cdot 0.5 + 0.9^2 \cdot 0.2 + 0.9^3 \cdot 0.1 = 0.1151$$

Approximately:

$$E[S] \approx 1-(1-p)^{\alpha\binom{K}{2}} = 1-0.9^{1.2} = 0.1188$$

5(a) Note that each choice of 3 nodes corresponds to precisely 3 edge pairs that share a node, and there are thus 3 times as many such edge pairs as there are node triplets. As there are $\binom{N}{3}$ ways to choose 3 nodes we get

$$N_S = 3\binom{N}{3}$$

Alternatively, use the formula

$$N_S = \sum_{j=1}^{N}\binom{d_j}{2}$$

where d_j is the degree of the jth node. Since all nodes have degree $d_j = N-1$ we get

$$N_S = N\binom{N-1}{2} = \frac{N(N-1)(N-2)}{2}$$

which also equals $3\binom{N}{3}$.

For N_D, in a similar way note that each choice of 4 nodes corresponds to precisely 3 edge pairs that do not share a node and hence

$$N_D = 3\binom{N}{4}$$

Alternatively, note that $N_D = \binom{N_E}{2} - N_S$ where $N_E = \binom{N}{2}$.

(b) No edge pairs share a node so $N_S = 0$ and $N_D = \binom{N/2}{2} = \frac{N(N-2)}{8}$.

(c) Here $N = 12$ and $N_E = 21$ which gives

$$N_S = \sum_{j=1}^{N}\binom{d_j}{2} = 10\binom{3}{2} + 2\binom{6}{2} = 60$$

and

$$N_D = \binom{N_E}{2} - N_S = \binom{21}{2} - 60 = 150$$