$$G^2 = 2\sum_{j=1}^{k} X_j \log\left(\frac{X_j}{n\hat{\pi}_j}\right)$$
$$G^2 = 2\sum_{j} O_j \log\left(\frac{O_j}{E_j}\right)$$

Sneha Krishna Kumaran

The Multinomial Distribution Model

The Basic Model

First we look at the set $X = \{1, ..., N\}$ which denotes all plant species.

On some day t, only a subset of X is available at one particular meadow. The availability of plants at this meadow on day t can be denoted by a matrix A_t of size $N \times 1$ where the availability of plant i is denoted by $A_t(i) \in \{0, 1\}$.

The score function $\phi : X \mapsto \mathbb{R}$ can be denoted as a matrix ϕ of size $N \times M$ where M is the number of all pollinator species and $\phi(i, j) \in \mathbb{R}$. This score function gives a real-valued score to plant i depending on the pollinator j's preference for i. Suppose that pollinator j makes a visit to an arbitrary plant denoted by $v_t(j)$ on a particular day, where $v_t(j) \in \{1, ..., N\}$. Then the probability that plant i was visited by the pollinator j is

$$P(v_t(j) = i | A_t, \phi) = \frac{A_t(i)exp(\phi(i, j))}{\sum_{i'=1}^{N} A_t(i')exp(\phi(i', j))}$$

Now, say that the pollinator j makes a sequence of visits on day t denoted by $V_t = \{v_1, v_2, ..., v_K\}$, where we have dropped j for notational convenience. Then the number of times j visits i is

$$N_t(i) = \sum_{k=1}^{K_t} \mathbb{1}(V_t(k) = i)$$

and the likelihood of this particular sequence of plants visited on day t is

$$L(V_t; \phi) = \frac{K_t!}{N_t(1)!...N_t(N)!} \prod_i P(v_t(j) = i)^{N_t(i)}$$

Over the entire summer, the likelihood is

$$L(V_t; \phi) = \prod_t \left(\frac{K_t!}{N_t(1)!...N_t(N)!} \prod_i P(v_t(j) = i)^{N_t(i)} \right)$$
$$= \prod_t \left(\frac{K_t!}{N_t(1)!...N_t(K)!} \prod_i \left(\frac{A_t(i)exp(\phi(i, j))}{\sum_{i'=1}^{N} A_t(i')exp(\phi(i', j))} \right)^{N_t(i)} \right)$$

Fitting the Model to Our Data

Our data is different from the model in that we do not have frequency data. In order to fit binary data to the multinomial model, we will denote the usage matrx by U_t of size $N \times M$ where M is the number of all pollinator species and where $U_t(i, j) \in \{0, 1\}$.

Now the probability that pollinator j used plant i on day t is

$$P(U_t(i,j) = 1 | A_t, \phi) = \frac{A_t(i) exp(\phi(i,j))}{\sum_{i'=1}^{N} A_t(i') exp(\phi(i',j))}.$$

Now, $N_t(i) \in \{0,1\}$ because we do not know how many times the plant was used, only if it was used or not. Therefore the denominator of the previous likelihood function becomes 1. Because $K_t = \sum_i U_t(i,j)$ our new likelihood is

$$L(\mathbf{V};\phi) = \prod_t \left(\left(\sum_i U_t(i,j) \right)! \prod \left(\frac{A_t(i) exp(\phi(i,j))}{\sum_{h=1}^{N} A_t(h) exp(\phi(h,j))} \right)^{U_t(i,j)} \right).$$

The log likelihood over the entire summer for one pollinator j would be

$$\begin{aligned} LL(V_t;\phi) &= \log \left(\prod_t \left(\left(\sum_i U_t(i,j) \right)! \prod_i \left(\frac{A_t(i) exp(\phi(i,j))}{\sum_{h=1}^{N} A_t(h) exp(\phi(h,j))} \right)^{U_t(i,j)} \right) \right) \\ &= \sum_t \left(\log \left(\left(\sum_i U_t(i,j) \right)! \right) + \sum_i U_t(i,j) \log \left(\frac{A_t(i) exp(\phi(i,j))}{\sum_{i'=1}^{N} A_t(i') exp(\phi(i',j))} \right) \right). \end{aligned}$$