1 Calculating the Hessian Matrix

$$f = \sum_i p_i \log p_i$$

$$p_n = 1 - \sum_i^{n-1} p_i$$

$$\frac{df}{dp_k} = \log p_k + - \log(1 - \sum_i^{n-1} p_i)$$

$$\frac{d^2 f}{dp_a dp_b} = \frac{1}{p_a} \delta_{ab} + (1 - \sum_i^{n-1} p_i)^{-1}$$

2 Fast calculation of LogDet(H)

H has the form $H = D + S$, where D is a diagonal matrix and S is a constant matrix.

$$\begin{pmatrix} d_1 + s & s & s & \dots & s \\ s & d_2 + s & s & \dots & s \\ s & s & d_3 + s & \dots & s \\ . & . & . & \dots & . \\ s & s & s & \dots & d_n + s \end{pmatrix}$$

We subtract the first column from all the others, which leaves the determinant unchanged:

$$\begin{pmatrix} d_1 + s & -d_1 & -d_1 & \dots & -d_1 \\ s & d_2 & 0 & \dots & 0 \\ s & 0 & d_3 & \dots & 0 \\ . & . & . & \dots & . \\ s & 0 & 0 & \dots & d_n \end{pmatrix}$$

To calculate the determinant, we expand in minors via the last column. This produces two terms, the first of which is given by:

$$d_n * \det \begin{pmatrix} d_1 + s & -d_1 & -d_1 & -d_1 \\ s & d_2 & 0 & 0 \\ . & . & . & . \\ s & 0 & 0 & d_{n-1} \end{pmatrix}$$

This term is simply the determinant of a similarly structured matrix (H_{n-1}) of size $(n-1) \times (n-1)$. The other term is the determinant

$$-d_1 * (-1)^n * \det \begin{pmatrix} s & d_2 & 0 & 0 & 0 \\ s & 0 & d_3 & 0 & 0 \\ . & . & . & . & . \\ s & 0 & 0 & 0 & d_{n-1} \\ s & 0 & 0 & 0 & 0 \end{pmatrix}$$

By permuting columns and repeated expansion of minors, this determinant evaluates to

$$d_1 d_2 d_3 ... d_{n-1} s$$

Putting together the two terms gives:

$$\det H_n = d_n \det H_{n-1} + s d_1 d_2 ... d_{n-1}$$

This can be written as

$$\det H_n = \prod_{i=1}^{n} d_i + s \sum_i \prod_{j \neq i} d_j$$

$$\det H_n = (1 + s \sum_i \frac{1}{d_i}) \prod_j d_j$$

$$\log \det H_n = \log(1 + s \sum_i \frac{1}{d_i}) + \sum_j \log d_j$$

Now, we insert p_i from above and simplify, leading to:

$$\log \det H_n = -\sum_{j=1}^{n} \log p_j$$

The log likelihood term will thus be

$$\frac{1}{2} \sum_{j=1}^{n} \log p_j$$

3   Alternative derivation using log form

The idea here is to work in log space, because that makes the most sense when it comes time to do integrals over all space. This should reduce any approximation error incurred because of the finite domain of the unit simplex.

Suppose that

$$\rho_i = \rho_i^0 \exp(-u_i)$$

Then for $i < m$

$$\frac{\partial \rho_i}{\partial u_i} = -\rho_i$$

Also,

$$\rho_n = 1 - \sum_j^{m-1} \rho_j$$

Therefore,

$$\frac{\partial \rho_n}{\partial u_i} = +\rho_i$$

Let $f = \sum_{j=1}^{m} \rho_j \log \rho_j$.

$$\frac{\partial f}{\partial u_a} = -\rho_a \log \rho_a + \rho_a \log \rho_n$$

and

$$\frac{\partial^2 f}{\partial u_a \partial u_b} = \delta_{ab}(\rho_a \log \rho_a + \rho_a - \rho_a \log \rho_n) + \rho_a \rho_b \rho_n^{-1}$$

Thus, the hessian matrix can be represented by

$$H = D + uv^t$$

where

$$d_a = \rho_a \log \rho_a + \rho_a + \rho_a \log \rho_n$$

$$u_a = \rho_a$$

$$v_a = +\rho_a \rho_n^{-1}$$

The matrix determinant lemma provides the value of determinant:

3

$$\det(H) = (1 + v^t D^{-1} u) \det(D) = \frac{1}{\rho_n}(\rho_n + \rho^T D^{-1} \rho) \det(D)$$

Now, we need to work out the Jacobian of the coordinate transformation into u. For now, we neglect the constraint on $\phi = \sum_j f_j \rho_j$. In this case, the matrix of coordinate derivatives is diagonal with entries $-\rho_i = -\exp(-u_i)$.

We therefore have the integral:

$$-\int \exp(-\sum_j u_j - \frac{\lambda}{2} u^T H u) du$$

The integral can be evaluated as

$$-(2\pi)^{\frac{1}{2}} \det(\lambda H)^{\frac{-1}{2}} \exp(-\frac{\lambda}{2} 1^t H^{-1} 1)$$

The Sherman-Morrison formula gives:

$$H^{-1} = D^{-1} - \frac{D^{-1} u v^T D^{-1}}{1 + v^T D^{-1} u}$$

However, it does not look like this formula will simplify into something nice, as we have complicated terms in both the determinant and the exponential.

4   Using powers

Suppose

$$\rho_i = \rho_i^0 \Delta^s$$

Then

$$\frac{\partial \rho_i}{\partial \Delta_i} = \frac{\rho_i s}{\Delta_i}$$

$$\frac{\partial \rho_n}{\partial \Delta_i} = -\frac{\rho_i s}{\Delta_i}$$

$$\frac{\partial f}{\partial \Delta_a} = \frac{s}{\Delta_a}(\rho_a \log \rho_a - \rho_a \log \rho_n)$$

$$\frac{\partial^2 f}{\partial \Delta_a \partial \Delta_b} = \delta_{ab} \frac{s^2}{\Delta_a^2} \rho_a - \frac{\rho_a \rho_b s^2}{\rho_n \Delta_a \Delta_b}$$

The challenge here is the Jacobian term, which will depend on the value of Δ and thus not drop out of the integral for most cases. Some of these integrals might be have closed form solutions in terms of gamma functions, but they will likely not have nice interpretations for use with our model.

5 Using (first) powers

$$\frac{\partial^2 f}{\partial \Delta_a \partial \Delta_b} = \delta_{ab} (\rho_a^0)^2 \frac{1}{\rho_a} + \rho_a^0 \rho_b^0 \frac{1}{\rho_n}$$

When we evaluate this at the maxent solution, we obtain:

$$\frac{\partial^2 f}{\partial \Delta_a \partial \Delta_b} = \delta_{ab} \rho_a^0 + \rho_a^0 \rho_b^0 \frac{1}{\rho_n^0}$$

The determinant eventually becomes:

$$\frac{1}{\rho_n} \prod \rho_j$$

The Jacobian is

$$\prod \rho_j$$

This should lead to a slightly different result by adjusting the exponent on the $\prod \rho_j$ term.