Notations and basic tools

Let $\mu = \sum n_i p_i$, and $V = \sum m_i (\text{diag}(p_i) - p_i p_i')$, where $m_i = n_i (1 + \lambda_i c_i)$ and $c_i = C_i (n_i - 1) / n_i$, with $C_i$ the average cluster size. The vectors $p_i$, where $i = 1, \ldots, I$, are vectors of transition probabilities with the last category removed. Thus they are of dimension $J - 1$ and describe the behavior at election 2 of voters of party $i$ at election 1, with the reference category omitted. The log likelihood for a generic local unit may be written as

$$2L(\beta) = -\log[\det(V)] - (y - \mu)' V^{-1} (y - \mu),$$

where $\beta$ is the vector of all independent parameters, with one component related to the conditional probabilities $p_i$ and one to the over-dispersion parameters $\lambda_i$. In particular we assume that

$$\beta_i = \log(p_{ij} / p_i), \text{ and } \beta_i = \log(\lambda_i (1 - \lambda_i)],$$

where $t$ stands for transition and $o$ for over dispersion. It can be shown that the corresponding reconstruction formulas are

$$\mu = \frac{\exp(\beta_i)}{1 + 1' \exp(\beta_i)} \text{ and } \lambda_i = \frac{\exp(\beta_i)}{1 + \exp(\beta_i)}.$$

The following general derivation rules will be used:

$$\frac{\partial \det(Y)}{\partial x} = \det(Y) \text{Tr} \left[ Y^{-1} \frac{\partial Y}{\partial x} \right],$$

$$\frac{\partial Y^{-1}}{\partial x} = -Y^{-1} \frac{\partial Y}{\partial x} Y^{-1}.$$

Let $\Omega_i = \text{diag}(p_i) - p_i p_i'$; it can be shown that $\Omega^{-1} = \text{diag}(p_i)^{-1} + 11' / p_{ij}$.

Score vector

The general expression for the derivative of $2L(\beta)$ is

$$s_x = -\text{Tr}(V^{-1} V x) + (y - \mu)' V^{-1} V x V^{-1} (y - \mu) + 2(y - \mu)' V^{-1} \mu_x,$$

where $\mu_x$ and $V_x$ denote the derivative, respectively, of the mean vector and the variance matrix by a generic parameter. Let $e_j$ be a vector of $J - 1$ elements which are 0 except for the $j$th element which is 1, then we have

$$\frac{\partial \mu}{\partial p_i} = n_i \Omega_{i-1}, \quad \frac{\partial p_i}{\partial (\beta_i)} = \Omega_i,$$

$$\frac{\partial \mu}{\partial p_{ij}} = m_i (e_j e_j' - e_i p_i' - p_i e_j),$$

$$\frac{\partial V}{\partial \lambda_i} = n_i \epsilon \Omega_i, \quad \frac{\partial \lambda_i}{\partial \beta_i} = \lambda_i (1 - \lambda_i).$$

From now on let $H = V^{-1}$, $d(A)$ be the vector made of the diagonal elements of $A$, $\epsilon = y - \mu$, $K = H - H e \epsilon' H$. We will also use $h_{jj}$ to denote the $j$th diagonal element of $H$ and $h_{jj}$ will denote the $j$th column and recall that

$$\frac{\partial H}{\partial x} = -H V_x H.$$

Note that both terms involving the derivative with respect to $V$ are part of a symmetric expression, so that $e_j p_i'$ and $p_i e_j'$ are equivalent; we also use the fact that $\text{Tr}(A d A') = a' A d A$.

$$\frac{\partial L(\beta)}{\partial p_{ij}} = -m_i [\text{Tr}(H (e_j e_j' - 2 e_j p_i') - e' (H e_j e_j' H - 2 H p_i e_j' H) e'_j / 2 + n_i e_j' H \epsilon]

= -m_i [h_{jj} - 2 h_{jj'} p_i - h_{jj'} \epsilon' h_{jj} + 2 h_{jj'} \epsilon' H p_i] / 2 + n_i h_{jj'} \epsilon

= -m_i [(h_{jj} - h_{jj'} \epsilon' h_{jj}) - 2 (h_{jj'} - h_{jj'} \epsilon' H) p_i + n_i h_{jj'} \epsilon].$$

The vector derivative is obtained by putting these elements one above the other; let $K = H - H e \epsilon' H$

$$\frac{\partial L(\beta)}{\partial p_i} = -m_i [d(K) / 2 - K p_i] + n_i H \epsilon.$$
As concerns the derivative with respect to λi, by using the property that \( e'\mathbf{H}\Omega,He = \mathbf{H}\epsilon'\mathbf{H}\Omega, \) we have

\[
\frac{\partial L(\beta)}{\partial \lambda_i} = -n_i c_i \left[ \text{Tr}(\mathbf{H}\Omega, i) - e'\mathbf{H}\Omega, i\epsilon / 2 \right] \\
= -n_i c_i \text{Tr}[(\mathbf{H} - \mathbf{H}\epsilon'\mathbf{H})\Omega, i] / 2 \\
= -n_i c_i \text{Tr}(\mathbf{K}\Omega, i) / 2.
\]

When there is no modeling restriction on \( \beta, \) the score vector may be written as

\[
s(\beta) = \Omega_i \{-m_i[\mathbf{d}(\mathbf{K})/2 - \mathbf{Kp}, i] + n_i \epsilon\},
\]

\[
s(\beta) = -n_ic_i \lambda_i (1 - \lambda_i) \text{Tr}(\mathbf{K}\Omega, i) / 2.
\]

With restrictions, there will be a matrix \( \mathbf{X}' \) for transition parameters and \( \mathbf{X}'' \) for over dispersion parameters which will left multiply the corresponding score vector.

**Expected information matrix**

When computing the expected information matrix, the following result follows from the fact that \( E(\epsilon) = 0 \)
and \( E(\epsilon') = \mathbf{H}^{-1}, \)

\[
E \left( \frac{\partial \mathbf{K}}{\partial x} \right) = E(\mathbf{H} - \mathbf{H}\epsilon'\mathbf{H} - \mathbf{H}\epsilon'\mathbf{H} - \mathbf{H}\epsilon'\mathbf{H}) = -\mathbf{H}_x.
\]

Note, in addition, that \( E(\mathbf{K}) = 0, \) that the \( \mathbf{d}(\cdot) \) operator can be interchanged with expectation and differentiation, that \( \mathbf{d}(\mathbf{ab}) = \mathbf{a} \times \mathbf{b}, \) where \( \times \) denotes the product element by element,

\[
E \left( \frac{\partial \mathbf{d}(\mathbf{K})}{\partial p_{kj}} \right) = -E \left( \frac{\partial \mathbf{d}(\mathbf{H})}{\partial p_{kj}} \right) = m_k [h_i h'_j h_p - h_j h'_p H_p h'_j] = m_k [h_i \times h_j - 2h_j \times (\mathbf{H}p_k)],
\]

because of the symmetry of the \( \mathbf{d}(\cdot) \) operator; the derivative with respect to \( p'_k \) is obtained by placing these vectors one beside the other, giving

\[
E \left( \frac{\partial \mathbf{d}(\mathbf{K})}{\partial p_k} \right) = m_k [\mathbf{H} \times \mathbf{H} - 2\text{diag}(\mathbf{H}p_k)\mathbf{H}].
\]

In a similar way we obtain

\[
E \left( \frac{\partial \mathbf{Kp}}{\partial p_{kj}} \right) = -E \left( \frac{\partial \mathbf{H}p}{\partial p_{kj}} \right) = m_k [h_i h'_j h_p - h_j h'_p H_p h'_j] = m_k [\mathbf{H} \times \mathbf{H} - 2\text{diag}(\mathbf{H}p_k)\mathbf{H}].
\]

In doing so we have omitted differentiating \( p_k \) with respect to \( b_p \) because \( E(\mathbf{K}) = 0. \) In order to combine the derivative with respect to different elements of \( \mathbf{p}, \) note that \( p'_k \mathbf{H} \mathbf{p} \) is a constant scalar and \( h'_k \mathbf{p} \) is a scalar depending on \( j \) which multiplies each row, so that we have

\[
E \left( \frac{\partial \mathbf{Kp}}{\partial p_k} \right) = -m_k [\mathbf{H} \times \text{diag}(\mathbf{H}p_k) - \mathbf{H} (p'_k \mathbf{H} \mathbf{p}) - (\mathbf{H}p'_k)(\mathbf{H}p_k)',]
\]

Finally we can put together the different pieces and obtain the expected information matrix relative to \( \mathbf{p}, p'_k \)

\[
\mathbf{F}(\mathbf{p}, p'_k) = m_i m_k [\mathbf{H} \times \mathbf{H} / 2 - \text{diag}(\mathbf{H}p_k)\mathbf{H} - \mathbf{H} \times \text{diag}(\mathbf{H}p_k) + \mathbf{H} (p'_k \mathbf{H} \mathbf{p}) + (\mathbf{H}p'_k)(\mathbf{H}p_k)',] + n_i n_k \mathbf{H}.
\]

To compute the second derivative with respect to elements of \( \lambda, \) note that

\[
E \left( \frac{\partial L(\beta)}{\partial \lambda_i \partial \lambda_k} \right) = -n_i c_i E \left[ \text{Tr} \left( \frac{\partial \mathbf{K}\lambda}{\partial \lambda_k} \right) \right] / 2 = n_i c_i \text{Tr} \left( \frac{\partial \mathbf{H}}{\partial \lambda_k} \Omega_i \right) / 2 \\
= -n_i n_k c_i c_k \text{Tr}(\Omega_i, \mathbf{H}\Omega, i) / 2.
\]

The mixed derivative may be computed as follows by exploiting the properties of the trace operator

\[
E \left( \frac{\partial L(\beta)}{\partial \lambda_i \partial p_{kj}} \right) = -n_i E \left[ \text{Tr} \left( \frac{\partial \mathbf{K}\lambda}{\partial p_{kj}} \right) \right] / 2 = n_i c_i \text{Tr} \left( \frac{\partial \mathbf{H}}{\partial p_{kj}} \Omega_i \right) / 2 \\
= -n_i c_i m_k \text{Tr}(h_i h'_j \Omega_i - h_j h'_p (\mathbf{H}p_k \Omega_j) - (\mathbf{H}p_k h'_j \Omega_i))/2 \\
= -n_i c_i m_k (h'_j \Omega_i, h_j - 2h'_j \Omega_k, \mathbf{H}p_k)/2.
\]
so that we finally have

$$\mathbf{F}(\mathbf{p}_k, \lambda_i) = n_i c_i m_k [d(\mathbf{H} \Omega, \mathbf{H})/2 - \mathbf{H} \Omega, \mathbf{H} \mathbf{p}_k].$$

The corresponding components of the information matrix relative to $\beta$ are given below where $\mathbf{D}_k$ is a shorthand for $\text{diag}(\mathbf{H} \Omega, \mathbf{H})$,

$$F[\beta'_i, (\beta'_k)'] = m_i m_k \Omega_i [\mathbf{H} \ast \mathbf{H} / 2 - \mathbf{D}_k \mathbf{H} - \mathbf{H} \mathbf{D}_k + \mathbf{H} (\mathbf{p}'_k \mathbf{H} \mathbf{p}_k) + (\mathbf{H} \mathbf{p}_k)(\mathbf{H} \mathbf{p}_k)' \Omega_k + n_i n_k \Omega_i \mathbf{H} \Omega_k,$$

$$F(\beta'_i, \beta'_k) = m_i m_k c_i \Omega_k [d(\mathbf{H} \Omega, \mathbf{H})/2 - \mathbf{H} \Omega, \mathbf{H} \mathbf{p}_k] \lambda_k (1 - \lambda_k),$$

$$F(\beta'_i, \beta'_k) = n_i n_k c_i c_k \lambda_k (1 - \lambda_k) \lambda_i (1 - \lambda_i) \text{Tr}(\mathbf{H} \Omega_k \mathbf{H} \Omega_i)/2.$$