PMEDM - The nitty gritty derivations. PMEDM as a likelihood problem

Assume that microdata represent a histogram. Then the problem is to use the histogram to estimate the unknown probability density.

Suppose that each person is selected into the sample with probability q_{ij}, and that each person is independently sampled, and that $w_{ij} = np_{ij}$ is the expected sample count of people in each histogram bin. Note that w_{ij} here is different than in Nagle et al (2013); there, $w_{ij} = Np_{ij}$, i.e. the expected *population* in each bin, here it is the expected *sample* in each bin. The multinomial density of these histogram bins are

$$\frac{n!}{\Pi_{ij} w_{ij}!} \prod_{ij} q_{ij}^{w_{ij}}$$

Multiply this by the Gaussian density of the tract-level and block group-level errors (assuming independence, which isn't so, but we'll fight that fight another day):

$$\prod_{j \in J_T} \prod_{k \in K_T} \frac{1}{2\pi\sigma_j^2} \exp\left(-\frac{e_{jk}^2}{2\sigma_{jk}^2}\right) \prod_{j \in J_B} \prod_{k \in K_B} \frac{1}{2\pi\sigma_j^2} \exp\left(-\frac{e_{jk}^2}{2\sigma_{jk}^2}\right)$$

We thus maximize the joint likelihood

$$\frac{n!}{\Pi_{ij}(w_{ij})!} \prod_{ij} q_{ij}^{w_{ij}} \prod_{j \in J_T} \prod_{k \in K_T} \frac{1}{2\pi\sigma_j^2} \exp\left(-\frac{e_{jk}^2}{2\sigma_{jk}^2}\right) \prod_{j \in J_B} \prod_{k \in K_B} \frac{1}{2\pi\sigma_j^2} \exp\left(-\frac{e_{jk}^2}{2\sigma_{jk}^2}\right)$$

subject to the constraints that:

$$\sum_{j'} A_{j,j'} w_{ij} X_{ik} = Y_{jk} + e_{jk}$$

for all census tracts ($j \in J_T$) and block groups ($j \in J_B$).

From likelihood to log-likelihood

Taking the log of the likelihood, we get

$$\log(n!) - \sum_{ij} \log w_{ij}! + \sum_{ij} w_{ij} \log(q_{ij}) + \sum_{\ell \in (B,T)} \sum_{j \in J_\ell} \sum_{k \in K_\ell} (-\log(2\pi) - \log(\sigma_{jk}) - \frac{e_{jk}^2}{2\sigma_{jk}^2}$$

Assume that Stirling's Approximation ($\log(n!) \sim n \log n - n$) is appropriate. Then the log-likelihood is

$$(n \log(n) - n) - (\sum_{ij} (w_{ij} \log w_{ij} - w_{ij})) + \sum_{ij} w_{ij} \log(q_{ij}) + \sum_{\ell \in (B,T)} \sum_{j \in J_\ell} \sum_{k \in K_\ell} (-\log(2\pi) - \log(\sigma_{jk}) - \frac{e_{jk}^2}{2\sigma_{jk}^2})$$

Rearrange terms a bit:

$$n\log(n) - n - \sum_{ij} w_{ij}\log w_{ij} + \sum_{ij} w_{ij} + \sum_{ij} w_{ij}\log(q_{ij}) + \sum_{\ell\in(B,T)}\sum_{j\in J_\ell}\sum_{k\in K_\ell}(-\log(2\pi) - \log(\sigma_{jk}) - \frac{e_{jk}^2}{2\sigma_{jk}^2})$$

Make the substitution: $w_{ij} = np_{ij}$

$$n\log(n) - n - n\sum_{ij} p_{ij}(\log n + \log p_{ij}) + n\sum_{ij} p_{ij} + n\sum_{ij} p_{ij}\log(q_{ij}) + \sum_{\ell\in(B,T)}\sum_{j\in J_\ell}\sum_{k\in K_\ell}(-\log(2\pi) - \log(\sigma_{jk}) - \frac{e_{jk}^2}{2\sigma_{jk}^2})$$

Rearrange:

$$(n\log n)(1 - \sum_{ij} p_{ij}) - n(1 - \sum_{ij} p_{ij}) - n\sum_{ij} p_{ij}\log\frac{p_{ij}}{q_{ij}} + \sum_{\ell\in(B,T)}\sum_{j\in J_\ell}\sum_{k\in K_\ell}(-\log(2\pi) - \log(\sigma_{jk}) - \frac{e_{jk}^2}{2\sigma_{jk}^2})$$

And, since $\sum_{ij} p_{ij} = 1$, a whole bunch of things cancel out:

$$-n\sum_{ij} p_{ij}\log\frac{p_{ij}}{q_{ij}} + \sum_{\ell\in(B,T)}\sum_{j\in J_\ell}\sum_{k\in K_\ell}(-\log(2\pi) - \log(\sigma_{jk}) - \frac{e_{jk}^2}{2\sigma_{jk}^2})$$

Maximizing this will be equivalent to maximizing:

$$-n\sum_{ij} p_{ij}\log\frac{p_{ij}}{q_{ij}} - \sum_{\ell\in(B,T)}\sum_{j\in J_\ell}\sum_{k\in K_\ell}\frac{e_{jk}^2}{2\sigma_{jk}^2}$$

Or, if you rather, use design weights $d_{ij} = Nq_{ij}$, and population weights $w'_{ij} = Np_{ij}$

$$-\frac{n}{N}\sum_{ij} w'_{ij}\log\frac{w'_{ij}}{d_{ij}} - \sum_{\ell\in(B,T)}\sum_{j\in J_\ell}\sum_{k\in K_\ell}\frac{e_{jk}^2}{2\sigma_{jk}^2}$$

which is the form given in Nagle et al 2013

Some Matrix Notation

In a simple matrix notation, the pynophylactic contraints for tracts and block groups are:

$$Y_T = A_T w' X_T + e_T$$

and

$$Y_B = A_B w' X_B + e_B,$$

where e_T and e_B are error terms with variance V_T and V_B.

See http://rpubs.com/nnnagle/PMEDM_1 (http://rpubs.com/nnnagle/PMEDM_1) for a visual demonstration of these matrices.

Rewrite these at

$$\mathrm{vec}(Y_T) = \mathrm{vec}(A_T w' X_T) = (X'_T \otimes A_T)\mathrm{vec}(w')$$

and similarly

$$\mathrm{vec}(Y_B) = \mathrm{vec}(A_B w' X_B) = (X_B' \otimes A_B)\mathrm{vec}(w')$$

This means that

$$\begin{bmatrix} \mathrm{vec}(Y_T) \\ \mathrm{vec}(Y_B) \end{bmatrix} = \begin{bmatrix} (X_T' \otimes A_T) \\ (X_B' \otimes A_B) \end{bmatrix} \mathrm{vec}(w')$$

or (really overloading the tilde operator!!!)

$$\tilde{Y} = \tilde{X}'\tilde{w}$$

Similarly, we may rewrite the maximum likelihood objective as

$$-n\mathrm{vec}(p)' \log(\mathrm{vec}(p)) - 0.5\mathrm{vec}(e)'(\mathrm{diag}(\mathrm{vec}(\sigma^{-2})))\mathrm{vec}(e)$$

Or, redefining everying as a vector/matrix $-np' \log p - .5e'\Sigma^{-1} e$, where Σ is the variance-covariance matrix.

The Primal Problem

Formulation

The Max Entropy Problem is max:

$$L \sim -n\tilde{p}' \log \tilde{p} - .5\tilde{e}'\Sigma^{-1} \tilde{e}$$

subject to

$$\sum_{ij} p_{ij} = 1$$

and

$$\tilde{X}'\tilde{p} = \tilde{Y}/n + \tilde{e}/n$$

The Lagrangian

I’m going to stop with the tilde’s for now. Everything is ‘tilde’ now. Solving the constrained problem is equivalent to the unconstrained problem

$$L = np' \log p/q - n\lambda(X'p - Y/N - e/N) - .5e'\Sigma^{-1} e - n\mu(1'p - 1)$$

The Gradient

The derivative of the log likelihood is

$$\frac{dL}{dp} = -n \log p - n + n \log q - nX\lambda - n\mu$$

$$\frac{dL}{d\lambda} = -nX'p + nY/N + ne/N$$

$$\frac{dL}{de} = n\lambda/N - \Sigma^{-1} e$$

The Solution

Solve for p

$$\log p - \log q = -X\lambda - 1 - \mu$$

$$p/q = \exp(-X\lambda)\exp(-1-\mu)$$

$$p = \frac{q \odot \exp -X\lambda}{q' \exp(-X\lambda)}$$

Solve for e

$$e = \Sigma\lambda n/N$$

Forming the dual problem

Substitute $p(\lambda)$ and $e(\lambda)$ into the objective, and minimize rather than maximize

$$n^{-1} M(\lambda, p(\lambda)) = -p(\lambda)' \log\left(\frac{q \odot \exp(-X\lambda)}{q(q' \exp(-X\lambda))}\right) - .5\frac{n}{N^2}\lambda'\Sigma\lambda$$

which, breaking apart the fraction in the logarithm, is:

$$p(\lambda)'X\lambda + p(\lambda)'1 \log(q' \exp(-X\lambda)) - .5\frac{n}{N^2}\lambda'\Sigma\lambda$$

Substitue $p'X = Y'/N + e'/N$ and $p'1 = 1$:

$$(Y'/N + e'/N)\lambda + \log(q' \exp(-X\lambda)) - .5\frac{n}{N^2}\lambda'\Sigma\lambda$$

Substitute $e' = \lambda'\Sigma n/N$

$$(Y'/N + n/N^2\lambda'\Sigma)\lambda + \log(q' \exp(-X\lambda)) - .5\frac{n}{N^2}\lambda'\Sigma\lambda$$

And collect terms

$$n^{-1} M(\lambda) = Y'\lambda/N + \log(q' \exp(-X\lambda)) + .5\frac{n}{N^2}\lambda'\Sigma\lambda$$

This is the Dual objective! The inputs are q, X, Y/N and $n\Sigma/N^2$

The Dual Gradient

The differential

$$n^{-1} dM = (Y'/N + \lambda'\Sigma n/N^2)(d\lambda) + \frac{1}{q' \exp(-X\lambda)} q' \odot \exp(-X\lambda)'(-X)(d\lambda)$$

$$n^{-1}\frac{dM}{d\lambda} = (Y/N + \Sigma\lambda n/N^2) - X'p$$

This is one of the nice things about MaxEnt: it's gradient is so easy to calculate!

The Dual Hessian

Calculate the Differential of the Gradient

$$n^{-1}\, d\frac{dM}{d\lambda} = \Sigma n/N^2(d\lambda) - X' dp$$

$$dp = -\frac{q \odot \exp(-X\lambda)}{(q' \exp(-X\lambda))^2}(q \odot \exp(-X\lambda))'(-Xd\lambda) + \frac{q \odot \exp(-X\lambda)}{q' \exp(-X\lambda)}(-Xd\lambda)$$

Which means that

$$n^{-1}\, \frac{d^2M}{d\lambda d\lambda'} = -(X'p)(p'X) + X'\mathrm{diag}(p)X + \Sigma n/N^2$$

This is a pain: - $\Sigma n/N^2$ is diagonal (trivial) - $X'\mathrm{diag}(p)X$ is sparse (easy) - $-(X'p)(p'X)$ is rank one and dense (boo. hiss)

In linear algebra terms, this is a sparse matrix with a rank-1 downdate.

In it's full glory, it is:

$$\begin{bmatrix} (X_T' \otimes A_T) \\ (X_B' \otimes A_B) \end{bmatrix} \mathrm{diag}(\mathrm{vec}(p')) \begin{bmatrix} (X_T' \otimes A_T) \\ (X_B' \otimes A_B) \end{bmatrix}' + \Sigma n/N^2 - \begin{bmatrix} (X_T' \otimes A_T) \\ (X_B' \otimes A_B) \end{bmatrix} \mathrm{vec}(p')\mathrm{vec}(p')' \begin{bmatrix} (X_T' \otimes A_T) \\ (X_B' \otimes A_B) \end{bmatrix}'$$

```
#Stuff in r
x<-3
```