STAT 770 Nov. 16 Lecture 22

Iterative Proportional Fitting and Survey Raking

Reading and Topics for this lecture: Chapter 9 Sections 9.7.2-9.7.4, Wikipedia article on "Raking", web sources on "Alternating Minimization".

(1) Iterative Proportional Fitting Algorithm (Deming & Stephan)

(2) Origin in Maximum Likelihood with Constrained Parameters

(3) Survey Weight-Adjustment with Calibration Conditions

(4) "Proof" of Convergence

Iterative Proportional Fitting (Sec.9.7.2-9.7.4)

Start with counts (summed weights, in survey contexts) $\mu_a^{(0)}$

$a \leftrightarrow (i, j, k, \cdots)$ to be marginalized (some factor indices summed)

- restrict to example (i,j,k) in model (XZ, W)
- target counts $\mu_{xz+} = Y_{xz+}$, $\mu_{++w} = Y_{++w}$
- in iterative passes $m \geq 0$: hit targets exactly by multiplication using respective factors f_{xz}, f_w

$$\mu_{ijk}^{(2m+1)} = \mu_{ijk}^{(2m)} \frac{Y_{ij+}}{\mu_{ij+}^{(2m)}} \ , \quad \mu_{ijk}^{(2m+2)} = \mu_{ijk}^{(2m+1)} \frac{Y_{++k}}{\mu_{++k}^{(2m+1)}}$$

Convergence occurs under general conditions (usually quickly) when all $\mu_a^{(0)}$ and marginals are positive

MLE in Loglinear Models

Recall:

• loglinear models are natural exponential families with coefficients λ^{XZ} for included interactions appearing in logLik multiplying sufficient statistics Y_{xz+}

• in natural exponential families, MLEs are Generalized Method of Moments solutions of equations $E_\lambda(T_k) = T_k$. Thus the (unique, if they exist) ML solutions have cell means $\mu_{xzw}(\lambda)$ satisfying $\mu_{xz+} = Y_{xz+}$, etc.

• if H is the design matrix of dummy columns, then the Poisson regression logLik to be maximized is

$$\sum_a Y_a \, (H\beta)_a - \sum_a \text{exp}((H\beta)_a)$$

Reformulation of the Minimization Problem

Using the information on the last slide: MLE $\widehat{\beta}$ solves

Loglinear ML Problem: $\min_{\mu_a = \exp((H\beta)_a)} \sum_a \left\{ \mu_a - Y_a \log(\mu_a) \right\}$

subject to all $\mu_{xz+} = Y_{xz+}$, etc. (all terms like XZ in model)

Rewrite constraints: (*) $\sum_{a=1}^{M} (Y_a - \mu_a)\, H_{a,b} = 0, \quad b = 1, \ldots, d$

In the loglinear ML: rewrite objective function as

$$\sum_{a=1}^{M} \left\{ \mu_a - \mu_a \log(\mu_a) \; - \; (Y_a - \mu_a)(H\beta)_a \right\}$$

Under the constraint (*), last term drops out, and ML problem is equivalent to:

$\min_{\mu_a = \exp((H\beta)_a)} \sum_a \left\{ \mu_a - \mu_a \log(\mu_a) \right\}$ subject to (*)

Alternative Form of ML Problem

Alternative Problem: $\min_{\mu_a} \sum_a \left\{ \mu_a - \mu_a \log(\mu_a) \right\}$
over all $\mu_a > 0$ subject to same constraints (*)

The restriction on parametric form of μ_a has been dropped. In the loglinear ML problem the constraints were redundant; in the 2nd problem the assumption of loglinear form is redundant!

Assume that the marginals $\sum_a Y_a \, H_{a,b}$ are all nonzero in what follows, and also in the IPF/Raking iterations on first slide. Find additional discussion of tables with sparseness and 0's in Agresti Sec. 10.6.2.

Lagrange Multipliers for the Alternative Problem

Use Lagrange multipliers $\tau \in \mathbb{R}^d$ to define unconstrained problem

$$\min_{\underline{\mu}, \tau} \sum_{a=1}^{M} \left\{ \mu_a - \mu_a \log(\mu_a) - (Y_a - \mu_a)(H\tau)_a \right\} \qquad \textbf{(A)}$$

Setting $\nabla_\tau = 0$ just gives back the constraints (*)

Setting $\partial / \partial \mu_a = 0$ implies $\log(\mu_a) = (H\tau)_a$

This shows the loglinear form is the only possible form for the minimizer of the alternative problem, so the loglinear ML problem is equivalent to it!

We also find: the Lagrange multipliers τ are precisely the λ log-linear-model coefficients (in which side-conditions prevent redundant coefficients).

Generalized Problem – Survey Calibration

(Multi-way) table with cells a and aggregated design weights d_a (from units i collected in a survey cross-classified and found to fall in cell a). **Question:** how to perturb these weights cell by cell as little as possible to achieve weights so that for each of a set of dummy columns $\{H_{a,b},\ a = 1,\ldots,M\}$ indexed by $b = 1,\ldots,d$

$$(\dagger) \quad \sum_{a=1}^{M} w_a\, H_{a,b} = t_b \quad \text{known totals,} \quad b = 1,\ldots,d$$

Formal Statement: $\min_{\underline{w}} \sum_{a=1}^{M} d_a\, G(w_a/d_a)$ subject to $(\dagger)$

where $G(z)$ is a function with $G(1) = 0$, $G'(1) = 0$, $G'' > 0$.

Examples: (a) $G(z) = (z-1)^2/2$ **linear calibration**,

(b) $G(z) = z \log z - z + 1$ **raking calibration**

Deville and Särndal 1992 Jour. Amer. Statist. Assoc.

Survey Calibration & Generalized Raking, cont'd

In contingency-table iid data setting, $d_a = Y_a$ are the observed counts, and $t_b = \sum_{a=1}^{M} w_a \, H_{a,b}$ the desired marginal totals.

'Marginal totals' may overlap information about more than one factor, e.g. `SEX` $\times$ `RACE` and `SEX` $\times$ `AGE-Group` as separate t_b.

Lagrangian objective: $\sum_a d_a \, G(\frac{w_a}{d_a}) + \sum_{b=1}^{d} \gamma_b \left(\sum_{a=1}^{M} w_a H_{a,b} - t_b \right)$

Gradient equation: $G'(w_a/d_a) = (H\gamma)_a \;\Rightarrow\; w_a = d_a \cdot (G')^{-1}((H\gamma)_a)$

In raking example (b): $G'(z) = \log z \;\Rightarrow\; w_a = d_a \, e^{(H\gamma)_a}$

In surveys may have $M \sim 10^4$, $d \sim 30$; resulting weight-vector $w \in \mathbb{R}^M$ is **parametric**, loglinear with $\log d_a$ as **offsets**.

Features of the IPF Algorithm

Key observation: if the algorithm converges, the only possible limit satisfies the constraints (*)

So if convergence is shown, limit is unique fitted loglinear ML.

Sketch Proof of Convergence. Associate subvector β_{sub} with columns $H_{\cdot,b}$ for $b \in$ sub for single marginal factor (eg XZ).

Take gradient with respect to β_{sub} of parametric $-logLik$.

$$\nabla_{\beta_{sub}} \sum_{a=1}^{M} \left\{ e^{(H\beta)_a} - Y_a (H\beta)_a \right\} = \sum_{a=1}^{M} H_{a,sub} \left(\mu_a - Y_a\right)$$

Alternating Minimization and Raking

Setting this to 0 finds partial minimum (wrt β_{sub}). Doing this repeatedly (alternating minimization) converges to minimum of convex objective function.

Each raking pass finds a unique partial minimum for the overall convex $-logLik$

Each raking pass preserves the form $\log \mu_a = (H\beta)_a$,
maps subvector $\beta_{sub} \mapsto \beta^{(*)}_{sub} = \beta_{sub} + \delta_{sub}$
achieving constraint $\sum_{a=1}^{M} (Y_a - \exp((H\beta^{(*)})_a) H_{a,b} = 0,\ \ b \in \text{sub}.$

Next Lecture – GLMs with Random Effects

This is material from Chapter 13, also see the Technical Report & Handout linked at Handouts topic (1)(ii) on the STAT 770 course web-page.

General framework: observations $Y_{a,c}$ now observed for cells a in clusters c

$P(Y_{a,c} = y \mid \epsilon_c)$ a GLM with predictors $X_{a,c}$, and shared family, link and parameters β and offset ϵ_c, where ϵ_c are iid cluster-effects, usually $\mathcal{N}(0, \sigma_e^2)$