Multinomial, Interval-Censoring, and Smoothing

Jutta Gampe

May 22, 2018

1 Notation

- ▷ Number of observations (individuals): n
  Number of bins: m
  Number of columns in B-spline matrix: r
  Number of covariates (in PH-regression): p
- ▷ Smoothing parameter: λ
  Lagrange multiplier: κ
- ▷ B-spline basis: matrix $B \in \mathbb{R}^{n\times r}$
  Coefficients B-splines: $\alpha \in \mathbb{R}^r$
  (if $B = I$, then $r = m$)
- ▷ Matrix covariates: $X \in \mathbb{R}^{n\times p}$
  Regression parameters: $\beta \in \mathbb{R}^p$

2 Density estimation: Grouped data and smoothing

2.1 Grouped multinomial distribution, standard EM

Start with a multinomial distribution over m categories with probabilities $\pi = (\pi_1, \ldots, \pi_m)^{\mathrm{T}}$ and $\sum_{j=1}^m \pi_j = 1$. Observed are n individuals and the data (in their raw format) are indicators I_{ij} which are equal to 1, if observation i is in category j, and zero otherwise. Clearly the row sums $\sum_j I_{ij} = 1$ and $\sum_i \sum_j I_{ij} = n$. The column sums $y_j = \sum_{i=1}^n I_{ij}$ are the numbers of observations in category j; they are the sufficient statistics for the parameter vector π, see below.

The log-likelihood is

$$\ell(\pi) = \ln L(\pi) = \ln\left(\prod_{i=1}^n \prod_{j=1}^m \pi_j^{I_{ij}}\right) = \sum_{i=1}^n \sum_{j=1}^m I_{ij} \ln \pi_j = \sum_{j=1}^m y_j \ln \pi_j \tag{1}$$

subject to $\sum_{j=1}^m \pi_j = 1$.

Introducing a Lagrange multiplier κ, we obtain

$$\ell(\pi) = \sum_{i=1}^n \sum_{j=1}^m I_{ij} \ln \pi_j - \kappa\left(\sum_{j=1}^m \pi_j - 1\right) \tag{2}$$

and the first derivatives

$$\frac{\partial \ell}{\partial \pi_s} = \sum_{i=1}^{n} I_{is} \frac{1}{\pi_s} - \kappa \stackrel{!}{=} 0 \tag{3}$$

Multiplying by π_s and summing over all categories we have

$$\sum_{s=1}^{m} \sum_{i=1}^{n} I_{is} = \kappa \sum_{s=1}^{m} \pi_s \quad \Rightarrow \quad \kappa = n. \tag{4}$$

Equations (3) & (4) lead to the standard multinomial MLE

$$\hat{\pi}_s = \frac{\sum_i I_{is}}{n} = \frac{y_s}{n}. \tag{5}$$

All that is needed are the counts y_s in each of the m bins, the sufficient statistics.

If observations are censored, they are only known to lie in a particular subset A_i of the categories $\{1, \ldots, m\}$ (notation inspired by Turnbull's 1976 paper). The censoring set A_i can differ across individuals.

The indicators I_{ij} are now replaced by values c_{ij}, but here several of the c_{ij} have a value 1 in each row, indicating the categories that observation i might be found in.[1]

Applying the EM algorithm: In the E-step we determine the expected values of the sufficient statistics, given the current values of the π_j and the data. In more detail, instead of the observed I_{ij} we calculate the $\tilde{I}_{ij} = E(I_{ij} \,|\, \text{data, current } \pi)$, which for current values $\pi_j^{(l)}$ are

$$\tilde{I}_{ij}^{(l)} = \frac{c_{ij} \pi_j^{(l)}}{\sum_{k=1}^{m} c_{ik} \pi_k^{(l)}}. \tag{6}$$

Each $\tilde{I}_{ij}^{(l)}$ gives the expected value of observation i to lie in category j in view of the current values $\pi^{(l)}$ and the 'data' c_{ij}. Summing over the rows for each category completes the E-step:

$$\tilde{y}_j^{(l)} = \sum_{i=1}^{n} \tilde{I}_{ij}^{(l)} \,. \tag{7}$$

The M-step simply is

$$\pi_j{}^{(l+1)} = \frac{\tilde{y}_j^{(l)}}{n} \,. \tag{8}$$

[1] We could also use $\tilde{c}_{ij} = c_{ij} / \sum_j c_{ij}$, that is, we normalize the values c_{ij} by their row sum $c_{i+} = \sum_j c_{ij}$. The $\tilde{c}_{ij}$ are the (uniformly) distributed probabilities for observation i to lie in category j. This is appealing if we we think about non-uniform uncertainty distributions. However, since these normalization constants cancel in the subsequent calculations, see equation (6), we stick to the c_{ij}.

Combining the E- and the M-step we have the following iteration:

$$\pi_j{}^{(l+1)} = \frac{1}{n} \sum_{i=1}^{n} \frac{c_{ij}\pi_j^{(l)}}{\sum_{k=1}^{m} c_{ik}\pi_k^{(l)}} \,. \tag{9}$$

The same iteration can be obtained without an explicit reference to the EM-setup: Using the same notation as before, the probability that the (censored) observation i contributes to the likelihood is

$$p_i = \sum_{j=1}^{m} c_{ij}\pi_j \,.$$

Adding a Lagrange term as before, the log-likelihood is

$$\ell(\pi) = \sum_{i=1}^{n} \ln p_i - \kappa \left(\sum_{j=1}^{m} \pi_j - 1 \right) \,.$$

First derivatives are

$$\frac{\partial \ell}{\partial \pi_s} = \sum_{i=1}^{n} \frac{c_{is}}{\sum_{j=1}^{m} c_{ij}\pi_j} \; - \; \kappa \,.$$

Multiplying by π_s and summing over s we obtain

$$\sum_{i=1}^{n}\sum_{s=1}^{m} \frac{c_{is}\pi_s}{\sum_{j=1}^{m} c_{ij}\pi_j} = \sum_{i=1}^{n} \frac{p_i}{p_i} = n = \kappa \sum_{s=1}^{m} \pi_s = \kappa \qquad \Rightarrow \; \kappa = n$$

and therefrom

$$\frac{1}{n} \sum_{i=1}^{n} \frac{c_{is}\pi_s}{\sum_{j=1}^{m} c_{ij}\pi_j} = \pi_s \,, \tag{10}$$

which leads to the same system as equation (9).

Finally, equation (9) can also be written as

$$\sum_{i=1}^{n} \frac{c_{ij}\pi_j}{\sum_{k=1}^{m} c_{ik}\pi_k} \cdot \frac{n}{n} = \sum_{i=1}^{n} \frac{c_{ij}\gamma_j}{\sum_{k=1}^{m} c_{ik}\gamma_k} = \gamma_j \,, \tag{11}$$

where $\gamma_j = n \cdot \pi_j$ is the expected number of observations in category j in a sample of size n. This version facilitates the comparison with the composite link model (CLM).

2.2 Comparing with the CLM

So is this a composite link model or not? To keep it simple, we do not consider additional smoothing or additional covariates.

In its Poisson version the core equations of the CLM are as follows:

The $y_i, i = 1, \ldots, n$, are independently Poisson with means μ_i and

$$\mu_i = \sum_{j=1}^{m} c_{ij}\,\gamma_j, \quad \gamma_j = e^{\eta_j}, \quad \eta_j = \sum_{k=1}^{p} x_{jk}\beta_k$$

The resulting likelihood equations are, see (3.11) in Eilers (2007),

$$\sum_{i=1}^{n}(y_i - \mu_i)\,\breve{x}_{ik} = 0 \qquad \text{with} \qquad \breve{x}_{ik} = \frac{\sum_{j=1}^{m} c_{ij}x_{jk}\gamma_j}{\mu_i}\,. \tag{12}$$

For $y_i \equiv 1$ and $X = I$, that is, $x_{jk} = 1$ if $j = k$ and $= 0$ otherwise, equation (12) becomes

$$\sum_{i=1}^{n}(1 - \mu_i)\,\frac{c_{ik}\gamma_k}{\mu_i} = 0$$

or

$$\sum_{i=1}^{n}\frac{c_{ik}\gamma_k}{\mu_i} = \sum_{i=1}^{n} c_{ik}\gamma_k = \gamma_k \underbrace{\sum_{i=1}^{n} c_{ik}}_{c_{+k}}\,. \tag{13}$$

Comparing (13) with (11) we see that the two are equivalent *only* if the column sums c_{+k} equal 1. This is true, e.g., in a histogram setting with equal non-overlapping groups for all observations. It is not true, however, with individually different and overlapping censoring intervals. The resulting iterations are very similar, but not equivalent.

> *Note! Insight after some hours of derivation: Row sums are not column sums, so the row sums may be 1 (if the $\tilde{c}_{ij}$ are used), but the column sums are not . . .*