Lindsey's Method Review

Stephen Bates

July 2, 2019

Setting

Suppose we have an exponential family with density

$$P_\beta(x) = e^{\beta^\top T(x) - \Lambda(\beta)} m(x)$$

where $T(x) \in \mathbb{R}^p$ is a vector of sufficient statistics and $\beta \in \mathbb{R}^p$ is the unknown parameter of interest. We often wish to estimate β using maximum likelihood, which is possible in principle since the log-likelihood is concave, as exponential family theory tells us that $\Lambda(\beta)$ is convex. This means any convex optimization solver will work, even something as simple as gradient descent. However, in some cases we may not know $\Lambda(\beta)$ or it may be difficult to evaluate it or its derivatives. **Lindsey's method is a technique for estimating the parameter β, even when the CGF Λ is unknown.**

Step 1: discretizing the likelihood

The first step in Lindsey's method is discretizing the likelihood. Take a partition $B_1, \ldots, B_K$ of the sample space, with associated measures $m(B_1), \ldots, m(B_K)$. Let $N_1, \ldots, N_K$ be the number of observations in these bins, respectively, and let N be the total number of observations. Now after this discretization, we have the probability model:

$$(N_1, \ldots, N_k) \sim \text{mult}(N, (\pi_\beta(B_1), \ldots, \pi_\beta(B_k)))$$

with

$$\begin{aligned}\pi_\beta(B_j) &= e^{\beta^\top T(x) - \Lambda(\beta)} m(x) \\ &\approx m(B_j) e^{\beta^\top T(x_j) - \Lambda(\beta)}\end{aligned}$$

where x_j is some point in B_j, usually taken to be the center. We then have

$$\pi_\beta(B_j) \approx \tilde{\pi}_\beta(B_j) := \frac{m(B_j) e^{\beta^\top T(x_j)}}{\sum_{k=1}^K m(B_j) e^{\beta^\top T(x_k)}}.$$

Importantly, in the final expression, the $\Lambda(\beta)$ term has canceled out from both the numerator and denominator. In what follows, we will fit the discretized approximate model

$$(N_1, \ldots, N_k) \sim \text{mult}(N, \tilde{\pi}_\beta(B_j)), \tag{1}$$

because the likelihood of this model no longer depends on $\Lambda(\beta)$.

Step 2: fitting with Poisson regression

Next, we fit the model (1) using Poisson regression. For convenience, we will use the Poisson trick, i.e. we take the model $N \sim \text{pois}(\lambda)$. In that case, we can write the model as

$$N_j \sim \text{pois}(\lambda \tilde{\pi}_\beta(B_j)), \tag{2}$$

independently for $j = 1, \ldots, k$, and then we will reparameterize as

$$N_j \sim \text{pois}(e^\alpha m(B_j) e^{\beta^\top T(x_j)}),$$

for convenience. This gives us the likelihood

$$l(\beta, \alpha) = \sum_{j=1}^K N_j(\beta^\top T(x_j) + \alpha + \log(m(B_j))) - \sum_{j=1}^K e^\alpha m(B_j) e^{\beta^\top T(x_j)}. \tag{3}$$

Solving for α gives $e^{\hat{\alpha}} = \frac{N}{\sum_{j=1}^K m(B_j) e^{\beta^\top T(x_j)}}$. We then obtain

$$\begin{aligned} l(\beta, \hat{\alpha}) &= \sum_{j=1}^K N_j(\beta^\top T(x_j) + \log(m(B_j))) - N \log(\sum_{j=1}^K m(B_j) e^{\beta^\top T(x_j)}) \\ &= \sum_{j=1}^K N_j \log(\tilde{\pi}_\beta(B_j)). \end{aligned}$$

This last expression is also the log-likelihood of the multinomial model (1), so we see that MLE of β from (3) is exactly the same as the MLE of β from (1). Furthermore, the MLE can be obtained from (3) using standard Poisson regression software, such as the `glm` command in R.

Note that unlike the point estimates, the confidence intervals from the models and (1) and (2) are not exactly the same, since the CIs from the poisson regression are also taking into account the variability in N. For large enough N, however, the CIs for β will be approximately the same from either the binomial model or the Poisson model.