Stat 8931 (Aster Models)
Lecture Slides Deck 8

Charles J. Geyer

School of Statistics
University of Minnesota

June 7, 2015

Conditional Aster Models

A **conditional aster model** is a submodel parameterized

$$\theta = a + M\beta$$

An **unconditional aster model** is a submodel parameterized

$$\varphi = a + M\beta$$

There is a subtle but profound difference.

Conditional Aster Models (cont.)

Both are exponential families, but

- An unconditional aster model is a regular full exponential family.
- A conditional aster model is a curved exponential family.

Curved exponential families have some nice properties (asymptotics always work for sufficiently large sample sizes), but none of the nice properties we talked about for unconditional aster models.

Conditional Aster Models (cont.)

Review. Unconditional aster models have

- concave log likelihood,
- MLE unique if they exist,
- MLE characterized by "observed = expected",
- observed and expected Fisher information the same,
- submodel canonical statistic is sufficient,
- maximum entropy property,
- multivariate monotone relationship between canonical and mean value parameters.

Curved exponential families don't, in general, have any of these properties.

Conditional Aster Models (cont.)

Conditional aster models have two of these

- concave log likelihood,
- MLE unique if they exist.

Conditional Aster Models (cont.)

The log likelihood is (from deck 2)

$$
\begin{aligned}
l(\theta) &= \sum_{j \in J} \bigl[y_j \theta_j - y_{p(j)} c_j(\theta_j) \bigr] \\
&= \langle y, \theta \rangle - \sum_{j \in J} y_{p(j)} c_j(\theta_j)
\end{aligned}
$$

and the conditional canonical affine submodel is

$$
l(\beta) = \langle M^T y, \beta \rangle - \sum_{j \in J} y_{p(j)} c_j(\theta_j)
$$

Conditional Aster Models (cont.)

$$l(\beta) = \langle M^T y, \beta \rangle - \sum_{j \in J} y_{p(j)} c_j(\theta_j)$$

We see we get almost no sufficient dimension reduction.

The likelihood is a function of $M^T y$ and the set of all predecessors. That typically is not a dimension reduction at all (when the dimension of $M^T y$ is more than the number of terminal nodes).

Conditional Aster Models (cont.)

$$l(\theta) = \sum_{j \in J} \left[y_j \theta_j - y_{p(j)} c_j(\theta_j) \right]$$

Each term in square brackets is strictly concave.

The sum of strictly concave functions is strictly concave.

The composition of a strictly concave function and an affine function is strictly concave.

Hence the log likelihood for a conditional canonical affine submodel is strictly concave. Hence the MLE is unique if it exists.

Conditional Aster Models (cont.)

$$l(\theta) = \sum_{j \in J} \left[y_j \theta_j - y_{p(j)} c_j(\theta_j) \right]$$

The observed Fisher information matrix for θ for a saturated aster model is

$$J_{\text{sat}}(\theta) = -\nabla^2 l(\theta)$$

is a diagonal matrix whose j, j component is

$$y_{p(j)} c_j''(\theta_j)$$

where the double prime indicates ordinary second derivative.

Conditional Aster Models (cont.)

The expected Fisher information matrix for θ, denoted $I_{\text{sat}}(\theta)$, is the expectation of the observed Fisher information matrix.

So it too is diagonal, and its i, i component is

$$\mu_{p(j)} c_j''(\theta_j)$$

Conditional Aster Models (cont.)

Then conditional canonical affine submodel observed and expected Fisher information matrices are

$$
\begin{aligned}
J(\beta) &= M^T J_{\text{sat}}(a + M\beta) M \\
I(\beta) &= M^T I_{\text{sat}}(a + M\beta) M
\end{aligned}
$$

Conditional Aster Models (cont.)

The maximum entropy argument only works for full exponential families, not for curved exponential families.

Conditional Aster Models (cont.)

We do have the saturated model multivariate monotone relationships $\mu \longleftrightarrow \varphi$ and $\xi \longleftrightarrow \theta$.

But that doesn't tell us anything about canonical affine submodels.

Conditional Aster Models (cont.)

Unconditional canonical affine submodels have the property that changing φ_j changes θ_k for all $k \succ j$.

Conditional canonical affine submodels do not have this property. Changing θ_j only changes θ_j.

Thus conditional canonical affine submodels tend to need many more parameters to fit adequately.

Conditional Aster Models (cont.)

So if conditional canonical affine submodels don't have any nice properties, why do they even exist?

One reason is just because they do exist as abstract mathematical objects, and they weren't that much extra code to implement, and — who knows? — maybe they will find an important use someday.

Just because they exist does not mean we actually recommend them for anything.

Conditional Aster Models (cont.)

One issue that would be a good reason to use conditional aster models is if you do not want the property that changing φ_j changes θ_k for all $k \succ j$.

For example, you might require that the conditional expectation of survival given survival to the previous year be the same for all years.

An unconditional aster model wouldn't do that, but a conditional aster model could.