

Stat 8931 (Aster Models)
Lecture Slides Deck 8
Conditional Aster Models

Charles J. Geyer

School of Statistics
University of Minnesota

October 3, 2018

- The version of R used to make these slides is 3.5.1.
- The version of R package aster used to make these slides is 1.0.2.
- This work is licensed under a Creative Commons Attribution-ShareAlike 4.0 International License (<http://creativecommons.org/licenses/by-sa/4.0/>).

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.)

The log likelihood is (from deck 2)

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

and the conditional canonical affine submodel is

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

On the right-hand side θ is a function of β through $\theta = a + M\beta$ even though the notation does not explicitly indicate this.

Conditional Aster Models (cont.)

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

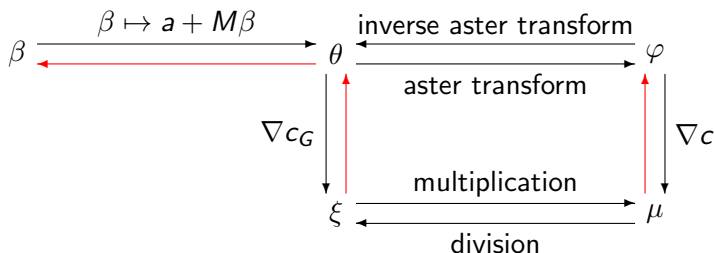
We see we get almost no sufficient dimension reduction.

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).

Because conditional aster models do not have the sufficient dimension reduction property, there is no submodel canonical sufficient statistic.

A Plethora of Parameterizations (cont.)

A conditional aster model only has five parameterizations.



Like with unconditional aster models, all of the parameters and arrows in the square on the right are the same as for saturated aster models.

Also like with unconditional aster models, if we know that θ has the form $\theta = a + M\beta$ for some β , then we know or can find that β and that defines the red horizontal arrow.

A Plethora of Parameterizations (cont.)

Unlike the case with unconditional aster models where the MLE for each of the six parameterizations $(\hat{\beta}, \hat{\tau}, \hat{\varphi}, \hat{\mu}, \hat{\theta}, \hat{\xi})$ is a vector sufficient statistic, with conditional aster models — because they do not have the sufficient dimension reduction property — the MLE for **no** parameterization is a vector sufficient statistic.

Conditional Aster Models (cont.)

Conditional aster models do have two of the aforementioned properties of regular full exponential families

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

(They do not have any of the other properties.)

Conditional Aster Models (cont.)

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

Each term in square brackets is concave and strictly concave if there are no multinomial dependence groups.

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 concave and strictly concave if there are no multinomial dependence groups. Hence the MLE is unique if it exists in case of no multinomial dependence groups.

Conditional Aster Models (cont.)

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

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

$$J(\beta) = M^T J_{\text{sat}}(a + M\beta)M$$

$$I(\beta) = M^T I_{\text{sat}}(a + M\beta)M$$

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. The preceding sentence was in the 2013 version of the course slides, and we have preserved it to show that things change.

We have since found a situation where unconditional aster models do not work and conditional aster models do.

Conditional Aster Models (cont.)

A paper by Shaw, Wagenius, and Geyer (*Journal of Ecology*, 2015) uses unconditional aster models for some analyses but also uses conditional aster models for a situation where unconditional aster models do not work.

Conditional Aster Models (cont.)

Unconditional aster models do not work — they cannot be scientifically interpreted — when there are **time-dependent covariates**.

The reason is that the aster transform means that increasing φ_j holding other components of φ fixed changes not only θ_j but also $\theta_{p(j)}$, $\theta_{p(p(j))}$, $\theta_{p(p(p(j)))}$, and so forth. (This was discussed in deck 3, slides 75 ff.)

And this means that — in an unconditional aster model — it is impossible for a time-dependent covariate to act at a given time. If it acts at node j , then it also acts at node $p(j)$, node $p(p(j))$, node $p(p(p(j)))$, and so forth.

Thus if one has time-dependent covariates one must use a conditional aster model.

Conditional Aster Models (cont.)

The issue Shaw, et al. (2015) were interested in was whether aphid load (aphids are herbivores of echinacea plants) in a specific year was related to components of fitness expressed in the following year (and of course this is for each year aphid load was measured).

The issue was complicated by aphid choice. A quote from the abstract of that paper

Further, flowering individuals generally harboured more aphids than non-flowering plants. In analyses of overall plant fitness, within each genotypic class, fitness was greatest for plants with the greatest aphid-loads, consistent with the preference of aphids for flowering individuals.

So the fact that aphid load was not a “treatment” controlled by the experimenters means we have a correlation is not causation problem.

Conditional Aster Models (cont.)

Nevertheless, the conditional aster analysis was suggestive.

Another quote

To distinguish the role of aphid choice from the effect of aphid herbivory in the relationship between plant fitness and aphid-load, we evaluated how components of fitness varied with prior aphid-load. Notably, [inbred] plants with high aphid-loads the previous year produced far fewer achenes per flower head than those that carried fewer aphids.

Conditional Aster Models (cont.)

Another issue that would be a good reason to use conditional aster models is if one wanted some form of **stationarity** in an aster model.

If one wanted for some reason that components of ξ and hence of θ do not change over time for a certain kind of node, then conditional aster models can be made to have this property but unconditional aster models cannot.

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