Aster Models for Life History Analysis

Charles J. Geyer
School of Statistics
University of Minnesota

Stuart Wagenius
Institute for Plant Conservation Biology
Chicago Botanic Garden

Ruth G. Shaw
Department of Ecology, Evolution, and Behavior
University of Minnesota

http://www.stat.umn.edu/geyer/aster/
http://echinacea.umn.edu/

The Organism

Echinacea angustifolia, common name purple coneflower. It is in the aster family *Asteraceae*.

Photo by Jennifer Ison

http://echinacea.umn.edu/people.htm

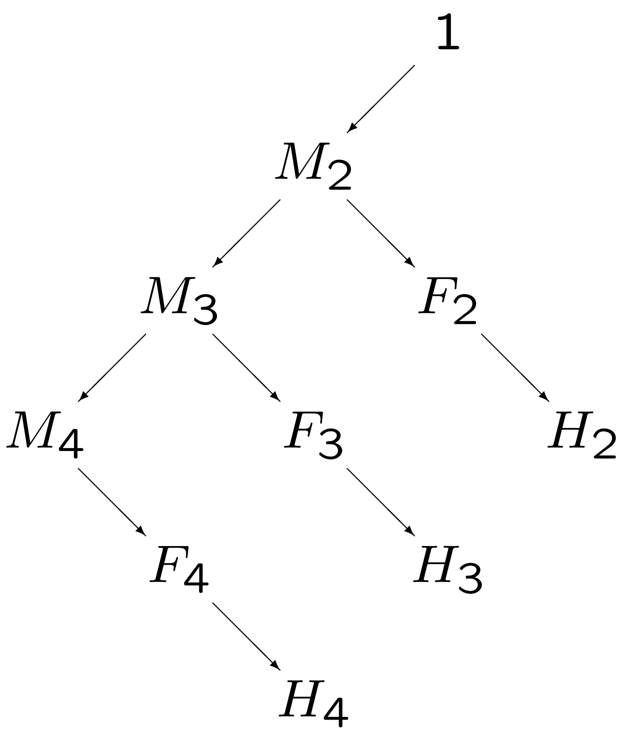
Another Organism

Salvia lyrata, common name lyre-leaf sage. It is in the mint family *Lamiaceae*.

Photo by R. Harrison Wiegand, Maryland Department of Natural Resources, Wildlife and Heritage Service.

http://www.nps.gov/plants/pubs/chesapeake/plant/325.htm

The Graphical Model



Graph for *Echinacea* aster data.

Arrows go from parent to child node. Nodes labels are associated variables. Only founder node is constant 1.

M_j is *mortality status*, F_j is *flowering status*, and H_j is *flower head count* in year $2000 + j$.

M_j and F_j are Bernoulli conditional on parent variable being one (and zero otherwise). H_j are zero-truncated Poisson conditional on parent variable being one (and zero otherwise).

A Little Bit of Exponential Family Theory

A statistical model is an *exponential family* if the log likelihood has the form

$$l(\boldsymbol{\varphi}) = \langle \mathbf{x}, \boldsymbol{\varphi} \rangle - \psi(\boldsymbol{\varphi})$$

where

- $\mathbf{x}$ is a vector statistic, called the *canonical statistic*.
- $\boldsymbol{\varphi}$ is a vector parameter, called the *canonical parameter*.
- $\langle \mathbf{x}, \boldsymbol{\varphi} \rangle$ is an inner product.
- $\psi(\boldsymbol{\varphi})$ is a function of the parameter only, called the *cumulant function* of the family.

A Little Bit of Exponential Family Theory: Example

Binomial Distribution

$$
\begin{aligned}
l(p) &= x \log(p) + (n - x) \log(1 - p) \\
&= x \log\left(\frac{p}{1 - p}\right) + n \log(1 - p)
\end{aligned}
$$

This is now exponential family form with canonical statistic x, canonical parameter

$$
\varphi = \log\left(\frac{p}{1 - p}\right)
$$

and cumulant function

$$
\begin{aligned}
\psi(\varphi) &= -n \log(1 - p) \\
&= n \log\bigl(1 + \exp(\varphi)\bigr)
\end{aligned}
$$

A Little Bit of Exponential Family Theory: Examples

Binomial Distribution

Poisson Distribution

Zero-Truncated Poisson Distribution

Normal Distribution

Conventional (Normal Theory) Linear Models

Generalized Linear Models (if not quasi-likelihood)

Log-Linear Models

A Little Bit of Exponential Family Theory (Cont.)

Independent and identically distributed (IID) sampling induces another exponential family.

$$l_n(\boldsymbol{\varphi}) = \left\langle \sum_{i=1}^{n} \mathbf{x}_i, \boldsymbol{\varphi} \right\rangle - n\psi(\boldsymbol{\varphi})$$

New canonical statistic is $\sum_{i=1}^{n} \mathbf{x}_i$.

New canonical parameter is same as old $\boldsymbol{\varphi}$.

New cumulant function is n times old cumulant function.

A Little Bit of Exponential Family Theory (Cont.)

Differentiation under the integral sign (Bartlett identities) gives

$$\begin{aligned} E_\varphi\Bigl\{\nabla l(\varphi)\Bigr\} &= 0 \\ \operatorname{var}_\varphi\Bigl\{\nabla l(\varphi)\Bigr\} &= -E_\varphi\Bigl\{\nabla^2 l(\varphi)\Bigr\} \end{aligned}$$

Applied to exponential family gives

$$\begin{aligned} E_\varphi(\mathbf{X}) &= \nabla \psi(\varphi) \\ \operatorname{var}_\varphi(\mathbf{X}) &= \nabla^2 \psi(\varphi) \end{aligned}$$

This allows us to rewrite the likelihood equations as

$$\nabla l(\varphi) = \mathbf{x} - E_\varphi(\mathbf{X}) = 0$$

The MLE is found by setting "observed equals expected" for the canonical statistic vector.

A Little Bit of Exponential Family Theory (Cont.)

$\boldsymbol{\tau} = E_{\boldsymbol{\varphi}}(\mathbf{X}) = \nabla \psi(\boldsymbol{\varphi})$ is called the *mean value parameter*.

Map $\nabla \psi : \boldsymbol{\varphi} \mapsto \boldsymbol{\tau}$ is invertible because its Jacobian matrix

$$\nabla^2 \psi(\boldsymbol{\varphi}) = \text{var}_{\boldsymbol{\varphi}}(\mathbf{X})$$

is everywhere positive definite.

This map is also monotone

$$\frac{\partial \tau_j}{\partial \varphi_j} = \frac{\partial}{\partial \varphi_j} E_{\boldsymbol{\varphi}}(X_j) = \frac{\partial^2 \psi(\boldsymbol{\varphi})}{\partial \varphi_j^2} = \text{var}_{\boldsymbol{\varphi}}(X_j) > 0$$

Elementary interpretation: increasing one canonical parameter, holding the others fixed, increases the expectation of the corresponding canonical statistic (and does something, increase, decrease, you don't know which, to the others).

Conditional Aster Model Log Likelihood

$$l(\boldsymbol{\theta}) = \sum_{j \in J} x_j \theta_j - x_{p(j)} \psi_j(\theta_j)$$

where

- J is set of non-founder nodes.
- $p(j)$ is parent of node j.
- ψ_j is cumulant function for one-parameter exponential family for j-th node.

Distribution of x_j given $x_{p(j)}$ is one-parameter exponential family with canonical statistic x_j, canonical parameter θ_j, and sample size $x_{p(j)}$.

Unconditional Aster Model Log Likelihood

$$l(\boldsymbol{\theta}) = \sum_{j \in J} x_j \theta_j - x_{p(j)} \psi_j(\theta_j)$$

Collect terms with same x_j

$$\sum_{j \in J} x_j \left[\theta_j - \sum_{m \in S(j)} \psi_m(\theta_m) \right] - \sum_{j \in S(F)} x_{p(j)} \psi_j(\theta_j)$$

where $S(j)$ is set of children of node j and $S(F)$ is set of children of founders.

Unconditional Aster Model Log Likelihood (Cont.)

$$\sum_{j \in J} x_j \left[\theta_j - \sum_{m \in S(j)} \psi_m(\theta_m) \right] - \sum_{j \in S(F)} x_{p(j)} \psi_j(\theta_j)$$

Recognize *unconditional exponential family* with new *canonical statistic vector* same as old (components x_j), new *canonical parameter vector* with components

$$\varphi_j = \theta_j - \sum_{m \in S(j)} \psi_m(\theta_m), \qquad j \in J,$$

and new *cumulant function*

$$\psi(\boldsymbol{\varphi}) = \sum_{j \in S(F)} x_{p(j)} \psi_j(\theta_j).$$

Unconditional Aster Model Log Likelihood (Cont.)

$$\varphi_j = \theta_j - \sum_{m \in S(j)} \psi_m(\theta_m) \qquad (*)$$

System of equations ($*$) can be solved for the θ_j in terms of the φ_j in one pass through the equations in any order that finds θ_j for children before parents.

Hence ($*$) defines an invertible mapping between the conditional parameter vector $\boldsymbol{\theta}$ and the unconditional parameter vector $\boldsymbol{\varphi}$.

Yet Another Bit of Exponential Family Theory

Canonical statistic of exponential family is *minimal sufficient*.

Linear transformation

$$\boldsymbol{\varphi} = \mathbf{M}\boldsymbol{\beta}$$

($\mathbf{M}$ is "model matrix" and $\boldsymbol{\beta}$ are "regression coefficients") gives new exponential family with log likelihood

$$\begin{aligned} l(\boldsymbol{\beta}) &= \langle \mathbf{x}, \mathbf{M}\boldsymbol{\beta} \rangle - \psi(\mathbf{M}\boldsymbol{\beta}) \\ &= \langle \mathbf{M}^T \mathbf{x}, \boldsymbol{\beta} \rangle - \psi(\mathbf{M}\boldsymbol{\beta}) \end{aligned}$$

new *canonical statistic vector* $\mathbf{M}^T\mathbf{x}$ (which is minimal sufficient statistic for new family), new *canonical parameter vector* $\boldsymbol{\beta}$, and new *cumulant function*

$$\psi_{\mathbf{M}}(\boldsymbol{\beta}) = \psi(\mathbf{M}\boldsymbol{\beta}).$$

Yet Another Bit of Exponential Family Theory (Cont.)

Linear transformation induces dimension reduction by minimal sufficiency for unconditional exponential families only.

Does not work for conditional exponential families.

Basis of conventional intuition about regression, linear models, generalized linear models, log-linear models!

Unconditional exponential families follow conventional intuition. Algebraically complicated, but statistically simple.

Conditional exponential families don't. Algebraically simple, but statistically complicated.

Independent and Identically Modeled (IIM)

So far, data x_j are measurements at different life stages for same individual. Now make data x_{ij} with i indexing different individuals.

$$
\begin{aligned}
\langle \mathbf{x}, \boldsymbol{\varphi} \rangle &= \sum_{i \in I} \sum_{j \in J} x_{ij} \varphi_{ij} \\
\psi(\boldsymbol{\varphi}) &= \sum_{i \in I} \sum_{j \in S(F)} x_{ip(j)} \psi_j(\theta_{ij}) \\
\varphi_{ij} &= \theta_{ij} - \sum_{m \in S(j)} \psi_m(\theta_{im}) \\
\varphi_{ij} &= \sum_{k \in K} m_{ijk} \beta_k \\
y_k &= \sum_{i \in I} \sum_{j \in J} x_{ij} m_{ijk}
\end{aligned}
$$

Last two are $\boldsymbol{\varphi} = \mathbf{M}\boldsymbol{\beta}$ and $\mathbf{y} = \mathbf{M}^T \mathbf{x}$.

R Package Aster

R package `aster` has usual interface for regression-like stuff.

Function `aster` fits models using R formula mini-language to specify models.

Function `summary.aster` prints regression coefficients, standard errors, etc.

Function `anova.aster` does likelihood ratio tests for comparison of nested models.

Function `predict.aster` does "prediction" (evaluates functions of parameter estimates) with standard errors.

Echinacea Example

Fit many models (http://www.stat.umn.edu/geyer/aster/tr644.pdf).

Scientific interest focuses on the model comparison

```
1: resp ~ varb + level:(nsloc + ewloc)
2: resp ~ varb + level:(nsloc + ewloc) + hdct * pop - pop
3: resp ~ varb + level:(nsloc + ewloc) + hdct * pop
4: resp ~ varb + level:(nsloc + ewloc) + level * pop
```

Model Number	Model d. f.	Model Deviance	Test d. f.	Test Deviance	Test p-value
1	15	2728.72			
2	21	2712.54	6	16.18	0.013
3	27	2684.86	6	27.67	0.00011
4	33	2674.70	6	10.17	0.12

Echinacea Example (Cont.)

```
1: resp ~ varb + level:(nsloc + ewloc)
2: resp ~ varb + level:(nsloc + ewloc) + hdct * pop - pop
3: resp ~ varb + level:(nsloc + ewloc) + hdct * pop
4: resp ~ varb + level:(nsloc + ewloc) + level * pop
```

Variable `resp` is data matrix x_{ij} strung out as vector.

Variable `varb` indicates node j.

Term `level:(nsloc + ewloc)` is spatial effect.

Variable `hdct` indicates head count H_j nodes.

Variable `level` indicates type of node (M_j, F_j or H_j).

Variable `pop` indicates remnant population (surviving fragment of tall-grass prairie) of origin.

Echinacea Example (Cont.)

```
1: resp ~ varb + level:(nsloc + ewloc)
2: resp ~ varb + level:(nsloc + ewloc) + hdct * pop - pop
3: resp ~ varb + level:(nsloc + ewloc) + hdct * pop
4: resp ~ varb + level:(nsloc + ewloc) + level * pop
```

Model 2 is model of most scientific interest. Puts **pop** effect in only at head count **hdct** nodes.

Head count is best surrogate of *fitness*: contribution over life span in descendants to the next generation.

Unconditional aster model takes into account contributions of mortality and flowering status to fitness. Likelihood equations are *observed fitness* = *expected fitness* in each **pop**.

Echinacea Example (Cont.)

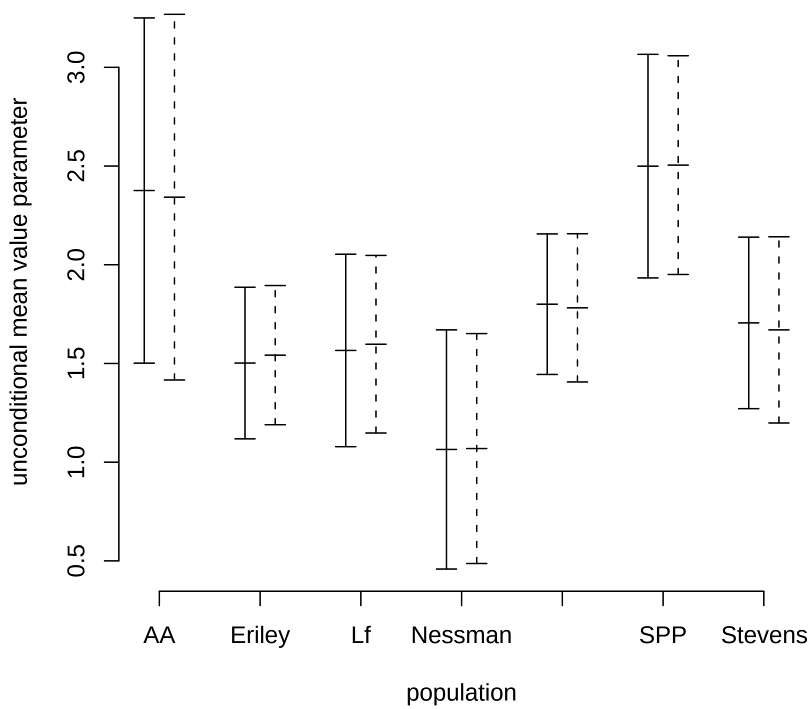
```
1: resp ~ varb + level:(nsloc + ewloc)
2: resp ~ varb + level:(nsloc + ewloc) + hdct * pop - pop
3: resp ~ varb + level:(nsloc + ewloc) + hdct * pop
4: resp ~ varb + level:(nsloc + ewloc) + level * pop
```

Model 3 is goofy. Seems to make sense, but actually doesn't.

Puts in `pop` effects at "non-`hdct`" level, in effect adds mortality and flowering status indicator variables. You score one point for being alive and two points for being alive with flowers. Why does that make sense?

Model 4 does fit significantly better than Model 2 ($P = 0.00016$). There is something going on with mortality and flowering that is not captured by Model 2.

Echinacea Example (Cont.)



Does Model 3 predict (best surrogate of) fitness better than Model 2?

Confidence Intervals for Total Head Count. 95% (non-simultaneous) confidence intervals for unconditional expectation of total flower head count for individuals from different `pop` and central spatial location. Solid bars Model 2, dashed bars Model 3.

Summary

Any joint analysis is better than separate analyses of each variable (thanks, Sandy).

Conditional aster models are simple algebraically, complicated statistically.

Unconditional aster models are simple statistically, complicated algebraically.

R package `aster` is even handed. Conditional and unconditional both implemented. Your choice.

Jaynes, maximum entropy. Choose canonical statistic vector to create scientifically meaningful exponential family model (Geyer and Thompson, 1992).

Other applications areas?

The Name of the Game

Photo from http://en.wikipedia.org/wiki/Sunflower

Aster models are named after the family *Asteraceae*, which is huge (20,000 species) and contains *Echinacea* (type genus *Aster*, common name aster).

They also come with a neat motto: *per aspera cum astris*, a take-off on the motto of the sunflower state.

Credit Where Credit is Due

co-authors

Janis Antonovics (Ruth's thesis advisor)

NSF DMS-0083468

Echinacea project, Julie Etterson, David Heiser, and Stacey Halpern

Helen Hangelbroek (got Ruth and me restarted on aster models)

The R project (www.r-project.org), especially Robert Gentleman
Sweave (Friedrich Leisch)

LaTeX and `pdflatex`

KDE, the Konqueror and KPresenter

Using Iterated Expectation

$$E(X) = E\{E(X \mid Y)\}$$

Use recursively

$$E_{\boldsymbol{\varphi}}\{X_{ij}\} = X_{if(j)} \prod_{\substack{m \in J \\ j \preceq m \prec f(j)}} \psi'_m(\theta_{im})$$

where $f(j)$ denotes founder ancestor of node j and $j \preceq m \prec f(j)$ has m run over node j and its ancestors, back to but not including $f(j)$.

Using Iterated Variance and Covariance

$$\operatorname{var}(X) = E\{\operatorname{var}(X \mid Y)\} + \operatorname{var}\{E(X \mid Y)\}$$
$$\operatorname{cov}(X, Y) = E\{\operatorname{cov}(X, Y \mid Z)\} + \operatorname{cov}\{E(X \mid Z), E(Y \mid Z)\}$$

$$\begin{aligned} \operatorname{var}_{\varphi}\{X_{ij}\} &= E_{\varphi}\Big[\operatorname{var}_{\varphi}\{X_{ij}|X_{ip(j)}\}\Big] + \operatorname{var}_{\varphi}\Big[E_{\varphi}\{X_{ij}|X_{ip(j)}\}\Big] \\ &= \psi_j''(\theta_{ij}) E_{\varphi}\{X_{ip(j)}\} + \psi_j'(\theta_{ij})^2 \operatorname{var}_{\varphi}\{X_{ip(j)}\} \end{aligned}$$

If j' is not a descendent of j

$$\begin{aligned} \operatorname{cov}_{\varphi}\{X_{ij}, X_{ij'}\} &= \operatorname{cov}_{\varphi}\Big\{E_{\varphi}(X_{ij}|X_{ip(j)}), X_{ij'}\Big\} \\ &= \psi_j'(\theta_{ij}) \operatorname{cov}_{\varphi}\Big\{X_{ip(j)}, X_{ij'}\Big\} \end{aligned}$$