

# Aster Models for Life History Analysis

BY CHARLES J. GEYER

*School of Statistics, University of Minnesota, 313 Ford Hall, 224 Church Street S. E.,  
Minneapolis, Minnesota, 55455, U. S. A.*

charlie@stat.umn.edu

STUART WAGENIUS

*Institute for Plant Conservation Biology, Chicago Botanic Garden, 1000 Lake Cook Road,  
Glencoe, Illinois, 60022, U. S. A.*

wagenius@chicagobotanic.org

AND RUTH G. SHAW

*Department of Ecology, Evolution, and Behavior, University of Minnesota, 100 Ecology  
Building, 1987 Upper Buford Circle, St. Paul, Minnesota, 55108, U. S. A.*

rshaw@superb.ecology.umn.edu

## Abstract

We present a new class of statistical models designed for life history analysis of plants and animals. They allow joint analysis of data on survival and reproduction over multiple years, allow for variables having different statistical distributions, and correctly account for the dependence of variables on earlier variables (for example, that a dead individual stays dead and cannot reproduce). We illustrate their utility with an analysis of data taken from an experimental study of *Echinacea angustifolia* sampled from remnant prairie populations in western Minnesota. Statistically, they are graphical models with some resemblance to

generalized linear models and survival analysis. They have directed acyclic graphs with nodes having no more than one parent. The conditional distribution of each node given the parent is a one-parameter exponential family with the parent variable the sample size. The model may be heterogeneous, each node having a different exponential family. We show that the joint distribution is a flat exponential family and derive its canonical parameters, Fisher information, and other properties. These models are implemented in an R package ‘aster’ available from CRAN.

*Some key words:* Conditional Exponential Family; Curved Exponential Family; Flat Exponential Family; Generalized Linear Model; Graphical Model; Maximum Likelihood; Nuisance Variable.

## 1. INTRODUCTION

This article introduces a class of statistical models that we call *aster models*. They were invented for life history analysis (LHA) of plants and animals, and are best introduced by example. Archetypal data for these models are about perennial plants censused at various times. For each individual planted, we record whether it is still alive, whether it has flowered, and how many flowers it has. These data are complicated, especially when recorded for several years, but when considered conditionally, simple models may suffice. We consider mortality status (dead or alive) to be Bernoulli given the preceding mortality status. Similarly for flowering status given mortality status. Given flowering, the number of flowers may have a Poisson distribution conditioned on being nonzero. Figure 1 gives a graphical representation of this kind of data.

[Figure 1 about here.]

The most important feature of these models is the simplest one. A simultaneous analysis that models the joint distribution of all the variables in a life history analysis can answer

questions that cannot be addressed when one does a separate analysis of each variable (conditional on the values of the others).

Although we called the Figure 1 data ‘archetypal’, there is nothing special about the three particular measurements in that example. We could add a fourth variable, seed number, modelled conditional on flower number. And so forth. Nor is there anything special about plants or even living organisms. This methodology applies to any similar conditional modelling.

Our models have some resemblance to discrete time Cox regression (Cox, 1972; Breslow, 1972, 1974) when the graph is linear and all the responses are Bernoulli, but does not have exactly the same likelihood (so those aster models are competitors rather than generalizations of Cox models). Of course, if the graph is a general forest or the responses are not all Bernoulli, then the resemblance to survival analysis is faint.

Our models also have some resemblance to generalized linear models (GLM; McCullagh & Nelder, 1989) when the graph has only one node, but we do not allow arbitrary link functions or quasi-likelihood, using only canonical exponential family parameters as linear predictors (so those aster models are specializations of GLM). Of course, if the graph is a general forest and especially if the responses do not all have the same exponential family, then the resemblance to GLM is faint.

Our models are graphical but of the simplest kind, associated with directed acyclic graphs (Lauritzen, 1996, Section 3.2.2), in which the joint density is a product of conditionals as in equation (1) below. Readers need no knowledge of graphical model theory to understand aster models.

The main innovative aspect of aster models is the interplay between two canonical parameterizations described in Sections 1.2 and 1.3 below. One, the *conditional canonical parameterization*, arises when each distribution in the product of conditionals (1) belongs

to an exponential family and we use the canonical parameterization for each. These are the conditional exponential family (CEF) aster models. The other, the *unconditional canonical parameterization*, arises from observing that the joint model is a full flat exponential family (Barndorff-Nielsen, 1978, Chapter 8) and using the canonical parameters for that family, defined by equation (5) below. These are the flat exponential family (FEF) aster models. And we see that CEF could also stand for ‘curved exponential family’ since, considered unconditionally, that is what they are.

We named our models after flowers because the name is short and much nicer than *forest graph exponential family conditional or unconditional canonical statistic models* or any other descriptive name we could think of. The particular name was chosen for the organism in our example, the purple coneflower *Echinacea angustifolia*, which is in the family *Asteraceae* of which *Aster* is the type genus. They also come with a neat motto: *per aspera cum astris*, a take-off on the motto of the sunflower state (sunflowers are also *Asteraceae*).

### 1.1. Forest Graph Models

We describe conditional dependence structure graphically, as in Figure 1. Each node in the graph is associated with a variable (in Figure 1 we labelled the nodes by their variable names). Aster models have *forest graphs*: directed acyclic graphs in which each node has at most one predecessor and there are no isolated nodes. Each edge (arrow) in the graph represents a conditional distribution of the variable at the arrow head given the variable at the arrow tail (its predecessor). Thus the joint distribution of all the variables is the product of conditionals, equation (1) below, one conditional for each arrow in the graph.

When  $m$  is the *predecessor* of  $j$ , we also say  $j$  is a *successor* of  $m$ . Nodes that have no predecessors are called *root* nodes. Nodes that have no successors are called *leaf* nodes. We also use an alternative terminology saying *parent* instead of predecessor, *child* instead of successor, *founder* instead of root, and *childless* instead of leaf.

We divide the nodes of the graph into disjoint sets  $F$  and  $J$ , the founder and non-founder nodes, respectively, and introduce a function  $p : J \rightarrow J \cup F$  that maps nodes  $j$  to their predecessors  $p(j)$ . The diagram of  $p$  is just like the graph of the graphical model (like Figure 1) except that all the arrows are reversed. With this notation, we can write the joint distribution of all the variables as the product of conditionals

$$\prod_{j \in J} \Pr\{X_j | X_{p(j)}\}. \quad (1)$$

The fact that the variables  $X_j$  associated with root nodes do not appear ‘in front of the bar’ in (1) means these variables are nonrandom (or at least are treated as nonrandom in that we are conditioning on them and not modelling their distributions).

It will simplify notation in what follows if we define the partial order relations determined by the graph. The predecessor function  $p$  determines a relation  $\{(j, p(j)) : j \in J\}$ . We denote its transitive closure by  $\prec$ , so  $j \prec m$  means  $m = p(j)$ , or  $m = p(p(j))$ , or  $\dots$ . And we denote its transitive reflexive closure by  $\preceq$ , so  $j \preceq m$  means  $j \prec m$  or  $j = m$ . We call  $\prec$  the *ancestor relation*, and read  $j \prec m$  as  $m$  is an *ancestor* of  $j$ . We call  $\preceq$  the *ancestor-or-self relation*.

It will also simplify notation if we define the function  $f : J \rightarrow F$  that maps a node  $j$  to its (unique) founder ancestor  $f(j)$ .

## 1.2. Conditional Exponential Families

We take each of the conditional distributions in (1) to be a one-parameter exponential family (perhaps a different such family for each  $j$ ) with  $X_j$  the canonical statistic and the dependence on  $X_{p(j)}$  being that  $X_j$  is the sum of  $X_{p(j)}$  i. i. d. (independent and identically distributed) random variables. In order that this make sense the variables  $X_j$  for  $j \in p(J)$ , where  $p(J)$  denotes the range of the predecessor function (the set of non-leaf nodes), must be nonnegative-integer-valued. No such restriction is placed on  $X_j$  for  $j \notin p(J)$  (for leaf

nodes).

Then the log likelihood for the whole family has the form

$$\sum_{j \in J} X_j \theta_j - X_{p(j)} \psi_j(\theta_j) \quad (2)$$

where  $\theta_j$  is the canonical parameter for the  $j$ -th conditional family and  $\psi_j$  is the so-called *cumulant function* for that family (Barndorff-Nielsen, 1978, pp. 105, 139, and 150) that satisfies

$$E_{\theta_j} \{X_j | X_{p(j)}\} = \psi'_j(\theta_j) \quad (3a)$$

$$\text{var}_{\theta_j} \{X_j | X_{p(j)}\} = \psi''_j(\theta_j) \quad (3b)$$

(for examples of cumulant functions, see Appendix B of the technical report at <http://www.stat.umn.edu/geyer/aster>).

When we have *independent and identically modelled* (i. i. m.) observations (independent and come from the same model but may have different parameter values) of data from such families, the log likelihood becomes

$$\sum_{i \in I} \sum_{j \in J} X_{ij} \theta_{ij} - X_{ip(j)} \psi_j(\theta_{ij}) \quad (4)$$

where  $I$  is a finite set (indexing individuals). Note that  $\psi_j$  is not  $\psi_{ij}$  so each node of the graph has only one exponential family associated with it, an idea required by the rows of the data matrix being i. i. m.

### 1.3. Unconditional Exponential Families

Introducing the notation

$$S(j) = \{ m \in J : j = p(m) \}$$

for the set of successors of  $j$  and collecting terms with the same  $X_j$  in (2), we get

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

where  $S(F)$  is the set of children of founders.

Now we see that by introducing new parameters

$$\varphi_j = \theta_j - \sum_{m \in S(j)} \psi_m(\theta_m), \quad j \in J \quad (5)$$

we have an unconditional exponential family with canonical statistics  $X_j$  and canonical parameters  $\varphi_j$  for  $j \in J$ .

Collecting the canonical statistics and parameters into vectors  $\mathbf{X}$  and  $\boldsymbol{\varphi}$  we can write the log likelihood of this unconditional family as

$$l(\boldsymbol{\varphi}) = \langle \mathbf{X}, \boldsymbol{\varphi} \rangle - \psi(\boldsymbol{\varphi}) \quad (6a)$$

where  $\langle \mathbf{X}, \boldsymbol{\varphi} \rangle$  denotes the inner product  $\sum_i X_i \varphi_i$  and where the cumulant function of this family is

$$\psi(\boldsymbol{\varphi}) = \sum_{j \in S(F)} X_{p(j)} \psi_j(\theta_j) \quad (6b)$$

Note that all of the  $X_{p(j)}$  in (6b) are at founder nodes and are constant so  $\psi$  is a deterministic (not random) function, and also note that, although it is not obvious that the right hand side of (6b) is (as the left hand side says) a function of  $\boldsymbol{\varphi}$ , this must be true by the logic of exponential families (Barndorff-Nielsen, 1978, pp. 105 ff.).

A little more thought shows that the system of equations (5) can be solved for the  $\theta_j$  in terms of the  $\varphi_j$  in one pass through the equations in any order that finds  $\theta_j$  for children before parents. Thus (5) determines an invertible change of parameter.

The analog of (6a) and (6b) when we have i. i. m. replication requires we interpret matrices as vectors, elements of the finite-dimensional vector space  $\mathbb{R}^{I \times J}$ . Equation (6a) remains, although we reinterpret its inner product as  $\sum_{ij} X_{ij} \varphi_{ij}$ , and (6b) gets additional indices

$$\psi(\boldsymbol{\varphi}) = \sum_{i \in I} \sum_{j \in S(F)} X_{ip(j)} \psi_j(\theta_{ij}) \quad (6c)$$

as does (5)

$$\varphi_{ij} = \theta_{ij} - \sum_{m \in S(j)} \psi_m(\theta_{im}), \quad i \in I, j \in J. \quad (6d)$$

#### 1.4. Sufficient Statistics

As is well known (Barndorff-Nielsen, 1978, p. 111) the canonical statistic of an exponential family is minimal sufficient. Since we have both conditional and unconditional families in play, we stress that this well-known result is about *unconditional* families.

One of the desirable aspects of exponential family GLM defined by reparameterization of the form

$$\varphi = \mathbf{M}\boldsymbol{\beta}, \quad (7)$$

where  $\mathbf{M}$  is a known matrix (the *model matrix*), is that from the identity  $\langle \mathbf{X}, \mathbf{M}\boldsymbol{\beta} \rangle = \langle \mathbf{M}^T \mathbf{X}, \boldsymbol{\beta} \rangle$  we see that the result is a new exponential family with canonical statistic  $\mathbf{M}^T \mathbf{X}$  and canonical parameter  $\boldsymbol{\beta}$ . The dimension of this new family will be the dimension of  $\boldsymbol{\beta}$ , if  $\mathbf{M}$  has full rank.

If  $\mathbf{X}$  is a matrix interpreted as a vector, then  $\mathbf{M}$  is an array representing a linear operator  $\mathbb{R}^K \rightarrow \mathbb{R}^{I \times J}$ . So (7) means

$$\varphi_{ij} = \sum_{k \in K} m_{ijk} \beta_k, \quad (8)$$

and  $\mathbf{Y} = \mathbf{M}^T \mathbf{X}$  means

$$Y_k = \sum_{i \in I} \sum_{j \in J} X_{ij} m_{ijk}. \quad (9)$$

This ‘dimension reduction’ to sufficient statistics  $Y_k$  does not occur when the conditional parameters  $\boldsymbol{\theta}$  are ‘generalized linear modelled’ in a similar way, and this suggests that generalized linear models for the unconditional parameterization may be scientifically more interesting despite their more complicated structure.



## 1.5. Mean Value Parameters

Canonical parameters, although interesting both theoretically and practically because they are the GLM ‘linear predictors’, have no real-world interpretation. The parameters that do are the *mean value parameters*.

The conditional mean value parameters are the conditional means

$$\xi_{ij} = E_{\theta_{ij}}\{X_{ij}|X_{ip(j)}\} = X_{ip(j)}\psi'_j(\theta_{ij}). \quad (10)$$

Strictly speaking, the  $\xi_{ij}$  are not *parameters* because they contain random data  $X_{ip(j)}$ . Nevertheless, they do play the role of mean value parameters when one is thinking conditionally, treating  $X_{ip(j)}$  as constant. Standard exponential family theory (Barndorff-Nielsen, 1978, p. 121) says that  $\psi'_j$  is an invertible change of parameter.

Similarly, the unconditional mean value parameters are the unconditional means

$$\boldsymbol{\tau} = E_{\boldsymbol{\varphi}}\{\mathbf{X}\} = \nabla\psi(\boldsymbol{\varphi}) \quad (11)$$

( $\nabla$  denotes the vector of partial derivatives). Again, standard theory says that  $\nabla\psi : \boldsymbol{\varphi} \mapsto \boldsymbol{\tau}$  is an invertible change of parameter. The unconditional expectation in (11) can be calculated using the iterated expectation theorem (from probability theory) or the chain rule (from calculus)

$$E_{\boldsymbol{\varphi}}\{X_{ij}\} = X_{if(j)} \prod_{\substack{m \in J \\ j \preceq m \prec f(j)}} \psi'_m(\theta_{im}), \quad (12)$$

where the  $\theta_{ij}$  are determined from  $\boldsymbol{\varphi}$  by solving (5).

## 2. THEORY

### 2.1. Conditional Models

The score for conditional canonical parameters is particularly simple

$$\frac{\partial l(\boldsymbol{\theta})}{\partial \theta_{ij}} = X_{ij} - X_{ip(j)}\psi'_j(\theta_{ij})$$

and, if these parameters are modelled linearly in terms of other parameters, equation (8) with  $\varphi$  replaced by  $\theta$ , then we have

$$\frac{\partial l(\boldsymbol{\beta})}{\partial \beta_k} = \sum_{i \in I} \sum_{j \in J} \left[ X_{ij} - X_{ip(j)} \psi'_j(\theta_{ij}) \right] m_{ijk} \quad (13)$$

The observed Fisher information for  $\boldsymbol{\theta}$  is diagonal with

$$-\frac{\partial^2 l(\boldsymbol{\theta})}{\partial \theta_{ij}^2} = X_{ip(j)} \psi''_j(\theta_{ij})$$

and zero for mixed second derivatives. And from this we see the observed Fisher information for  $\boldsymbol{\beta}$  is

$$-\frac{\partial^2 l(\boldsymbol{\beta})}{\partial \beta_k \partial \beta_{k'}} = \sum_{i \in I} \sum_{j \in J} X_{ip(j)} \psi''_j(\theta_{ij}) m_{ijk} m_{ijk'}. \quad (14a)$$

The expected Fisher information is just the (unconditional) expectation of the observed Fisher information. So it is

$$-E_{\boldsymbol{\beta}} \left\{ \frac{\partial^2 l(\boldsymbol{\beta})}{\partial \beta_k \partial \beta_{k'}} \right\} = \sum_{i \in I} \sum_{j \in J} E_{\boldsymbol{\beta}} \{ X_{ip(j)} \} \psi''_j(\theta_{ij}) m_{ijk} m_{ijk'} \quad (14b)$$

the unconditional expectation on the right hand side being evaluated by using (12).

## 2.2. Unconditional Models

The score for unconditional canonical parameters is, as in every (unconditional) exponential family, ‘observed minus expected’

$$\frac{\partial l(\boldsymbol{\varphi})}{\partial \varphi_{ij}} = X_{ij} - E_{\boldsymbol{\varphi}} \{ X_{ij} \}$$

the unconditional expectation on the right hand side being evaluated by using (12), and, if these parameters are modelled linearly in terms of other parameters (8), then we have

$$\frac{\partial l(\boldsymbol{\beta})}{\partial \beta_k} = \sum_{i \in I} \sum_{j \in J} \left[ X_{ij} - E_{\boldsymbol{\beta}} \{ X_{ij} \} \right] m_{ijk}. \quad (15)$$

Second derivatives with respect to the (unconditional) canonical parameters of an exponential family are nonrandom. Hence there is no difference between observed and expected

Fisher information. The information for  $\varphi$  is given by either of the expressions  $\nabla^2\psi(\varphi)$  or  $\text{var}_\varphi(\mathbf{X})$ , the former being the matrix of second partial derivatives of the cumulant function (6c) and the latter being the variance-covariance matrix of the random vector with components  $X_{ij}$  (Barndorff-Nielsen, 1978, p. 150). We choose to work on the latter. The iterated variance formula gives

$$\begin{aligned}\text{var}_\varphi\{X_{ij}\} &= E_\varphi[\text{var}_\varphi\{X_{ij}|X_{ip(j)}\}] + \text{var}_\varphi[E_\varphi\{X_{ij}|X_{ip(j)}\}] \\ &= \psi''_j(\theta_{ij})E_\varphi\{X_{ip(j)}\} + \psi'_j(\theta_{ij})^2 \text{var}_\varphi\{X_{ip(j)}\}\end{aligned}\tag{16a}$$

Since we already have (12) to calculate expectations, (16a) allows calculation of all variances (diagonal elements of the Fisher information matrix) recursively. Now we work on the covariance of  $X_{ij}$  and  $X_{ij'}$  with  $j \neq j'$ , in which case we may assume without loss of generality that  $j' \not\leq j$  so

$$\text{cov}_\varphi\{X_{ij}, X_{ij'}|X_{ip(j)}\} = 0 = \text{cov}_\varphi\{X_{ij}, X_{ij'}|X_{ip(j)}, X_{ij'}\}$$

because  $X_{ij}$  is conditionally independent given  $X_{ip(j)}$  of all variables except its descendants, which do not include  $X_{ij'}$ . Then the iterated covariance formula gives

$$\begin{aligned}\text{cov}_\varphi\{X_{ij}, X_{ij'}\} &= \text{cov}_\varphi\{E_\varphi(X_{ij}|X_{ip(j)}), X_{ij'}\} \\ &= \psi'_j(\theta_{ij}) \text{cov}_\varphi\{X_{ip(j)}, X_{ij'}\}\end{aligned}\tag{16b}$$

and this allows us to determine all of the covariances recursively. The calculation of the information for  $\beta$  from that for  $\varphi$  is analogous to that in the preceding section and is omitted.

### 2.3. Prediction

By ‘prediction’, we mean no more than evaluation of a function of estimated parameters, one job the `predict` function in R does for linear or generalized linear model fits. In aster models we have five different parameterizations of interest ( $\beta$ ,  $\theta$ ,  $\varphi$ ,  $\xi$ , and  $\tau$ ). The Fisher

information for  $\beta$ , already described, tells us what we need to know about predicting  $\beta$ . So this section is about ‘predicting’ the remaining four.

One often predicts for new individuals having different covariate values from the observed individuals. Then the model matrix  $\mathbf{M}$  used for the prediction is different from that used for fitting. So in this section  $\mathbf{M}$  denotes this possibly different model matrix. However when we use Fisher information  $I(\beta)$ , either observed or expected (Sections 2·1 and 2·2), this is based on the original model used to obtain parameter estimates  $\hat{\beta}$  and hence on the original model matrix.

Let  $\eta$  be the linear predictor ( $\eta = \theta$  for conditional models and  $\eta = \varphi$  for unconditional models). Let  $\zeta$  be any one of  $\theta$ ,  $\varphi$ ,  $\xi$ , or  $\tau$ . Let  $f_{\eta,\zeta}$  denote the map  $\eta \rightarrow \zeta$ , and suppose we wish to predict the parameter

$$g(\beta) = h(\zeta) = h(f_{\eta,\zeta}(\mathbf{M}\beta)) \quad (17)$$

Then by the chain rule (17) has derivative

$$\nabla g(\beta) = \nabla h(\zeta) \circ \nabla f_{\eta,\zeta}(\eta) \circ \mathbf{M} \quad (18)$$

(where here  $\mathbf{M}$  denotes the linear operator associated with the model matrix) and by the ‘usual’ asymptotics of maximum likelihood and the delta method, the asymptotic distribution of the prediction  $h(\hat{\zeta}) = g(\hat{\beta})$  is

$$\text{Normal}(g(\beta), [\nabla g(\hat{\beta})]I(\hat{\beta})^{-1}[\nabla g(\hat{\beta})]^T)$$

where  $\nabla g(\hat{\beta})$  is given by (18) with  $\hat{\eta} = \mathbf{M}\hat{\beta}$  plugged in for  $\eta$  and  $\hat{\zeta} = f_{\eta,\zeta}(\hat{\eta})$  plugged in for  $\zeta$

The point of writing ‘predictions’ in this complicated form is to separate the parts of the specification, the functions  $h$  and  $\nabla h$  and the model matrix  $\mathbf{M}$ , that are easy but change from application to application from the hard part  $\nabla f_{\eta,\zeta}$  that does not change and can be

done by computer (see Appendix A of the technical report at <http://www.stat.umn.edu/geyer/aster> for details).

For mean value parameters the user must also specify new ‘response’ data  $X_{ij}$  as well as new ‘covariate’ data in  $\mathbf{M}$ . This is one way that aster models differ from GLM. Unconditional mean value parameters  $\boldsymbol{\tau}$  will depend only on the root elements  $X_{ij}$ ,  $j \in F$ . Conditional mean value parameters  $\boldsymbol{\xi}$  will depend on all elements  $X_{ij}$ ,  $j \in J \cup F$ .

This makes conditional mean value parameters as described almost useless. Thus we envisage that users will usually specify  $X_{ij} = 1$  for all hypothetical individuals  $i$  and all nodes  $j$  so that then  $\xi_{ij} = \psi'_j(\theta_{ij})$  and hence are really parameters.

### 3. SOFTWARE

We have developed an R (R Development Core Team, 2004) package `aster` that fits, tests, and predicts aster models. It uses the R formula mini-language, originally developed for GENSTAT and S (Wilkinson & Rogers, 1973; Chambers & Hastie, 1992) so that model fitting is much like that for linear or generalized linear models. Similarly, R functions `summary.aster`, `anova.aster`, and `predict.aster` provide regression coefficients with standard errors,  $z$  statistics, and  $p$ -values, likelihood ratio tests for model comparison, and the predictions with standard errors described in Section 2.3. It is available from CRAN (<http://www.cran.r-project.org>).

The current version of the `aster` package has two limitations of the general model described in this article.

- In predictions, the only  $h$  allowed in (17) are linear, that is,  $h(\boldsymbol{\zeta}) = \mathbf{A}^T \boldsymbol{\zeta}$ , for some matrix  $\mathbf{A}$ .
- In modelling, the only families allowed are
  - Bernoulli,

- Poisson, and
- Poisson conditioned on being nonzero.

The package contains 1500 lines of C source code, only 200 lines of which are R-specific, and 925 lines of R. It is provided under a permissive X11-like open source license so the 1300 lines of C that are the computational core could be reused in another implementation, even a proprietary one. Adding another one-parameter exponential family requires only implementation of the  $\psi$ ,  $\psi'$ , and  $\psi''$  functions for the family.

#### 4. EXAMPLE

Data having the aster model structure shown in Figure 1 were collected on 570 individuals of *Echinacea angustifolia*. These plants were sampled as seeds from seven remnant populations that are surviving fragments of the tall-grass prairie that a century ago covered western Minnesota and other parts of the Great Plains of North America. The plants were experimentally randomized at the time of planting into a field within 6.5 km of all populations of origin. The data set contains three other predictor variables: `ewloc` and `nsloc` are spatial coordinates, east-west and north-south positions, of the individuals within the field and `pop` is the predictor variable of scientific interest. It is categorical giving the remnant population of origin.

The greatest advantage of aster models over previously available models for analyzing these data is that the aster model is a joint model for all the variables so a single analysis can account for all effects of predictors and relationships among responses. A secondary advantage is that unconditional aster models directly model marginal distributions by controlling their unconditional mean value parameters.

In order to use the R formula mini-language it is necessary to create some artificial variables. The variable `resp` is a vector comprising the nine response variables (the  $M_j$ ,

$F_j$ , and  $H_j$ ). The variable `varb` is categorical naming these response variables (so they can still be distinguished within `resp`). The variable `level` is categorical naming the type of response variable ( $M$ ,  $F$ , or  $H$ ). The variable `hdct` is an indicator variable indicating the  $H_j$  responses (a convenient shorthand for `level = H`).

We fit many models. See Appendix D of the technical report at <http://www.stat.umn.edu/geyer/aster> for details. Scientific interest focuses on the model comparison shown in Table 1.

[Table 1 about here.]

All models contain the quantitative spatial effect `level:(nsloc + ewloc)`, which was chosen after extensive model comparison (details the technical report). We explain here only the differences among the models.

The terms that differ are all categorical. Such categorical terms include dummy variables in the model (one for each category) and (in an unconditional aster model, which these are) require the maximum likelihood mean value parameters for each category (summed over all individuals in the category) to match the observed values ('observed equals expected').

The models are nested, numbered in increasing order, so Model 1 has no terms not in the others. Model 2 adds `hdct * pop - pop`, which makes observed equal expected for *total flower head count* for each of the 7 populations. Model 3 adds `pop`, which makes observed equal expected for *total non-head-count* ( $\sum_j M_j + F_j$ ) for each of the 7 populations. Model 4 adds `level * pop`, which splits total non-head-count into *total survival* ( $\sum_j M_j$ ) and *total flowering* ( $\sum_j F_j$ ) and makes observed equal expected for them for each of the 7 populations.

From purely statistical considerations, Model 3 is the best of these four nested models. Model 4 does not fit significantly better. Model 2 fits significantly worse. According to Model 1, the life history depends only on position within the field in which the plants are

growing. Model 2 fits differences among populations in total headcount. From scientific considerations Model 3 is rather difficult to interpret, because it fits differences among populations in “non-head-count” ( $M_j + F_j$ ), scoring each individual 0 for dead, 1 for alive without flowers, or 2 for alive with flowers. Model 4 fits differences among populations in each of survival, flowering, and headcount.

Model 2 is the model of primary interest for reasons we now explain. Evolutionary biologists are very interested in fitness. For our purposes here the fitness of an individual may be defined as its contribution over its lifespan in descendants to the next generation (see Beatty, 1992; Keller, 1992; Paul, 1992, for further discussion). Fitness is notoriously difficult to measure. One reason for this is that it is expressed over the lifespan, rather than instantaneously. For these data the most direct surrogate measure for fitness is *total flower head count* ( $\sum_j H_j$ ). The currently available data represent a small fraction of this plant’s lifespan. To obtain more complete measures of fitness, we are continuing these experiments and collecting these data for successive years.

Biologists call all our measured variables (the  $M_j$ ,  $F_j$ , and  $H_j$ ) *components of fitness*. Since  $M_j$  and  $F_j$  contribute to fitness (descendants) only through  $H_j$ , in an aster model the unconditional expectation of  $H_j$  (its mean value parameter) completely accounts for the contributions of  $M_j$  and  $F_j$ . Strictly speaking, this is not quite true, since we do not have  $H_j$  measured over the whole life span, so the last  $M_j$  contains the information that future reproduction is possible, but it becomes truer as more data are collected in future years. Moreover, we have no data about life span and do not wish to inject subjective opinion about future flower head count into the analysis.

The statistical point of this is that the  $M_j$  and  $F_j$  are only in the model to produce the correct stochastic structure. If we could directly model the marginal distribution of the  $H_j$  (but we can’t), we would not need the other variables. They are ‘nuisance variables’ that



must be in the model but are of no interest in this particular analysis. (This is similar to division of parameters into ‘interest’ and ‘nuisance’. In fact the mean value parameters for those variables are nuisance parameters.)

Statisticians seem not to have studied this kind of nuisance variable (there is an established usage meaning ‘confounded with treatment effect’ that is not what we mean). Model 3 is the best according to the likelihood ratio test, but does it fit the variables of interest better than Model 2? We do not know of an established methodology addressing this issue, so we propose looking at confidence intervals for the mean value parameters for total flower head count shown in Figure 2 (page 23).

[Figure 2 about here.]

Although we have no formal test to propose, we claim it is obvious that Model 3 is no better than Model 2 at ‘predicting’ the best surrogate of expected fitness. We take this as justification for using Model 2 in scientific discussion and infer from it significant differences among the populations in flower head count and, thus, fitness.

As we said above, Model 3 is difficult to interpret scientifically. Model 4 is the next larger readily interpretable model. The fact that Model 4 fits significantly better than Model 2 ( $P = 0.00016$ ) implies that there are differences among populations in mortality and flowering (the  $M_j$  and  $F_j$ ) that may be of scientific interest even though they make no direct contribution to fitness (since Model 2 already fully accounts for their contributions through  $H_j$ ).

Note that we would have gotten very different results had we used a conditional model (not shown, see Section D.3.2 of the technical report). The parameters of interest are *unconditional* expectations of total flower head count. This alone suggests an unconditional model. Furthermore, we see in (5) that unconditional aster models ‘mix levels’ passing information up from children to parents. This is why Model 2 in our example was successful

in predicting total head count while only modelling `pop` effects at head count nodes. By not mixing levels in this way, a *conditional* aster model *must* model all levels and so usually needs many more parameters than an unconditional model.

## 5. DISCUSSION

The key idea of aster models (as we see it) is the usefulness of what we have called unconditional aster models (FEF), which have low-dimensional sufficient statistics (9). Following Geyer and Thompson (1992), who argued in favour of exponential family models with the ‘right’ sufficient statistics (chosen to be scientifically interpretable), an idea they attributed to Jaynes (1978), we argue that scientists are likely to find among these models the ones of scientific interest.

We don’t insist, though. The `aster` R package is even-handed with respect to conditional and unconditional models and conditional and unconditional parameters. Users may use whatever seems best to them. Any joint analysis is better than any separate analyses of different variables.

But we have one warning for naive users. Conditional models, though algebraically simple, are deceptively complicated statistically. As with all exponential family models, the map from canonical parameters to mean value parameters is monotone. So with sufficient statistics  $Y_k$  given by (9) we have

$$-\frac{\partial^2 l(\boldsymbol{\beta})}{\partial \beta_k^2} = \frac{\partial E_{\boldsymbol{\beta}}\{Y_k\}}{\partial \beta_k} > 0 \quad (19a)$$

in an unconditional model. The middle term gives the regression coefficients their simple interpretation: an increase in  $\beta_k$  causes an increase in  $E_{\boldsymbol{\beta}}\{Y_k\}$ , other betas being held constant. The analog for a conditional model is

$$-\frac{\partial^2 l(\boldsymbol{\beta})}{\partial \beta_k^2} = \sum_{i \in I} \sum_{j \in J} \frac{\partial}{\partial \theta_{ij}} E_{\theta_{ij}}\{X_{ij} | X_{ip(j)}\} m_{ijk}^2 > 0. \quad (19b)$$

There is no way to give the middle term in (19b) an interpretation as a single conditional or unconditional expectation and hence no corresponding simple interpretation of regression coefficients.

Unconditional aster models seem algebraically complicated. If one attempts to understand them by understanding (5) and (16a) and (16b) and similarly complicated equations relegated to the technical report, one will find them very confusing. One can only understand them by understanding the big picture. They are the flat exponential families with the desired sufficient statistics. They are the aster models that behave according to the intuitions one has developed from linear and generalized linear models. They are statistically simple. In contrast, conditional aster models are algebraically simple but statistically complicated.

Aster models are, once one accepts the restriction to one-parameter exponential families for the single-node models, simply the right thing. There is no reason for the restriction to one-parameter exponential families other than our desire to keep the complexity manageable. Aster models are already very complicated and take some time to digest. We were afraid that allowing, for example, leaf nodes to be two-parameter normal would overwhelm both implementers (us) and users.

We saw in our example that aster models allowed us to successfully model the surrogate of fitness. In humans, Darwinian fitness is not the *summum bonum* that it is for organisms lacking language, memes, and minds. But neither is mere survival! Survival analysis, by its very name, rules out any consideration of quality of life. Aster models do permit such consideration. We understand that application of aster models to clinical trials involves medical and ethical considerations we are incompetent to address. We merely raise the point.

## ACKNOWLEDGEMENTS

Our key equation (5), at least in the case where the graph is linear so each node has at most one successor, was figured out about 1980 when the third author was a graduate student and the first author not yet a graduate student. The third author's adviser Janis Antonovics provided some funding for development (from a now forgotten source, but we thank it), but we did not publish then, and other work intervened. High quality life history data generated recently by the second author with funding from NSF (DMS-0083468) and also by Julie Etterson, David Heiser, and Stacey Halpern finally motivated implementation of these ideas. Also Helen Hangelbroek did what in hindsight we would call a conditional aster model analysis of mortality data, and questions about her analysis led us back to the full aster theory. Were it not for Janis's help and encouragement back then, we might not have been able to do this now. We should also acknowledge the R team. Without R we would never have made software as usable and powerful as the `aster` package. We also thank Robert Gentleman for pointing out the medical implications.

## REFERENCES

- BARNDORFF-NIELSEN, O. E. (1978). *Information and Exponential Families*. Chichester: John Wiley.
- BEATTY, J. (1992). Fitness: Theoretical contexts. In *Keywords in Evolutionary Biology*, Ed. E. F. Keller and E. A. Lloyd, pp. 115–119. Cambridge, Mass: Harvard University Press.
- BRESLOW, N. (1972). Discussion of the paper by Cox (1972). *J. R. Statist. Soc. B* **34**, 216–17.
- BRESLOW, N. (1974). Covariance analysis of censored survival data. *Biometrics*, **30**, 89-99.

- CHAMBERS, J. M. & HASTIE, T. J. (1992). Statistical models. In *Statistical Models in S*, J. M. Chambers and T. J. Hastie, eds. Pacific Grove, Calif.: Wadsworth & Brooks/Cole.
- COX, D. R. (1972). Regression models and life-tables (with discussion). *J. R. Statist. Soc. B* **34**, 187–220.
- GEYER, C. J. & THOMPSON, E. A. (1992). Constrained Monte Carlo maximum likelihood for dependent data, (with discussion). *J. R. Statist. Soc. B* **54** 657–99.
- JAYNES, E. T. (1978). Where do we stand on maximum entropy? In *The Maximum Entropy Formalism*, Ed. R. D. Levine and M. Tribus, pp. 15-118. Cambridge, Mass: MIT Press.
- KELLER, E. F. (1992). Fitness: Reproductive ambiguities. In *Keywords in Evolutionary Biology*, Ed. E. F. Keller and E. A. Lloyd, pp. 120–121. Cambridge, Mass: Harvard University Press.
- LAURITZEN, S. L. (1996). *Graphical Models*. New York: Oxford University Press.
- MCCULLAGH, P. & NELDER, J. A. (1989). *Generalized Linear Models*, 2nd ed. London: Chapman & Hall.
- PAUL, D. (1992). Fitness: Historical perspective. In *Keywords in Evolutionary Biology*, Ed. E. F. Keller and E. A. Lloyd, pp. 112–114. Cambridge, Mass: Harvard University Press.
- R DEVELOPMENT CORE TEAM (2004). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. <http://www.R-project.org>.
- WILKINSON, G. N. & ROGERS, C. E. (1973). Symbolic description of factorial models for analysis of variance. *Applied Statistics* **22**, 392–99.

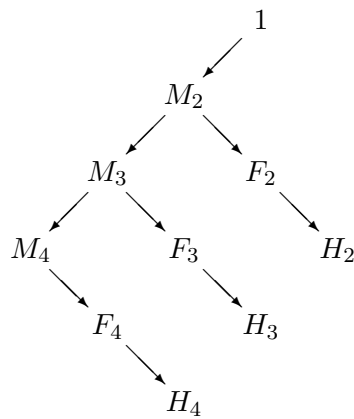


Figure 1: Graph for Archetypal Aster Data. Arrows go from nodes to their successors. Nodes are labelled by their associated variables. The only founder node is associated with the constant variable 1.  $M_j$  is the mortality status in year  $2000 + j$ .  $F_j$  is the flowering status in year  $2000 + j$ .  $H_j$  is the flower head count in year  $2000 + j$ . The  $M_j$  and  $F_j$  are Bernoulli conditional on their predecessor variables being one (and zero otherwise). The  $H_j$  are Poisson conditioned on being nonzero conditional on their predecessor variables being one (and zero otherwise).

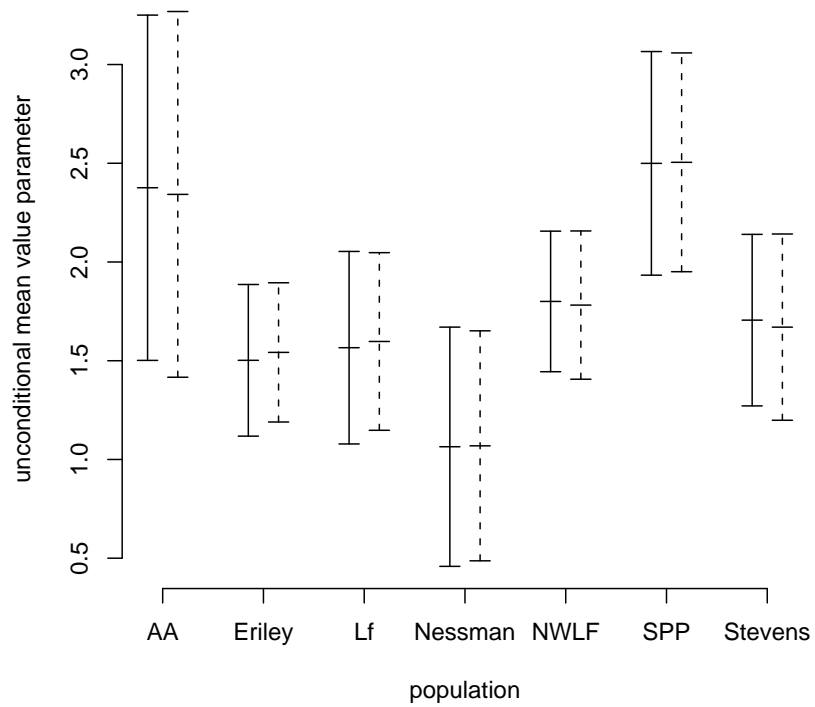


Figure 2: Confidence Intervals for Total Head Count. 95% (non-simultaneous) confidence intervals for the unconditional expectation of total flower head count (all three years) for individuals from different populations and central spatial location. Solid bars based on Model 2 in Table 1. Dashed bars based on Model 3 in Table 1.

Table 1: Tests for Model Comparison. The model formulae are given above and the analysis of deviance below (deviance is twice log likelihood).

- 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 <i>p</i> -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