Stat 8931 Fall 2018 Notes
Regular Full Aster Models

Charles J. Geyer

September 11, 2018

1 Introduction

The purpose of this note is to rewrite Appendix E of the technical report (Geyer, Wagenius, annd Shaw, 2005) that backed up the first paper on aster models (Geyer, Wagenius, and Shaw, 2007). That appendix proved a theorem about when aster models without dependence groups are regular and full. The reason it did not do general aster models is because at the time that technical report was written aster models had not yet been generalized as far as they were by Geyer, et al. (2007) as can be seen from the non-appendix part of that technical report, which is essentially the version of Geyer, et al. (2007) that was originally submitted to the journal. The technical report was not updated to match the published version of the paper because R package `aster` (Geyer, 2018) did not then and does not now do aster models with dependence groups and R package `aster2` (Geyer, 2017, first appeared on CRAN November 2010), which does do aster models with dependence groups, did not yet exist.

Aster models are exponential familes of distributions. What this document discusses is when they have the additional properties of fullness and regularity, which are used everywhere in the theory of aster models. All families implemented by R packages `aster` and `aster2` are regular full exponential families. Hence it follows from this document that every unconditional aster model is a regular full exponential family.

2 Basic Exponential Family Theory

2.1 Definitions

Here is a really short definition of exponential families introduced in Geyer (2009). An *exponential family of distributions* is a statistical model having log likelihood of the form

$$l(\theta) = \langle y, \theta \rangle - c(\theta), \tag{1}$$

where y is a vector statistic, θ is a vector parameter, and $\langle \cdot\,, \cdot \rangle$ is the bilinear form placing the vector space where y lives and the vector space where θ lives in duality. Readers who like to think of vectors as being matrices with one column will write

$$\langle y, \theta \rangle = y^T \theta = \theta^T y,$$

but we like our notation with angle brackets, which comes from physics and functional analysis via Geyer (1990, 2009), because it makes it clear that $\langle y, y \rangle$ and $\langle \theta, \theta \rangle$ are mistakes by definition of $\langle \cdot\,, \cdot \rangle$, whereas $y^T y$ and $\theta^T \theta$ are not clearly mistakes.

A change of statistic or a change of parameter may be necessary to get a log likelihood of this form. There may also be an additive term in the log likelihood that is a function of data only, but such terms can be dropped from a log likelihood and need to be dropped to get the form (1).

To indicate the special vector statistic y and vector parameter θ that give a log likelihood of the form (1), they are called the *canonical statistic* and *canonical parameter* (alternate terminology is *natural statistic* and *natural parameter*). The function c is called the *cumulant function* of the family.

The canonical statistic, canonical parameter, and cumulant function are not unique. Any one-to-one affine function of a canonical statistic is another canonical statistic, any one-to-one affine function of a canonical parameter is another canonical parameter, and any cumulant function plus any (real-valued) affine function is another cumulant function. These alterations are not algebraically independent. Changing any one requires changes in the others to maintain the form (1). Usually no fuss is made about this nonuniqueness. One fixes a choice of canonical statistic, canonical parameter, and cumulant function and leaves it at that.

2.2 Cumulant Functions, Fullness, and Regularity

2.2.1 Cumulant Functions

One can only determine the cumulant function directly from the log likelihood when the family is full (a concept that will be defined presently). In case the family is not full, one must use equation (5) in Geyer (2009)

$$c(\theta) = c(\psi) + \log E_\psi\bigl(e^{\langle y, \theta - \psi \rangle}\bigr) \tag{2}$$

to define the cumulant function up to an arbitrary additive constant $c(\psi)$. In (2), ψ is any valid value of the canonical parameter, held fixed, and θ varies over the whole vector space containing the canonical parameter space. We define $c(\theta) = \infty$ when the expectation does not exist.

2.2.2 Full Families

Then

$$\Theta = \{\, \theta : c(\theta) < \infty \,\} \tag{3}$$

is the parameter space of the *full* exponential family containing the originally given exponential family. Densities in the full family with respect to the probability distribution for parameter value ψ are given by equation (4) in Geyer (2009)

$$f_\theta(\omega) = e^{\langle Y(\omega), \theta - \psi \rangle - c(\theta) + c(\psi)}, \tag{4}$$

where ω, as usual in probability theory, is the whole data (recall that Y is just a statistic). These agree with the original definition for the original family but extend to the full family if the original family was not full.

2.2.3 Regular Full Families

The full family is *regular* if Θ is an open subset of the vector space containing it.

2.2.4 Moment and Cumulant Generating Functions

For any θ in the interior of Θ the distribution of Y for parameter value θ has moment generating function

$$M_\theta(t) = e^{c(\theta + t) - c(\theta)}$$

and cumulant generating function

$$K_\theta(t) = c(\theta + t) - c(\theta).$$

For θ on the boundary of Θ the distribution of Y for parameter value θ does not have a moment generating function.

2.2.5 Moments and Cumulants

Consequently, from the theory of moment generating functions, the cumulant function is infinitely differentiable at θ in the interior of Θ, and derivatives at θ are cumulants of the distribution of Y for parameter value θ. In particular, the first two derivatives are

$$\nabla c(\theta) = E_\theta(Y) \tag{5}$$

$$\nabla^2 c(\theta) = \operatorname{var}_\theta(Y) \tag{6}$$

2.2.6 One Point of Regularity

If the full family is regular, then the boundary of the full canonical parameter space (3) is empty, hence every distribution in the family had moments and cumulants of all orders given by derivatives of the moment and cumulant generating functions evaluated at zero.

2.2.7 Another Point of Regularity

Cumulant functions are convex by Hölder's inequality. Hence the full canonical parameter space of an exponential family is a convex set, and the log likelihood of the family is a concave function (Rockafellar and Wets, 1998, Exercise 2.3). Consequently, every local maximizer of the log likelihood is a global maximizer (Rockafellar and Wets, 1998, Theorem 2.6). Better, every point where the first derivative of the log likelihood is zero is a global maximizer (Rockafellar and Wets, 1998, Theorem 2.14, part (b)).

For a regular full exponential family, the first derivative of the log likelihood is

$$\nabla l(\theta) = y - \nabla c(\theta) = y - E_\theta(Y) \tag{7}$$

by (5), and this derivative exists at every point of the canonical parameter space. From multivariable calculus we know that a necessary condition for a local maximum in the interior of the domain of a function is that the first derivative is zero. Concavity, discussed above, says that it is also a sufficient condition.

Hence, for a regular full exponential family, maximum likelihood estimates satisfy the *observed equals expected property:* a point θ is an MLE if and only if

$$y = E_\theta(Y)$$

In words, the observed value of the canonical statistic is equal to its expected value under the MLE.

If a full exponential family is not regular, then MLE can occur on the boundary of the full canonical parameter space where the first derivative does not exist (and one-sided derivatives are not zero). An example of this is the Strauss process, a spatial point process (Geyer and Møller, 1994).

2.3 Canonical Affine Submodels

As in LM and GLM we consider linear submodels in which the canonical parameters are expressed as a linear function of other parameters, called *regression coefficients*. Actually, LM and GLM as implemented by the R functions `lm` and `glm` allow submodels to express canonical parameters as affine functions of regression coefficients through the `offset` argument. Thus, strictly speaking, they should be called "affine models" and "generalized affine models." But offsets are rarely used, so the terminology LM and GLM with L for linear persists. In aster model theory we do call affine submodels "affine" rather than "linear."

A *canonical affine submodel* has parameterization

$$\theta = a + M\beta, \tag{8}$$

where a is a known vector and M is a known matrix; a is called the *offset* vector and M is called the *model matrix* by the R functions `lm` and `glm`. M is called the *design matrix* by others. We use the terminology favored by R.

The log likelihood for the canonical affine submodel is

$$\begin{aligned} l(\beta) &= \langle y, a + M\beta \rangle - c(a + M\beta) \\ &= \langle M^T y, \beta \rangle + \langle y, a \rangle - c(a + M\beta) \end{aligned}$$

and the term $\langle y, a \rangle$ that does not contain the parameter can be dropped giving the log likelihood

$$l(\beta) = \langle M^T y, \beta \rangle - c(a + M\beta) \tag{9}$$

and this shows the canonical affine submodel is again an exponential family with canonical statistic vector $M^T y$, canonical parameter vector β, and cumulant function defined by

$$c_{\text{sub}}(\beta) = c(a + M\beta).$$

The proof that we get a log likelihood of this form is trivial

$$\langle y, M\beta \rangle = y^T (M\beta) = y^T M\beta = (M^T y)^T \beta = \langle M^T y, \beta \rangle.$$

The full canonical parameter space of the submodel is

$$
\begin{aligned}
B &= \{\, \beta \in R^Q : c(a + M\beta) < \infty \,\} \\
&= \{\, \beta \in R^Q : a + M\beta \in \Theta \,\}
\end{aligned}
$$

where Q is the finite set indexing β. In the worst case, if a and M are chosen badly, B could be empty, and the submodel does not exist. But if B is nonempty, it is the full canonical parameter space of the canonical affine submodel.

In order to distinguish the originally given exponential family model with parameter space Θ from its canonical affine submodels, we call this originally given model the *saturated* model.

Since the preimage of an open set is open for any continuous function (this is the definition of continuous function in general topology) and affine functions are continuous, a full canonical affine submodel is regular if the saturated model was regular.

3 Aster Models

3.1 Vectors and Subvectors

In *aster model* theory (Geyer, et al., 2007) we consider vectors and random vectors to be elements of $\mathbb{R}^J$ for some abstract finite set J rather than $\mathbb{R}^d$ for some positive integer d. That means indices j for components y_j of some vector y range over J rather than from 1 to d. We also consider subvectors y_A for subsets $A \subset J$.

As in abstract set theory we can also consider elements of $\mathbb{R}^J$ to be functions $J \to \mathbb{R}$. If $y \in \mathbb{R}^J$, then $y_A \in \mathbb{R}^A$ is the restriction of y to A. If $y \in \mathbb{R}^J$, then $y_j \in \mathbb{R}$ is function evaluation but we write y_j, as is usual in discussing vectors, rather than $y(j)$.

We have nothing to distinguish components y_j from subvectors y_A except the convention that lower case is for elements of sets and upper case for sets.

We also have the convention that lower case is for ordinary scalars and vectors like y, y_j, and y_A, and upper case is for random scalars and random vectors like Y, Y_j, and Y_A.

3.2 Aster Models and their Graphs

An aster model is a parametric statistical model for a random vector Y taking values in $\mathbb{R}^J$ for some finite set J. The model is described as follows.

- There is a partition $\mathcal{G}$ of J.
- There is a function $q : \mathcal{G} \to N$, where $J \subset N$.
- There is a total ordering on $\mathcal{G}$ such that $G, H \in \mathcal{G}$ and $q(G) \in H$ imply G precedes H in the ordering.
- The probability mass-density function of Y factors as

  $$f_\theta(y) = \prod_{G \in \mathcal{G}} f_{\theta_G}(y_G \mid y_{q(G)}) \tag{10}$$

  where the total ordering condition assures this is a valid factorization (what purport to be conditional distributions actually are).
- The conditional distribution of Y_G given $Y_{q(G)}$ is the distribution of the sum of $Y_{q(G)}$ independent and identically distributed (IID) random vectors whose distribution does not depend on Y. This is called the *predecessor is sample size* property.
- Write $Y_G = Z_1 + \cdots + Z_{Y_{q(G)}}$, where the terms of this sum are IID and where, by convention, the sum of zero terms is zero, that is $Y_{q(G)} = 0$ implies $Y_G = 0$. The distribution of the Z_k is in an exponential family of distribution having canonical statistic Z_k, canonical parameter θ_G, and cumulant function c_G.

From the theory of exponential families and the last item, the conditional distribution of Y_G given $Y_{q(G)}$ is also in an exponential family having canonical statistic Y_G, canonical parameter θ_G, and cumulant function $Y_{q(G)} c_G$.

It follows from these assumptions that the log likelihood of the family is

$$\begin{aligned} l(\theta) &= \sum_{G \in \mathcal{G}} \bigl[\langle y_G, \theta_G \rangle - y_{q(G)} c_G(\theta_G) \bigr] \\ &= \left(\sum_{j \in J} y_j \theta_j \right) - \sum_{G \in \mathcal{G}} y_{q(G)} c_G(\theta_G) \end{aligned} \tag{11}$$

Aster models are a special case of chain graph models (Lauritzen, 1996, Section 3.2.3), which are the most general models that factorize as products of marginals and conditionals. Hence we give aster models a graph as described by Lauritzen (1996). For all $G \in \mathcal{G}$ and for all $j \in G$, the graph has a directed edge (arrow) from $q(G)$ to j. For all $G \in \mathcal{G}$ and for all pairs $i, j \in G$, there is an undirected edge (line) between i and j. There are no other edges, directed or undirected.

3.3 Drawing Graphs

We make an exception to the definition in the preceding section when drawing graphs. We allow omission of some lines so long as the graph of lines (omitting arrows) has the same maximal connected sets. We take those maximal connected sets to be the elements of $\mathcal{G}$.

In the theory in Lauritzen (1996) omission of some lines would mean that the probability mass-density function could be further factored going from his equation (3.23) to his equation (3.24). But in aster models we do not factorize further. We keep (10). Hence the omission of some lines is harmless so long as we recover the same partition $\mathcal{G}$ from the graph.

Note that a node with no incoming lines is an element of $\mathcal{G}$ all by itself, or, more precisely, if j is such a node $\{j\} \in G$.

3.4 Infinite Divisibility

If there is an arrow $j \to k$, then we say j is the *predecessor* of k and k is a *successor* of j. Thus $q(G)$ is the predecessor of every element of G.

In order for the predecessor-is-sample-size property to hold, we must have every variable $Y_{q(G)}$ at a predecessor node be nonnegative-integer-valued. Otherwise Y_G being a random sum of $Y_{q(G)}$ IID terms makes no sense.

Except Geyer, et al. (2007) note an exception. The exponential family of distributions having cumulant function c_G is *infinitely divisible* if and only if rc_G is a cumulant function for any $r > 0$. Examples of such distributions that are currently implemented in R package `aster` are the Poisson family of distributions and the normal location family of distributions.

If we have this infinite divisibility property for all G such that $q(G) = j$, then we can allow Y_j to have any nonnegative-real-valued distribution that obeys the aster model assumptions. We just define the conditional distribution of Y_G given $Y_{q(G)}$ to be the exponential family of distributions having canonical statistic Y_G, canonical parameter θ_G, and cumulant function $Y_{q(G)} c_G$.

If the conditional distribution of Y_G given $Y_{q(G)}$ is not infinitely divisible, then $Y_{q(G)}$ must be nonnegative-integer-valued. But if the conditional distribution of Y_G given $Y_{q(G)}$ is infinitely divisible, then $Y_{q(G)}$ must be nonnegative-real-valued. The aster log likelihood is still given by (11) when some families are infinitely divisible and have nonnegative-real-valued predecessors.

3.5 The Aster Transform

Looking at (11) we see that the log likelihood is linear in y, so the aster model is an exponential family and y is the canonical statistic. But θ is not the canonical parameter, that is, subvectors of θ are canonical parameters for the conditional distributions in the factorization (10), but θ is not the canonical parameter of the (joint) distribution of the aster model.

To find the canonical parameter, we rewrite (11) collecting terms that multiply the same component of y

$$l(\theta) = \left(\sum_{j \in J} y_i \left[\theta_j - \sum_{\substack{G \in \mathcal{G} \\ q(G) = j}} c_G(\theta_G) \right] \right) - \sum_{\substack{G \in \mathcal{G} \\ q(G) \notin J}} y_{q(G)} c_G(\theta_G)$$

Then matching this up with the general exponential family log likelihood (1), we see that the components of the aster model canonical parameter vector φ are the terms in square brackets above

$$\varphi_j = \theta_j - \sum_{\substack{G \in \mathcal{G} \\ q(G) = j}} c_G(\theta_G) \tag{12}$$

and the cumulant function is the terms that are left over

$$c(\varphi) = \sum_{\substack{G \in \mathcal{G} \\ q(G) \notin J}} y_{q(G)} c_G(\theta_G) \tag{13}$$

Observe that in (13) every $q(G)$ is an initial node so every $y_{q(G)}$ is a constant random variable, so (13) defines a deterministic (nonrandom) function, as it must so that this is the cumulant function of an exponential family.

- The parameter vector θ is called the *conditional canonical parameter vector* (its subvectors are canonical parameter vectors of the conditional distributions in the factorization (10)).
- The parameter vector φ is called the *unconditional canonical parameter vector* (it is the canonical parameter vector of the (joint, unconditional) distribution of the aster model).

The aster transform is invertible. To invert it just rewrite (12) solving for θ_j

$$\theta_j = \varphi_j + \sum_{\substack{G \in \mathcal{G} \\ q(G) = j}} c_G(\theta_G) \tag{14}$$

and use (14) in any order that visits successors before predecessors, so that when calculating θ_j all components of θ_G for G such that $q(G) = j$ have already been calculated.

Since cumulant functions are infinitely differentiable (assuming a regular full exponential family), the aster transform and its inverse are both infinitely differentiable. The aster transform is a C^∞ diffeomorphism.

3.6 The Aster Transform and Fullness

Theorem 1. *The aster transform of the Cartesian product of the full canonical parameter spaces of the conditional exponential families in the aster model factorization* (10) *is the full canonical parameter space of the saturated aster model. For φ in this full canonical parameter space, the cumulant function is given by* (13)*.*

In symbols, if Θ_G is the full canonical parameter space of the exponential family having cumulant function c_G and

$$\Theta = \prod_{G \in \mathcal{G}} \Theta_G$$

and h denotes the aster transform $\theta \mapsto \varphi$, then $\Phi = h(\Theta)$ is the full canonical parameter space of the aster model.

Proof. Let $G_1, G_2, \ldots, G_m$ an enumeration of $\mathcal{G}$ such that $q(G_i) \in G_j$ implies $i < j$, the existence of such an enumeration being asserted by the third item in the list of assumptions in Section 3.2. Let $\prec$ be a total order on $\mathcal{G}$ defined by $G_1 \prec G_2 \prec \cdots \prec G_m$. Define

$$\mathcal{L}_i = \{\, G \in \mathcal{G} : G \prec G_i \,\}, \qquad i = 0, 1, \ldots, m$$

so $\mathcal{L}_0$ is empty and $\mathcal{L}_m = \mathcal{G}$.

We also use the notation in the comments between the theorem statement and proof. Fix θ^* in Θ, and let φ^* be its aster transform.

We prove the theorem by mathematical induction using the induction hypothesis that

$$e^{c(\varphi) - c(\varphi^*)} = E_{\varphi^*} \left\{ \prod_{\substack{G \in \mathcal{L}_i \\ q(G) \notin \mathcal{L}_i}} e^{Y_{q(G)} \left[c_G(\theta_G) - c_G(\theta_G^*) \right]} \prod_{H \in \mathcal{G} \setminus \mathcal{L}_i} e^{\langle Y_H, \varphi_H - \varphi_H^* \rangle} \right\} \tag{15}$$

provided that $\theta_G \in \Theta_G$ for all $G \in \mathcal{L}_i$, and otherwise $c(\varphi) = \infty$.

In case $i = 0$ or $i = m$, one of the products in (15) is empty. By convention, an empty product is equal to one.

The case $i = 0$ of the induction hypothesis is true by (2). The case $i = m$ of the induction hypothesis is the theorem statement, because then $\mathcal{L}_i = \mathcal{G}$, so the second product is empty, every $Y_q(G)$ in the first product is a constant random variable, the expectation is trivial, and (15) is equivalent to

$$c(\varphi) - c(\varphi^*) = \sum_{\substack{G \in \mathcal{G} \\ q(G) \notin \mathcal{G}}} Y_{q(G)} \left[c_G(\theta_G) - c_G(\theta^*_G) \right]$$

which agrees with (13). Moreover, the "provided" part of the induction hypothesis, case $i = m$, says $c(\varphi) < \infty$ if and only if $\theta \in \Theta$.

So it only remains to establish the induction step. Suppose case i of the induction hypothesis is true, where $0 \le i < m$. By (12) we have

$$\varphi_{G_{i+1}} - \varphi^*_{G_{i+1}} = \theta_{G_{i+1}} - \theta^*_{G_{i+1}} - \sum_{\substack{G \in \mathcal{G} \\ q(G) \in G_i}} \left[c_G(\theta_G) - c_G(\theta^*_G) \right]$$

and

$$\begin{aligned}
e^{c(\varphi) - c(\varphi^*)} &= E_{\varphi^*} \left\{ \prod_{\substack{G \in \mathcal{L}_i \\ q(G) \notin \mathcal{L}_i}} e^{Y_{q(G)} \left[c_G(\theta_G) - c_G(\theta^*_G) \right]} \prod_{H \in \mathcal{G} \setminus \mathcal{L}_i} e^{\langle Y_H, \varphi_H - \varphi^*_H \rangle} \right\} \\
&= E_{\varphi^*} \left\{ \prod_{\substack{G \in \mathcal{L}_i \\ q(G) \notin \mathcal{L}_{i+1}}} e^{Y_{q(G)} \left[c_G(\theta_G) - c_G(\theta^*_G) \right]} \right. \\
&\qquad \left. \times e^{\langle Y_{G_{i+1}}, \theta_{G_{i+1}} - \theta^*_{G_{i+1}} \rangle} \times \prod_{H \in \mathcal{G} \setminus \mathcal{L}_{i+1}} e^{\langle Y_H, \varphi_H - \varphi^*_H \rangle} \right\} \\
&= E_{\varphi^*} \left\{ E_{\varphi^*} \left[\prod_{\substack{G \in \mathcal{L}_i \\ q(G) \notin \mathcal{L}_{i+1}}} e^{Y_{q(G)} \left[c_G(\theta_G) - c_G(\theta^*_G) \right]} \right. \right. \\
&\qquad \left. \left. \times e^{\langle Y_{G_{i+1}}, \theta_{G_{i+1}} - \theta^*_{G_{i+1}} \rangle} \times \prod_{H \in \mathcal{G} \setminus \mathcal{L}_{i+1}} e^{\langle Y_H, \varphi_H - \varphi^*_H \rangle} \, \middle| \, Y_{q(G_{i+1})} \right] \right\}
\end{aligned}$$

and, because $Y_{G_{i+1}}$ is conditionally independent of $Y_{\mathcal{G} \setminus \bigcup \mathcal{L}_{i+1}}$ given $Y_{q(G_{i+1})}$ by the factorization (10), the conditional expectation above factors as

$$
\begin{aligned}
& E_{\varphi^*} \left[e^{\langle Y_{G_{i+1}}, \theta_{G_{i+1}} - \theta^*_{G_{i+1}} \rangle} \middle| Y_{q(G_{i+1})} \right] \\
& \quad \times E_{\varphi^*} \left[\prod_{\substack{G \in \mathcal{L}_i \\ q(G) \notin \mathcal{L}_{i+1}}} e^{Y_{q(G)} \left[c_G(\theta_G) - c_G(\theta^*_G) \right]} \prod_{H \in \mathcal{G} \setminus \mathcal{L}_{i+1}} e^{\langle Y_H, \varphi_H - \varphi^*_H \rangle} \middle| Y_{q(G_{i+1})} \right]
\end{aligned}
$$

and

$$
E_{\varphi^*} \left[e^{\langle Y_{G_{i+1}}, \theta_{G_{i+1}} - \theta^*_{G_{i+1}} \rangle} \middle| Y_{q(G_{i+1})} \right] = e^{Y_{q(G_{i+1})} \left[c_{G_{i+1}}(\theta_{G_{i+1}}) - c_{G_{i+1}}(\theta^*_{G_{i+1}}) \right]}
$$

Putting this all together gives case $i + 1$ of the induction hypothesis, so that finishes the proof. □

3.7 The Aster Transform and Regularity

Theorem 2. *An aster model is a regular full exponential family of distributions if each conditional family in the factorization* (10) *is a regular full exponential family.*

Proof. The aster transform, being a diffeomorphism, maps open sets to open sets. Using the notation established following the statement of Theorem 1, if each Θ_G is an open set, then Θ is an open set, hence Φ is an open set, and Theorem 1 says that Φ is the full canonical parameter space. □

3.8 Canonical Affine Submodels of Aster Models

3.8.1 Two Kinds

In aster models, canonical affine submodels come in two kinds, which are called *conditional* and *unconditional*.

3.8.2 Unconditional

The latter models the unconditional canonical parameter as an affine function of regression coefficients

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

Since φ is the canonical parameter vector of the aster model (of its joint distribution), the theory of Section 2.3 applies and an unconditional canonical affine submodel is a regular full exponential family if the saturated model is. And Theorems 1 and 2 tell us that happens if each conditional family for each dependence group is regular and full.

And every family implemented in R packages `aster` and `aster2` is regular and full. So every unconditional aster model that can be fitted using these packages is regular and full.

3.8.3 Conditional

R packages `aster` and `aster2` also allow submodels that parameterize the conditional canonical parameter as an affine function of regression coefficients

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

Since θ is not vectorwise canonical, none of the theory above applies to these models. Since the aster transform is smooth (infinitely differentiable), these submodels are smooth submodels of the saturated model. Hence they are so-called *curved exponential families*. But curved exponential families have no nice exponential family properties. In particular, they do not have the observed-equals-expected property. Indeed, they do not have submodel canonical statistics. In general, the full exponential family generated by a canonical affine submodel is the saturated model. Thus discussion of fullness and regularity in the context of conditional aster models is pointless.

References

Barndorff-Nielsen, O. (1978). *Information and Exponential Families.* Wiley, Chichester.

Geyer, C. J. (1990). *Likelihood and Exponential Families.* PhD thesis, University of Washington. `http://purl.umn.edu/56330`.

Geyer, C. J. and Møller, J. (1994). Simulation procedures and likelihood inference for spatial point processes. *Scandinavian Journal of Statistics*, **21**, 359–373.

Geyer, C. J., Wagenius, S., and Shaw, R. G. (2005). Aster models for life history analysis. Technical Report No. 644. School of Statistics, University of Minnesota. `http://www.stat.umn.edu/geyer/aster/`.

Geyer, C. J., Wagenius, S., and Shaw, R. G. (2007). Aster models for life history analysis. *Biometrika*, **94**, 415–426.

Geyer, C. J. (2009). Likelihood inference in exponential families and directions of recession. *Electronic Journal of Statistics*, **3**, 259–289.

Geyer, C. J. (2017). R package `aster2` (Aster Models), version 0.3. `http://cran.r-project.org/package=aster2`.

Geyer, C. J. (2018). R package `aster` (Aster Models), version 1.0.2. `http://cran.r-project.org/package=aster`.

Lauritzen, S. L. (1996). *Graphical Models.* Oxford University Press, New York.

Rockafellar, R. T. (1970). *Convex Analysis.* Princeton University Press, Princeton, NJ.

Rockafellar, R. T., and Wets, R. J.-B. (1998). *Variational Analysis.* Springer-Verlag, Berlin. (The corrected printings contain extensive changes. We used the third corrected printing, 2010.)