

Square Root Graphical Models: Multivariate Generalizations of Univariate Exponential Families that Permit Positive Dependencies

Samson Koelle

Department of Statistics, University of Washington Seattle, WA, 98195, USA

1 Introduction

Univariate exponential families include many of the most commonly encountered probability distributions, such as Normal, Poisson, Bernoulli, exponential, and Chi-squared, and are of interest because they "minimize the amount of prior information built into the distribution", are often the limits of physical systems over time, encapsulate diverse behaviors, and have sufficient statistics of fixed size that summarize all observational information relevant to inferring their parameters. Given their major role in univariate statistics, and particularly their concise optimization using sufficient statistics, extending these distributions to a multivariate setting is a logical goal. While no definition of "multivariate exponential family" distribution has solidified, there is significant work indicating the form such a distribution might take. Square Root Graphical Models, proposed by Inouye et al, dominate previous multivariate exponential family proposals, and enable construction of interactions that were previously precluded.

The factorization of multivariate probability measures over the cliques of an associated graph, shown via the Hammersley-Clifford Theorem, guarantees that we can construct a global probability measure given clique-specific sufficient statistics, or "compatibility functions". Since we often understand what univariate exponential family distribution with which to model a particular parameter, we may desire that node-conditional distributions are from a univariate exponential family. On the other hand, if univariate marginal distributions are not exponential families, and therefore (by the theorem of Pitman-Koopman-Darmois) do not have bounded sets of node-specific sufficient statistics, then the number of potential interactions between which will explode since the dimensionalities of the sufficient statistics

are each themselves growing. For these reasons, the calculation of compatibility functions from univariate node-conditional exponential family distributions is therefore the crucial issue when constructing multivariate distributions over a graph.

This approach in fact has its root in node-wise regression methods for learning these models (Meinshausen and Bühlmann, 2006; Ravikumar et al., 2010; Jalali et al., 2011), in which for each site s in a graph $G_{p,e}$ with p vertices, the interaction terms $\Phi_{-s} \in \mathbb{R}^{p-1}$ are learned through the regression $X_s = X_{-s}\Phi_{-s}$ where $X_s \in \mathbb{R}^n$. Since we often assume sparsity in our edge set, it can be desirable to limit the number of edges through use of either an information theoretic method like BIC, or a penalization. In the second case, an l_1 penalty on interaction complexity corresponding to a laplace prior on each interaction has seen particular use due to its convex geometry, which enables rapid computation of reasonably accurate estimates for the existence of an interaction between two node-specific sufficient statistics. "node-wise fitting methods have been shown to have strong computational as well as statistical guarantees" (look into further), and the two estimates for each interaction are guaranteed to asymptotically converge.

Dependencies between parameters are often interesting in their own right or as part of the regression problem of optimizing θ in $Y \approx \mathbb{X}\theta$ where $\mathbb{X} = \{X_1, \dots, X_n\}$. Note that inferring dependencies locally is also a regression problem $X_a \approx \mathbb{X}_{-a}\theta_a$, where \mathbb{X}_{-a} are adjacent sites of parameter a and θ_a are their edge weights [?]. Previous work has focused on learning these models in the multivariate Gaussian, multivariate Ising, and discrete settings, as well as mixed settings in which certain variables are Gaussian and others are Ising [Friedman et al., 2008] [Lau] [Bento and Montanari, 2011].

The probability density of the univariate exponential family is $p_x(\theta) = \exp \eta(\theta)T(x) + B(x) - A(\theta)$ where $T(x)$ is the sufficient statistic, $\eta(\theta)$ is the "natural parameter", $B(x)$ is the "base measure", and $A(\theta)$ is the normalizing constant. The initial proposal of [?] for creating a multivariate exponential family distribution was to set the natural parameter of each node-conditional univariate exponential family distribution equal to a linear combination of

interaction terms as follows:

$$\begin{aligned}
P(X_r|X_{vnr}) = & \exp T(X_r)(\theta_r + \sum_{t \in N(r)} (\theta_{rt}T(X_t)) + \sum_{t_1, t_2 \in N(r)} (\theta_{rt_1 t_2}T(X_{t_1})T(X_{t_2})) + \dots \\
& + \sum_{t_1, \dots, t_k \in N(r)} (\theta_{rt_1 \dots t_k}T(X_{t_1}) \dots T(X_{t_k})) + B(X_r) - A_{node}(X_{vnr}))
\end{aligned}$$

These conditionals then specify a unique joint distribution through the Hammersley-Clifford Theorem (Lauritzen 1996, Besag 1974). Within this general framework, restricting to pairwise graphical models, whose compatibility functions factor into cliques of size at most two, enables computation of the previously characterized multivariate Ising/Potts and Gaussian Markov Random Field models using the structure $P(x|\Theta, \Phi) = h(x) \exp(\Theta^T T(x) + T(x)^T \Phi T(x) - A(\Theta, \Phi))$ where $T(x) \in \mathbb{R}^p$ and $\Phi \in \mathbb{R}^{p \times p}$. When the sufficient statistic is a linear function $T(X) = X$, then the order two multivariate exponential family is in fact a generalized linear model, and this linear effect node-conditional model can be extended to graphical models in which each site in the graph may have a probability density corresponding to different exponential families, such as Bernoulli and Normal.

Though the parameterization in the previous paragraph has a desirable simplicity, it is not clear that this is in fact the ideal definition of a multivariate exponential family. Due to the normalizability constraint $A(\Theta, \Phi) < \infty$, this model restricts interactions $\Theta_{rt} \leq 0$ where r and t are sites with exponential or Poisson condition distributions. This can be seen in the exponential case since the interaction term is order $T(X)^2$, and through a similar argument for Poisson. Other attempts to create a multivariate Poisson distribution have been unsuccessful for alternate reasons such computational complexity, restrictions to positive-interactions, or unnatural transforms applied to sufficient statistics (Inouye 2013, 11-16).

Given that the major restrictions in defining a multivariate exponential family are the ability to recover independent univariate exponential families, univariate node-conditional exponential family distributions, and inference using node-wise regression (conditions 2 and 3 are synonymous in the pairwise-graphical model setting?), Inouye et al have proposed the square root graphical model $P(x|\Theta, \Phi) = h(x) \exp(\Theta^T \sqrt{T(x)} + \sqrt{T(x)}^T \Phi \sqrt{T(x)} - A(\Theta, \Phi))$. While these distributions admit node and radial factorizations analogous to Yang's, a key

difference that independent order $T(X)$ effects are now encoded in the self-interaction term $\Phi_{s,s}$, while the square-root self-interaction term (this is the main thing! still working on it).

Notes: I am damn curious if the univariate properties mentioned in the first sentence are satisfied by the Yang MVEF or SRGMs count data AND MORE! My favorite bedtime reading is ? . I also like to keep a copy of [Bento and Montanari, 2011] on my coffee table.

2 Methods

The multivariate sufficient statistics introduced by

$$\sum_{i=1}^n X_i \text{ for Poisson and exponential}$$

3 Results

4 Discussion

[[

References

LauWer89_Graphical_models.pdf.

Jose Bento and Andrea Montanari. Which graphical models are difficult to learn?, 2011.

Jerome Friedman, Trevor Hastie, and Robert Tibshirani. Sparse inverse covariance estimation with the graphical lasso. *Biostatistics*, 9(3):432–441, July 2008.

Eunho Yang, Pradeep Ravikumar, Genevera I Allen, and Zhandong Liu. On graphical models via univariate exponential family distributions. 17 January 2013.