

A note on the hyper Dirichlet distribution

Alcune osservazioni sulla distribuzione iper Dirichlet

Valentina Leucari^{1 2}
Dipartimento di Matematica
Università di Pavia
vl@dimat.unipv.it

Abstract: Dawid and Lauritzen (1993) introduce the hyper Dirichlet family for the parameter of a discrete decomposable graphical model. They describe an algorithm for the construction of such a family, which requires specification of prior densities for the parameters of clique marginal models. We consider a three vertex decomposable graphical model with binary variables, and show that the hyper Dirichlet family can be derived from the standard conjugate family on the canonical parameter of the underlying natural exponential family, exploiting conjugacy properties of the parametrisations.

Keywords: Conjugate parametrisations; Decomposable models; Hyper Dirichlet; Hyper Markov property; Natural exponential family.

1. Introduction

Consider the decomposable undirected graph $\mathcal{G} = (V, E)$ where $V = \{1, 2, 3\}$ and $E = \{(1, 2), (2, 3)\}$. Let $C_1 = \{1, 2\}$, $C_2 = \{2, 3\}$, and $S = \{2\}$ denote the two cliques and the separator of $\mathcal{G}$, respectively. Binary random variables $X = (X_1, X_2, X_3)$ are associated to vertices in $\mathcal{G}$, and we denote with $\mathbf{p} = \{p_{ijk}, (i, j, k) \in \{0, 1\}^3\}$ the set of joint probabilities for X , where $p_{ijk} = \Pr\{X_1 = i, X_2 = j, X_3 = k | \mathbf{p}\}$ and $p_{000} = 1 - \sum_{(i,j,k) \neq (0,0,0)} p_{ijk}$. Similarly, for any pair of subsets U, W of V , $\mathbf{p}^U$ and $\mathbf{p}^{W|U}$ are the set of U -marginal and $W|U$ -conditional probabilities respectively. We consider a parametric family of distributions for X , $\{f_X(\mathbf{x} | \mathbf{p}), \mathbf{p} \in \mathcal{P}\}$, such that each probability mass function (pmf) in the family satisfies the Markov property w.r.t. $\mathcal{G}$, namely $X_U \perp\!\!\!\perp X_W | X_{U \cap W}, \mathbf{p}$ for any decomposition (U, W) of $\mathcal{G}$. Assume that a parametric family of priors, indexed by α , is assigned on $\mathbf{p}$. We consider prior families satisfying the strong hyper Markov property w.r.t. $\mathcal{G}$, namely $\mathbf{p}^U \perp\!\!\!\perp \mathbf{p}^{W|U} | \alpha$ for any decomposition (U, W) of $\mathcal{G}$. Dawid and Lauritzen (1993) describe an algorithm for the construction of a hyper Dirichlet prior for $\mathbf{p}$ based on specification of hyper compatible Dirichlet densities on the parameters indexing the clique marginal models. We show that this prior coincides with that induced from a standard conjugate prior on the canonical parameter of the exponential family describing the graphical model. For an account of the theory of graphical models needed in this paper see Lauritzen (1996), Chapter 1-3, whereas Gutiérrez-Peña and Smith (1995) contains relevant material on the theory of exponential families and conjugate priors.

¹Indirizzo per corrispondenza: Dipartimento di Matematica, Università di Pavia, Via Ferrata 1, 27100 Pavia.

²Il lavoro è stato finanziato nell'ambito del progetto di ricerca PRIN03, "Reti bayesiane ed inferenza causale: metodologia e applicazioni". Ringrazio Alberto Roverato, Università di Modena e Reggio Emilia, per i suggerimenti relativi alla dimostrazione del Lemma 1.

The structure of the paper is as follows. Section 2 describes the NEF associated with $\mathcal{G}$ and derives the hyper Dirichlet family, while Section 3 contains some remarks and conjectures for future work.

2. Exponential family representation and hyper Dirichlet prior

Consider the clique and separator marginal probabilities $\mathbf{p}^{C_1} = \{p_{ij+}, (i, j) \in \{0, 1\}^2\}$, $\mathbf{p}^{C_2} = \{p_{+jk}, (j, k) \in \{0, 1\}^2\}$, $\mathbf{p}^S = \{p_{+j+}, j \in \{0, 1\}\}$, where p_{00+} , p_{+00} and p_{+0+} are redundant. By the Markov property $p_{ijk} = \frac{p_{ij+}p_{+jk}}{p_{+j+}}$, so the joint probability mass function for X can be written

$$\begin{aligned} f_X(\mathbf{x}|\mathbf{p}) &= \prod_{(i,j,k) \in \{0,1\}^3} p_{ijk}^{\mathbf{1}_i(x_1)\mathbf{1}_j(x_2)\mathbf{1}_k(x_3)} \\ &= \frac{f_{X_{C_1}}(x_1, x_2|\mathbf{p}^{C_1})f_{X_{C_2}}(x_2, x_3|\mathbf{p}^{C_2})}{f_{X_S}(x_2|\mathbf{p}^S)} \end{aligned}$$

where $\mathbf{1}_z$ is the indicator function for z , $\mathbf{1}_0(x_1) = 1 - \mathbf{1}_1(x_1)$ (similarly for $\mathbf{1}_0(x_2)$ and $\mathbf{1}_0(x_3)$), and $f_{X_{C_1}}$, $f_{X_{C_2}}$ and f_{X_S} are the clique and separator marginal pmfs. Given a sample (X_{1i}, X_{2i}, X_{3i}) , $i = 1, \dots, n$, the pmf above can be written as a five dimensional natural exponential family

$$f_Y(\mathbf{y}|\boldsymbol{\theta}) = \exp\{\boldsymbol{\theta}^\top \mathbf{y} - nM(\boldsymbol{\theta})\} \quad (1)$$

where

$Y = (Y_1, Y_2, Y_3, Y_4, Y_5) = (\sum_{i=1}^n X_{1i}, \sum_{i=1}^n X_{2i}, \sum_{i=1}^n X_{3i}, \sum_{i=1}^n X_{1i}X_{2i}, \sum_{i=1}^n X_{2i}X_{3i})$ and $\boldsymbol{\theta}$ are the canonical statistic and parameter respectively, and $M(\boldsymbol{\theta}) = \log \int e^{\boldsymbol{\theta}^\top \mathbf{y}} d\mathbf{y}$. The mean parameter $\mathbb{E}(Y|\boldsymbol{\theta})$ is denoted by $\boldsymbol{\mu}$. More precisely

$$\begin{aligned} \theta_1 &= \log \frac{p_{10+}}{p_{00+}} & \mu_1 &= n(p_{11+} + p_{10+}) \\ \theta_2 &= \log \frac{p_{01+}}{p_{00+}} + \log \frac{p_{+10}}{p_{+00}} - \log \frac{p_{+1+}}{p_{+0+}} & \mu_2 &= n(p_{11+} + p_{01+}) \\ & & &= n(p_{+11} + p_{+10}) \\ \theta_3 &= \log \frac{p_{+01}}{p_{+00}} & \mu_3 &= n(p_{+11} + p_{+01}) \\ \theta_4 &= \log \frac{p_{11+}p_{00+}}{p_{10+}p_{01+}} & \mu_4 &= np_{11+} \\ \theta_5 &= \log \frac{p_{+11}p_{+01}}{p_{+10}p_{+01}} & \mu_5 &= np_{+11}. \end{aligned}$$

Similarly, any clique and separator marginal pmf can be written as a NEF, where (with the obvious notation) $Y^{C_1} = (Y_1, Y_2, Y_3)$, $\boldsymbol{\theta}^{C_1} = (\theta_1, \theta_2^{C_2}, \theta_3)$, $\boldsymbol{\mu}^{C_1} = (\mu_1, \mu_2, \mu_3)$, $Y^{C_2} = (Y_2, Y_3, Y_5)$, $\boldsymbol{\theta}^{C_2} = (\theta_2^{C_2}, \theta_3, \theta_5)$, $\boldsymbol{\mu}^{C_2} = (\mu_2, \mu_3, \mu_5)$, $Y^S = Y_2$, $\boldsymbol{\theta}^S = \theta_2^S$, $\boldsymbol{\mu}^S = \mu_2$, and $\theta_2^{C_1} = \log \frac{p_{01+}}{p_{00+}}$, $\theta_2^{C_2} = \log \frac{p_{+10}}{p_{+00}}$, $\theta_2^S = \log \frac{p_{+1+}}{p_{+0+}}$. In order to prove the subsequent results, we point out that Y can be viewed as a vector of random variables associated to a graph $\mathcal{G}' = (V', E')$ where $V' = V \cup \{4, 5\}$ and $E' = E \cup \{(1, 4), (2, 4), (2, 5), (3, 5)\}$. Equivalently, $(Y_1, Y_4) \perp\!\!\!\perp (Y_3, Y_5) | Y_2, \boldsymbol{\theta}$.

Assume that $\boldsymbol{\theta}$ is assigned a standard conjugate family $\mathcal{C}_\theta$, namely

$$\pi_\theta(\boldsymbol{\theta}) \propto \exp\{\boldsymbol{\theta}^\top \mathbf{y}^0 - n^0 M(\boldsymbol{\theta})\}, \quad n^0 \in \mathbb{R}, \quad \mathbf{y}^0 \in \mathbb{R}^5.$$

Denote with $\mathcal{C}_\mathbf{p}$ the standard conjugate family for $\mathbf{p}$, and with $\mathcal{C}_\mathbf{p}^\theta$ the family of priors induced by $\mathcal{C}_\theta$ on $\mathbf{p}$, see Gutiérrez-Peña and Smith (1995) for definitions. Note that $\mathcal{C}_\mathbf{p}$ is the hyper Dirichlet family. We show that $\mathcal{C}_\mathbf{p} = \mathcal{C}_\mathbf{p}^\theta$. We first need the following result. Denote with $V(\boldsymbol{\mu})$, $V(\boldsymbol{\mu}^{C_i})$, $V(\boldsymbol{\mu}^S)$ the variance function for Y , Y^{C_i} , $i = 1, 2$ and Y^S .

Lemma 1. Let $\mathcal{G} = (V, E)$ be a three vertex decomposable graph with cliques $C_1 = \{1, 2\}$ and $C_2 = \{2, 3\}$. Let $J_\theta(\boldsymbol{\mu})$ be the Jacobian matrix for the transformation from θ to $\boldsymbol{\mu}$. Let $J_{\theta^{C_i}}(\boldsymbol{\mu}^{C_i})$, $i = 1, 2$, and $J_{\theta^S}(\boldsymbol{\mu}^S)$ be the Jacobian matrices for the corresponding transformations in the clique and separator marginal models. Then

$$\det(J_\theta(\boldsymbol{\mu})) = \frac{\det(J_{\theta^{C_1}}(\boldsymbol{\mu}^{C_1})) \det(J_{\theta^{C_2}}(\boldsymbol{\mu}^{C_2}))}{\det(J_{\theta^S}(\boldsymbol{\mu}^S))}. \quad (2)$$

Proof. Consider the matrix $J_\theta(\boldsymbol{\mu}) = V^{-1}(\boldsymbol{\mu})$ and notice that $\frac{\partial \theta_1}{\partial \mu_4} = \frac{\partial \theta_1}{\partial \mu_5} = \frac{\partial \theta_2}{\partial \mu_4} = \frac{\partial \theta_2}{\partial \mu_5} = 0$. Therefore, the Jacobian matrix is symmetric and has zero entries corresponding to conditional independencies in the graph $\mathcal{G}'$, so that Lemma 5.5 in Lauritzen (1996) can be applied to the decomposition $(V'_1 = \{1, 2, 4\}, V'_2 = \{2, 3, 5\})$, yielding

$$\begin{aligned} \det(J_\theta(\boldsymbol{\mu})) &= \frac{\det([J_\theta^{-1}(\boldsymbol{\mu})]_{V'_1 \cap V'_2})}{\det([J_\theta^{-1}(\boldsymbol{\mu})]_{V'_1}) \det([J_\theta^{-1}(\boldsymbol{\mu})]_{V'_2})} \\ &= \frac{\det([V(\boldsymbol{\mu})]_{V'_1 \cap V'_2})}{\det([V(\boldsymbol{\mu})]_{V'_1}) \det([V(\boldsymbol{\mu})]_{V'_2})} \end{aligned} \quad (3)$$

where the notation $[A]_{V'_i}$ identifies the submatrix of A with entries $(i, j) \in V'_1 \times V'_1$, and similarly for the other submatrices. It can be shown that each subvector Y^{C_1} , Y^{C_2} , Y^S is a cut for Y , so that submatrices of $V(\boldsymbol{\mu})$ corresponding to C_1 , C_2 , S are just the marginal variance functions. Moreover, notice that V'_1 , V'_2 and $V'_1 \cap V'_2$ are the vertices of the subgraphs induced by C_1 , C_2 and S respectively. Therefore, (3) becomes

$$\begin{aligned} \det(J_\theta(\boldsymbol{\mu})) &= \frac{\det(V^{-1}(\boldsymbol{\mu}_{C_1})) \det(V^{-1}(\boldsymbol{\mu}_{C_2}))}{\det(V^{-1}(\boldsymbol{\mu}_S))} \\ &= \frac{\det(J_{\theta^{C_1}}(\boldsymbol{\mu}^{C_1})) \det(J_{\theta^{C_2}}(\boldsymbol{\mu}^{C_2}))}{\det(J_{\theta^S}(\boldsymbol{\mu}^S))} \end{aligned}$$

and the proof is complete.

The following facts are also useful: i) for simple quadratic variance function (SQVF) NEFs the canonical parameter and mean parameter are conjugate (i.e. $\mathcal{C}_\mu = \mathcal{C}_\mu^\theta$), and ii) $\mathcal{C}_\mu = \mathcal{C}_\mu^\theta \Leftrightarrow J_\theta(\boldsymbol{\mu}) \propto L(\boldsymbol{\mu}|\mathbf{y}) = f_Y(\mathbf{y}|\boldsymbol{\theta}(\boldsymbol{\mu}))$.

We are now ready to prove the main result. We write $L_{C_i}(\boldsymbol{\mu}^{C_i}|\mathbf{y}^{C_i})$, $i = 1, 2$, and $L_S(\boldsymbol{\mu}^S|\mathbf{y}^S)$ for the marginal likelihood functions.

Proposition 2. Let $\mathcal{G} = (V, E)$ be a three vertex decomposable graph with cliques $C_1 = \{1, 2\}$ and $C_2 = \{2, 3\}$. Suppose the standard conjugate family $\mathcal{C}_\theta$ is assigned on θ . Then $\mathcal{C}_p^\theta$ is the hyper Dirichlet family.

Proof. Since marginal models corresponding to C_1 , C_2 and S are all SQVF NEFs, results (i) and (ii) above hold, and from (2) we can write

$$\det(J_\theta(\boldsymbol{\mu})) \propto \frac{L_{C_1}(\boldsymbol{\mu}^{C_1}|\mathbf{y}^{C_1}) L_{C_2}(\boldsymbol{\mu}^{C_2}|\mathbf{y}^{C_2})}{L_S(\boldsymbol{\mu}^S|\mathbf{y}^S)} = L(\boldsymbol{\mu}|\mathbf{y}).$$

Using (ii), it follows that θ and $\boldsymbol{\mu}$ are conjugate parametrisations, i.e. the family induced on $\boldsymbol{\mu}$ is the standard conjugate. Notice that the support of the hyper Dirichlet distribution is spanned by a subset $\tilde{\mathbf{p}}$ of $\{\mathbf{p}^{C_1}, \mathbf{p}^{C_2}\}$. Consider now the transformation from $\boldsymbol{\mu}$ to $\tilde{\mathbf{p}}$, and note $\boldsymbol{\mu}$ is a linear function of $\tilde{\mathbf{p}}$. The determinant $J_\mu(\tilde{\mathbf{p}})$ is therefore constant and the family induced on $\tilde{\mathbf{p}}$ from $\boldsymbol{\mu}$ is the standard conjugate, i.e the hyper Dirichlet. The latter coincides with $\mathcal{C}_p^\theta$ since the property of conjugate parametrisation is transitive. The conclusion follows upon noticing that the distribution of $\mathbf{p}$ is described in terms of $\tilde{\mathbf{p}}$.

3. Discussion

We have shown that, for a three vertex binary decomposable graphical model, the hyper Dirichlet family coincides with the family induced from the standard conjugate on θ . This is based on the following facts: i) the determinant of the Jacobian matrix can be decomposed according to (2), and ii) clique and separators marginal models are SQVF NEFs. We remark that a non complete decomposable discrete graphical model is *not* a quadratic variance function family. Yet we have proved that, in our setting, θ and μ are conjugate parametrisations, thus providing a partial answer to a query raised in Gutiérrez-Peña and Smith (1995).

We conjecture that the results above can be extended to more general models. If the Markov property is assumed, as it is usually the case, for an arbitrary decomposable graphical model with random variables $X = (X_1, \dots, X_m)$, cliques $(C_1, \dots, C_k)$ and separators $(S_2, \dots, S_k)$, the following factorisation holds

$$f_X(\mathbf{x}|\mathbf{p}) = \frac{\prod_{i=1}^k f_{X_{C_i}}(\mathbf{x}_{C_i}|\mathbf{p}^{C_i})}{\prod_{i=2}^k f_{X_{S_i}}(\mathbf{x}_{S_i}|\mathbf{p}^{S_i})}.$$

Also, recall that the hyper Dirichlet for a general model (see Dawid and Lauritzen (1993)) is of the form

$$\pi_{\mathbf{p}}(\mathbf{p}) = \frac{\prod_{i=1}^k \mathcal{D}(\mathbf{p}^{C_i}|\alpha^{C_i})}{\prod_{i=2}^k B(\mathbf{p}^{S_i}|\alpha^{S_i}, \beta^{S_i})}$$

where $\mathcal{D}(\mathbf{x}|\alpha)$ denotes the Dirichlet distribution with parameter α , and $B(\mathbf{x}|\alpha, \beta)$ denotes the Beta distribution with parameters α and β . We point out that for complete graphs the joint pmf is always a SQVF NEF, and conjugacy of parametrisations is automatically satisfied. Therefore, the idea is to work locally on cliques and separators where marginal pmfs are all SQVF NEFs and properties of the variance function can be used as in Proposition 1 and 2. In order to proceed in this way, we want to obtain a factorisation for the determinant of the Jacobian matrix as

$$\det(J_{\theta}(\mu)) = \frac{\prod_{i=1}^k \det(J_{\theta^{C_i}}(\mu^{C_i}))}{\prod_{i=2}^k \det(J_{\theta^{S_i}}(\mu^{S_i}))}.$$

Moreover, we would like to identify $\tilde{\mathbf{p}}$.

References

- Dawid A.P. and Lauritzen S.L. (1993) Hyper markov laws in the statistical analysis of decomposable models, *The Annals of Statistics*, 21, 1272–1317.
 Gutiérrez-Peña E. and Smith A.F.M. (1995) Conjugate parameterizations for natural exponential families, *Journal of the American Statistical Association*, 90, 1347–1356.
 Lauritzen S.L. (1996) *Graphical Models*, Oxford.