Reference priors for exponential families with simple quadratic variance function

Guido Consonni, a Piero Veronese, b,* and Eduardo Gutiérrez-Peña c,1

a University of Pavia, Italy
b L. Bocconi University, Institute of Quantitative Methods, Viale Isonzo, 25 I-20135 Milan, Italy
c IIMAS-UNAM, Mexico

Received 24 January 2001

Abstract

Reference analysis is one of the most successful general methods to derive noninformative prior distributions. In practice, however, reference priors are often difficult to obtain. Recently developed theory for conditionally reducible natural exponential families identifies an attractive reparameterization which allows one, among other things, to construct an enriched conjugate prior. In this paper, under the assumption that the variance function is simple quadratic, the order-invariant group reference prior for the above parameter is found. Furthermore, group reference priors for the mean- and natural parameter of the families are obtained. A brief discussion of the frequentist coverage properties is also presented. The theory is illustrated for the multinomial and negative-multinomial family. Posterior computations are especially straightforward due to the fact that the resulting reference distributions belong to the corresponding enriched conjugate family. A substantive application of the theory relates to the construction of reference priors for

*Corresponding author. Fax: +39-02-58365634.
E-mail address: piero.veronese@uni-bocconi.it (P. Veronese).
1GC’s research was partially supported by the European Union TMR Grant ERB-FMRX-CT96-0096, by MURST, Rome, (9813041044 009),CNR (98.01329.CT10) and by the University of Pavia. PV’s research was partially supported by MURST, Rome (9813041044 003) and by L. Bocconi University. The research of EGP was supported by MURST (Cofin98). This work was carried out while E. Gutiérrez-Peña was visiting the Department of Economics and Quantitative Methods, University of Pavia, and the Institute of Quantitative Methods, L. Bocconi University. He is grateful to these institutions for their hospitality.

0047-259X/03/$-see front matter © 2003 Elsevier Inc. All rights reserved.
doi:10.1016/S0047-259X(03)00095-2
the Bayesian analysis of two-way contingency tables with respect to two alternative parameterizations.

AMS 2000 subject classifications: primary 62F15; 62E15; secondary 62H17

Keywords: Bayesian inference; Conditional reducibility; Contingency table; Enriched conjugate prior; Multinomial family; Negative-multinomial family; Noninformative prior

1. Introduction

In the context of Bayesian inference, the objective of the so-called reference analysis is to specify a prior distribution such that, even for moderate sample sizes, the information provided by the data dominates that available a priori, because of the “vague” nature of the prior knowledge. Reference analysis, introduced by Bernardo [7] and further developed by Berger and Bernardo [5], provides one of the most successful general methods to derive noninformative prior distributions. In practice, however, reference priors are typically difficult to obtain. In this paper, we find reference priors for an important class of natural exponential families (NEFs), namely those having a simple quadratic variance function (SQVF). In the next section, we briefly describe how to find reference priors when certain simplifying conditions on the sampling model are met. In Section 3 we review some basic facts concerning natural exponential families and quadratic variance functions. We also describe the concept of conditional reducibility and extend the notion of enriched conjugate priors introduced in [13]. The main results are presented in Section 4. More precisely, we derive the general expression of the reference prior for several parameterizations, including the mean- and the natural-parameter, and find conditions for the reference posterior to be proper. In particular, we explicitly identify reference priors for the basic NEF-SQVFs (which include the multinomial and negative-multinomial families) and prove that the corresponding reference posteriors are always proper. A brief discussion of the frequentist coverage properties is included. Section 5 presents a substantive application of the theory to the analysis of two-way contingency tables and, finally, Section 6 offers some concluding remarks. To ease the flow of ideas all proofs are collected in the Appendix.

2. Ordered-group reference priors

Let \( X \) be an observable random quantity with density function \( p(x|\phi) \), where \( \phi \in \Phi \subseteq \mathbb{R}^d \) denotes an unknown parameter.

Berger and Bernardo [5] motivate and describe a general algorithm to find reference priors for the parameter \( \phi \). Such an algorithm is greatly simplified if the posterior distribution of \( \phi \) is asymptotically normal (the so-called regular case). We
now briefly review the main points of their methodology. Let
\[
    H(\phi) = -E_{X|\phi} \left[ \frac{\partial^2 \ln p(X|\phi)}{\partial \phi^T \partial \phi} \right]
\]
denote the Fisher information matrix for model \( p(x|\phi) \).

We assume that \( \phi \) is decomposed into \( r \) groups \((\phi_1, \ldots, \phi_r)\) with \( \phi_k \) having dimension \( d_k \); we also let \( d^* = \sum_{j=1}^{r} d_j \) and define \( \phi[x] = (\phi^T_1, \ldots, \phi^T_{(k)} \) and \( \phi[x] = (\phi^T_{(k+1)}, \ldots, \phi^T_{(r)}) \), for all \( k = 1, \ldots, r \). The elements of \( \phi \) are usually ordered according to inferential importance; in particular, the parameters of interest should come first.

The algorithm of Berger and Bernardo [5] typically requires one to specify a nested sequence \( \{ \Phi^l \in \mathcal{N} \} \) of compact subsets of \( \Phi \) such that \( \bigcup_{l=1}^{\infty} \Phi^l = \Phi \). (This step is not necessary if the reference priors turn out to be proper.)

Let \( S(\phi) = H(\phi)^{-1} \). For each \( k = 1, \ldots, r \), denote by \( S_k(\phi) \) the upper-left \( d_k^* \times d_k^* \) submatrix of \( S(\phi) \), and let \( h_k(\phi) \) be the lower-right \( d_k \times d_k \) submatrix of \( S_k(\phi)^{-1} \). For the regular case, if \( \Phi = \Phi_1 \times \cdots \times \Phi_r \) and if there exist positive functions \( a_k(\cdot) \) and \( b_k(\cdots) \) such that
\[
    \det \{ h_k(\phi) \} = a_k(\phi(k))b_k(\phi[k], \phi[\sim k]) \quad \forall k \in \{1, \ldots, r\},
\]
then the density—with respect to the Lebesgue measure—of the \( r \)-group reference prior on \( \phi \), relative to the order \((\phi_1, \ldots, \phi_r)\), is given by
\[
    \phi(1), \ldots, \phi(r) \pi_\phi(\phi(1), \ldots, \phi(r)) \propto \prod_{k=1}^{r} a_k(\phi(k))^{1/2}
\]
for all sequences of compact subsets of the form \( \Phi^l = \Phi^l_1 \times \cdots \times \Phi^l_r \), \( l \in \mathcal{N} \); see [8,22].

The special case for which \( H(\phi) = \text{Diag}\{H_{11}(\phi), \ldots, H_{rr}(\phi)\} \) with \( H_{kk}(\phi) \) a \( d_k \times d_k \) matrix, was considered in [18]. Since \( h_k(\phi) = H_{kk}(\phi) \), it follows that the prior in (2) does not depend on the ordering of the \( r \)-groups, i.e. \( \phi(i_1), \ldots, \phi(i_r) \pi_\phi(\phi(1), \ldots, \phi(r)) = \phi(1), \ldots, \phi(r) \pi_\phi(\phi(1), \ldots, \phi(r)) \), for all permutations \((i_1, \ldots, i_r)\) of \((1, \ldots, r)\). Whenever the ordering of the variables does not matter we shall simply write \( \pi_\phi(\phi(1), \ldots, \phi(r)) \).

We emphasize that the grouping and ordering of the components of the parameter are crucial for the definition of the reference prior, in the sense that a modification of either grouping or ordering will typically lead to a different prior. The following well-known example illustrates this point. Let \( p(x|\phi_1, \phi_2) = N(x|\phi_1, \phi_2^{-1}) \), that is, a normal family with mean \( \phi_1 \) and precision \( \phi_2 \). Then
\[
    H(\phi_1, \phi_2) = \begin{bmatrix}
        \phi_2 & 0 \\
        0 & \frac{1}{2} \phi_2^{-2}
    \end{bmatrix},
\]
and so the 1-group reference prior is \((\phi_1, \phi_2) \pi_\phi(\phi_1, \phi_2) \propto \phi_2^{-1/2} \), while the 2-group reference priors are given by \( \phi_1, \phi_2 \pi_\phi(\phi_1, \phi_2) = \phi_2, \phi_1 \pi_\phi(\phi_1, \phi_2) \propto \phi_2^{-1} \) (notice that the order is immaterial because \( H(\phi_1, \phi_2) \) is diagonal).
The reference prior may, in general, depend on the choice of the particular sequence of compact subsets employed in the constructing algorithm. The importance of result (2) stems from the fact that it holds for any sequence of compact subsets of the form $\Phi' = \Phi'_1 \times \cdots \times \Phi'_r$. The latter requirement, however, is most natural when $\Phi$ is the Cartesian product of the $r$-block-parameter-spaces.

We shall show in Section 4 that condition (1) holds for an important subclass of natural exponential families suitably parameterized, so that the reference prior can be easily computed using formula (2).

3. Conditionally reducible NEFS and enriched conjugate priors

3.1. NEFs having a quadratic variance function

This subsection reviews some basic facts about exponential families. For a general treatment see [2,10].

Let $\nu$ be a $\sigma$-finite positive measure on the Borel sets of $\mathbb{R}^d$. Suppose $\nu$ is not concentrated on an affine subspace of $\mathbb{R}^d$ and consider an exponential family $\mathcal{F}$ whose densities with respect to $\nu$ are of the form

$$p(x|\theta) = \exp\{\theta^T x - M(\theta)\}, \quad \theta \in \Theta,$$

with $\Theta$ nonempty. When $\Theta$ is the interior of the natural parameter space $\{\theta \in \mathbb{R}^d: \int \exp\{\theta^T x\} \nu(dx) < \infty\}$, the family $\mathcal{F}$ is said to be a natural exponential family (NEF).

Any NEF can be reparameterized in terms of the mean $\mu$ where

$$\mu = \mu(\theta) = E[X|\theta] = \frac{\partial M(\theta)}{\partial \theta}.$$

We shall assume in the paper that the family $\mathcal{F}$ is steep, so that in particular $\Omega = \mu(\Theta)$ coincides with the interior of the convex hull of the support of $\nu$. The function

$$V(\mu) = \text{Var}[X|\theta(\mu)] = \frac{\partial^2 M(\theta)}{\partial \theta^T \partial \theta}\bigg|_{\theta = \theta(\mu)}, \quad \mu \in \Omega$$

is called the variance function of the family $\mathcal{F}$. Here $\theta(\cdot)$ denotes the inverse of the mapping $\mu(\theta) = \frac{\partial M(\theta)}{\partial \theta}$. The pair $(V(\cdot), \Omega)$ characterizes the natural exponential family $\mathcal{F}$; see, for example, [25,28].

When the family $\mathcal{F}$ is real, its variance function is said to be quadratic if

$$V(\mu) = q\mu^2 + l\mu + c$$

for some $q, l, c \in \mathbb{R}$ such that $V(\mu) > 0$ for all $\mu \in \Omega$. The class of real natural exponential families with quadratic variance function includes some of the most widely used families of distributions, such as the normal (with known variance),
binomial, Poisson, gamma and negative-binomial. See [28,29] for a thorough
discussion of the properties of these models.

When $\mathcal{F}$ is defined on $\mathbb{R}^d$, there are various ways in which this concept can be
extended [25]. Let $\mathcal{M}_d$ denote the space of real symmetric $d \times d$ matrices. A variance
function $V(\mu)$, defined on $\Omega \subseteq \mathbb{R}^d$, is said to be quadratic if it has the form

$$V(\mu) = Q(\mu, \mu) + L(\mu) + C,$$

where the map $Q : \Omega \times \Omega \rightarrow \mathcal{M}_d$ is symmetric bilinear, $L : \Omega \rightarrow \mathcal{M}_d$ is linear and
$C \in \mathcal{M}_d$ is a constant matrix. An important particular case is the simple quadratic variance function (SQVF)

$$V(\mu) = q\mu\mu^T + \sum_{i=1}^d \mu_i L_i + C,$$

where $\mu = (\mu_1, \ldots, \mu_d)^T$, $q$ is a real constant and $L_i, C \in \mathcal{M}_d$ ($i = 1, \ldots, d$) are
constant matrices.

Casalis [11] has shown that any natural exponential family having a simple quadratic variance function can be obtained, via a nonsingular affine transforma-
tion, from one of the so-called $(2d + 4)$ basic families. She also provided a detailed study of the latter families, which can be grouped into five broad classes, namely: Poisson/normal, multinomial, negative-multinomial, negative-multinomial/gamma/
normal and negative-multinomial/hyperbolic secant.

The variance function of the basic SQVF families can be conveniently described as

$$V(\mu) = q\mu\mu^T + \text{Diag}(L_0\mu) + C,$$

where $q \in \mathbb{R}$, $L_0$ and $C$ are $d \times d$ matrices, and $\text{Diag}(u)$ denotes the diagonal matrix
whose entries are the elements of the vector $u$.

Kokonendji and Seshadri [24] find, for each of the basic SQVF families, the expression for the determinant of the variance function and show that it is
proportional to $\exp \{ \theta(\mu)^T z - v M(\theta(\mu)) \}$ for some $z \in \mathbb{R}^d$ and $v \in \mathbb{R}$.

In the sequel, we shall consider arbitrary NEF-SQVF's whose variance function is
described in (4). Since these families are related to the basic ones via a nonsingular affine transformation, it follows that the determinant of their variance function
admits a representation structurally identical to that of the basic families. Nevertheless, we find useful for further elaboration to relate explicitly $z$ and $v$ to the coefficients $q, L_i$ and $C$ describing $V(\mu)$ in (4).

**Proposition 1.** Let $\mathcal{F}$ be a NEF on $\mathbb{R}^d$ having a simple quadratic variance function as
in (4). Then

$$\det \{V(\mu)\} \propto \exp \{ \theta(\mu)^T z - v M(\theta(\mu)) \},$$

with $z = \sum_{i=1}^d l_i^T_1$ and $v = -q(d + 1)$, where $l_i^T$ denotes the ith column of the
matrix $L_i$. 
In particular, if $\mathcal{F}$ is one of the basic families, i.e. if it has a variance function of the form (5), then $z = \text{diag}(L_0)$, where $\text{diag}(L_0)$ is the vector of the diagonal elements of $L_0$.

### 3.2. Conditionally reducible families

We shall now describe a subfamily of NEFs, called \textit{conditionally reducible}, introduced and discussed in detail by Consonni and Veronese [13], that admits a reparameterization which enjoys useful structural properties, especially for conjugate Bayesian analysis. Such a reparameterization will be shown to be also expedient for the construction of reference priors. Conditionally reducible NEFs are a generalization of reducible NEFs; see [1,23]. As before, if $x$ is decomposed as $x = (x_T(1), \ldots, x_T(r))^T$ with $x(r) \in \mathbb{R}^{d_k}$, we set $x[k] = (x_T(1), \ldots, x_T(k))^T$, and write $d_k = d_1 + \cdots + d_k$.

**Definition 1.** Let $X$ be a random vector distributed according to a NEF $\mathcal{F}$ on $\mathbb{R}^d$, whose density with respect to $v$ is given in (3). $\mathcal{F}$ is \textit{conditionally $r$-reducible} if, for some $r \in \{1, \ldots, d\}$, there exists a decomposition $X = (X_T(1), \ldots, X_T(r))^T$ such that, for each $k = 1, \ldots, r$, the conditional distribution of $X(k)|X_{[k-1]} = x_{[k-1]}$, is an exponential family on $\mathbb{R}^{d_k}$, with respect to a transition kernel $K_k(dx(k), x_{[k-1]})$ from $(\mathbb{R}^{d_k-1}, \mathcal{B}^{d_k-1})$ to $(\mathbb{R}^{d_k}, \mathcal{B}^{d_k})$, where $\mathcal{B}^m$ denotes the $\sigma$-algebra of Borel sets of $\mathbb{R}^m$, $v(dx(1), \ldots, dx(r)) = \prod_{k=1}^r K_k(dx(k), x_{[k-1]})$, with the understanding that $K_1$ represents the dominating measure for the marginal density of $X(1)$.

It follows that the joint density of a conditionally $r$-reducible NEF can be factorized as

$$p_\theta(x_1, \ldots, x_d|\theta) = p_\phi(x_1, \ldots, x_d|\phi(\theta))$$

$$= \prod_{k=1}^r p_{\phi(k)}(x(k)|x_{[k-1]}; \phi(k)(\theta))$$

$$= \prod_{k=1}^r \exp\{\phi(k)(\theta)^T x(k) - M_k(\phi(k)(\theta); x_{[k-1]})\}$$

where $\phi = (\phi_T(1), \ldots, \phi_T(r))^T$ is a one-to-one function from $\Theta$ onto $\phi(\Theta) = \Phi$. Furthermore, it can be shown that $\Phi = \Phi_1 \times \cdots \times \Phi_r$, with $\phi(k) \in \Phi_k$, $k = 1, \ldots, r$, so that the $\phi(k)$'s are variation independent.

Notice that $\phi(k) \in \Phi_k$ represents the natural parameter associated to the $k$th conditional distribution; on the other hand, $\Phi_k$ does not necessarily coincide with the natural parameter space $\mathcal{N}_k(x_{[k-1]})$ associated to the exponential family generated through $K_k(\cdot, \cdot)$, which will in general depend on $x_{[k-1]}$. However, for the basic SQVF families one can show that $\mathcal{N}_k(x_{[k-1]})$ does not depend on $x_{[k-1]}$ and is equal to $\Phi_k$. 
This result, together with the relationship between the $\theta$ and $\phi$ parameterizations, are described in detail in Theorem 3 and Table 1 of Consonni and Veronese [13].

**Example 1** (Multinomial family). Consider the multinomial distribution on $\mathbb{R}^d$

$$p_\theta(x|\theta) = \exp\{\theta^T x - M(\theta)\} \left( \begin{array}{c} N \\ x_1, \ldots, x_{d+1} \end{array} \right),$$

where $M(\theta) = N \ln(1 + \sum_{k=1}^d e^{\theta_k})$, $\Theta = \mathbb{R}^d$ and $x_{d+1} = N - \sum_{k=1}^d x_k$, $\sum_{k=1}^d x_k \leq N$, with $x_k$ a nonnegative integer and $\theta_k = \ln(p_k/(1 - \sum_{r=1}^d p_r))$, with $p_k$ the probability of the $k$th outcome, $k = 1, \ldots, d$.

If $X = (X_1, \ldots, X_d)^T$ is distributed according to (8) then it is well known that the conditional distribution of any subset of $(X_1, \ldots, X_d)$ given the rest is still a multinomial distribution. In particular, the distribution of $(X_k|X_1 = x_1, \ldots, X_{k-1} = x_{k-1})$, $k = 2, \ldots, d$, is Binomial$(N - \sum_{j=1}^{k-1} x_j, p_k/(1 - \sum_{j=1}^{k-1} p_j))$, whereas the marginal distribution of $X_1$ is Binomial$(N, p_1)$. Since the Binomial family is a NEF with natural parameter equal to the logit of the probability of success, one can factorize the family as in (7) with $r = d$ and

$$\phi_k = \ln\frac{p_k}{1 - \sum_{j=1}^k p_j}, \quad k = 1, \ldots, d.$$  

Notice that the natural parameter space associated to the $k$th conditional distribution is $\mathbb{R}$ and so it coincide with $\Phi_k$, $k = 1, \ldots, d$.

Consider now a permutation of $(X_1, \ldots, X_d)$: the resulting distribution is of course still multinomial, and one can repeat the above argument obtaining a different “$\phi$-parameterization”. One can thus conclude that the parameter $\phi$ is specific to a given order of the vector components. This aspect can be usefully exploited to construct reference priors as we shall illustrate in the sequel.

**Remark 1.** Given a conditionally $r$-reducible family, if there exists no linear transformation which allows the family to become conditionally $r_1$-reducible, with $r_1 > r$, the family is said to be *fully* conditionally $r$-reducible. Clearly, if the family is fully conditionally $r_1$-reducible, then it is also conditionally $r$-reducible for all $r < r_1$. Because of this remark our results will be stated for a family regarded as $r$-conditionally reducible, with arbitrary $r \leq r_1$. This will provide a flexible framework for the construction of reference priors relative to various choices of ordering and grouping of the parameter components.

**Remark 2.** There exists a close relationship between the notion of conditional reducibility and that of a cut, see [2, p. 50]. For a detailed analysis the reader is referred to Consonni and Veronese [13], who show in particular that $\mathcal{F}$ is conditionally $r$-reducible if and only if the marginal distribution of $X_{[k]}$ is a NEF on $\mathbb{R}^{d_k}$; equivalently if and only if the principal $d_k^1 \times d_k^*$ matrix of the variance function does not depend on $\mu_{(k+1)}, \ldots, \mu_{(r)}$, $k = 1, \ldots, r - 1$. 


From factorization (7) it follows immediately that the Fisher information matrix for the $\phi$-parameterization, $H(\phi)$, is block-diagonal. Moreover,

$$H(\phi) = \text{Diag}\{H_{11}(\phi_{(1)}), \ldots, H_{kk}(\phi_{[k]}), \ldots, H_{rr}(\phi)\}$$

with the $k$th block only depending on $\phi_{[k]}$. This holds because: (i) the matrix of second derivatives of the log-likelihood is block diagonal with the $k$th block only depending on $\phi_{(k)}$ and $X_{[k-1]}$; (ii) the expectation of the $k$th block relative to $p_{\phi}(x|\phi)$ only involves the marginal distribution of $X_{[k-1]}$, which trivially only depends on $\phi_{[k-1]}$. The pattern of $H(\phi)$ exhibited in (9) implies that the $\phi_{(k)}$'s are totally orthogonal, see [16], and will prove very useful when constructing reference priors for conditionally reducible families, see Section 4.

A further useful property of conditionally $r$-reducible families relates to the structure of $M_k(\phi_{(k)}; x_{[k-1]})$ which can be written as

$$M_k(\phi_{(k)}; x_{[k-1]}) = \sum_{j=1}^{k-1} (A_{kj}(\phi_{(k)}))^T x_{(j)} + B_k(\phi_{(k)})$$

$$= x_{[k-1]}^T A_{k[k-1]}(\phi_{(k)}) + B_k(\phi_{(k)}),$$

for some functions $A_{kj}$ and $B_k$, with $A_{k[k-1]} = (A_{k1}^T, \ldots, A_{k(k-1)}^T)^T$. As a consequence, the conditional expectation of $X_{(k)}$ given $x_{[k-1]}$ is linear in $x_{[k-1]}$, because it is the gradient of (10).

Several useful relations have been established between pairs of alternative parameterizations for conditionally $r$-reducible families.

For example, it follows from (10) that

$$\mu_{(k)} = \frac{\partial B(\phi_{(k)})}{\partial \phi_{(k)}} + \mu_{[k-1]}^T \frac{\partial A_{k[k-1]}(\phi_{(k)})}{\partial \phi_{(k)}},$$

so that $\mu_{(k)}$ depends on $\phi$ only through $(\phi_{(1)}, \ldots, \phi_{(k)})$. When $F$ is a basic NEF-SQVF, (11) can be solved to find an explicit formula for $\mu$ in terms of $\phi$.

Furthermore, it can be checked, using (7) and (10), that

$$\theta_{(k)} = \phi_{(k)} - \sum_{u=k+1}^{r} A_{uk}(\phi_{(u)})$$

and

$$M(\theta) = \sum_{k=1}^{r} B_k(\phi_{(k)}(\theta)).$$

A consequence of (12) is that

$$\phi_{(k)} = \theta_{(k)} + g_{(k)}(\theta_{(k+1)}, \ldots, \theta_{(r)}),$$

for some vector-valued function $g_{(k)}$.

Of course, the previous formulas hold true for $k = 1, \ldots, r$, with the understanding that components that lose meaning for a specific $k$ are set to zero. The same convention will be used throughout the paper.
The situation where a NEF on $\mathbb{R}^d$ is conditionally $d$-reducible is especially interesting, because each conditional density in (7) is univariate, whence $\phi_{(k)}$ is a scalar and will be written as $\phi_k$. Each of the $(2d + 4)$ basic NEF-SQVFs is conditionally $d$-reducible.

3.3. Enriched standard conjugate families

We now propose a definition of enriched conjugate family for $\phi$, generalizing that contained in Section 4.3 of Consonni and Veronese [13].

Let $X_1 = x_1, \ldots, X_n = x_n$ be a random sample from a conditionally $r$-reducible NEF on $\mathbb{R}^d$. Using (7) and (10), the likelihood function can be written as

$$L_\phi(\phi|s,n) = \prod_{k=1}^r \exp\{\phi^T_{(k)} s_{(k)} - [s^T_{(k-1)} A_{k|k-1}(\phi_{(k)}) + n B_k(\phi_{(k)})]\},$$

(14)

where $s = (s^T_1, \ldots, s^T_r)^T$ and $s_{(j)} = \sum_{i=1}^n x_{(j)}^i$.

The structure of (14) suggests to take a separate conjugate prior for each term in order to recover a joint distribution for $\phi$.

**Definition 2.** Let $\mathcal{F}$ be a conditionally $r$-reducible NEF on $\mathbb{R}^d$. A family of measures on the Borel sets of the space $\Phi$ whose densities with respect to Lebesgue measure are of the form

$$\pi^*_\phi(\phi'|s',n') \propto \prod_{k=1}^r \exp\{\phi'^T_{(k)} s'_{(k)} - [s'^T_{(k-1)} A_{k|k-1}(\phi_{(k)}) + n' B_k(\phi_{(k)})]\},$$

(15)

where $s' = (s'^T_1, \ldots, s'^T_r)^T$, $s'^{k'} = (s'^{k'}_1, \ldots, s'^{k'}_r)^T$, $k = 1, \ldots, r; s'^{k'} \in \mathbb{R}^{d^r}$ and $n' = (n'_1, \ldots, n'_r)^T, n' \in \mathbb{R}^r$, is called the *enriched standard conjugate* family for $\mathcal{F}$ (relative to $\phi$) and denoted with $\mathcal{E}_\phi(\mathcal{F})$.

It is worth noticing that under the enriched standard conjugate family the parameters $\phi_{(k)}$ are stochastically independent.

The family of priors on an arbitrary parameter $\lambda$ induced by (15), $\mathcal{E}_\phi(\mathcal{F})$, is called the *induced enriched standard conjugate family* (relative to $\lambda$) and has density

$$\pi^*_\lambda(\lambda'|s',n') = \pi^*_\phi(\phi(\lambda)|s',n')|J_\phi(\lambda)|,$$

(16)

where $|J_\phi(\lambda)|$ is the absolute value of the Jacobian of the transformation $\phi(\lambda)$.

An important case occurs when $\lambda = \mu$, the mean parameter. For example, for the multinomial and negative multinomial families, (15) induces a prior on the “cell probabilities”-parameter which generalizes the standard conjugate Dirichlet family, see [13].

Another important instance is given by $\lambda = \theta$, the natural parameter. In this case it follows from Eq. (13) that $|J_\phi(\theta)| = 1$. If in particular $s'_{(j)} = s'_{(j)}$, $j = 1, \ldots, k$;
\( k = 1, \ldots, r \), one gets

\[
\pi_0^\phi(\theta | s', n') \propto \exp \left\{ \sum_{k=1}^{r} \left[ \phi_k(\theta)^T s'_k - s'_{k-1}^T A_{k|k-1}(\phi_k(\theta)) - n_k' B_k(\phi_k(\theta)) \right] \right\}
\]

\[
= \exp \left\{ \sum_{k=1}^{r} \left[ \phi_k(\theta)^T \sum_{u=k+1}^{r} A_{uk}(\phi_u(\theta)) s'_k - \sum_{k=1}^{r} n_k' B_k(\phi_k(\theta)) \right] \right\}
\]

\[
= \exp \left\{ \sum_{k=1}^{r} \left[ \theta_{(k)}^T s'_k - n_k' W_k(\theta) \right] \right\},
\]

where the last equality is based upon (12) and \( W_k(\theta) = B_k(\phi_k(\theta)) \).

Conditions for the enriched conjugate densities (15) to be proper can be obtained by extending Theorem 4 of Consonni and Veronese [13], namely for \( k = 1, \ldots, r \):

(C1). \( n_k' > 0 \),

(C2). \( s_k'/n_k' \in (X_{[k]})^\circ \),

where \((X_{[k]})^\circ \) denotes the interior of the convex-hull of the support of the marginal measure on \( X_{[k]} \).

Before concluding this subsection it is worth remarking that prior-to-posterior analysis under the enriched standard conjugate priors defined in (15) is straightforward. Multiplying (14) by (15) one immediately obtains that the posterior distribution still belongs to the class defined in (15) with hyperparameters \( s''_{[k]} = s_{[k]} + s'_{[k]} \) and \( n''_k = n + n_k' \), \( k = 1, \ldots, r \).

4. Reference analysis for conditionally reducible NEF-SQVFS

4.1. Ordered-group reference priors

In this section we shall determine the form of the reference prior for the parameter \( \phi \) when the natural exponential family \( F \) has a simple quadratic variance function.

As a preliminary step, we show that the condition for the regular case obtains. It is well known that the posterior distribution of the natural parameter of an exponential family is asymptotically normal (see, for example, [9, Section 5.3]). On the other hand, Mendoza [27] has shown that the asymptotic normality of posterior distributions is preserved under smooth transformations of the parameters. These results jointly imply that the posterior distribution of the \( \phi \) parameter indexing a conditionally reducible family is also asymptotically normal.

The next lemma establishes the validity of (1), which will prove to be instrumental in the construction of references priors.

**Lemma 1.** Let \( F \) be a conditionally \( r \)-reducible NEF on \( \mathbb{R}^d \) having a simple quadratic variance function as in (4). Then the determinant of the \( k \)th block of the Fisher
information matrix for $\phi$ can be factorized as
\[
\text{det}\{H_{kk}(\phi_{[k]})\} = a_k(\phi_{(k)})b_k(\phi_{[k-1]})
\]
for all $k = 1, \ldots, r$, with
\[
a_k(\phi_{(k)}) \propto \exp\{\phi^T_{(k)}z^k_{(k)} - \frac{1}{2}([z^k_{(k-1)}]^T A_{k[k-1]}(\phi_{(k)}) + v_k B_k(\phi_{(k)}))\},
\]
where $A_{k[k-1]}$ and $B_k$ are defined through (10), $v_k = -q(d_k^2 + 1)$ and $z^k = ((z^k_{(1)})^T, \ldots, (z^k_{(k)})^T)^T$ corresponds to the first $k$ blocks of the vector $\sum_{i=1}^{d_k} l_i^T$, where $l_i$ is the $i$th column of the matrix $L_i$ appearing in (4).

**Remark 3.** When $\mathcal{F}$ is one of the basic natural exponential families having a simple quadratic variance function written as in (5), then $\mathcal{F}$ is conditionally $d$-reducible and one can check that $z^k = z_{[k]} = (\text{diag}(L_0))[k]$ for all $k = 1, \ldots, d$.

**Proposition 2.** Let $\mathcal{F}$ be a conditionally $r$-reducible NEF on $\mathbb{R}^d$ having a simple quadratic variance function as in (4). Then the $r$-group reference prior for $\phi$ relative to the ordering $(\phi_{(1)}, \ldots, \phi_{(r)})$

(i) has density
\[
\pi_{\phi}(\phi_{(1)}, \ldots, \phi_{(r)}) \propto \prod_{k=1}^{r} a_k(\phi_{(k)})^{1/2}
\]
\[
= \prod_{k=1}^{r} \exp\left\{\frac{1}{2}(\phi_{(k)})^T z^k_{(k)} - \frac{1}{2}([z^k_{(k-1)}]^T A_{k[k-1]}(\phi_{(k)}) + v_k B_k(\phi_{(k)}))\right\},
\]
\[
(18)
\]
where $z^k$ and $v_k$ are defined in Lemma 1;

(ii) is invariant with respect to the ordering of the $r$-groups, and accordingly its density will be simply denoted by $\pi_{\phi}(\phi_{(1)}, \ldots, \phi_{(r)})$;

(iii) belongs to the enriched standard conjugate family (15).

**Corollary 1.** If $\mathcal{F}$ is one of the basic NEFs on $\mathbb{R}^d$ with simple quadratic variance function given in (5), then the $d$-group reference prior for $\phi$ has density
\[
\pi_{\phi}(\phi) \propto \exp\left\{\frac{1}{2} \phi^T z + q \sum_{k=1}^{d} B_k(\phi_k)\right\},
\]
\[
(19)
\]
where $z = \text{diag}(L_0)$.

We emphasize that formula (18), which stems from a family regarded as $r$-conditionally reducible, is especially useful in at least two important contexts:

(i) when the inferential objective suggests to group the parameters into blocks of varying importance. In this case, even if the family is a basic one, i.e. conditionally $d$-reducible, it is appropriate to view it as an $r$-conditionally
reducible family and correspondingly one must use (18) rather than (19), (see Examples 1 and 1 (ctd.)), in accordance with Remark 1.

(ii) the parameter of interest may be such that, in order to apply our method of construction, one cannot restrict attention only to basic families so that again (18) becomes the only applicable formula (see Examples 3 and 3 (ctd.)).

**Example 1** (ctd). The Multinomial family is one of the basic NEF-SQVFs and consequently is conditionally $d$-reducible, with $A_{kj}(\phi_k) = A(\phi_k) = -\ln(1 + e^{\phi_k})$, $B_k(\phi_k) = -NA(\phi_k)$ and $\phi_k = \theta_k - \ln(1 + \sum_{u=k+1}^d e^{\phi_u})$, see [13, Table 1]. Furthermore, since $z = \text{diag}(L_0) = (1, \ldots, 1)^T$, the $d$-group reference prior (19) for $\phi$ is given by

$$\pi(\phi) \varpropto \prod_{k=1}^d \exp\left\{ \frac{1}{2} \phi_k - \ln(1 + e^{\phi_k}) \right\} = \prod_{k=1}^d \frac{e^{\phi_k/2}}{1 + e^{\phi_k}}, \quad \phi \in \mathbb{R}^d. \quad (20)$$

Notice that under (20) the distribution of each $\phi_k$ is proper, and thus $\pi(\cdot)$ is also proper.

In accordance with point (i) following Corollary 1, we may regard the multinomial family as 2-conditionally reducible with $X_1 = (X_1, \ldots, X_m)^T$ distributed according to a multinomial family with natural parameter $\hat{\phi}(1)$ (say) while the conditional family of $X_2 = (X_{m+1}, \ldots, X_n)^T$ given $X_1$ is a multinomial distribution parameterized by $\hat{\phi}(2)$ (say). In order to obtain the reference priors for $(\hat{\phi}(1), \hat{\phi}(2))$, using (18), we observe that (i) both $z^1$ and $z^2$ are unit vectors, respectively, of dimension $m$ and $d$, because of Remark 3; (ii) $v_1 = (m+1)/N$ while $v_2 = (d+1)/N$ because of Lemma 1; (iii) $B_1(\hat{\phi}(1)) = N \ln(1 + \sum_{j=1}^m e^{\hat{\phi}(1)_j})$ since $B_1(\cdot)$ corresponds to the cumulant transform of an $m$-dimensional multinomial family; $A_{2j}(\hat{\phi}(2)) = A_2(\hat{\phi}(2)) = -\ln(1 + \sum_{j=1}^{d-m} e^{\hat{\phi}(2)_j})$, $j = 1, \ldots, m$, and $B_2(\hat{\phi}(2)) = -NA_2(\hat{\phi}(2))$, using the fact the conditional distribution of $X_2|X_1$ belongs to a multinomial family. Note that the functions $B_1$, $A_2$ and $B_2$ depend on the degree of conditional reducibility assumed for the family, and thus differ from the corresponding ones used to derive (20) under $d$-reducibility.

As a consequence one obtains

$$\pi(\hat{\phi}(1), \hat{\phi}(2)) \propto \prod_{j=1}^m e^{\hat{\phi}(1)_j} \prod_{l=1}^{d-m} e^{\hat{\phi}(2)_l} \left( 1 + \sum_{j=1}^m e^{\hat{\phi}(1)_j} \right)^{-m+1} \left( 1 + \sum_{l=1}^{d-m} e^{\hat{\phi}(2)_l} \right)^{-d-m+1}. \quad (21)$$

The reference prior (21) is useful to induce reference priors on alternative parameterizations grouped into two blocks, provided the first one is a bijective function of $\hat{\phi}(1)$, see Section 4.2.1; a notable illustration will be provided in Example 1 (ctd.) relative to the mean parameter grouped as $(\mu_1, \ldots, \mu_m), (\mu_{m+1}, \ldots, \mu_d)$. 

---

**ARTICLE IN PRESS**
Example 2 (Negative-multinomial family). Consider the negative-multinomial distribution on \( \mathbb{R}^d \)

\[
p_\theta(x|\theta) = \exp\{\theta^T x - M(\theta)\} \left( R - 1 + \sum_{k=1}^d x_k \right),
\]

where \( M(\theta) = -R \ln(1 - \sum_{k=1}^d e^{\theta_k}), \Theta = \{\theta \in \mathbb{R}^d : \sum_{k=1}^d e^{\theta_k} < 1\} \) and \( x_k = 0, 1, \ldots; R > 0 \) and \( \theta_k = \ln(p_k) \), with \( p_k \) the probability of the \( k \)th outcome, \( k = 1, \ldots, d \).

This family is also one of the basic NEF-SQVFs and consequently is conditionally \( d \)-reducible, with \( A_{kj}(\phi_k) = A(\phi_k) = -\ln(1 - e^{\phi_k}), \ B_k(\phi_k) = RA(\phi_k) \) and \( \phi_k = \theta_k - \ln(1 - \sum_{\mu=k+1}^d e^{\phi_\mu}) \), see [13, Table 1]. Since \( z = \text{diag}(L_0) = (1, \ldots, 1)^T \), the \( d \)-group reference prior (19) for \( \phi \) is given by

\[
\pi_\phi(\phi) \propto \prod_{k=1}^d \exp\left\{ \frac{1}{2} \phi_k - \ln(1 - e^{\phi_k}) \right\} = \prod_{k=1}^d \frac{e^{\phi_k/2}}{1 - e^{\phi_k}}, \quad \phi \in (- \infty, 0)^d. \tag{22}
\]

Notice that \( \pi_\phi(\cdot) \) in (22) is improper, because each \( \phi_k \) has an improper distribution.

Example 3 (A nonbasic NEF-SQVF). Consider a NEF-SQVF on \( \mathbb{R}^3 \) having variance function

\[
V(\mu) = \begin{bmatrix} \mu_1 & \mu_2 & 0 \\ \mu_2 & \mu_2 & 0 \\ 0 & 0 & \mu_3 \end{bmatrix}, \quad \mu_1 > \mu_2 > 0, \quad \mu_3 > 0.
\]

This family is 3-conditionally reducible because of the last sentence of Remark 2. However it is not a basic family; moreover, it can be checked that the marginal distribution of \( X_1 \) is Poisson(\( \mu_1 \)), the conditional distribution of \( X_2 | X_1 = x_1 \) is Binomial(\( x_1, \mu_2/\mu_1 \)), while \( X_3 \) is independent of \( (X_1, X_2) \) with distribution Poisson(\( \mu_3 \)).

Straightforward calculations lead to

\[
\phi_1 = \ln \mu_1, \quad B_1(\phi_1) = e^{\phi_1},
\]

\[
\phi_2 = \ln \frac{\mu_2}{\mu_1 - \mu_2}, \quad A_{21}(\phi_2; y_1) = \ln(1 + e^{\phi_2}), \quad B_2(\phi_2) = 0,
\]

\[
\phi_3 = \ln \mu_3, \quad A_{31}(\phi_3; y_1) = A_{32}(\phi_3; y_2) = 0, \quad B_3(\phi_3) = e^{\phi_3}.
\]

From the variance function we deduce that \( q = 0 \), so that \( v_k = 0, k = 1, 2, 3 \). Moreover \( z^1 = 1, z^2 = (2, 1)^T \), and \( z^3 = (2, 1, 1)^T \). Using (18) we derive the 3-group
reference prior
\[ \pi_{\phi}(\phi_1, \phi_2, \phi_3) \propto e^{\frac{1}{2} \psi_1} \exp \left\{ \frac{\phi_2}{2} - \frac{1}{2} \left[ 2 \ln(1 + e^{\phi_2}) \right] \right\} e^{\frac{1}{2} \psi_3} \]
\[ \propto e^{\frac{1}{2} (\phi_1 + \phi_2 + \phi_3)} (1 + e^{\phi_2})^{-1}, \quad (\phi_1, \phi_2, \phi_3) \in \mathbb{R}^3. \]  

(23)

4.2. Properties of the ordered-group reference prior

4.2.1. Reference priors on alternative parameterizations

It is known that, in general, reference priors are not invariant under arbitrary reparameterizations of the model. Specifically suppose that \( \phi \) and \( \lambda \) are two alternative parameterizations of \( F \). Given a reference prior on \( \phi \), a reference prior for \( \lambda \) cannot always be induced from that of \( \phi \). This is not surprising since the reference prior method takes explicitly into account the inferential importance of the parameters. Specifically, let \( \phi = (\phi_1^T, \ldots, \phi_r^T)^T \) and \( \lambda = (\lambda_1^T, \ldots, \lambda_r^T)^T \) where the blocks are arranged in increasing order of importance and \( \phi_k \) has the same dimension of \( \lambda_k \), \( k = 1, \ldots, r \). If, for each \( k \), \( \lambda_k \) is a function of \( (\phi_1, \ldots, \phi_k) \), then Yang [31] and Datta and Ghosh [19] show that the reference prior on \( \lambda \) can be obtained from that of \( \phi \), via the usual change-of-variable technique. Notice that the transformation from \( \phi \) to \( \lambda \) is block-lower triangular: we can therefore conclude that reference priors are invariant with respect to such transformations, whose feature is to preserve the order of inferential importance of the groups under the two parameterizations. An important instance occurs when \( \lambda \) represents the mean parameter of the NEF \( F \). When \( \lambda \) equals the natural parameter \( \theta \), block-lower triangularity does not hold; nevertheless we show that the reference prior on \( \theta \) can still be recovered from that of \( \phi \).

**Proposition 3.** Let \( F \) be a conditionally r-reducible NEF-SQVF on \( \mathbb{R}^d \). If \( z_{(j)}^k = z_{(j)}, \ j = 1, \ldots, k; \ k = 1, \ldots, r \), then

(i) the r-group reference prior of the natural parameter \( \theta \) relative to the ordering \( \theta_{(1)}, \ldots, \theta_{(r)} \) has density

\[ \theta_{(1)}, \ldots, \theta_{(r)}, \pi_{\theta}(\theta) \propto \exp \left\{ \frac{1}{2} \left[ \theta^T z - \sum_{k=1}^{r} v_k W_k(\theta) \right] \right\}; \]  

(24)

(ii) the r-group reference prior of the mean parameter \( \mu \) relative to the ordering \( \mu_{(1)}, \ldots, \mu_{(r)} \) has density

\[ \mu_{(1)}, \ldots, \mu_{(r)}, \pi_{\mu}(\mu) \propto \exp \left\{ \left[ -\frac{1}{2} \theta(\mu)^T z - \sum_{k=1}^{r} \left( \frac{1}{2} v_k - v \right) W_k(\theta(\mu)) \right] \right\}. \]  

(25)
We remark that the reference priors for $\theta$ and $\mu$ are not order-invariant, unlike that for $\phi$. Furthermore, all basic NEF-SQVFs satisfy the conditions on $z^{(j)}_{(j)}$ of Proposition 3, because of Remark 3.

**Example 1** (ctd.). To obtain the $d$-group reference prior for $\mu$ it suffices to recall that $\phi_k = \theta_k - \ln(1 + \sum_{u=k+1}^d e^{\theta_u})$ and $\theta_k(\mu) = \ln(\mu_k) - \ln(N - \sum_{j=1}^d \mu_j)$. As a consequence, $W_k(\theta) = B_k(\phi_k(\theta)) = N \ln(1 + \frac{e^{\theta_k}}{1 + \sum_{u=k+1}^d e^{\theta_u}})$, so that $W_k(\theta(\mu)) = N \ln(N - \sum_{j=1}^k \mu_j)$. Since $v_k = (k + 1)/N$ and $v = (d + 1)/N$, (25) yields

$$
\pi_1, \ldots, \pi_d \pi_\mu(\mu) \propto \prod_{k=1}^d \left( \frac{N - \sum_{j=1}^d \mu_j}{\mu_k} \right)^{1/2} \left( \frac{N - \sum_{j=1}^{k-1} \mu_j}{N - \sum_{j=1}^k \mu_j} \right) \left( \frac{N - k \mu_j}{N - \sum_{j=1}^k \mu_j} \right)^{d - \frac{k + 1}{2}} \propto \prod_{k=1}^d \left( \mu_k \left( N - \sum_{j=1}^k \mu_j \right) \right)^{-1/2}.
$$

Consider now the cell-probabilities parameter, $p = (p_1, \ldots, p_d)^T$. Since $p_k = \mu_k/N$, the $d$-group reference prior on $p$ is easily derived from (26), i.e.

$$
\pi_1, \ldots, \pi_d \pi_p(p) \propto \prod_{k=1}^d \left( \frac{1 - \sum_{j=1}^k p_j}{p_k} \right)^{-1/2}.
$$

Berger and Bernardo [6] obtained ordered group reference priors for the parameter $p$ and discussed their properties. Of course, (27) coincides with their $d$-group reference prior relative to the ordering $(p_1, \ldots, p_d)$. However, our derivation is straightforward because of the structure of $d$-conditional reducibility afforded by the Multinomial family. Notice that the reference prior (27) on $p$ is a generalized Dirichlet distribution; see [12].

Suppose now that one is interested in obtaining the reference prior on $p$ relative to the ordered-grouping $p_{(1)} = (p_1, \ldots, p_m)^T$, $p_{(2)} = (p_{m+1}, \ldots, p_d)^T$. This might occur, for instance, in an electoral contest, where $p_1, \ldots, p_m$ are the probabilities of voting in favor of the $m$ most prominent parties, while the remaining probabilities refer to smaller political groups. The ordered-grouping above implies that the focus of inference is on $(p_1, \ldots, p_m)$, without establishing any privileged hierarchy among these parties.

As already remarked, the above prior cannot be obtained from (27); on the other hand one can derive it from (21) since the mapping $(\phi((1), \phi,(2)) \rightarrow (p_{(1)}, p_{(2)})$ is...
block-lower triangular. Alternatively, since the conditions of Proposition 3 are satisfied and \( p_{(k)} = (1/N) \mu_{(k)} \), one can use directly formula (25) with \( r = 2 \).

Using the expressions for \( v_k \) and \( W_k(\theta) = B_k(\phi_k(\theta)) \) recalled in the latter part of Example 1 (ctd.), one readily deduces

\[
p_{(1):p_{(2)} \pi_p(p) \propto \left( \prod_{k=1}^{d} p_{k}^{-1/2} \right) \left( 1 - \sum_{j=1}^{m} p_{j} \right)^{-1/2} \left( 1 - \sum_{j=1}^{d} p_{j} \right)^{-1/2}. \]

**Example 2** (ctd.). To obtain the \( d \)-group reference prior on \( \mu \) it suffices to recall that \( \phi_k(\theta) = \theta_k - \ln(1 - \sum_{u=k+1}^{d} e^{\theta_u}) \) and \( \theta_k(\mu_k) = \ln(\mu_k) - \ln(R + \sum_{j=1}^{d} \mu_j) \). Consequently, \( W_k(\theta) = B_k(\phi_k(\theta)) = -R \ln\left( 1 - \frac{\theta_k}{1 - \sum_{u=k+1}^{d} e^{\theta_u}} \right) \), so that \( W_k(\theta(\mu)) = -R \ln\left( \frac{\mu_k}{R + \sum_{j=1}^{d} \mu_j} \right) = -R \ln\left( \frac{R + \sum_{j=1}^{d} \mu_j}{R + \sum_{j=1}^{k} \mu_j} \right) \).

Since \( v_k = -(k + 1)/R \) and \( v = -(d + 1)/R \), (25) yields

\[
\pi_{\mu}(\mu) \propto \prod_{k=1}^{d} \left( \frac{R + \sum_{j=1}^{d} \mu_j}{\mu_k} \right)^{1/2} \left( \frac{R + \sum_{j=1}^{d} \mu_j}{R + \sum_{j=1}^{k} \mu_j} \right)^{d-k+1/2} \left( \mu_k \left( R + \sum_{j=1}^{k} \mu_j \right) \right)^{-1/2}.
\]

(28)

Consider now the reference prior on \( p \). This distribution cannot be derived from the reference prior on \( \mu \), as in the multinomial case, because the mapping from \( \mu \) to \( p \) is not lower triangular. On the other hand, one can check that \( p_k = e^{\theta_k} \), so that the reference prior on \( p \) can be induced from that of \( \theta \). Using (24) we first obtain the reference prior on \( \theta \)

\[
\pi_{\theta}(\theta) \propto \prod_{k=1}^{d} \left( e^{\theta_k} \left( 1 - \sum_{u=k}^{d} e^{\theta_u} \right) \right)^{\gamma_k},
\]

where \( \gamma_1 = -1 \) and \( \gamma_k = -1/2, k = 2, \ldots, d \).

Since \( |\lambda(\theta)| = \prod_{k=1}^{d} p_{k}^{-1} \), the \( d \)-group reference prior on \( p \) becomes

\[
p_{(1):p_{(2)} \pi_p(p) \propto \prod_{k=1}^{d} \left( p_{k}^{-1} \left( 1 - \sum_{u=k}^{d} p_{u} \right) \right)^{\gamma_k}.
\]

**Example 3** (ctd.). Let \( \tilde{X}_i, i = 1, 2, 3 \) be independent Poisson variables with expectations \( \tilde{\mu}_i \) so that the corresponding family is a basic NEF-SQVF. Suppose
we want to make inference on $\mu_1 = \tilde{\mu}_1 + \tilde{\mu}_2$, $\mu_2 = \tilde{\mu}_2$, $\mu_3 = \tilde{\mu}_3$ in this order of importance.

Using (25) one can obtain the reference prior on $\tilde{\mu} = (\tilde{\mu}_1, \tilde{\mu}_2, \tilde{\mu}_3)^T$; however this cannot be used to obtain the reference prior on $\mu = (\mu_1, \mu_2, \mu_3)^T$ since the transformation $\tilde{\mu} \rightarrow \mu$ is not lower triangular.

A natural way to circumvent this problem is to transform $\tilde{X}$ to $X = (X_1, X_2, X_3)^T$, with $X_1 = \tilde{X}_1 + \tilde{X}_2$, $X_2 = \tilde{X}_2$, $X_3 = \tilde{X}_3$, whose distribution belongs to the 3-conditionally reducible, nonbasic NEF-SQVF, described in Example 3. One can now use the reference prior $\pi_\phi(\phi)$, obtained in (23), to induce the reference prior on $\mu$. Since $|J_\phi(\mu)| = \{(\mu_1 - \mu_2)\mu_2\mu_3\}^{-1}$ one readily obtains

$$\mu_1, \mu_2, \mu_3 \pi_\mu(\mu) \propto \{(\mu_1 - \mu_2)\mu_2\mu_3\}^{-1/2}, \quad \mu_1 > \mu_2 > 0, \quad \mu_3 > 0.$$ 

It is interesting to remark that the reference prior on $\mu$ cannot be obtained using formula (25) since the conditions on the $z_{(j)}$s in Proposition 3 are not satisfied because $z_1^1 = 1 \neq z_1^2 = 2$, see Example 3.

4.2.2. Reference posteriors

Since the reference prior for $\phi$ belongs to the enriched standard conjugate family, the $\phi_{(k)}$’s will also be independent under the reference posterior. More precisely, if $s$ denotes the realization of the sufficient statistic for a random sample of size $n$, the reference posterior will be obtained from (18) upon replacing $\{z^k/2, v^k/2\}$ with $\{z^k/2 + s_{[k]}, v^k/2 + n\}$, so that

$$\pi_\phi(\phi|s) \propto \prod_{k=1}^r \exp \left\{ \phi_{(k)}^T \left( \frac{z^k_{(k)}}{2} + s_{(k)} \right) - \sum_{j=1}^{k-1} A^j_{kj}(\phi_{(k)}) \left( \frac{z^k_{(j)}}{2} + s_{(j)} \right) + \left( \frac{v^k}{2} + n \right) B_k(\phi_{(k)}) \right\}.$$ 

(29)

In particular, if the family is one of the basic NEF-SQVFs on $\mathbb{R}^d$, with variance function given in (5), formula (29) holds with $r = d$, $v_k = -q(k + 1)$ and $z^k = z_{[k]} = \text{diag}(L_0)_{[kl]}$. Furthermore, for all such families, one has

$$\pi_\phi(\phi|s) \propto \exp \left\{ \phi^T \left( \frac{z}{2} + s \right) - \sum_{k=1}^d \left[ n + q \left( \sum_{j=1}^{k-1} s_j - 1 \right) \right] B_k(\phi_k) \right\},$$ 

(30)
except for the negative-multinomial/gamma/normal family for which one has instead

$$\pi_\phi(\phi|s) \propto \exp\left\{ \phi^T \left( \frac{Z}{2} + s \right) \right\} \times \exp\left\{ -\sum_{k=1}^{m+1} \left[ n + q \left( \sum_{j=1}^{k-1} s_j - 1 \right) \right] B_k(\phi_k) - s_{m+1} \sum_{k=m+2}^{d} B_k(\phi_k) \right\}. \quad (31)$$

Sufficient conditions for the reference posterior (29) on $\phi$ to be proper can be immediately derived from conditions (C1) and (C2) of Section 3.3, namely, for $k = 1, \ldots, r$

$$(C1^*) \quad v_k/2 + n > 0, \quad (C2^*) \quad (z_k^2/2 + s_{\{k\}})/(v_k/2 + n) \in (X^{X_{\{k\}}})^c.$$ 

The previous conditions do not allow one to conclude that for all basic NEF-SQVF$s the posterior reference distribution on $\phi$ is always proper. On the other hand, the result is actually true as stated below.

**Proposition 4.** Let $\mathcal{F}$ be a basic NEF-SQVF on $\mathbb{R}^d$. The $d$-group reference posterior on $\phi$ is proper for all sample sizes and all sample values.

Clearly, Proposition 4 is trivially satisfied for the multinomial family since the reference prior (20) is proper.

An interesting aspect of default or noninformative priors concern their frequentist properties, especially with respect to coverage of posterior probability intervals. Specifically, let $\Pi$ be a prior for $\lambda = (\lambda_1, \ldots, \lambda_d)^T$. Let $t_1(\lambda), \ldots, t_s(\lambda), s \leq d$, be real-valued twice continuously differentiable parametric functions of interest. If, given a sample of size $n$, $\Pi$ produces a posterior one-sided credibility interval of probability $\alpha$ for each $t_j(\lambda)$ whose confidence level, over repeated sampling, is $\alpha + O(n^{-1})$, then $\Pi$ is said to be a *simultaneous marginal* probability matching prior. For a concise discussion and extensive references see [18]; Datta [17] also deals with the notion of *joint* probability matching priors.

Under the assumption that the components of $\lambda$ are orthogonal, and if $t_j(\lambda) = \lambda_j$, one can characterize the prior on $\lambda$ that is simultaneous marginal probability matching (this prior is actually also joint probability matching). It is immediate to verify that the reference prior on $\phi$ given in Proposition 2 is simultaneous marginal probability matching for $\phi_1, \ldots, \phi_d$. This result is of particular interest when the component $\phi_k$ has a substantial interpretation. For instance in the multinomial case (see Example 1), $\phi_k$ represents the logit of the probability of an outcome in cell $k$ given the number of outcomes in the previous cells, namely $\phi_k = \logit(p_k/(1 - \sum_{j=1}^{k-1} p_j))$. On the other hand, in the negative-multinomial case (see Example 2), with $R$ a positive integer, the conditional distribution of $X_k|X_1 = x_1, \ldots, X_{k-1} = x_{k-1}$ is negative-binomial with number of occurrences outside cell $k$ equal to $R + \sum_{j=1}^{k-1} x_j$.
and probability of an occurrence in cell $k$ given by $p_k/(1 - \sum_{j=k+1}^d p_j)$. It can be checked that $\phi_k$ corresponds to the logarithm of this latter probability.

The above examples refer to the more general problem of invariance of matching probability priors to reparameterization. For an extensive treatment of this and related issues see [19].

As a final remark, suppose that $\mu_1$ represents the parameter of interest. Notice that $\mu_1$ is a function of $\phi_1$ alone, see (11). Then the joint density of $(\mu_1, \lambda)$ can be deduced from (18) to be equal to

$$
\mu_1, \lambda \pi_{\mu_1, \lambda}(\mu_1, \lambda) \propto a_1(\phi_1(\mu_1))^{1/2} \left| \frac{\partial \phi_1}{\partial \mu_1} \right| \prod_{k=2}^d a_k(\phi_k)^{1/2}
$$

$$
\propto \{H^\mu_{11}(\mu_1)\}^{1/2} \prod_{k=2}^d a_k(\phi_k)^{1/2},
$$

where $H^\mu_{11}(\mu_1)$ represents the first element of the block-diagonal information matrix relative to $(\mu_1, \lambda)$. Recalling that $\lambda$ is orthogonal to $\mu_1$, one can thus conclude, using Eq. (10) of Datta and Ghosh [18], see also [30], that the prior for $(\mu_1, \lambda)$ is also probability matching for $\mu_1$.

5. Reference priors for two-way contingency tables

A useful application of the theory described above concerns the analysis of a two-way contingency table under multinomial sampling.

Let $X_{ij}$ denote the counts in row $i$ and column $j$ of an $r \times c$ table, with $X_{rc} = N - \sum_{i=1}^{r-1} \sum_{j=1}^c X_{ij} - \sum_{j=1}^{c-1} X_{ij}$.

The joint distribution of $X = (X_{ij})$ is multinomial on $\mathbb{R}^{c-1}$ and, apart from the double-index notation, can be written as in (8) of Example 1.

As a consequence, the reference prior on $\phi = (\phi_{ij}; i = 1, \ldots, r - 1, j = 1, \ldots, c; \phi_{ij}; j = 1, \ldots, c - 1)$ is as in (20), namely

$$
\pi_\phi(\phi) \propto \left( \prod_{i=1}^{r-1} \prod_{j=1}^c \frac{e^{\phi_{ij}/2}}{1 + e^{\phi_{ij}}} \right) \left( \prod_{j=1}^{c-1} \frac{e^{\phi_{ij}/2}}{1 + e^{\phi_{ij}}} \right),
$$

whereas the reference prior on the cell probabilities $p_{ij}$, relative to the ordering $(p_{11}, p_{12}, \ldots, p_{rc-1})$, is given by, see (27)

$$
p_{ij} \pi_p(p) \propto \prod_{i=1}^{r-1} \prod_{j=1}^c \left[ p_{ij} \left( 1 - \sum_{k=1}^{i-1} \sum_{j=1}^c p_{kj} - \sum_{l=1}^{j-1} p_{il} \right) \right]^{-1/2}
$$

$$
\times \prod_{j=1}^{c-1} \left[ p_{ij} \left( 1 - \sum_{i=1}^{r-1} \sum_{j=1}^c p_{ij} - \sum_{l=1}^{j-1} p_{il} \right) \right]^{-1/2},
$$

Some properties of $p \pi_p$ will be described later on in this section.
One can factorize $p_{ij}$ as $p_i p_{ij|i}$, where $p_i$ represents the $i$th row-marginal probability and $p_{ij|i}$ the conditional probability of column $j$ given row $i$. This gives rise to the reparameterization $\tilde{p} = (p_i, p_{1|1}, \ldots, p_{c-1|1}, i = 1, \ldots, r - 1; p_{i|r}, \ldots, p_{c-1|r})$ which is more natural, for example, in the context of Bayesian networks, where rows index configurations of “parent” nodes, whereas columns index states of the “children” nodes; see [15,20].

Berger and Bernardo [6] recommended, for typical use, to employ a reference prior having a maximal number of groupings. Accordingly, we construct the $(rc - 1)$-group reference prior on $\tilde{p}$. On the other hand, the latter cannot be deduced from that of $p$, because the mapping $p \mapsto \tilde{p}$ is not lower triangular.

To solve the problem it is expedient to consider the transformed vector of observables $\tilde{X} = (X_i, X_{i1}, \ldots, X_{ic}, i = 1, \ldots, r - 1; X_r, \ldots, X_{rc-1})$. Of course, the distribution of $\tilde{X}$ is still NEF-SQVF because $\tilde{X} = \mathbf{G} \mathbf{X}$, with $\mathbf{G}$ a suitable $(rc - 1) \times (rc - 1)$ matrix.

**Proposition 5.** The distribution of $\tilde{X}$ is a conditionally $(rc - 1)$-reducible NEF-SQVF. If $\psi = (\psi_i, \psi_{i1}, \ldots, \psi_{ic}, i = 1, \ldots, r - 1; \psi_r, \ldots, \psi_{rc-1})$ denotes the corresponding conditionally reducible parameter, then the reference prior on $\psi$ is given by

$$
\pi_\psi(\psi) \propto \prod_{i=1}^{r-1} \left( \frac{e^{\psi_i/2}}{1 + e^{\psi_i}} \right) \left( \prod_{i=1}^{r-1} \prod_{j=1}^{c-1} \frac{e^{\psi_{ij}/2}}{1 + e^{\psi_{ij}}} \right). \tag{34}
$$

The $(rc - 1)$-reference prior on $\tilde{p}$ is given by

$$
\tilde{p}_{\text{ref}}(\tilde{p}) \propto \prod_{i=1}^{r-1} p_i \left( 1 - \sum_{k=1}^{i} p_k \right)^{-1/2} \prod_{i=1}^{r} \prod_{j=1}^{c-1} \left[ p_{ij} \left( 1 - \sum_{l=1}^{j} p_{lj} \right) \right]^{-1/2}. \tag{35}
$$

It is interesting to notice that (34) is structurally equivalent to (32), although $\phi$ and $\psi$ are distinct functions of the primary parameters $p_{ij}$.

Letting $p_i = (p_i, i = 1, \ldots, r - 1)$, $p_{ji|i} = (p_{1|1}, \ldots, p_{c-1|1})$ and $p_{ji|i} = (p_{j|i}, i = 1, \ldots, r)$, we deduce that the marginal distribution of $p_i$ and of each of the $p_{ji|i}$ under (35) is a (proper) generalized Dirichlet.

Consider now the reference prior on $\tilde{p}$ according to the order $p_i, p_{ji|i}$. This coincides with (35) because of the order invariance of the reference prior on $\psi$—a consequence of Proposition 2(ii)—and the fact that there exists a suitable permutation of $\psi$ such that its mapping onto $(p_i, p_{ji|i})$ is lower triangular.

Of course, one might have factorized $p_{ij}$ as $p_j p_{ij}$ and repeated the procedure leading to a reference prior on $p^\star = (p_j, p_{1|j}, \ldots, p_{c-1|j}, j = 1, \ldots, c - 1; p_{1|c}, \ldots, p_{r-1|c})$ which, with obvious modifications, will be of the same type as (35). We remark that the reference priors on $\tilde{p}$ and $p^\star$ are incompatible, i.e. they give rise to two distinct distributions on the joint cell probabilities $(p_{ij})$. This follows from Theorem 2 of Geiger and Heckerman [20] who show that only the Dirichlet distribution on $(p_{ij})$ induces a distribution on $\tilde{p}$ and $p^\star$ enjoying both global and local
and independence of $p_I$ and $p_{J|i}$ as well as of $p_I$ and $p_{J|J}$. In other words, choosing rows, rather than columns, as a criterion to assess parameter importance, does matter.

As far as the inferential aspects are concerned, it follows from Proposition 2(iii) that, under the reference posterior, $p_I$, and $p_{J|i}$, $i = 1, \ldots, r$ are still independent, each vector having a generalized Dirichlet distribution. Specifically, writing $\text{GDir}(x_1, \ldots, x_d, \gamma_1, \ldots, \gamma_d)$ for the joint density

$$f(t_1, \ldots, t_d|x_1, \ldots, x_d, \gamma_1, \ldots, \gamma_d) \propto \prod_{i=1}^d t_i^{x_i-1} \left(1 - \sum_{j=1}^i t_j\right)^{\gamma_i},$$

with $t_i > 0$, and $\sum_{i=1}^d t_i < 1$ we get $p_I|x \sim \text{GDir}(x_1 + \frac{1}{2}, \ldots, x_r - 1 + \frac{1}{2}, \gamma_r = -\frac{1}{2}, i = 1, \ldots, r - 2; \gamma_{r-1} = N - \sum_{i=1}^{r-1} x_i - \frac{1}{2})$, and independently $p_{J|i}|x \sim \text{GDir}(x_{i1} + \frac{1}{2}, \ldots, x_{i,c-1} + \frac{1}{2}, \gamma_{i,c-1} = -\frac{1}{2}, j = 1, \ldots, c - 2; \gamma_{i,c-1|i} = x_i - \sum_{j=1}^{c-1} x_{ij} - \frac{1}{2})$.

Formula (3.1) of Lochner [26] can be used to compute marginal moments of a generalized Dirichlet. In particular, we get

$$E_\hat{p}(p_i|x) = \left(\prod_{l=1}^{i-1} \frac{N - x_1 - \ldots - x_l + 1/2}{N - x_1 - \ldots - x_l + 1}\right) \left(\frac{x_i + 1/2}{N + 1}\right)$$

and

$$E_\hat{p}(p_{j|i}|x) = \left(\prod_{m=1}^{i-1} \frac{x_i - x_{i1} - \ldots - x_{im} + 1/2}{x_i - x_{i1} - \ldots - x_{im} + 1}\right) \left(\frac{x_{ij} + 1/2}{x_i + 1}\right).$$

We close this section with a couple of remarks on the comparison between the reference prior on $p$, described in (33)—which only employs the multinomial structure of the observations and is therefore essentially equivalent to that derived and discussed in [6]—and the reference prior on $\hat{p}$ given in (35).

First of all, both priors share the property of marginalization, see Berger and Bernardo [6, Section 3.3]. Relative to $\hat{p}$ this means for example that collapsing rows $r$ and $r - 1$ in the contingency table would lead to a reference prior on $p_I$ and $p_{J|i}$, $i = 1, \ldots, r - 2$, which coincides with that obtained through marginalization from (35).

On the other hand, a significant distinction between the reference prior on $p$ and that on $\hat{p}$ concerns expectations of cell probabilities $p_{ij}$. More precisely, under the former prior we have

$$E_p(p_{ij}) = \left(\frac{1}{2}\right)^{c(i-1)+j},$$

whereas, under the latter,

$$E_\hat{p}(p_{ij}) = E_\hat{p}(p_I)E_\hat{p}(p_{j|i}) = \left(\frac{1}{2}\right)^{i+j}.$$

Thus $E_p(p_{11}) = 1/2$ and then expectations progressively decrease following the ordering of $(i,j)$, with $j$ running faster than $i$. On the other hand, $E_\hat{p}(p_{11}) = 1/4$ and
then expectations progressively decrease along the first row; for the remaining rows a similar pattern occurs starting from $E_p(p_{11}) = (1/2)^{i+1}$.

In conclusion, the prior for $p$ is strongly unbalanced in favor of the cells in the first row, whereas that for $p$ distributes the mass more evenly. The difference between the two priors may lead to noticeably different inferences in the presence of sparse contingency tables.

6. Concluding remarks

In this paper we have obtained the reference prior on the parameter $\phi$ of a conditionally $r$-reducible NEF-SQVF. Its derivation is particularly straightforward thanks to the attractive properties enjoyed by the $\phi$ parameterization. We have shown that the reference prior is order invariant and belongs to the enriched standard conjugate family. Moreover, we have derived the reference priors for the natural and mean parameter.

Concerning posterior inference, we show that, for basic NEF-SQVFs, the $d$-group reference posterior distributions are proper for all data sets and sample sizes. Furthermore, the posterior expectation of $\mu$ can be obtained by applying results in Theorem 5 of Consonni and Veronese [13]; while posterior moments of $\phi$ and $\theta$ can be derived, whenever the normalizing constant is explicitly available, using Proposition 1 of Gutierrez-Pena [21].

The theory presented in this paper can be applied also to conditionally $r$-reducible NEFs whose variance function is not necessarily SQVF. An important instance is represented by NEFs having a homogeneous quadratic variance function such as the Wishart family. When sampling from an $l$-dimensional multivariate zero-mean normal distribution, the previous family arises both as the sampling distribution of the sum of cross-products matrix, or, from a Bayesian perspective, as the standard conjugate prior on the population precision- (i.e. inverse covariance-) matrix. Because of symmetry constraints on the matrix, the actual size of the family is $d = l(l + 1)/2$. Consonni and Veronese [14] have shown that the Wishart family on symmetric cones is fully conditionally $l$-reducible and that a reference prior for the $\phi$ parameters can be easily constructed. Such parameters admit a simple interpretation in terms of regression coefficients and variances associated to the $l$ conditional distributions, whose product gives the joint density of the Gaussian population.

Acknowledgments

The authors thank José M. Bernardo for the suggestion to apply the results of Section 4 to the analysis of contingency tables. The paper has greatly benefited from the insightful comments of the referees.
Appendix A. Proofs

Proof of Proposition 1. We first recall that if $X$ is distributed according to a NEF-SQVF on $\mathbb{R}^d$ then $X = G\tilde{X} + a$ where $\tilde{X}$ follows a basic NEF-SQVF distribution on $\mathbb{R}^d$ with variance function given in (5), $G$ is a nonsingular square matrix and $a$ is a constant vector.

From Propositions 2.3 and 3.2 of Gutiérrez-Peña and Smith [23], it follows that

\[
\prod_{V} \{ \theta(\mu)^T z - vM(\theta(\mu)) \},
\]

where $z = G\text{diag}(L_0) - q(d + 1)a$ and $v = -q(d + 1)$.

We now show that $z = \sum_{i=1}^{d} l_i$ where $l_i$ denote the $i$th column of the matrix $L_i$ appearing in (4).

The variance function of $\tilde{X}$, $\tilde{V}(\tilde{\mu})$ say, can be written as

\[
\tilde{V}(\tilde{\mu}) = q\tilde{\mu}\tilde{\mu}^T + \sum_{i=1}^{d} \tilde{\mu}_i\tilde{L}_i + \tilde{C}
\]

with $\tilde{L}_i = \text{Diag}(l_i)$ where $l_i$ corresponds to the $i$th column of the matrix $L_0$ in (5). The specific form of $L_0$ for each of the basic families can be found in [23, p. 18].

We initially consider the case of a linear transformation $\tilde{X} = G\tilde{X}$ and set $\tilde{\mu} = G\tilde{X}$. Then

\[
\tilde{V}(\tilde{\mu}) = G\tilde{V}(G^{-1}\tilde{\mu})G^T = q\tilde{\mu}\tilde{\mu}^T + \sum_{i=1}^{d} [G^{-1}\tilde{\mu}_i]G\tilde{L}_iG^T + GC\tilde{G}^T.
\]

(38)

Denoting by $g^{ij}$ the general element of the matrix $G^{-1}$, the linear term in (38) becomes

\[
\sum_{i=1}^{d} \sum_{j=1}^{d} g^{ij}\tilde{\mu}_jG\tilde{L}_iG^T = \sum_{j=1}^{d} \tilde{\mu}_j \left( \sum_{i=1}^{d} G\tilde{L}_i g^{ij}G^T \right).
\]

(39)

Because $\tilde{V}(\tilde{\mu})$ must be of the form (4), we have $\tilde{L}_j = G(\sum_{i=1}^{d} \tilde{L}_i g^{ij})G^T = G U_j G^T$, where $U_j = \text{Diag}(L_0 g^j)$ with $g^j$ the $j$th column of $G^{-1}$.

In particular, if $\tilde{L}_j$ denotes the $j$th column of $L_j$, we have

\[
\tilde{L}_j = \sum_{r=1}^{d} u_{jr} g_r g_{jr}
\]

(40)

where $u_{jr}$ is the $r$th element of the diagonal of $U_j$ and $g_r$ is the $r$th column of $G$. Now it is simple to check that, because of the specific form of $L_0$, we have $\sum_{j=1}^{d} \tilde{L}_j = \sum_{r=1}^{d} u_{jr} g_r g_{jr} = G\text{diag}(L_0)$.
Finally consider $X = \tilde{X} + a$. With obvious notation, we have

$$V(\mu) = \tilde{V}(\mu - a) = q\mu \mu^T - q(\mu a^T + a \mu^T) + \sum_{i=1}^{d} \tilde{L}_i \mu_i + C$$

$$= q\mu \mu^T + \sum_{i=1}^{d} L_i \mu_i + C,$$

with $C = \tilde{C} - \sum_{i=1}^{d} \tilde{L}_i a_i + qa^T$ and $L_i = \tilde{L}_i - qA_i$ where $A_i$ is the matrix with elements $a_{ik}' = a_{ki}' = a_k$, $k \neq i$, $a_{ii}' = 2a_i$ and the remaining $a_{jk}'$’s are zero.

As a consequence,

$$\sum_{j=1}^{d} I_j' = \sum_{j=1}^{d} \tilde{I}_j' - q(d + 1)a = G \text{diag}(L_0) - q(d + 1)a = z.$$

**Proof of Lemma 1.** Consider the partition $X = (X_{[r-1]}^T, X_{(r)}^T)^T$, with corresponding partitions for $\theta$, $\mu$ and $\phi$. Recall from Eq. (13) that $\theta_{(r)} = \phi_{(r)}$. Now, by (9), the Fisher information matrix relative to $\phi$ is block-diagonal, i.e.

$$H(\phi) = \text{Diag}\{H_{11}(\phi_{(1)}), \ldots, H_{kk}(\phi_{[k]}), \ldots, H_{rr}(\phi_{(r)})\}.$$

From Remark 2 the distribution of $X_{[r-1]}$ is a natural exponential family with natural parameter $\phi_{(r)}^{-1}$ and cumulant transform $M_{(r)}^{-1}$ say; alternatively we can also parameterize this family in terms of the corresponding mean parameter $\mu_{[r-1]}$ or in terms of $\phi_{(r)}$. Now consider the mixed parameterization $(\mu_{[r-1]}, \phi_{(r)})$ for the family of $X$. It is easy to see that the block of the Fisher information matrix (relative to this parameterization) corresponding to $\phi_{(r)}$ is given by

$$H_{rr}^{\mu_{[r-1]}, \phi_{(r)}}(\mu_{[r-1]}, \phi_{(r)}) = H_{rr}(\phi_{[r-1]}(\mu_{[r-1]}), \phi_{(r)}).$$

Since $X_{[r-1]}$ is a cut of $\mathcal{F}$, see [13], it follows from Lemma 3.3 of Barndorff-Nielsen and Bøesild [3] that

$$H_{rr}^{\mu_{[r-1]}, \phi_{(r)}}(\mu_{[r-1]}, \phi_{(r)}) = V_{rr}(\mu_{[r-1]}, \mu_{(r)}(\mu_{[r-1]}, \phi_{(r)}))$$

$$- \left\{ V_{rr}[r-1](\mu_{[r-1]}, \mu_{(r)}(\mu_{[r-1]}, \phi_{(r)})) V_{[r-1][r-1]}(\mu_{[r-1]})^{-1} \times V_{[r-1][r-1]}(\mu_{[r-1]}, \mu_{(r)}(\mu_{[r-1]}, \phi_{(r)})) \right\},$$

whence, using standard results on the determinant of partitioned matrices,

$$\det\{H_{rr}^{\mu_{[r-1]}, \phi_{(r)}}(\mu_{[r-1]}, \phi_{(r)})\}$$

$$= \det\{V(\mu_{[r-1]}, \mu_{(r)}(\mu_{[r-1]}, \phi_{(r)}))\} \det\{V_{[r-1][r-1]}(\mu_{[r-1]})\}^{-1}. $$
Furthermore, again from the fact that $X_{[r-1]}$ is a cut of $\mathcal{F}$, it follows from Barndorff-Nielsen and Koudou [4] that

$$p(x_{[r-1]}, x_{(r)} | \mu_{[r-1]}, \phi_{(r)}) = p(x_{[r-1]} | \mu_{[r-1]} ) p(x_{(r)} | \phi_{(r)}, x_{[r-1]}),$$

(41)

where the marginal density of $X_{[r-1]}$ is a NEF and that of $X_{(r)} | X_{[r-1]}$ is an exponential family.

From (6) and (41), we can now see that, setting $v = v_r$ and $z = z'$

$$\det \{ V(\mu_{[r-1]}, \mu_{(r)} | \mu_{[r-1]}, \phi_{(r)}) \}$$

$$\propto \exp \{ (z'_{[r-1]})^T \phi_{r}^{-1}(\mu_{[r-1]}) - v_r M_{r}^{-1}(\phi_{r}^{-1}(\mu_{[r-1]})) \}$$

$$\times \exp \{ (z_{(r)})^T \phi_{(r)} - [(z'_{[r-1]})^T A_{r[r-1]}(\phi_{(r)}) + v_r B_r(\phi_{(r)})] \}.$$

Therefore

$$\det \{ H_{rr}(\phi_{[r-1]}(\mu_{[r-1]}), \phi_{(r)}) \} = a_r(\phi_{(r)}) b_r(\phi_{[r-1]}),$$

where

$$a_r(\phi_{(r)}) \propto \exp \{(z'_{[r-1]})^T \phi_{r}^{-1} - [(z'_{[r-1]})^T A_{r[r-1]}(\phi_{(r)}) + v_r B_r(\phi_{(r)})].$$

This proves the case $k = r$. Since the distribution of $X_{[r-1]}$ belongs to a natural exponential family one can apply the previous argument to prove the case $k = r - 1$, replacing $X$ with $X_{[r-1]} = (X_{[r-2]}^T, X_{(r-1)})$, $\theta$ with $\phi_{r-1}^{-1} = (\phi_{[r-2]}^{-1}, \phi_{(r-1)})$ and $z'$ and $v_r$, respectively, with $z'^{-1}$ and $v_{r-1}$. Notice that $z'^{-1}$ can be computed applying Proposition 1 to the NEF-SQVF corresponding to $X_{(1)}, ..., X_{(r-1)}$ whose variance function can be obtained from (4) deleting the rows and columns corresponding to $X_{(r)}$. Repeating the argument for $k = r - 2, ..., 1$ completes the proof. □

**Proof of Proposition 2.** Because of Lemma 1, one can use (2) so that items (i) and (ii) follow. Item (iii) follows immediately upon noticing that (18) can be obtained from (15) by setting $s^k = \frac{1}{2} z^k$ and $n_k' = \frac{1}{2} v_k$. □

**Proof of Corollary 1.** Formula (19) follows directly applying (18) to each of the basic families using the fact that $z^k = z_{[k]}$. □

**Proof of Proposition 3.** (i) Recall from (12) that the transformation from $\phi$ to $\theta$ is block upper-triangular. Let $\tilde{\phi} = (\tilde{\phi}_{(1)}, ..., \tilde{\phi}_{(r)})^T = (\phi_{(1)}, ..., \phi_{(r)})^T$, i.e. the $\phi_{(k)}$’s listed in reverse order. Deriving the Fisher information for $\tilde{\phi}$ one establishes, because of the result mentioned in Section 2, that the reference prior for $\tilde{\phi}$ is also order-invariant; furthermore one has

$$\pi_{\tilde{\phi}_{(1)}, ..., \tilde{\phi}_{(r)}} (\tilde{\phi}_{(1)}, ..., \tilde{\phi}_{(r)}) = \pi_{\phi_{(1)}, ..., \phi_{(r)}} (\tilde{\phi}_{(r)}, ..., \tilde{\phi}_{(1)}).$$
Notice that the transformation from $\hat{\phi}$ to $\theta$ is block lower triangular, and so the reference prior for $\theta$ can be easily derived from that of $\hat{\phi}$. Specifically,

$$\theta_1, \ldots, \theta_\nu(\theta) = \pi_\phi(\theta) = \pi_\phi(\hat{\phi}(\theta)) = \pi_\phi(\theta) \mid J_\phi(\theta) = \pi_\phi(\theta),$$

since $|J_\phi(\theta)| = |J_\phi(\theta)||J_\phi(\theta)| = 1$ because $|J_\phi(\theta)| = |J_\phi(\theta)| = 1$.

Since $\pi_\phi(\cdot)$ belongs to the enriched standard conjugate family (15) with $s_k = s' = k_{j_2}/2$ and $n_k = v_k/2$, and since $\pi_\phi(\theta) = \pi_\phi(\theta)$, one can use (17) to derive the required result.

(ii) Recalling that the transformation from $\phi$ to $\mu$ is block-lower triangular, see (11), the reference prior for $\mu$ can be obtained from that of $\phi$. In fact, the reference prior $\pi_\mu(\mu)$ can be more easily derived via a double transformation, namely

$$\mu_1, \ldots, \mu_\nu(\mu) = \pi_\phi(\mu) = \pi_\phi(\theta) \mid J_\phi(\theta(\mu)) = \pi_\phi(\theta) \mid J_\phi(\theta)|\mu(\mu)||J_\mu(\mu)|.$$

Since $|J_\phi(\theta)| = 1$ and $|J_\mu(\mu)| \propto \det(V(\mu))^{-1}$, see [23, p. 34], the result follows using (24) and Proposition 1.

**Proof of Proposition 4.** Recalling that the $\phi_k$’s are independent a posteriori, it is enough to prove that each $\phi_k$ has a proper distribution.

For each basic class (or family), we shall identify, using the notation and results appearing in Table 1 of Consonni and Veronese [13], $z = \text{diag}(L_0)$ (see Eq. (5)), the parameter space $\Phi_k$, $B_k(\phi_k)$, $q$ as well as the range of $s_k$.

(i) **Poissonnormal class:** In this case conditions (C1*) and (C2*) of Section 5.2.2 hold. Indeed $v_k = 0$, because $q = 0$, so that (C1*) reduces to $n > 0$ which is trivially true. Moreover, (C2*) is also satisfied because $z_j = 1$, $s_j \in \{0, 1, 2, \ldots\}$ $(j = 1, \ldots, m)$; while $z_j = 0$ and $s_j \in \mathbb{R}$ $(j = m + 1, \ldots, d)$ and finally $(\mathcal{X}_n)^{(r)} = (0, \infty)^k$, for $(k = 1, \ldots, m)$; while $(\mathcal{X}_n^{(r)}) = (0, 0)^m \times \mathbb{R}^{k-m}$, for $(k = m + 1, \ldots, d)$.

(ii) **Multinomial family:** Conditions (C1*) and (C2*) do hold in this case, although checking them is somewhat elaborate. However the result follows immediately because the reference prior (20) for this family is proper, see Example 1.

(iii) **Negative-multinomial family:** In this case $q = 1/R$, $(R > 0)$, whence $v_k = -(k + 1)/R$, so that (C1*) is not satisfied for all $n$. We shall thus consider formula (30). Substituting $z_k = 1$, and $B_k(\phi_k) = -R \ln(1 - e^{\phi_k})$, $k = 1, \ldots, d$, the marginal posterior on $\phi_k$ becomes

$$\pi_\phi(\phi_k | s) \propto \exp\left\{\left(\frac{1}{2} + s_k\right)\phi_k\right\} + (1 - e^{\phi_k})^{Rn + \sum_{j=1}^{k-1} s_j - 1}, \quad \phi_k < 0,$$

where $s_k \in \{0, 1, 2, \ldots\}$.

Since distribution (42) induces a beta distribution on $\lambda_k = e^{\phi_k}$ with parameters $1/2 + s_k > 0$ and $Rn + \sum_{j=1}^{k-1} s_j > 0$, the result follows.

(iv) **Negative-multinomial/hyperbolic secant family:** The reference posterior on $\phi_k$, $k = 1, \ldots, d - 1$, coincides with the one given in (iii) above and consequently is always proper. Since $z_d = 0$, $q = 1/R$ $(R > 0)$, and $B_d(\phi_d) = -R \ln(\cos(\phi_d))$ the...
reference posterior on $\phi_d$ is

$$\pi_{\phi_d}(\phi_d|s) \propto \exp\{s_d \phi_d\} \cos(\phi_d) R^n + \sum_{j=1}^{d-1} s_j - \phi_d \in (-\pi/2, \pi/2),$$

while $s_d \in \mathbb{R}$.

Since $\exp\{s_d \phi_d\}$ is finite and $\cos(\phi_d)^{c-1}$, $c > 0$, is integrable on $(-\pi/2, \pi/2)$ the results follows.

(v) **Negative-multinomial/gamma/normal class**: Recall that for this class the reference posterior is provided in (31). As a consequence the reference posterior on $\phi_k$, $k = 1, \ldots, m$, coincides with the one given in (iii) above and consequently is always proper. Since $z_{m+1} = 0$, $q = 1/R$, $(R > 0)$, and $B_{m+1}(\phi_{m+1}) = -R \ln(-\phi_{m+1})$ the reference posterior on $\phi_{m+1}$ is

$$\pi_{\phi_{m+1}}(\phi_{m+1}|s) \propto \exp\{s_{m+1} \phi_{m+1}\}(-\phi_{m+1}) R^n + \sum_{j=1}^{m} s_j - \phi_{m+1} < 0,$$

with $s_k \in \{0, 1, \ldots\}$, $k = 1, \ldots, m$, while $s_{m+1} > 0$. This posterior is always proper because it corresponds to a gamma distribution on $-\phi_{m+1}$, with shape parameter $R n + \sum_{j=1}^{m} s_j > 0$ and scale parameter $s_{m+1} > 0$.

We finally consider the posterior reference on $\phi_k$ for $k = m+2, \ldots, d$. In this case $z_k = 0$ so that from (31) it follows that

$$\pi_{\phi_k}(\phi_k|s) \propto \exp\{s_k \phi_k - s_{m+1} \phi_k^2/2\} \quad \phi_k \in \mathbb{R},$$

with $s_{m+1} > 0$ and $s_k \in \mathbb{R}$, which represents the kernel of a normal distribution with mean $s_k/(s_{m+1})$ and variance $1/s_{m+1}$ and therefore is always proper.

**Proof of Proposition 5.** Let $Bin(t|n, \gamma)$ denote the binomial density with $n$ trials and probability of success $\gamma$ evaluated at $t$. The joint density of $\mathbf{X}$, expressed as a product of univariate conditional densities (given the preceding variables), can be written as

$$\prod_{i=1}^{r-1} \left\{ Bin \left( x_i \mid N - \sum_{l=1}^{i-1} x_l, \frac{p_i}{1 - \sum_{l=1}^{i-1} p_l} \right) \times \left[ \prod_{j=1}^{c-1} Bin \left( x_{ij} \mid x_i - \sum_{l=1}^{j-1} x_l, \frac{p_{j|i}}{1 - \sum_{l=1}^{j-1} p_{l|i}} \right) \right] \times \prod_{j=1}^{c-1} Bin \left( x_{jr} \mid N - \sum_{l=1}^{r-1} x_l - \sum_{m=1}^{j-1} x_{rm}, \frac{p_{j|r}}{1 - \sum_{l=1}^{r-1} p_{l|r}} \right) \right\}. \tag{43}$$

To simplify the application of Proposition 2 to our context, it is convenient to replace the double-index notation $(i,j)$ with a single index $k$, say, according to the rule: $(i,j) \mapsto (i-1) + 1; (i,j) \mapsto c(i-1) + 1 + j$, $i = 1, \ldots, r-1$, $j = 1, \ldots, c - 1; (r,j) \mapsto c(r-1) + j$, $j = 1, \ldots, c - 1$. In other words, $k$ acts as a counter along rows. We denote by $\mathbf{Y}$ the $(rc-1)$ vector corresponding to a relabeling of $\mathbf{X}$ according to the index mapping above.

Recalling that for a binomial density $Bin(\cdot|n, \gamma)$ the natural parameter $\psi$ is $\ln(\gamma/(1-\gamma))$ and the cumulant transform is $M(\psi) = n \ln(1 + e^{\psi})$, the factorization (43) identifies $(rc-1)$ parameters $\psi_k$’s together with the corresponding functions
From the structure of $M_k$ given in (10), one obtains:

$$B_k(\psi_k) = \begin{cases} 
N \ln(1 + e^{\psi_k}) & k = c(i - 1) + 1, \ i = 1, \ldots, r - 1 \\
0 & \text{otherwise}, 
\end{cases}$$

and, for $i = 1, \ldots, r - 1$ and $j = 1, \ldots, c - 1$,

$$A_{kl}(\psi_k) = \begin{cases} 
-N \ln(1 + e^{\psi_k}) & k = c(i - 1) + j \wedge (l = c(i - 1) + 1 + m), \ m = 1, \ldots, j - 1 \\
\ln(1 + e^{\psi_k}) & k = c(r - 1) + j \wedge (l = c(i - 1) + 1), \ m = 1, \ldots, r - 1 \\
0 & \text{otherwise}. 
\end{cases}$$

Of course, the pattern of zeros appearing in $A_{kl}$ merely reflects the conditional independence structure embodied in (43).

We finally identify the vectors $z_k$ and the values $v_k$ appearing in Lemma 1. Notice that $v_k = (k + 1)/N$, although its value matters only when $B_k(\psi_k) \neq 0$.

To identify $z^k$, notice first that $Y = GX$, with $G = \text{Diag}(G_1, \ldots, G_{r-1}, I_{c-1})$ where $G_1 = \cdots = G_{r-1} = G_0$, where

$$G_0 = \begin{bmatrix} 1 & \cdots & 1 & 1 \\
& I_{c-1} & \ddots \\
& & & 0 
\end{bmatrix}$$

and $I_{c-1}$ is the identity matrix of order $(c - 1)$.

Let $\mu$ denote the expected value of $X$ and write $\eta = E(Y|\mu) = G \mu$. Then the variance function of $Y$, $V_Y(\eta)$, is given by $GV(G^{-1} \eta)G^T$, where $V(\cdot)$ refers to the variance function of $X$.

Clearly, $V_Y(\cdot)$ is a SQVF, and thus admits a representation as in (4). Writing the linear component as $\sum_{k=1}^{c-1} \eta_k L_k$ and denoting with $l_k^k$ the $k$th column of $L_k$, it follows from (40) that

$$l_k^k = \sum_{i=1}^{c-1} u_k^i g_{ki},$$

where $u_k^i$ is the $i$th element of the diagonal of $U_k = \text{Diag}(L_0 g^k) = \text{Diag}(g^k)$. The last equality holds because $X$ has a multinomial distribution, so that $L_0$—defined
in (5)—corresponds to the identity matrix. It can be checked that

\[
(k^k_c)_{m} = \begin{cases} 
1 & \text{if} \quad k = m = c(i - 1) + 1 + j, \quad (i = 1, \ldots, r - 1; \quad j = 1, \ldots, c - 1) \\
or & \text{if} \quad k = m = c(i - 1) + 1, \quad (i = 1, \ldots, r - 1) \\
or & \text{if} \quad k = m = c(r - 1) + j, \quad (j = 1, \ldots, c - 1) \\
or & \text{if} \quad k \neq m = c(i - 1) + 1, \quad (i = 1, \ldots, r - 1) \\
0 & \text{otherwise}.
\end{cases}
\]

To evaluate (18) one needs to compute

\[
\sum_{j=1}^{k-1} A_{kj}(\psi_k) z^k_j + v_k B_k(\psi_k). \tag{44}
\]

Because \( B_k(\psi_k) \) and \( A_{kj}(\psi_k) \), when different from zero, are equal up to a constant, expression (44) reduces to evaluating a linear combination of \( v_k \) and the sum of the first \( k - 1 \) elements of \( z_k \). Recalling that \( z^k \) corresponds to the first \( k \) components of the vector \( \sum_{i=1}^{k} l_i \), straightforward algebra leads to conclude that (44) equals

\[
2 \ln(1 + e^{\psi_k}), \text{ for each } k.
\]

Furthermore since \( z^k_k = 1 \), the reference prior on \( \psi \) is

\[
\pi(\psi) \propto \prod_{k=1}^{r-1} \frac{e^{\psi_k/2}}{1 + e^{\psi_k}}. \tag{45}
\]

The final step concerns obtaining the reference prior for \( \tilde{p} \). However this is immediate upon realizing that the mapping \( \psi \mapsto \tilde{p} \) is lower triangular, so that one merely transforms (45), obtaining

\[
\tilde{p} \pi(\tilde{p}) = \pi(\psi) \propto \prod_{i=1}^{r-1} \left[ p_i \left( 1 - \sum_{k=1}^{i} p_k \right) \right]^{-1/2} \times \prod_{i=1}^{r} \prod_{j=1}^{c-1} \left[ p_{j|i} \left( 1 - \sum_{l=1}^{j} p_{l|i} \right) \right]^{-1/2}.
\]

References