

## ePub<sup>WU</sup> Institutional Repository

Kurt Hornik and Bettina Grün

On standard conjugate families for natural exponential families with bounded natural parameter space.

Article (Published)  
(Refereed)

*Original Citation:*

Hornik, Kurt [ORCID: https://orcid.org/0000-0003-4198-9911](https://orcid.org/0000-0003-4198-9911) and Grün, Bettina

(2014)

On standard conjugate families for natural exponential families with bounded natural parameter space.

*Journal of Multivariate Analysis*, 126.

pp. 14-24. ISSN 0047-259X

This version is available at: <https://epub.wu.ac.at/5451/>

Available in ePub<sup>WU</sup>: March 2017

*License:* [Creative Commons Attribution 3.0 Austria \(CC BY 3.0 AT\)](#)

ePub<sup>WU</sup>, the institutional repository of the WU Vienna University of Economics and Business, is provided by the University Library and the IT-Services. The aim is to enable open access to the scholarly output of the WU.

This document is the publisher-created published version.



# On standard conjugate families for natural exponential families with bounded natural parameter space

Kurt Hornik<sup>a</sup>, Bettina Grün<sup>b,\*</sup>

<sup>a</sup> Institute for Statistics and Mathematics, WU Wirtschaftsuniversität Wien, Welthandelsplatz 1, 1020 Vienna, Austria

<sup>b</sup> Department of Applied Statistics, Johannes Kepler University Linz, Altenbergerstraße 69, 4040 Linz, Austria

## ARTICLE INFO

### Article history:

Received 19 September 2012

Available online 13 January 2014

### AMS subject classifications:

62E10

62F15

62H11

### Keywords:

Bayesian analysis

Conjugate prior

Elliptical distribution

Exponential family

Linear posterior expectation

Spherical distribution

## ABSTRACT

Diaconis and Ylvisaker (1979) give necessary conditions for conjugate priors for distributions from the natural exponential family to be proper as well as to have the property of linear posterior expectation of the mean parameter of the family. Their conditions for propriety and linear posterior expectation are also sufficient if the natural parameter space is equal to the set of all  $d$ -dimensional real numbers. In this paper their results are extended to characterize when conjugate priors are proper if the natural parameter space is bounded. For the special case where the natural exponential family is through a spherical probability distribution  $\eta$ , we show that the proper conjugate priors can be characterized by the behavior of the moment generating function of  $\eta$  at the boundary of the natural parameter space, or the second-order tail behavior of  $\eta$ . In addition, we show that if these families are non-regular, then linear posterior expectation never holds. The results for this special case are also extended to natural exponential families through elliptical probability distributions.

© 2014 The Authors. Published by Elsevier Inc. Open access under [CC BY license](https://creativecommons.org/licenses/by/4.0/).

## 1. Introduction

Let  $\eta$  be a  $\sigma$ -finite measure on  $\mathbb{R}^d$ , and consider the natural exponential family (NEF)  $\mathcal{F}$  through  $\eta$ , with densities

$$f(x|\theta) = e^{\theta'x - M(\theta)}$$

with respect to  $\eta$ , where the cumulant generating function  $M(\theta)$  is defined by

$$e^{M(\theta)} = \int_{\mathbb{R}^d} e^{\theta'x} d\eta(x)$$

(e.g., [1]). Let  $\Theta = \{\theta : M(\theta) < \infty\}$  be the natural parameter space of  $\mathcal{F}$ . The family of standard conjugate distributions for  $\mathcal{F}$  (relative to the natural parameter) has densities

$$\pi(\theta|s, v) \propto e^{s'\theta - vM(\theta)}$$

with respect to the Lebesgue measure on  $\Theta$  (e.g., [6]).

\* Corresponding author.

E-mail addresses: [Kurt.Hornik@wu.ac.at](mailto:Kurt.Hornik@wu.ac.at) (K. Hornik), [Bettina.Gruen@jku.at](mailto:Bettina.Gruen@jku.at) (B. Grün).

Let  $J(s, \nu) = \int_{\Theta} e^{s\theta - \nu M(\theta)} d\theta$ . Then the hyperparameters  $s$  and  $\nu$  giving proper standard conjugate distributions are the ones for which  $J(s, \nu) < \infty$ . For Bayesian inference on  $\theta$  it is natural to employ priors from the standard conjugate family, and it is important to know when these are proper, or yield proper posteriors.

For regular NEFs (i.e.,  $\Theta$  is open) where  $\Theta$  is non-empty, Diaconis and Ylvisaker [4] show in Theorem 1 that if  $\mathcal{X}$ , the interior of the convex hull of the support of  $\eta$ , is a non-empty open set in  $\mathbb{R}^d$ , then  $J(s, \nu) < \infty$  if  $\nu > 0$  and  $s/\nu \in \mathcal{X}$ , and conversely, if  $\Theta = \mathbb{R}^d$ , then  $J(s, \nu) < \infty$  implies that  $\nu > 0$  and  $s/\nu \in \mathcal{X}$ . (Note that the reference uses  $\nu s$  where we use  $s$ .) This gives a complete characterization of all proper conjugate distributions for the case  $\Theta = \mathbb{R}^d$ , leaving open the cases where  $\Theta \subset \mathbb{R}^d$ .

In this paper, we prove that if  $\Theta$  is bounded, there exists  $-\infty \leq \nu_0 \leq 0$  such that for arbitrary  $s$ , the conjugate priors with hyperparameters  $\nu$  and  $s$  are proper for  $\nu > \nu_0$  and improper for  $\nu < \nu_0$ . We provide examples showing that all values for  $\nu_0$  in the range  $-\infty \leq \nu_0 \leq 0$  are possible.

More specific results are obtained when  $\eta$  is a (non-degenerate) spherical probability distribution on  $\mathbb{R}^d$ , i.e., a distribution invariant to orthogonal transformations. In this case,  $\mathcal{X}$  is of the form  $\{x : \|x\| < \sigma\}$ , where  $\sigma$  is finite if and only if  $\eta$  has bounded support, and  $\Theta$  is an open or closed ball with radius  $\rho$  for some  $0 < \rho \leq \infty$ . For  $\rho = \infty$ ,  $\Theta = \mathbb{R}^d$ , and the result of Theorem 1 in Diaconis and Ylvisaker [4] yields that the hyperparameters giving proper conjugate priors are those for which  $\nu > 0$  and  $\|s\| < \sigma\nu$ . For  $\rho < \infty$ , our characterization applies, and we show that lower (and/or upper) bounds for  $\nu_0$  can be derived if the behavior of the moment generating function of  $\eta$  at the boundary of the natural parameter space can be characterized via asymptotic lower (and/or upper) bound functions. In addition we establish that  $\nu_0$  can be related to the “second order tail behavior” of  $\eta$ .

If  $\theta \in \text{int}(\Theta)$ ,  $\mu(\theta) = \nabla M(\theta) = \int_{\mathbb{R}^d} x f(x|\theta) d\eta(x)$  is the mean parameter of the NEF. Diaconis and Ylvisaker [4] show in Theorem 2 that if  $\Theta$  is open and  $\theta$  has a distribution which corresponds to a proper conjugate prior with hyperparameters  $s$  and  $\nu$  satisfying  $s/\nu \in \mathcal{X}$  and  $\nu > 0$ , then  $\mathbb{E}(\nabla M(\theta)) = s/\nu$ . Clearly, in this case the posterior from an observation  $x$  is a conjugate distribution with parameters  $s + x$  and  $\nu + 1$ , so that  $\mathbb{E}(\nabla M(\theta)|x) = (s + x)/(\nu + 1)$  is linear in  $x$ . For NEFs through a spherical probability distribution with bounded  $\Theta$ , we show that  $\mathbb{E}(\nabla M(\theta))$  does not exist for  $\nu \leq 0$  if  $\Theta$  is open, and exists for all  $s$  and  $\nu$  if  $\Theta$  is closed, where in this case  $\mathbb{E}(\nabla M(\theta)) \neq s/\nu$  unless  $s = 0$  and  $\nu \neq 0$ . Finally, we show that if  $\Theta$  is closed, linear posterior expectation never holds when using canonical priors.

These results for  $\eta$  a (non-degenerate) spherical probability distribution on  $\mathbb{R}^d$  are extended to the case of elliptical distributions as given in Fang et al. [5, p. 31f]. We show that propriety of conjugate priors is only possible if the matrix in the linear transformation is a square matrix of full rank and that if the natural parameter space is bounded the value of  $\nu_0$  and the characterization of propriety for  $\nu = \nu_0$  are the same as for the corresponding spherical probability distribution. Similarly linear posterior expectation only holds in the regular case for  $\nu > 0$  and never holds in the non-regular case when canonical priors are used.

## 2. General NEFs with bounded natural parameter space

We first establish a general result on the propriety of conjugate priors for NEFs with bounded natural parameter space.

**Theorem 1.** *Let  $\eta$  be a  $\sigma$ -finite measure on  $\mathbb{R}^d$ . Suppose the natural parameter space  $\Theta$  of the NEF through  $\eta$  is bounded and non-empty. Then  $\nu \geq 0$  and arbitrary  $s$  give proper conjugate distributions, and there exists  $\nu_0 = \nu_0(\eta)$  with  $-\infty \leq \nu_0 \leq 0$  such that for arbitrary  $s$ , the conjugate distributions with parameters  $s$  and  $\nu$  are proper for  $\nu > \nu_0$ , and improper for  $\nu < \nu_0$ .*

**Proof.** If  $\Theta$  is bounded, then clearly  $J(s, \nu) < \infty$  if and only if  $K(\nu) = \int_{\Theta} e^{-\nu M(\theta)} d\theta < \infty$ , and  $K(0) = \int_{\Theta} d\theta < \infty$ . To establish the theorem, it suffices to show that if  $K(\nu_1) < \infty$ , then  $K(\nu) < \infty$  for all  $\nu > \nu_1$ . Now  $M$  is convex; the assumptions on  $\Theta$  are readily seen to imply that  $M$  is proper in the sense of Rockafellar [9]. By Corollary 12.1.2 of Rockafellar [9], there are  $x \in \mathbb{R}^d$  and  $\alpha \in \mathbb{R}$  such that  $M(\theta) \geq x'\theta + \alpha$  for all  $\theta$  (i.e.,  $M$  can be bounded below by a hyperplane). Hence, writing  $\gamma = |\alpha| + \|x\| \sup_{\theta \in \Theta} \|\theta\| < \infty$ ,  $-M(\theta) \leq \gamma$  for all  $\theta \in \Theta$ .

Now suppose  $K(\nu_1)$  is finite and  $\nu > \nu_1$ . Clearly, for all  $\theta \in \Theta$ ,  $-\nu M(\theta) = -\nu_1 M(\theta) + (\nu - \nu_1)(-M(\theta)) \leq -\nu_1 M(\theta) + (\nu - \nu_1)\gamma$  so that

$$K(\nu) \leq \int_{\Theta} e^{-\nu_1 M(\theta)} e^{(\nu - \nu_1)\gamma} d\theta = e^{(\nu - \nu_1)\gamma} K(\nu_1)$$

and hence  $K(\nu)$  is finite as well. Taking  $\nu_0 = \inf\{\nu : K(\nu) < \infty\}$ , the proof is complete.  $\square$

**Remark.** In contrast to the case where the natural parameter space is equal to  $\mathbb{R}^d$ , negative values of  $\nu$  also give proper prior distributions. In this case the parameter  $\nu$  cannot be interpreted as a prior sample size. Furthermore, the mean for the prior distribution does not necessarily exist as indicated in the example given by Diaconis and Ylvisaker [4, p. 275].

**Remark.** If  $\Theta$  is not bounded,  $J(s, \nu_1) < \infty$  does not necessarily imply that  $J(s, \nu) < \infty$  for all  $\nu \geq \nu_1$ . This can straightforwardly be seen for  $\Theta = \mathbb{R}^d$ , taking, e.g.,  $\eta$  to have the density with respect to the Lebesgue measure given by  $f(x) \propto e^{-\|x\|^2}$  for  $\min(x) := \min(x_1, \dots, x_d) > 1$ , and zero otherwise. Then clearly  $\Theta = \mathbb{R}^d$  and  $\mathcal{X} = \{x : \min(x) > 1\}$ . By Theorem 1 of Diaconis and Ylvisaker [4],  $J(s, \nu) < \infty$  if and only if  $\nu > 0$  and  $s/\nu \in \mathcal{X}$ , or equivalently, if and only if  $0 < \nu < \min(s)$ .

As a counterexample for unbounded  $\Theta \subset \mathbb{R}^d$ , take  $d = 1$  for simplicity (the case of general  $d$  can be handled by taking products as above), and let  $\eta$  have density  $f(x) \propto (x - \mu)^p e^{-x}$  for  $x > \mu > 0$  and zero otherwise, where  $p > -1$ . Then  $\mathcal{X} = (\mu, \infty)$  and

$$\begin{aligned} e^{M(\theta)} &= \int_{\mu}^{\infty} e^{\theta x} e^{-x} (x - \mu)^p dx = e^{(\theta-1)\mu} \int_{\mu}^{\infty} e^{-(1-\theta)(x-\mu)} (x - \mu)^p dx \\ &= e^{(\theta-1)\mu} \int_0^{\infty} e^{-(1-\theta)t} t^p dt = e^{(\theta-1)\mu} \frac{\Gamma(p+1)}{(1-\theta)^{p+1}} \end{aligned}$$

for  $\theta < 1$ , so that  $\Theta = (-\infty, 1)$ . Thus,

$$J(s, \nu) = \frac{1}{\Gamma(p+1)^\nu} \int_{-\infty}^1 e^{s\theta} (1-\theta)^{(p+1)\nu} e^{-\nu\mu(\theta-1)} d\theta.$$

For convergence, we need  $(p+1)\nu > -1$  (otherwise the singularity at  $\theta = 1$  integrates to infinity), i.e.,  $\nu > -1/(p+1) = \nu_0(p)$ , and  $s > \nu\mu$ , or equivalently,  $\nu_0(p) < \nu < s/\mu$ . Note that  $p \mapsto \nu_0(p)$  is increasing for  $p > -1$  with  $\lim_{p \rightarrow -1+} \nu_0(p) = -\infty$  and  $\lim_{p \rightarrow \infty} \nu_0(p) = 0$ . Hence, by suitably choosing  $p$ , we can achieve that for arbitrary negative  $\nu_0$  we have proper priors for hyperparameters satisfying  $\nu_0 < \nu < s/\mu$  (containing the range where  $\nu > 0$  and  $s/\nu \in \mathcal{X}$ , i.e.,  $\nu < s/\mu$ , as a proper subset).

### 3. NEFs through spherical probability distributions

In what follows, we restrict our attention to natural exponential families through spherical probability distributions.

Suppose that  $\eta$  is an orthogonally invariant probability measure on  $\mathbb{R}^d$ . Then if  $X$  is distributed according to  $\eta$ , it has a representation  $X = {}^dRU$  where  $U$  is uniformly distributed on the unit hypersphere  $\mathbb{S}^{d-1}$ ,  $R$  is a non-negative scalar, and  $R$  and  $U$  are independent (e.g., [5], p. 30). Equivalently, if we consider the polar decomposition  $X = \|X\|U$ , its polar part  $U$  is uniform, and independent of  $\|X\|$ . Write  $\eta_R$  for the probability measure on  $[0, \infty)$  with distribution function  $F_{\eta_R}(r) = \eta(\{x : \|x\| \leq r\})$ , and  $\eta_U$  for the uniform distribution on  $\mathbb{S}^{d-1}$ . Following [5], we will say that  $\eta$  is the spherical probability distribution generated by  $\eta_R$ .

Clearly, if  $r$  is in the support of  $\eta_R$ , then all points on the hypersphere with radius  $r$  are in the support of  $\eta$ . Hence, if we write  $\sigma = \sup\{r : F_{\eta_R}(r) < 1\}$  for the supremum of the support of  $F_{\eta_R}$ , the interior  $\mathcal{X}$  of the convex hull of the support of  $\eta$  is the open ball with radius  $\sigma$  (and hence  $\mathbb{R}^d$  if  $\eta_R$  has unbounded support).

Using the fact that

$$\int_{\mathbb{S}^{d-1}} e^{\gamma' u} d\eta_U(u) = {}_0F_1(; d/2; \|\gamma\|^2/4),$$

where

$${}_0F_1(; \nu; z) = \sum_{n=0}^{\infty} \frac{\Gamma(\nu)}{\Gamma(\nu+n)} \frac{z^n}{n!}$$

is a generalized hypergeometric series (e.g., [7], p. 168, by rewriting the modified Bessel function  $I_\nu$  in terms of  ${}_0F_1$ ) we obtain that the moment generating function of  $\eta$  satisfies

$$\begin{aligned} \int_{\mathbb{R}^d} e^{\theta' x} d\eta(x) &= \int_0^{\infty} \int_{\mathbb{S}^{d-1}} e^{r\theta' u} d\eta_U(u) d\eta_R(r) = \int_0^{\infty} {}_0F_1(; d/2; r^2\|\theta\|^2/4) d\eta_R(r) \\ &= \int_0^{\infty} \left( \sum_{n=0}^{\infty} \frac{\Gamma(d/2)}{\Gamma(d/2+n)} \frac{(r\|\theta\|)^{2n}}{4^n n!} \right) d\eta_R(r) = \sum_{n=0}^{\infty} c_n m_{2n} \|\theta\|^{2n}, \end{aligned}$$

where

$$c_n = \frac{\Gamma(d/2)}{\Gamma(d/2+n)} \frac{1}{4^n n!}, \quad m_{2n} = \int_0^{\infty} r^{2n} d\eta_R(r),$$

so that if  $d = 1$ ,  $c_n = 1/(2n)!$ ; if  $d = 2$ ,  $c_n = 1/(4^n (n!)^2)$ ; if  $d = 3$ ,  $c_n = 1/(2n+1)!$ .

Thus, the moment generating function of  $\eta$  is a function of the maximal invariant function  $\|\theta\|^2$  for the group of orthogonal transformations acting on  $\mathbb{R}^d$ , and (where finite) admits a power series representation with coefficients given by the moments of  $\eta_R$ . (Clearly,  $\int_{\mathbb{R}^d} e^{\theta' x} d\eta(x)$  can be finite for  $\theta \neq 0$  only if all moments of  $\eta_R$  are finite.)

Let

$$c(z) = \sum_{n=0}^{\infty} c_n m_{2n} z^n$$

so that  $e^{M(\theta)} = c(\|\theta\|^2)$ . Note that for  $r \geq 0$ ,  $r \mapsto c(r^2)$  is non-decreasing with  $c(0) = c_0 m_0 = 1$ . Let  $\rho^2$  be the radius of convergence of  $c(z)$ . Then clearly, if  $c(\rho^2-) = \infty$ ,  $\Theta$  is the open ball with radius  $\rho$ , and the NEF is regular; otherwise,  $\Theta$  is the closed ball with radius  $\rho$ . In analogy to the notion of the characteristic generator (for which the characteristic function at  $\theta$  equals  $\phi(\|\theta\|^2)$ ) employed by Fang et al. [5], we will refer to  $c$  as the *moment generator* of  $\eta$ .

### 3.1. Propriety of canonical priors

For spherical distributions where the parameter space  $\Theta$  is bounded Theorem 1 is applicable, but does not provide any insight on specific values of  $v_0$ . In the following we use the familiar Bachmann–Landau notations defined in the following way:

$$\begin{aligned}
 f(x) = O(g(x)) &\Leftrightarrow \exists C > 0 : \lim_{x \rightarrow a} \left| \frac{f(x)}{g(x)} \right| \leq C \\
 f(x) = \Omega(g(x)) &\Leftrightarrow \exists C > 0 : \lim_{x \rightarrow a} \left| \frac{f(x)}{g(x)} \right| \geq C \\
 f(x) \asymp g(x) &\Leftrightarrow \exists C_u, C_l > 0 : \lim_{x \rightarrow a} \left| \frac{f(x)}{g(x)} \right| \leq C_u, \lim_{x \rightarrow a} \left| \frac{f(x)}{g(x)} \right| \geq C_l \\
 f(x) = o(g(x)) &\Leftrightarrow \forall C > 0 : \lim_{x \rightarrow a} \left| \frac{f(x)}{g(x)} \right| \leq C.
 \end{aligned}$$

Using this notation the following theorem provides a characterization of  $v_0(\eta)$  for spherical  $\eta$  in terms of the behavior of the moment generator of  $\eta$  at its convergence radius.

**Theorem 2.** Let  $\eta$  be a spherical probability distribution on  $\mathbb{R}^d$  with moment generator  $c$  and  $\rho^2$  the radius of convergence of  $c$ . Suppose that  $0 < \rho < \infty$ . Let  $\beta > 0$ .

- (a) If  $\lim_{r \rightarrow \rho-} c(r^2)$  is finite,  $v_0(\eta) = -\infty$ .
- (b) If  $c(r^2) = O((\rho - r)^{-\beta})$  as  $r \rightarrow \rho-$ ,  $v_0(\eta) \leq -1/\beta$ .
- (c) If  $c(r^2) = \Omega((\rho - r)^{-\beta})$  as  $r \rightarrow \rho-$ ,  $v_0(\eta) \geq -1/\beta$ .
- (d) If  $c(r^2) \asymp (\rho - r)^{-\beta}$  as  $r \rightarrow \rho-$ ,  $v_0(\eta) = -1/\beta$ .

**Proof.** If the moment generator  $c$  has radius of convergence equal to  $\rho^2$  with  $0 < \rho < \infty$ , then the natural parameter space  $\Theta$  of the NEF through  $\eta$  satisfies  $\{\theta : \|\theta\| < \rho\} \subseteq \Theta \subseteq \{\theta : \|\theta\| \leq \rho\}$  and hence is bounded and non-empty.

Using the polar decomposition  $\theta = ru$ ,  $d\theta = a_d r^{d-1} dr d\eta_U(u)$ , where  $a_d$  is the area of  $\mathbb{S}^{d-1}$ ,

$$\begin{aligned}
 K(v) &= \int_{\Theta} e^{-vM(\theta)} d\theta = \int_0^{\rho} \int_{\mathbb{S}^{d-1}} c(r^2)^{-v} a_d r^{d-1} dr d\eta_U(u) \\
 &= a_d \int_0^{\rho} r^{d-1} c(r^2)^{-v} \int_{\mathbb{S}^{d-1}} d\eta_U(u) dr = a_d \int_0^{\rho} r^{d-1} c(r^2)^{-v} dr.
 \end{aligned}$$

By Theorem 1,  $K(v)$  is finite for  $v \geq 0$ . Hence, suppose  $v < 0$ . As  $r \mapsto c(r^2)$  is non-decreasing on  $[0, \rho]$  with  $c(0) = 1$ ,  $K(v)$  is finite if and only if  $K_{\epsilon}(v) = \int_{\rho-\epsilon}^{\rho} c(r^2)^{-v} dr$  is finite for some  $\epsilon > 0$ . This implies assertion (a), i.e., if  $\lim_{r \rightarrow \rho-} c(r^2)$  is finite,  $K_{\epsilon}$  is finite for all  $v$ .

Let  $\beta > 0$ . If  $c(r^2)$  is  $\leq$  (or  $\geq$ )  $C(\rho - r)^{-\beta}$  on  $(\rho - \epsilon, \rho)$  for positive  $\epsilon$  and  $C$ ,  $-v > 0$  implies that  $K_{\epsilon}(v)$  is  $\leq$  (or  $\geq$ )

$$\int_{\rho-\epsilon}^{\rho} (C(\rho - r)^{-\beta})^{-v} dr = C^{-v} \int_0^{\epsilon} r^{v\beta} dr,$$

respectively, which converges if and only if  $v\beta > -1$ , or equivalently, if and only if  $v > -1/\beta$ . Thus, if  $c(r^2) = O((\rho - r)^{-\beta})$  as  $r \rightarrow \rho-$ ,  $K(v)$  is finite for all  $v > -1/\beta$ , and hence  $v_0(\eta) \leq -1/\beta$  (assertion (b)). Conversely, if  $c(r^2) = \Omega((\rho - r)^{-\beta})$  as  $r \rightarrow \rho-$ ,  $K(v)$  is infinite for all  $v \leq -1/\beta$ , and hence  $v_0(\eta) \geq -1/\beta$  (assertion (b)). Assertion (d) follows by combining these two results.  $\square$

Our next result shows how for spherical  $\eta$ ,  $v_0(\eta)$  can also be characterized in terms of the tail behavior of the generating distribution  $\eta_R$ .

**Theorem 3.** Let  $\eta$  be a spherical probability distribution on  $\mathbb{R}^d$  such that  $\eta_R((r, \infty)) = \eta(\{x : \|x\| > r\}) \asymp e^{-\rho r} r^{\delta}$  as  $r \rightarrow \infty$ , for some  $\rho > 0$  and  $\delta \in \mathbb{R}$ . Then the natural parameter space  $\Theta$  of the NEF through  $\eta$  satisfies  $\{\theta : \|\theta\| < \rho\} \subseteq \Theta \subseteq \{\theta : \|\theta\| \leq \rho\}$  and hence is bounded and non-empty, and prior distributions from the standard conjugate family are proper for all  $s \in \mathbb{R}^d$  and  $v \in \mathbb{R}$  if  $\delta \leq (d - 3)/2$  (corresponding to  $v_0(\eta) = -\infty$ ), and for all  $s \in \mathbb{R}^d$  and  $v > v_0(\eta) = 1/((d - 3)/2 - \delta)$  if  $\delta > (d - 3)/2$ .

**Proof.** For  $x > 0$ ,

$$\int_0^\infty e^{-\rho r} r^{x-1} dr = \int_0^\infty e^{-u} (u/\rho)^{x-1} du / \rho = \Gamma(x) / \rho^x.$$

For  $n > 0$ ,

$$\begin{aligned} m_{2n} &= \int_0^\infty s^{2n} d\eta_R(s) = \int_0^\infty \left( \int_0^\infty 2n 1_{[0,s)}(r) r^{2n-1} dr \right) d\eta_R(s) \\ &= \int_0^\infty 2nr^{2n-1} \left( \int_0^\infty 1_{(r,\infty)}(s) d\eta_R(s) \right) dr = 2n \int_0^\infty r^{2n-1} \eta_R((r, \infty)) dr, \end{aligned}$$

interchanging the order of integration being justified by Fubini's theorem.

By assumption, there exist positive and finite constants  $C_l$  and  $C_u$  such that for all  $r$  sufficiently large,

$$0 < C_l \leq \frac{\eta_R((r, \infty))}{e^{-\rho r} r^\delta} \leq C_u < \infty.$$

By possibly modifying these constants, we can in fact assume that these inequalities hold for all  $r \geq 1$ . Then for all  $n > 0$  such that  $2n + \delta > 0$ ,

$$\begin{aligned} m_{2n} &\leq 2n \int_0^1 r^{2n-1} dr + 2n \int_1^\infty r^{2n-1} C_u e^{-\rho r} r^\delta dr \leq 1 + 2nC_u \int_0^\infty e^{-\rho r} r^{2n+\delta-1} dr \\ &= 1 + C_u \frac{2n}{2n + \delta} (2n + \delta) \frac{\Gamma(2n + \delta)}{\rho^{2n+\delta}} = 1 + C_u \frac{2n}{2n + \delta} \frac{\Gamma(2n + \delta + 1)}{\rho^{2n+\delta}}. \end{aligned}$$

Thus, there is a  $C_u^* < \infty$  such that for all  $n$  sufficiently large,

$$m_{2n} \leq C_u^* \frac{\Gamma(2n + \delta + 1)}{\rho^{2n+\delta}}.$$

Conversely,

$$\begin{aligned} m_{2n} &\geq 2n \int_1^\infty r^{2n-1} C_l e^{-\rho r} r^\delta dr = 2nC_l \left( \int_0^\infty e^{-\rho r} r^{2n+\delta-1} dr - \int_0^1 e^{-\rho r} r^{2n+\delta-1} dr \right) \\ &\geq 2nC_l \frac{\Gamma(2n + \delta)}{\rho^{2n+\delta}} - 2nC_l \int_0^1 r^{2n+\delta-1} dr = C_l \frac{2n}{2n + \delta} \frac{\Gamma(2n + \delta + 1)}{\rho^{2n+\delta}} - C_l \frac{2n}{2n + \delta}. \end{aligned}$$

Thus, there is  $C_l^* > 0$  such that for all  $n$  sufficiently large,

$$m_{2n} \geq C_l^* \frac{\Gamma(2n + \delta + 1)}{\rho^{2n+\delta}}.$$

Therefore,  $m_{2n} \asymp \Gamma(2n+\delta+1)/\rho^{2n}$  as  $n \rightarrow \infty$ , and the moment generator  $c(z) = \sum_n c_n m_{2n} z^n$  has the same convergence radius as  $\tilde{c}(z) = \sum_n c_n \Gamma(2n + \delta + 1) (z/\rho^2)^n$ , and the same asymptotic behavior as  $z$  converges to the convergence radius from below.

Using Pochhammer's symbol  $(x)_n = \Gamma(x + n) / \Gamma(x)$ ,

$$c_n = \frac{\Gamma(d/2)}{\Gamma(d/2 + n)} \frac{1}{4^n n!} = \frac{1}{(d/2)_n 4^n n!}$$

and for  $\delta \geq -1$ ,

$$\begin{aligned} \frac{\Gamma(2n + \delta + 1)}{\Gamma(\delta + 1)} &= \prod_{i=0}^{n-1} (\delta + 1 + 2i) \prod_{i=0}^{n-1} (\delta + 2 + 2i) = 4^n \prod_{i=0}^{n-1} ((\delta + 1)/2 + i) \prod_{i=0}^{n-1} ((\delta/2 + 1) + i) \\ &= 4^n \frac{\Gamma((\delta + 1)/2 + n)}{\Gamma((\delta + 1)/2)} \frac{\Gamma((\delta/2 + 1) + n)}{\Gamma(\delta/2 + 1)} = 4^n ((\delta + 1)/2)_n (\delta/2 + 1)_n. \end{aligned}$$

Thus,

$$\tilde{c}(z) = \Gamma(\delta + 1) \sum_n \frac{((\delta + 1)/2)_n (\delta/2 + 1)_n}{(d/2)_n n!} \left( \frac{z}{\rho^2} \right)^n = \Gamma(\delta + 1) {}_2F_1((\delta + 1)/2, \delta/2 + 1; d/2; z/\rho^2),$$

where  ${}_2F_1$  is the Gaussian (or ordinary) hypergeometric function. From well-known results on such functions, the radius of convergence  $\tilde{c}$  equals  $\rho^2$ . This implies that the natural parameter space  $\Theta$  of the NEF through  $\eta$  satisfies  $\{\theta : \|\theta\| < \rho\} \subseteq \Theta \subseteq \{\theta : \|\theta\| \leq \rho\}$  and hence is bounded and non-empty.

The behavior of  $\tilde{c}(z)$  at  $\rho^2$  and therefore also of  $c(z)$  can be obtained, e.g., from Olver et al. [8], Section 15.4(ii), <http://dlmf.nist.gov/15.4.ii>. We have the following three cases.

- If  $0 < d/2 - (\delta + 1)/2 - (\delta/2 + 1) = (d - 2\delta - 3)/2$ , or equivalently,  $\delta < (d - 3)/2$ ,  $\lim_{r \rightarrow \rho^-} \tilde{c}(r^2)$  exists. In this case, **Theorem 2** implies that  $v_0(\eta) = -\infty$ .
- If  $\delta > (d - 3)/2$ ,  $\tilde{c}(r^2) \sim (1 - (r/\rho))^{(d-2\delta-3)/2}$  as  $r \rightarrow \rho^-$ . We can thus use **Theorem 2** with  $\beta = -(d - 2\delta - 3)/2$  to obtain  $v_0(\eta) = -1/\beta = 2/(d - 2\delta - 3)$ .
- Finally, if  $\delta = (d - 3)/2$ ,  $\tilde{c}(r^2) \sim -\log(1 - (r/\rho))$  as  $r \rightarrow \rho^-$ , and **Theorem 2** cannot be applied directly. From the proof of **Theorem 2**, clearly,  $K(\nu) < \infty$  if and only if  $\int_{1/2}^1 (-\log(1 - r))^{-\nu} dr$  is finite. Substituting  $-\log(1 - r) = u$  so that  $r = 1 - e^{-u}$ ,

$$\int_{1/2}^1 (-\log(1 - r))^{-\nu} dr = \int_{\log(2)}^{\infty} u^{-\nu} e^{-u} du$$

which again is finite for all  $\nu$ . Hence,  $v_0(\eta) = -\infty$ .

The above assumed that  $\delta \geq -1$ . Otherwise, for  $0 \leq r < \rho$ ,  $\tilde{c}(r^2)$  is majorized by the series for  $\delta = -1$ , which always has a finite limit as  $r \rightarrow \rho^-$ . Thus  $\delta < -1$  gives  $v_0(\eta) = -\infty$ , and the proof of the theorem is complete.  $\square$

In the following examples different distributions for  $\eta_R$  are considered, with the Poisson and the negative binomial distributions as examples for discrete distributions and the Gamma distribution as an example for a continuous distribution. For these distributions it is investigated for which parameter values  $s$  and  $\nu$  prior distributions from the standard conjugate family are proper using either **Theorem 1** in Diaconis and Ylvisaker [4] or **Theorem 3**.

**Example.** Suppose  $\eta_R$  is a Poisson distribution with mean parameter  $\lambda$ . Then

$$\eta_R((r, \infty)) = 1 - \sum_{k=0}^a \frac{\lambda^k}{k!} e^{-\lambda} = \frac{1}{\Gamma(a + 1)} \int_0^\lambda t^a e^{-t} dt,$$

where  $a = \lfloor r \rfloor$ . As in the integrand,  $e^{-\lambda} \leq e^{-t} \leq 1$  for all  $0 \leq t \leq \lambda$ , it holds that

$$\eta_R((r, \infty)) \sim \frac{1}{\Gamma(a + 1)} \int_0^\lambda t^a dt = \frac{\lambda^{a+1}}{\Gamma(a + 2)} = \exp(-a(\log(a) + O(1)))$$

as  $a \rightarrow \infty$ . Hence, in this case  $\rho = \infty$  and **Theorem 1** in Diaconis and Ylvisaker [4] can be used to conclude that the prior distributions from the standard conjugate family are proper if and only if  $\nu > 0$ .

**Example.** Suppose  $\eta_R$  is a negative binomial distribution with size parameter  $n$  and success probability  $p$ , where the negative binomial distributed random variable indicates the number of failures before the first success. Then

$$\eta_R((r, \infty)) = \frac{1}{B(a + 1, n)} \int_0^p t^a (1 - t)^{n-1} dt,$$

where  $a = \lfloor r \rfloor$  and  $B(a + 1, n)$  is the Beta function. As in the integrand,  $(1 - p)^{n-1} \leq (1 - t)^{n-1} \leq 1$  for all  $0 \leq t \leq p$ , it holds that

$$\eta_R((r, \infty)) \sim \frac{1}{B(a + 1, n)} \int_0^p t^a dt = \binom{a + n}{n - 1} p^{a+1} \sim p^a a^{n-1}$$

as  $a \rightarrow \infty$ . Thus, **Theorem 3** can be applied with  $\rho = \log(p)$  and  $\delta = n - 1$  to obtain that  $v_0(\eta) = -\infty$  if  $n \leq (d - 1)/2$ , and  $v_0(\eta) = 1/((d - 1)/2 - n)$  otherwise.

**Example.** Suppose  $\eta_R$  is a Gamma distribution with shape parameter  $\alpha$  and rate parameter  $\rho$ . Then

$$\eta_R((r, \infty)) = \int_r^\infty \frac{\rho^\alpha}{\Gamma(\alpha)} e^{-\rho s} s^{\alpha-1} ds = \int_{\rho r}^\infty \frac{e^{-u} u^{\alpha-1}}{\Gamma(\alpha)} du \sim \frac{e^{-\rho r} (\rho r)^{\alpha-1}}{\Gamma(\alpha)}$$

as  $r \rightarrow \infty$  (see for example Olver et al. [8], Section 8.11(i), <http://dlmf.nist.gov/8.11.i>). Thus, **Theorem 3** can be applied with  $\rho$  and  $\delta = \alpha - 1$  to obtain that  $v_0(\eta) = -\infty$  if  $\alpha \leq (d - 1)/2$ , and  $v_0(\eta) = 1/((d - 1)/2 - \alpha)$  otherwise. If  $\eta_R$  is a shape mixture of Gamma distributions with fixed rate parameter  $\rho$ , where the mixture density is not compactly supported, the convergence radius is still  $\rho$ , but  $v_0(\eta)$  must satisfy  $v_0 > 1/((d - 1)/2 - \alpha)$  for all  $\alpha$  sufficiently large, so that  $v_0(\eta) = 0$ .

**Remark.** The example where  $\eta_R$  is a shape mixture of Gamma distributions shows that in fact all values of  $v_0$  in the range  $[-\infty, 0]$  are possible. In fact, this can also be inferred directly from **Theorem 3** by fixing  $r_0 \geq 0$  and taking  $\eta_R$  supported on  $(r_0, \infty)$  with  $\eta_R((r, \infty)) \propto e^{-\rho r} r^\delta$  for  $r \geq r_0$  ( $\delta < -1$  needs  $r_0 > 0$ ). This again gives  $v_0$  values which monotonically cover the range  $[-\infty, 0)$  as  $\delta$  varies from  $-\infty$  to  $\infty$ ;  $v_0 = 0$  can be obtained by using tails which are  $\asymp \exp(-\rho r(1 + o(1)))$  as  $r \rightarrow \infty$  but are heavier than  $e^{-\rho r} r^\delta$  for all  $\delta$ , such as tails proportional to  $\exp(-\rho r + \delta\sqrt{r})$ .

Combining the example where  $\eta_R$  is a shape mixture of Gamma distributions and the above remark we obtain the following.

**Corollary 1.** For every  $\nu_0$  with  $-\infty \leq \nu_0 \leq 0$  a  $\sigma$ -finite measure  $\eta$  on  $\mathbb{R}^d$  can be found, for which the natural parameter space  $\Theta$  of the NEF through  $\eta$  is bounded and non-empty, and for which the conjugate priors are proper for hyperparameters  $\nu > \nu_0$  and arbitrary  $s$  and improper for hyperparameters  $\nu < \nu_0$  and arbitrary  $s$ .

**Remark.** It would certainly be interesting to investigate in more detail whether and how  $\nu_0(\eta)$  can be related to the tail behavior of  $\eta$  in the general (non-spherical) case of bounded  $\Theta$  covered by Theorem 1 (assuming for example that  $\eta$  has a sufficiently nice density). We leave this for future research.

### 3.2. Linearity of posterior expectation

Next, we investigate how Theorem 2 in Diaconis and Ylvisaker [4] can be extended in the spherical case where regularity of the NEF is not necessarily given and  $\nu$  is allowed to be negative.

**Theorem 4.** Let  $\eta$  be a spherical probability distribution on  $\mathbb{R}^d$  with moment generator  $c$ , and  $0 < \rho^2 < \infty$  the radius of convergence of  $c$ . Suppose  $\theta$  follows a proper conjugate prior with parameters  $s$  and  $\nu$  for the NEF through  $\eta$ .

- (a) If  $c(\rho^2-) = \infty$  (regular case),  $\mathbb{E}(\nabla M(\theta)) = s/\nu$  for all  $s$  and  $\nu > 0$ . If  $\nu \leq 0$ , the expectation does not exist.
- (b) If  $c(\rho^2-) < \infty$  (non-regular case), then

$$\mathbb{E}(\nabla M(\theta)) = \begin{cases} \frac{s}{\nu} \left( 1 - \frac{c(\rho^2-)^{-\nu} J(s, 0)}{J(s, \nu)} \right) & \text{if } \nu \neq 0, s \in \mathbb{R}^d, \\ \frac{s}{J(s, 0)} \int_{\Theta} \log \left( \frac{c(\|\theta\|^2)}{c(\rho^2-)} \right) e^{s'\theta} d\theta & \text{if } \nu = 0, s \in \mathbb{R}^d. \end{cases}$$

**Proof.** We have

$$\mathbb{E}(\nabla M(\theta)) = \frac{1}{J(s, \nu)} \int_{\Theta} \nabla M(\theta) e^{s'\theta - \nu M(\theta)} d\theta =: E(s, \nu).$$

In the spherical case, we have  $M(\theta) = \log(c(\|\theta\|^2))$  so that for  $\theta \in \text{int}(\Theta)$

$$\nabla M(\theta) = 2 \frac{c'(\|\theta\|^2)}{c(\|\theta\|^2)} \theta$$

and

$$E(s, \nu) = \frac{2}{J(s, \nu)} \int_{\Theta} \frac{c'(\|\theta\|^2)}{c(\|\theta\|^2)^{1+\nu}} e^{s'\theta} \theta d\theta,$$

where the integrand is a positive scalar function times  $\theta$ , and can be uniformly bounded away from zero and infinity on  $\|\theta\| < \rho - \epsilon$  for all  $\epsilon > 0$ . Thus, the integral exists if and only if

$$\int_{\rho-\epsilon \leq \|\theta\| < \rho} \frac{c'(\|\theta\|^2)}{c(\|\theta\|^2)^{1+\nu}} \|\theta\| e^{s'\theta} d\theta < \infty,$$

which upon transforming to polar coordinates is easily seen to be equivalent to

$$\int_{\rho-\epsilon}^{\rho} \frac{c'(\kappa^2)}{c(\kappa^2)^{1+\nu}} 2\kappa d\kappa < \infty.$$

But the last integral is just  $c(\kappa^2)^{-\nu} / (-\nu) |_{\rho-\epsilon}^{\rho}$  for  $\nu \neq 0$  and  $\log(c(\kappa^2)) |_{\rho-\epsilon}^{\rho}$  for  $\nu = 0$ , and hence infinite if and only if  $c(\rho^2-) = \infty$  and  $\nu \leq 0$ . Hence,  $\mathbb{E}(\nabla M(\theta))$  exists unless  $c(\rho^2-) = \infty$  and  $\nu \leq 0$ .

To actually compute  $E(s, \nu)$  (if it exists), we can use  $\nabla(e^{s'\theta - \nu M(\theta)}) = (s - \nu \nabla M(\theta)) e^{s'\theta - \nu M(\theta)}$  for  $\theta \in \text{int}(\Theta)$  to obtain that for  $\nu \neq 0$ ,

$$E(s, \nu) = \frac{1}{\nu} \left( s - \frac{1}{J(s, \nu)} \int_{\Theta} \nabla e^{s'\theta - \nu M(\theta)} d\theta \right),$$



where  $\int_{\Theta} \nabla e^{s'\theta - \nu M(\theta)} d\theta$  exists if and only if  $E(s, \nu)$  exists. Write  $\theta_i$  for component  $i$  of  $\theta$ , and  $\theta_{-i}$  for  $\theta$  without component  $i$ . Then

$$\begin{aligned} \int_{\Theta} \frac{\partial}{\partial \theta_i} e^{s'\theta - \nu M(\theta)} d\theta &= \int_{\|\theta_{-i}\| < \rho} \int_{-\sqrt{\rho^2 - \|\theta_{-i}\|^2}}^{\sqrt{\rho^2 - \|\theta_{-i}\|^2}} \frac{\partial}{\partial \theta_i} e^{s'\theta - \nu M(\theta)} d\theta_i d\theta_{-i} = \int_{\|\theta_{-i}\| < \rho} e^{s'\theta - \nu M(\theta)} \left[ \frac{\sqrt{\rho^2 - \|\theta_{-i}\|^2}}{-\sqrt{\rho^2 - \|\theta_{-i}\|^2}} d\theta_{-i} \right. \\ &= c(\rho^2 - \nu)^{-\nu} \int_{\|\theta_{-i}\| < \rho} e^{s'_{-i}\theta_{-i}} \left( e^{s_i \sqrt{\rho^2 - \|\theta_{-i}\|^2}} - e^{-s_i \sqrt{\rho^2 - \|\theta_{-i}\|^2}} \right) d\theta_{-i}. \end{aligned}$$

If  $c(\rho^2 - \nu) = \infty$  and  $\nu > 0$ ,  $c(\rho^2 - \nu)^{-\nu} = 0$  and hence  $E(s, \nu) = s/\nu$  (which could also have been obtained from Theorem 2 in Diaconis and Ylvisaker [4]), establishing (a).

If  $c(\rho^2 - \nu) < \infty$ , we obtain that  $\int_{\Theta} \nabla e^{s'\theta - \nu M(\theta)} d\theta = c(\rho^2 - \nu)^{-\nu} h(s)$ , where  $h$  does not depend on  $\nu$ ; hence,  $h(s) = \int_{\Theta} \nabla e^{s'\theta} d\theta = s \int_{\Theta} e^{s'\theta} d\theta = sJ(s, 0)$ , establishing (b) for the case where  $\nu \neq 0$ . Finally,  $E(s, 0)$  can be obtained as  $\lim_{\nu \rightarrow 0} E(s, \nu)$ , where interchanging integration and taking limits is justified as the above shows that the integrals of the absolute values are uniformly bounded on compact  $\nu$  intervals. Now with  $a = c(\rho^2 - \nu)$ ,

$$E(s, \nu) = \frac{s}{\nu} \left( 1 - \frac{a^{-\nu} J(s, 0)}{J(s, \nu)} \right) = sJ(s, 0) \frac{g(s, 0) - g(s, \nu)}{\nu},$$

where  $g(s, \nu) = a^{-\nu} / J(s, \nu)$ , so that

$$\begin{aligned} \lim_{\nu \rightarrow 0} E(s, \nu) &= sJ(s, 0) \left. \frac{\partial}{\partial \nu} \frac{a^{-\nu}}{J(s, \nu)} \right|_{\nu=0} = sJ(s, 0) \left. \frac{-\log(a) a^{-\nu} J(s, \nu) - a^{-\nu} \partial J(s, \nu) / \partial \nu}{J(s, \nu)^2} \right|_{\nu=0} \\ &= sJ(s, 0) \left( \frac{-\log(a)}{J(s, 0)} + \frac{\int_{\Theta} M(\theta) e^{s'\theta} d\theta}{J(s, 0)^2} \right) = s \left( \frac{1}{J(s, 0)} \int_{\Theta} M(\theta) e^{s'\theta} d\theta - \log(a) \right) \\ &= \frac{s}{J(s, 0)} \int_{\Theta} (M(\theta) - \log(a)) e^{s'\theta} d\theta. \end{aligned}$$

Note that this can also be written as

$$E(s, 0) = \frac{s}{J(s, 0)} \int_{\|\theta\| < \rho} \log \left( \frac{c(\|\theta\|^2)}{c(\rho^2 - \nu)} \right) e^{s'\theta} d\theta$$

where the integral is always negative. □

**Remark.** More generally, for  $\theta \in \text{int}(\Theta)$

$$\nabla(e^{s'\theta - \nu M(\theta)} g(\theta)) = ((s - \nu \nabla M(\theta))g(\theta) + \nabla g(\theta)) e^{s'\theta - \nu M(\theta)},$$

so that if  $\theta$  follows a proper conjugate prior with parameters  $s$  and  $\nu$ ,

$$\mathbb{E}((\nu \nabla M(\theta) - s)g(\theta)) = \mathbb{E}(\nabla g(\theta)) - \frac{1}{J(s, \nu)} \int_{\Theta} \nabla(e^{s'\theta - \nu M(\theta)} g(\theta)) d\theta.$$

This is used by Chou [2] to show if  $\nu > 0$  and  $\Theta$  is open (so that  $e^{s'\theta - \nu M(\theta)}$  tends to 0 as  $\theta$  approaches the boundary of  $\Theta$ ) and  $g$  satisfies certain integrability conditions, the integral vanishes so that  $\mathbb{E}((\nu \nabla M(\theta) - s)g(\theta)) = \mathbb{E}(\nabla g(\theta))$ . If  $\Theta$  is not open, the integral does not necessarily vanish. In the spherical case, if  $g$  is radial (i.e.,  $g(\theta) = \phi(\|\theta\|^2)$ ), proceeding as above one can show that the integral equals  $c(\rho^2 - \nu)^{-\nu} \phi(\rho^2) sJ(s, 0)$ .

If  $\theta$  follows a proper conjugate prior with parameters  $s$  and  $\nu$ , in the regular case its posterior expectation is given by

$$\mathbb{E}(\nabla M(\theta)|x) = \frac{s + x}{\nu + 1}$$

provided that  $\nu + 1 > 0$ , and hence is linear in  $x$ . In the non-regular case, the following result shows that posterior linear expectation never occurs.

**Theorem 5.** Let  $\eta$  be the spherical probability distribution on  $\mathbb{R}^d$  generated by  $\eta_R$ , with moment generator  $c$ ,  $0 < \rho^2 < \infty$  the radius of convergence of  $c$ , and  $c(\rho^2 - \nu) < \infty$ . Suppose  $\theta$  follows a proper conjugate prior with parameters  $s$  and  $\nu$  for the NEF through  $\eta$ . Then linearity of posterior expectation does not hold.

**Proof.** From Theorem 4, if  $\theta$  follows a conjugate prior with parameters  $s$  and  $\nu$ ,  $\mathbb{E}(\nabla M(\theta)) = E(s, \nu)$  is of the form  $s\gamma(s, \nu)$  where  $\gamma$  is scalar. Linearity of posterior expectation thus holds if and only if there are  $\alpha \in \mathbb{R}$  and  $b \in \mathbb{R}^d$  such that

$$E(s + x, \nu + 1) = (s + x)\gamma(s + x, \nu + 1) = \alpha x + b$$

for all  $x$  in  $\text{supp}(\eta)$ , the support of  $\eta$ . Rewrite this as

$$(s + x)(\gamma(s + x, \nu + 1) - \alpha) = b - \alpha s.$$

If the right hand side were non-zero, necessarily  $\gamma(s + x, \nu + 1) \neq \alpha$  and hence

$$x = \frac{1}{\gamma(s + x, \nu + 1) - \alpha} (b - \alpha s) - s,$$

so that  $\text{supp}(\eta)$  were contained in the line through  $-s$  with direction  $b - \alpha s$ , which is impossible as  $\eta$  is non-degenerate and spherical (so that its support certainly contains a hypersphere with positive radius). Thus, the right hand side must be zero, and hence  $\gamma(s + x, \nu + 1) = \alpha$  for all  $x$  in  $\text{supp}(\eta)$  for which  $s + x \neq 0$ . If  $x = -s$  is in  $\text{supp}(\eta)$ , it can be approximated by elements with the same length different from  $-s$ ; thus, as  $\gamma$  is clearly continuous in its first argument,  $\gamma(s + x, \nu + 1) = \alpha$  for all  $x$  in  $\text{supp}(\eta)$ .

To simplify notations, take  $c(\rho^2) = c(\rho^2-)$ , so that  $r \mapsto c(r^2)^{-\nu}$  is continuous on  $[0, \rho]$  for all  $\nu$ , and  $r \mapsto \log(c(r^2)/c(\rho^2))$  is continuous on  $[0, \rho]$ . If  $h$  is continuous on  $[0, \rho^2]$ , transformation to polar coordinates gives

$$\begin{aligned} \int_{\Theta} e^{s'\theta} h(\|\theta\|^2) d\theta &= \int_0^\rho \int_{S^{d-1}} h(r^2) e^{(rs)'\theta} a_d r^{d-1} dr d\eta_U(u) = a_d \int_0^\rho h(r^2) {}_0F_1(; d/2; r^2 \|s\|^2/4) r^{d-1} dr \\ &= a_d \int_0^\rho h(r^2) \left( \sum_{n=0}^\infty c_n (r^2 \|s\|^2)^n \right) r^{d-1} dr = a_d \sum_{n=0}^\infty c_n \left( \int_0^\rho h(r^2) r^{2n+d-1} dr \right) \|s\|^{2n}, \end{aligned}$$

all rearrangements justified by absolute convergence. As  $h$  is bounded on  $[0, \rho^2]$ , the series

$$f_h(z) = a_d \sum_{n=0}^\infty c_n \left( \int_0^\rho h(r^2) r^{2n+d-1} dr \right) z^n$$

converges for all  $z$ , and hence defines an entire function.

If  $\nu + 1 \neq 0$ ,  $\gamma(s + x, \nu + 1) = \alpha$  for all  $x$  in  $\text{supp}(\eta)$  is equivalent to  $\Delta(x) = J(s + x, 0) - \delta J(s + x, \nu + 1) = 0$  for all  $x$  in  $\text{supp}(\eta)$ , where  $\delta = c(\rho^2)^{\nu+1} (1 - \alpha(\nu + 1))$ . Now  $J(s, \nu) = f_h(\|s\|^2)$  for  $h(r^2) = c(r^2)^{-\nu}$  and hence  $\Delta(x) = f_h(\|s + x\|^2)$  for  $h(r^2) = 1 - \delta c(r^2)^{-(\nu+1)}$ . If  $\nu + 1 = 0$ ,  $\gamma(s + x, 0) = \alpha$  for all  $x$  in  $\text{supp}(\eta)$  is equivalent to  $\Delta(x) = \int_{\Theta} \log(c(\|\theta\|^2)/c(\rho^2)) e^{(s+x)'\theta} d\theta - \alpha J(s + x, 0) = f_h(\|s + x\|^2) = 0$  for  $h(r^2) = \log(c(r^2)/c(\rho^2)) - \alpha$ . In both cases,  $h$  is monotone and does not vanish identically. As  $\rho < \infty$  is only possible if  $\eta$  has unbounded support,  $f_h(\|s + x\|^2) = 0$  for all  $x \in \text{supp}(\eta)$  implies that there is a sequence  $\sigma_k$  of non-negative reals with  $\lim_k \sigma_k \rightarrow \infty$  and  $f_h(\sigma_k) = 0$ , and the proof will be completed by showing that this is not possible.

Let us first show that we can always find an  $n_0$  such that the coefficients  $\gamma_n(h) = \int_0^\rho h(r^2) r^{2n+d-1} dr$  all have the same non-zero sign for  $n \geq n_0$ . This is trivial if  $h$  is non-negative or non-positive. If  $h$  changes from negative to positive, there exists  $0 < \beta < \rho$  such that  $h(\beta^2) > 0$ . Hence,

$$\begin{aligned} \gamma_n(h) &= \int_0^\beta h(r^2) r^{2n+d-1} dr + \int_\beta^\rho h(r^2) r^{2n+d-1} dr \geq h(0) \int_0^\beta r^{2n+d-1} dr + h(\beta^2) \int_\beta^\rho r^{2n+d-1} dr \\ &= \frac{h(\beta^2) \rho^{2n+d}}{2n+d} \left( 1 + \left( \frac{h(0)}{h(\beta^2)} - 1 \right) \left( \frac{\beta}{\rho} \right)^{2n+d} \right), \end{aligned}$$

where the term in parentheses tends to one as  $n \rightarrow \infty$ . Hence,  $\gamma_n(h) > 0$  for all  $n$  sufficiently large. Similarly, if  $h$  changes from positive to negative,  $\gamma_n(h) < 0$  for all  $n$  sufficiently large.

If  $\gamma_n(h) > 0$  for all  $n \geq n_0$  and  $\sigma > 0$ ,

$$\begin{aligned} f_h(\sigma) &= a_d \sum_{n=0}^\infty c_n \gamma_n(h) \sigma^n \geq a_d \sum_{n=0}^{n_0} c_n \gamma_n(h) \sigma^n \\ &= a_d c_{n_0} \gamma_{n_0}(h) \sigma^{n_0} \left( 1 + \sum_{0 \leq n < n_0} \frac{c_n \gamma_n(h)}{c_{n_0} \gamma_{n_0}(h)} \sigma^{n-n_0} \right) \end{aligned}$$

where the term in parentheses tends to one as  $\sigma \rightarrow \infty$ . Hence,  $f_h(\sigma) > 0$  for all  $\sigma$  sufficiently large and positive. Similarly, if  $\gamma_n(h) < 0$  for all  $n$  sufficiently large, we have  $f_h(\sigma) < 0$  for all  $\sigma$  sufficiently large and positive, and the proof is complete.  $\square$

**Remark.** The results of Theorems 3 and 4 in Diaconis and Ylvisaker [4] and Chou [3] characterize the canonical priors as those achieving linear posterior expectation, assuming that  $\Theta$  is open. Theorem 5 shows that if  $\Theta$  is closed, then the canonical priors never achieve linear posterior expectation in the case of NEFs through spherical probability distributions. This leaves open the question whether linear posterior expectation could be achieved by other priors.

#### 4. NEFs through elliptical probability distributions

Following Fang et al. [5, p. 31f], we say that a probability distribution  $\eta$  on  $\mathbb{R}^d$  is *elliptical* if for some  $k \in \{1, \dots, d\}$  there are  $m \in \mathbb{R}^d$ , a  $k \times d$  matrix  $A$  of rank  $k$ , and a spherical probability distribution  $\eta_S$  on  $\mathbb{R}^k$  such that  $\eta(B) = \eta_S(\{y : A'y + m \in B\})$  for all Borel sets  $B$  in  $\mathbb{R}^d$ . In this section, we discuss how the results in the previous section can be generalized from NEFs through spherical distributions to NEFs through elliptical distributions.

We have

$$\int_{\mathbb{R}^d} e^{\theta'x} d\eta(x) = \int_{\mathbb{R}^k} e^{\theta'(A'y+m)} d\eta_S(y) = e^{\theta'm} \int_{\mathbb{R}^k} e^{(A\theta)'y} d\eta_S(y).$$

Hence, with  $M$  and  $M_S$  the cumulant generating functions of  $\eta$  and  $\eta_S$ , respectively, we have

$$M(\theta) = \theta'm + M_S(A\theta).$$

The NEF through  $\eta$  has densities given by

$$f(x|\theta) = e^{\theta'x - M(\theta)} = e^{\theta'(x-m) - M_S(A\theta)},$$

natural parameter space  $\Theta = \{\theta : M_S(A\theta) < \infty\}$ , and standard conjugate distributions with densities

$$\pi(\theta|s, \nu) \propto e^{s'\theta - \nu M(\theta)} = e^{(s-\nu m)'\theta - \nu M_S(A\theta)}$$

with normalizing constants

$$J(s, \nu) = \int_{\Theta} e^{(s-\nu m)'\theta - \nu M_S(A\theta)} d\theta$$

if proper (i.e.,  $J(s, \nu) < \infty$ ). If  $k < n$ , the null space of  $A$  must at least have dimension one, so that  $\Theta$  is always unbounded and  $J(s, \nu)$  is never finite. Thus, in what follows we only consider the case where  $k = d$  and hence  $A$  is invertible, in which case  $\Theta$  is bounded if and only if the natural parameter space  $\Theta_S$  of the NEF through  $\eta_S$  is bounded, and

$$\begin{aligned} J(s, \nu) &= \int_{\theta: M_S(A\theta) < \infty} e^{(s-\nu m)'A^{-1}A\theta} e^{-\nu M_S(A\theta)} d\theta = |\det(A)|^{-1} \int_{\gamma: M_S(\gamma) < \infty} e^{(s-\nu m)'A^{-1}\gamma} e^{-\nu M_S(\gamma)} d\gamma \\ &= |\det(A)|^{-1} J_S(\tau(A)(s - \nu m), \nu) \end{aligned}$$

where  $\tau(A) = (A^{-1})'$ . Thus, the set of parameters  $s$  and  $\nu$  giving proper conjugate priors for the NEF through  $\eta$  can straightforwardly be obtained from the corresponding set for the NEF through  $\eta_S$ .

The above computations did not use the fact that  $\eta_S$  is spherical. Actually doing so and writing  $c_S$  for the moment generator of  $\eta_S$ , we have  $M_S(A\theta) = c_S(\|A\theta\|^2) = c_S(\theta'A'A\theta)$ . Thus,  $M(\theta)$  and hence also the distribution of  $\eta$  depend on  $A$  only through  $\Sigma = A'A$ , and if  $k = d$  we can take  $A = \Sigma^{1/2}$  as the symmetric root of  $\Sigma$  (in which case  $\tau(A) = \Sigma^{-1/2}$ ).

If  $c_S$  has finite radius of convergence  $\rho_S$  (so that  $\Theta$  and  $\Theta_S$  are bounded), the above relation between  $J$  and  $J_S$  implies that  $\nu_0(\eta) = \nu_0(\eta_S)$  and hence can be characterized using [Theorems 2 and 3](#).

For hyperparameters  $s$  and  $\nu$  giving a proper conjugate prior for the NEF through  $\eta$ , write

$$E(s, \nu) = \mathbb{E}(\nabla M(\theta)|s, \nu)$$

(where the expectation is taken with respect to this prior), and write  $E_S$  for the corresponding quantity for  $\eta_S$ . As clearly  $\nabla M(\theta) = m + A'\nabla M_S(A\theta)$  on  $\text{int}(\Theta)$ , we have

$$\begin{aligned} E(s, \nu) &= \frac{1}{J(s, \nu)} \int_{\Theta} \nabla M(\theta) e^{s'\theta - \nu M(\theta)} d\theta = m + \frac{1}{J(s, \nu)} \int_{\Theta} A'\nabla M_S(A\theta) e^{(s-\nu m)'A^{-1}A\theta} e^{-\nu M_S(A\theta)} d\theta \\ &= m + A'E_S(\tau(A)(s - \nu m), \nu), \end{aligned}$$

where in turn  $E_S$  can be obtained using [Theorem 4](#). In particular, in the regular case (a), we have  $E_S(s, \nu) = s/\nu$  for  $\nu > 0$ , and hence also

$$E(s, \nu) = m + A'\tau(A)(s - \nu m)/\nu = m + (s - \nu m)/\nu = s/\nu.$$

Finally, let us show that if the NEF through  $\eta$  is non-regular (i.e., if  $\Theta = \{\theta : \theta'\Sigma\theta \leq \rho_S^2\}$ ), posterior linear expectation never occurs. As the proof of [Theorem 5](#) shows, linearity of posterior expectation is equivalent so the existence of  $\alpha$  and  $b$  such that

$$\alpha x + b = E(s + x, \nu + 1) = m + A'E_S(\tau(A)(s + x - (\nu + 1)m), \nu)$$

or equivalently,

$$E_S(\tau(A)(s + x - (\nu + 1)m), \nu) = \tau(A)(\alpha x + (b - m))$$

for all  $x$  in the support of  $\eta$ . But clearly,  $x$  is in the support of  $\eta$  if and only if  $y = \tau(A)(x - m)$  is in the support of  $\eta_S$ , so that the above is equivalent to having

$$E_S(\tau(A)(s - \nu m) + y, \nu + 1) = \alpha y + \tau(A)(b - m + \alpha m)$$

for all  $y$  in the support of  $\eta_S$ . This in turn is equivalent to linearity of posterior expectation for the conjugate prior for the NEF through  $\eta_S$  with parameters  $\tau(A)(s - \nu m)$  and  $\nu$ , which by [Theorem 5](#) cannot hold if this family (and equivalently the NEF through  $\eta$ ) is non-regular.

## Acknowledgments

We thank an anonymous reviewer for suggesting to also investigate the case of NEFs through elliptical probability distributions. This research was funded by the Austrian Science Fund (FWF): V170-N18.

## References

- [1] O. Barndorff-Nielsen, *Information and Exponential Families in Statistical Theory*, in: *Wiley Series in Probability and Mathematical Statistics*, John Wiley & Sons, Ltd., 1978.
- [2] J.-P. Chou, An identity for multidimensional continuous exponential families and its applications, *J. Multivariate Anal.* 24 (1988) 129–142.
- [3] J.-P. Chou, Characterization of conjugate priors for discrete exponential families, *Statist. Sinica* 11 (2001) 409–418.
- [4] P. Diaconis, D. Ylvisaker, Conjugate priors for exponential densities, *Ann. Statist.* 7 (2) (1979) 269–281.
- [5] K.-T. Fang, S. Kotz, K.-W. Ng, *Symmetric Multivariate and Related Distributions*, Chapman and Hall, London, 1990.
- [6] E. Gutiérrez-Peña, A.F.M. Smith, Exponential and Bayesian conjugate families: review and extensions, *Test* 6 (1997) 1–90.
- [7] K.V. Mardia, P.E. Jupp, *Directional Statistics*, in: *Probability and Statistics*, Wiley, 1999.
- [8] F.W.J. Olver, D.W. Lozier, R.F. Boisvert, C.W. Clark (Eds.), *NIST Handbook of Mathematical Functions*, Cambridge University Press, 2010.
- [9] R.T. Rockafellar, *Convex Analysis*, Princeton University Press, Princeton, NJ, 1997.