The Annals of Statistics
1994, Vol. 22, No. 3, 1555–1586

SAMPLING MODELS WHICH ADMIT A GIVEN GENERAL EXPONENTIAL FAMILY AS A CONJUGATE FAMILY OF PRIORS

By Shaul K. Bar-Lev, Peter Enis and Gérard Letac

University of Haifa, State University of New York at Buffalo and Université Paul Sabatier

Let $\mathcal{K} = \{K_\lambda: \lambda \in \Lambda\}$ be a family of sampling distributions for the data x on a sample space $\mathcal{X}$ which is indexed by a parameter $\lambda \in \Lambda$, and let $\mathcal{F}$ be a family of priors on Λ. Then $\mathcal{F}$ is said to be conjugate for $\mathcal{K}$ if it is closed under sampling, that is, if the posterior distributions of λ given the data x belong to $\mathcal{F}$ for almost all x. In this paper, we set up a framework for the study of what we term the dual problem: for a given family of priors $\mathcal{F}$ (a subfamily of a general exponential family), find the class of sampling models $\mathcal{K}$ for which $\mathcal{F}$ is conjugate. In particular, we show that $\mathcal{K}$ must be a general exponential family dominated by some measure Q on $(\mathcal{X}, \mathcal{B})$, where $\mathcal{B}$ is the Borel field on $\mathcal{X}$. It is the class of such measures Q that we investigate in this paper. We study its geometric features and general structure and apply the results to some familiar examples.

1. Introduction and background. We begin by stating, in an informal manner, the problem we address in this paper. A formal and rigorous approach will be made in subsequent sections. Let $\mathcal{K}$ be a family of sampling distributions for the data x on a sample space $\mathcal{X}$, which is indexed by a parameter $\lambda \in \Lambda$. A family $\mathcal{F}$ of prior distributions on Λ is said to be a conjugate family for $\mathcal{K}$ if it is closed under sampling, that is, if the posterior distributions of λ given the sample observations x belong to $\mathcal{F}$.

In addition to permitting easy calculations of posterior distributions, conjugate priors have the intuitively attractive property [Berger (1980), page 70] "of allowing one to begin with a certain functional form for the prior and end up with a posterior of the same functional form, but with parameters updated by the sample information."

A subject which has been of some interest in this area—and one which has been discussed frequently in the literature [e.g., Raiffa and Schlaifer (1961), Chapter 3, and Lindley (1972), pages 22–23]—concerns the following problem: given a family $\mathcal{K}$ of sampling distributions, determine a family of prior distributions $\mathcal{F}$ which is conjugate for $\mathcal{K}$. For the case where $\mathcal{K}$ is a natural exponential family (NEF), Diaconis and Ylvisaker [(1979), Theorem 1] provide an explicit expression for a conjugate family of prior distributions for $\mathcal{K}$, which turns out to be a general exponential family (GEF).

In this paper, we set up a framework for the study of what we term a "dual

Received October 1990; revised November 1993.
AMS 1991 *subject classifications.* Primary 62A15; secondary 60E99.
Key words and phrases. General exponential family, natural exponential family, conjugate family of priors, variance function, Diaconis–Ylvisaker family.

problem" to that considered by Diaconis and Ylvisaker (1979): given a GEF $\mathcal{F}$ of prior distributions on Λ, find all families of sampling distributions $\mathcal{K}$ for which $\mathcal{F}$ is conjugate, and determine their structure.

The "dual problem" proposed here can be looked upon as a decoding problem; that is, roughly speaking, knowing the prior (or, equivalently, the posterior) family of distributions employed by an individual who is known to use conjugate priors, we wish to determine some aspects of the structure and form of the unknown sampling distribution of the phenomenon being studied.

In order to provide some motivation for notions to be introduced in the sequel, we now briefly exemplify and sketch in a nonrigorous manner, for the special case where $\mathcal{F}$ is a GEF of order 2, some of the ideas we adopt to solve the dual problem. Rigorous arguments and definitions for the general case will be presented in subsequent sections.

Let $\mathcal{F}$ be family of priors on Λ which constitutes a GEF of order 2 with densities

$$\exp\{\theta_1 u_1(\lambda) + \theta_2 u_2(\lambda) - r(\theta_1, \theta_2)\}\nu(d\lambda), \tag{1.1}$$

where $\theta := (\theta_1, \theta_2) \in D \subset \mathbb{R}^2$ and D is the parameter space of $\mathcal{F}$ determined by ν and (u_1, u_2). In the terminology of Barndorff-Nielsen (1978), θ and $u(\lambda) := (u_1(\lambda), u_2(\lambda))$ are called the canonical parameter and canonical statistic of $\mathcal{F}$, respectively. If there exists a sampling model $\mathcal{K} = \{K_\lambda(dx) : \lambda \in \Lambda\}$ on $\mathcal{X}$, for which $\mathcal{F}$ is conjugate, then, for any $\theta \in D$ and almost all x, the posterior density of λ given the data x has the form (1.1) with $\theta^* = (\theta_1^*, \theta_2^*)$ replacing θ, where $\theta^* = h(\theta, x) \in D$ may depend on both θ and x. Consequently, $K_\lambda(dx)$ necessarily has the form

$$K_\lambda(dx) = \exp\{(\theta_1^* - \theta_1)u_1(\lambda) + (\theta_2^* - \theta_2)u_2(\lambda) + r(\theta_1^*, \theta_2^*) - r(\theta_1, \theta_2)\}\mathbf{r}(dx),$$

for some measure $\mathbf{r}$. Under mild conditions, we show (Theorem 3.2) that there exists a measure Q on $\mathcal{X}$ such that, for all $\theta \in D$ and Q-almost all x, $h(\theta, x) - \theta := h(x)$ does not depend on θ and that $K_\lambda(dx) = \exp\{h_1(x)u_1(\lambda) + h_2(x)u_2(\lambda)\}Q(dx)$. Accordingly, to solve the dual problem, it is sufficient to find all measures Q on $\mathcal{X}$ such that

$$\int_{\mathcal{X}} \exp\{h_1(x)u_1(\lambda) + h_2(x)u_2(\lambda)\}Q(dx) = 1. \tag{1.2}$$

The set of all measures Q on $\mathcal{X}$ satisfying (1.2) is rather large. Fortunately, the resulting $\mathcal{K} = \{\exp\{h_1(x)u_1(\lambda) + h_2(x)u_2(\lambda)\}Q(dx) : \lambda \in \Lambda\}$ can be seen to constitute a GEF which then can be reduced to an NEF on $\mathbb{R}^2$ by the map $x \mapsto h(x)$. To obtain a better understanding of the latter statements, we consider, for example, the case where $\mathcal{F}$ is a family of gamma priors. (We shall return to this example in Section 6.)

EXAMPLE 1.1. Here,

$$\begin{aligned}\mathcal{F} = \Big\{&\exp\{\theta_1 \log\lambda - \theta_2\lambda + \theta_1 \log\theta_2 - \log\Gamma(\theta_1)\} \\ &\times \lambda^{-1}\mathbf{1}_{\mathbb{R}^+}(\lambda)\,d\lambda : \theta \in D = \mathbb{R}^+ \times \mathbb{R}^+\Big\};\end{aligned} \tag{1.3}$$

$\mathcal{F}$ is conjugate for Poisson models of sampling distributions $\mathcal{K} := \mathcal{K}_n = \{K_{\lambda,n}(dx): \lambda \in \mathbb{R}^+\}$, $n \in \mathbb{N}$, defined as follows. If $x = (x_1, \dots, x_n)$ and the x_i's are i.i.d. and follow a Poisson distribution with mean λ, then, for $n \in \mathbb{N}$,

$$K_{\lambda,n}(dx) := \exp\left\{\sum_{i=1}^{n} x_i \log\lambda - n\lambda\right\} Q_n(dx),$$

where $Q_n(dx) = \eta_n(dx)/\prod_{i=1}^n x_i!$ and η_n is the counting measure on $\mathbb{N}_0^n$. Accordingly, the set of all measures Q satisfying (1.2) will contain, in this example, the subset $\{Q_n(dx), n \in \mathbb{N}\}$, or, equivalently, the set of all sampling models for which $\mathcal{F}$ in (1.3) is conjugate will contain the subset $\{\mathcal{K}_n, n \in \mathbb{N}\}$. Note that, for $n \in \mathbb{N}$, $\mathcal{K}_n$ is a GEF. The NEF associated with the GEF $\mathcal{K}_n$ is composed of the marginal distributions of the minimal sufficient statistic $\sum_{i=1}^n x_i$. If we let q be the image of Q_n under the map $x \mapsto \sum_{i=1}^n x_i := t$, then, for all $n \in \mathbb{N}$, $q(dt) = \eta_1(dt)/t!$, and the NEF associated with $\mathcal{K}_n$ is $\{\exp\{\xi_n t - \exp(\xi_n)\}q(dt), \xi_n := \log n\lambda \in \mathbb{R}\}$. A substantial reduction is therefore achieved, since the subset $\{Q_n(dx): n \in \mathbb{N}\}$, whose elements satisfy (1.2), can be looked upon as being represented by the single measure q.

Accordingly, we let q denote the image of Q by the map $x \mapsto h(x)$. Since $h(x) = h(\theta, x) + \theta$ and $h(\theta, x) \in D$ for all $\theta \in D$ and Q-almost all x, it follows that $h(x) \in G := \{t \in \mathbb{R}^2: t + D \subset D\}$ and hence the support of q is contained in G. (Note that such a G is a closed additive semigroup of $\mathbb{R}^2$.) Also, let S denote the image of Λ by the map $\lambda \mapsto u(\lambda)$. The GEF $\mathcal{K}$ is then associated with the NEF generated by q whose elements have the form $\exp\{h_1u_1 + h_2u_2\}q(dh)$, $h \in G$, $u \in S$. Accordingly, the dual problem can now be reduced to finding the set $\mathcal{H}(S, G)$ of all measures q concentrated on G such that, for all $u \in S$,

$$\int_{\mathbb{R}^2} \exp\{h_1u_1 + h_2u_2\}q(dh) = 1. \tag{1.4}$$

One of the key tools for studying $\mathcal{H}(S, G)$ is linked with the use of the variance function of an NEF. [The variance function characterizes the NEF within the class of NEF's; cf. Morris (1982) and Letac and Mora (1990).] To illustrate this use, consider a special case of (1.4). Assume, for instance, that q is concentrated on an affine hyperplane $\{h \in G: h_1 = ah_2 - b,\ a \in \mathbb{R}, b \neq 0\}$ of $\mathbb{R}^2$. For this special case, (1.4) reduces to

$$\int_{\mathbb{R}} \exp\Big\{h_2\big[au_1(\lambda) + u_2(\lambda)\big] - bu_1(\lambda)\Big\}q(dh_2) = 1. \tag{1.5}$$

For simplicity, assume that $\xi := au_1(\lambda) + u_2(\lambda)$ is one-to-one on Λ. Let Ξ denote the image of Λ by the map $\lambda \mapsto au_1(\lambda) + u_2(\lambda)$, and let $k(\xi)$ denote $bu_1(\lambda)$ expressed in terms of ξ. If (1.5) is valid, then $\{\exp\{h_2\xi - k(\xi)\}q(dh_2): \xi \in \Xi\}$ constitutes an NEF on $\mathbb{R}$, and $\exp\{k(\xi)\}$ is the Laplace transform of q. (Note that the canonical parameter and Laplace transform of such an NEF, if it exists, are functionally related to the canonical statistic u of $\mathcal{F}$.) Does there exist such a Laplace transform? A positive answer will provide, of course, an element of $\mathcal{H}(S, G)$, in which case $k'(\xi)$ and $k''(\xi), \xi \in \operatorname{int}\Xi$ will be the mean and variance, respectively, of the NEF generated by q. Writing $m := k'(\xi), M := k'(\operatorname{int}\Xi)$ and $V(m)$ for $k''(\xi)$ expressed in terms of m, then the pair (V, M) is the variance function corresponding to such an NEF. In many instances it would appear that determining whether (V, M) is a variance function of an NEF is easier than determining whether $\exp\{k(\xi)\}$ is a Laplace transform. Indeed, we are fortunate to have lists of pairs (V, M) which are variance functions on NEF's on $\mathbb{R}$ and lists of pairs which are not. A list of all variance functions in which V is either quadratic or cubic can be found in Morris (1982) and Mora (1986) [see also Letac and Mora (1990)]. Additional lists can be found in Bar-Lev and Bshouty (1989, 1990), Bar-Lev, Bshouty and Enis (1991, 1992) and Jørgensen (1987). Such lists include cases in which the variance functions have simple functional form whereas the corresponding Laplace transform $\exp\{k(\xi)\}$ cannot be expressed explicitly in terms of ξ [in the above special case such a situation occurs when the equation $\xi = au_1(\lambda) + u_2(\lambda)$ cannot be solved for λ, although the mapping $\lambda \mapsto \xi$ is assumed to be one-to-one]. We now illustrate the above special case, where $h_1 = ah_2 - b$, for the family of gamma priors given by (1.3).

EXAMPLE 1.2. Here, $D = \mathbb{R}^+ \times \mathbb{R}^+, (u_1(\lambda), u_2(\lambda)) = (\log\lambda, -\lambda), G = D$ and S is the set of points on the curve $\{(\log\lambda, -\lambda), \lambda \in \mathbb{R}^+\}$. Setting $\xi = a\log\lambda - \lambda$, we consider, for simplicity, two special cases: (i) $a = 0$; and (ii) $a < 0$.

(i) We have $\xi = -\lambda, \Xi = \mathbb{R}^-, k(\xi) = b\log(-\xi), m = -b/\xi$; $M = \mathbb{R}^+$, if $b > 0$, and $M = \mathbb{R}^-$, if $b < 0$; and $V(m) = b^{-1}m^2$. Since V is positive on M, the case $b < 0$ is excluded. The remaining case $(V, M) = (b^{-1}m^2, \mathbb{R}^+), b > 0$, is, by Morris' (1982) classification, the variance function of the gamma NEF composed of gamma distributions with shape parameter b and (inverse) scale parameter bm^{-1}. The dominating measure $q(dh_2)$ of such an NEF is easily obtained. One then still has to check, however, that the resulting q belongs to $\mathcal{H}(S, G)$.

(ii) We have $\xi = a\log\lambda - \lambda$ and $\Xi = \mathbb{R}$. Here, λ and hence $k(\xi)$ are not expressible in terms of ξ, although the map $\lambda \mapsto \xi$ is one-to-one. The variance function, though, has a simple functional form. Since $m = k'(\xi) = (bu_1(\lambda))'/\xi' = b(a-\lambda)^{-1}$, we find that $M = (b/a, 0)$, if $b > 0$, and $M = (0, b/a)$, if $b < 0$. Also, $V(m) = k''(\xi) = b\lambda(a-\lambda)^{-3} = ab^{-2}m^2(m - b/a)$, hence the case $b < 0$ is excluded since otherwise V is negative. The remaining pair $(V, M) = (ab^{-2}m^2(m-b/a), (b/a, 0))$ is not a variance function by the Mora (1986) classification of cubic variance functions.

The family of Poisson sampling distributions for which the gamma family of priors is conjugate can be delineated in a manner similar to the above by considering other affine hyperplanes on which q is concentrated. We will realize, however, in subsequent sections that there is no need to analyze all possible affine hyperplanes as some of these are equivalent in some sense.

Having illustrated some of the notions and ideas which will be used in the sequel, we present in Section 2 formal definitions of NEF's, GEF's, Markov kernels and posterior distributions and give a rigorous formulation of the dual problem. A study of the general structure of the sampling models $\mathcal{K}$ for which $\mathcal{F}$ is conjugate is given in Section 3. Section 4 is devoted to a general study of the set of measures $\mathcal{H}(S,G)$. In particular, we study in Section 3 and 4 the general case where a subfamily of $\mathcal{F}$ is considered as our family of priors. This is equivalent to considering $\mathcal{F}$ restricted to a proper subset of the natural parameter space D. This, is in turn, may have an effect on the structure of G—the closed additive semigroup which contains the support of q—and, consequently, on the structure of $\mathcal{H}(S,G)$. Section 5 applies such a study to the case where $\mathcal{F}$ is a Diaconis–Ylvisaker family [see definition in (2.15) below] and related cases. Section 6 is devoted to some familiar examples. The use of the variance function as a tool, as described above, is demonstrated in Propositions 5.2 and 5.7 and is applied in the examples of Section 6.

2. Notation and definitions. We first introduce some notation. We then recall some definitions of NEF's, GEF's, Markov kernels, posterior distributions and what we term the Diaconis–Ylvisaker family of prior distributions. These will be needed for the presentation of our results throughout the paper.

We denote by E and V finite-dimensional real linear spaces; by dim E, the dimension of E; and by E^* the linear space dual to E, that is, E^* is the linear space of linear forms on E. The canonical bilinear form on $E^* \times E$ is denoted by $(\theta,x) \mapsto \langle\theta,x\rangle$. For $H \subset E$, $\operatorname{int} H$ designates the interior of H. If $(A,\mathcal{A})$ and $(B,\mathcal{B})$ are measurable spaces, ν a measure defined on $(A,\mathcal{A})$ and $t: A \mapsto B$ an $\mathcal{A}$-measurable map, then the measure $\nu(t^{-1}(B^*))$, $B^* \in \mathcal{B}$, which is the image of ν by the latter map, is denoted by $t_*\nu$. The support of ν is denoted by $S(\nu)$ and the closed convex hull of $S(\nu)$ by $\operatorname{conv}(S(\nu))$. The Dirac mass on $x \in E$ is denoted by δ_x.

Let μ be a positive measure on E. The *Laplace transform* L_μ of μ and the *effective domain* $D(\mu)$ of L_μ are define, respectively, by

$$L_\mu(\theta) := \int_E \exp\{\langle\theta,x\rangle\}\mu(dx) \le \infty, \tag{2.1}$$

and

$$D(\mu) := \{\theta \in E^*: L_\mu(\theta) < \infty\}. \tag{2.2}$$

Let $\Theta(\mu) := \operatorname{int} D(\mu)$. Two sets $D(\mu)$ and $\Theta(\mu)$ are convex, and the map

$$k_\mu := \log L_\mu: D(\mu) \to \mathbb{R} \tag{2.3}$$

is convex on $D(\mu)$ and real analytic on $\Theta(\mu)$. Define

$$\overline{\mathcal{M}}(E) := \{\text{Measures } \mu \text{ on } E: \Theta(\mu) \neq \phi\} \tag{2.4}$$

and

$$\mathcal{M}(E) := \{\mu \in \overline{\mathcal{M}}(E): \mu \text{ is not supported on an affine hyperplane of } E\}. \tag{2.5}$$

Let $\mu \in \mathcal{M}(E)$ and $\theta \in D(\mu)$. Define a probability measure on E by

$$P(\theta,\mu)(dx) := \exp\{\langle\theta,x\rangle - k_\mu(\theta)\}\mu(dx). \tag{2.6}$$

The two sets of probabilities

$$F(\mu) := \{P(\theta, \mu) \colon \theta \in \Theta(\mu)\} \quad \text{and} \quad \overline{F}(\mu) := \{P(\theta, \mu) \colon \theta \in D(\mu)\} \tag{2.7}$$

are called the NEF and full NEF (FNEF) generated by $\mu \in \mathcal{M}(E)$. For an FNEF $\overline{F}(\mu)$, the effective domain $D(\mu)$ of L_μ is also called the natural parameter space. Consider an NEF $F(\mu)$. The map $k'_\mu := \partial k_\mu / \partial\theta \colon \Theta(\mu) \to \mathbb{R}$ is injective and called the *mean function* of $F(\mu)$. The image $M_{F(\mu)} := k'_\mu(\Theta(\mu))$ is called the *mean domain* of $F(\mu)$. An FNEF $\overline{F}_\mu$ is called *regular* if $D(\mu) = \Theta(\mu)$ and *steep* if $M_{F(\mu)} = \operatorname{int}\operatorname{conv}(S(\mu))$ [see Barndorff-Nielsen (1978), Theorems 8.2 and 9.2].

Having defined NEF's, GEF's are defined on an abstract measurable space $(\Lambda, \mathcal{A})$ as follows. Let ν be a measure defined on $(\Lambda, \mathcal{A})$ and $u \colon \Lambda \to E$ a measurable map for which $\mu := u_*\nu \in \mathcal{M}(E)$. Then, for $\theta \in D(\mu)$,

$$P(\theta, u, \nu)(d\lambda) := \exp\{\langle \theta, u(\lambda)\rangle - k_\mu(\theta)\}\nu(d\lambda) \tag{2.8}$$

defines a probability on $(\Lambda, \mathcal{A})$. The two sets of probabilities

$$F(u, \nu) := \{P(\theta, u, \nu) \colon \theta \in \Theta(\mu)\} \quad \text{and} \quad \overline{F}(u, \nu) := \{P(\theta, u, \nu) \colon \theta \in D(\mu)\} \tag{2.9}$$

are called, respectively, the GEF and the full GEF (FGEF) generated by u and ν. In this case, the NEF $F(\mu)$ is called the NEF associated with GEF $F(u, v)$. Note that important properties of GEF's are actually properties of the associated NEF's.

We now recall definitions of some Bayesian concepts. We first define a Markov kernel.

DEFINITION 2.1. Let $(\Lambda, \mathcal{A})$ and $(\mathcal{X}, \mathcal{B})$ be measurable spaces. A *Markov kernel* $\mathcal{K}$ from Λ to $\mathcal{X}$ is a map $(\lambda, B) \mapsto K_\lambda(B)$ from $(\Lambda, \mathcal{B})$ into the interval [0, 1] such that the following hold: (i) for every $\lambda \in \Lambda, K_\lambda(\cdot)$ is a probability distribution on $(\mathcal{X}, \mathcal{B})$; (ii) for every $B \in \mathcal{B}, K_\lambda(B)$ is $\mathcal{A}$-measurable.

It is important to note that a Markov kernel $\mathcal{K}$ from Λ to $\mathcal{X}$ forms a family $\mathcal{K} = \{K_\lambda \colon \lambda \in \Lambda\}$ of probability distributions on $(\mathcal{X}, \mathcal{B})$. Now, fix a Markov kernel $\mathcal{K}$ from Λ to $\mathcal{X}$ and let π be a probability distribution on $(\Lambda, \mathcal{A})$. Then

$$\eta_\pi(d\lambda, dx) := \pi(d\lambda)K_\lambda(dx) \tag{2.10}$$

is a probability distribution on the product space $(\Lambda \times \mathcal{X}, \mathcal{A} \otimes \mathcal{B})$. Let γ_π denote the image of η_π by the map $(\lambda, x) \mapsto x$, that is, $\gamma_\pi = x_*\eta_\pi$. Then, under mild conditions of regularity which will always be met in this paper (such as, Λ and $\mathcal{X}$ are metric separable spaces equipped with Borel fields), there exists a Markov kernel $\mathcal{J}$ from $\mathcal{X}$ to Λ (note again that $\mathcal{J}$ is a family $\{J_{x,\pi} \colon x \in \mathcal{X}\}$ and $J_{x,\pi}$ depends on π) such that

$$\eta_\pi(d\lambda, dx) = J_{x,\pi}(d\lambda)\gamma_\pi(dx). \tag{2.11}$$

In Bayesian statistical terminology, $\mathcal{X}$ represents the *sample space* of the data $x \in \mathcal{X}$; Λ, the *parameter space* of the *mixing parameter* λ; π, the *prior distribution* of λ; $\mathcal{K} = \{K_\lambda : \lambda \in \Lambda\}$, the family of sampling distributions; γ_π, the *marginal distribution* of x (also called the mixture distribution); and $J_{x,\pi}$, the *posterior distribution* (of λ given the data x).

Consider a family $\mathcal{F}$ of prior distributions π on $(\Lambda, \mathcal{A})$ and a Markov kernel $\mathcal{K}$ from Λ to $\mathcal{X}$. Then $\mathcal{F}$ is said to be conjugate for $\mathcal{K}$ if $J_{x,\pi} \in \mathcal{F}$, for every $\pi \in \mathcal{F}$ and γ_π-almost all x.

The family of priors $\mathcal{F}$ we consider throughout the sequel is defined by means of the following assumption.

ASSUMPTION 2.1. Let $(\Lambda, \mathcal{A})$ and $(\mathcal{X}, \mathcal{B})$ be, as above, the parameter space and sample space, respectively. Also, let $\overline{F}(u, \nu)$ be an FGEF generated by a measure ν on $(\Lambda, \mathcal{A})$ and a measurable map $u: \Lambda \to E$, and let $\mu := u_* \nu$. For a subset $D \subset D(\mu)$, closed or open in E^*, define a subfamily $\mathcal{F}$ of $\overline{F}(u, v)$ by

$$\mathcal{F} := \{P(\theta, u, v): \theta \in D\}. \tag{2.12}$$

The dual problem we study in this paper can now be stated as follows. Which Markov kernels $\mathcal{K}$ from Λ to $\mathcal{X}$ [or, equivalently, which families of sampling distributions $\mathcal{K} = \{K_\lambda : \lambda \in \Lambda\}$ defined on $(\mathcal{X}, \mathcal{B})$] are such that $\mathcal{F}$ in (2.12) is conjugate for $\mathcal{K}$? We provide the general structure of such $\mathcal{K}$'s in Theorem 3.2.

Famous examples of conjugate families are presented in Diaconis and Ylvisaker [(1979, Theorem 2.1]. They begin by considering an NEF $\mathcal{K}$ as a given family of sampling distributions and obtain a family $\mathcal{F}$ of conjugate priors for $\mathcal{K}$. Their result can be stated, in our terminology, as follows. Let $F(\rho)$ be a given regular NEF [i.e., $\Theta(\rho) = D(\rho)$] on some linear space V, and consider a Markov kernel $\mathcal{K}$ from $\Theta(\rho)$ to E defined by

$$K_\lambda(dx) := \exp\{\langle \lambda, x\rangle - k_\rho(\lambda)\}\rho(dx), \qquad \lambda \in \Theta(\rho). \tag{2.13}$$

[Hence, the family $\mathcal{K}$ of sampling distributions considered in Diaconis and Ylvisaker (1979) constitutes an NEF $F(\rho)$ so that $\Lambda = \Theta(\rho)$.] They showed that, for any $p > 0$ and $v \in M_{F(\rho)} := k'_\rho(\Theta(\rho))$ [i.e., v belongs to the mean domain of $F(\rho)$], there exists a real number $A(v,p)$ such that

$$\pi_{v,p}(d\lambda) := A(v,p)\exp\Big\{p\big(\langle \lambda, v\rangle - k_\rho(\lambda)\big)\Big\}\mathbf{1}_{\Theta(\rho)}(\lambda)\,d\lambda \tag{2.14}$$

is a probability distribution. Consequently, the family

$$\mathcal{F} := \{\pi_{v,p}(d\lambda): v \in M_{F(\rho)}, p > 0\} \tag{2.15}$$

is a family of conjugate priors for $\mathcal{K}$ [$= F(\rho)$] in (2.13). We shall call $\mathcal{F}$ in (2.15) the Diaconis–Ylvisaker family associated with the NEF $F(\rho)$. It should be noted, however, that this family was first discovered and treated by Barndorff-Nielsen [(1978), pages 131–132]. We now make some remarks concerning the Diaconis–Ylvisaker family.

REMARK 2.1.

(i) Note that the set of parameter values on which the Diaconis–Ylvisaker family is defined is

$$D := \{(vp, p) \in V \times \mathbb{R}: v \in M_{F(\rho)}, p > 0\}. \tag{2.16}$$

Accordingly, if $\nu(d\lambda) = d\lambda$ is the Lebesgue measure on $\Theta(\rho)$, u: $\Theta(\rho) \to E := V \times \mathbb{R}$, $\lambda, \mapsto (\lambda, -k_\rho(\lambda))$ and $\mu := u_*\nu$, then

$$D \text{ in (2.16) is contained in } D(\mu), \tag{2.17}$$

that is, the FGEF $F(u, \nu)$, generated by such u and ν, contains the Diaconis–Ylvisaker family as a subfamily. Indeed, Diaconis and Ylvisaker (1979) indicated, in terms of a specific example (see the end of their Section 2), that the set of all parameter values of the form (pv, p) for which $\pi_{v,p}$ in (2.14) is a probability distribution may contain, in addition to D in (2.16), pairs (vp, p) in which p is negative. This is exactly our statement (2.17), which will be proved in Proposition 5.5.

(ii) Of course, we tackle the Diaconis–Ylvisaker problem in reverse (the dual problem). Starting with $\mathcal{F}$ in (2.15) as our family of priors, solving for those sampling models for which $\mathcal{F}$ is conjugate, and following our procedure, we will end up with $F(\rho)$ [or $\mathcal{K}$ in (2.13)] as one of the solutions. Adopting the terminology of Barndorff-Nielsen (1978), note that the *order* of the NEF $F(\rho)$, the sampling model in (2.13), is 1 less than the *order* of $\mathcal{F}$ in (2.15) associated with it. (The *order* of an NEF is the dimension of the space on which it is concentrated. The *order* of a GEF is the *order* of the NEF with which it is associated.)

(iii) Diaconis and Ylvisaker assumed that $F(\rho)$ is regular. In Theorem 5.4, we restate, in our terminology, their Theorem 2.1, while relaxing their assumption by only requiring that $F(\rho)$ is steep. [Such a relaxation of the assumption has been made previously by Barndorff-Nielsen (1978).] In Proposition 5.5 we present some properties of the set D in (2.16). In Proposition 5.7 we consider the case where $V = \mathbb{R}$ and present general results on sampling models, concentrated on hyperplanes, for which $\mathcal{F}$ in (2.15) is conjugate. Proposition 5.7 is then applied in Section 6 to some familiar examples.

(iv) We henceforth reserve the Greek letter ρ to denote a measure generating an NEF $F(\rho)$ with which the Diaconis family $\mathcal{F}$ in (2.15) is associated.

3. Characterizations of sampling models $\mathcal{K}$ for which $\mathcal{F}$ is conjugate. We first present a proposition and definition which are required to state the main result of this section (Theorem 3.2). The introduction section helps to understand the need for such a proposition and definition and, thereby, the statement of the theorem.

PROPOSITION 3.1. *Let E be a real finite-dimensional linear space, let E^* be the dual of E, let D be a nonempty subset of E^* and let*

$$G(D) := \{h \in E^*: h + D \subset D\}. \tag{3.1}$$

Then the following hold: (i) $G(D)$ *is an additive semigroup of* E^*; (ii) $G(D) = G(h + D)$ *for all* $h \in E^*$; (iii) $E^*\backslash G(D) = (E^*\backslash D) - D$; (iv) $G(D) = \{0\}$ *if* D *is bounded*; (v) $G(D)$ *is closed if* D *is either open or closed*; (vi) *if* D *is an open convex cone, then* $G(D)$ *is the closure of* D.

PROOF. Statements (i) and (ii) are obvious. Statement (iii) is easily checked. Statements (iv) and (v) follow from (iii). To prove (vi), let $\overline{D}$ be the closure of D. Then $\overline{D} + D$ is open and $\overline{D} + D \subset D = \operatorname{int} \overline{D}$, and thus $\overline{D} \subset G(D)$. Conversely, from (iii), we have to prove that $E^*\backslash\overline{D} \subset (E^*\backslash D) - D$. Letting $h \in E^*\backslash D$, there exists an open set U containing 0 such that $(h + U) \cap D = \phi$. Since D is a cone, there exists some $u \in U \cap D$ such that $h + u \in E^*\backslash D$. Hence, $h \in E^*\backslash D$, and the other inclusion is proved. □

Let G be a closed additive semigroup of E^*, and let S be a closed subset of E. We define $\mathcal{H}(S, G)$ to be the set of all positive measures q on E^*, concentrated on G [i.e., $S(q) \subset G \subset E^*$], such that, for all $u \in S$,

$$\int_{E^*} \exp\{\langle h, u\rangle\} q(dh) = 1. \tag{3.2}$$

THEOREM 3.2. *Assume that Assumption* 2.1 *holds. Let* $\mathcal{K}$ *be a Markov kernel from* Λ *to* $\mathcal{X}$. *Then* $\mathcal{F}$, *as defined by* (2.12), *is conjugate for* $\mathcal{K}$ *iff there exist a measure* Q *on* $(\mathcal{X}, \mathcal{B})$ *and a measurable map* $h: \mathcal{X} \to G(D)$ *such that* $q := h_*Q \in \mathcal{H}(S(\mu), G(D))$ *and such that, for* ν*-almost all* λ,

$$K_\lambda(dx) = \exp\Big\{\big\langle h(x), u(\lambda)\big\rangle\Big\} Q(dx). \tag{3.3}$$

REMARK. We have commented previously that properties of an FGEF are actually properties of the associated FNEF. Here, $\mathcal{K} = \{K_\lambda : \lambda \in \Lambda\}$, as defined by (3.3), if it exists, has the form of an FGEF generated by h and Q and associated with the FNEF $F(q)$. This observation reduces the search for the set of all families of sampling distributions $\mathcal{K}$ satisfying (3.3) to the smaller set $\mathcal{H}(S(\mu), G(D))$ [given q, the set of possible pairs (h, Q) is rather large—as has been demonstrated in Example 1.1]. This is the rationale for our contention: the set of solutions to the dual problem is essentially parameterized by $\mathcal{H}(S(\mu), G(D))$.

PROOF OF THEOREM 3.2. ($\Leftarrow$.) Let $Z := \{u \in D(q): k_q(u) = 0\}$. Since $q \in \mathcal{H}(S(\mu), G(D))$, we have $E\backslash Z \subset E\backslash S(\mu)$. Hence, $\mu(E\backslash S(\mu)) = 0$ implies $\mu(E\backslash Z) = 0$ and, for ν-almost all λ, $\int_{E^*} \exp\{\langle h, u(\lambda)\rangle\} q(dh) = 1$. Hence K_λ, defined by (3.3), is a probability, ν-almost all λ. Choose $\theta_0 \in D$, and let $\pi := P(\theta_0, u, \nu)$ be the prior distribution on $(\Lambda, \mathcal{A})$ and let

$$\eta_\pi(d\lambda, dx) := \exp\Big\{\big\langle h(x), u(\lambda)\big\rangle + \big\langle \theta_0, u(\lambda)\big\rangle - k_\mu(\theta_0)\Big\} Q(dx)\nu(d\lambda)$$

be the probability distribution on the product space $(\Lambda \times \mathcal{X}, \mathcal{A} \otimes \mathcal{B})$. Then, since $h(x) \in G(D)$, the posterior distribution $\exp\{\langle h(x) + \theta_0, u(\lambda)\rangle - k_\mu(\theta_0 + h(x))\}\nu(d\lambda) \in \mathcal{F}$. This implies that $\theta_0 + h(x) \in D$.

($\Rightarrow$.) For $\theta \in D$, let $Q_\theta(dx)$ denote the image of $P(\theta, u, \nu)(d\lambda)K_\lambda(dx)$ by the map $(\lambda, x) \mapsto x$. Since $\mathcal{F}$ is conjugate for $\mathcal{K}$, for all $\theta \in D$, there exists $h(\theta, x) \in D$ such that $x \mapsto h(\theta, x)$ is $\mathcal{B}$-measurable and

$$\begin{aligned}&\exp\Big\{\big\langle h(\theta,x), u(\lambda)\big\rangle - k_\mu\big(h(\theta,x)\big)\Big\}\nu(d\lambda)Q_\theta(dx)\\ &\quad= \exp\Big\{\big\langle \theta, u(\lambda)\big\rangle - k_\mu(\theta)\Big\}\nu(d\lambda)K_\lambda(dx).\end{aligned}\tag{3.4}$$

Let $f(\theta, x) := \theta - h(\theta, x)$ and $g(\theta, x) := k_\mu(\theta) - k_\mu(h(\theta, x))$. Then (3.4) implies that, for ν-almost all λ,

$$K_\lambda(dx) = \exp\Big\{g(\theta,x) - \big\langle f(\theta,x), u(\lambda)\big\rangle\Big\}Q_\theta(dx).\tag{3.5}$$

Since the left-hand side of (3.5) does not depend on θ, one gets, by fixing $\theta = \theta_0$ in (3.5), that for ν-almost all λ,

$$\exp\Big\{\big\langle f(\theta,x) - f(\theta_0,x), u(\lambda)\big\rangle\Big\}Q_{\theta_0}(dx) = \exp\Big\{\big\langle g(\theta,x) - g(\theta_0,x)\big\rangle\Big\}Q_\theta(dx).\tag{3.6}$$

Since the right-hand side of (3.6) does not depend on λ, and, by assumption, $\mu = u_*\nu \in \mathcal{M}(E)$ is not concentrated on an affine hyperplane, we get that $f(\theta, x) - f(\theta_0, x) = 0$, Q_{θ_0}-almost all x. Let $h(x) := h(\theta_0, x) - \theta_0$, then $h(\theta, x) = \theta + h(x)$ and, from (3.5), for ν-almost all λ, we get (3.3) with $Q(dx) = \exp\{g(\theta_0, x)\}Q_{\theta_0}(dx)$. We still have to check that $q := h_*Q \in \mathcal{H}(S(\mu), G(D))$. Since K_λ is a probability, ν-a.e., we have

$$\int \exp\Big\{\big\langle h, u(\lambda)\big\rangle\Big\}q(dh) = 1, \qquad \nu\text{-a.e.}\tag{3.7}$$

Let $u \in S(\mu)$. Then by (3.7) there exists a sequence $(u_n)_{n=1}^{\infty}$, $u_n \in S(\mu)$, $n \in \mathbb{N}$, such that $u_n \to_{n\to\infty} u$ and $k_q(u_n) = 0$ for all $n \in \mathbb{N}$. From Fatou's lemma, $k_q(u) \le 0$ and $u \in D(q)$. Since the subset $\{(u, y): u \in D(q), y \ge k_q(u)\} \subset E^* \times \mathbb{R}$ is closed [see Barndorff-Nielsen (1978), Theorem 7.1], there exists $y_n \ge k_q(u_n) = 0$ such that $(u_n, y_n) \to_{n\to\infty} (u, k_q(u))$. Since $y_n \to_{n\to\infty} k_q(u)$, $k_q(u)$ cannot be negative, and hence $k_q(u) = 0$. Thus, we proved that $\int \exp\{\langle h, u\rangle\}q(dh) = 1$, for al $u \in S(\mu)$. Since $h(\theta, x) = \theta + h(x)$ and $h(\theta, x) \in D$ for all $\theta \in D$ and Q-almost all x, we get $h(x) \in G(D)$, Q-almost all x. This implies that $q(E^* \backslash G) = 0$. □

4. A general study of $\mathcal{H}(S,G)$. We have realized that $\mathcal{H}(S, G)$ is the basic object to describe all solutions of the dual problem. This section is devoted to some useful remarks concerning its structure. We assume throughout this section that G is a closed additive semigroup of E^* and that $S \subset E$ is not concentrated on some hyperplane of E, that is, int conv$(S) \neq \emptyset$. [Any such S can serve as the support of some measure $\mu \in \mathcal{M}(E)$.] The following proposition shows that we may assume, without loss of generality, that S is contained in the boundary of conv(S).

PROPOSITION 4.1. *If* int conv$(S) \cap S \neq \emptyset$, *then* $\mathcal{H}(S, G)$ *contains only the Dirac mass* δ_0.

PROOF. Assume there exists $y \in \operatorname{int}\operatorname{conv}(S) \cap S$ and let $\dim E = d$. Then, by the Caratheodory theorem [see Phelps (1966), page 10], there exist $s_j \in S$, $j = 0, \dots, d$, affinely independent [i.e., $(s_j - s_0)_{j=1}^d$ are linearly independent], and $\lambda_j > 0, j = 0, \dots, d$ satisfying $\sum_{j=0}^d \lambda_j = 1$, such that $y = \sum_{j=0}^d \lambda_j s_j$. Accordingly, if $q \in \mathcal{H}(S, G)$, then $0 = k_q(\sum_{j=0}^d \lambda_j s_j) = \sum_{j=0}^d \lambda_j k_q(s_j)$. Hence, we are in the situation where an equality holds in the Hölder's inequality. Thus, there exist $a_i \in \mathbb{R}$, $i = 1, \dots, d$, such that, for q-almost all h, $\langle h, s_0 \rangle = \langle h, s_i \rangle + a_i$, for $i = 1, \dots, d$. This, along with the fact that $(s_j - s_0)_{j=1}^d$ is a basis of E, implies that $a_1 = \cdots = a_d = 0$, and therefore $q = \delta_0$. □

The following result, a corollary to Proposition 4.1, is useful for showing that there are no interesting elements of $\mathcal{H}(S, G)$ in some linear subspaces. Such a corollary will be applied in Example 6.1 and the proof of Theorem 5.1.

COROLLARY 4.2. *Let E_1 be a linear subspace of E with* $\dim E_1 < \dim E$, *and assume that S is such that, for any projection p_1 of E onto E_1,* $\operatorname{int} p_1(S) \neq \emptyset$. *If $q \in \mathcal{H}(S, G)$ is concentrated on some linear subspace E_1^* of E^* with* $\dim E_1^* = \dim E_1$, *then $q = \delta_0$.*

PROOF. Let E_0^* be a supplementary space of E_1^* in E^*. Define, for $i = 0, 1$,

$$E_i := \{u \in E: \langle h, u \rangle = 0, \forall h \in E_{1-i}^*\},$$

and let p_1 be the projection of E onto E_1 parallel to E_0. Since $\langle h, u \rangle = \langle h, p_1(u) \rangle$, for all $(h, u) \in E_1^* \times E$, it follows that, for all $u \in S$, $\int_{E_1^*} \exp\{\langle h, p_1(u) \rangle\} q(dh) = 1$. Since $\operatorname{int} p_1(S) \neq \emptyset$, then, by applying Proposition 4.1 to E_1^*, we obtain $q = \delta_0$. □

In the next proposition we gather some elementary properties of $\mathcal{H}(S, G)$.

PROPOSITION 4.3. (i) *The set of measures $\mathcal{H}(S, G)$ is convex*; (ii) $\mathcal{H}(S, G) \subset \overline{\mathcal{M}}(E^*)$; *and* (iii) *$\mathcal{H}(S, G)$ is closed under convolution.*

PROOF. Part (i) is obvious. (ii) Let $q \in \mathcal{H}(S, G)$. Then, by the definition of $\mathcal{H}$, $S \subset D(q)$ and $\operatorname{int}\operatorname{conv}(S) \subset \Theta(q)$. By assumption, $\operatorname{int}\operatorname{conv}(S) \neq \emptyset$; hence $\Theta(q) \neq \emptyset$, and therefore, by (2.4), $q \in \overline{\mathcal{M}}(E^*)$. (iii) Recall that if $\mu_i \in \overline{\mathcal{M}}(E)$, $i = 1, 2$, such that $\Theta(\mu_1) \cap \Theta(\mu_2) \neq \emptyset$, then the convolution of μ_1 and μ_2 is the element $\mu_1 * \mu_2 \in \overline{\mathcal{M}}(E)$ defined by $k_{\mu_1 * \mu_2} := k_{\mu_1} + k_{\mu_2}$. Now, if, $q_i \in \mathcal{H}(S, G)$, for $i = 1, 2$, then $\Theta(q_i) \supset \operatorname{int}\operatorname{conv}(S) \neq \emptyset$. Therefore, $\Theta(q_1) \cap \Theta(q_2) \neq \emptyset$ and $q_1 * q_2$ is well defined. Moreover, since G is an additive semigroup and $k_{q_1 * q_2}(u) = 0$ for all $u \in S$, it follows that $q_1 * q_2(E^* \backslash G) = 0$, and hence $q_1 * q_2 \in \mathcal{H}(S, G)$. □

We remark also that if $0 \in S$, then $\mathcal{H}(S, G)$ is a set of probabilities. In such a case, $\mathcal{H}(S, G)$ can be equipped with the topology of convergence of laws. However, $\mathcal{H}(S, G)$ is not necessarily stochastically bounded, preventing $\mathcal{H}(S, G)$ from being compact. If $0 \notin S$, we can easily translate back so that S will contain 0.

We finally introduce a concept of equivalence between pairs (S, G) which saves some trivialities. This concept is demonstrated in Example 4.1 and applied in Examples 6.1–6.3. Let E and E' be two linear spaces with $\dim E = \dim E'$; let S and S' be closed subsets of E and E', not concentrated on affine hyperplanes; and let G and G' be additive closed semigroups of E^* and $(E')^*$, respectively. We shall say that (S, G) and (S', G') are *equivalent* if there exists a linear isomorphism φ: $E \to E'$ such that $\varphi(S) = S'$ and ${}^t\varphi(G') = G$, where ${}^t\varphi$ is the transposed map $(E')^* \to E^*$ of φ. In this case, if $q' \in \mathcal{H}(S', G')$, then ${}^t\varphi * q' \in \mathcal{H}(S, G)$, since, for all $u \in S$,

$$\int_G \exp\{\langle h, u\rangle\}({}^t\varphi * q')(dh) = \int_{G'} \exp\Big\{\big\langle {}^t\varphi(h'), u\big\rangle\Big\}q'(dh)$$
$$= \int_{G'} \exp\Big\{\big\langle h', \varphi(u)\big\rangle\Big\}q'(dh') = 1.$$

Clearly, $q' \mapsto {}^t\varphi * q'$ is an affine bijection between the two convex sets $\mathcal{H}(S', G')$ and $\mathcal{H}(S, G)$. If (S, G) and (S', G') are built from families $\mathcal{F}$ and $\mathcal{F}'$ of type (2.12), we shall say (with some abuse) that $\mathcal{F}$ and $\mathcal{F}'$ are equivalent, and the search for sampling models $\mathcal{K}$ for which $\mathcal{F}$ or $\mathcal{F}'$ is conjugate is intrinsically the same problem. An excellent example of such an equivalence is provided by two types of Dirichlet families described in the following example.

EXAMPLE 4.1 (Dirichlet families of first and second kind). Denoting $B(a_0, \dots, a_d) := \prod_{i=0}^d \Gamma(a_i)/\Gamma(\sum_{i=0}^d a_i), a_i > 0$, $i = 0, \dots, d$, and introducing the simplex

$$T_d := \left\{\lambda = (\lambda_1, \dots, \lambda_d) \in (0,1)^d \colon \sum_{i=1}^d \lambda_i < 1\right\},$$

we consider two kinds of Dirichlet families:

$$\mathcal{F} = \left\{\left(1 - \sum_{j=1}^d \lambda_j\right)^{a_0 - 1}\left(\prod_{j=1}^d \lambda_j^{a_j - 1}\right)\mathbf{1}_{T_d}(\lambda)\prod_{j=1}^d \frac{d\lambda_j}{B(a_0, \dots, a_d)}\colon \right.$$
$$\left. a_i > 0,\ i = 0, \dots, d\right\},$$

$$\mathcal{F}' = \left\{\left(1 + \sum_{j=1}^d \lambda'_j\right)^{-a'_0}\left(\prod_{j=1}^d \lambda_j^{\prime a'_j - 1}\right)\mathbf{1}_{(0,\infty)^d}(\lambda')\frac{\left(\prod_{j=1}^d d\lambda'_j\right)}{B\left(a'_0 - \sum_{i=1}^d a'_i, a'_1, \dots, a'_d\right)}\colon \right.$$
$$\left. a'_j > 0,\ j = 1, \dots, d,\ \sum_{j=1}^d a'_j < a'_0\right\}.$$

Here, $E = E' = \mathbb{R}^{d+1}$,

$$G = [0,\infty)^d, \qquad G' = \left\{(a'_0, \dots, a'_d)\colon a'_j \ge 0,\ j = 1, \dots, d,\ \sum_{j=1}^d a'_j < a'_0\right\};$$

S is the hypersurface in $\mathbb{R}^{d+1}$ which is the image of T_d by the map $(\lambda_1,\dots,\lambda_d)\mapsto(\log(1-\sum_{i=1}^d\lambda_i), \log\lambda_1,\dots,\log\lambda_d)$; S' is the image of $(0,\infty)^d$ by the map $(\lambda'_1,\dots,\lambda'_d)\mapsto(-\log(1+\sum_{i=1}^d\lambda'_i), \log\lambda'_1,\dots,\log\lambda'_d)$; and $\mathcal{H}(S,G)$ and $\mathcal{H}(S',G')$ are equivalent through the linear map φ: $\mathbb{R}^{d+1}\to\mathbb{R}^{d+1}$, $(x_0,\dots,x_d)\mapsto(x'_0,\dots,x'_d)$, defined by $x'_0=x_0, x'_j=x_j-x_0, j=1,\dots,d$, with a transposed map ${}^t\varphi$: $(a'_0,\dots,a'_d)\mapsto(a_0,\dots,a_d)$, defined by $a_0=a'_0-\sum_{j=1}^d a'_j$, $a_j=a'_j$, $j=1,\dots,d$.

It should be noted that $\mathcal{F}$ and $\mathcal{F}'$ above have a close relationship with two Diaconis–Ylvisaker families [see (2.15)], which are associated with the negative binomial NEF (in $\mathbb{R}^d$) and multinomial NEF as follows. Consider in $\mathbb{R}^d$ the canonical basis $e_1,\dots,e_d$, with $e_j=(0,\dots,0,1,0,\dots,0)$, where 1 is in the j-th position. Consider also the measure $\alpha=\delta_{e_1}+\cdots+\delta_{e_d}$, the two measures in $\mathbb{R}^d$: $\rho=\sum_{n=0}^\infty\alpha^{*n}$ and $\rho'=\delta_0+\alpha$, and the two NEF's $F(\rho)$ and $F(\rho')$. For such NEF's we have $\Theta(\rho)=\{\theta\in\mathbb{R}^d: s(\theta):=\sum_{i=0}^d\exp(\theta_i)<1\}$, $\Theta(\rho')=\mathbb{R}^d$, $k_\rho(\theta)=-\log(1-s(\theta))$ and $k_{\rho'}(\theta)=\log(1+s(\theta))$. The Diaconis–Ylvisaker families associated with the NEF's $F(\rho)$ and $F(\rho')$, are, respectively, given by

$$\mathcal{F}_1=\left\{A(v,p)\exp\left\{p\Big(\langle\theta,v\rangle+\log\big(1-s(\theta)\big)\Big)\right\}\mathbf{1}_{\Theta(\rho)}(\theta)\prod_{i=1}^d d\theta_i\colon v\in(0,\infty)^d, p>0\right\},$$

$$\mathcal{F}'_1=\left\{A'(v,p)\exp\left\{p\Big(\langle\theta,v\rangle-\log\big(1+s(\theta)\big)\Big)\right\}\mathbf{1}_{\Theta(\rho')}(\theta)\prod_{i=1}^d d\theta_i\colon v\in T_d, p>0\right\}.$$

[Here, $A(v,p)$ and $A'(v,p)$ are suitable normalizing constants easily expressed in terms of the gamma function.] The image of $\mathcal{F}'_1$ by exp: $(\theta_1,\dots,\theta_d)\mapsto(\lambda_1,\dots,\lambda_d):=(e^{\theta_1},\dots,e^{\theta_d})$ is exactly $\mathcal{F}'$. The image of $\mathcal{F}_1$ by exp is strictly smaller than $\mathcal{F}$, but the D of $\exp\mathcal{F}_1$ is only a translate of the D of $\mathcal{F}$. Thus, from Proposition 3.1(ii), $\mathcal{F}$ and $\mathcal{F}_1$ share the same G and their $\mathcal{H}(S,G)$ coincide. To recapitulate, $\mathcal{F}$ and $\mathcal{F}_1$, $\mathcal{F}'$ and $\mathcal{F}'_1$ are conjugate for the same families and the four are equivalent.

The natural thing now would be to give a complete description of $\mathcal{H}(S,G)$ by looking, for instance, at extreme points of $\mathcal{H}(S,G)$ and hoping to represent any element as a suitable mixing of these extreme points. However, except for the trivial case of Example 4.3, we are not able to catch all extreme points of $\mathcal{H}(S,G)$.

We conclude this section with two rather academic examples. The more interesting ones result from Diaconis–Ylvisaker families and will be further considered in Section 6.

EXAMPLE 4.2 ($\mathcal{F}$ is the Fisher–von Mises family). Let E be an Euclidean space with $\dim E=d\geq 1$; let S be the unit sphere; let ν be the uniform measure on S; and let $\mathcal{F}=F(\nu)$, the familiar NEF called the Fisher–von Mises family. Here, $\Lambda=E$, $u(\lambda)=\lambda$, $D(\nu)=E^*=D=G$; E^* is identified with E, since E is Euclidean; and $\mathcal{H}(S,E)$ is the set of positive measures q on E such that

$$\int_E\exp\{\langle h,u\rangle\}q(dh)=1,\qquad \|u\|=1. \tag{4.1}$$

It is easy to exhibit elements of $\mathcal{H}(S,G)$. If $r \geq 0$, define

$$A_0(r) := \int_S \exp\{r\langle h,u\rangle\}\nu(dh), \qquad \|u\| = 1. \tag{4.2}$$

Clearly, $A_0(r)$ does not depend on u, since ν is invariant under rotations of E. Define ν_r to be the image of ν by the dilation $h \mapsto rh$, and denote $q_r := (A_0(r))^{-1}\nu_r$. Then, clearly, $q_r \in \mathcal{H}(S,G)$. In fact, q_r is, for any $r \geq 0$, an extreme point of $\mathcal{H}(S,G)$. For brevity, we prove the latter statement for $r = 1$. Suppose that there exist q and, and $q' \in \mathcal{H}(S,G)$ and $\lambda \in (0,1)$ such that $q_1 := \lambda q + (1-\lambda)q'$. Then q and q' are concentrated on S. We now use the following proposition to show that q_1 is an extreme point. (We have not found a clear proof of such a proposition and we sketch one in the Appendix.)

PROPOSITION 4.4. *Let ℓ be a signed measure on the unit sphere S of a Euclidian space such that $\int_S \exp\{\langle h,u\rangle\}\ell(dh) = 0$, for all $u \in S$. Then $\ell = 0$.*

EXAMPLE 4.2 (Continued). For the case $\ell = q_1 - q$, this proposition yields $\int_S \exp\{\langle h,u\rangle\}\ell(dh) = 1 - 1 = 0$, for all $u \in S$, and $q_1 = q = q'$. Hence, q_1 is extreme. Taking mixtures of $\{q_r: r \geq 0\}$ provides elements of $\mathcal{H}(S,G)$ which are invariant under rotation [Eaton (1981) is a good reference for this classical fact]. However, it is probably not true that all elements of $\mathcal{H}(S,G)$ are invariant under rotation, although we have not worked out an explicit example for $d \geq 2$. For $d = 1$, note that $\int_{\mathbb{R}} \exp\{h\}q(dh) = \int_{\mathbb{R}} \exp\{-h\}q(dh) = 1$ does not imply that q is symmetric. For example, $q = \frac{16}{35}\delta_{-\log 2} + \frac{9}{35}\delta_{\log 3}$ is not symmetric. Hence, the set of extreme points of $\mathcal{H}(S,G)$, defined by (4.1), is probably not exhausted by $\{q_r: r \geq 0\}$.

EXAMPLE 4.3 (S is concentrated on the vertices of a simplex). Suppose that $S = \{e_0, e_1, \dots, e_d\}$, where $d = \dim E$ and $(e_j - e_0)_{j=1}^d$ is a basis of E. Without loss of generality we assume that $e_0 = 0$ and, for simplicity, we take $D = G = E^*$. A family $\mathcal{F}$, of the form (2.12), which corresponds to the latter situation stems from a multinomial trial with $d+1$ cells $0, 1, \dots, d$ and corresponding positive probabilities $p_0, \dots, p_d$, satisfying $\sum_{j=0}^d p_j = 1$. Here, $\Lambda = \{0, \dots, d\}$; ν can be taken as the uniform measure on Λ; $u: \Lambda \to E$ is defined by $u(j) = e_j$; and $\mathcal{F}$ can be written as

$$\mathcal{F} = \left\{ \sum_{j=0}^{d} \delta_j \exp\left[\langle \theta, e_j\rangle - k(\theta)\right]: \theta \in E^* \right\}, \tag{4.3}$$

or, equivalently, as $\mathcal{F} = \{\sum_{j=0}^d p_j\delta_j: \sum_{j=0}^d p_j = 1, p_j > 0, j = 0, \dots, d\}$. [Note that the assumption that $e_0 = 0$ implies that all elements of $\mathcal{H}(S,G)$ are probabilities]. Let $q \in \mathcal{H}(S,G)$; let $\tilde{q} := \varphi * q$ be the image of q by the map $\varphi: E^* \to \mathbb{R}^d$; $h \mapsto (\exp\{\langle h, e_i\rangle\})_{i=1}^d$; and let $\varphi(\mathcal{H}(S,G))$ denote the set of probabilities $\tilde{q}$ on $(0,\infty)^d$ which are the solutions of the moment problem: $\int_{(0,\infty)^d} x_i\,\tilde{q}(dx_1, \dots, zdx_d) = 1$, $i = 1, 2, \dots, d$. Since $q \mapsto \varphi * q$ is an affine bijection, the searches for extreme points of $\mathcal{H}(S,G)$ and $\varphi(\mathcal{H}(S,G))$ are equivalent. This statement enables us to find all extreme points of $\mathcal{H}(S,G)$, and this is the only example in the present paper in which we are successful. To achieve this goal, we consider, for $j = 0, 1, \dots, d$, the family $\mathcal{C}_j$ of subsets $T = \{t_0, t_1, \dots, t_j\} \subset (0,\infty)^d$ such that (i) $t_0, t_1, \dots, t_j$ are affinely independent and (ii) T contains the point $(1,1,\dots,1) \in \mathbb{R}^d$ in its relative open convex hull; that is, for any $j = 0, 1, \dots, d$, there exists a unique sequence $(\lambda_0, \lambda_1, \dots, \lambda_j)$, $\lambda_i \in (0,1]$, $i = 0, \dots, j$ such that $\sum_{i=0}^{j} \lambda_i = 1$ and $\sum_{i=0}^{j} \lambda_i t_i = (1,1,\dots,1)$. For $T \in \mathcal{C}_j$, we let $\tilde{q}_T := \sum_{i=0}^{j} \lambda_i \delta_{t_i}$ (note that, for $j = 0$, $\tilde{q}_T = \delta_{(1,\dots,1)}$. Clearly, $\tilde{q}_T \in \varphi(\mathcal{H}(S,G))$ is an extreme point. Moreover, the set of extreme points of $\varphi(\mathcal{H}(S,G))$ [and thereby of $\mathcal{H}(S,G)$] is exhausted by the set of measures of the form $\tilde{q}_T$. The latter statement follows by the following proposition.

PROPOSITION 4.5.

(i) *For every* $\tilde{q} \in \varphi(\mathcal{H}(S,G))$, *there exist measures* α_j *on* $\mathcal{C}_j$ *such that*

$$\tilde{q} = \sum_{j=0}^{d} \int_{\mathcal{C}_j} \tilde{q}_T \alpha_j(dT). \tag{4.4}$$

(ii) *The set of sampling models for which the multinomial family* $\mathcal{F}$, *defined by* (4.3), *is conjugate is the set of NEF's* $F(\tilde{q})$, *where* $\tilde{q}$ *is defined by* (4.4).

Part (ii) of this proposition follows from part (i). We are not giving a proof of part (i) as it is rather tedious (induction on d) and uninteresting. We shall just briefly comment on it for $d = 1$. If $T = \{a, b\} \in \mathcal{C}_1$, with $a < 1 < b$ (say), we have $\tilde{q}_T = [(b-1)/(b-a)]\delta_a + [(1-a)/(b-a)]\delta_b$, and Proposition 4.5 states that if $\tilde{q}$ is a probability on $\mathbb{R}^+$ such that $\int_0^\infty x\tilde{q}(dx) = 1$, there exists a measure α_1 on $(0,1) \times (1,\infty)$ such that

$$\tilde{q} = \tilde{q}(\{1\})\delta_1 + \int_{(0,1)\times(1,\infty)} \left(\frac{(b-1)}{(b-a)}\delta_a + \frac{(1-a)}{(b-a)}\delta_b\right)\alpha_1(da, db).$$

Here, α_1 is by no means unique. Let

$$r := \int_{(1,\infty)} (x-1)\tilde{q}(dx) = \int_{(0,1)} (1-x)\tilde{q}(dx) > 0;$$

then a possible choice for α_1 is

$$\alpha_1(da, db) = r^{-1}(b-a)\mathbf{1}_{(0,1)}(a)\mathbf{1}_{(1,\infty)}(b)\tilde{q}(da)\tilde{q}(db).$$

5. Elements of $\mathcal{H}(S,G)$ concentrated on affine manifolds. As we have seen in Example 4.2, characterizing all extreme points of $\mathcal{H}(S,G)$ may be too hard a task. In this section, we will show that $\mathcal{H}(S,G)$ can contain a remarkable set of measures q which are supported on small affine manifolds contained in

E^*. Such q generate exponential families (the sampling models) admitting $\mathcal{F}$ in (2.12) as a conjugate family. These q will also provide extreme points of $\mathcal{H}(S, G)$.

More specifically, we will examine the case where S is an analytic manifold on E with $\dim S < \dim E$. Such a case is frequently met in the literature. The case $\dim S = \dim E - 1$ occurs when $\mathcal{F}$ is the Diaconis–Ylvisaker family in (2.15) (see Remark 2.1) associated with an NEF $F(\rho)$ on a linear space V with $\dim V = \dim S$. The case $\dim S = 1$ (i.e., when S is a curve or a skew curve if $\dim E > 2$) occurs also. For instance, consider the "hypergeometric" family. For fixed $z \in (-\infty, 1)$, define

$$\mathcal{F} := \left\{ \frac{\lambda^{b-1}(1-\lambda)^{c-b-1}(1-\lambda z)^{-a}}{\mathrm{B}(b, c-b)F(a, b; c; z)} \mathbf{1}_{(0,1)}(\lambda)\, d\lambda \colon a > 0,\ b > 0 \right\},$$

where B is the beta Eulerian function and F is the hypergeometric function. Here, $u(\lambda) = (\log \lambda, \log(1-\lambda), \log(1-\lambda z))$ and S is the image of $\wedge = (0, 1)$ under u. A similar example is provided by the family of generalized inverse Gaussian distributions [see (5.7)].

For such an S, we shall examine in Theorem 5.1 elements $q \in \mathcal{H}(S, G)$ which are concentrated on affine manifolds in E^* with dimensions equaling $\dim S$. Specializations of Theorem 5.1 for the case $\dim S = 1$ are made in Proposition 5.2 and Theorem 5.3, and for the case $\dim S = \dim E - 1$ in Theorem 5.4. Throughout this section we shall assume that S satisfies the following assumption.

ASSUMPTION 5.1. (i) The manifold S is a closed subset of E. Let $E_0 \oplus E_1$ be any direct decomposition of E with $\dim E_1 = \dim S$, and let p_1 be a projection of E onto E_1 parallel to E_0. Then $\operatorname{int} p_1(S) \neq \emptyset$. (ii) The manifold S has the induced topology from E; that is, for any $s \in S$ there exist an open set $U \subset S$ containing s, an open set $W \subset \mathbb{R}^{\dim S}$ and an analytic injection $f\colon W \to E$ such that $f(W) = U$ and the rank of the differential f' is $\dim S$ for all $w \in W$.

Part (i) of this assumption implies that S is not contained in some affine hyperplane of E and it is not concentrated, for instance, on the surface of a cylinder. Part (ii) corresponds to the definition of a regular analytic manifold [see Lelong-Ferrand (1963), page 108] and avoids certain pathologies (like double points for curves).

Before presenting Theorem 5.1, we impose two additional assumptions which are needed there. One concerns the properties of the linear spaces considered. The second relates to the properties of S and the family of priors being used.

ASSUMPTION 5.2.

(i) The space E is a real linear space with $\dim E = d$, and E^* is its dual.

(ii) The space E_1^* is a fixed linear subspace of E^*, E_0^* is a supplementary subspace of E_1^*, (i.e., $E^* = E_0^* \oplus E_1^*$), so that if $e_0^* \in E_0^*$, then $e_0^* \in E_1^*$ is an affine manifold of E^*.

(iii) The spaces E_0 and E_1 are the respective orthogonal spaces of E_1^* and E_0^* in E, defined by $E_{1-i} = \{u \in E\colon \langle h, u\rangle = 0,\ \forall\, h \in E_i^*\}, i = 0, 1$.

(iv) For $i = 0, 1$, p_i is the projection of E onto E_i parallel to E_{1-i}.

REMARK. Note that the dual of E_1^* is identified with E_1 as follows: if $(h_1, u_1) \in E_1^* \times E_1, \langle h_1, u_1 \rangle_1$ is defined as $\langle h, u \rangle$, for $u \in E$ such that $p_1(u) = u_1$. This definition does not depend on the particular chosen u, since $p_1(u) = p_1(u') = u_1$ implies that $u - u' \in E_0$.

ASSUMPTION 5.3.

(i) The manifold S is an analytic manifold contained in E and satisfies Assumption 5.1.

(ii) The map $u: \wedge \to S, \lambda \mapsto u(\lambda)$ is an onto map defined by $u_i(\lambda) = p_i(u(\lambda)), i = 0, 1, \nu$ is a measure on $(\wedge, \mathcal{A})$ such that $\mu := u_* \nu \in \mathcal{M}(E), D \subset D(\mu), G := G(D)$ is a closed additive semigroup in E^* and $\mathcal{F}$ is a family of priors on $(\wedge, \mathcal{A})$, a subfamily of $\overline{F}(u, \nu)$, defined in Assumption 2.1.

(iii) $\dim S = \dim E_1^*$, where E_1^* is defined in Assumption 5.2.

We now state our next theorem, whose proof is relegated to the Appendix.

THEOREM 5.1. *Let Assumptions* 5.2 *and* 5.3 *hold.*

(a) *Assume that* $q \in \mathcal{H}(S, G)$ *is concentrated on* $e_0^* + E_1^*$, *where* $e_0^* \in E_0 \backslash \{0\}$ *is fixed. Let* q_1 *denote the image in* E_1^* *of* q *by the translation map* $E^* \to E^*, h \mapsto h - e_0^*$.

(i) *Then* q *is unique and is an extreme point of* $\mathcal{H}(S, G)$. *Furthermore,* $q_1 \in \mathcal{M}(E_1^*)$ *is determined by the relationship*

$$k_{q_1}\big(u_1(\lambda)\big) = -\big\langle e_0^*, u(\lambda)\big\rangle \quad \text{for all } \lambda \in \wedge. \tag{5.1}$$

(ii) *Define*

$$K_\lambda(dh) := \exp\Big\{\big\langle h, u_1(\lambda)\big\rangle_1 + \big\langle e_0^*, u_0(\lambda)\big\rangle\Big\} q_1(dh). \tag{5.2}$$

Then $\mathcal{F}$ *is conjugate for* $\mathcal{K} = \{K_\lambda: \lambda \in \wedge\}$.

(b) *Conversely, let* V *be a real linear space;* $\delta \in \mathcal{M}(V)$; *let* $F(\delta) := \{P(\theta, \delta): \theta \in \Theta(\delta)\}$ *be the NEF generated by* δ; *and let* $\theta: \wedge \to \Theta(\delta), \lambda \mapsto \theta(\lambda)$, *be a map such that* $\theta_* \nu$ *is not concentrated on an affine hyperplane of* V^*. *If* $\mathcal{F}$ *is conjugate for* $\{P(\theta(\lambda), \delta): \lambda \in \wedge\}$, *then* $\dim V = \dim S$ *and there exist an injective linear map* $\varphi: V \to E, e_0^* \in E^*$ *and a measure* q *concentrated on* $e_0^* + E_1^* = e_0^* + \varphi(V)$ *such that* $\varphi_* P(\theta(\lambda), \delta) = K_\lambda$, *for all* $\lambda \in \wedge$, *where* K_λ *is defined by* (5.2).

In Proposition 5.2 we specialize Theorem 5.1 to the most important case where S is a curve; that is, $\dim S = 1$ and $\wedge$ is an interval (see Example 1.2 and the examples presented in Section 6). Recall that in this case $\dim E_1^* = 1$. Our best tool to study the existence of a measure q_1 satisfying (5.1) and then to compute it is the variance function of the NEF $F(q_1)$ (see relevant references in the paragraph preceding Example 1.2). Theoretically, this tool is also available for the case where $\dim S > 1$, but the study of variance functions of NEF's in the multivariate case is presently in its infancy. We recall here that the variance function of an NEF $F(\mu)$ in $\mathbb{R}$ is the positive function on the mean domain $M_{F(\mu)} := k'_\mu(\Theta(\mu))$ of $F(\mu)$, defined by

$$V_{F(\mu)}\big(k'_\mu(\theta)\big) := \int_{\mathbb{R}} \big(x - k'_\mu(\theta)\big)^2 P(\theta, \mu)(dx) = k''_\mu(\theta). \tag{5.3}$$

To specialize Theorem 5.1 for the case $\dim S = 1$, we continue to assume that Assumptions 5.1, 5.2 and 5.3 hold and that $\dim E_1^* = 1$. Here, if e_1^* is a basis of $E_1^* = \{te_1^*: t \in \mathbb{R}\}$, we define

$$\theta_i(\lambda) := \langle e_i^*, u_i(\lambda)\rangle, \qquad i = 0, 1. \tag{5.4}$$

In the next proposition we use the notation of Theorem 5.1. In particular, we let $q \in \mathcal{H}(S, G)$ be concentrated on $e_0^* + E_1^*$ and let q_1 be the image of q by the translation map $E^* \to E^*, h \mapsto h - e_0^*$.

PROPOSITION 5.2. *Let Assumptions* 5.2 *and* 5.3 *hold*; *let* $\dim E_1^* = 1$; *and let* $\wedge \subset \mathbb{R}$ *be an open interval such that* θ_i, $i = 0, 1$, *defined by* (5.4), *are analytic on* $\wedge$.

(i) *Let* q_0 *be the image of* q_1 *by* $E_1^* \to \mathbb{R}, te_1^* \mapsto t_0$. *Then* $k_{q_0}(\theta_1(\lambda)) = -\theta_0(\lambda)$, $\forall\, \lambda \in \wedge$.

(ii) *The map* $\lambda \mapsto -\theta'_0(\lambda)/\theta'_1(\lambda) := m(\lambda)$ *is analytic on* $\widetilde{\wedge} = \{\lambda \in \wedge: \theta_1(\lambda) \in \Theta(q_0)\}$.

(iii) *The variance function of* $F(q_0)$ *is given on* $\widetilde{M} := m(\widetilde{\wedge})$ *by*

$$V_{F_{q_0}}\big(m(\lambda)\big) = \big[\theta'_0(\lambda)\theta''_1(\lambda) - \theta'_1(\lambda)\theta''_0(\lambda)\big] / \big(\theta'_1(\lambda)\big)^3 = k''_{q_0}\big(\theta_1(\lambda)\big). \tag{5.5}$$

PROOF. (i) This is a reformulation of (5.1).

(ii) It follows from (i) that $\theta_1(\wedge) \subset D(q_0)$. Since $D(q_0)\backslash\Theta(q_0)$ has at most two points, $\wedge$ is open. Taking derivatives of both sides of the relation in (i) and observing that $m(\lambda) = k'_{q_0}(\theta_1(\lambda))$ is analytic as a composition of two analytic functions, proves (ii).

(iii) This is obtained by using (A.2) in the Appendix and differentiating both sides of the relation $k'_{q_0}(\theta_1(\lambda)) = -\theta'_0(\lambda)/\theta'_1(\lambda)$ with respect to λ. □

We now apply Proposition 5.2 to a frequently occurring situation. We assume that the curve S in E, with $\dim E = d$, is parametrized by the open interval $\wedge$ and the map $u: \wedge \to E$ such that there exists a continuous function $T: \wedge \to \mathbb{R}^+$ for which $T(\lambda)u'(\lambda)$ is a polynomial in λ with degree less than d; that is, for any basis $(f_1, \dots, f_d)$ of E, there exist real polynomials $P_1, \dots, P_d$ with degree less than d such that $T(\lambda)u'(\lambda) = \sum_{j=1}^d P_j(\lambda) f_j$. Note that, since S is not concentrated in an affine hyperplane, the polynomials $P_1, \dots, P_d$ must be independent. Therefore, it is always possible to choose a basis $(f_1, \dots, f_d)$ such that $P_j(\lambda) = \lambda^{d-j}$, $j = 1, \dots, d$. For instance, if $E = \mathbb{R}^2$, consider the family of beta distributions of the first kind:

$$\Big\{\big[\mathrm{B}(a, b)\big]^{-1}\lambda^{a-1}(1-\lambda)^{b-1}\mathbf{1}_{(0,1)}(\lambda)\, d\lambda: a > 0, b > 0\Big\}. \tag{5.6}$$

Here, $\wedge = (0, 1)$ and $u(\lambda) = (\log\lambda, \log(1-\lambda))$. Since $u'(\lambda) = (\lambda^{-1}, -(1-\lambda)^{-1})$, we are in the above situation with $T(\lambda) = \lambda(1-\lambda)$. Another example, for $E = \mathbb{R}^3$, stems from the family of generalized inverse Gaussian distributions:

$$\left\{A(a,b,c)\lambda^{b-1}\exp\left\{a\lambda - c\lambda^{-1}\right\}\mathbf{1}_{\mathbb{R}^+}(\lambda)\,d\lambda\colon a < 0, c < 0\right\}. \tag{5.7}$$

Here, $\wedge = \mathbb{R}^+, u(\lambda) = (\lambda, \log\lambda, -\lambda^{-1}), u'(\lambda) = (1, \lambda^{-1}, \lambda^{-2})$ and $T(\lambda) = \lambda^2$. The hypergeometric family, introduced in the beginning of this section, provides another example with $T(\lambda) = \lambda(1-\lambda)(1-\lambda z)$. We have also found in Morlat [(1956), Sections 2.4 and 2.6] two more examples of such a type: one corresponds to $T(\lambda) = \lambda$ and is due to Etienne Halphen; the other corresponds to $T(\lambda) = \lambda^3$ and is due to M. Larcher.

Suppose we are in the realm of Proposition 5.2. Then $V_{F(q_0)}(m(\lambda))$, defined by (5.5), is expressed there in terms of λ. For practical considerations we have to express it in terms of m, the mean function of $F(q_0)$, and then check, being helped by the dictionary of variance functions of NEF's on $\mathbb{R}$, whether the resulting pair $V_{F(q_0)}(m), \widetilde{M})$ is a variance function of some NEF. If it is, then q_0 is a legitimate measure. For expressing $V_{F(q_0)}$ in terms of m, we consider the following technique, which is applicable to various examples (such as those in Section 6). Assume that $T(\lambda)u'(\lambda)$ is a polynomial in λ of degree less than $\dim E$. If e_0^* and e_1^* are independent, as in Proposition 5.2, we define two real polynomials $P_i(\lambda) := \langle e_i^*, T(\lambda)u'(\lambda)\rangle$, $i = 0, 1$. Then $\theta_i'(\lambda)$, $i = 0, 1$, $m(\lambda)$ and $V_{F(q_0)}(m(\lambda))$ of Proposition 5.2 can be expressed as $\theta_i'(\lambda) = P_i(\lambda)/T(\lambda), i = 0, 1$, $m(\lambda) = -P_0(\lambda)/P_1(\lambda)$ and

$$V_{F(q_0)}\big(m(\lambda)\big) = \big[P_1'(\lambda)P_0(\lambda) - P_0'(\lambda)P_1(\lambda)\big]T(\lambda)/\big(P_1(\lambda)\big)^3. \tag{5.8}$$

Hence, for computing $V_{F(q_0)}(m)$, we solve the equation $P_0(\lambda) + mP_1(\lambda) = 0$ for λ in terms of m and substitute the solution in the right-hand side of (5.8). Such a procedure works nicely for the two cases described in the next theorem.

THEOREM 5.3. *Consider the above notation and assumptions.*

(i) *If* $\dim E = 2$ *and* $T(\lambda)$ *is a polynomial of degree less than or equal to* 3, *then* $V_{F(q_0)}(m)$ *is a polynomial with degree less than or equal to* 3.

(ii) *If* $\dim E = 3$ *and* $T(\lambda)$ *is a polynomial with degree less than or equal to* 4, *then*

$$V_{F(q_0)}(m) = P(m)\Delta(m) + Q(m)\sqrt{\Delta(m)}, \tag{5.9}$$

where P, Q *and* Δ *are polynomials in* m *with degrees less than or equal to* 1, 2 *and* 2, *respectively.*

PROOF. (i) Taking $E = \mathbb{R}^2$ and letting $P_i(\lambda) = a_i + b_i\lambda$, $i = 0, 1$, then $a_0b_1 - a_1b_0 \neq 0$, since e_0^* and e_1^* are independent. The solution for λ in terms of m is $\lambda = -(a_0 + a_1m)/(b_0 + b_1m)$, and hence, by (5.8),

$$V_{F(q_0)}(m) = -\frac{(b_0 + b_1m)^3}{(a_0b_1 - a_1b_0)^2}T\left(\frac{-(a_0 + a_1m)}{b_0 + b_1m}\right).$$

Since $T(\lambda)$ is a polynomial of degree less than or equal to 3, the desired result follows.

(ii) This part is proved in the Appendix. $\square$

COMMENTS.

(i) Good examples satisfying the premises of this theorem are given in (5.6) and (5.7). Applications of this theorem appear in Examples 6.1–6.3.

(ii) In part (i) of the theorem, the resulting $V_{F(q_0)}(m)$ is a polynomial with degree less than or equal to 3. Hence, one may use the classification of cubic variance functions appearing in Mora (1986) or Letac and Mora (1990) to check whether $(V_{F(q_0)}(m), \widetilde{M})$ is a variance function. In part (ii) of the theorem, $V_{F(q_0)}(m)$ has the form (5.9). A list of functions of such a form which are variance functions and which are not can be found in Bar-Lev, Bshouty and Enis (1991), Letac (1992) and Seshadri (1991).

The remainder of this section is devoted to applications of Theorem 5.1 to the case where $\mathcal{F}$ is the Diaconis–Ylvisaker family [defined in (2.14) and (2.15)] associated with an NEF $F(\rho)$, $\rho \in \mathcal{M}(V)$. (Remark 2.1 might be helpful in this respect.) Theorem 5.4 restates, in our terminology, Theorem 2.1 of Diaconis and Ylvisaker (1979). Proposition 5.5. presents some properties of the parameter set D, in (2.16), of $\mathcal{F}$ and its respective closed additive semigroup G. In Proposition 5.6 we apply Theorem 5.1 to find elements of the corresponding set $\mathcal{H}(S, G)$ which are concentrated on affine hyperplanes (subsets of D). The latter proposition is specialized in Proposition 5.7 to the case $V = \mathbb{R}$. In particular, we present there relations existing among the variance function of the NEF $F(\rho)$ and the variance functions of the NEF's generated by elements of $\mathcal{H}(S, G)$. A practical procedure for applying Proposition 5.7 is presented at the end of this section. Section 6 is devoted then to applications of such a procedure to familiar examples.

We first make an assumption, which we assume to hold throughout the sequel, and present a definition.

ASSUMPTION 5.4.

(i) Let V be a finite-dimensional real linear space, let V^* be the dual of V and let $\rho \in \mathcal{M}(V)$ [see definition of $\mathcal{M}$ in (2.5)]. We assume that the NEF $F(\rho)$, generated by ρ, is steep, that is, $M_{F(\rho)} = \operatorname{int}\operatorname{conv}(S(\rho))$.

(ii) Let Λ be a convex subset of $\Theta(\rho)$; let u be the map $\Lambda \to E$, $\lambda \mapsto (\lambda, -k_\rho(\lambda))$; $\nu(\lambda) = \mathbf{1}_\Lambda(\lambda)\,d\lambda$; $\mu := u_*\nu$; let $\mathcal{F}$ be the Diaconis–Ylvisaker family generated by u and ν and associated with $F(\rho)$; and let D be the parameter set of $\mathcal{F}$ defined in (2.16).

(iii) Let S be the closure of the graph in $E = V \times \mathbb{R}$ of the function from Λ to $\mathbb{R}$ defined by $\lambda \mapsto -k_\rho(\lambda)$, and let $G := \{h \in V^* \times \mathbb{R}\colon h + D \subset D\}$.

DEFINITION 5.4. Let $\rho \in \mathcal{M}(V)$. The Jørgensen set of ρ, or of $F(\rho)$, is defined to be the set of all positive p such that there exists $\rho_p \in \mathcal{M}(V)$ with $\Theta(\rho_p) = \Theta(\rho)$ and $k_{\rho_p} = pk_\rho$.

A reference to the Jørgensen set can be found in Jørgensen (1986, 1987). If p is in the Jørgensen set of $F(\rho)$, then the log-Laplace transform of ρ_p and the NEF $F(\rho_p)$ are given, respectively, by pk_ρ and $\{\exp\{\theta x - pk_\rho(\theta)\}\rho_p(dx)\colon \theta \in \Theta(\rho) = \Theta(\rho_p)\}$. Naturally, the Jørgensen set is, by convolution, an additive semigroup and contains the set of all positive integers. Note that the Diaconis–Ylvisaker family $\mathcal{F}$, defined in Assumption 5.4, will not change if we replace ρ by some ρ_p with p in the Jørgensen set. The latter statement implies that no generality is lost, for instance, by considering the Diaconis–Ylvisaker family associated with the Bernoulli NEF generated by $\rho_1 := \delta_0 + \delta_1$, rather than the Diaconis–Ylvisaker family associated with the binomial NEF generated by $\rho_n := \sum_{i=0}^{n} \binom{n}{i} \delta_i$, since $k_{\rho_n} = nk_{\rho_1}$.

THEOREM 5.4 [Diaconis and Ylvisaker (1979), Theorem 2.1]. *Let ν_ρ be the measure defined by* $\nu_\rho(d\theta) := \exp\{-pk_\rho(\theta)\}\mathbf{1}_{\Theta(\rho)}(\theta)\,d\theta, p > 0$. *Then, under Assumption* 5.4, $\nu_\rho \in \mathcal{M}(V^*)$ *and* $\Theta(\nu_p) = pM_{F(\rho)}$.

PROPOSITION 5.5. *Under Assumption* 5.4, (i) $D \subset D(\mu)$ *and* (ii) G *is the closure of* D.

PROOF. (i) Let $(pv, v) \in D$. Then $\int_\Lambda \exp\{p(\langle \lambda, v\rangle - k_\rho(\lambda))\}\,d\lambda \le \int_{\Theta(\rho)} < \infty$.
(ii) Since $F(\rho)$ is steep, then $M_{F(\rho)}$ is convex and D is an open convex cone, and Proposition 3.1(vi) applies. □

We now apply Theorem 5.1 to $\mathcal{F}$ of Assumption 5.4. We distinguish between two kinds of affine hyperplanes in $E^* = V \times \mathbb{R}$ which deserve separate treatments: (a) hyperplanes which are not parallel to the line $\{0\} \times \mathbb{R}$, and (b) others. We omit the proof of the next proposition as it is analogous to that of Proposition 5.2.

PROPOSITION 5.6. *Suppose that Assumption* 5.4 *holds.*

(a) *Let* $q \in \mathcal{H}(S, G)$ *be concentrated on the hyperplane* $\{(v,p)\colon p = \langle \lambda_0, v\rangle + p_0\}$, *where* $(0,0) \neq (\lambda_0, p_0) \in E := V^* \times \mathbb{R}$. *Let* q_1 *denote the image of* q *by the projection of the hyperplane on* V: $(v,p) \mapsto v$. *Then*

$$k_{q_1}\big(\lambda - \lambda_0 k_\rho(\lambda)\big) = p_0 k_\rho(\lambda). \tag{5.10}$$

(b) *Let* $q \in \mathcal{H}(S, G)$ *be concentrated on the hyperplane* $\{(v,p)\colon \langle \lambda_0, v\rangle + 1 = 0\}$, *where* $\lambda_0 \neq 0$. *Let* $V_1 := \{v\colon \langle \lambda_0, v\rangle = 0\}$ *and let* q_0 *be the image of* q *in* $V_1 \times \mathbb{R}$ *by* $(v,p) \mapsto (v - v_0, p)$. *Then, for all* $\lambda \in \Lambda$,

$$k_{q_1}\big(\lambda + \langle \lambda, v_0\rangle \lambda_0,\ -k_\rho(\lambda)\big) = -\langle \lambda, v_0\rangle. \tag{5.11}$$

We now specialize Proposition 5.6 to the case $V = \mathbb{R}$, getting more information than in Proposition 5.2.

PROPOSITION 5.7. *Let Assumption* 5.4 *hold with* $V = \mathbb{R}$, *and let* Λ *be an open interval contained in* $\Theta(\rho)$.

(a) *Let* $q \in \mathcal{H}(S, G)$ *be concentrated on the line*

$$\{(v,p): p = \lambda_0 v + p_0\}, \qquad (\lambda_0, p_0) \neq (0,0), \tag{5.12}$$

and let q_1 *be the image of* q *by* $(v,p) \mapsto v$. (i) *If* $\lambda_0 = 0$, *then* p_0 *belongs to the Jørgensen set of* ρ. (ii) *If* $\lambda_0 \neq 0$, *then* $\lambda_0^{-1} \notin k'_\rho(\Lambda)$. *Moreover, the variance functions of* $F(q_1)$ *and* $F(\rho)$ *satisfy*

$$V_{F(q_1)}(m) = p_0^{-2}(p_0 + \lambda_0 m)^3 V_{F(\rho)}\big(m/(p_0 + \lambda_0 m)\big), \tag{5.13}$$

on $\widetilde{M}$, *the image of* Λ *by* $\lambda \mapsto m(\lambda) := p_0 k'_\rho(\lambda)/(1 - \lambda_0 k'_\rho(\lambda))$.

(b) *Let* $q \in \mathcal{H}(S, G)$ *be concentrated on the half-line* $\{(v,p): v = v_0, p \geq 0\}$, $v_0 \neq 0$. *Then, for all* $\lambda \in \Lambda$, $-k_{q_1}(-k_\rho(\lambda)) = v_0\lambda$ *and* $0 \notin k'_\rho(\Lambda)$. *Moreover, the variance functions of* $F(q_1)$ *and* $F(\rho)$ *satisfy*

$$V_{F(q_1)}(m) = v_0^{-2} m^3 V_{F(\rho)}(v_0/m) \tag{5.14}$$

on $\widetilde{M}$, *the image of* Λ *by* $\lambda \mapsto m(\lambda) := v_0/k'_\rho(\lambda)$.

PROOF. (i) This follows from (5.10). (ii) Introducing the map $\theta\colon \Lambda \to \mathbb{R}$, $\lambda \mapsto \theta(\lambda) := \lambda - \lambda_0 k_\rho(\lambda)$, and taking derivatives in (5.10), we get

$$k'_{q_1}\big(\theta(\lambda)\big)\theta'(\lambda) = p_0 k'_\rho(\lambda). \tag{5.15}$$

Hence, $\lambda \mapsto k'_{q_1}(\theta(\lambda))$ is analytic on Λ, as a composition of two analytic functions. Also, letting $H(x) := p_0 x/(1 - \lambda_0 x)$, we get from (5.15) that $k'_{q_1}(\theta(\lambda)) = H(k'_\rho(\lambda))$; hence, $k'_\rho(\Lambda)$ does not contain λ_0^{-1}. The remainder is obvious.

The proof of part (b) is similar. □

A practical procedure for applying Proposition 5.7 is described by the following steps. This procedure is applied in the examples of the next section.

1. Consider a given $\rho \in \mathcal{M}(\mathbb{R})$. Find $\Theta(\rho)$, $k_\rho = \log L_\rho$ (the log-Laplace transform of ρ) and $(V_{F(\rho)}, M_{F(\rho)})$ [the variance function of $F(\rho)$]. Use these in (2.14) and construct the Diaconis–Ylvisaker family $\mathcal{F}$ associated with $F(\rho)$ by

$$\mathcal{F} = \Big\{A(v,p)\exp\{pv\theta - pk_\rho(\theta)\}\mathbf{1}_{\Theta(\rho)}(\theta)\,d\theta\colon p > 0,\ v \in M_{F(\rho)}\Big\} \tag{5.16}$$

Such an $\mathcal{F}$ is taken to be the family of priors.

2. To search for the elements q_1 of the $\mathcal{H}(S,G)$ corresponding to $\mathcal{F}$ in (5.16), which are concentrated on lines, use Proposition 5.7. In particular, use (5.13) and (5.14), which relate the variance function of $F(q_1)$ with the given variance function of $F(\rho)$. Since $V_{F(q_1)}$ depends on λ_0 and v_0, such relations will enable one to determine which values of λ_0 and v_0 are such that $V_{F(q_1)}$ is a genuine variance function, and thereby whether $q_1 \in \mathcal{H}(S,G)$.
3. For practical convenience, one might use the equivalence between families in the sense of Section 4, and thus, by making a suitable transformation, replace $\mathcal{F}$ by an equivalent family $\mathcal{F}'$. If such a replacement is made, then apply steps 1 and 2 to $\mathcal{F}'$ (see Examples 6.1 and 6.2).

6. Examples: Diaconis–Ylvisaker families associated with NEF's in the Morris class. In this section, we apply the procedure described above to the Diaconis–Ylvisaker families in (5.16) associated with all NEF's in the Morris class [more specifically, the $F(\rho)$'s in the above procedure are taken to be the NEF's in the Morris class], and we determine explicitly the extreme points of the corresponding $\mathcal{H}(S,G)$ which are concentrated on lines of the form $\{(v,p)\colon p = \lambda_0 v + p_0\}$. Recall that Morris (1982) classified all NEF's in $\mathbb{R}$ having polynomial variance functions with degree less than or equal to 2. His classification shows that there are exactly six types of NEF's namely, the binomial, negative binomial, Poisson, gamma, normal and hyperbolic cosine NEF's. However, we need not examine all six types of such NEF's, but only four types, since the phenomenon of equivalence (Section 4) reduces the number of really distinct $\mathcal{H}(S,G)$. Specifically, if $\mathcal{F}_1$ and $\mathcal{F}_2$ are the Diaconis–Ylvisaker families associated with the Poisson and gamma NEF's (respectively, binomial and negative binomial NEF's), then the problem of studying their corresponding sets of measures $\mathcal{H}(S_1,G_1)$ and $\mathcal{H}(S_2,G_2)$ is the same, because one can be obtained from the other by a suitable affine bijection (see the paragraph after the proof of Proposition 4.3). Further explanation is given in Example 6.1.

In studying the following examples we utilize such an equivalence, as well as Corollary 4.2, Theorem 5.3 and, in particular, Proposition 5.7. For simplicity, we do not compute in these examples the normalizing constant $A(v,p)$ appearing in (5.16). We give complete details in Example 6.1 and skip most of such details in the other examples. We also consider in Examples 6.1, 6.3 and 6.4 the phenomenon of truncation of $\mathcal{F}$. Details are provided therein.

EXAMPLE 6.1 (The Diaconis–Ylvisaker family associated with binomial and negative binomial NEF's). Let $F(\rho)$ be the Bernoulli NEF. It is generated by $\rho = \delta_0 + \delta_1 \in \mathcal{M}(\mathbb{R})$. Here, $k_\rho(\theta) = \log(1+e^\theta)$, $\Theta(\rho) = \mathbb{R}$ and $(V_{F(\rho)}, M_{F(\rho)}) = (m - m^2, (0,1))$. From (5.16), the Diaconis–Ylvisaker family associated with $F(\rho)$ is

$$\mathcal{F}_1 = \left\{A(v,p)\exp\{pv\theta - p\log(1+e^\theta)\}\,d\theta\colon v \in (0,1), p > 0\right\}.$$

By a change of a variable to $\lambda = e^\theta$, one gets an equivalent family

$$\mathcal{F} = \left\{A(v,p)\left[\lambda^{vp-1}/(1+\lambda)^p\right]\mathbf{1}_{\mathbb{R}^+}(\lambda)\,d\lambda\colon v \in (0,1), p > 0\right\}, \tag{6.1}$$

TABLE 1
The set of extreme points of $\mathcal{H}(S, G)$, corresponding to $\mathcal{F}$ in (6.1) which are concentrated on the line $\{(v,p)\colon p = \lambda_0 v + p_0\}$

Line	$q \in \mathcal{H}(S,G)$	$(V_{F(q)}, M_{F(q)})$	**NEF's Name**
(1) $\{(v,p)\colon p = p_0\}$, $p_0 \in \mathbb{N}$	$\sum_{i=0}^{p_0} \binom{p_0}{i} \delta_i$	$(m(1 - m/p_0), (0, p_0))$	Binomial
(2) $\{(v,p)\colon p = v + p_0\}$, $p_0 > 0$	$\sum_{i=0}^{\infty} \frac{\Gamma(i+p_0)}{\Gamma(p_0)i!} \delta_i$	$(m(1 + m/p_0), \mathbb{R}^+)$	Negative binomial
(3) $\{(v,p)\colon v = 0\}$	δ_0	—	"Degenerate"
(4) $\{(v,p)\colon v = v_0\}$, $v_0 > 0$	$\delta_{v_0}^* \sum_{i=0}^{\infty} \frac{\Gamma(i+v_0)}{\Gamma(v_0)i!} \delta_i$	$(m(m/v_0 - 1), (v_0, \infty))$	Translation of negative binomial

which, by writing $a = vp$ and $p = a + b$, is the family of beta distributions of the second kind. (Another change of variable will lead to the beta family of the first kind, which is also equivalent to $\mathcal{F}_1$ or $\mathcal{F}$.) Similarly, starting with $\rho = \sum_{i=0}^{n} \binom{n}{i} \delta_i$, $n \in \mathbb{N}$, which generates the binomial NEF, or with $\rho = \sum_{i=0}^{\infty} \delta_i \Gamma(r + i)/[\Gamma(r) i!], r > 0$, which generates the negative binomial NEF [see Morris (1982), Table 1], we will end up, after a suitable change of variable, with $\mathcal{F}$ in (6.1). Accordingly, we can claim that $\mathcal{F}$ covers the Diaconis–Ylvisaker families associated with the binomial and negative binomial NEF's. Moreover, by applying Theorem 5.3(i) with $\Lambda = \mathbb{R}^+$, $u(\lambda) = (\log \lambda, -\log(1 + \lambda))$, and $T(\lambda) = \lambda(1 + \lambda)$, we can conclude that the variance functions corresponding to the solution measures of our problem are polynomials with degree less than or equal to 3, and therefore their corresponding NEF's are in the Morris-Mora class.

We now show that the set of all extreme points of $\mathcal{H}(S, G)$, with $G = \{(v,p)\colon 0 < v < p\}$ and $S = u(\mathbb{R}^-)$, which are concentrated on lines, include only four elements presented in Table 1. Naturally, ρ will be included in this set.

To show that these are the only extreme points on lines, we proceed as follows. Apply Proposition 5.7 to $\rho = \delta_0 + \delta_1$, compute $(V_{F(\rho)}, M_{F(\rho)})$ and construct $\mathcal{F}$ as in (6.1). Assume that $q \in \mathcal{H}(S, G)$ is concentrated on the line $p = \lambda_0 v + p_0$. Since, $\lambda_0^{-1} \notin M_{F(\rho)} = (0, 1)$, λ_0 cannot be greater than 1. If $\lambda_0 < 1$, then the intersection of the line $p = \lambda_0 v + p_0$ with G has compact support, and therefore the support of q_1 of Proposition 5.7(a) must be bounded. Since $V_{F(\rho)} = m - m^2$, (5.13) yields

$$V_{F(q_1)}(m) = m\big(1 + \lambda_0 m/p_0\big)\big(1 + (\lambda_0 - 1)m/p_0\big) \quad \text{and} \quad \widetilde{M} = \big(0, p_0/(1 - p_0)\big).$$

If $\lambda_0 < 1$ and $\lambda \neq 0$, then $V_{F(q_1)}$ is a third-degree polynomial corresponding to an NEF with bounded support. By the Mora (1986) classification of cubic variance functions, or by Bar-Lev and Bshouty [(1989), Theorem 2.1], there exists no such variance function. Hence, $\lambda_0 = 0$ or $\lambda_0 = 1$. The case $\lambda_0 = 0$ gives $V_{F(q_1)}(m) = m(1 - m/p_0)$ and $\widetilde{M} = (0, p_0)$. By the Morris classification, p_0 must be a positive integer and the latter variance function is then that of the binomial NEF. Hence, (1) of Table 1 follows. The case $\lambda_0 = 1$ gives $V_{F(q_1)}(m) = m(1 + m/p_0)$ and $\widetilde{M} = \mathbb{R}^+$. By Morris' classification, the latter variance function is that of the negative binomial NEF. Hence, (2) of Table 1 follows. We now apply Proposition 5.7(b). If q is concentrated on $\{(v,p): v = v_0\}$, then $v_0 \geq 0$, since this line must have a nonempty intersection with G. If $v_0 = 0$, then, by Corollary 4.2, $q = \delta_0$. Hence, (3) of Table 1 follows. If $v_0 > 0$, (5.11) gives $V_{F(q_1)}(m) = m(m/v_0 - 1)$ and $\widetilde{M} = (v_0, \infty)$. Let F_2 be the image of $F(q_1)$ by the translation map $x \mapsto x - v_0$. Then $V_{F_2}(m) = m(1 + m/v_0)$ on $\mathbb{R}^+$, which is the variance function of the negative binomial NEF. Hence, (4) of Table 1 follows. This completes the demonstration.

REMARK 6.1. We illustrate here the phenomenon of truncation for the family of priors $\mathcal{F}$ in (6.1). Consider

$$\rho = \sum_{i=0}^{\infty} \left[\binom{2i}{i} \Big/ (i+1)! \right] \delta_i.$$

Then $\Theta(\rho) = (-\infty, -\log 4)$, $k_\rho(\theta) = \log\left[1 - (1 - 4e^\theta)^{1/2}\right]/2e^\theta$, $V_{F(\rho)}(m) = m(m + 1)(2m + 1)$ and $M_{F(\rho)} = \mathbb{R}^+$. Such an $F(\rho)$ is a special case of the Takacs NEF's having cubic variance functions presented in Letac (1986) and Mora (1986) [see also Letac and Mora (1990)] under the name of fluctuation families. Except for such a special case, no other measures generating the Takacs NEF's have explicit Laplace transforms. The Diaconis–Ylvisaker family associated with $F(\rho)$ is

$$\mathcal{F} = \left\{ A(v,p) e^{\theta v p} \left[\left(1 - (1 - 4e^\theta)\right)^{1/2} / 2e^\theta \right]^p \mathbf{1}_{(-\infty, -\log 4)}(\theta)\, d\theta \colon p > 0,\ v > 0 \right\}.$$

The change of variable to $\lambda = (1 - (1 - 4e^\theta)^{1/2})/2$ shows that $\mathcal{F}$ is equivalent to

$$\mathcal{F}' = \left\{ A'(v,p) \lambda^{pv - 1} (1 - \lambda)^{p(v+1) - 1} (1 - 2\lambda) \mathbf{1}_{(0, 1/2)}(\lambda)\, d\lambda \colon p > 0,\ v > 0 \right\}.$$

Consider the maps $u\colon (0,1) \to \mathbb{R}^2$ and $\lambda \mapsto (\log \lambda, \log(1 - \lambda))$ and define $\mu := u_* \nu$. The NEF which is associated with $\mathcal{F}$ is $F(\mu)$. Such an NEF is concentrated on the same S as the NEF generated by beta distributions of the first kind given in (5.6). No wonder that $\mathcal{F}'$ is also a conjugate family for the Bernoulli NEF although it is not a Diaconis–Ylvisaker family. Thus, the study made on the family of beta distributions of the second kind in (6.1), which is equivalent to the family of beta distributions of the first kind, provides extreme points (those in Table 1) for the $\mathcal{H}(S', G')$ corresponding to such an $\mathcal{F}'$. Careful analysis of Proposition 5.7 will provide other extreme points to $\mathcal{H}(S', G')$. Such extreme points generate NEF's in the Morris–Mora class, since here, the $T(\lambda)$ of Theorem 5.3 is a polynomial in λ.

EXAMPLE 6.2 (The Diaconis–Ylvisaker family associated with Poisson and gamma NEF's). For $\rho = \sum_{i=0}^{\infty} \delta_i / i!$, $F(\rho)$ is the Poisson NEF for which $\Theta(\rho) = \mathbb{R}$, $k_\rho(\theta) = e^\theta$ and $(V_{F(\rho)}, M_{F(\rho)}) = (m, \mathbb{R}^+)$. The Diaconis–Ylvisaker family (5.16) associated with $F(\rho)$ is $\mathcal{F}_1 = \{(A(v,p) \exp\{\theta v p - p e^\theta\}\, d\theta \colon v > 0, p > 0\}$. A change

TABLE 2
The set of extreme points of $\mathcal{H}(S, G)$ corresponding to $\mathcal{F}$ in (6.2) which are concentrated on lines

Line	$q \in \mathcal{H}(S,G)$	$(V_{F(q)}, M_{F(q)})$	**NEF's Name**
(1) $\{(v,p): p = p_0\}$, $p_0 > 0$	$\sum_{i=0}^{\infty} \delta/i!$	$(m, \mathbb{R}^+)$	Poisson
(2) $\{(v,p): v = v_0 > 0$	$h^{v_0 - 1} \mathbf{1}_{\mathbb{R}^+} \, dh$	$(m^2/v_0,\ \mathbb{R}^+)$	Gamma with shape parameter v_0

of variable to $\lambda = e^{\theta}$ leads to an equivalent family

$$\mathcal{F} = \{A(v,p)\lambda^{vp-1}e^{-p\lambda}: v > 0, p > 0\}, \tag{6.2}$$

which is the family of gamma distributions. [For $\rho(dx) = x^{\alpha-1}\mathbf{1}_{\mathbb{R}^+}(x)\,dx$, $F(\rho)$ is the gamma NEF. Its associated Diaconis–Ylvisaker family can easily be shown to be equivalent to $\mathcal{F}$.] Here, $u(\lambda) = (\log \lambda, -\lambda)$ and $\Lambda = \mathbb{R}^+$. This case is covered by Theorem 5.3(i) since $T(\lambda) = \lambda$ and $\dim E = 2$. Hence, the variance functions of the solution measures are polynomials of degree less than or equal to 3. The set of extreme points of the corresponding $\mathcal{H}(S, G)$, with $G = (0, \infty)^2$, which are concentrated on lines is presented in Table 2.

Indeed, by applying Proposition 5.7(a) we realize that the case $\lambda_0 > 0$ is impossible, since $k'_\rho(\Theta(\rho)) = \mathbb{R}^+$ cannot contain λ_0^{-1}. Noting that q is concentrated on $G = (0, \infty)^2$, the case $\lambda_0 < 0$ implies that $\underset{\sim}{q}$ has a compact support. However, (5.13) yields $V_{F(q_1)}(m) = m(1 + \lambda_0 m/p_0)^2$ and $\widetilde{M} = (0, p_0/(1 - \lambda_0))$, which is not a variance function by the Mora classification. Hence, the case $\lambda_0 < 0$ is impossible. This leaves us with the case $\lambda_0 = 0$ and $p_0 > 0$, which yields (1) of Table 2. Considering the line $\{(v,p): v = v_0\}$, $v_0 > 0$, and utilizing (5.14) of Proposition 5.7(b) yields $V_{F(q_1)}(m) = m^2/v_0$ and $\widetilde{M} = \mathbb{R}^+$, which corresponds to (2) in Table 2.

EXAMPLE 6.3 (The Diaconis–Ylvisaker family associated with the normal NEF). Let ρ be the $N(0,1)$ law. Then $F(\rho)$ is the family of $N(\theta, 1)$ laws. Here, $\Theta(\rho) = \mathbb{R}$, $k_\rho(\theta) = \theta^2/2$ and $(V_{F(\rho)}, M_{F(\rho)}) = (1, \mathbb{R})$. The Diaconis–Ylvisaker family in (5.16) can be rewritten as $\mathcal{F} = \{(p/2\pi)^{1/2} \exp\{-p(\theta - v)^2/2\}: v \in \mathbb{R}, p > 0\}$, which is the family of $N(\theta,p)$ laws. Here $\Lambda = \Theta(\rho) = \mathbb{R}$, $u(\lambda) = (-\lambda^2/2, \lambda)$ and, by Theorem 5.3, $T(\lambda) = 1$. Hence, the variance functions of the solution measures must be polynomials of degree less than or equal to 3. An analysis analogous to those used in previous examples shows that there is only one extreme point, corresponding to the case $\lambda_0 = 0$, which is concentrated on the line $\{(v,p): p = p_0\}, p_0 > 0$. The corresponding variance function is $(p_0^{-1}, \mathbb{R})$, which is that of the NEF composed of $N(\theta, p_0^{-1})$ laws. We make two remarks concerning such a small set of extreme points.

(i) Generally speaking, for $\mathcal{F}$, a Diaconis–Ylvisaker family, one can raise the following question. Is it true that *all* extreme points of $\mathcal{H}(S, G)$ are concentrated on lines? This is not true for this particular case of the family $\mathcal{F}$ of normal laws. For this case, the statement in the latter question is equivalent to the following one. Assume that (X, Y) is a random variable on $\mathbb{R} \times \mathbb{R}^+$ such that, for all $u \in \mathbb{R}$, $\mathbf{E}[\exp\{uXY - u^2Y^2/2\}] = 1$. Then X and Y are independent and the distribution of X is $N(0,1)$. A counterexample to the latter statement is easily obtained, however, by considering a standard Brownian motion B on $\mathbb{R}$ and defining $\sqrt{Y} = \inf\{t; |B(t)| = 1\}$ and $X = B(\sqrt{Y})$.

(ii) Consider the truncated normal family

$$\mathcal{F}' = \left\{A(v,p)\exp\{\theta v - p\theta^2/2\}\mathbf{1}_{\mathbb{R}^+}(\theta)\,d\theta\colon v \in \mathbb{R}, p > 0\right\},$$

where $A(v,p)$ is a normalizing constant. Then Proposition 5.7 still applies with $\Lambda = \mathbb{R}^+$. Here, $k'_\rho(\Lambda) = \mathbb{R}^+$ and Proposition 5.7(a) provides a larger set of extreme points, namely, to the previous case where $\lambda_0 = 0$, one adds the case $\lambda_0 < 0$, for which $V_{F(q_1)}(m) = p_0^{-2}(p_0 + \lambda_0 m)^3$ and $\widetilde{M} = (0, -p_0/\lambda_0)$. In this case $F(q_1)$ is the image of an inverse Gaussian family having variance function $(m^3, \mathbb{R}^+)$ [see Jørgensen (1987)], by the map $x \mapsto (x - p_0)/\lambda_0$. Proposition 5.7(b) provides an extreme point which corresponds to a scale transformation of the inverse Gaussian NEF, since (5.10) becomes $V_{F(q_1)}(m) = v_0^{-2}m^3$ and $\widetilde{M} = \mathbb{R}^+$.

EXAMPLE 6.4 (The Diaconis–Ylvisaker family associated with the hyperbolic-cosine NEF). To cover the last type in the Morris class, consider $\rho(dx) = (\cosh \pi x/2)^{-1}\,dx$. Here, $\Theta(\rho) = (-\pi/2, \pi/2)$, $k_\rho(\theta) = -\log\cos\theta$, $(V_{F(\rho)}, M_{F(\rho)}) = (m^2 + 1, \mathbb{R})$, $D = (0,\infty)^2$ and $\mathcal{F} = \{A(v,p)e^{\theta vp}/(\cos\theta)^p\,d\theta\colon v \in \mathbb{R}, p > 0\}$. As in the normal case, Proposition 5.7 provides only one set of extreme points corresponding to $\lambda_0 = 0$. However, as in Example 6.3, if we consider the truncated family $\mathcal{F}' = \{A'(v,p)[e^{\theta vp}/(\cos\theta)^p]\mathbf{1}_{\mathbb{R}^+}(\theta)\,d\theta\colon v \in \mathbb{R}, p > 0\}$, the resulting set of extreme points is larger and its study is left to the reader. Note, however, that since here we are in the realm of Theorem 5.3 with $\dim E = 2$ and $T(\lambda) = \lambda^2 + 1$, such extreme points are in the Morris–Mora class; in particular, a "strict arcsine" family appears as an NEF corresponding to one of these extreme points [see Letac and Mora (1990) and Mora (1986)].

7. Conclusions. We make some comments on the nature of difficulties that one may face in trying to extend and apply the results of this paper, and those of Section 5, in particular, to the case of Diaconis–Ylvisaker families associated with multivariate NEF's $F(\rho)$, $\rho \in \mathcal{M}(\mathbb{R}^n)$, say. Such difficulties occur, for instance, when $F(\rho)$ is the family of Wishart distributions, or when $\mathcal{F}$ is the family of Dirichlet distributions. If we try to apply Proposition 5.5, we are led to characterizations of q by variance functions in several dimensions, for which the results are very sparse.

Similarly, if we consider the case where S is concentrated on a skew curve (as in Proposition 5.2 or Theorem 5.3), we get variance functions outside of the familiar frame of the Morris–Mora class. A good theory of variance functions of the form $P\Delta + Q\sqrt{\Delta}$, as in Theorem 5.3, is desirable. This will enable use to solve some tantalizing cases, such as the hypergeometric family or the generalized inverse Gaussian family.

APPENDIX

PROOF OF PROPOSITION 4.4. Let β and r be nonnegative. Define the Bessel function $I_\beta(r) = \sum_{n=0}^{\infty}(r/2)^{\beta+2n}[n!\Gamma(n+\beta+1)]^{-1}$, and let

$$A_n(r) := \Gamma(d/2)\,(r/2)^{-d/2+1} I_{n+d/2-1}(r), \qquad n \in \mathbb{N}_0, \tag{A.1}$$

where d is the dimension of the Euclidean space. For $n = 0$, (A.1) coincides with (4.2). Let E_n denote the finite-dimensional subspace of $C^\infty(S)$, the space of eigenfunctions of the Laplacian Δ_S of the sphere, associated with the eigenvalue $-n(n+k-2)$ [see Seeley (1966)]. For ν being the uniform measure on S, let $L^2(\nu)$ be the orthogonal direct sum $\oplus_{n=0}^{\infty}E_n$. Then it is known [e.g., Letac (1985), Proposition 6] that if ℓ is a signed measure on S and if g_ℓ, defined on $[0,\infty)\times S$ by $g_\ell(r,u) = \int_S \exp\{r\langle h, u\rangle\}\ell(du)$, then there exists a sequence $(C_n)_{n=0}^{\infty}$ of real numbers and a sequence $(e_n(u))_{n=0}^{\infty}$ of functions on S such that $e_n(u) \in E_n$, for all $u \in S$, $\sum_{n=0}^{\infty} C_n^2(A_n(r))^2 < \infty$, for all $r \ge 0$, and $g_\ell(r,u) = \sum_{n=0}^{\infty}C_nA_n(r)e_n(u)$, in the $L^2(\nu)$ sense. Suppose now that $g_\ell(1,u) = 0$, for all $u \in S$. From the orthogonality of $(E_n)_{n=0}^{\infty}$, we get $C_nA_n(1) = 0$, for all $n \in \mathbb{N}_0$. Since $A_n(1) \neq 0$, we get $C_n = 0$ and $g_\ell(r, u) = 0$, for all $(r,u) \in [0,\infty] \times S$. However, $g_\ell(r,u)$ is the Laplace transform of ℓ expressed in polar coordinates, that is, $L_\ell(r, u) = g_\ell(r, u)$. Hence $L_\ell(\theta) = 0$ for all θ in the Euclidean space and $\ell = 0$. □

PROOF OF THEOREM 5.1. (a) First note that taking $e_0^* = 0$ yields $q = \delta_0$ (Corollary 4.2). To avoid trivialities, we assumed $e_0^* \neq 0$. We now prove (i) and (ii).

(i) If $h_1 \in E_1^*$, then $\langle e_0^* + h_1, u_0(\lambda) + u_1(\lambda)\rangle = \langle h_1, u_1(\lambda)\rangle + \langle e_0^*, u_0(\lambda)\rangle$, so that $I = \int_E \exp\{\langle h, u(\lambda)\rangle\}q(dh) = \exp\{\langle e_0^*, u_0(\lambda)\rangle\}L_{q_1}(u_1(\lambda))$, where L_{q_1} is the Laplace transform of q_1. This implies (5.1). Since $\operatorname{int} p_1(S) \neq \phi$ [Assumption 5.1(i)], q_1 is determined by (5.1) from the uniqueness property of Laplace transforms. Hence, q is unique. It is also extreme, since if $q = \alpha q' + (1-\alpha)q''$, $\alpha \in (0,1), q', q'' \in \mathcal{H}(S, G)$, then q' and q'' are also concentrated on $e_0^* + E_1^*$, and $q = q' = q''$ from the uniqueness property. To see that $q_1 \in \mathcal{M}(E_1^*)$, assume to the contrary that q_1 is concentrated on an affine hyperplane of E_1^* of dimension smaller than $\dim S$. To avoid new notation, suppose, for now, that $\dim E_1^* = m < d_s = \dim S$ and that $q \in \mathcal{M}(E_1^*)$. We show that we get a contradiction. Note that the case $\ell = 0$ is excluded; if not, we would have that $q = \delta_{e_0^*}$, and $\exp\{\langle e_0^*, u\rangle\} = 1$, for all $u \in S$, so that $e_0^* = 0$. However, the latter case was excluded in the hypotheses of the theorem. Let $u_1 \in \operatorname{int} p_1(S)$ and $s \in S$ be such that $p_1(s) = u_1$. Consider the analytic map $f\colon W \to S$ mentioned in Assumption 5.1(ii), where W is a small open ball in $\mathbb{R}^{d_s}$. Then $k_{q_1}(p_1(f(w))) = -\langle e_0^*, f(w)\rangle$. Taking suitable coordinates, we get, with obvious notation, that $k_{q_1}(f_1(w), \dots, f_m(m)) = -(a_{m+1}f_{m+1}(w) + \cdots + a_d f_d(w))$, where the rank of the matrix $(\partial f_j/\partial w_i)i$, $j = 1, \dots, d_s$ is d_s, and then taking the differential with respect to w, we get

$$\sum_{j=1}^{m} \frac{\partial k_{q_1}}{\partial \theta_j} f_j'(w) + \sum_{j=m+1}^{d} a_j f_j'(w) = 0.$$

Since $q_1 \in \mathcal{M}(E_1^*)$, there exists $w \in W$ such that $(\partial k_{q_1}/\partial \theta_j)_{j=1}^m$ is not $(0)_{j=1}^m$, the zero vector. Hence, the rank of $(f_j')_{j=1}^{d_s}$ is not d_s, a contradiction.

(ii) This is an immediate consequence of (5.1) and Theorem 3.2.

(b) Let $\langle\ ,\ \rangle_V$ denote the canonical bilinear form on $V^* \times V$ and $\langle\ ,\ \rangle$ that on $E^* \times E$, as usual. From Theorem 3.2, there exist a map h: $V \to E^*$ and a measure Q on V such that, for ν-almost all λ,

$$\text{(A.2)}\qquad \exp\Big\{\big\langle h(x),\, u(\lambda)\big\rangle\Big\}Q(dx) = \exp\Big\{\big\langle \theta(\lambda),\, x\big\rangle_V - k_\delta(\theta(\lambda)\Big\}\delta(dx).$$

Hence, Q and δ are equivalent measures. Denote by $\mathbf{L}^0(\delta)$ the space of equivalence classes of measurable functions on V with the relation $f_1 \sim f_2$, if $f_1 - f_2 = 0$, δ-a.e. Consider $f(\lambda, x) := \langle h(x), u(\lambda)\rangle - \langle \theta(\lambda), x\rangle_V + k_\delta(\theta(\lambda))$. Then, by (A.2), the map $\lambda \mapsto f(\lambda, \cdot)$, $\Lambda \to \mathbf{L}^0(\delta)$, is constant ν-a.e and there exists $\lambda_0 \in \Lambda$ such that, for $\nu \otimes \delta$-almost all (λ, x),

$$\text{(A.3)}\qquad \big\langle h(x), u(\lambda) - u(\lambda_0)\big\rangle + k_\delta\big(\theta(\lambda)\big) - k_\delta\big(\theta(\lambda_0)\big) = \big\langle \theta(\lambda) - \theta(\lambda_0), x\big\rangle_V.$$

By an assumption, there exist $\lambda_1, \dots, \lambda_d \in \Lambda$ such that $(u(\lambda_j)) - (u(\lambda_0))_{j=1}^d$ is a basis in E and (A.3) holds for $\lambda \in \{\lambda_1, \dots, \lambda_d\}$. There exists $e_0^* \in E$ such that, for $j = 1, 2, \dots, d$, $-\langle e_0^*, u(\lambda_j) - u(\lambda_0)\rangle = k_\delta(\theta(\lambda_j)) - k_\delta(\theta(\lambda_0))$. Define the linear map φ: $V \to E$, by $\langle \varphi(x), u(\lambda_j) - u(\lambda_0)\rangle = \langle \theta(\lambda_j) - \theta(\lambda_0), x\rangle_V$, $j = 1, \dots, d$. Therefore, (A.3) implies that $h(x) = e_0^* + \varphi(x)$, δ-a.e. Since, for ν-almost all λ, $\langle \varphi(x), u(\lambda) - u(\lambda_0)\rangle = \langle \theta(\lambda) - \theta(\lambda_0), x\rangle_V$, we get, ${}^t\varphi(u(\lambda) - u(\lambda_0)) = \theta(\lambda) - \theta(\lambda_0)$, ν-a.e., where ${}^t\varphi$: $E \to V^*$ denotes the transpose of φ. The latter relation along with the assumption made on $u_*\nu$ and $\theta_*\nu$ imply that ${}^t\varphi(E) = V^*$. Hence, the rank of φ is $\dim V$ and φ is injective. Proving that $\dim V = d_S$ can be conducted in a manner analogous to that used to prove that $q_1 \in M(E_1^*)$ in part (a). We omit this proof for brevity. □

PROOF OF THEOREM 5.3(ii). For $\dim E = 3$, the computations are more complicated and we borrow some ideas from Letac (1992). Let $P_i(\lambda) := a_i\lambda^2 + b_i\lambda + c_i$, $i = 0, 1$, $A(m) := a_0 + a_1 m$, $B(m) := b_0 + b_1 m$, $C(m) := c_0 + c_1 m$, $A_1 := b_0c_1 - c_0b_1$, $B_1 := c_0a_1 - a_0c_1$, $C_1 := a_0b_1 - b_0a_1$. Note that here we have chosen a basis (f_1, f_2, f_3) of E such that $T(\lambda)u'(\lambda) = \lambda^2 f_1 + \lambda f_2 + f_3$. If we put on E the oriented Euclidean structure admitting (f_1, f_2, f_3) as a positive orthonormal basis, then the vectors $a_if_1 + b_if_2 + c_if_3$, $i = 0, 1$, must be independent, and $A_1f_1 + B_1f_2 + C_1f_3$ is their cross product. This implies that

$$\text{(A.4)}\qquad A_1A + B_1B + C_1C = 0 \quad \text{for all } m.$$

Letting $\Delta := B^2 - 4AC$, $D_1 := B_1^2 - A_1C_1$ and $T(\lambda) = t_2\lambda^4 + t_1\lambda^3 + t_0\lambda^2 + t_{-1}\lambda + t_{-2}$, we get from (5.8) that

$$\text{(A.5)}\quad V_{F(q_0)}(m) = \frac{-C_1\lambda + 2B_1 - A_1\lambda^{-1}}{\big(a_1\lambda + b_1 + c_1\lambda^{-1}\big)^3}\big(t_2\lambda^2 + t_1\lambda + t_0 + t_{-1}\lambda^{-1} + t_{-2}\lambda^{-2}\big).$$

Since $A(m)\lambda^2 + B(m)\lambda + C(m) = 0$ on Λ, we easily conclude that $\Delta(m)$ is the square of a first-degree polynomial if and only $D_1 = 0$, and this is equivalent to requiring that P_0 and P_1 have a common root. For brevity, we assume $D_1 \neq 0$. [The calculations required for the case $D_1 = 0$ are easy to carry out by using (5.9) and (A.5), since λ is a Möbius function of m.]

Denote by $\widetilde{M}$ the image of Λ by $\lambda \mapsto P_0(\lambda)/P_1(\lambda)$. Then, $\widetilde{M}$ is open, which implies that $\Delta > 0$ on $\widetilde{M}$. Fix ε in $\{-1, 1\}$ and two open intervals $M_1 \subset \widetilde{M}$ and $\Lambda_1 \subset \Lambda$ such that $\lambda(m) = (-B + \varepsilon\sqrt{\Delta})/2A$ defines a bijective mapping from M_1 to Λ_1. We will prove that (5.9) holds for all $m \in M_1$. The fact that $V_{F(q_0)}$ is real analytic on $\widetilde{M}$ will enable us to extend (5.9) to all $m \in \widetilde{M}$. Using the relation $\lambda^{-1} = (-B + \varepsilon\sqrt{\Delta})/2C$, a tedious computation gives

$$a_1\lambda + b_1 + c_1\lambda^{-1} = \left(C_1C - A_1A + \varepsilon B_1\sqrt{\Delta}\right)/2AC. \tag{A.6}$$

By using (5.8), we observe that $(C_1C - A_1A)^2 - B_1^2\Delta = 4ACD_1$, and hence we get from (A.6) that

$$\left(a_1\lambda + b_1 + c_1\lambda^{-1}\right)^{-1} = \left(C_1C - A_1A + \varepsilon B_1\sqrt{\Delta}\right)/2D_1. \tag{A.7}$$

Noting that

$$\frac{C_1\lambda - 2B_1 + A_1\lambda^{-1}}{a_1\lambda + b_1 + c_1\lambda^{-1}} = \frac{P_1'P_0 - P_0'P_1}{P_1} = -(P_0' + mP_0')$$

and that $(P_0' + mP_1')^2 - \Delta = (2A\lambda + B)^2 - \Delta = 4A(A\lambda^2 + B\lambda + C) = 0$, it follows that $-(P_0' + mP_1') = \varepsilon\sqrt{\Delta}$. Using the latter result, squaring (A.7) and employing the resulting equation in (A.5) yields

$$\begin{aligned} V_{F(q_0)}(m) = {} & \left(4D_1^2\right)^{-1}\left[-2B_1(CC_1 - AA_1)\Delta + \left((CC_1 - AA_1)^2 + B_1^2\Delta\right)\varepsilon\sqrt{\Delta}\right] \\ & \times \left(t_2\lambda^2 + t_1\lambda + t_0 + t_{-1}\lambda^{-1} + t_{-2}\lambda^{-2}\right). \end{aligned}$$

Note that the coefficient of t_0 is already of the required form. We still have to prove that the coefficients of t_1, t_2, t_{-1} and t_{-2} have the form $P\Delta + Q\sqrt{\Delta}$, where P and Q are polynomials with degree less than or equal to 1 and 2, respectively. We prove this for t_1 and t_2. (The proof for t_{-1} and t_{-2} can be obtained analogously by exchanging A with C and ε with $-\varepsilon$.) Since such a proof is a problem of formal algebra, we assume that $A_1, B_1, C_1 \neq 0$ and translate the expressions in terms of polynomials in $a := AA_1/B_1$ and $c := CC_1/B_1$. We get from (A.4) that $-B = a + c$, $\Delta = (a + c)^2 - 4rac$ and $\lambda = A_1(a + c + \varepsilon\sqrt{\Delta})/2B_1$, where $r := B_1^2/A_1C_1$. Since the coefficient of t_0 is $(-2(c - a)\Delta + ((c - a)^2 + \Delta)\varepsilon\sqrt{\Delta})B_1^2/4D_1^2$, the coefficient of t_1 will be, after multiplication by $A_1B_1/8D_1^2$,

$$\begin{aligned} & a^{-1}\left(-2(c - a)\Delta + \left((c - a)^2 + \Delta\right)\varepsilon\sqrt{\Delta}\right)\left(a + c + 2\sqrt{\Delta}\right) \\ & = 4(a - rc)\Delta + 4\left(a^2 + ac(1 - 3r) + rc^2\right)\varepsilon\sqrt{\Delta}, \end{aligned}$$

which has the desired form. Similarly, the coefficient, of t_2 will be, after multiplication by $A_1^2/4D_1^2$

$$a^{-1}\Big((a-rc)\Delta+\big(a^2+ac(1-3r)+rc^2\big)\varepsilon\sqrt{\Delta}\Big)(a+c+\varepsilon\sqrt{\Delta})$$
$$=2\big(a+c(1-2r)\big)\Delta+\Big(a^2+2ac(1-2r)+c^2\big(1-2r+2r^2\big)\Big)\sqrt{\Delta},$$

which also have the desired form. This completes the proof. □

Acknowledgment. We are grateful to an Associate Editor for his careful criticism and many helpful comments which resulted in a considerably improved version of the original manuscript.

REFERENCES

BAR-LEV, S. K. and BSHOUTY, D. (1989). Rational variance functions. *Ann. Statist.* **17** 741–748.

BAR-LEV, S. K., BSHOUTY, D. (1990). A class of infinitely divisible variance functions with an application to the polynomial case. *Statist. Probab. Lett.* **10** 377–379.

BAR-LEV, S. K., BSHOUTY, D. and ENIS, P. (1991). Variance functions with meromorphic means. *Ann. Probab.* **19** 1349–1366.

BAR-LEV, S. K., BSHOUTY, D. and ENIS, P. (1992). On polynomial variance functions. *Probab. Theory Related Fields* **94** 69–82.

BARNDORFF-NIELSEN, O. (1978). *Information and Exponential Families in Statistical Theory.* Wiley, New York.

BERGER, J. O. (1980). *Statistical Decision Theory, Foundations, Concepts, and Methods.* Springer, New York.

DIACONIS, P. and YLVISAKER, D. (1979). Conjugate priors for exponential families. *Ann. Statist.* **7** 269–281.

EATON, M. (1981). On the projections of isotropic distributions. *Ann. Statist.* **9** 391–400.

JØRGENSEN, B. (1986). Some properties of exponential dispersion models. *Scand. J. Statist.* **13** 187–196.

JØRGENSEN, B. (1987). Exponential dispersion models. *J. Roy. Statist. Soc. Ser. B* **49** 127–162.

LELONG-FERRAND, J. (1963). *Géométrie Différentielle.* Masson, Paris.

LETAC, G. (1985). Une caractérisation élémentaire des solutions positives de l'équation de Helmholtz $\Delta f=f$. *Bull. Sci. Math. (2)* **109** 373–378.

LETAC, G. (1986). La réciprocité des familles exponentielles. *C. R. Acad. Sci. Paris Sér. I. Math.* **303** 61–64.

LETAC, G. (1992). *Lectures on Natural Exponential Families and their Variance Functions. Monografias de Matematica* **50**. Instituto de Matematica Pura e Aplicada, Rio de Janeiro.

LETAC, G. and MORA, M. (1990). Natural exponential families with cubic variance functions. *Ann. Statist.* **18** 1–37.

LINDLEY, D. V. (1972). *Bayesian Statistics, a Review.* SIAM, Philadelphia.

MORA, M. (1986). Classification des fonctions variances cubiques des familles exponentielles sur $\mathbb{R}^n$. *C. R. Acad. Sci. Paris Sér. I. Math.* **302** 587–590.

MORLAT, G. (1956). Les lois de probabilitiés de Halphen. *Rev. Statist. Appl.* **4** 21–46.

MORRIS, C. N. (1982). Natural exponential families with quadratic variance functions. *Ann. Statist.* **10** 65–80.

PHELPS, R. R. (1966). *Lectures on Choquet's Theorem.* Van Nostrand, New York.

RAIFFA, H. and SCHLAIFER, R. (1961). *Applied Statistical Decision Theory*, Graduate School of Business Administration, Harvard Univ.

SEELEY, R. T. (1966). Spherical harmonics. *Amer. Math. Monthly* **73** 115–121. (Slaught memorial paper number 11.)

SESHADRI, V. (1991). Finite mixtures of natural exponential families. *Canad. J. Statist.* **19** 437–445.

TWEEDIE, M. C. K. (1945). Inverse statistical variates. *Nature* **155** 453.

SHAUL K. BAR-LEV
DEPARTMENT OF STATISTICS
UNIVERSITY OF HAIFA
HAIFA, 31905
ISRAEL

PETER ENIS
DEPARTMENT OF STATISTICS
STATE UNIVERSITY OF NEW YORK
AT BUFFALO
BUFFALO, NEW YORK 14214

GÉRARD LETAC
LABORATOIRE DE STATISTIQUE
ET PROBABILITÉS
UNIVERSITÉ PAUL SABATIER
TOULOUSE 31062
FRANCE