HOW REGULAR ARE CONJUGATE EXPONENTIAL FAMILIES?

Ulrich MÜLLER-FUNK

Institut für Mathematische Statistik, Universitat Münster, Einsteinstr. 62, D-4400 Munster, FR Germany

Friedrich PUKELSHEIM *

Stanford University, Stanford, CA, USA

Abstract Given an exponential family of sampling distributions of order k, one may construct in a natural way an exponential family of conjugate (that is, prior) distributions depending on a k-dimensional parameter c and an additional weight $w > 0$. We compute the bias term by which the expectation of the sampling mean-value parameter under the conjugate distribution deviates from the conjugate parameter c. This bias term vanishes for regular exponential families, providing an appealing interpretation of the conjugate parameter c as a 'prior location' of the sampling mean-value parameter. By way of example we explore the extension of this approach to moments of higher order, in order to interprete the conjugate weight w as a 'prior sample size'.

1. Introduction

Conjugate exponential families were introduced by Raiffa and Schlaifer (1961, Section 3.2). A review, together with many references, is given by Dickey (1982). Conjugate families are also included in the monographs by Barndorff-Nielsen (1978, pp. 131f) and Brown (1986, pp. 112ff).

For an exponential sampling family with an open canonical parameter domain Θ Diaconis and Ylvisaker (1979, Theorem 2) established the remarkable result that the conjugate parameter c is the expected value of the sampling mean-value parameter $\tau(\theta)$ under the conjugate distribution $Q_c^{(w)}$,

$$E_{Q_c^{(w)}}[\tau(\theta)] = c;$$

see also Theorem 4.19 in Brown (1986, p. 113). In the present note we extend this result to arbitrary exponential families, and explore the validity of a similar statement for the conjugate weight w.

In Section 2 we review conjugate exponential families given a sampling family that is exponential and of order k, in the terminology of Barndorff-Nielsen (1973). The property of being closed under sampling suggests to introduce sample size n and conjugate weight w as a $k+1^{\text{st}}$ parameter, leading to the families $\mathscr{P}$ and $\mathscr{Q}$ of sampling distributions and conjugate distributions, respectively.

Conjugate densities are log-concave, as discused in Section 3. Barndorff-Nielsen (1973, p. 194; 1978, p. 93) provides a useful integrable majorant for log-concave densities. Based on this majorant a short-cut proof and an extension of Theorem 2 of Diaconis and Ylvisakar (1979) is given. This permits an exhaustive discussion of exponential families of order 1.

* On leave from the Universität Augsburg, Institut fur Mathematik, Memminger Str 6, D-8900 Augsburg, FR Germany. Research of F.P. supported by the Stiftung Volkswagenwerk, Hannover, and the Mathematical Sciences Research Institute, Cornell University.

In Section 4 we go one step further and outline a corresponding result for the conjugate weight w,

$$V_{Q_c^{(w)}}[\tau(\theta)] = \frac{1}{w} E_{Q_c^{(w)}}[I(\theta)],$$

that is, under the conjugate distribution $Q_c^{(w)}$ for the sampling parameter θ the variance–covariance matrix of the mean-value parameter $\tau(\theta)$ is equal to expected Fisher information $I(\theta)$ rescaled by the conjugate weight w. We show that even in classical regular families this result holds true only when the weight w is sufficiently large, and that there are other regular families for which it fails completely.

2. Conjugate exponential families

Let $\mathscr{P}^{(1)} = \{P_\theta^{(1)}: \theta \in \Theta\}$ be a family of distributions on $\mathbb{R}^k$ for a sample of size 1. Suppose that $\mathscr{P}^{(1)}$ forms a full and linear exponential family of order k, in the terminology of Barndorff-Nielsen (1978, Chapter 8). Then there exists a dominating measure μ such that the distribution $P_\theta^{(1)} \in \mathscr{P}^{(1)}$ admits a μ-density

$$p_\theta^{(1)}(t) \propto e^{\theta' t - \kappa(\theta)} \quad \text{for } t \in C,$$

while $p_\theta^{(1)}(t) = 0$ for $t \notin C$. Here $\theta' t = \Sigma_{j \leq k} \theta_j t_j$ is the Euclidean inner product of the column vectors θ and t, and $\kappa(\theta) = \log \int e^{\theta' t}\, d\mu(t)$ is the cumulant transformation. These distributions all share a common closed convex support

$$C = \text{cl conv supp } \mathscr{P}^{(1)}$$

which is the range of variation of t, see Barndorff-Nielsen (1978, p. 90).

The family of 'prior distributions' conjugate to $\mathscr{P}^{(1)}$ is defined to be a family of distributions on the parameter domain Θ. The parameter domain Θ is a Borel subset of $\mathbb{R}^k$ since it is the countable union of closed sets $\bigcup_{n \geq 1} \{\kappa \leq n\}$, due to the closedness of the cumulant transformation k, see Barndorff-Nielsen (1979, p. 103). The parameter θ (and the canonical statistic t) are unique only up to regular affine transformations, leaving k-dimensional Lebesgue measure λ as the only reasonable candidate for a dominating measure.

Convex duality determines the conjugate parameter, based on a result due to Barndorff-Nielsen (1978, p. 93; 1973, p. 194). The key point is that the function $f(\theta) = \kappa(\theta) - \theta' c$ is a convex function of θ, whence the functions $e^{-wf(\theta)} = (e^{\theta' c - \kappa(\theta)})^w$ are log-concave in θ for fixed weight $w > 0$.

Lemma 2.1. *For every weight $w > 0$ and for every vector $c \in \mathbb{R}^k$ one has*

$$\int_\Theta \left(e^{\theta' c - \kappa(\theta)}\right)^w d\lambda(\theta) < \infty \iff c \in \text{int } C.$$

Proof. The convex function $f(\theta) = \kappa(\theta) - \theta' c$ is closed and has an effective domain with nonempty interior. Thus the integral $\int_\Theta e^{-wf(\theta)}\, d\lambda(\theta)$ is finite if and only if

$$0 \in \text{int dom}(wf)^* = \text{int}(w((\text{dom } \kappa^*) - c))$$
$$= w((\text{int } C) - c),$$

where an asterisk indicates conjugate functions. The latter property means that c lies in the interior of C. Details of this reasoning are given in Barndorff-Nielsen (1978, pp. 132, 93, 140). Thus the proof is complete.

This justifies the definition of the family $\mathscr{Q}^{(w)}$ of *conjugate distributions* $Q_c^{(w)}$ *for weight* $w > 0$ on Θ by requiring that c lies in the interior of C and that $Q_c^{(w)}$ has Lebesgue density

$$q_c^{(w)}(\theta) \propto \left(e^{\theta' c - \kappa(\theta)}\right)^w \quad \text{for } \theta \in \Theta,$$

while $q_c^{(w)}(\theta) = 0$ for $\theta \notin \Theta$. Thus $\mathscr{Q}^{(w)}$ is an exponential family with canonical parameter c and canonical statistic $w\theta$. The union over varying positive weights $w > 0$,

$$\mathscr{Q} = \bigcup_{w>0} \mathscr{Q}^{(w)} = \{Q_c^{(w)}: c \in \text{int } C,\ w > 0\},$$

will be called the *family of conjugate distributions*. The weight w plays a role similar to sample size as to be discussed next.

Suppose $T_1, \ldots, T_n$ is a sample of size n from a distribution $P_\theta^{(1)} \in \mathscr{P}^{(1)}$. Then the sample average $T^{(n)} = \Sigma_{i \leq n} T_i / n$ is distributed according to the distribution $P_\theta^{(n)}$ that has density

$$p_\theta^{(n)}(t) \propto \left(e^{\theta' t - \kappa(\theta)}\right)^n \quad \text{for } t \in C,$$

while $p_\theta^{(n)}(t)=0$ for $t \notin C$, relative to some dominating measure $\mu^{(n)}$. The family $\mathscr{P}^{(n)} = \{P_\theta^{(n)}\colon \theta \in \Theta\}$ of *sampling distributions for sample size* n is an exponential family with canonical parameter θ and canonical statistic nt. For a seamless correspondence between $\mathscr{P}^{(n)}$ and $\mathscr{R}^{(m)}$ we let the 'sample size' n vary continuously over all positive numbers $(0, \infty)$ whenever possible, rather than restricting it to take integer values only. Thus the *family of sampling distributions* is

$$\mathscr{P} = \bigcup_{n>0} \mathscr{P}^{(n)} = \left\{P_\theta^{(n)}\colon \theta \in \Theta,\ n>0\right\}.$$

The most useful consequence of this correspondence is that the family $\mathscr{Q}$ of conjugate distributions is closed under sampling from $\mathscr{P}$: A prior distribution $Q_c^{(w)} \in \mathscr{Q}$ for the sampling parameter $\theta \in \Theta$ is transformed by an observed response t – distributed according to $P_\theta^{(n)} \in \mathscr{P}$ – into a posterior distribution of θ given t that again is a member of $\mathscr{Q}$, namely

$$\mathscr{L}_{Q_c^{(w)} \otimes P_\theta^{(n)}}(\theta \mid t) = Q^{(w+n)}_{(w/(w+n))c+(n/(w+n))t}$$

Thus the conjugate weight w and the sample size n are added, and the relative magnitudes $w/(w+n)$ and $n/(w+n)$ determine that mixture of the prior parameter c and the observed response t which yields the posterior parameter $(w/(w+n))c+(n(w+n))t$.

This construction emphasizes the similarities between the families $\mathscr{P}$ and $\mathscr{Q}$. However, the dissimilarities are more intriguing, and surface as soon as regularity of these families is considered.

3. Location interpretation of the conjugate parameter

The family of conjugate distributions $\mathscr{Q}$ is an exponential family of order $k+1$, as pointed out by Barndorff-Nielsen (1978, p. 132). However, little if anything seems to be gained by this fact. Notice that the family of sampling distributions $\mathscr{P}$ need not be exponential, even though it has been constructed as the union of exponential families.

More important, conjugate densities are always log-concave, as mentioned before Lemma 2.1 and used in its proof. Another aspect of this property is that they admit 'nice' integrable majorants based on the Euclidean norm $|\theta|$.

Lemma 3.1. *For every weight* $w>0$ *and for every vector* $c \in \operatorname{int} C$ *there exist scalars* $\alpha \in \mathbb{R}$ *and* $\beta > 0$ *such that*

$$q_c^{(w)}(\theta) \leqslant e^{w\alpha - w\beta|\theta|} \quad \text{for } \theta \in \mathbb{R}.$$

Proof. See Barndorff-Nielsen (1973, p. 194; 1978, p. 93).

The family $\mathscr{Q}^{(w)}$ of conjugate distributions for weight w has the interior of the set C for its canonical parameter domain and hence is regular. No regularity assumption has been made for the family $\mathscr{P}^{(n)}$ of sampling distributions for sample size n, but the boundary behaviour of the conjugate densities characterizes regularity, as follows. Regularity of the sampling family means that the canonical parameter domain Θ is open.

Lemma 3.2. *For every weight* $w>0$ *and for every vector* $c \in \operatorname{int} C$ *the conjugate density* $q_c^{(w)}$ *is finite on the boundary of* Θ; *moreover it vanishes on the boundary if and only if* Θ *is open.*

Proof. Convergence to the boundary splits into the cases when θ tends in norm to infinity or when it tends to a boundary point η of Θ, see Mäkeläinen, Schmidt and Styan (1981, p. 759). In the first case Lemma 3.1 forces $q_c^{(w)}(\theta)$ to converge to zero. In second case we have that $\limsup_{\theta \to \eta} q_c^{(w)}(\theta)$ is bounded by $e^{w\alpha - w\beta|\eta|}$ and hence finite.

More precisely the limit superior is proportional to $(e^{\eta' c - \kappa(\eta)})^w$ and hence vanishes if and only if $\kappa(\eta)$ equals $+\infty$. This prevents η from being a member of Θ. Therefore Θ does not include its boundary, that is, it must be open. Thus the proof is complete.

This lemma is the basis for a short-cut proof of the following result due to Diaconis and Ylvisakar (1979, Theorem 2) who circumvent Lemmas 3.1 and 3.2 by using direct arguments. Their result identifies the conjugate parameter c as the location of the sampling mean-value parameter $\tau(\theta) = \mathrm{E}_{P_\theta^{(n)}}[t]$ when it is averaged with respect to the conjugate distribution $Q_c^{(w)}$.

In fact, an analysis of their proof reveals that the result continues to hold when the conjugate densities are *constant on the boundary*. Their method carries over even to the most general case, resulting in a bias term based on the integral over the derivative $Dq_c^{(w)}(\theta)$ with respect to θ.

Theorem 3.3. *For every conjugate weight $w > 0$ and for every conjugate parameter $c \in \operatorname{int} C$ one has*

$$E_{Q_c^{(w)}}[\tau(\theta)] = c - \frac{1}{w}\int_\Theta Dq_c^{(w)}(\theta)\, d\lambda(\theta);$$

moreover if the conjugate densities are constant on the boundary then

$$E_{Q_c^{(w)}}[\tau(\theta)] = c.$$

Proof. For the purposes of Lebesgue integration over Θ we need only look at the interior of this set. Recall that for interior parameter vectors θ the first and second derivative of the cumulant transformation reproduces the mean-value parameter and the Fisher information matrix, $\tau(\theta) = D\kappa(\theta)$ and $I(\theta) = D^2\kappa(\theta)$.

The gradient of the conjugate density is

$$Dq_c^{(w)}(\theta) = w(c - \tau(\theta))q_c^{(w)}(\theta).$$

Thus the two statements:

- 'The gradient is integrable under Lebesgue measure.'
- 'The mean-value parameter has finite expectation under the conjugate distribution.'

imply each other and entail the first formula in the theorem. We establish Lebesgue integrability of the gradient. Without loss of generality consider its first component, $D_1 q_c^{(w)}(\theta)$, which is the partial derivative with respect to the component θ_1.

Fix the components $\theta_2, \ldots, \theta_k$. The partial function

$$h(\theta_1) = D_1 q_c^{(w)}(\theta) = w(c_1 - \tau_1(\theta))q_c^{(w)}(\theta)$$

has θ_1 varying over an interval from a_1 to b_1, say. Furthermore $h(\theta_1)$ is the derivative of $H(\theta_1) = q_c^{(w)}(\theta)$. But the vector θ with first component $\theta_1 = a_1$ or $\theta_1 = b_1$ lies on the boundary of Θ. Therefore Lemma 3.2 shows that the values $H(a_1)$ and $H(b_1)$ are finite.

The second partial derivative $D_1^2\kappa(\theta)$ is the diagonal element $I_{11}(\theta)$ of the information matrix and hence positive. Thus the first partial derivative $D_1\kappa(\theta) = \tau_1(\theta)$ is increasing in θ_1 and the function $h(\theta_1)$ changes sign at most once, and if so the sign change is from positive to negative. In case there is no sign change one has

$$\int_{a_1}^{b_1} |h(\theta_1)|\, d\theta_1 = |H(b_1) - H(a_1)| \leqslant H(a_1) + H(b_1).$$

In case there is one change of sign in the point c_1, one has

$$\int_{a_1}^{b_1} |h(\theta_1)|\, d\theta_1 = \int_{a_1}^{c_1} h(\theta_1)\, d\theta_1 - \int_{c_1}^{b_1} h(\theta_1)\, d\theta_1 = 2H(c_1).$$

In either case the function h is integrable.

Now allow the components $\theta_2, \ldots, \theta_k$ to vary. Lemma 3.1 provides the bounds

$$H(a_1) + H(b_1) \leqslant e^{w\alpha - w\beta\sqrt{a_1^2 + \theta_2^2 + \quad + \theta_k^2}} + e^{w\alpha - w\beta\sqrt{b_1^2 + \theta_2^2 + \quad + \theta_k^2}} \leqslant 2e^{w\alpha - w\beta\sqrt{0 + \theta_2^2 + \quad + \theta_k^2}},$$

$$2H(c_1) \leqslant 2e^{w\alpha - w\beta\sqrt{c_1^2 + \theta_2^2 + \quad + \theta_k^2}} \leqslant 2e^{w\alpha - w\beta\sqrt{0 + \theta_2^2 + \quad + \theta_k^2}}.$$

The common majorant is integrable with respect to $\theta_2, \ldots, \theta_k$, showing that $D_1 q_c^{(w)}(\theta)$ is integrable with respect to θ. This establishes integrability and the first formula.

Moreover the Fubini Theorem applies and yields

$$\int_\Theta D_1 q_c^{(w)}(\theta)\, d\lambda(\theta) = \int \cdots \int \left(\int_{a_1}^{b_1} h(\theta_1)\, d\theta_1 \right) d\theta_2 \cdots d\theta_k.$$

Since we have already verified integrability of h the inner integral equals $H(b_1) - H(a_1)$ and vanishes when the conjugate densities are constant on the boundary. This establishes the second formula. Thus the proof is complete.

We are now in a position to provide a complete discussion of what can happen in the one-dimensional case. Here Θ is an interval with endpoints a and b, say. The bias term becomes

$$-\frac{1}{w}\int_a^b Dq_c^{(w)}(\theta)\, d\theta = \frac{1}{w}\left(q_c^{(w)}(a) - q_c^{(w)}(b)\right).$$

In view of Lemma 3.2 three cases are possible:

If Θ is open then the bias term vanishes.

If Θ is half-open then the bias term is nonzero.

If Θ is compact then the bias term vanishes if and only if constancy on the boundary obtains.

The monograph of Barndorff-Nielsen (1978) provides a wealth of exponential families falling under the first two cases. The third case is also nonvacuous, since every compact convex set Θ appears as the canonical parameter domain of some exponential family (Brown, 1986, p. 26)

The third case has $q_c^{(w)}(a) = q_c^{(w)}(b)$ and therefore can hold only when the conjugate parameter has value

$$c = \frac{\kappa(a) - \kappa(b)}{a - b}.$$

Thus the bias term vanishes for precisely one value of c and is nonzero otherwise.

Example 3.4. The exponential family generated by the Laplace density $e^{-|t|}/2$ has canonical parameter domain $\Theta = (-1, +1)$, see Barndorff-Nielsen (1978, p. 168). Removing some mass from the tails of the densities, according to $e^{-|t|}/(1 + t^2)$, creates an exponential family with compact parameter domain $\Theta = [-1, +1]$. Due to symmetry one has $\kappa(-1) = \kappa(+1)$, so that here we have $E_{Q_c^{(w)}}[\tau(\theta)] = c$ if and only if $c = 0$.

The approach of integrating over the derivative of the conjugate density is put forward in Morgan (1970) as well. It clearly suggests to also look at the integrals over second (and higher) order derivatives, and to this we turn next.

4. Sample size interpretation of the conjugate weight

The Diaconis and Ylvisakar result is remarkable in that it covers all regular exponential families, which is an important and easily recognized class. This transparency gets lost when higher orders are considered. The Hesse matrix of the conjugate density $q_c^{(w)}(\theta)$ is

$$D^2 q_c^{(w)}(\theta) = w^2(\tau(\theta) - c)(\tau(\theta) - c)' q_c^{(w)}(\theta) - wI(\theta) q_c^{(w)}(\theta).$$

Integrability is no longer self-evident, and the analogue of Theorem 3.3 is burdened by the necessary provisos (leaving hardly anything to prove).

Theorem 4.1. *For every weight $w > 0$ and for every vector $c \in \operatorname{int} C$ any two of the following three statements imply the third:*

- *'The Hesse matrix is integrable under Lebesgue measure.'*
- *'The mean-value parameter has finite variance–covariance matrix under the conjugate distribution.'*
- *'The Fisher information matrix has finite expectation under the conjugate distribution.'*

If these integrability conditions are fulfilled then

$$V_{Q_c^{(w)}}[\tau(\theta)] = \frac{1}{w} E_{Q_c^{(w)}}[I(\theta)] + \frac{1}{w^2} \int_\Theta D^2 q_c^{(w)}(\theta)\, d\lambda(\theta);$$

moreover if the gradient $Dq_c^{(w)}(\theta)$ is constant on the boundary of Θ then

$$V_{Q_c^{(w)}}[\tau(\theta)] = \frac{1}{w} E_{Q_c^{(w)}}[I(\theta)].$$

Proof. Compare the proof of Theorem 3.3.

Integrability of the Hesse matrix holds if and only if the gradient stays finite on the boundary. This may fail even for regular exponential families.

Counterexample 4.2. The univariate logarithmic family (Barndorff-Nielsen, 1978, p. 118) has sampling density

$$p_\theta^{(1)}(t) = \frac{\pi^t}{t} \frac{1}{-\log(1-\pi)} = \frac{1}{t} e^{\theta t - \kappa(\theta)}$$

for $t = 1, 2, \ldots$.

The parameter domains are $(0, 1)$ for π, and $(-\infty, 0)$ for $\theta = \log \pi$. The cumulant transformation is $\kappa(\theta) = \log \log(1/\eta)$, with

$$\eta = 1 - e^\theta = 1 - \pi \in (0, 1).$$

The mean-value parameter

$$\tau(\theta) = (1 - \eta)/(-\eta \log \eta)$$

increases from 1 to ∞ as θ runs from $-\infty$ to 0. The conjugate family has parameter domain

$$\text{int } C = \text{int conv}\{1, 2, \dots\} = (1, \infty),$$

the density with weight $w > 0$ and parameter $c > 1$ being

$$q_c^{(w)}(\theta) = \mathrm{d}(w, c)\left(\mathrm{e}^{\theta c - \kappa(\theta)}\right)^w = \mathrm{d}(w, c)\frac{(1-\eta)^{wc}}{(-\log \eta)^w} \quad \text{for } \theta < 0.$$

The limit of the gradient for $\theta = -\infty$ is $\tau(-\infty)q_c^{(w)}(-\infty) = 1 \cdot 0 = 0$. At $\theta = 0$ we obtain the undetermined expression $\infty \cdot 0$, but a closer analysis leads to

$$\tau(\theta)q_c^{(w)}(\theta) = \mathrm{d}(w, c)\frac{(1-\eta)^{wc+1}}{\eta(-\log \eta)^{w+1}} = \frac{\mathrm{d}(w, c)}{(w+1)^{w+1}} \times \left(\frac{(1-\eta)(wc+1)/(w+1)}{-\eta^{1/(w+1)}\log\left(\eta^{1/(w+1)}\right)}\right)^{w+1}.$$

As θ and hence η tend to 0 these terms are positive except for the numerator which converges to $-0 \log 0 = 0$. In summary we obtain

$$\int_{-\infty}^{0} \mathrm{D}^2 q_c^{(w)}(\theta)\,\mathrm{d}\theta = \mathrm{D}q_c^{(w)}(\theta)\Big|_{-\infty}^{0} = -w\tau(\theta)q_c^{(w)}(\theta)\Big|_{-\infty}^{0} = -\infty.$$

Therefore in the logarithmic family the second derivative of the conjugate density fails to be Lebesgue integrable, implying that expected Fisher information under the conjugate distribution must be infinite.

In general we conjecture that there exists a *minimum weight* $w_1 \in [0, \infty]$ intrinsic to a given family so that for weights $w > w_1$ Theorem 4.1 applies successfully. This is reminiscent of the role played by the sample size: generally some minimum sample size must be exceeded before statements on higher order moments become feasible. We have no proof of our conjecture, but demonstrate it with the following set of examples.

Instances where the minimum weight is zero are met when the sampling family $\mathscr{P}^{(1)}$ is the normal location family, the binomial family, or the Poisson family.

Instances where the minimum weight is positive are the following. The gamma scale family with fixed shape parameter λ_0 has minimum weight $w_1 = 1/\lambda_0$; in particular the normal scale family has minimum weight $w_1 = 2$. In the negative binomial family with fixed shape parameter χ_0 the minimum weight is $w_1 = 1/\chi_0$. The exponential family generated by the Laplace distribution has minimum weight $w_1 = 1$. Of course, when $w = \infty$ then Theorem 4.1 is of no use at all, as is illustrated by the counterexample.

The correspondence between conjugate weight and sample size also suggests to consider the limiting behaviour as w tends to infinity. If the conjugate densities are constant on the boundary then under $Q_c^{(w)}$ the distribution of

$$\sqrt{w}\,(\tau(\theta) - c)$$

is centered at zero; moreover its variance–covariance matrix is expected Fisher information,

$$\mathrm{E}_{Q_c^{(w)}}[I(\theta)] = \mathrm{d}(w, c)\int_{\Theta} I(\theta)\left(\mathrm{e}^{\theta' c - \kappa(\theta)}\right)^w \mathrm{d}\lambda(\theta)$$

when Theorem 4.1 applies successfully. By the Laplace asymptotic method the behaviour of the integral as w tends to infinity is determined by the mode of the density, that is by

$$\sup_{\theta \in \Theta} (\theta' c - \kappa(\theta)) = \kappa^*(c).$$

The value $\hat{\theta}(c)$ that attains this supremum is the maximum likelihood estimate for θ when the observed response is c. In the limit we then obtain a k-variate normal distribution with mean zero and variance–covariance matrix $I(\hat{\theta}(c))$.

The discussion evidently hinges on the regularity properties of conjugate densities and their derivatives, in the sense of whether they are constant on the boundary and whether they are Lebesgue integrable. The exact domain of validity for this type of reasoning remains to be determined.

Acknowledgement

We thank Larry Brown, Persi Diaconis and Brad Efron for valuable discussions on the subject.

References

Barndorff-Nielsen, O. (1973), Unimodality and exponential families, *Communications in Statistics* **1**, 189–216.

Barndorff-Nielsen, O (1978), *Information and Exponential Families in Statistical Theory* (John Wiley & Sons, Chichester).

Brown, L.D. (1986), *Fundamentals of Statistical Exponential Families, with Applications in Statistical Decision Theory*, Institute of Mathematical Statistics Lecture Notes-Monograph Series Volume 9 (Institute of Mathematical Statistics, Hayward, California).

Diaconis, P. and D. Ylvisakar (1979), Conjugate priors for exponential families, *The Annals of Statistics* **7**, 269–281.

Dickey, J (1982), Conjugate families of distributions, in: S. Kotz and N.L. Johnson, eds., *Encyclopedia of Statistical Science*, Vol. 2 (Wiley Interscience, New York) pp. 135–145.

Mäkeläinen, T., K. Schmidt and G.P H. Styan (1981), On the existence and uniqueness of the maximum likelihood estimate of a vector-valued parameter in fixed-sized samples, *The Annals of Statistics* **9**, 758–767

Morgan, R.L. (1970), A class of conjugate prior distributions, Typescript, Department of Statistics, Oregon State University, Corvallis, Oregon 97331, 22 pp

Raiffa, H. and R. Schlaifer (1961), *Applied Statistical Decision Theory* (Devision of Research, Harvard Business School, Boston).