J. Japan Statist. Soc.
Vol. 37 No. 2 2007 307–325

CONJUGATE LOCATION-DISPERSION FAMILIES

Toshio Ohnishi* and Takemi Yanagimoto*

We make a conjugate analysis for the five location-dispersion families including the normal, the transformed gamma and the von Mises distributions. The five families are introduced through the requirement for the existence of conjugate prior densities. We show in a unified way that a Pythagorean relationship holds with respect to posterior risks, which clarifies the optimality of the posterior mode under a Kullback-Leibler loss. An explicit form of the posterior mode is given, and a type of linearity is observed. We construct an empirical Bayes estimator of a location vector explicitly.

Key words and phrases: Addition identity, conjugate prior, empirical Bayes estimator, Kullback-Leibler separator, location-dispersion family, posterior mode, Pythagorean relationship.

1. Introduction

A conjugate prior density, when it exists, provides us with a convenient tool for the Bayesian estimation problem. A primary interest has been taken in the conjugate analysis of the mean parameter of the natural exponential family. This is probably because a conjugate prior can be naively defined under mild regularity conditions.

In this paper we pursue a conjugate analysis of the location parameter. There are a considerable number of existing works on Bayesian analysis of the location family, which include Mardia and El-Atoum (1976), Diaconis and Ylvisaker (1985), Spiegelhalter (1985), Polson (1991), Bischoff (1993), Angers (1996), Garvan and Ghosh (1997), Leblanc and Angers (1999), Rodrigues *et al.* (2000) and Ohnishi and Yanagimoto (2003). Among them, Diaconis and Ylvisaker (1985) is one which discussed formally a conjugate prior for the location family. Their paper, however, concentrated on the analytical aspect; no explicit form of a conjugate prior density was given. Mardia and El-Atoum (1976) noted that a von Mises prior density is conjugate for a von Mises sampling density, which was followed by Rodrigues *et al.* (2000).

Some researchers are critical of the use of a conjugate prior since recent numerical development permits us to assume a general prior. In our view, the role of a conjugate prior corresponds to that of the exponential family with a quadratic variance function (Morris (1983b)) or that of the Tweedie distribution, i.e., the exponential family with a power variance function (Jørgensen (1997), Chapter 4). In fact, a statistical model having an error distribution of an analytically convenient form is useful in practice, though numerical computations are possible under a general error distribution. Our conjecture is that the search of the lo-

Received October 27, 2006. Revised March 13, 2007. Accepted April 9, 2007.
*Institute of Statistical Mathematics, 4-6-7 Minami-Azabu, Minato-ku, Tokyo 106-8569, Japan. Email: ohnishi@ism.ac.jp

cation family admitting a conjugate prior leads us to a practically useful model, and we obtain an affirmative result. That is, the derived density in the location family is a conjugate prior for the generalized linear model based on the Tweedie distribution as discussed in Ohnishi (2006).

The definition of a conjugate prior is something complicated, as discussed in Diaconis and Ylvisaker (1979) and Huang and Bier (1999). We can find two definitions in the literature. One is that a prior density is closed under sampling, that is, the posterior density has the same form as a prior density. The other is that the posterior mean of the parameter of interest is of linear form. In this study, we call a prior density conjugate when it is closed under sampling. Although the induced estimator of the location parameter is not always of linear form, we will find that a type of linearity is observed in all the cases. Rodrigues *et al.* (2000) noted this fact in the von Mises case.

Our interest will be focused on a location-dispersion family whose density is given by

$$p(x-\mu;\tau) = \exp\{-\tau d(x-\mu) + a(\tau)\}, \tag{1.1}$$

where μ and $\tau > 0$ are the location and the dispersion parameters, respectively, $d(t)$ is a non-negative function, and $\exp\{-a(\tau)\}$ is the normalizing constant. See Jørgensen (1997, p. 17) for a review of this family, where the function $d(t)$ is called the (unit) deviance function. This family presents us a general form of location families, which covers most of the existing ones such as the normal and the von Mises distributions.

We will seek a conjugate prior density for (1.1), though a Bayesian approach in relation to a noninformative prior was discussed by Garvan and Ghosh (1997). The prior density we assume has the form

$$\pi(\mu - m;\delta) = \exp\{-\delta d(m-\mu) + a(\delta)\} = p(m-\mu;\delta), \tag{1.2}$$

where m and δ are hyperparameters. An advantage of this prior density is that the normalizing constant depends only on δ. Our problem is to determine an explicit form of $d(t)$ such that (1.2) is conjugate for (1.1).

The organization of the rest of this paper is as follows. In Section 2, a differential equation is characterized by the conjugacy condition. By solving the differential equation, we present the five location-dispersion families allowing for a conjugate analysis. Section 3 shows three properties of these five location-dispersion families in a unified way. One of them is an addition identity which the deviance function satisfies. In Section 4 we discuss the Bayes estimation under a conjugate prior density. The Kullback-Leibler separator is adopted as a loss. Deriving a Pythagorean relationship with respect to posterior risks, we prove the optimality of the posterior mode. Section 5 proposes an empirical Bayes estimation of the location vector. In the final section, we discuss the conjugacy condition of prior densities. Also an interpretation of the addition identity in Section 3 is given from a viewpoint of statistical physics.

2. Conjugate location-dispersion families

In this section we introduce the five families of sampling densities each of which has a conjugate prior density. A differential equation is obtained as a necessary condition for the existence of a conjugate prior density. The five families are derived through its solution.

We deal with two cases of location-dispersion families: One is a family on $\mathbb{R}$ and the other is that on $I = [0, 2\pi)$. Our discussion includes the von Mises distribution which is defined on I. Let us introduce the following two families of (unit) deviance functions as

$$\mathcal{D}_{\mathbb{R}} = \left\{ d(t) \in \mathcal{D} \,\middle|\, \begin{array}{l} \int_{\mathbb{R}} \exp\{-\tau d(t)\} dt \text{ exists for any } \tau \in \mathbb{R}^+ \\ \lim_{t\to\pm\infty} d'(t) \exp\{-\tau d(t)\} = 0 \text{ for any } \tau \in \mathbb{R}^+ \end{array} \right\}, \tag{2.1a}$$

$$\mathcal{D}_{I} = \left\{ d(t) \in \mathcal{D} \,\middle|\, \begin{array}{l} \int_{I} \exp\{-\tau d(t)\} dt \text{ exists for any } \tau \in \mathbb{R}^+ \\ d(t) = d(t+2\pi) \text{ for any } t \in \mathbb{R} \end{array} \right\}, \tag{2.1b}$$

where

$$\mathcal{D} = \{d(t) \geq 0 \mid d(0) = d'(0) = 0 \text{ and } d''(0) = 1\}.$$

The latter condition in each of (2.1a) and (2.1b) guarantees

$$\int_{\mathbb{K}} \{-\tau d''(t)\} \exp\{-\tau d(t)\} dt = \int_{\mathbb{K}} \{\tau d'(t)\}^2 \exp\{-\tau d(t)\} dt,$$

where $\mathbb{K}$ denotes either $\mathbb{R}$ or I, and we shall use this notation throughout the present paper. The above equality is the so-called information unbiasedness and will be seen in the proof of Proposition 3.2.

Consider the sampling density (1.1) on $\mathbb{K}$ having a location parameter $\mu \in \mathbb{K}$ and a dispersion parameter $\tau \in \mathbb{R}^+$. The family of these densities is said to be the location-dispersion family generated by the deviance function $d(t)$. It follows from (2.1a) or (2.1b) that if $d(t)$ is in $\mathcal{D}_{\mathbb{K}}$ then $d(-t)$ is also in $\mathcal{D}_{\mathbb{K}}$.

We assume the prior density (1.2) for the location parameter μ in (1.1). This prior density belongs to the location-dispersion family generated by $d(-t)$. We investigate a necessary condition that the prior density (1.2) is conjugate.

Proposition 2.1. *Suppose that the prior density* (1.2) *is conjugate for the sampling density* (1.1). *Then, the following differential equation*

$$d''(t) = \alpha d(t) + \beta d'(t) + 1 \tag{2.2}$$

holds where α and β are some constants.

Proof. Set

$$g(\mu)(= g(\mu; x, m, \tau, \delta)) = \tau d(x-\mu) + \delta d(m-\mu). \tag{2.3}$$

Since the prior density (1.2) is conjugate, $g(\mu)$ is expressed as

$$g(\mu) = ad(b-\mu) + c, \tag{2.4}$$

where a, b and c are independent of μ. It follows from (2.3) that $\partial^2 g/\partial\tau\partial\delta = 0$. Differentiation of (2.4) with respect to τ and δ gives

$$Ad''(b-\mu) + Bd'(b-\mu) + Cd(b-\mu) + D = 0, \tag{2.5}$$

where

$$A = a\frac{\partial b}{\partial\tau}\frac{\partial b}{\partial\delta}, \qquad B = \frac{\partial a}{\partial\tau}\frac{\partial b}{\partial\delta} + \frac{\partial a}{\partial\delta}\frac{\partial b}{\partial\tau} + a\frac{\partial^2 b}{\partial\tau\partial\delta}, \qquad C = \frac{\partial^2 a}{\partial\tau\partial\delta} \quad \text{and}$$
$$D = \frac{\partial^2 c}{\partial\tau\partial\delta}.$$

Set $\mu = b$ in (2.5), and we have $D = -A$. Here we used the condition in the definition of $\mathcal{D}$. Differentiating both the sides of (2.5) with respect to μ and setting $\mu = b$, we obtain $B = -d^{(3)}(0)A$. Similarly, we get $d^{(4)}(0)A + d^{(3)}(0)B + C = 0$, therefore $C = [\{d^{(3)}(0)\}^2 - d^{(4)}(0)]A$. Setting $\alpha = -C/A = d^{(4)}(0) - \{d^{(3)}(0)\}^2$ and $\beta = -B/A = d^{(3)}(0)$, we obtain

$$d''(b-\mu) = \alpha d(b-\mu) + \beta d'(b-\mu) + 1.$$

Since this equality holds for any μ, the differential equation (2.2) is derived. □

The equation determining b is given by $\tau d'(x-b) + \delta d'(m-b) = 0$. If $g(\mu)$ has its minimum, the posterior mode $\hat{\mu}_{\text{map}}$ exists and satisfies this equation. Thus we write $b = \hat{\mu}_{\text{map}}$. We can also obtain the expression of a and c respectively as $g''(\hat{\mu}_{\text{map}})$ and $g(\hat{\mu}_{\text{map}})$.

Note that the solution to the differential equation (2.2) is not always in $\mathcal{D}_{\mathbb{K}}$, i.e., the function (1.1) with $d(t)$ being the solution to (2.2) is not always a density. The following proposition gives all the solutions in $\mathcal{D}_{\mathbb{R}}$ or $\mathcal{D}_I$. The proof is given in Appendix.

Proposition 2.2.

(i) *The deviance functions in* $\mathcal{D}_{\mathbb{R}}$ *which are solutions to the differential equation* (2.2) *are expressed as*

$$d(t) = \frac{t^2}{2}, \tag{2.6a}$$

$$d(t) = \frac{1}{\kappa}\left(\frac{e^{\kappa t}}{\kappa} - t - \frac{1}{\kappa}\right), \tag{2.6b}$$

$$d(t) = \frac{1}{\kappa+\gamma}\left(\frac{e^{\kappa t}}{\kappa} + \frac{e^{-\gamma t}}{\gamma} - \frac{1}{\kappa} - \frac{1}{\gamma}\right), \tag{2.6c}$$

$$d(t) = \frac{1}{\kappa}\left(\frac{e^{-\kappa t}}{\kappa} + t - \frac{1}{\kappa}\right), \tag{2.6d}$$

where κ and γ are positive constants.

(ii) *The deviance function in $\mathcal{D}_I$ which is a solution to the differential equation* (2.2) *is expressed as*

$$d(t) = \frac{1-\cos(\xi t)}{\xi^2}, \tag{2.6e}$$

where ξ is a positive integer.

It is interesting that the location-dispersion family generated by (2.6b), (2.6c) or (2.6d) can be used as a conjugate prior density for a generalized linear model. In fact, Ohnishi (2006) discussed a conjugate analysis of the logarithmic link regression model based on the Tweedie distribution (Jørgensen (1997), Chapter 4). It is our understanding that the conjugacy condition aids us with the search for a suitable likelihood function.

For notational convenience, let us write the set of the deviance functions in Proposition 2.2 as $\mathcal{D}_c$, i.e., $\mathcal{D}_c = \{(2.6\text{a}), (2.6\text{b}), (2.6\text{c}), (2.6\text{d}), (2.6\text{e})\}$. Correspondingly, we give the following definition:

DEFINITION 2.1. *Let $\mathcal{P}_c$ denote the set consisting of the five location-dispersion families generated by the five deviance functions in $\mathcal{D}_c$. Each member of $\mathcal{P}_c$ is said to be a conjugate location-dispersion family.*

Now, we prove that each member of $\mathcal{P}_c$ has a conjugate prior density by using the differential equation (2.2).

PROPOSITION 2.3. *Suppose that $d(t) \in \mathcal{D}_c$. Then, the posterior density corresponding to the prior density* (1.2) *is expressed as*

$$\begin{aligned} \pi(\mu \mid x; \tau, m, \delta) &= \exp[-g''(\hat{\mu}_{\text{map}}) d(\hat{\mu}_{\text{map}} - \mu) + a(g''(\hat{\mu}_{\text{map}}))] \\ &= \pi(\mu - \hat{\mu}_{\text{map}}; g''(\hat{\mu}_{\text{map}})), \end{aligned} \tag{2.7}$$

where $g(\mu)$ is the function (2.3) *and $\hat{\mu}_{\text{map}}$ is the posterior mode. Thus, the prior density* (1.2) *is conjugate for the sampling density* (1.1).

PROOF. Since the function $d(t)$ satisfies the differential equation (2.2), a mathematical induction shows that for each $k \geq 2$ there exist two constants α_k and β_k such that

$$d^{(k)}(t) = \alpha_k d(t) + \beta_k d'(t) + d^{(k)}(0). \tag{2.8}$$

It follows that $d^{(k)}(t) = \alpha_{k-1} d'(t) + \beta_{k-1} d''(t)$ for $k \geq 3$. This, together with the definition of $\hat{\mu}_{\text{map}}$, implies that

$$\begin{aligned} g^{(k)}(\hat{\mu}_{\text{map}}) &= (-1)^k \{\tau d^{(k)}(x - \hat{\mu}_{\text{map}}) + \delta d^{(k)}(m - \hat{\mu}_{\text{map}})\} \\ &= (-1)^k \{-\alpha_{k-1} g'(\hat{\mu}_{\text{map}}) + \beta_{k-1} g''(\hat{\mu}_{\text{map}})\} \\ &= (-1)^k \beta_{k-1} g''(\hat{\mu}_{\text{map}}). \end{aligned}$$

Expanding $g(\mu)$ in the Taylor series around $\mu = \hat{\mu}_{\mathrm{map}}$, we have

$$
\begin{aligned}
g(\mu) &= g(\hat{\mu}_{\mathrm{map}}) + \frac{1}{2!}g''(\hat{\mu}_{\mathrm{map}})(\mu - \hat{\mu}_{\mathrm{map}})^2 + \sum_{k\geq 3}\frac{1}{k!}g^{(k)}(\hat{\mu}_{\mathrm{map}})(\mu - \hat{\mu}_{\mathrm{map}})^k \\
&= g(\hat{\mu}_{\mathrm{map}}) + g''(\hat{\mu}_{\mathrm{map}})\left\{\frac{1}{2!}(\hat{\mu}_{\mathrm{map}} - \mu)^2 + \sum_{k\geq 3}\frac{\beta_{k-1}}{k!}(\hat{\mu}_{\mathrm{map}} - \mu)^k\right\}.
\end{aligned}
$$

Since $d^{(k)}(0) = \alpha_{k-1}d'(0) + \beta_{k-1}d''(0) = \beta_{k-1}$ for $k \geq 3$, the Taylor series of $d(t)$ is given by

$$
d(t) = \frac{1}{2}t^2 + \sum_{k\geq 3}\frac{\beta_{k-1}}{k!}t^k. \tag{2.9}
$$

Thus we obtain the expression

$$
g(\mu) = g''(\hat{\mu}_{\mathrm{map}})d(\hat{\mu}_{\mathrm{map}} - \mu) + g(\hat{\mu}_{\mathrm{map}}) \tag{2.10}
$$

as in (2.4), which completes the proof. □

Let us present explicitly the five conjugate location-dispersion families. We obtain (i) a normal, (ii) a gamma, (iii) a hyperbola, (iv) an inverted gamma and (v) a von-Mises family.

Example 2.1. Let $\mathcal{N}$ denote the location-dispersion family generated by (2.6a), which is the family of normal densities.

Example 2.2. We call the location-dispersion family generated by (2.6b) the gamma family and write it as $\mathcal{G}(\kappa)$. Let the random variable y have a gamma density

$$
\frac{\tau^\tau}{\Gamma(\tau)\theta^\tau}y^{\tau-1}\exp\left(-\frac{\tau y}{\theta}\right). \tag{2.11}
$$

Then $x = \log y$ has the density $\{\tau^\tau/\Gamma(\tau)\}\exp\{\tau(x - \mu - e^{x-\mu})\}$ with $\mu = \log\theta$. Thus we find that $\mathcal{G}(1)$ is the family of log-transformed gamma densities.

Example 2.3. The location-dispersion family generated by (2.6c) is called the hyperbola family, and the symbol $\mathcal{H}(\kappa, \gamma)$ is used. As presented in Barndorff-Nielsen (1978) and Jensen (1981), a density of the hyperbola distribution is given by

$$
\frac{1}{2K_0(\tau)}e^{-\tau\cosh(x-\mu)}, \tag{2.12}
$$

where $K_0(\tau)$ stands for the modified Bessel function of the third kind with index zero. The family of densities (2.12) is the location-dispersion family $\mathcal{H}(1, 1)$. Hence we use the term ‘hyperbola family.’

Example 2.4. We call the location-dispersion family generated by (2.6d) the inverted gamma family and use the symbol $\mathcal{IG}(\kappa)$. The variable transformation $s = -t$ in (2.6d) leads to (2.6b). It is found that $\mathcal{IG}(1)$ is the family of log-transformed inverted gamma densities. It should be noted that the location-scale family generated by (2.6d) with $\kappa = 1$ is the Gumbel distribution.

Example 2.5. Let $\mathcal{M}(\xi)$ denote the location-dispersion family generated by (2.6e). A member of the family $\mathcal{M}(1)$ is a von Mises density

$$\frac{1}{2\pi I_0(\tau)} e^{\tau\cos(x-\mu)}, \tag{2.13}$$

where $I_0(\tau)$ is the modified Bessel function of the first kind with index zero. We call $\mathcal{M}(\xi)$ the von Mises family. Analogies between the von Mises density (2.13) and the hyperbola density (2.12) were suggested in Barndorff-Nielsen (1978) and Jensen (1981). Here we note that the dispersion parameter τ can be zero or even be negative in the von Mises family. When $\tau = 0$, the density is that of the uniform distribution on I. A negative value of τ can be interpreted as follows. Set $\xi = 1$ for simplicity. Since $\tau\cos(x-\mu) = -\tau\cos(x-\mu-\pi)$, the two points in the parameter space, (μ, τ) and $(\mu+\pi, -\tau)$, mean the identical density. A negative dispersion parameter τ implies a shift in the location parameter μ. This fact will be recalled in Example 5.5 in Section 5.

3. Properties of the conjugate location-dispersion families $\mathcal{P}_c$

Three properties of the conjugate location-dispersion families are obtained in a unified way. A treatment using the differential equation (2.2) aids our deeper understanding of conjugate prior densities in $\mathcal{P}_c$. In this and subsequent sections we will assume that $d(t)$ is in the set $\mathcal{D}_c$ which is introduced in the previous section.

First, we show that an addition identity holds for $d(t)$.

Proposition 3.1. *The following addition identity*

$$d(s+t) = d(s)\tilde{d}(t) + d'(s)d'(t) + d(t) \tag{3.1}$$

holds where $\tilde{d}(t)$ is expressed as

$$\tilde{d}(t) = d''(t) - \beta d'(t) = 1 + \alpha d(t) \tag{3.2}$$

with α and β being constant coefficients in (2.2).

Proof. Expanding $d(s+t)$ in the Taylor series around $t = 0$ and using (2.8), we obtain

$$d(s+t) = d(s) + d'(s)t + \sum_{k\geq 2} \frac{d^{(k)}(s)}{k!} t^k$$

$$= d(s)\left(1 + \sum_{k\ge 2} \frac{\alpha_k}{k!} t^k\right) + d'(s)\left(t + \sum_{k\ge 2} \frac{\beta_k}{k!} t^k\right) + d(t).$$

First, let us show that $d''(t) - \beta d'(t) = 1 + \sum_{k\ge 2} \alpha_k t^k/k!$. Note that $d''(0) - \beta d'(0) = 1$ and $d'''(0) - \beta d''(0) = 0$. This is seen from the differential equation (2.2) and the definition of $\mathcal{D}$. Differentiating both sides of (2.8) and setting $t = 0$, we have $d^{(k+1)}(0) = \beta_k$ for $k \ge 2$. Similarly, we get from (2.2) that $d^{(k+2)}(0) = \alpha_k + \beta_k d'''(0)$ for $k \ge 2$. Noting that $d'''(0) = \beta$, we obtain $d^{(k+2)}(0) - \beta d^{(k+1)}(0) = \alpha_k$ for $g \ge 2$. Thus, the required Taylor expansion is derived. Secondly, the differentiation of both sides of (2.9) gives that $d'(t) = t + \sum_{k\ge 2} \beta_k t^k/k!$. Finally, the second equality in (3.2), another expression of $\tilde{d}(t)$, is derived directly through (2.2). $\square$

The explicit forms of the function $\tilde{d}(t)$ and the coefficients α and β are given in Table 1. Four familiar examples of the addition identity (3.1) are presented as follows.

(1) $\frac{1}{2}(x+y)^2 = \frac{1}{2}x^2 + xy + \frac{1}{2}y^2$,

(2) $e^{x+y} - (x+y) - 1 = e^x - x - 1 + (e^x - 1)(e^y - 1) + e^y - y - 1$,

(3) $1 - \cos(x+y) = (1 - \cos x)\cos y + \sin x \sin y + 1 - \cos y$,

(4) $\cosh(x+y) - 1 = (\cosh x - 1)\cosh y + \sinh x \sinh y + \cosh y - 1$.

The last two are essentially the same as the addition formulas for the cosine function and the hyperbolic cosine function (Abramowitz and Stegun (1974), p. 72 and p.83).

The identity (3.1) will be used in proving the optimality of the posterior mode. In fact, it makes simpler the proof of Proposition 4.1, as will be seen in the next section. Note that the identity (3.1) can be interpreted by using terms of the statistical physics. This interesting fact will be discussed in the final section.

Secondly, we calculate the Kullback-Leibler separator for $p(x-\mu;\tau)$ in $\mathcal{P}_c$.

Proposition 3.2. *The Kullback-Leibler separator* $\mathrm{D}_\tau(\mu_1, \mu_2)$ *from* $p(x - \mu_1; \tau)$ *to* $p(x - \mu_2; \tau)$ *is expressed as*

$$\mathrm{D}_\tau(\mu_1, \mu_2) = b(\tau) d(\mu_1 - \mu_2), \tag{3.3}$$

Table 1. Explicit forms of $\tilde{d}(t)$ and the coefficients in (2.2).

Family	$d(t)$	$\tilde{d}(t)$	α	β
Normal $\mathcal{N}$	(2.6a)	1	0	0
Gamma $\mathcal{G}(\kappa)$	(2.6b)	1	0	κ
Hyperbola $\mathcal{H}(\kappa,\gamma)$	(2.6c)	$\frac{\kappa\gamma}{\kappa+\gamma}\left(\frac{e^{\kappa t}}{\kappa} + \frac{e^{-\gamma t}}{\gamma}\right)$	$\kappa\gamma$	$\kappa - \gamma$
Inverted gamma $\mathcal{IG}(\kappa)$	(2.6d)	1	0	$-\kappa$
von Mises $\mathcal{M}(\xi)$	(2.6e)	$\cos(\xi t)$	$-\xi^2$	0

where $b(\tau)$ is the Fisher information due to the location parameter μ.

PROOF. The regularity conditions in (2.1a) or (2.1b) affirm the following equalities:

$$\int_{\mathbb{K}} d'(t)\exp\{-\tau d(t)+a(\tau)\}dt = 0, \tag{3.4}$$
$$\int_{\mathbb{K}} \{\tau d'(t)\}^2 \exp\{-\tau d(t)+a(\tau)\}dt = \int_{\mathbb{K}} \tau d''(t)\exp\{-\tau d(t)+a(\tau)\}dt.$$

When deriving the former equality, we used the fact that the density vanishes at $t=\pm\infty$ in the case of $\mathcal{D}_{\mathbb{R}}$ or has the same value both at $t=0$ and $t=2\pi$ in the case of $\mathcal{D}_I$. In the latter equality we used the integration by parts, and this is the Fisher information $b(\tau)$ of the density $\exp\{-\tau d(x-\mu)+a(\tau)\}$. The expression (3.2) together with these two equalities gives

$$\int_{\mathbb{K}} \tilde{d}(t)\exp\{-\tau d(t)+a(\tau)\}dt = \frac{b(\tau)}{\tau}. \tag{3.5}$$

Integrating by substituting $t=x-\mu_1$ and using the addition identity (3.1), we have an expression of $\mathrm{D}_\tau(\mu_1,\mu_2)$ as

$$\begin{aligned}\mathrm{D}_\tau(\mu_1,\mu_2) &= \tau\int_{\mathbb{K}}\{d(\mu_1-\mu_2+t)-d(t)\}\exp\{-\tau d(t)+a(\tau)\}dt\\ &= \tau\int_{\mathbb{K}}\{d(\mu_1-\mu_2)\tilde{d}(t)+d'(\mu_1-\mu_2)d'(t)\}\exp\{-\tau d(t)+a(\tau)\}dt.\end{aligned}$$

Therefore the expression (3.3) is derived from (3.4) and (3.5). □

This proposition will be used to derive a Pythagorean relationship which holds in $\mathcal{P}_c$ in Section 5.

The Kullback-Leibler separator from $\exp\{-\tau_1 d(x-\mu_1)+a(\tau_1)\}$ to $\exp\{-\tau_2 d(x-\mu_2)+a(\tau_2)\}$ is

$$\mathrm{D}((\mu_1,\tau_1),(\mu_2,\tau_2)) = a(\tau_1)-a(\tau_2)-a'(\tau_1)\left(1-\frac{\tau_2}{\tau_1}\right)+\frac{\tau_2}{\tau_1}b(\tau_1)d(\mu_1-\mu_2).$$

Compare the two estimators $(\hat{\mu}_1,\hat{\tau}_1)$ and $(\hat{\mu}_2,\hat{\tau}_1)$ where the estimator of τ is common. Then, we have

$$\mathrm{D}((\hat{\mu}_1,\hat{\tau}_1),(\mu,\tau))-\mathrm{D}((\hat{\mu}_2,\hat{\tau}_1),(\mu,\tau)) = \frac{\tau}{\hat{\tau}_1}b(\hat{\tau}_1)\{d(\hat{\mu}_1-\mu)-d(\hat{\mu}_2-\mu)\}.$$

As long as we estimate τ by $\hat{\tau}_1$, the loss function is essentially $d(\hat{\mu}-\mu)$. This fact reflects a certain orthogonality between μ and τ. Although this argument is not sufficient since the posterior mode contains τ, it seems natural to start from the estimation problem under the loss $d(\hat{\mu}-\mu)$.

Thirdly, we show that each density in the conjugate location-dispersion family has a maximum entropy property. The key of the proof is the addition identity (3.1) in Proposition 3.1.

PROPOSITION 3.3. *The density* (1.1) *is the one with the maximum entropy under the condition that the means of* $d(x)$ *and* $d'(x)$ *are given.*

PROOF. The addition identity (3.1) gives another expression of $p(x-\mu;\tau)$ as

$$\exp\{-\tau\tilde{d}(-\mu)d(x)-\tau d'(-\mu)d'(x)-\tau d(-\mu)+a(\tau)\}.$$

By applying Theorem 13.2.1 in Kagan *et al.* (1973), we obtain the required result. □

4. Conjugate analysis in $\mathcal{P}_c$

We discuss the Bayes estimation in $\mathcal{P}_c$ assuming a conjugate prior density. The optimality of the posterior mode under the Kullback-Leibler loss is shown in a unified way by using a Pythagorean relationship with respect to posterior risks. Since the location and the dispersion parameters are orthogonal with each other, we will focus on the estimation of the location parameter and assume that the dispersion parameter is known.

We adopt $d(\hat{\mu}-\mu)$ as a loss, though the squared error loss $(\hat{\mu}-\mu)^2$ is often used in the Bayes estimation. It is equivalent to the Kullback-Leibler loss $\mathrm{D}_\tau(\hat{\mu},\mu)$. Recall the expression (3.3) and that the Fisher information $b(\tau)$ is positive. The Kullback-Leibler loss is regarded as one of the intrinsic losses by Robert (2001, p. 82). This adoption seems natural especially in the von Mises family $\mathcal{M}(1)$ where $d(\hat{\mu}-\mu)=1-\cos(\hat{\mu}-\mu)$.

The following proposition states that the posterior mode satisfies a Pythagorean relationship with respect to posterior risks. The optimality of the posterior mode follows as its corollary. In this sense the Pythagorean relationship makes it clear how the posterior mode is superior to the other estimators.

PROPOSITION 4.1.

(i) *The posterior mode* $\hat{\mu}_{\mathrm{map}}$ *satisfies the equality*

$$\mathrm{E}_{\mathrm{post}}[d(\hat{\mu}-\mu)-d(\hat{\mu}-\hat{\mu}_{\mathrm{map}})\tilde{d}(\hat{\mu}_{\mathrm{map}}-\mu)-d(\hat{\mu}_{\mathrm{map}}-\mu)]=0 \tag{4.1}$$

for any estimator $\hat{\mu}$ *having a finite posterior risk, where* $\mathrm{E}_{\mathrm{post}}[\cdot]$ *denotes the posterior expectation and* $\tilde{d}(t)$ *is the function* (3.2).

(ii) *It holds that*

$$\mathrm{E}_{\mathrm{post}}[\tilde{d}(\hat{\mu}_{\mathrm{map}}-\mu)]=\frac{b(g''(\hat{\mu}_{\mathrm{map}}))}{g''(\hat{\mu}_{\mathrm{map}})}>0,$$

where $b(\tau)$ *is the Fisher information and* $g(\mu)$ *is the function defined by* (2.3).

PROOF. (i) Setting $s=\hat{\mu}-\hat{\mu}_{\mathrm{map}}$ and $t=\hat{\mu}_{\mathrm{map}}-\mu$ in (3.1), we have

$$\begin{aligned}&\mathrm{E}_{\mathrm{post}}[d(\hat{\mu}-\mu)-d(\hat{\mu}-\hat{\mu}_{\mathrm{map}})\tilde{d}(\hat{\mu}_{\mathrm{map}}-\mu)-d(\hat{\mu}_{\mathrm{map}}-\mu)]\\&\quad=d'(\hat{\mu}-\hat{\mu}_{\mathrm{map}})\mathrm{E}_{\mathrm{post}}[d'(\hat{\mu}_{\mathrm{map}}-\mu)].\end{aligned}$$

The regularity conditions in the definition (2.1a) or (2.1b) of $\mathcal{D}_{\mathbb{K}}$ gives the equality

$$\int_{\mathbb{K}} g'(\mu)e^{-g(\mu)}d\mu = 0. \tag{4.2}$$

Here we employed a similar calculation to that we had used when deriving (3.4). Differentiating both sides of (2.10) with respect to μ, we have

$$g'(\mu) = -g''(\hat{\mu}_{\text{map}})d'(\hat{\mu}_{\text{map}} - \mu). \tag{4.3}$$

Note that $g''(\hat{\mu}_{\text{map}})$ is positive and is independent of μ and also that the posterior density is proportional to $\exp\{-g(\mu)\}$. It follows from (4.2) and (4.3) that

$$\mathrm{E}_{\text{post}}[d'(\hat{\mu}_{\text{map}} - \mu)] = 0,$$

which yields the equality (4.1).

(ii) The representation (2.7) of the posterior density together with (3.5) completes the proof. □

COROLLARY 4.2. *The posterior mode* $\hat{\mu}_{\text{map}}$ *is optimum under the Kullback-Leibler loss* $\mathrm{D}_\tau(\hat{\mu}, \mu)$.

Recall that the posterior mode has a certain optimum property. See Corollary 1.2 in Lehmann and Casella (1998, Chapter 4). We find that the posterior mode is optimal in the two senses in $\mathcal{P}_c$.

Yanagimoto and Ohnishi (2005b) proposed a modified version of the posterior mode, which is invariant with respect to parameter transformation. This estimator, called the standardized posterior mode, is derived by discarding the Jacobian factor. In the location family case the standardized posterior mode coincides with the original posterior mode.

We conclude this section with five examples. In each example we make a comment on the assumed conjugate prior density and give an explicit form of the posterior mode. In order to emphasize the optimality of the estimator we will write $\hat{\mu}_{\mathrm{B}}$ in place of $\hat{\mu}_{\text{map}}$. Note that the linearity of $\hat{\mu}_{\mathrm{B}}$ is observed in all examples, though some modification may be necessary.

Example 4.1. Normal family $\mathcal{N}$;

The assumed conjugate prior density is a normal prior one, a usual choice. We have the well-known result $\hat{\mu}_{\mathrm{B}} = (\tau x + \delta m)/(\tau + \delta)$.

Example 4.2. Gamma family $\mathcal{G}(1)$;

A familiar choice of a prior density on the mean parameter θ of a gamma density (2.11) is that $1/\theta$ has a gamma density. See Carlin and Louis (2000, p. 86) and Robert (2001, p. 177) for examples. Our prior density for $\mathcal{G}(1)$ is in $\mathcal{IG}(1)$ and essentially the same as the familiar one. The optimal estimator is given by $\hat{\mu}_{\mathrm{B}} = \log\{(\tau e^x + \delta e^m)/(\tau + \delta)\}$.

Example 4.3. Hyperbola family $\mathcal{H}(1,1)$;

The assumed prior density has the same form as the sampling density (2.12). As will be seen in Example 4.5, there is a striking analogy between this and the von Mises cases. Using the addition formula for the hyperbolic cosine function, we obtain

$$\hat{\mu}_{\mathrm{B}} = \sinh^{-1}\left[\frac{\tau\sinh x + \delta\sinh m}{\{\tau^2+\delta^2+2\tau\delta\cosh(x-m)\}^{1/2}}\right].$$

In the general case $\mathcal{H}(\kappa,\gamma)$ the Bayes estimator satisfies $d'(\hat{\mu}_{\mathrm{B}})/\tilde{d}(\hat{\mu}_{\mathrm{B}}) = (\tau d'(x) + \delta d'(m))/(\tau\tilde{d}(x)+\delta\tilde{d}(m))$.

Example 4.4. Inverted gamma family $\mathcal{IG}(1)$;

The assumed prior density is in $\mathcal{G}(1)$. The optimal estimator is given by $\hat{\mu}_{\mathrm{B}} = \log\{(\tau+\delta)/(\tau e^{-x}+\delta e^{-m})\}$, which is to be compared with Example 4.2.

Example 4.5. von-Mises family $\mathcal{M}(1)$;

This example is the result in Mardia and El-Atoum (1976). Our prior density for $\mathcal{M}(1)$ is also in $\mathcal{M}(1)$. The von Mises prior density was employed in Guttorp and Lockhart (1988), Bagchi (1994) and Rodrigues *et al.* (2000). We obtain

$$\hat{\mu}_{\mathrm{B}} = \begin{cases} \sin^{-1}\left[\dfrac{\tau\sin x+\delta\sin m}{\{\tau^2+\delta^2+2\tau\delta\cos(x-m)\}^{1/2}}\right] & \text{if} \quad \tau\cos x+\delta\cos m \ge 0, \\ \pi - \sin^{-1}\left[\dfrac{\tau\sin x+\delta\sin m}{\{\tau^2+\delta^2+2\tau\delta\cos(x-m)\}^{1/2}}\right] & \text{otherwise,} \end{cases}$$

where the range of the arc sine function is chosen as $[-\pi/2,\pi/2]$. The addition formula for the cosine function was used to obtain the above result. This expression is analogous to the latter part of Example 4.3.

5. Estimation of the location vector in $\mathcal{P}_c$

The estimation of the vector parameter is becoming increasingly important in the analysis of models with high-dimensional vector parameters. We construct an estimator in each member of $\mathcal{P}_c$ through the parametric empirical Bayes method (Morris (1983a)). The key is a mean Pythagorean relationship in applying the method of moments.

The following mean Pythagorean relationship is known to hold for a density $p_e(x;\mu)$ with mean μ in an exponential family:

$$\begin{aligned}\mathrm{E}[\mathrm{D}(p_e(y;x),p_e(y;m)) - \mathrm{D}(p_e(y;x),p_e(y;\mu))\\ - \mathrm{D}(p_e(y;\mu),p_e(y;m)) \mid p_e(x;\mu)] = 0.\end{aligned}$$

In the above $\mathrm{E}[\cdot \mid p]$ and $\mathrm{D}(\cdot,\cdot)$ denote the expectation with respect to the density p and the Kullback-Leibler separator, respectively. The mean Pythagorean relationship in the estimation problem is discussed in Yanagimoto (2000), Ohnishi and Yanagimoto (2003) and Yanagimoto and Ohnishi (2005a). This relationship can be extended to the conjugate location-dispersion family. We have the following lemma.

LEMMA 5.1. *The following mean Pythagorean relationship holds for any* $m \in \mathbb{K}$.

$$\mathrm{E}[\tau d(x-m) - \tau d(x-\mu) - b(\tau)d(\mu - m) \mid p(x-\mu;\tau)] = 0. \tag{5.1}$$

PROOF. Recalling Proposition 3.2, we see that

$$b(\tau)d(\mu-m) = \mathrm{D}_\tau(\mu, m) = \tau\mathrm{E}[d(x-m) - d(x-\mu) \mid p(x-\mu;\tau)].$$

The second equality comes from the definition of the Kullback-Leibler separator. □

Let us obtain an equality, which enables us to construct an estimate of the dispersion parameter of the prior density.

PROPOSITION 5.2. *It holds that*

$$\mathrm{E}[\mathrm{E}[d(x-m) \mid p(x-\mu;\tau)] \mid \pi(\mu - m;\delta)] = a'(\tau) + \frac{b(\tau)}{\tau}h(\delta), \tag{5.2}$$

where $h(\delta)$ *is defined by*

$$h(\delta) = \int_{\mathbb{K}} d(t)\exp\{-\delta d(-t) + a(\delta)\}dt. \tag{5.3}$$

PROOF. A calculation yields $\mathrm{E}[d(x-\mu) \mid p(x-\mu;\tau)] = a'(\tau)$. And this quantity does not depend on μ. Noting that $d(\mu - m)$ does not depend on x, we see that

$$\mathrm{E}[\mathrm{E}[d(\mu-m) \mid p(x-\mu;\tau)] \mid \pi(\mu-m;\delta)] = h(\delta).$$

By combining these with the extended mean Pythagorean relationship (5.1), we obtain (5.2). □

We now proceed with the estimation of the location vector. Suppose that the p-dimensional random vector $\boldsymbol{x} = (x_1, \ldots, x_p)^T$ has the sampling density $p(\boldsymbol{x} - \boldsymbol{\mu};\tau) = \prod p(x_i - \mu_i;\tau)$. Suppose further that a location vector $\boldsymbol{\mu}$ has the prior density $\pi(\boldsymbol{\mu} - \boldsymbol{m};\delta) = \prod \pi(\mu_i - m_i;\delta)$. Here τ and $\boldsymbol{m} = (m_1, \ldots, m_p)^T$ are assumed to be known. In the parametric empirical Bayes method an estimate of the hyperparameter is usually obtained as a maximum likelihood estimate or a method of moments estimate (Carlin and Louis (2000), p. 62). Here we estimate the unknown hyperparameter δ based on the method of moments. The method of moments was used by Bagchi (1994) in the von Mises case, although his work is slightly different from ours. The expression (5.2) yields the equality

$$\mathrm{E}\left[\mathrm{E}\left[\frac{1}{p}\sum_{i=1}^{p} d(x_i - m_i) \,\middle|\, p(\boldsymbol{x}-\boldsymbol{\mu};\tau)\right] \,\middle|\, \pi(\boldsymbol{\mu}-\boldsymbol{m};\delta)\right] = a'(\tau) + \frac{b(\tau)}{\tau}h(\delta).$$

Thus the following estimating equation for δ is obtained:

$$\frac{1}{p}\sum_{i=1}^{p} d(x_i - m_i) = a'(\tau) + \frac{b(\tau)}{\tau} h(\delta). \tag{5.4}$$

When this estimating equation has no solution, we define the estimate of δ as follows:

$$\hat{\delta} = \arg\inf_{\delta} \left| \frac{1}{p}\sum_{i=1}^{p} d(x_i - m_i) - a'(\tau) - \frac{b(\tau)}{\tau} h(\delta) \right|.$$

Five examples of the estimation of δ are presented, which correspond respectively to Examples 4.1–4.5 in the previous section.

Example 5.1. Normal family $\mathcal{N}$;

The estimate of δ is given by $\hat{\delta}^{-1} = [\|\boldsymbol{x} - \boldsymbol{m}\|^2/p - \tau^{-1}]^+$ where $[x]^+ = \max\{0, x\}$.

Example 5.2. Gamma family $\mathcal{G}(1)$;

The estimating equation (5.4) is expressed as

$$\frac{1}{p}\sum_{i=1}^{p}(e^{x_i - m_i} - x_i + m_i - 1) - \psi(\tau) = \frac{1}{\delta - 1} - \psi(\delta),$$

where $\psi(t) = \log t - (d/dt)\log\Gamma(t)$. Here δ is assumed to be larger than one for the existence of the function $h(\delta)$ defined by (5.3). Note that the right-hand side in the above equation is strictly decreasing in $\delta > 1$. The adopted estimator $\hat{\delta}$ is given by

$$\left[\frac{1}{p}\sum_{i=1}^{p}(e^{x_i - m_i} - x_i + m_i - 1) - \psi(\tau)\right]^+ = \frac{1}{\hat{\delta} - 1} - \psi(\hat{\delta}).$$

If the left-hand side is equal to zero, then we set $\hat{\delta} = \infty$.

Example 5.3. Hyperbola family $\mathcal{H}(1,1)$;

The estimating equation (5.4) has a simple form

$$\frac{1}{p}\sum_{i=1}^{p}\cosh(x_i - m_i) = \frac{K_0'(\tau)}{K_0(\tau)}\frac{K_0'(\delta)}{K_0(\delta)}.$$

Since the right-hand side is strictly decreasing in $\delta > 0$ and has the infimum $-K_0'(\tau)/K_0(\tau)$, this equation has the unique solution when $\sum \cosh(x_i - m_i)/p > -K_0'(\tau)/K_0(\tau)$. Otherwise, we set $\hat{\delta} = \infty$.

Example 5.4. Inverted gamma family $\mathcal{IG}(1)$;

The hyperparameter δ is assumed to be larger than one for the same reason in Example 5.2. The adopted estimator $\hat{\delta}$ is given by

$$\left[\frac{1}{p}\sum_{i=1}^{p}(e^{m_i-x_i}+x_i-m_i-1)-\psi(\tau)\right]^{+}=\frac{1}{\hat{\delta}-1}-\psi(\hat{\delta}).$$

When the left-hand side is equal to zero, we set $\hat{\delta}=\infty$.

Example 5.5. von Mises family $\mathcal{M}(1)$;

We obtain a simple estimating equation with respect to δ as

$$\frac{1}{p}\sum_{i=1}^{p}\cos(x_i-m_i)=\frac{I_0'(\tau)}{I_0(\tau)}\frac{I_0'(\delta)}{I_0(\delta)},$$

which is to be compared with the estimating equation in Example 5.3. Our treatment here may be different from the treatments in Examples 5.1–5.4. Recall Example 2.5 where we noted that the dispersion parameter can be zero or negative. Note again that $\hat{\delta}\cos(\mu-m)=-\hat{\delta}\cos(\mu-m-\pi)$. A negative $\hat{\delta}$ implies that $\boldsymbol{m}+\pi\mathbf{1}$, with $\mathbf{1}\in\mathbb{R}^p$ being the vector of ones, is closer to $\boldsymbol{\mu}$ than $\boldsymbol{m}$ is. The right-hand side of the above estimating equation is strictly increasing in $\delta\in\mathbb{R}$, and it has the supremum and the infimum, $\pm I_0'(\tau)/I_0(\tau)$. Thus when the inequality

$$\left|\frac{1}{p}\sum_{i=1}^{p}\cos(x_i-m_i)\right|<\frac{I_0'(\tau)}{I_0(\tau)}$$

holds, the estimating equation has the unique solution, which defines $\hat{\delta}$. Otherwise, we set

$$\hat{\delta}=\begin{cases}\infty & \text{if } \dfrac{1}{p}\displaystyle\sum_{i=1}^{p}\cos(x_i-m_i)\geq\dfrac{I_0'(\tau)}{I_0(\tau)},\\ -\infty & \text{if } \dfrac{1}{p}\displaystyle\sum_{i=1}^{p}\cos(x_i-m_i)\leq-\dfrac{I_0'(\tau)}{I_0(\tau)}.\end{cases}$$

The performance of the resulting empirical Bayes estimator was examined by Ye and Ohnishi (2006). They showed that the empirical Bayes estimator is far superior to the maximum likelihood estimator in a practical situation.

6. Discussions

Two topics are discussed, both of which are related to Pythagorean relationships. One pertains to extensions of a conjugate prior density. The other is about an interpretation of the addition identity (3.1) in the light of statistical physics.

First, we show that the prior density assumed in this paper is derived through two different procedures. One is a simple extension of a conjugate prior density employed by Ibrahim and Chen (1998, 2000) in pursuing applications. Their prior density, which is called a power prior density, has the form

$$\pi(\theta;\delta) \propto \left\{\prod_{i=1}^{n_0} q(y_i;\theta)\right\}^{\delta} r(\theta) \tag{6.1}$$

for a general sampling density $q(x;\theta)$. Here $(y_1,\ldots,y_{n_0})$ can be regarded as a prior dataset, δ is called a power parameter and $r(\theta)$ is an appropriate function. In the case where the sampling density is of the form (1.1), the prior density (1.2) is a power prior density because it is proportional to $\{p(m-\mu,\tau)\}^{\delta/\tau}$.

The other procedure, proposed by Yanagimoto and Ohnishi (2005a), assumes the following prior density using the Kullback-Leibler separator

$$\pi_{\mathrm{KL}}(\theta;\theta_0,\delta) \propto \exp\{-\delta \mathrm{D}(q(y;\theta_0),q(y;\theta))\}r(\theta), \tag{6.2}$$

where θ_0 and δ are hyperparameters. This prior density can be looked upon as an extended conjugate prior density because it is conjugate in the case of an exponential family $\{p_e(x;\mu)\}$. It follows from Proposition 3.2 that our prior density in $\mathcal{P}_c$ is derived through this procedure.

We see that the two prior densities (6.1) and (6.2) coincide with each other if the normed log-likelihood is proportional to the Kullback-Leibler separator. In fact, in the exponential family the equality

$$\log\frac{\prod p_e(y_i;\check{\mu})}{\prod p_e(y_i;\mu)} = n_0\mathrm{D}(p_e(y;\check{\mu}),p_e(y;\mu)) \tag{6.3}$$

holds where $\check{\mu}=\sum y_i/n_0$. Let us show that the proportionality holds also in $\mathcal{P}_c$. Set $\tilde{g}(\mu)(=\tilde{g}(\mu;y_1,\ldots,y_{n_0}))=\sum d(y_i-\mu)$ and write the maximum likelihood estimate of μ based on $(y_1,\ldots,y_{n_0})$ as $\check{\mu}$. A similar treatment to the proof of Proposition 2.3 gives $\tilde{g}(\mu)=\tilde{g}(\check{\mu})+\tilde{g}''(\check{\mu})d(\check{\mu}-\mu)$. Therefore, Proposition 3.2 yields the proportionality as

$$\log\frac{\prod p(y-\check{\mu};\tau)}{\prod p(y-\mu;\tau)} = \frac{\tau\tilde{g}''(\check{\mu})}{b(\tau)}\mathrm{D}_\tau(\check{\mu},\mu). \tag{6.4}$$

It should be noted that the equalities (6.3) and (6.4) are Pythagorean relationships. Set $n_0=1$ for simplicity, and we have $\check{\mu}=y_1$. The two equalities can be written formally by using the Dirac δ-function $\delta(x)$ as

$$\begin{aligned}
&\mathrm{D}(\delta(z-\check{\mu}),p_e(z;\mu)) = \mathrm{D}(\delta(z-\check{\mu}),p_e(z;\check{\mu})) + \mathrm{D}(p_e(z;\check{\mu}),p_e(z;\mu)),\\
&\mathrm{D}(\delta(z-\check{\mu}),p(z-\mu;\tau)) = \mathrm{D}(\delta(z-\check{\mu}),p(z-\check{\mu};\tau))\\
&\qquad\qquad + \frac{\tau}{b(\tau)}\mathrm{D}(p(z-\check{\mu};\tau),p(z-\mu;\tau)).
\end{aligned}$$

Here we suggest that these Pythagorean relationships are closely related to the conjugacy of prior densities.

Next, we present a notable interpretation on the addition identity (3.1) from a viewpoint of statistical mechanics. Recall that the identity played an important role in proving the Pythagorean relationship in Proposition 4.1. Let k_{B} and T be respectively the Boltzmann constant and the absolute temperature, and set $\rho_\tau(t) = \exp\{-\tau d(t) + a(\tau)\}$. If we regard $d(t)$ as an energy function, then $\rho_\tau(t)$ with $\tau = 1/(k_{\mathrm{B}}T)$ is the density of the canonical distribution in statistical mechanics. The canonical distribution is a key concept in statistical mechanics, see Reif (1998, Chapter 4) for example.

An implication of (3.1) can be explained as follows: Consider the situation where we keep the absolute temperature of the system of interest at $T = 1/(\tau k_{\mathrm{B}})$ and the energy of the system located at x is given by $d(x - \mu_1)$ with μ_1 being a controllable parameter. At the equilibrium state the probability density of the system's being at the point x is given by the density $\rho_\tau(x - \mu_1)$ of the canonical distribution. Suppose that we change the parameter value from μ_1 to μ_2 instantaneously without changing the probability density. The addition identity (3.1) together with (3.4) and (3.5) permits us to calculate the increment of the average energy as

$$\int_{\mathbb{K}} \{d(x-\mu_2) - d(x-\mu_1)\}\rho_\tau(x-\mu_1)dx = \frac{b(\tau)}{\tau} d(\mu_1 - \mu_2). \tag{6.5}$$

The quantity $d(\mu_1 - \mu_2)$ is the energy of the system located at μ_1 under the new energy function $d(x - \mu_2)$. The equality (6.5) states that the increment of the average energy is proportional to $d(\mu_1 - \mu_2)$. It should be noted here that $x = \mu_1$ is the point at which the system is the most likely to be located under the probability density $\rho_\tau(x - \mu_1)$.

Appendix A

Proof of Proposition 2.2. (i) Since the deviance function is expressed as $d(t) = \int_0^t d'(s)ds$, we have only to solve the differential equation

$$d'''(t) = \alpha d'(t) + \beta d''(t) \quad \text{with} \quad d'(0) = 0 \text{ and } d''(0) = 1. \tag{A.1}$$

This differential equation is converted into an equivalent first order linear system of differential equations with a constant coefficient matrix

$$\frac{d}{dt}\begin{pmatrix} d'(t) \\ d''(t) \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ \alpha & \beta \end{pmatrix}\begin{pmatrix} d'(t) \\ d''(t) \end{pmatrix}.$$

We solve this differential equation according to Arrowsmith and Place (1982, Chapter 2). The characteristic equation of the above 2×2 matrix, $\lambda^2 - \beta\lambda - \alpha = 0$, determines the form of the solution. After deriving an explicit form of $d'(t)$, we check whether $d(t)$ is in $\mathcal{D}_{\mathbb{K}}$.

First, suppose that the quadratic equation $\lambda^2 - \beta\lambda - \alpha = 0$ has a single real root $\lambda = \lambda_1$. The solution to (A.1) is $d'(t) = te^{\lambda_1 t}$. Noting that $d(t) = \int_0^t se^{\lambda_1 s}ds$, we see that λ_1 must be zero in order for $d(t)$ to be in $\mathcal{D}_{\mathbb{R}}$ and obtain (2.6a).

Secondly, suppose that the quadratic equation has two different real roots $\lambda = \lambda_1$ and $\lambda = \lambda_2$, where the inequality $\lambda_1 < \lambda_2$ is assumed without loss of generality. The solution to (A.1) is $d'(t) = (e^{\lambda_2 t} - e^{\lambda_1 t})/(\lambda_2 - \lambda_1)$. A routine calculation leads us to the fact that $d(t) \in \mathcal{D}_{\mathbb{R}}$ exists only for the three cases, (1) $\lambda_1 = 0 < \lambda_2$, (2) $\lambda_1 < 0 < \lambda_2$ and (3) $\lambda_1 < 0 = \lambda_2$. In case (1) we set $\kappa = \lambda_2 > 0$ and obtain (2.6b). In case (2) we set $\kappa = \lambda_2 > 0$ and $\gamma = -\lambda_1 > 0$, which derives the deviance function (2.6c). In case (3) we set $\kappa = -\lambda_1 > 0$ and obtain (2.6d).

Finally, suppose that the quadratic equation has the two conjugate complex roots, $\lambda = \rho \pm \sqrt{-1}\xi$. Here we mean by $\sqrt{-1}$ the imaginary unit. Then we have $d'(t) = \xi^{-1} e^{\rho t} \sin(\xi t)$. Note, however, that $d(t)$ derived from this differential equation is not in $\mathcal{D}_{\mathbb{R}}$.

(ii) The deviance function $d(t) \in \mathcal{D}_I$ is sought in the same way. It can be shown that we have the desired function only when $\lambda^2 - \beta\lambda - \alpha = 0$ has two conjugate pure-imaginary roots, $\lambda = \pm\sqrt{-1}\xi$ with $\xi \in \mathbb{N}$. The function (2.6e) is thus obtained. $\square$

Acknowledgements

The authors are very grateful to the referees and Dr. Peter Dunn, University of Southern Queensland, for their valuable comments on the original version of the present paper.

References

Abramowitz, M. and Stegun, I. A. (1974). *Handbook of Mathematical Functions*, Dover Publications, New York.

Angers, J.-F. (1996). Fourier transform and Bayes estimator of a location parameter, *Statist. Probab. Lett.*, **29**, 353–359.

Arrowsmith, D. K. and Place, C. M. (1982). *Ordinary Differential Equations*, Chapman & Hall, London.

Bagchi, P. (1994). Empirical Bayes estimation in directional data, *J. Appl. Statist.*, **21**, 317–326.

Barndorff-Nielsen, O. (1978). Hyperbolic distribution and distribution on hyperbolae, *Scand. J. Statist.*, **5**, 151–157.

Bischoff, W. (1993). On the greatest class of conjugate priors and sensitivity of multivariate normal posterior distributions, *J. Multivariate Anal.*, **44**, 69–81.

Carlin, B. P. and Louis, T. A. (2000). *Bayes and Empirical Bayes Methods for Data Analysis*, 2nd ed., Chapman & Hall/CRC.

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

Diaconis, P. and Ylvisaker, D. (1985). Quantifying prior opinion (with discussions), *Bayesian Statist. 2* (eds. J. M. Bernardo *et al.*), North-Holland, Amsterdam, 133–156.

Garvan, C. W. and Ghosh, M. (1997). Noninformative priors for dispersion models, *Biometrika*, **84**, 976–982.

Guttorp, P. and Lockhart, R. A. (1988). Finding the location of a signal: A Bayesian analysis, *J. Am. Statist. Assoc.*, **83**, 322–330.

Huang, Y.-S. and Bier, V. M. (1999). A natural conjugate prior for the nonhomogeneous Poisson process with an exponential intensity function, *Commun. Statist. Theory Methods*, **28**, 1479–1509.

Ibrahim, J. G. and Chen, M.-H. (1998). Prior distributions and Bayesian computation for proportional hazard model, *Sankhya B*, **60**, 48–64.

Ibrahim, J. G. and Chen, M.-H. (2000). Power prior distributions for regression models, *Statistical Science*, **15**, 46–60.

Jensen, J. L. (1981). On the hyperboloid distribution, *Scand. J. Statist.*, **8**, 193–206.

Jørgensen, B. (1997). *The Theory of Dispersion Models*, Chapman & Hall, London.

Kagan, A. M., Linnik, Y. V. and Rao, C. R. (1973). *Characterization Problems in Mathematical Statistics*, John Wiley & Sons, New York.

Leblanc, A. and Angers, J.-F. (1999). Bayesian estimation of a location parameter using the Haar basis, *Bayesian Statist. 6* (eds. J. M. Bernardo *et al.*), Oxford University Press, Oxford, 803–812.

Lehmann, E. L. and Casella, G. (1998). *Theory of Point Estimation*, 2nd ed., Springer-Verlag, New York.

Mardia, K. V. and El-Atoum, S. A. M. (1976). Bayesian inference for the von Mises-Fisher distribution, *Biometrika*, **63**, 203–206.

Morris, C. N. (1983a). Parametric empirical Bayes inference: Theory and applications, *J. Am. Statist. Assoc.*, **78**, 47–55.

Morris, C. N. (1983b). Natural exponential families with quadratic variance functions: Statistical theory, *Ann. Statist.*, **11**, 515–529.

Ohnishi, T. (2006). Estimating a common slope of multiple strata in the Tweedie distribution using a conjugate prior, *Kōkyūroku*, Research Institute for Mathematical Science, Kyoto University, **1506**, 167–176.

Ohnishi, T. and Yanagimoto, T. (2003). Electrostatic views of Stein-type estimation of location vectors, *J. Japan Statist. Soc.*, **33**, 39–64.

Polson, N. G. (1991). A representation of the posterior mean for a location model, *Biometrika*, **78**, 426–430.

Reif, F. (1998). *Customized Complete Statistical Physics*, McGraw-Hill, New York.

Robert, C. P. (2001). *The Bayesian Choice*, 2nd ed., Springer-Verlag, New York.

Rodrigues, J., Leite, J. G. and Milan, L. A. (2000). An empirical Bayes inference for the von Mises distribution, *Austral. New Zealand J. Statist.*, **42**, 433–440.

Spiegelhalter, D. J. (1985). Exact Bayesian inference on the parameter of a Cauchy distribution with vague prior information, *Bayesian Statist. 2* (eds. J. M. Bernardo), North-Holland, Amsterdam, 743–749.

Yanagimoto, T. (2000). A pair of estimating equations for a mean vector, *Statist. Probab. Lett.*, **50**, 97–103.

Yanagimoto, T. and Ohnishi, T. (2005a). Extensions of the conjugate prior through the Kullback-Leibler separators, *J. Multivariate Anal.*, **92**, 116–133.

Yanagimoto, T. and Ohnishi, T. (2005b). Standardized posterior mode for the flexible use of a conjugate prior, *J. Statist. Plann. Inference*, **131**, 253–269.

Ye, X. and Ohnishi, T. (2006). Empirical Bayes estimation in von Mises distribution, *Tōkei-sūri*, Institute of Statistical Mathematics, **54**, 177–190 (in Japanese).