Morfismos, Vol. 8, No. 2, 2004, pp. 27–50

On information measures and prior distributions: a synthesis

Francisco Venegas-Martínez

Abstract

This paper suggests a new approach to reconciling, in a systematic way, all inferential methods that maximize a specific criterion functional to produce *non-informative* and *informative* priors. In particular, Good's (1968) Minimax Evidence Priors (MEP), Zellner's (1971) Maximal Data Information Priors (MDIP) and Bernardo's (1979) Reference Priors (RP) are seen as special cases of maximizing a more general criterion functional. In a unifying approach Good-Bernardo-Zellner's priors are introduced and applied to a number of Bayesian inference problems, including the Kalman filter and Normal linear model. Moreover, the paper focuses, under plausible conditions, on the existence and uniqueness of the solutions of the derived optimization problems.

2000 Mathematics Subject Classification: 62F15, 49K20.
Keywords and phrases: information measures, Bayesian inference.

1 Introduction

The distinctive task in Bayesian inference of deriving priors, in such a way that the inferential content of the data is minimally affected in the posterior, has been of great interest for more than two centuries since the early work of Bayes (1763). More current approaches to this problem, based on the maximization of a specific criterion functional, have been suggested by Good (1968), Zellner (1971) and Bernardo (1979), among others. It is also important to mention that recent literature has included inference procedures to provide a posterior without having a prior, like the Bayesian method of moments (BMOM) introduced by Zellner (1996) and (1998).

In Good's (1968) principle of maximum invariantized negative cross-entropy, the minimax evidence method of deriving priors was presented for the first time. In this approach the initial density is taken as the square root of Fisher's information. Zellner's (1971) book introduced a method to obtain priors through the maximization of the *total* information about the parameters provided by independent replications of an experiment (prior average information in the data minus the information in the prior). In Bernardo (1979) a procedure was proposed to produce *reference* priors by maximizing the expected information about the parameters provided by independent replications of an experiment (average information in the posterior minus the information in the prior). All of the above methods have comparative and absolute advantages in several respects and have been applied to a large number of inference problems:

(*i*) While Zellner's method is based on an exact finite sample criterion functional, Good's approach uses a limiting criterion functional, and Bernardo's procedure lies in asymptotic results. In Bernardo's proposal a reference prior (posterior) is defined as the limit of a sequence of priors (posteriors) that maximize finite-sample criteria. In a pragmatic approach in which results are most important, many reference prior algorithms have been developed. For instance, Berger, Bernardo and Mendoza (1989), and Berger and Bernardo (1989), (1992a), (1992b), Bernardo and Smith (1994 , ch. 5), and Bernardo and Ramón (1997).

(*ii*) The criterion functional used by Bernardo is a cross-entropy, which satisfies a number of remarkable properties, in particular, invariance with respect to one-to-one transformations of the parameters (Lindley 1956). In contrast, the total information functional employed by Zellner is invariant only for the location-scale family and under linear transformations of the parameters. To generate invariance under other relevant transformations, not necessarily one-to-one, side conditions could be needed, as suggested by Zellner (1971).

(*iii*) These methods have been tested by seeing how well they perform in particular examples. The evaluation is often based on contrasting the derived priors with Jeffreys' (1961) priors, usually improper. Even though improper priors can be associated with unbounded measures consistent with Renyi's (1970) axioms on probability measures, some technical difficulties remain, see: Box and Tiao (1973), p. 314; Akaike (1978), p. 58; and Berger and Bernardo (1992a), p. 37. It is also important to mention that Jeffreys' priors can lead to singularities producing inadequate results at certain values of the parameters; see Jeffreys (1967, p. 359). Of course, if MEP, MDIP, and RP priors were to be used to contrast the performance of other priors, the former priors could also produce unsatisfactory results under certain circumstances.

In this paper, we attempt to reconcile all inferential methods that maximize a criterion functional to produce *non-informative* and *informative* priors. In our general approach, Good's Minimax Evidence Priors (1968 and 1969), Zellner's Maximal Data Information Priors (1971, 1977, 1991, 1993, 1995, 1996a and 1996b) and Bernardo's Reference Priors (1979 and 1997) are seen as special cases of maximizing a more general indexed criterion functional. Thus, properties of the derived priors will depend on the choice of indexes from a wide range of possibilities, instead of on a few personal points of view with *ad hoc* modifications. In the spirit of Akaike (1978) and Smith (1979), we can say that this will look more like Mathematics than Psychology–without underestimating the importance of the latter in the Bayesian framework. This unified approach will enable us to explore a vast range of possibilities for constructing priors. It is worthwhile to note that our general method extends in a natural way Soofi's (1994) pyramid by adding more vertices and including their convex hull. In any event, a good choice will depend on the specific characteristics of the problem we are concerned with. Needless to say, the chosen method should also provide good predictions.

This work is organized as follows. In section 2, we will introduce an indexed family of information functionals. In section 3, on the basis of asymptotic normality, we will state a relationship between Bernardo's (1979) criterion functional and some members of the indexed family. In section 4, we will study a Bayesian inference problem associated with convex combinations of relevant members of the proposed indexed family. Here, we will introduce Good-Bernardo-Zellner's priors as well as their *controlled* versions as solutions of maximizing discounted entropy. We will pay special attention to the existence and uniqueness of the solution of the corresponding optimization problems. In section 5, we will study Good-Bernardo-Zellner's priors as Kalman Filtering priors. In section 6, we examine the relationship between Good-Bernardo-Zellner's priors and the Normal linear model. Finally, in section 7, we will draw conclusions, acknowledge limitations, and make suggestions for further research.

2 An indexed family of information functionals

In this section, we define an indexed family of information functionals and study some distinguished members. For the sake of simplicity, we will remain in the single parameter case. The extension to the multi-dimensional parameter case will lead to conceptual complications. This is not surprising when dealing with information measures and priors; see Jeffreys (1961), Zellner (1971), Box and Tiao (1973), and Berger and Bernardo (1992a).

Suppose that we wish to make inferences about an unknown parameter $\theta \in \Theta \subseteq \mathbb{R}$ of a distribution P_θ, from which there is available an observation, say, X. Assume that P_θ has density $f(x|\theta)$ (Radon-Nikodym derivative) with respect to some fixed dominating σ-finite measure λ on $\mathbb{R}$ for all $\theta \in \Theta \subseteq \mathbb{R}$, that is, $dP_\theta/d\lambda = f(x|\theta)$ for all $\theta \in \Theta \subseteq \mathbb{R}$, thus $P_\theta(A) = \int_A f(x|\theta)d\lambda(x)$ for all Borel sets $A \subset \mathbb{R}$.

The Bayesian approach is to assume that there is a prior density, $\pi(\theta)$, describing initial knowledge about the likelihood of the values of the parameter, θ. We will assume that $\pi(\theta)$ is a density with respect to some σ-finite measure μ on $\mathbb{R}$. Once a prior distribution, $\pi(\theta)$, has been prescribed, then the information provided by the data, x, about the parameter is used to modify the initial knowledge, as expressed in $\pi(\theta)$, *via* Bayes' theorem to obtain a posterior distribution of θ, namely, $f(\theta|x) \propto f(x|\theta)\pi(\theta)$ for every $x \in \mathbb{R}$ (using f generically to represent densities). The normalized posterior distribution is then used to make inferences about θ. Let us define an infinite system of nested functionals:

$$(1) \qquad \mathcal{V}_{\gamma,\alpha,\delta}(\pi) = \frac{1}{1-\gamma}\int \pi(\theta)G(\mathcal{I}(\theta),\mathcal{F}(\theta),\gamma,\alpha,\delta)d\mu(\theta)$$

where

$$\begin{aligned} &G(\mathcal{I}(\theta),\mathcal{F}(\theta),\gamma,\alpha,\delta) \\ &= \log\left\{\frac{\exp\{[\mathcal{F}(\theta)/\mathcal{I}(\theta)]^{1-\delta}[\mathcal{I}(\theta)]^{\frac{1-\gamma}{1+\alpha}} - \delta[\mathcal{I}(\theta)]^{1-\alpha}\}}{\pi(\theta)^{1-\gamma}}\right\},\end{aligned}$$

$0 \leq \gamma < 1,\ \alpha \in \{0,1\},\ \delta \in \{0,1\}$, and

$$\mathcal{I}(\theta) = \int \left(\frac{\partial}{\partial\theta} \log f(x|\theta) \right)^2 f(x|\theta) d\lambda(x) \tag{2}$$

is Fisher's information about θ provided by an observation X with density $f(x|\theta)$, and

$$\mathcal{F}(\theta) = \int f(x|\theta) \log f(x|\theta) d\lambda(x) \tag{3}$$

is the negative Shannon's information of $f(x|\theta)$, provided $\mathcal{I}(\theta)$ and $\mathcal{F}(\theta)$ exist. In the case that n independent observations of X are drawn from P_θ, say, $(X_1, X_2, ..., X_n)$, then $\mathcal{I}(\theta)$ and $\mathcal{F}(\theta)$ will still stand for the average Fisher's information and the average negative Shannon's information of $f(x|\theta)$ respectively. It is not unusual to deal with indexed functionals in inference problems about a distribution, as in Good (1968). It is worthwhile pointing out that for each triad (γ, α, δ) taking values in $0 \leq \gamma < 1,\ \alpha \in \{0,1\},\ \delta \in \{0,1\}$, then $\mathcal{V}_{\gamma,\alpha,\delta}(\pi)$ is a criterion functional that can be used to derive a prior $\pi(\theta),\ \theta \in \Theta$, belonging to a feasible set $\mathcal{C}$. Usually, $\mathcal{C}$ is defined by constraints in terms of potential values of θ.

Note now that for the location parameter family $f(x|\theta) = f(x - \theta),\ \theta \in \mathbb{R}$, with the properties

$$\int [f'(x)]^2/f(x)\ d\lambda(x) < \infty$$

and

$$\int f(x) \log f(x)\ d\lambda(x) < \infty,$$

where $\lambda = \mu$ stands for the Lebesgue measure, we have that both $\mathcal{I}(\theta)$ and $\mathcal{F}(\theta)$ are constant. Observe also that the scale parameter family $f(x|\theta) = (1/\theta)f(x/\theta),\ \theta > 0$, with the above properties, satisfies the following relationship:

$$\mathcal{F}(\theta) = \tfrac{1}{2} \log \mathcal{I}(\theta) + \text{constant}. \tag{4}$$

The indexed family in which we will be concerned with is given by:

$$\begin{aligned} \mathcal{A} &= conv[\ \overline{\{\mathcal{V}_{\gamma,\alpha,\delta}(\pi)\}}\ ] \\ &= \text{convex hull of the closure of the family} \{\mathcal{V}_{\gamma,\alpha,\delta}(\pi)\}. \end{aligned}$$

We readily identify a number of distinguished members of $\mathcal{A}$:

(*i*) Criterion for Maximum Entropy Priors (MAXENTP):

$$\mathcal{V}_{0,0,1}(\pi) = -\int \pi(\theta) \log \pi(\theta) d\mu(\theta),$$

which is just Shannon's information measure of a density $\pi(\theta)$, or Jaynes' (1957) criterion functional to derive maximum entropy priors. Notice also that (3) can be rewritten in a simpler way as $\mathcal{F}(\theta) = -\mathcal{V}_{0,0,1}(f(x|\theta))$.

(*ii*) Criterion for Minimax Evidence Priors (MEP):

$$(5) \quad \mathcal{V}_{1,1,1}(\pi) \overset{\text{def}}{=} \lim_{\gamma \to 1} \mathcal{V}_{\gamma,1,1}(\pi) = -\int \pi(\theta) \log \frac{\pi(\theta)}{p(\theta)} d\mu(\theta) - \log C,$$

which is Good's invariantized negative cross-entropy, taking as initial density $p(\theta) = C[\mathcal{I}(\theta)]^{\frac{1}{2}}$ with $C = \{\int [\mathcal{I}(\theta)]^{\frac{1}{2}} d\mu(\theta)\}^{-1}$, provided that $\int [\mathcal{I}(\theta)]^{\frac{1}{2}} d\mu(\theta) < \infty$. We can also write (5) as

$$(6) \qquad \mathcal{V}_{1,1,1}(\pi) - \mathcal{V}_{0,0,1}(\pi) = \int \pi(\theta) \log [\mathcal{I}(\theta)]^{\frac{1}{2}} d\mu(\theta).$$

(*iii*) Criterion for Maximal Data Information Priors (MDIP):

$$(7) \qquad \mathcal{V}_{0,0,0}(\pi) = \int\int f(x) f(\theta|x) \log \frac{\ell(\theta|x)}{\pi(\theta)} d\mu(\theta) d\lambda(x),$$

which is Zellner's criterion functional in his MDIP approach. Here, as usual,

$$f(\theta|x) = \frac{f(x|\theta)\pi(\theta)}{f(x)}, \qquad f(x) = \int f(x|\theta)\pi(\theta) d\mu(\theta),$$

and $\ell(\theta|x) = f(x|\theta)$ is the likelihood function. An alternative formulation of (7), which is often useful, is given by

$$(8) \qquad \mathcal{V}_{0,0,0}(\pi) - \mathcal{V}_{0,0,1}(\pi) = \int \pi(\theta) \mathcal{F}(\theta) d\mu(\theta).$$

Some members of $\mathcal{A}$ define new criterion functionals in which the information provided by the sampling model, $\mathcal{I}(\theta)$, plays a role:

(*iv*) Criterion for Maximal Modified Data Information Priors (MMDIP):

$$(9) \qquad \mathcal{V}_{0,1,0}(\pi) = \int\int f(x)f(\theta|x)\log\frac{[\ell(\theta|x)]^{[\mathcal{I}(\theta)]^{\frac{1}{2}}}}{\pi(\theta)}d\mu(\theta)d\lambda(x),$$

which is the prior average information in the data *modified* by Fisher's information minus the information in the prior. Note that when $\mathcal{I}(\theta)$ is constant, (9) reduces to Zellner's criterion functional (up to a constant factor).

(*v*) Criterion for Maximal Fisher Information Priors (MFIP):

$$(10) \qquad \mathcal{V}_{0,1,1}(\pi) = -\int \pi(\theta)\log\frac{\pi(\theta)}{\exp\{[\mathcal{I}(\theta)]^{\frac{1}{2}}\}}d\mu(\theta) - 1,$$

which is the prior average Fisher's information minus the information in the prior.

3 Revisiting Bernardo's reference priors

The maximization of Bernardo's (1979) criterion is usually a difficult problem to deal with. In order to get a simpler alternative procedure under specific conditions, we will derive a useful asymptotic approximation between Bernardo's criterion functional (or Lindley's information measure, 1956) and some members of the class $\mathcal{A}$. As stated in Bernardo (1979), the concept of reference prior is very general. However, in order to keep the analysis tractable, we will restrict ourselves to the continuous one-dimensional parameter case.

Suppose that there are available n independent observations, say, $(X_1, X_2, \dots, X_n)$, of a distribution P_θ, $\theta \in \Theta \subseteq \mathbb{R}$. Accordingly, the random vector $(X_1, X_2, \dots, X_n)$ has density

$$d\mathcal{P}_\theta/d\nu = f(\xi|\theta) = \prod_{k=1}^{n} f(x_k|\theta),$$

for all $\xi = (x_1, x_2, ..., x_n)$ and all $\theta \in \Theta \subseteq \mathbb{R}$, where

$$\mathcal{P}_\theta = \underbrace{P_\theta \otimes P_\theta \otimes \cdots \otimes P_\theta}_{n} \text{ and } \nu = \underbrace{\lambda \otimes \lambda \otimes \cdots \otimes \lambda}_{n}.$$

Following Lindley (1956), a measure of the expected information about θ of a sampling model $f(x|\theta)$ provided by a random sample of size n when the prior distribution of θ is $\pi(\theta)$, is defined to be

$$\mathcal{L}^{(n)}(\pi) = \int f(\xi) \int f(\theta|\xi) \log \frac{f(\theta|\xi)}{\pi(\theta)} d\mu(\theta) d\nu(\xi). \tag{11}$$

In order to obtain an asymptotic approximation of (11) in terms of $\mathcal{V}_{1,1,1}$ and $\mathcal{V}_{0,0,1}$, we state a limit theorem which justifies the passage of the limit under the integral signs in (11). The theorem rules out the possibility that the *essentials* of the statistical model, $f(\xi|\theta)$, change when samples grow in size. Let us rewrite (11) as:

$$\begin{aligned}\mathcal{L}^{(n)}(\pi) =& \mathcal{V}_{\gamma,0,1}(\pi) + \log\sqrt{n} \\ & - \int\int \log\Big(\int T_n(\omega) W_n(\omega) d\mu(\omega)\Big) f(\xi|\theta)\pi(\theta) d\nu(\xi) d\mu(\theta),\end{aligned} \tag{12}$$

where

$$T_n(\omega) = \frac{f(X_1, X_2, ..., X_n|\theta + \frac{\omega}{\sqrt{n}})}{f(X_1, X_2, ..., X_n|\theta)} \tag{13}$$

and

$$W_n(\omega) = \frac{\pi(\theta + \frac{\omega}{\sqrt{n}})}{\pi(\theta)}. \tag{14}$$

Throughout the paper, both λ and μ will stand for the Lebesgue measure on $\mathbb{R}$. Also, we will assume that all densities involved are Lebesgue measurable in both arguments, x and θ.

Theorem 3.1 *Assume that the following conditions hold:*

(I) Θ *is an open interval in* $\mathbb{R}$*;*

(II) The function $\sqrt{f(x|\theta)}$ *is absolutely continuous on* θ*, and*

$$\{x | f(x|\theta) > 0\}$$

is independent of θ*;*

(III) If $\theta, \theta' \in \Theta$*, then* $\theta \neq \theta'$ *implies* $\lambda\{x | f(x|\theta) \neq f(x|\theta')\} > 0$*;*

(IV) $\frac{\partial}{\partial\theta}\log f(x|\theta)$ *exists for all* $\theta \in \Theta$ *and every* x*;*

(V) $\mathcal{I}(\theta)$ *is a continuous and bounded function in* Θ*;*

(VI) For all $\delta > 0$*, and all* $\theta \in \Theta$

$$\int_{B_\delta(\frac{\omega}{\sqrt{n}})} \Big(\sqrt{f(x|\theta + \tfrac{\omega}{\sqrt{n}})} - \sqrt{f(x|\theta)}\ \Big)^2 d\lambda(x) = o(\tfrac{1}{n}),$$

where

$$B_\delta(\tfrac{\omega}{\sqrt{n}}) = \{x:\ |\sqrt{f(x|\theta + \tfrac{\omega}{\sqrt{n}})} - \sqrt{f(x|\theta)}\ | > \delta\sqrt{f(x|\theta)}\ \};$$

(VII) There exist $c > 0$ *and* $\tau > 0$ *such that*

$$\int \big|\ \pi(\theta + u) - \pi(\theta)\ \big| d\mu(\theta) \leq c|u|^\tau;$$

(VIII) For all $\rho > 0$

$$\int_{|\omega| > n^\rho} \Big(T_n(\omega)W_n(\omega) - T_n(\omega)\Big) d\mu(\omega) \xrightarrow{P} 0;$$

(IX) The sequence of random variables $\{\log U_n\}_{n=1}^\infty$ *where*

$$U_n = \int T_n(\omega)W_n(\omega)d\mu(\omega)$$

satisfies

$$\lim_{\varepsilon\to\infty} \sup_{n\geq 1} \int_{|\log U_n| \geq \varepsilon} |\log U_n| dP = 0,$$

where

$$P\{\xi \in A,\ \theta \in B\} = \int_B \pi(\theta) \int_A f(\xi|\theta)d\nu(\xi)d\mu(\theta),$$

for all $A \in \mathbb{R}^n$ *and* $B \in \Theta$*.*

Then, as $n \to \infty$*,*

$$\mathcal{L}^{(n)}(\pi) - \mathcal{V}_{1,1,1}(\pi) = -\mathcal{V}_{0,0,1}(\varphi) + \log C\sqrt{n} + o(1), \tag{15}$$

where $\varphi(z)$ *is the density of* $Z \sim \mathcal{N}(0,1)$*, and* C *is taken as in (4).*

Some comments are in order: (I)-(IV) are standard regularity conditions, (V) states desirable properties for $\mathcal{I}(\theta)$, (VI) is a bounded variance condition, (VII) is a smoothness condition, (VIII) is a convergence condition, and (IX) says that the sequence $\{\log U_n\}_{n=1}^{\infty}$ is uniformly integrable with respect to P. It can be shown that (I)-(VI) lead to

$$T_n(\omega) \xrightarrow{\mathcal{L}} \exp\big\{\omega\sqrt{\mathcal{I}(\theta)}\big[Z - \tfrac{1}{2}\omega\sqrt{\mathcal{I}(\theta)}\big]\big\}, \tag{16}$$

where $Z \sim \mathcal{N}(0,1)$, and (16) along with (VII)-(IX) imply

$$\log U_n = \log \int T_n(\omega) W_n(\omega) d\mu(\omega) \xrightarrow{\mathcal{L}} \log\sqrt{2\pi/\mathcal{I}(\theta)} + \tfrac{1}{2}Z^2,$$

from where the conclusion of the theorem follows. Notice that the right-hand side of (3.5) is independent of π. Thus, if conditions (I)-(IX) are fulfilled, instead of maximizing $\mathcal{L}^{(\infty)}(\pi)$, which is usually a difficult problem, we have as an alternative procedure maximizing $\mathcal{V}_{1,1,1}(\pi)$, which is independent of n. Notice that for maximization purposes the right-hand side of (15) becomes a constant.

Finally, it is worthwhile to note that the location parameter family $f(x|\theta) = f(x-\theta)$, with $\sqrt{f(x)}$ absolutely continuous on $\mathbb{R}$, and $\int [f'(x)]^2/f(x)\ d\lambda(x) < \infty$, fully satisfies the conditions of Theorem 3.1.

4 Good-Bernardo-Zellner priors

In this section we introduce Good-Bernardo-Zellner's priors as solutions of convex combination of relevant members of the class $\mathcal{A}$. Very often, there exist priors for which entropy becomes infinite, specially when dealing with the non-informative case. In order to overcome this difficulty, we suggest the concept of discounted entropy. We also introduce Good-Bernardo-Zellner's *controlled* priors as solutions of maximizing discounted entropy. We emphasize the existence and uniqueness of the solutions of the corresponding variational and optimal control problems.

Throughout this section, we will be studying a number of Bayesian inferential problems related to convex combinations of distinctive elements of $\mathcal{A}$. Let

$$\mathcal{M}_\phi(\pi) \overset{\text{def}}{=} \phi\mathcal{V}_{1,1,1}(\pi) + (1-\phi)\mathcal{V}_{0,0,0}(\pi),$$

$0 \le \phi \le 1$. Plainly, $\mathcal{M}_\phi(\pi) \in \mathcal{A}$. To see that $\mathcal{M}_\phi(\pi)$ is concave w.r.t. π, it is enough to observe, as in Zellner (1991), that

$$\mathcal{V}_{0,0,0}(\pi(\theta)) = \mathcal{L}^{(1)}(\pi(\theta)) + \mathcal{V}_{0,0,1}(\pi(\theta)) - \mathcal{V}_{0,0,1}(f(x)),$$

is a sum of concave functions w.r.t. π (up to the constant $\mathcal{V}_{0,0,1}(f(x))$). Since $\mathcal{V}_{1,1,1}(\pi)$ is concave w.r.t. π, $\mathcal{M}_\phi(\pi)$ is also concave w.r.t. π. Zellner (1996b) provides a criterion functional that agrees with $\mathcal{M}_\phi(\pi)$ given by

$$\begin{aligned} &G_\phi[\pi(\theta)] \\ &= \int \left[\phi\mathcal{F}(\theta) + (1-\phi)\log[\mathcal{I}(\theta)]^{\frac{1}{2}}\right]\pi(\theta)d\mu(\theta) - \int \pi(\theta)\log\pi(\theta)d\mu(\theta). \end{aligned}$$

Indeed, from (5), (6) and (8), we get

$$\begin{aligned} G_\phi[\pi(\theta)] =& \phi\left(\mathcal{V}_{0,0,0} - \mathcal{V}_{0,0,1}\right) + (1-\phi)\left(\mathcal{V}_{1,1,1} - \mathcal{V}_{0,0,1}\right) + \mathcal{V}_{0,0,1} \\ =& \phi\mathcal{V}_{0,0,0} + (1-\phi)\mathcal{V}_{1,1,1} - \mathcal{V}_{0,0,1} + \mathcal{V}_{0,0,1} \\ =& \mathcal{M}_\phi(\pi). \end{aligned}$$

Usually, in the absence of data *supplementary* information, in terms of expectations about the parameter, comes from additional knowledge of the experiment, or from the experience of the experimenter, namely,

$$\int a_k(\theta)\pi(\theta)d\mu(\theta) = \overline{a}_k, \quad k = 1,2,...,s, \tag{17}$$

where both the functions a_k and the constants $\overline{a}_k$, $k = 1,2,...,s$, are known. Hereafter, we will assume that (17) does not lead to any contradiction about $\pi(\theta)$. We will now concern with maximizing $\mathcal{M}_\phi(\pi)$ subject to *supplementary* information.

Proposition 4.1 *Consider the Good-Bernardo-Zellner problem:*

$$\textit{Maximize } \mathcal{M}_\phi(\pi) \quad (\textit{with respect to } \pi)$$

$$\textit{subject to } \mathcal{C} : \int a_k(\theta)\pi(\theta)d\mu(\theta) = \overline{a}_k, \quad k = 0,1,2,...,s, \quad a_0 \equiv 1 = \overline{a}_0.$$

Then a necessary condition for a maximum is

$$\pi^*_\phi(\theta) \propto [\mathcal{I}(\theta)]^{\frac{\phi}{2}} \exp\{(1-\phi)\mathcal{F}(\theta) + \sum_{k=0}^{s}\lambda_k a_k(\theta)\}, \tag{18}$$

where λ_k, $k = 0,1,...,s$, are the Lagrange multipliers associated with the constraints $\mathcal{C}$ (cf. Zellner 1995).

Notice that when no supplementary information is available, $\pi^*_\phi(\theta)$ is appropiate for an unprejudiced experimenter, otherwise it will be suitable for an informed experimenter who is in favor of $\mathcal{C}$. Observe also that $\pi^*_1(\theta)$ is Good-Bernardo's prior, and $\pi^*_0(\theta)$ is Zellner's prior. Consider the binomial distribution for a single observation, $f(x|\theta) = \theta^x(1-\theta)^{1-x},\ 0 \leq \theta \leq 1$. In such a case, $\pi^*_1(\theta) \propto \theta^{-\frac{1}{2}}(1-\theta)^{-\frac{1}{2}}$ and $\pi^*_0(\theta) \propto \theta^\theta(1-\theta)^{1-\theta}$ for $\theta \in [0,1]$, which are quite different. Notice that $\pi^*_1(\theta)$ becomes infinite at $\theta = 0$ and $\theta = 1$. On the other hand $\pi^*_0(\theta)$ rises monotonically to 1.6186 at $\theta = 0$ and $\theta = 1$. Yet, another view in this regard (Geisser, 1993) states that when the sample size is fairly large it does not matter which prior is employed, and the uniform prior may as well be used for θ.

Corolary 4.1 *Consider the location and scale parameter families,*

$$f(x|\theta) = f(x-\theta),\ \theta \in \mathbb{R},$$

and

$$f(x|\theta) = (1/\theta)f(x/\theta),\ \theta > 0,$$

respectively, both satisfying

$$\int [f'(x)]^2/f(x)\ d\lambda(x) < \infty$$

and $\int f(x)\log f(x)\ d\lambda(x) < \infty$. *Then, Good-Bernardo's and Zellner's priors agree regardless of the value of* $\phi \in (0,1)$.

The proof of the above corollary for the scale parameter case follows from (4). It is important to point out that when there is no supplementary information, we require $\mu(\Theta) < \infty$. Of course, the parameter space Θ can have bounds as large as needed to consider where the likelihood for θ is appreciable.

Notice that Proposition 4.1 can be used recursively when there is more supplementary information to be added, say,

$$\int a_k(\theta)\pi(\theta)d\mu(\theta) = \overline{a}_k, \quad k = s+1, s+2, ..., t. \tag{19}$$

In such a case, in a cross-entropy formulation (Kullback 1959), we take (18) as the initial density, and (19) as the additional information. Hence,

$$\pi^*_\phi(\theta) \propto \ [\mathcal{I}(\theta)]^{\frac{\phi}{2}} \exp\{(1-\phi)\mathcal{F}(\theta) + \sum_{k=0}^{s} \lambda_k a_k(\theta)\} \ \exp\{\sum_{k=s+1}^{t} \lambda_k a_k(\theta)\}$$

$$= [\mathcal{I}(\theta)]^{\frac{\phi}{2}} \exp\{(1-\phi)\mathcal{F}(\theta) + \sum_{k=0}^{t} \lambda_k a_k(\theta)\} \ .$$

To deal with the (local) uniqueness of the solution of the problem stated in Proposition 4.1, we rewrite the constraints, $\mathcal{C}$, as a function of the multipliers in the form $A(\Lambda) = [\int a_k(\theta)\pi^*_\phi(\theta)d\mu(\theta)]^s_{k=0} = \bar{A}$, where $\bar{A}^T = (\overline{a}_0, \overline{a}_1, ..., \overline{a}_s)$, and $\Lambda^T = (\lambda_0, \lambda_1, ..., \lambda_s)$ (the superindex T denotes the usual vector or matrix transposing operation).

Proposition 4.2 *Let $\pi^*_\phi(\theta)$ be as in (4.2), and suppose that a_k, $k = 0, 1, ..., s$, are linearly independent continuous functions in $L^2[\Theta, \pi^*_\phi d\mu]$ (the space of all $\pi^*_\phi d\mu$-measurable functions $a(\theta)$ defined on Θ such that $|a(\theta)|^2$ is $\pi^*_\phi d\mu$-integrable). Suppose that $A(\Lambda)$ is defined on an open set $\Delta \subset \mathbb{R}^{s+1}$, and let Λ_o be a solution of $A(\Lambda) = \bar{A}$ for a fixed value of $\bar{A} = \bar{A}_o$. Then, there exists a neighborhood of Λ_o, $N(\Lambda_o)$, in which Λ_o is the unique solution of $A(\Lambda) = \bar{A}_o$ in $N(\Lambda_o)$.*

The proof follows from the fact that $A(\Lambda)$ is continuously differentiable on Δ, with nonsingular derivative

$$A'(\Lambda) = [\int a_\jmath(\theta) a_\ell(\theta) \pi^*_\phi(\theta) d\mu(\theta)]_{0 \le \jmath, \ell \le s},$$

and from a straightforward application of inverse function theorem.

From (4.1) we may derive the following necessary condition, which is useful in practical situations.

Proposition 4.3 *The multipliers $\Lambda^T = (\lambda_0, \lambda_1, ..., \lambda_s)$ appearing in (18) satisfy the following non-linear system of $s+1$ equations:*

$$1 = \lambda_0 + log\Big\{\int [\mathcal{I}(\theta)]^{\frac{\phi}{2}} \ e^{(1-\phi)\mathcal{F}(\theta)} \ \prod_{k=1}^{s} e^{\lambda_k a_k(\theta)} d\mu(\theta)\Big\},$$

$$1 = \lambda_0 - log \ \bar{a}_k + log\Big\{\int a_k(\theta) [\mathcal{I}(\theta)]^{\frac{\phi}{2}} \ e^{(1-\phi)\mathcal{F}(\theta)} \ \prod_{u=1}^{s} e^{\lambda_u a_u(\theta)} d\mu(\theta)\Big\},$$

$k = 1, 2, ..., s.$

Moreover,

(i) if the integral in the first equality has a closed-form solution, then the rest of the multipliers can be found from the relations:

$$\frac{\partial \lambda_0}{\partial \lambda_k} = \bar{a}_k, \qquad k = 1, 2, ..., s,$$

(ii) the formula

$$\phi \mathcal{V}_{1,1,1}(\pi^*_\phi) + (1-\phi)[\mathcal{V}_{0,0,0}(\pi^*_\phi) - 2\mathcal{V}_{0,0,1}(\pi^*_\phi)] = 1 - \sum_{k=0}^{s} \lambda_k \bar{a}_k,$$

holds for all $0 \leq \phi \leq 1$.

Very often, experimenters are concerned with assigning weights $\bar{a}_k$, $k = 1, 2, ..., s$, to regions A_k, $k = 1, 2, ..., s$, to express, according to experience, how likely it is that θ belongs to each region. The following result, based on Proposition 4.3, characterizes Good-Bernardo-Zellner's priors when such a supplementary information comes in the form of quantiles, and both $\mathcal{I}(\theta)$ and $\mathcal{F}(\theta)$ are constant. Under such assumptions, the non-linear system of $s+1$ equations given in Proposition 4.3 is transformed into a homogeneous linear system of the same dimension as shown below:

Proposition 4.4 *Suppose that the sets* $A_k = (b_k, b_{k+1}]$, $k = 1, 2, ..., s-1$ *and* $A_s = (b_s, b_{s+1})$ *with* $b_1 < b_2 < \cdots < b_{s+1}$, $s \geq 2$, *constitute a partition of* Θ, $0 < \mu(\Theta) < \infty$. *Suppose also that both* $\mathcal{I}(\theta)$ *and* $\mathcal{F}(\theta)$ *are constant. Let* $\bar{a}_1, \bar{a}_2, ..., \bar{a}_s > 0$ *be such that* $\sum_{k=1}^{s} \bar{a}_k = 1$, *and* $\int I_{A_k}(\theta)\pi(\theta)d\mu(\theta) = \bar{a}_k$, $k = 1, 2, ..., s$. *If we define new multipliers:* $\omega_0 = e^{1-\lambda_0}/D$ *where* $D = [\mathcal{I}(\theta)]^{\frac{\phi}{2}}\ e^{(1-\phi)\mathcal{F}(\theta)}$, *and* $\omega_k = e^{\lambda_k}$, $k = 1, 2, ..., s$. *Then,* $\Omega = (\omega_0, \omega_1, ..., \omega_s)$ *can be found from the following homogeneous linear system:*

$$\begin{pmatrix} -1 & u_1 & u_2 & \ldots & u_s \\ -1 & v_1 & 0 & \ldots & 0 \\ -1 & 0 & v_2 & \ldots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ -1 & 0 & 0 & \ldots & v_s \end{pmatrix} \begin{pmatrix} \omega_0 \\ \omega_1 \\ \omega_2 \\ \vdots \\ \omega_s \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \tag{20}$$

where $u_k = \mu(A_k)$, *and* $v_k = \bar{a}_k^{-1} u_k$, $k = 1, 2, ..., s$.

Observe that the determinant, Δ, of the matrix in (20) is given by

$$\Delta = \left(\frac{\sum_{k=1}^{s} \bar{a}_k - 1}{\prod_{k=1}^{s} \bar{a}_k} \right) \prod_{k=1}^{s} u_k,$$

which guarantees that there exists a unique nontrivial solution since $\sum_{k=1}^{s} \bar{a}_k = 1$. In this case, the solution is $\Omega^{*T} = (1, v_1^{-1}, v_2^{-1}, ..., v_s^{-1})$, and $\pi_\phi^* = \sum_{k=1}^{s} v_k^{-1} I_{A_k}$.

The following proposition extends Good-Bernardo-Zellner's priors to a richer family by using the MMDIP and MFIP criteria:

Proposition 4.5 *Let*

$$\mathcal{N}_{\phi,\psi}(\pi) \overset{def}{=} \phi \mathcal{V}_{1,1,1}(\pi) + (1-\phi)(1-\psi)\mathcal{V}_{0,0,0}(\pi) + (\psi(1-\phi)/2)[\mathcal{V}_{0,1,1} + \mathcal{V}_{0,1,0}],$$

$0 \le \phi, \psi \le 1$. *Then*

(i) $\mathcal{N}_{\phi,\psi}(\pi) \in \mathcal{A}$ *and is concave w.r.t.* π.

(ii) A necessary condition for π *to be a maximum of the problem*

$$\textit{Maximize} \quad \mathcal{N}_{\phi,\psi}(\pi)$$

$$\textit{subject to} \quad \mathcal{C} : \int a_k(\theta)\pi(\theta)d\mu(\theta) = \bar{a}_k, \quad k = 0, 1, 2, ..., s,$$

where $a_0 \equiv 1 = \bar{a}_0$, *is given by*

$$\begin{aligned} \pi_{\phi,\psi}^*(\theta) \propto & [\mathcal{I}(\theta)]^{\frac{\phi}{2}} \exp\Bigg\{ (1-\phi)(1-\psi)\mathcal{F}(\theta) \\ & + \frac{\psi(1-\phi)}{2} \left[[\mathcal{I}(\theta)]^{\frac{1}{2}} + \frac{\mathcal{F}(\theta)}{[\mathcal{I}(\theta)]^{\frac{1}{2}}} \right] + \sum_{k=0}^{s} \lambda_k a_k(\theta) \Bigg\}, \end{aligned} \tag{21}$$

where λ_k, $k = 0, 1, ..., s$, *are the Lagrange multipliers associated with the constraints* $\mathcal{C}$.

The second term inside the exponential of (21) is the average between Fisher's information and the negative relative Shannon-Fisher's information. Notice that $\pi_{\phi,0}^*(\theta)$ is just Good-Bernardo-Zellner's prior.

In the following proposition, Good-Bernardo-Zellner type priors are derived as MAXENTP solutions by treating (5) and (8) as constraints (for the rationale of MAXENTP methods see Jaynes' 1982 seminal paper).

Proposition 4.6 *Consider the Jaynes–Good–Bernardo–Zellner problem:*

$$\textit{Maximize} \quad \mathcal{V}_{0,0,1}(\pi)$$

$$\textit{subject to:} \begin{cases} \mathcal{V}_{1,1,1}(\pi) - \mathcal{V}_{0,0,1}(\pi) = \bar{b}_1, \\ \mathcal{V}_{0,0,0}(\pi) - \mathcal{V}_{0,0,1}(\pi) = \bar{b}_2, \\ \displaystyle\int a_k(\theta)\pi(\theta)d\mu(\theta) = \overline{a}_k, \quad k = 0, 1, 2, ..., s, \ a_0 \equiv 1 = \overline{a}_0. \end{cases}$$

Then a necessary condition for a maximum is

$$\pi^*(\theta) \propto [\mathcal{I}(\theta)]^{\frac{\rho_1}{2}} \exp\{\rho_2\mathcal{F}(\theta) + \sum_{k=0}^{s} \lambda_k a_k(\theta)\}, \tag{22}$$

where $\rho_j,\ j = 1, 2,$ and $\lambda_k,\ k = 0, 1, ..., s,$ are the Lagrange multipliers associated with the constraints.

Unlike the coefficients ϕ and $1-\phi$ appearing in (4.6), the multipliers $\rho_j,\ j = 1, 2$, do not necessarily add up to 1.

There typically exist priors for which Shannon-Jaynes entropy becomes infinite. One way to overcome this problem consists of discounting entropy at a constant rate $\nu > 0$. The following proposition introduces Good-Bernardo-Zellner's controlled priors as solutions of maximizing discounted entropy:

Proposition 4.7 *Consider the discounted version of the problem stated in the preceding proposition:*

$$\textit{Maximize} \ -\int e^{-\nu\theta}\pi(\theta)\log\pi(\theta)d\mu(\theta),$$

subject to:

$$\begin{cases} \frac{1}{\pi(\theta)}\frac{dh_1(\theta)}{d\mu(\theta)} = \log[\mathcal{I}(\theta)]^{\frac{1}{2}}, \quad h_1(-\infty) = 0, \\ \qquad\qquad h_1(\infty) = \mathcal{V}_{1,1,1}(\pi) - \mathcal{V}_{0,0,1}(\pi) < \infty, \\ \\ \frac{1}{\pi(\theta)}\frac{dh_2(\theta)}{d\mu(\theta)} = \mathcal{F}(\theta), \quad h_2(-\infty) = 0, \\ \qquad\qquad h_2(\infty) = \mathcal{V}_{0,0,0}(\pi) - \mathcal{V}_{0,0,1}(\pi) < \infty, \\ \\ \frac{1}{\pi(\theta)}\frac{dg_k(\theta)}{d\mu(\theta)} = a_k(\theta), \ g_k(-\infty) = 0, \ g_k(\infty) < \infty, \ k = 0, 1, 2, ..., s \end{cases}$$

where $a_0 \equiv 1 = \overline{a}_0$. Then, a necessary condition for $\pi^*(\theta)$ to be an optimal control is given by

$$\pi^*(\theta) \propto [\mathcal{I}(\theta)]^{\frac{\rho_1(\theta)}{2}} \exp\{\rho_2(\theta)\mathcal{F}(\theta) + \sum_{k=0}^{s} \lambda_k(\theta)a_k(\theta)\}, \tag{23}$$

where $\rho_j(\theta) = \rho_{j0}e^{\nu\theta}$, $j = 1,2$, and $\lambda_k(\theta) = \lambda_{k0}e^{\nu\theta}$, $k = 0,1,...,s$, are the costate variables associated with the state variables $h_j(\theta)$, $j = 1,2$, and $g_k(\theta)$, $k = 0,1,...,s$, respectively. Furthermore, the constants ρ_{j0}, $j = 1,2$, and λ_{k0}, $k = 0,1,...,s$, can be computed from the following non-linear system of $s+3$ equations:

$$1 + \log h_1(\infty) = \log\left\{\int \log[\mathcal{I}(\theta)]^{\frac{1}{2}} m(\rho_{10}, \rho_{20}, \lambda_{00}, \lambda_{10}, ..., \lambda_{s0}; \theta)d\mu(\theta)\right\},$$
$$1 + \log h_2(\infty) = \log\left\{\int \mathcal{F}(\theta) m(\rho_{10}, \rho_{20}, \lambda_{00}, \lambda_{10}, ..., \lambda_{s0}; \theta)d\mu(\theta)\right\},$$
$$1 + \log g_k(\infty) = \log\left\{\int a_k(\theta) m(\rho_{10}, \rho_{20}, \lambda_{00}, \lambda_{10}, ..., \lambda_{s0}; \theta)d\mu(\theta)\right\},$$

$k = 0,1,2,...,s$

where

$$m(\rho_{10}, \rho_{20}, \lambda_{00}, \lambda_{10}, ..., \lambda_{s0}; \theta) = \left([\mathcal{I}(\theta)]^{\frac{\rho_{10}}{2}}\ e^{\rho_{20}\mathcal{F}(\theta)}\ e^{\lambda_{00}} \prod_{u=1}^{s} e^{\lambda_{u0}a_u(\theta)}\right)^{e^{\nu\theta}}.$$

5 Kalman filtering priors

In this section, we will study Good-Bernardo-Zellner's priors as Kalman Filtering priors (Kalman 1960, and Kalman and Bucy 1961). We will continue to work with the single parameter case, and focus our attention on both the location and scale parameter families.

Let $Y_1, Y_2, ..., Y_t$ be a set of indirect measurements, from a polling system or a sample survey, of an unobserved state variable β_t. The objective is to make inferences about β_t. The relationship between Y_t and β_t is specified by the measurement equation, sometimes also called the observation equation:

$$Y_t = A_t\beta_t + \varepsilon_t, \tag{24}$$

where $A_t \neq 0$ is known, and ε_t is the observation error distributed as $\mathcal{N}(0, \sigma^2_{\varepsilon_t})$ with $\sigma^2_{\varepsilon_t}$ known. Notice that the main difference between the measurement equation and the linear model is that, in the former, the coefficient β_t changes with time. Furthermore, we suppose that β_t is driven by a first order autoregressive process, that is,

$$\beta_t = Z_t \beta_{t-1} + \eta_{t-1}, \tag{25}$$

where $Z_t \neq 0$ is known, and $\eta_t \sim \mathcal{N}(0, \sigma^2_{\eta_t})$ with $\sigma^2_{\eta_t}$ known. In what follows, we will assume that β_0, ε_t, and η_t are independent random variables. We might state nonlinear versions of (24) and (25), but this would not make any essential differences in the subsequent analysis.

Suppose now, that at time $t = 0$, supplementary information is given by $\widehat{\beta}_0$ and $\widehat{\sigma}^2_0$, the mean and variance of β_0 respectively. That is,

$$\mathcal{C} \ : \ \begin{cases} \displaystyle\int_{-\infty}^{\infty} \pi(\beta_0) d\beta_0 = 1, \\ \displaystyle\int_{-\infty}^{\infty} \beta_0 \pi(\beta_0) d\beta_0 = \widehat{\beta}_0, \\ \displaystyle\int_{-\infty}^{\infty} (\beta_0 - \widehat{\beta}_0)^2 \pi(\beta_0) d\beta_0 = \widehat{\sigma}^2_0. \end{cases} \tag{26}$$

In this case, Good-Bernardo-Zellner's prior is given by

$$\pi^*_\phi(\beta_0) \propto [\mathcal{I}(\beta_0)]^{\frac{\phi}{2}} \exp\{(1-\phi)\mathcal{F}(\beta_0) + \lambda_0 + \lambda_1 \beta_0 + \lambda_2 (\beta_0 - \widehat{\beta}_0)^2\}, \tag{27}$$

where $\lambda_j, \ j = 0, 1, 2$, are Lagrange multipliers.

Suppose that, at time t, we wish to make inferences about the conditional state variable $\theta_t = \beta_t | I_t$, where $I_t = \{Y_1, Y_2, ..., Y_{t-1}\}$. To obtain a posterior distribution of θ_t, the information provided by the measurement Y_t, with density $f(Y_t|\theta_t)$, is used to modify the initial knowledge in $\pi^*_\phi(\theta_t)$ according to Bayes' theorem:

$$f(\theta_t|Y_t) \propto f(Y_t|\theta_t) \pi^*_\phi(\theta_t). \tag{28}$$

We are now in a position to state the Bayesian recursive updating procedure of the Kalman Filter (KF) for both the location and scale parameter families $f(Y_t|\theta) = f(Y_t - \theta)$, $\theta \in \mathbb{R}$, and $f(Y_t|\theta) = (1/\theta) f(Y_t/\theta)$, $\theta > 0$, respectively. To start off the KF procedure, we substitute (27) in (26), obtaining that Good-Bernardo-Zellner's prior at time $t = 0$, is given by $\mathcal{N}(\widehat{\beta}_0, \widehat{\sigma}^2_0)$, which is describing the initial knowledge of the system. Proceeding inductively, at time t, $\widehat{\beta}_{t-1}$ and

$\widehat{\sigma}^2_{t-1}$ become *supplementary* information, and therefore Good-Bernardo-Zellner's prior at time t is represented by

$$\theta_t = \beta_t | I_t \sim \mathcal{N}(Z_t \widehat{\beta}_{t-1}, M_t), \tag{29}$$

where

$$M_t = Z_t^2 \widehat{\sigma}^2_{t-1} + \sigma^2_{\eta_{t-1}}. \tag{30}$$

The sampling model (or likelihood function) is determined by

$$Y_t | \theta_t \sim \mathcal{N}(A_t \beta_t, \sigma^2_{\varepsilon_t}). \tag{31}$$

The posterior distribution, at time t, is then obtained by substituting both (29) and (30) in (28), so

$$f(\theta_t | Y_t) \propto \ \exp\bigl\{-\tfrac{1}{2}[(A_t \beta_t - Y_t)^2 \sigma^{-2}_{\varepsilon_t} + (\beta_t - Z_t \widehat{\beta}_{t-1})^2 M_t^{-1}]\bigr\}.$$

Noting that $\pi^*_\phi(\theta_t)$ is a natural conjugate prior, we may complete the squares to get

$$\theta_t | Y_t \sim \mathcal{N}\bigl[Z_t \widehat{\beta}_{t-1} + K_t(Y_t - A_t Z_t \widehat{\beta}_{t-1}), M_t - K_t A_t M_t\bigr],$$

where

$$K_t = M_t A_t (\sigma^2_{\varepsilon_t} + A_t^2 M_t)^{-1}. \tag{32}$$

This, of course, means that

$$\left\{ \begin{array}{l} \widehat{\beta}_t = Z_t \widehat{\beta}_{t-1} + K_t(Y_t - A_t Z_t \widehat{\beta}_{t-1}), \\ \widehat{\sigma}^2_t = M_t - K_t A_t M_t. \end{array} \right. \tag{33}$$

We then proceed with the next iteration. Equations (33), (30), and (32) are known in the literature as the KF.

The above analysis can be summarized in the following proposition:

Proposition 5.1 *Consider the state-space representation:*

$$\left\{ \begin{array}{l} Y_t = A_t \beta_t + \varepsilon_t, \\ \\ \beta_t = Z_t \beta_{t-1} + \eta_{t-1}, \end{array} \right.$$

defined as in (24) and (25). Suppose that supplementary information on the mean and variance of β_0 *is available. Let* $\theta_t = \beta_t | I_t$*, where*

$I_t = \{Y_1, Y_2, ..., Y_{t-1}\}$, *and consider the location and scale parameter families* $f(Y_t|\theta) = f(Y_t - \theta)$, $\theta \in \mathbb{R}$, *and* $f(Y_t|\theta) = (1/\theta)f(Y_t/\theta)$, $\theta > 0$, *respectively, along with the properties stated in Corollary 4.1. Then, under Good-Bernardo-Zellner's prior,* $\pi^*_\phi(\theta_t)$*, the posterior estimate of* β_t, $\widehat{\beta}_t$, *is given by*

$$\widehat{\beta}_t = \omega_t Z_t \widehat{\beta}_{t-1} + (1 - \omega_t)(Y_t/A_t),$$

where $\omega_t = \sigma^2_{\varepsilon_t}(\sigma^2_{\varepsilon_t} + A^2_t M_t)^{-1}$.

6 Revisiting the normal linear model

The results on Good-Bernardo-Zellner priors given so far can be easily extended to the multi-dimensional parameter case, namely, $\theta = (\theta_1, \theta_2, ..., \theta_m) \in \Theta \subseteq \mathbb{R}^m$, $m > 1$. Consider a vector of independent and identically distributed normal random variables $(X_1, X_2, ..., X_n)$ with common and known variance σ^2 satisfying

$$\text{(34)} \qquad \mathrm{E}(X_k) = a_{k1}\theta_1 + a_{k2}\theta_2 + \cdots + a_{km}\theta_m, \qquad k = 1, 2, ..., n,$$

where $A = (a_{ij})$ is a matrix of known coefficients for which $(A^T A)^{-1}$ exists.

Let X and θ stand for the column vectors of variables X_k and parameters θ_j, respectively. Then (34) can be written in matrix notation as, $\mathrm{E}(X) = A\theta$. In this case, we have

$$\text{(35)} \qquad f(\xi|\theta) = (\tfrac{1}{2\pi\sigma^2})^{\frac{n}{2}} \exp\{-\tfrac{1}{2\sigma^2}\|\xi - A\theta\|^2\},$$

where $\xi = (x_1, x_2, ..., x_n)$. Since σ^2 has been assumed known, only the location parameter is unknown. The analogue of (2) is now given by the matrix:

$$\begin{aligned}\mathcal{I}_n(\theta) \equiv &\left(\int \Big(\tfrac{\partial}{\partial\theta_\jmath} \log f(x|\theta)\Big)\Big(\tfrac{\partial}{\partial\theta_\ell} \log f(x|\theta)\Big) f(x|\theta) d\lambda(x)\right)_{1\leq \jmath,\ell \leq m} \\ = &\frac{1}{\sigma^2} A^T A,\end{aligned}$$

and so $\det[\mathcal{I}_n(\theta)]$ is constant, which implies that the Good-Bernardo-Zellner prior distribution $\pi^*_\phi(\theta)$, describing a situation of vague information on θ, must be a locally uniform prior distribution.

Let $\widehat{\theta}$ be the least squares estimate for θ, then it is known that $A^T A\widehat{\theta} = A^T X$, $\mathrm{E}(\widehat{\theta}\ ) = \theta$, and $\mathrm{Var}(\widehat{\theta}\ ) = \sigma^2(A^T A)^{-1}$. Noting from equation (35) that

$$f(\xi|\theta) = (\tfrac{1}{2\pi\sigma^2})^{\frac{n}{2}} \exp\{-\tfrac{1}{2\sigma^2}(\|\xi - A\widehat{\theta}\ \|^2 + \langle A^T A(\theta - \widehat{\theta}\ ), \theta - \widehat{\theta}\ \rangle)\},$$

and applying Bayes' theorem, we get as the posterior distribution of θ

$$f(\theta|\xi) = (2\pi)^{-\frac{m}{2}}(\det[\tfrac{1}{\sigma^2}A^T A])^{\frac{1}{2}} \exp\big\{-\tfrac{1}{2}\big\langle \tfrac{1}{\sigma^2}A^T A(\theta - \widehat{\theta}\ ), \theta - \widehat{\theta}\ \big\rangle\big\}.$$

If supplementary information in mean, c, and variance-covariance matrix, D, is now incorporated, then the (informative) Good-Bernardo-Zellner prior is given by

$$\pi^*_\phi(\theta) = (2\pi)^{-\frac{m}{2}}(\det[D])^{-\frac{1}{2}} \exp\big\{-\tfrac{1}{2}\langle D^{-1}(\theta - c), \theta - c\rangle\big\}.$$

The posterior distribution is now

$$\begin{aligned} f(\theta|\xi) &= (2\pi)^{-\frac{m}{2}}(\det[B])^{\frac{1}{2}} \\ &\times \exp\Big\{-\tfrac{1}{2}\langle B[\theta - ((DB)^{-1}c + \tfrac{1}{\sigma^2}B^{-1}A^T A\widehat{\theta}\ )], \\ &\qquad \theta - ((DB)^{-1}c + \tfrac{1}{\sigma^2}B^{-1}A^T A\widehat{\theta}\ )\rangle\Big\}, \end{aligned}$$

where $B = D^{-1} + \frac{1}{\sigma^2}A^T A$.

7 Summary and conclusions

We have presented, in a unified framework, a number of well-known methods that maximize a criterion functional to obtain non-informative and informative priors. Our general procedure is, by itself, capable of dealing with a range of interesting issues in Bayesian analysis. However, in this paper, we have limited our attention to Good-Bernardo-Zellner's priors as well as their application to some Bayesian inference problems, including the Kalman filter and the Normal linear model.

There exist priors for which Shannon-Jaynes entropy becomes infinite. In order to overcome this difficulty we proposed discounted entropy. We introduced Good-Bernardo-Zellner's controlled priors which maximize discounted entropy at a constant rate. Throughout the paper, we have emphasized the existence and uniqueness of the solutions of the corresponding variational and optimal control problems. There are, of course, many other members of the class $\mathcal{A}$ that deserve much more attention than that we have attempted here. Needless to say, more work will be required in this direction. Results will be reported elsewhere.

Acknowledgement

The author wishes to thank an anonymous referee for valuable guidance and numerous suggestions. The author is indebted to Arnold Zellner, José M. Bernardo, Jim Berger, and George C. Tiao for helpful comments on earlier drafts of this paper. As usual, the author bears sole responsability for opinions and errors.

Francisco Venegas-Martínez
Department of Finance,
Tecnológico de Monterrey,
Calle del Puente 222,
14380 México D. F.
fvenegas@itesm.mx

References

[1] Akaike H., *A new look at the Bayes procedure*, Biometrika **65** (1978), 53–59.

[2] Bayes T., *An essay towards solving a problem in the doctrine of chances*, Philos. Trans. R. Soc. London **53** (1763) 370–418. (Reprinted in Biometrika **45** (1958), 243–315.)

[3] Berger J. O.; Bernardo J. M., *Estimating a product of means: Bayesian analysis with reference priors*, J. Amer. Statist. Assoc. **84** (1989), 200–207.

[4] Berger J. O.; Bernardo J. M., *On the development of reference priors*, Bayesian Statistics **4** (1992a), 35–60.

[5] Berger J. O.; Bernardo J. M., *Ordered group reference priors with application to a multinomial problem*, Biometrika, **79** (1992b), 25–37.

[6] Berger J. O.; Bernardo J. M.; Mendoza M., *On priors that maximize expected information*, Recent Developments in Statistics and Their Applications (1989), 1–20.

[7] Bernardo J. M., *Noninformative priors do not exist*, J. Statist. Plann. Inference **B65** (1997), 177–189.

[8] Bernardo J. M., *Reference posterior distributions for Bayesian inference*, J. Roy. Statist. Soc. Ser. B Stat. Methodol. **41** (1979), 113–147.

[9] Bernardo J. M.; Ramón J. M., An Introduction to Bayesian Reference Analysis: Inference on the Ratio of Multinomial Parameters, Tech. Rep. 3–97, Universitat de València, Spain, 1997.

[10] Bernardo J. M.; Smith A. F. M., Bayesian Theory, John Wiley & Sons, Wiley Series in Probability and Mathematical Statistics, New York, 1994.

[11] Box G. E. P.; Tiao G. C., Bayesian Inference and Statistical Analysis, Addison-Wesley Series in Behavioral Science: Quantitative Methods, Massachusetts, 1973.

[12] Geisser S., Predictive Inference: An Introduction, Chapman & Hall, New York, 1993.

[13] Good I. J., *Utility of a distribution*, Nature **219** (1968), 1392.

[14] Good I. J., *What is the use of a distribution?*, Multivariate Analysis (Krishnaiah, ed.) Vol. II, Academic Press, New York (1969), 183–203.

[15] Jaynes E. T., *Information theory and statistical mechanics*, Phys. Rev. **106** (1957), 620–630.

[16] Jaynes E. T., *On the rationale of maximum-entropy methods*, Proc. of the IEEE **70** (1982), 939–952.

[17] Jeffreys H., Theory of Probability, 3rd. edition, Oxford University Press, Oxford, 1961.

[18] Kalman R. E., *A new approach to linear filtering and prediction problems*, Transactions ASME, Series D, J. of Basic Engineering **82** (1960), 35–45.

[19] Kalman R. E.; Bucy R., *New results in linear filtering and prediction theory*, Transactions ASME, Series D, J. of Basic Engineering **83** (1961), 95–108.

[20] Kullback S., Information Theory and Statistics, Wiley, New York, 1959.

[21] Lindley D. V., *On a measure of information provided by an experiment*, Ann. Math. Statist. **27** (1956), 986–1005.

[22] Rényi A., Foundations of Probability, Holdan-Day, San Francisco, 1970.

[23] Soofi E. S., *Capturing the intangible concept of information*, J. Amer. Statist. Assoc. **89** (1994), 1243–1254.

[24] Zellner A., An Introduction to Bayesian Inference in Econometrics, Wiley, New York, 1971.

[25] Zellner A., *Bayesian method of moments (BMOM) analysis of mean and regression models*, Bayesian Analysis in Econometrics and Statistics, Published by Edward Elgar Publishing Limited, UK (1997), 291–304. (Also available in Prediction and Modelling Honouring Seymor Geisser, Springer-Verlag, New York, Berlin, Heidelberg, (1996).)

[26] Zellner A., *Bayesian methods and entropy in economics and econometrics*, Maximum Entropy and Bayesian Methods, Dordrecht, Netherlands: Kluwer (1996), 17–31.

[27] Zellner A., *Maximal data information prior distributions*, New Developments in the Application of Bayesian Methods, Amsterdam: North-Holland (1977), 201–232.

[28] Zellner A., *Models, prior information and Bayesian analysis*, J. Econometrics **75** (1996a), 51–58.

[29] Zellner A., *Past and recent results on maximal data information priors*, J. Statist. Plann. Inference **49** (1996b), 3–8.

[30] Zellner A., *The finite sample properties of simultaneous equations' estimates and estimators: Bayesian and non-Bayesian approaches*, J. Econometrics **83** (1998), 185-212.