Variants of a natural conjugate prior for estimating a high-dimensional mean vector

Takemi Yanagimoto[1] and Toshio Ohnishi[2]
Institute of Statistical Mathematics[1] and
Graduate University for Advanced Studies[1),2)]
4-6-7 Minami-Azabu, Minato-ku, Tokyo, JAPAN
yanagmt@ism.ac.jp[1] and *ohnishi@ism.ac.jp*[2]

1. Introduction

The estimation problem of a high-dimensional mean vector is one of the most attractive ones from both the theoretical and the practical viewpoints. The empirical Bayes method provides us with a reasonable estimate, though a frequentist approach is also possible, see Yanagimoto (2000), for example.

A practical problem in applying the empirical Bayes method is concerned with the choice of a prior distribution. The empirical Bayes estimator can depend largely on the choice. A prior distribution can be chosen subjectively, or in some circumstances there may exist a reasonable evidence suggesting a prior distribution. It looks, however, more likely that other reasons such as numerical and analytical conveniencies are the key to choosing a prior distribution. A natural conjugate prior (Raiffa and Shlaifer, 1961) has been widely employed because of its simple form of the posterior mean. The aim of this talk is to pursue variants of a natural conjugate prior, which permits us to choose a prior distribution from a wider family of prior distributions having various convenient properties. In practical situations the most important property is probably its ease of interpretation on the assumed prior distribution.

2. Natural conjugate prior and its dual

Let $p(x;\, \mu)$ be a density (or probability) function in the exponential family with the form,

$$p(x;\, \mu) = \exp\{x\eta - \psi(\eta) + a(x)\}$$

where μ denotes the mean. A natural conjugate prior is defined, when it makes sense, as

$$\pi_n(\mu;\, \delta) \propto \exp \delta\{m\eta - \psi(\eta)\} \tag{2.1}$$

where m is a known mean. Recall that a natural conjugate prior is assumed on the natural parameter η rather than the mean parameter, though there exists one-to-one correspondence between the two parameters.

To derive another prior distribution dual to a natural conjugate prior we consider here the Kullback-Leibler separator between $p(x;\, \mu_1)$ and $p(x;\, \mu_2)$, which is expressed as $D(\mu_1,\, \mu_2) = E\{\log(p(x;\, \mu_1)/p(x;\, \mu_2)) \mid p(x;\, \mu_1)\}$. Suppose that the density function is in the exponential dispersion model, that is, $p(x;\, \mu,\, \tau_0) = \exp\{\tau_0(x\eta - \psi(\eta)) + a(x,\, \tau_0)\}$. Then it is shown that a natural conjugate prior is expressed as

$$\pi_n(\mu;\, \delta) \propto \exp\{-\delta D(m,\, \mu)\}\,. \tag{2.2}$$

Recall that there exists another separator, $D(\mu,\, m)$, which is dual to $D(m,\, \mu)$. Thus replacing $D(m,\, \mu)$ in (2.2) by $D(\mu,\, m)$, we obtain another prior density as

$$\pi_m(\mu;\, \delta) \propto \exp\{-\delta D(\mu,\, m)\}\,. \tag{2.3}$$

We will call this prior a mean conjugate prior. This is because this prior density is defined on the mean parameter, and also because the dual structure between the two separators, $D(\mu,\, m)$

and $D(m, \mu)$, corresponds to that between the canonical parameter and the mean parameter. Recall that a natural conjugate prior is also referred to as a canonical conjugate prior.

When the density function is also in the reproductive exponential distribution (Barndorff-Nielsen 1983), a mean conjugate prior density is written as

$$\pi_m(\mu;\, \delta) \propto p(x;\, m,\, \delta). \tag{2.4}$$

In other words the sample density function is equivalent to a prior density function. This notable property suggests further definitions of prior densities.

3. Variants of a mean conjugate prior

It looks naive that we assume prior densities of the form same as the sample density function. In the previous section we learned that such a naive prior density is related with a mean conjugate prior. This naive method provides us with a prior density in various situations, where the sample distribution has the dispersion parameter in addition to the mean parameter. This naive choice is available, when the sample and the mean parameter spaces are common.

Example 3.1. Suppose that the sample density is in a location-scale family with a known scale parameter σ_0, $p((x-\mu)/\sigma_0)/\sigma_0$. Then a naive choice is $\pi(\mu;\, \delta) = p((\mu - m)/\delta)/\delta$.

Example 3.2. Suppose that the sample density is a beta distribution, that is $p(x;\, \mu) = x^{\alpha-1}(1-x)^{\beta-1}/\mathrm{Be}(\alpha,\, \beta)$, where $\mu = \alpha/\alpha{+}\beta(>0)$ and $\alpha+\beta(>0)$ are known. Then a naive choice is given for $0 < m < 1$ as $\pi(\mu;\, \delta) = \mu^{\delta m-1}(1-\mu)^{\delta(1-m)-1}/\mathrm{Be}(\delta m,\, \delta(1-m))$.

Next, we discuss the case where the sample density does not contain a dispersion parameter case. Power prior distribution, which was introduced by Ibrahim and Chen (2000) to extend a natural conjugate prior, can be applied in this case. It results in $\pi(\mu;\, \delta) = p(\mu;\, m)^{\delta}$.

4. Advantages

A notable advantage of new sets of prior distributions is to enrich the selection of prior distributions, when an informative prior distribution is necessary. Although the choice of a prior distribution depends on various reasons, a variety of prior distributions are obviously convenient for the estimation of a high-dimensional mean vector.

Another advantage of newly introduced prior distributions is in its ease of interpretation on the distribution. Consider the case of Examples 3.2., for example. Suppose that we need a prior distribution on the mean μ in the beta distribution, and that there is no firm informations choosing a prior distribution. Then the beta distribution is probably a naive choice.

The most attractive property of a natural conjugate prior is its simple form of the posterior mean of μ and its relatively small amounts of computations. The computational difficulty is not serious because of recent improvements of the computational environments. Interestingly, in the empirical Bayes method the estimation of the hyperparameter δ under a mean conjugate prior can be easier than that under a natural conjugate prior.

REFERENCES

Barndorff-Nielsen, O.E. and Blaesild, P. (1983). Reproductive exponential families. Ann. Statist., 11, 770-782.

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

Raiffa, H. and Shlaifer, R. (1961). Applied Statistical Decision Theory. Graduate School of Business Administration, Harvard Univ. Boston.

Yanagimono, T. (2000). A pair of estimating equations for a mean vector. Statist. Prob. Letters, 50, 97-103.