INFERENCE FROM NONINFORMATIVE ML-II PRIORS

Paweł Kopciuszewski

Institute of Mathematics and Computer Science
Częstochowa University of Technology

Abstract. The type II maximum likelihood (ML-II) is considered in this paper. The problem of finding the ML-II prior is too complex, in many cases. But we propose some methods of approximation ML-II prior. Both noninformative and informative ML-II priors are considered. If no information is given about unknown prior then we will construct a proper density which is approximately ML-II prior. The theorem which let us approximate ML-II prior belonging to the given class of densities is formulated. The methods of approximation ML-II prior are simply and easy to applied. All required calculations are done by MCMC algorithms.

1. Introduction

Let $\mathcal{X} \in \mathcal{R}^n$ be an observable random variable with density $f(x|\theta) > 0$, for some unknown $\theta \in \Theta \subset \mathcal{R}^p$. The choice of prior distribution π for parameter $\theta \in \Theta$ is considered, here. π has a density with respect to a σ-finite measure ν

$$\pi(dx) = \pi(x)\nu(dx)$$

For simplicity, π will be used to denote both the distribution and the density of parameter $\theta \in \Theta$.

Definition 1. $\pi(\theta)$ *is called the improper density if* $\int_{\Theta} \pi(d\theta) = \infty$.

Methods of prior density choice are the most criticized point of Bayesian analysis. The choice of prior distribution is most often done subjectively. It is justified with the simplicity of calculations of some characteristics from posterior densities. There are noninformative and informative priors. Methods of noninformative prior choice are: Jeffreys prior, invariance under reparametrization and conjucate priors. Informative priors are chosen by maximum entropy method and ML-II method.
Jeffreys prior is based on Fisher Information given by $I(\theta)$,

$$I(\theta) = E[\frac{\partial \log f(x|\theta)}{\partial \theta}]^2 = -E[\frac{\partial^2 \log f(x|\theta)}{\partial \theta^2}]$$

Jeffreys prior $\pi(\theta)$ is given by

$$\pi(\theta) \propto [\det I(\theta)]^{0.5}$$

It is usually improper density.

Noninformative prior satisfies invariance under reparametrization if there exist functions Φ, Ψ, g such that, likelihood function $f(x|\theta)$ satisfies

$$f(x|\theta) = \Phi(\Psi(\theta) - q(x))$$

The uniform distribution is chosen for parameter $\Psi(\theta)$. Then the prior $\pi(\theta)$ is

$$\pi(\theta) \propto |\det(\tfrac{\partial \Psi^{-1}}{\partial \theta})|$$

A family of densities is said to be conjugate priors if for every prior density $\pi(\theta)$ from this class the posterior density $\pi(\theta|x)$ also belongs to this class. Maximum entropy method chooses prior distribution by maximizing functional $E(\pi)$

$$E(\pi) = -\int_{\Theta} \log[\pi(\theta)]\pi(d\theta)$$

This method will be able to applied if the likelihood function $f(x|\theta)$ is not given. It is used to choose prior distribution when some prior moments or prior quantiles are known.

In the paper ML-II method is considered. ML-II method chooses a prior distribution π by maximizing marginal density $f(x)$

$$f(x) = \int_{\Theta} f(x|\theta)\pi(d\theta) \tag{1}$$

It is a generalization of maximum likelihood method. The choice of prior distribution depends on the vector $x \in \mathcal{R}^n$. For that reason it is more objective method than the others. It was introduced by Good [4,5]. Berger and Berliner [1] analyzed robustness properties of ML-II posterior estimators. Sivaganesan and Berger [11] utilized ML-II method in Bayesian robustness studies. They determine the ranges of posterior quantities. Chaturvedi [2] utilized ML-II procedure to robust Bayesian analysis of the linear model. For ε-contamination class the ML-II posterior mean was estimated. Sivaganesan [10] found ranges of the posterior mean, posterior median and posterior mode and calculated the supremum of the posterior mean squared error for the ML-II posterior mean. Moreno and Carmona [8] considered the ML-II prior selection related to several ε-contamination classes. They considered the class of all distributions with known some quantiles and the class of unimodal distributions with some specified quantiles. Lee [7] applied ML-II procedure to estimating hyperparameters involved in conjugate priors. Waal and Nel [14] utilized ML-II method to derive a prior distribution from ε-contamination class multivariate distributions and multivariate normal distributions. Gosh and Kim [6] proposed some robust Bayes estimators of finite population mean using ML-II priors.

In the paper the theorems which allows us to approximate ML-II prior are formulated. The noninformative ML-II priors are considered, only.

2. An approximation of noninformative ML-II priors

If we have no information about parameter θ then Θ is the set on which likelihood function $f(x|\theta)$ is defined.

Definition 2. *Let $\widehat{\Theta}$ denote the set defined as follows*

$$\widehat{\Theta} = \{\widehat{\theta} \in \Theta : \ f(x|\widehat{\theta}) = \sup_{\theta \in \Theta} f(x|\theta)\} \tag{2}$$

From the general theorem [8] it results that for all prior distributions π we have

$$\int_{\Theta} f(x|\theta)\pi(d\theta) \leq f(x|\widehat{\theta}) \tag{3}$$

Equality holds in (3) if and only if prior distribution π is concentrated on the set $\widehat{\Theta}$. From this theorem it follows that If no information is given about prior distribution π, then ML-II prior is concentrated on the set $\widehat{\Theta}$. In most cases the set $\widehat{\Theta}$ consists of one point, only. Hence ML-II prior is concentrated on one point.

The set $\widehat{\Theta}$ is difficult to directly estimating, in many cases. From the idea of Simulated Annealing algorithm we propose the method of approximating ML-II prior.

Theorem 1. *If $\pi(\theta|\lambda)$ is a proper continuous density satisfying*

$$\pi(\theta|\lambda) \propto \exp\{\lambda f(x|\theta)\} \tag{4}$$

where $\lambda > 0$,
then the marginal density $f(x|\lambda) = \int_{\Theta} f(x|\theta)\pi(d\theta|\lambda)$ is the non-decreasing function of the parameter λ satisfying

$$\lim_{\lambda \to \infty} f(x|\lambda) = f(x|\widehat{\theta}) \tag{5}$$

Proof

From the assumption of the theorem it follows that the density $f(x|\lambda)$ is

$$f(x|\lambda) = \frac{1}{\int\limits_{\Theta} \exp\{\lambda f(x|\theta)\}\nu(d\theta)} \int\limits_{\Theta} f(x|\theta)\exp\{\lambda f(x|\theta)\}\nu(d\theta) \tag{6}$$

The derivative $f'(x|\lambda)$ after parameter λ is

$$\frac{\int\limits_{\Theta} f^2(x|\theta)\exp\{\lambda f(x|\theta)\}\nu(d\theta) \int\limits_{\Theta} \exp\{\lambda f(x|\theta)\}\nu(d\theta) - [\int\limits_{\Theta} f(x|\theta)\exp\{\lambda f(x|\theta)\}\nu(d\theta)]^2}{[\int\limits_{\Theta} \exp\{\lambda f(x|\theta)\}\nu(d\theta)]^2} \tag{7}$$

The derivative $f'(x|\lambda)$ may be written in the form

$$f'(x|\lambda) = E_\pi f^2(x|\theta) - [E_\pi f(x|\theta)]^2 \tag{8}$$

Hence

$$f'(x|\lambda) \geq 0 \tag{9}$$

We conclude that the density $f(x|\lambda)$ is non-decreasing function of parameter λ. Let $\widehat{\Theta}_k$ be some ν-measurable sets satisfying

$$\begin{array}{ccc} \widehat{\Theta} & \subset & \widehat{\Theta}_k \\ \nu(\widehat{\Theta}_k) > 0 & \text{for all} & k < \infty \\ \lim\limits_{k\to\infty} \nu(\widehat{\Theta}_k \setminus \widehat{\Theta}) & = & 0 \end{array} \tag{10}$$

Let the function $g(x|\lambda, k)$ be defined as follows

$$g(x|\lambda, k) = \frac{\int\limits_{\Theta\setminus\widehat{\Theta}_k} f(x|\theta)\exp\{\lambda[f(x|\theta) - f(x|\widehat{\theta})]\}\nu(d\theta) + f(x|\widehat{\theta})\nu(\widehat{\Theta}_k)}{\int\limits_{\Theta\setminus\widehat{\Theta}_k} \exp\{\lambda[f(x|\theta) - f(x|\widehat{\theta})]\}\nu(d\theta) + \nu(\widehat{\Theta}_k)} \tag{11}$$

From (11) we see that for all $\lambda < \infty$

$$\lim_{k\to\infty} g(x|\lambda, k) = f(x|\lambda) \tag{12}$$

Notice that $f(x|\theta) \leq f(x|\widehat{\theta})$ for all $\theta \in \Theta$. Thus for all $k < \infty$

$$\lim_{\lambda\to\infty} g(x|\lambda, k) = f(x|\widehat{\theta}) \tag{13}$$

Hence

$$\lim_{\lambda\to\infty} f(x|\lambda) = \lim_{\lambda\to\infty}\lim_{k\to\infty} g(x|\lambda,k) = \lim_{k\to\infty}\lim_{\lambda\to\infty} g(x|\lambda,k)$$
$$\lim_{k\to\infty} f(x|\widehat{\theta}) = f(x|\widehat{\theta}) \tag{14}$$

The theorem is thus proved.
Notice that theorem 1 will be able to applied when $f(x|\theta)$ is not differentiable at the point $\widehat{\theta} \in \widehat{\Theta}$. From theorem 1 two corollaries may be concluded.

Corollary 1. *If we have no information about prior distribution π then the density (4) is approximately ML-II prior.*

Corollary 2. *If (4) is improper density then $\pi^*(\theta|\lambda)$ is approximately ML-II prior*

$$\pi^*(\theta|\lambda) \propto \exp\{\lambda f(x|\theta)\}\delta_{\Theta^*}(\theta) \tag{15}$$

where Θ^* denote the set satisfying

$$\begin{array}{rcl} \widehat{\Theta} & \subset & \Theta^* \\ \int_{\Theta} \pi^*(d\theta|\lambda) & < & \infty \end{array} \tag{16}$$

3. An estimation of some characteristics from posterior density

If prior density $\pi(\theta|\lambda)$ satisfies condition (4) then posterior density $\pi(\theta|\lambda,x)$ will satisfy

$$\pi(\theta|\lambda,x) \propto f(x|\theta)\exp\{\lambda f(x|\theta)\} \tag{17}$$

We use Hastings-Metropolis algorithm to generate the sample from posterior density $\pi(\theta|\lambda,x)$. Suppose that $q(\theta^{(k)},\theta^{(k+1)})$ is a transition density from the state $\theta^{(k)} \in \Theta$ to the state $\theta^{(k+1)} \in \Theta$. In the k-th step we generate a point $\theta^{(k+1)}$ from the density $q(\theta^{(k)},\theta^{(k+1)})$, and next we accept the point $\theta^{(k+1)}$ with probability

$$\alpha(\theta^{(k)},\theta^{(k+1)}) = \min\{\frac{q(\theta^{(k+1)},\theta^{(k)})f(x|\theta^{(k+1)})\exp\{\lambda f(x|\theta^{(k+1)})\}}{q(\theta^{(k)},\theta^{(k+1)})f(x|\theta^{(k)})\exp\{\lambda f(x|\theta^{(k)})\}},1\} \tag{18}$$

If the density $q(\theta^{(k)},\theta^{(k+1)})$ satisfies

$$q(\theta^{(k)},\theta^{(k+1)}) \propto f(x|\theta^{(k+1)}) \tag{19}$$

then acceptance probability $\alpha(\theta^{(k)},\theta^{(k+1)})$ will be

$$\alpha(\theta^{(k)},\theta^{(k+1)}) = \min\{\exp\{\lambda[f(x|\theta^{(k+1)}) - f(x|\theta^{(k)})]\},1\} \tag{20}$$

But if the density $q(\theta^{(k)}, \theta^{(k+1)})$ satisfies

$$q(\theta^{(k)}, \theta^{(k+1)}) \propto \exp\{\lambda f(x|\theta^{(k+1)})\} \tag{21}$$

the the acceptance probability $\alpha(\theta^{(k)}, \theta^{(k+1)})$ will be

$$\alpha(\theta^{(k)}, \theta^{(k+1)}) = \min\{\frac{f(x|\theta^{(k+1)})}{f(x|\theta^{(k)})}, 1\} \tag{22}$$

Some characteristics from posterior density $\pi(\theta|\lambda, x)$ will be able to estimated if the sample $\{\theta^{(k)}\}$ is generated.

References

[1] Berger J., Berliner L., Robust Bayes and empirical Bayes analysis with ε-contaminated priors, Ann. Statist., (1986), 14, 461-486.

[2] Chaturvedi A., Robust Bayesian analysis of the linear regression model. J. Statist. Plann. Infer., (1993), 50, 175-186.

[3] Evans M., Swartz T., Methods for approximating integrals in statistics with special emphasis on bayesian integration problems, Statist. Sci., (1995), 10, 254-272.

[4] Good I.J., The estimation of probabilities, M.I.T. Press, Cambridge (1965).

[5] Good I.J., Good Thinking: The Foundations of Probability and Its Applications, University of Minesota Press, Minneapolis, (1983).

[6] Gosh M., Kim D.H., Robust Bayes competitors of the ratio estimator. Statist. Decisions, (1997), 15, 17-36.

[7] Lee C.B., Bayesian analysis of a change-point in exponential families with applications. Comp. Statist. Data Anal., (1998), 27, 195-208.

[8] Moreno E., Carmona A.G., Empirical Bayes analysis for ε-contaminated priors with shape and quantile constraints. Rebrape, (1990), 4, 177-200.

[9] Robert C.P., The Bayesian choice. Springer-Verlag, New York, (1997).

[10] Sivaganesan S., Sensitivity of some standard Bayesian estimates to prior uncertainty: A comparison. J. Statist. Plann. Infer., (1991), 27, 85-103.

[11] Sivaganesan S., Berger J., Ranges of posterior measures for priors with unimodal contaminations, Ann. Statist., (1989), 17, 868-889.

[12] Ross M.S., Statistical modelling and decision science. Academic Press, Berkeley (1997).

[13] Tierney L., Markov chain for exploring posterior distributions Ann. Statist., (1994) 22, 1701-1762.

[14] Waal D.J., Nel D.G., A procedure to select a ML-II prior in a multivariate normal case, Comm. Statist. Simulation Comput., (1988), 17, 1021-1035.