ELSEVIER

Statistics & Probability Letters 56 (2002) 171–175

STATISTICS & PROBABILITY LETTERS

www.elsevier.com/locate/stapro

On a Bayesian aspect for soft wavelet shrinkage estimation under an asymmetric Linex loss

Su-Yun Huang

Institute of Statistical Science, Academia Sinica, 128 Yen Chiu Yuan Road S2, Taipei, 11529, Taiwan, Republic of China

Received March 2001; received in revised form August 2001

Abstract

Consider an asymmetric Linex loss. We provide the soft wavelet shrinkage estimation of a Bayesian interpretation under such a loss. © 2002 Elsevier Science B.V. All rights reserved.

Keywords: Admissibility; Generalized Bayes estimator; Linex loss; Minimaxity; Soft threshold; Wavelet shrinkage estimation

1. Introduction

Consider the following discrete noisy signal model obtained from a discrete wavelet transform:

$$w = \theta + \varepsilon,$$

where $w = (w_1, \ldots, w_n)^{\mathrm{T}}$ are empirical wavelet coefficients, $\varepsilon = (\varepsilon_1, \ldots, \varepsilon_n)^{\mathrm{T}}$ are iid normal random errors with zero mean and variance σ^2, and $\theta = (\theta_1, \ldots, \theta_n)^{\mathrm{T}}$ are the true wavelet coefficients. Let $\delta(w) = (\delta_1(w), \ldots, \delta_n(w))^{\mathrm{T}}$ be an estimator for θ. The soft thresholding wavelet shrinkage estimation by Donoho and Johnstone (1994) is given by

$$\delta_i^{\mathrm{soft}}(w, \lambda) = \mathrm{sign}(w_i)(|w_i| - \lambda)\mathscr{I}(|w_i| \geqslant \lambda), \quad i = 1, \ldots, n, \tag{1}$$

where $\mathscr{I}(\cdot)$ is an indicator function and $\lambda > 0$ is a threshold parameter.

The wavelet estimation problem can be treated via the estimation of the mean vector θ from a multivariate normal distribution $w|_\theta \sim N(\theta, \sigma^2 I)$, where I is the $n \times n$ identity matrix. We employ an asymmetric Linex loss function (Varian, 1975; Zellner, 1986) as error criterion. Under such a loss

E-mail address: syhuang@stat.sinica.edu.tw (S.-Y. Huang).

PII: S0167-7152(01)00181-X

function, we derive a generalized Bayes estimator, which is also shown to be the unique admissible and minimax estimator. Then, we show that the soft wavelet shrinkage estimator (1) can be derived as an empirical version of the admissible-and-minimax generalized Bayes estimator.

2. Linex loss

Consider the following asymmetric Linex loss for δ:

$$L(\theta,\delta(w)) = \frac{1}{n}\sum_{i=1}^{n}(\mathrm{e}^{a_i\{\delta_i(w)-\theta_i\}} - a_i\{\delta_i(w)-\theta_i\} - 1), \quad a_i \neq 0. \tag{2}$$

We see that, for a positive a_i, the following term:

$$\mathrm{e}^{a_i\{\delta_i(w)-\theta_i\}} - a_i\{\delta_i(w)-\theta_i\} - 1 \tag{3}$$

increases exponentially for over-estimation of a component θ_i, as $\delta_i - \theta_i \to \infty$; on the other hand, the loss increases linearly for under-estimation of the component θ_i, as $\theta_i - \delta_i \to \infty$. Thus, the positivity of a_i discourages over-estimation and results in estimation shifting towards the left. For a negative a_i, this phenomenon is reversed; the single-term loss (3) increases linearly for over-estimation and exponentially for under-estimation. Thus, the negativity of a_i discourages under-estimation and results in estimation shifting towards the right.

The name "Linex loss" comes from the linearity–exponentiality phenomenon of loss. Applications of the Linex loss to several Bayesian estimation and prediction problems can be found in Zellner (1986).

3. The main result

Using Linex loss (2) as the error criterion, we have the following theorem.

Theorem 1. *Under loss* (2), *the estimator* $\delta^{\mathrm{GB}}(w)$ *given by*

$$\delta_i^{\mathrm{GB}}(w) = w_i - \frac{a_i\sigma^2}{2}, \quad i = 1,\ldots,n \tag{4}$$

is a generalized Bayes estimator for θ *with respect to the flat improper prior on* $\mathbb{R}^n$. *Moreover,* $\delta^{\mathrm{GB}}(w)$ *is the unique admissible and minimax estimator.*

For application of the Linex loss to wavelet estimation problem, we consider specifically the Linex loss with a_i values depending on signs of θ_i's

$$a_i = \begin{cases} c & \text{for } \theta_i \geqslant 0,\ i = 1,\ldots,n, \\ -c & \text{for } \theta_i < 0,\ i = 1,\ldots,n, \end{cases} \tag{5}$$

where $c > 0$ is some constant. Such an error criterion discourages estimators from over-estimation *in magnitude* (i.e. in absolute value) and results in shrinkage estimation towards zero. Under such a loss criterion, the *ideal* unique admissble and minimax estimator is given by

$$\delta_i^{\mathrm{GB}}(w) = w_i - \mathrm{sign}(\theta_i)\lambda, \quad \text{where } \lambda = \frac{c\sigma^2}{2}.$$

Often the signs of parameters θ_i's are not known. A natural approach is to use $\mathrm{sign}(w_i)$ to estimate $\mathrm{sign}(\theta_i)$ and make truncation at zero. We then have the following empirical version of δ^{GB}.

$$\delta_i^{\mathrm{soft}}(w) = \begin{cases} (w_i - \lambda) \vee 0 & w_i \geqslant 0, \\ (w_i + \lambda) \wedge 0 & w_i < 0, \end{cases}$$
$$= \mathrm{sign}(w_i)(|w_i| - \lambda)_+. \tag{6}$$

The above estimator is the renowned soft wavelet shrinkage estimator.

4. Proof for Theorem 1

4.1. Generalized Bayes estimator

Let π be the flat improper prior with probability density function $\pi(\theta) = 1$, $\theta \in \mathbb{R}^n$. Then, the posterior distribution is $\pi(\theta|w) \sim N(w, \sigma^2 I)$. It is straightforward to check that the posterior expected loss of an arbitrary estimator $\delta(w)$ is given by

$$\rho(\pi(\theta|w), \delta(w)) = \int L(\theta, \delta(w))\,\mathrm{d}\pi(\theta|w)$$
$$= \frac{1}{n}\sum_{i=1}^{n}(\mathrm{e}^{a_i(\delta_i - w_i) + a_i^2\sigma^2/2} - a_i\{\delta_i - w_i\} - 1). \tag{7}$$

A generalized Bayes estimator is an estimator δ which minimizes (7). First, we take derivatives with respect to δ_i and set them to zero. We get the system of equations.

$$\mathrm{e}^{a_i(\delta_i - w_i) + a_i^2\sigma^2/2} - 1 = 0, \quad i = 1, \ldots, n$$

or equivalently,

$$a_i(\delta_i - w_i) + a_i^2\sigma^2/2 = 0, \quad i = 1, \ldots, n.$$

The unique solution for the above system of equations, which is the generalized Bayes estimator with respect to the flat prior over $\mathbb{R}^n$, is given by (4).

4.2. Admissibility

Let $\Phi(w; \theta, \sigma^2 I)$ be the distribution function for $N(\theta, \sigma^2 I)$, then

$$R(\theta, \delta) = \frac{1}{n}\sum_{i=1}^{n}\int_{R^n}(\mathrm{e}^{a_i\{\delta_i(w) - \theta_i\}} - a_i\{\delta_i(w) - \theta_i\} - 1)\,\mathrm{d}\Phi(w; \theta, \sigma^2 I).$$

We see that $R(\theta, \delta)$ is continuous in θ for any δ. Suppose δ^{GB} is not admissible. Then, there exists an estimator δ such that $R(\theta, \delta) \leqslant R(\theta, \delta^{\mathrm{GB}})$, with strict inequality for some θ, say θ_0. Since $R(\theta, \delta)$

and $R(\theta,\delta^{\mathrm{GB}})$ are continuous in θ, there exist strictly positive constants c_1 and c_2 such that

$$R(\theta,\delta) < R(\theta,\delta^{\mathrm{GB}}) - c_1 \quad \text{for } \theta \in \{\theta: |\theta-\theta_0| < c_2\}.$$

Consider a sequence of priors $\pi_k(\theta) \sim N(0,\tau_k^2 I)$, with $\lim_{k\to\infty}\tau_k^2 = \infty$. Using the technique of minimizing posterior expected loss, the Bayes estimator under the prior π_k and Linex loss (2) can be shown by

$$\delta_i^{\pi_k}(w) = \frac{\tau_k^2}{\sigma^2+\tau_k^2}\left(w_i - \frac{a_i\sigma^2}{2}\right), \quad i=1,\ldots,n \tag{8}$$

with the Bayes risk

$$r(\pi_k,\delta^{\pi_k}) = \frac{\sigma^2\tau_k^2}{2n(\sigma^2+\tau_k^2)}\sum_{i=1}^n a_i^2. \tag{9}$$

One can also compute the Bayes risk for δ^{GB}, and the result is

$$r(\pi_k,\delta^{\mathrm{GB}}) = \frac{\sigma^2}{2n}\sum_{i=1}^n a_i^2. \tag{10}$$

Let $c_3 = \liminf_{k\to\infty}\int_{|\theta-\theta_0|<c_2}\pi_k(\theta)\,\mathrm{d}\theta$. Since $\lim_{k\to\infty}\tau_k^2=\infty$, we have

$$c_3 = \liminf_{k\to\infty}\int_{|\theta-\theta_0|<c_2}\pi_k(\theta)\,\mathrm{d}\theta > 0.$$

Therefore, for k is large enough

$$\begin{aligned} r(\pi_k,\delta^{\mathrm{GB}}) - r(\pi_k,\delta^{\pi_k}) &\geqslant r(\pi_k,\delta^{\mathrm{GB}}) - r(\pi_k,\delta) \\ &= \int_{R^n}(R(\theta,\delta^{\mathrm{GB}}) - R(\theta,\delta))\pi_k(\theta)\,\mathrm{d}\theta \\ &\geqslant \int_{|\theta-\theta_0|<c_2}(R(\theta,\delta^{\mathrm{GB}}) - R(\theta,\delta))\pi_k(\theta)\,\mathrm{d}\theta > c_1c_3 > 0. \end{aligned}$$

This contradicts with the fact that $\lim_{k\to\infty}\{r(\pi_k,\delta^{\mathrm{GB}}) - r(\pi_k,\delta^{\pi_k})\} = 0$.

4.3. Minimaxity

The minimaxity of δ^{GB} follows from its admissibility and the constant risk phenomenon $R(\theta,\delta^{\mathrm{GB}}) = \sigma^2\sum_{i=1}^n a_i^2/(2n)$. □

5. Remarks

The estimation of normal mean under certain Linex loss functions can also be found in the literature. Zellner (1986) has studied the estimation problem for the univariate case. It was assumed that $w_1,\ldots,w_n$ were iid univariate normal random variables with a common mean θ and variance σ^2. The estimator $\bar{w} - a\sigma^2/2$ was proposed therein. Later in Parsian (1990), the multivariate case was studied via a three-stage Bayesian framework. Therein, a generalized Bayes estimator was derived and also shown to be admissible. The generalized Bayes estimator in our Theorem 1 is different from Parsian's, as the Bayesian model setup in this article is different from his.

References

Donoho, D.L., Johnstone, I.M., 1994. Ideal spatial adaptation by wavelet shrinkage. Biometrika 81, 425–455.

Parsian, A., 1990. On the asmissibility of an estimator of a normal mean vector under a Linex loss function. Ann. Inst. Statist. Math. 42, 657–669.

Varian, H.R., 1975. A Bayesian approach to real estate assessment. In: Fienberg S.E., Zellner, A. (eds.), Studies in Bayesian Econometrics and Statistics in Honour of Leonard J. Savage, North-Holland, Amsterdam, pp. 195–208.

Zellner, A., 1986. Bayesian estimation and prediction using asymmetric loss functions. J. Amer. Statist. Assoc. 81, 446–451.