Statistics and Probability Letters 119 (2016) 21–29

Contents lists available at ScienceDirect

Statistics and Probability Letters

journal homepage: www.elsevier.com/locate/stapro

Pitman closeness properties of Bayes shrinkage procedures in estimation and prediction

Takeru Matsuda [a,*], William E. Strawderman [b]

[a] *Graduate School of Information Science and Technology, The University of Tokyo, Tokyo, Japan*

[b] *Department of Statistics and Biostatistics, Rutgers University, NJ, USA*

A R T I C L E   I N F O

Article history:
Received 28 April 2016
Received in revised form 8 July 2016
Accepted 8 July 2016
Available online 15 July 2016

MSC:
62C10
62F15
62F10

Keywords:
Pitman closeness
Predictive density
Point estimation
Bayes
Superharmonic prior
Stein prior

A B S T R A C T

Pitman closeness for Bayes shrinkage procedures in normal models are investigated. In point estimation, priors in the Strawderman class dominate the uniform prior. In predictive density estimation, spherically symmetric superharmonic priors dominate the uniform prior under log loss.

1. Introduction

Pitman (1937) proposed a criterion to compare point estimators based on the joint distribution of loss functions. Suppose we have an observation $X \sim p(x \mid \theta)$ and estimate the unknown parameter θ by some estimator $\hat{\theta}(x)$. An estimator $\hat{\theta}_1$ dominates another estimator $\hat{\theta}_2$ in the sense of Pitman under loss $L(\theta, \hat{\theta})$ if

$$\Pr_\theta[L(\theta, \hat{\theta}_1(x)) < L(\theta, \hat{\theta}_2(x))] > \frac{1}{2}$$

for every θ.

In this study, we apply the Pitman closeness criterion to compare estimators of predictive densities. Suppose that we have an observation $X \sim p(X \mid \theta)$ and predict the future observation $Y \sim \widetilde{p}(Y \mid \theta)$ by a predictive density $\hat{p}(Y \mid X)$. The plug-in predictive density with an estimator $\hat{\theta}(X)$ is defined as

$$\hat{p}_{\text{plug-in}}(Y \mid X) = \widetilde{p}(Y \mid \hat{\theta}(X)).$$

* Corresponding author.
E-mail address: Takeru_Matsuda@mist.i.u-tokyo.ac.jp (T. Matsuda).

http://dx.doi.org/10.1016/j.spl.2016.07.005
0167-7152/© 2016 Published by Elsevier B.V.

The Bayesian predictive density based on a prior $\pi(\theta)$ is defined as

$$\hat{p}_\pi(Y \mid X) = \int \widetilde{p}(Y \mid \theta)\pi(\theta \mid X)\mathrm{d}\theta = \frac{\int \widetilde{p}(Y \mid \theta)p(X \mid \theta)\pi(\theta)\mathrm{d}\theta}{\int p(X \mid \theta)\pi(\theta)\mathrm{d}\theta}.$$

Aitchison (1975) showed that Bayesian predictive densities are preferable under Kullback–Leibler risk comparison. We compare Pitman closeness of predictive density estimators under two different losses. The first comparison is based on the joint distribution of the Kullback–Leibler loss

$$D(p(y \mid \theta), \hat{p}(y \mid x)) = \int p(y \mid \theta) \log \frac{p(y \mid \theta)}{\hat{p}(y \mid x)} \mathrm{d}y. \tag{1}$$

Definition 1. $\hat{p}_1(y \mid x)$ dominates $\hat{p}_2(y \mid x)$ in the sense of Pitman under Kullback–Leibler loss if

$$\mathrm{Pr}_\theta[D(p(y \mid \theta), \hat{p}_1(y \mid x)) < D(p(y \mid \theta), \hat{p}_2(y \mid x))] > \frac{1}{2}$$

for every θ.

Note that the difference of Kullback–Leibler loss for $\hat{p}_1(y \mid x)$ and $\hat{p}_2(y \mid x)$ is

$$\begin{aligned} &D(p(y \mid \theta), \hat{p}_1(y \mid x)) - D(p(y \mid \theta), \hat{p}_2(y \mid x)) \\ &\quad = \int p(y \mid \theta)(-\log \hat{p}_1(y \mid x))\mathrm{d}y - \int p(y \mid \theta)(-\log \hat{p}_2(y \mid x))\mathrm{d}y. \end{aligned}$$

Therefore, Kullback–Leibler loss is essentially the same as the expectation of $-\log \hat{p}(y \mid x)$ with respect to $y \sim p(y \mid \theta)$.

The second comparison is based on the joint distribution of the log loss

$$-\log \hat{p}(y \mid x). \tag{2}$$

Here, we consider the distribution of log loss not averaged over y.

Definition 2. $\hat{p}_1(y \mid x)$ dominates $\hat{p}_2(y \mid x)$ in the sense of Pitman under log loss if

$$\mathrm{Pr}_\theta[-\log \hat{p}_1(y \mid x)) < -\log \hat{p}_2(y \mid x))] > \frac{1}{2}$$

for every θ.

See Grünwald and Dawid (2004) for a discussion of the applicability of log loss to data compression. Whereas the Kullback–Leibler loss concerns the quality of prediction averaged over y, the log loss involves each y individually. In other words, comparisons in terms of log loss are based on the joint distribution of X and Y.

We note that Matsuda and Strawderman (2016) also studied Pitman closeness properties of predictive densities in the context of univariate normal distributions with nonnegative mean. In this paper, we investigate multivariate normal distributions with unrestricted mean.

In this study, we consider the estimation and prediction in p dimensional normal models. In Section 2, we compare plug-in predictive densities and Bayesian predictive densities based on the uniform prior. In Section 3, after providing Pitman closeness properties of Bayes shrinkage point estimators, we compare Bayesian predictive densities based on spherically symmetric superharmonic priors, including the Stein prior (Stein, 1974), and those based on the uniform prior.

2. Plug-in predictive density and Bayesian predictive density based on the uniform prior

In this section, we compare the Bayesian predictive density based on the uniform prior and the standard plug-in predictive density.

Suppose that we have an observation $X \sim \mathrm{N}_p(\theta, \sigma^2 I)$ and predict the future observation $Y \sim \mathrm{N}_p(\theta, \tau^2 I)$ by a predictive density $\hat{p}(Y \mid X)$. Here, we assume that X and Y are independent conditionally on θ, and σ^2 and τ^2 are known. The plug-in predictive density with the maximum likelihood estimator and the Bayesian predictive density with respect to the uniform prior are defined as

$$\hat{p}_{\text{plug-in}}(y \mid x) = \mathrm{N}_p(x, \tau^2 I),$$
$$\hat{p}_{\mathrm{I}}(y \mid x) = \mathrm{N}_p(x, (\sigma^2 + \tau^2)I).$$

Aitchison (1975) showed that $\hat{p}_{\mathrm{I}}$ dominates $\hat{p}_{\text{plug-in}}$ under Kullback–Leibler risk. In this section, we show that this domination does not necessarily hold under Pitman closeness criterion. Indeed, Pitman closeness comparisons under Kullback–Leibler loss may favor either $\hat{p}_{\text{plug-in}}$ or $\hat{p}_{\mathrm{I}}$ depending on the dimension p and the ratio τ^2/σ^2. We denote the α quantile of the chi-squared distribution with p degrees of freedom as $\chi_p^2(\alpha)$.