第26卷 第2期 昆明理工大学学报 Vol.26 No.2
2001年4月 Journal of Kunming University of Science and Technology Apr. 2001

Average Risk Optimal

Huang Wen-liang
(The Faculty of Science, Kunming University of Science and Technology, Kunming 650093, China)

Abstract: At the beginning of this paper, we introduced two ways in which the risk function $R(\theta,\delta)$ can be minimized in some overall sense; minimizing a weighted-average risk and minimizing the maximum risk. But now we shall consider the first of these approaches.

Key words: minimizing the average risk; Bayes estimator; risk function
CLC number: O211 **Document code:** A **Article ID:** 1007-855X(2001)02-107-04

0 Introduction

At present, we shall drop some restrictions. The paper is concerned the problem of minimizing

$$E\left(\sum_{i=1}^{s}\mu_i\frac{\partial T_i(x)}{\partial x_j}|x\right)=\frac{\partial}{\partial x_j}\log m(x)-\frac{\partial}{\partial x_j}\log h(x) \tag{1.1}$$

Where we shall assume the weights represented by Λ add up to 1, that is

$$\int d\Lambda(\theta)=1 \tag{1.2}$$

So that Λ is a probability distribution. An estimator δ minimizing (1.1) is called a Bayes estimator with respect to Λ.

1 Frist Example

In constructing Bayes estimators, as function of the posterior density. Some choices are made. These choices will ultimately affect the properties of the estimators, in during not only risk performance but also more fundamental considerations. In this section, we look at an example to illustrate these points.

Example 1.1 Sequential binomial sampling. Consider a sequence of binomial trials with a stopping rule. Let X,Y,and N denote, respectively, the number of successes, the number of failures, and the total number of trials at the moment sampling stops. The probability of any sample path is the $p^x(1-p)^y$ and we shall again suppose that p is given by X and Y. The calculation shows that, as in the fixed sample size case, it is the beta distribution with parameters a and b given by $[1.11,\ a=a+x, b=b+n-x]$ so that, in particular, the Bayes estimator of P is given by

$$\left(1.12\quad \delta\Lambda(x)=E(p\,|\,x)=\frac{a'}{a'+b'}=\frac{a+x}{a+b+n}\right)$$

regardless of the stopping rule.

From a Bayesian view, estimators that are limits of Bayes estimators are some what more desirable than generalized Bayer estimators. This is because, by construction, a limit of Bayes estimators must be close to a proper Bayes estimator. In constrast, a generalized Bayes estimator may not be close to any proper Bayes estimator.

2 Single -Prior Bayes

As discussed at the end of section 1, the prior distribution is typically selected from a flexible family of prior densities indexed by one or more parameters. Instead of denoting the prior by Λ, as was done in section

Received date:2000-11-03;
Biography:HUANG Wen-liang (1975～), female, Master degree.

1, we shall denote its density by $\pi(\theta|\gamma)$ where the parameter γ can be real -or vector- valued.

We can then write a Bayes model in a general form as

$$x|\theta\cdots f(x|\theta)$$

$$\Theta|\gamma\cdots\pi(\theta|\gamma)$$

Thus, conditionally on θ, x has sampling density $f(x|\theta)$, and conditionally on γ, Θ has prior density $\pi(\theta|\gamma)$, from this model, we calculate the posterior distribution, $\pi(\theta|x,\gamma)$, from which all Bayesian answers would come. The exact manner in which we deal with the parameter γ or, more generally, the prior distribution $\pi(\theta|\gamma)$ will lead us to different types of Bayes analyses. In this section we assume that the functional form of the prior, and the value of γ, is known. So we have one completely specified prior.

Given a loss function $L(\theta,d)$.We then look for the estimator that minimizes

$$\int L(\theta,d(x))\pi(\theta|x,\gamma_0)\mathrm{d}\theta \tag{2.1}$$

where $\pi(\theta|x,\gamma_0)=f(x|\theta)\pi(\theta|\gamma_0)/\int f(x|\theta)\pi(\theta|\gamma_0)\mathrm{d}\theta$.

The calculation of single-prior Bates estimator, has already been illustrated in section 1. Here is another example.

Example 2.1 scale uniform

For estimation in the model

$$x_i|\theta\cdots\mu(0,\theta), I=1,2,\cdots,n$$

Where a and b are know $\frac{1}{\theta}a,b\cdots\text{Gamma}(a,b)$ (2.2)

sufficiency allows us to work only with the density of $y=\max_i x_i$ which is given by $g(y|\theta)=ny^{n-1}/\theta^n, 0<y<\theta$. We then calculate the single-prior Bayers estimator of θ under squared error loss, this is the posterior mean, given by

$$E(\Theta|y,a,b)=\frac{\int_y^\infty\theta\frac{1}{\theta^{n+a+1}}\mathrm{e}^{-1/\theta b}\mathrm{d}\theta}{\int_y^\infty\frac{1}{\theta^{n+a+1}}\mathrm{e}^{-1/\theta b}\mathrm{d}\theta} \tag{2.3}$$

Although the ratio of integral is not expressible in any simple form, calculation is not difficult.

In general, the Bayes estimator under squared error loss is given by

$$E(\Theta|x)=\frac{\int\theta f(x|\theta)\pi(\theta)\mathrm{d}\theta}{\int f(x|\theta)\pi(\theta)\mathrm{d}\theta} \tag{2.4}$$

Where $x\sim f(x|\theta)$ is the observed random variable and $\Theta\sim\pi(\theta)$ is the parameter of interest.While there is a certain appeal about expression (2.4). It is therefore important to find conditions under which it can be simplified. Such simplification is useful for two somewhat related purpose.

(I) Implementation

If a Bayes solution is deemed appropriate, and we want to implement it, we must be able to calculate (2.4). Thus, we need reasonably straight forward, and general, methods of evaluating these integrals.

(II) Performance

By construction, a Bayes estimator minimizes the posterior expected loss and, hence, the Bayes risk. Often, however, we are interested in its performance, and perhaps optimality under other measures. For example, we also might examine Bayesion measures using other prior, in an investigation of Bayesion robustness.

These latter considerations tend to lead us to look for either manageable expressions for or accurate approximations to the integrals in (2.4). On the other hand, the considerations in (I) are more numerical (or computation). However, even this path can involve statistical considerations, and often gives us insight into the performance of our estimators.

A simplification of (2.4) is possible when dealing with independent prior distributions. If $x_i \sim -f(x/\theta_i), i=1,\cdots,n$, are independent, and the prior is $\pi(\theta_1,\theta_2,\cdots,\theta_n)=\prod_i \pi(\theta_i)$, then the posterior mean of θ_i satisfies

$$E(\theta_i \mid x_1,\cdots,x_n)=E(\theta_i \mid x_i) \tag{2.5}$$

That is, the Bayes estimator of θ_i only depends on the data through x_i. Although the simplification provided by (2.5) may prove useful, at this level of generality it is impossible to go further.

However, for exponential families, evaluation of (2.4) is sometimes possible through alternate representa tions of Bayes estimaters. Suppose the distribution of $x=x(x_1,x_2,\cdots,x_n)$ is given by the multiparameter exponential family, that is,

$$p\eta(x)=\exp\left\{\sum_{i=1}^{s}\eta_i T_i(x)-A(\eta)\right\}h(x) \tag{2.6}$$

Then we can express the Bayes estimator as function of partial derivatives with respect to x. The following theorem present a general formula for the needed posterior expectation.

Theorem 2.1 If x has density (2.6), and η has prior density $\pi(\eta)$, then for $j=1,2,\cdots,n$

$$E\left(\sum_{i=1}^{s}\eta_i\frac{\partial T_i(x)}{\partial x_j}\middle| x\right)=\frac{\partial}{\partial x_j}\log m(x)-\frac{\partial}{\partial x_j}\log h(x) \tag{2.7}$$

where $m(x)=\int p\eta(x)\pi(\eta)\mathrm{d}\eta$ is the marginal distribution of x. Alternatively, the posterior expectation can be expressed in matrix form as

$$E(\tau\eta)=\nabla\log m(x)-\nabla\log h(x) \tag{2.8}$$

where $\tau=\{\partial T_i/\partial x_i\}$.

Proof Noting that $\partial\exp\{\sum\eta_i T_i\}/\partial x_j=\sum_i\eta_i(\partial T_i/\partial x_i)\exp\{\sum\eta_i T_i\}$, we can write

$$E\left(\sum\eta_i\frac{\partial T_i(x)}{\partial x_j}\middle| x\right)=\frac{1}{m(x)}\int\sum\left[\eta_i\frac{\partial T_i}{\partial x_i}\right]\mathrm{e}^{\sum\eta_i T_i-A(\eta)}h(x)\pi(\mu)\mathrm{d}\mu$$

$$=\frac{1}{m(x)}\int\left[\frac{\partial}{\partial x_j}\mathrm{e}^{\sum_i\eta_i T_i}-\frac{\partial}{\partial x_j}h(x)\mathrm{e}^{\sum_i\eta_i T_i}\right]\mathrm{e}^{-A(\eta)}h(x)\pi(\mu)\mathrm{d}\mu$$

$$=\frac{1}{m(x)}\frac{\partial}{\partial x_j}\int\left(\mathrm{e}^{\sum\eta_i T_i-A(\eta)}h(x)\right)\pi(\eta)\mathrm{d}\eta-\frac{\frac{\partial}{\partial x_j}h(x)}{h(x)}\frac{1}{m(x)}\int\mathrm{e}^{\sum\eta_i T_i-A(\eta)}h(x)\pi(\eta)\mathrm{d}\eta$$

$$=\frac{\frac{\partial}{\partial x_j}m(x)}{m(x)}-\frac{\frac{\partial}{\partial x_j}h(x)}{h(x)} \tag{2.9}$$

where , in the third equality, we have used the fact that

$$\left[\frac{\partial T_i}{\partial x_j}\right]\mathrm{e}^{\sum_i\eta_i T_i}h(x)=\frac{\partial}{\partial x_j}(\mathrm{e}^{\sum_i\eta_i T_i}h(x))-\mathrm{e}^{\sum_i\eta_i T_i}\left[\frac{\partial h(x)}{\partial x_j}\right]$$

In the fourth equality, we have interchanged the order of integration and differentiation, and used the definition of $m(x)$. Finally, using logarithms, $E\left(\sum\eta_i\frac{\partial T_i(x)}{\partial x_j}\middle| x\right)$ can be written as (2.7).

Although it may appear that this theorem merely shifts calculation form one integral to another, this shift brings advantages. These advantage stem form the facts that the calculation of the derivatives of $\log m(x)$ is often feasible and that, with the estimator expressed as (2.7), risk calculations may be simplified.

Theorem 2.1 simplifies further when $\tau(i)=x(i)$.

Corollary 2.1 If $x=(x_1,x_2,\cdots,x_p)$ has the density

$$p\eta(x)=e^{\sum_{i=1}^{p}\eta_i x_i-A(\eta)}h(x) \tag{2.10}$$

and η has prior density $\pi(\eta)$, the Bayes estimator of η under the loss $L(\eta,\delta)=\sum(\eta_i-\delta_i)^2$ is given by

$$E(\eta_i\,|\,x)=\frac{\partial}{\partial x_i}\log m(x)-\frac{\partial}{\partial x_i}\log h(x) \tag{2.11}$$

Example 2.2 Multiple normal model

For
$$x_i\,|\,\theta_i\sim N(\theta_i-\sigma^2),\ i=1,2,\cdots,p,\ \text{independent}$$
$$\Theta_i\sim N(\mu,\tau^2),\quad i=1,2,\cdots,p,\ \text{independent}$$

Where σ^2, and μ are know, $\eta(i)=\theta_i/\delta^2$ and the Bayes estimator of θ_i is

$$E(\Theta_i\,|\,x)=\sigma^2E(\eta_i\,|\,x)=\sigma^2\left[\frac{\partial}{\partial x_i}\log m(x)-\frac{\partial}{\partial x_i}\log h(x)\right]=\frac{\tau^2}{\sigma^2+\tau^2}x_i+\frac{\sigma^2}{\sigma^2+\tau^2}\mu$$

Since

$$\frac{\partial}{\partial x_i}\log m(x)=\frac{\partial}{\partial x_i}\log\left(e^{\frac{1}{2(\sigma^2+\tau^2)}\sum_i(x_i-\mu)^2}\right)=\frac{-(x_i-\mu)}{\sigma^2+\tau^2}$$

And

$$\frac{\partial}{\partial x_i}\log h(x)=\frac{\partial}{\partial x_i}\log\left(e^{-\frac{1}{2}\sum x_i^2/\sigma^2}\right)=-\frac{x_i}{\sigma^2}$$

An application of the representation (2.11) is to the comparison of the risk of the Bayes estimator, with the risk of the best unbiased estimator.

References:

[1] Lehmann E L, George Casella. Theory of Point Estimation [M]. Maple-Vail Book Manufacturing Group, York, PA, 1998.270～275.

[2] Guoxi Ni. Theory and Method of Usual Matrix [M]. Shanghai: Science Press,1984.

[3] Songgui Wang. The Theory of Linear Models and Its application [M]. Hefei: Anhui Education Press, 1987.

[4] Jianbao Chen, Jinlong Zhan. Admissibility of Linear Estimators of Regression Coefficients in the Class of All Estimators under Matrix Loss Function [J]. Systems Science and Mathematical Sciences,1991,(4): 145～150.

平均风险最优化

黄文亮

(昆明理工大学 理学院, 云南 昆明 650093)

摘要: 在本文开始,先介绍从某种整体意义上风险函数$R(\theta,\delta)$可最小化的两种方法:即加权的平均风险最小化和最大风险最小化.本文只介绍第一种方法.

关键词: 平均风险最小化;贝叶斯估计;风险函数