广西科学 Guangxi Sciences 2007,14(4):365～366

一种非对称损失下分布函数的最优不变估计

Estimation of Distribution Function Under an Unsymmetrical Loss Fuction

余慧敏

YU Hui-min

(广东海洋大学理学院,广东湛江　524088)

(College of Science,Guangdong Ocean University,Zhanjiang,Guangdong,524088,China)

摘要:给定来自一未知连续分布函数 F 的容量为 n 的子样 $x_1,x_2,\cdots,x_n$,考虑分布函数 F 的不变估计问题.在非对称损失函数 $L(F(t),d(t))=b\int(\exp\{a[d(t)-F(t)]\}-a[d(t)-F(t)]-1)\mathrm{d}F(t)$ 和单调变换群下得到 F 的最优不变估计为 $d(t,X)=\sum_{i=0}^{n}c_iI(x_{(i)}\leqslant t\leqslant x_{(i+1)})$,其中 $c_i=\frac{1}{a}\ln\frac{\int_0^1 t^i(1-t)^{n-i}\mathrm{d}t}{\int_0^1\exp\{-at\}t^i(1-t)^{n-i}\mathrm{d}t},a\neq 0,b>0$.

关键词:非对称损失　连续分布函数　不变估计

中图法分类号:O212.5　**文献标识码**:A　**文章编号**:1005-9164(2007)04-0365-02

Abstract:Given a random sample of $x_1,x_2,\cdots,x_n$ size n from an unknown continuous distribution function F, this paper considers the problem of invariant estimator of the continuous distribution function F. Under the unsymmetrical loss function $L(F(t),d(t))=b\int(\exp\{a[d(t)-F(t)]\}-a[d(t)-F(t)]-1)\mathrm{d}F(t)$, we get the best invariant estimator of the continuous distribution function F.

Key words:unsymmetrical loss,continuous distribution function,invariant estimator

在分布函数的估计问题中,很多文献讨论了在以下3种损失下分布函数的优良估计问题[1～6]:

$$L_1(F,a)=\int|F(t)-a(t)|^r h(F(t))\mathrm{d}(F(t));$$

$$L_2(F,a)=\int|F(t)-a(t)|^r h(F(t))\mathrm{d}(F(t));$$

$$L_3(F,a)=\sup_t|F(t)-a(t)|.$$

这3种损失都是对称损失,估计 d 和参数 F 在损失中地位相同,也就是说估计高估实际分布函数与估计低估实际分布函数所承担的风险是一样的.但是在实际应用中的很多情况下,对称损失函数并不适用,一些非对称损失函数逐渐引起了人们的注意.在参数的估计问题中,已有许多为人们熟知的非对称损失函数,其中 Varian 提出的线性指数损失函数 $L(\theta,\delta)=b\{\exp\{a(\delta-\theta)\}-a(\delta-\theta)-1\}$ 是受到广泛关注的非对称损失函数之一,其中 $b>0$ 为均衡因子,$a\neq 0$ 为谱轮廓参量[7～9].本文中我们将非对称损失推广到非参数问题中,即对分布函数 F 的估计中,引入损失函数

$$L(F(t),d(t))=b\int(\exp\{a[d(t)-F(t)]\}-a[d(t)-F(t)]-1)\mathrm{d}F(t).\qquad(1)$$

假设 $x_1,x_2,\cdots,x_n$ 为来自一未知连续分布函数 F 的样本,由 Aggawa[1] 可知,连续分布函数在单调群 G 下的不变估计为:$d(t,X)=\sum_{i=0}^{n}c_iI(x_{(i)}\leqslant t\leqslant x_{(i+1)})$,其中 $G=\{g_\varphi:g_\varphi(x_1,x_2,\cdots,x_n)=(g_\varphi(x_1),g_\varphi(x_2),\cdots,g_\varphi(x_n)),\varphi$ 是从实数域 R 到 R 的严格单调递增函数$\}$,而 $x_{(1)}<x_{(2)}<\cdots<x_{(n)}$ 是样本 $x_1,x_2,\cdots,x_n$ 的次序统计量,$x_{(0)},x_{(n+1)}$ 定义成 $-\infty$ 和 $+\infty$,$I(E)$ 表示集合 E 的示性函数,c_i 为常数,具体可参见文献[1].记 U 表示所有满足 $d(t,X)=\sum_{i=0}^{n}c_iI(x_{(i)}\leqslant t\leqslant x_{(i+1)})$ 的 F 的估计的集合.在损失函数(1)以及单调变换群 G 下,要找到 F 的最优不变估计,即要从集合

收稿日期:2007-05-29

作者简介:余慧敏(1980-),女,硕士,助教,主要从事概率统计研究。

U 中找风险最小的估计.

引理 在损失函数(1)条件下,对 $\forall\, d \in U$,则 d 的风险函数为 $R(F,d) = b\sum_{i=0}^{n} R_i(F,d)$,其中 $R_i = \int_0^1 \binom{n}{i}[\exp\{a(c_i - t)\} - a(c_i - t) - 1]t^i(1-t)^{n-i}\mathrm{d}t$.

证明 $L(F,d) = b\int(\exp\{a[d(t) - F(t)]\} - a[d(t) - F(t)] - 1)\mathrm{d}F(t) = b\sum_{i=0}^{n}\int_{x_{(i)}}^{x_{(i+1)}}(\exp\{a[d(t) - F(t)]\} - a[d(t) - F(t)] - 1)\mathrm{d}F(t) = b\sum_{i=0}^{n}\int_{F(x_{(i)})}^{F(x_{(i+1)})}\exp\{a(c_i - t)\} - a(c_i - t) - 1)\mathrm{d}t = b\sum_{i=0}^{n} L_i(F,d)$,

其中 $L_i(F,d) = \int_{F(x_{(i)})}^{F(x_{(i+1)})}\exp\{a(c_i - t)\} - a(c_i - t) - 1)\mathrm{d}t$.

令 $Z_i = F(x_{(i)})$,则有

$R(F,d) = E[L(F,d)] = b\sum_{i=0}^{n} E[L_i(F,d)] = b\sum_{i=0}^{n} E\int_{Z_i}^{Z_{(i+1)}}(\exp\{a(c_i - t)\} - a(c_i - t) - 1)\mathrm{d}t = b\sum_{i=0}^{n}\int_0^1\int_0^{Z_{i+1}}\int_{Z_i}^{Z_{i+1}}(\exp\{a(c_i - t)\} - a(c_i - t) - 1)\mathrm{d}t\mathrm{d}F_{Z_iZ_{i+1}}(Z_iZ_{i+1}) = b\sum_{i=0}^{n}\int_0^1\int_t^1\int_0^t(\exp\{a(c_i - t)\} - a(c_i - t) - 1)\mathrm{d}F_{Z_iZ_{i+1}}(Z_iZ_{i+1})\mathrm{d}t = b\sum_{i=0}^{n}\int_0^1[\exp\{a(c_i - t)\} - a(c_i - t) - 1]\int_t^1\int_0^t \mathrm{d}F_{Z_iZ_{i+1}}(Z_iZ_{i+1})\mathrm{d}t = b\sum_{i=0}^{n}\binom{n}{i}\int_0^1[\exp\{a(c_i - t)\} - a(c_i - t) - 1]t^i(1-t)^{n-i}\mathrm{d}t = b\sum_{i=0}^{n} R_i(F,d)$.

引理证毕.

定理 在非对称损失函数(1)和变换群 G 条件下,连续分布函数 F 的最优不变估计为 $d(t,X) = \sum_{i=0}^{n} c_i I(x_{(i)} \leqslant t \leqslant x_{(i+1)})$,其中 $c_i =$

$$\frac{1}{a}\ln\frac{\int_0^1 t^i(1-t)^{n-i}\mathrm{d}t}{\int_0^1 \exp\{-at\}t^i(1-t)^{n-i}\mathrm{d}t}, a \neq 0, b > 0.$$

证明 要在集合 U 中找不变估计,即要找出一不变估计是其风险在所有风险中最小,那么由引理知,要最小化 $R(F,d)$ 等价于极小化每一项 $R_i(F,d)$. 又 $R_i(F,d)$ 有如下的一阶和二阶导函数:

$$\frac{\partial}{\partial c_i}R_i(F,d) = \int_0^1\binom{n}{i}[a\exp\{a(c_i - t)\} - a]t^i(1-t)^{n-i}\mathrm{d}t, \quad (2)$$

$$\frac{\partial}{\partial c_i^2}R_i(F,d) = \int_0^1\binom{n}{i}a^2\exp\{a(c_i - t)\}t^i(1-t)^{n-i}\mathrm{d}t. \quad (3)$$

由(3)式知,对 $\forall\, i = 0,1,2,\cdots,n$,有 $\frac{\partial}{\partial c_i^2}R_i(F,d) > 0$,因此 c_i 为 $\frac{\partial}{\partial c_i}R_i(F,d)$ 的实根时,$R_i(F,d)$ 达到最小,同时 $R(F,d)$ 最小. 解 $\frac{\partial}{\partial c_i}R_i(F,d) = 0$ 得 $(\int_0^1 t^i(1-t)^{n-i}\mathrm{d}t)/(\int_0^1\exp\{-at\}t^i(1-t)^{n-i}\mathrm{d}t) = \exp\{ac_i\}$,则 $c_i = \frac{1}{a}\ln(\int_0^1 t^i(1-t)^{n-i}\mathrm{d}t)/(\int_0^1\exp\{-at\}t^i(1-t)^{n-i}\mathrm{d}t)$,且 $0 \leqslant c_i \leqslant 1$,因此得到 F 的最优不变估计 $d(t,X) = \sum_{i=0}^{n} c_i I(x_{(i)} \leqslant t \leqslant x_{(i+1)})$,其中 $c_i = \frac{1}{a}\ln(\int_0^1 t^i(1-t)^{n-i}\mathrm{d}t)/(\int_0^1\exp\{-at\}t^i(1-t)^{n-i}\mathrm{d}t)$.

定理证毕.

参考文献:

[1] AGGARWA O P. Some minimax invariant procedures for estimating a cumulative distribution function[J]. Ann Math Statist, 1995, 26: 450-462.

[2] YU Q. Minimax invariant estimator of a continuous distribution function[J]. Ann Inst Statist Math, 1992, 44: 729-735.

[3] YU Q. Admissibility of the best invariant estimator of a discrete distribution function[J]. Statistica Sinica, 1998, 8: 377-392.

[4] YU Q, CHOW M S. Minimaxity of the empirical distribution function in invariant estimation[J]. Ann Statist, 1991, 19: 935-951.

[5] BROWN L D. Admissibility in discrete and continuous invariant nonparametric estimation problems and in their multinomial analogs[J]. Ann Statist, 1988, 16: 1567-1593.

[6] FRIEDMAN Y, GELMAN A, PHADLA E. Best invariant estimator of a distribution function under the Kolmogorov-Smirnov loss function[J]. Ann Statist, 1988, 16: 1254-1261.

[7] TAKAGI Y. Local asymptotic minimax risk bounds for asymmetric loss functions[J]. Ann Statist, 1994, 22: 39-48.

[8] CAIN M, JANSSEN C. Real estate price prediction under asymmetric loss[J]. Ann Inst Statist Math, 1995, 47: 401-414.

[9] ZOU G H. Admissible estimation for finite population under the LINER loss function[J]. J Statist Plann Inference, 1997, 61: 373-384.

(责任编辑:韦廷宗)