Frontiers in Business, Economics and Management
ISSN: 2766-824X | Vol. 14, No. 3, 2024

Empirical Bayes Likelihood for Distrubition Family Based on Ranked Set Sampling

Naiyi Li[1], Yongming Li[2], Juan Huang[1, *]

[1]School of Mathematics and Computer, Guangdong Ocean University, Zhanjiang 524088, China
[2]School of Mathematics and Computer Science, Shang Rao Normal University Shang Rao, 334001, China
[*] Corresponding Author: Juan Huang

Abstract: In empirical likelihood methods, the key to constructing a likelihood function is to find a suitable weight scheme that reflects the importance of data. These weights are usually limited by some constraints. By maximizing this likelihood function, we can obtain estimates of the parameters and perform corresponding hypothesis tests, In this paper, we get the empirical Bayes likelihood test rules for distribution based on ranked set sampling. Its asymptotic optimality is obtained.

Keywords: Empirical Bayes likelihood, ranked set sampling.

1. Introduction

Empirical Likelihood Method is a statistical inference method aimed at constructing the

likelihood function of data in a non parametric manner for parameter estimation and hypothesis testing. This method was initially proposed by Thomas and Grunkemeier and further developed by Owen. The main advantage of the empirical likelihood method is that it does not require any assumptions about the distribution of data, thus having robustness. [1-7].

Ranked Set Sampling (RSS) is a statistical sampling method that improves sampling efficiency to a certain extent, reduces experimental costs, and is particularly suitable for situations where samples are easy to sort but not easy to quantify[8].

The sorting set sampling method is divided into balanced sorting set sampling and unbalanced sorting set sampling. The process of balanced sorting set sampling is as follows: Randomly select samples with a capacity of the k^2 sample from the population and divide them into k groups, each containing k individuals.Sort each group of sample individuals in ascending order using visual or other intuitive information.Extract individuals from each group of samples arranged in order: select the individual with the smallest order from the first group of samples, select the individual with the second order from the second group of samples, and so on, until the individual with the highest order is selected from the k-th group of samples.After completing the above steps, samples with a capacity of k were selected from the k^2 sample individuals, which is a complete cycle.

The application examples of empirical Bayesian methods are indeed quite extensive. For example, in the field of signal processing, empirical Bayesian methods can be used for signal detection and classification. Through the prior probability and the conditional probability, we can calculate the posterior probability of different signals. These are just some examples of the application of empirical Bayesian methods. In fact, with the deepening of the research, the empirical Bayesian method also shows its potential and value in more fields. For example, in the field of machine learning, natural language processing, speech recognition and so on, empirical Bayesian methods have been widely used. It helps us understand and process data more accurately, leading to more informed decisions.empirical Bayes approach haved been stuied [9-14]

In this paper, we obtain empirical Bayes likelihood test for distribution in ranked set sampling.

Let X have a conditional density function for given θ

$$f(x\mid\theta)=\mathrm{e}^{-x}\theta(1-\mathrm{e}^{-x})^{\theta-1} \tag{1.1}$$

where θ is unknown parameterand $\Theta=\{\theta>0\}$ is parameter space.

We discuss the following test problem:

$$H_0:\theta=\theta_0 \leftrightarrow H_1:\theta\neq\theta_0 \tag{1.2}$$

where $\theta 0$ is given constants.

We choose loss function $L_0(\theta,d_0)=\frac{(\theta-\theta_0)^2}{\theta}I(\theta>\theta_0)$, $L_1(\theta,d_1)=\frac{(\theta-\theta_0)^2}{\theta}I(\theta\leq\theta_0)$.

Where d={d0,d1} is action space, d0 and d1 imply acceptance and rejection of $H0$ respectively.

Suppose that the prior distribution G(θ) of parameter the θ is unknown.

We have random decision function

$$\delta(x)=P(\text{accept } H0 \mid X=x). \tag{1.3}$$

Then, the risk function of δ(x) is shown by

$$\begin{aligned}R(\delta(x),G(\theta))&=\int_\Theta\int_\Omega\big[L(\theta,d_0)f(x\mid\theta)\delta(x)+\\&\quad L(\theta,d_1)f(x\mid\theta)(1-\delta(x))\big]dxdG(\theta)\\&=\int_\Omega\beta(x)\delta(x)dx+C_G\end{aligned}$$

where

$$C_G=\int_\Omega L(\theta,d_1)dG(\theta),\ \ \beta(x)=\int_\Theta(\theta-\theta_0)^2 f(x\mid\theta)dG(\theta).\tag{1.4}$$

The marginal density function of X is shown by

$$f_G(x)=\int_{\Theta} f(x\mid\theta)\mathrm{d}G(\theta), \beta(x)=\int_{\Theta}\mathrm{e}^{-x}(1-\mathrm{e}^{-x})^{\theta-1}\mathrm{d}G(\theta),$$

where

$$f_G^{(1)}=-f_G(x)+\int_{\Theta}\mathrm{e}^{-2x}\theta(\theta-1)(1-\mathrm{e}^{-x})^{\theta-2}\mathrm{d}G(\theta) \quad (1.5)$$

By (1.5), we have $\beta(x)=u_1(x)f_G^{(1)}(x)+u_2(x)f_G(x)$ where $u_1(x)=\mathrm{e}^{2x}-\mathrm{e}^{x}+3$, $u_2(x)=(\mathrm{e}^{x}-1)(\mathrm{e}^{x}-\theta_0)$. Using (1.5), Bayes test function is obtained as follows

$$\delta_G(x)=\begin{cases}1,\beta(x)\le 0\\ 0,\beta(x)>0\end{cases}$$

Further, we obtian the minimum Bayes risk as follows

$$\begin{aligned}R(G)&=\inf_{\delta} R(\delta,G)=R(\delta_G,G)\\ &=\int_{\Omega}\beta(x)\delta_G(x)dx+C_G\end{aligned} \quad (1.6)$$

From above that δ(x)=δG(x) and R(G) can be obtained when the prior distribution of G(θ) is given. If not, we apply the empirical Bayes likelihood test method. The rest of this paper is organized as follows. Section 2 presents an empirical Bayes likelihood Test under ranked set sampling. In section 3, we obtain asymptotic optimality of convergence of the empirical Bayes likelihood test in ranked set sampling.

2. Construction of Empirical Bayes Likelihood Test under Ranked Set Sampling

Supposed that $X_{(1)1},X_{(1)2},\cdots,X_{(1)m},X_{(2)1},X_{(2)2}\cdots,X_{(2)m},\cdots,X_{(k)1},X_{(k)2}\cdots,X_{(k)m}$ be a balanced

ranked set sample from population which has the common marginal density function $f_G(x)$. We assume perfect ranking. Denote that$X_{(1)1}$,$X_{(1)m}$,$X_{(2)1}$,$X_{(2)m}$, $X_{(k)1}$,$X_{(k)m}$ are ranked set historical samples, and X is present sample. Assume $f(x)\in C_{s,\alpha}, x\in R_1$, where

$C_{s,\alpha}$={g(x)|g(x) is a probability density function. The s−he order derivative $g^{(s)}(x)$ is continuous with $|g^{(s)}(x)|\le\alpha, s\ge3, \alpha>0$}, $n=km$.

Supposed that $K_r(x)$ be a Borel measurable bounded function vanishing off (0,1) such that (A1):

$$\frac{1}{t!}\int_0^1 v^t K_r(v)dv=\begin{cases}1,t=0\\ 0,t=1,\cdots,s-1\end{cases} \quad (2.1)$$

empirical likelihood is established by

$$R(f^{(r)}(x))=\sup\left\{\prod_{i=1}^n np_i, p_i\ge 0,\sum_{i=1}^n p_i=1,\sum_{i=1}^n p_i\Phi_i=0\right\}$$

empirical likelihood ratio is established by

$$l(f^{(r)}(x))=-2\log R(f^{(r)}(x))=2\sum_{i=1}^n\log(1+s\Phi_i).$$

where $s\in R$, s is determined by

$$\Pi(s)=\frac{1}{n}\sum_{i=1}^n\frac{\Phi_i}{1+s\Phi_i}=0.$$

Kernel estimator of f(x) is defined by

$$\Phi_i=K_r(x-X_{(i)1}/h_n)-f^{(r)}(x)$$

where h_n is a positive and smoothing bandwidth, and $\lim_{n\to\infty}h_n=0$.

$$f_n^{(r)}(x)=\arg\max R(f^{(r)}(x))$$

Thus, the estimator of β(x) is shown by

$$\beta_n(x)=u_1(x)f_n^{(1)}(x)+u_2(x)f_n(x) \quad (2.2)$$

And, the empirical Bayes likelihood test function is defined as follows

$$\delta_n(x)=\begin{cases}1,\beta_n(x)\le 0\\ 0,\beta_n(x)>0\end{cases} \quad (2.3)$$

Let E stand for mathematical expectation with respect to the joint distribution of $X_{(1)1}$, $X_{(1)m}$,$X_{(2)1}$, $X_{(2)m}X_{(k)1}$, $X_{(k)m}$. Then, the overall Bayes risk of $\delta_n(x)$ is shown by

$$R(\delta_n(x),G)=a\int_{\Omega}\beta(x)E[\delta_n(x)]dx+C_G$$

If $\lim_{n\to\infty}$ $R(\delta_n,G)=R(\delta_G,G)$,{$\delta_n(x)$} is called asymptotic optimality of empirical Bayes likelihood test function.. Before proving the theorems, we need the following lemmas. Supposed that c,c1 be different constants in different cases even in the same expression.

Lemma [15]. $R(\delta_G,G)$ and $R(\delta_n,G)$ are defined by above, then

$0\le R(\delta_n,G)-R(\delta_G,G)\le c\int|\beta(x)|P(|\beta_n(x)-\beta(x)|\ge|\beta(x)|)dx\Omega$.

3. Asymptotic Optimality of Empirical Bayes Likelihood Test in Ranked Set Sampling

Theorem 3.1. Assume (A1) and the following regularity conditions hold.

(1) $h_n>0$, $\lim_{n\to\infty}h_n=0$,

(2) $\int\theta^3$ dG(θ)<∞,

(3) $f(x)$ is continuous function,

Then, $\lim_{n\to\infty}$ $R(\delta_n, G)=R(\delta_G,G)$.

Proof. Lemma 1 shows that $0\le R(\delta_n,G)-R(\delta_G,G)\le a\int|\beta(x)|P(|\beta_n(x)-\beta(x)|\ge|\beta(x)|)dx\Omega$.

Applying Fubini theorem, we have

$$\int_{\Omega}|\beta(x)|)\mathrm{d}x=\int_{\Omega}\int_{\Theta}\left|(\theta_0-1)^3+(3-\theta_0)(\theta-2)+(\theta+1)(\theta-2)\right|f(x|\theta)\mathrm{d}G(\theta)\,\mathrm{d}x\leqslant$$

$$\int_{\Omega}\int_{\Theta}\left|(\theta_0-1)^2\right|f(x|\theta)\mathrm{d}G(\theta)\,\mathrm{d}x+\int_{\Omega}\int_{\Theta}\left|(3-2\theta_0)(\theta-2)\right|f(x|\theta)\mathrm{d}G(\theta)\,\mathrm{d}x+$$

$$\int_{\Omega}\int_{\Theta}\left|(\theta+1)(\theta-2)\right|f(x|\theta)\mathrm{d}G(\theta)\,\mathrm{d}x\leqslant$$

$$(\theta_0-1)^3+\left|(3-2\theta_0)\right|\int_{\Theta}\left|\theta-1\right|\mathrm{d}G(\theta)+\int_{\Theta}\left|(\theta+1)(\theta-2)\right|\mathrm{d}G(\theta)<+\infty.$$

Denote $\Omega_n(x)$=$|\beta(x)|P(|\beta_n(x)-\beta(x)|\ge|\beta(x)|)$.

Obviously, $\Omega_n(x)\le|\beta(x)|$. Then, by domain convergence theorem, we get

$$0\le\lim_{n\to\infty} R(\delta_n, G)-R(\delta_G,G)\le\int[\lim_{n\to\infty}\Omega_n(x)]\Omega dx. \quad (3.1)$$

Next, we need prove that $\lim_{n\to\infty}\Omega_n(x)=0$ holds almost everywhere. By Markov's and Jensen's inequality,

$$\lim_{n\to\infty}\Omega_n(x)=0 \quad (3.2)$$

Substituting (3.2) into (3.1), the proof of theorem 3.1 is finished.

Acknowledgment

This paper was financially supported by Natural Science Foundation of China (12161075), Natural Science Foundation of Guangdong Province (2022A1515010978; 2024A1515011258) Natural Science Foundation of Jiangxi Province (20122ABC201006) and Natural Science Foundation of Guangdong Ocean University (R17083, C17201,P16091).

References

[1] Zhou, W and Jing, B. Y, Smoothed empirical likelihood confidence intervals for the difference of quantiles, Statist. Sinica, 2003, Vol.13, p83-96.

[2] Keziou, A and Leoni-Aubin, S, On empirical likelihood for semiparametric two-sample density ratio models, J. Statist. Plann. Inference, 2008, Vol.138,p 915-928.

[3] Liu, Y and Yu, C. W, Bartlett correctable two-sample adjusted empirical likelihood, J. Multivariate. Anal, 2010, Vol.101,p 1701-1711.

[4] Qin, Y. S and Zhang, S. C, Empirical likelihood confidence intervals for difference between two data sets with missing data, Pattern. Recognit. Lett, 2008, Vol.29,p 903-812.

[5] Shen, J. S & He, S.Y. Empirical likelihood for the difference of quantiles under censorship, Stat. Papers, 2007, 48:437-457.

[6] Zhou, Y. & Liang, H. Empirical-likelihood-based semi-parametric inference for the treatment effect in the two-sample problem with censoring, Biometrika, 2005, 92:271-282.

[7] Su, H.& Liang H. An empirical likelihood-based method for comparison of treatment effects-test of equality of coefficients in linear models, Comput. Stat. Data. An, 2010, 54: 1079-1088.

[8] McIntyre G A. A Method for Unbiased Selective Sampling, Using Ranked Sets (Aust J Agric Res 3, 1952) p.385-390.

[9] Chen Z H, Bai Z D, Sinha B K. Ranked Set Sampling: Theory and Applications (Springer 2003).

[10] Robbins, H. An empirical Bayes approach to statistics, Proc. Third Berkeley Symp. Math. Statist. Prob, Vol. 1 (1955), p.157--163.

[11] Karunamuni, R. J, Li, J, Wu, J. Robust empirical Bayes tests for continuous distributions, Journal of Statistical Planning and Inference, Vol. 140 (2010) No.1, p.268-282.

[12] Xinyi Xu, Dunke Zhou. Empirical Bayes predictive densities for high-dimensional normal models, Journal of Multivariate Analysis, Vol. 102 (2011) No.10, p.1417-1428.

[13] Johns M VJr, Van Ryzin J R. Convergence Rates in Empirical Bayes Two-action Problems 2: Continuous Case. Ann. Math. Statist. Vol. 43 (1972), p.934-937.

[14] Lichun Wang, Radhey S. Singh. Linear Bayes estimator for the two-parameter exponential family under type II censoring. Computational Statistics & Data Analysis,Vol. 71 (2014), p.633-642.

[15] Naiyi Li, Yuan Li, Yongming Li, Yang Liu. Empirical Bayes Inference for the Parameter of Power Distribution Based on Ranked Set Sampling. Discrete Dynamics in Nature and Society, (2015), p.1-5