

## Empirical Bayes two-sided test for the parameter of pareto distribution based on ranked set sampling

Liqun Xiao<sup>1</sup>, Naiyi Li<sup>2,\*</sup>

<sup>1</sup>School of Economics and Statistics, Guangzhou University, Guangzhou 510006, China;

<sup>2</sup>School of Mathematics and Computer, Guangdong Ocean University, Zhanjiang 524088, China.

\*linaiyi1979@163.com (Corresponding Author)

### Abstract

**In the paper, we propose the empirical Bayes test rules based on ranked set sampling under Pareto distribution. Its asymptotic optimality and convergence rate are proved strictly.**

### Keywords

**Ranked set sampling, empirical Bayes, Pareto distribution, asymptotic optimality, convergence rates.**

### 1. Introduction

Ranked set sampling (RSS) was first proposed by McIntyre (1952, [1]) to estimated pasture yields. It is a two-stage sampling plan where a number of sampling units are first ranked without taking actual measurements at a small cost, and then, measurements are taken from a fraction of the ranked units. This can improve the precision of statistical inference when the actual measurements are difficult or expensive to obtain, but sampling units can be easily ranked by some means without actual quantification. For this advantage, RSS has been applied successfully in many areas such as environment, ecology, industrial statistics and sociology. The reader is referred to the monograph by Chen et al. [2].

Since Robbins, H [3-4] introduced empirical Bayes (EB) approach, it has been developed in the literature[5-9]. However, most EB methods are based on simple random sampling (SRS). A heuristic idea is to develop EB methods based on RSS. Recently, in our another work (Li et al. [10]), empirical Bayes test rule and its asymptotical property for the parameter of power distribution based on RSS has been established. In this paper, we will construct empirical Bayes test rule for the parameter of Pareto distribution based on RSS.

Let  $X$  have a conditional density function for given  $\theta$

$$f(x|\theta) = \frac{\theta \alpha^\theta}{x^{\theta+1}}, \quad (1.1)$$

where  $\alpha$  is known parameter,  $\theta$  is unknown parameter,  $\Omega = \{x|x > \alpha\}$  is the sample space and  $\Theta = \{\theta|\theta > 0\}$  is parameter space. We then discuss the following two-sided test problem:

$$H_0: \theta_1 \leq \theta \leq \theta_2 \Leftrightarrow H_1: \theta < \theta_1 \text{ or } \theta > \theta_2 \quad (1.2)$$

where  $\theta_1$  and  $\theta_2$  are given constants. Let  $\theta_0 = \frac{\theta_1 + \theta_2}{2}$  and  $\gamma_0 = \frac{\theta_2 - \theta_1}{2}$ , then the two-sided test problem (1.2) is equivalent with:

$$H_0^* : |\theta - \theta_0| \leq \gamma_0 \Leftrightarrow H_1^* : |\theta - \theta_0| > \gamma_0. \quad (1.3)$$

For testing (1.3), we take loss function

$$L_i(\theta, d_i) = (1 - i)a[(\theta - \theta_0)^2 - \gamma_0^2]I_{\{|\theta - \theta_0| > \gamma_0\}} + ia[\gamma_0^2 - (\theta - \theta_0)^2]I_{\{|\theta - \theta_0| \leq \gamma_0\}},$$

$i = 0, 1$ , where  $a > 0$ ,  $d = \{d_0, d_1\}$  is action space,  $d_0$  and  $d_1$  imply acceptance and rejection of  $H_0^*$  respectively.

Suppose that the prior distribution  $G(\theta)$  of parameter the  $\theta$  is unknown.

We can get random decision function

$$\delta(x) = P(\text{accept } H_0^* | X = x). \tag{1.4}$$

Then, the risk function of  $\delta(x)$  is given by

$$\begin{aligned} R(\delta(x), G(\theta)) &= \int_{\Theta} \int_{\Omega} [L_0(\theta, d_0)f(x|\theta)\delta(x) + L_1(\theta, d_1)f(x|\theta)(1-\delta(x))] dx dG(\theta) = \\ &a \int_{\Omega} \beta(x)\delta(x)dx + C_G, \tag{1.5} \end{aligned}$$

where

$$C_G = \int_{\Theta} L_{1(\theta, d_1)} dG(\theta), \quad \beta(x) = \int_{\Theta} [(\theta - \theta_0)^2 - \gamma_0^2] f(x|\theta) dG(\theta). \tag{1.6}$$

The marginal density function of X is

$$f_{G(x)} = \int_{\Theta} f(x|\theta) dG(\theta) = \int_{\Theta} \theta \alpha^{\theta} x^{-(\theta+1)} dG(\theta). \tag{1.7}$$

By (1.6), we have

$$\beta(x) = A(x)f_G^{(2)}(x) + B(x)f_G^{(1)}(x) + Cf_G(x), \tag{1.8}$$

where  $A(x) = x^2, B(x) = (2\theta_0 + 3)x, C = (\theta_0 + 1)^2 - \gamma_0^2$ , and  $f_G^r(x)$  is the r-th order derivative of  $f_{G(x)}$ , for  $r = 0, 1, 2$ . Using (1.5), Bayes test function is obtained as follows

$$\delta_{G(x)} = \begin{cases} 1, & \beta(x) \leq 0 \\ 0, & \beta(x) > 0 \end{cases} \tag{1.9}$$

Further, we can get the minimum Bayes risk

$$R(G) = \inf_{\delta} R(\delta, G) = R(\delta_G, G) = \int_{\Omega} \beta(x)\delta_G(x)dx + C_G. \tag{1.10}$$

From the issue above that  $\delta(x) = \delta_G(x)$  and  $R(G)$  can be achieved when the prior distribution of  $G(\theta)$  is given. If not, we can use the EB method. The rest of this paper is organized as follows. Section 2 presents an EB test based on RSS. In section 3, we obtain asymptotic optimality and the optimal rate of convergence of the EB test.

## 2. Construction of EB test Based on Ranked Set Sampling

A balanced RSS procedure can be described as follows:

Randomly select k independent SRS samples from the population of interest.

The k units in each sample are ranked visually or by any negligible cost method that does not need actual measurements.

Only measure the smallest unit for each r size sample, r runs from 1 to k.

Repeat (i)-(iii) m times, then we obtain  $k \times m$  independent observations

$X_{(r)i}, r = 1, \dots, k, i = 1, \dots, m$ , that collected a balanced RSS sample of total size  $n = k \times m$ .

We then construct the EB test function. Let

$$X_{(1)1}, X_{(1)2}, \dots, X_{(1)m}, X_{(2)1}, X_{(2)2}, \dots, X_{(2)m}, \dots, X_{(k)1}, X_{(k)2}, \dots, X_{(k)m}$$

be a balanced ranked set sample from population which has the common marginal density function  $f_{G(x)}$ . We assume perfect ranking. Denote that

$$X_{(1)1}, X_{(1)2}, \dots, X_{(1)m}, X_{(2)1}, X_{(2)2}, \dots, X_{(2)m}, \dots, X_{(k)1}, X_{(k)2}, \dots, X_{(k)m}$$

are historical samples, and X is present sample. Assume  $f_{G(x)} \in C_{s,\alpha}, x \in R^1$ , where

$C_{s,\alpha} = \{g(x)|g(x) \text{ is a probability density function};$

the  $s - \text{th}$  order derivative  $g^{(s)}(x)$  is continuous with  $|g^{(s)}(x)| \leq \alpha, s \geq 3, \alpha > 0\}$ .

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

$$(C1): \frac{1}{t!} \int_0^1 y^t K_r(y) dy = \begin{cases} (-1)^t, & \text{when } t = r \\ 0, & \text{when } t \neq r, \end{cases} \quad t = 0, 1, 2, \dots, s - 1.$$

Kernel estimator of  $f_G^{(r)}(x)$  is defined by

$$f_n^{(r)}(x) = \frac{1}{m k h_n^{(1+r)}} \sum_{i=1}^k \sum_{j=1}^m K_r \left( \frac{x - X_{(i)j}}{h_n} \right) \tag{2.1}$$

where  $h_n$  is a positive and smoothing bandwidth, and  $\lim_{n \rightarrow \infty} h_n = 0$ . Denote  $f_G^{(0)}(x) = f_G(x), f_G^{(r)}(x)$  is the  $r$ -th order derivative of  $f_G(x)$ , for  $r = 0, 1, 2$ . Thus, the estimator of  $\beta(x)$  is

$$\beta_n(x) = A(x) f_n^{(2)}(x) + B(x) f_n^{(1)}(x) + C f_n(x). \tag{2.2}$$

And, the EB test function is defined as

$$\delta_n(x) = \begin{cases} 1, & \beta_n(x) \leq 0 \\ 0, & \beta_n(x) > 0 \end{cases} \tag{2.3}$$

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

$$R(\delta_n(x), G) = a \int_{\Omega} \beta(x) E[\delta_n(x)] dx + C_G. \tag{2.4}$$

If  $\lim_{n \rightarrow \infty} R(\delta_n, G) = R(\delta_G, G), \{\delta_n(x)\}$  is called asymptotic optimality of EB test function. If  $R(\delta_n, G) - R(\delta_G, G) = O(n^{-q})$ , where  $q > 0, O(n^{-q})$  is asymptotic optimality convergence rates of EB test function  $\{\delta_n(x)\}$ . Before proving the theorems, we need the following lemmas. Let  $c, c_1, c_2, c_3$  be different constants in different cases even in the same expression.

Lemma 2.1. Let  $X_{(1)1}, X_{(1)2}, \dots, X_{(1)m}, X_{(2)1}, X_{(2)2}, \dots, X_{(2)m}, \dots, X_{(k)1}, X_{(k)2}, \dots, X_{(k)m}$  be balanced ranked set samples. Suppose that (C1) hold,  $\forall x \in \Omega$ ,

When  $f_G^{(r)}(x)$  is continuous function,  $\lim_{n \rightarrow \infty} h_n = 0$ , and  $\lim_{n \rightarrow \infty} n h_n^{2r+1} = \infty$ , we have

$$\lim_{n \rightarrow \infty} E \left| f_n^{(r)}(x) - f_G^{(r)}(x) \right|^2 = 0.$$

When  $f_G^{(r)}(x) \in C_{s,\alpha}$ , putting  $h_n = n^{-\frac{1}{2+r}}$ , for  $0 < \lambda \leq 1$ , we have

$$E \left| f_n^{(r)}(x) - f_G^{(r)}(x) \right|^{2\lambda} \leq c \cdot n^{-\frac{\lambda(s-2r+1)}{2+s}}.$$

*Proof.* Proof of (I): Using  $C_r$  inequation, we have

$$\begin{aligned} E \left| f_n^{(r)}(x) - f_G^{(r)}(x) \right|^2 &\leq 2 \left| E f_n^{(r)}(x) - f_G^{(r)}(x) \right|^2 + 2 \text{Var} \left( f_n^{(r)}(x) \right) \\ &:= 2(A_1^2 + A_2), \end{aligned} \tag{2.5}$$

where

$$E f_n^{(r)}(x) = \sum_{i=1}^k \sum_{j=1}^m n^{-1} h_n^{-(r+1)} E \left[ K_r \left( \frac{x - X_{(i)j}}{h_n} \right) \right] = h_n^{-(r+1)} E \left[ K_r \left( \frac{x - X}{h_n} \right) \right]$$

$$\begin{aligned}
 &= h_n^{-(r+1)} \int_0^\infty K_r\left(\frac{x-X}{h_n}\right) f_G(y) dy \\
 &= h_n^{-r} \int_0^1 K_r(u) f_G(x - h_n u) du.
 \end{aligned}$$

The Taylor expansion shows

$$f_G(x - h_n u) - f_G(u) = \frac{f_G^{(1)}(x)}{1!} (-h_n u) + \frac{f_G^{(2)}(x)}{2!} (-h_n u)^2 + \dots + \frac{f_G^{(s)}(x - \xi h_n u)}{s!} (-h_n u)^s,$$

where  $0 < \xi < 1$ . Because  $f_G^{(r)}(x)$  is continuous in  $x$ , then

$$\begin{aligned}
 0 \leq \lim_{n \rightarrow \infty} |E f_n^{(r)}(x) - f_G^{(r)}(x)| &= \lim_{n \rightarrow \infty} \left| \frac{1}{h_n^r} \int_0^1 K_r(u) f_G(x - h_n u) du - f_G^{(r)}(x) \right| \\
 &\leq \frac{1}{r!} \int_0^1 |K_r(u)| \lim_{n \rightarrow \infty} |f_G^{(r)}(x - \xi h_n u) - f_G^{(r)}(x)| du = 0,
 \end{aligned}$$

and

$$\lim_{n \rightarrow \infty} A_1^2 = \lim_{n \rightarrow \infty} |E f_n^{(r)}(x) - f_G^{(r)}(x)|^2 = 0. \tag{2.6}$$

It is easy to see that

$$\begin{aligned}
 A_2 &= 2Var\left(f_n^{(r)}(x)\right) = 2 \sum_{i=1}^k \sum_{j=1}^m n^{-2} h_n^{-2(r+1)} Var\left[K_r\left(\frac{x - X_{(i)j}}{h_n}\right)\right] \\
 &\leq 2n^{-2} h_n^{-2(r+1)} \sum_{i=1}^k \sum_{j=1}^m E\left[K_r\left(\frac{x - X_{(i)j}}{h_n}\right)\right]^2 \leq c \cdot (nh_n^{2r+1})^{-1}.
 \end{aligned} \tag{2.7}$$

When  $h_n \rightarrow 0$  and  $nh_n^{2r+1} \rightarrow \infty$ , we get

$$\lim_{n \rightarrow \infty} A_2 = \lim_{n \rightarrow \infty} Var\left(f_n^{(r)}(x)\right) = 0. \tag{2.8}$$

Substituting (2.6) and (2.8) into (2.5), proof of (I) is finished.

Proof of (II): Similar to (2.5), we can show that

$$\begin{aligned}
 E\left|f_n^{(r)}(x) - f_G^{(r)}(x)\right|^{2\lambda} &\leq 2\left|E f_n^{(r)}(x) - f_G^{(r)}(x)\right|^{2\lambda} + 2Var\left(f_n^{(r)}(x)\right)^\lambda \\
 &:= 2(B_1^{2\lambda} + B_2^\lambda),
 \end{aligned} \tag{2.9}$$

Due to  $A_1$  and  $f_G^{(r)}(x) \in C_{s,\alpha}$ ,

$$E\left|f_n^{(r)}(x) - f_G^{(r)}(x)\right| \leq \int_0^1 |K_r(v)| h_n^{s-r} v^s \left| \frac{f_G^{(r)}(x - \xi h_n v)}{s!} \right| dv \leq c \cdot h_n^{s-r}.$$

Therefore, taking  $h_n = n^{-\frac{1}{2+s}}$ , we have

$$B_1^{2\lambda} = 2\left|E f_n^{(r)}(x) - f_G^{(r)}(x)\right|^{2\lambda} \leq c \cdot n^{-\frac{2\lambda(s-r)}{2+s}}. \tag{2.10}$$

And by (2.8), we obtain

$$B_2^\lambda \leq [c_1 \cdot (nh_n^{2r+1})^{-1}]^\lambda \leq c \cdot n^{-\frac{\lambda(s-2r+1)}{2+s}}. \tag{2.11}$$

Substituting (2.10) and (2.11) into (2.9), the proof of (II) is finished.

Lemma 2.2.  $[^8] R(\delta_G, G)$  and  $R(\delta_n, G)$  are defined by (1.10) and (2.4), then

$$0 \leq R(\delta_n, G) - R(\delta_G, G) \leq a \int_{\Omega} |\beta(x)| P(|\beta_n(x) - \beta(x)| \geq |\beta(x)|) dx.$$

### 3. Asymptotic Optimality and Convergence Rates of Empirical Bayes test Based on Ranked Set Sampling

Theorem 3.1. Assume (C1) and the following regularity conditions

$$h_n > 0, \lim_{n \rightarrow \infty} h_n = 0, \lim_{n \rightarrow \infty} nh_n^5 = \infty,$$

$$\int_{\Theta} \theta^2 dG(\theta) < \infty,$$

$f_G^{(2)}(x)$  is continuous function,

hold. Then,

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

*Proof.* Lemma 2.2 shows that

$$0 \leq R(\delta_n, G) - R(\delta_G, G) \leq a \int_{\Omega} |\beta(x)| P(|\beta_n(x) - \beta(x)| \geq |\beta(x)|) dx.$$

Applying (1.6) and Fubini theorem, we have

$$\int_{\Omega} |\beta(x)| dx \leq |\theta_0^2 - \gamma_0^2| + \int_{\Theta} \theta^2 dG(\theta) + 2|\theta_0| \int_{\Theta} \theta dG(\theta) < \infty.$$

Denote  $T_n(x) = |\beta(x)| P(|\beta_n(x) - \beta(x)| \geq |\beta(x)|)$ . Obviously,  $T_n(x) \leq |\beta(x)|$ . Then, by domain convergence theorem, we can get

$$0 \leq \lim_{n \rightarrow \infty} R(\delta_n, G) - R(\delta_G, G) \leq \int_{\Omega} \left[ \lim_{n \rightarrow \infty} T_n(x) \right] dx. \tag{3.1}$$

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

$$T_n(x) \leq |A(x)| \left[ \left| E f_n^{(2)}(x) - f_G^{(2)}(x) \right|^2 \right]^{\frac{1}{2}} + |B(x)| \left[ \left| E f_n^{(1)}(x) - f_G^{(1)}(x) \right|^2 \right]^{\frac{1}{2}} + |C| \left[ \left| E f_n(x) - f_G(x) \right|^2 \right]^{\frac{1}{2}}.$$

For fixed  $x \in \Omega$ ,  $r = 0,1,2$  and  $\lambda = 1$ , we have

$$0 \leq \lim_{n \rightarrow \infty} T_n(x) \leq A(x) \left[ \lim_{n \rightarrow \infty} E \left| f_n^{(2)}(x) - f_G^{(2)}(x) \right|^2 \right]^{\frac{1}{2}} + B(x) \left[ \lim_{n \rightarrow \infty} E \left| f_n^{(1)}(x) - f_G^{(1)}(x) \right|^2 \right]^{\frac{1}{2}} + |C| \left[ \lim_{n \rightarrow \infty} E \left| f_n(x) - f_G(x) \right|^2 \right]^{\frac{1}{2}} = 0, \tag{3.2}$$

by (I) in lemma 2.1 and lemma 2.2. Substituting (3.2) into (3.1), the proof of theorem 3.1 is finished.

Theorem 3.2. Assume (C1) and the following regularity conditions

$f_G(x) \in C_{s,\alpha}$ , where  $s \geq 4$ ,

$$h_n = n^{-\frac{1}{2+s}},$$

$\int_{\Omega} x^{m\lambda} |\beta(x)|^{1-\lambda} dx < \infty$ , for  $0 < \lambda \leq 1$ , and  $m = 0,1,2$ ,

hold. Then,

$$R(\delta_n, G) - R(\delta_G, G) = o \left( n^{-\frac{\lambda(s-2)}{2(1+s)}} \right).$$

*Proof.* Using Lemma 2.2 and Markov's inequality, we have

$$\begin{aligned}
 0 \leq R(\delta_n, G) - R(\delta_G, G) &\leq c_1 \int_{\Omega} |\beta(x)|^{1-\lambda} |A(x)|^\lambda E \left| f_n^{(2)}(x) - f_G^{(2)}(x) \right|^\lambda dx \\
 &\quad + c_2 \int_{\Omega} |\beta(x)|^{1-\lambda} |B(x)|^\lambda E \left| f_n^{(1)}(x) - f_G^{(21)}(x) \right|^\lambda dx \\
 &\quad + c_3 \int_{\Omega} |\beta(x)|^{1-\lambda} |C|^\lambda E |f_n(x) - f_G(x)|^\lambda dx \\
 &= A_n + B_n + C_n.
 \end{aligned} \tag{3.3}$$

Applying (II) in Lemma 2.1 and the conditions (4)-(5) in the Theorem (3.2), we have

$$A_n \leq c_1 n^{-\frac{\lambda(s-2)}{2(1+s)}} \int_{\Omega} |\beta(x)|^{1-\lambda} |A(x)|^\lambda dx \leq c_4 n^{-\frac{\lambda(s-2)}{2(1+s)}}. \tag{3.4}$$

$$B_n \leq c_2 n^{-\frac{\lambda(s-1)}{2(1+s)}} \int_{\Omega} |\beta(x)| |A(x)|^\lambda dx \leq c_5 n^{-\frac{\lambda(s-1)}{2(1+s)}}. \tag{3.5}$$

$$C_n \leq c_3 n^{-\frac{\lambda s}{2(1+s)}} \int_{\Omega} |\beta(x)|^{1-\lambda} |C|^\lambda dx \leq c_6 n^{-\frac{\lambda(s-12)}{2(1+s)}}. \tag{3.6}$$

Substituting (3.4)-(3.6) into (3.3), we have

$$R(\delta_n, G) - R(\delta_G, G) = O\left(n^{-\frac{\lambda(s-2)}{2(1+s)}}\right).$$

The proof of theorem 3.2 is finished.

Remark 3.1. When  $\lambda \rightarrow 1$ ,  $O\left(n^{-\frac{\lambda(s-2)}{2(1+s)}}\right)$  nears  $O\left(n^{-\frac{1}{2}}\right)$ .

### Acknowledgements

This work was partially supported by the National Social Science Foundation of China(15BTJ031), National Natural Science Foundation of China (11601097; 11461057; 11561010), Natural Science Foundation of Guangdong Province (2016A030313812) and Natural Science Foundation of Guangdong Ocean University (C17201).

### References

- [1] McIntyre G A. *A Method for Unbiased Selective Sampling, Using Ranked Sets* (Aust J Agric Res 3, 1952) p.385-390.
- [2] Chen Z H, Bai Z D, Sinha B K. *Ranked Set Sampling: Theory and Applications* (Springer 2003).
- [3] Robbins, H. An empirical Bayes approach to statistics, Proc. Third Berkeley Symp. Math. Statist. Prob, Vol. 1 (1955), p.157--163.
- [4] Robbins, H. The empirical Bayes approach to statistical decision problem, Ann. Math. Statist. Vol. 35 (1964), p.1-20.
- [5] 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.
- [6] 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.
- [7] Xiuchun Li, Yimin Shi, Jieqiong Wei, Jian Chai. Empirical Bayes estimators of reliability performances using Linear loss under progressively Type-II censored samples. Mathematics and Computers in Simulation, Vol. 73 (2007) No.5, p.320-326.
- [8] 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.
- [9] 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.

- [10]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.
- [11]Pareto V. *Cour Economic Politique*. (Lausanne and pairs ,Rouge and Cie 1987).