〈**Research Paper**〉

Journal of the Korean
Statistical Society
Vol. 13, No. 1, 1984

Jackknife Estimation of the Coefficient of Variation in the Pareto Distribution

Jungsoo Woo and Suk Bok Kang*

ABSTRACT

In this paper, the means of the estimators for the coefficient of variation (CV) in an underlying Pareto distribution are expressed in terms of confluent hypergeometric functions. The numerical values of the biases for the CV estimators in the Pareto distribution are also obtained.

1. Introduction

We consider the Pareto distribution with two parameters defined by the density function

$$f(x;\xi,\ \theta)=\xi\theta^{\xi}/x^{\xi+1},\quad x\geq\theta>0,\quad \xi>0, \qquad (1.1)$$
$$=0,\quad x<\theta,$$

where ξ is a shape parameter and θ is a scale parameter. The Pareto distribution has been used in several situations (see Johnson and Kotz(1970, Chap. 19)). Some remarks and the basic properties of the Pareto distribution can be found in Johnson and Kotz (1970). The means and covariances of the order statistics from the Pareto distribution are given by Malik(1966).

As a general method for reduction of bias in the parametric model, jackknife technique has been discussed by many authors. Sharot(1976) has proposed a new family of jackknives which include the first and second order jackknife estimators as special cases. In this paper, we investigate the bias reduction properties of the jackknife estimators of the coefficient of variation (CV) in the Pareto distribution.

* Department of Statistics, Yeungnam University, Gyongsan 632, Korea

2. The Jackknife Estimations of the Coefficient of Variation

We wish to know whether the jackknife method is effective or not in reducing bias of CV estimators in the Pareto distribution.

Let $X_1, X_2, \ldots, X_n$ be a random sample of size n from the Pareto distribution (1. 1). Then $\hat{\theta}=\min(X_1, X_2, \ldots, X_n)$ and $\hat{\xi}=n/\sum_{j=1}^{n}\log(X_j/\hat{\theta})$ are the MLE's of θ and ξ in (1. 1), respectively. Suppose $X_{(1)}, X_{(2)}, \ldots, X_{(n)}$ are the corresponding order statistics of the above sample. Since $\hat{\xi}=n[\sum_{j=1}^{n}\log(X_j/\theta)]^{-1}$ is stochastically independent of $\hat{\theta}=\min(X_1, X_2, \ldots, X_n)$, here we may consider the case that the scale parameter is known.

We take $\hat{\xi}=n[\sum_{j=1}^{n}\log(X_j/\theta)]^{-1}$ as the maximum likelihood estimator of the shape parameter when the scale parameter is known. The density function of $\hat{\xi}$ is

$$f_{\hat{\xi}}(x)=\frac{\xi^n n^n}{\Gamma(n)}x^{-(n+1)}e^{-\xi n/x},\quad x>0. \tag{2. 1}$$

Note that $2\xi n/\hat{\xi}$ is distributed as χ^2 with $2n$ degrees of freedom.

Let $\xi^*=(n-1)\hat{\xi}/n$. Then, by Lehmann-Scheffe theorem, ξ^* is the uniformly minimum variance unbiased estimator (UMVUE) of the shape parameter in the Pareto distribution under fixed known θ. The coefficient of variation in the Pareto distribution under fixed known θ is $[\xi(\xi-2)]^{-1/2}$ $(\xi>2)$, and its MLE is expressed by

$$\hat{CV}=[\hat{\xi}(\hat{\xi}-2)]^{-1/2}.$$

From the density function of $\hat{\xi}$ in (2. 1) and the formula 3. 471 in Gradshteyn and Pyzhik(1965),

$$E[\hat{CV}]=\xi^n n^{n+1}\Gamma(1/2)\,{}_1F_1(n+1;\ n+3/2;\ -\xi n/2)/2^{n+1}\Gamma(n+3/2), \tag{2. 2}$$

where ${}_1F_1(a;b;c)$ is the Kummer's function.

Let us consider the R-estimator which is defined by $CV^*=[\xi^*(\xi^*-2)]^{-1/2}$, where ξ^* is the uniformly minimum variance unbiased estimator of ξ when the scale parameter is known. Then $CV^*=n[\hat{\xi}(\hat{\xi}-2n/(n-1))]^{-1/2}/(n-1)$.

By applying formula 3. 471 in Gradshteyn and Pyzhik(1965), we obtain

$$E[CV^*]=\xi^n\, n(n-1)^n\Gamma(1/2)\,{}_1F_1(n+1;\ n+3/2;\ -\xi(n-1)/2)/2^{n+1}\Gamma(n+3/2). \tag{2. 3}$$

Since $\hat{CV}=\{n^2[\sum_{j=1}^{n}\log(X_j/\theta)]^{-2}-2n[\sum_{j=1}^{n}\log(X_j/\theta)]^{-1}\}^{-1/2}$,

the jackknife estimators for $\hat{CV}$ are given by

$$J(\hat{CV})=\sqrt{n}\,\{n[\sum_{j=1}^{n}\log(X_j/\theta)]^{-2}-2[\sum_{j=1}^{n}\log(X_j/\theta)]^{-1}\}^{-1/2}$$

$$-\sqrt{n-1}n^{-1}\sum_{i=1}^{n}\{(n-1)[\sum_{\substack{j=1\\j\neq i}}^{n}\log(X_j/\theta)]^{-2}-2[\sum_{\substack{j=1\\j\neq i}}^{n}\log(X_j/\theta)]^{-1}\}^{-1/2},$$

$$J^{(2)}(\hat{CV})=n^2\{n^2[\sum_{j=1}^{n}\log(X_j/\theta)]^{-2}-2n[\sum_{j=1}^{n}\log(X_j/\theta)]^{-1}\}^{-1/2}/2$$

$$-\frac{(n-1)^2}{n}\sum_{i=1}^{n}\{(n-1)^2[\sum_{\substack{j=1\\j\neq i}}^{n}\log(X_j/\theta)]^{-2}$$

$$-2(n-1)[\sum_{\substack{j=1\\j\neq i}}^{n}\log(X_j/\theta)]^{-1}\}^{-1/2}$$

$$+(n-2)^2\sum_{i<j}\{(n-2)^2[\sum_{\substack{l=1\\l\neq i,j}}^{n}\log(X_l/\theta)]^{-2}$$

$$-2(n-2)[\sum_{\substack{l=1\\l\neq i,j}}^{n}\log(X_l/\theta)]^{-1}\}^{-1/2}/n(n-1).$$

And the jackknife estimators for CV^* are given by

$$J(CV^*)=n\{[\sum_{j=1}^{n}\log(X_j/\theta)]^{-2}-2[\sum_{j=1}^{n}\log(X_j/\theta)]^{-1}/(n-1))\}^{-1/2}/(n-1)$$

$$-(n-1)\sum_{i=1}^{n}\{[\sum_{\substack{j=1\\j\neq i}}^{n}\log(X_j/\theta)]^{-2}-2[\sum_{\substack{j=1\\j\neq i}}^{n}\log(X_j/\theta)]^{-1}/(n-2)\}^{-1/2}/n(n-2),$$

$$J^{(2)}(CV^*)=n^2\{[\sum_{j=1}^{n}\log(X_j/\theta)]^{-2}-2[\sum_{j=1}^{n}\log(X_j/\theta)]^{-1}/(n-1)\}^{-1/2}/2(n-1)$$

$$-(n-1)^2\sum_{i=1}^{n}\{[\sum_{\substack{j=1\\j\neq i}}^{n}\log(X_j/\theta)]^{-2}-2[\sum_{\substack{j=1\\j\neq i}}^{n}\log(X_j/\theta)]^{-1}/(n-2)\}^{-1/2}/n(n-2)$$

$$+\frac{(n-2)^2}{n(n-1)(n-3)}\sum_{i<j}\{[\sum_{\substack{l=1\\l\neq i,j}}^{n}\log(X_l/\theta)]^{-2}-2[\sum_{\substack{l=1\\l\neq i,j}}^{n}\log(X_l/\theta)]^{-1}/(n-3)\}^{-1/2}.$$

The means of these jackknife estimators are as follows;

$$E[J(\hat{CV})]=\xi^{n-1}2^{-n}\Gamma\left(\frac{1}{2}\right)[\xi n^{n+2}{}_1F_1(n+1;\ n+3/2;\ -\xi n/2)/(2n+1)$$

$$-(n-1)^{n+1}{}_1F_1(n;\ n+1/2;\ -\xi(n-1)/2)]/\Gamma(n+1/2), \qquad (2.4)$$

$$E[J^{(2)}(\hat{CV})]=\frac{\xi^{n-2}\Gamma(1/2)}{2^{n-1}\Gamma(n-1/2)}[\frac{\xi^2n^{n+3}}{2(4n^2-1)}{}_1F_1(n+1;\ n+3/2;\ -\xi n/2)$$

$$-\frac{\xi(n-1)^{n+2}}{2n-1}{}_1F_1(n;n+1/2;\ -\xi(n-1)/2)$$

$$+(n-2)^{n+1}{}_1F_1(n-1;\ n-1/2;\ -\xi(n-2)/2)], \qquad (2.5)$$

$$E[J(CV^*)]=\frac{\xi^{n-1}\Gamma(1/2)}{2^n\Gamma(n+1/2)}[\frac{\xi n^2(n-1)^n}{2n+1}{}_1F_1(n+1;n+3/2;\ -\xi(n-1)/2)$$

$$-(n-1)^2(n-2)^{n-1}{}_1F_1(n;n+1/2;\ -\xi(n-2)/2)], \qquad (2.6)$$

and

$$E[J^{(2)}(CV^*)]=\frac{\xi^{n-2}\Gamma(1/2)}{2^{n-1}\Gamma(n-1/2)}[\frac{\xi^2n^3(n-1)^n}{2(4n^2-1)}{}_1F_1(n+1;\ n+3/2;\ -\xi(n-1)2)$$

$$-\frac{\xi(n-1)^3(n-2)^{n-1}}{2n-1}{}_1F_1(n;\ n+1/2;\ -\xi(n-2)/2)$$
$$+\frac{1}{2}(n-2)^3(n-3)^{n-2}{}_1F_1(n-1;\ n-1/2;\ -\xi(n-3)/2)]. \qquad (2.7)$$

From the results (2.2) to (2.7), the numerical values of the relative biases for the estimators of CV in the Pareto distribution are evaluated for $n=4(2)30$ and $CV=2, 3, 4, 5$. The values are tabulated in Table 1. From Table 1, in the sence of bias the following results are obtained;

a) The jackknife estimator is better than the R-estimator and MLE of CV in the given distribution.

b) The MLE of CV is better than the R-estimator.

c) The second order jackknife estimator for the MLE of CV is better than the other estimators.

d) The second order jackknife estimator for the R-estimator is better than the first order jackknife estimator for the R-estimator and MLE of CV.

e) The absolute relative bias increases whenever the coefficient of variation increases. Hence, the jackknifing is effective in reducing bias for the estimators of CV in the Pareto distribution when the scale parameter is known.

Table 1. The absolute relative biases of the R-estimator, the MLE of CV in the Pareto distribution, and 1st & 2nd order jackknife estimators for the R-estimator and the MLE of CV.

CV	n	$\hat{CV}$	CV^*	$J(\hat{CV})$	$J^{(2)}(\hat{CV}^*)$	$J(CV^*)$	$J^{(2)}(CV^*)$
2	4	0.7280	0.7915	0.6468	0.6093	0.6860	0.6467
	6	0.6866	0.7407	0.5911	0.5399	0.6251	0.5691
	8	0.6546	0.7029	0.5486	0.4884	0.5785	0.5125
	10	0.6281	0.6722	0.5135	0.4463	0.5402	0.4667
	12	0.6052	0.6460	0.4833	0.4104	0.5073	0.4278
	14	0.5849	0.6231	0.4566	0.3788	0.4783	0.3938
	16	0.5666	0.6025	0.4326	0.3504	0.4522	0.3633
	18	0.5498	0.5838	0.4106	0.3247	0.4285	0.3357
	20	0.5344	0.5666	0.3904	0.3010	0.4067	0.3103
	22	0.5200	0.5507	0.3715	0.2790	0.3864	0.2869
	24	0.5065	0.5358	0.3539	0.2586	0.3674	0.2651
	26	0.4938	0.5218	0.3374	0.2394	0.3497	0.2446
	28	0.4818	0.5086	0.3217	0.2214	0.3329	0.2254
	30	0.4704	0.4961	0.3069	0.2043	0.3170	0.2073
3	4	0.8218	0.8665	0.7700	0.7464	0.7999	0.7757
	6	0.7956	0.8348	0.7353	0.7037	0.7627	0.7284

CV	n	$\hat{CV}$	CV^*	$J(\hat{CV})$	$J^{(2)}(\hat{CV})$	$J(CV^*)$	$J^{(2)}(CV^*)$
	8	0.7756	0.8113	0.7092	0.6722	0.7344	0.6944
	10	0.7590	0.7924	0.6878	0.6468	0.7113	0.6670
	13	0.7448	0.7764	0.6694	0.6250	0.6915	0.6438
	14	0.7323	0.7624	0.6532	0.6059	0.6741	0.6235
	16	0.7210	0.7499	0.6386	0.5888	0.6585	0.6054
	18	0.7108	0.7385	0.6253	0.5732	0.6443	0.5885
	20	0.7013	0.7281	0.6130	0.5588	0.6313	0.5737
	22	0.6925	0.7185	0.6016	0.5454	0.6193	0.5596
	24	0.6842	0.7095	0.5908	0.5328	0.6078	0.5464
	26	0.6765	0.7011	0.5807	0.5210	0.5971	0.5340
	28	0.6691	0.6931	0.5712	0.5098	0.5870	0.5223
	30	0.6621	0.6855	0.5621	0.4992	0.5774	0.5112
4	4	0.8673	0.9015	0.8292	0.8119	0.8526	0.8350
	6	0.8481	0.8783	0.8040	0.7811	0.8257	0.8010
	8	0.8335	0.8612	0.7852	0.7586	0.8053	0.7767
	10	0.8215	0.8475	0.7698	0.7404	0.7888	0.7572
	12	0.8112	0.8360	0.7567	0.7250	0.7747	0.7408
	14	0.8022	0.8259	0.7245	0.7115	0.7624	0.7266
	16	0.7941	0.8169	0.7348	0.6994	0.7514	0.7138
	18	0.7867	0.8088	0.7253	0.6885	0.7414	0.7023
	20	0.7799	0.8014	0.7167	0.6784	0.7322	0.6917
	22	0.7736	0.7745	0.7086	0.6690	0.7237	0.6818
	24	0.7677	0.7881	0.7011	0.6602	0.7157	0.6727
	26	0.7622	0.7821	0.6940	0.6520	0.7082	0.6641
	28	0.7569	0.7764	0.6873	0.6442	0.7012	0.6559
	30	0.7520	0.7711	0.6809	0.6368	0.6945	0.6482
5	4	0.8937	0.9214	0.8633	0.8496	0.8825	0.8685
	6	0.8784	0.9030	0.8434	0.8253	0.8612	0.8417
	8	0.8668	0.8895	0.8286	0.8076	0.8453	0.8227
	10	0.8574	0.8786	0.8165	0.7934	0.8323	0.8075
	12	0.8492	0.8695	0.8062	0.7814	0.8213	0.7948
	14	0.8421	0.8616	0.7972	0.7709	0.8115	0.7836
	16	0.8357	0.8545	0.7891	0.7615	0.8030	0.7738
	18	0.8299	0.8483	0.7817	0.7530	0.7953	0.7648
	20	0.8246	0.8423	0.7750	0.7452	0.7881	0.7566
	22	0.8197	0.8369	0.7687	0.7380	0.7815	0.7491
	24	0.8151	0.8319	0.7629	0.7312	0.7754	0.7420
	26	0.8107	0.8273	0.7574	0.7249	0.7696	0.7354
	28	0.8066	0.8228	0.7522	0.7187	0.7641	0.7292
	30	0.8028	0.8187	0.7413	0.7132	0.7589	0.7233

REFERENCES

(1) Efron, B. (1982). *The Jackknife, the Bootstrap and Other Resampling Plans*, Society for Industrial and Applied Mathematics.

(2) Gradshteyn, I.S. and Pyzhik, I.M. (1965). *Tables of Integrals Series and Products*, Academic Press, New York.

(3) Gray, H.L. and Schucany, W.R. (1972). *The Gemeralized Jackknife Statistic*, Marcel Dekker. New York.

(4) Johnson, N.L. and Kotz, S. (1970). *Continuous Univariate Distribution*-1, John Wiley and Sons, New York.

(5) Malik, H.J. (1966). Exact Moments of Order Statistics from the Pareto Distridution, *Skand. Aktuarietidskr*, 144-57.

(6) Setô, N. and Iwase, K. (1982). Uniformly Minimum Variance Unbiased Estimation of Quantiles and Probabilities for the Pareto Distribution of the First Kind, *J. Japan Statist. Soc.*, Vol. 12, No. 2, 105-112.

(7) Sharot, T. (1976). Sharpening the Jackknife, *Biometrika*, Vol. 63, No. 2, 315-321.

(8) Vännman, K. (1976). Estimators Based on Order Statistics from a Pareto Distribution, *J. Amer. Statist. Assoc.*, Vol. 71, No. 355, 704-708.