Kyungpook Math, J.
Volume 23, Number 2
December, 1983

ESTIMATION OF THE SCALE PARAMETER IN THE PARETO DISTRIBUTION BY THE QUICK MEASURES

By Jungsoo Woo and Kgoang Hoo Lee

1. Introduction

Assume that n observations are made on a random variable with a Pareto distribution

$$F(x)=1-\left(1+\frac{x-\alpha}{\beta}\right)-\gamma \qquad \text{for } x\geq\alpha>0, \qquad (1.1)$$

where, $\beta>0$ ad $\gamma>0$. The Pareto distribution has been used in connecion with studies of income, insurance risk, migration, size of cities and firms, word frequencies, and business morality etc. [4].

In this paper, we shall consider the problems of estimating the scale parameter in the distribution (1.1) by use of the quick measures. Quick measures whose principal merits are their simplicity, and it has not been influenced by the censored data. Here we shall consider the range, the quasi-range, the thickened range, and the Downton's unbiased estimator as the quick measures to estimate the scale parameter in the distribution (1.1).

A number of authers, including Gupta and Singh [3], Lee and Kapadia [6], considered the statistical properties of the range and quasi-range. The thickened range was introduced by Jones, advocated by Prescott, and further studied by D'Agotino and Cureton [2] only when the population is normal distribution. Barnett et. al. [1] studied the Downton's unbiased estimator, and pointed out that it had high efficient and not so influenced by outliers as the range for the normal case.

The relative mean square error will be defined as the mean square error divided by the unknown scale parameter.

2. Mean square error of the quasi-range

The i-th quasi-range, say W_i, of a sample of size n is defined as the range of $n-2(i-1)$ sample values deleting the $(i-1)$ largest and the $(i-1)$ smallest samples. That is,

$$W_i=X_{(n+1-i)}-X_{(i)}, \quad (i=1, 2, \cdots, [(n+1)/2]),$$

where $X_{(1)}, X_{(2)}, \cdots, X_{(n)}$ are the order statistics from a random sample of size n from the distribution (1.1). Obviously the 1st quasi-range is a range. We shall denote the range by W. These quasi-ranges will be useful in the censored samples, and obviously have some robustness against outliers.

Now, in order to obtain the mean square error of the quasi-range of random samples from the distribution (1.1), the mean and the covariance of the order statistic of a random sample from the Pistribution (1.1) can be easily checked as follows;

$$E(X_{(i)})=\alpha+(A_i-1)\beta, \text{ if } \gamma>(n-i+1)^{-1}, \tag{2.1}$$

and

$$Cov(X_{(i)}, X_{(j)})=A_j(A_i^{-1}B_i^{-1}-A_i)\beta^2, \tag{2.2}$$

if $\gamma>\max\{(n-i+1)^{-1}, (n-j+1)^{-1}\}$, ($i\leq j$ and $i, j=1, 2, \cdots, n.$)
where,

$$A_i=\Gamma(n-i+1-1/\gamma)\Gamma(n+1)/\{\Gamma(n-i+1)\Gamma(n+1-1/\gamma)\}$$
$$B_i=\Gamma(n-i+1)\Gamma(n+1-2/\gamma)/\{\Gamma(n-i+1-2/\gamma)\Gamma(n+1)\}.$$
$$(i=1, 2, \cdots, n.)$$

It follows that

$$MSE(W_i)=(B_{n+1-i}^{-1}-A_{n+1-i}^{2}+B_i^{-1}-A_i^2-2A_{n+1-i}(A_i^{-1}B_i^{-1}-A_i)+(A_{n+1-i}-A_i-1)^2)\beta^2, \tag{2.3}$$

where $i=1, 2, \cdots, [(n+1)/2]$.

From (2.3), the exact numerical values of the mean square error of the quasi-ranges of random samples from the distribution will be evaluated by use of a I.B.M. computer for the sample size $n=4$, 8, 16, 24, 32, and the shape parameter $\gamma=2.5$, 3.0, 3.5, 4.0, and 4.5.

Throughout the table 1, the sample quasi-range of the following cases will be considered as good estimators in the sense of mean square error;

a) For a given shape parameter $\gamma=2.5$, the 2nd, 3rd, 4th, 5th, and 7th quasi-ranges are good estimators for the sample size $n=4$, 8, 16, 24, 32, respectively.

b) For a given shape parameter $\gamma=3.0$, the 2nd, 3rd, 4th, and 5th quasi-ranges are good estimators for the sample size $n=4$(and 8), 16, 24, 32, respectively.

c) For a given shape parameter $\gamma=3.5$, the 2nd, 3rd, and 4th quasi-ranges are good estimators for the sample size $n=4$(and 8), 16(and 24), 32, respectively.

d) For a given shape parameter $\gamma=4.0$, the range, 2nd, and 3rd quasi-ranges are good estimators for the sample size $n=4, 8$(and 16), 24(and 32), respectively.

e) For a given shape parameter $\gamma=4.5$, the range 2nd, and 3rd quasi-ranges are good estimators for the sample size $n=4, 8$(16 and 24), 32, respectively.

3. Mean square errors of the range and it's jackknife estimator.

It is well known that the range is widely used for an estimating problem because of it's computational simplicity. But, the table 1 shows that the range is worse than some quasi-ranges in the sense of mean square error.

It is natural that we compare the mean square error of the range with that of the jackknife estimator for the range of random samples from the distribution (1.1). As the mean square error of the range has been considered in section 1, now we consider only the mean square error of the jacknife estimator for the range.

The jackknife estimator for the range is defined by

$$J(W)=(2n-1)W/n-(n-1)W_2/n.$$

From the results (2.1) and (2.2), the mean squares error of the jackknife estimator for the range can be represented as follows;

$$\begin{aligned}MSE(J(W))=&[(2n-1)^2\{B_n^{-1}-A_n^2+B_1^{-1}-A_1^2-2A_n(A_1^{-1}B_1^{-1}\\&-A_1)\}/n^2+(n-1)^2\{B_{n-1}^{-1}-A_{n-1}^2+B_2^{-1}-A_2^2\\&-2A_{n-1}(A_2^{-1}B_2^{-1}-A_2)\}/n^2-2(n-1)(2n-1)\\&\{(A_2-A_{n-1})(A_1^{-1}B_1^{-1}-A_1)+A_n(A_{n-1}^{-1}B_{n-1}^{-1}\\&-A_2^{-1}B_2^{-1}-A_{n-1}+A_2\}/n^2+\{(2n-1)(A_n-A_1)/n\\&-(n-1)(A_{n-1}-A_2)/n-1\}^2]\beta^2.\end{aligned}$$

Therefore, we can obtain the exact numerical values of the mean square errors of the jackknife estimator for the range of the random samples from the distribution (1.1) for the shape parameter $\gamma=2.5$, 3.0, 3.5, 4.0, 4.5 and the sample size $n=4$. 8, 16, 24, 32 by use of the I.B.M. computer.

It will be shown in table 2 that the jackknife estimator for the range is worse than the range of the random samples from the distribution in the sense of the mean square error.

4. Efficiencies for estimators of the scale parameter.

Although the quasi-ranges are generally useful in the censored samples and have some robustness against outliers, their efficiencies are not very high in complete samples, but a suitable linear combination of the range and the quasi-ranges may provide an efficient estimator. A simple way doing this is to use the thickened range which is defined by

$$J_i=W+W_2+\cdots+W_i,\ \ i=1,2,\cdots,[(n+1)/2].$$

From the results (2.1) and (2.2), the mean square error of the thickened range of random samples from the distribution (1.1) can be represented as follows;

$$\begin{aligned}MSE(J_i)=\Big[&\sum_{j=1}^{i}(B_{n+1-j}^{-1}-A_{n+1-j}^{2}+B_j^{-1}-A_j^2)\\&+2\sum_{j<j'}^{i}\{A_{n+1-j}(A_{n+1-j'}^{-1}\ B_{n+1-j}^{-1}-A_{n+1-j'})\\&+A_{j'}(A_j^{-1}B_j^{-1}-A_j)\}-2\sum_{j=1}^{i}\sum_{j'=1}^{i}A_{n+1-j}(A_{j'}^{-1}B_{j'}^{-1}\\&-A_{j'})+\Big\{\sum_{j=1}^{i}(A_{n+1-j}-A_j)-1\Big\}^2\Big]\beta^2.\end{aligned}$$

The exact numerical values of the mean square error for the thickened range of random samples from the given distribution will be evaluated by use of the I.B.M. computer for the sample size $n=4,8,16,24,32$ and the shape parameter $\gamma=2.5, 3.0, 3.5, 4.0$, and 4.5.

It will be shown in table 3 that the thickened ranges of the following cases are considered as good estimators in the sense of the mean square error;

a) For a given shape parameter $\gamma=2.5$, the thickened ranges J_2, J_3, J_4, and J_5 are good estimators for the sample size $n=4$, 8, 16(and 24), 32, respectively.

b) For a given shape parameter $\gamma=3.0$, the thickened ranges J_2, J_3, and J_5 are good estimators for the sample size $n=4$, 8(16 and 32), 24 respectively.

c) For a given shape parameter $\gamma=3.5$, the thickened ranges J_2 and J_3 are good estimators for the sample size $n=4$(and 8), 16(24 and 32), respectively.

d) For a given shape parameter $\gamma=4.0$, the thickened ranges J_1, J_2 and J_3 are good estimators for the sample size $n=4$, 8, 16(24 and 32), respectively.

e) For a given shape parameter $\gamma=4.5$, the thickened ranges J_1 and J_3 are good estimators for the sample size $n=4$(and8), 16(24 and 32), respectively.

Next, the efficiency of a given estimator T, denoted by $Eff(T)$, will be defined as the ratio of the mean square error of the given estimator to the mean square error of the best linear unbiased estimator.

The best linear unbiased estimator of the scale parameter in the distribution (1.1) when the location and the shape parameters are known can be obtained by using the well-known Gauss-Markoff theorem [5]. Provied $\gamma>2$, the best liner unbiased estimator of the scale parameter becomes

$$\tilde{\beta}=((\gamma+1)\sum_{i=1}^{n-1} B_i X_{(i)}+(\gamma-1)B_n X_{(n)}-d\alpha)/(n\gamma-2-d),$$

where,

$$d=(\gamma+1)\sum_{i=1}^{n} B_i+(\gamma-1)B_n, \text{ and } B_i \text{ is given in (2.2).}$$

The variance of the best linear unbiased estimator of the scale parameter in the distribution (1.1) can be represented by

$$\mathrm{Var}(\tilde{\beta})=(n\gamma-2-d)^{-1}\beta^2. \tag{4.1}$$

Now we can define another unbiased estimator of the scale parameter in the distribution (1.1) as follows;

$$D=(\gamma-1)(2\gamma-1)\sum_{i=1}^{n} (2i-n-1)X_{(i)}/\{n(n-1)\gamma\}.$$

This estimator has been called Downton's unbiased estimator. It is well known that the Downton's unbiased estimator is highly efficient and not so influenced by the outliers as either the range or the root-mean square in the case of the normal population [1].

From (2.1) and (2.2), the variance of the Downton's unbiased estimator can be easily checked as follows;

$$\begin{aligned}\mathrm{Var}(D)=&\{(\gamma-1)(2\gamma-1)/(n(n-1)\gamma)\}^2\Big\{\sum_{i=1}^{n}(2i-n-1)^2\\ &(B_i^{-1}-A_i^2)+2\sum_{i<j}^{n}(2i-n-1)(2j-n-1)(A_i^{-1}B_i^{-1}\\ &-A_i)A_j\Big\}\beta^2.\end{aligned} \tag{4.2}$$

From (4.1) and (4.2), the exact numerical values for efficiencies of the Downton's unbiased estimator will be evaluated by use of the I. B. M. computer for the shape parameter $\gamma=2.5$, 3.0, 3.5, 4.0 and the sample size $n=4$, 8, 16, 24, 32.

Finally, we may consider the optimal quasi-range and the optimal thickened range, say W_0 and J_0, respectively. Here the optimal quasi-range means a quasi-range which has the least mean square error for a given shape parameter and a given sample size n. For an example, the optimal quasi-range for $\gamma=3.0$ and $n=16$ is the 3rd quasi-range. The optimal thickened range has the same meaning as the optimal quasi-range.

From table 1, table 3, and the result (4.1), the exact numerical values of the mean square error of the optimal quasi-range and the optimal thickened range can be easily obtained, and the efficiencies for those ranges can be evaluated for the shape parameter $\gamma=2.5$, 3.0, 3.5, 4.0, and the sample size $n=4$, 8, 16, 24 and 32.

It will be shown in table 4 that the optimal quasi-range and the optimal thickened range are more efficient than the Downton's unbiased estimator, but the optimal thickened range is not always more efficient than the optimal quasi-range for the shape $\gamma=2.5$, 3.0, 3.5, 4.0 and the sample size $n=4$, 8, 16, 24 and 32. These results for estimation of the scale parameter in the Pareto distribution are very dfferent from those of [8] for the scale parameter in the normal distribution.

TABLE 1. The relative MSE of the quasi-ranges of random samples from Pareto distribution.

n	i	$\gamma=2.5$	$\gamma=3.0$	$\gamma=3.5$	$\gamma=4.0$	$\gamma=4.5$
4	1	7.3253	2.0464	0.9278	0.5873	0.4761
	2	0.6437	0.6633	0.6993	0.7328	0.7612
8	1	14.4167	3.8115	1.4969	0.7446	0.4616
	2	0.5329	0.3291	0.3165	0.3524	0.3990
	3	0.4090	0.4732	0.5367	0.5903	0.6344
	4	0.7589	0.8024	0.8335	0.8564	0.8740
16	1	28.0723	7.3223	2.7849	1.2692	0.6617
	2	1.4822	0.5199	0.2494	0.1868	0.1955
	3	0.3234	0.1842	0.1976	0.2488	0.3062
	4	0.1930	0.2371	0.3094	0.3788	0.4392
	5	0.2805	0.3709	0.4502	0.5149	0.5673
	6	0.4423	0.5313	0.5987	0.6502	0.6909
	7	0.6419	0.7067	0.7525	0.7363	0.8122
	8	0.8722	0.8974	0.9144	0.9267	0.9359
	1	41.0899	10.6417	4.0373	1.8214	0.9172

	2	2.8176	0.9636	0.3836	0.1935	0.1447
	3	0.7191	0.2398	0.1342	0.1400	0.1809
	4	0.2414	0.1280	0.1523	0.2098	0.2717
	5	0.1350	0.1589	0.2288	0.3012	0.3662
	6	0.1576	0.2375	0.3220	0.3956	0.4573
24	7	0.2360	0.3356	0.4204	0.4890	0.5443
	8	0.3427	0.4431	0.5208	0.5809	0.6283
	9	0.4667	0.5566	0.6224	0.6719	0.7103
	10	0.6035	0.6752	0.7258	0.7631	0.7917
	11	0.7522	0.7995	0.8320	0.8586	0.8735
	12	0.9137	0.9309	0.9425	0.9508	0.9571
	1	62.3980	19.5109	9.3830	5.5845	3.7989
	2	4.2543	2.9473	2.1702	1.6408	1.2448
	3	1.2865	0.4082	0.1589	0.1056	0.1207
	4	0.4778	0.1504	0.0994	0.1285	0.1803
	5	0.1980	0.0982	0.1284	0.1894	0.2536
	6	0.1100	0.1197	0.1869	0.2601	0.3273
	7	0.1075	0.1729	0.2563	0.3328	0.3985
	8	0.1484	0.2413	0.3300	0.4050	0.4669
32	9	0.2135	0.3177	0.4054	0.4759	0.5327
	10	0.2834	0.3989	0.4814	0.5457	0.5965
	11	0.3832	0.4832	0.5578	0.6146	0.6587
	12	0.4806	0.5701	0.6347	0.6830	0.7204
	13	0.5844	0.6596	0.7125	0.7516	0.7813
	14	0.6944	0.7520	0.7917	0.8207	0.8427
	15	0.8103	0.8470	0.8729	0.8910	0.9046
	16	0.9349	0.9480	0.9568	0.9630	0.9677

TABLE 2. The relative MSE of the range and the jackknife estimator of the range of random samples from the Pareto distribution.

n	γ	Rel. MSE(W)	Rel. MSE(J(W))
	2.5	7.3255	22.9455
	3.0	2.0464	6.4813
4	3.5	0.9278	2.7353
	4.0	0.5878	1.4302
	4.5	0.4761	0.8813
	2.5	14.4167	48.3071
	3.0	3.8155	12.9759

8	3.5	1.4569	5.2211
	4.0	0.7446	2.5531
	4.5	0.4616	1.4222
	2.5	28.0723	94.2702
	3.0	7.3223	24.1714
16	3.5	2.7849	9.4404
	4.0	1.2692	4.4875
	4.5	0.6617	2.4032
	2.5	41.0899	136.5192
	3.0	10.6417	34.0003
24	3.5	4.0373	13.0634
	4.0	1.8214	6.1409
	4.5	0.9171	3.2521
	2.5	62.3980	175.6694
	3.0	19.5109	43.3130
32	3.5	9.3830	16.9484
	4.0	5.5845	8.3474
	4.5	3.7980	4.7830

TABLE 3. The relative MSE of the thicken ranges of random samples from the Pareto distribution.

n	i	$\gamma=2.5$	$\gamma=3.0$	$\gamma=3.5$	$\gamma=4.0$	$\gamma=4.5$
4	1	7.3253	2.0464	0.9278	0.5873	0.4761
	2	1.4627	0.9012	0.8742	0.9021	0.9231
8	1	14.4167	3.8115	1,4969	0.7446	0.3542
	2	2.0461	0.9088	0.4024	0.4021	0.4121
	3	0.2294	0.2271	0.4471	0.5502	0.4321
	4	0.8246	0.2907	0.5067	0.5972	0.6107
16	1	28.0723	7.3223	2.7849	1.2692	0.6617
	2	4.7211	1.4774	0.3806	0.3906	0.4824
	3	1.6022	0.2901	0.2771	0.2708	0.4226
	4	0.2112	0.3048	0.2940	0.2942	0.4302
	5	0.3272	0.3671	0.3144	0.3333	0.4771
	6	0.3842	0.3990	0.4506	0.3806	0.5024
	7	0.4192	0.4112	0.4607	0.4129	0.5764
	8	0.5207	0.4802	0.5598	0.7146	0.8002
	1	41.0899	10.6417	4.0374	1.8214	0.9171

	2	4.7122	2.8027	0.9108	0.1402	0.2047
	3	2.0018	1.0927	0.2084	0.1054	0.1154
	4	0.3120	0.2172	0.2701	0.1342	0.1630
24	5	0.3471	0.2109	0.2909	0.1667	0.2031
	6	0.3506	0.3274	0.3114	0.1907	0.2122
	7	0.4172	0.4006	0.3202	0.2204	0.2797
	8	0.4490	0.5471	0.3406	0.2906	0.3214
	1	62.3980	19.5109	9.3330	5.5845	3.7985
	2	7.0101	3.0572	1.9406	1.5024	1.1024
	3	3.9072	0.5104	0.4409	0.9027	0.2036
	4	0.4072	0.5248	0.4772	0.2671	0.3288
32	5	0.1142	0.5806	0.5021	0.2906	0.3655
	6	0.2284	0.6706	0.5353	0.3112	0.3910
	7	0.3762	0.6706	0.5807	0.3302	0.4114
	8	0.4027	0.7221	0.6001	0.4105	0.5067

TABLE 4. The efficiencies of the optimal quasi-range, the optimal thickened range and the Downton's unbiased estimator with respect to the B. L. U. E. of the scale parameter in the Pareto distribution.

n	Eff.	$\gamma=2.5$	$\gamma=3.0$	$\gamma=3.5$	$\gamma=4.0$	$\gamma=4.5$
	Wo	1.164	1.199	1.572	1.423	1.219
4	Jo	1.642	1.854	1.570	1.423	1.219
	D	4.721	7.810	6.884	6.485	6.230
	Wo	1.643	1.460	1.512	1.784	2.114
8	Jo	0.822	1.007	1.923	2.03	1.762
	D	5.182	5.972	4.979	4.483	4.251
	Wo	1.681	1.698	1.901	1.938	2.112
16	Jo	1.785	2.675	2.730	2.812	3.023
	D	7.249	5.244	4.216	3.670	3.367
	Wo	1.741	1.794	2.014	2.190	2.366
24	Jo	3.021	3.044	3.115	1.654	1.191
	D	5.522	5.278	4.109	3.505	3.184
	Wo	2.230	2.408	2.687	2.944	3.167
32	Jo	2.601	2.814	3.001	3.244	3.516
	D	5.877	5.230	4.466	3.724	3.333

Department of Statistcs,
Yeungnam University

REFERENCNS

[1] Barnett, F.C., Mullen, K, and Saw, J.G(1967), *Linear estimates of a population scale parameter.* Biometerika 54, pp. 551—54.

[2] D'Agostino, R.B. and Cureton, E.F. (1973), *A class of simple simple linear estimators of the standard deviation of the normal distribution,* J. Amer. Stat. Ass. 68, pp. 207—10.

[3] Gupta, R.P. and Singh, C. (1981), *Distribution of the ratio of two range of samples from nonnormal population.* Commun. Stat-Simula. Computa., B10(6), pp. 629—637.

[4] Kulldorff, G. and Vännman, K. (1973), *Estimation of the location and scale parameter of a Pareto distribution by linear function of order statistic.* J. Amer. Stat. Ass. 68. pp. 218—27.

[5] Kendall, M.G. and Stuart, A. (1961), *The Advanced Theory of Statistics,* Vol. 2, Charles Gritten & Co. Ltd.

[6] Lee, K.R. and Kapadia, C.H. (1982), *Statistical properties of quasi-range in sample from a gamma density.* Commun. Statist-Simula. Computa., pp. 175—95.

[7] Malik, H.J. (1980), *Exact formula for the cumulative distribution function of the quasi-range from the logistic distribution.* Commun. Stat.-Thear. A9(14), pp. 1527—34.

[8] Nair, K.R. (1967), *Efficiencies of certain linear systematic statistics for estimating dispersion from normal samples.* Biometrika 37, pp. 182—83.