Pakistan Journal of Applied Sciences 2(12): 1119-1122, 2002

Power Study for Empirical Distribution Function Tests for Generalized Pareto Distribution

[1]M. Arshad, [2]M.T.Rasool and [3]M.I. Ahmad
[1]University College of Administrative Sciences, Kotli, Azad Kashmir
[2]Government College, Samundri, Pakistan
[3]Department of Mathematics and Statistics, College of Science, Sultan Qaboos University Masqat, Oman

Abstract: For the power study of empirical distribution function tests, the Generalized Pareto distribution is considered. Four empirical distribution function tests i.e., Kolmogorov Smirnov, Cramer Von Mises, Anderson Darling and Modified Anderson Darling tests are compared for Generalized Pareto distribution. In the study four symmetrical and four skewed distributions are used as an alternative. The power comparisons are made using Monte Carlo methods at 5% and 10% significance levels for various sample sizes.

Key Words: Empirical Distribution Function, Generalized Pareto Distribution

Introduction

The Generalized Pareto (GP) distribution has applications in a number of fields including reliability studies, in the modeling of large insurance claims, as a failure time distribution (Hosking and Wallis, 1987). Pareto distribution is frequently used as a model in the study of income distributions (Aigner and Goldberger, 1970). The applications of GP-distribution include use in the analysis of extreme events (Hosking and Wallis, 1987). e.g., for the analysis of the precipitation data, in the flood frequency analysis, in the analysis of greatest wave heights or sea levels, maximum winds loads on building, in the maximum rainfall analysis, in the analysis of greatest values of yearly floods, breaking strength of materials etc. The GP distribution has been quite popular not only for flood frequency analysis but for fitting the distribution of extreme natural events in general.

Once a distribution function is assumed or selected for study at hand, it remains to estimate its parameters and to test the goodness of fit. There are so many methods of estimation but as a comparative study of the estimates of the GP distribution by Moharram *et al.*, (1993) and Hosking and Wallis (1987) that the Probability Weighted Moments (PWM) estimates will probably preferred because of their low bias, so we have used the PWM, a generalization of the usual moments of a probability distribution introduced by Greenwood *et al.*, (1979). The important goodness of fit techniques are, tests of Chi-squares types, moment ratio techniques, tests based on correlation and tests based on Empirical Distribution Function (EDF). Several power studies have revealed EDF tests to be more powerful than other tests of fit for a wide range of sample sizes (Stephens, 1974 and 1976). Until recently satisfactory use of EDF tests has been difficult due to lack of readily available tables of significance points for the case where the parameters of the assumed distribution have to be estimated from the sample data. Massy (1951) has suggested that if the test is used in this case, the results will be conservative in the sense that the probability of a type I error will be smaller than as given by the standard tables of the EDF statistics. Thus the use of these critical values which are for a specified parameters case to assess the agreement of a theoretical distribution when parameters are estimated from the data may result in accepting fitted distribution that ought to be rejected.

Arshad (1994) derived the tail area probabilities by simulation for the EDF tests for the GP distribution when the parameters are estimated by PWM. The EDF tests are Kolmogorov-Smirnov (KS) test, Cramer Van Mises (CVM) test, Anderson Darling (AD) test and Modified Anderson Darling (MAD) test. Stephens (1976) showed that among the EDF tests, the AD test is one of the most powerful, in contrast with the KS test whose power is relatively low. Hence the AD test should provide better discrimination between distributions. So it was necessary to perform the power study to evaluate the ability of the test to discriminate among competing distributional alternatives and to enhance our understanding of why the GP distribution often appears to provide such a good fit to observed maximum rainfall data. So an empirical power study of the assessment of the GP distribution is made to compare KS, CVM, AD and MAD test statistics.

Materials and Methods

As defined by Van-Montfort and Witter (1985) a random variable X is said to be distributed as Generalized Pareto (GP) distribution:

$$X(F) = b + a/c\,[1 - (1-F)^{c}] \qquad c \neq 0$$
$$= b - a\,\log(1-F) \qquad c = 0$$

where a = scale parameter, b = location parameter and c = shape parameter

The cumulative distribution function of GP distribution is given by:

$$F(x) = 1 - (1-cz)^{1/c} \qquad c \neq 0$$
$$= 1 - \exp(z) \qquad c = 0$$

where $z = (x-b)/a$

Method of Estimation: As a comparative study of the estimation of the GP distribution by Moharram *et al.*, (1993) and Hosking and Wallis (1987) that the PWM estimates will probably preferred because of their low bias, so we have used the PWM method.

Greenwood (1979) defined the probability weighted moments of X as:

$$M_{p,r,s} = E[X^{p}\{F(x)\}^{r}\{1-F(x)\}^{s}]$$
$$= \int X^{p}\,\{F(x)\}^{r}\,\{1-F(x)\}^{s}\,dF(x)$$

$$= \int_0^1 \{X(F)\}^p F^r \{1-F\}^s \, dF$$

where p, r, s are real numbers and X(F) is the quantile function of X.

Empirical Distribution Function (EDF): Suppose a given random sample of size n is $X_1, X_2, \ldots, X_n$ and let $X_{(1)}<X_{(2)}<\ldots<X_{(n)}$ be the order statistic and also suppose that the cumulative distribution function of X is F(x). The EDF is defined as:

$$F_n(X) = \frac{\text{No. of observations} \le x}{n}$$

More specifically the definition is

$$F_n(X) = 0 \qquad x < X_{(i)}$$
$$F_n(X) = i/n \qquad X_{(I)} \le x < X_{(i+1)} \qquad I = 1,2,\ldots,n-1$$
$$F_n(X) = 1 \qquad X_n \le x$$

The $F_n(x)$ is a step function which shows the proportion of observations less than or equal to x for any x. While F(x) is the probability of an observation less than or equal to x $[(F(x)=P(X\le x)]$. Empirical distribution function is calculated from the sample which estimates the population distribution function and it is infact a consistent estimator of F(x); as $n\to\infty$, $|F_n(x) - F(x)|$ decreases to zero with probability one.

EDF Statistics: A statistic measuring the difference between $F_n(x)$ and F(x) will be called an EDF statistic. There are so many EDF statistics which are based on vertical differences between $F_n(x)$ and F(x) and are conveniently divided into two classes, the supremum class and the quadratic class where $F_n(x)$ is the empirical distribution function.

i. The Spectrum Statistics: The first two EDF statistics D^+ and D^- are respectively, the largest vertical difference when $F_n(x)$ is greater than F(x) and the largest vertical difference when $F_n(x)$ is smaller than F(x); formally

$$D^+ = \sup\{F_n(x) - F(x)\}$$
$$D^- = \sup\{F(x) - F_n(x)\}$$

and the most well known EDF statistic is D introduced by Kolmogorov Smirnov (1933):

$$D = \sup|F_n(x) - F(x)| = \max(D^+, D^-)$$

From the basic definition of the supremum statistics given above, suitable computing formulas must be found. This is done by using the probability integral transformation Z=F(x); when F(x) is the true distribution of X; the new random variable Z is uniformly distributed between 0 and 1 (D'Agostino and Stephens, 1986). Then Z has distribution function

$$F^*(z) = Z \quad 0 \le x \le 1$$

Suppose that a sample $X_1, X_2, \ldots, X_n$ gives values $Z_i=F(X_i)$; i=1,2,...,n and let $F_n^*(z)$ be the EDF of the values Z_i.

EDF statistics can now be calculated from a comparison of $F_n^*(z)$ with the uniform distribution for z. It is easily shown that, for values z and x related by z=F(x), the corresponding vertical differences in the EDF diagrams for X and for Z are equal. i.e.,

$$F_n(x) - F(x) = F_n^*(z) - F^*(z)$$

Consequently EDF statistics calculated from the EDF of the Z_i compared with the uniform distribution will take the same values as if they were calculated from the EDF of the X_i, compared with F(x). This leads to the following formulae for calculating EDF statistics from the z values.

The formulas involve the z values arranged in ascending order, $z_{(1)}<z_{(2)}<\ldots<z_{(n)}$. Then

$$D^+ = \max(i/n - z_{(i)})$$

and $D^- = \max(z_{(i)} - (i-1)/n)$

then $D = \max(D^+, D^-)$

ii. The Quadratic Statistic: Anderson and Darling (1954) proposed a second and wide class of measures of di screpancy which is given as:

$$W^2_n = n \int_{-\infty}^{\infty} \{Fn(x) - F(x)\}^2 \, \psi(x) \, dF(x)$$

where $\psi(x)$ (≥ 0) is a preassigned weight function. When $\psi(x)=1$, W^2n reduces to nW^2 where W^2 is the Cramer-Von Mises statistic and when $\psi(x) = [\{F(X)\{1-F(X)\}]^{-1}$ the statistic is the Anderson Darling, called A^2_n.

The Anderson Darling test is one of the more powerful empirical distribution function based tests of fit in a wide range of circumstances (Stephen, 1974 and 1976). However, it gives equal weight to differences between empirical and theoretical distribution functions corresponding to all the observations. When the objective of the fitting process is to predict a quantile at the top end of the distribution, a goodness of fit statistic that gives more weight to the larger values would be more helpful. Since design engineers are primarily concerned with estimates of flood magnitudes with return periods higher than 50 years. In such cases the interest is concentrated on the upper tail of the distribution. So the Sinclair *et al.,* (1990) proposed a modified form of the Anderson Darling test (AU^2_n) using the weight function $\psi(x) = [1-F(x)]^{-1}$. For computational purposes the Cramer Von Mises (W^2), Anderson Darling (A^2_n) and Modified Anderson Darling (AU^2_n) statistics are:

$$W^2 = \Sigma\{Z_{(i)} - (2i-1)/(2n)\}^2 + \frac{1}{12n}$$

$$A^2_n = -n - \frac{1}{n} \Sigma(2i-1) \{\log Z_{(i)} + \log(1-Z_{(n+1-i)})\}$$

and

$$AU^2_n = \frac{n}{2} - 2\sum_{i=1}^{n} Z_{(i)} - \sum_{i=1}^{n} [2 - \frac{(2i-1)}{n}] \log[1 - Z_{(i)}]$$

Power Studies: The power of the test is the frequency with which the null hypothesis is actually rejected at the given significance level when samples are drawn from the specified alternative distribution (Chowdhury *et al.,* 1991). In order to investigate the behaviour of EDF tests of fit for the GP distribution we have carried out computations to determine the power for several alternative families of distributions. Chowdhury *et al.,* (1991) performed power studies to evaluate probability plot correlation tests for the GEV (Generalized Extreme Value) hypothesis and revealed that the results of power study, by two parameter non-GEV alternatives and the GEV family alternatives are similar. The power of the test depends upon the shape of the alternative distributions. So for simplicity, the two parameter distributions are used, as an alterative.

We have compared the power of the four EDF statistics using eight alternative distributions. These power comparisons are made using Monte Carlo simulation. We generated 1000 pseudo-random samples of size n from each of the eight alternative distributions considered. Since all the alternatives considered involve unknown location and scale parameters. As described by Dahiya and Gurland (1973), the power will be the same for all the possible values of the location and scale parameters. So we have generated the samples by using the appropriate values of the parameters. We then calculated each of the four EDF statistics and compared them to their respective critical values and counted the number of rejections of the null hypothesis. The computer recorded the number of samples out of 1000 for which the test criterion lay beyond the 5% level and 10% level appropriate for sampling from a GP distribution. These counts out of 1000 samples, when divided by ten, give estimates of the power of the respective tests. The power calculations have been carried out for sample size n = 10, 25 and 40.

In computing the power, we formulate the null hypothesis H_o that the random variable X has commutative distribution function

$$H_o: F(X/a,b,c) = 1 - (1-cZ)^{1/c} \qquad c \neq 0$$
$$= 1 - \exp(Z) \qquad c = 0$$

where $Z = \dfrac{x-b}{a}$

For the power studies the following eight alternative distributions are considered in which the four distributions are symmetrical and the other four are skewed and the random samples are generated by Minitab (1985).

The symmetrical distributions are Logistic, Normal, Cauchy and Beta. These all distributions are defined on the real line. While the GP distribution is only defined on the positive half of the real line. Hence we computed Y_i where $Y_i=e^{X_i}$ as recommended by Little *et al.*, (1979) and tested for goodness of fit of these Y_i to the GP distribution. The skewed distributions are Exponential, Gamma, Gumbel and Weibull. These distributions are defined on the positive half of the real line and therefore are directly comparable to the GP distribution.

Results and Discussion

We compared the power of the Kolmogorov Smirnov (D), Cramer Von Mises (W^2), Anderson Darling (A^2_n) and Modified Anderson Darling (AU^2_n) test statistics for the Generalized Pareto distribution, using the eight alternatives i.e., logistic, normal, Cauchy, Beta, Exponential, Gamma, Gumble and Weibull at 5% and 10% levels of significance, taking the sample size, n=10, 25 and 40 and the random samples are generated by Minitab (1985). The results of the power study are presented in Table 1 for the symmetrical alternatives and in Table 2 for the skewed alternatives. It is evident from Table 1 and 2 that the power of all EDF tests is rather high against the Cauchy alternative. As a matter of fact, it is the highest among all the alternatives considered here. It is also evident that the power increases as the sample size n increases. In contrast to the other alternatives, the power pertaining to Gamma alternative decreases as the sample size increases especially for the Kolmogorov Smirnov and the Cramer Von Mises statistics. Moreover, the power for the Gamma alternative is substantially lower than the all other alternatives. This could be expected, since the Gamma is skewed distribution which resembles the GP distribution much more than the other symmetric distributions. By comparing Table 1 and 2 it is seen that the EDF statistics are more powerful for the symmetrical alternatives than the skewed alternatives. This is not surprising, however, it is known that the Generalized Pareto distribution is also skewed distribution.

A glance over Table 1 reveals that Anderson Darling test is the most powerful for all the symmetric alternatives and all the sample sizes under consideration. The Modified Anderson Darling test appears to be less powerful as compared to Anderson Darling test but better than the other tests. Kolmogorov Smirnov test seems to be least sensitive to the symmetrical alternatives. In scanning Table 2, it is noted that the Anderson Darling and Modified Anderson Darling tests tend to be more powerful than Kolmogorov Smirnov test and Cramer Von Mises test slightly better than Kolmogorov Smirnov test. In practice, it would always seem worth while to look at Cramer Von Mises, Anderson Darling and Modified Anderson Darling.

Table 1: Symmetrical Populations: Estimates of Power (in Percent) of EDF Tests at 5% and 10% Significance Levels

Alternative Distribution	n	D		W^2		A^2_n		AU^2_n	
		0.10	0.05	0.10	0.05	0.10	0.05	0.10	0.05
Logistic	10	26	21	27	20	56	30	30	16
	25	35	29	39	28	74	45	39	25
	40	40	33	44	34	76	54	41	30
Normal	10	22	14	20	8	32	14	17	5
	25	26	16	27	13	50	24	18	5
	40	30	22	31	14	69	31	19	6
Cauchy	10	55	50	56	50	80	58	56	40
	25	80	74	80	74	92	82	74	70
	40	81	80	88	82	97	88	83	81
Beta	10	10	6	10	5	33	6	19	16
	25	21	13	18	10	34	11	37	32
	40	22	16	19	11	36	13	39	35

Table 2: Skewed Populations: Estimates of Power (in Percent) of EDF Tests at 5% and 10% Significance Levels

Alternative Distribution	n	D		W^2		A^2_n		AU^2_n	
		0.10	0.05	0.10	0.05	0.10	0.05	0.10	0.05
Exponential	10	16	10	19	14	40	18	25	13
	25	19	11	28	14	47	14	26	13
	40	22	13	30	15	47	16	34	15
Gamma	10	12	7	13	8	30	10	19	6
	25	18	8	12	7	33	11	35	16
	40	10	6	8	10	35	13	38	20
Gumbal (EV1)	10	19	13	19	15	28	22	32	18
	25	58	42	58	33	65	52	60	35
	40	67	62	70	60	80	73	70	45
Weibull	10	16	10	15	8	35	21	27	16
	25	18	14	18	11	47	39	32	20
	40	35	21	29	15	49	32	39	27

Historically, Kolmogorov Smirnov has been the most commonly used EDF statistics, for its simplicity, it tends to be the least powerful among the four EDF tests considered here. Anderson Darling and Modified Anderson Darling appears to be the best pair of the EDF statistics. A very little difference is found between the power of Cramer Von Mises and Modified Anderson Darling in overall power. These findings are similar to those found by Little *et al.,* (1979) for the two parameter Weibull distribution. It is found that the powers of EDF tests are broadly in the following order of descending power.

$$A^2_n \rightarrow AU^2_n \rightarrow W^2 \rightarrow D$$

So with no prior knowledge of the alternative distribution ,it may be advisable to use the Anderson Darling, Modified Anderson Darling and Cramer Von Mises test statistics for testing the Generalized Pareto distribution, since their power was reasonably good for all alternatives.

Conclusion

In order to investigate the behaviour of EDF tests of fit for Generalized Pareto distribution, a power study is made by using eight alternative families of distributions. The power comparisons are made using Monte Carlo methods, at 5% and 10% significance levels, for sample sizes n = 10, 25 and 40. The Anderson Darling and Modified Anderson Darling appears to be the best pair of EDF test statistics. Kolmogorov Smirnov tends to be the least powerful among the four EDF tests considered here. It is concluded that the powers of EDF tests are broadly in the following order of decreasing power.

$$A^2_n \rightarrow AU^2_n \rightarrow W^2 \rightarrow D$$

So it may be advisable to use the Anderson Darling, Modified Anderson Darling and Cramer Von Mises test statistics for testing the Generalized Pareto distribution.

References

Adnerson, T. W. and D.A. Darling, 1954. A test of goodness of fit. J. Am. Stat. Assoc. 49: 765-769.

Aigner, D. J. and A.S. Goldberger, 1970. Estimation of Pareto's law from grouped observations. J. Amer. Stat. Assoc., June: 713-727.

Arshad, M., 1994. Empriical distribution function test for Generalized Pareto Distribution and its applications in rainfall intensity estimation. M.Phil. thesis. Dept. of Math. and Stat. Univ. of Agri. Faisalabad, Pak.

Chowdhury, J. U., J. R. Stedinger and L. H. Lu, 1991. Goodness of fit tests for regional Generalized Extreme value flood distributions. Water Resour. Res., 27: 1765-1776.

D'Agostino, R. B. and M.A. Stephens, 1986. Goodness of fit techniques. Marcel Dekker, New York.

Dahiya, R.C. and J. Gurland, 1973. How many classes in the Pearson Chi-square test? Theory and Method, 68: 707-712.

Greenwood, J. A., J. M. Landwehr and W.C. Matalas, 1979. Probability weighted moments. Definition and relation to parameters of distributions expressible in inverse form. Water Resour. Res., 15: 1049-1054.

Hosking, J. R. M. and J. R. Wallis, 1987. Parameter and quantile estimators for Generalized Pareto distribution. Technometrics, 29: 339-349.

Kolmogorov, A. N., 1933. Sulla determinzione empirica di una legge de distribuzione Giorn. 1st. Ital. Attvari, 83-91.

Littel, R. G., J.T. McClave and W.W. Offen, 1979. Goodness fit tests for the two parameter Weibull distribution. Common. Statist. Simul. Computa., 138: 257-269.

Massey, F. J., 1951. The Kolmogorov-Smirnov test for goodness of fit. J. of the Amer. Stat. Assoc. 46: 68-78.

Minitab, 1985. Minitab Inc., State College PA 16801, USA.

Moharram, S. H., A. K. Gosain and P.N. Kapoor, 1993. A comparative study for the estimators of the Generalized Pareto distribution. J. Hydrolog, 150: 169-185.

Sinclair, C. D., B. D. Spurr and M. I. Ahmad, 1990. Modified Anderson Darling test; Commun. Statist. Theory and Math. 19: 3677-3686.

Stephens, M. A., 1974. EDF statistics for goodness of fit and some comparisons. J. Amer. Stat. Assoc. 69: 730-737.

Stephens, M. A., 1976. Asymptotic power of EDF statistic for exponentiality against Gamma and Weibull alternatives. Tech. Report No.297, Dept. of Stat. Stanford Univ.

Van Montfort, M. A. J. and J. V. Witter, 1985. Testing exponentiality against Generalized Pareto distribution. J. Hydrol., 78: 305-315.