8-1976

Statistical Models for Precipitation

Z. Govindarajulu
University of Kentucky

Peter Purdue
University of Kentucky

Frank Cioch
University of Kentucky

Research Report No. 96

STATISTICAL MODELS FOR PRECIPITATION

by

Z. Govindarajulu

Peter Purdue

Principal Investigators

and

Frank Cioch

Graduate Student Assistant

Project Number: A-055-KY (Completion Report)

Agreement Numbers: 14-31-0001-5017 (FY 1975)
14-34-0001-6018 (FY 1976)

Period of Project: July 1974 - June 1976

University of Kentucky
Water Resources Research Institute
Lexington, Kentucky

The work on which this report is based was supported in part by funds provided by the Office of Water Research and Technology, United States Department of the Interior, as authorized under the Water Resources Research Act of 1964.

August, 1976

PREFACE

This report is Part I of the technical completion report for the research project entitled "Stochastic and Statistical Models for Precipitation". The project was supported by funds provided by the United States Department of Interior to the University of Kentucky Water Resources Institute as authorized by the Water Resources Act of 1964, Public Law 88-379, as Office of Water Resources Research Project No. A-055-KY.

ACKNOWLEDGEMENTS

Our sincere thanks are extended to Dr. Robert Grieves, Director of the Kentucky Water Resources Institute for supporting this project and Mr. Ralph Huffsey, Assistant Director of the Kentucky Water Resources Institute for his great help and cooperation. We also thank Mr. Doyle Cook and Mr. Jerry Hill of the National Weather Service for several discussions we had with them and for their generous cooperation throughout this investigation. Appreciation is expressed to Mr. Douglas Griffin of the Kentucky Department of Natural Resources and Environmental Protection for the use of rainfall data tapes. The University of Kentucky Computer Center was extremely helpful and cooperative throughout this investigation.

0. ABSTRACT

The available data consists of daily rainfall for the past 24 years (1948-1972) for Lexington, Louisville and Paducah. However, for Ashland, the data is available for the period of 40 years (1932-1972). The problem is to find an appropriate family of distributions indexed by a suitable number of parameters that fits the maximum daily rainfall. Further, there might be seasonal variations. The following seasons were considered: (i) Dry Convective season: August 1-October 30; (ii) Early Convective season: May 1-July 31; (iii) Late Convective season: November 1-April 30. After extracting the yearly maximums for each of these seasons (see Appendix 1), we set about fitting three distributions for each station; one for each season.

Our first temptation was to fit Gumbel's type I extreme value distributions differing in location and scale parameters, to the maximum daily rainfall. However, the type I model for Louisville was rejected. Next, we considered the Cauchy type of extreme value distribution given by

$$F(x;\delta,\beta) = \exp[-(\frac{x-\theta}{\delta})^{-\beta}], \qquad \theta < x < \infty. \tag{1}$$

where θ, δ and β are unknown parameters. One can easily verify that the distribution given by (1) is unimodal. The goodness of fit procedure of Mann, Scheuer and Fertig (1973) (MSF) can be carried out either when θ and δ are unknwon and β is known, or when θ is known and δ and β are unknown. For the rainfall data, it is not unreasonable to set θ=0. Using this assumption, we found that the

distribution of the form (1) fits all the stations.

Next we were concerned with estimation of the parameters δ and β. These were estimated via the maximum likelihood method and the variances of these estimates were evaluated using Fisher's information matrix. The variances of δ's range from 1% to 2% whereas the variances of β's range from 5% to 10%. Estimators based on least squares method came very close to the maximum likelihood estimators. We also have developed certain test procedures for the parameters; we infer that a separate model be fitted for each of the stations, namely Lexington, Louisville, Ashland and Paducah. We also feel that a separate model be used for each season. It seems reasonable to combine into a single season the Early and Late seasons for Lexington, the Dry and Early seasons for Louisville and Paducah, and all the seasons into a single season for Ashland. When once a model is fitted to the maximum rainfall data, one can answer several questions like, what are the chances of maximum daily rainfall at a particular station exceeding a certain specified amount.

DESCRIPTORS

Maximum rainfall.

IDENTIFIERS

Extreme value distributions, goodness of fit tests.

TABLE OF CONTENTS

1. INTRODUCTION

The study of extreme rainfall is of much interest, especially to the National Weather Service (NWS). There are long term prediction problems which would be of use in such design consideration as the drainage system for shopping-area parking lots; in the short term case an ability to predict when flood conditions may occur in some community would help local emergency services. Daily rainfall data for the last several years is available for several stations in Kentucky. It is of interest and importance to subject the available daily rainfall data to statistical analysis so as to provide answers to some typical questions of the form:
(i) what extreme rainfall can be predicted at certain station with a specified degree of confidence?
(ii) Do the extreme rainfalls in Stations A and B differ significantly?

Let $Y_j^{(i)}$ denote the amount of rainfall on jth day of the ith year for a certain station ($j=1,\ldots,365$, $i=1,\ldots,n$, n denoting the number of years for which the recorded rainfall data is available for the station under consideration). Also, let $X_i=\max(Y_1^{(i)},\ldots,Y_{365}^{(i)})$, $(i=1,\ldots,n)$. That is, X_i denotes the extreme daily rainfall during the ith year ($i=1,\ldots,n$). We can reasonably assume that the distribution of X_i is free of i (that is $P(X_i\leq x)=F(x)$. It is well known that if the $Y_1^{(i)},\ldots,Y_{365}^{(i)}$ are independent, the approximate distribution of X_i (since 365 is sufficiently large) follows one of the forms of the extreme value distribution (see Gumbel (1958)). David (1970, p. 214) remarks that the limiting distribution of the maximum of a certain dependent (stationary m-dependent) variables, under certain regularity conditions is the same as for the independent variables (see, for instance, Watson (1954)). Hence, it seems reasonable to assume that the X_i follow an extreme value distribution. The typical forms for the extreme value distribution are

(i) Type I: $F(x) = \exp\{-\exp(-(x-\theta)/\sigma)\}, \quad -\infty<x<\infty,$

(ii) Cauchy type: $F(x) = \exp[-(\frac{x-\theta}{\sigma})^{-\beta}], \quad \theta<x<\infty,$

where θ, σ and β are unknown parameters to be determined. In case (ii) we can reasonably assume, (at least for the rainfall data) that $\theta=0$. First we determine the adequacy of the model given by (i) or (ii) and then estimate the parameters of the model. Further, in order to reflect the seasonal variations in extreme rainfall, we propose to fit different distributions to data pertaining to (a) Dry Convective season: August 1-October 30; (b) Early Convective season: May 1-July 31; and (c) Late Convective season: November 1-April 30.

We take a sample of four stations in Kentucky, namely, Ashland, Lexington, Louisville and Paducah and analyse the maximum daily rainfall at these stations.

2. Goodness of Fit Tests.

The daily rainfall data for Ashland, Lexington, Louisville and Paducah has been obtained from the tapes made by the Division of Water, Kentucky Department of Natural Resources. For Ashland, the daily rainfall data is available for a period of 40 years and for the other stations, it is available for 24 years. Maximum daily rainfall data has been extracted from the daily rainfall data for the various seasons. (See Appendix 1 for the data). We first thought of fitting Gumbel's (1958) Type I distibution to the maximum daily rainfall data. If X_i denotes the maximum daily rainfall for ith year at a particular station, then the Type I extreme value distribution is given by

$$P(X_i<x) = \exp\{-(-x-\theta)/\sigma)\}, \quad -\infty<x<\infty, \tag{2.1}$$

where θ is the location parameter and σ is the scale parameter.

First we obtained the least squares estimates of θ and σ based on ordered observations drawn from the Type I distribution (see Appendix 5) for the (i) yearly maximum daily rainfall (ii) maximum daily rainfall for the Late Convective Season covering the period: November 1 to April 30, (iii) maximum daily rainfall for the Early Convective Season covering the period: May 1 to July 31, and (iv) maximum daily rainfall for the Dry Convective Season covering the period: August 1 to October 30. For Ashland, we needed to divide the data into two groups since the variances and the covariances of order statistics are not available for sample sizes up to 40. The following numerical results are obtained.

Table 2.1

Least Squares estimates of Location and Scale Parameters of the Type I Extreme Value Distribution,

		Lexington	Louisville	Paducah	Ashland (a)	(b)
Yearly	$\hat{\theta}$	2.41	2.44	2.67	2.05	2.09
	$\hat{\sigma}$	.48	.68	.77	.47	.46
Late	$\hat{\theta}$	1.79	1.86	2.21	1.48	1.41
	$\hat{\sigma}$	.50	.60	.78	.46	.46
Early	$\hat{\theta}$	1.76	1.53	1.89	1.36	1.45
	$\hat{\sigma}$	.57	.69	.65	.49	.75
Dry	$\hat{\theta}$	1.52	1.51	1.68	1.37	1.36
	$\hat{\sigma}$	.64	.68	.57	.58	.43

The variance-covariance matrix of $\hat{\theta}$ and $\hat{\sigma}$ for Lexington, Louisville and Paducah is $\begin{matrix}\hat{\theta}\\ \hat{\sigma}\end{matrix}\begin{pmatrix} .047 & .010 \\ .010 & .027 \end{pmatrix}$ and for Ashland it is $\begin{matrix}\hat{\theta}\\ \hat{\sigma}\end{matrix}\begin{pmatrix} .056 & .012 \\ .012 & .033 \end{pmatrix}$.

After standardizing the data by means of $\hat{\theta}$ and $\hat{\sigma}$ for the yearly maximum daily rainfall at each of the four stations we applied the one-sample Kolmogorov-Smirnov goodness of fit test (see Appendix 6) and we find the Type I distribution was accepted in each of the four cases. Since the parameters are estimated from the data, the inferences based on the test are of doubtful nature. Hence, the invariant goodness of fit test of Mann-Scheuer and Fertig (MSF) (1973) (see Appendix 6) was carried out and to our surprise the test rejected the Type I distribution for the Louisville station.

Next, we consider fitting the Cauchy type of extreme value distribution given by

$$P(Y \leq y) = \exp[-(\frac{y-\theta}{\delta})^{-\beta}], \quad y \geq \theta. \tag{2.2}$$

If we let $\theta=0$, we see that extreme value distribution is obtained from the Cauchy type by the change of variable $X=\ln Y$. So if Y has the Cauchy type distribution with δ for scale and β for the shape parameter then X has the extreme value distribution with location $\ln\delta$ and scale parameter β^{-1} (see appendix 2). So, after taking logarithms of the maximum daily rainfall, Mann-Scheuer and Fertig (MSF) test was carried out and the Cauchy type distribution was accepted for all stations. Notice that we need not know δ and β in order to carry out the MSF goodness of fit test, since it is invariant under location shifts and scale changes.

We tried to fit the Weibull distribution given by

$F(x) = 1-e^{-(x/\delta)^{\beta}}$, $x \geq 0$ and the MSF test rejects this distribution for all stations except Lexington. Notice that the Weibull distribution becomes the Type I extreme value distribution by the transformation $x=-\ln Y$. The following numerical values for the test criterion L (see Appendix 6 (ii) for the definition of L) are obtained.

TABLE 2.2

Giving the values of the MSF test criterion.

	Lexington	Louisville	Paducah	Ashland
Type I Extreme Value	.45	3.19*	1.28	1.06
Weibull	1.01	7.00*	2.76*	2.34*
Cauchy Type	.33	1.89	.81	0.73

The critical values at the 5% level of significance for the F distribution are $F_{.95}$ (40, 40) = 1.69, $F_{.95}$ (22, 24) = 2.03. The values of the test criterion bearing an asterisk lead to rejection of the hypothesis.

3. Estimation of parameters of the Cauchy type of extreme value distribution

Since the type I extreme value distribution fits the data which is log of the maximum daily rainfall, one can obtain least squares estimates of θ and σ. Then one can obtain the estimates of δ and β by

$$\hat{\delta} = \exp(\hat{\theta}) \text{ and } \hat{\beta} = 1/\hat{\sigma}. \tag{2.3}$$

However, $\hat{\delta}$ and $\hat{\beta}$ will neither be unbiased, nor linear in the log data, nor will have the smallest variances. One can also obtain the maximum likelihood estimates (MLE's) of δ and β from the original data (ie. without taking logs). (For the MLE procedure see Appendix 3). These results are presented in Table 3.1.

TABLE 3.1

Estimates of δ and β, the parameters of the Cauchy type of extreme value distribution.

		Lexington		Louisville		Paducah		Ashland (a)		Ashland (b)	
		δ	β	δ	β	δ	β	δ	β	δ	β
Based on (2.3)	Yearly	2.36	2.44	2.34	4.17	2.56	3.45	1.95	4.35	2.05	4.55
	Late	1.73	3.57	1.77	3.45	2.08	2.86	1.46	3.57	1.35	2.86
	Early	1.68	3.13	1.39	2.22	1.79	2.86	1.28	3.03	1.28	1.92
	Dry	1.40	2.44	1.35	1.89	1.60	3.13	1.27	2.56	1.30	3.13
								δ		β	
MLE	Yearly	2.38	5.25	2.36	4.30	2.57	3.40	2.01		4.55	
	Late	1.41	2.55	1.32	1.67	1.61	3.31	1.38		3.11	
	Early	1.73	3.62	1.39	2.17	1.79	2.95	1.29		2.34	
	Dry	1.69	3.31	1.78	3.49	2.09	2.94	1.29		2.89	

As one can see, the estimates of δ and β by both the methods are fairly close. In order to get a better idea as to how reliable these estimates are, one should obtain the variances and covariances of the $\hat{\delta}$'s and $\hat{\beta}$'s that are based on the "information matrix" and are given by equations (13)-(15) of Appendix 3.

TABLE 3.2

Giving the asymptotic variances and covariances of $\hat{\delta}$'s and $\hat{\beta}$'s for the yearly maximum daily rainfall data.

	Lexington	Louisville	Paducah	Ashland
$\hat{\beta}$	5.25	4.30	3.40	4.55
$\hat{\delta}$	2.38	2.36	2.57	2.01
Var $\hat{\beta}$	.52	.35	.22	.23
Var $\hat{\delta}$	.009	.014	.026	.005
Cov$(\hat{\beta},\hat{\delta})$	-.019	-.019	-.020	-.010

Notice that

$$\left.\begin{aligned} \text{Estimate of var } \hat{\beta} &= 0.45\ \hat{\beta}^2/n \\ \text{Estimate of var } \hat{\delta} &= 1.08\ \hat{\delta}^2/n\hat{\beta}^2 \\ \text{Estimate of cov}(\hat{\beta},\hat{\delta}) &= -0.19\ \hat{\delta}/n. \end{aligned}\right\} \qquad (2.4)$$

From Table 3.2 we surmise that the $\hat{\delta}$'s are stable since variances of the $\hat{\delta}$'s are very small. The variances of $\hat{\beta}$'s are also fairly small compared to $\hat{\beta}$'s themselves. Hence, the MLE's are fairly stable.

A Justification for Taking θ=0

Consider the three-parameter density given by

$$f_X(x;\theta,\delta,\beta) = (\beta/\delta)\left(\frac{x-\theta}{\delta}\right)^{-\beta-1} e^{-\{(x-\theta)/\delta\}^{-\beta}}, \quad x>\theta.$$

Let $Y=(x-\theta)/\delta$. Then the density of Y is given by

$$f_Y(y;\beta) = \beta\, y^{-\beta-1}\, e^{-y^{-\beta}}, \quad y>0.$$

Consider

$$EY = \beta \int_0^\infty y^{-\beta} e^{-y^{-\beta}} dy = \int_0^\infty u^{-(1/\beta)} e^{-u} du = \Gamma(1- \frac{1}{\beta}).$$

That is

$$(EX-\theta)/\delta = \Gamma(1- \frac{1}{\beta}).$$

Since $\bar{X}$ is the best estimate of EX, an estimate of θ is

$$\hat{\theta} = \bar{X}-\hat{\delta}\ \Gamma(1- \frac{1}{\hat{\beta}}).$$

We shall compute $\hat{\theta}$ for the yearly maximum daily rainfall data for the four stations.

TABLE 3.3

Giving values of $\hat{\theta}$ based on the yearly maximum daily rainfall data.

	Lexington	Louisville	Paducah	Ashland
$\hat{\beta}$	5.00	4.17	3.45	4.55
$\hat{\delta}$	2.36	2.34	2.56	2.01
$\bar{X}$	2.67	2.90	3.10	2.33
$\hat{\theta}$	-.07	-.07	-.18	-.07

As one can see, the $\hat{\theta}$'s are fairly negligible. However, it should be pointed out that $\hat{\theta}$ is the estimate based on the first moment and $\hat{\theta}$ is not based on the MLE procedure. However, if one were to solve the likelihood equations given in Appendix 4, he would get different estimates for θ, β and δ than given in Table 3.3.

4. Testing Hypotheses about Parameters.

Since we have shown that Cauchy type of extreme value distribution fits the data for all stations, we will be interested in knowing whether, a common distribution can be fitted for all the seasons, or a certain distribution will do the job for more than one station. We shall do this in two stages. First, we shall focus our attention on combining certain seasons for each station. We shall guess at a certain common β value based on MLE's (see Table 3.1) and by setting up confidence intervals for δ as indicated in Appendix 7. We infer whether a common distribution will adequately serve for two or more seasons.

Lexington.

TABLE 4.1

	Common guessed value of β	95% confidence interval for δ
Late	3.5	(1.47, 1.86)
Early	3.5	(1.53, 1.93)
Dry	3.5	(1.17, 1.47)
Late	3.25	(1.48, 1.89)
Early	3.25	(1.53, 1.97)
Dry	3.25	(1.17, 1.51)

In both the cases, (ie differing in β values) the confidence limits for δ for the Dry season differs from those for Late and Early seasons.

Hence we surmise that the distributions for Late and Early seasons can be combined, whereas the dry seasons needs a separate model since its β is also different from the β's for the Late and Early seasons. (See Table 3.1).

Louisville. TABLE 4.2

	Common β (guessed)	95% confidence interval for δ
Late	1.75	(1.47, 2.34)
Early	1.75	(1.13, 1.79)
Dry	1.75	(1.01, 1.60)
Late	2.00	(1.51, 2.25)
Early	2.00	(1.13, 1.70)
Dry	2.00	(0.98, 1.47)

β = 1.75 or 2.00 seems to be most reasonable. Early and Dry seasons can be combined. Model for Late season should be different (since $\hat{\beta}$ for Late season is 3.49).

Paducah. TABLE 4.3

	Common β	95% confidence interval for δ	Remark
Late	3.00	(1.80, 2.35)	Does not include $\hat{\delta}$ for Dry or Early
Early	3.00	(1.54, 2.01)	
Dry	3.00	(1.40, 1.84)	

Conclusion. Models for Dry and Early seasons are the same, however, the model for the Late season is different.

Ashland. TABLE 4.4

	Common β	95% confidence interval for δ
Late	2.75	(1.24, 1.56)
Early	2.75	(1.10, 1.38)
Dry	2.75	(1.15, 1.44)

Conclusion. One model can be fitted to all the three seasons.

Tests for β.

From Table 3.1, we feel that the δ's are essentially the same for certain seasons within each station. Hence, we will be tempted to use a single model for one or more seasons for a particular station. Before we do this we should test for the equality of the corresponding β's, which can be done by the procedures outlined in Appendix 8.

First, let us consider only the seasons in all four stations which has δ's different from other seasons. These are the Dry season for Lexington, the Late season for Louisville and Paducah and the Early season for Ashland. The following results are obtained.

TABLE 4.5

Giving the results of the tests for β for certain seasons where H_0: $\beta = \beta_0$ and H_1: $\beta \neq \beta_0$

Station	Season	δ	β_0	Degrees of freedom	Test Criterion S
Lexington	Dry	1.4	3.5	48	56.34
Louisville	Late	1.8	2.0	48	40.77
Paducah	Late	2.1	3.0	48	46.01
Ashland	Early	1.3	3.0	80	95.98

The critical value is obtained by using the asymptotic expression (for large degrees of freedom)

$$\chi^2_{k,\ 1-\alpha} \simeq \frac{1}{2} \{z_{1-\alpha} + (2k-1)^{1/2}\}^2,$$

where $z_{1-\alpha}$ denotes the $(1-\alpha)$th point on the standard normal distribution and k denotes the degrees of freedom. (See Thompson (1941)).

When $k = 48$, $\alpha = .01$, we obtain

$$\chi^2_{48,\ .99} \simeq 68.52$$

when $k = 40$, $\alpha = .01$, $\chi^2_{80,\ .99} \simeq 106.13$.

Hence, we accept H_0: $\beta = \beta_0$ in each case.

<u>We also carried out the following tests.</u>

<u>Lexington</u>. We combined the data for the Early and Late seasons

and tested the hypothesis that it constitutes a random sample from the Cauchy type of extreme value distribution, by carrying out the S-M-F test with m = 48, r = 23 and s = 24. Then

$$F = R/S = \frac{(\Sigma \ell_j)/23}{(\Sigma \ell_j)/24} = \frac{.20}{.35} = .57,$$

which leads to acceptance of the null hypothesis.

Louisville. We combined the data for the Dry and Early seasons, took the logarithms of the observations and carried out the S-M-F tests with m = 48, r = 23, s = 24 and obtained

$$F = R/S = \frac{.29}{.55} = .53,$$ which is again not significant.

Paducah. We combined the data for the Dry and Early seasons and after taking logarithms, carried out the M-S-F goodness of fit test with m = 48, r = 23, s = 24 and obtained

$$F = \frac{.24}{.38} = 0.63,$$

which is not statistically significant.

Ashland. Since a common δ can be used for all the seasons, the data for the three seasons is combined, logarithms of the observations are obtained and S-M-F test is carried out with m = 120, r = 59, s = 60. We obtain

$$F = .25/.50 = 0.5$$

which is not statistically significant. Hence, it is possible to combine all the seasons into a single season and consequently a single distribution can be fitted to the yearly maximum daily rainfall data in Ashland.

5. Application of the Results.

(i) Suppose we are interested in finding out the chance of getting a maximum rainfall of 3 inches or more during the Dry Convective season in Lexington. This is approximately given by

$$1 - e^{-(3.00/1.69)^{-3.31}} = 1-.86 = .14,$$

since the MLE's of δ and β are 1.69 and 3.31 respectively.

(ii) What extreme daily rainfall during the Dry season in Lexington can be predicted with 95% confidence? We wish to solve the equation

$$e^{-(x/1.69)^{-3.31}} = .95,$$

or

$$(x/1.69)^{-3.31} = -\ell n(.95) = .05129.$$

That is

$$(3.31)\ [\ell nx - \ell n\ 1.69] = -\ell n\ .05129$$

yielding

$$\ell nx = 1.422, \text{ hence } x = 4.14.$$

6. Summary and conclusions.

On the basis of the results of Section 4 we infer that Cauchy type of extreme value distribution can be fitted to the maximum daily rainfall data. Obviously a single distribution does not fit all the seasons in each station. Hence, for Lexington it is recommended to have one distribution for the Dry season and another for the combined Early and Late seasons. For Louisville, one distribution for Late season and another for the combined Early and Dry seasons, can be fitted. For Paducah, Dry and Early seasons can be combined into one season; hence, one distribution for the combined season and another for the Late season can be fitted. For Ashland, we can fit a single distribution for all the seasons. Thus it suffices to look at the yearly maximum daily rainfall data for Ashland.

REFERENCES

[1] DAVID, H.A. (1970). Order Statistics, John Wiley & Sons, New York.

[2] GRIFFIN, D.G. (1974). Reference Manual for Daily Climitological Observations in Kentucky. Engineering Memorandum No. 4, Division of Water, Department of Natural Resources, Commonwealth of Kentucky.

[3] GUMBEL, E.J. (1958). Statistics of Extremes. John Wiley & Sons, New York.

[4] HASSANEIN, K.M. (1969). Estimation of the parameters of the extreme value distribution by use of two or three order statistics. Biometrika, Vol. 56 pp. 429-436.

[5] GRADSHTEYN, I.S. and RYSHIK, I.W. (1965). Table of Integrals, Series and Products. Academic Press, Fourth Edition, New York, pp. 573-574.

[6] MANN, N. SCHEUER, E. and FERTIG, K. (1973). A new goodness of fit test for the two-parameter Weibull or Extreme value distribution with unknown parameters. Communications in Statistics, Vol. 2 No. 5, pp. 383-400.

[7] NATIONAL BUREAU OF STANDARDS (1951). Table of the first moment of ranked extremes. Report No. 1167, Special Report submitted to the National Advisory Committee for Aeronautics.

[8] PYKE, R. (1965). Spacings. Journal of the Royal Statistical Society, Series B Vol. 27, pp. 395-449.

[9] SARHAN, A.E. and GREENBERG, B.G. (1962). Contributions to Order Statistics. John Wiley & Sons, New York.

[10] THOMPSON, C.M. (1941). Table of percentage points of chi-square distribution. Biometrika Vol. 32, Part 2, pp. 187-191.

[11] WATSON, G.S. (1954). Extreme values in samples from m-dependent stationary processes. Annals of Mathematical Statistics , Vol. 25, pp. 798-800.

[12] WHITE, J.S. (1967). The Moments of Log-Weibull Order Statistics. General Motors Research Publication GMR-717, General Motors Research Laboratories, Warren, Michigan.

[13] WHITE, J.S. (1964). Least squares unbiased censored linear estimates for the log Weibull distribution. SIAM, <u>Journal of the Industrial Mathematics Society</u>, Vol. 14, pp. 21-60.

APPENDIX 1: Maximum Daily Rainfall Data

Maximum daily rainfall data for each year in inches during the following seasons for four stations.

I. Late Convective Season: November 1 to April 30,

II. Early Convective Season: May 1 to July 31,

III. Dry Convective Season: August 1 to October 30

Name of Station: Ashland: 1932-1972

Dry	Early	Late
3.97	2.52	1.90
1.64	1.27	1.24
0.84	1.88	1.68
1.46	1.28	2.67
1.21	0.98	1.94
1.74	3.38	1.50
2.30	1.76	2.63
1.89	1.12	1.20
1.05	1.37	0.96
1.91	1.91	2.06
1.52	1.92	1.46
0.91	0.86	1.71
0.80	1.12	2.31
1.15	2.46	1.10
3.03	2.55	0.95
1.72	1.06	1.91
1.60	0.99	2.09
2.41	1.27	1.45
2.53	1.44	2.02
0.96	2.00	1.86
2.23	1.26	1.50
2.72	2.91	1.59
1.44	3.21	1.26
1.46	1.85	2.02
1.41	0.57	2.35
1.70	2.22	0.95
1.12	1.44	1.45
1.40	1.56	1.34
1.47	3.02	1.16
1.16	4.09	1.63
2.16	1.24	2.25
2.29	0.77	1.64

Ashland (continued)

Dry	Early	Late
1.70	2.84	1.80
1.09	2.04	1.69
1.17	2.00	2.20
1.80	1.45	1.97
2.35	0.93	1.26
1.25	1.30	2.03
0.74	2.09	0.71
1.32	0.98	2.11

Lexington: 1948-1972

Dry	Early	Late
1.26	1.23	2.07
0.80	2.59	2.96
1.30	1.57	2.18
1.31	1.08	2.81
1.49	2.24	2.71
3.08	2.10	1.99
2.06	2.33	1.63
1.66	1.46	2.39
2.30	2.09	2.57
1.73	2.34	1.34
1.66	2.38	1.37
3.05	3.89	1.56
0.86	3.80	1.77
2.39	1.48	2.90
0.96	2.06	1.42
2.49	1.93	3.54
1.19	1.74	1.11
2.21	1.95	1.74
1.27	1.93	1.88
3.56	1.58	1.89
2.94	1.85	1.48
2.96	1.80	2.76
1.23	2.96	1.13
1.70	2.36	2.87

Louisville: 1948-1972.

Dry	Early	Late
1.83	2.24	1.67
1.09	2.23	1.84
2.27	1.24	2.56
1.03	2.25	1.95
0.37	1.60	1.87
3.19	0.57	1.51
2.33	1.90	2.23
1.47	1.23	1.94
1.61	2.66	3.07
2.36	1.32	1.06
1.85	1.16	2.15
1.20	5.12	1.68
1.48	4.60	2.04
2.36	1.64	1.94
2.58	1.65	2.72
1.20	1.43	6.97
2.95	1.77	1.49
1.77	1.40	2.56
2.40	1.51	1.75
1.61	1.00	2.14
1.72	1.17	1.68
2.84	2.35	4.08
1.73	3.35	1.58
1.19	2.35	1.73

Paducah: 1948-1972

Dry	Early	Late
2.77	2.09	1.54
2.14	1.96	3.55
1.66	1.78	4.06
3.16	1.19	3.68
1.34	1.46	1.51
3.38	1.93	2.51
1.30	3.01	2.26
1.23	1.26	2.29
2.28	2.69	4.67
1.52	2.75	2.78
2.17	2.07	2.49
1.96	2.67	1.22
1.64	2.39	2.68
1.90	2.48	1.91
1.02	1.78	1.42
1.84	2.02	5.28
4.74	3.20	2.35
1.17	3.29	3.40
1.63	3.76	2.27
1.41	2.11	2.38
2.19	1.08	2.83
1.99	1.30	2.57
2.40	2.61	2.25
1.86	2.93	2.05

APPENDIX 2: Relation Between the Type I Extreme Value Distribution and the Cauchy Type of Extreme Value Distribution

Let X denote the random variable having Cauchy type of extreme value distribution with zero for location parameter, δ for scale and β for the shape parameter. That is, its distribution function is given by

$$F_X(x;\beta,\delta) = \exp(-(x/\delta)^{-\beta}), \quad x,\beta,\delta>0. \tag{1}$$

Let us define $Y = \ell nX$. Then

$$F_Y(y) = P(Y\leq y) = P(X\leq e^y) = \exp[-(e^y/\delta)^{-\beta}] = \exp\{-\exp[-\beta(y-\ell n\delta)]\}$$

which is Type I extreme value distribution having $\ell n\delta$ for the location parameter and $1/\beta$ for the scale parameter. Notice that the range of variation of y is from $-\infty$ to ∞.

APPENDIX 3: Maximimum Likelihood (ML) Estimates (MLEs) of the Scale and Shape Parameters of the Cauchy Type of Extreme Value Distribution.

The probability density function of the Cauchy type of extreme value distribution is given by

$$f(x;\delta,\beta) = (\beta/\delta)\ (x/\delta)^{-(\beta+1)}\ e^{-(x/\delta)^{-\beta}}, \quad x,\delta,\beta>0.$$

The likelihood of δ and β based on a random sample of size n is given by

$$\ell(\delta,\beta) = \ell n\left\{\prod_{i=1}^{n} f(x_i;\delta,\beta)\right\} = n\ell n\beta - n\ell n\delta + \sum_{i=1}^{n} \ell n \left(\frac{x_i}{\delta}\right)^{-(\beta+1)} - \sum_{i=1}^{n} (x_i/\delta)^{-\beta}.$$

$$\frac{\partial\ell}{\partial\delta} = 0 \text{ yields} \quad \delta^{\beta} = n/\sum_{1}^{n} x_i^{-\beta}. \qquad (1)$$

$$\frac{\partial\ell}{\partial\beta} = 0 \text{ yields} \quad \beta = [n^{-1}\Sigma\ell n x_i - \frac{\Sigma x_i^{-\beta}\ \ell n x_i}{\Sigma x_i^{-\beta}}], \qquad (2)$$

We start with an initial value of β and solve (2) iteretively and then solve for δ from (1).

Asymptotic variances and covariance of MLEs

It is well know that the variance-covariance matrix of $(\hat{\beta},\hat{\delta})$ is given by n

$$n^{-1}\ I^{-1} \qquad (3)$$

where

$$I = \begin{pmatrix} -E\,\frac{\partial^2 \ell n f}{\partial\beta^2} & -E\,\frac{\partial^2 \ell n f}{\partial\beta\partial\delta} \\ -E\,\frac{\partial^2 \ell n f}{\partial\beta\partial\delta} & -E\,\frac{\partial^2 \ell n f}{\partial\delta^2} \end{pmatrix} . \qquad (4)$$

where

$$\ell nf(X;\beta,\delta) = \ell n\beta - \ell n\delta - (\beta+1)(\ell nX - \ell n\delta) - \left(\frac{X}{\delta}\right)^{-\beta}.$$

Hence

$$\frac{\partial^2 \ell nf}{\partial\delta^2} = \frac{-\beta}{\delta^2} - \frac{\beta(\beta-1)}{\delta^2}\left(\frac{X}{\delta}\right)^{-\beta}, \tag{5}$$

$$\frac{\partial^2 \ell nf}{\delta\beta^2} = \frac{-1}{\beta^2} - \left(\frac{X}{\delta}\right)^{-\beta}\left(\ell n\left(\frac{X}{\delta}\right)\right)^2, \tag{6}$$

and

$$\frac{\partial^2 \ell nf}{\partial\delta\partial\beta} = \frac{1}{\delta} - \frac{1}{\delta}\left(\frac{X}{\delta}\right)^{-\beta} + \frac{\beta}{\delta}\left(\frac{X}{\delta}\right)^{-\beta}\ell n\left(\frac{X}{\delta}\right). \tag{7}$$

Straight forward computations yeild

$$E\left(\frac{\partial^2 \ell nf}{\partial\delta^2}\right) = \frac{-\beta}{\delta^2} - \frac{\beta(\beta-1)}{\delta^2} = \frac{-\beta^2}{\delta^2}, \tag{8}$$

and

$$E\left(\frac{\partial^2 \ell nf}{\delta\beta^2}\right) = -\frac{1}{\beta^2} - \frac{1}{\beta^2}\left(\frac{\pi^2}{6} + \gamma^2 - \gamma\right)$$

$$= -\frac{1}{\beta^2}\left(1 + \frac{\pi^2}{6} - \gamma^2 - \gamma\right), \tag{9}$$

where γ denotes the Euler's constant = .5772. Also

$$E\left(\frac{\partial^2 \ell nf}{\partial\beta\partial\delta}\right) = \frac{1}{\delta} - \frac{1}{\delta} - \left(\frac{\beta}{\delta}\right)\left(\frac{1-\gamma}{\beta}\right) = -\left(\frac{1-\gamma}{\delta}\right). \tag{10}$$

Thus

$$I = \begin{pmatrix} \beta^{-2}\left(1-\gamma+\gamma^2+\frac{\pi^2}{6}\right) & \left(\frac{1-\gamma}{\delta}\right) \\ \left(\frac{1-\gamma}{\delta}\right) & \frac{\beta^2}{\delta^2} \end{pmatrix} \tag{11}$$

Consequently

$$I^{-1} = (\gamma + \frac{\pi^2}{6})^{-1} \begin{pmatrix} \beta^2 & -\delta(1-\gamma) \\ -\delta(1-\gamma) & \frac{\delta^2}{\beta^2}(1-\gamma+\gamma^2+\frac{\pi^2}{6}) \end{pmatrix} . \quad (12)$$

That is,

$$\text{Var } \hat{\beta} = \frac{\beta^2}{n} (\frac{\pi^2}{6} + \gamma)^{-1} = 0.45 \frac{\beta^2}{n}, \quad (13)$$

$$\text{Var } \hat{\delta} = \frac{\delta^2}{n\beta^2} (1-\gamma+\gamma^2 + \frac{\pi^2}{6}) (\frac{\pi^2}{6} + \gamma)^{-1} = \frac{1.08\delta^2}{n\beta^2}, \quad (14)$$

and

$$\text{Cov } (\hat{\beta},\hat{\delta}) = - \frac{\delta}{n} (1-\gamma) (\frac{\pi^2}{6} + \gamma)^{-1} = - \frac{.19\delta}{n} . \quad (15)$$

<u>Integrals used for evaluating the variance-covariance matrix.</u>

Let Z denote the standard Cauchy extreme value random variable. That is, $Z = X/\delta$. Then

$$E(Z^{-\beta}) = \beta \int_0^\infty z^{-\beta} z^{-\beta-1} e^{-z^{-\beta}} dz = \int_0^\infty u e^{-u} du = 1,$$

after using the substitution $z^{-\beta} = u$.

$$E(Z^{-\beta} \ln Z) = \beta \int_0^\infty z^{-\beta} \ln z \,.\, z^{-\beta-1} e^{-z^{-\beta}} dz$$

$$= - \beta^{-1} \int_0^\infty u \ln u\, e^{-u} du$$

$$= - \beta^{-1}[\int_0^\infty e^{-u} \ln u\, du + \int_0^\infty e^{-u} du] = - \beta^{-1}(-\gamma+1),$$

after performing integration by parts. Next, consider

$$E(Z^{-\beta}(\ell nZ)^2) = \beta \int_0^\infty z^{-\beta} (\ell nz)^2 z^{-\beta-1} e^{-z^{-\beta}} dz$$

$$= \beta^{-2} \int_0^\infty u (\ell nu)^2 e^{-u} du$$

$$= \beta^{-2} [\int_0^\infty (\ell nu)^2 e^{-u} du + \int_0^\infty (\ell nu) e^{-u} du]$$

$$= \beta^{-2} [\frac{\pi^2}{6} + \gamma^2 - \gamma].$$

<u>Integrals taken from a book on integrals</u>

$$\int_0^\infty e^{-x} (\ell nx)^2 dx = \frac{\pi^2}{6} + \gamma^2$$

$$\int_0^\infty e^{-x} (\ell nx) dx = -\gamma = -.5772.$$

(See Gradshteyn and Ryzhik (1965, pp. 573-574).

APPENDIX 4: ML Estimation of Location, Scale and Shape Parameters of the Cauchy Type of Extreme Value Distribution

The three parameter Cauchy type of extreme value distribution is given by

$$f(x;\theta,\delta,\beta) = (\beta/\delta)\ (\tfrac{x-\theta}{\delta})^{-\beta-1}\ \exp\{-(\tfrac{x-\theta}{\delta})^{-\beta}\},\quad \delta,\beta>0 \text{ and } x\geq\theta.$$

The likelihood of θ, δ and β based on a random sample of size n from the above density is given by

$$\ell(\theta,\delta,\beta) = \ell n\{\prod_{i=1}^{n} f(x_i;\ \theta,\delta,\beta)\}$$

$$= n\ell n\beta - n\ell n\delta - (\beta+1)\ \{\sum_{1}^{n} \ell n(x_i-\theta) - n\ell n\ \delta\} - \sum_{i=1}^{n} (\tfrac{x_i-\theta}{\delta})^{-\beta}\ .$$

$\frac{\delta\ell}{\partial\delta} = 0$ implies that

$$\delta^{-\beta} = n^{-1} \sum_{i=1}^{n} (x_i-\theta)^{-\beta}, \qquad (1)$$

$\frac{\partial\ell}{\partial\beta} = 0$ implies that

$$\beta^{-1} = n^{-1}\ \Sigma\ \ell n\ (x_i-\theta) - \frac{\Sigma(x_i-\theta)^{-\beta}\ \ell n(x_i-\theta)}{\Sigma(x_i-\theta)^{-\beta}}\ , \qquad (2)$$

$\frac{\partial\ell}{\partial\theta} = 0$ implies that

$$(\beta+1) \sum_{i=1}^{n} (x_i-\theta)^{-1} = \frac{\beta\ n\ \Sigma\ (x_i-\theta)^{-\beta-1}}{\Sigma\ (x_i-\theta)^{-\beta}}\ . \qquad (3)$$

One can iteratively solve for θ and β from (2) and (3). Using these estimates in (1) one can solve for δ. When $\theta=0$, equations (1) and (2) reduce to (1) and (2) of Appendix 3. As in Appendix 3, one can obtain explicit expression for the asymptotic variance-covariance matrix of the maximum likelihood estimates of θ, δ and β.

APPENDIX 5: Least Squares Estimates of Location and Scale Parameters Based on Ordered Observations

Let $X_{1n} \leq \ldots \leq X_{nn}$ denote the ordered observations in a random sample of size n drawn from a density of the form $\frac{1}{\sigma} f(\frac{x-\theta}{\sigma})$. Let

$$U_{in} = (X_{in}-\theta)/\sigma, \quad i=1,\ldots,n,$$

Then the distribution of the U_i will be free of the parameters θ and σ. Let

$$\mu_{in} = EU_{in}, \qquad \sigma_{ijn} = \text{Cov}(U_{i,N}, U_{j,N}), \; 1\leq i, \; j\leq n.$$

$$\underline{\mu} = (\mu_{1n}, \ldots, \mu_{nn})' \text{ and } \Sigma = ((\sigma_{ijn}))$$

and $\Omega = \Sigma^{-1}$. Then the best linear unbiased estimates of θ and σ together with their variances are

$$\hat{\theta} = -\underline{\mu}' \, \Gamma \, \underline{X}, \quad \text{var } \hat{\theta} = (\underline{\mu}' \Omega \underline{\mu})\sigma^2/\Delta$$

$$\hat{\sigma} = \underline{1}' \, \Gamma \, \underline{X}, \quad \text{var } \hat{\sigma} = (\underline{1}' \Omega \underline{1})\sigma^2/\Delta$$

$$\text{Cov}(\hat{\theta}, \hat{\sigma}) = -(\underline{1}' \Omega \underline{\mu})\sigma^2/\Delta$$

where

$$\underset{\sim}{X} = (X_{1n}, \ldots, X_{nn})', \quad \underset{\sim}{1}' = (1, \ldots, 1),$$

$$\Gamma = \Omega(\underline{1}\underline{\mu}' - \underline{\mu}\underline{1}')\Omega/\Delta$$

and

$$\Delta = (\underset{\sim}{1}' \Omega \underline{1})(\underline{\mu}' \Omega \underline{\mu}) - (\underset{\sim}{1}' \Omega \underline{\mu})^2.$$

For details see Sarhan and Greenberg (1962 pp. 21-22).

Remark 5.1. $\hat{\theta}$ and $\hat{\sigma}$ can be evaluated even if one has a censored sample (that is, certain observations from the ordered sample are missing).

Remark 5.2. If f(x) is symmetric about zero and the sample is not censored, then $\underset{\sim}{1}'\underset{\sim}{\mu} = 0$ and $\underset{\sim}{1}'\Omega\underset{\sim}{\mu} = 0$. Thus

$$\hat{\theta} = \underset{\sim}{1}'\Omega\underset{\sim}{X}/\underset{\sim}{1}'\Omega\underset{\sim}{1}, \quad \hat{\sigma} = \underset{\sim}{\mu}'\Omega\underset{\sim}{X}/\underset{\sim}{\mu}'\Omega\underset{\sim}{\mu}, \quad \text{var}\ (\hat{\theta}) = \sigma^2/\underset{\sim}{1}\Omega\underset{\sim}{1}$$

$$\text{var}\ (\hat{\sigma}) = \sigma^2/\underset{\sim}{\mu}'\Omega\underset{\sim}{\mu} \quad \text{and} \quad \text{cov}(\hat{\theta}, \hat{\sigma}) = 0.$$

The expected values of Type I extreme value order statistics are given up to n=100 by the National Bureau of Standards (1951) and White (1967). The variances and covariances of the Type I extreme value order statistics are given up to n = 20 by White (1964).

APPENDIX 6: Goodness of Fit Tests

(i) Kolmogorov-Smirnov test.

Let $X_1, \ldots, X_n$ denote a random sample from the continuous (unknown) distribution function $F(x)$. On the basis of this sample we wish to test the null-hypothesis

$$H_0: \; F(x) = F_0(x), \text{ for all } x,$$

where $F_0(x)$ is completely specified. The alternative hypothesis is H_1: $F(x) \neq F_0(x)$ for at least some x. The test criterion of Kolmogorov and Smirnov (K-S) is: Rejected H_0 when

$$T = \max_i \; |F_n(X_i) - F_0(X_i)| > k_\alpha$$

where k_α is a constant which depends on α and n, and $F_n(x)$ denotes the empirical distribution function defined by

$$F_n(x) = (\text{number of } X_i \leq x)/n.$$

The K-S is distribution-free in the sense that k_α is free of $F_0(x)$. Critical points of T are available and the asymptotic distribution of T is also well known.

Since we are testing H_0 against a broad class of alternatives, we should not expect too much power from this test. Quite often we may be able to specify $F_0(x)$ except for certain parameters, for instance, the location and scale parameters. Then, one can carry out the K-S test after the unknown parameters are estimated. Then the true level of significance may not coincide with the nominal level of significance. Usually it is less than the nominal level of significance. Alternatively, we can look for a goodness of fit test which is invariant under location and scale parameters. Such tests are available and one of them will be considered next.

(ii) Invariant Goodness of Fit Test of Mann-Scheuer-Fertig (1973)

This test is based on standaradized spacings and the test is asymptotic in nature. (That is, it is applicable for large values of n).

Let $X_{1n} \leq X_{2n} \leq \ldots \leq X_{nn}$ be the order statistics in a random sample of size n drawn for F(x). As before, we wish to test H_0: $F = F_0$ against H_1: $F \neq F_0$.

Let $\ell_j = (X_{j+1,n} - X_{j,n}) / E_0(X_{j+1,n} - X_{j,n})$ where E_0 denotes expectation taken when F_0 is the true underlying distribution. When F_0 is the negative exponential, that is

$$F_0(x) = (1/\sigma) \exp\{-(x-\theta)/\sigma\}, \qquad x \geq \theta,$$

the ℓ_j $(j = 0,1,\ldots,n-1)$ are independent and $2\ell_j$ has a chi-square distribution with two degrees of freedom. Pyke (1965) has shwon that for all continuous F_0, the ℓ_i are asymptotically independent and $2\ell_i$ is asymptotically distributed as chi-square with 2 degrees of freedom. If the sample is censored above at m (that is, only $X_{1n}, \ldots, X_{m,n}$ are available) then the statistic

$$L = r^{-1} \sum_{j=m-r}^{m-1} \ell_j / [s^{-1} \sum_{j=1}^{s} \ell_j]$$

is asymptotically distributed as the Snedecor F with 2r and 2s degrees of freedom.

Since the right-hand tail of the extreme value density is "shorter" than the usual appropriate alternative distributions, while the left tail is "larger", the upper "gaps", $X_{i+1,n} - X_{i,n}$ (upper meaning that i is closer to m-1 than to 1) will tend to be smaller than the "lower" gaps. That is, $(X_{m,n} - X_{m-1,n})/(X_{2,n} - X_{1,n})$ will tend to be smaller under the hypothesis that the sample was drawn from an extreme value distribution than under the alternatives of interest. Also, it is optimal to use $r+1 = s = m/2$ or $(m-1)/2$ according as m is even or odd respectively. The test for fit based on L can be used for samples of size as large as n=100 by using the tables of expected values of the reduced extreme value order statistics given by White (1967). Further, the convergence of the distribution of L to its asymptotic one is rapid compared to other goodness of fit tests.

APPENDIX 7: Tests of Hypotheses About δ When β is Known

Let us assume that the maximum daily rainfall data conforms to the two-parameter Cauchy type of extreme value distribution, the density of which is given by

$$f_X(x;\delta,\beta) = (\beta/\delta)\ (x/\delta)^{-\beta-1}\ e^{-(x/\delta)^{-\beta}}, \quad x>0.$$

Lemma 1. Let $Y = (X/\delta)^{-\beta}$. Then Y has the negative exponential distribution given by

$$F_Y(y) = 1 - e^{-y}, \qquad y>0.$$

Proof.
$$P[Y \leq y] = P((X/\delta)^{-\beta} \leq y) = P(X/\delta > y^{-\frac{1}{\beta}})$$
$$= 1 - e^{-y}.$$

Result. The uniformly most powerful test of H_0: $\delta > \delta_0$ when β is known, is to reject H_0 when

$$2 \sum_{i=1}^{n} (Y_i/\delta_0)^{-\beta} < k_\alpha,$$

where n denotes the random sample size, $k_\alpha = \chi^2_{2n,\alpha}$, the αth point on the chi-square distribution having 2n degrees of freedom.

Proof. Let us consider a fixed alternative $\delta = \delta_1 > \delta_0$. Then according to Neyman-Pearson theory, the most powerful test is given by : reject H_0 when Λ_n; the likelihood ratio exceeds a certain constant, where

$$\Lambda_n = \frac{(\beta/\delta_1)^n \prod_{i=1}^{n} (y_i/\delta_1)^{-(\beta+1)} e^{-\sum_{i=1}^{n} (y_i/\delta_1)^{-\beta}}}{(\beta/\delta_0)^n \prod_{i=1}^{n} (y_i/\delta_0)^{-(\beta+1)} e^{-\sum_1^n (y_i/\delta_0)^{-\beta}}}$$

$$= (\delta_0/\delta_1)^{n-\beta-1} e^{-\sum_{i=1}^{n} [(y_i/\delta_1)^{-\beta} - (y_i/\delta_0)^{-\beta}]} .$$

That is, we reject H_0 when

$$\sum_{i=1}^{n} [(y_i/\delta_1)^{-\beta} - (y_i/\delta_0)^{-\beta}] < k ,$$

$$\sum_{i=1}^{n} (y_i/\delta_0)^{-\beta} [(\frac{\delta_0}{\delta_1})^{-\beta} - 1] < k ,$$

or

$$\sum_{i=1}^{n} (y_i/\delta_0)^{-\beta} < k' \text{ since } \delta_0 < \delta_1 .$$

Now, since $2(y_i/\delta_0)^{-\beta}$ is distributed as chi-square variable with 2 degrees of freedom when H_0 is true, the conclusion of the result readily follows with $k' = \chi^2_{2n,\ \alpha}$, where α denotes the level of significance. Now, since the test is free of the simple alternative, namely δ_1, the above test is uniformly most powerful.

Next, inverting the exceptance region, we obtain the uniformly most accurate lower confidence bound for δ as

$$[\chi^2_{2n,\alpha}/2 \sum_{i=1}^{n} y_i^{-\beta}]^{1/\beta} .$$

Also, a $(1-\alpha)$ confidence interval for δ is given by

$$\left[\frac{\chi^2_{2n,\alpha/2}}{2\ \Sigma y_i^{-\beta}}\right]^{1/\beta} \leq \delta \leq \left[\frac{\chi^2_{2n,1-\alpha/2}}{2\ \Sigma y_i^{-\beta}}\right]^{1/\beta} .$$

APPENDIX 8: Tests for the Parameter β.

Suppose that δ is known and we wish to test H_0: $\beta=\beta_0$ against the alternative H_1: $\beta\neq\beta_0$. Let $X_1,\ldots,X_n$ be a random sample available for the test. Recall that δ is known and that $(X_i/\delta)^{-\beta}$ has a negative exponential distribution with mean 1. Let $X_{1,n} < X_{2,n} < \ldots < X_{n,n}$ denote the order X_i's. Then define $Y_{in} = (X_{in}/\delta)^{-\beta}$ $i=1,\ldots,n$. Then

$$\ell_i = \frac{(X_{i+1,n}/\delta)^{-\beta_0} - (X_{i,n}/\delta)^{-\beta_0}}{E(Y_{i+1,n}) - E(Y_{i,n})}$$

$$= (n-i)\ [Y_{i+1,n}^{-\beta} - Y_{i,n}^{-\beta}]$$

(since $EY_{i,n} = n^{-1} + (n-1)^{-1} + \ldots + (n-i+1)^{-1}$).

Let $S = 2 \sum_{i=1}^{n-1} \ell_i \stackrel{d}{=} \chi^2_{2(n-1)}$ under H_0. We reject H_0 when $S > \chi^2_{2(n-1),\ 1-\alpha}$.

Remark 8.1. In fitting Cauchy type of extreme value distribution, if the δ's are reasonably the same for two or more seasons, we can test the hypothesis whether a common specified β will do the job for the seasons in question. First all the data should be divided by the common δ agreed upon and then a common β_0 should be hypothesized before the above test is carried out.

Remark 8.2. If the common δ is unknown, we can combine the data for two or more seasons, take logarithms of the observations and then carry out the M.S.F test procedure outlined in Appendix 6.