Zeitschrift: Schweizer Ingenieur und Architekt
Herausgeber: Verlags-AG der akademischen technischen Vereine
Band: 106 (1988)
Heft: 37

Artikel: On outlier detection tests
Autor: Shen, Hsieh Wen / Tabios, G.Q. / Obeysekera, J.
DOI: https://doi.org/10.5169/seals-85803

On Outlier Detection Tests

Ein Problem für die Auswertung von Hochwasser mittels Wahrscheinlichkeitsmodellen sind Ausreisser in den Messreihen. In dieser Arbeit werden einige Tests zur Identifikation von Ausreissern mit Hilfe der Monte Carlo Simulation untersucht. Es zeigt sich, dass die meisten gebräuchlichen Identifikationstests bei transformierten Stichproben dieser Verteilung unbefriedigende Resultate liefern.

BY HSIEH WEN SHEN,
G.Q. TABIOS III AND
J. OBEYSEKERA,
USA

In the summer, 1987, severe storms hit many parts of Switzerland and caused numerous landslides. The first writer was stopped by police for attempting to enter Sustenpass from Wassen due to landslides. He had to drive around lake of Lucerne, through Brünigpass and finally reached Sustenpass before dark. The questions of the return periods and outliers in analysing these storms occured in his mind. Since Professor *Th. Dracos* has always stressed the need to use computer and statistics in analyzing hydrologic problems, the results of this study on outliers are presented here to honour Professor Dracos' birthday.

For outliers, a variety of definitions exists in both statistical and hydrologic literature. Often the terms such as "discordant", "inconsistent", "unrepresentative", etc. are used to describe outliers. The definition of Grubbs [6] is typical: an outlier "is one that appears to deviate markedly from other members of the sample in which it occurs". This definition excludes the "foreign" data which contaminate a sample, yet appears to be homogeneous with the rest of the samples. Of course, such contaminations is also of interest to us in estimations although their detection is extremely difficult.

In general, outliers occur because of [6,1]: (a) measurement errors; (b) execution errors (imperfect collection of data); and (c) inherent variability. The first two causes may be characterized as deterministic, although the resulting outliers may be random. Clearly, the outliers resulting from gross errors in measurement must be discharded or corrected. Often in flood data, large measurement errors are associated with the highest values (due to errors in stage-discharge relations or other estimation procedures). The correction of such data is difficult. In this case, discarding the outliers may also result in a valuable loss of information.

Inherent variability is the natural variation of the process as reflected in its population characteristics. Since the outliers may be caused by many types of inherent variability, their detection and modeling are difficult and often requires assumptions regarding the underlying model for outlier generation.

The statistical tests for outlier detection usually require assumptions regarding a model for outlier generation. In these tests, the null hypothesis is that the entire sample arises from a common distribution, say F. The alternative hypothesis can take several forms depending on the form of the model for outlier generation. Barnett and Lewis [3] present the following alternative hypotheses: (a) deterministic alternative which includes the measurement and execution errors; (b) inherent alternative which favors the rejection of model F for alternative model G for the entire sample; (c) mixture alternative which allows a proportion of the sample to come from an alternate distribution G: (d) slippage alternative which allows a certain small number of observations to arise from a modified version of the initial model F, typically with an increased location or an increased scale parameter; and (e) exchangeable alternative which assumes that the outlier arises from an alternate model G but its index is equally likely to be (1, 2, ..., n) where n is the sample size. The reader is referred to a paper by Barnett [2] for the mathematical formulations of the above alternative hypotheses.

Tests for Outlier Detection

The following four tests for detecting outliers are selected for the study:

Barnett and Lewis Test

Barnett and Lewis [3] presented a test of discordancy for a single upper outlier in a gamma or exponential sample based on a maximum-likelihood ratio test.

Studentized Test

This test is applicable for a single upper outlier in a normal sample with unknown mean and standard deviation. The test statistic is also called internally studentized extreme deviation from the mean [3]. There is also a maximum likelihood ratio test for a location slippage alternative in which one observation arises from a normal distribution with a higher mean. The acceptance rule is similar to the first test by Barnett and Lewis [3] and its critical values are tabulated by these authors as a function of r and sample size n.

Skewness Test

Tables giving the critical values of these statistics for a normal population are used based on the characterization that the sample skewness statistic has an asymptotic variance of $6n(n-1)/[(n-2)(n+1)(n+3)]$, where n is the sample size.

Kurtosis Test

Critical values of the statistic are also computed for a normal population. The kurtosis statistic has an asympotic variance of $24/n$.

Analysis of the Problems

Both skewness and kurtosis tests can be applied for one or more upper or lower outliers in a normal sample with unknown mean and variance. These two tests are locally best unbiased invariant tests against a location-slippage alternative provided that the ratio of contaminants to the total number of samples is less than 50 percent and 20 percent for the skewness test and kurtosis test, respectively [3].

In this study, the four tests given above are applied to samples taken from gamma distribution and log-Pearson Type III distribution. The following transformations are used to suitably apply the different tests.

In the case where the gamma variables ($y_1, \ldots, y_n$) have a common density function given by [4]:

$$f(y) = \frac{1}{|\lambda|\Gamma(r)}\left[\frac{y-y_o}{\lambda}\right]^{r-1}\exp\left[-\frac{y-y_o}{\lambda}\right] \tag{1}$$

Where y_o, λ and r are respectively the location, scale and shape parameters. The Barnett-Lewis Test is used after the transformation

$$x_j = y_j - y_o \tag{2}$$

Where y_o, λ and r are respectively the location, scale and shape parameters. The Barnett-Lewis Test is used after the transformation

$$x_j = (y_j - y_o)^{1/2} \tag{3}$$

such that x_j is approximately normally distributed with mean $\sqrt{\lambda(r-1/4)}$ and variance $\lambda/4$. An alternative transformation is the cube-root transformation Kottegoda [7]:

$$x_j = (y_j - y_o)^{1/3} \tag{4}$$

then x_j is approximately normally distributed with mean $(r\lambda)^{1/3}[1-1/(9r)]$ and variance $(\lambda^2/r)^{1/3}/9$.

When the samples ($z_1, \ldots, z_n$) are log-Pearson Types III distributed with density function given by [4]

$$f(z) = \frac{1}{|\lambda|\Gamma(r)}\left[\frac{\log(z)-y_o}{\lambda}\right]^{r-1}\exp\left[-\frac{\log(z)-y_o}{\lambda}\right] \tag{5}$$

in which λ, r and y_o are in the log-domain; the four different tests are applicable using the logarithmic transformation

$$y_j = \log(z_j) \tag{6}$$

followed by Eq. (2) and either one of the two transformations given in Eqs. (3) and (4).

Simulation Study

The simulation experiments commence with generation of 1000 traces of uncontaminated data each of sample sizes 25, 50 and 100. The parameters of the underlying distribution used are $y_o = 5650$, $\lambda = 3000$ and r = 2.8 for gamma; and $y_o = 8.0$, $\lambda = 0.05$ and $r = 25$ for log-Pearson Type III. These parameters correspond approximately to the annual flood series (1916–1975) of St. Mary's River at Stillwater, Canada. It is noted that such a high shape parameter in the log domain estimated for this data implies that the log transformed annual flood can be fitted well by a normal distribution also. In terms of the moment estimates of the mean, standard deviation, and skewness, the gamma distribution has values of 14 000, 5000 and 1.2 respectively and the log-Pearson distribution has (in the log domain) 9.25, 0.25 and 0.4, respectively. Schemes for generating gamma and log-Pearson samples are presented in Bratley, et al [5], and Wallis, et al [9]. Two other data sets are then derived using the generated samples above by deliberately contaminating each trace with a single upper outlier generated from the same underlying distributions but with different location parameters. The location parameters of the contaminants are 15 650 and 20 650 for the gamma samples and 8.5 and 8.75 for the log-Pearson samples. These correspond to their means plus two- and three-standard deviations, respectively. The contamination is accomplished by replacing one observation at random in a given sample trace by a contaminant higher than the maximum value observed for that sample trace.

The four tests for detecting outliers are applied to the different data sets generated above. For brevity in the discussions below, the following labels are adopted. For gamma samples: Test 1 corresponds to the use of Barnett-Lewis test together with Eq. (2); Test 2, Test 3 and Test 4 correspond to using studentized, skewness and kurtosis tests, respectively, with the square-root transformation (Eq. (3)); and, Test 5, Test 6 and Test 7 corresponding to the three latter tests above, respectively, but with the cube-root transformation (Eq. (4)). The test labels for log-Pearson samples are similar to the above but in addition, each sample is logarithmically transformed beforehand (Eq. (6)).

The three distributional parameters of each sample are estimated using the maximum likelihood method. Also, the quantiles corresponding to 50, 500 and 2000 return periods are estimated. Details and program algorithms of the estimation techniques are given in Salas and Smith [8]. In the discussions below, the quartiles are simply referred to as Q50, Q500 and Q1000 for return periods of 50, 500 and 1000, respectively. Also, to differentiate the three data sets for each sampling distribution UCDT refers to uncontaminated data, CNDT1 for the contaminated data with mean plus two standard deviations, and CNDT2 for the contaminated data with mean plus three standard deviations.

Summary of Results

The numbering system for different tests are discribed in the previous section under «Simulation Study».

1. Gamma Distribution (Pearson)

A. Test 4 appears to be the only test to give consistent results in terms of the significance level. Theoretically, 95% of the samples should be passing the test at α = 5% and likewise 99% should be passing at α = 1%. On the other hand, Test 3 and Test 2, in that order give slightly better results as compared to the remaining tests. It may be noted that Tests 2, 3, and 4 are formed using the square-root transformation. Test 1 seems to give the worst results especially at sample size N = 25; in some cases the percentage of passing for uncontaminated data is less than those of the contaminated data. Although Test 1 is specifically designed for gamma samples with two parameters, the estimation of the location parameter appears to have an effect on its performance, or this test is not powerful against the type of slippage alternative introduced in generating the outliers.

B. In most cases, the power of the tests increases as the sample size increases. In view of the generation and contamination schemes, this may be explained by the fact that extreme values are more likely to appear in a sample trace as the sample size increases. This is especially true in the case of the contaminated data since a contaminant has a value always greater than the maximum value obtained in that sample.

C. All tests show that the power of the tests increases only slightly as the degree of contamination increases (from CNDT1 to CNDT2). On the basis of Test 4 alone, it can be observed that, as the sample size increases, the maximum increase in the power is about 12% for α = 5% whereas it is about 6% for α = 1%. However, the effect of increasing the contamination (i. e. CNDT1 to CNDT2) by adding one standard deviation to the mean results in a maximum increase in the power less than 3% for both $\alpha = 5\%$ and $\alpha = 1\%$.

D. High percentages for the case of passing the test by both uncontaminated and contaminated samples (P-P) reaching as much as 100% for some tests. This is further evidence of the poor performance of some of the tests. It is also interesting to note that for Test 1, often the cases of (F-P) are greater than those of (P-F) which confirms its worse performance relative to others.

2. Log-Pearson Type III Distribution

A. In this case, Test 2 show the most consistent results while Test 1 has the worst results as before.

B. For Test 2 the effect of contamination (i. e., from UCDT to CNDT1) reduces the number of passes in the range of 5 to 15%, while the effect of increasing the degree of contamination (i. e. from CNDT1 to CNDT2) is in the order of 10%. This latter effect is significantly different from those of gamma samples.

Conclusions

This is a preliminary study which addresses the general topics of outlier detection and accommodation. Many tests available for detecting outliers are compared by the Monte Carlo Simulation. The underlying population distribution were the three parameter gamma and log-Pearson Type III. A particular scheme of outlier generation which mimics approximately the unlabeled slippage alternative was employed. A single outlier replaces at random a data point in a sample whose sizes are 25, 50 and 100.

Many tests which are particularly based on goodness of fit of normal distribution do not appear to perform satisfactorily for gamma and log-Pearson Type III samples with square-root and cube-root transformations. In particular, for gamma samples, only the kurtosis test with square-root transformation gave results consistent with the significance levels used. The tests designed for the normal distribution (studentized, skewness, and kurtosis) but applied to gamma samples with cube-root transformation yielded poor results. Even the Barnett-Lewis test which is based on the two parameter gamma distribution, does not appear to perform satisfactorily for the three parameter gamma distribution. This discrepancy is suspected to be a consequence of ignorance of the location parameter which is estimated. It is also possible that the test is less powerful against the type of slippage alternative used for outlier generation.

In case of log-Pearson Type III distribution, the studentized test with square-root transformation applied to the logarithms of data is the only one found to be satisfactory. Also, the tests appeared to be more powerful for log-Pearson samples than for gamma samples in the case of the slippage alternative under consideration. This difference in results between gamma and log-Pearson (which is also a gamma in the log domain) is deemed to be due to different population parameters used in Monte Carlo Simulation.

References

[1] *Barnett, V.*, 1978; The Study of Outliers: Purpose and Model, Applied Statistics, Vol. 27, No. 3, pp. 242–250

[2] *Barnett, V.:* 1983; Principles and Methods for Handling Outliers in Data Sets. Statistical Methods and the Improvement of Data Quality, T. Wright (ed.), Academic Press, New York, pp. 131–166

[3] *Barnett, V.* and *T. Lewis*, 1978; Outliers in Statistical Data, John Wiley & Sons, New York

[4] *Bobee, F.* 1975; The Log-Pearson Type III distribution and Its Application in Hydrology, Water Resources Research, Vol. 11, No. 5, pp. 681–689

[5] *Bratley, P., B. L. Fox* and *L. E. Schrage*, 1983; A Guide to Simulation, Springer-Verlag, New York

[6] *Grubbs, F. E.*, 1969; Procedures for Detecting Outlying Observations in Samples, Technometrics, Vol. 11, No. 1, February, pp. 1–21

[7] *Kottegoda, N. T.*, 1984; Investigation of Outliers in Annual Maximum Flow Series, Journal of Hydrology, Volume 72, pp. 105–137

[8] *Salas, J. D.* and *R. A. Smith*, 1985; Statistical Computer Techniques in Hydrology and Water Resources (to be published)

[9] *Wallis, J. R., N. C. Matalas*, and *J. R. Slack*, 1974; Just a Moment!, Water Resources, Vol. 10, No. 2, April, pp. 211–219

Author's address: *Dres. H. W. Shen* and *G. Q. Tabios*, University of California, Berkley, California 94720, U.S.A. and *Dr. J. Obeysekera*, South Florida Water Management District, West Palm Beach, Florida 33402, U.S.A.

Korrosion der Stahlbewehrung in Beton

Berichterstattung über die Arbeitsgruppe Korrosion

Am 28. und 29. April 1988 trafen sich in Stuttgart unter dem Vorsitz von Herrn Dr. B. Isecke Korrosionsexperten, die Mitglieder der neu geschaffenen Arbeitsgruppe Korrosion von Armierungseisen in Beton sind. Bei dieser ersten Sitzung wurde über die Ziele und das Vorgehen der Arbeitsgruppe diskutiert.

Die *Europäische Föderation Korrosion (EFC)* will über Korrosion und Korrosionsschutz im europäischen Rahmen informieren. Sie koordiniert damit die Arbeiten und Ideen der Korrosionsexperten aus den verschiedenen Ländern. Zu diesem Zweck werden Tagungen veranstaltet, wo die neuesten Forschungsergebnisse vorgetragen werden. So wird zum Beispiel diesen Herbst die EUROCOR 1988, vom 3. bis 8. Oktober 1988 im Brighton, England, stattfinden.

Die EFC setzt auf verschiedenen Gebieten Arbeitsgruppen ein. Einige seien im folgenden genannt:

- Korrosion in Meerwasser
- Mikrobiologische Korrosion
- Inhibitoren
- Surface science and the mechanisms of corrosion and protection
- Corrosion education

Bewehrungseisen in Beton werden bekanntlicherweise durch zwei verschiedene Mechanismen angegriffen:

Alter, karbonatisierter Beton, der nicht mehr alkalisch ist, kann bei mittleren Luftfeuchtigkeiten die Korrosion des Bewehrungseisens nicht mehr verhindern. Diese Korrosionsart tritt aber erst nach vielen Jahrzehnten auf und kann durch verschiedene Massnahmen bekämpft werden. Die zweite Korrosionsart tritt auf, wenn Beton höhere Mengen an Chlorid, zum Beispiel aus dem