MACHAKOS UNIVERSITY JOURNAL OF SCIENCE AND TECHNOLOGY, ISSN2707-6741 VOL. 2, ISSUE 3, JUNE 2021

Change point analysis in the generalized Pareto distribution

Mwelu Susan[1], A.G Waititu[1], P.N Mwita [2]

[1]Department of Statistics and Actuarial Sciences, Jomo Kenyatta University of Agriculture and Technology

[2]Department of Mathematics, Statistics and Actuarial Sciences, Machakos University

Abstract

The general goal of data statistical analysis is to find a near perfect translation to reality such that minimal information is lost in the approximating model. In this paper change point problem is viewed as a model selection problem where the point in time that model parameters change is estimated. This paper develops a change point estimator of the shape parameter of the generalized Pareto distribution which is shown to be consistent through simulations. The likelihood ratio test statistic based on the Kullback-Leibler divergence is used to detect a change point under the assumption that the model is correctly specified. The maximum likelihood estimation method is used to estimate the change point. The estimator is then used to detect a change point within extreme events with a climatic application in mind.

Keywords: Change point, Kullback-Leibler, Generalized Pareto distribution

INTRODUCTION

Time series data can be viewed as data that is recorded sequentially with reference to time. An entire data array can initially be analysed with the goal of finding a statistical model that can adequately describe the underlying data generating process through the estimation of model parameters. Since reality cannot be fully exhibited through a model, we seek to minimize information loss through an approximating model that gives a near perfect translation to reality.

In the study of time series, it is natural to assume time-shift invariance of the data probability distribution i.e. stationarity. However, some properties of time series data such as mean, variance or higher order moments may change with time. The basic assumption would be that estimated model parameters remain unchanged throughout time. Truong et al. (2018) states that may real world data are made up of consecutive regimes that are separated by abrupt changes. Prior to model estimation, the statistical hypothesis of stochastic homogeneity of the data can be checked for the purpose of parameter estimation in each segment of the data separately if any changes are detected. In its simplest form, change-point detection is the name given to the problem of estimating the point at which the statistical properties of a sequence of observations change (Killick & Eckley, 2014). The overall behaviour of observations can change over time due to internal systemic changes in distribution dynamics or due to external factors. Change-point analysis entails finding both the number and the location of such changes. The detection of change points has been increasingly taken into consideration as opposed to the classical stationary assumptions as change points can drastically alter inferences made from a time series.

Fundamental features in an extreme value analysis are captured by the tail behaviour. When the overall distribution changes, what happens to the mean may be different from what happens to the extremes at either ends of the distribution (Field et al., 2012). This guides the study to the changes in the tail behaviour of any given distribution which is characterized by the extreme values. Extreme values on the other hand can be summarized by the existing extreme value distributions. This means that by imposing a fixed parametric density form, then any changes in distribution parameters can be monitored. Changes observed in extremes could be linked to changes in the mean, variance or shape of the probability distribution or all of the three.

Rationale: Review of previous research works

EVT focuses on the tail distribution of time series data and provides flexible, simple parametric models for capturing tail related behaviors. By the use of asymptotic limit theorems, models that can be used to make inferences about the tails of a given underlying distribution are derived. Extreme values have been defined in two ways: block maxima whose limiting distribution is the generalized extreme value distribution and threshold exceedances whose limiting distribution is the generalized Pareto.

In the study of time series, it is natural to assume time-shift invariance of the data probability distribution i.e. stationarity. The standard EVT methods assume stationarity (Coles et al., 2001). In contrast, non-stationary processes are characterized by changes through time. In certain aspects such as environmental processes and climatology, the characteristics of the underlying process is often non-stationary and depends on changes in large-scale processes, seasonality or long-term trends (Rust et al., 2009; Cheng et al., 2014; De Paola et al., 2018). The presence of temporal dependence challenges the utility of standard EVT models. This implies that model parameters cease to be constants rather become functions of time, covariates or other underlying processes.

In its simplest form, change-point detection is the name given to the problem of estimating the point at which the statistical properties of a sequence of observations change (Killick & Eckley, 2014). Change-point analysis entails finding both the number and the location of such changes. The detection of change points has been increasingly taken into consideration as opposed to the classical stationary assumptions as change points can drastically alter inferences made from a time series. Generally, it is of interest to segment a time series into homogeneous segments for better informed inferences.

Jarušková and Rencová (2008) investigate change points in temperature extremes based on the assumption that the maxima and minima follow the generalized extreme value distribution. In order to decide whether a series is stationary then the principles of mathematical statistics lead to hypothesis testing. The null hypothesis corresponds to stationarity of the series whereas the alternative corresponds to the type of change that is being sought. Their test statistic based on the likelihood ratio considered a change in all the three parameters of the GEV but was modified to a trimmed maximum type to ensure that a change cannot occur at the beginning or end points of a time series. Since they required to compute the maximum likelihood estimates before and after all possible change points, asymptotic theory was used and the authors recommend that good parameter estimates are calculated from a sample size of at least fifty observations. They indicate that in practice it is better to choose an alternative hypothesis that corresponds better to the type of the change expected to occur. Practically, it may be difficult to clearly articulate the expected change as it is usually unknown.

According to Dierckx and Teugels (2010) extreme values of a distribution might be of greater importance than the mean values when performing change point analysis as the mean and variance do not often adequately describe the tail of a distribution. Their study is based on the likelihood approach of Csörgö and Horváth (1997) which can be adapted to extreme values. Through simulations, their study revealed that the test was more powerful for large sample sizes and when the position of the change point is not close to the sample end points. Changes in the tail distribution and variance can lead to more extreme events as compared to changes in only in the mean (Dupuis et al., 2015). Their work focused on changes in the tail behaviour based on multiple cross-sectional time series where they sought multi-year seasonal change points.

The distribution of the likelihood ratio statistic cannot be obtained in analytic form for small sample sizes and the ML estimates of the change point does not satisfy regularity conditions required to apply standard asymptotic likelihood ratio theory (Chen et al., 2017) . Instead they used a Bayesian approach to compare the models of no change point versus change point models that considered a change in both the scale and tail index parameters of the GPD. When assessing the accuracy of change point detection as well as model selection, the proposed models exhibited lower accuracy as when compared to cases where the main interest was change point detection regardless of the model selected. Since this points out to the possibility of a change point to be adequately detected regardless of an incorrectly chosen model, then model selection uncertainty becomes important.

Detection of change points is critical to statistical inference as a near perfect translation to reality is sought through model selection and parameter estimation. Stationarity in the strict sense, implies time-invariance of the distribution underlying the process. Stationarity is arguably a very strong assumption in many real-world applications as process characteristics evolve over time. Opposed to the assumption of stationarity, standard extreme value distributions have been modified to account for non-stationarity but this does not always imply that abrupt changes within the underlying process will be accounted for. Change points within a parametric setting can be attributed to change in the parameters of the underlying data distribution. Thus the main problem would be to find the point in time that the changes occur.

In section 3, the test statistic is constructed based on the Kullback-Leibler divergence introduced by Kullback and Leibler (1951) that will test whether the shape and/or scale parameters of the generalized Pareto distribution change over time. Simulation studies are carried out in section 4 with application of the change point estimator to real life data having a climatology context.

METHODOLOGY

The generalized Pareto distribution was introduced by Pickands III et al. (1975) to model exceedances over thresholds and is widely used in the analysis of extreme events, modelling of large insurance claims, hydrology, climatology and as a failure time distribution. The GPD can also be used in any situation in which the exponential distribution might be used but in which some robustness is required against heavier tailed or lighter tailed alternatives (Hosking & Wallis, 1987).

Definition 3.1. The Generalized Pareto distribution function is defined by;

$$H(y) = \begin{cases} 1 - \left(1 + \frac{\xi y}{\sigma}\right)^{-\frac{1}{\xi}}, & \xi \neq 0 \\ 1 - \exp\left(-\frac{y}{\sigma}\right), & \xi = 0 \end{cases} \quad (1)$$

Where,

$$y \in \begin{cases} [0, \infty), & \xi \geq 0 \\ \left[0, -\frac{\sigma}{\xi}\right], & \xi < 0 \end{cases}$$

More specifically, given that $X \sim GP(\sigma, \xi)$ then the probability density function is;

$$h(y) = \begin{cases} 1 - \left(1 + \frac{\xi y}{\sigma}\right)^{-\frac{1}{\xi}}, & \xi \neq 0 \\ 1 - \exp\left(-\frac{y}{\sigma}\right), & \xi = 0 \end{cases} \quad (2)$$

The parameter ξ is the extreme value index and its sign is the dominant factor in describing the tail of the underlying distribution. The larger the ξ the heavier the tail. σ is referred to as the scale parameter and characterizes the spread of the distribution.

The GP distribution contains a number of specific distributions under its parametrization. When $\xi > 0$ the corresponding distribution is the usual heavy tailed Pareto distribution. If $\xi < 0$ the corresponding distribution is a type II Pareto distribution whereas $\xi = 0$ yields the exponential distribution.

Single change point estimation

Consider a non-stationary process $X = \{x_1, \ldots, x_n\}$. A change point is said to occur when there exists a time $\tau \in \{2, \ldots, n-1\}$ such that the statistical properties of $\{x_1, \ldots, x_\tau\}$. and $\{x_{\tau+1}, \ldots, x_n\}$. are different. Then the hypothesis of no change in model parameter(s) versus change in model parameter(s) at an unknown time τ would be stated as:

$$\begin{aligned} H_0 &: \theta_1 = \theta_2 = \ldots = \theta_n \; versus \\ H_1 &: \theta_1 = \cdots = \theta_\tau \neq \theta_{\tau+1} = \ldots = \theta_n \end{aligned} \quad (3)$$

At a given level of significance, if the null hypothesis is rejected, then the process X is said to be locally piecewise-stationary and can be approximated by a sequence of stationary processes that may share certain features such as the general functional form of the distribution F. Assuming that the time of change τ is unknown and there are K $(1 < K < n)$ admissible change points, then the test statistic $T_\tau = T_{max} = \max k \in K\, T_k$. Then the estimated time of change τ would be the value of k that maximizes the test statistic T_k. This would naturally mean that the null hypothesis is rejected if T_τ is large.

The hypothesis being tested is as in equation 4

$$\begin{aligned} &H_0 : X_t \sim GP(\sigma, \xi) \\ &\text{Against} \\ &H_1 : X_t \sim GP(\sigma_1, \xi_1) \; t \leq \tau \end{aligned} \quad (4)$$

$: X_t \sim GP(\sigma_2, \xi_2)\ t > \tau$

Where $\xi_1 \neq \xi_2$ and $\sigma_1 \neq \sigma_2$ are unknown before and after the change point τ

Information measures provide descriptions of the long term behavior of random processes (Garrido, 2009). For this reason, measures of distance between probability distributions are central to the problems of inference and discrimination. Statistical inference starts with a set of observations $\{x_1, \dots, x_n\}$ which are assumed to have been generated by an unknown model, say $p(x)$. By means of estimation methods an approximated model, say $q(x)$, is obtained. The overall aim would be to have $q(x)$ as similar as possible to $p(x)$. Then probabilistic distances (divergences) would be computed to assess how 'close' the two probability distributions are from one another.

Definition: Let P and Q be two finite distributions on X with probability density functions $p(x)$ and $q(x)$ respectively. The Kullback-Leibler divergence of p relative to$q$, a measure of information lost when $p(x)$ is used to approximate $q(x)$ is defined by (Kullback & Leibler, 1951)

$$\begin{aligned} I(p||q) &= \int_x p(x) \log \frac{p(x)}{q(x)} dx \\ &= \int_x p(x) \log p(x) dx - \int_x p(x) \log q(x) dx \\ &= E_{p(x)}[\log p(x)] - E_{p(x)}[\log q(x)] \end{aligned} \tag{5}$$

where E_{p_i} denotes the expectation operator over the distribution p_i. x represents the random vector with density p and q. Despite not being a true distance, this expectation provides relevant information on how close q is from p. Kullback and Leibler (1951) coined the term "directed divergence" to distinguish it from the divergence defined by 6 which is a symmetric measure as opposed to 5 which is asymmetric and fails to satisfy the triangular inequality since $I(p||q) \neq I(q||p)$.

$$D(p||q) = I(p||q) + I(q||p) \tag{6}$$

The KL information is applicable for both continuous and discrete distributions. It is also known as relative entropy, informational divergence, or information for discrimination. Being a one dimensional measure, it is easy to comprehend and thus forms a deep theoretical basis for data-based model selection as it takes into account the entire range of the distribution.

Consider two GP densities $f_1(x)$ and $f_2(x)$ with density function as in 2 with scale and shape parameters $\xi_1 \neq \xi_2$ and $\sigma_1 \neq \sigma_2$ respectively. The KLD can be expressed as a function of the two distribution

$$\begin{aligned} I(f_1||f_2) &= \int_x f_1(x)\log\frac{f_1(x)}{f_2(x)}dx \\ &= \int_x f_1(x)\log f_1(x)dx - \int_x f_1(x)\log f_2(x)dx \\ &= E_{f_1(x)}[\log f_1(x)] - E_{f_1(x)}[\log f_2(x)] \end{aligned} \tag{7}$$

$$\begin{aligned} I(f_2||f_1) &= \int_x f_2(x)\log\frac{f_2(x)}{f_1(x)}dx \\ &= \int_x f_2(x)\log f_2(x)dx - \int_x f_2(x)\log f_1(x)dx \\ &= E_{f_2(x)}[\log f_2(x)] - E_{f_2(x)}[\log f_1(x)] \end{aligned} \tag{8}$$

Equation 9 gives one of the properties of the GP distribution (Embrechts et al., 2013)

$$E\left[\log\left(1+\frac{\xi y}{\sigma}\right)\right] = \xi^k k!, \quad k \in N \tag{9}$$

An application of property 9, numerical computations and methods of integration equation 7 can be reduced to

$$\begin{aligned} &log\left(\frac{1}{\sigma_1}\right) - \left(\frac{1}{\xi_1}+1\right)E\left[log\left(1+\frac{\xi_1}{\sigma_1}x\right)\right] - \log\left(\frac{1}{\sigma_2}\right) - \left(\frac{1}{\xi_2}+1\right)E\left[log\left(1+\frac{\xi_2}{\sigma_2}x\right)\right] \\ &= \ log\left(\frac{1}{\sigma_1}\right) - \left(\frac{1}{\xi_1}+1\right)\xi_1 - \log\left(\frac{1}{\sigma_2}\right) - \left(\frac{1}{\xi_2}+1\right)\int_x \log\left(1+\frac{\xi_2}{\sigma_2}x\right)f_{1(x)}dx \\ &= \log\left(\frac{1}{\sigma_1}\right) - log\left(\frac{1}{\sigma_2}\right) - (1+\xi_1) - \left(\frac{1}{\xi_2}+1\right)\frac{\sigma_1}{\sigma_2}\xi_1\int_x \left(1+\frac{\xi_1}{\sigma_1}x\right)^{-\frac{1}{\xi_1}}\left(\frac{1}{\xi_2}+1\right)^{-1} dx \end{aligned} \tag{10}$$

The KL divergence as in 10 is a function of the parameters of the two densities $f_1(x)$ and $f_2(x)$. The maximum likelihood estimation method is adapted to estimate the parameters as they have desirable characteristics such as consistency, efficiency and asymptotic normality.

Suppose that $\{x_1, \ldots x_{\tau,}, x_{\tau+1}, \ldots, x_n\}$ follow a GP distribution with parameters $\theta_i = (\sigma_i, \xi_i)\,, i = 1{,}2$.

Then the log-likelihood functions are as in equation 11 for the likelihood estimators $(\sigma_\tau^-, \xi_\tau^-)$ and $(\sigma_\tau^+, \xi_\tau^+)$ for the data sets $x_1, \dots x_{\tau,}$ and $x_{\tau+1}, \dots, x_n$ respectively. The log-likelihood functions are maximized with respect to the parameters to obtain the estimates before and after the change point τ.

$$L_\tau^-(\hat{\theta}) = -\tau \log \hat{\sigma}_\tau^- \left(\frac{1}{\xi_\tau^-} + 1\right) \sum_{i=1}^{\tau} \log\left(1 + \frac{\hat{\xi}_\tau^-}{\hat{\sigma}_\tau^-} x_i\right)$$
$$L_\tau^+(\hat{\theta}) = -(n-\tau) \log \hat{\sigma}_\tau^+ \left(\frac{1}{\xi_\tau^+} + 1\right) \sum_{i=\tau+1}^{n} \log\left(1 + \frac{\hat{\xi}_\tau^+}{\hat{\sigma}_\tau^+} x_i\right) \qquad (11)$$

Although asymptotically efficient, the MLEs are computationally difficult and can have convergence problems (Hosking & Wallis, 1987). The ML estimates sometimes cannot be obtained for the generalized Pareto distribution as a result of nonexistence of a local maximum of the likelihood function. For instance, the ML estimates do not exist for $\xi \geq 1$ because the likelihood function has no local maximum. For $\xi \leq -0.5$, the ML estimates do not exist asymptotically since $var(x) = \infty$ (Zhang, 2007). For $\xi > -0.5$, the maximum regularity conditions are achieved and thus the maximum likelihood estimates are asymptotically normally distributed, consistent and asymptotically efficient.

RESULTS

Consider a simulation from the GP density with parameters (1, 0.1) and (3, 0.2) for the scale and shape respectively with n = 500 and τ = 250. The first 250 data come from the GP (1, 0.1) and the next 250 come from the GP (3, 0.2). These values are taken to represent excesses from two different densities with varying parameters. Setting the shape parameter at $\xi_1 = 0.1$ then ξ_2 varies across the values 0.2, 0.3 and 0.4. The change point as estimated by equation 6 is given at the time when the maximum value of the divergence is observed.

Some of the simulation results are shown in figure 2. For the simulation studies the values of $\tau \in \left(\frac{n}{4}, \frac{n}{2}, \frac{3n}{4}\right)$

are varied corresponding considering different locations of the change point for different sample sizes (200, 500, 1000). The divergence between the two samples is estimated and the detected change point results presented in table 1.

n	τ	$\xi_2 = 0.2$ $\hat{\tau}$	$\xi_2 = 0.3$ $\hat{\tau}$	$\xi_2 = 0.4$ $\hat{\tau}$
500	150	167	162	142
	250	234	237	239
	350	369	356	353
1000	250	171	174	180

	500	428	424	431
	750	720	724	728
2000	500	310	303	300
	1000	821	795	786

Table 1: Estimation by simulation with 500 replications

n	τ	$\xi_2 = 0.2$	$\xi_2 = 0.3$	$\xi_2 = 0.4$
500	150	3125.85	3146.48	3182.77
	250	8324.46	8130.28	7672.63
1000	250	50275.75	30837.72	30312.39
	500	50963.03	42619.13	41780.05
2000	500	171739.4	177772.9	179237.1
	1000	176082.5	195395.4	200240.9

Table 2: Mean squared errors (MSE)

Table 2 gives the mean squared errors used to assess the performance of the estimator considering three generalized Pareto distributions with different sample sizes. Regardless of the location of the change point, the overall performance of the estimator improves as the change in the shape parameter becomes larger for smaller samples. This can be seen from the trend in the values of the mean squared errors. As the sample size gets larger the estimator's sensitivity to small changes in the shape parameter improves. This implies that very large sample sizes are required to detect very small changes in non-negative shape parameters.

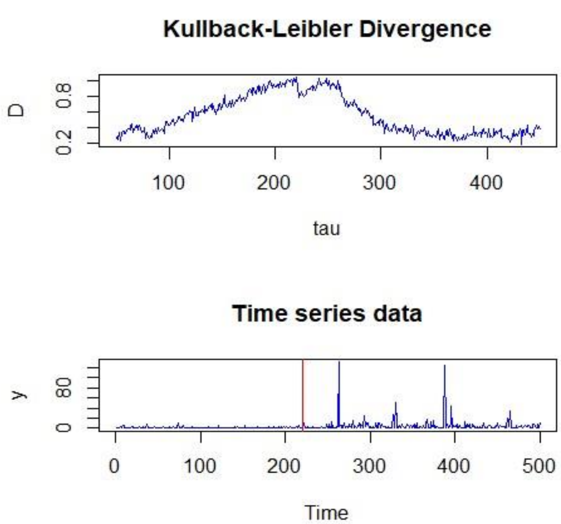

Figure 1: n = 500, $\tau = 250, \hat{\tau} = 221$

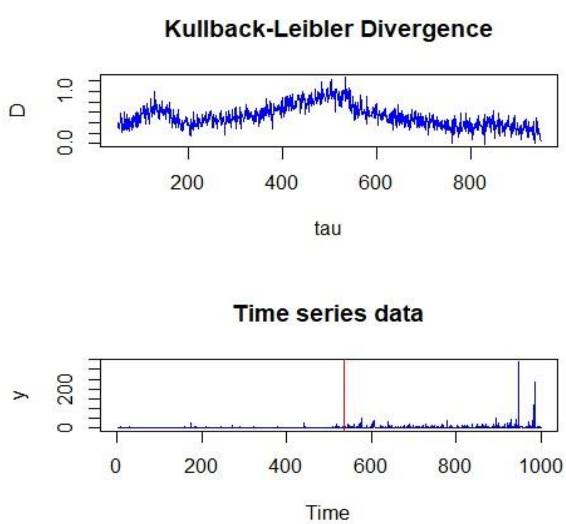

Figure 2: n = 1000, $\tau = 500, \hat{\tau} = 535$

Application to data

Before fitting the GP density function to data, it is first necessary to choose a threshold. It is important to choose a sufficiently high threshold in order that the theoretical justification applies thereby reducing bias. However, the higher the threshold, the fewer available data remain. Thus, it is important to choose the threshold low enough in order to reduce the variance of the estimates. The Fort Collins precipitation data as shown in figure 3 is used. The data contains daily precipitation amount records over the period 1900 to 1999. A threshold value of 0.4 is selected and extremes classified as the values above this. It is assumed that the values are independently and identically distributed. The exceedances are considered to follow the generalized Pareto distribution. The KLD values are computed and shown in figure 4

The change point estimator detects a change at point $\hat{\tau}$= 979 (March, 14 1996). On July 28, 1997, an extreme flood disaster hit Fort Collins, Colorado with the heaviest rains ever documented in an urbanized area of the state. It was one of the major urban floods of recent years in the United States and was labelled a "500- year event" by the media, causing major impacts on the city and its people (Grigg et al., 1999). The change point could thus be accounted for by this extreme flooding event. Figure 5 shows the extreme precipitation

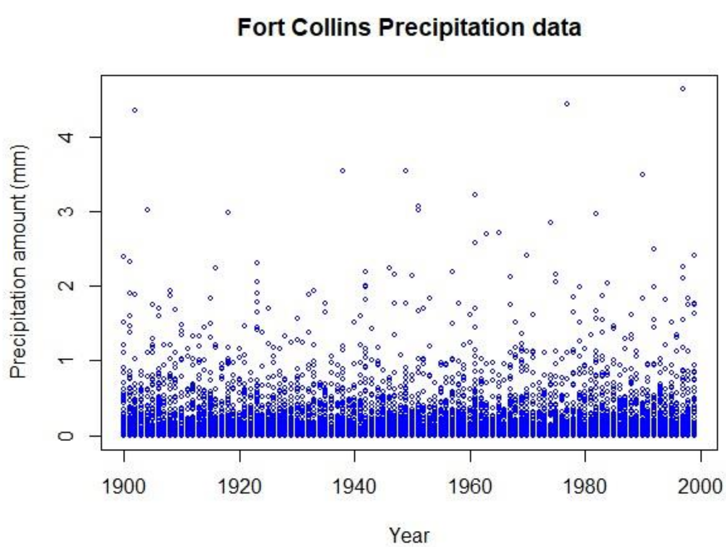

Figure 3: Fort Collins precipitation time series

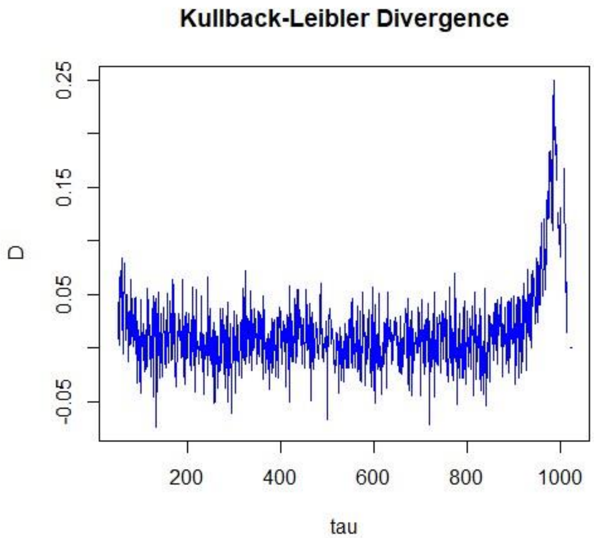

Figure 4: Kullback- Leibler divergence

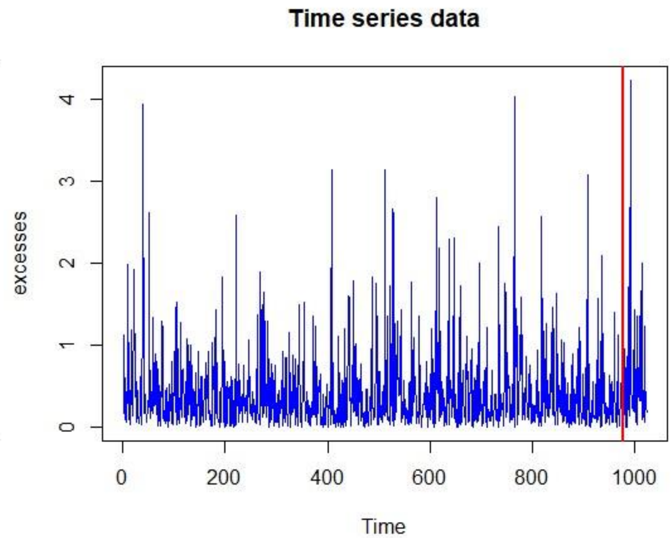

Figure 5: Excesses change point estimate

data with a superimposed line (red) indicating the point in time when the change point was detected. Looking at the values of the estimated parameters of the GP before and after the change point an increase in the scale parameter is accompanied by a decrease in the shape.

DISCUSSION AND CONCLUSION

The Kullback-Leibler divergence has been used to distinguish between two distributions belonging to the same family. A change point is then defined as the point in time where the distribution parameters change characterized by maximum divergence. This change can be attributed to many factors including internal and systemic changes. Since fundamental features of extreme values can be captured by the tail behavior described by the generalized Pareto distribution, then the method can be used explicitly to examine the parameter changes. The method has shown difficulties that are present when estimating small changes in the shape parameter given small sample sizes. This then implies that very large samples are crucial in estimating small changes since the performance and sensitivity of the estimator improves with larger samples. Hence, the estimator is seen to be consistent with respect to the magnitude of change. The method is tested within a simulation setting as well as real life example and works quite well.

REFERENCES

Chen, S., Li, Y., Kim, J., & Kim, S. W. (2017). Bayesian change point analysis for extreme daily precipitation. *International Journal of Climatology*, *37*(7), 3123–3137.

Cheng, L., AghaKouchak, A., Gilleland, E., & Katz, R. W. (2014). Non-stationary extreme value analysis in a changing climate. *Climatic change*, *127*(2), 353–369.

Coles, S., Bawa, J., Trenner, L., & Dorazio, P. (2001). *An introduction to statistical modeling of extreme values* (Vol. 208). Springer.

Csörgö, M., & Horváth, L. (1997). *Limit theorems in change-point analysis* (Vol. 18). John

Wiley & Sons Inc.

De Paola, F., Giugni, M., Pugliese, F., Annis, A., & Nardi, F. (2018). Gev parameter estimation and stationary vs. non-stationary analysis of extreme rainfall in african test cities. *Hydrology*, *5*(2), 28.

Dierckx, G., & Teugels, J. L. (2010). Change point analysis of extreme values. *Environmetrics*, *21*(7-8), 661–686.

Dupuis, D., Sun, Y., & Wang, H. J. (2015). Detecting change-points in extremes. *Statistics and Its Interface*.

Embrechts, P., Klüppelberg, C., & Mikosch, T. (2013). *Modelling extremal events: for insurance and finance* (Vol. 33). Springer Science & Business Media.

Field, C. B., Barros, V., Stocker, T., Qin, D., Dokken, D., Ebi, K., ... others (2012). Ipcc, 2012: Managing the risks of extreme events and disasters to advance climate change adaptation. a special report of working groups i and ii of the intergovernmental panel on climate change. *Cambridge University Press, Cambridge, UK, and New York, NY,*

USA, *30*(11), 7575–7613.

Garrido, A. (2009). About some properties of the kullback-leibler divergence. *Advanced*

Modeling and Optimization, *11*, 4.

Grigg, N. S., Doesken, N. J., Frick, D. M., Grimm, M., Hilmes, M., McKee, T. B., & Oltjenbruns, K. A. (1999). Fort collins flood 1997: Comprehensive view of an extreme event. *Journal of water resources planning and management*, *125*(5), 255–262.

Hosking, J. R., & Wallis, J. R. (1987). Parameter and quantile estimation for the generalized pareto distribution. *Technometrics*, *29*(3), 339–349.

Jarušková, D., & Rencová, M. (2008). Analysis of annual maximal and minimal temperatures for some european cities by change point methods. *Environmetrics: The official journal of the International Environmetrics Society*, *19*(3), 221–233.

Killick, R., & Eckley, I. (2014). changepoint: An r package for changepoint analysis. *Journal of statistical software*, *58*(3), 1–19.

Kullback, S., & Leibler, R. A. (1951). On information and sufficiency. *The annals of mathematical statistics*, *22*(1), 79–86.

Pickands III, J., et al. (1975). Statistical inference using extreme order statistics. *the Annals of Statistics*, *3*(1), 119–131.

Rust, H., Maraun, D., & Osborn, T. (2009). Modelling seasonality in extreme precipitation.

The European Physical Journal Special Topics, *174*(1), 99–111.

Truong, C., Oudre, L., & Vayatis, N. (2018). A review of change point detection methods. *arXiv preprint arXiv:1801.00718*.

Zhang, J. (2007). Likelihood moment estimation for the generalized pareto distribution.

Australian & New Zealand Journal of Statistics, *49*(1), 69–77.