UDC 519.246.5

Estimation of heavy tail dependence based on copulas for the precipitation analysis

N. D. Rassakhan*, E. Yu. Shchetinin†

* *Department of Applied Mathematics*
Moscow State Technology University "STANKIN"
3a Vadkovsky Ln., Moscow, 127055, Russia

† *FGU "All-Russian research institute on problems of civil defence and emergencies" of Emergency Contorl Ministry of Russia*
7 Davydkovskaya St., Moscow, 121352, Russia

Email: rassahan@gmail.com, riviera-molto@mail.ru

Consideration of tail dependence is a very important part of risk analysis in many applied sciences that is measured in order to estimate the risk of simultaneous extreme events. Usually the tail dependence coefficient is the measurement in question. Pearson correlation coefficient unfortunately is not a suitable measure for estimating dependencies between two quantities in the context of simultaneous occurrence of extreme events when these events are of interest for the researcher because it takes extreme events into account with the same weight as it takes "normal" events although extreme events' dependence may slightly differ.

Present work emphasizes the importance of taking into consideration tail dependencies in bivariate statistical analysis using copulas. Due to increasing frequency of environmental cataclysms the issue of analyzing risks (e.g. economic losses) and their consequences comes to the fore. Moreover, researchers should take into account consequences of their joint occurrence. Three non-parametric estimators of tail dependence coefficients were compared in order to estimate correlation between daily cumulative rainfall totals recorded in central European part of Russia. Major part of existing estimators depends on threshold k and thus there is a trade-off between variance and bias during the calculation of the best value for k. For balancing an algorithm is presented that is based on using moving average and then searching the "stable" part of tail dependence coefficient. Estimate of tail dependence coefficient is assumed to be equal to mean value on the "stable" part.

Key words and phrases: extreme value theory, spatial modelling, extreme precipitation, spatial structures of statistical dependence, tail dependence coefficient.

1. Introduction

One of the most important parts of multivariate extreme value analysis is exploration of extremal dependencies; basically tail dependence coefficient is used for this purpose. For bivariate vector (X_1, X_2) upper tail dependence coefficient has the following form:

$$\lambda_U = \lim P\left(F_1(X_1) > t | F_2(X_2) > t\right), \quad t \to 1^-, \tag{1}$$

where F_1, F_2 are distribution functions of random variables X_1, X_2 respectively, $0 < t \leq 1$ is the threshold. We can define lower tail dependence coefficient in the same way:

$$\lambda_L = \lim P\left(F_1(X_1) < t | F_2(X_2) < t\right), \quad t \to 0^+. \tag{2}$$

Using the copula function [] equations (1),(2) can be written in alternative form:

$$\lambda_U = \lim_{t \to 1^-} \frac{1 - 2t + C(t,t)}{1 - t}; \quad \lambda_L = \lim_{t \to 0^+} \frac{C(t,t)}{t}.$$

2. Main section

Tail dependence coefficient estimation methods are essential analysis tools for extremal precipitation structures that are studied in this paper. Onward we will describe some of them. Foremost such estimators are non-parametric estimators based on empirical copula $C^{(n)}(u,v)$ concept:

$$C^{(n)}(u,v) = F^{(n)}\left(F^{-1}_{(n)1}(u), F^{-1}_{(n)2}(v)\right),$$

where $F_{(n)}(\cdot)$ is empirical distribution function.

Let $\left(X_1^{(1)}, X_2^{(1)}\right), \dots, \left(X_1^{(n)}, X_2^{(n)}\right)$ be independent identically distributed copies of bivariate random vector (X_1, X_2). Using the fact that their joint distribution function is defined as

$$F(x_1,x_2) = P\left(F_1(X_1) \le F_1(x_1), F_2(X_2) \le F_2(x_2)\right) = C\left(F_1(x_1), F_2(x_2)\right)$$

and equation

$$\lambda = 2 - \lim_{t\to 1^-} \frac{1-C(t,t)}{1-t} = 2 - \lim_{t\to 0+} \frac{1-C(1-t,1-t)}{t}, \tag{3}$$

an estimator for upper tail dependence (1) coefficient can be derived:

$$\hat{\lambda}^{SEC} \equiv \hat{\lambda}^{SEC}(k) = 2 - \frac{1-\hat{C}\left(1-\frac{k}{n}, 1-\frac{k}{n}\right)}{\frac{k}{n}}, \quad 1 \le k < n. \tag{4}$$

Then taking into consideration $\log(1-t) \sim -t, \quad t \approx 0$ next estimator can be obtained:

$$\hat{\lambda}^{LOG} \equiv \hat{\lambda}^{LOG}(k) = 2 - \frac{\log \hat{C}\left(1-\frac{k}{n}, 1-\frac{k}{n}\right)}{\log\left(1-\frac{k}{n}\right)}, \quad 1 \le k < n.$$

Here $\hat{C}$ denotes empirical copula defined as

$$\hat{C}\left(1-\frac{k}{n}, 1-\frac{k}{n}\right) = \frac{1}{n}\sum_{i=1}^{n} 1_{\left\{\hat{F}_1\left(X_1^{(i)}\right)\le 1-\frac{k}{n}, \hat{F}_2\left(X_2^{(i)}\right)\le 1-\frac{k}{n}\right\}}, \quad 1 \le k < n,$$

where 1 is indicator function and $\hat{F}_j, \quad j = 1,2$ are marginal empirical distribution functions of X_1 and X_2 respectively. To increase estimation's accuracy it is considered that

$$\hat{F}_j(u) = \frac{1}{n+1}\sum_{i=1}^{n} 1_{\left\{X_j^{(i)} \le u\right\}}, \quad j = 1,2.$$

Note that both estimators depend on choice of threshold k and thereafter k^{th} order statistic. It is very important to choose right value for k while not being an easy task due to trade-off variance and bias.

Another estimator for upper tail dependence coefficient is suggested in works [],[]:

$$\hat{\lambda}^{CFG} = 2 - 2\left[\frac{1}{n}\sum_{i=1}^{n} \log\left\{\frac{\sqrt{\log\frac{1}{X_1} log \frac{1}{X_2}}}{\log\frac{1}{\max(X_1,X_2)^2}}\right\}\right].$$

Main advantage of this equation is that $\hat{\lambda}$ doesn't depend on k. However, copula $C(X_1, X_2)$ must be well approximated with extreme-value copulas for correctness of the estimator.

It follows from the equations (3), (4) that estimators depend on choice of threshold k choice of which is determined by balancing variance and bias for estimator according to stability theorem for λ_U []. Increasing the value of k leads to reduction of bias and increase in variance; it goes the same the other way around. For big enough data sample size n balance between bias and variance is described by the "stable" part of λ_U plot. An algorithm for finding this "stable" part is presented in paper []:

1. Empirical estimation is smoothed with moving average filter which window size is equal to $b = int(0.05n)$. Sequence $\hat{\lambda}_1, \ldots, \hat{\lambda}_{n-2b}$ is obtained as a result.
2. Now we can find vector $\left(\hat{\lambda}_k, \ldots, \hat{\lambda}_{k+m-1}\right)$, $k = 1, \ldots, n-2b+m-1$, $m = int\left(\sqrt{n-2b}\right)$ from the sequence $\hat{\lambda}_1, \ldots, \hat{\lambda}_{n-2b}$ by a sequential search.
3. If the current vector satisfies

$$\sum_{i=k+1}^{k+m-1} |\bar{\lambda_i} - \bar{\lambda_k}| \leq 2\sigma,$$

where σ is standard deviation of $\hat{\lambda}_1, \ldots, \hat{\lambda}_{n-2b}$ then final evaluation takes the form of

$$\lambda_U = \frac{1}{m} \sum_{i=1}^{m} \bar{\lambda}_{k+i-1}.$$

If the condition is not satisfied after sequential searching, then

$$\lambda_U = 0.$$

Implementation of this algorithm is presented in Fig.2.

In this study precipitation data of the All-Russian Research Institute of Hydrometeorological Information - the World Data Center of the Russian Federation is used, which is daily precipitation in 14 cities of the European part of Russia. The data is freely available on the website `http://aisori.meteo.ru/ClimateR` and is represented by a set of tables (a separate table for each city); each table contains daily rainfall value for the period 1966-2016 years.

3. Conclusions

This paper highlights the importance of taking into account the tail-dependence coefficient in the context of multivariate frequency analysis using copulas. The three following nonparametric estimators, have been compared. The aim of this comparison was to choose the best estimator in the context of our application. No estimator works in every case. Specifically they show a poor performance in the case of tail independence. It is therefore important to pursue research in this field.

Most of the nonparametric estimators have to deal with the choice of the number k of order statistics to be considered in the production of an estimate. This is not an easy task since it requires a tradeoff between variance and bias (small values of k cause large variance and large values of k increase the bias). An optimal choice of k that leads to the smallest mean squared error is difficult to derive and, in practice, this is frequently solved through intensive simulation studies.

Since the nonparametric estimators yield a characteristic plateau while plotting the estimates for successive k Frahm et al. [10] introduced a simple plateau finding algorithm after smoothing the latter plot by some box kernel in order to find the optimal threshold k. Here we have applied this heuristic procedure to estimators of the TDC in (1). Since this very simple heuristic procedure revealed some potential, we intend

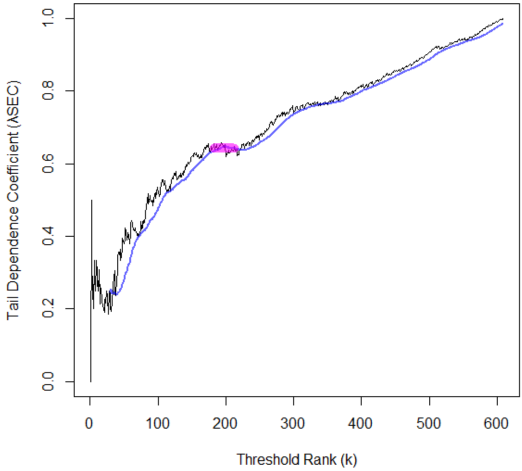

Figure 1. Some text

to develop it further and compare with other heuristic methods like, for instance, the graphical method in de Sousa and Michailidis [20] and bootstrap methods (see, e.g., Peng and Qi [21] 2008 and Gomes and Oliveira [22])

References

1. L. Parker, S. M. Christensen, MathTensor: a system for doing tensor analysis by computer, Addison-Wesley, 1994.
2. W. T. Jones, R. J. Fogelin, The Twentieth Century to Quine and Derrida, A History of Western Philosophy, Harcourt Brace College Publishers, 1997.
3. G. M. Sheldrick, A Short History of SHELXL, International Union of Crystallography and Oxford University Press, 2006.
4. A. J. Arduengo, III, R. L. Harlow, M. Kline, A stable crystalline carbene, J. Am. Chem. Soc. 113 (1) 361–363. doi:10.1021/ja00001a054.
5. G. Booth, J. Chatt, The reactions of carbon monoxide and nitric oxide with tertiary phosphine complexes of iron(II), cobalt(II), and nickel(II), J. Chem. Soc. 2099–2106. doi:10.1039/JR9620002099.
6. E. Hope, J. Bennett, A. Stuart, Fluorous zirconium phosphonates: novel inorganic supports for catalysis, in: Pacifichem (International Chemical Congress of Pacific Basin Societies), no. 961, Pacific Basin Chemical Societies.

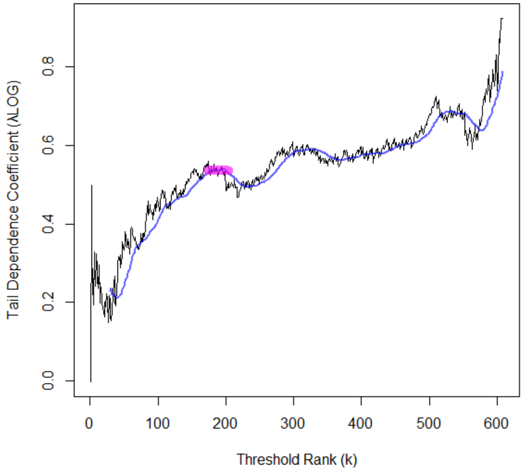

Figure 2. Some text

УДК 519.246.5

Шаблон оформления рукописи доклада на конференцию «ITTMM»

Н. Д. Рассахан*, Е. Ю. Щетинин†

* *Кафедра прикладной математики*
Московский государственный технологический университет "Станкин"
Вадковский пер., д.3а, Москва, Россия, 127055
† *Всероссийский научно-исследовательский институт*
по проблемам гражданской обороны и чрезвычайных ситуаций МЧС России
ул. Давыдковская, д. 7, Москва, Россия, 121352

Email: rassahan@gmail.com, riviera-molto@mail.ru

Измерение хвостовой зависимости является важной задачей во многих прикладных науках для оценивания риска совместного наступления экстремальных событий. Обычно мерой зависимости является коэффициент хвостовой зависимости. Корреляция Пирсона не является подходящей мерой для оценивания зависимости двух величин в контексте совместного наступления экстремальных событий, когда они представляют интерес для исследователя, так как она учитывает экстремальные события так же (с тем же весом), что и "рядовые"события, хотя зависимость между экстремальными событиями может сильно отличаться от общей картины.

Данная работа подчеркивает важность учитывания хвостовой зависимости в контексте двумерного анализа при помощи копул. В связи с учащающимися природными катаклизмами резко вставет вопрос об оценивании разоичных рисков (в т.ч. экономических) и последствий их совместного наступления с учетом пространственных связей

между наблюдениями. Сравниваются 3 непараметрические оценки коэффициента хвостой зависимости для оценки зависимости между ежедневными наблюдениямми осадков в городах Европейской части России. Большинство существующих оценок зависит от порога k и, следовательно, при выборе используемого значения k происходит трейдофф между смещением и вариацией. Для установаления баланса в работе представлен алгоритм, основанный на использовании скользящего среднего и поиска "стабильного участка"коэффициента хвостовой зависимости. Именно среднее значение на "стабильном"участке и принимается за значения оценки коэффициента хвостовой зависимости.

Ключевые слова: теория экстремальных величин, пространственное моделирование, экстремальные осадки, пространственные структуры статистической зависимости, коэффициент хвостовой зависимости.