

**МИНОБРНАУКИ РОССИИ**

**федеральное государственное бюджетное образовательное учреждение**

**высшего образования**

**«Московский государственный технологический университет «СТАНКИН»**

**(ФГБОУ ВО «МГТУ «СТАНКИН»)**

**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

Отчет

о выполнении научно-исследовательской работы

МАГИСТРАНТ 1 курса

РАССАХАН НИКИТА ДМИТРИЕВИЧ

|  |  |  |
| --- | --- | --- |
| Магистрант | \_\_\_\_\_\_\_\_\_\_\_\_\_ | Рассахан Н.Д. |
| Научный руководитель | \_\_\_\_\_\_\_\_\_\_\_\_\_ | доцент Красикова Е.М. |
| Заведующий кафедрой | \_\_\_\_\_\_\_\_\_\_\_\_\_ | к.ф.-м.н., профессор Уварова Л.А. |

Москва

2017

**Оглавление**

[**Введение** 3](#_Toc502077213)

[**Основная часть** 3](#_Toc502077214)

[Масштабная инвариантность 4](#_Toc502077215)

[Асимптотическая масштабная инвариантность 5](#_Toc502077216)

[**Одномерная теория экстремальных величин** 5](#_Toc502077217)

[Оценка Хилла для параметра формы 6](#_Toc502077218)

[Оценка Деккерса-Эйнмаля для параметра формы 6](#_Toc502077219)

[Оценка Пикендса для параметра формы 6](#_Toc502077220)

[**Программный код для нахождения хвостовых индексов** 7](#_Toc502077221)

[**Пример визуализации результатов** 10](#_Toc502077222)

[**Заключение** 11](#_Toc502077223)

Введение

В рамках осеннего семестра 2017 года, соответствующего первому семестру обучения в магистратуре, основными целями проводимой научно-исследовательской работы стали: актуализация информации из выпускной квалификационной работы бакалавра, поиск перспективных направлений развития исследуемой проблемы, а также закрепление полученных результатов в виде научной публикации.

Продолжается работа над материалами бакалаврского диплома с их дальнейшим расширением до магистерской работы. В рамках данного семестра внимание было уделено одномерной теории экстремальных величин, в частности, более подробному изучению тяжелохвостых распределений и различных оценок хвостовых индексов.

Была разработана программная реализация нахождения оценок хвостовых индексов, примененная к рассматриваемым данным по осадкам в ряде городов Европейской части России. В дальнейшем программный код можно будет интегрировать в уже существующие наработки.

Основная часть

В теории вероятностей тяжелохвостые распределения - это распределения вероятностей, чьи хвосты экспоненциально неограничены, то есть те распределения, чьи хвосты тяжелее, чем у экспоненциального распределения. В большинстве приложений интерес представляет правый хвост распределения, но распределение может и иметь и тяжелый левый хвост, или даже оба хвоста могут быть тяжелыми. Понятие тяжелохвостости не является до конца устоявшимся, поэтому иногда (но достаточно редко) к тяжелохвостым относят те распределения, хвосты которых тяжелее хвостов нормального распределения.

Основные распределения с тяжелыми хвостами:

* Распределение Парето
* Лог-нормальное распределение
* Распределение Леви
* Распределение Вейбулла с положительным параметром формы
* Распределение Бурра
* Распределение Стьюдента
* Распределение Коши

Масштабная инвариантность

F масштабно инвариантно, если существуют такие и , что

для таких, что . Выражение выше называют изменением масштаба.

*Теорема*. Распределение масштабно инвариантно только если это распределение Парето.

*Пример*.

Асимптотическая масштабная инвариантность

F асимптотически масштабно инвариантно, если существует непрерывная ограниченная функция такая, что

Одномерная теория экстремальных величин

Теория экстремальных величин начинает отсчет своего основания с теоремы Фишера и Типпета о существовании предельного распределения последовательности нормализованных максимумов последовательности независимых одинаково распределенных случайных величин. Если такое распределение  существует и является невырожденным, тогда оно должно удовлетворять требованиям распределений устойчивых максимумов, т.е.

 (1)

для n > 1 и некоторых  и , . Такие распределения имеют следующий вид

 , (2)

где a + = max(a,0) , параметр положения −∞ < µ < ∞, параметр масштаба σ > 0, параметр формы −∞ < ξ < ∞; случай ξ = 0 интерпретируется как предельный ξ → 0. Распределение  называется также обобщенным распределением экстремальных величин, т.к. включает распределение Вейбулла (ξ < 0), Гумбеля (ξ = 0) и Фреше (ξ > 0).

С этим результатом тесно связан и другой подход к моделированию порядковых статистик, известный как пороговый. Следуя теории, разработанной Пикендсом, при известных условиях выбора порога u, распределение надпороговых значений может быть аппроксимировано обобщенным распределением Парето *(GPD)*

, (3)

 , где τ +ξu > 0, −∞ < ξ < ∞. Наиболее важным в рамках наших исследований является параметр  и различные его оценки в связи с тем, что он не только показывает принадлежность к тому или иному классу распределений (2), но и характеризует такое широко известное явление как тяжелохвостость эмпирических распределений. Далее мы рассмотрим некоторые наиболее известные оценки .

Оценка Хилла для параметра формы

Пусть  - последовательность независимых и одинаково распределенных случайных величин, удовлетворяющих (1). Пусть также  – некоторый высокий порог для статистики. Оценка Хилла, основана на том, что при ,  имеет вид

. (4)

Заметим, что при , ,  при .

Оценка Деккерса-Эйнмаля для параметра формы

Деккерс, Эйнмаль и др. предложили более общее выражение для оценки 

 , (5)

где . Тогда при условиях при  одновременно с ,  справедливо .

Оценка Пикендса для параметра формы

Наконец, известна оценка Пикендса

 , (6)

так что при  одновременно с , .

Программный код для нахождения хвостовых индексов

Программный код на языке R:

readdata <- function (cline) {

df2 <- read.csv(cline, header=FALSE, sep=';')

df2 <- df2[1:5]

colnames(df2) <- c('st', 'year', 'month', 'day', 'value')

df2$date <- as.Date(with(df2, paste(year, month, day,sep="-")), "%Y-%m-%d")

df <- df2[c(6,5)]

dm <- mean(df2$value[is.na(df2$value)==FALSE])

df$value[is.na(df$value)]<- dm

df2$value[is.na(df2$value)]<- round(dm,1)

#threshold <- findthresh(df$value, ne = 900)#length(df$value)\*0.05)

df\_x <- df$value#[df$value > threshold]

return (df2)

}

shorten <- function (row){

return(row[1:900])

}

monthmax <- function(df){

x\_max <- sqldf('select year,month,1 day, max(value) value from df group by year,month')

x\_max$date <- as.Date(with(x\_max, paste(year, month, day,sep="-")), "%Y-%m-%d")

x\_max <- x\_max[c(5,4)]

return(x\_max)

}

make\_df <- function(x1, x2, x4, x5, x6, x8, x9, x10, x11, x12, x13){

df <- data.frame(x1$date, x1$value, x2$value, x4$value, x5$value, x6$value,

x8$value, x9$value, x10$value, x11$value, x12$value, x13$value)

colnames(df) <- c('Date', 'SPb', 'Pskv', 'Smlnsk', 'Brnsk', 'Kstrm', 'NNvgrd', 'Mzhsk',

'Msk', 'Klmn', 'Rzn', 'Tmbv')

return(df)

}

pairswork <- function (row1, row2){

X=cbind(row1, row2)

U=rank(X[,1])/(nrow(X)+1)

V=rank(X[,2])/(nrow(X)+1)

Lemp=function(z) sum((U<z)&(V<z))/sum(U<z)

Remp=function(z) sum((U>1-z)&(V>1-z))/sum(U>1-z)

u=seq(.001,.5,by=.001)

L=Vectorize(Lemp)(u)

R=Vectorize(Remp)(rev(u))

plot(c(u,u+.5-u[1]),c(L,R),type="l",ylim=0:1,xlab="LOWER TAIL UPPER TAIL")

abline(v=.5,col="grey")

tau=cor(X,method="kendall")[1,2]

paramgauss=sin(tau\*pi/2)

copgauss=normalCopula(paramgauss)

Lgaussian=function(z) pCopula(c(z,z),copgauss)/z

Rgaussian=function(z) (1-2\*z+pCopula(c(z,z),copgauss))/(1-z)

u=seq(.001,.5,by=.001)

Lgs=Vectorize(Lgaussian)(u)

Rgs=Vectorize(Rgaussian)(1-rev(u))

lines(c(u,u+.5-u[1]),c(Lgs,Rgs),col="red")

paramgumbel=max(1/(1-tau), 1)

copgumbel=gumbelCopula(paramgumbel, dim = 2)

Lgumbel=function(z) pCopula(c(z,z),copgumbel)/z

Rgumbel=function(z) (1-2\*z+pCopula(c(z,z),copgumbel))/(1-z)

u=seq(.001,.5,by=.001)

Lgl=Vectorize(Lgumbel)(u)

Rgl=Vectorize(Rgumbel)(1-rev(u))

lines(c(u,u+.5-u[1]),c(Lgl,Rgl),col="blue")

}

}

pairswork2 <- function (row1, row2, row1\_max, row2\_max){

par(mfrow=c(2,1))

pairswork(row1, row2)

pairswork(row1\_max, row2\_max)

}

\*\*\*

Msk <- df$Tmbv

Msks <- Msk[Msk>3]

Msks <- sort(Msks)

logMsks <- rev(log(Msks))

n = length (Msks)

#plot(log(Msks), log((n:1)/(n+1)))

xi=1/(1:n)\*cumsum(logMsks)-logMsks

xise=1.96/sqrt(1:n)\*xi

plot(1:n,xi,type="l", xlab="",ylab="", xlim=c(1,1000), ylim=c(-0.05,0.6))

lines(1:n,xi+xise,col="red",lwd=1.5)

lines(1:n,xi-xise,col="red",lwd=1.5)

lines(1:n,xi,lwd=1.5)

abline(h=0,col="grey")

alpha=1/xi

alphase=1.96/sqrt(1:n)/xi

YL=c(0,5)

plot(1:n,alpha,type="l",ylim=YL,xlab="",ylab="", xlim=c(1,1000))

lines(1:n,alpha+alphase,col="red",lwd=1.5)

lines(1:n,alpha-alphase,col="red",lwd=1.5)

lines(1:n,alpha,lwd=1.5)

abline(h=0,col="grey")

Msk <- df$Msk

#Msks <- Msk[Msk>0]

Msks <- sort(Msks)

Xs=rev(sort(Msks))

n = length(Xs)

xi=1/log(2)\*log( (Xs[seq(1,length=trunc(n/4),by=1)]-

Xs[seq(2,length=trunc(n/4),by=2)])/

(Xs[seq(2,length=trunc(n/4),by=2)]-Xs[seq(4,

length=trunc(n/4),by=4)]) )

xise=1.96/sqrt(seq(1,length=trunc(n/4),by=1))\*sqrt( xi^2\*(2^(xi+1)+1)/((2\*(2^xi-1)\*log(2))^2))

plot(seq(1,length=trunc(n/4),by=1),xi,type="l",

ylim=c(0,3),xlab="",ylab="",xlim=c(1,1000))

lines(seq(1,length=trunc(n/4),by=1),

xi+xise,col="red",lwd=1.5)

lines(seq(1,length=trunc(n/4),by=1),

xi-xise,col="red",lwd=1.5)

lines(seq(1,length=trunc(n/4),by=1),xi,lwd=1.5)

abline(h=0,col="grey")

Пример визуализации результатов

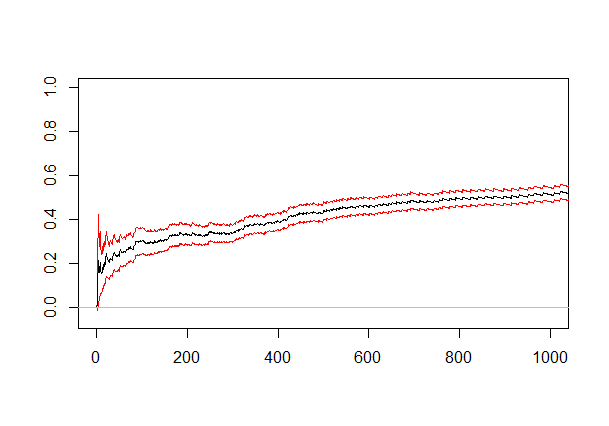
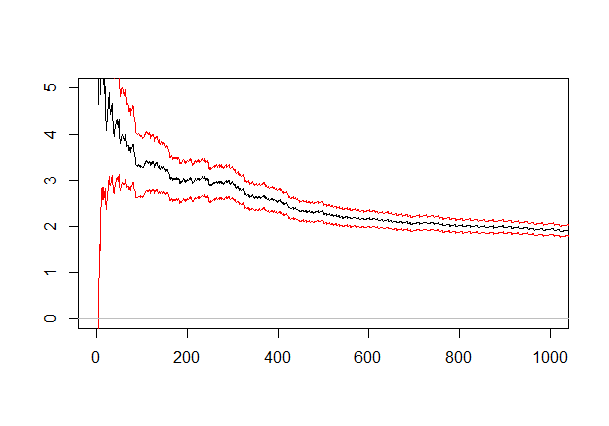


Рис. Оценки хвостового индекса для г. Рязань.

Заключение

Основными результатами научно-исследовательской работы в этом семестре стоит считать изучение тяжелохвостых распределений и оценок хвостовых индексов, а также оценок для параметра формы в экстремальных распределениях, который тесно связан с хвостовым индексом; полученные результаты войдут в текст магистерской работы, а также готовую к публикации в 1 квартале 2018 в журнале "*Вестник РУДН. Серия Математика. Информатика. Физика*" статью "*Extreme precipitation fields modeling on the territory of the European part of Russia*" на английском языке.

|  |  |  |
| --- | --- | --- |
| Магистрант | \_\_\_\_\_\_\_\_\_\_\_\_\_ | Рассахан Никита Дмитриевич |

Оценка за НИР: ОТЛИЧНО/ХОРОШО/УДОВЛЕТВОРИТЕЛЬНО

Дата: \_\_. \_\_.\_\_\_\_

|  |  |  |
| --- | --- | --- |
| Научный руководитель | \_\_\_\_\_\_\_\_\_\_\_\_\_ | доцент Красикова Е.М. |
| Заведующий кафедрой | \_\_\_\_\_\_\_\_\_\_\_\_\_ | к. ф.-м.н., профессор Уварова Л.А. |